

А.Н. Кричевец, А.А. Корнеев, Е.И. Рассказова

ОСНОВЫ СТАТИСТИКИ ДЛЯ ПСИХОЛОГОВ

УЧЕБНИК

А.Н. Кричевец, А.А. Корнеев, Е.И. Рассказова

ОСНОВЫ СТАТИСТИКИ ДЛЯ ПСИХОЛОГОВ

**Акрополь
Москва
2019**

УДК 159.9.07:004(075.80)

ББК 88.9с51я73

К828

Кричевец А.Н., Корнеев А.А., Рассказова Е.И.
К828 Основы статистики для психологов.
М.: Акрополь, 2019. – 286 с.

ISBN 978-5-98807-086-3

Цель данного учебника – ввести читателя в мир статистических инструментов, необходимых для профессиональной деятельности психолога. В нем на различных примерах рассматриваются основы теории вероятности и математической статистики. Главное внимание уделено базовым статистическим методам, обычно применяемым в рамках методологии проверки нулевой гипотезы: параметрическим и непараметрическим методам сравнения средних значений корреляционный и регрессионный анализ. Также в учебнике кратко представлены современные тенденции методологии статистической обработки данных в научных исследованиях.

Для каждого метода разбираются примеры и предлагаются задания для самостоятельной обработки смоделированных данных психологических исследований в пакете SPSS.

Учебник адресован студентам психологических специальностей, исследователям, преподавателям основ статистической обработки психологических данных и всем читателям, интересующимся вопросами применения количественных методов в психологии.

УДК 159.9.07:004(075.80)

ББК 88.9с51я73

ISBN 978-5-98807-086-3

© Кричевец А.Н., Корнеев А.А.,
Рассказова Е.И. 2019 г.

Оглавление

Предисловие	9
0.1. Структура учебника	9
0.2. Ориентирующий пример	11
Глава 1. Измерение и шкалы	13
1.1. Введение в теорию измерений	13
1.1.1. Измерение как расширение поля возможных операций	13
1.1.2. Типы шкал	16
1.1.3. Измерение с практической точки зрения	20
1.2. Чем интервальная шкала отличается от порядковой	21
1.3. Статистический анализ данных на компьютере. Пакет SPSS	24
Глава 2. Постановка проблемы: случайность, вероятность и принятие решений	29
2.1. Элементарные сведения	29
2.1.1. Что такое «случайность»? Что в психологии может считаться случайным?	29
2.1.2. Генеральная совокупность и распределение	30
2.1.3. Случайность и вероятность	33
2.1.4. Выборки и выборочные распределения	34
2.2. Точные определения и техника обращения с вероятностями	35
2.2.1. Элементарные события и классическая вероятность	35
2.2.2. Ситуации с отсутствием равновозможности	36
2.2.3. Формулы алгебры событий.	
Несовместимые и независимые события	37
2.2.4. Биномиальные вероятности	39
2.3. Вероятности и частоты в SPSS	41
Глава 3. Случайные величины	47
3.1. Элементарные сведения о случайных величинах	47
3.1.1. Случайная величина. Распределение	47
3.1.2. Математическое ожидание и выборочное среднее	49
3.1.3. Выборочная дисперсия	51
3.1.4. Случайная величина и генеральная совокупность	51
3.1.5. Непрерывные случайные величины.	
Распределения и их характеристики	52

3.1.6. Зачем нужны математическое ожидание и дисперсия . . .	54
3.1.7. Простейшие операции над случайными величинами и выборками. Стандартизация.	55
3.2. Случайные величины (уточняющее продолжение)	57
3.2.1. Операции над случайными величинами	57
3.2.2. Математическое ожидание и выборочное среднее	58
3.2.3. Дисперсия и среднеквадратическое отклонение	59
3.2.4. Стандартизация случайной величины	62
3.2.5. Сумма случайных величин	63
3.3. Случайные величины. Практикум	68
Глава 4. Нормальное распределение	75
4.1. Начальные сведения о нормальном распределении	75
4.1.1. Стандартное нормальное распределение	75
4.1.2. Нормальное распределение в общем виде	77
4.1.3. Задачи на нормальное приближение распределений	78
4.2. Нормальное распределение. Преобразования шкал и соответствующие преобразования плотностей	80
4.2.1. Изменение единиц измерения	80
4.2.2. Перенос начала координат	83
4.3. Нормальное распределение. Практикум	85
Глава 5. Статистическое оценивание средних.	89
5.1. Одновыборочный критерий Стьюдента и доверительный интервал для среднего	89
5.1.1. Стандартное отклонение как единица измерения индивидуального результата	89
5.1.2. Стандартная ошибка как единица измерения точности оценки по выборке	90
5.1.3. Определения основных понятий. Одновыборочный Т-критерий (Стьюдента)	91
5.1.4. Значимость 0.05 как социально закрепленная граница принятия решения в психологии	95
5.1.5. Статистическая значимость — не единственное, что нам нужно от результата. Другие виды значимости	97
5.2. Одновыборочная статистика Стьюдента и доверительные интервалы	98
5.2.1. Среднее выборочное как случайная величина	98
5.2.2. Доверительный интервал для среднего и значимость. Точные формулы	100
5.2.3. Доверительный интервал для среднего. Точный смысл	101
5.3. Практикум. Доверительный интервал	103

Глава 6. Оценивание средних. Продолжение	107
6.1. Т-критерий: сравнение средних	107
6.1.1. Т-критерий для парных выборок. Что такое парные, или связанные, выборки.	107
6.1.2. Т-критерий для независимых выборок. Что такое независимые выборки.	109
6.2. Т-критерий. Углубление в тему	112
6.2.1. Дополнительные замечания о Т-критерии для парных выборок	112
6.2.2. Условия применимости Т-критерия для независимых выборок	112
6.3. Практикум. Сравнение средних	114
Глава 7. Однофакторный дисперсионный анализ	125
7.1. Однофакторный дисперсионный анализ. Примеры и разъяснения.	125
7.1.1. Почему недостаточно парных сравнений	125
7.1.2. Числовой пример. Сравнение условий обучения	126
7.1.3. Зависимость значимости от объема выборки	130
7.1.4. Апостериорные сравнения	132
7.1.5. Односторонние и двухсторонние критерии	132
7.2. Эквивалентность Т-критерия и дисперсионного анализа для двух выборок	135
7.3. Однофакторный дисперсионный анализ. Практикум	136
Глава 8. Двухфакторный дисперсионный анализ	143
8.1. Двухфакторный дисперсионный анализ. Примеры и качественные разъяснения	143
8.1.1. Двухфакторная схема эксперимента. Пример сложения воздействий	143
8.1.2. Двухфакторная схема эксперимента. Пример сильного взаимодействия	144
8.1.3. Двухфакторная схема эксперимента. Комбинация влияния факторов и взаимодействия. Интерпретация	146
8.1.4. Пример. Доминантность	147
8.1.5. Пример. Закон Йеркса-Додсона	148
8.2. Расчеты статистических оценок.	149
8.2.1. Пример расчета межгрупповых сумм квадратов.	149
8.2.2. Расчеты степеней свободы.	152
8.3. Двухфакторный дисперсионный анализ в SPSS. Практикум	152
Глава 9. Гипотезы о связи переменных.	157
9.1. Коэффициент корреляции и линейная регрессия.	157

9.1.1. Коэффициент корреляции Пирсона	157
9.1.2. Проверка статистических гипотез о связи переменных	160
9.1.3. Простая линейная регрессия	162
9.1.4. Определение регрессионной прямой	163
9.1.5. Стандартизованный регрессионный коэффициент. Значимость	165
9.1.6. Корреляция, регрессия и причинность	166
9.2. Корреляция и регрессия. Продолжение	166
9.2.1. Корреляция и регрессия. Степени свободы. Сравнение с дисперсионным анализом. Мощность критерия	166
9.2.2. Ковариация и корреляция между случайными величинами	169
9.2.3. Гипотезы, связанные с ненулевой теоретической корреляцией. Преобразование Фишера	171
9.3. Линейная регрессия. Практикум	173
Глава 10. Регрессия в случае многих переменных	183
10.1. Регрессионная модель для нескольких независимых переменных (множественная регрессия)	183
10.1.1. Трехмерный пример	183
10.1.2. Значимость для модели множественной регрессии в целом	183
10.2. Иерархическая регрессия, анализ медиации и модерации	185
10.2.1. Основные идеи и примеры	185
10.2.2. Иерархическая регрессия (расчеты)	189
10.2.3. Анализ модерации (расчеты)	190
10.2.4. Медиация, модерация и опосредствование	192
10.3. Регрессия в случае многих переменных. Практикум	193
Глава 11. Непараметрические критерии	209
11.1. Основные сведения о наиболее часто используемых непараметрических критериях	209
11.1.1. Критерий знаковых рангов Вилкоксона	209
11.1.2. Критерий Манна-Уитни для независимых выборок	211
11.1.3. Критерий Краскелла-Уоллиса для n независимых выборок	213
11.1.4. Критерий Джонкхиера	214
11.1.5. Коэффициент корреляции Спирмена	215
11.1.6. Таблицы сопряженности	217

11.1.7. Таблицы сопряженности более высоких размерностей	220
11.2. Условия применимости параметрических и непараметрических критериев и последствия их нарушения	222
11.2.1. Критерий согласия	223
11.2.2. Проверка нормальности распределения перед применением дисперсионного анализа	224
11.2.3. Другие методы проверки применимости дисперсионного анализа.	225
11.3. Расчет непараметрических критериев в SPSS	227
11.3.1. Критерий Вилкоксона — непараметрический аналог Т-критерия для парных выборок	227
11.3.2. Критерий Манна-Уитни	228
11.3.3. Критерии Краскелла-Уоллиса и Джонкхиера	230
11.3.4. Коэффициент корреляции Спирмена	232
11.3.5. Таблицы сопряженности	232
Заключительные методологические замечания	237
Приложение 1. Основы работы в SPSS.	241
Приложение 2. Дисперсия суммы независимых случайных величин.	255
Приложение 3. Степени свободы	257
П3.1. Двумерное стандартное нормальное распределение	258
П3.2. Свойства двумерного стандартного распределения	259
П3.3. Свойства случайной величины, являющейся квадратом модуля двумерной стандартной нормальной случайной величины	260
П3.4. Оценка дисперсии.	261
П3.5. Разложение сумм квадратов в случае больших размерностей.	261
П3.6. Оценка дисперсии. Последнее замечание	261
П3.7. Другие разложения сумм квадратов	262
Приложение 4. Величина статистического эффекта	263
П4.1. Зачем и как оценивать статистический эффект	263
П4.2. Статистический эффект как оценка силы связи	265
П4.3. Статистический эффект как стандартизованная разница между средними	267
П4.4. Величина эффекта в случаях непараметрических статистик	271

Приложение 5. Гипотезы и их вероятности.	
Формула полной вероятности. Формула Байеса	273
П5.1. Условная вероятность. Формула полной вероятности	273
П5.2. Формула Байеса	275
П5.3. Геометрическая интерпретация формулы полной вероятности и формулы Байеса.	276
П5.4. Обобщение байесовского подхода	276
П5.5. Методологическое замечание	278
Приложение 6. Комбинаторика. Бином Ньютона.	279
П6.1. Размещения	279
П6.2. Сочетания	280
П6.3. Бином Ньютона	282
Список литературы.	285

Наш учебник призван ввести читателя в мир статистических инструментов, необходимых для профессиональной деятельности психолога. Чтобы «жить» в этом мире надо знать, зачем нужны эти инструменты, как они устроены, как их применять, и иметь навыки их использования. Наш учебник дает возможность читателю сделать первые шаги в решении всех этих задач.

Мы также знакомим читателя с современными тенденциями работы с данными в мировом научном сообществе, в частности, с критикой важных аспектов традиционного подхода к проверке статистических гипотез и с переориентацией на массовую обработку результатов в противовес единичным проверкам¹. Мы стараемся, таким образом, подготовить учащегося к будущим условиям работы в нашей области.

0.1. Структура учебника

Для того чтобы дать столь широкую ориентировку, мы структурировали учебник специальным образом. Он разделен на одиннадцать глав, каждая из которых, в свою очередь, поделена на три параграфа.

Первый параграф каждой главы предназначен для читателей, которые хотят овладеть базовыми знаниями по статистике. Здесь излагаются минимальные теоретические сведения по теме данной главы, описываются типичные ситуации и самым элементарным образом затрагивается общетеоретический контекст, в котором данный материал полезно понимать.

Во втором параграфе, который мы рекомендуем читать тем, кто собирается овладеть статистикой на хорошем, а не только элементарном уровне, или тем, кто имеет достаточно хорошую математическую подготовку (для них освоение материала не составит труда), мы излагаем все необходимые для теоретического осмысления материала сведения из теории вероятностей и математической статистики.

В третьем параграфе каждой главы разбираются примеры обработки данных в статистическом пакете SPSS, который в настоящее время

¹ Мы подробнее разбираем этот вопрос в «Заключительных методологических замечаниях», поскольку понимание проблемы требует, вообще говоря, освоения значительной части материала учебника.

используется большинством психологов во всем мире. В этом параграфе мы стараемся также дать материалы для развития способности суждения, т.е. помочь читателю научиться выделять в многообразных практических ситуациях те общие черты, которые позволяют прилагать те или иные методы к реальным ситуациям психологического исследования. Здесь читатель найдет минимально адаптированные в дидактических целях описания реальных исследований, сделанных классиками психологической науки. Файлы данных для практических занятий с пакетом SPSS можно найти на приложенном к учебнику диске, а также скачать на сайте факультета психологии МГУ им. М.В.Ломоносова по адресу: <http://mathpsy.com/handbook>

Таким образом, читатель в принципе может выбирать для себя достаточно разнообразные уровни и траектории освоения материала. В минимальном случае можно ограничиться первым и третьим параграфами каждой главы (мы бы оценили такое знание как удовлетворительное по старой шкале оценивания знаний). В максимальном случае, разумеется, следует освоить полностью все три параграфа каждой главы (отличная оценка), а хорошая оценка соответствует разнообразным промежуточным комбинациям из вторых параграфов. Для тех, кто собирается сделать статистику своим профессиональным занятием, в небольших дополнениях в конце книги даются материалы, если можно так выразиться, к экстраординарному уровню освоения.

Мы всегда начинаем изложение новых тем с нескольких типичных и часто встречающихся модельных примеров. При этом мы надеемся, что примеры послужат лишь первой опорой в освоении идей и принципов употребления методов данной темы. Сами эти идеи нуждаются в весьма общем и независимом от конкретных примеров понимании. Мы специально обращаем внимание читателя на это обстоятельство. **Без опоры на примеры знание будет слишком формальным и потому бесполезным. Без схватывания общих принципов (которые должны быть закреплены на теоретическом уровне) знание инструкторий по употреблению методов будет часто приводить к грубым ошибкам,** временами обесценивающим даже весьма добротный труд по сбору и первичной обработке данных.

Эти начальные примеры весьма просты и естественны. Иногда мы будем ссылаться на конкретные работы, в которых проведены подобные исследования, однако для удобства понимания статистического существа дела мы их обычно упрощаем, о чем и уведомляем читателя.

Пример, который мы теперь рассмотрим, позволяет сформулировать три вопроса, которые связаны с использованием статистики.

В дальнейшей работе мы постоянно будем иметь их в виду, как и то, что это три разных вопроса.

0.2. Ориентирующий пример.

Пример 0.2(1)². С помощью тестовой методики мы можем за короткое время оценить тревожность³ респондента. В нашем распоряжении имеется также тренинговая методика, претендующая на то, что ее прохождение снижает тревожность. Наша задача — разобраться, обоснованно ли утверждение о полезности представленной тренинговой методики, прежде чем мы начнем ее использовать.

Схема проверки такова: берем некоторое количество испытуемых, допустим 10 человек⁴, измеряем уровень тревожности каждого участника исследования, затем проводим с ними тренинг и еще раз измеряем тревожность⁵.

Каждый испытуемый получает две оценки — одну до и другую после прохождения тренинга. Разность между вторым и первым измерением дает представление о динамике тревожности: она может быть положительной (испытуемый стал тревожнее) или отрицательной (испытуемый стал спокойнее). Если тренинг эффективен, каким будет количественный результат? Нас не удивит, что испытуемые по-разному среагируют на тренинг, возможно даже, что некоторые из них станут тревожнее. Мы можем охарактеризовать эффективность тренинга в среднем, взяв среднее арифметическое разностей между оценками до и после тренинга, характеризующих динамику. Предположим, что в среднем тревожность снизилась на два тестовых балла. Что это значит? Вопросы, которые естественно задать в связи с этим результатом, таковы:

1. Что такое два балла по тесту — много это или мало, с чем их можно сравнить?

² Номера примеров, упражнений, рисунков и т.д. включают разделенные точками номер главы, номер параграфа (и подпараграфа, если есть) и затем заключенный в скобки порядковый номер объекта в данном параграфе (подпараграфе) независимо от категории объекта.

³ Здесь мы не будем останавливаться на важных деталях, связанных с психологическим содержанием понятия тревожности.

⁴ 10 человек — очень маленькая выборка, крайне редко встречающаяся в исследованиях. Однако мы первое время будем для удобства расчетов вероятностей использовать в примерах небольшие объемы выборок.

⁵ Здесь надо иметь в виду, что повторное использование теста может дать искаженные результаты. Предположим, однако, что наша тестовая методика свободна от этого недостатка.

2. Одинаково ли хорошо уменьшить тревожность на два балла у невротика, для которого тревожность служит серьезным препятствием для нормальной жизни, и у чрезмерно оптимистичного человека?

3. Насколько мы можем доверять полученному на данной выборке результату, можно ли по этому результату делать общий вывод об эффективности методики?

Ответ на первый вопрос связан с единицей измерения тревожности, которая используется в опроснике. Для того чтобы понимать значение этой единицы, разумно было бы, например, сопоставить ее с нижней и верхней границами нормы по данному показателю или другими характеристиками разброса значений показателя в интересующей нас популяции. Если характеристики популяции неизвестны, то в качестве опоры можно взять характеристики самой полученной нами выборки.

На второй вопрос статистика не может дать ответ, ждать его следует от психологии. Почти во всех случаях, с которыми мы будем иметь дело, разница в два балла будет иметь разный смысл для разных участков шкалы. Это значит, что психолог должен быть внимательным, характеризуя какие-то тенденции «в среднем». Об этом мы будем напоминать постоянно. Подробнее о первых двух вопросах в следующей главе.

Третий вопрос полностью относится к компетенции статистики и будет центральным в нашем учебнике. Нам надо научиться выделять закономерности из «шума», порождаемого случайностью, для чего нам послужит введение в теорию вероятностей — в том объеме, который позволит нам понимать и контролировать неизбежное влияние случайности на получаемые нами результаты.

Эти и подобные им очень важные вопросы требуют обращения к статистической теории и теории вероятностей, и далее мы их будем обсуждать в самых разных теоретических и практических контекстах.

Глава 1.

Измерение и шкалы

1.1. Введение в теорию измерений

1.1.1. Измерение как расширение поля возможных операций

Измеряя что-то, мы по определенному правилу ставим в соответствие измеряемым объектам или их свойствам числа. Однако не само по себе приписывание чисел привлекает нас в измерении, а то, что с этими числами можно производить разнообразные операции.

Пример 1.1.1(1). Землекоп экстра-класса за сутки выкапывает 30 метров канавы. Обычный землекоп выкапывает 20 метров канавы в сутки. Сколько обычных землекопов надо нанять олигарху, чтобы заменить заболевшего землекопа экстра-класса, который должен был выкопать тридцатиметровую канаву в установленный суточный срок?

Решение. Делим 30 метров на скорость 20 метров в сутки. Получаем 1.5. Ответ: Надо нанять полтора землекопа.

Хотя слова «полтора землекопа» режут здравомыслящему человеку слух, вряд ли кто-либо будет оспаривать корректность проведенных операций. Мы не будем отбрасывать этот ответ как заведомо неправильный, оставив приемлемую интерпретацию результата заказчику работы (он, например, может нанять землекопа на полставки). То, что в предметном мире нашему вычислению может не соответствовать никакая предметная конфигурация, не должно нас, как видим, останавливать.

Измерение, придающее количественные значения характеристикам интересующих нас предметов, позволяет значительно расширить и усилить средства получения и обоснования знаний. Это в полной мере относится и к психологии.

Однако это не значит, что как бы мы ни приписывали числовые значения нашим объектам и что бы ни делали потом с полученными числами, всякий раз получится хорошо. Рассмотрим простую «психологическую» модификацию предыдущего примера.

Пример 1.1.1(2). Старший научный сотрудник имеет 100 баллов по шкале интеллекта. Сколько надо нанять младших научных сотрудников с коэффициентом интеллекта 50 баллов, чтобы заменить заболевшего старшего научного сотрудника?

Рассуждая аналогично, можно посоветовать нанять двух младших научных сотрудников. Однако будет ли осмысленным такое рассуждение? Хорошо, например, если шкала интеллекта измеряет скорость решения задач определенного типа (как в примере с землекопами) и именно такого типа задачи и требуется решить с помощью научных сотрудников. Но тесты интеллекта, по которым рассчитывают упомянутые коэффициенты, ориентируются не столько на скорость, сколько на качественные показатели сложности решенных задач. В таком случае предложенное решение бессмысленно.

Пример 1.1.1(3). Для измерения температуры в России используют шкалу Цельсия, а в Соединенных Штатах — шкалу Фаренгейта. Температура кипения воды по Цельсию составляет 100 градусов, а по Фаренгейту — 212 градусов, точка замерзания соответственно 0 и 32 градуса. Сегодня температура в Мадриде 30 градусов по Цельсию (86 по Фаренгейту), в Лондоне 20 (68), а в Осло 10 (50). Эти данные представлены в таблице 1.1.1(4).

Таблица 1.1.1(4). Температура по Цельсию и Фаренгейту в городах Европы

Температура в городах			
	Осло	Лондон	Мадрид
По Цельсию	10	20	30
По Фарингейту	50	68	86

Можно сказать, что по Цельсию температура в Лондоне в 1.5 раза меньше, чем в Мадриде, и вдвое больше, чем в Осло. По Фаренгейту отношения температур другие: в Лондоне в 1.26 раза холоднее, чем в Мадриде, и в 1.36 раза теплее, чем в Осло. Зато утверждения «В Лондоне на столько же градусов холоднее, чем в Мадриде, на сколько градусов теплее, чем в Осло», оказываются одинаково истинными при измерении в обеих шкалах.

Обсуждаемые шкалы температур не приспособлены для оценивания отношений, но вполне адекватны при оценивании интервалов. Если мы сольем стакан мадридской тридцатиградусной воды со стаканом десятиградусной воды из Осло, то перемешанная вода будет иметь температуру Лондона независимо от того, какими градусниками мы измеряем температуру. Многие другие практические примеры покажут, что при измерении интервалов обе шкалы температур будут в равной степени полезны. То, что они будут одинаково полезны, следует из формулы перевода $T_f = T_c \cdot 1.8 + 32$. Равные температурные интервалы будут оценены как равные обеими шкалами.

Температурные шкалы дают пример хорошо обоснованных шкал интервалов (определение будет дано в следующем разделе). Они однако не приспособлены для ответов на вопросы «во сколько раз температура Лондона выше температуры Осло?», поскольку шкалы Цельсия и Фаренгейта дадут на него разные ответы. В то же время в задаче о землекопах мы использовали информацию о том, во сколько раз одни землекопы работают быстрее других. Эти различия возможностей надо четко зафиксировать.

Что касается единиц измерения, то иногда от их выбора существенно зависят практические результаты.

Пример 1.1.1(5). Два абитуриента претендуют на одно место на психологическом факультете. Их оценки по ЕГЭ по предметам «математика», «биология» и «русский язык» составляют у первого (80, 80, 80), а у второго (70, 80, 95). Кто будет принят в число студентов?

Решение. Будет принят второй абитуриент с суммарной оценкой 245 баллов против 240 у первого абитуриента.

Возможно, преподаватели математики сочли бы такой выбор ошибкой, полагая, что 10 баллов по математике для будущего студента-психолога «стоят дороже», чем 15 баллов по русскому языку. Они могли бы предложить, например, оценку по математике удваивать при суммировании (или, что то же самое, оценивать математику вдвое более мелкой единицей).

Еще одно, менее явное обстоятельство состоит в том, что диапазон возможных оценок важен не сам по себе. Важно, как варьирует оценка интересующих нас абитуриентов. Если подбор заданий ЕГЭ таков, что фактически оценка по математике у большинства наших абитуриентов заключена между 60 и 70 баллами, оценка по русскому языку — между 30 и 40 баллами, а оценка по биологии достаточно равномерно заполняет интервал между 30 и 95 баллами, то при суммировании баллы по математике и русскому языку будут оказывать ничтожное влияние на общий результат, а отбор будет вестись в основном по биологии⁶. Однако, так или иначе, суммирование представляется единственным простым и разумным способом отбора абитуриентов, которые должны обладать знаниями по нескольким отдельно оцениваемым предметам. В психологической практике, а не только в исследованиях, задачи одновременно учета показателей, измеренных в трудно соизмеримых шкалах, встречаются очень часто. Способы решения этой задачи мы обсудим в третьей главе.

⁶ На самом деле при формировании заданий для ЕГЭ соображения такого рода учитывались.

Между примерами 1.1.1(2) и 1.1.1(5) есть существенное различие, которое делает процедуру суммирования оценок во втором из них хотя и не безупречной, но все же более приемлемой, чем процедуру сложения усилий работников в первом. В случае с научными сотрудниками мы используем суждения типа «оценка старшего научного сотрудника вдвое выше», а в примере с абитуриентами суждения типа «оценка второго абитуриента на 10 баллов выше». Если мы прибавим к оценке по ЕГЭ каждого школьника 100 баллов и будем по-прежнему отбирать абитуриентов с лучшей суммой, то в вузы попадут те же самые соискатели студенческого билета. Если прибавить 100 баллов к оценке интеллекта научных сотрудников, то соотношение «старший научный сотрудник вдвое продуктивнее младшего» нарушится, и процедура замены одного старшего на двух младших потеряет числовое равенство. С другой стороны, мы видим, что оценка скорости работы землекопов не допускает прибавления константы, поскольку метры канавы выглядят гораздо более твердыми единицами измерения, чем баллы ЕГЭ и подобных оценок, и 2 метра канавы в самом прямом смысле вдвое больше, чем 1 метр. Это значит, что мы должны научиться различать системы оценивания и их употребления на практике, чтобы не делать серьезных ошибок.

1.1.2. Типы шкал

Дадим **Определение**: измерение есть приписывание чисел, обладающих определенными свойствами, отдельным характеристикам объектов или событий в соответствии с определенными правилами.

Множество характеристик объектов (будем называть его эмпирическим множеством) обладает определенными свойствами, множество чисел (будем называть его шкалой) также обладает свойствами. В каком отношении должны находиться те и другие? С. Стивенс предложил классификацию шкал, которая проясняет эти отношения.

Все шкалы состоят в принципе из одних и тех же чисел, но различаются по набору операций, которые с ними можно производить.

1.1.2(1) Номинативная шкала (ее также называют номинальной или шкалой наименований) — самая слабая (в том смысле, что набор возможных действий с числовыми значениями здесь минимальный) — это всего лишь именование объектов числами. Например, можно кодировать в анкете пол испытуемых числами 0 и 1 или использовать вместо названий субъектов Российской Федерации их коды. Понятно, что разность между кодами 77 и 50, которыми соответственно обозначены Москва и Московская область, не имеет никакого смысла. Никаких арифметических или иных операций, кроме проверки тождества или различия, такая шкала не предусматривает.

Не обязательно каждый объект имеет свой номер. Иногда в анкетах респондентов просят теперь указать место постоянного проживания именно в указанной выше кодировке. Это значит, что все москвичи имеют одно и то же значение этого признака — 77.

Характерным свойством номинативной шкалы является возможность произвольного переименования объектов. На множестве чисел эта возможность реализуется любым отображением, при котором никакие два разных числа не получают одно и то же новое значение. Например, новые коды регионов получаем по формуле $K_{new} = 50 - K_{old}$. Хотя некоторые новые коды окажутся отрицательными⁷, это не помешает установлению тождества и различия во всех возможных случаях. Говорят: «номинативная шкала определена с точностью до взаимнооднозначного отображения множества значений одной шкалы в множество значений другой шкалы», т.е. такого отображения, что никакие два возможных значения первой шкалы не отображены в одно и то же значение второй. Такое преобразование является *допустимым для данной шкалы*. Осуществляя допустимое преобразование шкалы, мы получаем другую шкалу с такими же свойствами, что и исходная. Как мы увидим дальше, такое преобразование может сделать шкалу более удобной для интерпретации и обработки.

1.1.2(2) Шкала порядка. Измерив какие-то качества и отобразив их числами, мы используем в этой шкале только одно свойство чисел — заданное на числах отношение «больше-меньше». Это имеет смысл, когда на эмпирическом множестве (т.е. на множестве характеристик объектов) мы различаем большую или меньшую выраженность характеристики. Например, мы можем попросить эксперта проранжировать тревожность студентов в группе, т.е. упорядочить студентов по степени тревожности и наименее тревожному присвоить ранг 1, следующему — ранг 2 и т.д. После этого можно достаточно уверенно использовать информацию о большей тревожности студента X по сравнению со студентом Y, выражаемую отношением «больше-меньше» на рангах сравниваемых студентов, но нельзя использовать выражаемую разностью рангов информацию о том, насколько первый тревожнее второго.

Если мы используем только отношение «больше-меньше» на нашем числовом множестве, то любое преобразование этого числового множества, сохраняющее данное отношение (такие преоб-

⁷ Хотя в принципе можно употреблять в качестве значений номинативной шкалы и отрицательные и вообще любые другие числа, далее мы ограничимся только натуральными, т.е. целыми неотрицательными числами.

разования называются монотонными)⁸, позволяет сохранить всю определяемую исходным отношением информацию. Например, если вместо ранга студента в группе использовать куб этого ранга, то более тревожные по-прежнему будут иметь более высокую оценку по новой шкале. Если из рангов испытуемых вычесть какое-то число, прибавить число, взять логарифм от ранга, то большее значение новой характеристики по-прежнему будет отражать большую выраженность тревожности.

Говорят: «шкала порядка определена с точностью до монотонно-го преобразования числового множества» или «допустимым преобразованием для порядковой шкалы является любое монотонное преобразование числового множества».

В шкале порядка не имеют точного смысла производные характеристики, получаемые с помощью арифметических операций, например среднее значение по группе. Действительно, если две методики приписывают трем студентам показатели тревожности (1, 7, 10) и (1, 3, 10) соответственно, то в смысле порядка результаты оценивания тревожности эквивалентны. Однако второй студент оказывается тревожнее, чем «в среднем группа из первого и третьего студента» в первой шкале, и менее тревожным, чем эта группа, во второй.

1.1.2(3) Шкала интервалов. Следующая по «силе» шкала — более для нас привычная — шкала интервалов. Здесь, кроме отношений «быть одинаковыми или различными» и «быть большими, равными или меньшими» мы используем еще одно свойство: то, что расстояние между двумя числами на числовой прямой измеряется разностью этих чисел. Отображая множество измеряемых объектов в числовое множество, мы задаем тем самым расстояние между объектами. Если А получил 15 баллов по тесту тревожности, а В 10 баллов, то можно сказать, что А на 5 баллов тревожнее, чем В. Также на 5 баллов тревожнее студент, получивший 45 баллов, чем студент, получивший 40 баллов.

В физических измерениях можно указать объект, отображающий эти 5 баллов разницы. Для измерений температуры это порция энергии, которая нагреет на 5 градусов данный предмет, независимо

⁸ На практике иногда приходится менять порядок на противоположный. Например, если в одном исследовании часть переменных измеряют количество ошибок выполнения заданий, а другие — продуктивность различных видов деятельности, то для применения многих методов их надо привести к одному порядку: например, для того чтобы во всех переменных большему значению соответствовал лучший результат.

от того, какова его исходная температура⁹. В психологии подобные обоснования невозможны. Однако во многих случаях использование интервальных шкал выглядит вполне разумным, надо только не забывать о том, что «цены» интервалов на разных участках шкалы могут быть разными, и не допустить ошибки, когда различие «цен» окажется существенным при принятии решений.

Соотношения между интервалами сохраняются, если числовые оценки умножить на какое-либо число и прибавить или вычесть какое-либо число. Т.е. применив к шкале произвольное линейное преобразование $K_{new} = a \cdot K_{old} + b$, мы сохраняем всю информацию, которая могла быть получена в старой шкале¹⁰. Говорят: «шкала интервалов определена с точностью до линейного преобразования» или «допустимым преобразованием шкалы интервалов является произвольное линейное преобразование». В частности, положение нуля на шкале может меняться, если мы проводим допустимое преобразование с ненулевой константой b .

В той степени, в какой обосновано, что шкала имеет интервальный тип, в ней имеют точный смысл такие характеристики, как «среднее значение по группе». В интервальной шкале получает естественный смысл вопрос «на сколько более выражено данное качество у объекта А, чем у объекта В», хотя надо иметь в виду, что ответ зависит от выбранной единицы измерения.

1.1.2(4) Шкала отношений. Следующий тип шкалы в нашей иерархии — шкала отношений. Шкала интервалов допускала сдвиги, т.е. изменения начала отсчета. Именно поэтому для этой шкалы не имели смысла вопросы «во сколько раз больше?» В шкале отношений нуль выражает отсутствие измеряемого свойства, поэтому не может сдвигаться, и вопрос приобретает смысл. «Во сколько раз больше?» можно, например, спрашивать об объективных весах предметов и о субъективных их оценках, измеренных методом Стивенса. Допустимым преобразованием шкалы отношений является только изменение масштаба, приводящее к умножению значений на определенное число.

С точки зрения применения статистических методов различия между шкалами интервалов и отношений не являются существенными, а различия с остальными типами весьма важны. Для

⁹ Надо, конечно, оговориться, что предмет не меняет при этом нагревании фазового состояния.

¹⁰ Чтобы избежать трудностей с отношением «больше-меньше», коэффициент надо взять положительным, отрицательный коэффициент может понадобиться в случаях, описанных в предыдущем примечании.

шкал интервалов и более сильных — шкал отношений — имеют смысл арифметические операции над шкальными значениями, на которых основаны так называемые параметрические методы статистики. Для шкал порядка (и, разумеется, для шкал интервалов и отношений, которые «сильнее» шкалы порядка, т.е. среди прочих свойств обеспечивают также информативность отношений порядка на числовых множествах) имеют смысл операции сравнения и ранжирования, на которых основаны непараметрические (или порядковые) методы статистики. Для номинативных шкал набор методов еще более сужается.

1.1.3. Измерение с практической точки зрения

Для корректного употребления статистических методов важно установить, является ли данная шкала номинативной, порядковой или интервальной, а разграничение интервальных шкал и шкал отношений задача практически не важная для применимости тех или иных статистических методов.

Разграничение номинативных и порядковых шкал обычно не вызывает проблем — достаточно убедиться, что вместо чисел для различения полученных в результате измерений значений можно использовать слова, лишенные каких-либо указаний на порядковые отношения. В таком случае данный вид измерений уверенно должен быть отнесен к номинативным шкалам.

Упражнение 1.1.3(1). Можно ли использовать в качестве шкалы порядка и что будет измерять такая шкала

- а) номера студентов в алфавитном списке группы?
- б) номера курсов, на которых студенты учатся?
- в) номера их зачетных книжек?
- г) год рождения студентов?

Весьма важная задача разграничения порядковых и интервальных шкал оказывается принципиально значительно более трудной. Если мы убедились, что отношения порядка заданы на степенях выраженности измеряемого признака и отражены в шкале, то законным, как мы сказали, становится тогда применение лишь небольшой части статистических методов — порядковых или непараметрических методов. Большая и важная часть методов требует обоснования того, что измерение проведено в интервальной шкале. Однако доказательств такой принадлежности предъявить обычно не удастся.

Некоторые современные зарубежные учебные пособия (Howitt, Cramer, 2008, p. 5) советуют своим читателям игнорировать сомнительное с практической точки зрения различие и спокойно применять разнообразные методы обработки данных. Мы относимся

к такой рекомендации с осторожностью, хотя и соглашаемся с трудностями различения шкал¹¹.

Мы разберем несколько типичных ситуаций, в которых шкалы, которые кажутся на первый взгляд интервальными, не могут считаться таковыми с полным правом, и ситуации, в которых порядковые шкалы могут претендовать на более высокий статус.

Пример 1.1.3(2). Быстроту чтения у младших школьников можно характеризовать двумя способами: количеством прочитанных в минуту знаков и временем (измеренным в минутах), затраченным на чтение 1000 знаков. Оба показателя вполне осмысленные.

Однако можно заметить, что количество прочитанных знаков (в минуту) n и затраченное время на 1000 знаков (выраженное в минутах) t связаны соотношением. Действительно, время, затраченное на чтение одного знака, равно, с одной стороны, $1/n$, а с другой стороны, $t/1000$. Таким образом, $t=1000/n$. Предположим, что три ученика А, В и С прочитали в минуту 200, 400 и 900 знаков соответственно. При таком представлении результатов А и В оказываются по своим показателям близки, а С показывает сильно отличающийся от них хороший результат.

Если же характеризовать быстроту чтения вторым способом, то результаты испытуемых окажутся таковы: 5, 2.5 и 1.1 минуты на 1000 знаков. При таком подходе кажется, что С и В показывают близкие результаты, а результат А — экстремально плохой.

В педагогических исследованиях употребляются оба вида показателей и для обоих иногда применяются параметрические методы, соответствующие интервальным шкалам. Ясно, однако, что обе шкалы не могут быть одновременно интервальными, поскольку в одной из них интервал АВ меньше интервала ВС, а в другой АВ > ВС, чего для пары интервальных шкал быть не должно¹². В этом случае обе шкалы, строго говоря, надо рассматривать как порядковые и применять арифметические операции к шкальным значениям с некоторой осторожностью. Например, по первой шкале В показывает результат хуже среднего, а по второй — лучше среднего, если средним считать среднее арифметическое показателей.

¹¹ По крайней мере, если методы, ориентированные на порядковые и интервальные шкалы, дают существенно разные результаты, то необходимо разобраться в причинах этого расхождения.

¹² Интервальными с большим правом можно было бы считать показатели $\ln(t)$ и $\ln(n)$. Они связаны допустимым преобразованием $\ln t = \ln(1000) - \ln(t)$, поэтому могут считаться интервальными оба одновременно.

1.2. Чем интервальная шкала отличается от порядковой

Поскольку параметрические методы статистики дают значительно больше возможностей, чем непараметрические, приведение шкал к интервальной форме имеет практический смысл. Так, при разработке тестов, предназначенных для широкого использования, делается предположение, имеющее основания как в теории вероятностей, так и в практике разнообразных измерений: считается, что распределение результатов измерения данного параметра на всех индивидах, подлежащих данному измерению, должно быть похоже на нормальное распределение (нормальное распределение будет введено в главе 4). С этим связан метод преобразования первоначально порядковой шкалы в интервальную путем некоторой монотонной деформации шкалы (допустимого преобразования), которое приводит к нормальному распределению измеряемого показателя по всей совокупности индивидов, потенциально подлежащих данной измерительной процедуре¹³.

Пример 1.2(1). Грамотность школьников оценивается количеством ошибок в стандартном диктанте, имеющем объем 1000 знаков (примерно полстраницы). Можно ли считать шкалу «количество ошибок» интервальной?

Вопрос поставлен не вполне правильно. Мы можем считать такую шкалу интервальной, но при этом некоторые наши выводы могут оказаться сомнительными. Основания для сомнения таковы: школьники А и В, сделавшие 0 и 5 ошибок в диктанте соответственно, отличаются, если рассматривать их в контексте педагогических задач, сильнее, чем слабые ученики С и D, сделавшие 50 и 60 ошибок соответственно. Если мы хотим охарактеризовать интервальной шкалой степень грамотности, нам не обойтись без экспертных оценок.

Упражнение 1.2(2). Попробуйте себя в роли эксперта. Разбейте школьников на 5 групп по числу ошибок в диктанте: «очень сильные», «сильные», «средние», «слабые», «очень слабые». Например, так: «очень сильные (5 баллов), от 0 до 20 ошибок», «сильные (4 балла), 21 — 26 ошибок», «средние (3 балла), 27 — 29 ошибок», «слабые (2 балла), 30 — 44 ошибок», «очень слабые (1 балл), 45 и больше

¹³ Некоторые авторы (Наследов, 2004, с. 59) считают нормальное распределение показателя главным или даже единственным аргументом в пользу интервальности данной шкалы. Однако этот аргумент имеет лишь ограниченную применимость. Например, шкалы субъективных оценок вообще не связаны с распределениями. К таковым относятся психофизические шкалы, шкалы, используемые в психосемантике, и т. д.

ошибок»¹⁴. Сделайте это так, чтобы разделяющие группы интервалы были, на ваш взгляд, равными с точки зрения педагогических задач. Сравните свои предложения с предложениями коллег-экспертов. Если вы можете достичь согласия, это серьезный аргумент в пользу принятия такой шкалы в качестве интервальной.

Упражнение 1.2(3). Возьмите с помощью калькулятора натуральные логарифмы от верхних границ первых четырех классов в вашем разбиении. Насколько последовательность логарифмов отличается от арифметической прогрессии?

Если в последнем задании вы получили нечто похожее на арифметическую прогрессию, то, возможно, в качестве интервальной разумно взять шкалу $y = \ln(x+1)$ (где x — количество ошибок), поскольку она хорошо согласуется с вашим интуитивным преобразованием шкалы ошибок. Подумайте, не использовали ли вы соображение такого рода: равный сдвиг оценки должен соответствовать пропорциональному увеличению числа ошибок (например, если ученики А, Б и В сделали 5, 10 и 20 ошибок соответственно, то интервалы между их оценками — равные). Этим соображениям соответствуют логарифмические шкалы. В подобных случаях логарифмы от сырых баллов используются достаточно часто.

Пример 1.2(4). Порядковые шкалы иногда называют ранговыми¹⁵. Предположим, группа из 8 экспертов оценивала 10 педагогов некоторой школы по степени их влияния на учеников. Эксперт должен был расположить педагогов в ряд по степени их влиятельности, начав с наиболее влиятельного (ему мы припишем ранг 1) и заканчивая наименее влиятельным (ранг 10). Результаты каждого эксперта, несомненно, принадлежат к порядковой шкале. Однако, используя эти результаты ранжирования, мы можем построить шкалу более похожую на интервальную. Если мы для каждого педагога рассчитаем его суммарный ранг по всем экспертам, то смысл этих сумм будет достаточно ясным, хотя с порядковыми шкалами производить такие операции не вполне корректно. Вычисленные суммы уже будут указывать на нечто, подобное расстоянию между педагогами. Действительно, если предположить, что влиятельность есть объективная характеристика педагога, а эксперты с некоторой точностью оценивают этот показатель, и если педагог А имеет существенно большую влия-

¹⁴ Нашу классификацию мы сделали заведомо неразумной, чтобы не навязывать читателю свое мнение.

¹⁵ Это не совсем правильно, поскольку приписывание упорядоченных чисел может производиться и не в связи с ранжированием. Например, порядковые, но не связанные с ранжированием результаты дают обычные опросники.

тельность, чем следующий за ним по влиятельности педагог В, то все эксперты, несмотря на вариации в оценках, все же предпочтут в списке поставить А выше В. Это значит, что разность суммарных рангов А и В будет равна 8 (количеству экспертов). Если же А более влиятелен, чем В, но разница между ними на грани различимости, то эксперты будут ставить А прежде В и В прежде А почти одинаково часто, поэтому суммарные ранги у А и В будут близки.

На таких соображениях основаны многие методы, которые, предполагая наличие неизвестной интервальной шкалы для данного показателя, ищут шкальные значения каких-то объектов с помощью оценок экспертов или испытуемых.

1.3. Статистический анализ данных на компьютере. Пакет SPSS

В настоящее время статистическая и математическая обработка данных и расчеты производятся на компьютерах с помощью специализированных программных пакетов. Наиболее распространенными являются Matlab, Mathcad и Octave, в основном предназначенные для математических расчетов, а также SPSS, Statistica, SAS и R¹⁶, предназначенные для статистической обработки данных. В нашем учебнике в качестве основного инструмента для обработки данных мы будем использовать статистический пакет SPSS. Этот пакет создан для обработки результатов исследований в социальных науках и является одним из наиболее распространенных программных продуктов, используемым при обработке данных, полученных при социологических, экономических и психологических исследованиях. SPSS обладает достаточно богатым набором возможностей: в нем удобно вводить и преобразовывать данные исследования, проводить разнообразные процедуры статистической обработки данных, от элементарных расчетов среднего значения ряда данных до использования сложных многомерных методов. Также с помощью SPSS можно визуализировать полученные данные, строить различные типы диаграмм.

В задачу данного пособия не входит исчерпывающее описание программного обеспечения, предназначенного для статистиче-

¹⁶ R является не просто программным пакетом, а языком, предназначенным для статистического анализа. Это превращает его в очень мощное средство, позволяющее реализовать практически любую статистическую процедуру. Его изучение осложняется необходимостью понимания скриптового языка, но заинтересованного читателя мы можем адресовать к учебникам А.Б. Шипунов, Е.М. Балдин, П.А. Волкова и др., 2012 и Джеймс Г., Уиттон Д., Хасти Т., Тибширани Р., 2016.

ских расчетов вообще и пакета SPSS в частности. Однако в дальнейшем мы будем сопровождать теоретический материал практическими примерами обработки данных, благодаря чему у читателя сможет сложиться представление о практическом применении знаний по математической статистике в исследовательской деятельности. Также в последующих главах будут предложены разнообразные практические задачи, для самостоятельного выполнения. Все статистические расчеты, которые будут включаться в такие практические примеры и задания, могут быть выполнены в статистическом пакете SPSS. Вначале мы остановимся на описании только самых базовых возможностей этой программы. В дальнейшем, по мере обсуждения различных аспектов и процедур статистической обработки данных, мы будем подробнее останавливаться на соответствующих функциях SPSS. При этом следует подчеркнуть, что приведенное в данном учебнике описание не является полным и в целях углубления знания статистического пакета SPSS читатель может обратиться к более специализированным источникам. Перечислим некоторые из них:

Наследов Д.А. IBM SPSS Statistics 20 и AMOS: профессиональный статистический анализ данных. — Издательский дом «Питер», 2012.

Бюль А., Цёфель П. SPSS. Искусство обработки информации, Москва, Diasoft, 2005

Из зарубежных источников мы можем порекомендовать следующие книги:

Field A. Discovering statistics using IBM SPSS statistics. — Sage, 2013. (<https://books.google.ru/books?id=c0Wk9IuBmAoC>)

Brace N., Snelgar R., Kemp R. SPSS for Psychologists. — Palgrave Macmillan, 2012. (<https://books.google.ru/books?id=-N2OCgAAQBAJ>)

Основные элементы интерфейса SPSS и правила ввода данных описаны в приложении 1. Читателю, не знакомому с этой программой, мы рекомендуем изучить этот раздел прежде, чем двигаться дальше. Мы же перейдем к примерам и упражнениям, позволяющим получить опыт практической работы с данными, измеренными в различных шкалах.

Пример 1.3(1). Рассмотрим модельные данные, соответствующие Примеру 1.1.3(2). Напомним, что в этом примере обсуждается два способа оценки скорости чтения: количество прочитанных в минуту знаков и время, затраченное на чтение 1000 знаков. Допустим, что в исследовании скорости чтения приняло участие 19 учеников 2-го класса начальной школы. Модельные результаты такого экспери-

мента содержатся в файле *ReadingSpeed.sav*.¹⁷ Запустите имеющуюся у вас программу SPSS. С помощью главного меню откройте файл *ReadingSpeed.sav*. На экране возникает таблица данных, содержащая четыре столбца, названия которых подписаны вверху таблицы. Столбцы называются переменными, а надписи — их именами. При вводе данных исследования каждый столбец содержит информацию об одном из регистрируемых параметров, а каждая строка — информацию об одном испытуемом. В нашем примере первая переменная — *chars* — это количество знаков, прочитанных испытуемым за минуту, а вторая — *t* — это время, затрачиваемое испытуемым на чтение 1000 знаков. Показатели связаны соотношением $t = 1000/n$ (см. объяснения в Примере 1.1.3(2)). Третья и четвертая переменные представляют собой натуральные логарифмы от первой и второй переменных (*charLN* и *tLN* для *char* и *t* соответственно). Шкала «число знаков, прочитанных за минуту» с формальной точки зрения интервальная, однако если обратиться к содержательной стороне вопроса, то можно спросить: одинаковы ли в интересующем нас смысле различие между детьми, прочитавшими 200 и 300 знаков за минуту, и различие между детьми, прочитавшими 900 и 1000 знаков? По всей видимости, нет. Такое же рассуждение можно применить и к шкале «время, затрачиваемое на прочтение 1000 знаков».

Еще одну проблему мы уже обсуждали в Примере 1.1.3(2). Здесь мы имеем следующие результаты: самый быстрый ученик (номер 13) прочитал 890 знаков в минуту, а на 1000 знаков потратил 1.12 минуты. Самый медленный ученик (номер 1) прочитал 205 знаков за минуту и потратил на 1000 знаков 4.88 минуты. Занимающий среднюю позицию ученик номер 16 показал результат 454 и 2.20 соответственно.

Упражнение 1.3.(2). К какому из двух полюсов скорости чтения в классе окажется ближе средний ученик (номер 16), если его скорость чтения измерять первым способом и вторым способом?

Не получилось ли у нас, что две «интервальные» шкалы оценивают при таком подходе расстояния между одними и теми же парами испытуемых по-разному?

Упражнение 1.3.(3). Если рассмотреть эти, формально претендующие на интервальный тип, шкалы как порядковые, то две шкалы дадут один и тот же результат (так как связаны между собой монотонным преобразованием).

¹⁷ Таким курсивом мы будем помечать имена файлов. Жирным курсивом с разделяющим тире помечены последовательности выборов, начинающиеся в главном меню SPSS.

Чтобы убедиться в этом, проведем следующую операцию. В SPSS есть возможность автоматически ранжировать данные. Для этого надо зайти в пункт **Преобразовать (Transform)** главного меню и выбрать из списка пункт **Ранжировать наблюдения (Rank Cases)**. Далее последовательность действий с меню мы будем коротко записывать так: Преобразовать — **Ранжировать наблюдения (Transform — Rank Cases)** — в скобках указывая название пунктов в англоязычных версиях SPSS. В появившемся окне надо указать, какую переменную (или переменные) мы хотим проранжировать, перетащив с помощью мыши из левого поля, содержащего полный список переменных в таблице, интересующую нас переменную в поле **Переменные (Variables)**. Проранжируем испытуемых по числу знаков, прочитанных за минуту — перенесем переменную *chars* в рабочее окно. Также следует обратить внимание на направление ранжирования, оно возможно от наибольшего значения исходной переменной к наименьшему и наоборот. Если мы хотим, чтобы первый ранг получил испытуемый, читающий быстрее всех, то ранг 1 в данном случае надо присвоить наибольшему значению числа прочитанных знаков. Поэтому внизу, под общим списком переменных, выберем вариант *Ранг 1 присвоить наблюдению с максимальным значением (Assign Rank 1 to Largest value)*. После этого проведем ранжирование, нажав кнопку *OK*. В таблице данных должна появиться переменная *Rchars*, содержащая ранги испытуемых по числу прочитанных знаков.

Упражнение 1.3(4). Проранжируйте испытуемых по времени, затрачиваемому на чтение 1000 знаков. Будьте внимательны, определите направление ранжирования так, чтобы ранг 1 получил наиболее быстро читающий испытуемый. Сопоставьте ранги, получаемые при ранжировании двух шкал, убедитесь, что ранги испытуемых совпадают.

Задание 1.3(5). Еще один вариант, позволяющий сделать оценки по двум шкалам более согласованными — это логарифмирование полученных данных. В переменных *charsLN* и *tLN* содержатся натуральные логарифмы значений переменных *chars* и *t* соответственно. Как мы писали в примечании к Примеру 1.1.3(2), обе логарифмические шкалы с полным правом могут считаться интервальными. Вопрос: Какое допустимое преобразование переводит первую логарифмическую шкалу во вторую? (Указание: убедитесь, что сумма этих переменных постоянна для всех испытуемых с точностью до округления.)

Задание 1.3(6). Испытуемых можно упорядочить по значениям какой-то переменной. Для того, чтобы это сделать, выберите последо-

вательность пунктов *Данные — Сортировать наблюдения (Data — Sort cases)* и перетащите с помощью мышки переменную `chars` в окно *Сортировать по: (Sort by)*. Нажмите *OK*. Убедитесь, что упорядоченности по всем переменным и по рангам согласованы.

Глава 2.

Постановка проблемы: случайность, вероятность и принятие решений

2.1. Элементарные сведения

2.1.1. Что такое «случайность»? Что в психологии может считаться случайным?

Мы легко пользуемся словом «случайность» в обыденной речи. Случайным обычно называется результат влияния неконтролируемых побочных обстоятельств на какой-то интересующий нас процесс.

Иногда случайность специально культивируется. В азартных играх это — перемешивание карт, встряхивание костей в стаканчике и т.п. Бросание жребия — еще один пример организации случайного результата. В таких случаях оказывается, что в массовых процессах, связанных с организованной случайностью, выявляются некоторые закономерные тенденции, которые в принципе можно учитывать при планировании деятельности. Например, используя случайность, можно с достаточной надежностью делать выводы о результатах выборов, опрашивая не всех проголосовавших, а только некоторую часть из них, что и делают социологические службы, чтобы оценить результаты до официального их объявления. Оценки такого рода опираются на относительно узкое понятие случайности, связанное с повторениями одинаково организованных процедур, приводящих к результатам из определенной совокупности возможностей. Именно такое использование случайности в психологических исследованиях мы и будем рассматривать в нашей книге.

Если исследователь хочет узнать, как влияют на время реакции человека различные экспериментальные условия, он берет в качестве испытуемых не всех живущих на Земле людей, а группу «случайных» представителей человеческого рода. Как можно реализовать такой случайный выбор? Можно, например, сделать огромный барабан (похожий на тот, который используется в телеигре «Поле чудес», только несравненно больший по размерам), по краю которого выписать фамилии всех жителей Земли. Раскрутив его достаточно сильно, будем ждать, на кого укажет стрелка. Если список граждан хранится в какой-нибудь компьютерной базе данных, то случайный выбор еще проще организовать, используя специальные компьютерные средства. Однако физическая реализуемость — не главное для нас. Нам важно сейчас представить себе

идеальную равновозможность появления имени каждого гражданина по нашему запросу и извлечь статистические следствия из этого представления.

На самом деле исследователю приходится довольствоваться совсем не безупречными процедурами случайного выбора. Например, он может повесить объявление в коридоре университета и привлечь студентов-добровольцев. Если по времени реакции студенты еще могут считаться случайными представителями человечества в данных условиях, то по уровню интеллекта они, разумеется, таковыми считаться уже не могут, так как прошли солидные образовательные процедуры и отбор. Мы не будем в нашем учебнике специально обсуждать эти вопросы, но при проведении исследований и обследований законность и обоснованность выбора испытуемых надо иметь в виду.

Если нам удастся приемлемым образом решить вопрос выбора для исследования в удовлетворительном смысле случайных людей, то возникает второй вопрос: какие выводы о психологии человека мы можем тогда делать из наших экспериментальных результатов? Чтобы узнать ответ на этот вопрос, надо изучить статистику.

2.1.2. Генеральная совокупность и распределение

Начнем с уточнений и идеализаций. Пусть у нас есть множество людей, некоторая характеристика которых нас интересует. Например, мы исследуем тревожность россиян, используя определенную тестовую методику.

Определение 2.1.2(1). Будем называть *генеральной совокупностью* множество значений определенной характеристики, принадлежащей представителям некоторой совокупности индивидов¹⁸.

Если нас интересует тревожность россиян, то генеральной совокупностью будут все граждане России, точнее, их тревожность.

Определение 2.1.2(2). *Выборкой* будем называть некоторую, обычно относительно небольшую и обозримую часть этого множества.

Определение 2.1.2(3) Будем называть выборку *случайной*, если члены выборки представляют собой случайно отобранных представителей интересующей нас генеральной совокупности.

Если бы каким-то образом нам стали известны результаты тестирования каждого представителя генеральной совокупности, каждого нашего согражданина, то с такой информацией нам было бы очень трудно работать — она необозрима. Можно было бы попросить того, от кого мы ее получили, дать более компактное ее представление. Напри-

¹⁸ Иногда для простоты изложения мы будем называть генеральной совокупностью само множество индивидов, а не множество значений интересующей нас характеристики.

мер, если результаты тестирования нашей методикой измерения тревожности могут изменяться от 0 (не тревожный) до 4 (очень тревожный), то одним из таких компактных представлений будет *таблица частот*, где для каждой из возможных оценок рассчитан процент людей, которые эту оценку получили. Возможный результат представлен в таблице 2.1.2(4).

Таблица 2.1.2(4). Возможный результат тестирования

Распределение результатов тестирования					
Результаты тестирования	0	1	2	3	4
Процент получивших оценку	3	24	40	26	7

Эти результаты можно представить в графическом виде. Такой способ представления называется *гистограммой (Histogram)* (рис. 2.1.2(5)а).

Упражнение 2.1.2(6). Пусть мы опросили 100 случайно выбранных граждан России. Что можно было бы назвать гистограммой распределения результатов тестирования тревожности по этой выборке?

Упражнение 2.1.2(7). Насколько гистограмма по выборке была бы похожа на гистограмму распределения результатов у всех граждан РФ?

Ответ на первый вопрос понятен. Надо вычислить проценты получивших те или иные баллы по тесту, нарисовать прямоугольники соответствующей величины и добавить соответствующие надписи.

Второй вопрос гораздо интереснее. Возможность статистической оценки параметров опирается на тот факт, что чаще всего выборочная гистограмма будет похожа на гистограмму по генеральной совокупности. На рис. 2.1.2(5)б, 2.1.2(5)в и 2.1.2(5)г приводятся три результата компьютерного моделирования гистограмм выборок по 100 человек из 100 000 000, имеющих процентное распределение оценок тревожности, отображенное на гистограмме (см. рис. 2.1.2(5)а).

Фундаментальный постулат, который хорошо согласуется с опытом, определяет наше отношение к данному сходству: наблюдаемые частоты при увеличении количества испытаний стремятся к частотам в генеральной совокупности. В случае испытаний тревожности граждан РФ постулат можно конкретизировать так: мы видим, что 3% генеральной совокупности характеризуются нулевой оценкой тревожности; если брать все увеличивающиеся выборки из данной генеральной совокупности, то процент минимально тревожных людей в них будет все ближе подходить к этим 3%.

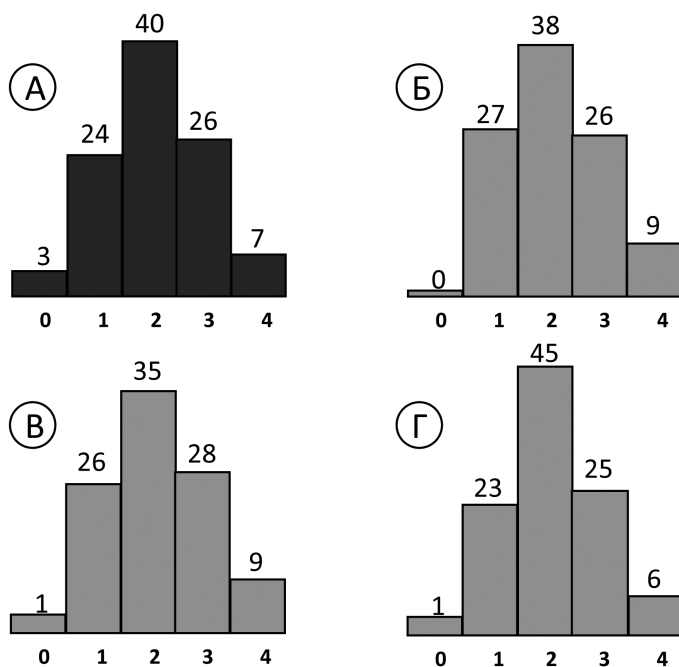


Рис. 2.1.2(5) Гистограмма распределения результатов тестирования тревожности (а) у всех граждан РФ, (б), (в) и (г) - в случайных выборках объема 100 индивидов (в процентах).

Обычно мы не знаем распределение интересующего нас показателя у всех граждан России, и узнать его точный вид можем, только если проделаем огромную и дорогостоящую работу. **Однако если выборочные гистограммы в каком-то смысле похожи на гистограммы распределения по всей генеральной совокупности, то ведь верно и обратное: если мы знаем выборочное распределение, то распределение генеральной совокупности должно быть на него похоже.** Чтобы делать выводы о генеральной совокупности на основании выборки, мы должны научиться как-то оценивать «степень возможной похожести».

Пример 2.1.2(8). Предположим, нам надо выяснить, кто показывает более высокий балл по нашему тесту тревожности — мужчины или женщины. Выбрав случайно 50 мужчин и 50 женщин, мы просим их заполнить опросник и убеждаемся, что средний балл мужчин 2.4, в то время как у женщин средний балл 2.2. Мы можем надеяться, что вы-

борочные значения похожи на средние по соответствующим частям генеральной совокупности. Однако не может ли быть, что нам случайно попались женщины спокойные, а мужчины столь же случайно попались очень тревожные, что и обусловило полученный результат?

Мы никогда не застрахованы полностью от принятия ошибочных решений, если опираемся на эмпирические данные, хотя бы часть вариации которых обеспечивается случайностью. Но в таких случаях мы можем оценивать надежность нашего вывода благодаря знанию того, как эта случайность может «себя вести» при многократных испытаниях.

Сделать это можно, изучая ситуации «хорошо организованных случайностей». Это случайности, о которых мы точно знаем нечто очень важное — вероятности событий, а другие важные вещи можем уверенно вывести средствами математики. Таким образом мы узнаем, насколько выборочные характеристики должны быть похожи на характеристики генеральных совокупностей, и сможем судить о вторых на основании первых. Сверх того, мы получим возможность решать разнообразные задачи, не сводимые к генеральным совокупностям, об этом речь пойдет ниже.

2.1.3. Случайность и вероятность

Говоря о случайном выборе из генеральной совокупности, мы использовали обыденное представление о равной возможности исходов некоторого испытания. Например, раскрутив барабан с написанными на нем именами, мы ожидаем, что у этих имен одинаковые шансы быть выбранными.

Классическое определение вероятности использует именно это представление. В простейшем случае двух равновероятных исходов какого-то заранее не определенного события вероятность каждого исхода считают равной 50 процентам. В теории вероятностей обычно не пользуются процентами, а выражают вероятность дробями или десятичными числами от 0 (невозможное событие) до 1 (достоверное событие)¹⁹. Например, для игральной кости вероятность каждого из 6 равновероятных результатов будет равна 1/6. В виде формулы это записывается так: $p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = 1/6$ (вероятность выпадения единицы равна вероятности выпадение двойки... равна 1/6). Латинская буква p для обозначения вероятностей предпочитается примерно так же, как буква x в алгебраических уравнениях. Так же, как в алгебре, могут использоваться и другие буквы.

Если нас интересует вероятность сложного события, например, выпадения на игральной кости четного числа, то поскольку три простых равновероятных исхода соответствуют этому сложно-

¹⁹ Теория, правда, делает оговорки — в некоторых теоретических моделях могут наступать события, имеющие нулевую вероятность (см. параграф 3.1.5).

му событию (выпадение 2, 4 и 6), то и вероятность его вычисляется:
 $p(2 \text{ или } 4 \text{ или } 6) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2}$ или 0.5.

Пусть имеется некоторое множество равновероятных исходов какого-то испытания. Пусть некоторые из них выделены как предпочтительные в каком-то смысле (например, при игре в рулетку мы поставили рубль «на красное»), и нас интересует событие «выпадает один из предпочтительных исходов» (т.е. в нашем примере шарик остановился на красной цифре). Тогда вероятность этого события равна отношению числа предпочтительных исходов к общему числу равновероятных исходов (в нашем примере число красных клеток рулетки 18, всего же имеется 37 клеток, следовательно, вероятность выиграть равна $18/37$). Это так называемое **классическое определение вероятности**.

Для такой случайности и связанных с ней представлений о вероятности можно сформулировать свой вариант постулата о частотах: если мы имеем несколько равновероятных исходов какого-то испытания, которое можно повторять неограниченно, то при увеличении количества испытаний частоты появления всех исходов будут стремиться к равенству. Этот постулат является всего лишь уточнением практической интуиции, которой обладают люди до всякого знакомства с теорией вероятностей и статистикой.

Если монету изогнуть или опилить до формы усеченного конуса, она перестанет быть симметричной, и говорить о равных шансах выпадения герба и цифры становится невозможно. Можем ли мы в этом случае считать, что какие-то вероятности все же могут быть приписаны этим возможным событиям — выпадению герба и цифры? Математики считают — да, хотя теоретически рассчитать эти вероятности невозможно. Сформулированный выше постулат относится и к этому случаю: если проводить все более длинные серии подбрасываний подпиленной монеты, то частоты выпадения герба будут сходиться к этой неизвестной вероятности. Таким образом, неизвестные вероятности можно оценивать с помощью длинных серий испытаний.

2.1.4. Выборки и выборочные распределения.

Что будет, если мы подбросим сразу две одинаковые монеты? В результате такого опыта на столе могут оказаться две цифры, цифра и герб либо два герба. Вероятности этих результатов отображаются диаграммой, в которой столбцы подписаны нулем, единицей и двойкой — количеством выпавших гербов (рис. 2.1.4(1)а). Возможны четыре равновероятных исхода: «герб, герб», «цифра, герб», «герб, цифра» и «цифра, цифра» Результат «при двух бросках монеты выпал один герб» является суммой результатов двух из них («герб, циф-

ра» и «цифра, герб»)), поэтому соответствующая вероятность равна $1/4+1/4=1/2$.

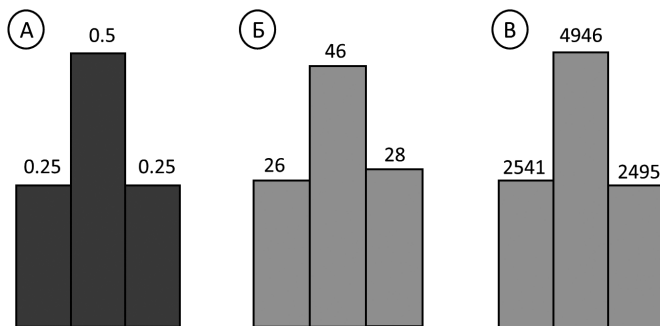


Рис 2.1.4(1) Диаграмма распределения вероятностей количества выпавших гербов при двукратном подбрасывании монеты (а), гистограмма частот количества выпавших гербов при 100 испытаниях (б) и при 10000 испытаниях (в) двукратного подбрасывания монеты.

Результаты испытаний похожи на теоретические диаграммы, причем можно сказать, что распределение вероятностей как бы «притягивает» частоты при увеличении числа испытаний. На рис. 2.1.4(1)б приведена гистограмма выборочного распределения результатов 100 подбрасываний пары монет. На рис. 2.1.4(1)в аналогичная гистограмма, полученная по результатам 10 000 подбрасываний²⁰. Легко заметить, что вторая гистограмма почти не отличается от диаграммы теоретического распределения.

Замечание. Мы будем всегда использовать слово «диаграмма» для изображений теоретического распределения и «гистограмма» — для выборочного распределения.

2.2. Точные определения и техника обращения с вероятностями

2.2.1. Элементарные события и классическая вероятность

Как уже обсуждалось выше, наиболее простым способом определяется вероятность в случае, когда имеется совокупность заведомо равно-возможных исходов некоторого испытания.

Элементарными событиями будем называть появление в результате испытания соответствующих исходов из этой совокупности, а *сложными*

²⁰ Мы не подбрасывали настоящие монеты, а смоделировали процесс на компьютере.

событиями какие-то совокупности этих элементарных событий. Теперь можно дать более четкое, чем в предыдущем параграфе, определение.

Определение 2.2.1(1). Классическая вероятность.

Пусть n — число всех равновозможных элементарных событий, а m — число элементарных событий, составляющих сложное событие A . Вероятность события A (обозначение $P(A)$) определяется следующим образом

$$P(A) = \frac{m}{n}$$

Примеры.

1) Вероятность того, что на игральной кости выпадет число больше двух, равна $4/6$, или $2/3$.

2) Вероятность того, что из тщательно перемешанной колоды из 36 игровых карт наугад вытянутая карта окажется черной масти, равна $18/36$, или $1/2$.

3) Вероятность того, что у первой проехавшей мимо вас завтрашним утром машины последняя цифра номера будет нечетной, равна $5/10$ или $1/2$.

4) Вероятность того, что при троекратном бросании монеты появится последовательность ГГЦ — «герб, герб, цифра» — (элементарное событие) равна $1/8$, а появление события «выпала ровно одна цифра» (сложное событие, состоящее из трех элементарных: ЦГГ, ГЦГ и ГГЦ) имеет вероятность $3/8$.

2.2.2. Ситуации с отсутствием равновозможности

В алгебре мы составляем уравнение, рассчитывая, например, исходя из неизвестного пути известную скорость, а затем решая его, находим неизвестную величину. Похожим образом в прикладной статистике, исходя из неизвестных вероятностей, мы рассчитываем вероятности других событий (которые уже произошли) и затем делаем выводы об этих исходных гипотетических вероятностях, почти столь же определенные, как извлечение значений неизвестных переменных при решении уравнений в алгебре.

Чтобы решать такие задачи, мы нуждаемся в довольно абстрактном определении вероятностей событий.

Определение 2.2.2(1). *Вероятностью случайного события в некоторой определенной ситуации называется число, к которому, как мы ожидаем, будет сходиться частота появления этого события в серии испытаний, в которых исходную ситуацию удастся точно повторить.*

Поскольку относительная частота определяется как отношение числа выпадений данного исхода к общему числу испытаний, вероятность может принимать значения между нулем и единицей включительно.

Для равновероятных событий, в полном согласии с опытом, мы ожидаем, что частоты их появления будут стремиться к равенству, поэтому таким событиям присваиваем вероятности, равные соответствующим долям единицы.

2.2.3. Формулы алгебры событий. Несовместимые и независимые события

Если определены вероятности элементарных событий, можно переходить к вычислению вероятностей более сложных событий, являющихся комбинацией определенных ранее элементарных.

Определение 2.2.3(1). Событием, противоположным событию A , назовем событие, состоящее в том, что A не произошло (обозначение: \bar{A} , «не A »).

Вероятности этих событий связаны естественным соотношением $p(A) + p(\bar{A}) = 1$.

Пример 2.2.3(2). Пусть испытанием является десятикратное подбрасывание монеты. Если событие $A =$ «монета все десять раз выпала гербом», тогда $\bar{A} =$ «монета хотя бы раз выпала цифрой».

Определение 2.2.3(3). Предположим, в некоторой ситуации возможны события A и B . Назовем их *суммой* событие, заключающееся в том, что произошло хотя бы одно из событий — A или B или оба сразу (сумма обозначается $A + B$). Как и в арифметике, сумма может состоять и из большего числа слагаемых.

Некоторые примеры вполне просты.

Пример 2.2.3(4). Пусть событие $A =$ «игральная кость выпала двойкой», событие $B =$ «игральная кость выпала четверкой». Тогда $(A+B) =$ «игральная кость выпала двойкой или четверкой».

Заметим, что в этом случае каждое из событий исключает другое, т.е. они не могут произойти вместе. Такие события называют *несовместимыми*. В этом случае вероятности связаны простой формулой $p(A + B) = p(A) + p(B)$.

Пример 2.2.3(5). Пусть событие $A =$ «игральная кость выпала числом, делящимся на два», событие $B =$ «игральная кость выпала числом, делящимся на три». Тогда событие $A+B =$ «игральная кость выпала числом, делящимся на два или на три».

В этом случае шестерка входит в оба события-слагаемых. В сумму входят четыре элементарных исхода: двойка, тройка, четверка и шестерка, следовательно, $p(A + B) = 4/6 \neq p(A) + p(B)$.

Определение 2.2.3(6). *Произведением* событий A и B назовем событие, состоящее в совместном наступлении обоих этих событий (обозначение AB).

Общая формула вероятности суммы событий такова: $p(A + B) = p(A) + p(B) - p(AB)$. Действительно, элементарные события,

из которых состоит AB , входят и в A , и в B , т.е. в $p(A) + p(B)$ посчитаны дважды. Вычитание $p(AB)$ корректирует сумму.

В нашем последнем примере $p(A) = 1/2$, $p(B) = 1/3$, $p(AB) = 1/6$, и AB состоит из единственного события — выпадения шестерки, т.е. $p(AB) = 1/6$, поэтому $p(A) + p(B) - p(AB) = 1/2 + 1/3 - 1/6 = 4/6$.

Пример 2.2.3(7). Мы уже писали, что при бросании двух неразличимых монет вероятность получить сочетание «герб-цифра» равна $1/2$. Этот факт является следствием равенства $p(AB) = p(A)p(B)$, которое верно в случае произведения независимых событий. Предположим, мы бросаем две легко различимые монеты — рубль и юань.

Пусть событие A = «юань выпал гербом», событие B = «рубль выпал гербом», и событие AB = «обе монеты выпали гербом». Пространство элементарных исходов для первой монеты: {«рубль выпал гербом», «рубль выпал цифрой»}, поэтому $p(A) = 1/2$, аналогично для второго: {«юань выпал гербом», «юань выпал цифрой»}: $p(B) = 1/2$.

Элементарные исходы для броска двух монет мы можем изобразить в виде таблицы.

Таблица 2.2.3(8). Возможные исходы броска 2 монет

	юань выпал гербом	юань выпал цифрой
рубль выпал гербом	ГГ	ГЦ
рубль выпал цифрой	ЦГ	ЦЦ

Все четыре исхода равновероятны. Действительно, если сначала мы подбросим рубль, то результат определит строку с вероятностью $1/2$. Далее, результат подбрасывания юаня определит с той же вероятностью столбец таблицы²¹. Таким образом, мы видим, что равенство $p(AB) = p(A)p(B)$ здесь не случайно, а связано с тем, что $p(AB)$ состоит из половины (определяемой выбором столбца таблицы) от половины (определяемой выбором строки таблицы) элементарных исходов для события AB . Заметим, что равенство вероятностей не изменится, если мы сначала подбросим юань, а потом рубль, или если мы бросим обе монеты одновременно. Больше того, если мы бросим одну и ту же монету два раза, то и тогда все четыре последовательности ГГ, ГЦ, ЦГ и ЦЦ будут иметь равные вероятности $1/4$. Именно поэтому, если мы, не различая монеты или броски, будем считать только количество выпавших гербов, то событие «из двух монет только одна выпала гербом»

²¹ Если с обеих сторон юаня изображены иероглифы, мы даже можем не знать, какая сторона монеты является гербом, что дополнительно свидетельствует в пользу равной вероятности исходов.

будет включать два элементарных события из четырех и, следовательно, иметь вероятность $1/2$.

Такого рода соотношение выполняется в том случае, если А и В — физически независимые события, т.е. мы понимаем и имеем основания так считать, что результат одного испытания (юаня) никак не зависит от результата другого испытания (рубля). Точно так же, бросая две игральные кости, мы имеем основания считать, что число на первой не зависит от числа, выпавшего на второй, и тогда результат «сумма равна 12 очкам» будет иметь вероятность $1/36$ (для обоснования надо будет нарисовать таблицу 6×6). В математической теории вероятностей формула перемножения вероятностей является основанием определения: события А и В независимы, если $p(AB) = p(A)p(B)$. Для физически независимых событий²² она верна по изложенным выше соображениям, однако она выполняется также и для событий, которые мы не можем признать физически независимыми. Например, для событий А = «игральная кость выпадает четной гранью» и В = «игральная кость выпадает числом, делящимся на три» $p(AB) = p(A)p(B)$, т.е. события независимы. Если бы мы использовали игральную кость с семью гранями²³, то эти события оказались бы зависимыми. Для семи граней событие А состоит из суммы трех элементарных событий «игральная кость выпадает двойкой», «игральная кость выпадает четверкой» и «игральная кость выпадает шестеркой», событие В = «игральная кость выпадает тройкой» и «игральная кость выпадает шестеркой». В этом случае $p(A) = 3/7$, $p(B) = 2/7$ и $p(AB) = 1/7$. Равенство нарушено.

Далее мы будем говорить только о физически независимых событиях, поэтому не вступим в конфликт ни с интуицией, ни с математическим определением.

Заметим, что аналогичные формулы перемножения вероятностей верны и для большего числа независимых событий. Например

$$p(ABC) = p(A)p(B)p(C)$$

для независимых А, В и С.

2.2.4. Биномиальные вероятности

В ситуациях, связанных с «хорошо организованной» случайностью, не важно, как именно реализован процесс с двумя возможными исходами и априори равными шансами реализации их обоих. Если вместо монеты мы будем подбрасывать игральную кость и ставить плюс, когда на выпавшей грани четное количество точек, и минус, когда нечетное, то поскольку

²² Однако дать математически строгое определение физической независимости событий невозможно, поскольку невозможно записать на математическом языке факт отсутствия причинных связей.

²³ Таковую «кость» можно реализовать, сделав длинную правильную семигранную призму (семигранный карандаш).

вероятность получить плюс при однократном бросании кости равна 0.5, то при 10 испытаниях вероятность получить 10 плюсов равна тому же самому числу $1/1024$, что и вероятность выпадения 10 монет гербом вверх. При этом также безразлично, будем ли мы бросать одну и ту же кость или монету 10 раз последовательно или подкинем одновременно 10 костей или монет. Также не изменится вероятность, если все 10 монет не одинаковы, а, напротив, легко различимы, например, принадлежат разным странам. Достаточно убедиться в том, что на каждой из них можно уверенно различить цифру и нечто, скорее напоминающее герб, чтобы распределение вероятностей совпадало с описанным выше.

Предположим, что в генеральной совокупности граждан России количество лиц мужского пола составляет ровно половину, а вторую половину составляют граждане женского пола. Тогда вероятность того, что 10 случайно выбранных граждан окажутся мужчинами, также равна $1/1024$. Как практически выбрать 10 совершенно случайных граждан — другой вопрос, причем не очень простой. Однако в конкретных случаях его можно удовлетворительным образом решить.

Пусть мы бросаем монету три раза и подсчитываем количество выпавших гербов. Вероятность того, что все три монеты выпадут гербом, равна $(1/2)^3$ по правилу перемножения вероятностей для независимых событий. Точно так же $1/8$ равна вероятность того, что сначала выпадут два герба, а затем цифра. Подобным же образом мы убеждаемся в том, что все восемь возможных последовательностей: ГГГ, ГГЦ, ГЦГ, ГЦЦ, ЦГГ, ЦГЦ, ЦЦГ, ЦЦЦ могут появиться с равной вероятностью $1/8$. Это элементарные события для нашего тройного испытания. Мы видим, что один из этих результатов дает три герба, три результата дают по два, три — по одному гербу и один результат дает ноль гербов. Таким образом, вероятность получить один герб за три броска равна $3/8$, и такова же вероятность получить два герба.

Мы собираемся обобщить этот результат на большее число испытаний, поэтому посмотрим на полученные числа с иной точки зрения. Чтобы получить последовательность с одним гербом и двумя цифрами, мы должны в последовательности ЦЦЦ заменить одну Ц на Г. Сделать это можно тремя способами — столькими способами, сколькими из цифр 1, 2, 3 можно выбрать одну. Это число в точности соответствует определению числа сочетаний, а именно, числа сочетаний из трех по одному C_3^1 . Поскольку (элементарные) события ГЦЦ, ЦГЦ и ЦЦГ несовместимы, то чтобы получить вероятность события «при трех бросках выпал один герб», надо вероятность одной последовательности (которая для любой последовательности равна $1/8 = 1/(2^3)$) умножить на количество таких последовательностей, откуда получим $C_3^1/(2^3)$.

Подбросим теперь четыре монеты. Чтобы получить ровно два герба, в последовательности ЦЦЦЦ две буквы Ц надо заменить на Г. Таких (разных) последовательностей будет столько, сколько существует способов выбрать из цифр 1, 2, 3, 4 любые две. Это количество равно числу сочетаний из четырех по два C_4^2 . Поскольку всего разных последовательностей из четырех гербов или цифр 2^4 , то вероятность получить два герба и две цифры равна $C_4^2/2^4$.

Это рассуждение можно провести для любых количеств бросков и числа ожидаемых выпадений гербов. Таким образом, если мы бросаем монету n раз, то вероятность получить ровно k гербов равна

$$\frac{C_n^k}{2^n}$$

Вероятности такого вида называются биномиальными по следующей причине: если мы рассмотрим несимметричную монету, которая выпадает гербом с вероятностью p , а цифрой с вероятностью $q=1-p$, то, используя аналогичные рассуждения, можем показать, что вероятность получить при n бросках ровно k гербов равна

$$C_n^k p^k q^{n-k}.$$

а это выражение есть k -й член²⁴ в разложении бинома $(p + q)^n$.

Формулу числа сочетаний и ее вывод читатель может найти в приложении 6.

2.3. Вероятности и частоты в SPSS

Пример 2.3(1) Соотношение данных в генеральной совокупности и выборке. В файле *IntroProb.sav* содержатся модельные данные о частоте выраженности различных типов темперамента, полученные в результате проведения опроса по методике Айзенка. Таблица данных состоит из единственной переменной, названной *introversion*, содержащей данные о выраженности экстраверсии или интроверсии испытуемых (три градации переменной: 1 — интроверт, 2 — экстраверт, 3 — неопределенный тип (так называемый амбиверт)). Поскольку амбиверт не является промежуточным между экстравертом и интровертом типом, то шкалу можно считать номинативной. Таблица содержит очень большое количество измерений этих двух параметров (426 679), статистика соответствует данным, накопленным на сайте http://psyline.retter.ru/auzeng/st_e.php (по состоянию на май 2016 года).

Будем в рамках этого примера рассматривать данные как исчерпывающие генеральную совокупность. Скажем, нас интересует такая генеральная совокупность, как население города, в котором живет 426 679 че-

²⁴ Точнее, $(k + 1)$ -й, если считать от вероятности получить 0 гербов, равной $C_n^0 p^0 q^n = q^n$.

людей. Такие полные данные о генеральной совокупности обычно либо не существуют вообще, либо засекречены. Допустим, что это секретная база, собранная сотрудниками спецслужб. В таком случае мы можем практически точно определить вероятность событий типа «случайный житель города — экстраверт» или «случайный житель города — интроверт». Для этого надо рассчитать относительную частоту появления того ли иного типа личности в нашей условной генеральной совокупности. Чтобы сделать это, зайдём в меню *Анализ — Описательные статистики — Частоты (Analyze — Descriptive statistics — Frequencies)*. Также, для получения гистограммы, в дополнительном меню «*Диаграммы (Charts)*» следует выбрать пункт *Гистограммы (Histograms)*. В качестве анализируемой переменной укажем *Introversion*. В результате получим таблицу 2.3(2).

Таблица 2.3(2). Распределение типов личности в условной «генеральной совокупности»

Интроверсия/экстраверсия

		Частота	Проценты	Валидный процент	Накопленный процент
Валидные	Интроверт	203 423	47.7	47.7	47.7
	Экстраверт	184 524	43.2	43.2	90.9
	неопределенный (амбиверт)	38 732	9.1	9.1	100.0
	Всего	426 679	100.0	100.0	

Из нее следует, что вероятность интроверсии составляет 0.477 (47.7%), экстраверсии — 0.432, а амбиверсии — 0.091.

Полученная гистограмма (соответствующая таблице) изображена на рис. 2.3(3).

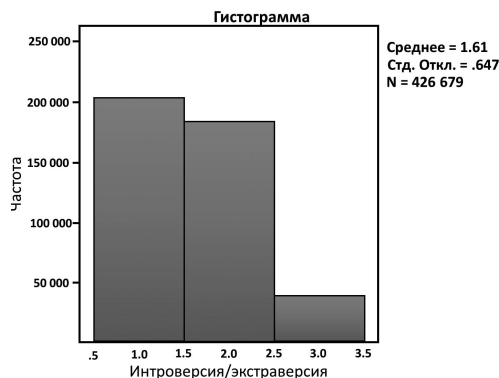


Рис. 2.3(3). Гистограмма распределения интроверсии/экстраверсии

Упражнение 2.3(4). В файле *TempProb.sav* содержатся данные о выраженности темперамента у жителей того же города. В переменной *temp* приведена принадлежность жителей к одному из четырех типов темперамента: 1 — холерик, 2 — сангвиник, 3 — флегматик, 4 — меланхолик или 5 — к неопределенному типу. Исходя из положения, что эти данные соответствуют генеральной совокупности, рассчитайте вероятности того, что случайный житель будет обладать тем или иным типом темперамента.

Пример 2.3(5). Смоделируем ситуацию выборочной оценки распределения вероятностей в генеральной совокупности. Допустим, в городе, о котором шла речь в Примере 2.3(1), проводится исследование, направленное на оценку соотношения интровертов и экстравертов. Исследование проводится неким институтом и этим проектом занимаются 10 исследователей. Каждый из исследователей тестирует 100 случайно отобранных жителей города. Таким образом, каждый из них получает выборку размером 100, а при объединении всех набранных данных можно получить большую выборку размером 1000. В файле *IntroProbResearch.sav* содержатся модельные данные такого исследования тысячи человек десятками исследователями. В первом столбце так же, как и в предыдущем примере, отмечена принадлежность испытуемых к интровертам, экстравертам или амбивертам, а во втором — *ResearcherId* — условный номер исследователя, который собрал эти данные (от 1 до 10). Сравним результаты, полученные отдельными исследователями, между собой и с данными из Примера 2.3(1). Для этого можно использовать функцию разбиения файла в SPSS: последовательно выбрать пункты меню **Данные** — **Расщепить файл (Data — Split file)** и в появившемся диалоговом окне выбрать пункт **Организовать вывод по группам (Organize output by)**, а затем вставить в поле **Группы образуются по: (Group based on:)** переменную *ResearcherId* и нажать кнопку **ОК**. После этого SPSS будет проводить любой анализ отдельно для каждой из групп с разными номерами исследователей.

При повторении расчета частот значений переменной *introversion*, описанного в Примере 2.3(1), в этом случае мы получим 10 различных таблиц. При этом различия между данными, полученными различными исследователями, будут довольно заметны. Например, доли интровертов, экстравертов и амбивертов по данным исследователя номер 3 будут 0.55, 0.42 и 0.03, а если мы возьмем данные, полученные исследователем номер 10, то доли будут 0.5, 0.4 и 0.1 соответственно.

Упражнение 2.3(6). Соотнесите результаты, полученные каждым из 10 исследователей с данными «генеральной совокупности». Насколько они отличаются от данных, полученных в генеральной совокупности?

Как будет соотноситься результат каждого из исследователя с «истинным» распределением типов личности в генеральной совокупности, если каждый исследователь обследует не 100, а только 10 жителей города? А если каждый сумеет обследовать 1000 жителей? В каком случае полученные данные будут более схожими с генеральной совокупностью?

Пример 2.3(7). Если мы объединим выборки, полученные 10 исследователями в одну большую «мета-выборку», то мы получим данные в общей сложности на 1000 случайных респондентов. Проведение частотного анализа соотношения типов личности в объединенной выборке даст результат, приведенный в таблице 2.3.4(1).

Таблица 2.3(8). Распределение типов личности в объединенной выборке (1000 человек)

Интроверсия

	Частота	Проценты	Валидный процент	Накопленный процент
Валидные	1.00	494	49.4	49.4
	2.00	423	42.3	91.7
	3.00	83	8.3	100.0
Всего	1 000	100.0	100.0	

Как видно из таблицы, полученные частоты того или иного типа личности уже достаточно близки к полученным в генеральной совокупности, о чем свидетельствует и форма гистограммы общих данных (рис. 2.3(9)).

Данный пример наглядно демонстрирует, что недостаточный объем выборки может приводить к искажению оценки вероятности того или иного события (или значения параметра) в генеральной совокупности. В дальнейшем мы обсудим закон больших чисел, которые строго показывают, что увеличение выборки приводит к все более точной оценке «истинного» значения любого параметра в генеральной совокупности.

Пример 2.3(10) Вызовите файл *IntroProb.sav* и попробуйте самостоятельно смоделировать исследование, в которое включено 100 случайных жителей города.

Упражнение 2.3(11). Для отбора случайной выборки заданного размера в матрице данных в диалоговом окне отбора переменных **Данные — Отобразить наблюдения (Data — Select Cases)** надо выбрать вариант *Случайная подвыборка (Random sample of cases)* и, нажав кнопку *Подвыборка (Sample)*, указать *Точно k наблюдение из первых n наблюдений (Exactly k cases from the first n cases)*, где k — размер подвыборки (100 для этого упражнения), а n — размер всей выборки (426 679 в нашем примере). После этого SPSS случайным образом от-

берет k наблюдений и будет включать в анализ только их до тех пор, пока фильтр не будет изменен или отключен (вариант **Все наблюдения (All cases)** в диалоговом окне отбора наблюдений). Рассчитайте таблицу частот значений переменной *introversion*. Насколько полученные результаты соответствуют данным генеральной совокупности?

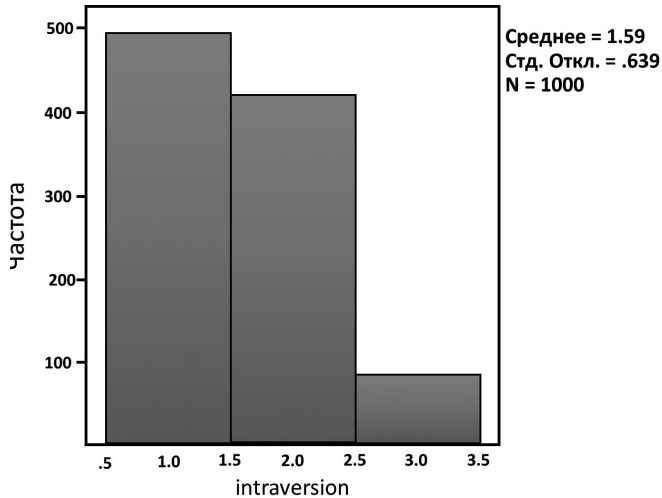


Рисунок 2.3(9). Гистограмма распределения интроверсии/экстраверсии в объемной выборке

Упражнение 2.3(12). Повторите эту процедуру для 1000 случайно отобранных испытуемых. Проверьте, насколько ваш ответ на заданный выше вопрос соответствует полученному в этом случае результату. Повышается или понижается согласованность данных в генеральной совокупности и в выборке большего размера по сравнению с небольшой выборкой?

Пример 2.3(13). Рассмотрим ситуацию биномиального эксперимента, описанного в подпараграфе 2.2.4, в котором подбрасывают три монеты и считают количество выпавших гербов. Теоретические вероятности выпадения 0, 1, 2 и 3 в случае симметричной монеты равны $1/8$, $3/8$, $3/8$ и $1/8$ соответственно (см. 2.2.4). В файле *3FairCoins.sav* приведены смоделированные данные эксперимента, в котором три монеты подбрасывают 1000 раз и каждый раз фиксируют количество гербов (единственный столбец — *heads*). Результаты расчета частоты появления каждого из возможных исходов представлены в таблице 2.3(14).

Таблица 2.3(14). Частота выпадения k гербов при подбрасывании трех монет

Heads		Частота	Проценты	Валидный процент	Накопленный процент
Валидные	0.00	127	12.7	12.7	12.7
	1.00	377	37.7	37.7	50.4
	2.00	375	37.5	37.5	87.9
	3.00	121	12.1	12.1	100.0
	Всего	1 000	100.0	100.0	

Из таблицы видно, что относительная частота достаточно близка к рассчитанным выше теоретическим вероятностям.

Упражнение 2.3.6(2). В файле *3UnknownCoins.sav* содержатся модельные результаты трех серий экспериментов, аналогичных описанному в предыдущем примере. Однако в одном из экспериментов использовались симметричные монеты, а в двух других искривленные так, что в одной серии более вероятным было выпадение герба, а в другой — выпадение цифры. Данные представлены в трех столбцах (*heads1*, *heads2* и *heads3*). Порядок серий с симметричными и асимметричными монетами неизвестен. Мы предлагаем читателю самостоятельно рассчитать частоты выпадения того или иного числа гербов и определить, в какой переменной содержатся данные об эксперименте с симметричной монетой, а в каких — данные об искривленных в пользу орла или цифры монетах.

Глава 3. _____

Случайные величины

3.1. Элементарные сведения о случайных величинах

3.1.1. Случайная величина. Распределение

Определение 3.1.1(1). Если каждому возможному в результате некоторого испытания случайному событию определенным образом поставлено в соответствие некоторое число, говорят, что этим самым задана *случайная величина*. Иными словами, случайная величина — это величина, принимающая в результате испытания то или иное случайное числовое значение.

Мы будем обозначать случайные величины большими латинскими буквами X , Y и т.д.

С каждой случайной величиной связано некоторое множество чисел — значений, которые она может принимать. *Закон распределения* случайной величины — это множество ее возможных значений вместе с соответствующими вероятностями их появления.

Примеры, которые рассматривались в предыдущей главе, могут интерпретироваться как случайные величины. Таковы: количество выпавших гербов при бросании определенного количества монет, суммы очков на верхних гранях брошенных одной или нескольких игральных костей, уровень тревожности случайно выбранного человека, измеренный с помощью определенной методики.

Далеко не всегда само случайное событие естественно и очевидно характеризуется возможными числовыми значениями. Гораздо чаще для наших целей приходится придумывать специальные полезные числовые характеристики случайных событий.

Пример 3.1.1(2). Разрабатывается методика по развитию функций зрительно-пространственной памяти у детей с отставанием в развитии. Для проверки эффективности методики предполагается давать детям тестовое задание «Рисунок дома», входящее в одну из багатель нейропсихологического обследования, перед началом использования развивающей методики и через несколько недель — после завершения работы с детьми. Полученные рисунки предполагается предъявить эксперту, не указывая, какой из рисунков нарисован испытуемым раньше. Эксперт выберет лучший из двух рисунков, а лаборант напишет

в протоколе единицу, если лучшим окажется более поздний рисунок, и нуль в противном случае.

Если методика абсолютно неэффективна, то два рисунка ребенка могут с равными вероятностями продемонстрировать сдвиг к лучшему или к худшему — в силу случайных обстоятельств, например, лучшего или худшего самочувствия ребенка в момент тестирования. Таким образом, в этом случае количество единиц в итоговом протоколе будет случайной величиной. Она будет иметь разобранное в подпараграфе 2.2.4 биномиальное распределение с количеством испытаний n , равным числу тестируемых детей, и равными 0.5 вероятностями того, что второй рисунок окажется лучше или хуже первого. Это распределение совпадает с распределением числа выпавших гербов при n -кратном подбрасывании симметричной монеты. Если методика эффективна, то мы будем иметь случайную величину с иным распределением, в котором единиц будет ожидаемо больше половины, хотя совсем не обязательно, что их будет n .

Перейдем теперь к рассмотрению того, каким образом может быть задан закон распределения случайной величины. Наиболее просто задается распределение в случае, когда случайная величина принимает лишь конечное число значений. В этом случае ее распределение удобно задать таблицей.

Пример 3.1.1(3). Для игральной кости вероятности принять любое из шести значений равны между собой. Таблица распределения соответствующей случайной величины (обозначим ее через X) выглядит так

X	1	2	3	4	5	6
	1/6	1/6	1/6	1/6	1/6	1/6

В верхней строке записаны возможные значения случайной величины, в нижней, под каждым значением, соответствующая этому значению вероятность.

Пример 3.1.1(4). Подбрасывая три монеты, будем фиксировать количество выпавших гербов. Расчет вероятностей (см. 2.2.4) дает для нашей случайной величины Y следующую таблицу распределения вероятностей:

Y	0	1	2	3
	1/8	3/8	3/8	1/8

Отметим важное обстоятельство. Задание распределения случайной величины подобной таблицей подразумевает, что в результате испытания появится одно и только одно из значений, помещенных в верхнюю строку таблицы. Это значит, что сумма вероятностей, составляющих нижнюю строку, равна единице.

Случайная величина как теоретическая идеализация однозначно определяется своим законом распределения. При этом реализовать ее можно различными способами, как мы уже неоднократно говорили. Все выводы, которые можно сделать об абстрактной, заданной таблицей распределения случайной величине, будут верны и для любой реализации. Именно поэтому имеет смысл развивать общую теорию случайных величин. Вследствие этого мы и будем часто отождествлять случайную величину с ее законом распределения.

Случайную величину, описанную в последнем примере, можно задать и диаграммой распределения (Рис. 3.1.1(5)).

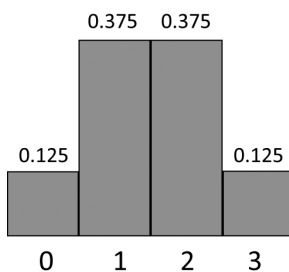


Рис. 3.1.1 (5). Диаграмма распределения результатов испытания неизвестного процесса.

3.1.2. Математическое ожидание и выборочное среднее

Предположим, что некий игровой клуб предлагает своим посетителям, заплатив за вход 2000 рублей, 1000 раз сыграть в игру с автоматом, который с указанными в Примере 3.1.1(4) вероятностями выдает за одну игру 0, 1, 2 или 3 рубля. Стоит ли соглашаться на такие условия?

За 1000 игр мы можем ожидать появления примерно 125 нулей (1/8 от 1000), по 375 единиц и двоек (по 3/8 от 1000) и 125 троек. Это значит, что мы выиграем примерно $125 \cdot 0 + 375 \cdot 1 + 375 \cdot 2 + 125 \cdot 3 = 1500$ рублей, в среднем 1.5 рубля за игру. Мы видим, что соглашаться играть не стоит. Равные шансы выиграть и проиграть будут у нас, если за вход мы заплатим 1500 рублей.

Мы получили средство решать большое количество похожих проблем, причем не только связанных с азартными играми²⁵. Эти средства окажутся важной составляющей математических методов в психологических исследованиях. Ответ «полтора рубля за игру в среднем», как легко видеть, получается перемножением значений и их вероятностей: $0.125 \cdot 0 + 0.375 \cdot 1 + 0.375 \cdot 2 + 0.125 \cdot 3 = 1.5$. По

²⁵ Заметим, что первоначально теория вероятностей развивалась прежде всего на примерах задач о разумных ставках в аналогичных играх.

сути, выше мы и сделали это вычисление, только в два этапа: сначала по вероятностям определили ожидаемые количества результатов каждого вида, а потом, поделив на общее количество испытаний, получили произведения значений на ожидаемые частоты, которые и суть вероятности.

Определение 3.1.2(1). *Математическим ожиданием случайной величины* называется сумма произведений возможных значений этой величины на соответствующие этим значениям вероятности.

Если задать случайную величину X общей таблицей

X	x_1	...	x_n
	p_1	...	p_n

то математическое ожидание X (которое обозначается символом M_X) задается формулой:

$$M_X = x_1 p_1 + \dots + x_n p_n$$

Математическое ожидание дает среднее ожидаемое значение результата испытаний, как мы видели в предыдущем примере.

В предыдущей главе мы определили случайную выборку из генеральной совокупности. Поскольку случайные величины не обязательно связаны с выбором из генеральной совокупности, то определение выборки надо расширить.

Определение 3.1.2(2). *Выборка значений случайной величины, имеющая объем n* , это результат n последовательных независимых испытаний случайной величины.

Определение 3.1.2(3). *Средним выборочным значением* (коротко — *выборочным средним*) называется среднее арифметическое значений, составляющих выборку.

Если выборка представляет собой реализацию испытаний случайной величины с известным распределением, то выборочное среднее будет лежать вблизи теоретического среднего — математического ожидания. Говорят, что выборочное среднее *оценивает* математическое ожидание соответствующей случайной величины.

Поскольку частоты появления исходов испытаний приближаются по мере увеличения объема выборки к вероятностям, постольку и выборочное среднее приближается к математическому ожиданию.

Хотя случайность дает свои коррективы и количества результатов каждого вида будут несколько отличаться от ожидаемых, все же с помощью теории вероятностей можно рассчитать, что практически всегда выборочное среднее по 1000 испытаний для описанной выше игры с автоматом будет лежать в интервале от 1.42 до 1.58 (вероятность такого события равна 0.998, см. параграф 4.2).

Выборка объема n испытаний некоторой случайной величины X обозначается $\{x_1, x_2, \dots, x_n\}$, а выборочное среднее:

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \dots + x_n)$$

3.1.3. Выборочная дисперсия

Следующие важные характеристики случайной величины оценивают разброс колебаний ее значений. Рассмотрим две выборки: $\{90, 100, 110\}$ и $\{0, 100, 200\}$. Как легко видеть, они имеют одно и то же среднее значение — 100. Однако колебания значений первой составляют 20 единиц, а второй — 200 единиц. Наиболее удобной характеристикой разброса оказалась дисперсия.

Определение 3.1.3(1). *Выборочная дисперсия*, которая обозначается s^2 рассчитывается для выборки $\{x_1, x_2, \dots, x_n\}$, по формуле

$$s^2 = \frac{1}{n-1}((x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2)$$

Так же, как и в случае пары «выборочное среднее — математическое ожидание», значение выборочной дисперсии при испытаниях нашей случайной величины располагается вблизи теоретической дисперсии распределения, формулу для расчета которой мы приведем в параграфе 3.2.

3.1.4. Случайная величина и генеральная совокупность

Если интересующий нас параметр в генеральной совокупности принимает какой-то набор значений с соответствующими частотами, то случайный выбор одного представителя из данной совокупности и измерение данного параметра будут тождественны испытанию случайной величины, имеющей то же самое распределение. Рис. 3.1.4(1) одинаковыми графиками отображает распределение параметра (в нашем гипотетическом случае — тревожности) в генеральной совокупности и распределение случайной величины, которая реализуется измерением того же параметра у случайно выбранного представителя генеральной совокупности.

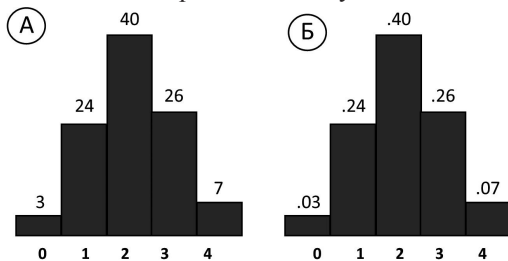


Рис. 3.1.4(1). а) Гистограмма распределения результатов тестирования тревожности у всех граждан РФ (в процентах), б) Диаграмма распределения случайной величины «значение по тесту тревожности случайно выбранного гражданина РФ».

Однако, как мы видим, случайная величина может и не порождаться никакой генеральной совокупностью. Самый простой пример мы рассматривали выше — количество выпавший гербов при подбрасывании монеты. Другой, очень важный пример — выборочное среднее. Даже если мы набираем случайную выборку из генеральной совокупности, распределение возможных значений выборочного среднего не соответствует распределению признака в какой-то простой генеральной совокупности²⁶.

3.1.5. Непрерывные случайные величины. Распределения и их характеристики

Мы измеряем величины всегда с некоторой точностью. Если длина измеряется рулеткой, на ленте которой нанесены сантиметры, то результатом будет натуральное число, выражающее длину измеряемого предмета в сантиметрах. Более точная рулетка с миллиметровыми делениями даст результаты более точные, но выражающиеся целым числом миллиметров. Если предмет исследования — рост человека, то вариации в генеральной совокупности этого параметра (*размах* — разность максимального и минимального значения) составят около двух метров. Это значит, что случайная величина «рост в миллиметрах случайно выбранного человека» может принимать любое из примерно 2000 значений. Хотя определения, которые мы ввели выше, позволяют в принципе работать с таким количеством значений, все же это явно неудобно. Если принять во внимание, что точность измерения может увеличиваться неограниченно и при каждом таком увеличении количество значений интересующей нас случайной величины также будет расти, то имеет смысл рассмотреть разумную альтернативу.

Вместо большого количества значений можно рассмотреть сразу бесконечное количество. Эта идеализация вполне наглядна. Предположим, мы, вооружившись ружьем, целимся и стреляем в центр мишени. Нас интересует расстояние от центра мишени до отверстия от пули. Эта случайная величина может меняться от нуля до нескольких метров, а проще всего «с запасом» считать в каких-то случаях возможным, хотя и очень маловероятным, отклонение на любое, как угодно большое расстояние. Мы абстрагируемся от того, что в каждой начальной ситуации обладаем при измерении минимальной единицей, а считаем, что реальное отклонение может быть любым положительным действительным числом, хотя мы и не можем его измерить с точностью, превышающей наши текущие возможности.

²⁶ В этом случае генеральная совокупность все же может быть определена, но как весьма абстрактный математический объект.

Какова в этой идеализации вероятность того, что пуля попадет точно в центр мишени, т.е. случайная величина примет значение в точности 0 при любой точности измерения? Похожий вопрос: какова вероятность того, что пуля отклонится ровно на 1 см? на 3/4 см? на $\sqrt{2}$ см? На произвольное фиксированное число сантиметров? Такая вероятность не может измеряться положительным числом, поскольку возможных значений слишком (бесконечно) много, поэтому и сумма даже очень маленьких вероятностей в бесконечном числе окажется (бесконечно) больше единицы²⁷.

Таким образом, вероятность появления каждого значения должна быть равной нулю. Однако нуль нулю рознь. Попасть с пяти метров близко к центру мишени все же вероятнее, чем на миллион километров в сторону или выше. В последнем случае ствол пришлось бы поднять почти вертикально, а такую ошибку не допустит даже очень плохой стрелок.

Для того чтобы различать «большие» нули от «маленьких» вводится функция: так называемая *плотность распределения*. Рис. 3.1.5(1) показывает плотность распределения для нашего примера. На вопрос «Какова вероятность отклонения пули от центра ровно на 1 см?» плотность распределения дает тривиальный ответ — нуль.

Но зато на вопросы типа «Какова вероятность того, что отклонения пули от центра будут лежать между 1 см и 2 см?» ответ будет вполне разумным: эта вероятность измеряется площадью криволинейной трапеции, которая на рисунке заштрихована. Похожим образом вычисляются вероятности отклонения пули от центра в любой интервал между двумя числами²⁸. Вся площадь под графиком функции, естественно, должна быть равна единице, поскольку какое-то значение отклонения от центра обязательно будет иметь место, а значит, с вероятностью единица случайная величина примет какое-то неотрицательное значение.

²⁷ Ситуация на самом деле несколько сложнее. Для обоснования утверждения надо сослаться на то, что точки на плоскости в каком-то смысле равноправны: например, вероятности попадания в точки окружности, равноотстоящие от центра прицеливания, равны. Тогда окажется бесконечно много точек с некоторой определенной вероятностью, и сумма будет заведомо бесконечной.

²⁸ В области отрицательных значений аргумента (левее начала координат) случайная величина значений принимать не может, поскольку отрицательных расстояний не бывает. Плотность распределения в этой области равна нулю. В отличие от положительных значений аргумента, вероятность также равна нулю и для попадания отклонения пули в интервал между любыми отрицательными числами.

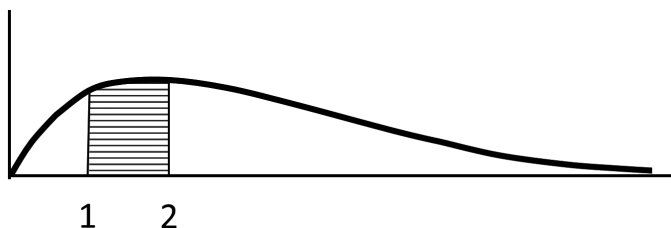


Рис. 3.1.5(1). Плотность распределения случайной величины «Расстояние от точки попадания пули в мишень до центра мишени (в сантиметрах)». Заштрихована площадь, соответствующая расстоянию от 1 до 2 сантиметров.

Для выборки $\{x_1, x_2, \dots, x_n\}$, объема n испытаний непрерывной случайной величины формулы расчета выборочного среднего и выборочной дисперсии те же самые, что и в дискретном случае:

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \dots + x_n), \quad s^2 = \frac{1}{n-1}((x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2),$$

хотя формулы расчета теоретических величин — математического ожидания и дисперсии — задаются не суммами, а интегралами. Мы не будем их здесь приводить. Интересующиеся могут обратиться к учебным пособиям по теории вероятностей. Как и в случае дискретных величин, и для непрерывных случайных величин при увеличении объема выборки выборочные среднее и дисперсия стремятся к теоретическим значениям математического ожидания и дисперсии.

3.1.6. Зачем нужны математическое ожидание и дисперсия

Квадратный корень из дисперсии играет важную роль во многих случаях, поэтому введены специальные термины: корень из теоретической дисперсии называется *среднеквадратическим отклонением* и обычно обозначается строчной греческой буквой σ (читается «сигма»), корень из выборочной дисперсии называется *стандартным отклонением* и обычно обозначается латинской строчной буквой s . Стандартное отклонение оценивает среднеквадратическое отклонение с тем большей точностью, чем больше выборка. Когда последнее неизвестно, используется эта оценка. Для любой случайной величины с помощью математического ожидания и среднеквадратического отклонения можно рассчитать область наиболее вероятных значений. Если на числовой прямой отложить от математического ожидания вправо и влево отрезки длиной по 3σ , то с вероятностью не меньше чем $8/9$ случайная величина будет принимать значения из этой области. Если обозначить математическое ожидание случайной величины X через M_X , то сказанное выше можно переформулировать так: с вероятностью не меньше $8/9$ (т.е. в среднем не реже чем в восьми случаях из девяти) случайная величина X будет принимать

значения из интервала $[M_X - 3\sigma, M_X + 3\sigma]$ (где M_X и σ — математическое ожидание и среднеквадратическое отклонение этой случайной величины). Эта оценка верна для любой случайной величины²⁹, но для тех, с которыми мы обычно имеем дело, вероятность получить значение внутри этого интервала еще значительно ближе к единице.

3.1.7. Простейшие операции над случайными величинами и выборками. Стандартизация

Если на каждой грани игральной кости вырезать еще по одной точке, т.е. увеличить каждое выпадающее значение на единицу, то в терминах случайных величин это будет означать прибавление к случайной величине единицы. Аналогично операция прибавления/вычитания постоянного числа может быть проведена для любой случайной величины: все значения надо изменить на данное число, а вероятности оставить прежними. Математическое ожидание новой случайной величины будет равно прежнему плюс/минус добавляемое число:

$$M_{X+a} = M_X + a$$

Если к каждому члену выборки $\{x_1, x_2, \dots, x_n\}$, составленной из испытаний случайной величины X , прибавить число a , то новое среднее значение будет равно $\bar{x} + a$. Эту выборку можно считать испытаниями случайной величины $X+a$, и выборочное среднее будет близко к математическому ожиданию новой величины в том же смысле, в каком близко среднее \bar{x} и математическое ожидание M_X .

При изменении случайной величины на константу ее дисперсия не меняется.

Чтобы умножить/поделить случайную величину на константу, надо умножить/поделить на эту константу каждое её значение, а вероятности оставить прежними. Для игральной кости операция выглядит похоже на операцию прибавления, но вместо вырезания новых точек удобнее будет выгравировать на гранях новые числа, равные соответствующим произведениям.

При умножении случайной величины на константу ее дисперсия умножается на квадрат этой константы. При делении случайной величины на константу ее дисперсия делится на квадрат этой константы.

Есть более полезные примеры этой операции. Вместо количества выпавших гербов при 15-кратном бросании симметричной монеты можно рассчитывать частоты выпадения гербов, деля количество гербов на 15. Математическое ожидание также поделится на 15 и станет равным 0.5, как и в случае одной монеты. В этом нет ничего удивительного:

²⁹ Она следует из неравенства Чебышева. Доказательство можно найти в (Кричевец, Шикин, Дьячков, 2003).

средняя частота выпадения герба в пятнадцати бросаниях и должна колебаться вокруг 0.5.

Если вычесть из случайной величины ее математическое ожидание, то математическое ожидание результата станет равным нулю. Аналогично, если вместо выборки $\{x_1, x_2, \dots, x_n\}$, рассмотреть выборку $\{(x_1 - \bar{x}), (x_2 - \bar{x}), \dots, (x_n - \bar{x})\}$, т.е. вычесть из каждого члена среднее выборочное, то среднее новой выборки будет равно нулю. Такая выборка называется *центрированной*, а операция вычитания среднего *центрированием*.

Если поделить случайную величину на среднее квадратическое отклонение, то у новой случайной величины дисперсия будет равна единице (математическое ожидание также поделится на дисперсию). Аналогично, если вместо выборки $\{x_1, x_2, \dots, x_n\}$, рассмотреть выборку $\{x_1/s, x_2/s, \dots, x_n/s\}$, т.е. поделить каждый член выборки на стандартное отклонение, то выборочная дисперсия новой выборки станет равной единице, а среднее значение поделится на s .

Объединяя эти две операции, можно получить стандартизованную случайную величину: сначала вычесть математическое ожидание, затем поделить на среднее квадратическое отклонение. Полученная случайная величина будет иметь математическое ожидание равное нулю и дисперсию равную единице. Аналогично можно получить стандартизованную выборку: сначала из всех членов выборки вычесть среднее значение, затем поделить на стандартное отклонение. Полученная выборка

$$\left\{ \frac{x_1 - \bar{x}}{s}, \frac{x_2 - \bar{x}}{s}, \dots, \frac{x_n - \bar{x}}{s} \right\}$$

будет иметь среднее значение равное нулю и выборочную дисперсию равную единице. Все доказательства будут приведены в параграфе 3.2.

Стандартизация важна как составная часть разнообразных операций. Поясним ее смысл. В параграфе, посвященном измерению, сообщалось, что для интервальных шкал допустимо линейное преобразование $x' = ax + b$. Поскольку ни нуль шкалы, ни единица измерения не имеют чаще всего определенного смысла, то предпочтительной единицей оказывается стандартное отклонение генеральной совокупности (или ее выборочная оценка). В самом деле, что значит оценка двенадцать баллов по шкале тревожности некоторого нового теста? Если же окажется, что среднее значение тестовых баллов, оцененное по достаточно большой выборке, т.е. с достаточной точностью, равно шести, а стандартное отклонение равно четырем, то респондент, получивший двенадцать сырых баллов, имеет стандартизованную оценку $(12 - 6)/4 = 1.5$, а это означает, что он входит в 10 процентов наиболее тревожных людей в ге-

неральной совокупности. Вывод опирается на тот почти повсеместный факт, что тестовые оценки распределены в генеральной совокупности по закону, близкому к нормальному (см. следующую главу), а для него вероятности, относимые к стандартизованным баллам, постоянны и записаны в хорошо известных таблицах. Стандартизованные баллы называются также *Z-баллами* или *Z-оценками* (по-английски *Z-scores*). Ими измеряют не только оценки конкретных лиц, но и разности средних оценок по группам и многое другое.

3.2. Случайные величины (уточняющее продолжение)

3.2.1. Операции над случайными величинами

Случайные величины могут подвергаться различным преобразованиям. Наиболее простые операции: сложение с константами и умножение на константу. Пусть имеется случайная величина X , распределение которой задано таблицей

X	x_1	...	x_n
	p_1	...	p_n

Случайная величина $X+a$ задается в таком случае следующей таблицей:

$X+a$	x_1+a	...	x_n+a
	p_1	...	p_n

Случайная величина $X \cdot a$ задается такой таблицей:

$X \cdot a$	$x_1 \cdot a$...	$x_n \cdot a$
	p_1	...	p_n

Если случайная величина X реализована, например, игральной костью, то прибавление числа 3 к этой случайной величине реализуется переписыванием значений на соответствующих гранях (1 превращается в 4, 2 — в 5 и т.д.). Естественно, что при такой операции вторая строка таблицы распределения остается без изменений. Умножение на три реализуется заменой единицы на тройку, двойки — на шесть и т.д.

Мы можем применять к случайным величинам и более сложные функции, например, функцию $y = x^2$. Действуя аналогично предыдущей операции, возведем в квадрат каждое значение, оставив вторую строку без изменений. Для игральной кости результат будет такой:

X	1	4	9	16	25	36
	1/6	1/6	1/6	1/6	1/6	1/6

Однако в некоторых случаях при таком преобразовании может возникнуть трудность. Пусть у нас есть случайная величина X

X	-1	0	1	2
	0.1	0.2	0.3	0.4

Таблица для случайной величины X^2 получится такой:

X^2	1	0	1	4
	0.1	0.2	0.3	0.4

В первой строке есть совпадающие значения. Их следует объединить, сложив соответствующие вероятности:

X^2	0	1	4
	0.2	0.4	0.4

Таблицу распределения случайной величины $Y = f(X)$ для любой функции f можно построить аналогично. В психологических измерениях нередко разумным оказывается преобразование шкал $x' = \ln(x)$. Ему будет соответствовать логарифмическое преобразование соответствующих случайных величин (например, логарифм количества ошибок в каком-то задании). Естественно, распределение такой случайной величины будет задаваться таблицей:

$\ln(X)$	$\ln(x_1)$...	$\ln(x_n)$
	p_1	...	p_n

3.2.2. Математическое ожидание и выборочное среднее

Пусть имеется случайная величина X с заданной таблицей распределения

X	x_1	x_2	...	x_n
	p_1	p_2	...	p_n

Согласно определению ее математическое ожидание задается формулой

$$M_X = x_1 p_1 + x_2 p_2 + \dots + x_n p_n \text{ или } M_X = \sum_{i=1}^n x_i p_i.$$

Предположим, что проведено k испытаний случайной величины X , при этом k_1 раз она приняла значение x_1 , k_2 раз — значение x_2 , ..., k_n раз — значение x_n :

$$k_1 + k_2 + \dots + k_n = k.$$

Найдем среднее арифметическое всех этих k значений (т.е. выборочное среднее). Имеем

$$\begin{aligned} & \frac{\overbrace{x_1 + \dots + x_1}^{k_1} + \overbrace{x_2 + \dots + x_2}^{k_2} + \dots + \overbrace{x_n + \dots + x_n}^{k_n}}{k} = \\ & = \frac{x_1 k_1 + x_2 k_2 + \dots + x_n k_n}{k} = x_1 \frac{k_1}{k} + x_2 \frac{k_2}{k} + \dots + x_n \frac{k_n}{k}. \end{aligned}$$

Дробь $\frac{k_i}{k}$ представляет собой относительную частоту появления в k

испытаниях события «случайная величина X приняла значение x_i » (в k испытаниях событие « $X = x_i$ » произошло k_i раз). При больших значениях k относительная частота примерно равна вероятности события « $X = x_i$ », т.е. p_i , поэтому

$$x_1 \frac{k_1}{k} + x_2 \frac{k_2}{k} + \dots + x_n \frac{k_n}{k} \approx x_1 p_1 + x_2 p_2 + \dots + x_n p_n = M_X.$$

Таким образом, в серии из большого количества испытаний среднее арифметическое полученных в этой серии значений случайной величины (выборочное среднее) будет приближаться к ее математическому ожиданию по мере того, как частоты значений будут приближаться к соответствующим вероятностям.

Этот факт имеет два важных следствия.

Следствие 3.2.2(1). Математическое ожидание случайной величины, распределение которой нам неизвестно, можно оценить средним арифметическим значений в достаточно большой серии ее последовательных испытаний. Чем длиннее серия, тем точнее эта оценка³⁰.

Следствие 3.2.2(2). В практически интересных случаях можно оценивать наиболее вероятный результат серии испытаний, исходя из математического ожидания случайной величины.

Пример 3.2.2(3). Предлагается следующая азартная игра: Бросают два игральных кубика. Если полученная сумма больше 10, то игрок выигрывает 10 копеек, в противном случае проигрывает 1 копейку. Имеет ли ему смысл играть в эту игру 12 000 000 партий?

Из 36 возможных исходов выпадения двух различных кубиков в трех случаях выпадает благоприятствующая игроку сумма. Это значит, что 10 копеек он выигрывает с вероятностью $3/36 = 1/12$, а одну копейку проигрывает (скажем так: выигрывает (-1) копейку) с вероятностью $33/36 = 11/12$. В таком случае

$$M_X = 10 \cdot \frac{1}{12} + (-1) \cdot \frac{11}{12} = -\frac{1}{12}$$

Проигрывая в среднем $1/12$ копейки за партию, за 12 000 000 партий игрок проиграет около 1 000 000 копеек, или 10 000 рублей.

3.2.3. Дисперсия и среднееквадратическое отклонение

Следующий простейший пример показывает, что случайные величины с равным математическим ожиданием могут существенно различаться по степени близости к нему.

³⁰ Строго говоря, это верно не для всякой случайной величины. Например, случайная величина может иметь распределение, но не иметь математического ожидания. Здесь мы не будем обсуждать эти вопросы.

Рассмотрим две случайные величины:

X	-1	0	1
	0.25	0.5	0.25

и

Y	-100	0	100
	0.25	0.5	0.25

Нетрудно видеть, что $M_x = M_y = 0$. Но если для величины X отклонение от нуля незначительно, то для величины Y оно весьма заметно.

Если представить, что значения этих случайных величин — это предполагаемый выигрыш/проигрыш в рублях, то хотя обе игры справедливы (такие игры называются *играми с нулевой суммой*), игра Y может быть сочтена более рискованной, чем игра X.

Показателем разброса значений случайной величины служит ее дисперсия.

Определение 3.2.3(1). Если величина X задана таблицей

X	x_1	x_2	...	x_n
	p_1	p_2	...	p_n

то дисперсия случайной величины X может быть вычислена по формуле

$$D_X = (x_1 - M_X)^2 p_1 + (x_2 - M_X)^2 p_2 + \dots + (x_n - M_X)^2 p_n,$$

или сокращенно $D_X = \sum_{i=1}^n (x_i - M_X)^2 p_i.$

Можно рассчитать дисперсии приведенных выше примеров.

$$D_X = (-1 - 0)^2 \cdot 0.25 + (0 - 0)^2 \cdot 0.5 + (1 - 0)^2 \cdot 0.25 = 0.5,$$

$$D_Y = (-100 - 0)^2 \cdot 0.25 + (0 - 0)^2 \cdot 0.5 + (100 - 0)^2 \cdot 0.25 = 5000.$$

Значения случайных величин различаются в 100 раз, а их дисперсии — в $100^2 = 10000$ раз, что, разумеется, является следствием возведения разностей, характеризующих отклонение от среднего значения, в квадрат.

Разброс случайных величин можно было бы характеризовать и другими показателями, например

$$|x_1 - M_X|p_1 + |x_2 - M_X|p_2 + \dots + |x_n - M_X|p_n$$

но квадратичная мера разброса оказалась исключительно эффективной. Приведенная ниже Теорема 3.2.5(5) — лишь один пример такой эффективности.

Следующий пример показывает, что у случайных величин с одинаковыми значениями также могут быть разные дисперсии, если различаются вероятности одинаковых значений:

X	-1	0	1
	0.25	0.5	0.25

и

Y	-1	0	1
	0.4	0.2	0.4

Дисперсия первой равна 0.5, как мы уже убедились, дисперсия второй равна 0.8.

Замечание 3.2.3(2). Математическое ожидание может быть любым числом, а дисперсия всегда неотрицательна.

Выше шла речь об операциях над случайными величинами. В случае линейных преобразований случайной величины X (т.е. преобразований вида $Y = aX + b$, где a и b — некоторые числа) математическое ожидание и дисперсию получившейся случайной величины Y можно вычислить, исходя из этих же числовых характеристик величины X . Именно, справедливы следующие формулы:

$$M_Y = aM_X + b, D_Y = a^2M_X$$

Рассмотрим пример: случайная величина X задана таблицей

X	1	2	3
	1/3	1/3	1/3

Она имеет математическое ожидание, равное 2, и дисперсию $2/3$. Распределение случайной величины $4X+5$ задается таблицей

$4X+5$	9	13	17
	1/3	1/3	1/3

Легко видеть:

$$M_Y = 13, \text{ т.е. } M_Y = 4 \cdot 2 + 5 = 4 \cdot M_X + 5, \text{ а } D_Y = 4^2/3 + 0 + 4^2/3 = 32/3, \\ \text{т.е. } D_Y = 4^2 \cdot D_X.$$

В общем случае утверждение легко доказать. Мы советуем читателю обратить внимание на тот факт, что употребление алгебраических переменных только увеличивает ясность, а вовсе не усложняет понимание. Действительно, числа из примеров могут не вполне отчетливо указывать на то, откуда они ведут свое происхождение. Хотя мы стараемся сделать происхождение чисел прозрачным, это не всегда просто осуществить. В последней строке только что разобранный пример встречаются три выражения « 4^2 ». Первые два — это отклонения от математического ожидания двух разных значений случайной величины, третье — квадрат множителя в линейной функции. Понять настоящее значение констант поэтому не так просто. В общем доказательстве эта проблема преодолевается.

Пусть дана случайная величина

X	x_1	x_2	...	x_n
	p_1	p_2	...	p_n

с математическим ожиданием

$$M_X = x_1p_1 + x_2p_2 + \dots + x_np_n$$

и дисперсией $D_X = (x_1 - M_X)^2p_1 + (x_2 - M_X)^2p_2 + \dots + (x_n - M_X)^2p_n$.

Случайная величина $Y = aX + b$, задается таблицей

$Y = aX + b$	$ax_1 + b$	$ax_2 + b$...	$ax_n + b$
	p_1	p_2	...	p_n

Ее математическое ожидание

$$M_Y = (ax_1 + b)p_1 + (ax_2 + b)p_2 + \dots + (ax_n + b)p_n.$$

Продолжаем, раскрывая скобки, а затем группируя слагаемые:

$$\begin{aligned} M_Y &= ax_1p_1 + bp_1 + ax_2p_2 + bp_2 + \dots + ax_np_n + bp_n = \\ &= ax_1p_1 + ax_2p_2 + \dots + ax_np_n + bp_1 + bp_2 + \dots + bp_n = \\ &= a(x_1p_1 + x_2p_2 + \dots + x_np_n) + b(p_1 + p_2 + \dots + p_n) \end{aligned}$$

В последней строке в первой скобке стоит математическое ожидание M_X , а во второй скобке сумма всех вероятностей p_i , равная единице. Таким образом, последняя строка равна тому самому выражению $aM_X + b$, которое мы доказывали. Первая часть завершена.

Далее, рассчитываем дисперсию случайной величины $Y = aX + b$, подставляя вместо M_Y равное ему $aM_X + b$:

$$D_Y = (ax_1 + b - aM_X - b)^2p_1 + (ax_2 + b - aM_X - b)^2p_2 + \dots + (ax_n + b - aM_X - b)^2p_n.$$

В скобках сокращаются b и $-b$, а константу a можно вынести за скобки:

$$D_Y = (ax_1 + b - aM_X - b)^2p_1 + (ax_2 + b - aM_X - b)^2p_2 + \dots + (ax_n + b - aM_X - b)^2p_n = [a(x_1 - M_X)]^2p_1 + [a(x_2 - M_X)]^2p_2 + \dots + [a(x_n - M_X)]^2p_n.$$

Последняя строка преобразуется далее:

$$\begin{aligned} D_Y &= [a(x_1 - M_X)]^2p_1 + [a(x_2 - M_X)]^2p_2 + \dots + [a(x_n - M_X)]^2p_n = \\ &= a^2(x_1 - M_X)^2p_1 + a^2(x_2 - M_X)^2p_2 + \dots + a^2(x_n - M_X)^2p_n = \\ &= a^2[(x_1 - M_X)^2p_1 + (x_2 - M_X)^2p_2 + \dots + (x_n - M_X)^2p_n] = a^2D_X. \end{aligned}$$

Обе формулы доказаны.

Напомним, что выборочная дисперсия сходится при увеличении объема выборки к теоретической дисперсии. Стандартное отклонение s сходится при этом к величине \sqrt{D} , которая получила специальное наименование среднеквадратическое отклонение и обозначается греческой буквой σ .

3.2.4. Стандартизация случайной величины

Определение 3.2.4(1). Случайная величина, у которой математическое ожидание равно 0, а дисперсия равна 1, называется *стандартной* или *стандартизованной* случайной величиной.

Пусть имеется случайная величина X с математическим ожиданием m и среднеквадратическим отклонением σ . Нетрудно показать, что случайная величина

$$Y = \frac{X - m}{\sigma}$$

является стандартной случайной величиной. Действительно, вышеприведенную формулу можно переписать в виде

$$Y = \frac{1}{\sigma}X - \frac{m}{\sigma}.$$

Теперь, вспомнив, что $M_{aX+b} = aM_X + b$ и $D_{aX+b} = a^2D_X$ и заменяя

a на $\frac{1}{\sigma}$, а b на $\frac{m}{\sigma}$, получаем

$$M_Y = \frac{1}{\sigma}m - \frac{m}{\sigma} = 0 \text{ и, поскольку } D_X = \sigma^2, D_Y = \frac{\sigma^2}{\sigma^2} = 1.$$

Это именно то, что мы хотели доказать.

Замечание 3.2.4(2). Если случайная величина моделирует какие-то процессы реального мира, то ее значения имеют размерность. Например, эти значения могут измеряться в метрах, килограммах и т.п. При этом математическое ожидание случайной величины имеет ту же размерность, что и сама случайная величина, а размерность дисперсии равна квадрату размерности случайной величины. Например, если случайная величина измеряется в рублях, то ее дисперсия — в рублях в квадрате. Однако среднеквадратическое отклонение σ имеет ту же размерность, что и случайная величина. Более того, в некоторых случаях саму σ имеет смысл выбрать в качестве более удобной единицы измерения, отражающей вероятностные свойства исходной шкалы. Именно это и делается посредством стандартизации $x' = (x - m) / \sigma$ (заметим, что такое преобразование допустимо для интервальных шкал — см. параграф 1.1). После этой операции точка m (математическое ожидание) получает значение 0 а точка, которая лежала на расстоянии σ справа от математического ожидания, получает значение 1.

3.2.5. Сумма случайных величин

Точно так же, как операции над числами делают из системы пересчета весьма продуктивную систему расчетов, так и операции над случайными величинами превращают их в продуктивную систему описаний и предсказаний случайных событий. Сумма случайных величин — узловое понятие в этой системе.

Если проводятся испытания двух случайных величин, X_1 и X_2 , то можно определить их сумму, просто складывая результаты испытаний. Например, бросая две игральные кости, мы можем суммировать выпавшие результаты. Другой пример: при подсчете тестовых баллов мы собираемся суммировать результаты по отдельным вопросам.

Если в результате испытания величина X принимает значение x , а величина Y — значение y , то случайная величина $X+Y$ принимает значение $x+y$.

В некоторых случаях нам надо будет складывать случайные величины, не являющиеся независимыми, но пока мы займемся только сложением независимых случайных величин. Независимость случайных величин определяется через независимость событий. Как и в случае независимости событий, яснее всего ситуация для физически независимых случайных величин: например, при подбрасывании двух монет количество гербов при бросании первой и второй монеты — независимые случайные величины. В этом будет уверен каждый, кто считает, что выпадение герба на первой монете и выпадение герба на второй монете — независимые события.

Определение 3.2.5(1). Суммой случайных величин $X+Y$ называется случайная величина, результат испытания которой получается проведением испытаний случайных величин-слагаемых X и Y и суммированием полученных результатов.

Сначала рассмотрим пример.

Пример 3.2.5(2). Пусть случайные величины X и Y задаются таблицами распределения следующего вида:

X	-1	0	1
	0.2	0.5	0.3

и

Y	-1	1
	0.4	0.6

Пусть также у нас есть основания считать, что случайные величины физически не зависят одна от другой. Так как X принимает три различных значения, а Y — два, то для суммы $X + Y$ получаем шесть возможностей. Выпишем их и вычислим попутно вероятности, используя независимость событий «случайная величина X принимает i -е значение» и «случайная величина Y принимает j -е значение» и перемножая соответствующие вероятности, как мы это уже делали для независимых событий (подпараграф 2.2.3):

$$p(X = -1) p(Y = -1) = 0.2 \cdot 0.4 = 0.08, \text{ при этом } X + Y = -2;$$

$$p(X = -1) p(Y = 1) = 0.2 \cdot 0.6 = 0.12, \text{ при этом } X + Y = 0;$$

$$p(X = 0) p(Y = -1) = 0.5 \cdot 0.4 = 0.2, \text{ при этом } X + Y = -1;$$

$$p(X = 0) p(Y = 1) = 0.5 \cdot 0.6 = 0.3, \text{ при этом } X + Y = 1;$$

$$p(X = 1) p(Y = -1) = 0.3 \cdot 0.4 = 0.12, \text{ при этом } X + Y = 0;$$

$$p(X = 1) p(Y = 1) = 0.3 \cdot 0.6 = 0.18, \text{ при этом } X + Y = 2.$$

Таким образом, для величины $X+Y$ получаем следующую таблицу распределения:

X+Y	-2	-1	0	1	2
	0.08	0.2	0.25	0.3	0.18

Легко проверить, что математическое ожидание суммы $X + Y$ равно сумме математических ожиданий X и Y :

$$M_{X+Y} = M_X + M_Y.$$

Этот факт нетрудно доказать и в общем случае. Мы проведем доказательство для независимых случайных величин, каждая из которых принимает по два значения. Разобравшийся в нем читатель легко докажет теорему для произвольных наборов значений (отличие только в размере таблиц и употребляемых в таких случаях многоточиях вместо произвольного числа слагаемых).

Теорема 3.2.5(3). Математическое ожидание суммы независимых случайных величин равно сумме математических ожиданий слагаемых:

$$M_{X+Y} = M_X + M_Y.$$

Пусть случайные величины X и Y задаются таблицами распределения следующего вида:

X	x_1	x_2	и	Y	y_1	y_2
	p_1	p_2			q_1	q_2

Их математические ожидания равны соответственно $x_1 p_1 + x_2 p_2$ и $y_1 q_1 + y_2 q_2$.

Распределение их суммы задается следующей таблицей:

X+Y	$x_1 + y_1$	$x_1 + y_2$	$x_2 + y_1$	$x_2 + y_2$
	$p_1 + q_1$	$p_1 + q_2$	$p_2 + q_1$	$p_2 + q_2$

Математическое ожидание $X+Y$ вычислим, расположив четверку слагаемых в виде таблицы

$$M_{X+Y} = \begin{array}{r} (x_1 + y_1)p_1 q_1 + (x_1 + y_2)p_1 q_2 \\ + \\ (x_2 + y_1)p_2 q_1 + (x_2 + y_2)p_2 q_2 \end{array}$$

Раскрыв скобки, получим восемь слагаемых:

$$\begin{array}{r} x_1 p_1 q_1 + y_1 p_1 q_1 + x_1 p_1 q_2 + y_2 p_1 q_2 \\ + \\ x_2 p_2 q_1 + y_1 p_2 q_1 + x_2 p_2 q_2 + y_2 p_2 q_2. \end{array}$$

Поменяем теперь местами второй и третий столбцы:

$$\begin{array}{r} x_1 p_1 q_1 + x_1 p_1 q_2 + y_1 p_1 q_1 + y_2 p_1 q_2 \\ + \\ x_2 p_2 q_1 + x_2 p_2 q_2 + y_1 p_2 q_1 + y_2 p_2 q_2. \end{array}$$

В новой таблице плюсы показывают порядок суммирования: в первой четверке мы сначала сложим элементы строк, а во второй четверке — сначала сложим элементы столбцов. Это позволит нам далее вынести за скобки нужные множители. Проведим группировку с помощью квадратных скобок и выносим за эти скобки множители:

$$M_{X+Y} = [x_1p_1q_1 + x_1p_1q_2] + [x_2p_2q_1 + x_2p_2q_2] + [y_1p_1q_1 + y_1p_2q_1] + [y_2p_1q_2 + y_2p_2q_2] = \\ = x_1p_1[q_1 + q_2] + x_2p_2[q_1 + q_2] + y_1q_1[p_1 + p_2] + y_2q_2[p_1 + p_2].$$

Но $p_1+p_2=1$ и $q_1+q_2=1$, поэтому квадратные скобки вместе с их содержимым можно опустить. В результате получаем выражение $M_{X+Y} = x_1p_1 + x_2p_2 + y_1q_1 + y_2q_2$ где первые два слагаемые это M_x , а вторые два — M_y .

Теорема доказана.

Замечание 3.2.5(4). Сумму случайных величин можно определить и для случайных величин, которые не являются независимыми. В этом случае нельзя будет вычислять вероятность совместного наступления событий как произведение вероятностей этих событий. Однако теорема о сложении математических ожиданий верна и в этом случае, хотя доказательство и оказывается чуть сложнее.

В случае суммирования n дискретных случайных величин, где $n > 2$, таблицу распределения можно строить последовательным суммированием. Как и в привычной арифметике, результат не будет зависеть от порядка слагаемых. И в этом случае математические ожидания слагаемых в сумме дают математическое ожидание суммарной случайной величины:

$$M(X_1 + \dots + X_n) = MX_1 + \dots + MX_n.$$

Следующая теорема служит весомым аргументом в пользу употребления дисперсии в качестве меры разброса случайной величины.

Теорема 3.2.5(5). Если X и Y — независимые случайные величины, то

$$D_{X+Y} = D_X + D_Y.$$

Никакие другие меры разброса случайной величины не дают такую простую формулу сложения. Доказательство теоремы мы даем в приложении 2. В следующих главах будут рассматриваться чрезвычайно важные следствия из этой теоремы, а сейчас разберем один пример сложения дисперсий.

Рассмотрим две одинаково распределенные случайные величины.

X	-1	1
	0.5	0.5

и

Y	-1	1
	0.5	0.5

Эти распределения можно назвать симметричными, поэтому нет ничего удивительного в том, что их математические ожидания равны нулю. Поскольку отклонения от математического ожидания каждого из двух значений случайной величины X равны единице, то и дисперсия X равна единице: $D_X = 1^2 \cdot 0.5 + 1^2 \cdot 0.5 = 1$. Также единице равна дисперсия Y .

Сумма $X+Y$ будет иметь следующее распределение:

$X+Y$	-2	0	2
	0.25	0.5	0.25

Ее математическое ожидание также равно нулю, а дисперсия $D_{X+Y} = 2^2 \cdot 0.25 + 0 + 2^2 \cdot 0.25 = 2$ т.е., естественно, равна сумме дисперсий слагаемых.

Рассмотрим теперь среднее арифметическое этих двух случайных величин

$$\frac{X+Y}{2}$$

Таблица его распределения такова:

$(X+Y)/2$	-1	0	1
	0.25	0.5	0.25

а дисперсия $D_{(X+Y)/2} = 1^2 \cdot 0.25 + 0 + 1^2 \cdot 0.25 = 0.5$. т.е. вдвое меньше дисперсии каждого из слагаемых. Это пример того, что разброс среднего арифметического существенно понижается по сравнению с разбросом слагаемых.

Упражнение 3.2.5(6). Равны ли дисперсии случайных величин $2X$ и $X+X$?

Ответ. Запись $X+X$ не вполне корректна, обычно мы пишем X_1+X_2 с оговоркой, что эти случайные величины распределены так же, как X . После этого ясно, что дисперсия X_1+X_2 равна $2D_x$, а дисперсия $2X$ равна $4D_x$. Случайную величину $2X$ можно интерпретировать как умноженную на константу 2 величину X , но и как сумму максимально сильно связанных случайных величин: если первая из них принимает значение a , то и вторая должна принять это же значение (например, так или почти так связаны роста однойцевых близнецов). При более слабой связи дисперсия суммы будет принимать промежуточные значения — в данном случае между $2X$ и $4X$.

Замечание 3.2.5(7). Отметим важный случай, когда величины X_1, X_2, \dots, X_n имеют один и тот же закон распределения. Иногда вместо термина «независимые одинаково распределенные случайные величины» употребляют термин «независимые наблюдения (испытания) данной случайной величины».

Замечание 3.2.5(8). Для случайной величины X , принимающей значения 0 и 1 с вероятностями соответственно q и p , математическое ожидание равно $0 \cdot q + 1 \cdot p = p$. Имея в виду, что $q=1-p$ и, подставив в формулу дисперсии возможные значения случайной величины с соответствующими вероятностями и p в качестве математического ожидания, получаем

$$(0 \cdot p)^2 \cdot q + (1 \cdot p)^2 \cdot p = p^2q + q^2p = pq(p + q) = pq.$$

Сумма двух экземпляров такой случайной величины X есть биномиальная величина с двумя испытаниями и вероятностью p исхода, число выпадений которого мы суммируем. Аналогично, биномиальная величина с n испытаниями есть сумма n экземпляров X . Эта величина обозначается обычно $V(n,p)$.

По теореме о сложении математических ожиданий, математическое ожидание числа выпавших единиц в n -кратном испытании данной случайной величины, т.е. математическое ожидание биномиальной случайной величины $V(n,p)$ равно np (что совершенно не удивительно: при 100-кратном подбрасывании монеты, выпадающей гербом с вероятностью 0.6, среднее количество выпавших гербов равно $100 \cdot 0.6 = 60$). По теореме о сложении дисперсий дисперсия биномиальной случайной величины $V(n,p)$ равна npq . Для случая равных вероятностей $V(n,1/2)$ математическое ожидание и дисперсия равны соответственно $n/2$ и $n/4$.

3.3. Случайные величины. Практикум

Пример 3.3(1) Рассмотрим пример, иллюстрирующий смысл выборочного среднего на материале ситуации азартной игры. Допустим, нам предлагают сыграть на игральном автомате, причем одна игра стоит 5 рублей. Выигрыш может составлять 0, 10, 100 и 1000 рублей. Нам точно не известен закон, по которому распределяются выигрыши, но из надежных источников мы получили их статистику. Данные о результатах 1000 игр содержатся в файле *Play.sav* в переменной *profit*. Имея эти данные, мы можем оценить, насколько выгодно играть в такую игру. Для этого рассчитаем выборочное среднее по имеющимся у нас данным. Для этого в SPSS надо зайти в меню **Анализ — Описательные статистики — Описательные (Analyze — Descriptive statistics — Descriptives)**. В диалоговом окне в поле *Переменные (Variable(s))* перенесем переменную *profit*. После нажатия кнопки *OK* в окне вывода появится таблица.

Таблица 3.3(2). Описательная статистика выигрышей

Описательные статистики

	N	Минимум	Максимум	Среднее	Среднекв. отклонение
Profit	1 000	.00	1 000.00	6.3600	35.58675
N валидных (по списку)	1 000				

В столбце «Среднее» находится интересующее нас число — выборочная оценка математического ожидания — в среднем за игру мы будем получать 6.36 рубля, т.е. при достаточно долгой игре, платя за каждую попытку 5 рублей, в среднем мы будем выигрывать 1.36 рубля. Нам эта игра выгодна.

Замечание. Мы не должны забывать, что выборочная оценка не совпадает с истинным математическим ожиданием. Однако в практических случаях мы можем использовать наши оценки, считая их удовлетворительно точными (чем больше выборка, тем точнее). Об оценках точности оценок см. главу 5.

Упражнение 3.3(3). Выведите таблицу частот выигрышей по данной выборке (*Описательные статистики — Частоты (Descriptive statistics — Frequencies)*); затем перенести имя *profit* в поле *Переменные (Variables)*). Рассматривая частоты как оценки вероятностей, подставьте полученные значения в формулу математического ожидания вместе с величинами соответствующих выигрышей. Убедитесь, что полученный результат совпадает с выведенным средним в предыдущей таблице, равным 6.36.

Пример 3.3(4). Обратимся теперь к использованию оценок математического ожидания и дисперсии случайной величины в рамках простого психологического исследования. Допустим, мы исследуем состояние так называемых управляющих функций у учащихся двух школ: обычной и инклюзивной. Мы измеряем состояние этих функций с помощью нейропсихологического обследования, в рамках которого ребенку дается ряд заданий, оценивающих функции произвольной регуляции деятельности. На основании их выполнения высчитывается интегральный показатель состояния интересующих нас функций. Этот полученный балл для конкретного человека, разумеется, не случаен — он зависит от уровня развития когнитивной сферы, созревания соответствующих мозговых структур и множества других факторов. Однако мы можем предполагать, что две группы наших испытуемых — дети из обычной школы и дети из инклюзивной школы — это случайные выборки

из двух разных генеральных совокупностей, которые характеризуются, возможно, разными средними значениями и дисперсиями. Если связать каждую из генеральных совокупностей со случайной величиной, реализующейся случайным выбором представителя из данной совокупности, то наши выборки являются реализациями испытаний этих случайных величин, а мы оцениваем их параметры.

В файле *Executive.sav* приведены модельные данные такого исследования. В первом столбце *index_exec* содержатся интегральные оценки управляющих функций 50 детей-первоклассников, 25 из которых учатся в обычной школе, и еще 25 — в инклюзивной школе. Переменная *index_exec* содержит штрафные баллы при выполнении заданий, поэтому низкие оценки соответствуют хорошему состоянию функций, а высокие — плохому. Принадлежность ученика к школе закодирована переменной *school*: 1 — учащийся обычной школы, 2 — учащийся инклюзивной школы.

Рассчитаем среднее и дисперсию в двух выборках. Чтобы сделать это отдельно для двух групп испытуемых в SPSS, можно зайти в меню *Анализ — Сравнение средних — Средние (Analyze — Compare Means — Means)*. В диалоговом окне в поле *Список зависимых переменных (Dependent List)* перенесем оцениваемую переменную *index_exec*, а в поле *Список независимых переменных (Independent List)* — переменную *school*. Далее, в дополнительном меню *Параметры (Options)* перенесем в поле *Статистики в ячейках (Cell Statistics)* из полного списка возможных параметров *Дисперсию (Variance)*. Закроем это окно и запустим анализ (кнопка *OK*). В результате будет получена таблица.

Таблица 3.3(5) Таблица средних по группам.

Отчет

exec_index				
school	Среднее	N	Стд. Отклонение	Дисперсия
обычная школа	5.9251	25	1.05820	1.120
инклюзивная школа	6.0520	25	3.31392	10.982
Итого	5.9886	50	2.43547	5.932

Из таблицы видно, что выборочные средние в двух группах различаются незначительно (5.9251 и 6.0520 в обычной и инклюзивной школе, соответственно). Мы можем надеяться, что реальные математические ожидания случайных величин «индекс управляющих функций учащегося обычной школы» и «индекс управляющих функций учащегося инклю-

живной школы» также мало отличаются друг от друга. С другой стороны, дисперсии в двух выборках отличаются очень сильно: 1.03 и 10.982 в обычной и инклюзивной школах соответственно.

Для прояснения ситуации полезно построить гистограммы для двух выборок: *Графика — Устаревшие диалоговые окна — Гистограмма (Graphics — Legacy Dialogs — Histogram)*. Чтобы построить два отдельных графика для каждой выборки, в поле *Строки (Rows)* следует перенести переменную *school*. Получившаяся гистограмма приведена на рисунке 3.3(5).

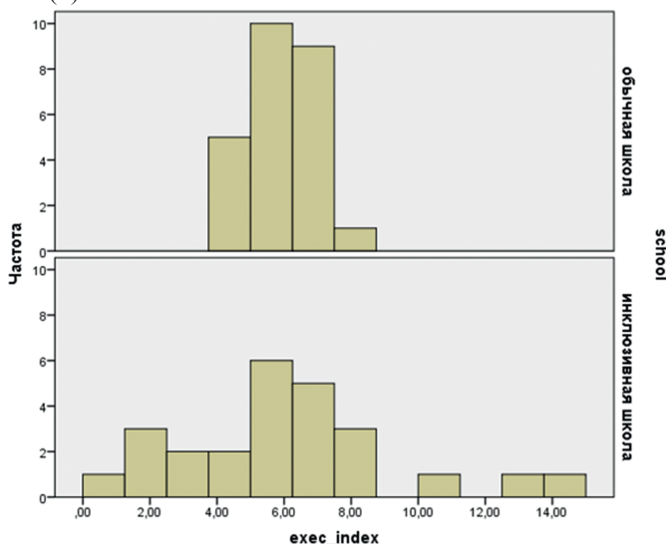


Рис. 3.3(5). Гистограмма распределения индекса управляющих функций для обычной и инклюзивной школ

На графиках отчетливо видно, что в обычной школе почти все дети получили баллы, близкие к среднему (5.92), а в инклюзивной школе при практически таком же среднем разброс данных намного больше. Таким образом, в этом примере различия между выборками проявились не в среднем, а в дисперсии. Мы можем думать, что выборочные оценки отражают характеристики генеральных совокупностей (и связанных с ними случайных величин).

Содержательно различие дисперсий объясняется тем, что в инклюзивной школе состав учеников более разнороден по сравнению с обычной школой, что вполне объяснимо: в инклюзивной школе дети с нарушениями в развитии могут демонстрировать дефицит различных когнитив-

ных функций, а управляющие функции могут быть у них не только не нарушены, но и хорошо развиты — помогая (компенсаторно) справиться с другими трудностями.

Пример 3.3(6). Стандартизация переменных.

В некоторой незадачливой стране перевод школьного образования на единый экзамен происходит поэтапно. В 2016 году знания по математике оценивались традиционно, по пятибалльной шкале, а по русскому языку проводился ЕГЭ, оценивавшийся по стобальной шкале. На факультете психологии столичного университета школьников принимают по результатам этих двух дисциплин, суммируя оценки.

Вопрос 3.3(7). Какой из предметов окажет большее влияние на суммарный балл в этом университете?

В файле *EgeSum.sav* содержатся модельные данные результатов 100 школьников. В переменной *Math* приведены оценки по математике (в шкале от 1 до 5), а в переменной *Rus* — по русскому языку (в шкале от 1 до 100).

Упражнение 3.3(8). Вычислите переменную $Sum = Math + Rus$. Произведите сортировку школьников по переменной *Sum* (**Данные** — **Сортировать наблюдения (Data — Sort cases)**), затем перенесите имя *Sum* в поле **Переменные (Variables)**.

Убедитесь, что среди лучших двадцати есть несколько школьников, получивших по математике два и три. Таким образом, оценка по русскому языку дает больший вклад.

Есть универсальный способ сделать вклад дисциплин равным — провести стандартизацию данных. Тогда сырые значения окажутся приведены к единому масштабу — со средним значением 0 и дисперсией равной 1. Рассчитаем сначала среднее и стандартное отклонение для двух переменных (**Анализ** — **Описательные статистики** — **Описательные (Analysis — Descriptive Statistics — Descriptives)**). Результат представлен в таблице.

Таблица 3.3(9). Описательная статистика исходных оценок по ЕГЭ.

Описательные статистики

	N	Минимум	Максимум	Среднее	Стд. отклонение
Math	100	2.00	5.00	4.0900	.90000
Rus	100	61.00	86.00	74.3400	4.68874
N валидных (целиком)	100				

Из таблицы видно, что стандартные отклонения двух переменных значительно отличаются и поэтому суммирование и привело к большему влиянию переменной *Rus* с большим стандартным отклонением. Для того чтобы можно было объединить две оценки, мы стандартизуем обе переменные, т.е. вычтем из значений каждой переменной ее среднее и поделим на ее стандартное отклонение. Деление на стандартное отклонение придаст слагаемым одинаковый масштаб (стандартное отклонение станет равно единице). Сначала стандартизуем переменную *Math*. Среднее по ней составляет 4.09, стандартное отклонение — примерно 0.9. Выберем в меню пункты **Преобразовать — Вычислить (Transform — Compute)**. В открывшемся окне в строке **Вычисляемая переменная (Target variable)** слева сверху введем название, которые мы хотим дать новой, стандартизованной переменной — *MathSt*. Далее справа сверху нам нужно вписать формулу для расчета этой переменной, а именно $(Math - 4.09)/0.9$. Нажатие кнопки *OK* приведет к появлению нового столбца в файле с данными — стандартизованных значений. Если теперь мы рассчитаем среднее и стандартное отклонение новой переменной *MathSt* (так же, как описано выше — при помощи описательной статистики), среднее составит 0, а стандартное отклонение 0.99. Тот факт, что стандартное отклонение у нас получилось приблизительно, а не точно, равным единице, объясняется просто: тем, что мы делили на округленное стандартное отклонение.

Вторую переменную *Rus* мы можем стандартизовать тем же способом. Однако существует и более простой путь: в знакомом нам окне расчета описательных статистик (**Анализ — Описательные статистики — Описательные, Analysis — Descriptive Statistics — Descriptives**), перенесем в поле **Переменные (Variables)** те переменные, которые хотим стандартизовать (*Math* и *Rus*) и отметим внизу галочкой пункт **Сохранить стандартизованные значения в переменных (Save standardized values as variables)**. После нажатия *OK* в таблице данных появятся две колонки: *ZMath* — стандартизованные значения переменной *Math* и *ZRus* — стандартизованные значения переменной *Rus*. Можно заметить, что значения испытуемых по рассчитанной нами переменной *Math.st* и рассчитанной компьютером переменной *ZMath* очень похожи и различаются только из-за того, что мы округлили значения стандартного отклонения.

Рассчитаем средние и стандартные отклонения новых стандартизованных переменных (выберем **Анализ — Описательные статистики — Описательные** и введем в поле **Переменные** *ZMath* и *ZRus*). Результаты представлены в таблице 3.3(10) Как видим, теперь средние равны 0, а стандартные отклонения 1 — мы привели переменные к единой шкале.

Таблица 3.3(10). Описательная статистика стандартизированных переменных.

Описательные статистики

	N	Минимум	Максимум	Среднее	Среднекв. отклонение
Zscore(Math)	100	-2.32222	1.01111	0.0000000	1.0000000
Zscore(Rus)	100	-2.84512	2.48681	0.0000000	1.0000000
N валидных (по списку)	100				

Теперь, после такого преобразования, мы можем рассчитать итоговый суммарный балл по двум предметам. Используя знакомую процедуру вычисления переменной, вычислим новую переменную, сложив переменные Z_{Math} и Z_{Rus} , назовем ее $SumZ$.

Вопрос 3.3(11). Чему будет равно среднее новой переменной?

Упражнение 3.3(12). Произведите ранжирование по переменным Sum и $SumZ$ (**Преобразовать — Ранжировать наблюдения (Transform — Range cases)**). В поле переменных можно перенести обе переменных разом. Убедитесь, что ранги не совпадают.

Упорядочите наблюдения по переменной $SumZ$. Убедитесь, что среди лучших двадцати процентов присутствуют абитуриенты лишь с верхними оценками по русскому языку и математике.

Глава 4.

Нормальное распределение

4.1. Начальные сведения о нормальном распределении

4.1.1. Стандартное нормальное распределение

Среди распределений случайных величин особо выделяется так называемое *нормальное распределение*. Его статус связан с тем, что нормальное распределение является предельным случаем распределения самых разнообразных массовых процессов. Например, если 1000 раз подкинуть монету, то вероятности того, что количество выпавших гербов попадет в определенный интервал, можно вычислить с хорошей точностью, предполагая, что результат имеет нормальное распределение.

Не только подобные экспериментальные, но и разнообразные практические ситуации дают похожие на нормальное распределения. Например, если станок штампует семидесятимиллиметровые гвозди, то их реальная длина будет колебаться вокруг семидесяти миллиметров приблизительно по закону нормального распределения. Очень часто приблизительно нормально распределены количественно измеряемые признаки в биологических видах — например, такое распределение будет иметь размер листа на растущей во дворе березе.

Все эти совпадения не случайны. Одна из важнейших теорем теории вероятностей утверждает, что сумма большого числа независимых случайных величин при достаточно большом количестве слагаемых и не слишком сильно отличающихся дисперсиях имеет приблизительно нормальное распределение. Эта теорема объясняет нормальность распределения разнообразных количественных признаков в природе и технике: если признак зависит от большого количества независимых факторов, то о нем можно говорить как о результате *суммирования случайных вариаций* каждого из этих факторов, что и приводит к нормальному распределению. В частности, многие переменные в психологии также можно считать суммой независимых вкладов различных факторов.

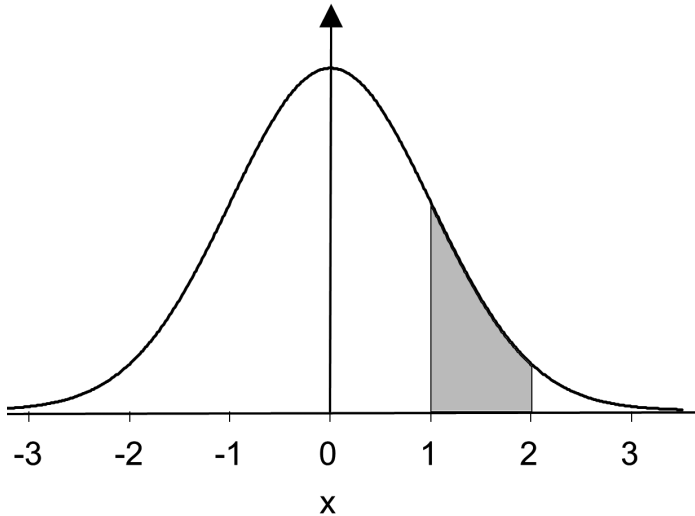


Рис.4.1.1(1). График плотности стандартного нормального распределения.

Мы начнем с так называемого *стандартного нормального распределения*. Его плотность выражается формулой

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

График этой функции изображен на рис. 4.1.1(1).

На рисунке выделена область под графиком плотности, ограниченная слева и справа значениями аргумента 1 и 2 соответственно. Площадь этой области и есть вероятность того, что испытание данной случайной величины приведет к появлению числа между единицей и двойкой. Вся площадь под графиком равна, разумеется, единице, поскольку этой площади соответствует вероятность того, что случайная величина примет какое-нибудь значение, т.е. единица.

Несмотря на то что значение $f(x)$ можно вычислить для любого x , фактически при $x < -4$ и $x > 4$ значения функции так малы, что для наших целей ими можно пренебрегать. Это значит, что результат испытания стандартной нормально распределенной случайной величины практически достоверно содержится в интервале от -4 до 4 (вероятность этого равна примерно 0.9998).

Достаточно очевидно, что математическое ожидание стандартной нормально распределенной случайной величины (которую мы далее будем обозначать $N(0, 1)$) равно нулю, поскольку график плотности распределения симметричен относительно нуля. Дисперсия

стандартной нормальной величины $N(0, 1)$ также простейшая из возможных — она равна единице.

4.1.2. Нормальное распределение в общем виде

Если подбрасывать монетку 100 раз, то, как мы говорили, количество выпадений гербов будет иметь приблизительно нормальное распределение (это утверждение называется теоремой Муавра-Лапласа). Математическое ожидание биномиальной случайной величины $B(100, 1/2)$, которая описывает данный процесс, равно 50, дисперсия равна 100-кратно увеличенной дисперсии³¹, рассчитанной для одного броска (равной $1/4$), т.е. равна 25, а корень из дисперсии (среднеквадратическое отклонение) равен 5. Нормальное распределение с такими параметрами задается формулой

$$f(x) = \frac{1}{\sqrt{2\pi} \cdot 5} e^{-\frac{(x-50)^2}{2 \cdot 5^2}}.$$

Число 5 здесь — это среднеквадратическое отклонение, а 50 — математическое ожидание. График этой функции симметричен относительно вертикальной прямой $x=50$. По сравнению со стандартным, график растянут в 5 раз вдоль оси абсцисс и во столько же раз сжат вдоль оси ординат.

Замечание 4.1.2(1). Формула становится понятней, если среднеквадратическое отклонение обозначить, как мы уже делали, буквой σ , а математическое ожидание буквой a :

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

График этой функции симметричен относительно вертикальной прямой $x = a$, растянут по сравнению со стандартным в σ раз³² и во столько же раз сжат сверху вниз.

Поскольку равномерные сжатие и растяжение сохраняют соотношения площадей частей, то мы получаем простой способ оценки вероятностей для близких к нормальному распределений, который мы можем использовать для решения большого числа разнообразных задач.

Без доказательств, которые желающие могут провести самостоятельно, приведем важные свойства нормально распределенных случайных величин.

³¹ По теоремам 3.2.5(3) и 3.2.5(5) математическое ожидание и дисперсия суммы независимых случайных величин равны сумме математических ожиданий и дисперсий слагаемых.

³² Если $\sigma < 1$, то «растяжение в σ раз» означает сжатие. Например, растяжение в 0.5 раза — это сжатие в 2 раза.

Свойство 4.1.2(2). Если Y нормально распределенная случайная величина, то $Y + C$ также имеет нормальное распределение (C — любая положительная или отрицательная константа). Математические ожидания этих величин связывает естественное соотношение

$$M_{(Y+C)} = M_Y + C.$$

Дисперсии при этом остаются неизменными:

$$D_{(Y+C)} = D_Y.$$

Свойство 4.1.2(3). Если Y нормально распределенная случайная величина, C положительная константа, то Y/C также имеет нормальное распределение (для отрицательных C это также верно, но нас такие случаи не интересуют). Дисперсии этих величин связывает соотношение

$$D_{\frac{Y}{C}} = \frac{D_Y}{C^2}$$

Если математическое ожидание Y было равно нулю, то Y/C так же будет иметь нулевое математическое ожидание.

На этих свойствах основаны методы приближенного расчета вероятностей путем сведения к стандартной нормальной величине.

4.1.3. Задачи на нормальное приближение распределений

Упражнение 4.1.3(1). Какова вероятность того, что при 100-кратном подбрасывании симметричной монеты количество выпавших гербов превысит 60?

Решение. Обозначим нашу случайную величину через Y . Как уже говорилось, эта случайная величина имеет близкое к нормальному распределение с математическим ожиданием 50, дисперсией, равной 25, и среднеквадратическим отклонением 5. Случайная величина $Y-50$ будет иметь приближенно нормальное распределение с нулевым математическим ожиданием, а случайная величина $\frac{Y-50}{5}$ к тому же будет иметь дисперсию, равную единице (при сохранении нулевого математического ожидания), т.е. будет иметь приближенно стандартное нормальное распределение. Количеству гербов больше 60 для исходного условия будет соответствовать событие «испытание стандартной нормально распределенной случайной величины N привело к появлению результата, превышающего $(60 - 50)/5 = 2$ ».

Ниже приведена таблица распределения стандартной нормальной величины. Напротив чисел от 0.1 до трех стоят вероятности того, что эта случайная величина примет значение от нуля до данного числа — значение функции $\Phi(x)$. Например, вероятность того, что $N(0, 1)$ примет значение между нулем и единицей равна 0.341.

Таблица 4.1.3(2). Распределение стандартной нормальной случайной величины

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0.1	0.040	1.1	0.364	2.1	0.482
0.2	0.079	1.2	0.385	2.2	0.486
0.3	0.118	1.3	0.403	2.3	0.489
0.4	0.155	1.4	0.419	2.4	0.492
0.5	0.192	1.5	0.433	2.5	0.494
0.6	0.226	1.6	0.445	2.6	0.495
0.7	0.258	1.7	0.455	2.7	0.497
0.8	0.288	1.8	0.464	2.8	0.497
0.9	0.316	1.9	0.471	2.9	0.498
1.0	0.341	2.0	0.477	3.0	0.499

В таблице дана вероятность попадания в промежуток $(0; 2)$, она равна 0.477 (для непрерывных распределений включение или не включение границ интервалов, естественно, не меняет вероятностей, поэтому мы не будем уточнять, замкнутый имеется в виду интервал или открытый). Поскольку вероятность попадания $N(0, 1)$ в интервал от нуля до бесконечности равна 0.5, то вероятность попадания в промежуток от 2 до бесконечности равна $0.5 - 0.477 = 0.023$ (рис. 4.1.3(3)а).

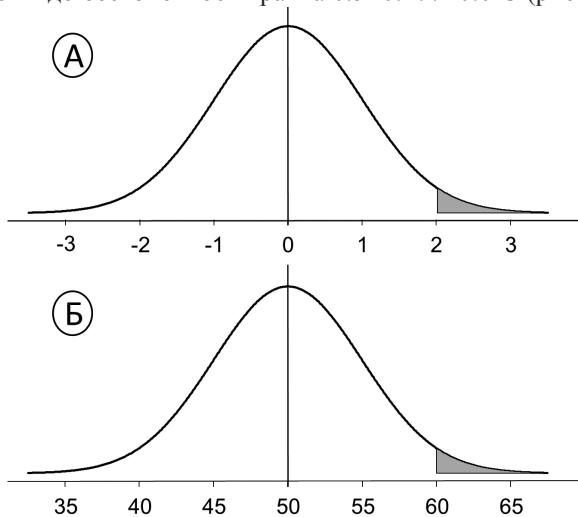


Рис. 4.1.3(3) Нормальное приближение частот биномиального распределения.

Пример 4.1.3(4). В исследовании проверяется гипотеза о том, что просмотр телевизионных новостей увеличивает тревожность. У 100 испытуемых измеряется уровень ситуативной тревожности перед и после просмотра новостей. Если в результате получено, что у 60 испытуемых уровень тревожности повысился после просмотра, то вряд ли этот результат можно считать случайным, поскольку тогда результат «60 испытуемых стали более тревожными» соответствует выпадению 60 гербов при 100 бросаниях симметричной монеты. При переходе к нормальному приближению вероятность того, что число гербов будет меньше 60, приравнивается приблизительно к вероятности попадания $N(0, 1)$ в интервал $(-\infty; 2)$, которая равна 0.023 (рис. 4.1.3(3б)).

Замечание 4.1.3(5) Вероятно, для читателя этот результат довольно неожиданный: обыденная интуиция говорит нам, что 65 из 100 не слишком удивительный результат. В таких случаях теория вероятностей нас уверенно поправляет.

Замечание 4.1.3(6). Когда мы используем нормальное приближение дискретного распределения, для полной точности надо вводить специальную поправку. Ведь для нормального распределения включение или не включение граничного значения интервал не меняет вероятности попадания в интервал, а для дискретного распределения события «больше 65» и «больше либо равно 65» имеют разные вероятности.

4.2. Нормальное распределение. Преобразования шкал и соответствующие преобразования плотностей

4.2.1. Изменение единиц измерения

Плотность нормального распределения в общем случае задается следующей формулой

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}.$$

Мы говорили в подпараграфе 4.1.2, что график этой функции симметричен относительно вертикальной прямой $x=a$, растянут по сравнению со стандартным в σ раз и во столько же раз сжат сверху вниз. Сейчас мы обсудим эти утверждения подробнее.

На графике рис. 4.1.3(4) видно, что математическое ожидание стандартной нормальной случайной величины является центром распределения и точкой максимума плотности. Предположим, что наша стандартная нормально распределенная случайная величина реализуется нами в игре в боулинг и отражает отклонение пущенного нами шара от центральной кегли, измеренное в дециметрах, причем откло-

нение влево будет отображаться отрицательными числами.

Упражнение 4.2.1(1). Какова вероятность того, что пущенный нами шар прокатится ближе, чем в дециметре от центральной кегли?

Решение. Эта вероятность равна площади заштрихованной фигуры на рис. 4.2.1(2). Правая половина этой фигуры (между нулем и единицей) имеет площадь, равную значению функции $\Phi(x)$ при $x = 1$. По таблице 4.1.3(2) находим, что $\Phi(1) = 0.341$, значит вероятность отклонения нашего шара меньше чем на дециметр равна $0.341 \cdot 2 = 0.682$.

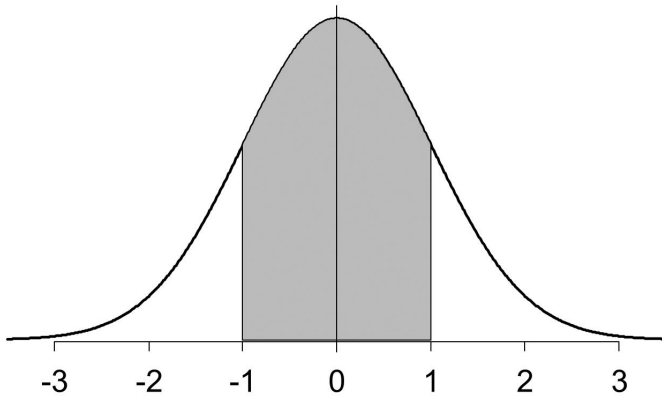


Рис. 4.2.1(2). Вероятность, рассчитываемая как площадь фигуры под графиком плотности распределения.

Упражнение 4.2.1(3). Что произойдет с распределением, если отклонение мы будем измерять в готских дюймах (1 готский дюйм равен 5 см или 0.5 дм)?

Замена единиц измерения есть допустимое преобразование интервальной шкалы. Все значения исходной шкалы следует в нашем случае умножить на 2 (отметим, что уменьшение единицы измерения в 2 раза приводит к увеличению во столько же раз шкальных значений). Если на рис. 4.2.1(4а), на котором изображено стандартное нормальное распределение, описывающее отклонение нашего боулингового шара, просто заменить подписанные под осью абсцисс числа на вдвое большие, мы получим первый вариант решения. Однако если потребовать, чтобы единицы длины на графиках всегда отображались равными отрезками, то график придется растянуть в 2 раза вдоль оси абсцисс (рис. 4.2.1(4б)). Это, однако, приведет к тому, что площадь под графиком станет больше в 2 раза, а это для плотности рас-

предела случайной величины недопустимо. Чтобы привести площадь к требуемой (равной единице) величине, просто сожмем график в 2 раза по оси ординат (рис. 4.2.1(4)в).

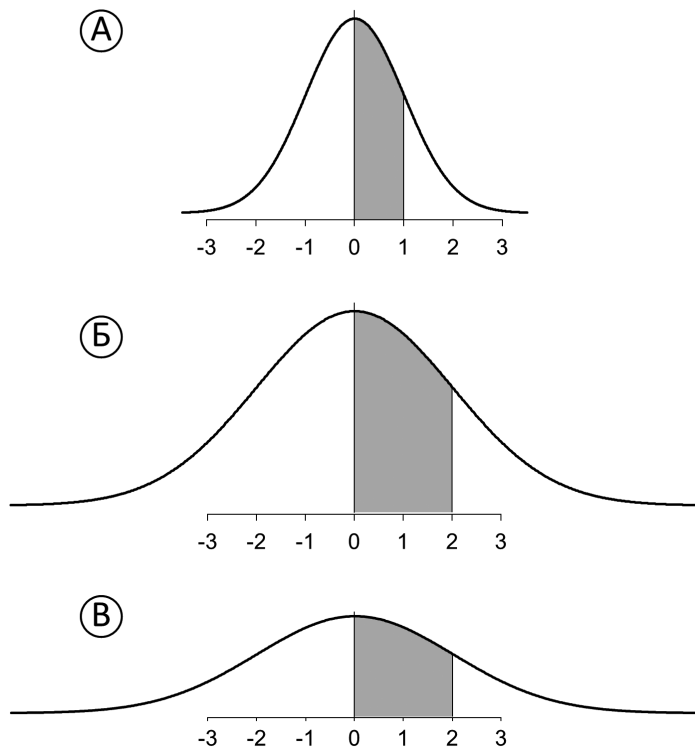


Рис. 4.2.1(4). Уменьшение вдвое единицы измерения: поведение плотности распределения.

Нас теперь интересует уравнение, описывающее функцию, график которой дан на рис. 4.2.1(4)в, но сначала найдем, что надо поменять в формуле

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},$$

чтобы получить уравнение функции, график которой дан на рис. 4.2.1(4)б и представляет собой растянутый в два раза по горизонтали первый график. В точке $x=2$ вторая функция принимает такое

значение, которое первая функция принимает при $x=1$ (это и есть суть операции растяжения в два раза). И более общо, при любом c в точке $x=2c$ вторая функция принимает такое значение, которое первая функция принимает при $x=c$. Именно такое соответствие значений обеспечивает формула

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x/2)^2}{2}}.$$

Действительно, подставив во вторую формулу $x=2c$, мы получим такое же выражение, которое будет получено подстановкой $x=c$ в первую формулу.

Замечание. Эти рассуждения выглядят проще в самом общем виде. Пусть дана функция $f(x)$. Будем обозначать через $f(x/2)$ функцию, задаваемую правилом: «сначала поделить аргумент на 2, затем применить функцию f ». Примерами таких функций могут служить пара $\sin(x)$, $\sin(x/2)$ или пара x^2 ; $(x/2)^2$. График функции $f(x/2)$ будет представлять собой результат растяжения в два раза графика функции $f(x)$. Доказательством является повторение вышеприведенного рассуждения: значения $f(x)$ при $x=c$ и $f(x/2)$ при $x=2c$ совпадают, а это и есть указанное растяжение.

Таким образом, график на рис. 4.2.1(4)б есть график функции

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x/2)^2}{2}}.$$

Совсем просто из этого следует, что график на рис. 4.2.1(4)в задается уравнением

$$f(x) = \frac{1}{2\sqrt{2\pi}} e^{-\frac{(x/2)^2}{2}}.$$

Аналогично разъясняется, что график функции $f(x/\sigma)$ есть в σ раз растянутый график функции $f(x)$, причем при «растянуть в σ раз» означает «сжать в $1/\sigma$ раз».

Это значит, что график функции

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x/\sigma)^2}{2}} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}.$$

есть растянутый в σ раз по горизонтали и одновременно сжатый в σ раз по вертикали график стандартного нормального распределения.

Отметим в дополнение, что заштрихованные на двух графиках рис. 4.2.1(4)а и 4.2.1(4)в фигуры имеют равные площади: вторая фигура получена из первой растягиванием в два раза по горизонтали с одновременным сжатием в два раза по вертикали.

4.2.2. Перенос начала координат

Перенос начала координат — это второе преобразование, которое нам понадобится. Предположим, мы решили отсчитывать расстояние не от центральной кегли, а от левой из тех, что расставлены на площадке. Допустим, она стоит левее центральной на 2 готских дюйма. Пересчитанные координаты подписаны под исходными на рис. 4.2.2(1) и выделены курсивом. Нас теперь интересует такая модификация формулы, чтобы она описывала график на рис. 4.2.2(1) в новой системе координат.

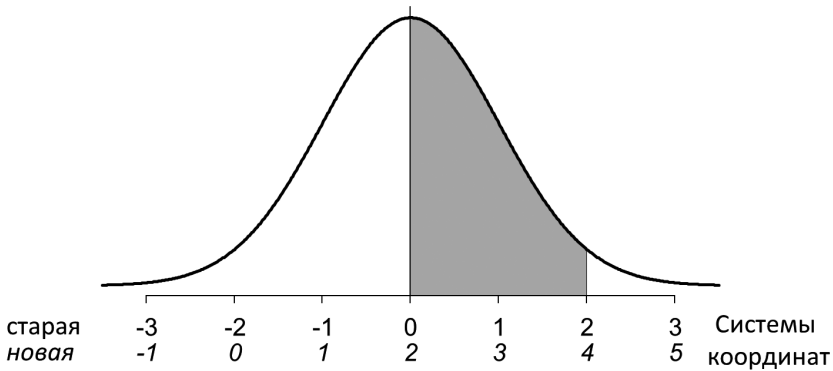


Рис. 4.2.3(1). Сдвиг начала координат

Эта задача еще легче. На графике сразу видно, что функция принимает одинаковые значения для s в старой системе координат и для $s+2$ в новой, выделенной курсивом. Если для второй системы мы запишем формулу в виде

$$f(x) == \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-2)^2}{2\sigma^2}},$$

то, глядя из начала этой новой системы координат, мы будем видеть график сдвинутым вправо на 2 единицы.

Аналогичным рассуждением мы можем получить, что при положительном a график функции

$$f(x) == \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

получается сдвигом исходного графика вправо на a единиц, а график функции

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x+a)^2}{2\sigma^2}}$$

получается сдвигом на столько же единиц влево.

Окончательно: график

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{((x-a)/\sigma)^2}{2}}$$

получается из графика стандартного нормального распределения растяжением по горизонтали в σ раз, сжатием во столько же раз по вертикали и сдвигом вправо на a единиц при положительном a и влево на $|a|$ единиц, если a отрицательно.

Пусть случайная величина X имеет плотность распределения, заданную предыдущей общей формулой. Обратное преобразование шкалы $x' = (x - a)/\sigma$, переведет график в график плотности стандартного нормального распределения (подставив x вместо $(x-a)$, мы сдвинем график влево, а заменив x/σ на x , сожмем его в σ раз, т.е. действительно проделаем обратные операции). Следовательно, операция над случайной величиной $X' = (X - a)/\sigma$ даст в результате стандартную нормально распределенную случайную величину, график плотности которой задается знакомой нам формулой.

4.3. Нормальное распределение. Практикум

Пример 4.3(1). Рассмотрим пример нормального распределения в психологическом исследовании. Допустим, мы измерили коэффициент интеллекта (по стандартной шкале IQ) у нормальных взрослых испытуемых. Модельные данные такого исследования представлены в файле *NormIQ.sav*. В единственной переменной IQ содержатся баллы IQ 150 испытуемых. Известно, что шкала IQ строилась так, что средний балл этого показателя равен 100, а стандартное отклонение — 15 баллам³³.

Сначала рассмотрим распределение полученных данных. Нарисуем гистограмму данных с помощью меню *Графика — Устаревшие диалоговые окна — Гистограмма (Graphics — Legacy Dialogs — Histogram)*. Полученный результат представлен на рис. 4.3(2).

³³ См. подробнее Neisser U. Rising scores on intelligence tests: Test scores are certainly going up all over the world, but whether intelligence itself has risen remains controversial // American scientist. — 1997. — Т. 85. — №. 5. — С. 440-447.

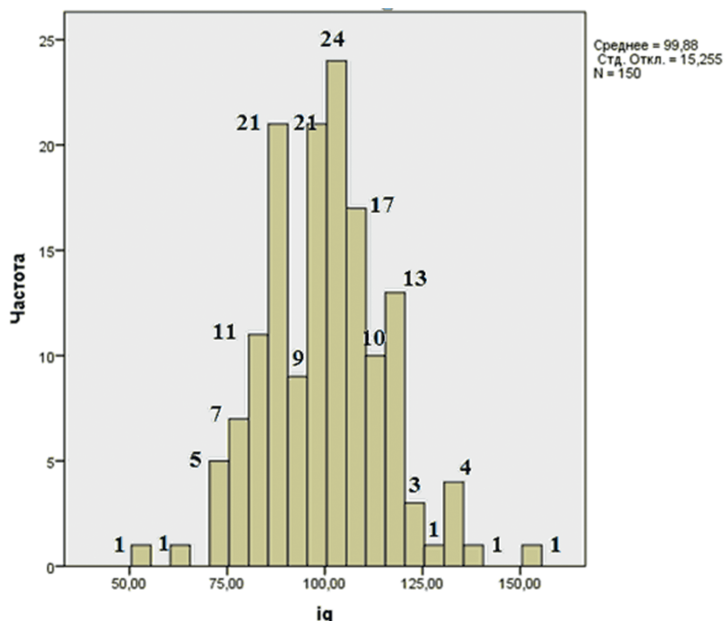


Рис. 4.3(2). Распределение баллов IQ у 150 испытуемых. Для удобства читателя мы написали частоты появления значений над столбиками гистограммы

Можно заметить, что выборочное среднее практически совпадает со средним по генеральной совокупности, равным 100, и стандартное отклонение также слабо отличается от 15. Будем, тем не менее, пользоваться выборочными оценками для упражнений.

Теперь мы сравним выборочные частоты отклонений с теоретически предсказанными нормальным распределением.

Можно видеть, что имеется семь наблюдений меньше или равных 75. Для оценки возьмем значение 75.5, лежащее в середине отрезка (75, 76). Отклонение от выборочного среднего, измеренное в стандартных отклонениях, равно $(75.5 - 98.9)/15.26 = -1.53$.

В таблице распределения стандартной нормальной величины (см. 4.1.3) против 1.5 стоит значение 0.433, числу 1.53 можно считать соответствующим значение 0.44. На долю больших, чем 1.53 значений случайной величины (а значит и меньших, чем -1.53) приходится доля 0.06 от выборки. Наша выборка дает 7 из 150 (см. гистограмму), что составляет 0.047 — довольно близкое к 0.06 значение.

Упражнение 4.3(3). Проведите аналогичные оценки для значений 125, 110, 90.

Пример 4.3(4). В подпараграфе 4.2.1. рассматривается пример случайной величины — отклонения брошенного шара для боулинга от центральной кегли, измеренного в дециметрах. В файле *Bowling.sav* в переменной *distance* содержатся модельные данные такой случайной величины (результаты 1000 бросков шара). Оценим по этим данным частоту наблюдения события «шар отклонится от центральной кегли больше, чем на 1 дециметр». Для этого в SPSS можно вычислить новую переменную, которая будет принимать значение 1 при условии, что расстояние меньше единицы. Для этого надо зайти в меню **Преобразование — Вычислить переменную (Transform — Compute Variable)**. В появившемся диалоговом окне зададим в поле **Целевая переменная (Target Variable)** имя новой переменной *less1*, а в поле **Цифровое выражение (Numeric Expression)** — значение переменной «1». Затем, в дополнительном окне **Если (If)** выберем пункт **Выбрать наблюдения, удовлетворяющие условию (Include if case satisfies condition)** и зададим условие $ABS(distance) \leq 1$ (абсолютное значение переменной *distance* меньше или равно нулю). После выполнения расчета в таблице данных появится новая переменная *less1*.

Упражнение 4.3(5). Рассчитайте количество единиц в этой переменной (с помощью знакомого нам расчета частот) и сравните с теоретической вероятностью, полученной в упражнении 4.2(4).

Пример 4.3(6). Этот пример призван проиллюстрировать сходимость к нормальному распределению сумм независимых случайных величин. В файле *CenLimTh* содержатся результаты испытаний пяти очень разных случайных величин (переменные *V1*, *V2*, *V3*, *V4* и *V5*).

Упражнение 4.3(7). Постройте гистограммы каждой из них. Вычислите их сумму $S = V1+V2+V3+V4+V5$ и построьте гистограмму суммы. Чтобы сравнить с нормально распределенной случайной величиной, присвойте переменной *SN* с помощью генератора случайных чисел, имеющих нормальное распределение $N(24, 5)$, случайные значения и построьте гистограмму. Сравните две гистограммы.

Глава 5. _____

Статистическое оценивание средних

5.1. Одновыборочный критерий Стьюдента и доверительный интервал для среднего

5.1.1. Стандартное отклонение как единица измерения индивидуального результата

Рост голландского футболиста Робина ван Перси 183 см. Для расчета высоты дверных проемов в строящейся квартире этого достаточно. Для ответа на вопрос, высок ли Робин или нет, у нас нет достаточной информации. Если мы узнаем, что средний рост мужчины-европейца в настоящее время равен 178 см, то мы сможем сказать, что рост ван Перси выше среднего. Если мы найдем информацию о том, что стандартное отклонение роста мужчин по всей генеральной совокупности мужчин-европейцев равно 5 см, то получим уже довольно ясное представление о месте ван Перси в распределении ростов. А именно, рассматривая это распределение как приблизительно нормальное, мы видим, что рост ван Перси отклоняется от среднего европейского на 5 см, т.е. ровно на одно стандартное отклонение. По таблице 4.1.3(2) единице отклонения в стандартном нормальном распределении соответствует вероятность 0.341. Это значит, что ниже ван Перси оказывается примерно европейского мужского населения, а выше него примерно 0.159 (15.9%) мужчин-европейцев. Оба показателя равно информативны: можно сказать, что Z -оценка (Z -score) роста ван Перси равна 1, можно сказать, что рост ван Перси отсекает от распределения верхний «хвост» весом в 0.159.

Однако среднее 178 и стандартное отклонение 5 в данном случае, как и практически всегда в подобных случаях, не были получены измерением всех европейцев, а были оценены по выборкам. Предположим, мы хотим оценить средний рост жителей какого-то достаточно большого города, например Томска. Мы набираем случайную выборку из взрослых мужчин-томичей объемом 100 человек. Поскольку мы не знаем не только средний рост томичей, но и стандартное отклонение распределения в генеральной совокупности, мы оцениваем оба показателя по выборке.

5.1.2. Стандартная ошибка как единица измерения точности оценки по выборке

Пусть $\bar{x}=177$, а оцененное по выборке стандартное отклонение $s=4$. Как всегда в таких случаях, мы можем быть уверены, что полученный результат $\bar{x}=177$ не совпадает в точности со средним ростом томичей, а лишь более или менее близок к нему. Опираясь на оценки среднего и дисперсии, которые дает наша выборка, мы построим *доверительный интервал* для среднего, т.е. определим промежуток возможных значений неизвестного среднего роста томичей, которые согласуются с нашей выборкой.

Определение 5.1.2(1). Для выборки объема n *стандартной ошибкой среднего* (обозначается se) называется выборочное стандартное отклонение, деленное на \sqrt{n} :

$$se = s/\sqrt{n}.$$

Деление на \sqrt{n} вытекает из теорем о дисперсии третьей главы. Здесь мы даем только алгоритм вычисления доверительного интервала, дальнейшие разъяснения в следующем параграфе.

Доверительный интервал среднего — это наглядное представление возможных значений оцениваемого настоящего математического ожидания, которые в определенной степени согласуются с полученным результатом. Степень согласованности выражается в процентах. Можно сказать так: ноль-процентный доверительный интервал — это в точности выборочное среднее, а стопроцентный — вся числовая ось, поскольку абсолютно достоверно мы можем утверждать лишь, что настоящее математическое ожидание лежит где-то на числовой оси. Обычно используются 95 и 99 процентов степени согласованности. Доверительный интервал легко построить, но точный смысл его не так прост. Мы обсудим это в параграфе 5.2.

Радиус 95-процентного доверительного интервала для среднего вычисляется умножением стандартной ошибки se , в нашем случае равной $4/\sqrt{100}$, на число 1.96, полученное из таблицы нормального распределения (интервал (-1.96; 1.96) вырезает 95% площади под кривой стандартного нормального распределения).

Доверительный интервал среднего по данной выборке получается откладыванием радиуса в обе стороны от выборочного среднего. В результате получим интервал радиуса $0.4 \cdot 1.96 = 0.78$ с центром в 177, т.е. (176.22; 177.78). Нам тем самым даны границы: если реальный средний рост томичей лежит в данном интервале, то в появлении результата 177 (при объеме выборки 100) нет ничего удивительного, результат 177 согласуется с любым из значений из этого интервала. Значения среднего роста томичей вне этого интервала менее правдоподобны,

и чем дальше от границ вовне, тем они согласуются с результатом 177 меньше.

Замечание 5.1.2(2). Мы рассчитывали доверительные границы по таблице нормального распределения. Здесь необходимо уточнить, что для выборок объема $n < 100$ (и чем меньше выборка, тем это существеннее) доверительные границы следует рассчитывать с помощью таблиц *распределения Стьюдента с $n-1$ степенью свободы*. По сравнению с нормальным распределением, распределение Стьюдента имеет несколько более тяжелые хвосты³⁴ и они тем тяжелее, чем меньше объем выборки, поэтому границы доверительных интервалов оказываются тем шире, по сравнению с вычисленными по нормальным таблицам, чем меньше n . Для $n=100$ различия распределений Стьюдента и нормального почти незаметны (рис. 5.1.3(2)).

5.1.3. Определения основных понятий. Одновыборочный Т-критерий (Стьюдента)³⁵

Мы уже фактически использовали квантили распределений, не называя их. Теперь мы дадим определение.

Определение 5.1.3(1). Пусть задано распределение некой случайной величины, и это распределение симметрично относительно нуля. *Двухсторонний квантиль уровня α* данного распределения это пара чисел $-k_\alpha$ (оно называется нижний двухсторонний квантиль) и k_α (оно называется верхний двухсторонний квантиль)³⁶, таких, что вероятность попадания данной случайной величины в интервал между нижним и верхним квантилями или, что одно и то же, площадь под кривой плотности распределения между нижним и верхним квантилями равна $1 - \alpha$.

Например, -1.96 и 1.96 — нижний и верхний двухсторонние квантили стандартного нормального распределения на уровне 0.05 . Такие квантили называют еще 5%-й. Для краткости будем говорить: 1.96 — 5%-й двухсторонний квантиль нормального распределения, подразумевая, что это верхний двухсторонний квантиль, а нижний имеет симметричное отрицательное значение.

³⁴ «Более тяжелые хвосты» — статистический термин, означающий, что вероятности больших отклонений от нуля для распределения Стьюдента больше, чем для стандартного нормального распределения. Подробнее в параграфе 5.2.

³⁵ Буква Т в названии критерия в литературе встречается как заглавная, так и строчная. Более того, в русской версии SPSS употребляется строчная буква, а в английской — заглавная. Мы всюду, кроме отсылок к пунктам меню русской версии, употребляем заглавную букву в названии Т-критерия и строчную в сочетании «t-статистика».

³⁶ В литературе слово «квантиль» иногда относят к женскому роду. Нам не удалось найти авторитетных указаний на этот счет.

Как мы уже говорили, у распределения Стьюдента более тяжелые хвосты, поэтому для выборки объема 21 двухсторонний 5%-й квантиль для распределения Стьюдента с 20 степенями свободы³⁷ это 2.08, а для 3 степеней свободы (для выборки объема 4) — 3.18 (рис. 5.1.3(2)). При уменьшении числа степеней свободы графики распределения Стьюдента «уплощаются», а квантили монотонно удаляются от начала координат.

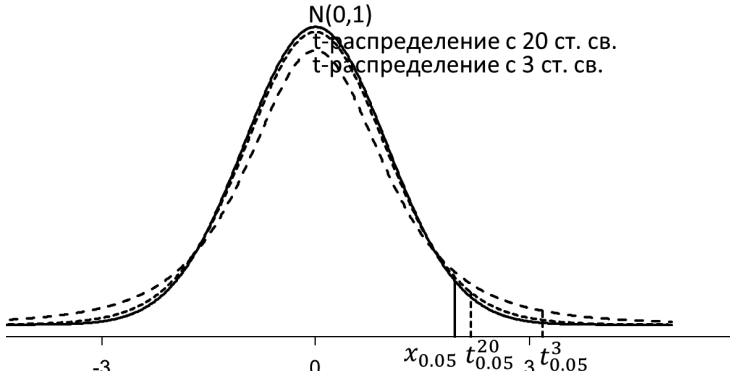


Рис. 5.1.3(2). Графики стандартного нормального распределения и t-распределений (Стьюдента) с 20 и 3 степенями свободы; $x_{0.05} = 1.96$ — верхний двухсторонний 5%-й квантиль нормального распределения, $t_{0.05}^{20} = 2.08$ — верхний двухсторонний 5%-й квантиль t-распределения с 20 степенями свободы, $t_{0.05}^3 = 3.18$ — верхний двухсторонний 5%-й квантиль t-распределения с 3 степенями свободы.

Для построения 95%-ых и 99%-ых доверительных интервалов мы берем соответствующие квантили соответствующих распределений.

Можно решить обратную задачу. Если мы имеем ориентированный на данное распределение результат (пока подразумеваем распределение Стьюдента, совпадающее в пределе с нормальным, далее мы рассмотрим и другие полезные распределения), то мы можем оценить этот результат удвоенным весом отсекаемого им хвоста распределения. Вернемся к первому примеру предыдущего пункта. Если мы хотим оценить соответствие гипотезы «средний рост жителей Томска 178 сантиметров» полученному выборочному среднему 177 см по описанной выше выборке объема 100, то следует поделить абсолютную величину³⁸ раз-

³⁷ Пока будем понимать степени свободы чисто рецептурно — разъяснения будут даны в главе 8 и подробнее в приложении 4.

³⁸ В некоторых случаях можно учитывать знак разности и применять односторонние критерии, связанные с односторонними хвостами распределения. Мы обсудим этот вопрос в главе 7.

ности 178 – 177 на стандартную ошибку, равную 0.4. Получив результат 2.5, мы можем оценить по таблице нормального распределения вес верхнего хвоста — по таблице это , вес двух симметричных хвостов 0.012.

Определение 5.1.3(3). Так рассчитанный вес отсекаемых данным результатом хвостов называется *значимостью* этого результата относительно данной гипотезы. Можно также называть это число значимостью гипотезы относительно результата или значимостью различия полученного результата и гипотезы. Чем *меньше* вес отсекаемых хвостов, тем *больше* у нас оснований считать не случайным отклонение полученного результата от значения, заданного гипотезой.

Еще раз вернемся к примеру подпараграфа 5.1.2. Каждая гипотеза оценивается своей значимостью по отношению к полученному нами результату (среднее 177 при стандартном отклонении 4 на выборке 100 человек). Для гипотезы «средний рост томичей равен 176.4» разность, измеренная стандартной ошибкой 0.4, дает результат $(176.4 - 177) / 0.4 = -1.5$. Взяв модуль результата 1.5, рассмотрим верхний хвост (рис. 5.1.3(4)), площадь которого определим по таблице 4.1.3(2) стандартного нормального распределения. Она равна $0.5 - 0.433 = 0.067$. Удвоив, получаем 0.134. Это значимость, соответствующая гипотезе «средний рост томичей равен 176.4» по отношению к результату 177 со стандартным отклонением 4 по выборке объема 100.

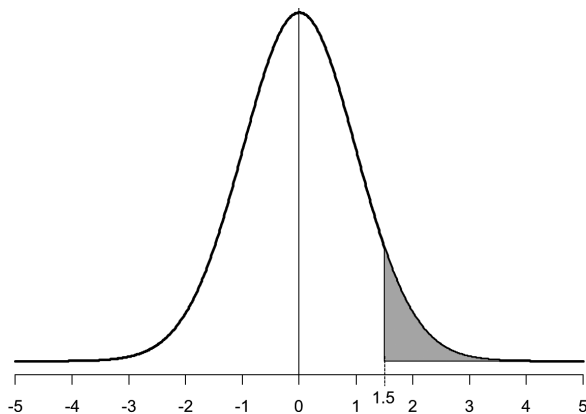


Рис. 5.1.3(4). Оценка значимости результата «среднее выборочное равно 177» по выборке объема 100 со стандартным отклонением 4 относительно гипотезы «среднее по генеральной совокупности равно 176.4». Статистика $t = (177 - 176.4) / 0.4 = 1.5$. Хвост, отсекаемый значением 1.5 под стандартным нормальным распределением, имеет площадь 0.067. Значимость равна $0.067 * 2 = 0.134$.

95%-й доверительный интервал накрывает в нашем случае гипотезы, значимость которых по отношению к этому результату выражается числами, превосходящими $1 - 0.95 = 0.05$.

Посмотрим теперь, что произойдет, если объем выборки увеличить до 400. Предположим, что среднее выборочное и стандартное отклонение остались теми же (в реальности они, конечно, изменились бы, но, скорее всего, не существенно). Новые вычисления дали бы другое значение \sqrt{n} , теперь $4/\sqrt{100} = 0.2$, и радиус доверительного интервала становится вдвое меньше. Гипотезе «средний рост томичей равен 178» соответствует вдвое большее, чем по выборке объема 100, Z-значение отклонения, оно теперь равно 5. Значимость (вес соответствующих хвостов) для отклонения 5 меньше 0.000001 (в нашем распоряжении нет таблиц, которые бы точно оценили вероятности для таких больших отклонений, но эти вероятности легко рассчитать с помощью специализированных программ или на общедоступных сайтах).

Техника построения доверительных интервалов и оценки значимости, которую мы здесь разобрали, называется T-критерием или критерием Стьюдента. Итак, в общем случае для выборки объема n в задачах, подобных разобранной выше, для построения доверительных интервалов и оценки значимости сначала вычисляется стандартная ошибка se . Затем берется двухсторонний квантиль (5%-й для 95%-го доверительного интервала, 1%-й для 99%-го) распределения Стьюдента $n-1$ с степенями свободы (вручную это можно сделать с помощью таблиц, SPSS вычисляет его самостоятельно). Радиус R соответствующего доверительного интервала получается умножением квантиля на стандартную ошибку. Сам доверительный интервал — это промежуток $(\bar{x} - R; \bar{x} + R)$. Он дает представление о диапазоне согласующихся с полученной выборкой на соответствующем уровне доверия (95, 99%-м или ином) вариантов гипотетического среднего в генеральной совокупности.

Операция оценки значимости такова: разность выборочного среднего и проверяемого гипотетического среднего $\bar{x} - a$ делится на стандартную ошибку. Результат $t = (\bar{x} - a)/se$ называется значением t-статистики. Значимость определяется так: абсолютная величина t откладывается по оси абсцисс графика распределения Стьюдента с соответствующим числом степеней свободы. Затем вычисляется площадь верхнего хвоста распределения, отсекаемого этим значением. Удвоенная площадь хвоста и объявляется значимостью результата относительно данной гипотезы.

Приведем в заключение рецепт вычисления t-статистики в виде формулы:

$$t = \frac{\bar{x} - a}{se} = \frac{\bar{x} - a}{\sqrt{\frac{1}{n-1} \sum (x - \bar{x})^2}} \cdot \sqrt{n}.$$

Знак суммы здесь и везде далее подразумевает суммирование по всем элементам выборки.

5.1.4. Значимость 0.05 как социально закрепленная граница принятия решения в психологии

В исследовательской практике и в обычаях принятия результатов к публикации значимость 0.05 служит границей принятия решения. Если гипотеза имеет значимость по отношению к данной выборке меньше или равную 0.05, то она считается отвергнутой как маловероятная при полученных результатах, если больше 0.05, то не отвергнутой.

В значительном числе исследовательских ситуаций одна из гипотез выделена как наиболее важная опорная точка. Например, в исследовании подпараграфа 5.1.2 нас интересовало, меньше или не меньше рост томичей, чем рост европейцев. Гипотеза «Средний рост томичей равен среднему росту европейцев» выделяется среди других, обозначается H_0 и называется нулевой гипотезой или нуль-гипотезой.

Наиболее часто исследование, связанное со сравнением средних, строится именно так: формулируется нулевая гипотеза (о равенстве среднего какому-то числу или равенстве двух средних), формулируется альтернативная гипотеза (одно среднее больше другого и т.п. или просто «средние не равны») и предполагается, что в случае, если значимость нулевой гипотезы по отношению к полученной выборке меньше 0.05 (это значит, что разность средних достаточно велика — она превосходит приблизительно две стандартные ошибки), будет принята альтернатива, если больше 0.05 (разность средних невелика) — будет принята нулевая гипотеза о равенстве. Обычно интересным научным результатом оказывается как раз альтернативная гипотеза «голландцы выше средних европейцев», или «тренинг оптимизма в среднем увеличивает оптимизм участников», или «опосредствование запоминания слов карточками увеличивает количество воспроизведенных слов», и результат в таком случае формулируется, например, так: принимается (альтернативная) гипотеза на уровне значимости³⁹ 0.019. Это означает только, что значимость нулевой гипотезы («голландцы не выше средних европейцев», или «тренинг

³⁹ Правильнее говорить: «принимается альтернативная гипотеза, значимость равна...». Слова «уровень значимости» исходно относились к заранее, еще до проведения эксперимента (или иного измерения) выбранному критерию.

оптимизма в среднем не увеличивает оптимизм участников», или «опосредствование запоминания слов карточками не увеличивает количество воспроизведенных слов») по отношению к выборке равна 0.019, или, что почти то же самое, 98%-й доверительный интервал не включает нулевое значение, а 99%-й уже включает.

Если задача стоит так, что мы *должны* принять *одну* из гипотез, то тогда необходимо заранее договориться о том, какая значимость гипотезы H_0 по отношению к полученной выборке будет говорить в ее пользу. Как мы говорили, в настоящее время принято считать такой границей 0.05. Большинство журналов в России⁴⁰ принимает статьи, в которых значимость нулевой гипотезы (H_0) меньше 0.05, и она отвергается в пользу интересующей автора альтернативной гипотезы H_1 . Если мы принимаем такое условие выбора (равно как и любое другое), то нам грозят естественные ошибки: отвергнуть H_0 , когда она верна, и принять ее, когда она не верна.

Определение 5.1.4(1). *Ошибкой первого рода* называется отвержение H_0 , когда она верна. *Ошибкой второго рода* называется принятие H_0 , когда она не верна.

Например, наши стандартные 0.05 оказываются при таком правиле принятия решения *вероятностью ошибки первого рода*. Для единственного эксперимента это означает следующее: мы выбираем такой уровень значимости (0.05), что вероятность отвергнуть нуль-гипотезу, если она верна, при проведении единичного эксперимента невелика (0.05).

Однако в массовых исследованиях ситуация меняется: *истинная* гипотеза H_0 может быть в конце концов отвергнута, если исследователи проявят достаточное упорство. Например, гипотеза о том, что телепатии не существует, если ее примутся проверять 100 исследовательских коллективов, в которых заведомо нет телепатов, будет тем не менее отвергнута по принятому правилу примерно пятью из этих ста групп. Если остальные 95 никак не сообщат о своих результатах, то в журналах могут появиться пять статей со статистическими «доказательствами». В любом случае такие результаты науке полезны не будут.

Подобный сбой метода оценивания называется *публикационный сдвиг* (*publication bias* в англоязычной литературе, где термин был введен).

Современная тенденция решения проблемы публикационного сдвига такова:

⁴⁰ Ситуация в мире начинает меняться.

1. В том или ином виде информация о неудачных экспериментах (т.е. тех, в которых нуль-гипотеза не была отвергнута) должна быть доступна сообществу.

2. Одно исследование, как правило, ничего не доказывает. Не обязательно объявлять гипотезу отвергнутой или принятой — вполне можно ограничиться более мягкими формулировками, оставив сообществу развитие темы.

3. Правильнее предоставить максимально полную информацию об исследовании, которая даст возможность полностью корректно сопоставлять все исследования в данной области — удачные и неудачные. Такое сопоставление называется метаанализом и является в настоящее время очень важным инструментом исследований.

В исследованиях, связанных с оценками средних значений, для возможности мета-анализа достаточно публиковать выборочные средние, объемы выборок и стандартные отклонения. В настоящее время в различных сферах исследований создаются базы данных, где такие результаты публикуются вне зависимости от того, больше или меньше 0.05 оказываются значимости гипотез. Мы полностью на стороне этой практики.

Обратим в заключение внимание читателя на то, что статистика может предъявить *практически* надежное доказательство. Если значимость нулевой гипотезы равна 0.00000001, то авторы учебника готовы считать альтернативную гипотезу доказанной (в той степени, в какой в науке вообще можно что-нибудь доказать) — надо только убедиться, что экспериментальные процедуры совершенно свободны от ошибок. Нередко метаанализ доводит результаты до такой степени убедительности.

5.1.5. Статистическая значимость — не единственное, что нам нужно от результата. Другие виды значимости

Статистическая значимость результата не означает, что он содержательно значим — значим для психологической теории или практики. Например, важна ли для биолога (антропометриста) средняя разница в росте между томичами и европейцами в 0.1 см (177.9 против 178 см), даже если на наших, довольно больших, выборках она достигает уровня статистической значимости? Следует ли считать тренинг оптимизма эффективным с психологической точки зрения, если в среднем на выборке в 100 человек он привел к увеличению оптимизма в среднем на 5 баллов (по некоторой шкале), но у двух человек из этих 100 — к его ухудшению (нарастанию пессимизма) на 30 — 40 баллов (в той же шкале)? С точки зрения психолога ухудшение (хотя это и исключение) может иметь больший содер-

жательный смысл, нежели небольшое улучшение в среднем. Делая психологический вывод из исследования на основе статистической значимости, важно дополнять его другими, нестатистическими мерами оценки значимости результатов. То, в какой степени результаты соответствуют (не противоречат) теории (или одной из теорий) или важны для решения научной проблемы, называют *содержательной значимостью*. Так, доказанная разница в росте в 0.1 см содержательной значимостью для практика вряд ли обладает. Наоборот, если задача исследователя — сравнить в эксперименте предсказания двух альтернативных теорий, результаты, согласующиеся с одной из них и противоречащие другой, даже если они не подтверждаются статистикой вполне убедительно, — важная пища для размышлений. Возможно, продолжив эксперимент или видоизменив его, психолог сможет достичь более сильного статистического подтверждения своих предположений.

Важность результатов для какой-либо области практики называют *практической значимостью* (в медицине и клинической психологии ее обозначают как клиническую значимость). Например, если по итогам специальной программы обучения детей коррекционного класса несколько из них показали результаты, которые позволили им учиться в общеобразовательной школе, эффект программы может быть значим практически, даже если не достигает статистической значимости. Возможно, психолог-практик испробует свою программу на других школьниках или в других условиях или задастся вопросом, чем отличаются эти дети от тех, кому программа не помогла, — в обоих случаях сопоставление статистической и практической значимости полезно для науки.

Значимость, связанную с включением результата в контекст научной жизни, мы бы назвали *метааналитической*. Напомним пример про телепатию: насколько мы были бы готовы опубликовать статистически значимые результаты в пользу телепатии, если знаем, что 20 предыдущих похожих исследований не дали никаких интересных результатов? Другой пример: изменится ли наша интерпретация некоторого результата, значимость которого относительно нуль-гипотезы 0.08, если мы знаем, что из десяти похожих работ в семи была получена значимость менее 0.05, а в остальных трех — менее 0.10? Еще раз обратим внимание, что в методологии проверки нулевой гипотезы речь идет о единичном исследовании, но результаты нескольких исследований могут быть далее сопоставлены содержательно (теоретически) или при помощи статистики (метаанализ).

5.2. Одновыборочная статистика Стьюдента и доверительные интервалы

5.2.1. Среднее выборочное как случайная величина

Напомним, что для случайной величины, заданной таблицей⁴¹

X	a_1	a_2	...	a_n
	p_1	p_2	...	p_n

математическое ожидание и дисперсия задаются формулами:

$$M_X = a_1 p_1 + a_2 p_2 + \dots + a_n p_n$$

$$D_X = (a_1 - M_X)^2 p_1 + (a_2 - M_X)^2 p_2 + \dots + (a_n - M_X)^2 p_n.$$

Напомним также формулы выборочного среднего и выборочной дисперсии:

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \dots + x_n),$$

$$D_x = \frac{1}{n-1}((x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2).$$

В предыдущем параграфе мы говорили, что стандартная ошибка среднего для X , которой измеряется точность оценок, есть стандартное отклонение s , деленное на \sqrt{n} .

Теперь мы разъясним смысл сказанного.

Когда мы планируем оценить средний рост томичей по выборке объема 100, мы заранее заготавливаем форму обработки выборки, которую еще предстоит получить. Можно сказать, что мы предполагаем совершить испытание случайной величины

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_{100}}{100},$$

которая представляет собой среднее арифметическое 100 экземпляров случайной величины X : «рост случайного взрослого мужчины-жителя Томска» (обратите внимание, что формула составлена из заглавных букв, которыми мы обозначаем случайные величины). По теореме о сложении дисперсий для суммы независимых случайных величин (см. подпараграф 3.2.5 и приложение 2), дисперсия числителя в правой части в 100 раз больше дисперсии X . Легко также увидеть непосредственно из вида формулы математического ожидания и дисперсии случайной величины, что для любой случайной величины Y/n (в том

⁴¹ В таблице мы обозначили *возможные* значения случайной величины буквами a , чтобы отличать их от *результатов испытаний* этой случайной величины (выборочных значений), по-прежнему обозначаемых буквами x .

числе и для стократной суммы одинаковых величин, которая помещена в числителе последней формулы)

$$M_{Y/n} = \frac{1}{n} M_Y$$

и

$$D_{Y/n} = \frac{1}{n^2} D_Y.$$

Первая формула достаточно очевидна. Используя ее, выводим вторую:

$$\begin{aligned} D_{Y/n} &= \frac{1}{n-1} [(a_1/n - M_Y/n)^2 p_1 + (a_2/n - M_Y/n)^2 p_2 + \dots + (a_n/n - M_Y/n)^2 p_n] = \\ &= \frac{1}{n-1} \left[\left(\frac{1}{n}(a_1 - M_Y)\right)^2 p_1 + \left(\frac{1}{n}(a_2 - M_Y)\right)^2 p_2 + \dots + \left(\frac{1}{n}(a_n - M_Y)\right)^2 p_n \right] = \\ &= \frac{1}{n^2} \frac{1}{n-1} [(a_1 - M_Y)^2 p_1 + (a_2 - M_Y)^2 p_2 + \dots + (a_n - M_Y)^2 p_n] = \frac{1}{n^2} D_Y \end{aligned}$$

В таком случае $M_{\bar{X}} = M_X$ и $D_{\bar{X}} = \frac{1}{n} D_X$, а следовательно, среднеквадратическое отклонение делится на корень из n : $\sigma_{\bar{X}} = \sigma_X/\sqrt{n}$, т.е. среднеквадратическое отклонение среднего по n испытаниям случайной величины в \sqrt{n} раз меньше, чем среднеквадратическое отклонение данной случайной величины. Мы можем сказать теперь, что стандартная ошибка среднего оценивает среднеквадратическое отклонение случайной величины \bar{X} .

Поскольку для любого нормального распределения с математическим ожиданием M и среднеквадратическим отклонением α вероятность попадания в интервал $(M - \alpha, M + \alpha)$ постоянна (как и вероятности попадания в интервалы с радиусом, равным любому другому фиксированному кратному среднеквадратического отклонения), то среднее арифметическое при увеличении выборки стягивается сужающейся рамкой $(M - \sigma/\sqrt{n}, M + \sigma/\sqrt{n})$ к реальному математическому ожиданию M .

5.2.2. Доверительный интервал для среднего и значимость. Точные формулы

Пусть для некоторой выборки объема n мы получили $\bar{x} = a$ и стандартное отклонение, равное s . Чтобы вычислить радиус доверительного интервала, надо сначала найти соответствующий квантиль распределения Стьюдента с $(n-1)$ степенями свободы t_{α}^{n-1} ($\alpha = 1 - 0.95 = 0.05$ (для 95%-го интервала, аналогично для интервалов других «номиналов»). Затем квантиль надо умножить на стандартную ошибку.

Итак, доверительный интервал уровня $1 - \alpha$ для среднего по выборке объема n это интервал $(\bar{x} - R_{\alpha}, \bar{x} + R_{\alpha})$, где

$$R_{\alpha} = t_{\alpha}^{n-1} \cdot se = t_{\alpha}^{n-1} \cdot \frac{\sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2}}{\sqrt{n}}$$

Для оценки значимости гипотезы «истинное среднее равно a » необходимо рассчитать t -статистику по формуле

$$t = \frac{\bar{x} - a}{se} = \frac{\bar{x} - a}{\sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2}} \cdot \sqrt{n}$$

и оценить вес отсекаемого полученным результатом хвоста распределения Стьюдента с $n-1$ степенями свободы (и затем его удвоить, чтобы получить двухстороннюю значимость)⁴².

Упражнение 5.2.2(1). Убедиться, что если $\bar{x} - a = R_{\alpha}$, то значимость соответствующей гипотезы «истинное среднее равно a » равна в точности α .

5.2.3. Доверительный интервал для среднего. Точный смысл

Возьмем выборку объема 121 испытаний нормально распределенной случайной величины с математическим ожиданием 177 и среднеквадратическим отклонением 11. 5-процентный квантиль распределения Стьюдента для 120 степеней свободы равен 1.98 (почти не отличаясь от квантиля нормального распределения, как мы говорили). Стандартная ошибка для получаемых случайных выборок будет колебаться около $11/\sqrt{121}$, т.е. вокруг единицы. Таким образом, радиус доверительного интервала будет колебаться около 1.98. На рис. 5.2.3(1) показаны результаты 100 испытаний такой процедуры: порождалась выборка объема 121, оценивалось стандартное отклонение, вычислялся доверительный интервал и он изображался вертикальным отрезком. Жирными линиями выделены те экземпляры доверительных интервалов, которые не накрыли математическое ожидание 177.

В литературе встречаются следующие интерпретации доверительного интервала (не все они правильные):

1. 95%-й доверительный интервал для среднего с вероятностью 0.95 покрывает истинное математическое ожидание. 2. Истинное математическое ожидание с вероятностью 0.95 содержится в доверительном интервале.

3. 95% средних значений новых выборок попадут в полученный нами интервал.

Истинность первого утверждения мы можем проиллюстрировать нашим примером на рис. 5.2.3(1). Действительно, если математическое ожидание *нам известно*, то 95%-й доверительный интервал будет накрывать его в среднем в 95% случаев.

⁴² Если в последней формуле заменить строчные буквы x заглавными, то полученная формула задаст случайную величину, подобно тому, как в предыдущем параграфе рассматривалась как случайная величина. Напомним: при большом объеме выборки эта случайная величина имеет стандартное нормальное распределение. Если выборка невелика, то распределение «расплывается» (как изображено на рис. 5.1.3(2)) и квантили надо определять по соответствующим распределениям Стьюдента.

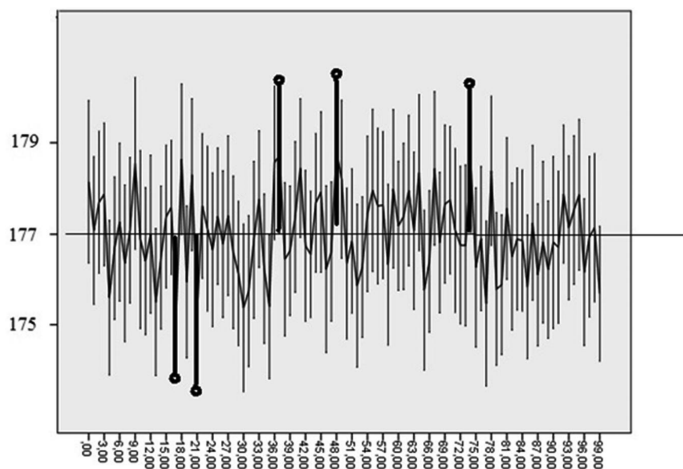


Рис. 5.2.3(1). Результат моделирования последовательного порождения выборок объема 121 испытаний нормальной случайной величины с математическим ожиданием 177 и среднеквадратическим отклонением 11. Построено 100 доверительных интервалов. Жирной линией выделены те, которые не накрывают математическое ожидание 177

Вторая формулировка тем не менее не истинна: если мы получили какой-то доверительный интервал, то ничего о вероятности попадания истинного математического ожидания в этот интервал мы сказать не можем. Прежде всего, потому что истинное математическое ожидание не является случайной величиной⁴³. Далее, если вы знаете истинное математическое ожидание роста случайно выбранного жителя Томска, а я оцениваю его по выборке, то вы всегда *знаете*, получил ли я «счастливый» или «несчастливый» интервал, а я не могу этого знать. В «несчастливом» случае на мое заявление «истинное математическое ожидание с вероятностью 0.95 содержится в (полученном мной) доверительном интервале» вы со стопроцентной уверенностью скажете, что оно там не содержится и никогда не будет содержаться.

Третья формулировка ложна в силу тех же обстоятельств. «Несчастливый» интервал накрывает меньше даже той половины оси, в которой он размещился. Вероятность попадания среднего в другую половину оси равна 0.5. Несчастливый интервал не обещает даже 50-процентной вероятности попадания в него будущих средних выборочных.

⁴³ Только в искусственных ситуациях можно *сделать* истинное математическое ожидание случайной величиной, но это потребует задать его априорное распределение.

Итак, верно лишь первое утверждение — в том смысле, который прояснен на рис. 5.2.3(1): известное кому-то (в худшем случае, одному богу) математическое ожидание покрывается доверительным интервалом с соответствующей вероятностью.

Следующая формулировка еще точнее: если мы собираемся оценить истинное среднее по выборке, то вероятность того, что 95%-й доверительный интервал, который мы по выборке построим, некроет истинное среднее, равна 0.05. Здесь важно, что вероятность вообще касается только тех событий, которые еще не произошли.

5.3. Практикум. Доверительный интервал

Пример 5.3(1). Рассмотрим простой пример, демонстрирующий, как можно рассчитать доверительный интервал для среднего в SPSS. В файле *HeightTomsk.sav* содержатся модельные данные измерения роста у 100 жителей города Томск, измеренного в сантиметрах (см. пример в подпараграфе 5.1.1). Рассчитать точечную оценку среднего, доверительный интервал и значимости относительно данной гипотезы в SPSS можно, используя Т-критерий для одной выборки. Для расчета нужно вызвать диалоговое окно — *Анализ* — *Сравнение средних* — *Одновыборочный t-критерий (Analyze — Compare Means — One-Sample T-test)*, перенесем интересующую нас переменную в окно «*Проверяемые переменные (Test Variable(s))*». Результаты такого анализа приведены в таблицах 5.3(2). и 5.3(3).

Таблица 5.3(2). Статистики для одновыборочного t-критерия

Статистики для одновыборочного t-критерия

	N	Среднее	Стд. отклонение	Стд. ошибка среднего
height	100	177.0050	4.00157	0.40016

Таблица 5.3(3). Одновыборочный t-критерий (сравнение с 0)

Одновыборочный t-критерий

	Проверяемое значение = 0					
	t	ст. св.	Зн. (2-сторонняя)	Разность средних	95% доверительный интервал разности средних	
					Нижняя граница	Верхняя граница
height	442	99	0.00	177.005	176.2	177.799

В первой из них мы находим среднее значение роста, стандартное отклонение и стандартную ошибку, во второй — доверительный интервал для среднего (по умолчанию 95-процентный, но можно заменить его «номинал» на любой другой во вкладке *Параметры (Options)*).

В столбце *Значимость (Sig)* мы находим в этой таблице значимость результата по отношению к гипотезе «средний рост равен нулю» — еще раз откройте вкладку и найдите поле с маркировкой *Проверяемое значение (Test Value)*. Если мы хотим узнать значимость другой гипотезы, нуль надо заменить. Поставим в поле число 178, соответствующее обсуждавшейся в подпараграфе 5.1.3 гипотезе о среднем росте в популяции. В таблице 5.3(4) показана значимость нашей гипотезы, несколько отличная от рассчитанной в подпараграфе 5.1.3. В таблице результат более точный, поскольку использовалось распределение Стьюдента, а не нормальное, как делали мы в указанном подпараграфе.

К сожалению, доверительный интервал также поменял свое представление (но не существо). Столбцы подписаны *Доверительный интервал разности средних (Confidence Interval of the Difference)*, если мы прибавим к 178 обозначенные в качестве нижней границы — 1.79, мы и получим 176.21 таблицы 5.3(3). Аналогично для верхней границы.

Таким образом, таблица 5.3(3) лучше представляет доверительный интервал среднего значения переменной, а таблица 5.3.3 — значимость и саму t-статистику. Можно выводить обе таблицы или пересчитывать доверительный интервал используя результаты таблицы 5.3(4).

Таблица 5.3(4). Одновыборочный t-критерий (сравнение с 178)

Одновыборочный t-критерий

	Проверяемое значение = 178					
	Т	ст.св.	Значимость (2-сторонняя)	Разность средних	95% доверительный интервал разности средних	
					Нижняя граница	Верхняя граница
height	-2.48	99	.015	-.995	-1.7890	-.2010

Упражнение 5.3(5). Предположите, как изменится доверительный интервал при увеличении степени доверия до 99% и при уменьшении до 90%. В каком случае интервал будет более узким, а в каком — широким? Проведите расчеты и проверьте свои соображения.

Пример 5.3(6). В подпараграфе 5.1.3 обсуждался смысл доверительного интервала при сравнении выборочного среднего с каким-ли-

бо теоретическим значением. Допустим, что мы хотим проверить предположение, что интеллект поступивших на психологический факультет первокурсников отличается от среднего по стране⁴⁴. Для этого мы измерили с помощью стандартного теста IQ интеллект 30 случайно выбранных студентов.

Упражнение 5.3(7). Модельные данные содержатся в файле *IqPsy.sav*. Известно, что средний балл IQ в генеральной совокупности составляет 100 баллов.

Проведите расчет одновыборочного t-критерия. Найдите значимость результата относительно гипотезы о том, что средний балл IQ у психологов равен 100 баллам. Постройте 95-процентный доверительный интервал. Сравните его с 99-процентным и 90-процентным доверительными интервалами, внося изменения во вкладке **Параметры (Options)**.

Пример 5.3(8). Пусть результат, полученный в примере 5.3(7) не убедил нас, как исследователей. Для более точной оценки мы проводим опрос большей выборки, всех поступивших на первый курс, и получаем данные о 150 респондентах. И, допустим, что среднее и стандартное отклонение в расширенной выборке оказались почти такими же, как в маленькой. Предположите, что происходит с доверительным интервалом. Будет ли меняться соответствующий уровень значимости, и если да, то как?

Упражнение 5.3(9). В файле *IqPsy150.sav* содержатся такие результаты. Проведите анализ, аналогичный описанному в примере 5.3(7), и сделайте вывод, какую гипотезу можно принять при наличии таких данных.

Значимость по новой выборке равна 0.011. Войдет ли 100 в 99-процентный доверительный интервал? Как будет меняться значимость при увеличении уровня доверия для интервала до 99%?

Пример 5.3.4. Средний балл ЕГЭ по математике в 2016 году составил 4.14 балла по пятибалльной системе (информация взята со страницы официального сайте ЕГЭ — http://ege.edu.ru/tu/news/News/?id_4=22583).

Упражнение 5.3.4(1). В файле *egeMath.sav* содержатся модельные данные результатов такого экзамена для 1000 учеников из 10 разных школ по 100 человек в каждой. Первая переменная *Math* содержит

⁴⁴ Логично предположить, что он выше среднего, но мы в данном случае специально формулируем неопределенную гипотезу (они называются ненаправленными), то есть допускаем оба варианта, что IQ поступивших может быть и выше, и ниже популяционного среднего.

оценки, а вторая, *School* — условный код школы, от 1 до 10. Используя функцию разбиения файла по переменной (см. пример 2.3.3), рассчитайте 95%-й доверительный интервал средней оценки для каждой из школ. Ориентируясь на него, оцените, есть ли случаи и в каких из школ, когда в доверительный интервал среднего балла по школе не попадает известное нам популяционное среднее. Проинтерпретируйте результат.

Прочитайте (или перечитайте) подпараграф 5.2.3. Еще раз проинтерпретируйте результат, учитывая точный смысл доверительного интервала для среднего.

Отменив разбиение по школам, рассчитайте доверительный интервал среднего по всей выборке в 1000 человек. Насколько он соответствует данным по популяции? Отследите, как меняются стандартное отклонение и стандартная ошибка среднего в маленьких и объединенной выборках?

Глава 6. _____

Оценивание средних. Продолжение

6.1. Т-критерий: сравнение средних

6.1.1. Т-критерий для парных выборок. Что такое парные, или связанные, выборки

Довольно часто один и тот же показатель измеряется у испытуемого дважды, например, при двух разных экспериментальных условиях: до и после воздействия, в спокойном состоянии и в состоянии стресса и т.п. В качестве модельного возьмем следующий пример: у группы испытуемых измеряют тревожность до и после участия в тренинге. Можно сказать, дана выборка, состоящая из пар чисел:

$$\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$$

Исследователя при этом интересует, насколько в среднем отличаются показатели испытуемых в этих двух замерах. Обычно в таких случаях употребляют иное обозначение: говорят не о выборке, состоящей из пар чисел, а о *связанных* (или *парных*) *выборках* $\{x_1, \dots, x_n\}, \{y_1, \dots, y_n\}$.

Задачу оценки значимости изменения состояния испытуемых до и после воздействия можно легко свести к уже известным нам задачам. Если взять разность между первым и вторым замером, т.е. выборку $\{(x_1 - y_1), (x_2 - y_2), \dots, (x_n - y_n)\}$, то она и будет характеризовать воздействие: изменилось ли и если да, то насколько, состояние испытуемых. Если разность средних значений⁴⁵ отрицательна, значит, значение показателя возрастает в среднем при данном воздействии, если положительна, убывает. Таким образом, нас интересует оценка среднего значения одной переменной — в данном случае разности, т.е. та самая задача, которую мы решали в предыдущей главе.

В SPSS имеется возможность применить в данном случае Т-критерий для парных выборок или вычислить разность и применить одновыборочный Т-критерий, который мы обсуждали в предыдущей главе. Статистические значимости и доверительные интервалы полностью совпадут.

⁴⁵ Упражнение: проверьте, что разность средних значений равна среднему значению разностей.

Есть, однако, существенный аспект, различающий два метода, — это оценка величины эффекта. Величина эффекта — это оценка существа дела, которую в разных ситуациях приходится определять по-разному. Вспомним об оценках роста, которые мы обсуждали в предыдущей главе. Уточним теперь: мы решали фактически две задачи — (1) оценить средний рост какой-то группы людей и (2) оценить точность нашей оценки. Вторая задача решается оценками значимостей и доверительными интервалами. Первая — отнесением к выборочным средним и стандартным отклонениям. В случае ван Перси мы относили его рост к нормальному распределению генеральной совокупности роста голландцев и нашли, что он выше среднего голландца на одно стандартное отклонение. Похожая задача возникает, если мы хотим найти место среднего роста всей сборной Голландии в генеральной совокупности ростов голландцев.

Небольшое отличие возникнет, если доступные нам данные о росте голландцев не содержат стандартного отклонения распределения в генеральной совокупности. Тогда для оценки последнего мы можем взять выборочную оценку дисперсии у игроков сборной и извлечь из нее квадратный корень. Полученное выборочное стандартное отклонение будет давать нам единицу измерения разности между средним ростом сборной и остальных голландцев.

Подобные оценки называются оценками величины эффекта. По отношению к последнему примеру слова «величина эффекта» звучат несколько странно. Но упомянутый выше пример с тренингом тревожности полностью соответствует данному термину: насколько уменьшилась тревожность после прохождения испытуемыми тренинга? Если в среднем она уменьшилась на 20 тестовых баллов по сравнению со стартовым замером, то интерес представляет отношение этих 20 баллов к параметрам распределения в генеральной совокупности. Действительно, эти 20 баллов будут выглядеть совершенно по-разному, если стандартное отклонение измеренной с помощью данной методики тревожности равно 200 баллам или 20 баллам. В первом случае влияние тренинга можно считать практически ничтожным, во втором достаточно ощутимым.

Итак, если мы не знаем стандартное отклонение в генеральной совокупности, то в качестве его оценки разумно взять выборочное стандартное отклонение. Однако поскольку мы имеем две оценки — до воздействия и после воздействия, — вопрос не решается автоматически. Из практических соображений лучше взять оценку по первому измерению, поскольку эта выборка представляет «нетронутую» генеральную совокупность. Тренинг может стабилизировать выбор-

ку (уменьшать разброс) или, напротив, дестабилизировать выборку (увеличивать разброс). Этот аспект действия тренинга, кстати, может представлять и отдельный интерес.

Отметим, что для одновыборочного критерия в случае оценки роста базой соотнесения было стандартное отклонение, которому во втором случае соответствует стандартное отклонение разности. В последнем примере мы берем совершенно отличное от него стандартное отклонение первого замера тревожности. Различие в методах, предоставляемых SPSS, состоит в том, что таблицы вывода для одновыборочного критерия выводят только стандартное отклонение разности, а парный критерий выводит также и стандартные отклонения по каждой из исходных выборок.

Подробнее об оценках величины эффекта сообщается в приложении 4.

Описанный выше экспериментальный план называется обычно *план с повторными измерениями*. Однако парные выборки могут появляться и в других ситуациях. Подробнее об этом — в параграфе 6.2.

6.1.2. Т-критерий для независимых выборок. Что такое независимые выборки

Две независимые выборки требуются нам в ситуациях, когда мы хотим сравнить два каких-то условия, которые не могут быть отнесены к одному индивиду. Например, один индивид не может учиться сразу в двух средних школах, поэтому сравнивать эти школы можно, только сопоставляя разных учеников.

Другой вариант — оценка результативности воздействия, которое проходит на фоне процесса, который невозможно контролировать. Например, если мы придумали прекрасную методику повышения уровня оптимизма, то кроме подтверждения того, что у прошедших тренинг оптимизм увеличился, надо еще убедиться, что у людей, похожих на участников тренинга (точнее говоря, людей, принадлежащих к той же генеральной совокупности потенциальных адресатов методики), оптимизм не увеличился без участия методики — например, оптимизм россияне может массово вырасти, если вдруг цены на нефть станут устойчиво расти. Если наше исследование эффективности методики придется на такой период, то разделить рост оптимизма, связанный с воздействием тренинга, от роста, происходящего равномерно у всех или почти всех граждан РФ, можно, только сравнивая рост оптимизма экспериментальной группы (группы людей, прошедших тренинг) с ростом оптимизма контрольной группы, динамика которой связана с общими для обеих групп неконтролируемыми процессами.

Похожа на предыдущую ситуация, когда мы хотим сравнить влияния двух различных условий на какой-либо показатель. И в этом случае для каждого условия надо создать свою группу.

Еще пример: некоторые методики могут применяться как измерительный инструмент только один раз для каждого испытуемого. Предположим, мы оцениваем действие кофеина на продуктивность решения творческой задачи. Мы не можем дать эту задачу до и после приема кофеина, поскольку задача не может использоваться дважды. Это значит, что мы должны провести эксперимент с двумя группами — принявшими и не принявшими кофеин.

Предположим, в этой ситуации мы получили две выборки, содержащие измерения продуктивности решения задачи: выборку экспериментальной группы (x_1, x_2, \dots, x_n) , которая подвергалась воздействию таблеткой кофеина, и контрольную (y_1, y_2, \dots, y_n) , которая получала плацебо. Нас интересует вопрос, отличаются ли в среднем эти выборки.

Мы разберем ситуацию, когда выборки равны по объему, для которой формулы приобретают совершенно прозрачный вид. В случае неравных выборок усложняется система коэффициентов, но смысл формул — тот же самый. Как и в случае с тренингом тревожности, нас интересуют две задачи:

1. Оценить, насколько сильно сдвигается в среднем продуктивность при воздействии кофеина (величина эффекта).
2. Оценить, с какой точностью мы измерили эффект воздействия кофеина (точность оценки).

Для ответа на эти вопросы надо сначала вычислить средние и стандартные отклонения по выборкам: \bar{x} , \bar{y} , s_x , s_y , и стандартное отклонение разности

$$s = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2 + \frac{1}{n-1} \sum (y_i - \bar{y})^2}.$$

А затем рассчитать стандартную ошибку разности, которая получается делением стандартного отклонения разности на \sqrt{n} :

$$se = \frac{s}{\sqrt{n}} = \frac{\sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2 + \frac{1}{n-1} \sum (y_i - \bar{y})^2}}{\sqrt{n}}.$$

По поводу первого вопроса практически повторим то, что уже говорили: разность средних значений по экспериментальной и контрольной группам надо измерить стандартным отклонением. Разумным выглядит стандартное отклонение контрольной группы s_y , иногда предлагают также использовать корень из средней дисперсии $(D_x + D_y)/2$, который называют *объединенным* стандартным отклонением.

Надо иметь в виду, что при этом расчете мы приняли наши данные как достоверные, а на самом деле они являются результатом наложения вариаций показателей наших случайно выбранных испытуемых на ту тенденцию сдвига продуктивности, которую мы хотим оценить. Второй наш вопрос касался этого аспекта оценки.

Предположим, мы получили $\bar{x} = 4$; $\bar{y} = 5$; $s_x = 1$; $s_y = 1$. Тогда ответом на первый вопрос будет отношение $E = (\bar{x} - \bar{y})/s_y = -1$.

Как и в случае парного Т-критерия, Т-статистика вычисляется по формуле $T = (\bar{x} - \bar{y})/se$. Для больших выборок (если $2n > 120$) полученное значение вполне можно сравнивать с квантилями стандартного нормального распределения. Если в нашем примере положить $n=61$, то мы получим $se = s/\sqrt{61} = 0.19$; $T = (5 - 4)/0.19 = 5.3$. Такое Т указывает на практически достоверное отклонение разности от нуля, соответствующая значимость будет меньше, чем 0.000001, т.е. кофеин статистически надежно увеличивает продуктивность, хотя и не очень сильно, если измерять сдвиг продуктивности в стандартных отклонениях.

Мы можем построить доверительный интервал для сдвига среднего, и процедура будет теперь совершенно естественна: надо взять квантиль распределения Стьюдента и умножить на стандартную ошибку разности. Если мы хотим провести вычисления совершенно точно, то примем во внимание, что 5%-й квантиль для 120 степеней свободы все же немного отличается от квантиля нормального распределения и равен 1.98. Для 95%-го доверительного интервала радиус в таком случае будет равен $1.98 \cdot 0.19 = 0.38$. Соответствующий доверительный интервал разности $(1-0.38; 1+0.38) = (0.62; 1.38)$. Этот интервал оценивает действенность кофеина на продуктивность при решении задачи, если судить о ней по нашим выборкам

Если взять выборки вчетверо меньшего объема — по 15 человек, то радиус доверительного интервала увеличится вдвое и даже несколько больше: увеличение вдвое обеспечивает уменьшение в два раза, а дополнительное увеличение обеспечивает сдвиг квантиля в сторону увеличения, вызванный утяжелением хвостов распределения Стьюдента при уменьшении n .

Таким образом, при увеличении объема выборки выборочная величина эффекта будет колебаться вокруг константы — истинного значения величины эффекта данного воздействия, приближаясь к нему, а значимость нулевой гипотезы, как и радиус доверительного интервала, будет при увеличении выборки стремиться к нулю.

Дадим в заключение формулу, по которой вычисляется стандартная ошибка разности средних в случае неравенства объемов выборок.

Если выборки имеют объем n и m соответственно, то

$$se = \sqrt{\frac{1}{(n+m-2)}(\sum(x_i - \bar{x})^2 + \sum(y_i - \bar{y})^2)} / \sqrt{\frac{nm}{n+m}}.$$

Упражнение 6.1.2(1). Подставьте в формулу $m=n$ и убедитесь, что результат совпадает с приведенной выше формулой расчета стандартной ошибки.

Для вычисления t -статистики надо поделить разность средних на se :

$$t = (\bar{x} - \bar{y})/se$$

а радиус доверительного интервала уровня $1 - \alpha$ рассчитывается по формуле $R = t_{\alpha}^{n-1} \cdot se$ после чего строится доверительный интервал $(\bar{x} - R; \bar{x} + R)$.

6.2. Т-критерий. Углубление в тему

6.2.1. Дополнительные замечания о Т-критерии для парных выборок

Повторные измерения являются наиболее частым употреблением критерия для парных выборок. Однако возможны и другие ситуации. Например, мы можем использовать в каком-то исследовании пары близнецов, причем один из пары включается в экспериментальную группу, а другой — в контрольную. Если близнецы очень похожи, то результаты могут по структуре совпадать с повторными измерениями, а проблемы, о которых мы говорили во введении к подпараграфу 6.1.2, будут обойдены. Например, если измерение может быть проделано с каждым человеком только один раз, то «близнецовый» дизайн эксперимента, может дать значительный выигрыш по сравнению с Т-критерием для независимых выборок. Выигрыш будет тем меньше, чем меньше похожи близнецы (или иные пары, подобранные для эксперимента) и совсем исчезнет, когда мы возьмем пары случайных людей. О величине этого выигрыша можно судить по величине линейной связи (корреляции) между первой и второй переменными (они выводятся в специальной таблице SPSS). О корреляциях мы будем говорить в десятой главе.

6.2.2. Условия применимости Т-критерия для независимых выборок

Точная формула вычисления t -статистики для независимых выборок такова:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\sum(x_i - \bar{x})^2 + \sum(y_i - \bar{y})^2}} \cdot \sqrt{\frac{(n+m-2)nm}{n+m}}.$$

Если рассматривать проект исследования, то можно сказать так: мы наберем две выборки, предположительно объема m и n , и подста-

вим полученные данные в формулу t-статистики. Это значит, что мы предполагаем сделать m испытаний случайной величины X и n испытаний случайной величины Y , затем подставить полученные значения в формулу. Это в точности означает, что мы собираемся провести испытание случайной величины (заглавными буквами мы обозначаем случайные величины)

$$T = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sum(x_i - \bar{x})^2 + \sum(y_i - \bar{y})^2}{n+m}}} \cdot \sqrt{\frac{(n+m-2)nm}{n+m}}.$$

Эта случайная величина и имеет то распределение — распределение Стьюдента с $m+n-2$ степенями свободы, — по которому оцениваются квантили и радиусы доверительных интервалов. Если проводить эксперимент по данной схеме 1000 раз, то в случае, если $M_x = M_y$ (**именно в этом случае!**), гистограмма полученных 1000 результатов будет похожа на кривую теоретической плотности этого распределения.

Но теперь необходимо сделать важное уточнение: мы получим распределение Стьюдента, если наши случайные величины X и Y имеют нормальное распределение с одинаковой дисперсией. Именно исходя из таких предположений математики *доказали*, что данная случайная величина имеет распределение Стьюдента.

Распределение нужных показателей в наших генеральных совокупностях не бывает в точности нормальным, дисперсии в совокупностях, дающих нам «иксы» и «игреки», тоже могут быть различными. Это значит, что наши распределения отличаются от теоретических и квантили для оценивания параметров и гипотез мы берем неправильные. В таком случае важно знать, насколько неправильны наши квантили, как сильно отличается наше распределение от теоретического распределения Стьюдента.

В книге (Шеффе, 1980) и в статье (Корнеев, Кричевец, 2011) разобраны некоторые варианты нарушений нормальности и их последствия для пятипроцентных и однопроцентных квантилей. Коротко, результаты таковы:

Предложение 6.2.2(1). Равенство/неравенство дисперсий. При равенстве объема выборки неравенство дисперсий приводит лишь к небольшому ошибкам в расчете значимостей. При неравенстве объемов выборки ошибки тем более значительны, чем больше разница в объемах. При этом, если выборка с меньшей дисперсией имеет больший объем, статистический вывод ошибается «в нашу пользу», т.е. реальная значимость «хуже», чем дает T-критерий, если же выборка с меньшей дисперсией имеет меньший объем, то наш вывод недооценен, реальная значимость «еще лучше», чем показывает T-критерий. Причины этого легко видны: в знаменателе t-статистики в первом

случае будет большее значение (больше половины слагаемых малы) и сама t -статистика будет переоценена, а во втором случае по противоположной причине t -статистика недооценивается.

Кроме расчета, предполагающего равенство дисперсий, в SPSS есть вариант расчета, не предполагающий равенства (в нем предусмотрены коэффициенты, которые в знаменателе делают вклад оценок дисперсий двух выборок равным). Этот вариант применим *всегда*. В случае, если есть сомнения в том, что выборки сделаны из совокупностей с равными дисперсиями⁴⁶, мы рекомендуем пользоваться в SPSS именно этой опцией — «равенство дисперсий не предполагается». Чаще всего результаты обоих вариантов будут близки.

Предложение 6.2.2(2). Серьезные ошибки расчетов, по нашему опыту (см. также учебник Howit, Gramer, 2008), наблюдаются лишь в случае асимметрии распределений, причем разнонаправленной — гистограммы «скошены» (максимальные значения гистограммы (выборочная мода) расположены близко к одному краю распределения), причем в разные стороны (Корнеев, Кричевец, 2011). Ориентироваться на статистическую значимость отличия асимметрии от нуля не следует. На больших выборках будет сочтена значимой близкая к нулю асимметрия, а на малых не обнаружится существенная асимметрия распределений. Скорее, надо оценить это условие по гистограмме.

В главе 11, посвященной непараметрическим методам, мы добавим к сказанному еще несколько рекомендаций.

6.3. Практикум. Сравнение средних

В этом параграфе мы предлагаем читателю ознакомиться с процедурой сравнения средних значений в статистическом пакете SPSS. Наиболее распространенный (и понятный) пример — это ситуация повторного измерения каких-либо психологических параметров у одних и тех же испытуемых, например, изменений этих параметров во времени (возрастных изменений или связанных с каким-либо экспериментальным воздействием).

Рассматриваемые ниже примеры представляют собой данные, реконструированные по отчетам о реальных исследованиях.

⁴⁶ Решать, какой опцией воспользоваться, ориентируясь на приведенный в той же таблице вывода SPSS результат сравнения дисперсий по критерию Левиня, неразумно, поскольку статистика работает здесь против логики: на малых выборках равенство дисперсий не отвергается даже в случаях, если теоретические дисперсии отличаются сильно; а на больших выборках равенство, напротив, отвергается даже в случаях, когда теоретические дисперсии отличаются незначительно и вызванное различием отклонение от расчетной значимости ничтожно.

Пример 6.3(1). Расчет Т-критерия Стьюдента для связанных выборок в статистическом пакете SPSS.

Рассмотрим смоделированные данные исследования влияния алкоголя на время простой двигательной реакции. Эксперимент строился следующим образом. В начале у испытуемых измерялось фоновое время реакции: для этого им предъявляли визуальные стимулы и просили их как можно быстрее нажимать на кнопку при их предъявлении. При этом регистрировалось время, прошедшее от момента предъявления стимула до момента ответа испытуемого (в миллисекундах). Затем испытуемым вводили дозу алкоголя и спустя 20 минут проводили ту же пробу на время реакции повторно.

Результаты эксперимента приведены в файле *ReactionTime.sav*. Первый столбец (*num*) — это номер испытуемого, второй (*rt1*) — это среднее время реакции в 50 фоновых пробах, третий столбец (*rt2*) — это среднее время реакции в экспериментальных пробах (после приема алкоголя).

Проведем сравнение времени реакции в первой и второй пробе с помощью Т-критерия Стьюдента. Для вызова этой процедуры нужно выбрать в меню пункт *Анализ — Сравнение средних — Т-критерий для парных выборок (Analyze — Compare Means — Paired-Samples T-test)*. В появившемся окне необходимо задать по крайней мере одну пару сравниваемых переменных. В нашем случае это пара переменных *rt1 — rt2*. Переместим их в первую строку столбцов «*Переменная1*» (*Variable1*) и «*Переменная2*» (*Variable2*) соответственно. В случае, когда в исследовании есть несколько переменных, которые надо сравнить между собой, таких пар можно задать любое количество. После задания пар для сравнения нажмем кнопку *ОК*, после чего в окне вывода появится результат анализа. Таблица 6.3(2) содержит информацию об основных параметрах распределения сравниваемых переменных — средние значения, стандартное отклонение, стандартная ошибка среднего.

Таблица 6.3(2). Статистики парных выборок

	Среднее	N	Стд. отклонение	Стд. ошибка среднего
Фоновое время реакции	173.2739	70	29.80	3.56
Время реакции после принятия алкоголя	192.5388	70	26.94	3.22

Как видно из таблицы, среднее значение времени реакции после приема алкоголя увеличивается больше, чем на 19 мс.

Нет смысла сравнивать это число со стандартным отклонением разности, чтобы оценить величину эффекта. Действительно, представим ситуацию, когда все участники эксперимента увеличили время реакции ровно на 19 мс. Тогда разность замеров была бы просто константой, т.е. имела бы нулевую дисперсию. Если иметь в виду содержательную задачу – понять, что значит 19 мс для нашей выборки (величина эффекта), эту разность средних имеет смысл сравнить со стандартным отклонением фонового времени реакции, которое равно 29.8. В таком случае мы получим, что среднее время реакции после приема алкоголя отклоняется от среднего фонового на 0.65 стандартного отклонения. Это значит, например, что человек со средним фоновым временем реакции имеет сдвинутое алкоголем время реакции на уровне человека, быстрее которого $3/4$ популяции, а четверть – медленнее (см. таблицу нормального распределения в п. 4.1.3). Много это или мало — вопрос практический, а не статистический.

Статистическую значимость различий средних позволяют оценить результаты расчета Т-критерия, приведенные в таблице 6.3(3).

Таблица 6.3(3). Критерий парных выборок

		Парные разности				t	ст. св.	З н ч . (2-ст.)	
		Средн.	Стд. откл.	Стд. ошибка средн.	95% доверительный интервал разности средних				
					Нижняя граница	Верхняя граница			
Пара 1	Фоновое время реакции — Время реакции после принятия алкоголя	-19.26	38.05	4.54	-28.33	-10.19	-4.23	69	0.000

В начале этой таблицы приведены значения параметров распределения разностей оцениваемых переменных. Стандартная ошибка среднего получается делением стандартного отклонения среднего на \sqrt{N} , далее помещены доверительные интервалы разности (наша оценка -19.26 согласуется с истинными значениями в генеральной совокуп-

ности замедления реакции при приеме алкоголя, лежащими в интервале от -28.33 до -10.19), t -статистика Стьюдента есть результат деления средней разности на стандартную ошибку. В нашем случае $t = -4.23$. Следующий столбец содержит количество степеней свободы. Так как в эксперименте приняло участие 70 человек, количество степеней свободы в данном случае равно 69. Наконец, последний столбец — это двухсторонняя значимость полученного значения t . В таблице стоит число 0.000. Принимая во внимание округление, можно быть уверенными, что значимость результата меньше 0.0005. Однако есть способ получить более точное значение в таблице. Надо сделать двойной клик левой кнопкой мыши на само это число, что переведет таблицу в режим редактирования (кстати, в этом режиме можно редактировать подписи и т.п.), а затем сделать на это же число еще раз двойной клик. В результате появится более точное число. В нашем случае оно равно 0.000069. Так как значимость близка к нулю, мы можем вполне уверенно отвергнуть нулевую гипотезу и принять альтернативную, т.е. сделать вывод о том, что время реакции изменяется в результате приема алкоголя. Отчет о результате может выглядеть так: «Т-критерий Стьюдента для связанных выборок показал различие времени реакции в двух экспериментальных сериях $t(69) = -4.325$, $p = 0.000069$). Сдвиг среднего составляет 0.65 от стандартного отклонения фоновых значений времени реакции, равного 29.8».

Пример 6.3(4). В файле *Ego.sav* содержатся данные, моделирующие результаты исследований, проведенных на основе теории Дж. Левинджер⁴⁷. В теоретической модели Левинджер выделяется ряд последовательных стадий развития эго, которые диагностируются при помощи методики неоконченных предложений. Диапазон изменения оценки развития эго от двух⁴⁸ (импульсивная стадия) до девяти (стадия интеграции). В исследовании, проведенном в рамках этой концепции, испытуемых, учеников 10-го класса просили заполнить методику неоконченных предложений Вашингтонского университета. Эксперты рассчитывали средний уровень развития эго. Через пять лет (на четвертом курсе) тех же испытуемых просили заполнить методику повторно. Задачей исследования было выявить, переходят ли с возрастным испытуемые на более высокие уровни развития эго. В файле первый столбец содержит полученные оценки уровня развития эго в 10-м классе, во втором — оценки уровня эго на четвертом курсе.

⁴⁷ Hy L. X. et al. Measuring ego development. – Psychology Press, 2014.

⁴⁸ Согласно Дж. Левинджер, первая стадия развития характерна для младенчества и не встречается у детей школьного возраста и старше, поэтому методика не включает критериев ее оценки

Упражнение 6.3(5). Проверьте, есть ли изменения в уровне развития эго с возрастом. Оцените значимость с помощью Т-критерия Стьюдента

Пример 6.3(6). Файл *IllnessSelf.sav* содержит данные, полученные с помощью методики, в основе которой лежит понятие самоэффективности А. Бандуры⁴⁹.

В модели А. Бандуры самоэффективность — уверенность человека в своих силах достичь какой-либо цели. Самоэффективность в отношении болезни — уверенность человека в своих возможностях справиться с болезнью, вылечиться или (в случае хронического заболевания) — сохранять привычный уровень активности и круг интересов, несмотря на заболевание.

В исследовании оценивалась ретестовая надежность методики оценки самоэффективности в отношении болезни — устойчивость показателей с течением времени. Один из аспектов анализа устойчивости — оценка различий средних баллов выборки испытуемых при первом и втором замере. В файле приведены данные здоровых испытуемых, дважды отвечавших на вопросы методики с интервалом в три недели.

Упражнение 6.3(7). Проведите оценку изменения показателей испытуемых по двум замерам методики при помощи Т-критерия Стьюдента для связанных выборок. Подумайте, в каких единицах следует измерять степень (не)надежности методики. Проведите вычисления.

Далее мы рассмотрим Т-критерий Стьюдента для независимых выборок. Как уже говорилось выше, критерий используется для решения задачи сравнения какого-либо показателя, представленного в интервальной шкале, у двух групп испытуемых. При этом группы являются независимыми, т.е. испытуемые, входящие в первую группу, и испытуемые во второй группе — разные люди. Примером таких задач может быть сравнение интеллекта у мужчин и женщин, сравнение успеваемости у школьников двух параллельных классов, сравнение выраженности нарушений мышления у больных шизофренией и в группе нормы.

Пример 6.3(8). Расчет Т-критерия для независимых выборок

Данные, представленные в файле *Neuro.sav*, — результаты эксперимента, проводившегося в лаборатории нейropsихологии МГУ имени М.В. Ломоносова. Исследовалась взаимосвязь между уровнем

⁴⁹ Bandura A. Regulation of cognitive processes through perceived self-efficacy // Psychology of education. – 2000. – Т. 2. – С. 365-380.

успеваемости детей в школе и состоянием их нейропсихологических функций, связанных с так называемым энергетическим блоком мозга (по А.Р. Лурии). Состояние нейропсихологической функции измерялось с помощью методики быстрого автоматизированного названия⁵⁰ — испытуемым предлагались наборы различных объектов для названия, их задача состояла в как можно более быстром назывании объектов. Предполагается, что чем быстрее испытуемый справляется с заданием, тем лучше у него развиты нейропсихологические функции, отвечающие за общую активацию.

В файле *Neuro.sav* приведены результаты проведения описанной методики у младших школьников. Дети были поделены на две группы — хорошо успевающих и отстающих, второй столбец (*group*) содержит код группы, к которой отнесен каждый испытуемый, «0» обозначает отстающих, «1» — хорошо успевающих. Далее представлены времена выполнения трех проб: название цифр, название осмысленных изображений и название цветов (переменные *numbers*, *figures* и *colors* соответственно). Время измерялось в секундах.

Исследователи предполагали, что среднее время выполнения проб может различаться у двух выделенных групп детей: отстающие справятся с заданием за большее время, чем хорошо успевающие.

Сравним средние значения с помощью Т-критерия Стьюдента для несвязанных выборок. Для этого в меню надо выбрать пункт *Анализ — Сравнение средних — Т-критерий для независимых выборок (Analyze — Compare means — Independent-Sample T-test)*. В появившемся диалоговом окне в поле *Проверять переменные (Test Variable(s))* необходимо внести переменные, значения которых сравниваются у двух групп. Перенесем в окно переменную *numbers* (в это окно можно переносить сколько угодно переменных одновременно). В поле *Группировать по (Grouping Variable)* следует внести переменную *group*, кодирующую принадлежность испытуемых к той или иной группе. Затем, нажав кнопку *Задать группы (Define groups)* необходимо указать номера двух сравниваемых групп, в нашем случае это 0 и 1. Это дополнительное окно важно при попарном сравнении нескольких экспериментальных групп — в этом случае можно последовательно задавать различные пары для сравнения средних.

⁵⁰ Denckla M. B., Rudel R. Rapid «automatized» naming of pictured objects, colors, letters and numbers by normal children //Cortex. — 1974. — Т. 10. — №. 2. — С. 186-202.

Таблица 6.3(9). Групповые статистики

	Группа	N	Среднее	Стд. отклонение	Стд. ошибка среднего
Проба «цифры» — время выполнения	отстающие	15	38.9627	8.85589	2.28658
	хорошо успевающие	15	28.9627	8.92550	2.30455

После нажатия кнопки *OK* в окне вывода появляются результаты расчетов. Первая таблица содержит информацию о размере групп (столбец *N*), среднем значении, стандартном отклонении и стандартной ошибке сравниваемых параметров в двух экспериментальных группах (таблица 6.3(9)).

Если обратить внимание на средние показатели времени выполнения проб в двух группах, то становится видно, что отстающие справлялись с заданием за большее время, чем дети из группы нормально успевающих: среднее время в секундах составило 38.96 и 28.96 соответственно. На вопрос об оценке значимости гипотез отвечает таблица 6.3(10). Рассмотрим ее подробнее (для компактности таблица сжата по горизонтали по сравнению с тем видом, который она имеет в окне вывода SPSS).

Таблица 6.3(10). Т-критерий для независимых выборок

		Критерий равенства дисперсий Ливиня		t-критерий равенства средних							
										95% доверительный интервал разности средних	
		F	Знч.	t	ст. св.	Значимость (2-сторонняя)	Разность средних	Стд. ошибка разности	Нижняя граница	Верхняя граница	
Проба «цифры» — время выполнения	Предполагается равенство дисперсий	4.412	.045	3.08	28	.005	10.00	3.24645	3.34995	16.65005	
	Равенство дисперсий не предполагается			3.08	27.998	.005	10.00	3.24645	3.34994	16.65006	

В SPSS Т-критерий для независимых выборок рассчитывается всегда в двух вариантах: предполагающем равенство дисперсий показателей сравниваемого параметра в двух группах и не предполагающем равенство дисперсий. В данном случае дисперсии практически равны и результаты в двух строках почти не различаются. Мы уже указывали, что всегда можно использовать строку *Равенство дисперсий не предполагается (Equal of variances not assumed)* (см. подпараграф 6.2.2). Чаще всего результаты двух вариантов различаются мало. В следующем примере 6.3(13) мы продолжим обсуждение этой темы.

Столбец *t* содержит результат расчета t-статистики, далее приведены соответствующие степени свободы (при неравенстве дисперсий число степеней свободы перестает измеряться целым числом). Столбец *Знч. (Sig.)* содержит значимость полученного значения t-статистики, в данном случае она равна 0.005. Именно этой значимостью оцениваются различия среднего времени выполнения проб, характеризующих состояния энергетического блока мозга испытуемыми двух групп. Кратко результаты применения Т-критерия можно записать так: $t(28) = 3.08$, $p = 0.005$.

В таблице приведена не только разность средних, которая равна 10 секундам, но и 95%-й доверительный интервал для разности: от 3.35 до 16.65 секунд — это напоминает нам, что 10 секунд — лишь оценка различия средних. Остается, однако, вопрос, что значат эти секунды с точки зрения различия развития функций.

На этот вопрос отвечают, оценивая величину эффекта (SPSS ее не рассчитывает). В данном случае разумная мера — отношение разности средних к некоему среднему стандартному отклонению выборок (квадратному корню из средней дисперсии). Выборочные дисперсии также приходится вычислять вручную, возводя в квадрат стандартные отклонения 8.85589 и 8.92550. Средняя дисперсия равна 79.045, корень из нее — 8.89, а $10/8.89=1.125$. Это и есть величина эффекта⁵¹, которую в данном случае точнее называть мерой различия групп: средние по группам различаются на 1.125 стандартных отклонения. Некоторую визуализацию этой оценки дает рис. 6.3(11).

Упражнение 6.3(12). Повторите самостоятельно расчет Т-критерия с остальными двумя переменными из файла *Neuro.sav*.

⁵¹ Точнее, один из способов ее оценки, называющийся *d* Коэна

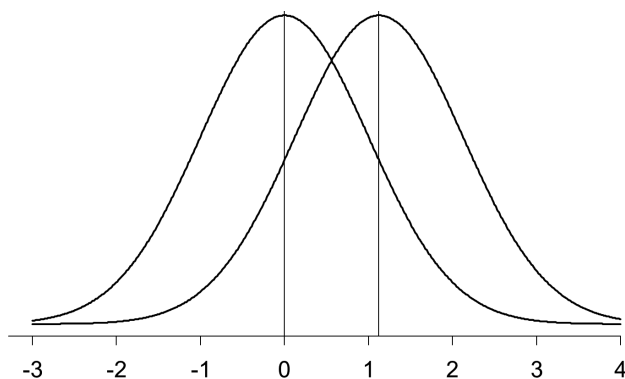


Рис. 6.3(11). Так выглядели бы два нормальных распределения, средние которых отличаются на 1.125 стандартных отклонения. По аналогии можно представлять себе соотношение распределений характеристики недоразвития энергетического блока мозга (чем больше время на выполнение задания, тем хуже = слева успевающие, справа отстающие).

Пример 6.3(13). Данные, представленные в файле *recital1.sav* моделируют результаты школьного изложения представителей двух параллельных учебных классов (кодируются переменной *group*). Переменная *Time* кодирует время, затраченное на работу, переменная *Score* – оценку содержания текста, переменная *Mistakes* – количество ошибок. В первой группе 10 человек, во второй 20. Между группами имеются весьма существенные различия в дисперсиях по двум из трех переменных. Наша задача – проиллюстрировать изложенное в предложении 6.2.2(1): как влияет на значимость неравенство дисперсий, когда сравниваемые выборки имеют разный объем.

Упражнение 6.3(14). Прделайте расчет Т-критерия для всех трех переменных сразу (их можно переместить в окно **Проверяемые переменные (Testing variables)** одновременно).

Дисперсия переменной *Time* в меньшей группе существенно больше, что фиксируется критерием равенства дисперсий Ливиния. Различия в оценках значимости между критериями, предполагающим и не предполагающим равенство дисперсий, достаточно серьезное.

Дисперсия переменной *Mistakes* в меньшей группе, напротив, значительно меньше, чем в большей. Различие оценок значимости также велико, но направлено противоположно.

Дисперсии переменной *Score* примерно равны, и оценки значимости практически не различаются.

Полезно представить себя одновременно докладчиком курсовой работы, сообщаящим эти результаты, и параллельно слушателем-оппонентом. Предположим, в докладе сообщалось только о значимости, полученной в *предположении равенства дисперсий* (первая строка таблицы). Что в каждом случае отметит критик? Попробуйте сформулировать ответ самостоятельно, прежде чем читать наш ответ.

Ответ. Грамотный оппонент скажет о переменной *Time*: «Вы переоценили свой результат. На самом деле значимость вашего результата хуже, поэтому, собственно говоря, никакого результата нет. Вы заслуживаете оценки “неудовлетворительно”».

О переменной *Mistakes* оппонент скажет: «Вы недооценили свой результат. Если бы Вы, заметив неравенство дисперсий, применили более адекватный критерий оценки, то получили бы, что он еще лучше обоснован. Вы заслуживаете снижения оценки до “хорошо” за несовершенное знание статистики».

По поводу третьей переменной *Score* замечаний бы вообще не последовало.

Пример 6.3(15). В файле *recital2.sav* содержатся данные, аналогичные описанным в предыдущем задании, с теми же средними и дисперсиями, но при равных объемах выборок.

Упражнение 6.3(16). Убедитесь, что критерий Ливиня сообщает о еще более близкой к нулю значимости различий дисперсий (там, где они есть), поскольку несколько увеличилась выборка. В то же время значимость различий средних оценивается критериями практически одинаково. Что изменилось бы в ситуации доклада результатов курсовой работы?

Ответ. Докладчику сделали бы только замечание, что он не сообщил о неравенстве дисперсий (действительно, в докладах такие вещи лучше говорить явно), но оценку бы не снизили.

Наши последние упражнения иллюстрируют рекомендацию всегда ориентироваться на вторую строку таблицы (равенство дисперсий не предполагается) и по возможности исследовать содержательно причины и следствия неравенства дисперсий, если они есть, поскольку это может оказаться неожиданным и интересным результатом исследования.

Глава 7.

Однофакторный дисперсионный анализ

7.1. Однофакторный дисперсионный анализ. Примеры и разъяснения

7.1.1. Почему недостаточно парных сравнений

Предположим, что мы планируем провести исследование влияния способа проведения дополнительных занятий по математике на экзаменационные оценки студентов. Пусть студентов разделили на три примерно равные по силе группы и *предложили* первой группе по желанию прослушать факультативный курс, разъясняющий существенные аспекты использования статистических методов. Вторую группу *обязали* прослушать такой курс, а третьей, контрольной, возможности прослушать этот курс *не предоставили*. После прослушивания курса первыми двумя группами состоялся экзамен для всех трех групп. Как оценить воздействие дополнительного курса при данных трех условиях?

Разумеется, в первую очередь следует вычислить средние экзаменационные оценки по группам. Предположим, некоторые интересные различия обнаружены: добровольные слушатели показали наиболее высокие оценки, а лишенные возможности слушать курс — самые низкие в среднем оценки. Как и в случае сравнения средних для двух групп, нас интересуют два вопроса: (1) оценка — насколько велики расхождения между группами? (величина эффекта) и (2) насколько надежны полученные нами результаты? (статистическая значимость). Требуется метод, обобщающий Т-критерий на случай трех и более групп.

Прежде всего, заметим, нельзя решить задачу, сравнивая попарно каждую группу с каждой — такое попарное сравнение может приводить к ошибкам тем более серьезным, чем больше количество сравниваемых групп. Если мы станем сравнивать 20 различных условий по влиянию на подобный или какой-то иной показатель, ошибка станет практически неизбежной.

В практикуме в конце главы вы убедитесь, что двадцать групп по пятьдесят испытуемых в каждой практически всегда продемонстрируют отличие лучшей группы от худшей на уровне значимости меньше 0.05, даже если группы эти набирались совершенно случайно и ни о каком воздействии, которое могло бы привести к систематическому сдвигу

гу среднего значения, речи не шло. Например, вы можете сравнить пассажиров двадцати вагонов метро по тесту интеллекта и убедиться, что значимость отличия лучшего вагона от худшего по Т-критерию, как правило, меньше 0.05.

Можно сказать, что $20(20-1)/2$ (число сочетаний из 20 по 2) попарных сравнений «всех со всеми» практически гарантируют хотя бы одну ошибку первого рода, т.е. отвержение гипотезы H_0 , когда она на самом деле верна. Вспомним, что уровень значимости критерия (например, 0.05 — см. 5.1.4) — это и есть вероятность ошибки первого рода, и повторив сравнения двадцать раз на разных парах групп, мы с большой вероятностью будем наблюдать хотя бы один случай события, вероятность которого равна 0.05 (или $1/20$), т.е. хотя бы раз совершим ошибку первого рода. В нашем случае сравнения «всех со всеми» вероятность еще больше. Чтобы избежать опасности такой ошибки, дисперсионный анализ охватывает всю совокупность групп общей оценкой — насколько похожа на случайную вся эта совокупность результатов.

Дисперсионный анализ сравнивает дисперсии двух видов — внутри групп (связанную со случайными, неконтролируемыми различиями между испытуемыми) и между группами (связанную с влиянием группирующей переменной, или фактора). Мы помним, что при сравнении двух групп t-статистика измеряет разность средних стандартной ошибкой. Дисперсионный анализ измеряет *квадрат* разности средних *квадратами* стандартной ошибки, т.е. результат для двух выборок равен квадрату рассчитанной по этим же данным t-статистики (подробнее подпараграфе 7.2.1). Далее мы приводим пример расчета этого показателя. Вместо дисперсий мы будем считать похожие на них суммы квадратов. В конце эти суммы будут поделены на соответствующие константы и превратятся в оценки дисперсии.

7.1.2. Числовой пример. Сравнение условий обучения

Естественного аналога разности средних, которая стоит в числителе Т-статистики, для нескольких групп придумать не удастся. Однако модифицированный вариант (детали в параграфе 7.2), сравнивающий дисперсии внутри групп с дисперсией средних значений по группам, обобщается на случай нескольких групп совершенно естественно. Если в различных группах влияние определяющего фактора (в нашем случае это способ вовлечения в прослушивание курса) на результат различается, то разброс между группами окажется относительно большим, если же влияние незначительно, то и разброс между группами будет примерно равен разбросу внутри групп.

Предположим, конечные результаты экзамена первой экспериментальной группы представлены выборкой $\{2, 4, 6\}$, результаты вто-

рой группы — $\{2, 3, 4\}$, а результаты контрольной группы — выборкой $\{1, 2, 3\}$. Разумеется, никогда такие исследования не проводятся на группах из трех испытуемых каждая. Наш пример таков исключительно из соображений удобства ручных расчетов. Мы рекомендуем, если возможно, параллельно с чтением текста выполнять пример 7.3(1).

Мы начнем с вопроса о статистической значимости результата, а оценкой эффекта займемся в девятой главе, где рассмотрим этот вопрос сразу и для регрессионного, и для дисперсионного анализа.

Мы оценим значимость относительно полученного результата гипотезы H_0 , утверждающей, что теоретические средние (математические ожидания) во всех группах равны, т.е. способ вовлечения в прослушивание курса не сказывается на результате⁵². Альтернативная гипотеза H_1 утверждает, что теоретические групповые средние в каждой группе свои⁵³, а варьирование условий подачи материала производит некоторый систематический сдвиг показателей у целой группы.

Для расчета значимости запишем данные в таблицу:

факультатив	обязательный	отсутствие
2	2	1
4	3	2
6	4	3

Далее мы разложим таблицу в сумму трех таблиц. Во-первых, выделим таблицу, в которой заполним все клетки общим для трех выборок средним. Заголовки столбцов опускаем для краткости.

$$\begin{array}{|c|c|c|} \hline 2 & 2 & 1 \\ \hline 4 & 3 & 2 \\ \hline 6 & 4 & 3 \\ \hline \end{array} = \begin{array}{|c|c|c|} \hline 3 & 3 & 3 \\ \hline 3 & 3 & 3 \\ \hline 3 & 3 & 3 \\ \hline \end{array} + \begin{array}{|c|c|c|} \hline -1 & -1 & -2 \\ \hline 1 & 0 & -1 \\ \hline 3 & 1 & 0 \\ \hline \end{array}$$

Каждая клетка таблицы в левой части равенства представлена суммой соответствующих клеток таблиц в правой части. Во второй таблице справа сумма всех чисел равна нулю, поскольку общее среднее «ушло» в предыдущую таблицу.

⁵² В простом варианте дисперсионного анализа также предполагается равенство групповых дисперсий, это означает, что все девять полученных значений имеют своим источником независимые испытания *одной и той же* нормально распределенной случайной величины X , т.е. различные способы подачи дополнительного материала вообще не влияют на итоговые оценки

⁵³ И здесь дисперсии предполагаются равными.

Теперь для последней таблицы рассчитаем в каждом столбце среднее значение (среднее по группе) и разложим ее в сумму таблицы, содержащей в каждом столбце соответствующее групповое среднее, и таблицы, содержащей остаток:

$$\begin{array}{|c|c|c|} \hline 2 & 2 & 1 \\ \hline 4 & 3 & 2 \\ \hline 6 & 4 & 3 \\ \hline \end{array} = \begin{array}{|c|c|c|} \hline 3 & 3 & 3 \\ \hline 3 & 3 & 3 \\ \hline 3 & 3 & 3 \\ \hline \end{array} + \begin{array}{|c|c|c|} \hline 1 & 0 & -1 \\ \hline 1 & 0 & -1 \\ \hline 1 & 0 & -1 \\ \hline \end{array} + \begin{array}{|c|c|c|} \hline -2 & -1 & -1 \\ \hline 0 & 0 & 0 \\ \hline 2 & 1 & 1 \\ \hline \end{array}$$

Каждая клетка левой таблицы представлена теперь суммой соответствующих клеток трех следующих таблиц. При этом сумма групповых средних во второй таблице в правой части равна нулю, а в третьей таблице равна нулю сумма в каждом столбце, поскольку из этих чисел вычтено среднее по столбцу. Посчитаем теперь для всех четырех таблиц суммы квадратов всех клеток таблицы.

$$\begin{aligned}
 S_{Total} &= 2^2 + 2^2 + 1^2 + 4^2 + 3^2 + 2^2 + 6^2 + 4^2 + 3^2 = \\
 &4 + 4 + 1 + 16 + 9 + 4 + 36 + 16 + 9 = 99.
 \end{aligned}$$

Для первой таблицы справа $S_{const} = 3^2 * 9 = 81$.

Для второй таблицы справа $S_{model} = 1^2 * 3 + 0^2 * 3 + (-1)^2 * 3 = 6$.

Для третьей таблицы справа

$$S_{error} = (-2)^2 + (-1)^2 + (-1)^2 + 0^2 * 3 + 2^2 + 1^2 + 1^2 = 12.$$

Наше разложение таблиц обладает замечательным свойством: разложились также и суммы квадратов: $99 = 81 + 6 + 12$.

Суммы имеют стандартное обозначение: $S_{total} = S_{const} + S_{model} + S_{error}$ (общая сумма, сумма константы, сумма модели⁵⁴ и сумма ошибки, которую также называют остатком). Константа выражает среднее значение признака по всей выборке, сумма модели — изменчивость признака в зависимости от фактора (различие между группами), сумма ошибки отражает случайные по отношению к влиянию фактора колебания признака.

Примечание для знакомых с линейной алгеброй. Если каждую, состоящую из 9 элементов таблицу представить 9-мерным вектором, то получится векторное равенство, причем попарные скалярные произведения векторов в правой части все равны нулю, что означает, что они ортогональны. Не удивительно, что при разложении вектора по ортогональным направлениям квадрат длины гипотенузы равен сумме ква-

⁵⁴ В литературе эта сумма часто обозначается S_{bg} (аббревиатура от «Between group» — межгрупповая сумма).

дратов обобщенных катетов (теорема Пифагора в многомерном случае), а это и есть наше разложение сумм квадратов.

Заметим, что первая таблица разложения (в правой части), состоящая из одинаковых чисел, психологов обычно не интересует, поскольку шкалы психологических переменных практически всегда интервальные, а значит, положение нуля в них неопределенно. Сделав же допустимое преобразование шкалы — изменив положение нулевой точки, мы получим другие константы, но две последние таблицы останутся без изменений. Они-то нас и интересуют.

Нам потребуется еще расчет так называемых степеней свободы (обозначается df — *Degrees of Freedom*) для каждой таблицы. Число степеней свободы таблицы — это число независимых чисел в этой таблице. В первой таблице все девять чисел могут быть какие угодно, поэтому для нее $df_{Total} = 9$. Для второй таблицы только одно число может изменяться независимо, остальные должны быть ему равны, поэтому $df_{const} = 1$. Далее, для таблицы, содержащей средние по группам, независимы два числа: первые числа в первом и втором столбце. Другие числа в этих столбцах совпадают с первыми, а числа в третьем столбце должны дополнять первые два до нуля, следовательно, $df_{model} = 2$. В последней таблице в каждом столбце два независимых числа, третье дополняет их до нуля, поэтому $df_{error} = 2 * 3 = 6$ (3 — число столбцов). Легко видеть, что и здесь имеет место равенство $df_{Total} = df_{const} + df_{model} + df_{error}$ ⁵⁵.

Значимость, характеризующая гипотезу H_0 по отношению к полученным данным, рассчитывается далее так.

Сначала составляется так называемое F-отношение и вычисляется его значение:

$$F = \frac{S_{model}/df_{model}}{S_{error}/df_{error}},$$

а затем по таблице или другим каким-нибудь способом (в статистическом пакете программа берет эту проблему на себя) находится вес соответствующего хвоста распределения правее данного значения F. Это распределение называется распределением Фишера. Деление на число степеней свободы делает из сумм квадратов оценки дисперсии (именно поэтому метод называется дисперсионным анализом). Если верна гипотеза о том, что настоящие средние по группам (может быть, понятнее будет сказать: по генеральным совокупностям, за группами стоящим) равны, то в числителе и знаменателе стоят оценки одной и той же дисперсии. Это значит, что F-отношение будет колебаться около единицы. К сожалению, распределение Фишера совсем не похоже

⁵⁵ Знакомые с линейной алгеброй уже догадались, что это равенство размерностей пространства и суммы размерностей его непересекающихся подпространств, сумма которых дает все пространство.

на нормальное распределение. Для каждого сочетания объема выборок в группах и количества групп распределение F будет свое, поэтому правильное будет, если количество степеней свободы мы подпишем в качестве индексов к F. Тогда в нашем случае

$$F_6^2 = \frac{6/2}{12/6} = 1.5.$$

В нашем случае хвосты надо будет смотреть по распределению Фишера с числом степеней свободы (2, 6). Примерный вид графика плотности распределения Фишера можно найти на рис. 7.1.2(1)⁵⁶.

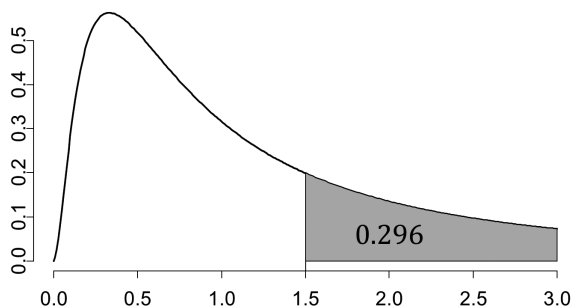


Рис. 7.1.2(1). Примерный вид распределения Фишера

Чем больше значение F, тем надежнее можно говорить о влиянии фактора на результат. Точная значимость в нашем случае равна 0.296 (см. рис. 7.1.2(1)), т.е. результат не позволяет сделать вывод о существенном влиянии способа подачи дополнительных занятий на результаты экзамена.

7.1.3. Зависимость значимости от объема выборки

Как и в случае T-критерия, в случае дисперсионного анализа увеличение объема выборки также приводит к повышению надежности вывода. Чтобы в этом убедиться, достаточно посмотреть внимательно на формулу

$$F = \frac{S_{model}/df_{model}}{S_{error}/df_{error}}$$

и приведенные ниже таблицы. Предположим, мы набрали еще по три испытуемых в каждую группу, причем их показатели имеют те же характеристики, которые имели и первые три представителя данной груп-

⁵⁶ При различных комбинациях числа степеней свободы числителя и знаменателя графики распределения Фишера демонстрируют большое разнообразие. Большинство из них похожи на наш рисунок. Для объяснения сути дела такого рисунка достаточно.

пы. Для удобства расчетов просто повторим уже имеющиеся числа. Поскольку нас интересуют две последние таблицы, запишем только их для обоих объемов выборок.

Для первоначального объема выборок:

1	0	-1	+	-2	-1	-1
1	0	-1		0	0	0
1	0	-1		2	1	1

Для удвоенных выборок:

1	0	-1	+	-2	-1	-1
1	0	-1		0	0	0
1	0	-1		2	1	1
1	0	-1		-2	-1	-1
1	0	-1		0	0	0
1	0	-1		2	1	1

В первом случае $S_{model1} = 6; S_{error1} = 12$.

Во втором случае $S_{model2} = 12; S_{error2} = 24$ — обе суммы удвоились.

Значения F будут различаться только из-за различий степеней свободы. В первом случае $df_{model} = 2; df_{error} = 6$ во втором случае $df_{model} = 2; df_{error} = 15$ (в каждом из трех столбцов по пять независимых чисел).

Таким образом, в первом случае

$$F_6^2 = \frac{6/2}{12/6} = 1.5.$$

во втором случае

$$F_{15}^2 = \frac{12/2}{24/15} = 30/8 = 3.75.$$

т.е. F увеличилось более чем в два раза (при удвоении больших выборок F бы почти точно удвоилось — убедитесь в этом, сравнив степени свободы для больших выборок). Соответствующая значимость равна 0.048, против 0.296 в первом случае.

Заметим еще, что отношение рассчитанных по двум таблицам сумм квадратов без учета степеней свободы оказалось постоянным при удвоении выборки. В реальных исследованиях тождества, конечно, не бывает, но, как и в случае Т-критерия, увеличение выборки заставляет это отношение приближаться к теоретическому отношению разбросов по ге-

неральным совокупностям, т.е. наблюдаемые значения этого отношения в той или иной степени близки к константе. Насколько надежно можно считать их близкими, мы судим по значимости, описанной выше. В главе, посвященной регрессии, мы дадим показатель, который обычно используют, чтобы характеризовать это отношение в духе величины эффекта.

7.1.4. Апостериорные сравнения

После того как мы оценили всю совокупность средних значений по группам общей оценкой значимости, мы можем поставить вопрос, какие группы отличаются уверенно, а какие нет. Для этого служат апостериорные сравнения: проводятся сравнения по Т-критерию каждой группы с каждой, но показываемая Т-критерием значимость умножается на зависящую от количества сравнений константу. В результате скорректированные значимости могут интерпретироваться, как интерпретировались бы значимости при отдельных сравнениях, но без риска завязать результат вследствие большого числа сравнений.

В случае нашего «удвоенного» примера значимость различия первой и третьей группы равна 0.046 (при использовании метода Бонферрони)⁵⁷, в то время как непосредственное сравнение по Т-критерию этих групп дает значимость 0.034. Значимости различий первой от второй и второй от третьей группы равна 0.573. Можно осторожно интерпретировать результат так: при повторениях эксперимента первая группа покажет, скорее всего, результат больший, чем третья, но со второй группой результат может сравняться и даже поменяться местами, как и у групп второй и третьей.

7.1.5. Односторонние и двухсторонние критерии

Напоминаем, что распределение Фишера, о котором идет речь в дисперсионном анализе, в общих чертах выглядит так, как показано на рис. 7.1.2(1). Для оценки значимости результата относительно нулевой гипотезы в дисперсионном анализе используется верхний *односторонний* квантиль, в то время как в Т-критерии используется двухсторонний. В данном случае причина этого выбора понятна. Например, при сравнении двух групп экстремально большим положительным и экстремально большим по модулю отрицательным значениям t-статистики

⁵⁷ Помимо метода Бонферрони, самого консервативного, т.е. наиболее строгого, существует множество других, среди которых можно выделить как самые распространенные методы Тьюки (Tukey) и Хольма (Holm). Более подробно математические проблемы решения задачи множественного сравнения обсуждаются, например в Lehmann E. L., Romano J. P. Chapter 9: Multiple testing and simultaneous inference // Testing statistical hypotheses. — 3rd ed.. — New York: Springer, 2005.

соответствуют только экстремально большие положительные значения статистики Фишера для дисперсионного анализа. Точные выкладки приведены в параграфе 7.2. Но вопрос о выборе одностороннего или двухстороннего критерия при анализе данных в других случаях требует более глубокого понимания сути дела.

Рассмотрим пример: пусть мы оценили рост оптимизма членов экспериментальной группы, которые прошли через тренинг, и сравнили его с ростом оптимизма членов контрольной группы, которые в тренинге не участвовали. Если мы заинтересованы в том, чтобы показать, что тренинг эффективен, то рассчитываем получить положительную разность средних между экспериментальной и контрольной группами.

Предположим, имея по 16 испытуемых в каждой группе, мы получили Т-статистику, равную 2.08. Наш результат имеет значимость 0.046. Однако в таблице SPSS столбец таблицы, где значимость приведена, озаглавлен «2-сторонняя значимость». Подумаем, что было бы, если бы мы получили значение Т-статистики, равное -2.08 , что свидетельствовало бы о том, что оптимизм экспериментальной группы, несмотря на тренинг, оказался хуже, чем у контрольной. В таблице появилась бы та же самая значимость 0.046. Дело в том, что 2.083 — верхний двухсторонний квантиль уровня 0.046 для распределения Стьюдента с 30 степенями свободы. Случайная величина, имеющая соответствующее распределение, с вероятностью 0.046 принимает значение, по модулю превосходящее 2.08, т.е. либо меньшее, чем -2.08 , либо большее, чем 2.08.

Если мы сопоставляем две гипотезы: «тренинг эффективен» и «тренинг неэффективен», вопрос о том, односторонний или двухсторонний критерий надо применять в данной ситуации, статистика предлагает решать в зависимости от того, что мы, планируя исследование, собираемся делать, получив в результате расчета экстремально большое по модулю отрицательное значение t -статистики, свидетельствующее о том, что тренинг ухудшает ситуацию.

В принципе возможны две позиции. (1) Мы можем интересоваться отбором эффективных методик. Получив такой результат, мы просто эту методику отбрасываем. (2) Нас интересуют всякие влияния тренингов на уровень оптимизма. Получив отрицательный результат, мы планируем написать статью о том, что некоторые тренинги ухудшают ситуацию.

Добросовестный исследователь в первом случае может использовать односторонний критерий. Тогда (односторонняя) значимость результата равна 0.023, т.е. половине от двухсторонней. Во втором случае надо пользоваться двухсторонним критерием. Будет нарушением научной этики задним числом заявить, что нас интересовали только отрицательные результаты, поэтому мы использовали односторонний критерий.

Если заранее был поставлен вопрос о влиянии тренинга на оптимизм с замыслом считать достойными внимания и положительный и отрицательный сдвиги, критерий должен браться двухсторонний (рис. 7.1.5(2)).

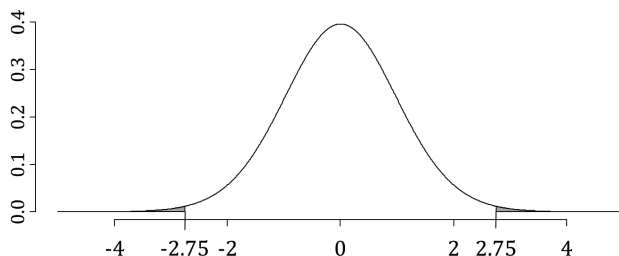


Рис.7.1.5(2). Двухсторонние и односторонние критериальные значения. На рисунке изображен график распределения Стьюдента с 30 степенями свободы. Одно и то же значение 2.75 является критериальным значением на уровне значимости 0.005 при односторонней (направленной) гипотезе исследования и верхним критериальным значением на уровне значимости 0.01 при двухсторонней (ненаправленной) гипотезе исследования

Таким образом, вопрос, который мы задаем природе и на который ждем ответа от статистики, должен формулироваться так, чтобы было ясно, какого рода альтернативу мы предполагаем для основной гипотезы. Если альтернатива явно односторонняя (оценить, достаточно ли выражено позитивное воздействие тренинга), то и критерий может использоваться односторонний (например, основная гипотеза отвергается, если получено аномально большое значение статистики Стьюдента). Если же ставится вопрос, предполагающий обе альтернативы (например, меняется ли с возрастом уровень оптимизма?), то критерий должен браться двухсторонний и гипотеза о неизменности должна отвергаться, если получен аномально большой или аномально низкий результат. В первом случае (односторонняя гипотеза) попадание результата в верхнюю критериальную область означает отвержение основной гипотезы на уровне 0.005, а во втором (двухсторонняя гипотеза) попадание в эту же критериальную область означает, что основная гипотеза отвергается на уровне значимости только 0.01 (рис. 7.1.5(2)).

Отметим, что критерий Фишера (дисперсионный анализ) при сравнении двух групп не может реализовать односторонний вариант, поскольку не различает знак разности средних. Если же групп три, невозможно даже определить понятие разности средних сразу для всех трех групп.

7.2. Эквивалентность T-критерия и дисперсионного анализа для двух выборок

Дисперсионный анализ для двух выборок эквивалентен T-критерию в том смысле, что значимости гипотезы о равенстве средних в обоих случаях равны.

Пусть у нас имеются две выборки $\{x_1, x_2, x_3\}$ и $\{y_1, y_2, y_3\}$, причем их общее среднее уже заранее равно нулю. В таком случае $\bar{x} = -\bar{y}$. Обозначим для удобства $\bar{x} = a, \bar{y} = b$ с условием $b = -a$. Таблица выборочных значений разложится в две таблицы (поскольку таблица констант состоит из нулей, мы ее опускаем):

$$\begin{array}{|c|c|} \hline x_1 & y_1 \\ \hline x_2 & y_2 \\ \hline x_3 & y_3 \\ \hline \end{array} = \begin{array}{|c|c|} \hline a & b \\ \hline a & b \\ \hline a & b \\ \hline \end{array} + \begin{array}{|c|c|} \hline x_1 - a & y_1 - b \\ \hline x_2 - a & y_2 - b \\ \hline x_3 - a & y_3 - b \\ \hline \end{array}$$

Для числителя число степеней свободы равно $2-1 = 1$, для знаменателя $2(3-1) = 4$.

F-отношение запишется так:

$$F_4^1 = \frac{3a^2 + 3b^2}{((x_1 - a)^2 + (x_2 - a)^2 + (x_3 - a)^2 + (y_1 - b)^2 + (y_2 - b)^2 + (y_3 - b)^2) / (2(3 - 1))}$$

Перенесем двойку в числитель главной дроби (эта двойка, заметим, выражает количество групп), а тройку в главном числителе вынесем за скобки и перенесем за пределы дроби:

$$F_4^1 = \frac{2 * (a^2 + b^2)}{((x_1 - a)^2 + (x_2 - a)^2 + (x_3 - a)^2 + (y_1 - b)^2 + (y_2 - b)^2 + (y_3 - b)^2) / (3 - 1)} * 3.$$

Теперь из этих же выборок составим формулу для T-критерия:

$$T_4 = \frac{a - b}{((x_1 - a)^2 + (x_2 - a)^2 + (x_3 - a)^2 + (y_1 - b)^2 + (y_2 - b)^2 + (y_3 - b)^2) / (3 - 1)} * \sqrt{3}.$$

Если мы возведем T в квадрат, то лишь числители главной дроби будут выглядеть по-разному.

Однако, поскольку $a = -b$, то $a - b = 2a$, то $2 * (a^2 + b^2) = 4a^2$ и по той же причине $(a - b)^2 = (2a)^2 = 4a^2$. Мы доказали, таким образом, что для одинаковых выборок $(T_4)^2 = F_4^1$. Понятно, что в общем случае для выборок объема n доказательство будет отличаться только многоточиями и буквой n вместо цифры 3 во всех ее вхождениях в формулы.

Технически сложнее доказательство для неравных выборок, но принципиальных трудностей нет и в нем.

Отсюда следует, например, что если $t_{0.05}^{17} = t$ является двусторонним 5-процентным квантилем для распределения Стьюдента с 17 степенями свободы, то для F_{17}^1 5%-й квантиль равен t^2 . Аналогично для любого объема выборок и любого «номинала» уровня значимости.

7.3. Однофакторный дисперсионный анализ. Практикум

Пример 7.3(1). В файле *EduCondition.sav* содержатся данные примера подпараграфа 7.1.2. об условиях прослушивания некоторого курса студентами (факультатив, обязательный курс, отсутствие курса).

Отличие от данных, предназначенных для расчета Т-критерия, состоит в большем количестве групп. В данном случае группирующая переменная *Condition* принимает три значения.

Для проведения однофакторного дисперсионного анализа в SPSS необходимо вызвать в меню **Анализ — Сравнение средних — Однофакторный дисперсионный анализ (Analyze — Compare means — One-way ANOVA)**. После этого на экране в возникшем диалоговом окне необходимо независимую переменную перенести в поле **Фактор (Factor)**, а зависимую — в поле **Список зависимых переменных (Dependent list)**⁵⁸. В нашем примере независимой переменной является группирующая переменная *Condition*⁵⁹, а зависимой — экзаменационная оценка *Score*.

Нажмем **OK**, и в файле вывода получим основную таблицу 7.3(2). В данной таблице указаны суммы квадратов, степени свободы и значение F, рассчитанные в подпараграфе 7.1.2, а также значимость результата.

⁵⁸ Если ввести в это поле несколько переменных, то в рамках одного дисперсионного анализа в SPSS можно проанализировать влияние фактора на несколько зависимых переменных. Однофакторный дисперсионный анализ можно провести также через пункты меню **Анализ — Общая линейная модель — ОЛМ-одномерная (Analyze — General Linear Model — Univariate)**. Интерфейс этого варианта несколько отличается от варианта, который мы использовали. Возможность анализа нескольких зависимых переменных там не предусмотрена, но зато анализируется также сумма квадратов константы.

⁵⁹ В дисперсионном анализе порядковые и даже интервальные переменные в качестве независимых факторов рассматриваются только как *именующие* группы, т.е. как номинативные переменные.

Таблица 7.3(2). Результаты однофакторного дисперсионного анализа в SPSS.

Однофакторный дисперсионный анализ

Score

	Сумма квадратов	ст. св.	Средний квадрат	F	Знач.
Между группами	6.000	2	3.000	1.500	0.296
Внутри групп	12.000	6	2.000		
Итого	18.000	8			

Пример 7.3(2). Откройте файл *EduCondition2.sav*, где находится второй вариант примера подпараграфа 7.1.2 (с удвоенной выборкой). Повторите выборы пунктов меню, описанные выше (кроме окончательного ОК).

При проведении дисперсионного анализа в SPSS можно также установить дополнительные параметры. В меню **Параметры (Options)** поставьте галочки в соответствующих квадратиках, чтобы

- задать вывод описательной статистики зависимых переменных по уровням фактора — пункт **Описательные (Descriptive)**;

- задать вывод графика средних значений зависимой переменной на различных уровнях независимой переменной — пункт **График средних (Means Plot)**;

- задать проверку равенства дисперсий у сравниваемых групп — пункт **Проверка однородности дисперсий (Homogeneity of variance test)**.

Выберите также пункт **Апостериорные (Post hoc)** и поставьте галочку в квадратик, помеченный **Бонферрони (Bonferroni)**, который задает вид поправки для значимости при множественных сравнениях (см. 7.1.4).

После нажатия **ОК** перед основной таблицей мы получаем таблицу описательных статистик, где приводятся по каждой группе средние, стандартные отклонения, стандартные ошибки и доверительные интервалы для среднего. Она организована так же, как аналогичная таблица для Т-критерия. После нее выводится таблица с результатами теста однородности дисперсий Ливиня — также аналогичного употреблявшемуся в расчете Т-критерия. (О последствиях нарушения равенства дисперсий см. 6.2.2 и 10.2.3).

В основной таблице суммы квадратов и степени свободы изменились по сравнению с предыдущей версией примера.

Упражнение 7.3(3). Найдите и объясните изменения. В случае затруднения вернитесь к п. 7.1.2.

Таблицу 7.3(4), в которой приведены результаты апостериорных сравнений разберем подробнее.

Таблица 7.3(4). Результаты множественных сравнений в SPSS.

Множественные сравнения

Зависимая переменная: Score

Бонферрони

(I) Condition	(J) Condition	Разность средних (I-J)	Стд. Ошибка	Знч.	95% доверительный интервал	
					Нижняя граница	Верхняя граница
1.00	2.00	1.00000	0.73030	0.573	-0.9672	2.9672
	3.00	2.00000*	0.73030	0.046	0.0328	3.9672
2.00	1.00	-1.00000	0.73030	0.573	-2.9672	0.9672
	3.00	1.00000	0.73030	0.573	-0.9672	2.9672
3.00	1.00	-2.00000*	0.73030	0.046	-3.9672	-0.0328
	2.00	-1.00000	0.73030	0.573	-2.9672	0.9672

* Разность средних значима на уровне 0.05.

Первая строка значащей части таблицы показывает, что сравниваются первая и вторая группы, разность средних составляет единицу (т.е. среднее в первой группе на единицу больше среднего во второй), стандартная ошибка разности, которую получают усреднением стандартных ошибок по группам, с увеличивающей поправкой (в данном случае поправкой Бонферрони), затем поправленная значимость и поправленный доверительный интервал. Если значимость оказывается меньше 0.05, разность средних помечается звездочкой (в нашем случае помечены вторая и пятая строки, где сравниваются первая и третья группы). Остальные строки интерпретируются аналогично.

Наконец, после таблиц выводится график средних, понимание которого также не вызывает затруднений.

Апостериорные сравнения обычно, но не обязательно, вводятся в дело, когда дисперсионный анализ уже дал приемлемую значимость различий средних в целом по группам. Поправки учитывают количество сравнений, проводимых при множественном сравнении. В списке методов апостериорных сравнений, которые выводятся в соответствующем окне SPSS, предлагается 18 методов, в тонкости различий между которыми мы не будем здесь входить. Более подробную информацию о различных методах попарных сравнений можно найти, например, в книге «Дисперсионный анализ в экспериментальной психологии» А.Н. Гусева (Гусев, 2000).

Упражнение 7.3(5). Проведите сравнение первой и третьей групп с помощью Т-критерия и убедитесь, что апостериорные сравнения оценивают значимость различия как более слабую.

Пример 7.3(6). Зависимость результатов дисперсионного анализа от внутригрупповой дисперсии.

Рассмотрим модельные результаты следующего эксперимента.

Исследовалось влияние просмотренного фильма на эмоциональное состояние испытуемых. Трем группам испытуемых были показаны три коротких видеоролика: первой группе демонстрировался небольшой научно-популярный фильм, второй группе — лирическая зарисовка на тему жизни в деревне, третьей — репортаж из криминальных новостей.

Исследование проводилось два раза: в одном случае испытуемые были обычные люди (данные в файле *emotion1.sav*), во втором отбирались испытуемые с эмоциональной нестабильностью (данные в файле *emotion2.sav*). Эмоциональное состояние измерялось с помощью методики самооценки эмоциональных состояний⁶⁰. Два файла организованы одинаково: первый столбец содержит информацию о принадлежности испытуемого к одной из трех экспериментальных групп, второй столбец — результат проведения методики оценки эмоционального состояния, большие баллы соответствуют лучшему эмоциональному состоянию.

Упражнение 7.3(7). Проведите дисперсионный анализ обоих файлов, выведя также описательные статистики. Сравните (внутригрупповые) стандартные отклонения в двух случаях, суммы квадратов и полученные в результате F-отношения. Сопоставьте средние по группам в двух случаях. Объясните различия самостоятельно, а затем проконтролируйте себя, прочитав ответ.

Ответ. В первом случае стандартное отклонение в каждой из трех групп равно 1.33, сумма квадратов ошибки равна 48, $F(2, 27) = 22.5$ ($p < 0.001$).

Во втором случае стандартное отклонение в каждой из групп равно 4.02 (в три раза больше, чем в первом), сумма квадратов ошибки равна 432 (в девять раз больше), $F(2, 27) = 2.466$ (в девять раз меньше, значимость ухудшилась: $p = 0.104$). Эмоционально нестабильные испытуемые дают больший разброс результатов измерений, поэтому эффект воздействия более «зашумлен» и труднее выделяется.

Исследовательский дизайн. Какие данные подходят для обработки дисперсионным анализом?

⁶⁰ А. Уэссман и Д. Рикс / Сонин В.А. Психодиагностическое познание профессиональной деятельности. — СПб., 2004. С.94-96

Приводимые далее примеры не сопровождаются файлами данных. В каждом случае прочитайте описание исследования. Вам далее предлагается

- a. Сформулировать гипотезу исследования.
- b. Обосновать, *применим ли* дисперсионный анализ для проверки гипотез данного типа.
- c. Выявить независимую и зависимую (зависимые) переменные. Определить, сколько уровней имеет независимая переменная.

Пример 7.3(8). В некотором вузе исследовался уровень самооценки студентов, обучающихся на различных факультетах. В исследовании принимали участие 4 группы студентов: учащиеся психологического, химического, математического и филологического факультетов. У всех этих испытуемых был измерен уровень самооценки с помощью стандартной методики Дембо-Рубинштейн в модификации Прихожан⁶¹.

Конкретная (и самая очевидная) эмпирическая гипотеза, которую можно сформулировать в рамках такого исследования, может быть, например, такой: студенты, обучающиеся на различных факультетах, имеют различный уровень самооценки по шкалам «умный-глупый» и «способный-неспособный». Уровень самооценки в данной методике измеряется в баллах от 0 до 100, эти шкалы можно считать интервальными. Параметр «принадлежность к факультету» является независимой переменной или фактором, измеренным в шкале наименований.

Пример 7.3(9). Исследовалось влияние типа личностного расстройства на понимание эмоций других людей. В качестве испытуемых были отобраны здоровые люди, а также люди с истерическим расстройством и с шизоидным расстройством личности. Восприятие эмоций испытуемыми оценивалось с помощью опросника Н. Холла⁶², по шкале «Распознавание эмоций других людей» (предполагаем, что шкала интервальная).

Пример 7.3(10). Исследовалось влияние конфликтов с родителями на уровень тревожности подростков. По частоте конфликтов подростков разделяли на тех, у кого практически не бывает конфликтов с родителями, тех, у кого они бывают иногда, и тех, кто конфликтует постоянно. Уровень тревожности оценивался при помощи опросника Спилбергер-Ханина (предполагаем, что шкала интервальная).

⁶¹ см., например, Прихожан А. М. Диагностика личностного развития детей подросткового возраста. — М.: АНО «ПЭБ», 2007.

⁶² см., например, Фетискин Н.П., Козлов В.В., Мануйлов Г.М. Социально-психологическая диагностика развития личности и малых групп. — М., Изд-во Института Психотерапии. 2002. — С.57-59.

Пример 7.3(11). В стране N исследовалось, как влияет пол избирателей (мужчины в отличие от женщин) на их предпочтения на выборах президента (в президенты баллотировались три кандидата). Одной из переменных исследования был пол, указанный в анкете, а второй — кандидатура, за которую этот человек голосовал на выборах.

Пример 7.3(12). Множественное сравнение большого количества групп.

Целью данного задания является демонстрация неизбежности искажения оценки значимостей и, как следствие, ложных выводов при использовании множественных попарных сравнений независимых групп с помощью Т-критерия Стьюдента. В качестве основы для проведения статистического эксперимента в рамках данного задания следует использовать шаблон для генерации выборки случайных чисел, содержащийся в файле *20groups.sav*.

Упражнение 7.3(12). Сначала следует сгенерировать данные для проведения анализа. Для этого необходимо в меню вызвать **Преобразовать — Вычислить переменную (Transform — Compute)**. В появившемся окне в качестве имени переменной следует указать *test*, а в окне формулы расчета вставить функцию генерации нормально распределенной случайной величины с показателями среднего и стандартного отклонения 100 и 15 соответственно (в окне **Группы функций (Group of function)** выбрать **Случайных чисел (Random numbers)** и в появившемся ниже списке выбрать функцию **RV.NORMAL** и переместить ее в окно **Числовое выражение (Numerical expression)**, где в образовавшиеся аргументные места (помеченные вопросительными знаками) вставить 100 и 15, чтобы получилось **RV.NORMAL(100, 15)**). После этого надо нажать **OK** и подтвердить замену переменной. В столбце *test* появятся значения нормально распределенной случайной величины со средним значением 100 и стандартным отклонением 15. В первом столбце (переменная *group*) содержатся 20 групп, обозначенных числами от 1 до 20. Вместе с полученными случайными значениями это соответствует примерно следующей ситуации: мы измерили коэффициент интеллекта у пассажиров 20 вагонов метро (в каждом вагоне — по 50 человек).

После генерации выборки проведите однофакторный дисперсионный анализ с группирующей переменной *group* и зависимой *test*, задав опцию **Описательные (Descriptives)** в пункте **Параметры (Options)**. В результате вы получите средние значения по всем 20 группам и общую оценку значимости результата. В нашем примере общая оценка значимости различий между средними значениями переменной *test* в 20 группах равна 0.628, а минимальное и максимальное значение среднего содержатся соответственно во второй и пятой группах.

Проведите теперь сравнение по Т-тесту групп с минимальным и максимальным значением среднего (в наших данных это — вторая и пятая группы). Убедитесь, что значимость отличия меньше 0.05 (в нашем случае 0.009)⁶³.

Если принять легенду с измерением интеллекта в вагонах метро, со-держательно этот результат может быть ошибочно интерпретирован так: в пятом вагоне едут более умные люди, чем во втором.

Чтобы избежать подобных ошибок, в таких ситуациях следует провести дисперсионный анализ с последующим апостериорным сравнением. Добавьте опцию *Апостериорные сравнения (Post Hoc)* к дисперсионному анализу и убедитесь в том, что ошибка не повторилась.

Заметим в заключение: поскольку программа SPSS использует генератор *псевдослучайных* чисел, а не реально случайных (для наших целей разницы нет), то в первом эксперименте вы, вероятно, получите те же числа, что и мы. Если у вас остаются сомнения, повторите присвоение случайных значений. Вы получите уже новые числа и сможете убедиться, что ошибка попарных сравнений снова воспроизвелась.

⁶³ Вообще говоря, мы не можем гарантировать, что вы получите значимость меньше 0.05. С очень маленькой вероятностью случайно вы можете получить и больший результат.

Глава 8. _____

Двухфакторный дисперсионный анализ

8.1. Двухфакторный дисперсионный анализ. Примеры и качественные разъяснения

Умение «читать» данные в табличном и графическом представлении — важное профессиональное качество психолога. В первом параграфе главы, посвященной двухфакторному дисперсионному анализу, мы сосредоточимся на этом аспекте работы с данными, оставив математические подробности второму параграфу.

8.1.1. Двухфакторная схема эксперимента. Пример сложения воздействий

Мы рассмотрим результаты эксперимента по восприятию кратковременно предъявляемых слов. Испытуемым на экране монитора последовательно предъявлялись слова из одного и того же набора в случайном порядке. Каждое слово появлялось четыре раза в четырех различных вариантах условий, образованных парами значений независимых переменных (факторов).

Фактор 1 — длительность предъявления (два уровня: 100 мс и 200 мс),

Фактор 2 — контрастность изображения (два уровня: низкая и высокая).

Зависимая переменная — процент правильно опознанных слов.

Средние значения по всему набору слов задаются таблицей

Таблица 8.1.1(1)

Пример сложения воздействий в двухфакторном дизайне

Процент правильных ответов		
	Длительность 100 мс	Длительность 200 мс
Низкая контрастность	10	15
Высокая контрастность	20	25

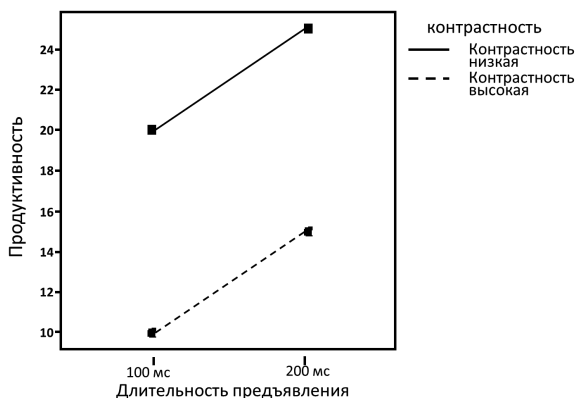


Рис. 8.1.1(2). Групповые средние для двух уровней двух факторов для примера подпараграфа 8.1.1.

В данном случае факторы действуют независимо друг от друга. При любом значении длительности предъявления увеличение контрастности прибавляет 10% правильных ответов. Также и независимо от контрастности увеличение длительности предъявления прибавляет 5% правильных ответов. На практике такая точность сложения воздействий, разумеется, не встречается.

Статистический пакет SPSS выводит средние групповые значения в виде графика (рис. 8.1.1(2)). Независимость действия факторов проявляется на графике параллельностью линий – расстояние между линиями одинаково для обоих значений аргумента «длительность».

8.1.2. Двухфакторная схема эксперимента. Пример сильного взаимодействия

Факторы предыдущего примера измерялись явно в шкале порядка или более сильной шкале интервалов: по крайней мере, вполне осмысленно говорить о большей контрастности и большем времени экспозиции. Для дисперсионного анализа достаточно, чтобы фактор измерялся в номинативной шкале, где отношения «больше-меньше» не имеют смысла.

Рассмотрим следующий пример. В учебном заведении, готовящем специалистов для текстильной промышленности, проводилось исследование связи креативности, которая измерялась с помощью теста Гилфорда, с доминантностью мозговых полушарий. Результаты свидетельствовали об отсутствии преимуществ в креативности как у левшей, так и у правшей. Однако при принятии во внимание специализации обследованных студентов картина оказалась и более сложной, и значительно более интересной. В каждой клетке таблицы представлены средние результаты десяти испытуемых.

Таблица 8.1.2(1)

Пример сильного воздействия в двухфакторном дизайне

Показатель креативности		
	Специальность «художник»	Специальность «механик»
Левши	10	20
Правши	20	10

Обратим внимание, что средние значения креативности равны 15 баллам как у левшей, так и у правой; как у механиков, так и у художников (среднее значение, например, креативности у художников – это среднее арифметическое чисел в соответствующем столбце таблицы; аналогично для других средних). Однако мы видим, что художники-левши оказываются менее креативными, чем их праворукие коллеги, а у механиков картина как раз обратная: правши менее креативны, чем левши.

В подобных случаях говорят о сильном взаимодействии факторов. Факторы доминантности и специализации по отдельности не влияют заметно на средние значения креативности (однофакторный дисперсионный анализ, как и другие статистические критерии, указал бы на полное отсутствие влияния), но оказывается, что при совместном рассмотрении обоих факторов обнаруживаются важные зависимости: более тщательный анализ тестового материала показал, что художники и механики добиваются одинаковых количественных показателей теста креативности разными средствами, что и дает преимущество в одном случае левшам, а в другом – правшам. На графике это сильное взаимодействие факторов при отсутствии воздействия факторов в среднем выглядит так, как показано на рис. 8.1.2(2).

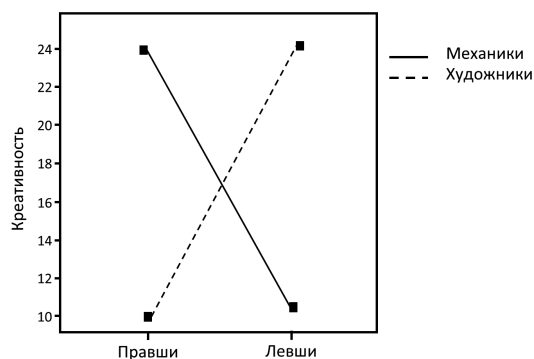


Рис. 8.1.2(2). Групповые средние для двух уровней двух факторов для примера в подпараграфе 8.1.2.

8.1.3. Двухфакторная схема эксперимента. Комбинация влияния факторов и взаимодействия. Интерпретация

В исследовательской практике чаще всего приходится встречаться с наложением более или менее сильного прямого влияния факторов и их взаимодействия. В данном примере описано исследование влияния силы мотивации студентов и опыта преподавателей на учебные результаты студентов. В начале семестра оценивалась мотивация студентов. Студенты были в случайном порядке поделены между тремя учебными группами так, чтобы в каждой учебной группе было равное количество студентов с низкой, средней и высокой мотивацией к обучению.

Три преподавателя — начинающий, обладающий средним опытом преподавания и опытный преподаватель — вели занятия каждый в своей группе. В конце семестра оценили знания студентов по предмету по десятибалльной шкале. Средние оценки по группам изображены на графике (рис. 8.1.3(1)).

Как мы говорили, способность «читать» подобные графики составляет важную часть умений психолога — исследователя и практика. Прежде чем перейти к чтению наших интерпретаций, попытайтесь дать интерпретацию самостоятельно.

По графику можно видеть, во-первых, что студенты с более высокой мотивацией в среднем учатся лучше при любом преподавателе. Во-вторых, опытный преподаватель достигает в среднем более высоких результатов при любой мотивации студентов. Эти результаты вполне согласуются с ожиданиями. В-третьих, весьма интересный результат дает взаимодействие факторов: можно видеть, что при низкой мотивации изменение квалификации преподавателя от низкой к средней и от средней к высокой изменяет результат в среднем на один балл (не желающему учиться студенту помочь трудно). То же самое можно сказать и влиянии квалификации на результат высоко мотивированных студентов (они достигают хороших результатов собственными усилиями). При средней мотивации квалификация преподавателя очень существенна и ее влияние на среднюю оценку студентов более чем в два раза сильнее, чем при низкой и высокой мотивации.

Определение 8.1.3(2). Если на разных уровнях первого фактора второй фактор оказывает на зависимую переменную различное воздействие, то говорят о *влиянии взаимодействия факторов*⁶⁴.

⁶⁴ Мы дали не вполне корректную формулировку с целью сделать ее максимально простой и понятной. На самом деле говорить о воздействии или влиянии факторов с полным правом можно только в ситуации причинной связи между независимой и зависимой переменными. Поскольку обоснование причинной

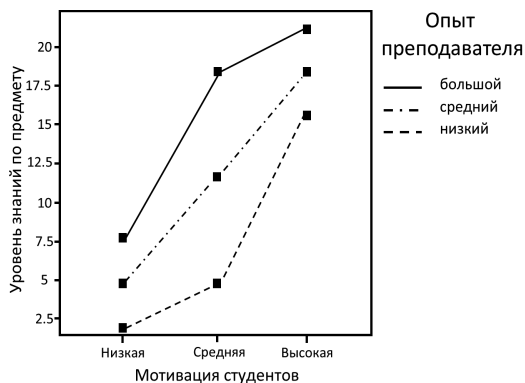


Рис. 8.1.3(1). Групповые средние для трех уровней трех факторов для примера в подпараграфе 8.1.3.

Не будет преувеличением сказать, что самые интересные результаты в экспериментальных психологических исследованиях, использующих двухфакторный дисперсионный анализ, есть результаты о взаимодействии факторов.

Далее мы разберем несколько примеров. Результаты будут даны в графическом виде, и в данном параграфе их интерпретация будет чисто качественной, без учета значимости. Параллельно эти же примеры будут разобраны в практикуме данной главы. Мы рекомендуем, если это возможно, выполнять задания практикума непосредственно после чтения текстов примеров в данном параграфе.

Отметим для читателей, которым сложность дальнейшего изложения покажется превышающей их силы, что принцип оценивания значимостей не отличается от аналогичного принципа для однофакторного дисперсионного анализа: воздействие факторов, как и их взаимодействие, будет характеризоваться подобранными показателями дисперсии, рост которых с очевидностью свидетельствует в пользу наличия соответствующего влияния фактора или взаимодействия. Эти дисперсии будут делиться на внутригрупповую дисперсию, которая остается постоянной как при наличии влияния, так и при его отсутствии. Результаты оцениваются в сопоставлении с соответствующими квантилями распределения Фишера. На выходе даются оцен-

связи невозможно сделать чисто статистически, то в общем случае в последнее время специалисты по математической психологии предлагают использовать более аккуратные слова, которые не вводят в заблуждение: «эффект» и «вклад» вместо «влияние» и «воздействие».

ки значимости влияния отдельно по каждому из двух факторов и взаимодействию между ними.

8.1.4. Пример. Доминантность

В книге Е.В. Сидоренко (Сидоренко, 2000) описано исследование уровня доминантности у взрослых мужчин и женщин. Кроме пола, исследователи учитывали порядок рождения испытуемых в семье: старшие дети, средние дети и младшие дети. Авторы предполагали, что доминантность должна быть выше у людей, которые были первенцами в своих семьях, и ниже у средних и тем более младших детей. Однако результат оказался более интересным (рис. 8.1.4(1)). Как и в предыдущем случае, рекомендуем переходить к нашей интерпретации после самостоятельной попытки.

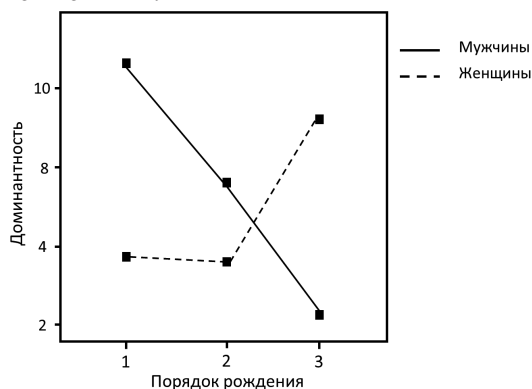


Рис. 8.1.4(1). Групповые средние для трех уровней трех факторов для примера в подпараграфе 8.1.4.

Гипотеза исследователей оправдалась только наполовину. Доминантность ожидаемо больше у мужчин и падает у более поздних детей — за одним очень интересным исключением. Третья по рождению девочка оказывается более доминантной, уступая только мальчику-первенцу. Интерпретацию этого факта мы оставляем читателю.

8.1.5. Пример. Закон Йеркса-Додсона

Один из самых интересных психологических законов гласит: наилучшее выполнение задания обеспечивается не сильной и не слабой, а некоторой средней мотивацией.

Вот пример, взятый из работы «Экспериментальная психология»⁶⁵. Приведены результаты исследования, проведенного на крысах. Корот-

⁶⁵ Фресс П., Пиаже Ж. (ред.), Экспериментальная психология. — Вып. 5. — М.: Прогресс, 1975/ — С. 119-125

ко схему исследования можно описать так. Задача животных состояла в различении двух объектов различной яркости, при этом один из ответов оценивался как неправильный. Выделялись три различных уровня сложности различения объектов. Кроме того, предусматривались три уровня мотивации, а именно слабый, средний или сильный электрический удар за ошибки. Общие результаты, полученные в исследованиях, представлены на рис. 8.1.5(1), взятом из упомянутой работы⁶⁶.



Рис. 8.1.5(1). Результат исследования Йеркса-Додсона. По оси абсцисс отложена сила мотивации (сила удара током). По оси ординат — количество проб, понадобившихся для формирования навыка различения

По рисунку видно, что оптимальная мотивация лежит в середине диапазона, причем для сложных задач оптимальна более слабая мотивация, чем для простых.

8.2. Расчеты статистических оценок

8.2.1. Пример расчета межгрупповых сумм квадратов

Вернемся к примеру из подпараграфа 8.1.3. Полные модельные данные представлены в таблице, где в каждой клетке по три результата испытуемых.

⁶⁶ В задании практикума 8.3(6) будут представлены модельные данные и построен график в формате SPSS.

	Начинающий преподаватель	Преподаватель со средним опытом	Преподаватель с большим опытом
Низкая мотивация студента	1	3	4
	2	4	6
	3	5	8
Средняя мотивация студента	3	7	13
	4	9	14
	5	11	15
Высокая мотивация студента	10	13	15
	12	14	16
	14	15	17

Далее мы оставляем в таблице только числа без заголовков строк и столбцов. Как и в случае однофакторного анализа, сначала вычитаем общее среднее (которое здесь нас не интересует), равное 9, и раскладываем оставшуюся таблицу в сумму модели и ошибки:

$$\begin{array}{|c|c|c|} \hline -8 & -6 & -5 \\ \hline -7 & -5 & -3 \\ \hline -6 & -4 & -1 \\ \hline -6 & -2 & 4 \\ \hline -5 & 0 & 5 \\ \hline -4 & 2 & 6 \\ \hline 1 & 4 & 6 \\ \hline 3 & 5 & 7 \\ \hline 5 & 6 & 8 \\ \hline \end{array} = \begin{array}{|c|c|c|} \hline -7 & -5 & -3 \\ \hline -7 & -5 & -3 \\ \hline -7 & -5 & -3 \\ \hline -5 & 0 & 5 \\ \hline -5 & 0 & 5 \\ \hline -5 & 0 & 5 \\ \hline 3 & 5 & 7 \\ \hline 3 & 5 & 7 \\ \hline 3 & 5 & 7 \\ \hline \end{array} + \begin{array}{|c|c|c|} \hline -1 & -1 & -2 \\ \hline 0 & -0 & 0 \\ \hline 1 & 1 & 2 \\ \hline -1 & -2 & -1 \\ \hline 0 & 0 & 0 \\ \hline 1 & 2 & 1 \\ \hline -2 & -1 & -1 \\ \hline 0 & 0 & 0 \\ \hline 2 & 1 & 1 \\ \hline \end{array}$$

В первой таблице правой части вместо индивидуальных показателей ставится среднее по соответствующей клетке, во второй таблице правой части — разница между индивидуальным значением и средним по соответствующей клетке (т.е. средним по данной группе). Соответствующее равенство сумм квадратов $684 = 648 + 36$ ($S_{m+e} = S_{model} + S_{error}$). Правая таблица (*таблица ошибки*) дает нам сумму квадратов остатка, которая пойдет в знаменатели F-отношений, а среднюю таблицу раскладываем в три таблицы, характеризующие разные аспекты вклада факторов.

-7	-5	-3
-7	-5	-3
-7	-5	-3
-5	0	5
-5	0	5
-5	0	5
3	5	7
3	5	7
3	5	7

 $=$

-3	0	3
-3	0	3
-3	0	3
-3	0	3
-3	0	3
-3	0	3
-3	0	3
-3	0	3
-3	0	3

 $+$

-5	-5	-5
-5	-5	-5
-5	-5	-5
0	0	0
0	0	0
0	0	0
5	5	5
5	5	5
5	5	5

 $+$

1	0	-1
1	0	-1
1	0	-1
-2	0	2
-2	0	2
-2	0	2
1	0	-1
1	0	-1
1	0	-1

Разложение сумм квадратов для этих таблиц $648 = 162 + 450 + 36$ ($S_{model} = S_{f_1} + S_{f_2} + S_{f_1 \cdot f_2}$).

Первая таблица в правой части равенства, описывающего это разложение, это *таблица вклада фактора опыта преподавателя*, содержащая средние по столбцам разлагаемой таблицы. Чтобы ее построить, рассчитываются средние по каждому столбцу (для каждого опыта преподавания) и ставятся в этот столбец. Эта таблица нужна для расчета основного эффекта опыта преподавания.

Следующее слагаемое — *таблица вклада фактора мотивации студента* — содержит средние по строкам. Для нее рассчитываются средние по каждой строке (для каждого уровня мотивации) и ставятся в эту строку. На основе этой таблицы рассчитывается основной эффект фактора мотивации студентов.

Третья *таблица взаимодействия факторов* содержит числа, которые нужны для равенства правой и левой частей. Например, -7 в верхнем левом углу левой таблицы разлагается на -3 первой таблицы, -5 второй таблицы и дополняющую 1 третьей таблицы в правой части. Эта таблица используется для расчета эффекта взаимодействия. Она может выглядеть довольно сложно. Единственное условие для проверки — суммы по всем строкам и по всем столбцам должны быть равны нулю.

Как и в примечании подпараграфа 7.1.2, можно заметить, что если каждую, состоящую из 27 элементов таблицу представить 27-мерным вектором, то получится векторное равенство, причем попарные скалярные произведения векторов в правой части все равны нулю, что означает, что они ортогональны. При разложении вектора по ортогональным направлениям квадрат длины гипотенузы равен сумме квадратов обобщенных катетов (теорема Пифагора в многомерном случае), а это и есть наше разложение сумм квадратов.

Именно слагаемые в правой части и будут характеризовать значимости влияния соответствующих факторов и их взаимодействия, попадая в числитель соответствующего F-отношения. Число степеней свободы в каждом случае (только для задачи 3*3) равно соответственно, 2, 2 и 4 (подробнее чуть ниже). В знаменателе всех трех дробей будет стоять сумма квадратов таблицы ошибки, деленная на число степеней свободы, которое равно $(3-1)*9 = 18$. Непростая теорема математической статистики говорит, что вклады каждой из трех таблиц могут оцениваться независимо, каждая по отношению к своей нулевой гипотезе о равенстве средних.

Для действия на средние отдельно опыта преподавателя

$$F_{18}^2 = \frac{162/2}{36/18} = 40.5, \quad p < 0.000001.$$

Для действия на средние отдельно мотивации студентов

$$F_{18}^2 = \frac{450/2}{36/18} = 112.5, \quad p < 0.000001.$$

Для влияния взаимодействия факторов

$$F_{18}^4 = \frac{36/4}{36/18} = 4.5, \quad p = 0.011.$$

(p во всех трех случаях измеряет вес хвоста соответствующего распределения Фишера, который отсекается данным значением F).

Сопоставление сумм квадратов показывает, что факторы дают каждый в отдельности весьма существенный вклад в уровень знаний студентов (мотивация более важна, чем опыт преподавателя). Взаимодействие факторов дает менее существенный вклад, но психологический смысл этого взаимодействия более интересен, чем вполне предсказуемый вклад факторов.

8.2.2. Расчеты степеней свободы

Как распределяются степени свободы по трем таблицам, очень просто разъяснить на примере. Пусть первый фактор в наборе данных имеет три уровня, а второй четыре уровня. Для расчета нам понадобится таблица:

Общее среднее	Фактор 1	Фактор 1
Фактор 2	взаимодействие	взаимодействие
Фактор 2	взаимодействие	взаимодействие
Фактор 2	взаимодействие	взаимодействие

Пересчитывая клетки, получаем для первого фактора 2 степени свободы, для второго — 3 степени свободы и для взаимодействия — 6 степеней свободы.

8.3. Двухфакторный дисперсионный анализ в SPSS. Практикум

Рассмотрим пример использования двухфакторного дисперсионного анализа в психологическом исследовании.

Пример 8.3(1). В файле *TeacherExp.sav* содержатся данные, моделирующие результаты исследования, описанного в подпараграфе 8.1.3.

В файле данных первый столбец (*TeacherExp*) содержит информацию об опыте преподавателя, второй (*StudMotivation*) — о мотивации студента и, наконец, третий — балл, характеризующий уровень знаний после прослушивания курса.

Теперь мы опишем, как проводится двухфакторный дисперсионный анализ в SPSS.

Для вызова этой процедуры следует выбрать пункты меню *Анализ* — *Общая линейная модель* — *ОЛМ-одномерная (Analyze — General Linear Model — Univariate)*. В появившемся диалоговом окне в поле *Зависимая переменная (Dependent Variable)* внесем нашу зависимую переменную (*Results*), а в поле *Фиксированные факторы (Fixed Factor(s))* два фактора: *TeacherExp* и *StudMotivation*⁶⁷.

Чтобы построить графики, выбираем *Графики (Plots)*, затем в поле *Горизонтальная ось (Horizontal Axis)* вносим переменную *TeacherExp*. После этого нажимаем кнопку *Добавить (Add)*. Этими действиями задается график вклада фактора опыта преподавателя. Для отображения вклада второго фактора в поле *Горизонтальная ось (Horizontal Axis)* вносим переменную *StudMotivation*, после чего нажимаем кнопку *Добавить (Add)*. И наконец, для построения графика, отображающего взаимодействие факторов, в поле *Горизонтальная ось (Horizontal Axis)* вносим переменную *StudMotivation*, а в поле *Отдельные линии (Separate Lines)* — переменную *TeacherExp*, после чего нажимаем кнопку *Добавить (Add)*. Для завершения процедуры задания графиков нажмем кнопку *Продолжить (Continue)*.

⁶⁷ В поле «Фиксированные факторы» технически можно внести и большее количество переменных (в SPSS до четырех), однако на практике в дисперсионный анализ редко включается больше трех факторов. Действительно, уже в случае трех независимых переменных придется рассмотреть семь возможных влияний (каждой переменной в отдельности, попарных взаимодействий и взаимодействия всех факторов). В случае же четырех факторов результаты статистической обработки становится крайне сложно интерпретировать. В рамках нашего учебника мы будем рассматривать примеры только с двумя независимыми переменными, однако при увеличении их числа результаты интерпретируются аналогично.

Заметим, что график взаимодействия факторов можно задать с переменной *TeacherExp* в поле **Отдельные линии**, а переменную *StudMotivation* поместить в поле **Горизонтальная ось**. Вся нужная информация содержится в обоих графиках, но иногда они не одинаково наглядно ее представляют. Можно посоветовать рисовать сначала оба графика, а затем выбрать более наглядный для представления результатов.

Если теперь нажать **OK** и запустить расчет, мы получим следующие таблицы и графики.

Первая таблица «Межгрупповые факторы» содержит информацию о количестве факторов, их уровнях и количестве испытуемых на каждом из уровней фактора. Она не нуждается в пояснениях.

Вторая таблица (8.3(2)) содержит информацию о суммах квадратов и значимостях.

Таблица 8.2(2). Результат двухфакторного дисперсионного анализа в SPSS

Оценка эффектов межгрупповых факторов

Зависимая переменная: Уровень знаний по предмету

Источник	Сумма квадратов типа III	ст. св.	Средний квадрат	F	Знч.
Скорректированная модель	648.000 ^a	8	81.000	40.500	0.000
Свободный член	2187.000	1	2187.000	1093.500	0.000
TeacherExp	162.000	2	81.000	40.500	0.000
StudMotivation	450.000	2	225.000	112.500	0.000
TeacherExp * StudMotivation	36.000	4	9.000	4.500	0.011
Ошибка	36.000	18	2.000		
Всего	2871.000	27			
Скорректированный итог	684.000	26			

^a. R квадрат = 0.947 (Скорректированный R квадрат = 0.924)

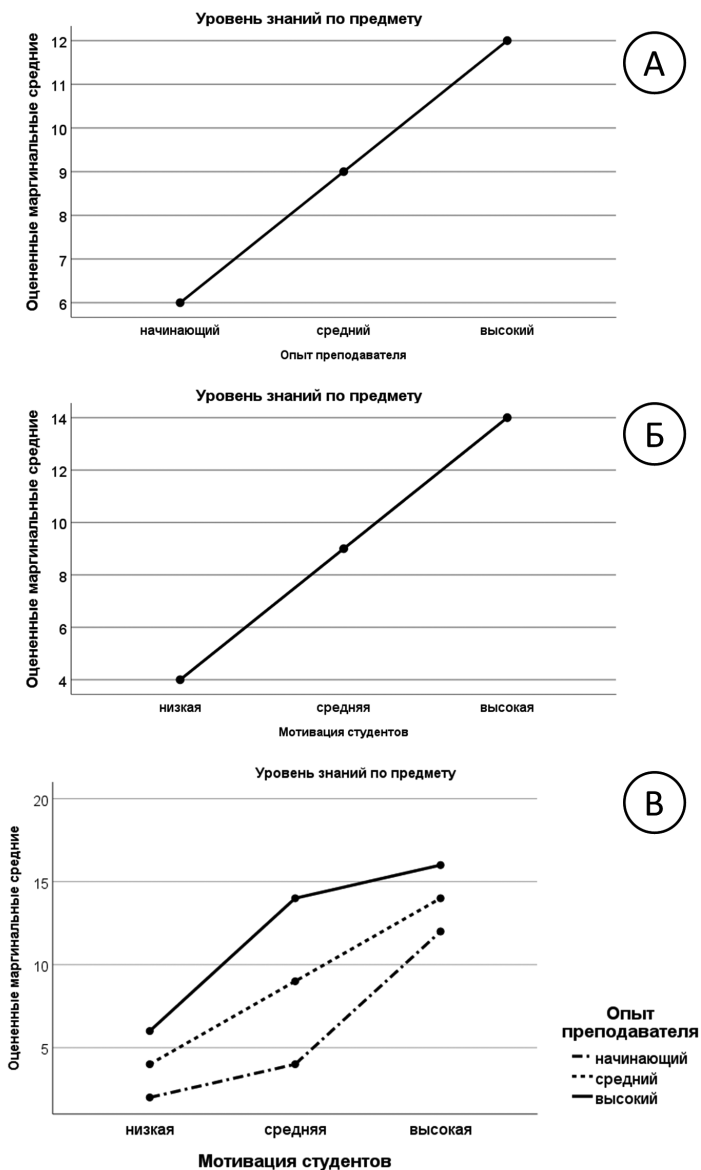


Рис. 8.3(3). Графики среднего для влияния факторов по отдельности (А, Б) и их взаимодействия (В).

Строка в таблице, описывающая значимость по переменной *TeacherExp*, соотносится с первым графиком аналогично однофакторному дисперсионному анализу. Надо только иметь в виду, что каждая точка графика отображает усредненные по всем значениям второго фактора *StudMotivation*, результаты в данной точке. Следующая строка относится ко второму графику, а третья строка, подписанная *TeacherExp*StudMotivation* относится к третьему графику. Интерпретацию этого графика мы давали в подпараграфе 8.1.3.

Можно добавить к выводу апостериорные сравнения (по каждому фактору в отдельности) аналогично однофакторному анализу, а с помощью дополнительного окна, вызываемого кнопкой **Параметры (Options)** можно задать расчеты различных показателей зависимой переменной на разных уровнях факторов. Чтобы задать расчет средних значений по всем параметрам на всех уровнях обоих факторов, надо отметить пункт **Описательная статистика (Descriptive Statistics)** и перенести «переменную» (*OVERALL*) в поле **Вывести средние для (Display Means for)**. Выводимая таблица средних проста и не нуждается в пояснениях.

Пример 8.3(4). В файле *Dominance.sav* представлены данные подпараграфа 8.1.4.

Упражнение 8.3(5). Проведите двухфакторный дисперсионный анализ. Нарисуйте графики. Найдите способ повторить график, изображенный на рис. 8.1.4(1). Найдите соответствующую каждому графику значимость результата.

Пример 8.3(6). В файле *YerksDodson.sav* приведены модельные данные для эксперимента, описанного в подпараграфе 8.1.5.

Упражнение 8.3(7). Проведите двухфакторный дисперсионный анализ и сопоставьте график рисунка 8.1.5(1) и полученный с помощью SPSS. Объясните различия.

Глава 9.

Гипотезы о связи переменных

9.1. Коэффициент корреляции и линейная регрессия

9.1.1. Коэффициент корреляции Пирсона

Задача о связи характеристик изучаемого предмета — одна из наиболее часто встречающихся в психологических исследованиях.

Типичная задача такого рода возникает в исследованиях, где каждый испытуемый характеризуется двумя или большим числом показателей. Пусть, например, 70 детей разного возраста решают задачи некоторого теста на запоминание несложного вербального материала. Результат оценивается в процентах воспроизведенного материала. Каждый испытуемый характеризуется парой чисел — возрастом и оценкой успешности воспроизведения материала, которые позволяют отобразить его характеристики на диаграмме рассеяния (рис. 9.1.1(1)), которая получается следующим образом. На плоскости в системе координат «возраст» (горизонтальная ось) — «успешность воспроизведения» (вертикальная ось) каждый испытуемый отображается точкой, координаты которой — это его показатели по указанным переменным.

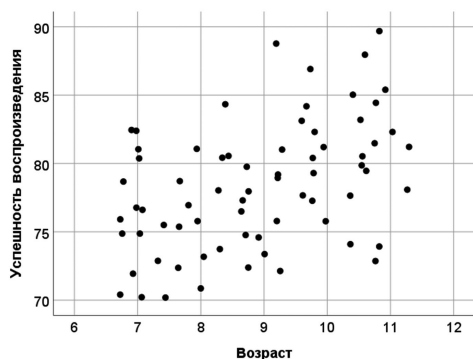


Рис. 9.1.1(1). Диаграмма рассеяния. По оси абсцисс отложен возраст испытуемого, по оси ординат — оценки по тесту

Мы видим по диаграмме, что имеется тенденция повышения результата тестирования при увеличении возраста испытуемого.

Другой пример: те же 70 испытуемых взвешены на весах. Диаграмма для этого показателя и возраста выглядит так (рис. 9.1.1(2)).

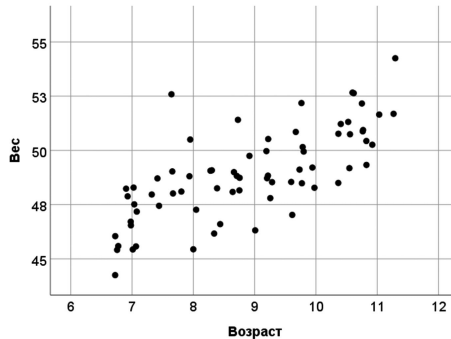


Рис. 9.1.1(2). Диаграмма рассеяния. По оси абсцисс отложен возраст испытуемого, по оси ординат — вес

На этой диаграмме наблюдается естественная тенденция увеличения веса при увеличении возраста испытуемого, однако мы видим различие в степени выраженности этих тенденций на двух диаграммах. Показатели, к изучению которых мы теперь переходим, предназначены для оценивания именно этой степени выраженности. В данной главе мы будем предполагать, что все переменные измерены в интервальной шкале. Поскольку речь идет о связи *переменных*, то показатель этой связи не должен зависеть от единиц, в которых мы измеряем, например, вес: связь веса и возраста должна быть описана независимо от того, измеряется ли вес в килограммах или фунтах. Таким инвариантным показателем связи является *коэффициент корреляции Пирсона*.

Вычислим сначала выборочные оценки \bar{x} , s_x , \bar{y} и s_y по уже знакомым нам формулам

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \dots + x_n) \text{ и } s_x = \sqrt{\frac{1}{n-1}((x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2)}$$

и (и аналогично для y).

Коэффициент корреляции Пирсона вычисляется по формуле

$$r_{xy} = \frac{\frac{1}{n-1}((x_1 - \bar{x})(y_1 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y}))}{s_x s_y}$$

Формулу можно преобразовать:

$$r_{xy} = \frac{1}{n-1} \left(\frac{(x_1 - \bar{x})}{s_x} \frac{(y_1 - \bar{y})}{s_y} + \dots + \frac{(x_n - \bar{x})}{s_x} \frac{(y_n - \bar{y})}{s_y} \right)$$

Если обозначить $x'_i = \frac{(x_i - \bar{x})}{s_x}$, $y'_i = \frac{(y_i - \bar{y})}{s_y}$, (стандартизованные переменные), то формула корреляции принимает простой вид

$$r_{xy} = \frac{1}{n-1} (x'_1 y'_1 + x'_2 y'_2 + \dots + x'_n y'_n)$$

На рис. 9.1.1(3) показаны диаграммы, соответствующие различным значениям коэффициента корреляции. При изменении единиц измерения коэффициент Пирсона не меняется, поэтому мы не указывали единицы измерения на осях.

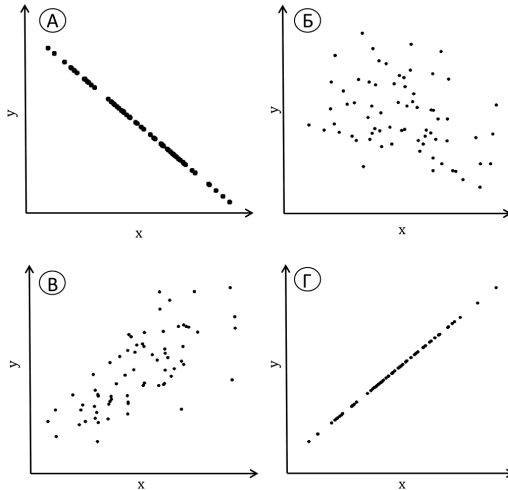


Рис. 9.1.1(3). Диаграммы рассеяния при некоторых значениях коэффициента корреляции Пирсона: а) $r=-1$; б) $r=-0.35$; в) $r=0.7$; г) $r=1$

Самая большая положительная корреляция, равная единице, соответствует линейной связи с положительным коэффициентом $y_i = a \cdot x_i + b$ при $a > 0$ (это не значит, что y зависимая переменная, а x независимая; для корреляции переменные совершенно равноправны; корреляция сама по себе вообще ничего не говорит о причинно-следственных связях между переменными — см. 9.1.6 и 9.3.2(2)). На другом полюсе — равная минус единице корреляция для линейно связанных выборочных значений с отрицательным коэффициентом: $y_i = a \cdot x_i + b$ при $a < 0$. Коэффициент корреляции не может превышать единицу и быть меньше минус единицы. При наличии линейной связи он равен по модулю единице при любых значениях a и b . Если сопоставляемые переменные на самом деле независимы, то выборочный коэффициент корреляции будет приблизительно равен нулю, а облако точек на ди-

аграмме рассеяния будет более или менее округлым или вытянутым вдоль одной или другой оси.

Если в первом примере вместо процента запомненных стимулов характеризовать испытуемых количеством не запомненных, то новая переменная будет связана с исходной соотношением «процент не запомненных стимулов» = 100 — «процент запомненных стимулов». Это допустимое для интервальных шкал преобразование меняет знак коэффициента корреляции, сохраняя его модуль.

Коэффициент корреляции Пирсона измеряет так называемую *линейную связь* переменных, которая на графиках выражается вытянутостью облака точек на диаграмме рассеяния вдоль не параллельной осям прямой линии. Для изображенных на рис. 9.1.1(4) конфигураций коэффициент корреляции Пирсона будет близким к нулю, несмотря на очевидность связи переменных.

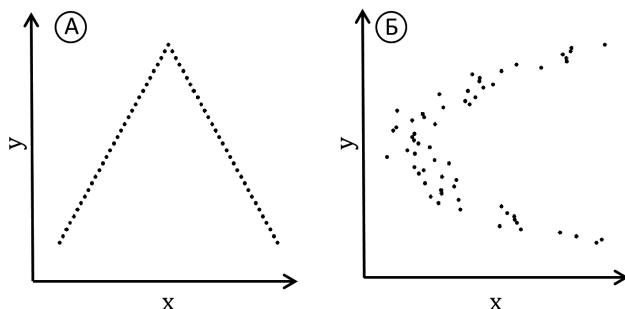


Рис. 9.1.1(4). Примеры пар выборок с очевидной связью значений, но близкой к нулю корреляцией

Числитель формулы корреляции Пирсона иногда представляет отдельный интерес и имеет собственное название — *коэффициент ковариации*:

$$cov_{xy} = \frac{1}{n-1} ((x_1 - \bar{x})(y_1 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y}))$$

При изменении единиц измерения он, в отличие от коэффициента корреляции, изменяется пропорционально выборочным значениям. Соответственно последним, коэффициент ковариации может принимать значения от минус до плюс бесконечности.

9.1.2. Проверка статистических гипотез о связи переменных

Выборочный коэффициент корреляции оценивает подразумеваемую исследователем реальную связь между переменными. Как и в случае

оценки среднего значения, нас интересуют два вопроса: (1) Насколько сильна связь между переменными; (2) Насколько надежна наша оценка. Сила связи между переменными по всей генеральной совокупности существует объективно. Если ее измерять корреляцией, то она будет выражаться числом от -1 до 1 . Выборочная корреляция этих переменных будет колебаться вокруг истинного показателя силы связи. Трудность состоит в том, что, получив выборочную корреляцию, мы не можем знать, ни насколько она отклоняется от истинного значения, ни даже в какую сторону. В случае корреляции оценка обычно выражается в терминах значимости.

Прделаем небольшое упражнение.

Упражнение 9.1.2(1). Возьмите две симметричные монеты достоинством в один рубль и один евро. Проведите серию четырех подбрасываний пары монет и запишите результаты в виде $(x_1, y_1), \dots, (x_4, y_4)$, полагая

$x_i = 0$, если рубль выпал цифрой;

$x_i = 1$, если рубль выпал гербом;

$y_i = 0$, если евро выпал цифрой;

$y_i = 1$, если евро выпал гербом.

Подсчитайте коэффициент корреляции Пирсона. Истинная корреляция между результатами двух монет равна, разумеется, нулю. Повторите процедуру несколько раз и убедитесь, что нулевое значение выборочного коэффициента корреляции выпадает примерно один раз из трех. При многократном повторении опыта можно убедиться, что его результат имеет некоторое *распределение*, симметричное относительно нуля. Это распределение зависит от объема выборки n : чем больше n , тем меньше дисперсия распределения, тем ближе к нулю ее вероятные значения.

В таблице 9.1.2(2) приведены двухсторонние квантили распределения выборочного коэффициента корреляции по Пирсону для $n = 10$. Они рассчитаны для выборок, полученных испытаниями двух нормально распределенных случайных величин, теоретическая корреляция между которыми равна нулю. Дихотомический результат подбрасывания монеты не распределен нормально, однако некоторое представление о возможных результатах наших испытаний табличный квантиль все же дает.

Таблица 9.1.2(2) Двусторонние квантили распределения коэффициента Пирсона для $n = 10$

α	0.05	0.025	0.01	0.005
$r_\alpha(10)$	0.497	0.576	0.658	0.709

Обычно при исследовании связи переменных статистической гипотезой H_0 будет гипотеза об отсутствии связи, т.е. о независимости переменных. Альтернативная гипотеза H_1 (т.е. гипотеза, к которой мы склоняемся, получив большие по модулю значения выборочной корреляции) будет утверждать только наличие связи⁶⁸. Можно оценить значимость относительно данного результата (полученной парной выборки) гипотез о других значениях теоретической корреляции, но это требует некоторых дополнительных усилий (см. подпараграф 9.2.3). Если истинна гипотеза H_0 , то выборочный коэффициент корреляции будет принимать значения, более или менее близкие к нулю. Если выборочная корреляция принимает достаточно большое по модулю значение, которому соответствует значимость, измеряемая маленьким числом, то мы склоняемся к гипотезе H_1 о наличии связи, но без указания точного значения теоретической корреляции.

Можно заметить, что если верна гипотеза H_0 об отсутствии зависимости между случайными величинами, то выборочный коэффициент при $n=10$ может принимать тем не менее довольно большие значения, так что уровень значимости 0.05 для принятия гипотезы о зависимости случайных величин требует, чтобы выборочный коэффициент корреляции достигал почти 0.5 (см. 9.1.2(2)). В связи с этим надо иметь в виду, что даже выборочная корреляция, например 0.6, вполне может согласовываться с истинной корреляцией, равной 0.2⁶⁹.

9.1.3. Простая линейная регрессия

Применение линейного регрессионного анализа имеет специфические черты по сравнению с другими методами обработки данных. Его непосредственное употребление ограничено, в основном, задачами о предсказании значений зависимой переменной по известным значениям аргумента (или аргументов), что в психологии задача не слишком востребованная. Однако, во-первых, линейная регрессия входит как часть во многие другие методы (например, анализ медиации и модерации, о которых речь пойдет в следующей главе), и, во-вторых, служит простым примером отыскания наилучших параметров для модели определенного типа, и психологу полезно понимать суть этого мето-

⁶⁸ Здесь также можно различать двухсторонние и односторонние гипотезы, как в случае Т-критерия (см. подпараграф 7.1.5).

⁶⁹ Для коэффициента корреляции так же, как и для других статистик, возможен расчет доверительных интервалов, показывающих, какие возможные значения истинной корреляции согласуются с выборочным. Смысл доверительного интервала тот же, что и в разобранных выше случаях, но техника расчета сложнее, поэтому мы не будем ее здесь давать.

да. Качество каждого набора параметров, а затем и модели в целом, оценивается процентом дисперсии, который остался вне предсказаний, сделанных моделью по данным значениям аргументов. Замечательным результатом для читателя будет здесь улавливание аналогий с двухфакторным дисперсионным анализом.

9.1.4. Определение регрессионной прямой

Простая линейная регрессия имеет дело с данными того же типа, что и корреляция, но придает им несколько иной смысл. Вот пример: в эксперименте шестилетнего ребенка оставляли на 15 минут одного в комнате перед вкусной шоколадной конфетой, пообещав дать две, если эту конфету он не тронет. Через два года оценивалась его успеваемость в школе. Оказалось, что время, в течение которого ребенок удерживался от соблазна, предсказывает его успеваемость во втором классе с точностью 60%. Уточним, что стоит за этим числом.

Пусть в результате эксперимента с участием n детей получены n пар чисел

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n).$$

Линейный регрессионный анализ исходит из гипотезы, что интегральный показатель успеваемости данного ребенка представляет собой сумму детерминированного вклада линейной функции от «времени терпения» и случайного вклада, зависящего от его неучтенных индивидуальных особенностей.

Таким образом, предполагается зависимость $y_i = ax_i + b + \alpha_i$, где a и b некоторые коэффициенты, а α_i случайная добавка (она называется «остаток»), для каждого испытуемого своя. Ясно, что задача не имеет очевидного однозначного решения. Всякая прямая может фигурировать в правой части равенства, от ее выбора будет зависеть только совокупность компенсирующих «остатков» α_i . «Лучшей» из этих прямых будем считать ту прямую, для которой эта совокупность в каком-то смысле минимальна, т.е. ту, которая «ближе» к точкам, представляющим испытуемых. С точки зрения математики это «ближе» разумнее всего понимать так: надо найти такую прямую, чтобы сумма квадратов «остатков» $\alpha_1^2 + \alpha_2^2 + \dots + \alpha_n^2$ была минимальна⁷⁰.

Каждую пару чисел из (x_i, y_i) можно рассматривать как точку на плоскости в системе координат (x, y) .

⁷⁰ В подпараграфе 9.3.1 практикума мы разберем эти операции на конкретном примере. Мы рекомендуем читателю сначала выполнить практическое задание, а затем вернуться к данному пункту.

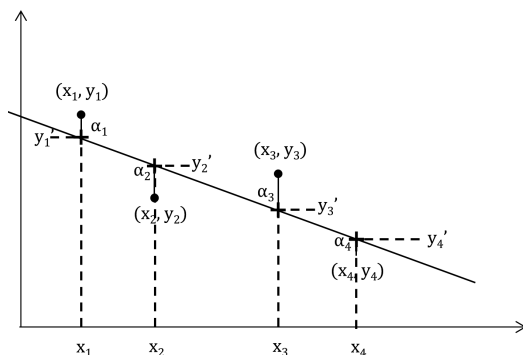


Рис.9.1.4(1). Выборочные значения, регрессионная прямая и ошибки регрессионной модели.

На рис. 9.1.4(1) изображены четыре точки, отображающие выборку $(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4)$, и регрессионная прямая с уравнением $y = ax + b$. Крестиками отмечены на прямой точки, которые имеют предсказанные регрессионным уравнением ординаты (т.е. ординаты $y_1' = ax_1 + b, y_2' = ax_2 + b, y_3' = ax_3 + b, y_4' = ax_4 + b$ соответственно). Вертикальное расстояние $y_i - y_i'$ от точки выборки, до точки, помеченной крестиком, и есть α_i — ошибка данной модели для данной точки. Сумма квадратов ошибок и будет характеризовать качество данной прямой как модели для данной выборки.

Операция по нахождению наилучшей в этом смысле прямой, описывающей подобную зависимость, называется *линейным регрессионным анализом*, а метод, обеспечивающий решение, которое характеризуется минимумом дисперсии случайной составляющей, называется *методом наименьших квадратов (Гаусса)*.

Наилучшая в указанном смысле прямая всегда проходит через точку (\bar{x}, \bar{y}) и имеет угловой коэффициент и свободный член, соответственно

$$a = \frac{r_{xy} \cdot s_y}{s_x}, b = \bar{y} - \bar{x} \cdot a,$$

где r_{xy} — коэффициент корреляции. Т.е. наилучшей в смысле минимума суммы квадратов отклонений является прямая

$$y = a \cdot x + \bar{y} - \bar{x} \cdot a.$$

Прикладной смысл регрессионной прямой заключается в том, что ордината точки на прямой над данным значением аргумента x и есть предсказание успеваемости для ребенка, показавшего x минут терпения. Что означает 60% точности, мы разьясим по рис. 9.1.4(1).

Общая дисперсия результатов детей есть дисперсия выборки $\{y_1, \dots, y_4\}$. Дисперсия предсказанных результатов есть дисперсия выбор-

ки $\{y_1, \dots, y_4\}$, состоящей из проекций Y_i на регрессионную прямую. Для расчетов вместо дисперсий будем использовать (как и в дисперсионном анализе) суммы квадратов, не деля их на $(4 - 1)$. Результат не изменится, поскольку качество предсказания будет измеряться отношением: $R^2 = S_{model}/S_{total}$, где S_{total} — сумма квадратов для выборки $\{y_1, \dots, y_4\}$ ⁷¹, а S_{model} — сумма квадратов для выборки $\{y_1, \dots, y_4\}$. Роль остатков в R^2 более понятна, если учесть, что $S_{model} = S_{total} - S_{error}$, где S_{error} — сумма квадратов остатков. Тогда $R^2 = 1 - S_{error}/S_{total}$.

R^2 — доля дисперсии выборки, приходящаяся на предсказанные регрессионным уравнением значения y , называется коэффициентом детерминации. Говоря «предсказывается с точностью 60%», мы имели в виду $R^2 = 0.6$.

Для простой линейной регрессии можно вычислить также, возведя в квадрат коэффициент корреляции: $R^2 = r_{xy}^2$. Доказательство этого факта мы здесь приводить не будем. Для множественной регрессии, о которой речь пойдет дальше, напротив, первичным оказывается R^2 , а множественная корреляция определяется через него.

9.1.5. Стандартизованный регрессионный коэффициент. Значимость

Если в задаче простой линейной регрессии стандартизовать зависимую и независимую переменные, т.е. преобразовать переменные так, чтобы их дисперсии стали равными единице, то регрессионный коэффициент a совпадет с коэффициентом корреляции, а свободный член будет равен нулю, что видно непосредственно из вышеприведенных формул. Статистическая оценка значимости не меняется при допустимом преобразовании шкал, описывающих переменные, поэтому мы будем вести речь о значимости для стандартизованных переменных, поскольку это наиболее удобно.

Во-первых, заметим, что оценка R^2 не меняется при стандартизации, как и при любом другом линейном преобразовании переменных, поскольку на одинаковые константы умножаются все суммы квадратов S_{total} , S_{model} и S_{error} . Для оценки значимости аналогично тому, как это делали в дисперсионном анализе, составляется F-отношение:

$$F = \frac{S_{model}/df_{model}}{S_{error}/df_{error}}$$

Чем больше это отношение, тем больше у нас оснований склонить-

⁷¹ В главе 7 S_{total} обозначала у нас сумму квадратов, включая сумму константы, здесь же S_{total} ее не включает. Это не наш недосмотр, так обозначаются суммы в соответствующих таблицах SPSS, на которые мы здесь ориентируемся. Чтобы уменьшить риск путаницы, мы в первом случае используем заглавную букву 'T'.

сы к тому, что наша модель надежна, и тем уже доверительный интервал вокруг полученного углового коэффициента⁷². Для числителя число степеней свободы равно единице, а для знаменателя $n - 2$, где n — объем выборки (подробнее о степенях свободы в следующем параграфе). Далее — знакомая уже процедура: по полученному значению F находится вес верхнего хвоста соответствующего распределения Фишера, который отсекается данным значением.

Разумеется, значимость регрессионного коэффициента в простой линейной регрессии совпадет со значимостью коэффициента корреляции.

9.1.6. Корреляция, регрессия и причинность

Корреляция и регрессия — инструменты исследования связи, или согласованности двух переменных. Их возможности ограничены. Сами по себе они никогда не смогут ничего сказать о направлении связи между переменными. Влияет ли уровень оптимизма на продолжительность жизни или, напротив, прогноз состояния организма, каким-то образом воспринимаемый его владельцем, влияет на уровень оптимизма — вопрос, на который нельзя ответить исходя только из корреляционных и регрессионных коэффициентов. Если ответ и возможен, то только с опорой на тонкие аспекты экспериментального дизайна. В третьем параграфе мы обсудим эту тему на примерах.

9.2. Корреляция и регрессия. Продолжение

9.2.1. Корреляция и регрессия. Степени свободы. Сравнение с дисперсионным анализом. Мощность критерия

Мы сравним сейчас два метода — регрессионный и дисперсионный анализ — на простом примере из шести точек. Предположим, наши данные — это оценки знания математических методов у студентов соответствующего курса (номер курса задан абсциссой x). Уместность первого или второго метода определяется содержательной гипотезой, которую мы хотим оценить в нашем исследовании. Если мы (преподаватели математических методов) предполагаем, что наши лекции и семинары должны приводить и приводят к постоянному увеличению знаний студентов, то нашей гипотезе будет соответствовать рост средних значений. Пусть результат выражается парной выборкой (2, 1); (2, 3); (4, 2); (4, 4); (6, 3); (6, 5) (рис. 9.2.1(1)). Регрессионная прямая имеет угловой коэффициент 0.5 и проходит через точки (2, 2), (4, 3) и (6, 4).⁷³

⁷² Не будем забывать, что наши данные содержат вклад случайных обстоятельств, поэтому при повторении исследования мы можем получить иные коэффициенты.

⁷³ Для лучшего понимания рекомендуем ввести данные в SPSS, проделать регрессионный анализ и найти в таблицах вывода все полученные ниже числа.

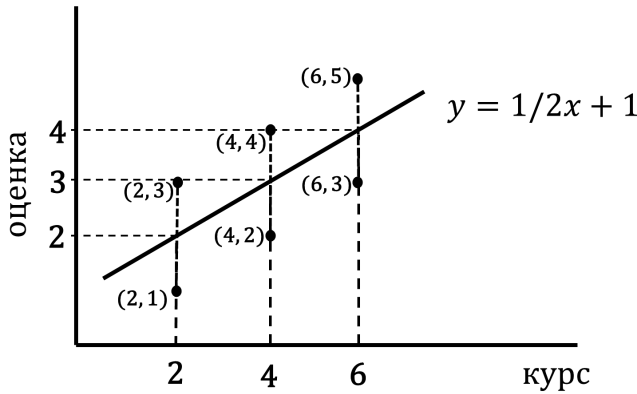


Рис. 9.2.1(1)). Регрессионная прямая для набора из 6 точек. Угловым коэффициентом равен 0.5, прямая проходит через точки (2, 2), (4, 3) и (6, 4).

Сумма квадратов задается ординатами проекций точек на регрессионную прямую (2, 2, 3, 3, 4, 4) относительно общего среднего, равного 3, и составляет $S_{model} = (-1)^2 + (-1)^2 + 0 + 0 + 1^2 + 1^2 = 4$ (каждая пара точек, имеющих одну абсциссу, проецируется в одну точку прямой, которая поэтому учитывается дважды). Остаток $S_{error} = 6 * 1^2 = 6$. Показатель качества модели $R^2 = S_{model}/S_{total} = 0.4$. Значимость нулевой гипотезы о равенстве нулю углового коэффициента регрессионной прямой определяется значением

$$F_4^1 = \frac{S_{model}/1}{S_{error}/4} = \frac{4}{1.5} = 2.667.$$

Суммы квадратов в числителе и знаменателе мы делим на соответствующее число степеней свободы. Исходно их шесть — столько же, сколько точек. После вычитания \bar{y} их остается пять — поскольку \bar{y} вместе с любыми пятью числами определяют всю исходную шестерку чисел. Далее эти пять степеней свободы делятся на две части. Одна степень свободы (модели) приходится на управляемую одним лишь угловым коэффициентом прямую, проходящую через точку (\bar{x}, \bar{y}) . Остальные четыре остаются остатку, или ошибке. Сумму квадратов в числителе мы делим на число степеней свободы модели, которое равно единице. На число четыре делится сумма остатка в знаменателе.

Мы привели здесь, скорее, мнемоническое правило, чем объяснение. Интересующихся более точными разъяснениями адресуем к приложению 4.

Однако мы здесь изложим эту важную тему компактно и без таблиц.

Вернемся к нашим данным. Прделаем теперь над этими же данными однофакторный дисперсионный анализ (три группы по два студента в каждой). Вычисляя, мы найдем те же самые суммы квадратов, что и в случае регрессии, поскольку групповые средние лежат на регрессионной прямой. Но F-отношение будет другим из-за иного соотношения степеней свободы:

$$F_3^2 = \frac{S_{model}/2}{S_{error}/3} = \frac{2}{2} = 1.$$

Почему? После вычитания общего среднего \bar{y} модель дисперсионного анализа управляется двумя параметрами — среднее третьей группы дополняет сумму средних двух групп до нуля⁷⁴. На остаток приходится теперь три степени свободы.

Подразумеваемое соображение, которое верно для *случайных выборок*, таково: для достаточно разумных моделей после вычитания общего среднего введение каждого нового параметра должно забирать из общей суммы квадратов примерно одинаковую порцию. Шесть параметров должны свести остаток к нулю, а всю сумму перевести в S_{model} (наилучшее приближение в этом случае сами числа). «Беда» модели дисперсионного анализа по сравнению с регрессионной в нашем случае в том, что, добавив параметр, позволяющий групповым средним не лежать на одной прямой, она не перевела ничего из суммы остатка в сумму модели. Это и означает, что для наших данных *регрессионная модель лучше модели дисперсионного анализа, ее F-отношение больше, и значимость лучше*.

Наши данные соответствуют гипотезе о монотонном и даже линейном росте уровня знаний у студентов. Если именно это мы и хотели доказать, то нам следовало заранее спланировать обработку результатов методом линейной регрессии. Если же неких нейтральных исследователей интересует всякая динамика знаний студентов и они заранее понимают, что и результат с максимальным уровнем знаний в середине времени обучения и забыванием материала к концу не менее интересен и также будет ими опубликован, то регрессионный анализ им недостаточен, поскольку процесс с обучением и забыванием не будет уловлен. В первом случае гипотеза более узкая и регрессионный метод ей адекватен.

⁷⁴ Если быть совсем точным, то описание модели дисперсионного анализа в нашем случае таково: для централизованной системы шести чисел, разбитых на три группы по два, подобрать централизованную систему чисел для каждой из групп, чтобы сумма квадратов остатков была минимальной. Решение регулируется двумя параметрами (поскольку система трех чисел должна быть централизована). Наилучшее приближение дает система средних групповых, которая является централизованной, поскольку централизована исходная система шести точек.

Во втором случае гипотеза шире, регрессионный метод для нее узок, а дисперсионный анализ адекватен. За эту адекватность в последнем случае мы платим тем, что, получив результат, как в разобранном примере, мы проигрываем в величине F-отношения и значимости.

Говорят, что по отношению к гипотезе о линейном росте знаний регрессионный метод *более мощный*, чем дисперсионный анализ⁷⁵.

«Мораль» нашего примера такова.

На языке статистической значимости: Совершенно честно и в то же время полезно для усиления наших результатов точно формулировать гипотезы исследования и подбирать соответствующий наиболее мощный критерий.

На языке моделей: Введение лишних параметров настройки модели может вообще не улучшить ситуацию с остатками, или может ее улучшить слабо, так что это улучшение не компенсирует рост числа степеней свободы модели (уменьшающий числитель F-отношения) и убывание числа степеней свободы остатка (увеличивающее знаменатель F-отношения). Выходя за пределы примера, можно сформулировать более общее утверждение. Введение бесполезных параметров (а также переменных, групп испытуемых и т.п.) в исследование может существенно ухудшить результат. Надежды на то, что сбор всевозможной информации без наличия разумных гипотез и направленного отбора позволит с помощью математической обработки получить хороший результат, при «честной игре» не оправдывается. Такой «неразборчивый» дизайн может работать только в пилотажном исследовании, результаты которого затем должны перепроверяться в исследовании с более точно сформулированными условиями и дизайном.

9.2.2. Ковариация и корреляция между случайными величинами

Пусть X и Y — дискретные случайные величины, распределение которых задано таблицами

X	-1	1
	0.5	0.5

и

Y	-1	1
	0.5	0.5

Если X и Y независимы, то вероятность одновременного наступления событий $X = -1$, $Y = -1$ равна произведению вероятностей наступления каждого из событий в отдельности, т.е. 0.25 (аналогично для остальных пар событий ($X = -1$, $Y = 1$), ($X = 1$, $Y = -1$) и ($X = 1$, $Y = 1$)). Однако так бывает далеко не всегда. Например, если $X=1$ кодирует собы-

⁷⁵ Мощность критерия измеряется вероятностью принять гипотезу H_1 , когда она верна. В нашем случае она выше для регрессионного анализа, если гипотеза H_1 предполагает линейный рост знаний.

тие «вес испытуемого выше среднего», а $X = -1$ — «вес ниже среднего», а случайная величина Y кодирует аналогичные высказывания про рост испытуемого, то понятно, что вероятность одновременного превышения среднего роста и среднего веса выше, чем вероятность наблюдать большой вес при маленьком росте. В этом случае имеет смысл говорить о совместном распределении случайных величин. Предположим, что для нашего примера оно может быть задано таблицей

$X = -1, Y = -1$	$X = -1, Y = 1$	$X = 1, Y = -1$	$X = 1, Y = 1$
0.4	0.1	0.1	0.4

Характеристикой степени отклонения этого распределения от распределения независимых величин служит *корреляция*.

Сначала вычислим *теоретическую ковариацию* R_{XY} . Для этого будем перемножать отклонения значений каждой из случайных величин от ее собственного математического ожидания и домножать произведение на вероятность совместного появления этих значений.

Первое слагаемое получается из первой клетки таблицы, где $X = -1$, $Y = -1$, а вероятность их совместного появления 0.4 берется из второй строки. В результате имеем произведение $((-1) - M_X)((-1) - M_Y) \cdot p(X = -1, Y = -1)$ (хотя $M_X = M_Y = 0$, мы оставляем их в формуле). Это слагаемое будет первым в формуле теоретической ковариации. Остальные слагаемые формируются аналогично.

$$\begin{aligned}
 R_{XY} = & \\
 = & ((-1) - M_X)((-1) - M_Y) \cdot p(X = -1, Y = -1) + \\
 & + ((-1) - M_X)((+1) - M_Y) \cdot p(X = -1, Y = 1) + \\
 & + ((+1) - M_X)((-1) - M_Y) \cdot p(X = 1, Y = -1) + \\
 & + ((+1) - M_X)((+1) - M_Y) \cdot p(X = 1, Y = 1)
 \end{aligned}$$

Подставляя значения вероятностей и учитывая, что $M_X = M_Y = 0$, получаем

$$R_{XY} = (-1)(-1) \cdot 0.4 + (-1)(1) \cdot 0.1 + (1)(-1) \cdot 0.1 + (1)(1) \cdot 0.4 = 0.6.$$

Упражнение 9.2.2(1). Подставьте в предыдущую формулу вероятности 0.25, соответствующие независимости случайных величин, и убедитесь, что значение ковариации в этом случае равно нулю.

Для того чтобы получить *теоретическую корреляцию* ρ_{XY} (ρ — буква греческого алфавита, читается «ро»), надо поделить ковариацию на среднеквадратические отклонения

$$\rho_{XY} = \frac{R_{XY}}{\sqrt{D_X} \sqrt{D_Y}}$$

В нашем случае $D_X = D_Y = 2$. Подставляя в формулу, получаем $\rho_{XY} = 0.3$.

В общем случае для случайных величин, заданных таблицами

X	x_1	x_2	...	x_m
	p_1	p_2	...	p_m

и

Y	y_1	y_2	...	y_m
	q_1	q_2	...	q_m

теоретические ковариация и корреляция задаются формулами

$$R_{XY} = \sum_{i,j} (x_i - M_X)(y_j - M_Y)p(X = x_i \text{ и } Y = y_j)$$

$$\rho_{XY} = \frac{R_{XY}}{\sigma_X \sigma_Y},$$

где $\sigma_X = \sqrt{D_X}$, $\sigma_Y = \sqrt{D_Y}$.

Замечание. Если вместо Y и y подставить в формулу ковариации X и x , соответственно, то формула ковариации превратится в формулу дисперсии X . Действительно, если $i \neq j$, то $p(X = x_i \text{ и } X = x_j) = 0$, поскольку случайная величина не может одновременно принимать два разных значения; если $i = j$, то $p(X = x_i, X = x_i)$ тавтологично равно $p(X = x_i)$. Таким образом, в формуле для ковариации будут отличны от нуля только «диагональные» члены, для которых $i = j$, т.е. $R_{XX} = \sum_{i=j} (x_i - M_X)(x_j - M_X)p(X = x_i \text{ и } X = x_j) = \sum_i (x_i - M_X)^2 p_i = D_X$.

9.2.3. Гипотезы, связанные с ненулевой теоретической корреляцией. Преобразование Фишера

В предыдущей главе обсуждались вопросы, связанные с противопоставлением гипотез о равенстве нулю или неравенстве нулю теоретической корреляции. В случае, если верна гипотеза H_0 о независимости переменных (тем самым их теоретическая корреляция равна нулю), выборочная корреляция принимает значения более или менее близкие к нулю. Большие отклонения от нуля рассматриваются как свидетельства в пользу альтернативной гипотезы.

Если же теоретическая корреляция случайных величин отлична от нуля, то задачи решаются с помощью *преобразования Фишера*. Наиболее распространенной из таких задач является задача сравнения корреляций. Например, проверка гипотезы о равенстве корреляций $\rho_1 = \rho_2$ против альтернативы $\rho_1 \neq \rho_2$.

Напомним формулу выборочной корреляции. Пусть (x_1, \dots, x_n) и (y_1, \dots, y_n) — связанные выборки (в данном случае показатели с одинаковыми номерами принадлежат одному испытуемому).

Формула выборочной корреляции в наиболее часто употребляемом виде:

$$r_{xy} = \frac{1}{n-1} \frac{((x_1 - \bar{x})(y_1 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y}))}{s_x s_y}.$$

Если теоретическая корреляция двух переменных не равна нулю, то выборочная корреляция этих переменных будет варьировать от испытания к испытанию не вокруг нуля, а приблизительно вокруг этой теоретической корреляции.

Если размер выборки достаточно велик ($n > 20$), то можно воспользоваться преобразованием Фишера.

Если корреляция реальных случайных величин, в результате испытаний которых были получены выборки (x_1, \dots, x_n) и (y_1, \dots, y_n) , равна некоторому числу ρ , то распределение статистики (напомним, что статистика является случайной величиной)

$$z_{xy} = \frac{1}{2} \ln \left(\frac{1 + r_{xy}}{1 - r_{xy}} \right)$$

будет приближенно нормальным с математическим ожиданием

$$M_z = \frac{1}{2} \ln \left(\frac{1 + \rho}{1 - \rho} \right) + \frac{\rho}{2n - 2}$$

и дисперсией

$$D_z = \frac{1}{n - 3}$$

Переход от r_{xy} к z_{xy} называется преобразованием Фишера.

Если мы теперь захотим, имея выборки (x_1, \dots, x_{20}) и (y_1, \dots, y_{20}) , проверить гипотезу H_0 о том, что реальная корреляция ρ равна, скажем, 0,2, то эта гипотеза будет эквивалентна гипотезе о математическом ожидании преобразования Фишера от выборочной корреляции, рассмотренной как случайная величина. Результат этого преобразования — нормально распределенная случайная величина с математическим ожиданием

$$M_z = \frac{1}{2} \ln \left(\frac{1 + 0.2}{1 - 0.2} \right) + \frac{0.2}{2 * 20 - 2}$$

и дисперсией $Dz = 1/(20 - 3)$. Для проверки гипотез мы можем использовать квантили нормального распределения.

Особенно важное применение преобразования Фишера — сравнение двух выборочных коэффициентов корреляции. (Например, мы хотим исследовать вопрос, одинаковы ли корреляции между ростом и весом у жителей Москвы и Лиссабона.)

Пусть и выборочные коэффициенты корреляции, полученные по выборкам размера и соответственно.

Если реальные корреляции ρ_1 и ρ_2 равны, то случайная величина $z_1 - z_2$ будет иметь нормальное распределение с математическим ожиданием, равным нулю, и дисперсией, равной сумме дисперсий D_1 и D_2 , т.е.

$$\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}$$

Тогда
$$z_{1-2} = \frac{z_1 - z_2}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}}$$

будет стандартной нормальной случайной величиной, и гипотеза о равенстве корреляций оценивается с помощью квантилей нормального распределения. В пользу неравенства будут свидетельствовать anomalно большие (положительные или отрицательные) значения Z_{1-2} .

9.3. Линейная регрессия. Практикум

Пример 9.3(1). Разложение сумм квадратов. Простейший пример

Рассмотрим простой пример, конкретизирующий пример подпараграфа 9.1.3. Предположим, ребенок А терпел и не брал конфету в течение 2 минут, а оценка его успехов во 2-м классе 3 балла; ребенок В терпел 3 минуты, а оценка его успехов 5 баллов; ребенок С терпел 4 минуты, а оценка его успехов 4 балла.

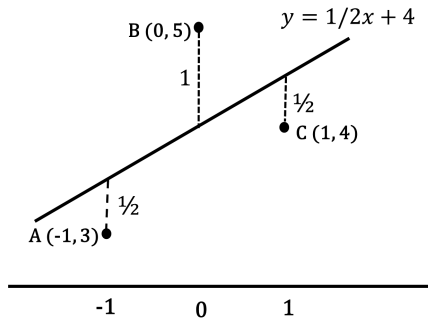


Рис.9.3(2). Регрессионная прямая для трех точек $(-1, 3)$, $(0, 5)$ и $(1, 4)$.

Для упрощения расчетов примем за нулевую точку среднее «время терпения» детей 3 минуты. В результате получаем три отображающих наши данные точки: $(-1, 3)$, $(0, 5)$ и $(1, 4)$ — они изображены на рис. 9.3(2).

Первый ребенок вытерпел на минуту меньше среднего значения, а третий на минуту дольше.

Среднее значение оценки успехов $\bar{y}=4$.

Обе наши выборки имеют дисперсию, равную единице, поэтому коэффициент ковариации равен коэффициенту корреляции $R_{xy} = r_{xy} = 1/2$. Напомним, что регрессионные коэффициенты рассчитываются по формуле:

$$a = \frac{r_{xy} \cdot s_y}{s_y}, b = \bar{y} - \bar{x} \cdot a.$$

В нашем случае наилучшая в смысле суммы квадратов отклонений по оси y прямая имеет уравнение $y = \frac{1}{2}x + 4$.

Составим суммы квадратов.

Первая сумма S_{total} отражает дисперсию ординат наших точек, т.е. это сумма квадратов их отклонений от среднего (мы не делим ее и другие суммы на $n-1$, поскольку будем рассматривать отношения сумм, как и в дисперсионном анализе). $S_{total} = (3 - 4)^2 + (5 - 4)^2 + (4 - 4)^2 = 2$.

Вторая сумма $S_{regression}$ (вместо S_{model} мы будем писать теперь $S_{regression}$, а вместо S_{error} — $S_{residual}$, чтобы согласовываться с заголовками таблиц вывода SPSS) связана с собственно регрессионной моделью. Эта сумма квадратов отражает дисперсию предсказанных моделью значений. На рисунке 9.3.1(1) соответствующие точки лежат на регрессионной прямой и помечены звездочками. Подстановка $x=-1$ в регрессионное уравнение дает первую звездочку $y = 3\frac{1}{2}$, $y = 3\frac{1}{2}$ и остальные точки (0, 4) и (1, 4). Искомая сумма

$$S_{regression} = (3\frac{1}{2} - 4)^2 + (4\frac{1}{2} - 4)^2 = \frac{1}{2}.$$

Третья сумма $S_{residual}$ отражает ошибку регрессионной модели, то, насколько реальные наблюдения отклоняются от предсказанных. На рисунке отклонения изображены жирными линиями, соединяющими точки А, В, С с регрессионной прямой, их длины подписаны рядом с линиями

$$S_{residual} = (\frac{1}{2})^2 + (\frac{1}{2})^2 + 1 = \frac{3}{2}.$$

Читатель уже не удивится, обнаружив, что $S_{total} = S_{regression} + S_{residual}$.

На каждую из сумм приходится по одной степени свободы, поэтому

$$F_1^1 = S_{regression}/S_{residual} = 1/3.$$

Можно добавить еще две суммы: сумму квадратов значений $3^2 + 4^2 + 5^2 = 50$ и $4^2 + 4^2 + 4^2 = 48$ и убедиться, что $50 = 48 + S_{total}$. Но, как и в случае дисперсионного анализа, константа нас не интересует, когда мы ставим вопрос о связи переменных. Она нужна, только если мы хотим реально предсказывать значения.

Пример 9.3(3). Разложение сумм квадратов относительно регрессионной прямой в SPSS

Теперь посмотрим, как справляется с нашей задачей SPSS. Откроем файл *Wtime.sav*. Он содержит наши три наблюдения (переменные *Wtime* — «время терпения» и *Success* — «успеваемость во втором классе»). Сначала рассчитаем корреляцию между переменными.

Последовательно выберем пункты меню **Анализ — Корреляции — Парные (Analyze — Correlation — Bivariate)**. В открывшемся окне переносим переменные *Wtime* и *Success* в поле **Переменные (Variables)**.

После нажатия *OK* получаем таблицу.

Корреляции

		Wtime	Success
Wtime	Корреляция Пирсона	1	.500
	Знч.(2-сторон)		.667
	N	3	3
Success	Корреляция Пирсона	.500	1
	Знч.(2-сторон)	.667	
	N	3	3

Собственно корреляция между нашими переменными имеет величину 0.5, а соответствующая значимость равна 0.667.

Теперь проведем линейную регрессию. Выберем в меню пункты *Анализ — Регрессия — Линейная (Analyze — Regression — Linear)*. В открывшемся окне перенесем переменную *Success* в окно *Зависимая переменная (Dependent)*, а переменную *Wtime* в окно *Независимые переменные (Independent(s))*. Нажав *OK*, получаем таблицы:

Сводка для модели

Модель	R	R-квадрат	Скорректированный R-квадрат	Стд. ошибка оценки
1	.500 ^a	.250	-.500	1.22474

a. Предикторы: (конст) W_time

Дисперсионный анализ^a

Модель		Сумма квадратов	ст.св.	Средний квадрат	F	Знч.
1	Регрессия	.500	1	.500	.333	.667 ^b
	Остаток	1.500	1	1.500		
	Всего	2.000	2			

^a. Зависимая переменная: Success

^b. Предикторы: (конст) W_time

Коэффициенты^a

В	Модель	Нестандартизованные коэффициенты		Стандартизованные коэффициенты	t	Знч.
		Стд. Ошибка	Бета			
1	(Константа)	4.000	.707		5.657	.111
	W_time	.500	.866	.500	.577	.667

^a. Зависимая переменная: Success

В первой таблице нам представлены коэффициент корреляции R и показатель качества модели R².

Во второй — рассчитанные нами выше суммы квадратов: регрессия (regression), остаток (residual), общая (total). Это те самые 0.5, 1.5 и 2, которые мы получили в расчетах предыдущего пункта. Далее с этими числами производятся операции, аналогичные тем, которые производились в дисперсионном анализе: суммы квадратов делятся на число степеней свободы (здесь они обе равны единице, в других примерах это будут другие числа), и затем 0.5 (регрессия) делится на 1.5 (остаток). Получается F-отношение 0.333, которому соответствует значимость 0.667 — та же самая, что и значимость корреляции, которую мы посчитали выше. Для простой линейной регрессии это всегда так.

В третьей таблице приведены регрессионные коэффициенты. В столбце, озаглавленном «В», первое число — это свободный член (в нашем случае, когда переменная *Wtime* центрирована, свободный член равен среднему по трем наблюдениям переменной *Success*), а второе число — угловой коэффициент регрессионного уравнения: 0.5.

В столбце, озаглавленном «Бета», даны регрессионные коэффициенты, как если бы уравнение составлялось после стандартизации переменных. Свободный член в этом случае равен всегда нулю и поэтому не приводится. Угловой коэффициент в нашем примере тот же самый, потому что дисперсии наших переменных равны единице и стандартизация бы их не изменила. В следующем примере мы получим различие стандартизованных и нестандартизованных коэффициентов. Далее, можно оценить значимость результата (т.е. рассчитанного регрессионного коэффициента) относительно нулевой гипотезы (регрессионный коэффициент равен нулю — т.е. успешность в школе не зависит линейно от времени ожидания)⁷⁶. Для этого, деля нестандартизованный коэффициент регрессии на его стандартную ошибку ($0.500/0.866 = 0.577$), по-

⁷⁶ Для свободного члена этот результат для нас неинтересен, это просто константа, связанная с разницей метрик независимой и зависимой переменной и почти всегда отличная от нуля, но он также рассчитывается в SPSS по умолчанию.

лучим показатель, сравнив который с t-распределением Стьюдента узнаем, какой хвост отсекают эти значения, или их значимость. В случае одной независимой переменной эти результаты совпадают с результатами дисперсионного анализа: F-отношение и t углового коэффициента связаны формулой $F = t^2$, а значимости одинаковы⁷⁷.

Чтобы графически представить полученный результат, построим диаграмму рассеяния для «времени выдержки» и успеваемости наших трех школьников: в меню **Графика — Устаревшие диалоговые окна — Рассеяния/Точки (Graphs — Legacy Dialogs — Scatter/Dot)** выбрать Простая диаграмма рассеяния (Simple Scatter) и перенести переменные в строки ось X и Y. Если дважды кликнуть по построенной диаграмме и выбрать в открывшемся окне в меню пункт **Добавить: линия аппроксимации для итога (Add Fit Line at Total)**, регрессионная прямая добавится на рисунок.

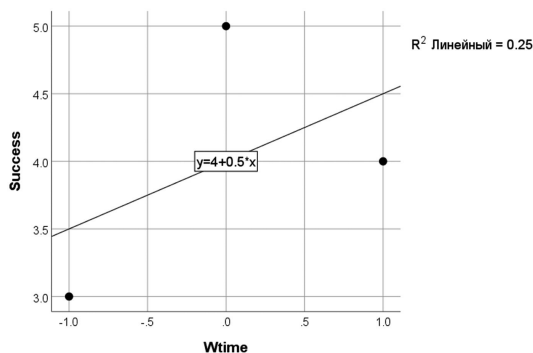


Рис. 9.3(4). График рассеяния с линией регрессии для модели из трех точек

Заметим в заключение, что показатель R^2 может быть получен как возведением в квадрат коэффициента корреляции, так и делением суммы регрессии (0.5) на общую сумму (2). Второй способ прямо связан с оценками моделей в других, более сложных случаях.

Упражнение 9.3(5). В файле *WtimeM.sav* добавлены еще две переменные, полученные из исходных преобразованиями $Wtime2 = Wtime * 2 + 2$, $Success3 = Success * 3$. Проведите регрессионный

⁷⁷ В главе 10 мы покажем, что когда независимых переменных несколько, F характеризует общее качество модели, а T-критерий направлен на проверку (не всегда удачную) того, насколько угловой коэффициент каждой из независимых переменных отличается от нуля.

анализ. Убедитесь и объясните, почему

- коэффициент корреляции не изменился;
- показатель качества модели R-квадрат не изменился;
- все суммы квадратов увеличились в 3×3 раз, а их отношение и соответствующая значимость не изменились;
- угловой коэффициент увеличился в $3/2$ раза;
- изменился свободный член.

Пример 9.3(6). В файле *simple.sav* содержатся три ряда данных: x , y_1 и y_2 .

Упражнение 9.3(7). Постройте графики рассеяния следующих пар переменных: x и y_1 , x и y_2 . Добавьте на графики регрессионные прямые.

По форме рассеяния точек оцените, в каком из двух случаев при построении регрессионной прямой коэффициент детерминации будет выше. Проверьте свои предположения, произведя расчет регрессии.

Убедитесь, что регрессионные прямые, построенные на графиках рассеяния y_1 по x и x по y_1 не совпадают. Проведите регрессию x по y_1 . Для взаимно обратных линейных функций угловые коэффициенты подчиняются отношению $k_1 * k_2 = 1$. Проверьте, что это соотношение в нашем случае не выполняется. Проверьте также, что показатели R и R^2 совпадают.

Пример 9.3(8). Предсказания с помощью регрессионного уравнения для новых данных. В файле *conflict.sav* содержатся модельные данные исследования, задачей которого было выявить индикаторы крепости супружеских отношений. Молодых супругов приглашали для исследования в помещение, где просили подождать начала эксперимента. Для того чтобы скоротать время ожидания, испытуемым предлагали посмотреть журналы мод с фотографиями моделей. Оценивалось количество времени, которое испытуемый смотрел на фотографии лиц противоположного пола (переменная *lookingtime* содержит время, измеренное в минутах). Через два года при помощи интервью по нескольким параметрам оценивали выраженность конфликтов в супружеской паре (переменная *conflictness*). В таблицу данных занесены также показатели времени для нескольких новых участников эксперимента, у которых данных о конфликтности еще нет.

Проведите регрессионный анализ. После заполнения именами переменных полей зависимых и независимых переменных нажмите кнопку **Сохранить** и поставьте галочку в клетку **Предсказанные значения — Нестандартизованные**. В результате выполнения регрессионного анализа в таблице данных появится еще один столбец, содержащий предсказанные значения, в том числе и для новых участников. Проверьте, что эти значения для всех получаются из регрессионного уравнения, коэффициенты которого даны в соответствующей таблице вывода в столбце

В. Обратите внимание, что регрессия «объясняет» не очень большую долю дисперсии переменной *conflictness* (R^2 равен 0.288), и хотя значимость регрессионного коэффициента (и коэффициента корреляции R) достаточно близка к нулю (0.001), все же интерпретировать результат надо с осторожностью. Мы продолжим обсуждение этих данным ниже.

Пример 9.3(9). Нелинейные связи в простой линейной регрессии. Предположим, классный руководитель решил проводить дополнительные занятия по русскому языку с учениками, которые сделали ошибки в сложном диктанте, а затем оценивал, насколько меньше ошибок ученики сделали после проведенных занятий. Учителя интересовало, кому из учеников эти занятия наиболее полезны — «хорошистам», делающим мало ошибок, «среднячкам» со средним количеством ошибок или неуспевающим детям, делающим много ошибок. В файле *ErrorsTraining.sav* смоделированы результаты такого исследования для 30 школьников: количество ошибок в первом диктанте зафиксировано в переменной *ErrorsBefore*. Изменение к следующему диктанту описывается переменной *ErrorsDiff* (отрицательные числа означают, что ошибок стало меньше, положительные — что больше).

Описательные статистики показывают, что среднее число ошибок после первого диктанта — чуть меньше пяти, а занятия в среднем уменьшают это значение почти на одну ошибку. Результаты регрессионного анализа показывают, что от количества ошибок в начале прогресс школьников не зависит (процент объясняемой дисперсии R^2 , оценка значимости $F=1.296$, $p=0.265$).

Описательные статистики

	N	Минимум	Максимум	Среднее	Среднекв. отклонение
ErrorsBefore	30	1.00	9.00	4.7333	1.85571
ErrorsDiff	30	-4.00	1.00	-0.9333	1.46059
N валидных (по списку)	30				

Сводка для модели

Модель	R	R-квадрат	Скорректированный R-квадрат	Стандартная ошибка оценки
1	0.210 ^a	0.044	0.010	1.45319

a. Предикторы: (константа), Errors_before

ANOVA^a

Модель		Сумма квадратов	ст.св.	Средний квадрат	F	Значимость
1	Регрессия	2.737	1	2.737	1.296	0.265 ^b
	Остаток	59.130	28	2.112		
	Всего	61.867	29			

a. Зависимая переменная: ErrorsDiff

b. Предикторы: (константа), ErrorsBefore

Однако по диаграмме рассеяния, изображенной на рис. 9.3(10), кажется, что закономерность все же есть: похоже, занятия хорошо работают в случае «среднячков» и малоэффективны у «хорошистов» и неуспевающих.

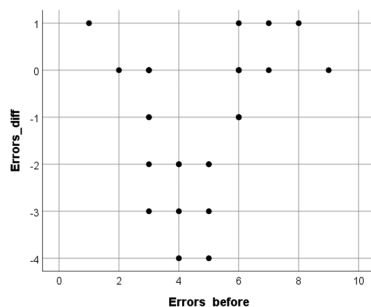


Рис. 9.3(10). График рассеяния исходного числа ошибок в диктанте и изменения числа ошибок во втором диктанте

Такого рода гипотезы — о нелинейных закономерностях — требуют применения нелинейной регрессии (в нашем учебнике мы не будем обсуждать эти методы). Тем не менее, оставаясь в рамках линейной регрессии, можно предварительно сориентироваться в ситуации. Разделим выборку на две части в соответствии со средним значением: тем, кто сделал 4 и менее ошибки, припишем «1», тем, что 5 и более, припишем «2» (переменная *Group*) и проведем регрессионный анализ отдельно в двух группах. Для этого выберем *Данные* — *Расщепить файл (Data — Split file)*, укажем *Сравнить группы (Compare groups)* и перенесем в поле справа (*Группы образуются по, (Groups based on)*) переменную *Group*. Весь дальнейший анализ будет проводиться отдельно в выделенных группах до тех пор, пока мы не уберем эту установку (выбрав *Анализировать все наблюдения, группы не создавать, (Analyze all cases, do not create groups)*).

Повторив теперь процедуру задания регрессионного анализа для переменной *ErrorsDiff*, получим две серии таблиц. Обе линейные модели для двух подвыборок теперь характеризуются близкими к нулю значениями и имеют достаточно высокие показатели R^2 : модель для хорошо успевающих учеников объясняет 54.9% дисперсии изменений в количестве ошибок, а у слабых учеников — 44.5%. Как видно в таблице коэффициентов, если рассматривать успевающих учеников, занятия лучше помогают тем, кто делает больше ошибок; если слабых — наоборот.

Коэффициенты^а

Group	Модель В		Нестандартизованные коэффициенты		Стандартизованные коэффициенты	t	Значимость
			Стандартная ошибка	Бета			
1		(Константа)	2.840	1.097	-0.741	2.589	0.024
	1	ErrorsBefore	-1.320	.346			
2		(Константа)	-6.078	1.633	0.667	-3.723	0.002
	1	ErrorsBefore	0.871	0.260			

а. Зависимая переменная: Errors_diff

Упражнение 9.3(11). Постройте две диаграммы рассеяния для успевающих и слабых учеников и рассчитайте для них регрессионную модель. Проинтерпретируйте результат.

Итак, как и в случае корреляции, незначимость регрессионной модели не означает отсутствия связи, и диаграммы рассеяния и описательная статистика — незаменимые помощники на предварительном этапе анализа данных.

Пример 9.3(12). В предыдущем примере саму связь между независимой и зависимой переменными было точнее описывать как нелинейную. Рассмотрим еще более сложную зависимость: случай, когда связь зависит от какой-то другой, внешней по отношению к исследованию переменной. В файле *conflictCB.sav* мы модифицировали данные обсуждавшегося выше исследования связи времени рассматривания фотографий лиц противоположного пола и конфликтов в браке через два года. Предположим, что в исследовании участвовали только женщины, недавно вступившие в брак, и за два года некоторые из них успели родить ре-

бенка. Эта информация отражена в переменной *ChildBirth*, где «0» — нет ребенка, «1» — есть.

Упражнение 9.3(13). Проведите линейную регрессию отдельно в этих двух группах и сравните результаты.

Обратите внимание, что ни по длительности рассматривания фотографий, ни по конфликтности в среднем обе группы женщин не различаются. Это можно проверить при помощи *t*-критерия Стьюдента: средняя длительность составляет 16.02 у женщин без детей и 15.81 у родивших ($t=0.151$, $p=0.881$), а средняя конфликтность составляет 17.06 и 17.25 соответственно ($t=-0.262$, $p=0.595$).

Сравнение двух простых линейных регрессий показывает, что найденная в целом по выборке связь (описанная в параграфе 9.3.2) есть и даже значительно более выражена только у бездетных женщин ($R^2 = 54.4\%$, по сравнению с 28.8% в общей выборке). У тех, кто родил ребенка, уровень конфликтности в семье такой же, но он не связан с просмотром фотографий два года назад.

Такая ситуация носит название *модерации*: эффект длительности просмотра фотографий на конфликтность в браке зависит от того, как развивался этот брак. Переменная *ChildBirth* выступает переменной-модератором связи независимой и зависимой переменными. Из-за модерации закономерности, полученные в общей выборке, могут быть слабее, чем в одной из групп (как в нашем случае), не выявляться или даже быть столь же бессмысленными, как и «средняя температура по больнице». В главе 10 анализ модерации и процедура его проведения описаны подробнее.

Глава 10.

Регрессия в случае многих переменных

10.1. Регрессионная модель для нескольких независимых переменных (множественная регрессия)

10.1.1. Трехмерный пример

Мы можем усложнить задачу о связи «времени выдержки» с последующей успеваемостью в школе (см. 9.1.4). Предположим, что кроме времени, которое ребенок выдержал наедине с конфетой, мы характеризуем его также максимальным временем выполнения какой-то монотонной работы.

Таким образом, у нас имеется две независимые переменные — время выдержки и время монотонной работы — и зависимая переменная — успеваемость в школе. Тогда вместо n пар чисел мы имеем n троек чисел $(x_1, y_1, z_1), (x_2, y_2, z_2), \dots, (x_n, y_n, z_n)$.

Пусть x_i — это время выдержки i -го испытуемого, y_i — время монотонной работы, а успеваемость — z_i . Допустим, мы хотим оценить параметры зависимости последующей успеваемости от времени выдержки и времени монотонной работы.

Каждую тройку чисел можно рассматривать как точку в пространстве. Уравнение регрессии — $(x_1, y_1, z_1), (x_2, y_2, z_2), \dots, (x_n, y_n, z_n)$.

Приведенные в предыдущей главе выкладки легко обобщаются на этот случай. Тройки чисел (x_i, y_i, z_i) можно представить точками в пространстве, а регрессионное уравнение задает плоскость в этом пространстве, которая наилучшим образом соответствует этому множеству точек. Чтобы вычислить S_{model} и S_{error} , надо спроецировать вертикальными отрезками точки на регрессионную плоскость. Наилучшей будет такая плоскость, для которой сумма квадратов этих отрезков от точек до плоскости будет минимальна. S_{error} — это минимальная сумма квадратов, а S_{model} — дисперсия проекций точек на плоскость (только не деленная на $n-1$) — полностью аналогично простой регрессии.

10.1.2. Значимость для модели множественной регрессии в целом

Коэффициент детерминации для множественной регрессии можно посчитать по той же формуле, что и для простой регрессии: $R^2 = S_{model}/(S_{error} + S_{model}) = S_{model}/S_{total}$. И здесь коэффициент детерминации характеризует долю дисперсии, определенную регрес-

сионной моделью, в дисперсии зависимой переменной. Заметим, что в данном случае он уже не вычисляется просто через квадраты коэффициентов корреляции.

Составим из сумм квадратов отношение Фишера:

$$F = \frac{S_{model}/df_{model}}{S_{error}/df_{error}}$$

Для числителя теперь число степеней свободы равно двум (два регрессионных коэффициента), а для знаменателя — (из числа точек вычитается число регрессионных коэффициентов и еще единица, приходящаяся на среднее выборочное).

Вес отсекаемого хвоста распределения Фишера и есть значимость модели в целом.

Если переменные «время выдержки» и «время монотонной работы» независимы (т.е. корреляция между ними равна нулю), то регрессионные коэффициенты вычисляются по той же формуле $a_x = r_{xz} * s_z/s_x$, $a_y = r_{yz} * s_z/s_y$, а свободный член b вычисляется из соотношения

$$\bar{z} = a_x \bar{x} + a_y \bar{y} + b.$$

Точно так же суммируются и их независимые вклады в показатель качества модели R^2 .

Однако в случае наличия корреляции между независимыми переменными (что имеет место практически всегда) ситуация усложняется.

Как мы разбирали в прошлой главе, коэффициент корреляции двух переменных может рассматриваться как мера силы связи между переменными. Тот же смысл можно придать и равному ему стандартизованному коэффициенту регрессии (безразлично, какую из двух переменных считать независимой, какую — зависимой).

Для множественной регрессии в случае некоррелирующих независимых переменных такая интерпретация стандартизованных регрессионных коэффициентов также вполне допустима.

Проблемы возникают тогда, когда о силе связи зависимой переменной с независимыми пытаются судить по стандартизованным регрессионным коэффициентам при наличии корреляции между независимыми переменными (такую ошибку можно довольно часто найти в публикациях). Отличие стандартизованных регрессионных коэффициентов от коэффициентов корреляции тем больше, чем больше корреляция между независимыми переменными. Ситуация, когда эта корреляция становится слишком велика, в литературе называется *мультиколлинеарностью* и считается препятствием применимости регрессионного анализа. Суть проблемы можно прояснить примером: если представить ситуацию так, что у двух независимых пе-

ременных есть «общая часть», которая и порождает корреляцию между ними, то на вопрос о силе связи между зависимой и независимыми переменными можно дать два ответа, исходя из разного отношения к этой общей части. Если не принимать во внимание то, что эта часть — общая, то силу связи с каждой переменной в отдельности правильнее характеризовать коэффициентом корреляции. Если иметь в виду зависимость в совокупности, то, характеризуя обе связи зависимой с двумя независимыми переменными коэффициентами корреляции, общую часть мы посчитаем дважды, а это может не соответствовать вопросу, ответ на который мы хотим найти в наших данных⁷⁸.

Отметим еще, что регрессионные коэффициенты сложным и трудно предсказуемым образом перераспределяют общую часть между независимыми переменными, так что положительная корреляционная связь с одной переменной может преобразоваться в отрицательный регрессионный коэффициент, отдав общую часть даже с избытком второй независимой переменной. В каких-то случаях дело обстоит еще сложнее, и стандартизованный регрессионный коэффициент для этой второй переменной может превзойти единицу, что удивит понимающего читателя (см. пример 10.3(9)).

10.2. Иерархическая регрессия, анализ медиации и модерации

10.2.1. Основные идеи и примеры

Большинство независимых переменных, которые используются в психологии, так или иначе коррелируют между собой. Это означает, что по причинам, рассмотренным в предыдущем параграфе, во множественной регрессии исследователь может основывать свои выводы на показателях значимости модели и коэффициенте детерминации, но ограничен при оценке вклада (эффекта)⁷⁹ каждой из независимых переменных. Дальнейшее развитие множественного регрессионного анализа привело к появлению трех новых методов, которые позволили проверять некоторые (довольно узкие) гипотезы о вкладах независимых

⁷⁸ Неприятные эффекты мультиколлинеарности могут наблюдаться и в отсутствие влияния общей части дисперсии независимых переменных на зависимую. Достаточно наличия этой общей части. Однако пример таких данных слишком сложен для нашего учебника.

⁷⁹ Далее мы используем термины «эффект» и «вклад» независимых переменных в зависимую, говоря о регрессионных коэффициентах и их значимости, для того, чтобы избежать иллюзии причинности, описанной выше (и которая создается при использовании слов «влияние» и «предсказание» в регрессионном анализе).

переменных, с минимальным риском искажений из-за мультиколлинеарности независимых переменных. Все эти методы основаны на проведении нескольких регрессионных анализов и сопоставлении между собой их результатов.

Иерархический регрессионный анализ направлен на проверку гипотезы о независимом вкладе одной из переменных в зависимую переменную, после статистического контроля других переменных. Для иллюстрации вернемся к примеру со «временем выдержки» дошкольников и последующей успеваемостью (подпараграф 9.1.4). Понятно, что успеваемость зависит и от исходного уровня интеллекта, который, вероятнее всего, связан со «временем выдержки» дошкольника наедине с конфетой. Критики могут предположить, что полученная нами связь «время выдержки — успеваемость» полностью объясняется интеллектом дошкольников, и эксперимент с отсрочкой вознаграждения не добавляет нового к нашему знанию об успеваемости. Изменим психологическую гипотезу так, чтобы она отвечала иерархическому регрессионному анализу: умение дошкольника отсрочить вознаграждение связано с последующей успеваемостью, после статистического контроля (или независимо от) его уровня интеллекта. В отличие от множественной регрессии, в иерархической регрессии заданы уровни независимых переменных: переменные первого уровня – те, эффект которых мы хотим проконтролировать, и переменные второго уровня – те, эффект которых мы хотим установить⁸⁰ (рис. 10.2.1(1)).

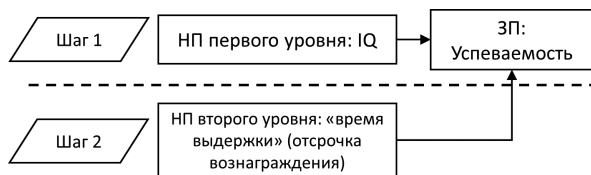


Рис. 10.2.1(1). Пример модели иерархического регрессионного анализа (НП — независимая переменная, ЗП — зависимая переменная)

Анализ медиации применяется для установления прямого и косвенного (непрямого) эффекта влияния независимой переменной на

⁸⁰ Уровней переменных может быть больше: на каждом следующем шаге независимая переменная уровня n — переменная, уникальную связь которой с зависимой переменной мы хотим установить, после контроля всех независимых переменных всех предыдущих уровней.

зависимую. Гипотеза звучит следующим образом: по крайней мере, часть эффекта влияния независимой переменной в отношении зависимой — прямая, а проходит через третью переменную-медиатор. Например, вряд ли «время выдержки» напрямую определяет успеваемость через несколько лет — разумно предположить, что умение ждать вознаграждения в дошкольном возрасте помогает чему-то, что позже определяет успеваемость, — скажем, концентрации внимания в случае монотонной деятельности в классе. Общий эффект независимой переменной на зависимую состоит из двух эффектов — прямого и косвенного (рис. 10.2.1(2)). При этом косвенный эффект состоит в том, что дети, которые дольше не берут конфету, позже лучше и дольше сосредотачиваются на предметах в школе, и именно этот навык помогает их успеваемости. В отличие от множественной регрессии, в анализе медиации установлена последовательность связей между переменными — что и через какие переменные мы пытаемся объяснить. В отличие от иерархической регрессии, в центре внимания здесь не «прирост» после добавления переменной (ведь общий эффект не изменится с добавлением медиатора), а содержание общего эффекта. Как мы покажем в подпараграфе 10.2.3, это означает, что если в иерархической регрессии ключевой интерес представляют коэффициенты детерминации и F-отношения, а также значимость модели, то в анализе медиации — коэффициенты регрессионного уравнения и их значимость.

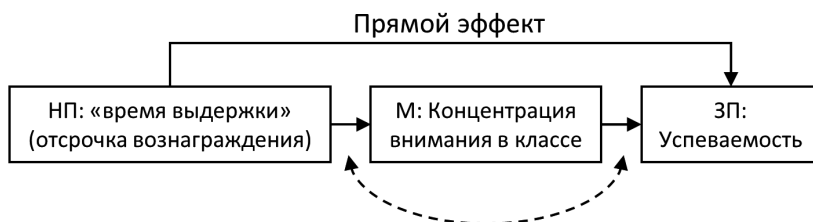


Рис. 10.2.1(2). Пример модели анализа медиации

Анализ модерации используется для проверки гипотезы о различиях во вкладе независимой переменной в зависимую при разных уровнях переменной-модератора. Например, можно предположить, что «время выдержки» у дошкольников по-разному сказывается на успеваемости в отношении конкретного предмета (скажем, русского языка и литературы или математики) в зависимости от отношения к этому предмету. Если ребенку предмет нравится, ему интересно, то для освоения предмета ему не требуется заставлять себя длительно концентрироваться на уроке и время, которое он может ждать награждения, не столь важно.

Если же предмет не нравится, возможности саморегуляции и терпение крайне важны — без них успеваемость вряд ли будет высока. На рис. 10.2.1(3) показан пример такой модели: всего в ней возможно три эффекта. Во-первых, «время выдержки» может способствовать успеваемости по всем предметам, как любимым, так и не любимым (основной эффект независимой переменной). Во-вторых, очевидно, что отношение к предмету само по себе связано с успеваемостью (основной эффект модератора). В-третьих, «время выдержки» может более сильно сказываться на успеваемости у тех детей, которые данный предмет не любят, и слабее — у тех, для кого этот предмет любимый (эффект взаимодействия независимой переменной и модератора). В анализе модерации исследователя интересует именно этот эффект взаимодействия — эффект модерации. Внимательный читатель, вероятно, заметил, что задачи анализа модерации те же, что и в двухфакторном дисперсионном анализе — в обоих методах речь идет о разделении основных эффектов и эффекта взаимодействия, однако задача эта решается разными способами. В дисперсионном анализе мы могли сравнить внутригрупповые и межгрупповые дисперсии; в регрессионном анализе нам придется создать новую переменную, описывающую взаимодействие независимой переменной и модератора, и проверить, добавляет ли она что-то к объяснению дисперсии зависимой переменной. В отличие от множественной регрессии, анализ модерации сосредоточен на поиске эффектов взаимодействия. Как мы покажем в подпараграфе 10.2.4, по процедуре анализ модерации проводится так же, как и иерархическая регрессия — с той разницей, что на шаге 2 добавляется не другая независимая переменная, а переменная, созданная из переменных шага 1. От анализа медиации анализ модерации отличается тем, что задачей является не разделение прямого и косвенного эффектов независимой переменной, а поиск различий в ее эффектах в зависимости от модератора.



Рис. 10.2.1(3). Пример модели анализа модерации

Далее мы кратко изложим математическую суть иерархического регрессионного анализа и анализа модерации. Подробное описание

анализа медиации (и модерации тоже) можно найти на сайте разработчиков современного варианта проведения этих методов Кристофера Причера и Эндрю Хейеса (<http://www.quantpsy.org/medn.htm> и <http://afhayes.com/>), которые содержат не только материалы, но и программы (в том числе для SPSS), позволяющие рассчитать косвенный эффект и оценить его значимость.

10.2.2. Иерархическая регрессия (расчеты)

Процедура иерархической регрессии требует проведения анализа несколько раз и сопоставления каждого следующего шага с предыдущим. Для примера на рис. 10.2.1(1) таких шагов два: на первом шаге проводится простая линейная регрессия для выявления связи «IQ — успеваемость», на втором шаге — множественная регрессия с двумя независимыми переменными IQ и «время выдержки». Результаты сравниваются, т.е. и нуль-гипотеза, и альтернативная гипотеза являются гипотезами о значимости различий: «изменение объясняемой моделью дисперсии зависимой переменной на шаге 2 не отличается от нуля» и «изменение объясняемой моделью дисперсии зависимой переменной на шаге 2 отличается от нуля» соответственно. Исходные данные и основные формулы остаются неизменными: измерения заданы тройками значений:

$$(x_1, y_1, z_1), \quad \dots, (x_n, y_n, z_n).$$

а формула регрессии

$$z = a_x x + a_y y + b.$$

На первом шаге регрессии мы полагаем $a_y = 0$, что превращает формулу в уравнение простой линейной регрессии. Иными словами, мы ищем наилучшую в смысле наименьших квадратов плоскость, параллельную оси y . На втором шаге мы рассчитываем, какое «улучшение» в отношении наименьших квадратов дает возможность расположить плоскость «под углом» к оси y . Изменение коэффициента детерминации характеризует изменение (улучшение) процента дисперсии зависимой переменной, объясняемой моделью (относительно общей дисперсии):

$$\Delta R^2 = R_2^2 - R_1^2 = \frac{S_{model1} - S_{model2}}{S_{total}}.$$

Заметим, что общую дисперсию зависимой переменной теперь можно записать двумя способами: как сумму дисперсии, объясняемой моделью 1, и ее дисперсии ошибки и как сумму дисперсии, объясняемой моделью 2, и ее дисперсии ошибки. Для суммы квадратов (без деления на число степеней свободы):

$$S_{total} = S_{model1} + S_{error1} = S_{model2} + S_{error2}$$

F-отношение для изменения модели аналогично обычному F-отно-

шению для регрессии, которое обсуждалось выше: но в числителе теперь стоит разность сумм квадратов двух моделей, деленных на разность их степеней свободы, а знаменатель считается по второй модели. Другими словами, F-отношение для изменения модели описывает «выигрыш» в объясняемой дисперсии относительно дисперсии ошибки:

$$\Delta F = \frac{(S_{model2} - S_{model1}) / (df_2 - df_1)}{S_{error2} / df_{error2}}$$

Значимость изменения модели можно теперь оценить как вес отсекаемого ΔF хвоста распределения Фишера, первая степень свободы которого равна $df_2 - df_1$ — разница между числом параметров во второй и первой моделях, а вторая df_{error2} — общее число испытуемых минус число независимых переменных минус 1. Для нашего примера $df_2 - df_1 = 2 - 1 = 1$, а $df_{error2} = n - 3$. Значимость изменения модели позволяет сделать выбор между нуль-гипотезой и альтернативной гипотезой, отвечая на вопрос о том, можно ли считать, что «время выдержки» важно для успеваемости, вне зависимости от исходного уровня интеллекта в дошкольном возрасте. Проблема мультиколлинеарности независимых переменных в иерархическом регрессионном анализе остается, но она перестает влиять на вывод: ведь вся общая дисперсия между независимыми переменными, которая разделена у них с зависимой переменной, приписывается переменным первого уровня («контролируется»). А определяется уникальный вклад переменной второго уровня. Регрессионные коэффициенты в иерархической регрессии не являются центральным предметом интереса для исследователя.

10.2.3. Анализ модерации (расчеты)

На рис. 10.2.1(3) приведен пример, в котором мы предполагаем, что вклад «времени выдержки» в успеваемость выше у тех, кто не любит данный предмет, и ниже у тех, кто любит данный предмет. Это означает, что, если мы рассматриваем тех, кому предмет не нравится, облако точек-измерений на графике «время выдержки — успеваемость» будет более вытянуто вдоль некоторой прямой (связь будет сильнее, а наше предсказание — точнее), а если тех, кому нравится, — оно будет более округлым. Во всех случаях анализа модерации речь идет о том, что линии существенно различаются при разных уровнях модератора.

Во множественной регрессии по тройкам значений $(x_1, y_1, z_1), (x_2, y_2, z_2), \dots, (x_n, y_n, z_n)$ в трехмерном пространстве подбирается плоскость, лежащая ближе всего к точкам, координаты которых заданы этими наборами значений. В анализе модерации мы, говоря образно, «разрешаем» этой плоскости быть изогнутой, иметь седловину — и оцениваем, дает ли нам это разрешение «выигрыш»

в объяснении дисперсии зависимой переменной. Нулевая гипотеза в данном случае такова: эффект независимой переменной в отношении зависимой не различается на разных уровнях (при разных значениях) модератора. Альтернативная гипотеза утверждает, что различие есть. Новая переменная, характеризующая модерацию, несет в себе информацию о сочетании разных значений независимой переменной и модератора: в случае если у испытуемого крайне высокие или крайне низкие значения, показатель по этой переменной должны быть высок. Если же одно из значений крайне низкое, а другое — крайне высокое, показатель должен быть низок. В нашем примере на одном полюсе оказываются школьники, «умеющие ждать и любящие предмет», а также «не умеющие ждать и не любящие предмет», а на другом — «умеющие ждать и не любящие предмет» и «не умеющие ждать, но любящие предмет»⁸¹.

Для вычисления новой переменной независимая переменная и модератор центрируются, а затем для каждого испытуемого (измерения) два его значения по этим переменным перемножаются. Затем проводится иерархическая регрессия: на первом ее шаге в модель добавляются центрированные независимая переменная и модератор, на втором — переменная их взаимодействия (говорят: «их продукт»). Изменение в проценте объясняемой дисперсии зависимой переменной и значимость изменения позволяют выбрать между нулевой гипотезой и альтернативой.

Центрирование необходимо, чтобы избежать проблемы мультиколлинеарности: если перемножить сырые значения, новая переменная может коррелировать с исходными и добавление ее во множественную регрессию исказит регрессионные коэффициенты независимой переменной и модератора (см. параграф 10.1). Коэффициенты детерминации R^2 и значимость модели и ее изменения останутся теми же, но интерпретировать результаты будет сложно из-за нестабильности коэффициентов. Центрируя переменные, мы достигаем того, что они не коррелируют со своим продуктом (переменной взаимодействия). Уравнение регрессии в анализе модерации выглядит следующим образом:

$$z = a_x x + a_y y + a_{xy} xy + b.$$

При этом коэффициент a_x — это то, насколько изменение x на

⁸¹ То, что первый полюс включает таких разных школьников, не должно смущать: роль умения ждать и отношения к предмету самих по себе оцениваются и учитываются в модели, и новая переменная лишь добавляет информацию о роли их соотношения.

единицу приводит к изменению z , когда y равен нулю (поскольку мы центрировали переменные, нуль — среднее значение). Т.е. при увеличении времени выдержки дошкольника на единицу при нейтральном (среднем) отношении к предмету успеваемость увеличивается на a_x .

Упражнение. Самостоятельно опишите, что означает a_y .

Коэффициент a_{xy} описывает то, насколько меняется эффект x при изменении y от 0 до 1. Это видно, если переписать уравнение регрессии:

$$z = (a_x + a_{xy}y)x + a_yy + b.$$

Для удобства понимания и иллюстрации полученного эффекта взаимодействия обычно применяют простые регрессии: разделяя испытуемых на группы с разными уровнями по переменной-модератору (например, любящих предмет школьников, не любящих и относящихся к нему нейтрально) и затем проводя анализ отдельно в этих группах (а графически: рисуя несколько прямых на плоскости). В третьем параграфе мы приведем пример такой иллюстрации.

10.2.4. Медиация, модерация и опосредствование

Медиацию и модерацию нередко путают — в русском языке еще и из-за того, что результаты обоих методов интерпретируется в исследованиях как «опосредствование». Медиацию применяют, если задача — установить природу некоторого эффекта независимой переменной на зависимую. Это поиск «путей»⁸², по которым происходит эффект, который можно выразить предложением: «Эффект X в отношении Y обусловлен его общей дисперсией с M ». Эффект модерации — это эффект взаимодействия между независимыми переменными, который можно описать так: «Эффект X в отношении Y различается при разных значениях M ».

Говоря метафорически, медиатором можно назвать провод между лампочкой и розеткой — именно по нему идет ток, и без этого лампочка не зажжется, а модератором — выключатель. Когда он выключен, связи между розеткой и лампочкой нет, когда включен — есть. Действия психолога в анализе медиации схожи с работой электрика, который в пучке проводов ищет единственный работающий провод («путь», благодаря которому лампочка горит). Анализ модерации же сродни проверке выключателя: изменяется ли освещение, если на него нажать?

Ранее для того, чтобы различить гипотезы о медиации и модерации в своих исследованиях, предлагалось проверить корреля-

⁸² Анализ медиации — наиболее простой пример так называемого путевого анализа, группы методов, направленных на исследование прямых и косвенных эффектов ряда переменных, проверку теоретических моделей, включающих регрессионные связи.

ции между переменными. Так, в случае медиации все три переменные коррелируют (чтобы объяснить связь независимых и зависимой переменных, нужно, чтобы она была), а модерацию часто применяют именно в тех случаях, когда ожидаемой корреляции нет или она различна в разных работах.

«Эмпирическая» постановка задач для обработки данных действительно имеет место в науке (например, получив связь между независимыми и зависимой переменными, исследователь задается вопросом, чем она объясняется, а отсутствие ожидаемой связи приводит к размышлениям, не зависит ли эта связь от каких-то других переменных). Однако это «правило» применимо далеко не всегда. Косвенный эффект (эффект медиации) может быть предметом исследования и достичь приемлемой значимости, даже если общий и/или прямой эффекты независимой переменной приемлемой значимости не достигают⁸³. Модерация также может проводиться и в случае, если связь между независимыми и зависимой переменными установлена.

Обратим внимание, что ни медиация, ни модерация не означают причинно-следственной связи — что характерно для любых методов обработки данных. Делая психологический вывод из исследования, медиацию нередко интерпретируют как свидетельство в пользу механизма действия независимой переменной (термин «опосредствование» употребляется тогда в смысле «путь» и «механизм»), а модерацию — как доказательство разницы в механизмах («опосредствование» в смысле «различий в механизмах», «различий в способах действия»). Однако вывод этот не относится к сфере статистики и не гарантируется ею; теория, дизайн и процедура исследования помогают исследователю быть более или менее уверенным и обосновать свои выводы. Интересно, что некоторые специалисты по статистике предлагают употреблять термин «медиация», только если речь идет об экспериментальном или лонгитюдном исследовании и все переменные измерены как минимум трижды⁸⁴, а в других случаях говорить только о косвенном эффекте.

⁸³ Один из примеров такой ситуации — если действуют другие медиаторы, не учтенные и не измеренные в исследовании, в частности так называемые супрессоры — переменные, подавляющие основной эффект (см. Rucker et al., 2011).

⁸⁴ Т.е. можно проверить, как изменения в независимой переменной и медиации с течением времени связаны с последующими изменениями в зависимой переменной (Little, 2013).

10.3. Регрессия в случае многих переменных. Практикум

Пример 10.3(1). Матрица корреляций. У 40 менеджеров по продажам крупной компании измерялись возраст (переменная *Age*), интеллект (*IQ*), опыт работы по специальности (*WorkExperience*), трудовая мотивация (*Motivation*), коммуникативные навыки (*CommunicationSkills*), а затем их начальники оценивали их продуктивность по 100-балльной системе (*WorkProductivity*). Данные приведены в файле *WorkProductivity.sav*. Задачи перед практическим психологом на основе этих данных могут стоять разные: выяснить связь продуктивности и психологических факторов, обнаружить индикаторы, по которым можно ее предсказать, разработать систему, ориентируясь на которую можно уже при приеме на работу оценить продуктивность будущих сотрудников, и т.п. В любом случае одна из первых задач — установление связи между переменными.

Корреляции между шестью нашими переменными удобно выводить в виде матрицы. Для этого в открывшемся окне после выбора **Анализ — Корреляции — Парные (Analyze — Correlation — Bivariate)** переносим в поле **Переменные (Variables)** все наши переменные. После нажатия **OK** получаем матрицу.

На главной диагонали матрицы проставлены единицы (корреляции переменных с собой), в симметричных относительно главной диагонали клетках стоят равные числа, поскольку при расчете корреляций порядок переменных не важен. Под корреляциями подписаны их значимости, а еще ниже количество учтенных парных данных. Заметим, что корреляции интеллекта, опыта работы, мотивации, коммуникативных навыков и возраста крайне слабы: максимальная корреляция составляет 0.15: более старшие работники имеют немного более длительный опыт работы по специальности, что закономерно. Ни в одном случае коэффициенты корреляции не достигают уровня значимости 0.05. Напротив, средние по силе положительные корреляции связывают продуктивность на рабочем месте с опытом работы, мотивацией и коммуникативными навыками.

Таблица, приведенная ниже, называется матрицей корреляций. По ней можно сразу оценить значимость каждой корреляции. Однако оценка эта проводится для каждой пары отдельно, а рассчитывается сразу много корреляций (для n переменных матрица содержит $n(n-1)/2$ различных корреляций). В этом случае необходима поправка значимости, похожая на ту, с которой мы знакомы в подпараграфе 7.1.4, посвященном апостериорным сравнениям в дисперсионном анализе. Однако до сих пор согласия в статистическом сообществе о такого рода поправках для корреляций нет⁸⁵.

⁸⁵ Тем не менее, к значимостям, полученным из больших матриц корреляций, в редакциях ведущих журналов отношение достаточно критическое.

Таблица 10.3(2). Матрица корреляций

		IQ	Work Experience	Motivation	Communication Skills	Age	Work Productivity
IQ	Корреляция Пирсона	1	0.060	-0.110	0.031	-0.095	0.246
	Знач. (двухсторонняя)		0.696	0.472	0.838	0.535	0.125
	N	45	45	45	45	45	40
Work Experience	Корреляция Пирсона	0.060	1	-0.038	-0.048	0.151	0.376*
	Знач. (двухсторонняя)	0.696		0.805	0.755	0.323	0.017
	N	45	45	45	45	45	40
Motivation	Корреляция Пирсона	-0.110	-0.038	1	-0.126	-0.093	0.342*
	Знач. (двухсторонняя)	0.472	0.805		0.408	0.545	0.031
	N	45	45	45	45	45	40
Communication Skills	Корреляция Пирсона	0.031	-0.048	-0.126	1	-0.064	0.354*
	Знач. (двухсторонняя)	0.838	0.755	0.408		0.677	0.025
	N	45	45	45	45	45	40
Age	Корреляция Пирсона	-0.095	0.151	-0.093	-0.064	1	-0.024
	Знач. (двухсторонняя)	0.535	0.323	0.545	0.677		0.885
	N	45	45	45	45	45	40
Work Productivity	Корреляция Пирсона	0.246	0.376*	0.342*	0.354*	-0.024	1
	Знач. (двухсторонняя)	0.125	0.017	0.031	0.025	0.885	
	N	40	40	40	40	40	40

*. Корреляция значима на уровне 0,05 (двухсторонняя)

Пример 10.3(3). Применение множественной регрессии в предсказании

Как говорилось в параграфе 9.1, именно задачи предсказания значенной зависимой переменной по значениям независимой и оценки точности этого предсказания — основная сфера прямого применения регрессионного анализа. При этом результаты особенно информативны, если независимые переменные не коррелируют. Данные примера 10.3.1 — удобная иллюстрация такой ситуации. Перед психологом в организации стояло две

задачи: выяснить, на основе каких индикаторов и насколько точно можно предсказать продуктивность менеджеров, и отобрать из пяти новых кандидатов на работу (номера 41 — 45) потенциально лучшего.

Упражнение 10.3(4). С теми же данными (файл *WorkProductivity.sav*) проведите множественный регрессионный анализ по той же схеме, что и в параграфе 9.3, но включив в список независимых переменных все пять индикаторов.

Регрессионная модель объясняет 48.2% дисперсии продуктивности на рабочем месте⁸⁶, а статистическая оценка значимости: $F = 6.336$, $p < 0.001$. Это не идеальное предсказание, но для практической сферы — неплохой результат.

Близость корреляций между независимыми переменными нулю в таблице 10.3(2) позволяет нам рассматривать таблицу коэффициентов регрессии с минимальным риском искажений из-за мультиколлинеарности. Обратите внимание, что значимость модели не означает, что все индикаторы одинаково «хороши»: например, возраст не связан с продуктивностью, а связь интеллекта с продуктивностью не достигает принятого уровня значимости 0.05.

Таблица 10.3(5). Регрессионные коэффициенты

Модель	Нестандартизованные коэффициенты		Стандартизованные коэффициенты	t	Значимость
	B	Стандартная ошибка	Бета		
1 (Константа)	-75.813	36.791		-2.061	0.047
IQ	0.518	0.288	0.225	1.797	0.081
WorkExperience	3.856	1.265	0.385	3.049	0.004
Motivation	0.884	0.263	0.418	3.356	0.002
Communication Skills	0.902	0.289	0.391	3.126	0.004
Age	-0.057	0.497	-0.014	-0.114	0.910

а. Зависимая переменная: WorkProductivity

Упражнение 10.3(6). Проведите регрессионный анализ повторно, убрав из списка независимых переменных сначала возраст, затем — возраст и IQ. Убедитесь, что знание возраста не помогает в предсказа-

⁸⁶ Скорректированный R^2 ниже, поскольку независимых переменных в модели довольно много; он составляет 40.6%.

нии — его удаление не сказывается на R^2 . Обоснуйте, почему изменяется F-отношение и скорректированный R^2 .

Удаление из модели переменной IQ приводит к ухудшению ее точности: $R^2 = 43.2\%$, даже если судить по скорректированному коэффициенту детерминации «выигрыш» в одну независимую переменную не «перекрывает» проигрыша в объясняемой дисперсии (скорректированный $R^2 = 38.5\%$). Важны ли эти 5% объясняемой дисперсии, по сравнению с затратами на проведение теста на интеллект, — решать практику.

Запишем уравнение регрессии в нестандартизованном виде для модели без учета возраста: предсказанную моделью продуктивность каждого менеджера можно рассчитать по формуле

$$y = 0.521x_1 + 3.832x_2 + 0.885x_3 + 0.907x_4 - 77.791,$$

где x_{1-4} — баллы каждого из менеджеров по интеллекту, опыту работы, мотивации и коммуникации, соответственно.

Упражнение 10.3(7). Сохраните предсказанные нестандартизованные значения для каждого менеджера, добавив галочку в пункте **Остатки — Нестандартизованные (Residuals — Unstandardized)**. В таблице с сырыми данными появились не только предсказанная продуктивность, но и индивидуальные отличия этого предсказания от реальной продуктивности каждого менеджера. Убедитесь, что остатки получаются вычитанием из действительной продуктивности предсказанных значений. Метод наименьших квадратов стремится свести сумму квадратов этих остатков по испытуемым (наблюдениям) к минимуму.

Заметим, что величина остатка, являющаяся для математика «ошибкой», может быть важна для психолога: например, у менеджера под номером 22 предсказанная продуктивность 44.37, а в действительности он работает много лучше — на 78. Вопрос о том, что именно, не учтенное в нашей модели, помогает ему в работе (проясненный, например, в интервью с этим человеком или наблюдении за ним), может помочь развитию и теории, и практики психологии.

Наконец, предсказанные значения новых пяти кандидатов можно взять за основу для сравнения: согласно регрессионному уравнению, лучшим будет кандидат № 44 с предсказанной продуктивностью в 57.2990 балла.

В завершение примера еще раз обратим внимание, что такое прямое применение множественной регрессии характерно для узких практических задач (предсказания), а интерпретация коэффициентов регрессионного уравнения не вызывает вопросов, только если корреляции независимых переменных близки к нулю, что в психологической реальности — крайне редкий случай. Далее рассмотрим, что происходит в более сложных, но значительно более распространенных ситуациях.

Пример 10.3(8). Интерпретация множественной регрессии при мультиколлинеарности. В примере о конфликтности в браке (пример 9.3(8)) мы не обсудили, как можно интерпретировать результат. Является ли долгое разглядывание картинок в экспериментальной ситуации причиной конфликтности отношений в семье через два года? Очевидно, нет, поскольку считанные минуты взаимодействия с картинками в экспериментальной ситуации не могут произвести значительное долговременное воздействие. Скорее, время разглядывания картинок является индикатором, характеризующим состояние какой-то более существенной характеристики супругов. Претендентом на эту роль может оказаться удовлетворенность браком. В файле *ConflictM.sav* приведены данные эксперимента о связи времени разглядывания картинок с устойчивостью брака, но с добавлением результата анкетирования, по которому оценивалась удовлетворенность браком.

Мы имеем теперь три переменные: *lookingtime*, *conflictiness*, *satisfaction*, измеряющие указанные параметры. Матрица корреляций между ними фактически содержит три корреляции: *satisfaction* дает корреляцию 0.708 с *conflictiness* и 0.598 с *lookingtime*. Корреляция между *conflictiness* и *lookingtime* равна 0.537. Иными словами, можно опасаться искажения регрессионных коэффициентов вследствие мультиколлинеарности.

Проведем регрессионный анализ вновь, добавив в окне регрессии к независимым переменным переменную *satisfaction*, и нажмем *OK*.

Рассмотрим три таблицы. В первой мы находим существенное увеличение показателя R^2 (с 0.288 в случае простой регрессии с переменной *lookingtime* — см. 9.3(8), до 0.521), свидетельствующее о возросшем качестве модели (ее регрессионное уравнение мы давали в подпараграфе 9.1.5).

Сводка для модели

Модель	R	R-квадрат	Скорректированный R-квадрат	Стд. ошибка оценки
1	0.722 ^a	0.521	0.489	1.47457

a. Предикторы: (конст) *satisfaction*, *lookingtime*

Во второй дана общая оценка значимости модели в целом (см. 10.1.2: F-отношение (16.329) и собственно значимость (0.000, что в силу правил округления говорит о том, что значимость по крайней мере меньше 0.0005).

Дисперсионный анализ^а

Модель	Сумма квадратов	ст.св.	Средн. квадрат	F	Знч.
Регрессия	71.012	2	35.506	16.32	.000 ^б
Остаток	65.231	30	2.174		
Всего	136.242	32			

^а. Зависимая переменная: conflictness

^б. Предикторы: (конст) satisfaction, lookingtime

В третьей таблице сначала даны нестандартизованные коэффициенты (свободный член 6.289 и угловые коэффициенты 0.095 для времени разглядывания изображений и 0.328 для удовлетворенности браком). По этим коэффициентам мы можем составить регрессионное уравнение и с точностью более 50% (R-квадрат равен 0.521) предсказывать конфликтность отношений, которую семья будет иметь через два года.

О силе связи независимых и зависимой переменных обычно судят по стандартизованным коэффициентам, однако чем выше корреляция между независимыми переменными (а в нашем случае она достаточно высока), тем меньше смысла в такой интерпретации результатов (см. 10.1.2). Ниже мы рассмотрим вопрос подробнее.

	Нестандартизованные коэффициенты		Стандартизованные коэффициенты	t	Знч.
	B	Стд. ошибка	Бета		
(Константа)	6.289	1.982		3.172	0.003
1 lookingtime	0.095	0.085	0.176	1.119	0.272
satisfaction	0.328	0.086	0.602	3.820	0.001

^а. Зависимая переменная: conflictness

Заметим, что наша регрессионная модель оставляет вопрос о статусе времени разглядывания изображений не проясненным. Как мы говорили, статистика сама по себе не может помочь определить направление

и характер связей между переменными⁸⁷. Приблизиться к их пониманию позволяет вовсе не статистическая обработка данных, а дизайн исследования. Например, если при проведении эксперимента исследователь сам управляет независимой переменной и пытается контролировать другие переменные — это дает ему некоторые гарантии, что именно то, на что он повлиял (независимая переменная) — причина тех изменений, которые произошли (зависимая переменная, см. Корнилова, 2002).

Пример 10.3(9). Переименуйте переменную *lookingtime*, присвоив ей имя *alcohol*. Пусть теперь интерпретация переменных такова: В эксперименте участвовали добровольцы, каждому из которых давали дозу алкоголя (количество в миллилитрах кодируется переменной *alcohol*). У них затем измеряли уровень ситуативной конфликтности в межличностных отношениях и показатель по тесту «удовлетворенность жизнью». Хотя результат регрессионного анализа будет тем же самым, но смысла он не будет иметь вовсе. При таком экспериментальном дизайне единственной независимой переменной является *alcohol*, а две другие переменные характеризуют коррелирующие между собой следствия ее изменения.

Важнейший вывод из сказанного: модель множественной регрессии в своих рамках хорошо служит для предсказания результатов и для общей оценки доли дисперсии зависимой переменной, объясняемой совокупностью независимых. Для понимания характера связей нужны содержательный анализ, формулирование разумных гипотез и построение более сложных моделей, учитывающих, что влияние переменной может быть опосредованным другими переменными, что она сама может быть таким посредником (что мы рассмотрим далее).

⁸⁷ Часто, поняв это, психологи задают следующий вопрос: зачем она тогда нужна? Существует два пути использования статистических методов в науках: описательный и объяснительный. Первый случай более «честный» с точки зрения математики — мы ищем модель, наилучшим образом описывающую полученные («реальные») данные, но не считаем, что эта модель — истина. Важно лишь то, что эта модель позволяет нам делать — например, в экономике мы можем надежно предсказать нечто на основе серии индикаторов, и нам не важно, как именно эти индикаторы «работают». Для психологии чаще важен второй случай — мы пытаемся привлечь статистику для «объяснения» реальности, для проверки того, «работает» ли наша теория. Такое использование статистики уязвимо для критики. Оно возможно, только если теория и дизайн исследования (например, эксперимент) позволяют исследователю обосновать (но все равно не доказать окончательно) свои выводы.

Пример 10.3(10). «Бессмысленные» интерпретации при мультиколлинеарности нередко имеют место в случаях, если сложные многоуровневые психологические модели проверяются с использованием линейной множественной регрессии. Рассмотрим следующий пример.

Одна из самых известных в психологии здоровья модель предсказания поведения человека — теория запланированного поведения⁸⁸. Она схематически изображена на рис. 10.3(11). Предполагается, что поведение (например, «заниматься спортом два раза в неделю по полчаса») зависит от намерения («Я собираюсь заниматься спортом два раза в неделю по полчаса») и воспринимаемого контроля («Я смогу заниматься спортом два раза в неделю по полчаса»). Намерение, в свою очередь, зависит от установок («Мне полезно заниматься спортом два раза в неделю по полчаса»), субъективных норм («Окружающие считают, что мне нужно заниматься спортом два раза в неделю по полчаса») и того же воспринимаемого контроля

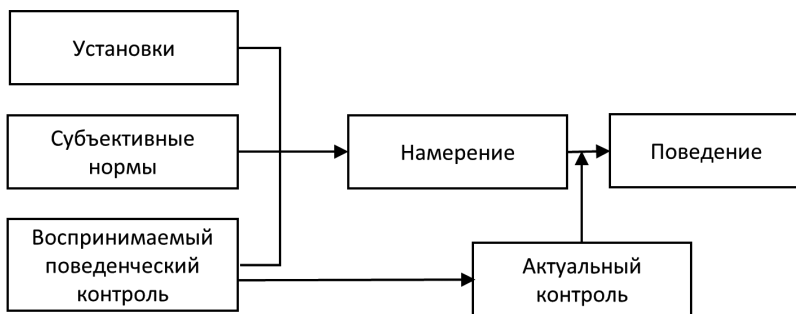


Рис. 10.3(11). Теория запланированного поведения

В файле *HealthBehavior.sav* представлены данные 100 человек, у которых оценивались установки (переменная *Attitude*), субъективные нормы (переменная *SocNorms*) и воспринимаемый контроль (переменная *Control*) в отношении физической зарядки; намерение ее делать (переменная *Intention*), а через неделю — действительная частота зарядки (переменная *Behavior*). Каждая переменная оценивалась при помощи оценки приведенных выше утверждений по 10-балльной шкале (0 — полностью не согласен, 10 — полностью согласен).

⁸⁸ Ajzen, I. (2001). Construction of a standard questionnaire for the theory of planned behavior. Электронный ресурс: <http://www-unix.oit.umass.edu/~ajzen/pdf/tpb.measurement.pdf>.

Упражнение 10.3(12). Рассчитайте матрицу корреляций для всех переменных. По ней предположите (и обоснуйте свое мнение), как будут результаты множественной регрессии в отношении поведения (зависимая переменная).

Проведите множественный регрессионный анализ, используя установки, социальные нормы, воспринимаемый контроль и намерение как независимые переменные, а поведение — как зависимую. Проинтерпретируйте результат.

С одной стороны, непосредственные связи «заглушают» действие опосредствованных. Хотя все независимые переменные коррелируют с поведением, в регрессионном анализе исследователь легко обнаружит эффекты намерения и контроля, но не установок и социальных норм, хотя именно последние могут быть крайне важны практически. Например, сказать школьнику «соберись бросить курить» чаще бесполезно, но если на его желание влияют субъективные нормы, можно повлиять на представления его и его друзей о курении и том, как другие реагируют на то, что он курит. С другой стороны, переменные, хорошо предсказывающие результат, могут оказаться хорошими индикаторами, а не факторами, влияющими на зависимую переменную.

Пример 10.3(13). Файл *terrible.sav* содержит пример, в котором один стандартизованный регрессионный коэффициент больше единицы, а второй — отрицательный несмотря на то, что корреляция зависимой и соответствующей независимой переменных положительная и достаточно большая. Постройте матрицу корреляции и проведите линейную регрессию. Оцените результаты. Пример показывает, что значительная корреляция независимых переменных (мультиколлинеарность) делает регрессионную модель практически бессмысленной. Если учесть, что для некоррелирующих независимых переменных регрессионные коэффициенты равны корреляциям, то под вопросом оказываются те изменения, которые происходят с регрессионными коэффициентами и делают их отличными от корреляций при росте корреляции между независимыми переменными. Также под вопросом оценивание стандартизованными коэффициентами силы связи между переменными.

Пример 10.3(13). Иерархическая регрессия.

Задача иерархической регрессии — проверить, отличается ли от нуля уникальный вклад независимых переменных в зависимую после статистического контроля (учета в модели) независимых переменных первого уровня.

В файле *WtimeIQ.sav* смоделированы данные примера, описанного в подпараграфе 10.2.1 для 30 школьников: у них измерялись «время выдержки» (*Wtime*), последующая успешность в школе (*Success*) и дополнительно — интеллект (*IQ*). Поскольку уровень интеллекта очевидно важный и хорошо изученный фактор успеваемости, основное внимание уделялось вопросу, добавляет ли «время выдержки» дошкольника что-то к предсказанию последующей успеваемости в школе, после учета *IQ*. Т.е. интеллект — независимая переменная первого уровня, «время выдержки» — второго уровня.

Проведя корреляционный анализ, убедимся, что как интеллект, так и «время выдержки» одинаково тесно связаны с успеваемостью⁸⁹ ($r=0.52$ и $r=0.51$), но связаны и между собой ($r=0.39$). Для иерархической регрессии в меню *Анализ — Регрессия — Линейная (Analyze — Regression — Linear)* найдем над строкой ввода независимых переменных кнопки *Предыдущее/Следующее (Next/Previous)*. Эти кнопки переключают шаги регрессии. Введем *IQ* в строку независимой переменной и нажмем *Next*. Теперь во вновь опустевшее окно независимых переменных нужно добавить переменные, дополнительно введенные на шаге 2 (*Wtime*). Зависимая переменная — *Success*. Наконец, чтобы оценить изменение процента объясняемой дисперсии на втором шаге иерархической регрессии, нажмем справа кнопку *Статистика (Statistics)* и выберем *Изменение R-квадрат (R squared change)*.

По результатам дисперсионного анализа обе модели (шаг 1 и после добавления «времени выдержки» на шаге 2) оцениваются приемлемой значимостью (см. таблицу 10.3(14)), но, что более важно — добавление новой переменной в модель привело к увеличению R^2 с 0.271 до 0.385, т.е. на 42%.

Это число указана во второй строке столбца *Изменение R-квадрат*, и при проверке гипотезы об отличии этого изменения от нуля F-отношению, равному 5.044, соответствует значимость 0.033. Можно предположить, что учет «времени выдержки» дополнительно к интеллекту позволяет точнее предсказать успешность младших школьников.

⁸⁹ Это модельный пример, поэтому связи намеренно преувеличены

Таблица 10.3(14). Сводки для двух моделей

Мо- дель	R	R- ква- драт	Скор- ректи- рованный R-квадрат	Стан- дарт- ная ошибка оценки	Статистика изменений				
					Изме- нение R квадрат	Изме- нение F	ст. св.1	ст. св.2	Знач. Изме- нение F
1	0.520 ^a	0.271	0.244	0.86920	0.271	10.385	1	28	0.003
2	0.621 ^b	0.385	0.340	0.81250	0.115	5.044	1	27	0.033

^a. Предикторы: (константа), IQ

^b. Предикторы: (константа), IQ, Wtime

Как видно по таблице коэффициентов (таблица 10.3(15)), эффекты IQ и «времени выдержки» положительные, хотя эффект IQ и снижается при добавлении «времени выдержки» в модель.

Обратим внимание, что исследователя в данном случае не интересует, является ли интеллект первопричиной или первопричина «время выдержки», действует ли одно через другое (медиация), — он лишь описывает наличие или отсутствие уникального вклада независимой переменной, например, чтобы обосновать важность учета «времени выдержки» в дальнейших исследованиях и концепциях.

Таблица 10.3(15). Коэффициенты в двух моделях

Модель		Нестандарти- зованные коэффициенты		Стандар- тизованные коэффи- циенты	t	Значимость
		B	Стандарт- ная ошибка	Бета		
1	(Константа)	-2.965	1.796		-1.651	0.110
	IQ	0.056	0.018	0.520	3.223	0.003
2	(Константа)	-2.317	1.704		-1.360	0.185
	IQ	0.041	0.018	0.379	2.319	0.028
	Wtime	0.061	0.027	0.367	2.246	0.033

^a. Зависимая переменная: Success

Пример 10.3(16). Часто имеет смысл статистически контролировать переменные, вклад которых в зависимую переменную известен и не обсуждается, и переменные других предметных областей (социодемографические факторы, клинические характеристики и т.п.). Еще

одна типичная ситуация — необходимость статистически учесть исходный уровень зависимой переменной в исследованиях динамики каких-либо процессов.

Упражнение 10.3(17). В файле *SatisfactionChange.sav* приведены показатели изменения в удовлетворенности жизнью людей в начале и середине (через полтора года) экономического кризиса в стране. Помимо этого, оценивались особенности совладания со стрессом по типу конфронтации (*Confrontation*) — склонность и готовность активно противостоять неприятностям и конфликтам, доказывать свое мнение при разногласиях.

Очевидно, что удовлетворенность жизнью людей — довольно стабильная характеристика; меняется она слабо и зависит от уровня при предыдущем замере. Проведите иерархическую регрессию, используя удовлетворенность при первом замере *Satisfaction1* как независимую переменную первого уровня, а *Confrontation* — как второго. Как легко ожидать, удовлетворенность через полтора года сильно зависит от исходной удовлетворенности (корреляция $r = 0.72, R^2 = 52.5\%$). Однако добавление конфронтации приводит к значимому улучшению процента объясняемой дисперсии (пишут: $\Delta R^2 = 8\%, p < 0.001$). Одна из возможных интерпретаций результата: в периоды кризиса именно конфронтация помогает противостоять жизненным трудностям, приводя к улучшению удовлетворенности жизнью.

10.3(18). Анализ медиации

Мы не включили в основной практикум примеры на медиацию, поскольку они не обеспечиваются стандартным набором функций пакета SPSS. Для интересующихся этот практический раздел мы поместили в электронное приложение на сайте <http://mathpsy.com/textbook>.

Пример 10.3(19). Анализ модерации. В примере 9.3(12) (файл *ConflictCB.sav*) мы модифицировали пример со временем просмотра фотографий лиц противоположного пола и конфликтностью, при этом учитывая, родились ли у испытуемых-женщин за два года дети. На уровне простых регрессий мы видели, что коэффициенты регрессии разные в двух группах (родивших и бездетных). Проверим теперь гипотезу о модерации: рождение детей в браке выступает модератором связи времени просмотра фотографий и конфликтности через два года. Конкретнее, связь есть только у тех, кто к моменту второго замера не успел родить детей. На первом шаге центрируем независимую переменную *lookingtime*. Для этого считаем ее среднее (15.9162) и, используя команду **Преобразование — Вычислить (Transform — Compute)**, считаем переменную *lookingtime_centered* = *lookingtime* – 15.9162.

Упражнение 10.3(20). Убедитесь, что среднее значение новой переменной равно нулю, а стандартное отклонение не изменилось.

Далее при помощи той же команды рассчитаем переменную, характеризующую модерацию, перемножив центрированную переменную *lookingtime_centered* и переменную *ChildBirth*, кодированную нулями и единицами⁹⁰: $lookingtimebirth = lookingtime_centered * child_birth$. Далее проведем иерархический регрессионный анализ для зависимой переменной *conflictiness*, добавив на первом шаге *lookingtime_centered* и *ChildBirth*, а на втором — *lookingtimebirth*.

Мо- дель	R	R- квадрат	Скорректи- рованный R-квадрат	Стандарт- ная ошибка оценки	Статистика изменений				
					Изме- нение R квадрат	Изме- нение F	ст. св. 1	ст. св. 2	Знач. Изме- нение F
1	0.540 ^a	0.292	0.245	1.79297	0.292	6.190	2	30	0.006
2	0.616 ^b	0.379	0.315	1.70802	0.087	4.059	1	29	0.053

^a. Предикторы: (константа), *lookingtime_centered*, *ChildBirth*

^b. Предикторы: (константа), *lookingtime_centered*, *ChildBirth*, *lookingtimebirth*

Коэффициенты^a

Модель		Нестандартизованные коэффициенты		Стандартизованные коэффициенты	t	Значимость
		B	Стандартная ошибка	Бета		
1	(Константа)	17.030	0.435		39.155	0.000
	<i>ChildBirth</i>	.251	0.625	0.062	0.401	0.691
	<i>lookingtime_centered</i>	.289	0.082	0.539	3.505	0.001
2	(Константа)	17.016	0.414		41.062	0.000
	<i>ChildBirth</i>	0.247	0.595	0.061	0.415	0.681
	<i>lookingtime_centered</i>	0.434	0.107	0.809	4.073	0.000
	<i>lookingtime-birth</i>	-0.318	0.158	-0.400	-2.015	0.053

^a. Зависимая переменная: *conflictiness*

⁹⁰ В результате половина данных стала нулями — однако, эта потерянная информация будет учтена в модели благодаря переменной *lookingtime_centered*.

Несмотря на общую значимость модели уже на первом шаге (29.2% объясненной дисперсии конфликтности), рождение ребенка не связано с конфликтностью ($\beta = 0.062$) при учете времени просмотра фотографий. Второй шаг добавляет 8.7% объясненной дисперсии, но оценка значимости для улучшения ($p = 0.053$) не позволяет сделать уверенный вывод о модерации. Если вопрос нас серьезно интересует, то исследования следует продолжить. Заметим, что регрессионный коэффициент переменной времени просмотра явно изменился, а коэффициент переменной рождения — нет. Итак, хотя в примере 9.3(12) сравнение простых регрессий позволило нам *предположить* различия между родившими и бездетными женщинами, анализ модерации не подтвердил эту гипотезу.

Если при продолжении исследований мы получили бы уверенное подтверждение гипотезы, то для интерпретации результата надо было бы рассмотреть коэффициенты регрессии, учитывая смысл нуля переменной-модератора:

- В нестандартизованном виде интерпретация была бы такой: модель предсказывает, что увеличение времени просмотра фотографий на единицу у не имеющих детей ($ChildBirth = 0$) приведет к увеличению конфликтности на 0.434 балла. Такое же увеличение времени просмотра фотографий у родивших детей ($ChildBirth = 1$) женщин должно (т.е. ожидается в соответствии с моделью) привести к меньшему увеличению конфликтности — на $0.434 - 0.318 = 0.116$ балла.
- В стандартизованном виде: модель предсказывает, что увеличение времени просмотра фотографий на одно стандартное отклонение у не имеющих детей ($ChildBirth = 0$) приведет к увеличению конфликтности в стандартизованном виде на 0.809 этого отклонения. Такое же увеличение времени просмотра фотографий у родивших детей ($ChildBirth = 1$) женщин должно (т.е. ожидается в соответствии с моделью) привести к меньшему увеличению стандартизованной конфликтности — на $0.809 - 0.400 = 0.409$ стандартного отклонения.

Глава 11.

Непараметрические критерии

11.1. Основные сведения о наиболее часто используемых непараметрических критериях

Непараметрические критерии предъявляют к используемым шкалам и распределениям переменных более низкие требования, чем аналогичные параметрические критерии, что в некоторых случаях делает их применение предпочтительным. В данной главе мы сосредоточимся на оценках значимости результатов. Доверительные интервалы для непараметрических методов имеют тот же смысл, что и в рассмотренных выше параметрических методах, но довольно сложные алгоритмы их построения мы здесь приводить не будем.

11.1.1. Критерий знаковых рангов Вилкоксона

Критерий знаковых рангов Вилкоксона используется в тех же ситуациях повторных измерений, что и соответствующий Т-критерий. В отличие от последнего, критерий Вилкоксона применим и в ситуациях, когда распределения не являются нормальными, однако данные должны быть измерены в интервальной шкале, поскольку сравниваются разности значений, принадлежащих разным участкам шкал.

Пример 11.1.1(1). Пусть наша выборка состоит из n чисел (X_1, X_2, \dots, X_n) , которые являются разностями между двумя тестовыми результатами каждого испытуемого по тесту оптимизма до и после тренинга. Для критерия знаковых рангов формулировка гипотезы H_0 такова: эти числа являются результатом случайных колебаний показателя оптимизма симметрично в обе стороны вокруг нулевого значения⁹¹ (тем самым математическое ожидание случайной величины, порождающей выборку, равно нулю, или — тренинг не оказывает систематического воздействия на уровень оптимизма участников).

Гипотеза H_1 утверждает, что есть тенденция к изменению уровня оптимизма и распределение показателя после тренинга отличается

⁹¹ Для симметричности распределения разностей достаточно того, что выборка состоит из разностей испытаний одинаково распределенных случайных величин. Условие не выполняется, например, если исходное распределение не симметрично, а распределение оптимизма после тренинга увеличивает дисперсию, т.е. «растягивает» распределение, сохраняя его общую картину.

сы от распределения стартового замера некоторым положительным сдвигом.

Предположение о нормальности распределения *не делается*.

Пусть, например, разности между результатами второго и первого тестированием таковы: $\{-3, -1, 2, 4, 5, 7, 8, 10, 13, 17\}$. Мы видим, что хотя уровень оптимизма увеличился не у всех участников тренинга, но, во-первых, большинство участников стали более оптимистичными, а во-вторых, у «неудачников» тренинга уровень оптимизма изменился слабее, чем у выигравшего от тренинга большинства. Критерий Вилкоксона чувствителен именно к такой ситуации.

Для того чтобы применить критерий Вилкоксона к данной выборке, надо:

1) расположить все члены выборки в порядке возрастания *абсолютной величины*, подписать под ними их ранги; в нашем примере это будет следующий порядок и ранги:

-1	2	-3	4	5	7	8	10	13	17
1	2	3	4	5	6	7	8	9	10

2) вычислить статистику Вилкоксона, для чего подсчитать сумму рангов, приписанных отрицательным членам выборки (в нашем случае эта сумма $1+3=4$);

3) оценить «хвостовую» вероятность, соответствующую полученному результату расчета статистики. Последнюю операцию мы проделаем с помощью приведенной ниже таблицы распределения Вилкоксона (таблица 11.1.1(2)). Мы найдем соответствующие нашему объему выборки критериальные значения, и сравним с полученным значением статистики Вилкоксона, равным 4.

Таблица 11.1.1(2). Распределение Вилкоксона. Нижние граничные значения.

n	одностороннее 1% двухстороннее 2%	одностороннее 2.5% двухстороннее 5%	одностороннее 5% двухстороннее 10%
5			1
6		1	2
7	0	2	4
8	2	4	6
9	3	6	8
10	5	8	11

Полученное по нашей выборке значение статистики Вилкоксона оказывается меньше граничного значения для выборки объема 10 и уровня значимости 0.02, которое равно 5 (для двухсторонних границ; о возможности применения одностороннего критерия см. 7.1.5). Значимость результата (относительно гипотезы H_0 о равенстве нулю среднего выборочного) меньше 0.02.

Мы использовали нижние граничные значения, поэтому рассчитывали статистику, суммируя ранги отрицательных выборочных значений. Полностью эквивалентный результат мы получили бы, суммируя ранги положительных выборочных значений и сопоставляя результат с верхними граничными значениями. Обычно в таблицах для симметричных распределений (а распределение Вилкоксона именно таково) приводятся только нижние или только верхние граничные значения. С нижними границами сопоставляется меньшая сумма рангов, с верхними — бóльшая, которая должна при этом превосходить граничные значения.

В пункте 11.3.1 проведен расчет примера с помощью пакета SPSS. Мы рекомендуем, если это возможно, проделать упражнение 11.3.1(1) сразу после чтения данного раздела.

Замечание 11.1.1(3). Парные значения в непараметрических критериях. Если используемая в исследовании шкала дискретна, тем более если она имеет лишь небольшое количество значений, то при использовании критерия Вилкоксона и других непараметрических критериев может возникнуть проблема: если два или больше чисел в выборке равны между собой, то как присваиваются ранги в этом случае?

Предположим, мы имеем ряд $\{1, 2, 2, 2, 2, 4, 5\}$. Правило присваивания рангов вполне естественное: группа двоек занимает места со второго по пятое, следующая четверка занимает шестое место. Всем двойкам присваивается ранг $(2+3+4+5)/4 = 3.5$, а четверке присваивается ранг, соответствующий ее месту в ряду, а именно 6.

При таком алгоритме присвоения рангов их сумма по всем членам ряда остается неизменной, зависящей только от того, сколько всего в ряду чисел, независимо от количества равных значений.

11.1.2. Критерий Манна-Уитни для независимых выборок

В параграфе 6.1.2 мы решали задачу сравнения средних значений двух выборок, используя Т-критерий, который требует, чтобы выборки были реализациями нормально распределенных случайных величин.

Если условие нормальности не гарантируется, тем более если измерения проведены не в интервальной, а в порядковой шкале, к тем же выборкам для решения той же задачи можно применить *критерий Манна-Уитни*. Требования к распределениям для этого кри-

терия слабее (подробнее в параграфе 11.2). Для абсолютно корректного применения требуется, чтобы выборочные распределения были подобны (отличались только сдвигом), но эти распределения не обязаны быть нормальными.

Пример 11.1.2(1). Проверяется гипотеза о том, что у родственников пациентов с шизофренией чаще, чем в норме, наблюдаются искажения мышления, характерные для шизофрении (хотя и гораздо реже, чем у больных). У молодых родственников пациентов с шизофренией (первая группа) и у испытуемых, у которых таких родственников не было (вторая группа), оценивалось количество ошибок по типу искажения мышления при выполнении набора заданий патопсихологического обследования. Предполагалось, что у родственников больных выраженность искажения мышления выше, чем в контрольной группе. Результаты таковы:

{10, 12, 14, 17, 18, 19, 20, 22, 25, 27} (первая группа).

{7, 8, 9, 11, 13, 15, 16, 21} (вторая группа).

Статистика Манна-Уитни — это не формула, а алгоритм вычисления. Чтобы рассчитать эту статистику для наших выборок, надо сначала записать члены обеих выборок в порядке возрастания, помечая значения признаком группы (мы выделяем курсивом первую и жирным шрифтом вторую группу):

7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 25, 27.

Если бы случайные величины, испытаниями которых получены выборки, имели совершенно одинаковое распределение (гипотеза H_0), то «жирные» и «курсивные» числа перемешивались бы более или менее равномерно. Если же первая группа имеет систематически больший балл, то «курсивные» числа должны оказаться в основном правее «жирных» — как оно и оказывается в нашем случае. Теперь мы пересчитаем это преимущество в значимость наших данных относительно гипотезы H_0 . Для этого мы должны рассчитать статистику.

Алгоритм расчета таков:

1) каждому числу из последовательности присваиваем ранг, который подписываем под соответствующим числом:

7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	25	27
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18

2) Теперь суммируем ранги отдельно для «курсивных» (S_1) и «жирных» (S_2) чисел.

Получаем $S_1 = 4 + 6 + 8 + 11 + 12 + 13 + 14 + 16 + 17 + 18 = 119$,

$S_2 = 1 + 2 + 3 + 5 + 7 + 9 + 10 + 15 = 52$.

Мы видим, что первая группа имеет сумму большую, чем вторая. Но поскольку выборки могут иметь сильно различающиеся объемы, то судить о том, какая группа преобладает, надо по средним рангам. В данном случае они равны соответственно 11.9 и 6.5. Первая выборка располагается в ряду правее второй.

Значимости можно определять по таблицам распределения. Для каждой пары объемов выборок распределение будет свое, поэтому таблицы достаточно громоздки и мы не будем их здесь приводить. SPSS укажет значимости в таблицах вывода. Как их понимать, мы разберем в подпараграфе 11.3.2.

Замечание 11.1.2(2). О расчете статистики. Имеется альтернативный способ расчета статистики Манна-Уитни через попарное сравнение членов выборок (каждого с каждым). Результаты отличаются зависящей от объемов выборок легко рассчитываемой константой⁹². Далее мы рассмотрим два обобщения критерия Манна-Уитни, первый из которых удобнее считать через ранги, а второй — через сравнения. В предыдущей таблице заменим нижнюю строку. Вместо рангов будем писать только для «жирных» чисел, сколько «курсивных» находится в ряду левее каждого из них.

7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	25	27
0	0	0		1		2		3	3					7			

Сумма чисел нижней строки и есть альтернативный результат. В нашем случае она равна 16. Для выборок, обозначенных $\{x\}$ и $\{y\}$, этот результат обозначается U_{xy} . Если к U_{xy} прибавить (8 — объем второй, «жирной» выборки), то полученные 52 и дают сумму рангов «жирной» выборки, посчитанную первым способом. Почему сумма рангов отличается от U_{xy} на $n(n+1)/2$, где n — объем выборки $\{y\}$, — это не слишком сложная комбинаторная задача, которую мы оставляем читателю.

11.1.3. Критерий Краскелла-Уоллиса для n независимых выборок

Критерий Краскелла-Уоллиса обобщает критерий Манна-Уитни точно так же, как однофакторный дисперсионный анализ обобщает Т-критерий для двух выборок. Рассмотрим пример, аналогичный использованному при сравнении регрессионного и дисперсионного анализа, но с другими числами, более удобными для непараметрических критериев.

Пример 11.1.3(1). Пусть уровень статистических знаний оценивался по стобалльной системе у студентов первого, второго и третьего

⁹² Этот способ расчета подробно описан в учебнике (Кричевец и др, 2003).

курса. Результаты первокурсников $x: \{5, 9, 15, 18, 19\}$, второкурсников $y: \{11, 12, 17, 22, 23\}$, третьекурсников $z: \{10, 16, 20, 24, 25\}$. Нас интересует, меняется ли уровень знаний в процессе обучения статистике.

Для читателей, уловивших основные идеи построения статистик, статистика Краскелла-Уоллиса будет выглядеть вполне естественным обобщением статистики Манна-Уитни. Все три выборки выстраиваются в один ряд:

5, 9, 10, 11, 12, 15, 16, 17, 18, 19, 20, 22, 23, 24, 25

(если групп больше, то и сортов написания цифр должно быть больше) и числам присваиваются ранги. Затем ранги суммируются отдельно для каждой группы и вычисляется средний для группы ранг \bar{R}_i . Если между группами нет систематических различий, то средние ранги будут колебаться вокруг общего среднего для рангов, равного половине от общего объема выборок, в нашем случае $15/2 = 7.5$. Рассматривается сумма квадратов отклонений: $H = (\bar{R}_1 - 7.5)^2 + (\bar{R}_2 - 7.5)^2 + (\bar{R}_3 - 7.5)^2$. Если она велика, это говорит в пользу наличия систематических (не случайных) различий между группами. Для небольших объемов групп распределение статистики рассчитывается для каждого случая отдельно, при увеличении объема групп распределение статистики можно аппроксимировать распределениями Фишера или Хи-квадрат.

11.1.4. Критерий Джонкхиера

Критерий Джонкхиера нацелен на более узкую задачу, чем критерий Краскелла-Уоллиса. Их отношения похожи на отношения регрессионного анализа к дисперсионному, о котором мы говорили в главе 9. В двух разных случаях: (1) если уровень знаний студентов устойчиво падает или (2) максимум знаний приходится на второй курс, а затем падает ниже уровня первого курса, критерий Краскелла-Уоллиса будет показывать результат в пользу отвержения гипотезы о равенстве средних. Любые перестановки групп не меняют значение статистики и значимость результата

Мы же имеем в нашей задаче вполне определенный порядок групп и вполне определенные ожидания, что интересующий нас показатель уровня знаний должен демонстрировать рост по мере увеличения стажа учебы в университете. Нечувствительность критерия к перестановкам групп выглядит в этом случае странно: мы хотим доказать, что уровень подготовки студентов растет, но критерий Краскелла-Уоллиса даст тот же самый ответ, если уровень, например, с той же скоростью падает.

Поскольку близкая к нулю значимость приписывается критерием Краскелла-Уоллиса результатам, нас не интересующим, значимость

(как вероятность ошибки первого рода) будет включать вероятности этих ненужных событий, т.е. будет переоцениваться не в нашу пользу.

Ситуацию можно улучшить, если применить другой критерий, более точно выделяющий нужную нам альтернативу — гипотезу о росте знания в процессе обучения. Этим свойством обладает критерий Джонкхиера.

Расчет статистики Джонкхиера проведем на том же примере. Напомним, результаты студентов таковы: у первокурсников $x: \{5, 9, 15, 18, 19\}$, у второкурсников $y: \{11, 12, 17, 22, 23\}$, у третьекурсников $z: \{10, 16, 20, 24, 25\}$. Вычисляются статистики Манна-Уитни для всех возможных пар (ставя справа данные курса с большим номером): U_{xy}, U_{yz}, U_{xz} (см. Замечание 10.1.2). Затем они складываются и рассчитывается общая статистика превосходства правых групп (с большим номером) над левыми (с меньшим номером). В нашем случае это $17+15+20 = 52$. Результат сравнивается со средним значением для ситуации с совпадающими групповыми средними. При равенстве групповых средних в сравнении двух групп по пять испытуемых примерно в половине из 25 пар левое число будет больше правого, а в половине — наоборот, т.е. средний балл на одно сравнение групп 12.5. В нашем случае три сравнения дадут $12.5 \cdot 3 = 37.5$. Наше значение статистики 52 больше среднего, следовательно, имеется тенденция к монотонному росту.

Значимость результата определяется по соответствующему распределению. Мы рекомендуем сейчас, если возможно, выполнить упражнение 11.3.3. Отметим, что значимость нашего результата при использовании критерия Краскелла-Уоллиса равна 0.289, а критерий Джонкхиера дает 0.127. Преимущество последнего, как мы говорили, вытекает из его более узкой направленности на монотонное изменение показателя.

11.1.5. Коэффициент корреляции Спирмена

Коэффициент корреляции Пирсона предназначен для расчетов меры связи нормально распределенных случайных величин. Если наши данные имеют иное происхождение (в частности, если они принадлежат шкале порядка), то корректно использовать непараметрическую меру связи — *коэффициент корреляции Спирмена*.

Рассмотрим пример. Группа из четырех испытуемых провела эксперимент на время реакции (ВР) на слуховые и зрительные стимулы. Показатели представлены в таблице 11.1.5(1) (время измерено в миллисекундах).

Таблица 11.1.5(1). Время реакции на слуховые и зрительные стимулы у четверых испытуемых.

Номер испытуемого	1	2	3	4
Среднее ВР на слуховые стимулы	497	576	658	709
Среднее ВР на зрительные стимулы	547	454	745	516

Для того чтобы посчитать коэффициент корреляции Спирмена для этих данных, сначала надо заменить исходные выборочные данные на их ранги⁹³ в соответствующей выборке (в новой таблице эти ранги представлены под соответствующими строками), затем для каждого испытуемого вычислить разность d_i рангов, затем посчитать квадраты разностей d_i^2 и, наконец сумму этих квадратов по всем испытуемым (таблица 11.1.5(2)).

Таблица 11.1.5(2). Ранги времени реакции

Среднее ВР на слуховые стимулы	497	576	658	709
Ранги ВР на слуховые стимулы	1	2	3	4
Среднее ВР на зрительные стимулы	447	554	745	526
Ранги ВР на зрительные стимулы	1	3	4	2
Разности рангов (d_i)	0	-1	-1	2
Квадраты разностей рангов d_i^2	0	1	1	4

Сумма квадратов $S = d_1^2 + d_2^2 + d_3^2 + d_4^2 = 0 + 1 + 1 + 4 = 6$.

Теперь остается только вычислить окончательный результат по формуле

$$\hat{r} = 1 - 6 \cdot \frac{S}{n(n^3 - 1)}.$$

Подставляя в формулу $S=6$ и размер выборки $n=4$, получаем

$$\hat{r} = 1 - \frac{36}{60} = 0.4.$$

Если в выборке имеются равные значения, то им приписываются равные средние ранги, как мы это делали в случаях статистик Вил-

⁹³ Ранги членов выборки — это их порядковые номера в ряду, где эти числа расставлены в порядке возрастания.

коксона и Манна-Уитни. Однако в этом случае в формулу подсчета надо вводить поправки, о которых мы здесь говорить не будем⁹⁴.

Можно показать, что если в выборках нет совпадающих значений, то число \hat{r} лежит на отрезке $[-1, +1]$, т.е. $-1 \leq \hat{r} \leq 1$.

Упражнение 11.1.5(3). Проверить, что для выборок $x: \{1, 2, 3\}$, $y: \{1, 2, 3\}$ выборочный коэффициент корреляции Спирмена равен 1, а для выборок $x: \{1, 2, 3\}$, $y: \{3, 2, 1\}$ и $x: \{1, 2, 3, 4\}$, $y: \{4, 3, 2, 1\}$ коэффициент равен -1 .

Коэффициент корреляции Спирмена имеет в общем тот же смысл, что и коэффициент Пирсона, но он «выдерживает» допустимые преобразования порядковых шкал. Вследствие этого одинаковые значения будут иметь посчитанные по Спирмену корреляции пар выборок, изображенных на рис. 11.1.5(4) (коэффициенты корреляции Пирсона будут для них различны), где слева (а) изображена линейная зависимость, а справа (б) квадратичная. Равенство объясняется тем, что возведенные в квадрат значения (если они положительны) располагаются в том же порядке друг относительно друга, что и исходные значения. Другой пример мы разберем в подпараграфе 11.3.4.

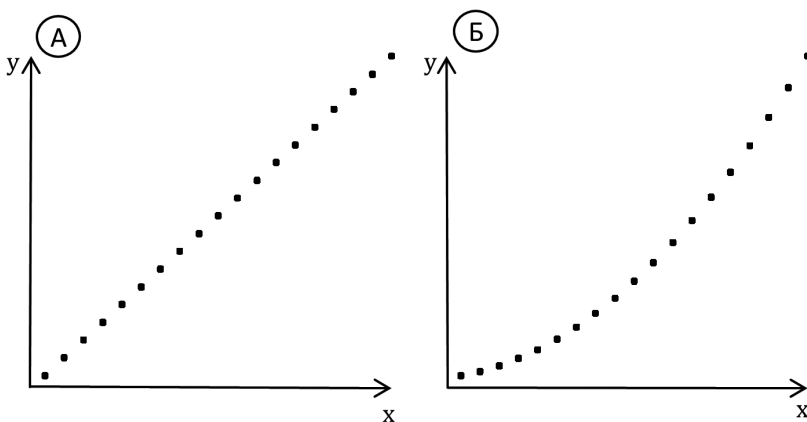


Рис. 11.1.5(4). Выборки, связанные линейной и квадратичной зависимостью, которым соответствуют равные единице коэффициенты корреляции Спирмена

⁹⁴ Предлагаемые в отечественной литературе (например, Сидоренко, 2000; Рубцова, Ленков, 2005) формулы для поправок не совпадают с используемыми в пакете SPSS.

Таблицы квантилей распределения коэффициента корреляции Спирмена мы не будем здесь приводить. Статистические пакеты дают точную оценку значимости результата относительно гипотезы о равенстве нулю теоретического коэффициента корреляции.

11.1.6. Таблицы сопряженности

Таблицы сопряженности 2x2 дают самый грубый метод статистического оценивания связи переменных. Его преимущество — независимость от распределения переменных и хорошая наглядность. Кроме того, разобравшись с таблицами 2x2, мы получим основу для работы с таблицами более высоких порядков, которые используются для выявления нелинейных зависимостей.

Если сопоставляемые переменные *дихотомические*, т.е. имеют всего два значения (например, 1 — «мужчина», 2 — «женщина» или 0 — «не согласен», 1 — «согласен»), то таблицы сопряженности 2x2 являются предпочтительным методом статистической обработки.

Пример 11.1.6(1). Эксперимент⁹⁵ был посвящен особенностям нарушения мышления у больных шизофренией. Одна из ключевых особенностей мышления при этом заболевании — нарушение избирательности внимания: больные в рассуждениях склонны опираться не на ведущие, основные признаки объектов, а на латентные, скрытые признаки предметов и явлений. Однако эти особенности могут помогать больным в тех случаях, когда нужно заметить и использовать не явное, а скрытое свойство предмета, — например, решить творческую задачу, требующую нестандартного подхода, им может быть легче, чем здоровым испытуемым.

Для проверки этой гипотезы и было проведено следующее исследование. Больным шизофренией и испытуемым контрольной группы предлагалось решить творческую задачу Л. Секея в варианте Л.И. Анцыферовой. Задача состояла в следующем. Испытуемым давали весы и несколько предметов (разновески, фонарик, баночка с солью, баночка с сухой ватой, карандаши, счетные палочки, ластик, свеча). Нужно было уравновесить весы таким образом, чтобы через некоторое время они сами вышли из равновесия. Единственное решение задачи состояло в том, чтобы сжечь свечу — по мере горения она теряет вес. Из 55 здоровых испытуемых самостоятельно решили задачу 25 человек, тогда как из 45 больных — 35. Данные запишем в таблицу:

⁹⁵ Критская В.П., Мелешко Т.К. и Поляков Ю.Ф. Патология психической деятельности при шизофрении: мотивация, общение, познание. — М.: Изд-во МГУ, 1991.

Наблюдаемые частоты

	Не решил	Решил	Сумма
Норма	30	25	55
Шизофрения	10	35	45
Сумма	40	60	100

Отражённые в таблице результаты говорят в пользу того, что больные имеют преимущество при решении задачи. Наша задача — дать количественные меры этого преимущества.

Рассмотрим таблицу, в которой отражены только суммы: в эксперименте участвовали 55 здоровых и 45 больных, задачу не решили 40 и решили 60 человек:

Число испытуемых

	Не решил	Решил	Сумма
Норма			55
Шизофрения			45
Сумма	40	60	100

Что должно было бы оказаться в пустых клетках, если бы процент решивших и не решивших задачу был бы одинаков для больных и здоровых? Таблица, которая называется «таблица ожидаемых частот», выглядела бы так:

Ожидаемые частоты

	Не решил	Решил	Сумма
Норма	22	33	55
Шизофрения	18	27	45
Сумма	40	60	100

В гипотетической ситуации равных возможностей решения (гипотеза H_0) задачу решили 60% как здоровых, так и больных. Автоматически пропорциональны доли здоровых (и больных) среди решивших и не решивших — по 55% и 45%.

Ожидаемые частоты получаются так: в верхней строке должно быть 55/100 участников, из них в левой верхней клетке 40/100. Тогда из 100 в левой верхней клетке должно оказаться $100 \cdot (55/100) \cdot (40/100) = 22$ участника эксперимента. Аналогично для других клеток.

Следующая таблица отображает разности между наблюдаемыми и ожидаемыми частотами:

Разности наблюдаемых и ожидаемых частот

	Не решил	Решил	Сумма
Норма	8	-8	0
Шизофрения	-8	8	0
Сумма	0	0	0

Чтобы оценить значимость, делим квадраты значений в клетках таблицы разностей на значения в соответствующих клетках таблицы ожидаемых частот и суммируем результаты: $(-8)^2/22 + 8^2/33 + 8^2/18 + (-8)^2/27 = 2.91 + 1.94 + 3.56 + 2.37 = 10.78$.

Если бы частоты соответствовали таблице ожидаемых частот, то результат проведенного расчета над случайной реализацией такой случайной величины имел бы распределение Хи-квадрат с одной степенью свободы. Полученные 10.78 — очень большое для такого распределения значение. Ему соответствует значимость 0.0011.

Замечание. В таблицах 2×2 на третьем этапе мы всегда получаем таблицу с равными по модулю значениями и одинаковыми знаками по диагонали (одна степень свободы). Это можно использовать для проверки.

Рекомендуем теперь, если это возможно, проделать упражнение 11.3.5(1).

11.1.7. Таблицы сопряженности более высоких размерностей

Таблицы сопряженности можно использовать для выявления более сложных отношений переменных, чем линейные зависимости.

Пример 11.1.7(1). В некотором банке проводится исследование уровня стресса у сотрудников в начале, середине и конце недели. Каждый раз обследуются 60 человек, выбираемых случайно из числа сотрудников банка:

	Начало	Середина	Конец	Сумма
Низкий стресс	15	0	15	30
Средний стресс	25	10	25	60
Высокий стресс	20	50	20	90
Сумма	60	60	60	180

Мы видим, что в середине недели доля сотрудников с высоким стрессом гораздо выше, чем в начале и конце. Для оценки значимости резуль-

тата составим таблицу произведений долей. Для всех столбцов мы получаем равные доли: по $1/3$ от общего числа испытуемых. По строкам: в первой строке $1/6$ от общего количества испытуемых, во второй $1/3$ и в третьей $1/2$ от общего количества испытуемых.

Для H_0 о равенстве частот наблюдений уровня стресса по периодам недели таблица ожидаемого распределения испытуемых по клеткам такова⁹⁶:

Таблица ожидаемых частот

	Начало	Середина	Конец	Сумма
Низкий стресс	10	10	10	30
Средний стресс	20	20	20	60
Высокий стресс	30	30	30	90
Сумма	60	60	60	180

Таблица разностей

	Начало	Середина	Конец	Сумма
Низкий стресс	5	-10	5	0
Средний стресс	5	-10	5	0
Высокий стресс	-10	20	-10	0
Сумма	0	0	0	0

Заметим, что таблица разностей для размеров больших, чем 2×2 не имеет такой простой структуры, как в случае 2×2 . Для контроля можно использовать равенство нулю сумм по строкам и столбцам.

Далее проделаем знакомую по предыдущему примеру операцию — сложим квадраты разностей, деленные на содержимое клеток таблицы ожидаемых частот. Сумма такова:

$$\begin{aligned} \frac{5^2}{10} + \frac{(-10)^2}{10} + \frac{5^2}{10} + \frac{5^2}{20} + \frac{(10)^2}{20} + \frac{5^2}{20} + \frac{(-10)^2}{30} + \frac{20^2}{30} + \frac{(-10)^2}{30} = \\ = 150/10 + 150/20 + 600/30 = 37.5. \end{aligned}$$

Число степеней свободы распределения χ^2 для таблицы размера $m \times n$ вычисляется по формуле $(m-1)(n-1)$. В нашем случае оно равно 4.

⁹⁶ Напомним: в верхней левой клетке стоит произведение доли испытуемых в левом столбце, доли испытуемых в первой строке и общего количества испытуемых; аналогично в других клетках.

37.5 — очень большое значение для χ^2 . Мы можем быть уверены, что уровень стресса сотрудников данного банка меняется в течение недели. Однако о том, как он меняется, таблицы сопряженности говорят так же недостаточно внятно, как и дисперсионный анализ. В последнем рассмотрены апостериорные сравнения для прояснения подробностей динамики средних. В таблицах сопряженности таких уточняющих опций нет, к сожалению. В нашем примере можно уверенно сказать, что в середине недели стресс максимален, а в начале и конце примерно одинаков (это последнее утверждение о сходстве вообще невозможно подтвердить статистически). При других конфигурациях может быть достаточно трудно сформулировать гипотезу, которую таблица сопряженности подтверждает.

Например, такой вариант таблицы:

	Начало	Середина	Конец	Сумма
Низкий стресс	18	0	12	30
Средний стресс	25	10	25	60
Высокий стресс	17	50	23	90
Сумма	60	60	60	180

в целом также не согласуется с гипотезой о равенстве частот, однако это не позволяет сделать вывод о том, что имеется различие по уровню стресса в начале и конце недели. Если такой вывод нам интересен, придется проводить специально нацеленный на это эксперимент.

Упражнение 11.3.5(2), где разобран данный пример, можно выполнить сразу после чтения приведенного выше текста.

11.2. Условия применимости параметрических и непараметрических критериев и последствия их нарушения

Непараметрическими называются статистики, которые работают на шкалах более слабых, чем интервальные. Первый из рассмотренных нами в начале данной главы критерий, вообще говоря, относится к непараметрическим не по праву. В действительности критерий знаковых рангов требует выборки, измеренных в интервальной шкале, поскольку использует разности показателей (для порядковой шкалы разность величин не имеет смысла). Его с большим основанием можно отнести к «свободным от распределения» критериям, которые работают с любыми распределениями исходных данных. Однако критерий Вилкоксона все же требует, чтобы распределение разности стартового

и финишного показателей было симметричным⁹⁷. Такому требованию удовлетворяет, например, случай, когда стартовые и финишные показатели имеют какое угодно, но одинаковое распределение. Доказательство достаточно просто, но мы не будем его здесь приводить.

Обсуждение непараметрических аналогов дисперсионного анализа начнем с замечания. Все варианты дисперсионного анализа (включая Т-критерий) рассчитаны исходя из предположения о нормальном распределении переменных, подвергающихся обработке. Стандартные варианты этих методов требуют также равенства внутрigrупповых дисперсий. Часто в руководствах по статистическим методам предлагается прежде, чем употребить один из перечисленных методов, проверить эмпирическое распределение — удовлетворяет ли оно перечисленным требованиям. Мы теперь рассмотрим влияние нарушения различных условий применимости на результаты критериев.

Прежде всего мы опишем один из распространенных методов проверки (имеются и другие методы проверки, но мы выбрали критерий χ^2 , поскольку он уже знаком читателю и узнать о нем больше будет полезно).

11.2.1. Критерий согласия χ^2

Рассмотрим сначала пример. Пусть мы имеем кривую монетку, выпадающую гербом с вероятностью p , цифрой — q и на ребро — r .

Если мы задумали бросить монету 100 раз, то случайная величина

$$\frac{(N_p - 100p)^2}{100p} + \frac{(N_q - 100q)^2}{100q} + \frac{(N_r - 100r)^2}{100r}$$

имеет приближенно распределение χ^2 с двумя степенями свободы.

В общем случае пусть в некотором испытании событие S_i происходит с вероятностью p_i и

$$\sum_{i=1}^m p_i = 1,$$

т.е. S_i представляют полную систему из m событий. Пусть мы предполагаем совершить N испытаний. Обозначим N_i количество появлений события S_i в данной серии. Тогда случайная величина

$$\frac{(N_1 - N_{p1})^2}{N_{p1}} + \dots + \frac{(N_m - N_{pm})^2}{N_{pm}}$$

⁹⁷ «Требуется» означает здесь, как и в других случаях, что квантили считаются именно для такого распределения. Насколько будут отличаться результаты, если требование так или иначе не удовлетворено — всегда предмет отдельного исследования. Свойство давать достаточно устойчивые результаты, если требование не удовлетворено, называется робастностью критерия.

имеет распределение χ^2 с $m - 1$ степенью свободы.

Если мы хотим проверить, возможно ли, чтобы полученное нами выборочное распределение было результатом испытаний некоторой случайной величины с известным нам распределением, то мы разбиваем весь диапазон возможных значений переменной на несколько ячеек (на сколько ячеек, мы либо решаем сами, либо перепоручаем это решение установкам по умолчанию в статистических пакетах). Далее, по тестируемому распределению рассчитываются теоретические вероятности попадания в i -ю ячейку и по приведенной формуле сравниваются с частотами попадания в эту ячейку. Чем больше результат, т.е. чем больше отклоняются частоты от предсказанных теоретически, тем больше у нас оснований сомневаться в том, что эмпирическое распределение получено испытанием данного теоретического. Если расчет по формуле статистики χ^2 дает экстремально большое значение, то мы отвергаем нулевую гипотезу о том, что выборка получена испытанием тестируемой случайной величины.

11.2.2. Проверка нормальности распределения перед применением дисперсионного анализа

Если мы применяем критерий согласия, чтобы проверить, применим ли дисперсионный анализ, надо иметь в виду следующее.

Мы уже обсуждали последствия бездумного употребления значимости 0.05 в качестве границы принятия решения. В данном случае ошибочность такой процедуры особенно очевидна. Читатель может убедиться в этом сам, проделав следующий опыт: надо взять уже знакомое биномиальное распределение с числом испытаний 5 и набирать с помощью генератора случайных чисел выборки все большего объема, имеющие такое распределение. Для небольших выборок гипотеза нормальности выдержит проверку, но при объеме около 50 гипотеза нормальности будет отвергнута согласно процедуре «отвергнуть, если значимость меньше 0.05» в подавляющем большинстве случаев. В то же время применимость дисперсионного анализа к большим биномиально распределенным выборкам не вызывает сомнений. Таким образом, при увеличении выборки пригодность дисперсионного анализа возрастает, но шансы распределения выдержать проверку, напротив, убывают. Это верно, по-видимому, для симметричных распределений самого разного вида. Мы адресуем читателя к нашей статье (Корнеев, Кричевец, 2011), краткое изложение которой доступно на сайте факультета психологии МГУ вместе с файлами для нашего практикума по SPSS.

Причина указанного дефекта метода состоит в следующем: если распределение случайной величины, испытаниями которой получают наши выборки, отличается от нормального самым незначительным об-

разом, то при любом фиксированном разбиении множества возможных значений эмпирические частоты попадания в ячейки разбиения будут все более достоверно отклоняться от ожидаемых по расчетам нормального распределения частот по мере роста объема выборки, т.е. критерий χ -квадрат (как и любой другой критерий согласия) рано или поздно уловит это различие на уровне значимости 0.05. Скажем в дополнение, что применимость критериев согласия также не безусловна. Важнейшим из этих условий является достаточно большой объем выборки: для χ -квадрат, например, рекомендуют выборку объема более 50. Но именно на таких объемах отвергается нормальность весьма похожих на нормальное биномиальных распределений.

Отметим, что условия применимости критерия Манна-Уитни не так просты, как часто представляется⁹⁸. Статистика Манна-Уитни применима к порядковым шкалам, но требует, чтобы распределения в сравниваемых группах были одинаковыми (с точностью до сдвига). В упомянутой статье разобраны последствия нарушения этого требования. Некоторые выглядят абсурдно: приведены три выборки, превосходство которых друг над другом по статистике Манна-Уитни циклично, т.е. нарушена транзитивность превосходства.

Отметим еще один аспект статистики Манна-Уитни и ее обобщений. Поскольку математическое ожидание и выборочное среднее не корректны по отношению к данным в порядковых шкалах, то для последних уместной экспликацией интуитивного представления о среднем выборочном значении служит медиана⁹⁹. Говорят, что медиана — мера центральной тенденции для порядковых шкал. Однако статистика Манна-Уитни не соотносит медианы, если распределения в группах не удовлетворяют строгому условию сдвига (там же). При выполнении же этого условия наличие сдвига говорит само за себя и без применения медианы.

Окончательно наша рекомендация такова: во всех сколько-нибудь сомнительных случаях применить оба критерия из соответствующей пары (Т-критерий/критерий Манна-Уитни; дисперсионный анализ/критерий Краскелла-Уоллиса) и сравнить значимости. Если они сильно различаются, не пожалеть сил на выяснение причин различий. Часто эти причины оказываются для психолога даже более интересными, чем отличия средних.

⁹⁸ То же относится и к его обобщениям — критериям Краскелла-Уоллиса и Джонкхиера. Нам не известны работы, в которых применимость последних проверялась бы в различных условиях.

⁹⁹ Мы не давали определения медианы распределения, поскольку нигде не использовали это понятие, что не отменяет его важности, читатель без труда найдет его в любом руководстве по статистике.

В заключение о статистике Джонкхиера. Эта статистика ориентирована на монотонный рост или монотонное убывание показателя, соответствующие номерам групп. Однако определить точнее, что именно измеряет эта статистика, не удастся, поскольку не слишком большие отклонения от монотонности статистика не выделяет. Таким образом, мы имеем здесь специфическую «джонкхиеровскую» экспликацию представления о систематическом росте (убывании), которые могут не выражаться в строго монотонной последовательности.

11.2.3. Другие методы проверки применимости дисперсионного анализа

В описаниях процедуры употребления дисперсионного анализа (включая Т-критерий) встречается рекомендация ориентироваться на критерий Ливиня, проверяющий однородность дисперсий. Кроме тех же возражений, которые мы предъявляли в предыдущем подпараграфе к критерию χ^2 , касающихся логической структуры проверки, уместно будет напомнить, что неоднородность дисперсии в одних случаях приводит к переоценке результата, а в других — к *недооценке*. Ниже приведена таблица из классической книги (Шеффе, 1980), на которую обычно ссылаются при обсуждении вопроса (таблица 11.2.3(1)).

Таблица 11.2.3(1). Влияние неравенства дисперсий ошибок и неравенства объемов групп на истинную вероятность того, что 95% доверительный интервал для $\mu_1 - \mu_2$ не покрывает истинное значение при больших n — отношение дисперсий, R — отношение объемов групп (таблица из работы Г. Шеффе, в которой приведены реальные уровни значимости для дисперсионного анализа при неравных дисперсиях выборок).

R	0	0*	1/5	1/2	1	2	5	∞^*
1		0.050	0.050	0.050	0.050	0.050	0.750	0.050
2		0.17	0.12	0.080	0.050	0.029	0.014	0.006
5		0.38	0.22	0.12	0.050	0.014	0.002	$1 \cdot 10^5$
∞^*		1.00	0.38	0.17	0.050	0.006	$1 \cdot 10^5$	0

* Недостижимые предельные случаи показывают границы изменения вероятности.

Таблица содержит результаты проверки нарушения однородности дисперсий для Т-критерия¹⁰⁰. В правом нижнем углу таблицы приведены результаты проверки в случаях, когда объем выборки, в которой дисперсия существенно больше, превышает объем выборки с меньшей дисперсией. В этом случае вероятности ошибки первого рода оценены как существенно меньшие, чем 0.05. Это значит, что результаты оказываются даже более достоверными, чем говорит Т-критерий.

Причину легко увидеть, посмотрев на формулу t-статистики. В случае большего объема выборки с большей дисперсией совокупная дисперсия переоценивается, что приводит к недооценке результата и занижению достоверности. Обе тенденции компенсируются, если применять модифицированный Т-критерий, результаты которого помещаются в нижней строке таблицы вывода SPSS для Т-критерия. Эти результаты мы и рекомендуем использовать во всех случаях. Результат теста Ливиня можно при этом не учитывать.

В наших собственных исследованиях обнаружилось, что наиболее сложный случай представляет собой для Т-критерия разнонаправленная асимметрия выборочных распределений. Подробности можно получить в упомянутой статье, к которой мы и отсылаем читателя.

11.3. Расчет непараметрических критериев в SPSS

11.3.1. Критерий Вилкоксона — непараметрический аналог Т-критерия для парных выборок

Упражнение 11.3.1(1). Расчет критерия Вилкоксона в статистическом пакете SPSS.

В подпараграфе 11.1.1 мы рассматривали результаты гипотетического тренинга оптимизма. Проведем теперь анализ с помощью пакета SPSS. В файле *OptimismTraining.sav* содержатся результаты оценки оптимизма участников тренинга до (переменная *First*) и после (переменная *Second*) тренинга.

Нужная процедура в SPSS вызывается следующим образом: **Анализ — Непараметрические критерии — Устаревшие диалоговые окна — Для двух связанных выборок (Analyze — Nonparametric Tests — Legacy Dialogs — 2 Related Tests)**. По умолчанию предусмотрен расчет критерия Вилкоксона (галочка в соответствующем квадратике). Окно ввода параметров анализа в данном случае устроено практически аналогично описанному выше диалоговому окну задания параметров Т-критерия Стьюдента (см. параграф 5.3). Зададим пару сравниваемых переменных

¹⁰⁰ Для достоверности цитирования мы сохранили опечатку в столбце, помеченном пятеркой.

Second — *First* (для этого в первое поле пары надо ввести имя *First*, а во второе *Second* — в отличие от Т-критерия, где порядок должен быть обратным) и нажмем кнопку *OK*.

Результаты расчета критерия Вилкоксона представлены двумя таблицами: таблицей рангов разностей первой и второй выборки и собственно статистикой критерия.

Ранги

		N	Средний ранг	Сумма рангов
Second — First	Отрицательные ранги	2 ^a	2.00	4.00
	Положительные ранги	8 ^b	6.38	51.00
	Связи	0 ^c		
	Всего	10		

^a Second < First

^b Second > First

^c Second = First

Статистики критерия^a

	Second — First
Z	-2.395 ^b
Асимпт. знч. (двухсторонняя)	0.017

^a Критерий знаковых рангов Уилкоксона.

^b Используются отрицательные ранги.

Мы видим, что результаты совпадают с полученными ручным расчетом: в первой таблице приведены суммы рангов, а во второй двухсторонняя значимость. Если вы тестируете одностороннюю альтернативу, значимость надо поделить на два.

В статистическом пакете SPSS критерий Вилкоксона реализован только как парный, т.е. на «вход» необходимо подать две переменные (в нашем случае результаты двух замеров). Если у вас по каким-то причинам введены разности замеров, а исходные результаты недоступны, можно выйти из положения, введя в качестве пары переменную, тождественно равную нулю. Как легко получить такую переменную, оставляем в качестве упражнения читателю.

Условия применимости критерия Вилкоксона слабее условий применимости Т-критерия — требуется, чтобы распределение разности было симметричным (подробнее см. 11.2). Если распределение разности силь-

но отличается от нормального, но остается симметричным, большего доверия заслуживает значимость, определенная по Вилкоксоу. Чаще всего значимости, полученные двумя методами, различаются мало.

Рекомендация. *Если различие велико, то наша рекомендация — постараться найти причины большого различия, проанализировав распределения каждой переменной и их разностей. Это касается всех непараметрических методов и их параметрических аналогов.*

Нередко за определяющими различия значимостей свойствами выборочных распределений стоят более интересные обстоятельства, чем исходная гипотеза о средних (см. упражнение 11.3.9).

11.3.2. Критерий Манна-Уитни

Пример 11.3.2(1). В подпараграфе 11.1.2 мы разобрали пример исследования искажений мышления у родственников шизофреников. Данные хранятся в файле *Distort.sav*. Откройте его и произведите обработку методом Манна-Уитни. Первая переменная *group* содержит номер группы, вторая, *distort* — количество ошибок.

Для расчета в SPSS надо выбрать в главном меню **Анализ — Непараметрические критерии — Устаревшие диалоговые окна — Для двух независимых выборок (Analyze — Nonparametric Tests — Legacy Dialogs — 2 Independent Samples)**. Появившееся окно устроено таким же образом, как и диалоговое окно параметров проведения Т-критерия. В поле **Список проверяемых переменных (Test Variable List)** следует указать сравниваемую переменную (число ошибок в нашем примере), в поле **Группирующая переменная (Grouping Variable)** — переменную, обозначающую принадлежность испытуемого к группе (переменная *group* в нашем примере). Также необходимо задать номера сравниваемых групп (1 и 2).

После нажатия кнопки **ОК** в окне вывода появляются результаты расчет критерия Манна-Уитни. Первая таблица содержит информацию о рангах сравниваемых переменных в двух экспериментальных группах:

Ранги

	group	N	Средний ранг	Сумма рангов
distort	1.00	10	11.90	119.00
	2.00	8	6.50	52.00
	Всего	18		

Суммы рангов совпадают с рассчитанными вручную в подпараграфе 10.1.3. Как можно заметить, средний ранг в экспериментальной груп-

пе (1) больше, чем в контрольной (2)¹⁰¹. Насколько это различие значимо, можно заключить, рассмотрев следующую таблицу, содержащую результат расчета критерия Манна-Уитни.

Статистики критерия^a

	mistakes
Статистика U Манна-Уитни	16.000
Статистика W Уилкоксона	52.000
Z	-2.132
Асимпт. знч. (двухсторонняя)	0.033
Точная знч. [2*(1-сторонняя Знач.)]	0.034 ^b

^a Группирующая переменная: group

^b Не скорректировано на наличие связей.

В первой строке статистика, рассчитанная способом, описанным в Замечании 11.1.2(2). Во второй строке всегда ставится меньшая из сумм рангов предыдущей таблицы. О том, у кого большие показатели, надо судить по средним рангам из предыдущей таблицы. Значимость можно найти в последней строке таблицы. Для не слишком больших выборок приводится точная значимость, если выборки велики, в последней строке оказывается асимптотическая значимость, а точная не рассчитывается. Условия применимости T-критерия и критерия Манна-Уитни обсуждались в подпараграфе 11.2.1. Практические рекомендации даны в конце подпараграфа 11.3.1.

11.3.3. Критерии Краскелла-Уоллиса и Джонкхиера

Упражнение 11.3.3(1). В подпараграфе 11.1.4. мы разбирали пример, описывающий исследование изменения уровня статистических знаний у студентов университета. Данные содержатся в файле *StatKnowledge.sav*. Обе статистики будем рассчитывать одновременно. Для этого выберем последовательность пунктов меню *Анализ — Непараметрические критерии — Устаревшие диалоговые окна — Для K независимых выборок (Analyze — Nonparametric Tests — Legacy Dialogs — K Independent Samples)*.

В отличие от дисперсионного анализа для группирующей переменной (class) здесь требуется задать диапазон изменения (от 1 до 3 в нашем случае). По умолчанию в окне задания параметров помечен квадратик

¹⁰¹ Обратите внимание, что сумма рангов зависит от количества испытуемых в группах. Если в группах разное количество испытуемых, делать содержательные выводы о различии между группами по этому показателю нельзя и следует анализировать показатель «средний ранг».

***H* Крускала-Уоллеса (Kruskal-Wallis *H*).** Поставьте галочку также в пункте **Джонкхиера-Терпстры (Jonckheere-Terpstra)**. Это и есть наш критерий Джонкхиера.

Результаты расчета критерия Краскелла-Уоллеса размещаются в двух таблицах. В первой приведены суммы рангов по группам, во второй нормированная сумма квадратов отклонений от среднего ранга, имеющая распределение Хи-квадрат при условии отсутствия систематического сдвига групповых средних. Асимптотическая значимость результата приводится в последней строке второй таблицы:

Ранги

	class	N	Средний ранг
Knowledge	1.00	5	5.60
	2.00	5	8.40
	3.00	5	10.00
	Всего	15	

Статистики критерия^{a,b}

	Knowledge
Хи-квадрат	2.480
ст.св.	2
Асимпт. знч.	0.289

a. Критерий Краскела-Уоллеса

b. Группирующая переменная: class

Результаты расчета статистики Джонкхиера приводятся в единственной таблице:

Критерий Джонкхиера-Терпстра^a

	Knowledge
Число уровней в class	3
N	15
Наблюденная статистика Джонкхиера-Терпстра	52.000
Среднее статистики Джонкхиера-Терпстра	37.500
Стд. отклонение статистики Джонкхиера-Терпстра	9.465
Нормир. статистика Джонкхиера-Терпстра	1.532
Асимпт. знч. (двухсторонняя)	0.126

a. Группирующая переменная: class

В отличие от таблицы результатов Манна-Уитни, о направлении изменения показателя надо судить по строкам *Наблюдённая статистика* и *Среднее статистики*. Если первая больше второй, то групповые средние растут с возрастанием номера группы, если меньше, то, наоборот, убывают. Значимость оценивается асимптотически, поскольку распределение статистики похоже на нормальное с параметрами (мы приводим формулы для равных по объёму групп)

$$MJ = \frac{K(K-1)n^2}{4}, DJ = \frac{(Kn)(2Kn+3) - Kn^2(2n+3)}{72}$$

где n — объём групп, а K — их количество.

В нашем случае $MJ = 3 * 2 * 5 * 5 / 4 = 37.5$, а дисперсия равна $((15 * 15) * (33) - 3 * (5 * 5)(13)) / 72 = (7425 - 975) / 72 = 89.58$.

Стандартное отклонение получаем, извлекая из этого числа корень. Оно равно 9.465. Нормированная статистика получается делением разности наблюдаемого значения и среднего на стандартное отклонение, после чего по нормальному распределению вычисляется значимость.

Как мы и ожидали, на наших данных критерий Джонкхиера оказывается более мощным, чем критерий Краскелла-Уоллиса, для гипотезы о монотонном изменении знаний студентов.

11.3.4. Коэффициент корреляции Спирмена

Упражнение 11.3.4(1). Как мы разъясняли в подпараграфа 11.3.2, непараметрические критерии работают на порядковых шкалах, в отличие от параметрических, которые требуют интервальности шкал. С этим связано их преимущество в ситуациях, когда мы сталкиваемся с выбросами, т.е. значениями показателя, аномально отличающимися от среднего по генеральной совокупности значения. Мы разберем сейчас пример, когда выброс меняет результат исследования кардинально. Предположим, школьникам предложили выполнить сразу два задания ЕГЭ в ограниченное время. В файле *CorrelationPS.sav* даны результаты. Переменная *RL* представляет оценку по русскому языку, а *GG* — по географии. Один из учеников почти все время потратил на sms-переписку, поэтому показал плохой результат по обоим предметам, однако его истинное занятие не было замечено преподавателем.

Рассчитайте коэффициенты корреляции Пирсона и Спирмена и сравните их.

Напомним, что для этого надо выбрать пункты меню *Анализ* — *Корреляции* — *Парные (Analyze — Correlation — Bivariate)*. Чтобы рассчитать обе корреляции, отметьте галочкой квадраты, подписанные *Пирсон (Pearson)* и *Спирман (Spearman)*. В результате получаем корреляции разных знаков — положительную по Пирсону и отрицательную по Спирмену. В данном случае у нас нет оснований удалить выброс, однако его влияние корреляция по Спирмену минимизирует.

Упражнение 11.3.4(2). Постройте график рассеяния и скажите, в каких пределах можно изменять данные последнего испытуемого, чтобы корреляция по Спирмену не менялась. Сделайте какие-то из этих изменений и посмотрите, как ведет себя корреляция по Пирсону.

11.3.5. Таблицы сопряженности

Упражнение 11.3.5(1). Проведем расчеты подпараграфа 11.1.6 о решении творческих задач. Откройте файл *PolyakovExp.sav*. Переменная *Diagnosis* кодирует принадлежность группе больных или здоровых, переменная *TaskSolving* содержит информацию о решении задачи.

Для того чтобы построить таблицы сопряженности и рассчитать критерий Хи-квадрат Пирсона, выберем пункт меню *Анализ — Описательные статистики — Таблицы сопряженности (Analysis — Descriptive statistics — Crosstabs)*. В появившемся окне одну из переменных нужно указать в поле *Строки (Rows)*, другую — в поле *Столбцы (Columns)*. Для того чтобы ориентация таблицы совпадала с приведенной в предыдущем параграфе, перенесите переменную *Diagnosis* в окно *Строки*, вторую переменную — в *Столбцы*. Для того чтобы рассчитать ожидаемые частоты, нажмите на кнопку *Ячейки (Cells)* справа и поставьте галочку на *Ожидаемые частоты (Expected Frequencies)*. Чтобы рассчитать Хи-квадрат Пирсона нажмите на кнопку *Статистики (Statistics)* справа и поставьте галочку в клетке *Хи-квадрат (Chi-square)*. Обратите внимание, что, если этот пункт не выбран, в выводе не будет оцениваться значимость.

Первая таблица результатов — «Сводка обработки наблюдений» — содержит вспомогательную информацию о количестве испытуемых и пропущенных результатов. Вторая таблица приведена ниже:

Таблица сопряженности Контрольная группа / шизофрения
* Решение задачи

			Решение задачи		Итого
			Не решил	Решил	
Контрольная группа / шизофрения	Норма	Частота	30	25	55
		Ожидаемая частота	22.0	33.0	55,0
	Шизофрения	Частота	10	35	45
		Ожидаемая частота	18.0	27.0	45,0
Итого		Частота	40	60	100
Ожидаемая частота		40,0	60.0	100.0	

Она повторяет нашу таблицы наблюдаемых и ожидаемых частот подпараграфа 11.1.7. Наконец, третья таблица предоставляет нам оценки значимости результата.

Критерии хи-квадрат					
	Значение	ст.св.	Асимпт. значимость (2-стор.)	Точная значимость (2-стор.)	Точная значимость (1-стор.)
Хи-квадрат Пирсона	10.774 ^a	1	0.001		
Поправка на непрерывность ^b	9.470	1	0.002		
Отношение правдоподобия	11.138	1	0.001		
Точный критерий Фишера				0.001	0.001
Линейно-линейная связь	10.667	1	0.001		
Кол-во валидных наблюдений	100				

^a. В 0 (0,0%) ячейках ожидаемая частота меньше 5. Минимальная ожидаемая частота равна 18,00.

^b. Вычисляется только для таблицы 2x2.

Для небольших выборок рассчитывается точная значимость результата по соответствующему (так называемому гипергеометрическому) распределению. В данном случае это двухсторонняя значимость 0.001. Вверху слева значение статистики Хи-квадрат, отличающееся от полученного нами ошибкой округления. Однако значимость по критерию Хи-квадрат (если выборка велика и нет точной значимости) надо брать из второй строки, где вводится необходимая для таблиц только размера 2x2 поправка.

Значимость близка к нулю, гипотеза исследования довольно надежно подтверждается.

Упражнение 11.3.5(2). Проведем расчеты примера 11.1.7(1) о стрессе сотрудников банка. Данные записаны в файл *BankStress.sav*. Переменная *Period* кодирует начало, середину и конец недели, переменная *StressGroup* кодирует уровень стресса. Расставьте самостоятельно переменные в соответствующих полях в окне задания таблицы

сопряженности так, чтобы ориентация таблицы совпадала с таблицами из примера 11.1.7(1). Поставьте также необходимые знаки, чтобы получить ожидаемые значения и статистику.

Таблица вывода, касающаяся значимости, в данном случае проще, чем в предыдущем: нужная нам значимость проставлена в первой строке, альтернатив нет.

Упражнение 11.3.5(3). На самом деле данные о стрессе получены первоначально в виде показателя, изменяющегося от 1 до 60 (переменная *StressScore* в том же файле *BankStress.sav*). Переменная *StressGroup* представляет собой некоторое «огрубление» балла. Разберитесь, как именно перекодированы данные *StressScore* в *StressGroup*.

Технически такую перекодировку можно выполнить только почти вручную. Мы предложим вам сделать похожую, но требующую минимум усилий перекодировку. Выберите пункты меню **Преобразовать — Ранжировать (Transform — Rank cases)**. В полученном окне переведите в поле **Переменные (Variables)** имя *StressScore*, а затем нажмите кнопку **Типы рангов (Rank types)**. В полученном окне выберите **N-разбиение (Ntiles)** на 3 группы. Другие галочки можно убрать. В результате получается переменная *NStressS*, в которой кодируется, к какой из третей выборки принадлежит результат данного испытываемого.

Проделайте анализ с помощью таблицы сопряженности 3x3 и сопоставьте результаты с предыдущим анализом.

Упражнение 11.3.5(4). Проведите дисперсионный анализ с зависимой переменной *StressScore* и фактором *Period*. Сравните результаты. Корректен ли такой вид анализа для таких данных?

Заключительное упражнение 11.3.5(5). Это последнее упражнение нашего учебника можно считать итоговым тестом. Откройте файл *LastExersize.sav*. Данные описывают динамику результативности юных спортсменов, тренирующихся в школе, ориентированной на высокие спортивные достижения и практикующей очень жесткий режим поощрений и наказаний, однако без отчисления за плохие показатели. Переменная *Year* кодирует год тренировки (оценивались разные учащиеся первого, второго и третьего года тренировок, по 60 человек каждого года). Переменная *Results* кодирует отнесенные к возрастным нормам спортивные результаты учеников, т.е. рост результатов означает прогресс более быстрый, чем средний прогресс в спортивных школах данного профиля.

Нарисуйте график рассеяния, примените уместные известные вам методы и опишите результаты. После завершения самостоятельной работы наше описание вы можете найти в файле *LastExersizeDescription.doc*.

Заключительные методологические замечания

Мы надеялись сделать статистические методы понятным и удобным инструментом оценивания разнообразных аспектов количественных данных. Насколько нам это удалось, судить вам.

Несколько заключительных замечаний. Мы всячески старались увести читателя от догматического следования предписаниям и алгоритмам обработки, в частности от пристрастия принимать и отвергать гипотезы, ориентируясь на пресловутый уровень значимости 0.05.

Лучшая, с нашей точки зрения, часть научного сообщества принимает в настоящее время очень серьезные усилия, чтобы переломить довольно грустную тенденцию последних десятилетий, заключающуюся в упрощающем подходе к обучению статистике и приведшую к снижению качества понимания материала даже на уровне экспертов и рецензентов. Вслед за вызвавшей много шума публикацией Американской психологической ассоциации о неудачных попытках повторить эксперименты, отчеты о которых был опубликованы в ведущих журналах¹⁰², Американская статистическая ассоциация выпустила бюллетень, текст которого начинается с очень точно выражающей суть дела шутки:

Вопрос: Почему так много колледжей и университетов учат о значимости 0.05?

Ответ: Потому что это все еще то, что используют научное сообщество и редакторы журналов.

Вопрос: А почему так много людей все еще используют значимость 0.05?

Ответ: Потому что этому учат в колледжах и университетах¹⁰³.

Мы хотим, чтобы читатели, которые захотят понять описанные выше проблемы, связанные с публикационным сдвигом, выдвижением и оцен-

¹⁰² Open Science Collaboration et al. Estimating the reproducibility of psychological science //Science. – 2015. – Т. 349. – №. 6251. – С. aac4716. В публикации дан отчет о репликациях 100 экспериментов, результаты которых опубликованы в ведущих журналах. Больше 60% репликаций не подтвердили статистическую значимость выводов на уровне 0.05.

¹⁰³ Wasserstein R. L. et al. The ASA's statement on p-values: context, process, and purpose //The American Statistician. – 2016. – Т. 70. – №. 2. – С. 129-133.

кой гипотез по уже имеющимся данным, некорректированными множественными сравнениями и т.п., могли найти в нашей книге достаточно простые и ясные, но не упрощающие разъяснения.

Для того чтобы правильно понять и употреблять описанные в нашем учебнике статистические методы, надо понять, во-первых, основные ограничения подхода к оцениванию результатов исследований с помощью статистических средств, а во-вторых, четко увидеть, что эти ограничения не суть недостатки именно данного подхода, а связаны с неустранимой проблемой индуктивного вывода — принятия общих утверждений на основании ограниченного эмпирического материала. Альтернативные подходы — например байесовский (см. приложение 5) — также не могут ни решить эту проблему, ни обойти ее.

Мы надеемся, что после наших настойчивых разъяснений понятно, что 95-процентный интервал определенным образом указывает на вероятные значения оцениваемого параметра, но полагаться на эти оценки можно лишь очень осторожно — у других исследователей оценки могут получиться совсем другие. Точно так же и отвержение нулевой гипотезы на уровне значимости 0.05 вполне реально может оказаться ошибкой первого рода.

Может возникнуть впечатление, что статистика вообще ничего не может доказать. Теоретически это так. Действительно, как пишут критики традиционного подхода¹⁰⁴, схема статистического вывода такова: (1) мы получаем результат (выборку) и оцениваем вероятность p получения такого результата¹⁰⁵ при условии, что истинна гипотеза H_0 , т.е. вероятность данной выборки при условии H_0 . Но вывод, который мы пытаемся сделать, звучит примерно так: вероятность гипотезы H_0 при условии данной выборки равна p , и тем самым вероятность альтернативной гипотезы равна $1 - p$. Это, конечно, логическая ошибка.

Критика совершенно правильна, однако в ситуациях, где альтернатива в принципе допустима¹⁰⁶, она может быть признана практически достоверной, если соответствующая значимость равна, например,

¹⁰⁴ По английски он называется «Null hypothesis significance testing» с общепринятой аббревиатурой NHST.

¹⁰⁵ Если выражаться точно: такого или более экстремального для гипотезы результата.

¹⁰⁶ Например, дальнейшее рассуждение не может считаться бесспорным в случае, если альтернативная гипотеза это гипотеза о существовании телепатии. Разногласия по поводу принципиальной возможности этой альтернативной гипотезе, утверждающей отсутствие телепатии, могут быть столь велики, что интерпретации результатов несовместимо разойдутся.

0.0000001, а экспериментальные процедуры признаются безупречными, то есть нет сомнения в добросовестности исполнителей и корректности экспериментального дизайна.

Однако единственный эксперимент редко (а скорее никогда не) удаётся такой чести, поскольку гораздо надежнее выглядят пусть не столь статистически убедительные, но согласованные результаты исследований, проведенных несколькими коллективами.

Для того чтобы понять существо дела, надо выйти в социальное пространство науки. Роль единственного эксперимента в методологии науки иногда сильно преувеличивается¹⁰⁷. Методология NHST как раз и опирается на это преувеличение. На самом деле нам очень редко (а в научных исследованиях — никогда не) требуется провести решающий эксперимент и окончательно принять одну или другую гипотезу. Если бы такая необходимость возникла, то у нас не было бы лучшего варианта, кроме принятия решения на основании процедуры NHST. Мы могли бы только менять уровень значимости критерия в зависимости от цены ошибок первого и второго рода и субъективных (или объективных, если они даны) вероятностей гипотез.

Однако таких ситуаций в научной деятельности не бывает. Если вопрос интересен, то десятки коллективов проводят исследования в сходных или совпадающих условиях, и их результаты могут быть собраны и соединены на уровне так называемого мета-анализа. Только на этом уровне может быть достигнута практически полная достоверность утверждений, поскольку только в этом случае реально возникает полное доверие воспроизведенной многими коллективами экспериментальной процедуре, что в случае единственного эксперимента недостижимо.

Но для проведения корректного метаанализа требуется публикация (т.е. доступность сообществу) результатов *всех* проведенных исследований, желательно также, чтобы были доступны и «сырые» данные.

Таким образом, если в нашем исследовании 95-процентный доверительный интервал не содержит нулевую гипотезу, то это — скромное основание поделиться с коллегами на страницах журнала радостью и надеждами на то, что интересный результат в будущем будет достоверно подтвержден. Если же доверительный интервал включает нуль-гипотезу, то поделиться своим огорчением в печати или специальных базах результатов — наш прямой долг перед научным сообществом. В обоих случаях будет хорошим тоном сделать доступными

¹⁰⁷ Этот промах делают обычно те, кто ориентируется на работы К. Поппера, хотя его последователи, в частности И. Лакатос, указали вполне адекватное место «решающему» эксперименту.

данные¹⁰⁸. Если мы этого не делаем, то описанный выше (5.1.4) публикационный сдвиг исказит общую картину для научного сообщества.

Кроме того, необходимо различать ситуации, когда гипотезы, проверившиеся в исследовании, были сформулированы до сбора данных или в результате работы с данными. Ценность результатов в последнем случае значительно ниже, поскольку умелый «обработчик» практически всегда может найти «гипотезу», задним числом подтвердившуюся уже имеющимися у него данными¹⁰⁹. Для того чтобы это различие стало рабочим, некоторые журналы предполагают регистрировать планируемые исследования по желанию авторов. В случае такой регистрации отчет получает специальный бейдж¹¹⁰.

Как нам кажется, эта практика станет общей уже в ближайшие годы. Наш учебник помогает подготовиться к этим переменам.

¹⁰⁸ Немотивированный отказ от предоставления доступа к данным уже считается серьезным недостатком статьи в некоторых журналах — Science, Шпрингеровские журналы соответствующего профиля (Munafò M.R., Nosek B.A., Bishop D.V. M., Button K.S., Chambers Ch.D. A manifesto for reproducible science // *Nature human behaviour*. – 2017. – Т. 1. – №. 1. – С. 0021).

¹⁰⁹ Имеются подобные ситуации, даже не требующие большого умения — это корреляционные исследования с большим числом переменных, когда в отчете указываются значимые корреляции из огромной таблицы, хотя гипотезы о конкретных связях заранее не формулировались.

¹¹⁰ Там же.

Приложение 1.

Основы работы в SPSS

Как мы уже писали, в задачи нашего учебника не входит подробное обсуждение работы в SPSS, однако, чтобы читатель, незнакомый с этим статистическим пакетом, мог выполнять предлагаемые нами упражнения и задания, остановимся на базовых элементах этой программы.

Главное рабочее окно программы организовано в виде матрицы данных (рис. П1(1)).

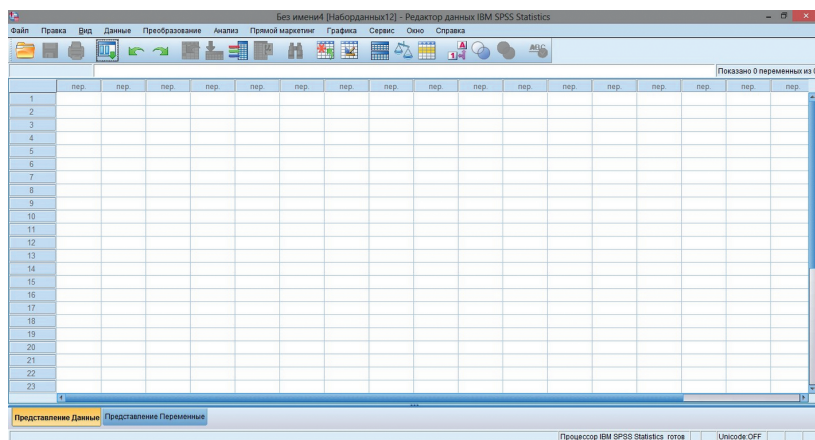


Рис. П1(1). Окно данных в SPSS

Столбцы в этой таблице — это измеряемые параметры (переменные), а строки — это наблюдения. Если рассматривать результаты психологических исследований, каждая строка, как правило, включает набор данных об одном испытуемом. В некоторых случаях, например в психосемантических исследованиях, строки могут содержать другую информацию, однако мы не будем подробно останавливаться на этом. Изменяемые параметры могут быть любого типа, например, имя испытуемого, его возраст, пол, результаты различных опросников, заполненных испытуемым, время реакции в каком-либо эксперименте и т.п. В зависимости от информации, содержащейся в переменных, они могут быть разных типов. Два основных типа, которые используются чаще

всего, — это числовые и символьные переменные. Для управления переменными, изменения их имен, типов и т.п. в SPSS существует режим Представление Переменные (Variable view)¹¹¹. Рассмотрим модельный пример ввода данных исследования в SPSS.

Пример П1(2) Организация ввода данных исследования в SPSS

Проведено тестирование уровня ситуативной и личностной тревожности 10 студентов (5 мужчин и 5 женщин). В результате тестирования у каждого испытуемого был рассчитан итоговый балл ситуативной и личностной тревожности. Чем выше балл, тем более выражен тот или иной вид тревожности, максимально возможный балл равнялся 30. Также исследователи регистрировали пол и возраст испытуемых. Помимо этого, студентам задавался следующий вопрос: «Насколько часто Вы волнуетесь перед экзаменами?». Испытуемые должны были обвести один из вариантов ответа: «Почти никогда», «Редко», «Иногда», «Часто/почти всегда».

Полученные в ходе исследования результаты представлены в таблице П1(3).

Таблица П1(3). Данные исследования личностной и ситуативной тревожности.

№	Имя	Пол	Возраст	Ответ на вопрос анкеты	Ситуативная тревожность	Личностная тревожность
1	Н.И.	Женский	20	Часто/почти всегда	20	26
2	К.Л.	Мужской	18	Почти никогда	18	16
3	П.Д.	Мужской	18	Иногда	15	20
4	У.Р.	Женский	18	Часто/почти всегда	17	21
5	И.К.	Мужской	19	Редко	19	17
6	Т.К.	Женский	20	Редко	17	18
7	А.Н.	Женский	21	Иногда	21	19
8	Г.М.	Мужской	21	Почти никогда	10	19
9	В.Н.	Женский	19	Иногда	19	18
10	Л.Н.	Мужской	20	Иногда	20	20

¹¹¹ Так как в настоящее время в нашей стране одинаково распространены английская и русифицированная версия программы, мы будем приводить русский вариант, а затем, в скобках, давать английский аналог. Перевод на русский язык может меняться в зависимости от версии программы, мы приводим перевод для версии 23. Снимки экрана и таблицы расчетов мы будем давать только в русской версии.

Рассмотрим, как корректно ввести полученные результаты в SPSS. При открытии программы появляется пустая таблица данных. Перед вводом данных необходимо разобрать, какие переменные включены в исследование и как их оптимально закодировать для ввода в SPSS. Выделим переменные в нашем исследовании: две переменные, измеренные в номинальной шкале: имя испытуемых и пол; ответ на вопрос анкеты (шкала порядка) и три интервальные шкалы: возраст, ситуативная тревожность, личностная тревожность. Следовательно, в нашей таблице будет 6 столбцов.

На закладке *Представление Переменные (Variable View)* можно задать имена переменных в первом столбце, озаглавленном *Имя (Name)*. В названиях переменных можно использовать буквы латиницы или кириллицы¹¹² и цифры, запрещены пробелы и знаки препинания (допустимо использовать нижние подчеркивания и точки). Еще одно требование к названию переменной в SPSS — первый символ не должен быть цифрой. Сразу после ввода переменной на листе появляются ее данные, настроенные по умолчанию: тип переменной, длина переменной (максимальное количество знаков), метка переменной (этот столбец по умолчанию пуст), значение, пропуски, колонки, выравнивание, шкала. После ввода названия переменной можно определить другие ее свойства. В столбце *Тип (Type)* можно выбрать тип данных. По умолчанию новая переменная имеет *числовой* формат (*Numeric*), однако можно указать и другие: числовой с запятой-разделителем, отделяющий десятые и сотые доли, с точкой-разделителем, научное обозначение, дату, доллары и иную валюту и текстовую строку. В практике психологических исследований, как правило, достаточно первого и последнего варианта: числового и текстового. В нашем случае переменная *Имя* является текстовой, поэтому для нее тип данных нужно изменить на *Текстовая (String)*.

Следующий показатель — *Ширина (Width)*, количество символов в переменной, ее можно варьировать, если данные (например, текст), не умещаются в ячейке.

Четвертый столбец — *Знаков после запятой (Decimals)* — количество отображаемых у числовых переменных знаков после запятой (при изменении этого свойства меняется именно отображение числа, округления в расчетах не происходит).

Пятый столбец — *Метка (Label)* — название переменной в исследовании. Здесь может быть указано любое название или пояснение к переменной без ограничений в символах — в том виде, в каком оно будет пред-

¹¹² Технически допустимо использование кириллицы в названиях переменных, однако по нашему опыту это в непредсказуемые моменты может приводить к сбоям в выводе результатов. Поэтому мы рекомендуем использовать только латиницу.

ставлено в таблицах после обработки данных. Это поле необязательно для заполнения, но мы рекомендуем использовать его для более полного и читаемого обозначения переменных, что упрощает форматирование результатов расчетов и делает более понятной саму таблицу данных. Впишем более подробные названия переменных в столбец (рис. П1(4)).

В следующем столбце — *Значения (Value)* — для переменных, измеренных в номинативной шкале, можно указать, какие обозначения для разных уровней переменной были использованы в исследовании. В нашем случае в номинативной шкале выражен показатель *Пол* (мужской/женский). Показатель *Анкета* в зависимости от целей и задач исследования можно считать порядковым (мы предполагаем, что ответы «Часто/почти всегда», «Иногда», «Редко», «Почти никогда» выстраиваются в линейную последовательность) и будем кодировать их натуральными числами от 1 до 4.

Чтобы присвоить значения переменной *Пол*, поставим курсор на вторую строку в столбце *Значения* и нажмем на многоточие. В появившейся таблице нужно указать числовое значение и его содержание и нажать на кнопку *Добавить*. Мы будем обозначать женский пол цифрой «1», мужской пол — цифрой «2». Введем эти обозначения в диалоговое окно. Теперь как в таблице данных, так и в таблицах результатов можно будет увидеть словесное обозначение группы испытуемого, а не только ее номер.

Имя	Тип	Ширина	Знаков	Метка	Значения	Пропущенн.	Ширина...	Выравнивание	Мера
name	Текстовая	16	0	Имя	Нет	Нет	8	По левом...	Номинальн...
gender	Числовой	8	2	Пол	(1,00, Женский ...	Нет	8	По право...	Номинальн...
age	Числовой	8	2	Возраст	Нет	Нет	8	По право...	Нет данных
survey	Числовой	8	2	Ответ на анкетный вопрос	(1,00, Почти ник...	Нет	8	По право...	Номина...
sit_anxiety	Числовой	8	2	Ситуативная тревожность	Нет	Нет	8	По право...	Нет данных
reg_anxiety	Числовой	8	2	Личностная тревожность	Нет	Нет	8	По право...	Нет данных

Рис. П1(4). Заполненное окно описания данных

На остальных параметрах переменных мы пока останавливаться не будем. Подробную информацию о них можно найти в специальных пособиях по использованию SPSS.

Вернемся к таблице данных, нажав на закладку *Представление Данные (Data View)* внизу рабочего окна. На листе данных появились названия переменных. Теперь мы можем ввести данные (рис. П1(5)). (После ввода каждого значения в соответствующую клетку можно нажать клавишу Enter или любую стрелку на клавиатуре.)

После того как данные исследования полностью введены в программу, можно приступать к обработке полученных результатов — проводить различные расчеты, строить диаграммы и т.п.

	name	gender	age	survey	sit_anxiety	pers_anxiety	pe
1	Н.И.	1	20	4	20	26	
2	К.Л.	2	18	1	18	16	
3	П.Д.	2	18	3	15	20	
4	У.Р.	1	18	4	17	21	
5	И.К.	2	19	2	19	17	
6	Т.К.	1	20	2	17	18	
7	А.Н.	1	21	3	21	19	
8	Г.М.	2	21	1	10	19	
9	В.Н.	1	19	3	19	18	
10	Л.Н.	2	20	3	20	20	
11							

Рис. П1(5). Заполненная таблица данных

Мы можем сохранить файл данных с помощью соответствующего пункта меню **Файл (File)**, а затем использовать этот и другие файлы для последующей работы. Далее, мы предлагаем читателю самостоятельно выполнить задание, подобное описанному выше.

Упражнение П1(6). В таблице П1(7) представлены показатели эмоционального интеллекта в трех группах испытуемых. Введите результаты эксперимента в SPSS.

Таблица П1(7). Данные исследования эмоционального интеллекта у испытуемых с истерическим расстройством

Номер испытуемого	Балл по шкале «Эмоциональная осведомленность»	Балл по шкале «Распознавание эмоций других людей»
Здоровые испытуемые		
1	4	6
2	6	5
3	6	3
4	3	2
5	4	2
Испытуемые с истерическим расстройством		
6	6	5
7	7	4
8	3	5
9	4	6
10	3	5
Испытуемые с шизоидным расстройством		
11	3	4
12	5	2
13	2	5
14	6	3
15	2	4

2. Статистическая обработка данных в SPSS

После ввода данных мы можем приступить к их статистическому анализу. Опишем общую схему проведения такого анализа в SPSS. Для вызова различных процедур по статистической обработке данных нужно зайти в меню *Анализ (Analyze)*. В разделах этого меню содержатся пункты по вызову различных процедур обработки данных (рис. П1(8)). При выборе каждого из них появляется соответствующее методу диалоговое окно, в котором необходимо задать параметры проводимого анализа. В каждом виде анализа эти параметры могут быть самыми различными, однако принцип заполнения диалогового окна может быть описан в достаточно общем виде.

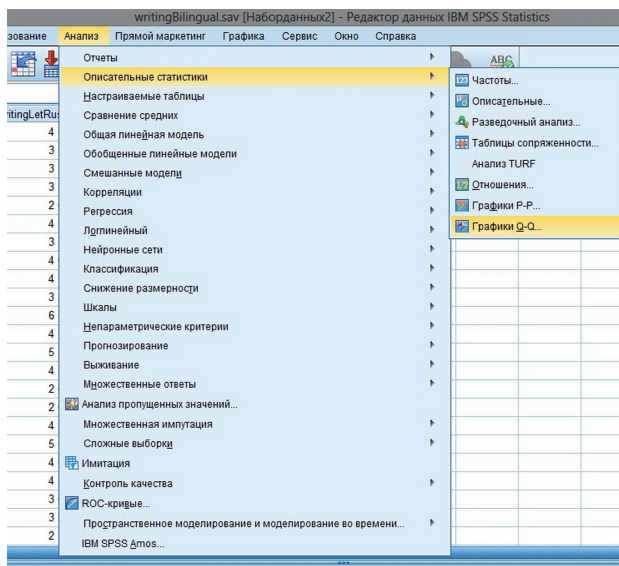


Рис. П1(8). Меню процедур статистического анализа

Количество статистических процедур, реализованных в SPSS, достаточно велико, в рамках нашего учебника мы рассмотрим только некоторую их часть. Основной целью этого приложения является описание техники проведения статистического анализа в SPSS в самом общем виде на материале простого примера.

Пример П1(9). Процедура статистического анализа в SPSS

Рассмотрим общую схему процедуры статистического анализа в SPSS на простом примере — расчете средних значений переменной. В качестве материала мы возьмем данные исследования скорости письма у

детей-билингвов младшего школьного возраста. Мы предлагали второклассникам, учащимся в Финско-русской школе города Хельсинки, списать два небольших текста — один на русском, второй на финском языках¹¹³. Процесс письма регистрировался с помощью графического планшета и затем подвергался обработке и анализу. В частности, оценивалась скорость письма на двух языках, основным показателем было среднее время, затрачиваемое ребенком на написание одной буквы текста. Результаты этого исследования содержатся в файле *WritingBilingual.sav*. Каждый испытуемый в этом наборе данных характеризуется четырьмя параметрами: идентификационным номером (переменная *id*), доминантностью русского или финского языка (в зависимости от основного «домашнего» языка, используемого в семье, переменная *group*), средним временем написания буквы на русском и на финском языках (переменные *WritingLetRus* и *WritingLetFin* соответственно).

Первая задача, которую можно решить с помощью SPSS — это расчет описательной статистики¹¹⁴ времени, затрачиваемого на письмо одной буквы на двух языках. Для этого необходимо войти в пункт Анализ, в нем выбрать пункт Описательные статистики и в подменю выбрать пункт «Описательные». Подобные последовательности действий с пунктами меню мы будем далее обозначать следующим образом:

Анализ — Описательные статистики — Описательные (*Analyze — Descriptive statistics — Descriptives*).

После этого на экране появится диалоговое окно расчета описательной статистики. Исходный вид окна изображен на рис. П1(10).

Как и во всех диалоговых окнах статистических процедур, слева в нем расположен список всех переменных, имеющихся в исходном файле данных. Справа находится поле, в которое необходимо внести переменные, включаемые в анализ, к ним будут применены выбранные статистические процедуры.

¹¹³ Более подробное описание этого исследования и его результатов можно найти, например, в статье: Корнеев А.А., Протасова Е.Ю. Письмо у финско-русских билингвов младшего школьного возраста // Психолингвистические аспекты изучения речевой деятельности. — Т. 13 из Труды Уральского психолингвистического общества. — Уральский государственный педагогический университет Екатеринбург, 2015. — С. 107–122.

¹¹⁴ Термин «Описательная статистика» в данном случае обозначает, что будут рассчитываться параметры, описывающие распределение данных (в первую очередь, по умолчанию — среднее значение и стандартное отклонение, подробнее см. параграф 3.1)

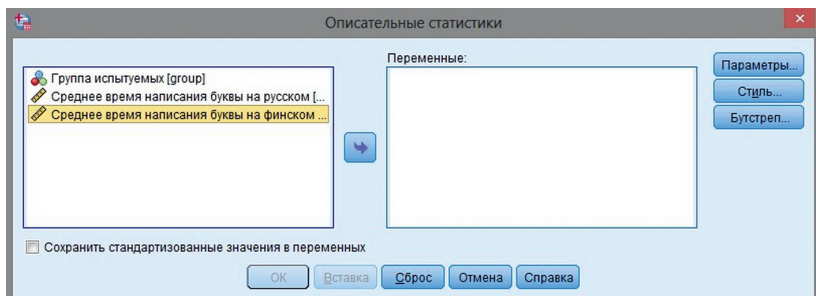


Рис. П1(10). Окно выбора переменных и установки параметров расчета описательной статистики

В нашем случае две переменные — *WritingLetRus* и *WritingLetFin*¹¹⁵. Перенесем их в поле **Переменные (Variables)**¹¹⁶. Затем мы можем задать те показатели распределения, которые будут рассчитаны для выбранной переменной. Для этого надо нажать на кнопку **Параметры (Options)**. Раскроется дополнительное окно, в котором приводится перечень параметров, расчет которых доступен в рамках данной процедуры (рис. П1(11)).

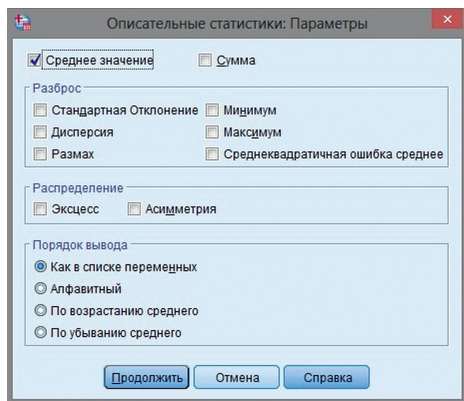


Рис. П1(11). Диалоговое окно для ввода дополнительных параметров процедуры расчета описательной статистики в SPSS

¹¹⁵ Обратим внимание, что в общем списке переменных при описании переменной сначала выводится ее метка, введенная нами на русском языке, а затем — собственно, название переменной

¹¹⁶ В это окно может быть перемещено любое количество имеющихся переменных.

Помимо среднего значения, по умолчанию рассчитываются минимальное и максимальное значения выбранных переменных, а также их стандартное отклонение. В рамках этого примера мы предлагаем не рассчитывать стандартное отклонение, минимум и максимум, нужно снять отметки рядом с этими параметрами. Любознательный читатель может выбрать другие параметры, поэкспериментировать с настройками, мы же обсудим наиболее простой вариант.

Такие меню настройки дополнительных параметров есть практически в каждой разновидности статистического анализа, представленного в SPSS. За подробной и исчерпывающей информацией читатель может обратиться к специализированным книгам по SPSS. В рамках этой книги при описании различных расчетов, проводимых в SPSS, мы будем указывать на параметры, имеющие наибольшую важность при базовом освоении темы.

После установки дополнительных настроек можно нажать кнопку *Продолжить (Continue)*, чтобы вернуться к основному диалоговому окну. Далее, для проведения расчетов и получения результата следует нажать кнопку *OK*. Результаты расчетов будут выведены в отдельное окно *Вывода (Output)*. Результаты дальнейших расчетов будут выводиться в это же окно в порядке выполнения различных процедур. При необходимости все результаты могут быть сохранены в отдельном файле или импортированы в различные форматы (MsWord, html и т.д.)

Рассмотрим полученные результаты (таблица П1(12)).

Таблица П1(12). Описательные статистики

Описательные статистики		
	N	Среднее
Среднее время написания буквы на русском	31	3 931.04
Среднее время написания буквы на финском	31	2 639.32
N валидных (по списку)	31	

В таблице столбец N содержит информацию о количестве испытуемых, далее идут столбцы, содержащие средние выбранных переменных. Как видно из результатов, среднее время написания буквы на русском составило 3931 мс, а на финском — 2639 мс. Таким образом, можно сделать вывод о том, что скорость письма у детей-второклассников на финском оказывается выше скорости письма на русском. Насколько эти различия заметны и отличимы от случайных колебаний можно определить с помощью специальных статистических критериев, о которых идет речь в основной части учебника.

Если нас интересует сравнение средних значений времени написания букв на двух языках у детей с разными доминантными языками и мы хотим оценить, насколько доминантность языка влияет на скорость письма, то необходимо ввести группирующую переменную. В рассматриваемом наборе данных переменная `group` обозначает принадлежность испытуемого к той или иной группе. Значения группирующих и других так называемых независимых переменных так или иначе контролируются исследователем при организации и проведении исследования. В данном случае мы отбирали детей таким образом, чтобы среди них было примерно равное число детей с доминантным русским и финским языками. Те переменные, на которые могут влиять независимые переменные, называются зависимыми. В нашем примере это среднее время написания буквы.

Рассмотрим, как в SPSS можно рассчитать средние значения для групп испытуемых, определенных с помощью такой группирующей переменной. Для этого надо выбрать в меню последовательность **Анализ — Сравнение средних — Средние (Analysis — Compare means — Means)**. В появившемся диалоговом окне для проведения расчетов необходимо выделить зависимые и независимые переменные. Заполненное окно изображено на рис. П1(13).

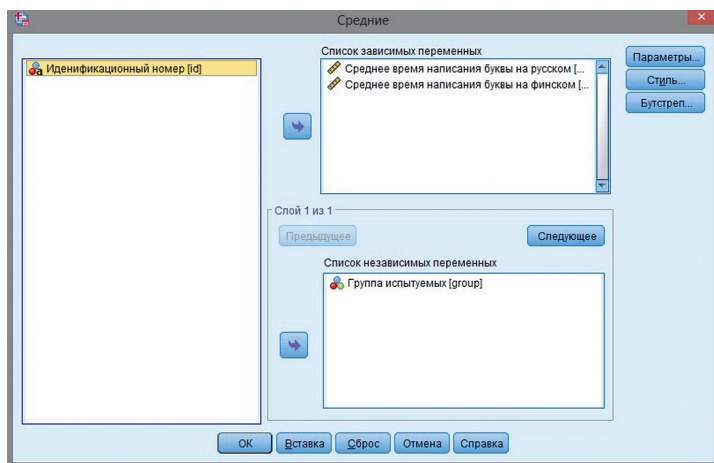


Рис. П1(13). Заполненное диалоговое окно для сравнения средних

После нажатия кнопки *OK* в окне вывода появляется новая таблица с результатами расчета средних для групп испытуемых (таблица П1(14)).

Таблица П1(14). Сравнение средних значений двух переменных

Группа испытуемых		Среднее время написания буквы на русском	Среднее время написания буквы на финском
Доминантный русский	Среднее	3 530,67	2 533,73
	N	16	16
Доминантный финский	Среднее	4 358,10	2 751,94
	N	15	15
Всего	Среднее	3 931,04	2 639,32
	N	31	31

Как видно, эта таблица организована несколько сложнее, чем предыдущая. В ней приведены средние значения показателей времени написания буквы как для всей выборки в целом (последние две строчки), так и для двух групп испытуемых по отдельности. Полученные результаты позволяют заметить, что между группами имеются различия в скорости письма, причем доминантность языка приводит ускорению письма: на русском относительно быстро пишут дети с русским доминантным языком, а при письме на финском — наоборот, дети с доминантным финским.

Еще один важный аспект анализа данных, на котором нужно остановиться, описывая основные функции SPSS — это визуализация данных, их графическое отображение. В SPSS есть возможность создавать разнообразные графики, сейчас мы приведем несколько простейших примеров, чтобы у читателя сложилось общее представление о способах графического отображения различных параметров имеющихся данных.

Рассмотрим общий принцип построения диаграмм в SPSS на примере уже описанных данных. Допустим, мы хотим графически отобразить различия во времени написания буквы на двух языках отдельно у детей с доминантным русским и финским языками, т.е. визуализировать приведенную выше таблицу.

Для отображения графиков в SPSS нужно открыть пункт меню **Графика (Graphics)**. Затем можно действовать двумя способами: либо использовать появившийся в сравнительно поздних (начиная с 15-й) версиях **конструктор диаграмм (Chart Builder)**, либо выбрать нужный тип диаграммы в подменю **Устаревшие диалоговые окна (Legacy Dialogs)**. Мы будем описывать наиболее простой и удобный способ в зависимости от задачи.

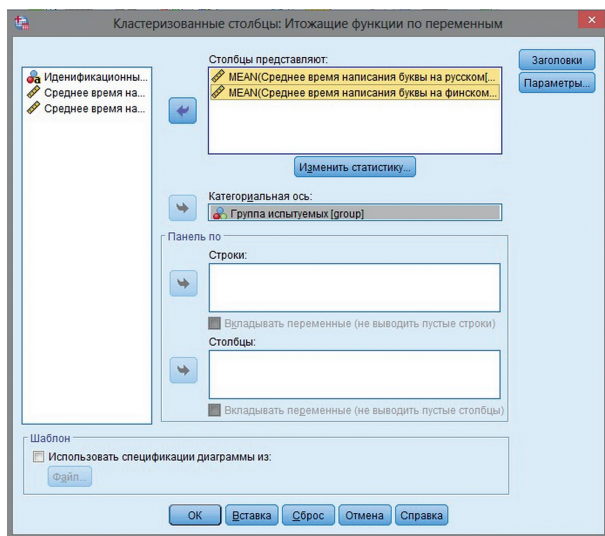


Рис. П1(15). Диалоговое окно выбора типа столбиковой диаграммы

В нашем примере проще нарисовать график распределения возрастов с помощью последовательности *Графика — Устаревшие диалоговые окна — Столбцы — Кластеризованные (Legacy Dialogs — Bar — Clustered)*. В качестве исходных данных зададим *Итоги по отдельным переменным (Summaries of separate variables)* (рис. П1(15)).

После нажатия кнопки *Задать (Define)* появится диалоговое окно построения диаграммы. Такие окна в SPSS устроены по тому же принципу, что и окна задания параметров статистических процедур: слева приведен список всех переменных, имеющихся в таблице данных, а в правой части — поля для ввода тех переменных, значения которых будут отображены в диаграмме. В данном случае нам необходимо заполнить поле *Столбцы представляют (Bars Represent)* — туда надо перенести две переменные, содержащие показатели времени написания буквы, и поле *Категориальная ось (Category Axis)* — в него надо внести переменную *Group*. При переносе переменных в первое поле по умолчанию будет отображаться среднее значение внесенных переменных. При необходимости можно изменить тип рассчитываемой статистики: если нажать кнопку *Изменить Статистику (Change Statistics)*, то появится дополнительное окно, где можно выбрать другие параметры: медиану, стандартное отклонение и др.

После нажатия кнопки *OK* в окне вывода появится диаграмма, изображенная на рис. П1(16).

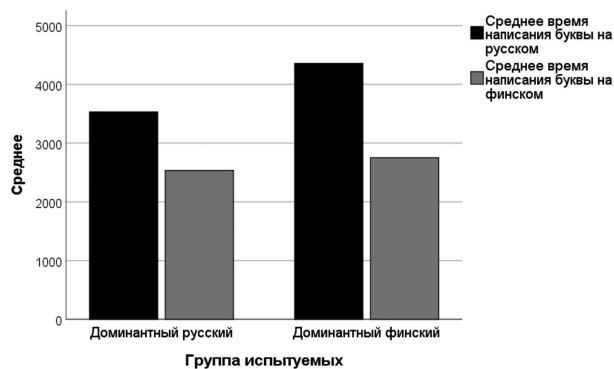


Рис. П1(16). Столбчатая диаграмма средних значений времени написания одной буквы на русском и финском языках в двух группах испытуемых

На этом графике достаточно хорошо видно, что время написания буквы на русском меньше в группе детей с доминантным русским языком, а на финском они пишут букву в среднем немного дольше, чем их сверстники с доминантным финским языком. При этом также можно отметить, что в целом время письма на финском меньше, чем на русском.

В данном случае диаграмма, как мы уже говорили, является практически прямым отображением рассчитанной выше таблицы средних, поэтому никакой новой информации из нее мы не извлечем. Однако такое наглядное представление данных в некоторых случаях бывает более удобным. Кроме того, достаточно часто построение диаграмм может дать некоторую дополнительную информацию о данных.

В SPSS можно строить как такие простые графики, так и намного более сложные диаграммы, которые позволяют более полно оценить и проинтерпретировать результаты, получаемые при применении различных статистических критериев. В тексте учебника по мере обсуждения решения тех или иных задач методами математической статистики мы кратко описываем способы построения разнообразных диаграмм средствами SPSS.

Еще одна важная процедура, которая часто используется при обработке данных, — это преобразование переменных и расчет новых параметров на основании имеющихся данных. Так, например, в нашем случае можно рассчитать разность времени написания одной буквы для каждого участника исследования. Для подобного рода расчетов в SPSS надо выбрать последовательность **Преобразование — Вычислить переменную (Transform — Compute)**. После этого появляется диалоговое окно, в котором надо ввести имя новой рассчитываемой переменной в поле **Целевая переменная**

(*Target variable*), а также формулу вычисления переменной в поле *Числовое выражение* (*Numeric expression*). Формула может быть и очень простой, и достаточно сложной, в нее можно включать имеющиеся переменные, а также использовать различные встроенные функции SPSS из списка Функции и специальные переменные (*Functions and special variables*) в правой части диалогового окна (например, функцию логарифмирования, извлечения квадратного корня, генерации случайных чисел и т.п.). В нашем случае формула расчета будет достаточно проста: нам надо из среднего времени написания буквы на русском вычесть среднее время написания буквы на финском: *WritingLetRus – WritingLetFin*. Новую переменную можно назвать *diffWritingLet*. Заполненное окно приведено на рис. П1(17).

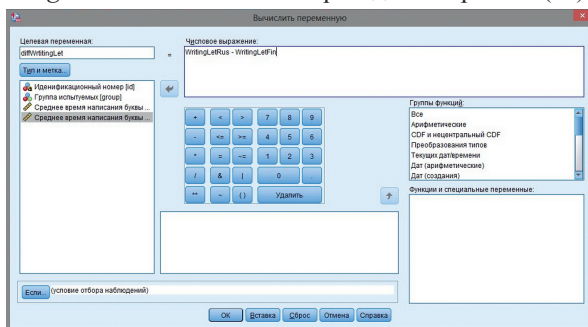


Рис. П1(17). Диалоговое окно для задания вычисления новой переменной

После нажатия кнопки *OK* в таблице данных появится новая переменная *diffWritingLet*, рассчитанная по введенной формуле, ее можно также подвергать статистическому анализу. Таким образом, данная процедура позволяет манипулировать с данными, вводя новые переменные, производные от имеющихся данных.

Упражнение П1(18). В файле *WritingSkills.sav* приведены результаты исследования письма у детей, учащихся первого и второго классов. В качестве экспериментального задания им давали списать небольшой текст. В качестве показателей уровня развития навыка письма регистрировались два показателя — время выполнения заданий в секундах и количество допущенных при письме ошибок. Файл данных, таким образом, содержит три показателя: класс, в котором обучался испытуемый (переменная *group*), время выполнения задания (переменная *time*) и количество допущенных ошибок (переменная *mistakes*).

- Вычислите средние показатели времени списывания и количества ошибок в первом и во втором классе.
- Нарисуйте график изменения этих двух показателей.

Приложение 2. _____

Дисперсия суммы независимых случайных величин

Теорема. Если X и Y — независимые случайные величины, то

$$D_{X+Y} = D_X + D_Y$$

Будем проводить доказательство для случайных величин, принимающих только по два значения:

X	x_1	x_2
	p_1	p_2

и

Y	y_1	y_2
	q_1	q_2

Предположим сначала, что их математические ожидания равны нулю, что означает, что равно нулю также и математическое ожидание суммы $X+Y$.

Распределение их суммы задается следующей таблицей:

X+Y	x_1+y_1	x_1+y_2	x_2+y_1	x_2+y_2
	p_1q_1	p_1q_2	p_2q_1	p_2q_2

Дисперсию суммы вычислим, расположив четверку слагаемых в виде таблицы (нам нет необходимости явно записывать в скобках равное нулю математическое ожидание):

$$D_{X+Y} = \begin{aligned} & (x_1 + y_1)^2 p_1 q_1 + (x_1 + y_2)^2 p_1 q_2 \\ & \quad + \\ & (x_2 + y_1)^2 p_2 q_1 + (x_2 + y_2)^2 p_2 q_2 \end{aligned}$$

Раскрыв скобки, получим двенадцать слагаемых (удвоенные произведения и мы собрали отдельно справа):

$$\begin{aligned} & x_1^2 p_1 q_1 + y_1^2 p_1 q_1 + x_1^2 p_1 q_2 + y_2^2 p_1 q_2 + 2x_1 y_1 p_1 q_1 + 2x_1 y_2 p_1 q_2 \\ & \quad + \\ & x_2^2 p_2 q_1 + y_1^2 p_2 q_1 + x_2^2 p_2 q_2 + y_2^2 p_2 q_2 + 2x_2 y_1 p_2 q_1 + 2x_2 y_2 p_2 q_2 \end{aligned}$$

Последние пары слагаемых в каждой строке рассмотрим отдельно:

$$2x_1 y_1 p_1 q_1 + 2x_1 y_2 p_1 q_2 = 2x_1 p_1 (y_1 q_1 + y_2 q_2) = 0;$$

$$2x_2 y_1 p_2 q_1 + 2x_2 y_2 p_2 q_2 = 2x_2 p_2 (y_1 q_1 + y_2 q_2) = 0$$

поскольку в скобках стоит математическое ожидание Y , равное нулю.

Это значит, что равна

$$\begin{array}{cccc} x_1^2 p_1 q_1 & + & y_1^2 p_1 q_1 & + & x_1^2 p_1 q_2 & + & y_2^2 p_1 q_2 \\ & & + & & + & & + \\ x_2^2 p_2 q_1 & + & y_1^2 p_2 q_1 & + & x_2^2 p_2 q_2 & + & y_2^2 p_2 q_2 \end{array}$$

Поменяем теперь местами второй и третий столбцы.

$$\begin{array}{cccc} x_1^2 p_1 q_1 & + & x_1^2 p_1 q_2 & + & y_1^2 p_1 q_1 & + & y_2^2 p_1 q_2 \\ & & + & & + & & + \\ x_2^2 p_2 q_1 & + & x_2^2 p_2 q_2 & + & y_1^2 p_2 q_1 & + & y_2^2 p_2 q_2 \end{array}$$

В новой таблице плюсы показывают порядок суммирования: в первой четверке мы сначала сложим элементы строк, а во второй четверке — сначала сложим элементы столбцов. Это позволит нам далее вынести за скобки нужные множители. Проводим группировку с помощью квадратных скобок и выносим за эти скобки множители:

$$\begin{aligned} & [x_1^2 p_1 q_1 + x_1^2 p_1 q_2] + [x_2^2 p_2 q_1 + x_2^2 p_2 q_2] + [y_1^2 p_1 q_1 + y_1^2 p_2 q_1] [y_2^2 p_2 q_2 + y_2^2 p_1 q_2] = \\ & = x_1^2 p_1 (q_1 + q_2) + x_2^2 p_2 (q_1 + q_2) + y_1^2 q_1 (p_1 + p_2) + y_2^2 q_2 (p_1 + p_2) \end{aligned}$$

Но $p_1 + p_1 = 1$, и $q_1 + q_1 = 1$, и, поэтому квадратные скобки вместе с их содержимым можно опустить. В результате получаем выражение $x_1^2 p_1 + x_2^2 p_2 + y_1^2 q_1 + y_2^2 q_2$ где первые два слагаемые это D_x , а вторые два — D_y .

Если исходно математические ожидания слагаемых не были равны нулю, центрируем их. Это не изменит дисперсии слагаемых и суммы, но проведенное выше доказательство будет к ним применимо.

Теорема доказана.

Приложение 3.

Степени свободы

В тех главах нашего учебника, где рассматривались различные оценки, вычисляемые с помощью сумм квадратов, возникал вопрос о числе степеней свободы для этих сумм. Первый раз такой вопрос возник, когда в выборочной оценке дисперсии сумма квадратов отклонений делилась не на n (количество слагаемых под знаком корня), а на $n - 1$. Затем более серьезные вопросы возникали, когда объявлялось, что в одновыборочной t -статистике в знаменателе формулы под корнем стоят n слагаемых, представляющих собой квадраты отклонений, но эта сумма имеет $n - 1$ степеней свободы. За всеми этими и другими аналогичными случаями, которые встретились нам в курсе статистических методов, стоит одна и та же суть дела, которую мы сейчас и рассмотрим.

Мы разберем простейший пример такого снижения числа степеней свободы и разъясним, как он должен обобщаться. Доказательство основного утверждения будет проведено для случая плоскости, требующего фактически только хорошего уровня школьных знаний, но обобщение доказательства требует понимания «устройства» n -мерного пространства, т.е. подразумевает знакомство с основами линейной алгебры.

Теорема ПЗ(1). Пусть у нас имеется стандартная нормально распределенная случайная величина X и ее n независимых испытаний $\{X_1, \dots, X_n\}$. Пусть случайная величина \bar{X} выражается через эти испытания по формуле $\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$, а случайная величина по формуле

$$S^2 = \frac{1}{n-1}((X_1 - \bar{X})^2 + \dots + (X_n - \bar{X})^2).$$

Тогда случайная величина S^2 имеет распределение Хи-квадрат с числом степеней свободы, равным $n-1$.

Прежде чем доказывать теорему, надо разъяснить, что представляет собой случайная величина, имеющая распределение Хи-квадрат с k степенями свободы. Напомним, что такое распределение имеет случайная величина

$$\frac{1}{k}(N_1^2 + \dots + N_k^2),$$

где N_i — независимые испытания стандартной нормальной случайной величины.

Мы рассмотрим сначала распределение Хи-квадрат с двумя степеня-

ми свободы, которое оказывается очень похоже на распределение квадрата длины расстояния от центра мишени до пулевого отверстия (см. 3.1.5).

Для понимания дальнейшего материала требуется составить понятие о двумерной стандартной нормальной величине, которая и описывает распределение пулевых отверстий нашего примера.

П3.1. Двумерное стандартное нормальное распределение

Двумерное стандартное нормальное распределение можно задать как распределение положения на плоскости случайной точки, каждая из координат которой независимо описывается стандартным нормальным распределением. Плотность распределения положения точки будет функцией, заданной на плоскости. Вероятность того, что результат испытания двумерной величины попадет в определенную область, вычисляется как интеграл этой функции по данной области. Чтобы понять суть дела, вспомним, что для обычного стандартного нормального распределения вероятность попадания в *очень* маленький отрезок длины l_x с левым концом в точке a оси X может быть приблизительно рассчитан как произведение длины отрезка на значение плотности распределения в точке a . Это произведение равно

$$l_x \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{a^2}{2}}.$$

Для нашего двумерного распределения с независимыми координатами вероятность попадания результата в очень маленький прямоугольник с левым нижним концом в точке (a, b) и со сторонами l_x и l_y , равна произведению вероятности попадания X -ой координаты в отрезок, начинающийся в точке a , и вероятности попадания Y -ой координаты в отрезок длины l_y , начинающийся в точке b . Это произведение записывается формулой

$$l_x \cdot l_y \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{a^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{b^2}{2}}.$$

т.е. равно площади прямоугольника (она равна $l_x \cdot l_y$), умноженной на произведение покоординатных плотностей, равное

$$\frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{a^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{b^2}{2}}.$$

Если превратить произведение экспонент в экспоненту суммы, то произведение плотностей равно

$$\frac{1}{2\pi} \cdot e^{-\frac{(a^2+b^2)}{2}}.$$

Функция

$$p(x, y) = \frac{1}{2\pi} \cdot e^{-\frac{(x^2+y^2)}{2}}$$

и задает плотность двумерного стандартного нормального распределения.

П3.2. Свойства двумерного стандартного распределения

Свойство П3.2(1). Проекция случайной точки, положение которой задается двумерной стандартной случайной величиной, на любую прямую, проходящую через начало координат, имеет стандартное нормальное распределение.

Нам потребуется частный случай этого свойства, когда прямая наклонена под углом 45 градусов к осям.

Сделаем двумерную замену переменных, определяемую уравнениями

$$x' = \frac{x + y}{\sqrt{2}}, y' = \frac{x - y}{\sqrt{2}}.$$

Новые координатные оси являются биссектрисами старых координатных углов. Старые координаты выражаются тогда через новые по формулам

$$x = \frac{x' + y'}{\sqrt{2}}, y = \frac{x' - y'}{\sqrt{2}}.$$

Подставим эти выражения в формулу плотности распределения и получим для новых координат

$$p(x', y') = \frac{1}{2\pi} \cdot e^{-\frac{\left(\frac{x'+y'}{\sqrt{2}}\right)^2 + \left(\frac{x'-y'}{\sqrt{2}}\right)^2}{2}}.$$

Преобразуем выражение

$$\left(\frac{x' + y'}{\sqrt{2}}\right)^2 + \left(\frac{x' - y'}{\sqrt{2}}\right)^2,$$

возведя в квадрат дроби и приведя подобные члены:

$$\left(\frac{x' + y'}{\sqrt{2}}\right)^2 + \left(\frac{x' - y'}{\sqrt{2}}\right)^2 = \frac{(x')^2 - 2x'y' + (y')^2 + (x')^2 + 2x'y' + (y')^2}{2} = (x')^2 + (y')^2.$$

Это значит, что плотность двумерного нормального распределения в новых осях выглядит точно так же, как и в старых:

$$p(x', y') = \frac{1}{2\pi} \cdot e^{-\frac{((x')^2 + (y')^2)}{2}}$$

В этом нет ничего удивительного — поскольку $x^2 + y^2 = C$ есть уравнение окружности, то наша плотность $p(x, y)$ имеет одинаковые значения на всякой окружности с центром в начале координат, т.е. обладает центральной симметрией, поэтому и выглядит одинаково при любых поворотах системы координат.

Свойство П3.1(1), стало быть, выражает тот факт, что, задав двумерное нормальное распределение какими-то осями, мы затем можем смотреть на это распределение, так сказать, наклонив голову (для нашего преобразования координат достаточно наклонить голову влево на 45 градусов, чтобы новые оси выглядели точно как старые при вер-

тикальном положении головы) и получая новые стандартно нормально распределенные проекции на новые оси.

П3.3. Свойства случайной величины, являющейся квадратом модуля двумерной стандартной нормальной случайной величины

Квадратом модуля мы здесь для краткости называли квадрат расстояния от точки на плоскости до начала координат. Обозначим двумерную стандартную нормальную случайную величину N_{sq}^2 . Обозначим случайные величины — одномерные компоненты, из которых она составлена, X_1 и X_2 . Изобразим возможный результат ее испытания точкой с координатами (X_1, X_2) (Рис. П3.3(1)). По теореме Пифагора квадрат расстояния равен сумме квадратов координат этой точки, которые могут быть получены возведением в квадрат «осевых» одномерных стандартных случайных величин. Обозначим эти квадраты $(X_1)^2$ и $(X_2)^2$.

По теореме Пифагора $N_{sq}^2 = (X_1)^2 + (X_2)^2$.

Но точно так же можно разложить в сумму квадратов и по новым осям X'_1 и X'_2 . Тогда мы получим другое справедливое равенство

$$N_{sq}^2 = (X'_1)^2 + (X'_2)^2.$$

Посмотрим теперь, что представляет собой X'_2 . Это проекция двумерной стандартной нормальной случайной величины на ось координат, т.е. (одномерная) стандартная нормальная случайная величина. Для нашей точки (X_1, X_2) проекция на ось X'_2 есть точка $(X_1 - \bar{X}, X_2 - \bar{X})$. Квадрат длины проекции соответствующего вектора мы можем вычислить по теореме Пифагора через ее координаты: он равен $(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2$.

Таким образом, квадрат *одномерной* стандартной нормальной случайной величины X'_2 можно вычислять по формуле $(X'_2)^2 = (X_1 - \bar{X})^2 + (X_2 - \bar{X})^2$ через исходные случайные величины X_1 и X_2 , и этот квадрат случайной величины имеет распределение Хи-квадрат *с одной степенью свободы*. Но в правой части стоит «заготовка» формулы для оценки дисперсии по двум наблюдениям. Вернемся к оценке дисперсии.

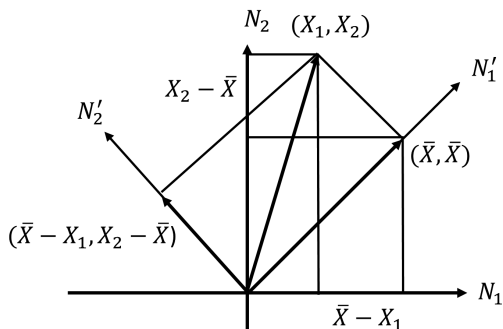


Рис. П3.3(1). Разложение сумм квадратов

П3.4. Оценка дисперсии

Предположим сначала, что математическое ожидание нормально распределенной случайной величины нам известно, а дисперсия неизвестна. Пусть дана случайная величина X с математическим ожиданием M_X . Мы хотим оценить ее дисперсию по выборке. Тогда нам подойдет естественная оценка дисперсии со знаменателем n :

$$S^2 = \frac{1}{n}((x_1 - M_X)^2 + \dots + (x_n - M_X)^2).$$

Для дискретной случайной величины доказательство того, что это — оценка дисперсии можно провести совершенно аналогично доказательству того, что среднее выборочное есть оценка математического ожидания — надо увидеть, что частоты каждого значения (в данном случае квадрата отклонения) сходятся к соответствующим вероятностям. Для непрерывной случайной величины доказательство технически усложняется, но по смыслу остается тем же.

Если же мы не знаем математическое ожидание и вынуждены использовать его оценку \bar{x} , то надо иметь в виду следующее: \bar{x} не попадает точно в центр распределения, а смещается в сторону случайного сдвига выборки, грубо говоря, вправо или влево в сторону большего числа членов выборки, случайно там оказавшихся. Это значит, что в среднем разность $(x_i - \bar{x})^2$ оказывается меньше, чем в среднем $(x_i - M_X)^2$. Замечательный факт, который мы доказали в частном случае, состоит в том, что этот сдвиг в точности соответствует уменьшению числа степеней свободы на единицу. Например, в разобранный частном случае распределение $(X_1 - M_X)^2$ в точности совпадает с распределением суммы $(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2$.

Сумма $X_1^2 + X_2^2$ разлагается в сумму $\bar{X}^2 + \bar{X}^2 + (X_1 - \bar{X})^2 + (X_2 - \bar{X})^2$.

П3.5. Разложение сумм квадратов в случае больших размерностей

Если теперь рассмотреть выборку объема n испытаний стандартной нормальной случайной величины, то надо разложить квадрат вектора, имеющего стандартное нормальное n -мерное распределение, в сумму квадрата одномерного распределения вдоль вектора $(1, 1, \dots, 1)$ (это будет вектор $(\bar{X}, \bar{X}, \dots, \bar{X})$), квадрат его длины есть $n \cdot \bar{X}^2$ — сравни с первой таблицей межгрупповых сумм в дисперсионном анализе; и квадрата $(n - 1)$ -мерного «остатка», который станет не чем иным, как случайной величиной Хи-квадрат с $n - 1$ степенью свободы, если ее разделить на это число $n - 1$. Число степеней свободы и есть размерность этого дополняющего подпространства.

П3.6. Оценка дисперсии. Последнее замечание

Если теперь мы захотим оценить неизвестную дисперсию нормально распределенной случайной величины с неизвестным математическим ожиданием, то мы должны воспользоваться формулой

$$s^2 = \frac{1}{n-1} ((x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2)$$

Если эта дисперсия равна какому-то числу D , то этому же числу будет равно в среднем наше S^2 , имеющее в этом случае распределение, пропорциональное Хи-квадрат с коэффициентом пропорциональности D .

П3.7. Другие разложения сумм квадратов

Каждое отдельное приложение представленного подхода требует особых разложений сумм квадратов и особых доказательств.

Во всех случаях подобных разложений сначала выделяют взаимно ортогональные подпространства в пространстве, размерность которого равна общему числу испытуемых, а затем производят разложение квадрата стандартного нормально распределенного многомерного вектора по этим направлениям (хотя доказательство ортогональности иногда представляет значительную трудность).

Например, разложение сумм квадратов в однофакторном дисперсионном анализе выделяет (1) одномерное подпространство, порожаемое вектором $(1, 1, \dots, 1)$, (2) $k - 1$ -мерное подпространство (где k число групп), порожаемое возможными групповыми средними¹¹⁷; (3) все остальное «остается» остатку или ошибке, и размерность соответствующего подпространства равна $n - k$. Если вся выборка взята из одной и той же генеральной совокупности, то пространство «за вычетом» одномерного подпространства средних будет иметь размерность $n - 1$, а соответствующая оставшейся сумме квадратов случайная величина имеет распределение Хи-квадрат с $n - 1$ степенями свободы. Выделенные ортогональные подпространства групповых средних и остатка определяют соответствующие распределения Хи-квадрат, частное которых и дает отношение Фишера.

В разных главах учебника мы говорили о величине эффекта и упоминали несколько способов его оценки. В данном приложении мы сведем эту картину воедино: обсудим, зачем и как оценивать статистический эффект при разных гипотезах и разных методах обработки данных.

¹¹⁷ Это сумма одномерных подпространств, заданных средними по каждой группе с условием после выделения общего среднего, что сумма этих средних равна нулю — что и делает размерность подпространства равной $k - 1$.

Приложение 4.

Величина статистического эффекта

П4.1. Зачем и как оценивать статистический эффект

В подпараграфе 6.1.2 приводился пример, в котором нас интересовало влияние кофеина на продуктивность и обсуждалось два вопроса:

1. Насколько сильно сдвигается продуктивность под действием кофеина?
2. Насколько мы можем быть уверены, что этот эффект оценен достаточно точно?

На практике первый вопрос, которым задается исследователь, еще более простой: влияет ли кофеин на эффективность вообще? К сожалению, как раз на этот вопрос статистика прямого ответа не дает, хотя предлагает косвенный ответ — в Главе 12 мы уже писали об этом. Используя Т-критерий Стьюдента и вычисляя значимость, мы отвечаем на противоположный вопрос: насколько вероятно, набрав случайную выборку, получить такое или более экстремальное значение t -статистики, если на самом деле кофеин не действует? Если вероятность маленькая, это дает основания отвергнуть нулевую гипотезу и принять альтернативную — сделать вывод, что скорее кофеин все же влияет на продуктивность.

Напоминание П4.1(1). Проверая нулевую гипотезу (или, другими словами, пользуясь методологией проверки нулевой гипотезы, NHST), важно помнить, что уровень значимости не означает вероятности гипотезы H_0 , а выбор его (например, 0.05) — результат условной договоренности (подпараграф 5.1.4).

Уровень значимости не поможет и в ответе на вопросы о силе эффекта и точности его оценки, поскольку он зависит от объема выборки. Например, для того, чтобы такая сильная линейная связь, как $r = 0.81$ достигла уровня значимости $p < 0.05$, достаточно выборки лишь в шесть человек, но при выборке в 100 испытуемых этого уровня значимости будет достигать и корреляция $r = 0.20$. Однако понятно, что корреляция 0.80 и корреляция 0.20 — это совсем разная по силе связь и для психолога она может иметь очень разное значение. Чтобы оценить силу связи, или силу сдвига продуктивности, требуется показатель, от размера выборки не зависящий.

По аналогии с тем, как ранее мы предполагали, что в генеральной совокупности существуют неизвестные нам математическое ожида-

ние и среднее квадратичное отклонение по данной переменной (например, продуктивности), и оценивали их в исследовании на основе среднего и стандартного отклонения выборки, предположим теперь, что для любого исследовательского вопроса, проверяющего нулевую гипотезу, в генеральной совокупности существует некая неизвестная нам величина эффекта, и попытаемся оценить ее с помощью статистики на основе данных выборки. Существует два ключевых подхода к пониманию величины статистического эффекта (статистической оценки эффекта). Первый подход предлагает оценивать эффект по силе связи между переменными примерно так, как силу линейной связи между двумя интервальными переменными можно оценить по величине корреляции ($r = 0.81$ в отличие от $r = 0.20$) или коэффициенту детерминации ($R^2 = 0.66$ в отличие от $R^2 = 0.04$, т.е. 66% общей дисперсии двух переменных в отличие от 4%). Теоретическую величину такого статистического эффекта обычно называют R , а способы ее оценки в разных методах обработки данных — оценками силы связи, или оценками с учетом дисперсии (или — «семья r »). Такие аналоги коэффициента корреляции предложены для самых разных методов обработки данных — и все они представляют собой разные оценки теоретического статистического эффекта как силы связи.

Пример второго подхода приведен в исследовании с кофеином в подпараграфе 6.1.2.: в нем статистический эффект понимается как стандартизованная разница между средними, или, другими словами, разница между средними, измеренная в стандартных отклонениях. Такую теоретическую величину обычно обозначают D , и для разных методов обработки данных придумано немало способов оценить теоретическую величину D по имеющейся выборке (их иногда называют «семья d »).

Замечание П4.1(2). Как самостоятельный метод, оценка величины статистического эффекта не добавляет исследователю уверенности в выборе между нулевой и альтернативной гипотезой или в точности оценки параметров (математического ожидания, среднее квадратичное отклонение) генеральной совокупности. Она лишь напоминает ему, что результат может быть значимым, но при этом весьма скромным или даже не имеющим смысла в данной области науки. Для сравнения, не добавляет этой уверенности и доверительный интервал: он лишь напоминает, что оценка может быть неточной.

Замечание П4.1(3). Оба понимания величины статистического эффекта применяются в психологии. Оценивая эффект как силу связи, удобнее интерпретировать результат по аналогии с коэффициентом корреляции. Оценка же эффекта как стандартизованной разницы сред-

них обладает другим преимуществом: одна из таких оценок (которая носит название *d* Коэна) используется в метаанализе, процедуре, позволяющей обобщить результаты разных исследований, рассчитав их общий, взвешенный эффект. Именно поэтому к настоящему моменту оценка статистического эффекта при помощи *d* Коэна стала настолько распространенной, что иногда, говоря «статистический эффект», имеют в виду именно этот критерий. Ниже мы покажем, что это очень упрощенное понимание.

Обратим также внимание читателя, что существуют формулы «перевода» одних способов оценки величины статистического эффекта в другие. Они полезны, если у исследователя нет доступа к «сырым» данным и по разным показателям в статьях он пытается провести единый метаанализ. Однако не во всех случаях это точные формулы, часто это лишь приблизительная «подгонка», экстраполяция, хотя и вполне приемлемая.

П4.2. Статистический эффект как оценка силы связи

Если нас интересует *попарная линейная связь* переменных (например, возраста детей и их успешности в запоминании некоторого материала), оценить статистический эффект легко двумя уже знакомыми нам способами (см. главу 9):

1. По коэффициенту корреляции r : чем дальше показатель от нуля, тем сильнее линейная связь.

2. По коэффициенту детерминации R^2 : в линейной регрессии он описывает долю суммы квадратов, учтенных в модели, по отношению к общей сумме квадратов ($R^2 = S_{model}/S_{total} = (S_{total} - S_{error})/S_{total}$).

Напомним, если переменных всего две, коэффициент детерминации равен квадрату коэффициента корреляции ($R^2 = r^2$) и описывает общую дисперсию двух переменных.

Во множественной линейной регрессии (например, если нас интересует связь возраста, успеваемости и поддержки со стороны родителей у детей с их успешностью запоминания) сила связи также описывается коэффициентом детерминации R^2 : чем ближе он к единице, тем больше величина статистического эффекта; чем ближе к 0 — тем меньше. В иерархической линейной регрессии и в анализе модерации (см. главу 10) изменение процента объясняемой дисперсии зависимой переменной ΔR^2 говорит об изменении «силы связи» (в смысле — объяснительной силы модели) при добавлении новых независимых переменных.

Упражнение П4.2(1). Вернитесь к примерам 10.3(13) и 10.3(20) и, опираясь на приведенные значения R^2 и ΔR^2 , сравните, каких изменений (в смысле объяснительной силы — значительных или неболь-

ших) удалось достичь добавлением новых независимых переменных. Обратите внимание, что в первом случае улучшение модели достигает принятого в психологии уровня значимости, а во втором — нет. Какие выводы из исследований следует сделать, если учитывать только уровень значимости? Достаточно ли велика разница между улучшением в 12% объясняемой дисперсии и улучшением в 9% объясняемой дисперсии, чтобы в одном случае признать ее имеющейся, а во втором — отсутствующей? Упражнение иллюстрирует, что выводы из исследования становятся более осмысленными, если помимо уровня значимости, учитывать величину эффекта.

Для дисперсионного анализа величина эффекта оценивается показателем η^2 , который при сравнении двух и более выборок параметрическими методами (t-критерий Стьюдента для несвязанных выборок, однофакторный дисперсионный анализ) совпадает с $\eta^2 = S_f / (S_f + S_{error})$.¹¹⁸

Упражнение П4.2(3). Вернитесь к примеру 7.3(1) и 7.3(2) и рассчитайте η^2 для вклада способа привлечения студентов к дополнительному курсу в экзаменационную оценку. Проверьте свои вычисления, используя SPSS. Обратите внимание, что хотя в 7.3(1) различия между группами далеки от приемлемого уровня значимости, а в 7.3(2) мы можем сделать вывод, что группы различаются (если ориентироваться на $p < 0.05$), величина статистического эффекта одинакова. Это не удивительно, потому что в 7.3(2) мы лишь удвоили выборку (что повлияло на уровень значимости), а сила связи сохранилась.

Если проводится многофакторный дисперсионный анализ, величина эффекта рассчитывается отдельно для каждого фактора и для каждого взаимодействия. В двухфакторном анализе таких показателей будет три: два фактора и взаимодействие. Формула при этом несколько изменится: вместо используются сумма по тому фактору (или взаимодействию), который нас интересует. В обозначениях параграфа 8.2 $\eta^2 = S_f / (S_f + S_{error})$. В первом случае в знаменателе мы учитываем всю дисперсию, а во втором — ту, которая объясняется одним из факторов (и которая не объясняется ни одним из других факторов). Поскольку в разных исследованиях дополнительные факторы могут быть разными, то выделение эффекта отдельного фактора становится вполне разумным, если мы сопоставляем эти исследования. В таких случаях используют дополнительную букву в обозначении — η_p^2 (от английского partial, частичная эта в квадрате).

¹¹⁸ Показатель R^2 выводится в таблице SPSS. Повторите какую-либо задачу параграфа 7.3 и убедитесь в этом.

Упражнение П4.2(4). Вернитесь к примеру 8.3(1) и рассчитайте величины статистического эффекта η^2 для вклада мотивации студента, опыта преподавателя и их взаимодействия в итоговый уровень знаний по предмету. Проверьте свои вычисления, используя SPSS (Воспользуйтесь инструкциями параграфа 8.3, затем перейдите по кнопке *Параметры (Options)* и отметьте пункт *Оценки размера эффекта (Estimates of effect size)*).

П4.3. Статистический эффект как стандартизованная разница между средними

Наиболее известный вариант расчета величины статистического эффекта как стандартизованной разницы между средними был предложен Дж. Коэном для сравнения двух несвязанных выборок и носит название *d* Коэна. Идея ее расчета проста: разность между средними двух выборок делится на их общее стандартное отклонение — в результате получается разность, измеренная в стандартных отклонениях, или стандартизованная разность. Однако идея может быть реализована по-разному. Например, в подпараграфе 6.1.2 мы вычитали из среднего времени реакции тех, кто не принимал кофеин, среднее время реакции тех, кто принимал, — и делили на стандартное отклонение контрольной группы. Такой показатель величины эффекта называется *дельтой Гласса*.

Предположим, что в каждой группе по 10 человек, при этом среднее время реакции у принимавших кофеин $\bar{x} = 4$, а стандартное отклонение $s_x = 1$, а у непринимавших (контрольная группа) $\bar{y} = 5$ — со стандартным отклонением $s_y = 1$. Тогда дельта Гласса вычисляется так:

$$\delta = \frac{\bar{x} - \bar{y}}{S_y} = \frac{5 - 4}{1} = 1$$

Соображения в пользу такого выбора приведены в 6.1.2.

Если группы равноправны (например, как мы писали, в одной группе пили кофе, а в другой принимали таблетку кофеина), то можно взять среднее стандартное отклонение $s = (1+2)/2 = 1.5$ (тогда величина эффекта будет равна 0.667) или рассчитывать среднюю дисперсию по выборкам, а затем извлекать из нее корень

$$s = \sqrt{(1^2 + 2^2)/2} = 1.58 \quad (\text{результат отличается от предыдущего}).$$

Надо иметь в виду, что интерес к величине эффекта не исчерпывается существом дела для данного конкретного эксперимента, важным аспектом этого интереса является желание унифицировать отчеты о результатах единичных исследований так, чтобы их можно было

использовать в метааналитических исследованиях. Упомянутый показатель d Коэна имеет в виду, прежде всего, эту цель.

С этой точки зрения становится понятным выбор Коэна: в общем случае неравных по объему выборок $\{x_n, y_m\}$ он предлагает такую формулу:

$$d = \frac{\bar{x} - \bar{y}}{S_{pooled}}$$

где

$$S_{pooled} = \sqrt{\frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}} = \sqrt{\frac{1}{n+m-2} \left(\sum (x_i - \bar{x})^2 + \sum (y_i - \bar{y})^2 \right)}$$

(правая часть равенства получается подстановкой вместо s_x^2 и s_y^2 их выражений).

При равенстве объемов формула приобретает понятный вид

$$S_{pooled} = \sqrt{\frac{s_x^2 + s_y^2}{2}},$$

но при неравенстве вклад каждой выборки пропорционален ее объему, что выглядит не оправданно, если иметь в виду существо дела. Однако такая формула позволяет получать статистический эффект, если из отчета об исследовании известны только объемы выборок и значение t -статистики. Действительно, формулы d Коэна и t Стьюдента отличаются только множителем, поскольку в знаменателе статистики Стьюдента стоит

$$se = \sqrt{\frac{1}{n+m-2} \left(\sum (x_i - \bar{x})^2 + \sum (y_i - \bar{y})^2 \right)} / \sqrt{\frac{nm}{n+m}},$$

а в знаменателе d Коэна

$$S_{pooled} = \sqrt{\frac{1}{n+m-2} \left(\sum (x_i - \bar{x})^2 + \sum (y_i - \bar{y})^2 \right)}.$$

Таким образом,

$$d = t \sqrt{\frac{nm}{n+m}}.$$

Всегда ли «объединенное» стандартное отклонение S_{pooled} — хорошая единица измерения? Предположим, из предыдущих масштабных исследований нам известно стандартное отклонение в скорости реакции на большой выборке испытуемых. Если мы подозреваем,

что эта оценка популяционного среднеквадратичного отклонения более точна, чем в наших двух небольших экспериментальных выборках, не разумнее ли воспользоваться ею и подставить в знаменатель в формуле?

Другой пример: что, если прием кофеина сказывается не только на средней скорости реакции, но и на ее дисперсии (а значит, и стандартном отклонении)? Например, предположим, что время реакции после кофеина становится более разнообразным — чаще попадают как очень быстрые ответы, так и медленные. «Объединяя» стандартные отклонения контрольной и экспериментальной групп так, как предложил Коэн, мы не приближаемся к оценке среднеквадратичного отклонения в скорости реакции в популяции, как нам того бы хотелось, а отдаляемся от нее! Происходит это потому, что стандартное отклонение в контрольной группе действительно описывает различия между людьми во времени реакции, а стандартное отклонение в экспериментальной группе принимавших кофеин — уже нет, поскольку само искажено действием кофеина. В таких случаях разумнее использовать формулу Дж. Гласса (см. выше). Обратите внимание, что и интерпретация этого эффекта будет другой: принимающие кофеин реагируют в среднем быстрее, чем не принимающие, на одно стандартное отклонение во времени реакции контрольной группы. Тот факт, что при приеме кофеина растет дисперсия времени реакции, представляет, заметим, отдельный интерес.

Замечание П4.3(1). Мы видим, что для решения одной и той же задачи — оценки статистического эффекта как стандартизованной разности средних при сравнении двух несвязанных выборок — исследователи могут использовать разные меры, если обоснуют, почему в их случае та или иная оценка среднеквадратичного отклонения в популяции более разумна. Однако, это меняет и сам результат, и его интерпретацию — и мы советуем читателю внимательно проверять, какую именно меру оценки величины эффекта использовали авторы того или иного исследования.

Замечание П4.3(2). Поскольку в «объединенном» стандартном отклонении стандартные отклонения каждой из групп учтены пропорционально количеству испытуемых в каждой группе, интерпретировать результат особенно трудно, если сильно отличаются не только стандартные отклонения в сравниваемых группах, но и их размеры. «Перекося» получается в пользу той группы, которая больше по размеру. Поэтому исследователи рекомендуют при разных дисперсиях (и стандартных отклонениях) в группах стараться, чтобы в большей по размеру группе было не более 60% всех испытуемых (а в меньшей, следовательно, не менее 40%).

Для сравнения двух связанных выборок расчеты, казалось бы, можно упростить: если для каждого человека рассчитать изменение его показателей. Так, в примере 6.3.1 это изменение в скорости реакции до и после приема алкоголя у каждого испытуемого.

Вариант 1. Легко рассчитать среднее и стандартное отклонение переменной «изменение скорости реакции». Разделив одно на другое, получим некоторую величину — можно ли ее считать оценкой величины эффекта?

Вариант 2. В подпараграфе 6.1.2 в качестве величины эффекта мы брали другой показатель: отношение средней разности к стандартному отклонению стартового замера и обосновывали этот выбор. Теперь мы обсудим новые детали вопроса.

В случае связанных выборок показатели первого и второго замеров, как правило, коррелируют между собой. Если «сдвиг» в скорости реакции после приема алкоголя у всех похожий (например, скорость реакции убывает), самые быстрые люди будут реагировать довольно быстро и после приема алкоголя, а самые медленные по-прежнему медленнее других. Т.е. корреляция между переменными «скорость реакции до алкоголя» и «скорость реакции после» будет положительной.

Как мы отмечали в подпараграфе 6.1.1, связанные выборки могут породиться не только в результате повторных измерений. Следующий фантастический пример показывает весь спектр возможностей для связанных выборок.

Пример П4.3(3). Предположим, что исследуется различие действия таблетки кофеина против плацебо на субъективный уровень бодрствования. Авторы выбрали следующий дизайн: взяли несколько пар однояйцевых близнецов, воспитывавшихся в разных семьях, и отнесли одного из каждой пары в контрольную, а другого в экспериментальную группу. Авторы хотят, так сказать, вычлест эффект плацебо из эффекта кофеина. Такой дизайн имеет тем больше смысла, чем сильнее плацебо-внушаемость определяется наследственностью — и тем сильнее будет тогда корреляция между близнецовыми замерами. Если же плацебо-внушаемость не имеет наследственной компоненты, то корреляция будет равна нулю.

Если бы корреляция была близка к единице — это означало бы¹¹⁹, что плацебо-эффект и действие кофеина практически суммируются. Тогда дисперсия разности была бы близка к нулю и вариант 1 расче-

¹¹⁹ Точно — лишь при равенстве дисперсий обеих выборок.

та величины эффекта (см. выше) дал бы очень большое число. Ясно, что в таком виде он нам не подходит.

Если же корреляция равна нулю, то мы фактически имеем дело с независимыми выборками, и вариант 1 даст число, похожее на d Коэна.

Коэн предложил единую формулу, которая дает разумный результат при всех промежуточных ситуациях:

$$S_{pooled} = \frac{S_{pgain}}{\sqrt{2(1 - r_{xy})}}$$

где — коэффициент корреляции, а — стандартное отклонение разности парных замеров.

В случае нулевой корреляции (при условии приблизительного равенства дисперсий замеров) S_{pooled} становится почти очевидно равно дисперсии экспериментальной (= контрольной) выборки, при промежуточных значениях дает приемлемые приближения к этим значениям. Для доказательства этого надо использовать равенство

$$s_{x-y}^2 = s_x^2 + s_y^2 - 2 \cdot r_{xy} \cdot s_x \cdot s_y,$$

которое верно как для случайных величин и их теоретических среднеквадратических отклонений и корреляции, так и для выборочных оценок.

Замечание П4.3(4). При сравнении двух выборок, зная t или F , всегда можно рассчитать оба показателя — и $\eta^2 = F/(F + df_{error}) = t^2/(t^2 + df_{error})$, и Коэна $d = t/\sqrt{n}$, воспользовавшись эквивалентностью Т-критерия и дисперсионного анализа при сравнении двух групп: $F = t^2$ (параграф 7.2) (число степеней свободы модели на единицу меньше числа уровней независимой переменной, т.е. $df = 2 - 1 = 1$).

П4.4. Величина эффекта в случаях непараметрических статистик

Для оценки силы эффекта для коэффициента корреляции Спирмена можно использовать сам коэффициент или его квадрат.

При непараметрических сравнениях групп, например, для статистик Вилкоксона, Манна-Уитни или Краскала-Уоллиса — распределения этих статистик аппроксимируются нормальным распределением (Z -распределением). Эту Z -оценку, аналог полученного результата в нормальном распределении, можно использовать для расчета статистического эффекта как силы связи: аналог коэффициента корреляции можно получить, если Z -оценку разделить на корень из объема выборки, а аналог коэффициента детерминации, или η^2 , — если возвести полученный результат в квадрат.

Аналогичные показатели используются для таблиц сопряженности.

В расширенном варианте этого приложения, помещенного там же, где файлы данных для практикума, можно познакомиться с дальнейшими деталями обсуждаемого предмета.

Приложение 5. _____

Гипотезы и их вероятности. Формула полной вероятности. Формула Байеса

Среди альтернатив, предлагаемых критиками подхода NHST (статистической проверки нулевой гипотезы) имеется так называемый байесовский подход. Он опирается на формулу Байеса, которую мы далее выведем и обсудим. Мы не являемся сторонниками этой альтернативы, однако считаем, что общее представление об этом подходе может быть полезно читателю.

П5.1. Условная вероятность. Формула полной вероятности

Пример 5.1(1). Предположим, что у нас имеется две одинаковые корзины. В одной содержится 8 белых и 2 черных шара, а в другой — 1 белый и 5 черных шаров. Процедура выбора шара проводится в два этапа. Сначала случайно выбирается одна из двух одинаковых корзин, затем из выбранной корзины наугад вынимается шар. Нас интересуют два вопроса:

1) Какова вероятность того, что вынутый по такой процедуре шар будет белым?

2) Если первый шар оказался белым, то какова вероятность того, что вытянутый наугад из той же корзины следующий шар будет белым?

Обсудим сначала второй вопрос. Интуитивно понятно, что результат «вытянут белый шар» может быть интерпретирован как аргумент в пользу того, что на первом этапе процедуры скорее всего была выбрана корзина номер 1, с большим количеством белых шаров. Эта интуиция может быть переведена в точные значения вероятностей с помощью так называемой формулы Байеса. После расчета соответствующих вероятностей ответ на второй вопрос будет достаточно легко дать.

Обозначим символом W событие «Достать белый шар из корзины, выбранной наугад из двух описанных корзин». Оно может наступить только вместе с одним из попарно несовместимых событий H_1 и H_2 и (будем называть их *гипотезами*: «Выбрать первую корзину» и «Выбрать вторую корзину»).

Для доказательства в общем случае мы будем использовать так называемую геометрическую интерпретацию вероятностей (рис. П5.1(2)). Внешний овал Ω изображает пространство всех возможных исходов. В испытании может появиться любой из них (одна точка внутри овала), причем все с равными вероятностями. Тогда вероятность наступления

сложного события W равна отношению «количества» точек внутри W к «количеству» точек во внешнем овале Ω , точнее говоря, отношению площадей W и Ω . Если положить площадь Ω равной единице, то вероятности всех событий, изображенных фигурами на рисунке, будет равна их площадям.

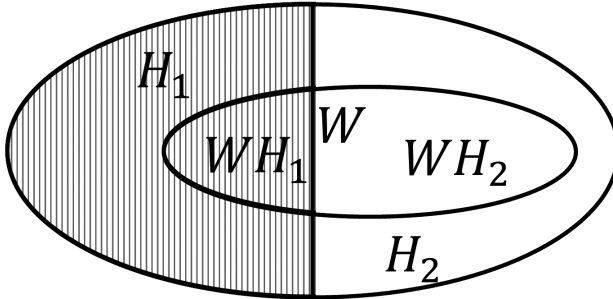


Рис. П5.1(2). Формула полной вероятности. Разложение события. H_1 представлена всей заштрихованной левой половиной большого овала, H_2 — не заштрихованной правой половиной большого овала, W представлен внутренним овалом, левая половина которого есть WH_1 , а правая WH_2 .

Можно разложить событие W в сумму двух несовместимых событий-произведений WH_1 и WH_2 , которые на рисунке изображаются долями внутреннего овала W .

Вероятность наступления W будет выражаться в таком случае формулой $P(W) = P(WH_1) + P(WH_2)$.

Если обозначить $P(W|H_1)$ *условную вероятность* вытянуть белый шар, если на первом шаге была выбрана первая корзина, то вероятность можно представить в виде произведения $P(W|H_1) \cdot P(H_1)$. Действительно, $P(W|H_1)$ это доля площади левой, заштрихованной стороны внутреннего овала на рис. П5(1), по отношению к площади внешнего овала, представляющего собой полную систему событий Ω . Отношение площадей $P(WH_1)/P(\Omega)$ можно преобразовать в вид,

$$P(WH_1)/P(\Omega) = (P(WH_1)/P(H_1)) \cdot (P(H_1)/P(\Omega)).$$

Первая дробь представляет собой $P(W|H_1)$ — долю площади WH_1 в H_1 , а вторая по аналогичным соображениям $P(H_1)$ (долю площади H_1 в Ω).

Таким образом, $P(WH_1) = P(W|H_1)P(H_1)$.

Произведя теперь подстановку, получаем формулу

$$P(W) = P(WH_1) + P(WH_2) = P(W|H_1)P(H_1) + P(W|H_2)P(H_2),$$

которая и представляет собой *формулу полной вероятности* для слу-

чая двух слагаемых. Если гипотез больше, то и сумма становится длиннее. В общем случае она такова:

$$P(W) = P(W|H_1)P(H_1) + P(W|H_2)P(H_2) + \dots + P(W|H_n)P(H_n).$$

Первая задача нашего примера решается так: с вероятностью $1/2$ мы выбираем первую корзину, из которой белый шар достается с вероятностью $8/10$; с вероятностью $1/2$ вторую корзину, из которой белый шар достается с вероятностью $1/6$. Общая вероятность вытянуть белый шар равна

$$\frac{8}{10} \cdot \frac{1}{2} + \frac{1}{6} \cdot \frac{1}{2} = \frac{29}{60}.$$

Теперь перейдем ко второму вопросу: если первый шар оказался белым, то какова вероятность того, что вытянутый наугад из той же корзины следующий шар будет белым?

П5.2. Формула Байеса

Для того чтобы на него ответить, надо рассчитать вероятности того, что у нас в руках первая или вторая корзина, при условии, что мы достали из этой имеющейся в руках корзины белый шар, а затем применить формулу полной вероятности для новых вероятностей гипотез.

Заметим, что $P(WH_1)$ может быть вычислена двумя симметричными способами $P(WH_1) = P(H_1|W)P(W)$ и $P(WH_1) = P(W|H_1)P(H_1)$. Если мы приравняем правые части обоих равенств, то получим $P(H_1|W)P(W) = P(W|H_1)P(H_1)$ откуда

$$P(H_1|W) = \frac{P(W|H_1)P(H_1)}{P(W)}.$$

Эта формула и называется *формулой Байеса*. Мы теперь можем ответить на вопрос о вероятности вынуть белый шар повторно. Учитывая, что ранее мы нашли $P(W) = \frac{29}{60}$, по формуле Байеса рассчитываем: после обнаружения, что вынутый первый шар — белый, вероятность того, что у нас в руках первая корзина равна

$$P(H_1|W) = \frac{P(W|H_1)P(H_1)}{P(W)} = \frac{\frac{8}{10} \cdot \frac{1}{2}}{\frac{29}{60}} = \frac{8}{10} \cdot \frac{1}{2} \cdot \frac{60}{29} = \frac{480}{580} = \frac{24}{29}.$$

Не приходится сомневаться, что вероятность того, что в руках у нас вторая корзина, равна $1 - \frac{24}{29} = \frac{5}{29}$.

Теперь вероятность достать белый шар из первой корзины равна $7/9$, поскольку один белый шар из корзины уже достали. Из второй корзины достать белый шар уже невозможно. Подставляя в формулу полной вероятности новые вероятности, получаем

$$P(W_2) = \frac{7}{9} \cdot \frac{24}{29} + 0 \cdot \frac{5}{29} = \frac{56}{87}.$$

П5.3. Геометрическая интерпретация формулы полной вероятности и формулы Байеса

Формула Байеса имеет другую форму, которая делает понятным механизм вычисления условной вероятности гипотезы. Мы опять будем разбирать ее на примере с двумя гипотезами. Если учесть, что

$$P(W) = P(W|H_1)P(H_1) + P(W|H_2)P(H_2),$$

и заменить $P(W)$ в знаменателе формулы Байеса на правую часть этой формулы (вспомним, что по ней мы и рассчитывали $P(W)$), то мы получим второй общеупотребительный вариант формулы Байеса

$$P(H_1|W) = \frac{P(W|H_1)P(H_1)}{P(W|H_1)P(H_1) + P(W|H_2)P(H_2)}.$$

Заменим теперь обратно произведение $P(W|H_1)P(H_1)$ на $P(WH_1)$, а произведение $P(W|H_2)P(H_2)$ на $P(WH_2)$. Формула тогда примет вид

$$P(H_1|W) = \frac{P(WH_1)}{P(WH_1) + P(WH_2)}.$$

Эта формула легко интерпретируется с помощью рис. П5.1(2). Если произошло событие W , то это значит, что случилось одно из элементарных событий, изображаемых внутренним овалом. Вероятность того, что происшедшее элементарное событие представляло правую, большую часть внутреннего овала, равна доле площади этой части в площади всего внутреннего овала, что и выражается последней нашей формулой.

В заключение параграфа дадим вторую форму формулы Байеса в общем случае:

$$P(H_i|W) = \frac{P(W|H_i)P(H_i)}{P(W|H_1)P(H_1) + P(W|H_2)P(H_2) + \dots + P(W|H_n)P(H_n)}.$$

Ее интерпретация совершенно аналогична приведенной выше интерпретации частного случая. Вся дробь равна доле слагаемого, помещенного в числитель, во всей сумме, стоящей в знаменателе.

Во многих практических случаях формула Байеса работает совершенно адекватно. Полученные вероятности могут проверяться статистически и выдерживают такую проверку. Т.е. если мы будем проводить многократно испытания по выниманию шаров из двух корзин, как это задано в примере, и фиксировать из них только случаи, когда первый вынутый шар был белым, и смотреть долю случаев, когда этот исход наблюдался в случае выбора первой корзины, то частота будет приближаться к вычисленной по формуле Байеса вероятности $P(H_1|W)$.

П5.4. Обобщение байесовского подхода

Пока речь у нас шла о «гипотезах», довольно скромных и вполне определенных. Их специфика состоит в том, что они выразимы на языке элементарных событий, т.е. сами являются законными событиями.

Однако байесовская схема кажется некоторым исследователям приложимой и в контексте научных исследований, поскольку, как они надеются, с ее помощью можно пересчитывать вероятности теоретических гипотез после единичного наблюдения. Действие экспериментальной выборки на вероятности гипотез можно свести к последовательному пересчету по каждому испытываемому, т.е., сначала пересчитать вероятности по результатам первого испытываемого, потом полученные вероятности гипотез пересчитать по результатам второго испытываемого и т.д. до исчерпания всего набора испытываемых. Можно доказать, что итог не будет зависеть от порядка испытываемых, если только все их результаты будут использованы.

В научных исследованиях мы ведь могли бы и ограничиться вероятностями гипотез, вместо принятия той или иной гипотезы в качестве окончательно истинной. Это могло бы звучать, например, так: «в настоящее время мы считаем, что пара теорий “динамика Ньютона” и “общая теория относительности Эйнштейна” делят единичную вероятность в соотношении $\frac{1}{100}$ против $\frac{99}{100}$ ». В зависимости от результатов назначенного на завтра эксперимента на синхрофазотроне в Дубне, эти вероятности будут пересчитаны».

Примем, однако, во внимание, следующую вещь. В примере 5.1(1) мы рассчитали вероятность того, что у нас в руках первая корзина, при условии, что извлеченный из этой корзины шар был белым. В условиях задачи было указано, что вероятности взять в руки первую и вторую корзины исходно равны $1/2$. Расчет проводился по формуле

$$\frac{\frac{8}{10} \cdot \frac{1}{2}}{\frac{8}{10} \cdot \frac{1}{2} + \frac{1}{6} \cdot \frac{1}{2}} = \frac{24}{29}$$

Если бы вероятности были другие, например P («взять в руки первую корзину») = $1/3$, P («взять в руки вторую корзину») = $2/3$, то итоговая вероятность (при условии извлечения белого шара из взятой в руки корзины) также была бы другой, а именно

$$\frac{\frac{8}{10} \cdot \frac{1}{3}}{\frac{8}{10} \cdot \frac{1}{3} + \frac{1}{6} \cdot \frac{2}{3}} = \frac{24}{34}$$

Разные априорные (до-опытные) вероятности приводят к разным результатам пересчета по Байесу.

Вопрос, стало быть, состоит в том, что еще до всякого исследования мы должны иметь некоторые априорные вероятности наших гипотез, чтобы с помощью эмпирических проверок получать какие-то новые вероятности гипотез.

Мы можем сформулировать теперь принципиальное утверждение: **Если перед исследованием не сделано никаких предположений об априорных вероятностях противопоставленных друг другу гипотез, то и после получения любых эмпирических результатов невозможно делать выводы об апостериорных вероятностях этих гипотез.** Неразрешимая проблема байесовского подхода лежит здесь: эти априорные вероятности нам взять неоткуда. Например, какова априорная вероятность того, что есть телепатия? Этот вопрос надо решить прежде, чем мы вознамеримся использовать байесовский подход для ориентированного на статистику доказательства ее существования. Впрочем, наше возражение касается не только байесовских выводов, но и любых других эмпирических проверок таких теорий.

П5.5. Методологическое замечание

Гипотезы, о которых шла речь в примере, как мы говорили, существенно отличаются от теоретических гипотез. Мы вполне можем организовать ситуацию так, что вероятность взять в руки первую из предложенных корзин реально будет равна $1/2$. И мы можем повторять экспериментирование с этими равновероятно оказывающимися в руках корзинами и убедиться, что к пересчитанным по формуле Байеса вероятностям сходятся эмпирические частоты. Но вероятности теоретических гипотез принадлежат другому уровню. Даже проделывая осмысленную работу по байесовскому уточнению каким-то разумным образом заданных априорных вероятностей теоретических гипотез, мы *никогда* не сможем сконструировать метод проверки, в котором какие-то эмпирические частоты сходились бы к этим вероятностям. Такие вероятности обычно называют *субъективными*. При их использовании надо иметь в виду существенное отличие от вероятностей, о которых говорит математическая теория вероятностей. Последняя говорит о вероятностях возможных событий, а первая о вероятностях истинности утверждений о возможных событиях, что является существенно более сложным предметом, как только мы выходим за пределы утверждений о совокупностях элементарных событий. Таким образом, байесовский подход не способен полноценно заменить подход NHST.

Приложение 6.

Комбинаторика. Бином Ньютона

П6.1. Размещения

При расчете вероятностей по «классическому» определению мы используем отношение количества благоприятных исходов к общему количеству возможных исходов. Здесь мы научимся считать эти количества исходов, как бы велики они ни были. Для того чтобы считать биномиальные (см. подпараграф 2.1.4) вероятности, мы должны освоить комбинаторные понятия и соответствующие им формулы.

Наш первый модельный пример, с которым будут соотноситься другие примеры, таков:

Пример П6.1(1). Сколькими способами можно разложить k пронумерованных шаров в n пронумерованных корзин ($n \geq k$) так, чтобы в каждой корзине оказалось не больше одного шара?

Решим задачу сначала для $n = 4, k = 2$. Всего возможно 12 различных размещений (рис. П6.1(2), справа). Следующее рассуждение показывает, почему вариантов именно 12.

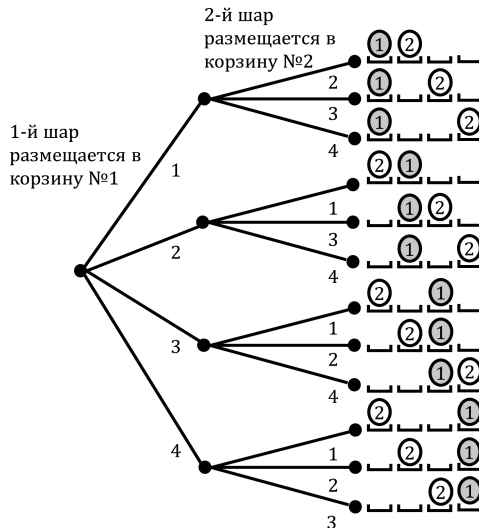


Рис. П6.1(2). Размещения двух различных шаров по четырем корзинам

Первый шар мы можем положить в любую из четырех имеющихся корзин, после чего второй шар может быть размещен в любой из оставшихся трех корзин. Можно представить выбор в виде дерева, каждая ветка которого оканчивается одним из вариантов размещения (рис. Пб.1(2), слева).

Это рассуждение легко распространяется на случай произвольных n и k :

первый шар может быть положен в любую из n корзин,

второй шар — в любую из оставшихся $n - 1$ корзин,

третий шар — в любую из оставшихся $n - 2$ корзин,

...

k -й шар в любую из оставшихся $n - (k - 1)$ корзин.

Всего получается $n \cdot (n - 1) \cdot (n - 2) \cdot \dots \cdot (n - (k - 1))$ размещений.

Количество **размещений** k элементов в n ячеек обозначается A_n^k .

Мы вывели формулу числа размещений:

$$A_n^k = n \cdot (n - 1) \cdot (n - 2) \cdot \dots \cdot (n - (k + 1)) = n \cdot (n - 1) \cdot (n - 2) \cdot \dots \cdot (n - k + 1)$$

Если $n = k$, то последняя строка будет оканчиваться словами «в любую из одной корзины». Этот последний частный случай размещений называется **перестановками** совокупности n элементов. Количество перестановок n элементов равно в точности $n!$. Напомним, что $n! = 1 \cdot 2 \cdot \dots \cdot n$ — (читается «эн факториал») есть произведение натуральных чисел от 1 до n .

Пб.2. Сочетания

Следующий вопрос, который вплотную подводит нас к формуле бинома Ньютона, звучит так: сколькими способами можно выбрать из n различных предметов k штук?

Пример Пб.2(1). Сколькими способами можно выбрать из четырех пронумерованных корзин две?

Мы свяжем этот пример с предыдущим следующим образом: выбранные корзины будем отмечать тем, что положим в них шары. Тот же рис. Пб.1(2) дает первый шаг решения — в первой строке обозначен выбор первой и второй корзин, во второй строке — первой и третьей и т.д.

Однако можно заметить, что каждый выбор пары корзин встречается в списке из двенадцати размещений дважды. Первый выбор можно найти также в четвертой строке, второй выбор — в шестой и т.д. В случае размещений нам существенно, какой шар оказался в данной корзине, в случае сочетаний — не существенно.

Это значит, что каждый случай выбора пары корзин считается столько раз, сколькими способами можно поменять местами шары в уже отмеченных корзинах. В данном случае таких перемен мест (перестановок) всего две.

Итак, выбрать две корзины из четырех можно шестью способами.

Пример П6.2(2). Сколькими способами можно выбрать из четырех пронумерованных корзин три?

Решим сначала задачу о размещении. По нашей формуле имеется 24 способа размещения трех шаров в четырех корзинах: каждое из приведенных на рис. П6.1(2) размещений двух шаров может быть дополнено двумя разными размещениями третьего шара в любую из оставшихся пустыми корзин.

Рассмотрим теперь, сколько раз мы считаем один и тот же выбор корзин, различая размещения разных шаров по одним и тем же корзинам (рис. П6.2(3)).

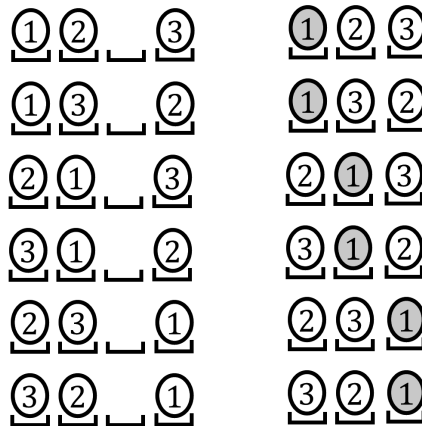


Рис. П6.2(3). Расчет количества перестановок для формулы числа сочетаний

Рассмотрим, например, размещение, изображенное на рис. П6.2(3) в первой строке слева. Под ним перечислены все возможные размещения в тех же самых корзинах, но с другим порядком шаров. В правой колонке приведены соответствующие перестановки из трех элементов. Понятно, что эти же самые перестановки будут соответствовать размещениям трех шаров в любых фиксированных трех корзинах. Это значит, что каждый выбор трех корзин из четырех считается шесть раз в формуле числа размещений. Таким образом, три корзины из четырех можно выбрать $\frac{24}{6} = 4$ способами.

Замечание П6.2(4). Отметим интересное равенство: точно так же четверем равно количество возможных выборов из четырех корзин по одной. Это очень важное совпадение объясняется тем, что каждый вы-

бор трех корзин оставляет одну корзину «невыбранной», и, наоборот, каждому выбору одной корзины можно поставить в соответствие выбор дополнительных к ней трех. Следовательно, сколькими способами можно выбрать одну корзину из четырех, столькими способами можно выбрать и три из четырех.

Теперь мы можем решить вопрос о выборе из совокупности в общем случае. Для того чтобы посчитать, сколькими способами можно выбрать k корзин из различимых n , надо сначала вычислить количество размещений k различимых шаров в n корзинах и полученное число поделить на количество перестановок различимых k шаров, или, что то же самое, на количество размещений k шаров в k корзинах. Результат, который называется *числом сочетаний из n элементов по k* и обозначается C_n^k , запишем сначала в следующем виде:

$$C_n^k = \frac{n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-(k+1))}{k \cdot (k-1) \cdot (k-2) \cdot \dots \cdot 2 \cdot 1}$$

Эту формулу проще запомнить, если домножить и числитель, и знаменатель на $(n-k)!$, тогда в числителе окажется $n!$ и формула примет вид

$$C_n^k = \frac{n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-(k+1)) \times (n-k) \cdot (n-k-1) \cdot \dots \cdot 2 \cdot 1}{k \cdot (k-1) \cdot (k-2) \cdot \dots \cdot 2 \cdot 1 \times (n-k) \cdot (n-k-1) \cdot \dots \cdot 2 \cdot 1} = \frac{n!}{k! \cdot (n-k)!}$$

Замечание П6.2(5). Из последнего варианта формулы сразу понятно, что $C_n^k = C_n^{n-k}$, поскольку если $a = n - k$, то $n - a = k$ и знаменатели в обеих формулах совпадут. На уровне нашей подразумеваемой интерпретации в терминах выбора корзин этому равенству соответствует следующий аргумент, повторяющий приведенный в предыдущем замечании: если мы выбрали k корзин из n , то тем самым мы выбрали и дополнительные, оставшиеся $n - k$ корзин. Всякому выбору k корзин соответствует единственный выбор $n - k$ корзин.

П6.3. Бином Ньютона

Рассмотрим выражение $(p + q)^n$ (оно называется биномом Ньютона на n -й степени). Что получится, если раскрыть скобки? Начнем с выражения $(p + q)^4 = (p + q)(p + q)(p + q)(p + q)$. После раскрытия скобок получатся 16 слагаемых, которые мы не будем сейчас выписывать, потому что нам гораздо важнее сформулировать правило перечисления этих слагаемых. Заметим, что каждое слагаемое — произведение четырех сомножителей — получается так: из каждой из четырех скобок выбирается p или q , затем они перемножаются.

Для того чтобы не потерять ни одного слагаемого, будем использовать так называемое двоичное дерево (рис. П6.3(1)). Каждая ветка

дерева заканчивается произведением p и q , которое получается по следующему правилу: если на первом шаге от корня дерева выбирается ветка p , то p становится первым членом произведения, если q , то первым членом произведения становится q . На втором шаге аналогично, в зависимости от выбора ветки, добавляется второй множитель p или q . Так же осуществляется выбор на третьем и четвертом шагах.

Легко видеть, что общее количество конечных вершин $16 = 2^4$ — два в степени, равной степени бинома. Наш следующий вопрос, сколько слагаемых из этих шестнадцати после перегруппировки множителей окажутся равными p^2q^2 . Рис. П6.3(1) показывает, что мы можем дать ответ на этот и подобные вопросы без непосредственного счета — достаточно установить соответствие между каждым из произведений, содержащих p и q во второй степени, и элементами нашей «корзинной модели»: каждому такому произведению соответствует выбор каких-то двух корзин из четырех. На рисунке справа, рядом с произведениями, содержащими квадраты p и q , изображены соответствующие этим произведениям выборы корзинок. Сколько способов выбора пары корзинок из четырех, столько будет членов, содержащих квадраты p и q , в биноме после раскрытия всех скобок.

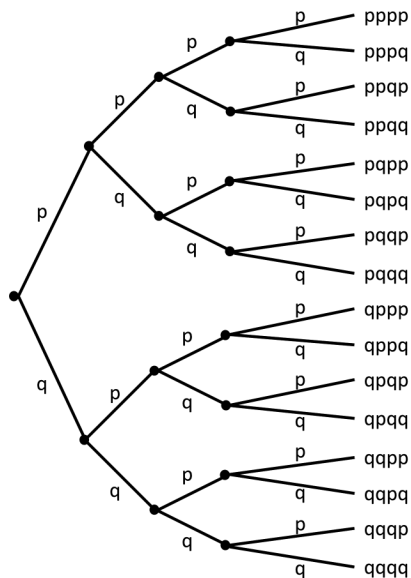


Рис. П6.3(1). Дерево слагаемых бинома Ньютона четвертой степени

Совершенно аналогично ведется рассуждение в общем случае: если нам надлежит раскрыть скобки в выражении $(p + q)^n$, то слагаемых, содержащих $p^k q^{n-k}$ будет столько, сколько существует различных выборов k (или, что то же самое, $n - k$) корзин из n возможных.

Таким образом,

$$(p + q)^n = C_n^0 p^n q^0 + C_n^1 p^{n-1} q^1 + C_n^2 p^{n-2} q^2 + \dots + C_n^{n-1} p^1 q^{n-1} + C_n^n p^0 q^n$$

В этой формуле присутствует член C_n^0 , который словами может быть описан по аналогии со своими соседями: сколькими способами можно выбрать 0 корзин из n . При всей сомнительности такой формулировки, из алгебраической формулы понятно, что $C_n^0 = 1$.

По крайней мере, нет никаких соображений, по которым выбрать из n корзин 0 можно каким-то другим, отличным от единицы количеством способов. В подобных ситуациях математики принимают соглашение¹²⁰ считать C_n^0 равным единице, а заодно равным единице и $0!$, который входит в формулу

$$C_n^0 = \frac{n!}{n! \cdot 0!}.$$

Коэффициенты C_n^k входящие в формулу бинома Ньютона, называются *биномиальными коэффициентами*.

¹²⁰ Это соглашение никак нельзя считать произвольным, поскольку дальнейшее использование показывает, что соглашение дает реальные удобства, причем не только в той области, в которой первоначально были введены такие конвенциональные определения, а противоречий или других каких-то неприятностей не обнаруживается. Читатель может продумать этот комментарий на примере хорошо ему известной «конвенции» считать произведение отрицательных чисел положительным числом. Такой конвенции невозможно придумать приемлемую альтернативу.

Список литературы

- Бююль А., Цёфель П.* SPSS. Искусство обработки информации. — Москва, Diasoft, 2005.
- Гусев А. Н.* Дисперсионный анализ в экспериментальной психологии. — М.: УМК «Психология». — 2000.
- Джеймс Г., Уиттон Д., Хастис Т., Тибширани Р.* Введение в статистическое обучение с примерами на языке R. — М.: ДМК Пресс. — 2016.
- Корнеев А. А., Кричевец А. Н.* Условия применимости критериев Стьюдента и Манна-Уитни //Психологический журнал. — 2011. — Т. 32. — №. 1. — С. 97–110.
- Корнилова Т.В.* Экспериментальная психология: Теория и методы: Учебник для вузов. — М.: Аспект Пресс, 2005.
- Кричевец А. Н., Шикин Е. В., Дьячков А. Г.* Математика для психологов. — Флинта, 2005.
- Наследов А. Д.* Математические методы психологического исследования. — С–Пб: Речь. 2004.
- Наследов Д.А.* IBM SPSS Statistics 20 и AMOS: профессиональный статистический анализ данных. — Издательский дом «Питер», 2012.
- Рубцова Н.Е., Ленъков С.Л.* Статистические методы в психологии: учеб. пособие: для студентов вузов, обучающихся по направлению и специальностям психологии. — М.: УМК «Психология», 2005.
- Сидоренко Е.В.* Методы математической обработки в психологии. — С.–Пб.: Речь, 2000
- Уэссман А., Рикс Д.* Психодиагностическое познание профессиональной деятельности. — С.–Пб., 2004. С.94–96

- Фетискин Н.П., Козлов В.В., Мануйлов Г.М.* Социально-психологическая диагностика развития личности и малых групп. — М.: Изд-во Института Психотерапии. — 2002. — С.57–59.
- Шеффе Г.* Дисперсионный анализ. — М.: Наука, 1980.
- Шурунов А. Б., Балдин Е. М., Волкова П. А., Коробейников А. И., Назарова С. А., Петров С. В., Суфиянов В. Г.* Наглядная статистика. Использование R. — М.: ДМК Пресс. — 2012.
- Brace N., Snelgar R., Kemp R.* SPSS for Psychologists. — Palgrave Macmillan, 2012.
- Field A.* Discovering statistics using IBM SPSS statistics. — Sage, 2013.
- Howitt D. Cramer D.* Introduction to Statistics in Psychology. — L.: Financial Times, 2008
- Open Science Collaboration et al.* Estimating the reproducibility of psychological science // Science. — 2015. — Т. 349. — №. 6251. — С. aac4716.
- Spector P.* Data Manipulation with R. — Springer, 2008.
- Wasserstein R. L. et al.* The ASA’s statement on p-values: context, process, and purpose //The American Statistician. — 2016. — Т. 70. — №. 2. — С. 129–133.

Кричевец А.Н., Корнеев А.А., Рассказова Е.И.
Основы статистики для психологов.

Верстка А.Троицкая

ООО “Акрополь”
г. Москва, ул. Краснобогатырская, 42, стр. 2

Подписано в печать 04.10.2019
Формат 60×84/16. Бумага офсетная.
Усл. печ. л. 16,62. Тираж 500. Заказ № 188.



Анатолий Николаевич Кричевец

кандидат физико-математических наук, доктор философских наук, профессор факультета психологии Московского государственного университета имени М.В. Ломоносова



Корнеев Алексей Андреевич

кандидат психологических наук, старший научный сотрудник лаборатории нейропсихологии факультета психологии МГУ имени М.В. Ломоносова.



Елена Игоревна Рассказова

кандидат психологических наук, доцент кафедры нейро- и патопсихологии факультета психологии МГУ имени М.В. Ломоносова.