

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
ТЮМЕНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИНСТИТУТ ПСИХОЛОГИИ И ПЕДАГОГИКИ

П. Е. ГРИГОРЬЕВ, И. В. ВАСИЛЬЕВА

СТАТИСТИЧЕСКИЕ МЕТОДЫ В ПСИХОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ

Учебное пособие

Тюмень
Издательство
Тюменского государственного университета
2018

УДК 159.9:51:681.3(075.8)

ББК Ю9В641я73

Г834

Авторы:

П. Е. Григорьев — доктор биологических наук, доцент, профессор кафедры общей и социальной психологии Института психологии и педагогики Тюменского государственного университета, заведующий кафедрой медицинской физики и информатики Крымского федерального университета им. В. И. Вернадского

И. В. Васильева — кандидат психологических наук, доцент, доцент кафедры общей и социальной психологии Института психологии и педагогики Тюменского государственного университета

Рецензенты:

О. И. Дубровина — кандидат психологических наук, доцент кафедры общей и социальной психологии Института психологии и педагогики Тюменского государственного университета

И. Г. Тимошук — кандидат психологических наук, доцент кафедры психологии Крымского инженерно-педагогического университета

Григорьев, П. Е.

Г834

Статистические методы в психологических исследованиях: учебное пособие / П. Е. Григорьев, И. В. Васильева ; Министерство науки и высшего образования Российской Федерации, Тюменский государственный университет, Институт психологии и педагогики. — Тюмень : Издательство Тюменского государственного университета, 2018. — 216 с.

ISBN 978-5-400-01480-2

В учебном пособии сделан акцент на тех аспектах статистической обработки данных исследования, которые недостаточно освещены в наиболее популярных учебниках: содержательная и статистическая значимость, расчет объемов выборки для подтверждающих исследований, расчет мощности для ретроспективных исследований и др., а также уделено внимание вопросам воспроизводимости результатов и проведению метаанализа аналогичных исследований. Пошаговые иллюстрации, приведенные в пособии, помогут корректно подобрать и использовать программы для обработки данных.

Предназначено для обучающихся по программам магистратуры направления «Психология». Может быть использовано студентами и аспирантами психологического направления, научно-педагогическими работниками, желающими освоить в исследовательских целях различные компьютерные программы статистического анализа.

УДК 159.9:51:681.3(075.8)

ББК Ю9В641я73

ISBN 978-5-400-01480-2 © Тюменский государственный университет, 2018
© Григорьев П. Е., Васильева И. В., 2018

ОГЛАВЛЕНИЕ

| | |
|---|-----|
| ВВЕДЕНИЕ | 5 |
| 1. ТЕОРИЯ ВЕРОЯТНОСТЕЙ КАК ОСНОВА ДЛЯ СТАТИСТИЧЕСКИХ МЕТОДОВ ОБРАБОТКИ ДАННЫХ. ВЫБОРОЧНЫЙ МЕТОД В ПСИХОЛОГИИ | 8 |
| 2. ТИПЫ ШКАЛ, В КОТОРЫХ ИЗМЕРЯЮТСЯ ДАННЫЕ ПСИХОЛОГИЧЕСКИХ ИССЛЕДОВАНИЙ | 15 |
| 3. ОПИСАТЕЛЬНЫЕ СТАТИСТИКИ. ТОЧЕЧНЫЕ И ИНТЕРВАЛЬНЫЕ ОЦЕНКИ СЛУЧАЙНОЙ ВЕЛИЧИНЫ | 19 |
| 4. СТАТИСТИЧЕСКИЕ КРИТЕРИИ И МЕРЫ СВЯЗИ ДЛЯ КОЛИЧЕСТВЕННЫХ И КАЧЕСТВЕННЫХ ДАННЫХ | 40 |
| 4.1. Виды статистических гипотез. Статистическая значимость. Статистический критерий | 44 |
| 4.2. Понятие о статистическом распределении данных | 51 |
| 4.2.1. Построение гистограмм | 52 |
| 4.2.2. Роль нормального распределения в психологических исследованиях. Проверка данных на нормальность распределения | 55 |
| 4.3. Определение различий и связей в количественных и порядковых данных | 67 |
| 4.3.1. Сравнение выборочного среднего с константой. Параметрический одновыборочный критерий Стьюдента | 68 |
| 4.3.3. Непараметрические критерии для оценки положения и сдвигов в уровне признака в двух независимых и зависимых выборках | 85 |
| 4.3.4. Параметрические и непараметрические коэффициенты корреляции для оценки уровня статистической связи между переменными | 92 |
| 4.3.5. Простая и множественная линейная регрессия | 100 |
| 4.3.6. Анализ различий уровня признака в трех и более независимых выборках с нормальным распределением | 111 |
| 4.3.7. Анализ различий уровня признака в трех и более независимых выборках с отличным от нормального законом распределения | 123 |

| | |
|--|-----|
| 4.3.8. Анализ различий уровня признака при повторных измерениях у одной и той же выборки | 127 |
| 4.4. Статистические критерии для определения различий и связей в качественных и категориальных данных | 143 |
| 4.4.1. Сравнение доли выраженности признака в двух независимых выборках | 145 |
| 4.4.2. Сравнение долей выраженности признака в двух связанных выборках | 151 |
| 4.4.3. Сравнение доли выраженности признака в более чем двух независимых выборках | 157 |
| 4.4.4. Выявление отличий наблюдаемой доли признака от эталонной (теоретически ожидаемой)..... | 161 |
| 4.5. Алгоритм выбора некоторых из возможных к применению статистических критериев для базовых задач психологических исследований | 167 |
| 5. ПРОСПЕКТИВНЫЕ ИССЛЕДОВАНИЯ. ПИЛОТАЖНЫЕ И ПОДТВЕРЖДАЮЩИЕ ИССЛЕДОВАНИЯ. РЕТРОСПЕКТИВНЫЕ ИССЛЕДОВАНИЯ..... | 170 |
| 5.1. Содержательная и статистическая значимость результатов. Расчет объема выборок для подтверждающего исследования | 176 |
| 5.2. Ретроспективные исследования. Отличия от проспективных исследований. Оценка мощности имеющихся выборок..... | 193 |
| 6. ВВЕДЕНИЕ В ПРОБЛЕМАТИКУ ВОСПРОИЗВОДИМОСТИ РЕЗУЛЬТАТОВ ИССЛЕДОВАНИЙ | 203 |
| ЗАКЛЮЧЕНИЕ..... | 211 |
| БИБЛИОГРАФИЧЕСКИЙ СПИСОК | 214 |

ВВЕДЕНИЕ

Перед исследователем в любой предметной области науки стоят, как правило, следующие задачи:

1) сформулировать гипотезы, которым впоследствии будут релевантны выводы;

2) выбрать методы исследования согласно используемой теории, и так, чтобы измеряемые признаки как можно точнее отражали предмет исследования и изучаемые свойства;

3) обеспечить точность измерений признаков на приемлемом уровне;

4) сделать как можно более точные и содержательные выводы;

5) обеспечить воспроизводимость результатов исследования или, по крайней мере, как можно точнее и полнее указать на причины их невоспроизводимости, что позволит при учете выявленных факторов невоспроизводимости организовать и провести последующие исследования.

Практическому психологу (не только психологу-исследователю) необходимо учитывать, что требования, предъявляемые современной наукой к исследованиям в гуманитарных и общественных науках, все больше приближаются к требованиям, которые существуют в естественных науках. Критерии, которые научные журналы прилагают к предоставляемому для публикации материалу, предполагают математическую обработку данных исследования. Это необходимо, чтобы иметь возможность оценить степень достоверности представленных в них данных и выводов, сделанных на их основе, а также в принципе отличать современную психологию как науку от сугубо гипотетических спекуляций или философствований.

При этом развитие программных средств обсчета эмпирических данных позволяет нивелировать традиционные трудности, испытываемые «гуманитариями» в работе с числовыми данными и методами их обработки. При планировании психологических исследований с некоторым числом измерений, проводимых над испытуемыми (количество слов в описании, количество опросников или

тестов, число шкал в них, количество сессий эксперимента и попыток в каждой из них и т. д.), становится необходимым заранее просчитывать число испытуемых и относящихся к ним данных. Без этого невозможно осуществить подтверждающие исследования, когда необходимо проверить ту или иную гипотезу, в частности, на заданном уровне статистической значимости и с заданным размером эффекта; и таким образом — выбрать определенный дизайн исследования, методы и результаты статистической обработки результатов, которые позволят грамотно и достоверно ответить на вопросы исследования.

Авторы пособия преследовали цель, прежде всего, предоставить простой и понятный инструментарий, который может быть полезен при математико-статистическом сопровождении психологических исследований. В то же время пособие следует читать не выборочно, как справочник, а изучать от начала и до конца, поскольку именно так формируется понимание материала, где важные теоретические сведения зачастую вводятся по мере изложения, именно там, где уровень читателя, который полагается начальным, постепенно повышается.

Представленный материал носит несколько эклектичный характер в отношении программных средств и статистических пакетов, на примере которых описаны алгоритмы действий для неспециалистов в областях математики и статистики. Это сделано намеренно, чтобы предложить читателю для конкретного типа и этапа психологического исследования максимально простой и вместе с тем эффективный «рецепт» его математико-статистического сопровождения, поскольку разные этапы и типы задач с неодинаковой наглядностью, «красотой» и простотой реализованы в том или ином пакете. При этом мы ориентировались на использование статистических пакетов, которые «работают» с максимально широким диапазоном задач, стоящих перед психологами-исследователями. Среди используемых для иллюстрации алгоритмов действий программ были использованы табличный процессор Excel, пакеты статистической обработки данных Statistica, SPSS, Primer of Biostatistics, «Медстат» (MedStat), StatMed, LePrep, BioStat.

Книга не предназначена для изучения самих методов психологии, дизайнов исследований, не предполагает строгости в описании используемых математико-статистических показателей и формул, для чего читатель адресуется к изучению соответствующих источников.

Примеры в настоящем учебном пособии носят обезличенный и зачастую упрощенный характер, чтобы внести ясность в применение того или иного статистического метода.

Настоятельно рекомендуем в паре к настоящему пособию учебное пособие Е. Л. Доценко и З. З. Вахитовой «Психосемантика» [7], где на высоком методическом и прекрасном иллюстративном уровне описаны наиболее применимые в исследованиях психологов многомерные статистические методы.

Авторы выражают надежду, что теоретические сведения и практические рецепты статистического анализа данных психологических исследований будут полезны широкому кругу обучающихся: аспирантам, магистрантам, бакалаврам, практическим психологам и психологам-исследователям, а также специалистам смежного профиля.

1. ТЕОРИЯ ВЕРОЯТНОСТЕЙ КАК ОСНОВА ДЛЯ СТАТИСТИЧЕСКИХ МЕТОДОВ ОБРАБОТКИ ДАННЫХ. ВЫБОРОЧНЫЙ МЕТОД В ПСИХОЛОГИИ

В основе всех статистических методов обработки информации, в том числе в психологии, лежит аппарат теории вероятностей. Базовые положения теории вероятности довольно просты для понимания.

Теория вероятностей изучает и отображает в количественной форме закономерности, присущие случайным событиям массового характера, например:

- количество студентов, поступивших в вузы для обучения по направлению «Психология»;
- количество учеников школ, выбравших для обучения естественнонаучный или гуманитарный профиль;
- количество детей дошкольного и младшего школьного возраста, у которых была диагностирована минимальная мозговая дисфункция (синдром дефицита внимания и гиперактивности);
- количество обращений родителей к школьным психологам по вопросам взаимоотношений с подростками.

Следует подчеркнуть, что такого рода закономерности вообще играют исключительно важную роль в психологии. Количество выпускников общеобразовательных школ, сдавших ЕГЭ на 90 баллов и более; количество медицинских сестер в городе, готовых трудоустроиться; количество назначаемых судебно-психологических экспертиз и т. п. — представляют примеры случайных событий, закономерности возникновения и предсказание которых требуют анализа большой группы событий. Изучение подобных явлений в психологии с помощью аппарата теории вероятностей обнаруживает новые своеобразные закономерности и законы, которые являются количественным выражением психологических закономерностей.

К изучению явлений теория вероятностей применяет математический метод и, следовательно, является одним из разделов математики, столь же логически точным и строгим, как и другие математические науки. Поэтому ее путь — точное рассуждение, а орудиями служат формулы, таблицы, графики и другие символические средства математического языка.

Под *случайным событием* понимают такое, которое может произойти или не произойти при осуществлении определенного фиксированного комплекса условий. Осуществление такого комплекса условий называется *испытанием* (опытом, экспериментом, операцией). Будем называть благоприятными такие исходы испытаний, в которых наблюдается изучаемое событие.

Приведем несколько примеров благоприятных исходов:

- 1) получение 50 и менее баллов в итоговой школьной аттестации (при задаче выявления наименее успешных выпускников);
- 2) возрастание доли оценок «отлично» и «хорошо» по предметам физико-математического цикла (в исследовании успеваемости учащихся после введения в образовательном учреждении специализированного обучения);
- 3) обнаружение нарушения процессов мышления у ребенка дошкольного возраста (исследование — перед поступлением в школу отбор детей с нарушениями познавательной сферы).

Каждое из перечисленных событий обладает некоторой степенью возможности: одни большей, другие меньшей, причем для некоторых из этих событий можно сразу решить, какое из них более, а какое менее возможно.

Частота. Вероятность

Отношение числа благоприятных испытаний, в котором появилось событие (или A , или B), к общему числу испытаний называется **частотой** появления данного события.

Количественной характеристикой устойчивой частоты служит *вероятность*. В общем случае вероятность события A — это число $P(A)$, вблизи которого колеблется частота события A при большом количестве испытаний.

В некоторых случаях вероятность наступления события можно определить и без проведения испытания, исходя из теоретических соображений.

Достоверные и невозможные события

Из определения частоты и вероятности следует, что их величины для различных событий и испытаний могут изменяться в определенных пределах. Если при каждом испытании обязательно наблюдается исследуемое событие A , то частота (вероятность) такого события равна единице, или 100%, если умножить на 100. Событие, которое обязательно наблюдается в каждом испытании, называется *достоверным* и его вероятность равна единице: $P(A) = 1$.

Например, при проведении процедуры психологического тестирования состоится оценивание интеллекта и будет получено числовое выражение в формате IQ уровня развития интеллекта человека определенного возраста. Поэтому получение того или иного числового выражения уровня развития интеллекта при проведении тестирования, явление достоверное, и его вероятность $P(A) = 1$.

Если рассматриваемое событие заведомо не может произойти при осуществлении заданного комплекса условий, то число благоприятных испытаний равно нулю, и соответственно частота и вероятность такого события тоже равны нулю. Событие, которое не может появиться ни в одном испытании, называется *невозможным*.

Например, при нарушении интеллектуального развития в стадии идиотии человек не способен успешно обучаться в общеобразовательных учреждениях. Это явление невозможное, и его вероятность $P(A) = 0$.

Из приведенных определений следует, что вероятные события располагаются между достоверными и невозможными, и, следовательно, вероятность любых мыслимых событий лежит в диапазоне: $0 \leq P(A) \leq 1$. Ясно, что когда вероятность умножается на 100, то мы получаем ее выражение в процентах.

Если $P(A) = 1$, такое событие достоверно, если $P(A) = 0$ — невозможно. Вероятность события не может быть больше 1, поскольку число «благоприятных» испытаний (в которых проявляется

изучаемое событие) не может быть больше общего количества испытаний.

Принято также различать *совместимые* (совместные) и *несовместимые* (несовместные) события.

Несколько событий называются несовместимыми в данном опыте, если никакие из них не могут появиться вместе, то есть взаимно исключаются. Если же при появлении события А возможно также и осуществление события В, то такие два события называются совместимыми.

Например, в результате психологического тестирования была оценена характеристика экстраверсии–интроверсии методикой ЕРІ (Г. Айзенка) и было получено количественное выражение этой шкалы 20 (это означает выраженную экстравертированность). Этот результат исключает получение аналогичных значений по показателю интровертированности, поскольку это полярные показатели одной шкалы. Это события в большинстве случаев несовместимые (несовместные).

Однако в результатах того же тестирования испытуемый может получить выраженные показатели по шкале эмоциональной стабильности — нейротизма. Испытуемый с выраженными показателями экстравертированности может иметь как высокий уровень нейротизма (и тогда диагностируется тип личности, по Г. Айзенку, — холерик), так и высокий уровень эмоциональной стабильности (диагностируется тип личности — сангвиник). Эти события, как правило, совместимы (являются совместными).

Место статистики в психологических исследованиях

В психологии часто приходится иметь дело с результатами наблюдений, полученными в ходе групповых исследований.

Например:

1. В октябре в школах проводятся массовые обследования учеников первых классов для оценки когнитивной и эмоциональной сфер на предмет готовности к школе. В результате школьный психолог получает данные об уровне развития наглядного, абстрактного мышления, словесной и зрительной памяти, уровня

выраженности тревожности, работоспособности, нервно-психического напряжения.

2. При поступлении на службу в силовые структуры (МВД, ФСИН, МЧС и т. д.) претенденты проходят процедуры оценки профессионально важных качеств. В результате оценивается уровень интеллекта, личностные характеристики, потенциальные риски психопатологических отклонений, психофизиологические особенности сенсорных систем.

3. На прием к психологу-консультанту приходят клиенты с проблемами детско-родительских отношений. Они описывают феноменологию проблемы, которую консультант фиксирует в терминах поведения: «не слушается», «огрызается в ответ», «пропускает уроки», «связался с дурной компанией». Эти поведенческие проявления являются данными, которые могут быть учтены и статистически обработаны.

При таких обследованиях получают массив результатов, которые называются статистическими данными, и они нуждаются в дальнейшей обработке методами статистики, в основе которых лежит теория вероятностей.

Задачами математической статистики являются: разработка способов сбора и группировки статистических сведений, разработка методов анализа статистических данных, в зависимости от целей исследования.

В результате решения этих задач в психологических исследованиях можно установить новые психологические закономерности, вскрыть психологические механизмы явлений, оценить динамику изменения состояния испытуемых, клиентов, подтвердить эффективность способов психокоррекции и т. п.

Генеральная и выборочная совокупности

Пусть требуется изучить совокупность однородных объектов относительно некоторого признака, характеризующего эти объекты. Например, уровень развития невербального интеллекта школьников общеобразовательных учреждений в Российской Федерации.

Генеральной совокупностью называется совокупность всех объектов, из которых производится выборка. Для вышеприведен-

ного примера генеральной совокупностью будут все школьники с 1 по 11 класс общеобразовательных учреждений Российской Федерации. На практике очень часто генеральная совокупность содержит такое большое количество объектов, что их сплошное обследование нерационально или вообще физически невозможно. В подобных случаях отбирают из всей совокупности ограниченное число объектов и подвергают их изучению. Например, проводят обследование школьников в одной школе на предмет оценки их невербального интеллекта.

Выборочной совокупностью, или просто *выборкой*, называется совокупность объектов, из которых состоит выборка. Для рассматриваемого примера такой выборкой могут быть школьники, обучающиеся в какой-нибудь одной школе. *Объемом совокупности* (выборочной или генеральной) называется число объектов n этой совокупности. Для нашего примера объемом генеральной совокупности будет N — число школьников в Российской Федерации, а объемом выборки — n — число обучающихся в школе, которая выбрана для проведения обследования.

Обычно к выборке предъявляют требования представительности (репрезентативности) и гомогенности в отношении к генеральной совокупности, что достигается правильной рандомизацией (случайным отбором испытуемых в выборку из генеральной совокупности), за счет чего исследователь получает неискаженные данные в достаточном количестве, чтобы доказательно подтвердить или отвергнуть выдвинутые гипотезы.

Репрезентативность выборки — это способность выборки представлять изучаемые явления в достаточной степени, чтобы выводы исследования можно было бы с допустимой погрешностью распространить на генеральную совокупность.

Чтобы обеспечить репрезентативность, применяют:

1. Простой случайный отбор. Каждый член генеральной совокупности имеет одинаковые шансы попасть в выборку; принимаются меры, исключающие появление какой-либо закономерности при отборе.

Например, изучая агрессивность подростков, исследователь может случайным образом остановить свой выбор на трех классах

разных школ и затем случайным образом отобрать по 10 учащихся из каждого класса. Если исследователь, например, просит испытуемого пригласить на обследование своих друзей, он грубо нарушает принцип случайности отбора.

2. Случайный отбор по свойствам генеральной совокупности (стратифицированный случайный отбор). Требуется предварительное определение свойств, которые могут влиять на изучаемые свойства (семейное положение, возраст, пол, конфессия, уровень дохода, сфера деятельности и проч.). После этого определяется соотношение долей (процентов) различающихся по этим свойствам групп в генеральной совокупности и обеспечивается аналогичное процентное соотношение подгрупп в выборке. Далее в каждую подгруппу испытуемые подбираются по принципу простого случайного отбора.

Например, исследователь может предположить, что мальчики и девочки различаются как по агрессивности, так и по восприимчивости к сценам насилия в компьютерных играх. Если исследователь планирует обобщить результат исследования влияния насилия в компьютерных играх на агрессивность всех подростков, то, исходя из социально-демографических данных, необходимо аналогичное генеральной совокупности соотношение мальчиков и девочек в выборке.

Еще одно важное условие репрезентативности — объем выборки. Однако в зависимости от задачи, стоящей перед исследователем, и типа исследования (пилотное, эксплораторное, подтверждающее) объем выборки может существенно разниться.

2. ТИПЫ ШКАЛ, В КОТОРЫХ ИЗМЕРЯЮТСЯ ДАННЫЕ ПСИХОЛОГИЧЕСКИХ ИССЛЕДОВАНИЙ

Психологу-исследователю необходимо понимать, к какой именно шкале по типу относятся данные, с которыми он собирается иметь дело или уже имеет дело. Рассмотрим эти шкалы, ведь они во многом определяют дальнейший выбор методов обработки данных.

Шкала наименований (категорий). Объекты различаются только названиями, и им могут быть присвоены условные числовые обозначения. Например, пол — мужской и женский. Условно можем обозначить мужской пол как 1, женский — как 2. Можем любым другим произвольным числом или символом. Однако мы не можем сказать — больше или меньше тот или иной объект другого, не можем складывать и вычитать параметры, присущие каждому объекту, а также делить их и умножать. Единственное, что мы можем, так это определить количество одних и других объектов, и в принципе различать их между собой. Например, в выборке может быть 20 женщин и 30 мужчин. Мы можем в данном случае лишь сказать, что количество мужчин больше, чем женщин, а значит, определить только такой статистический параметр, как *мода*. *Мода* показывает, какие объекты преобладают в выборке. В данном случае — *мода = мужчины*. Моды может не быть, если, например, количество объектов разного сорта в выборке одинаково: например, 20 женщин и 20 мужчин. С другой стороны, распределение признака в шкале наименований может иметь больше одной моды. Например, рассмотрим некоторую выборку представителей разных профессий: 20 плотников, 30 слесарей, 60 сантехников, 3 бухгалтера, 60 дворников. В этой выборке 2 моды — сантехники и дворники. Обычно моду обозначают как *Мо*.

Шкала порядка. В этой шкале мы уже не только можем различать объекты, но и упорядочивать их по степени выраженности того или иного свойства. Например, в выборке школьников у нас 4 троечника, 1 двоечник, 2 отличника и 3 хорошиста. Нам известно,

что троечник учится лучше, чем двоечник, а хорошист — хуже отличника. То есть мы можем упорядочить объекты по возрастанию уровня выраженности признака, однако не знаем, насколько конкретный троечник проявляет знания хуже, чем конкретный хорошист. Иными словами, мы не можем сказать, насколько именно значение признака объекта выражено больше другого значения, но зато однозначно можем их упорядочить. Поэтому шкала и называется шкалой порядка. Помимо моды, мы еще можем узнать *медиану* выборки, которая обозначается как *Me*. Медиана нам может дать информацию о срединном значении изучаемого признака в выборке. Чему равна медиана в нашем примере? Расположим значения признаков по возрастанию в ряд — проранжируем: 2, 3, 3, 3, **3**, **4**, 4, 4, 5, 5. В данном случае имеется 10 значений, середина попадает на значения, расположенные под номерами 5 и 6. У нас четное количество объектов в выборке, поэтому $Me = (3 + 4) : 2 = 3,5$, то есть среднему между значениями, попадающими в середину выборки. Если количество объектов нечетное, тогда мы берем просто одно значение, попадающее на середину ряда, например, если бы в нашем ряду не было бы двоечников и число объектов было бы равно девяти (3, 3, 3, 3, **4**, 4, 4, 5, 5), то мода равнялась бы 4 и по обе стороны этого значения было бы одинаковое число упорядоченных по возрастанию наблюдений. Медиана дает больше информации о выборке, чем мода, однако она не учитывает вклада *вариант* (элементов вариационного ряда признака, в котором значения выстроены от минимального до максимального), лежащих за пределами срединного значения. С другой стороны, поскольку медиана учитывает только одно число, она свободна от искажений за счет вклада крайних значений, которые могут иметь несоразмерно большие или малые значения — артефакты или выбросы. Следует отметить, однако, что в психологии именно эта шкала имеет особое значение, поскольку признаки, выраженные в этой шкале, часто ошибочно приписывают метрическим шкалам, о которых сейчас и пойдет речь.

Шкала интервалов. Более мощная в смысле получаемой информации, чем рассмотренные выше. Если первые две шкалы от-

носились к неметрическим шкалам, когда нам не было известно «расстояние» между значениями, то здесь появляется некоторая условная линейка. Она позволяет не только различать и упорядочивать объекты, но и находить разности или суммировать, рассчитывать расстояние между значениями признака. Например, по шкале IQ (стандартизированной со средним 100 и стандартным отклонением 15), мы можем сказать, на сколько единиц тот или иной тестируемый показал лучший результат, и, мало того, можем интерпретировать его с точки зрения нормы. Здесь помимо моды и медианы оказывается уместным применение среднего арифметического, часто обозначаемого как M . Однако, например, если испытуемый получил бы 0 баллов по тесту интеллекта, это еще ничего не говорит об отсутствии интеллекта вообще. И если мы попробовали бы уточнить в данном случае, во сколько раз больше этого нулевого значения интеллект испытуемого, получившего не важно какое значение — 1 или 130, соотношение было бы в бесконечное число раз выше. Ведь какое бы конечное число мы бы ни делили на нуль, то получим бесконечность, чего не может быть. Еще один простой пример из жизни: температура, выраженная в градусах Цельсия. Сегодня -2 градуса, а вчера было 0. Значит разность составила $0 - 2 = -2$ градуса. Действительно, сегодня на 2 градуса холоднее, но мы не можем узнать — во сколько раз сегодня холоднее: $-2 : 0 =$ минус бесконечность...

Еще раз коснемся следующей проблемы: психологи-исследователи могут ошибочно приписывать интервальность признаку, фактически выраженному в шкале порядка. Напомним пример из рассмотрения порядковой шкалы. Хотя нам понятно, что оценка 4 лучше, чем 3, мы не знаем, насколько реально лучше отвечал ученик. Иная тройка — это почти двойка, а иная четверка — почти пятерка... Поэтому если мы имеем дело с опросниками, анкетами или другими методами психологии, не прошедшими процедуру стандартизации, наши данные, возможно, являются порядковыми, и мы не можем их складывать и вычитать, но лишь упорядочивать. На это следует обратить особое внимание. Да, иногда применение среднего арифметического и в признаках, выраженных в шкале

порядка, оправданно, но при условии четкого и содержательного обоснования.

Шкала отношений — самая «мощная», позволяет не только складывать и вычитать, но еще делить и умножать. Например, мы изучаем время решения испытуемым неизвестной ему задачи и начинаем отсчет времени по секундомеру. Мы можем подсчитать, во сколько раз один испытуемый решил задачу быстрее другого, притом мы имеем нулевую точку — начало отсчета времени решения задачи. Еще один бытовой пример — масса. 6 килограмм конфет ровно в три раза больше, чем 2 килограмма.

Что же роднит эти два примера? То, что у нас нулевая точка соответствует полному отсутствию данного признака. Скажем, ноль килограмм — это отсутствие массы вообще, так же, как и ноль в приведенной выше задаче соответствует отсутствию времени, затраченного на решение задачи в принципе. Поэтому в этой шкале мы можем не только различать объекты (как в шкале наименований), упорядочивать их (как в шкале порядка), вычитать и складывать из значения (как в шкале интервалов), но еще, к тому же, делить и умножать.

3. ОПИСАТЕЛЬНЫЕ СТАТИСТИКИ. ТОЧЕЧНЫЕ И ИНТЕРВАЛЬНЫЕ ОЦЕНКИ СЛУЧАЙНОЙ ВЕЛИЧИНЫ

С некоторыми из описательных статистик мы начали знакомиться в предыдущих главах. Подытожим наши знания и приведем примеры расчетов с помощью одного из статистических пакетов.

Итак, каждой выборке можно приписать некоторые точечные оценки, то есть конкретное число, характеризующее параметр выборки.

Среднее арифметическое значение — характеристика центра статистического распределения, — равно сумме вариантов (наблюденных значений), деленной на их количество.

Медиана — характеристика центра статистического распределения, делит упорядоченный по возрастанию ряд исходных значений (вариационный ряд) на две равные части по числу вариантов.

Мода — характеристика центра статистического распределения, наиболее вероятное (чаще всего встречающееся) значение признака. Напомним, что мод может быть несколько (если есть несколько значений, встречающихся одинаково и наиболее часто) или вовсе не быть, когда объектов каждого сорта в выборке одинаковое количество.

Представленные выше характеристики также называют мерами центральной тенденции.

Минимум и максимум — соответственно минимальное и максимальное значения признака в выборке — крайние точки вариационного ряда.

Однако нам не обойтись при анализе данных лишь точечными оценками признака. Представим ситуацию, когда в одном классе много отличников и двоечников, но мало троечников и хорошистов, а в другом классе — много хорошистов и троечников и вовсе нет отличников и двоечников. У нас могут получиться в таком случае одинаковые значения медиан и средних, при том, что на самом деле эти классы существенно различаются по структуре успеваемо-

сти. Поэтому нам нужны помимо точечных еще интервальные оценки случайной величины, которые и будут показывать степень разброса данных относительно мер центральной тенденции.

Дисперсия. Величина одного и того же признака неодинакова у всех членов совокупности. Например, в классе есть дети, которые легко и быстро решают математические задачи (обладают развитым числовым мышлением) и в интеллектуальных тестах по разделу «числовое мышление» показывают высокие результаты, а есть те, кто показывают средние и низкие результаты. Тогда как в физико-математических классах, естественно, отсутствуют низкие результаты по показателю «числовое мышление». В этом проявляется разнообразие изучаемого параметра.

Один из способов измерения интервальных данных заключается в том, чтобы определить степень отклонения каждого наблюдения от среднего арифметического. Очевидно, что чем больше отклонение, тем больше изменчивость, вариабельность наблюдений. Однако невозможно использовать среднее этих отклонений как меру рассеяния, потому что положительные отклонения компенсируют отрицательные отклонения (их сумма равна нулю). Чтобы решить эту проблему, можно возвести в квадрат каждое отклонение и найти среднее возведенных в квадрат отклонений; эта величина — выборочная дисперсия:

$$D_x = \frac{\sum_{i=1}^n (x_i - M_x)^2}{n - 1},$$

где x — конкретное значение; n — количество значений в ряду признака (например, количество испытуемых, измерений и т. п. — вариант); i — номер конкретного измерения. Видно, что все отклонения конкретного значения x от среднего M возводятся в квадрат и суммируются, затем их сумма делится на количество вариантов за вычетом единицы.

Среднее квадратичное отклонение — характеристика степени разброса элементов выборки вокруг среднего значения. Квадратный корень из дисперсии:

$$\sigma = \sqrt{D}.$$

Это характеристика изменчивости значений самой выборки, и, как правило, мало изменяется с увеличением вариантов (значений) в выборке.

Стандартная ошибка среднего — используется для расчета интервальной оценки среднего арифметического генеральной совокупности. Уменьшается пропорционально квадратному корню из числа вариантов в выборке:

$$m = \sqrt{\frac{D}{n}}.$$

Поэтому с увеличением выборки разброс значений вокруг среднего уменьшается, и мы получаем все более точное представление об истинном среднем значении, сужаем интервал, в котором оно может находиться. Таким образом, увеличивая выборки по размеру, мы сможем с большей вероятностью отвергнуть ложную статистическую гипотезу.

Левая и правая границы 95% доверительного интервала оценки среднего — интервал значений признака, в который с вероятностью $p = 0,95$ попадает истинное значение математического ожидания генеральной совокупности: $X_l < M < X_n$ (M — среднее арифметическое, X_l — левая граница доверительного интервала, X_n — его правая граница). Внутри этого интервала лежит 95% значений признака. Однако такое определение не является строгим, поскольку точнее характеризует большие выборки, нежели малые, в этом легко убедиться на примерах (академическая успешность в классе — общеизвестно, что есть «слабые» и «сильные» классы; сплоченность группы — вследствие того, что группы могут быть разной степени сплоченности, специально организуются и проводятся социально-психологические тренинги). Применяются также и другие доверительные интервалы, например, 90% или 99%, в зависимости от целей исследования.

Приведенные интервальные оценки используются, когда в качестве меры центральной тенденции используется среднее арифметическое.

Первый (третий) квартиль — значение элемента, упорядоченного по возрастанию вариационного ряда, левее (правее) которого находится 25% элементов ряда. Пятьдесят процентов всех значений анализируемого признака лежит между первым и третьим квартилями. Понятие второго квартиля соответствует медиане ряда. Квартили широко используются в случае, когда в качестве меры тенденции мы используем медиану. Разность между квартилями называют межквартильным расстоянием.

Размах вариации — разность между максимальным и минимальным значениями в выборке:

$$R_x = x_{\max} - x_{\min}.$$

Реже используется *процентиль* — это 99 точек — значений признака, которые делят упорядоченную по возрастанию признака совокупность измерений на 100 частей, равных по численности. Для определения конкретного процентиля, например, десятого, все значения признака упорядочиваются по возрастанию, затем отсчитываются 10% значений со стороны меньших значений. Так, P_{10} соответствует значению признака, который отделяет эти 10% испытуемых от остальных 90%. Не стоит путать вероятность и процентиль, обозначаемый одной и той же буквой P , поскольку у процентиля всегда присутствует индекс, соответствующий тому проценту доли признака, который остался «позади».

Рассмотрим примеры расчета некоторых из этих значений в программе Statistica.

После запуска программы вводим или загружаем данные, необходимые для анализа. В этом примере будут анализироваться количественные данные эксперимента, в котором был получен ряд переменной «понимание своих эмоций» (из теста «Эмоциональный интеллект» Д. В. Люсина). Рассмотрим ввод данных вручную. В случае же загрузки готового файла в формате самой программы или табличного процессора (например, Excel), следует пользоваться опцией Open. Для начала в разделе File выбираем New (рис. 3.1), после чего появляется меню (рис. 3.2).

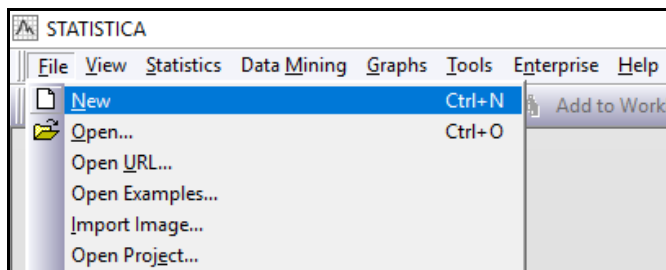


Рис. 3.1

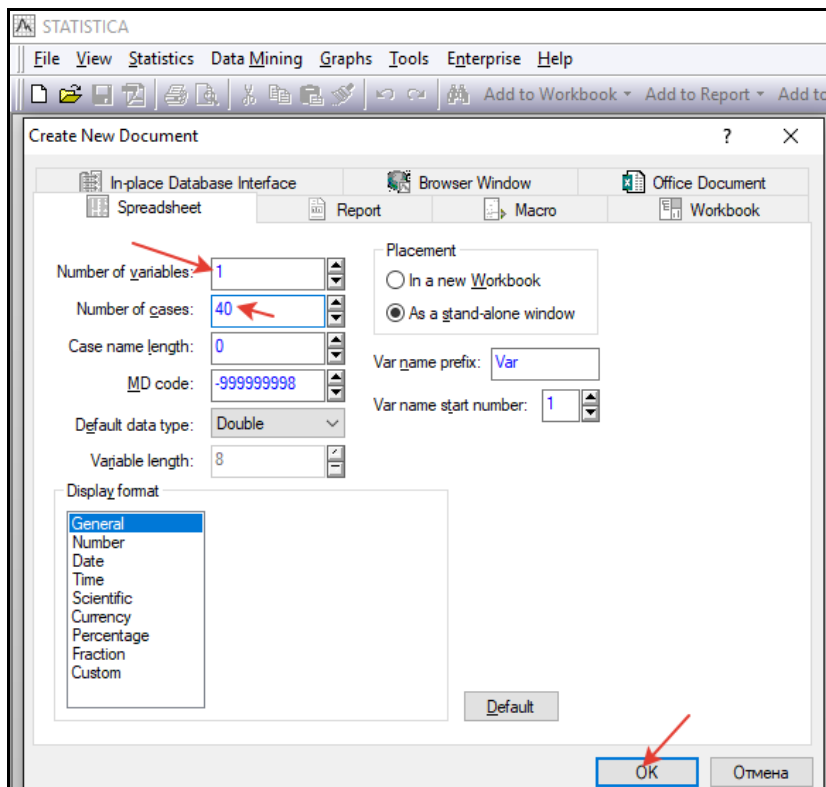


Рис. 3.2

В меню мы задаем количество переменных Variables, равное 1, поскольку мы будем иметь дело в данном случае лишь с одной переменной, и задаем количество случаев (Cases), соответствующему числу испытуемых. Пусть в нашем случае их было 40. Затем нажимаем на «OK». Далее открывается следующая вкладка (рис. 3.3). Для экономии места на скриншоте покажем лишь часть вариантов (10), на самом деле их 40.

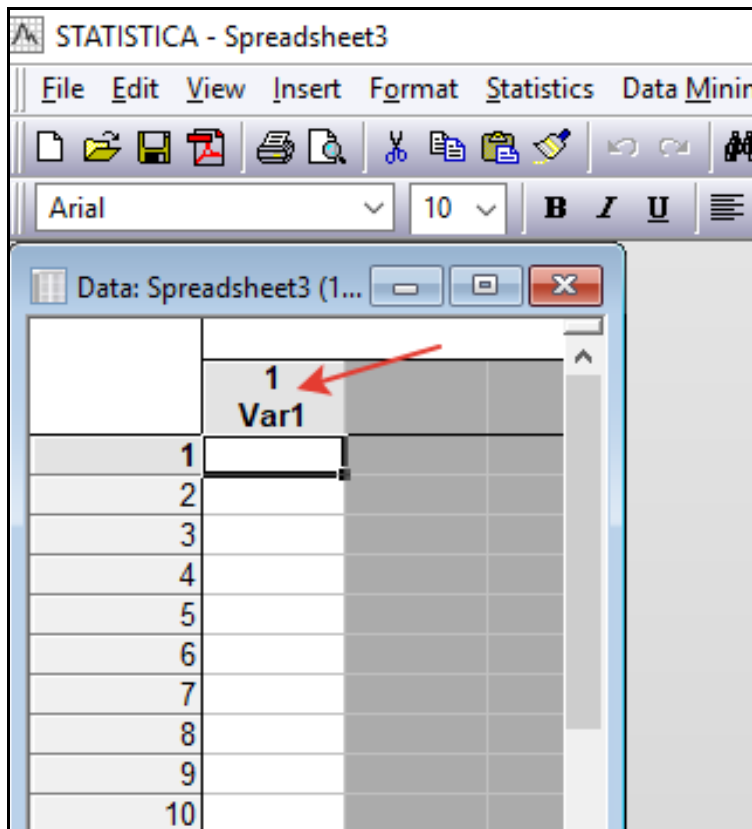


Рис. 3.3

Далее наша задача — назвать переменную и ее тип. Для этого нажимаем на указанное на рис. 3.3 место (Var1) — название переменной по умолчанию надо изменить на требуемое, а также определяем ее тип (рис. 3.4).

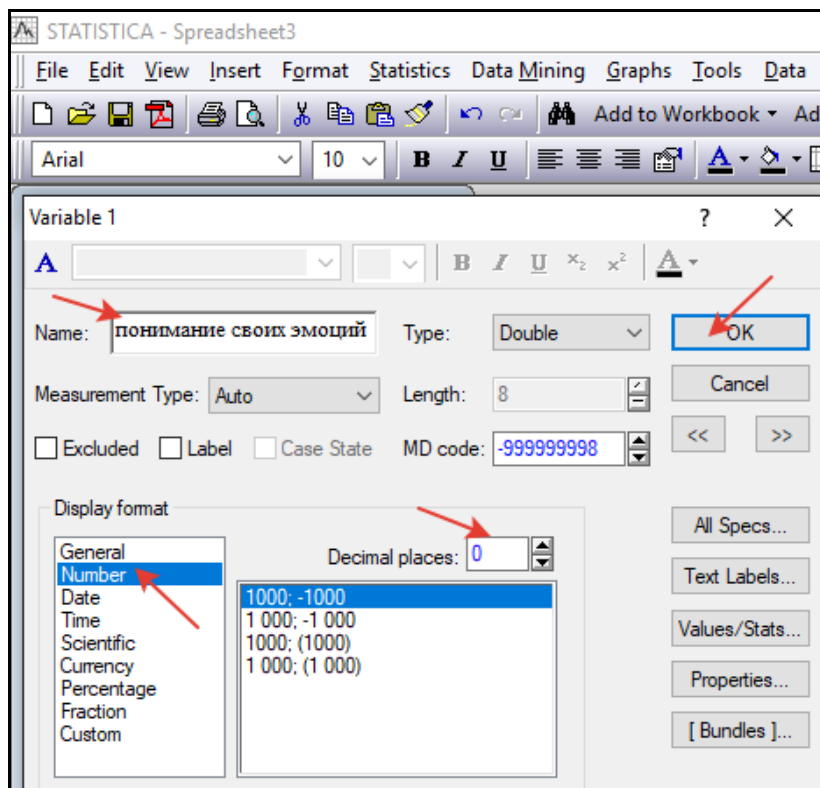


Рис. 3.4

Называем нашу переменную «Понимание своих эмоций», определяем ее тип Number (числовой) и указываем количество знаков после запятой (Decimal places). В данном случае наша переменная является натуральным числом, поэтому мы задали именно такие параметры. Затем нажимаем на «OK».

Теперь наша переменная обрела свое название (рис. 3.5).

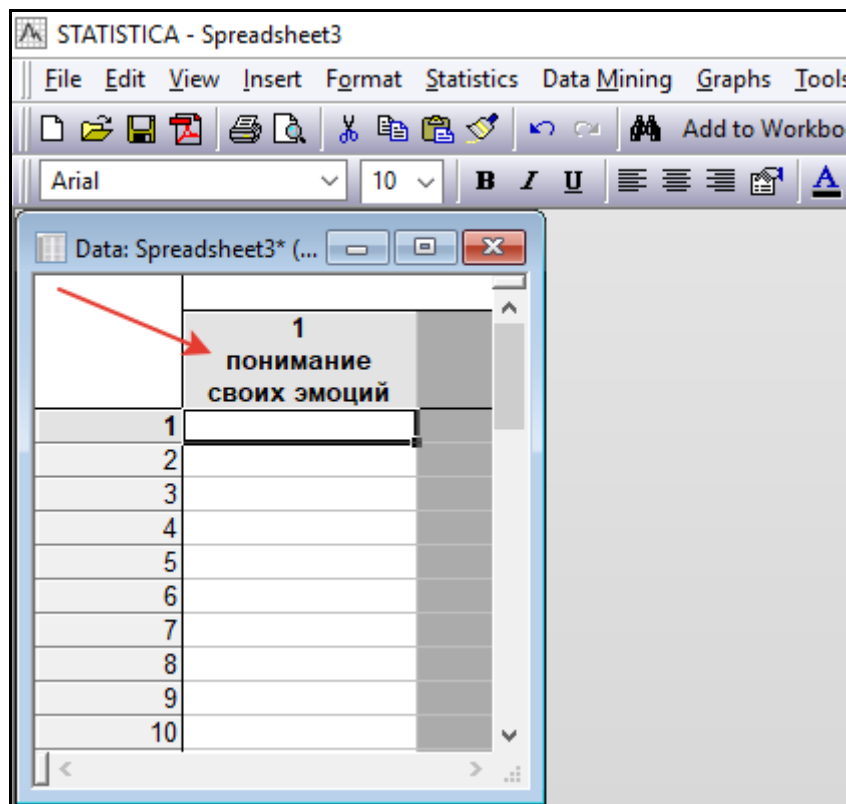
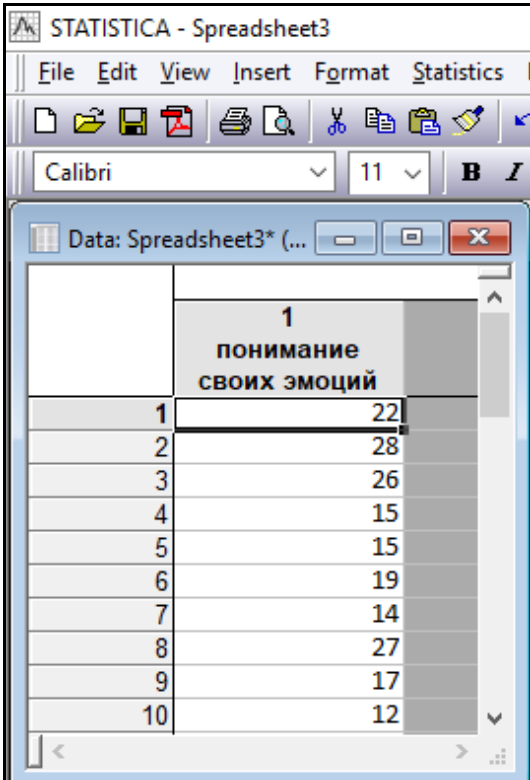


Рис. 3.5

Осталось внести соответствующие данные вручную или вставить их из буфера обмена, если они были набраны в другой программе (рис. 3.6).



The screenshot shows a window titled "STATISTICA - Spreadsheet3". The menu bar includes File, Edit, View, Insert, Format, and Statistics. The toolbar contains icons for file operations and editing. The font settings are set to Calibri, size 11, with bold and italic options. The data table is titled "Data: Spreadsheet3* (...)" and contains the following data:

| | 1 |
|----|------------------------|
| | понимание своих эмоций |
| 1 | 22 |
| 2 | 28 |
| 3 | 26 |
| 4 | 15 |
| 5 | 15 |
| 6 | 19 |
| 7 | 14 |
| 8 | 27 |
| 9 | 17 |
| 10 | 12 |

Рис. 3.6

Когда данные будут введены, запускаем раздел Statistics (Статистика), вкладка Basic Statistics/Tables (Базовые статистики/таблицы) (рис. 3.7).

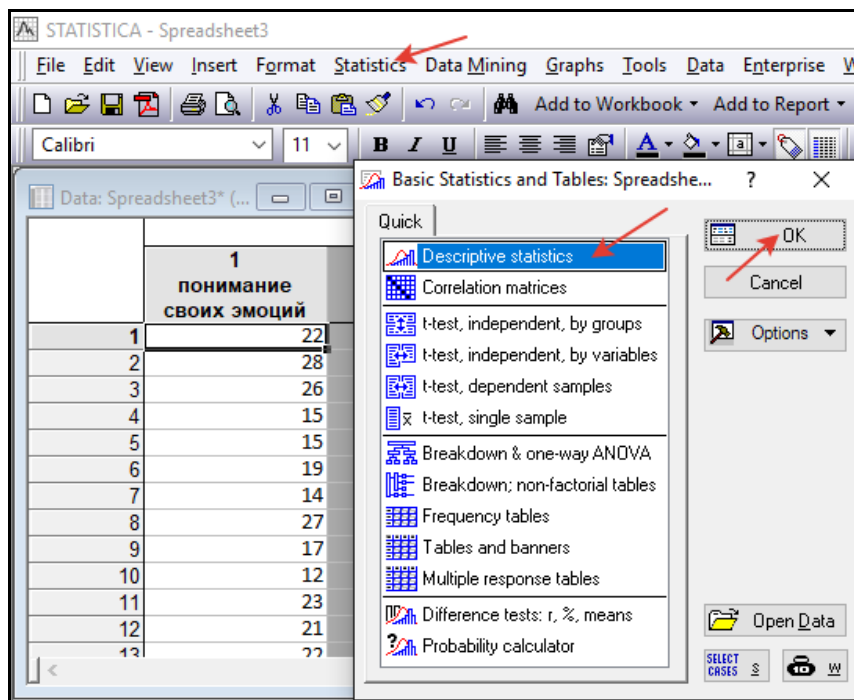


Рис. 3.7

После этого открывается меню для описательных статистик (рис. 3.8). Последовательно нажимаем на вкладку Advanced (Расширенный выбор, в данном случае), затем на кнопку Variables (Переменные), выбираем интересующие нас переменные, в данном случае — она одна, и нажимаем на «ОК».

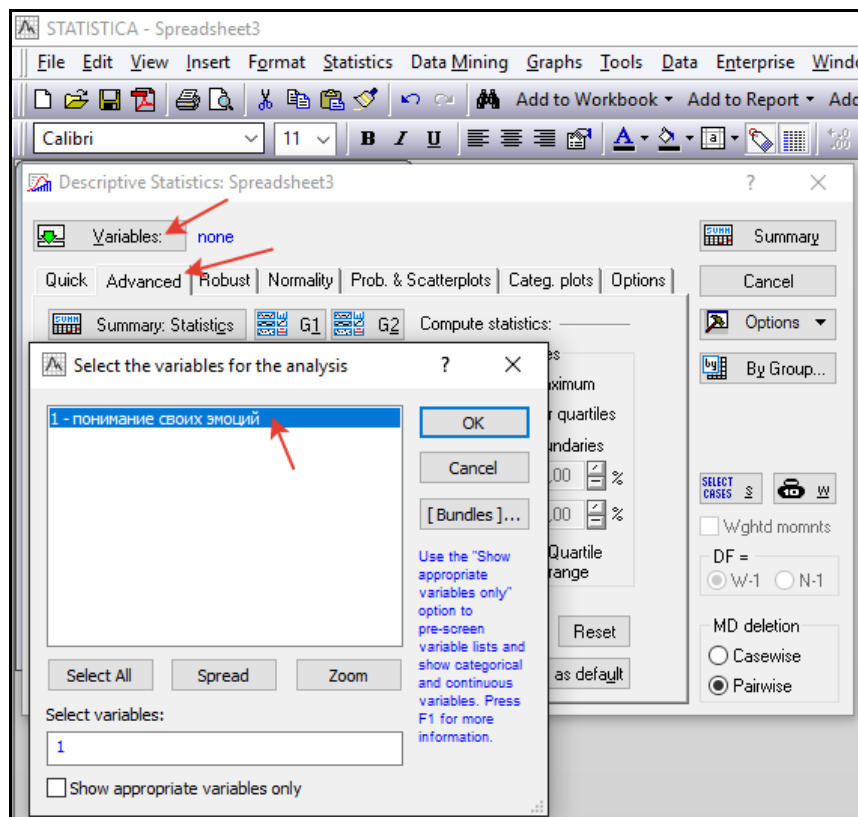


Рис. 3.8

В открывшемся меню (рис. 3.9) выберем описательные статистики, которые мы обсудили выше: Valid N — количество пригодных для анализа случаев (испытуемых), Mean — среднее арифметическое, Median — медиана, Mode — мода, Standard deviation — стандартное отклонение, CI for Sample SD — 95% доверительный интервал, Variance — дисперсия, Std. err. of mean — стандартная ошибка среднего, Conf. limits for means — доверительные интервалы для среднего, Minimum & maximum — минимальное и максимальное значения в выборке, Lower and upper quartiles — нижний и верхний квартиль, Range — размах вариации, то есть разность между максимальным и минимальным значениями, Quartile Range — разность между верхним и нижним квартилями.

После этого нажимаем на Summary — итог.

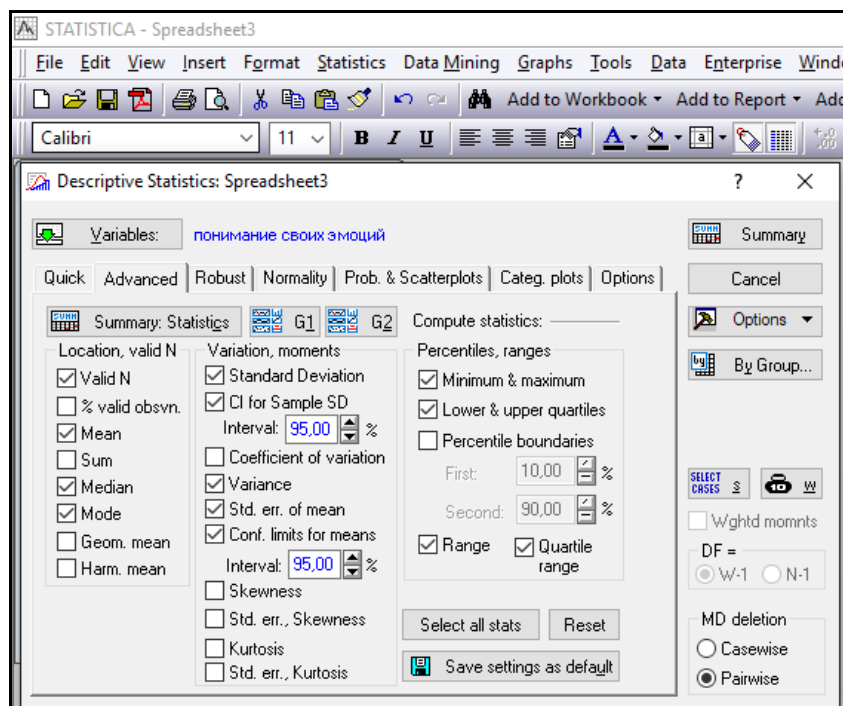
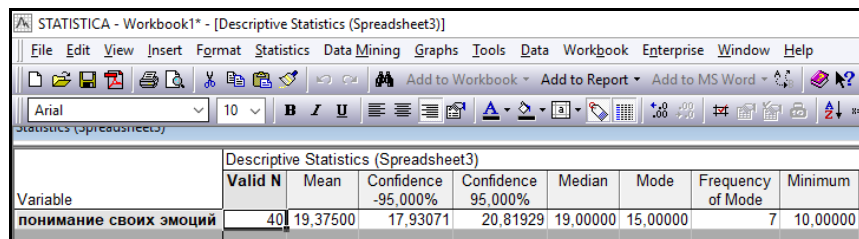


Рис. 3.9

Итог будет представлен в виде таблицы. В целях наглядности разобьем строку результатов по вертикали на два скриншота, поскольку иначе текст будет слишком мелким для печатного издания. Итак, первая часть результатов (рис. 3.10).



STATISTICA - Workbook1* - [Descriptive Statistics (Spreadsheet3)]

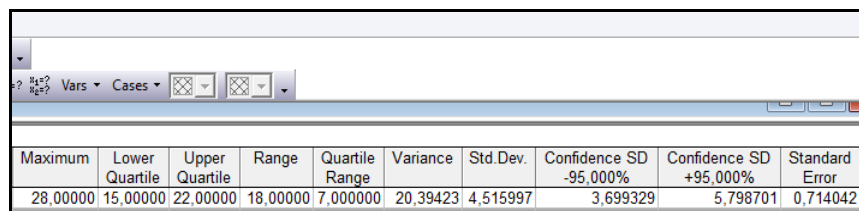
File Edit View Insert Format Statistics Data Mining Graphs Tools Data Workbook Enterprise Window Help

Statistics (spreadsheet3)

| Descriptive Statistics (Spreadsheet3) | | | | | | | | |
|---------------------------------------|---------|----------|------------------------|-----------------------|----------|----------|----------------------|----------|
| Variable | Valid N | Mean | Confidence -95,000% | Confidence 95,000% | Median | Mode | Frequency of Mode | Minimum |
| понимание своих эмоций | 40 | 19,37500 | 17,93071 | 20,81929 | 19,00000 | 15,00000 | 7 | 10,00000 |

Рис. 3.10

И вторая часть (рис. 3.11), которая на самом деле является продолжением первой.



Vars Cases

| Maximum | Lower Quartile | Upper Quartile | Range | Quartile Range | Variance | Std.Dev. | Confidence SD -95,000% | Confidence SD +95,000% | Standard Error |
|----------|-------------------|-------------------|----------|-------------------|----------|----------|---------------------------|---------------------------|-------------------|
| 28,00000 | 15,00000 | 22,00000 | 18,00000 | 7,000000 | 20,39423 | 4,515997 | 3,699329 | 5,798701 | 0,714042 |

Рис. 3.11

Теперь мы видим весь спектр полученных значений для данной выборки. Отсюда можно получить самую разнообразную информацию о любой выборке. Другое дело, как мы обсуждали выше, что не всякую выборку мы можем охарактеризовать всеми значениями. Например, количество испытуемых и мода — это единственные характеристики номинативных данных. По отношению к порядковым данным мы можем также применить максимум, минимум, размах, моду, квартили, размах квартилей. И лишь для данных, представленных в интервальной шкале или шкале отношений, применимы все описательные статистики, и то лишь при

определенных условиях, о которых мы скажем, когда речь пойдет о нормальном статистическом распределении.

Зачастую графическое представление данных является более наглядным, поэтому приведем для данной выборки некоторые возможности построения графиков в программе Statistica, построив так называемый «ящик с усами» (box and whiskers plot). Оговоримся, что графически можно изобразить практически любые из подсчитанных (и не только их) параметров, однако мы ограничимся построением среднего (в качестве точечной оценки) и стандартной ошибки в совокупности с доверительным интервалом. Для этого последовательно активируем вкладку Graphs (графики), 2D Graps (двумерные графики), затем — Box Plots (графики в виде «коробок») (рис. 3.12).

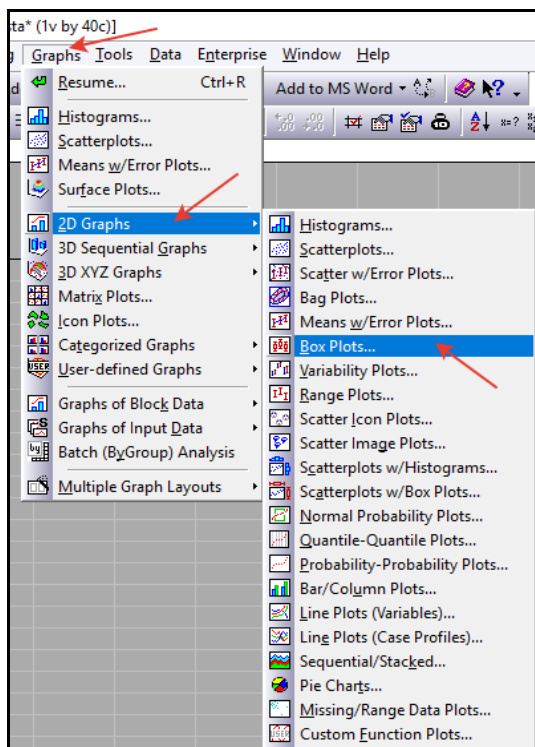


Рис. 3.12

Затем в открывшемся меню (рис. 3.13) выбираем стиль графика «ящик с усами», в качестве переменной указываем нашу переменную «понимание своих эмоций», в качестве точечной оценки — среднее, а в качестве интервальных — стандартную ошибку и доверительный интервал. Во вкладке Advanced доступны и другие варианты для выбора точечных и интервальных оценок (медиана, квартили, размах, среднеквадратичное отклонение и т. д.), каждое из которых, как мы увидели, имеет определенный смысл.

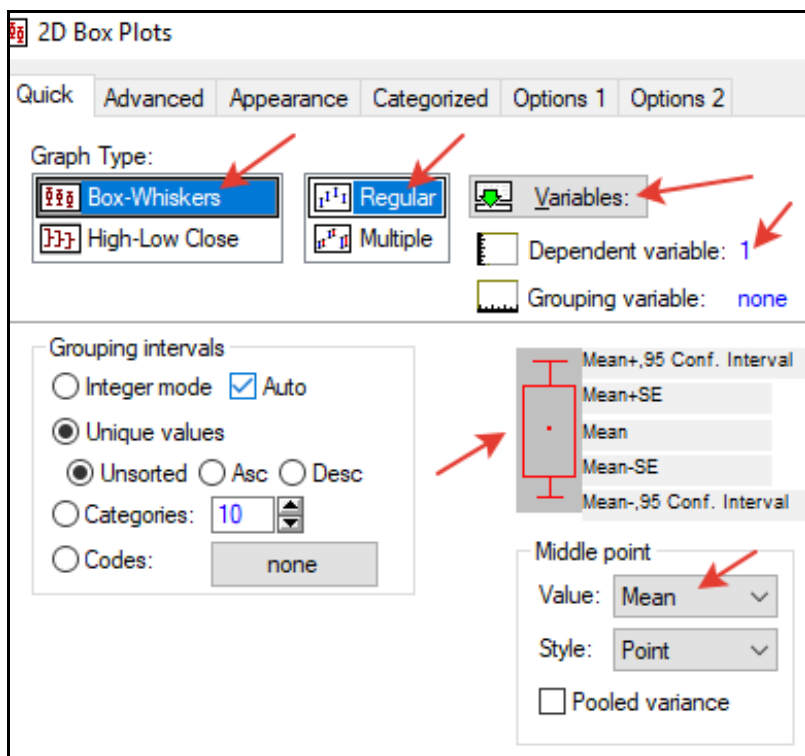


Рис. 3.13

Итак, мы получили итоговый график, на котором представлены искомые значения (рис. 3.14). Нам понадобится построение подобных графиков в дальнейшем при сравнении двух или более выборок.

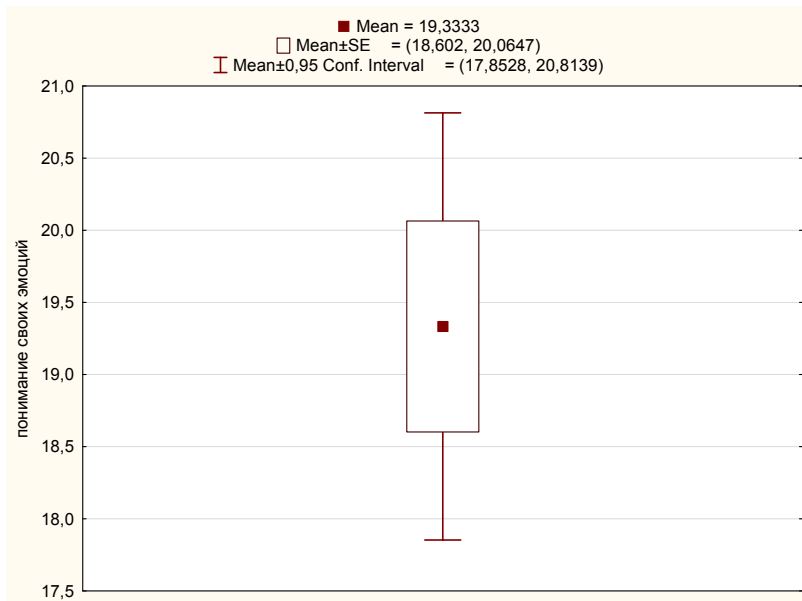


Рис. 3.14

Рассмотрим еще один часто встречающийся случай, когда распределение признака таково, что возможно только два исхода: «да» или «нет», или, говоря точнее, мы имеем дело с признаками, которые принимают всего два значения, и в совокупности все данные составляют полную систему событий. С помощью, например, *углового преобразования Фишера* (метода анализа) можно подсчитать значение доли признака (отношение количества случаев интересующего нас исхода к общему количеству случаев) и доверительные интервалы для этого частного случая.

Если данные представлены в виде дихотомии, то есть мы можем четко указать из общего числа испытуемых тех, кто обладает

одним свойством, и тех, кто обладает другим и только другим свойством и при этом не обладает первым, — у нас есть возможность определить точечные (доли) и интервальные (доверительные интервалы) оценки признака.

Важно отметить и то, что в дихотомических данных как ошибка, так и доверительный интервал в последнем примере рассчитываются иначе, нежели в данных, приведенных в предыдущем примере, поскольку в первом примере данные количественные, а во втором — качественные. Особенности обработки результатов, представленных в виде количественных и качественных данных, будут подробно обсуждаться далее.

Рассмотрим пример с помощью программы MedStat. Опять же воспользуемся вводом данных вручную, хотя можно загружать уже готовые данные, представленные в формате программы или вставлять данные, скопированные из табличного процессора.

После открытия программы выбираем раздел «Данные», после чего нажимаем на «Новые данные» и далее «Таблица: $k \times m$ » (рис. 3.15).

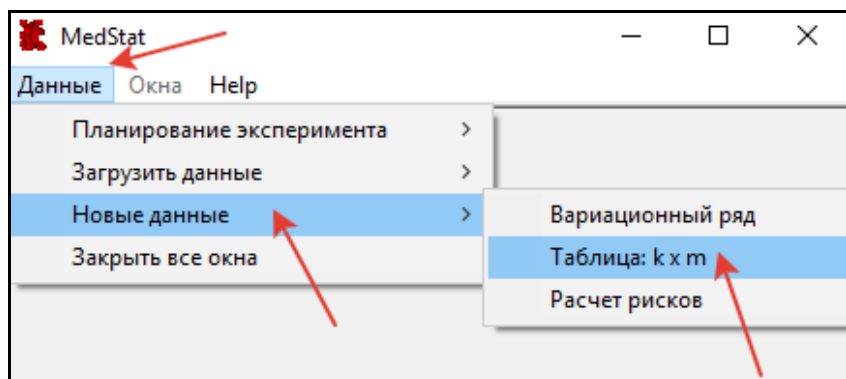


Рис. 3.15

Выбираем количество столбцов и строк. Для нашего примера нам понадобится две строки (количество градаций признака) и один столбец (количество переменных) (см. рис. 3.16).

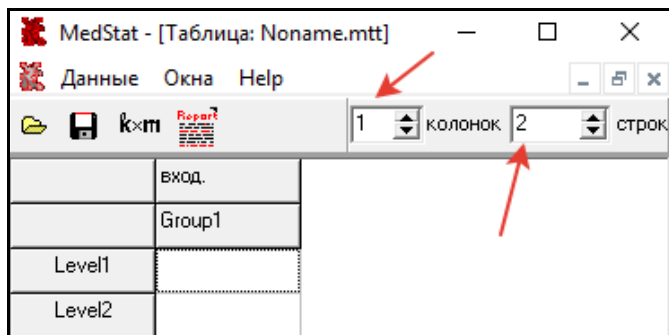


Рис. 3.16

Пусть в нашем примере выборку подростков из 50 человек после тестирования опросником ЕРІ Г. Айзенка разделили на экстравертов (35 человек) и интровертов (15 человек). Кликнув правой кнопкой мыши в указанное место (рис. 3.17), открываем место для ввода названия переменной. Level1 и Level2 — соответственно градации, которые принимает переменная. В нашем случае это количество экстравертов (35) и интровертов (15).

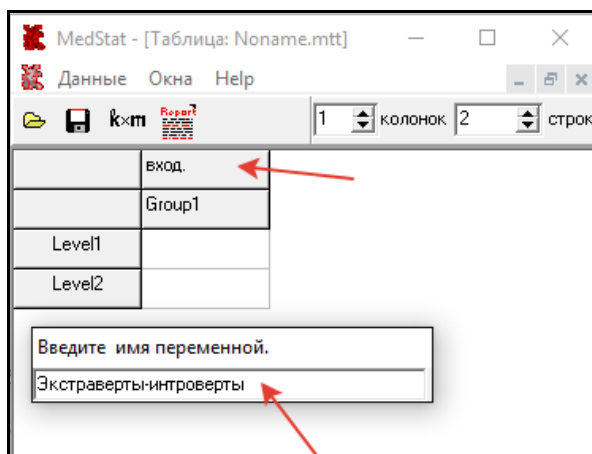


Рис. 3.17

Далее производим выбор типа анализа, нажав на меню «k x m», и выбираем пункт «Описательная статистика» (рис. 3.18).

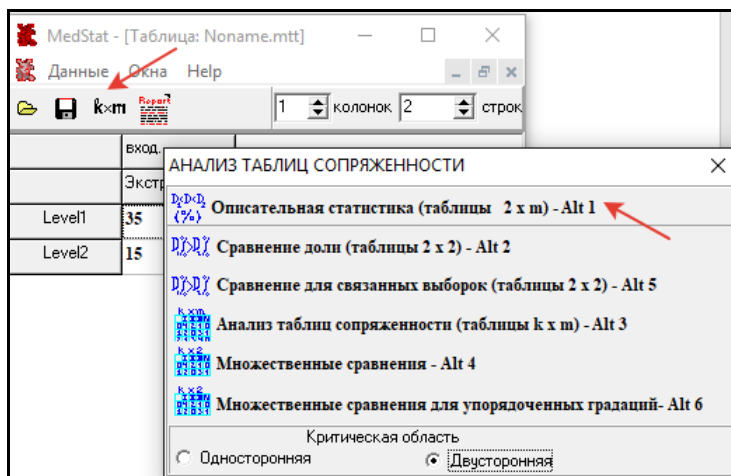


Рис. 3.18

Как видно, с помощью критерия «угловое преобразование Фишера» (определение критерию мы дадим ниже) можно найти точечную и интервальную оценку признака (рис. 3.19).

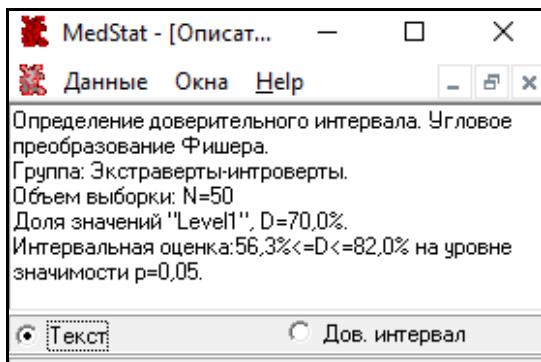


Рис. 3.19

Иными словами, при объеме выборке в 50 испытуемых, 70% из них — экстраверты, а доверительный интервал составляет от 56,3% до 82,0%, то есть следует ожидать, что 95% значений переменной «экстраверсия» попадут в соответствующий интервал. Если бы на первом месте располагалось количество испытуемых со свойством «интроверсия», мы получили бы аналогичные значения для данного свойства. Доверительный интервал можно изобразить и графически, активировав кнопку «Дов. интервал» (рис. 3.20).

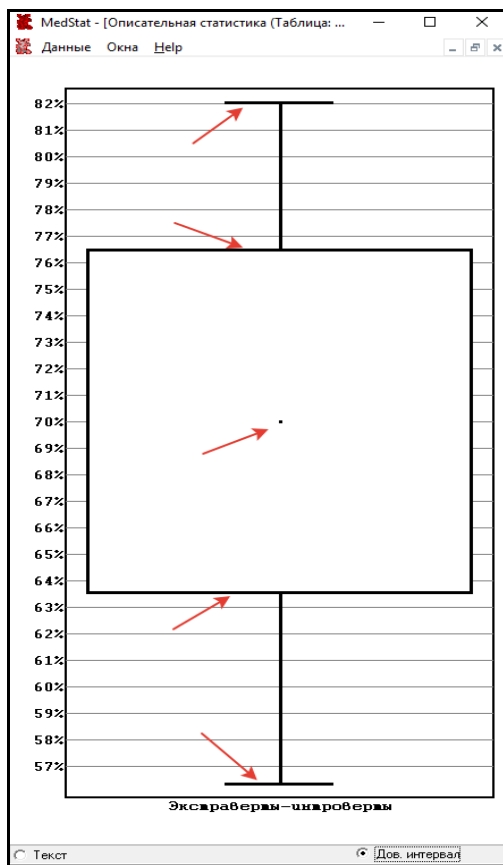


Рис. 3.20

На графике представлены точечная оценка распределения, а также интервальные оценки — стандартная ошибка (в виде «ящичка») и собственно доверительный интервал (в виде «усов»). Результаты, полученные в данном примере, ценны тем, что позволяют не только оценить саму долю испытуемых с данным свойством, но также и характер разброса значений относительно данной точечной оценки (доли), что позволяет получить дополнительную важную информацию о выборке. Если мы хотим добиться большей точности в оценке встречаемости данного свойства, нам следует увеличивать объем выборки.

Как мы упоминали выше, чем больше значений — тем более узок доверительный интервал, поскольку тем точнее мы приближаемся к его истинной оценке. Рассмотрим случай, когда объем выборки составляет не 50, а 50 000 человек, при этом соотношение между экстравертами и интровертами такое же. Нетрудно убедиться, что доверительный интервал существенно сузится и окажется в пределах от 69,6 до 70,4% для 50 000 испытуемых, а не от 56,3 до 82,0%, как для 50 испытуемых. Ширина доверительного интервала тем меньше, чем меньше стандартная ошибка, а величина последней убывает с ростом количества вариантов в выборке. Поэтому такими понятиями, как доверительный интервал и стандартная ошибка, следует оперировать крайне аккуратно, не забывая о том, что они «сужаются» при увеличении объема выборки.

4. СТАТИСТИЧЕСКИЕ КРИТЕРИИ И МЕРЫ СВЯЗИ ДЛЯ КОЛИЧЕСТВЕННЫХ И КАЧЕСТВЕННЫХ ДАННЫХ

Зачем нужна математическая статистика в психологии? Самый простой ответ будет заключаться в том, что без нее невозможно делать доказательные выводы о правильности тех или иных исследовательских гипотез.

В этом контексте рассмотрим элементарные ситуации, в которых исследователя может интересовать:

1. Поиск отличий, полученных в выборке значения признака в *одной выборке*, от некоторого заранее известного значения.

Например, освоенность навыка чтения у детей в начальной школе оценивают по скорости чтения текста в течение 1 минуты. В первом классе скорость чтения в норме должна быть не менее 60 слов в минуту. Всегда в классе есть дети, которые читают значительно быстрее, но также всегда есть дети, которые отстают в освоении навыка чтения, что затрудняет в комплексе их процесс обучения. Учителю важно знать, значительно ли отстают в освоении навыка чтения ученики его класса. К примеру, норма 60 слов, тогда 50 слов — значимое снижение скорости чтения от нормативного показателя?

Нас может интересовать, насколько полученное в эксперименте на выборке по исследованию успешно решенных задач определенного вида их доля (процент) будет больше или меньше значения, выявленного в гораздо большей по размеру нормативной выборке или генеральной совокупности, или относительно теоретически ожидаемой доли (процентного соотношения).

2. Отличия между собой точечных или интервальных параметров признака в *двух независимых выборках*. Эти выборки состоят из разных в отношении контролируемого фактора объектов, и при этом вероятность включения определенного объекта в исследование случайна.

Например, администрации школы необходимо понять, какими способностями обладают школьники, претендующие на поступление в профильные 10-е классы. Для этого может быть проведена процедура психологической оценки интеллекта тестом структуры интеллекта Р. Амтхауэра (TSI) для учащихся, приходящих в 10 класс и заявляющих об определенном выборе (к примеру, социогуманитарный или физико-математический профиль). Основной вопрос исследования: на самом ли деле в класс соответствующего профиля записываются учащиеся с выраженными вербальными или числовыми способностями?

Администрацию школы может интересовать задача оценки успеваемости различных классов не только по ее средним величинам (которые могут быть одинаковы), а насколько однородными являются классы по данному признаку. Так, в двух классах в среднем может быть одна и та же успеваемость (допустим, средний балл равен 3,5), но значения разброса в успеваемости могут различаться: в одном классе преобладают отличники и двоечники, а другой состоит почти полностью из троечников и небольшого количества хорошистов. Возможно, на этом основании администрация посчитает необходимым перераспределить учеников в классах так, чтобы в одном из них оказались дети с приблизительно одинаковой успеваемостью, с тем чтобы обучать их по наиболее эффективным методикам в одном и другом случаях.

3. Наличие сдвига в точечных или интервальных показателях признака, измеренного в *зависимых выборках* (чаще всего это одна и та же выборка, значения в которой повторно или многократно измеряются в определенных ситуациях, интересующих исследователя).

Например, группа студентов-психологов в самом начале обучения в университете проходит социально-психологический тренинг, задача которого — сплочение участников группы, облегчение процесса адаптации к обучению в вузе. Цель исследователя — оценить (измерить) качество сплоченности одной и той же методикой в этой группе до начала тренинга и после его завершения (обычно оценивают не сразу после завершения, чтобы избежать эффекта

посттренинговой эйфории, а примерно через недели две). Ожидаемый результат: увидеть наличие/отсутствие различий по показателю сплоченности.

Исследователя может интересовать уровень выраженности экзаменационного стресса студентов. Для этого соответствующие признаки измеряются в «спокойном периоде» обучения и непосредственно во время экзаменов, то есть на тех же испытуемых, но в разных ситуациях. Исходя из этого, можно далее исследовать причины соответствующих сдвигов в состоянии и принимать решение о необходимости тех или иных мероприятий, направленных на коррекцию состояния испытуемых.

4. Существование связи между двумя или несколькими показателями, измеренными на одной и той же выборке.

Например, исследователь задумал оценить связь между отсутствием нарушений дисциплины и уровнем развития интеллекта у школьников. Нарушения дисциплины будут фиксироваться как объективный показатель — в количестве дисциплинарных нарушений по данным социально-педагогической службы школы. Уровень развития интеллекта будет оцениваться по результатам выполнения соответствующего теста (например, теста структуры интеллекта Р. Амтхауэра, теста прогрессивных матриц Дж. Равена и т. д.).

Перед исследователем может стоять задача определения направления и выраженности связи между удовлетворенностью профессией у работников (определенное число испытуемых X_1 — удовлетворено, а другое X_2 — не удовлетворено) и производительностью труда (Y_1 — имеют высокую производительность труда, а Y_2 — низкую).

5. Зависимость одного показателя (результатирующего) от одного или нескольких (факторных) в одной и той же выборке.

Например, академическая успешность студентов может складываться из нескольких факторов: мотивированности к обучению, уровня развития интеллекта, установления эмоционально позитивного контакта с преподавателем, конфликтов с сокурсниками, семейными проблемами и т. д. Исследователь хотел бы выяснить,

какие из этих факторов вносят наибольший вклад в академическую успешность (как положительный, так и отрицательный).

6. Несоответствие полученного в эксперименте распределения признака или признаков распределению, полученному в другой выборке, или теоретическому распределению.

Например, нас может интересовать несогласованность предпочтений в режимах питания сотрудников, работающих на предприятии и страдающих желудочно-кишечными болезнями, и здоровыми людьми из гомогенной выборки (согласованной по возрасту, полу).

Исследователь может задаться вопросом, нет ли в исследуемой группе значительных отклонений (так называемых «выбросов») по показателям теста интеллекта. Чаще всего в однородной группе максимальное количество испытуемых со средними результатами и все убывающее их число в сторону лучших и худших значений. Однако может быть так, что в группе окажется некоторое количество испытуемых с аномально низкими (высокими) показателями, не характерными для группы в целом.

Нас может интересовать, насколько распределение результатов по тесту интеллекта в данной выборке соответствует популяционным нормативам, которые также характеризуются не определенным числом, а набором значений, средние из которых встречаются чаще, а все большие и все меньшие значения — реже. В эксперименте мы сможем наблюдать несоответствие или соответствие показателей распределения.

Очень важный случай, который будет обсуждаться в дальнейшем, — это определение соответствия или несоответствия имеющихся данных, выстроенных (по возрастанию) в виде вариационного ряда, закону определенного статистического распределения, прежде всего, нормального, которому придается особое значение не только в психологии, но и во всех науках.

4.1. Виды статистических гипотез.

СТАТИСТИЧЕСКАЯ ЗНАЧИМОСТЬ. СТАТИСТИЧЕСКИЙ КРИТЕРИЙ

Применяя методы математической статистики в психологии, исследователю следует в общем случае придерживаться следующей последовательности этапов проработки данных:

1. Психолого-математическая интерпретация данных. На этом этапе исследователь переводит психологические данные и/или гипотезы в статистические, с которыми он в дальнейшем будет работать.

2. Математико-статистическая обработка данных. На данном этапе исследователь обрабатывает математико-статистические данные с тем, чтобы получить информацию в более сжатом и понятном виде и/или получить доказательные сведения о том, какая из гипотез верна.

3. Математико-психологическая интерпретация результатов. На этом этапе психолог должен с помощью полученных математико-статистических результатов ответить на те психологические вопросы, гипотезы, которые изначально выдвигались, то есть перевести результаты с «языка математики» обратно на «язык психологии».

Данный этап далеко не так прост, как кажется, вследствие возможных спекуляций, чрезмерных обобщений на основе ограниченной выборки, вследствие непонимания самой сути математических методов, выводы которых сильно зависят от количества испытуемых и могут быть плохо воспроизводимы в последующих исследованиях, и т. п.

Еще одна распространенная ошибка — это подача в качестве результата психологического исследования его математического аналога, когда исследователь не учитывает принципиальное различие между психологическими и математико-статистическими гипотезами.

Теперь более подробно рассмотрим эти этапы и их содержание. Итак, на первом этапе исследователь чаще всего вносит в таблицу полученные данные или, в случае необходимости, присваивает им некоторые условно числовые или символичные значения (в частно-

сти, если признак измерен в шкале наименований). Затем психологические гипотезы переводятся на язык математической статистики. Чаще всего мы имеем дело с нулевой и альтернативной гипотезами.

Нулевая гипотеза (H_0) — гипотеза об отсутствии различий (или связей, если речь идет о корреляции, регрессии и т. п.).

Альтернативная гипотеза (H_1) — гипотеза о наличии различий (или связей, если речь идет о корреляции, регрессии и т. п.).

Следует различать направленные и ненаправленные гипотезы.

Например, *направленные гипотезы*: 1) высокий уровень интеллекта связан с высоким уровнем академической успеваемости; 2) низкий уровень тревожности связан с высоким уровнем успешности в социальных контактах; 3) агрессивность подростков из неполных семей выше, чем их сверстников из полных семей; 4) проблемы с коммуникациями в реальном мире со сверстниками выше у детей, злоупотребляющих компьютерными играми.

Ненаправленные гипотезы: 1) агрессивность ребенка зависит от типа взаимоотношений с родителями; 2) формирование познавательных процессов младшего школьника связано с установлением контакта с учителем; 3) частота посещения музеев и театров отличается у старшеклассников мужского и женского пола; 4) имеются изменения в самооценке нового сотрудника через неделю, месяц, год после пребывания в новом коллективе по сравнению с той, что была до поступления на работу.

Таким образом, при формулировке направленных гипотез предполагается и проверяется лишь одно направление различий признаков, а при формулировке ненаправленных — изначально допускаются различия признака, как в сторону большего, так и меньшего значения относительно другой выборки или сдвига признака.

Далее на основе таблиц, содержащих результаты исследований, в зависимости от этапа обработки, строятся распределения признаков в виде гистограмм, находятся описательные статистики. На данном этапе мы вместо множества исходных чисел имеем дело уже с одним или несколькими числами, характеризующими признак

в выборках. Например: среднее арифметическое, доверительные интервалы, стандартные отклонения и проч. Иногда для анализа данных в выборке нам вполне достаточно этих описательных статистик или наглядного представления данных на гистограммах (в основе которых тоже, естественно, числа).

В большинстве случаев, однако, далее применяется тот или иной *статистический критерий* как определенный способ обработки данных, обеспечивающий исследователя статистически обоснованным ответом на вопрос о том, какую гипотезу следует принять — нулевую или альтернативную.

Как же понять, какую гипотезу следует принять, а какую — отвергнуть, применяя статистические критерии? Для этого вводятся понятия *статистическая значимость*, обозначаемая буквой p , и *ошибка первого рода* (обозначается как буква «альфа» греческого алфавита — α). В определенном приближении это почти одно и то же: вероятность ошибочно принять альтернативную гипотезу, в то время как на самом деле верна нулевая. Как правило, граничный уровень статистической значимости, начиная с которого отвергается нулевая гипотеза и принимается альтернативная, принимают за 0,05 в терминах вероятности (или в процентах — 5%). Следует лишь отметить, во избежание путаницы, что статистическая значимость тем выше, чем она более приближается к нулю. Условно принято считать, что $p < 0,05$ соответствует статистически значимому результату, $p < 0,01$ — высоко значимому, $p < 0,001$ — очень высоко значимому. Если же $p \geq 0,05$, то результат является статистически незначимым, и мы не можем отвергнуть нулевую гипотезу.

Ошибка второго рода (бета, β), наоборот — вероятность ошибочно принять нулевую гипотезу, в то время как верна альтернативная. С понятием ошибки второго рода напрямую связано понятие *статистической мощности* (из единицы вычесть бета) — способности с помощью данного критерия и с данной выборкой или выборками при заданном уровне значимости доказать альтернативную гипотезу: чем больше мощность, тем больше эта способность. Обычно на практике мощность должна составлять не менее

80% (0,8 — в терминах вероятности), то есть в 80% случаев при фактической верности альтернативной гипотезы мы можем это доказать с помощью критерия. Иногда в качестве граничной мощности выбирают 90%, 95%, что, однако, предъявляет более строгие требования к размеру выборки.

Следует понимать, что в результате расчета статистического критерия мы получаем некоторое число (эмпирическое значение критерия), которое может быть сопоставлено с заранее рассчитанными числами — критическими значениями для данного критерия, соответствующими определенным уровням статистической значимости и количеству степеней свободы. Количество степеней свободы (часто обозначается как df) определяется, в сущности, числом значений или интервалов в выборках. Для расчета критерия вручную, когда находится его значение, затем сопоставляется с табличным, без подсчета количества степеней свободы (элементарного математического действия), дальнейшие действия по принятию истинной статистической гипотезы невозможны. Между тем, соотношение эмпирического и критического значений критерия позволяет принять одну из статистических гипотез (нулевую или альтернативную). В ряде случаев исследователь волен, исходя из гипотезы исследования, решить, применять ли ненаправленную или направленную гипотезу (соответственно, при расчете критерия использовать *двустороннюю* или *одностороннюю* критическую область). Если критическая область односторонняя, то при одном и том же значении критерия величина ошибки первого рода в два раза меньше, чем для двусторонней критической области.

Тем не менее, в большинстве случаев все же целесообразно пользоваться двусторонней критической областью, особенно когда мы заранее не можем с достаточным на то основанием предположить, в какую сторону друг относительно друга должны разниться значения в выборках или происходить их сдвиг. Например, если мы исследуем влияние употребления наркотиков подростками на открытость коммуникации с родителями, мы вправе предполагать, что следует воспользоваться направленным критерием (почти наверняка употребление наркотиков будет негативным фактором,

вопрос лишь в том — насколько). Мы просто-напросто не проверяем гипотезу «с другой стороны», когда употребление наркотиков улучшало бы открытость коммуникации. Тогда при меньшем объеме выборки мы статистически достоверно установим наличие данной связи. С другой стороны, если мы, например, исследуем влияние табакокурения среди педагогов высшей школы на успешность их профессиональной деятельности, то ответ вряд ли будет столь однозначным. В таком случае следует воспользоваться ненаправленной гипотезой, а следовательно, применять двустороннюю критическую область.

Как правило, в интерфейсе программ по статистике предлагается выбрать учет одно- или двусторонней критической области. Например, в программе «Медстат» в меню запуска метода обработки данных (имеются в виду критерии) сразу появляется опция выбора одно- или двусторонней критической области (рис. 4.1).

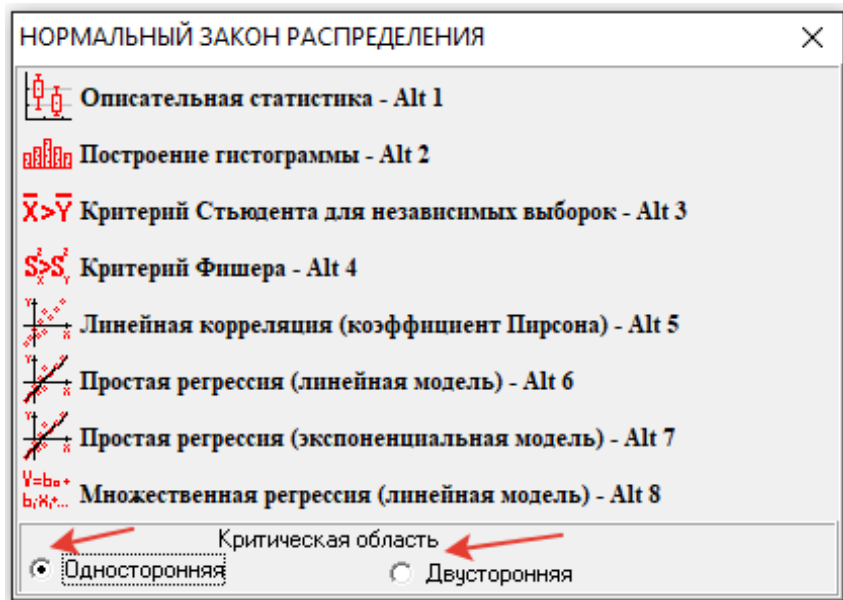


Рис. 4.1

Вопрос о выборе одно- или двусторонней критической области может стать решающим при принятии определенной статистической гипотезы. Например, при применении критерия с выбором двусторонней критической области мы можем получить статистически незначимый результат (например, $p = 0,065$) и принять нулевую гипотезу, но если выберем одностороннюю критическую область, то уровень значимости составит уже $p = 0,032$ — и принята будет альтернативная гипотеза о наличии неслучайных различий, сдвигов или связей между переменными или измерениями.

Следует отметить, что при выдвижении ненаправленных гипотез (с двусторонней критической областью) могут потребоваться большие объемы выборок для получения доказательства правильности альтернативной гипотезы при применении любого критерия. Зато при применении двусторонней критической области мы фактически доказываем обе альтернативные направленные гипотезы — в обе стороны, если они верны. Например, если мы предполагаем, что эмоциональный интеллект выше у женщин, чем у мужчин (направленная гипотеза), то при односторонней критической области мы изначально в принципе не допускаем противоположного (что эмоциональный интеллект выше у мужчин). Если же мы применяем двустороннюю критическую область, то мы изначально предполагаем две возможности и можем доказать ту или иную из них.

На этапе математико-психологической обработки результатов следует грамотно интерпретировать, как именно результаты, выраженные в числовой форме, отвечают на вопросы, которыми изначально задавался исследователь.

Мы уже касались вопроса о том, что для доказательства неслучайности результата (правильности альтернативной гипотезы) необходимо показать, что такие результаты могут быть случайно получены в менее чем 5% случаев, то есть в терминах вероятности, $p < 0,05$.

Перед исследователем всегда стоит непростая задача: как рассчитать количество анализируемых объектов так, чтобы, с одной стороны, обеспечить доказательность исследования и с достаточным на то основанием принять истинную гипотезу, с другой — сэкономить время и силы на сбор данных, поскольку с ростом объ-

ектов выборки увеличивается и статистическая значимость при том же размере эффекта.

С одной стороны, исследование может быть проведено впустую, показав большие численные различия между изучаемыми признаками в выборках, но при этом не обеспечив статистическую значимость соответствующих результатов из-за малых объемов выборок или большого разброса значений в выборках.

С другой стороны, можно потратить много усилий, средств, времени на сбор и обработку данных, в итоге получив тот же результат, что был бы получен и на значительно меньших по объему выборках (и достаточный в отношении статистической значимости). Или того печальнее, получить на больших группах статистически значимые различия, но не имеющие значения для целей исследования.

Далее мы увидим, как, опираясь на более строгие расчеты, правильно выбрать объем выборок для подтверждающего исследования. Пока предварительно сформулируем некоторые эмпирические, но определенным образом базирующиеся на теории вероятностей правила для эксплораторного или подтверждающего исследования в том случае, если будут проведены аналогичные повторные исследования (или проводились ранее этим или другими исследователями) для прояснения вопроса о воспроизводимости результатов:

- желательно, чтобы количество объектов в независимых выборках было одинаковым, насколько это возможно в данных условиях (в зависимых — их количество одинаково по определению);

- при сравнении двух выборок их общая численность должна быть 50 человек и более;

- при изучении взаимосвязи между какими-либо свойствами — не меньше 30-35 человек в выборке;

- при разработке диагностической методики объем выборки наибольший — от 200 до 1000-2500 и более человек (здесь вопрос о проведении повторных измерений на других выборках решается исходя из свойств генеральной совокупности, релевантной диагностической методике).

Следует различать достоверность и статистическую значимость. Достоверность относится к способу измерений, получению точных результатов.

Кроме того, измеряемый признак должен соответствовать тому свойству, которое мы изучаем. Так, изучая тревожность с помощью соответствующего опросника и получая некие значения, мы получаем отнюдь не значения тревожности, а именно баллы по опроснику, которые еще неизвестно как именно связаны с тревожностью в данном случае либо вообще в принципе, если вдруг исследователь выбрал неверный инструментарий для измерения изучаемого свойства. В отличие от школьной физики, где мы точно знаем, например, что линейкой мы измеряем именно длину отрезка и достоверность измерений зависит только от точности самой линейки, в психологии, говоря метафорически, во-первых, все линейки неточные как в отношении соответствия длин отрезков, так и в отношении соответствия измеряемого признака изучаемому свойству. Последняя ситуация может произойти, когда, продолжая наш пример, с помощью линейки мы пытаемся измерить массу.

4.2. ПОНЯТИЕ О СТАТИСТИЧЕСКОМ РАСПРЕДЕЛЕНИИ ДАННЫХ

Мы уже упоминали некоторые базовые понятия статистики в психологии (типы шкал, описательные статистики, понятие о критериях и уровне значимости и т. п.). Мы упоминали также и о понятии вариационного ряда, что имеет непосредственное отношение к распределению данных. Допустим, из генеральной совокупности извлечена ограниченная по объему выборка. Если все полученные данные одной переменной расположить по возрастанию от минимального до максимального, то мы получим вариационный ряд. Отдельные наблюдаемые значения вариационного ряда одной переменной называются вариантами и обозначаются как $x_1, x_2, x_i \dots x_n$, где n — количество вариантов в выборке. Если одна и та же варианта встречается более чем в одном случае, то ее частота будет равняться не единице, а тому числу, сколько раз она встретилась: 2, 3 и т. д. При этом относительной частотой встре-

чаемости варианты будет являться отношение ее частоты к общему количеству числа наблюдений. Например, если одно и то же значение встречается пять раз в выборке, состоящей из 100 вариантов, то относительная частота будет равняться $5 / 100 = 0,05$.

Если построить график, который будет показывать на оси абсцисс (горизонтальной оси) значение показателя, а по оси ординат (вертикальной оси) частоту его встречаемости, то мы увидим, как распределяются данные, то есть это и будет статистическим распределением данных. Форма и вообще особенности распределения крайне важны, ведь они являются наиболее наглядной характеристикой данных и зачастую сразу дают больше наглядной (пусть и не столь строгой) информации о выборке, чем все описательные статистики. Поэтому визуальное рассмотрение распределения данных является необходимым этапом их анализа.

Наиболее часто применяемым в психологии представлением распределения данных является гистограмма.

4.2.1. Построение гистограмм

Гистограмма — это наглядное представление распределения статистических данных, где по оси абсцисс находятся интервалы значений вариационного ряда признака, а по оси ординат — абсолютные или относительные частоты вариант, попадающих в соответствующий интервал.

Вообще говоря, построение статистического распределения «вручную» — это хоть и не сверхсложная, но довольно трудоемкая задача, особенно в условиях, когда статистические пакеты прекрасно справляются с этой задачей. Покажем на примере, откуда берется гистограмма.

Пусть пятьдесят учащихся решали тестовую задачу, при этом измерялось время, затраченное на правильное решение. Исходные данные мы опускаем ради экономии места. По некоторым правилам, с которыми можно ознакомиться, например, в пособии Остапенко [12], была построена таблица, представляющая результат подсчета сгруппированных частот по разрядам (интервалам) значений признака «время решения тестовой задачи» (табл. 4.1).

**Группировка частот по разрядам значений признака
«время решения тестовой задачи»**

| <i>Интервал времени, с</i> | <i>Абсолютная частота</i> | <i>Относительная частота</i> | <i>Накопленная частота</i> |
|----------------------------|---------------------------|------------------------------|----------------------------|
| 30-34 | 2 | 0,04 | 0,04 |
| 35-39 | 3 | 0,06 | 0,10 |
| 40-44 | 7 | 0,14 | 0,24 |
| 45-49 | 12 | 0,24 | 0,48 |
| 50-54 | 11 | 0,22 | 0,70 |
| 55-59 | 8 | 0,16 | 0,86 |
| 60-64 | 5 | 0,10 | 0,96 |
| 65-69 | 2 | 0,04 | 1 |
| Σ (сумма) | 50 | 1,000 | — |

Видно, что в каждом из интервалов содержится определенное количество значений (абсолютная частота), при этом также видно, какую долю занимают значения в каждом интервале (относительная частота), а также — какая часть вариационного ряда преодолена к определенному интервалу (накопленная частота). Путем суммирования убеждаемся, что всего у нас действительно 50 вариантов, которые в сумме частот дают единицу в терминах вероятности (то есть 100%).

Важно отметить, что границы интервалов (разрядов), вообще говоря, не пересекаются. Например, первый интервал у нас от 30 до 34 секунд, а второй — от 35 до 39 и т. д. Однако здесь мы измеряли время целыми значениями. Если же представить, что время измерялось не с точностью до секунд, а с точностью до миллисекунд, например, 34,999 секунды, а следующее значение было бы 35,000 секунд, что имеет место быть, например, в спорте высоких достижений, то, очевидно, интервалы бы обозначались как 30-35 и 35-40, но мы бы помнили, что все равно границы интервалов не пересекаются.

Для приведенного выше примера гистограмма относительных частот представлена на рис. 4.2.

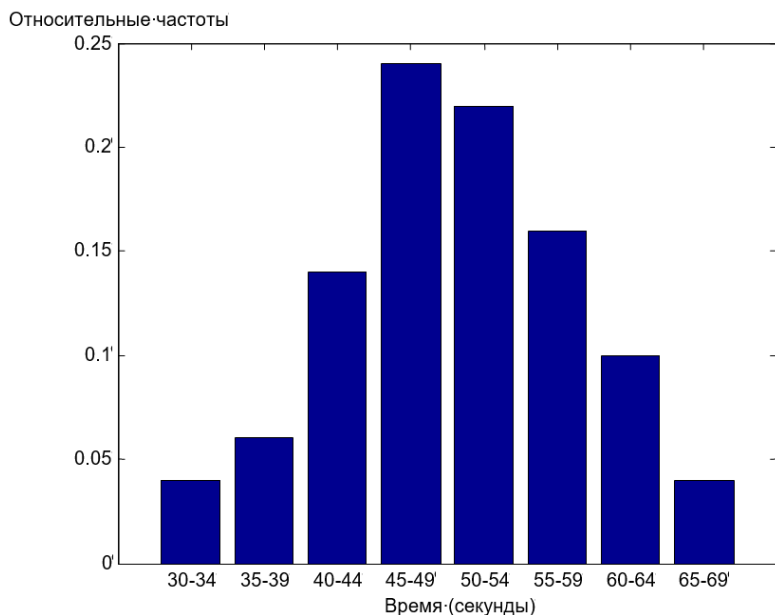


Рис. 4.2

В некоторых статистических пакетах (например, в «Медстат»), программа автоматически просчитывает число разрядов-интервалов, а в некоторых (Statistica), мы, кроме того, и самостоятельно можем задавать желаемое число интервалов.

При построении гистограмм рекомендуется пользоваться следующими правилами:

1. Число интервалов выбирается в зависимости от числа наблюдений согласно рекомендациям табл. 4.2.

2. Длины интервалов гистограммы, как правило, выбираются одинаковыми. Однако если распределение крайне неравномерно, то в области максимальной концентрации результатов наблюдений следует выбирать более узкие интервалы.

**Соответствие числа наблюдений и числа интервалов
разбиения гистограммы**

| <i>Количество вариантов или наблюдений (n)</i> | <i>Количество интервалов (разрядов) гистограммы (r)</i> |
|--|---|
| 40–100 | 7–9 |
| 100–500 | 8–12 |
| 500–1000 | 10–16 |
| 1000–10000 | 12–22 |

Иногда строят полигон частот, который отличается от гистограммы лишь внешним видом, когда вместо столбцов ставятся точки, объединяемые линиями, что, на наш взгляд, не столь наглядно.

**4.2.2. Роль нормального распределения
в психологических исследованиях.**

Проверка данных на нормальность распределения

На основании нормального закона статистического распределения разрабатываются тестовые шкалы и методики, работают наиболее информативные и мощные статистические критерии. Суть закона состоит в том, что в вариационном ряду чаще встречаются значения, близкие к среднему значению, а чем больше отклонения от него, тем реже встречаются такие значения.

Большинство свойств (характеристик, показателей психологических, биологических и т. д.) имеют одну и ту же «колоколообразную» форму, при условии, что этот показатель является результатом вклада (влияния) многих факторов, ни один из которых не является определяющим. Действительно, чаще всего так и происходит: как в генетике (например, при аддитивной полигении), в биологии (распределение роста, веса, продолжительности жизни), так и в психологии (например, скорость реакции, проявление различных способностей, характеристик темперамента, интеллекта и т. п.).

Важно еще раз подчеркнуть, что нормальное распределение описывает, прежде всего, генеральную совокупность, и то с ого-

ворками. На практике же при формировании выборок об этом можно с легкостью забыть. Это приведет к ошибкам, если в анализируемую выборку будут входить подгруппы, существенно отличающиеся по определяющим тот или иной параметр фактору или факторам. Например, распределение коэффициента интеллекта всех школьников России будет иметь формулу нормального распределения. Однако если составить выборку из подгрупп школ для умственно отсталых детей и одаренных, то нормального распределения уже не будет наблюдаться, поскольку мы увидим два максимума (средних арифметических для каждой из подгрупп) в частотах признака: один — для умственно отсталых (например, IQ = 80), другой — для одаренных (например, IQ = 130).

Еще один пример. Распределение роста отдельно мужчин и отдельно женщин будут подчиняться закону нормального распределения, однако в совокупности они будут давать опять же «двугорбое» распределение, как и в предыдущем случае, поскольку средний рост женщин меньше, чем мужчин. В этом легко убедиться практически.

Математически закон нормального распределения был сформулирован де Муавром:

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

где $f(x_i)$ — высота подъема кривой (фактически — вероятность); e — специальная константа — основание натурального логарифма (2,72); M и σ — среднее арифметическое и стандартное отклонение для переменной x .

Примечательно, что лишь два параметра: среднее (M) и стандартное отклонение (σ) *полностью* описывают особенности формы нормального распределения. Среднее определяет положение центрального, наиболее часто встречающегося значения на оси значений (абсцисс). Стандартное отклонение определяет ширину этой кривой и выступает в качестве масштаба измерения.

На рис. 4.3 показано семейство нормальных кривых, 1-е распределение отличается от 2-го стандартным отклонением ($\sigma_1 < \sigma_2$), 2-е от 3-го средним арифметическим ($M_2 < M_3$).

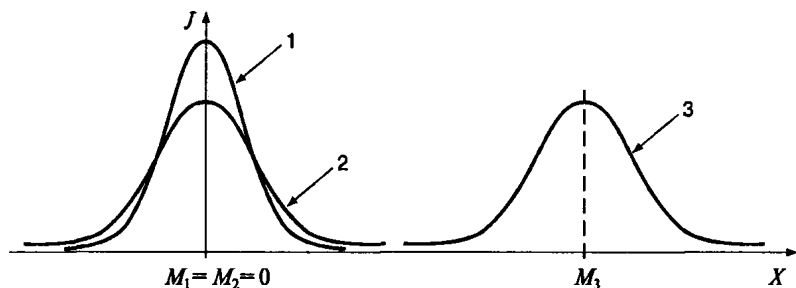


Рис. 4.3

Все многообразие нормальных распределений может быть сведено к одной единственной кривой, если применить z -преобразование.

Стандартизация, или z -преобразование данных, — это перевод измерений в стандартную Z -шкалу со средним $M = 0$ и стандартным отклонением $\sigma = 1$. Сначала для переменной, измеренной на выборке, вычисляют среднее M , стандартное отклонение σ . Затем все значения переменной x_i пересчитываются по формуле:

$$z_i = \frac{x_i - M_x}{\sigma_x}.$$

Индексом i обозначена конкретная варианта (значение), а индексом x — среднее и стандартное отклонения исходной выборки.

В результате преобразованные значения (z -значения) непосредственно выражаются в единицах стандартного отклонения от среднего. Если для одной выборки несколько признаков переведены в z -значения, то появляется возможность сравнения уровня выраженности разных признаков у того или иного испытуемого. Фигурально выражаясь, можно сравнивать «литры с килограммами», поскольку в z -шкале все признаки приводятся к единому масштабу при условии их соответствия нормальному распределению.

Подытожим некоторые свойства нормального распределения:

1. В нормальном распределении среднее арифметическое, медиана и мода совпадают.

2. Нормальное распределение симметрично относительно среднего значения.

3. Для описания формы нормального распределения достаточно всего лишь двух параметров — среднего арифметического и стандартного отклонения.

4. Кривая приближается к оси по краям асимптотически — никогда не касаясь ее.

Для того чтобы избавиться от отрицательных и дробных значений, можно для удобства перейти к любой другой известной шкале: IQ (среднее 100, стандартное отклонение 15); T -оценок (среднее 50, стандартное отклонение 10), 10-балльной — стенов (среднее 5,5, стандартное отклонение 2) и т. д. (рис. 4.4 [9]).

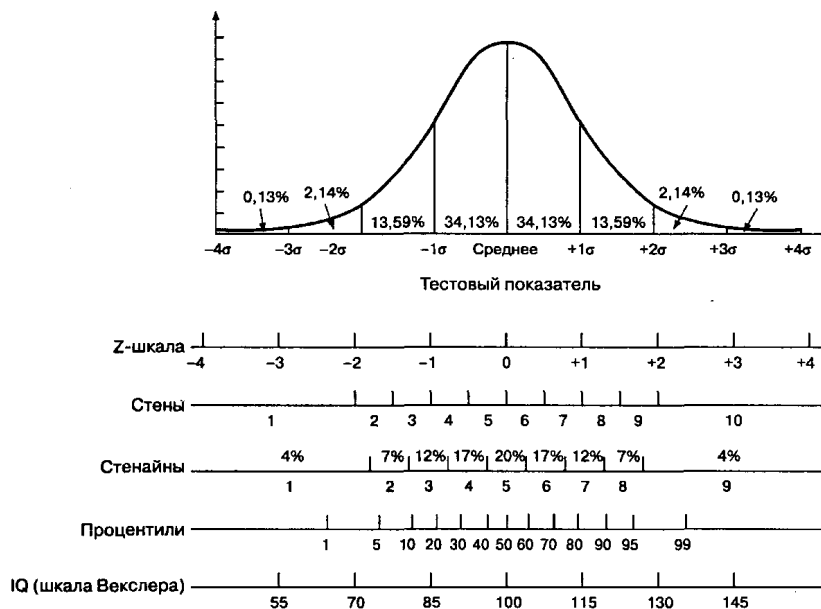


Рис. 4.4

Перевод в новую шкалу из z-значений осуществляется путем умножения каждого z-значения на заданное стандартное отклонение и прибавление среднего: $S_i = \sigma_s z_i + M_s$.

Существует специальная таблица, позволяющая определять вероятность встречаемости значений признака из любого диапазона z-значений, представленная во многих пособиях по статистике или онлайн (например, <https://www.resourcesystemsconsulting.com/2010/05/20/z-table>).

Такого рода таблицы или их автоматизированные аналоги (<http://www.fourmilab.ch/rpkp/experiments/analysis/zCalc.html>) широко используются при интерпретации данных тестирования, прежде всего, в задачах — насколько часто встречаются значения больше или меньше данного, переведенные в зет-шкалу.

Проверка данных на нормальность является чрезвычайно важной частью обработки данных психологического исследования. Ведь именно ее результаты будут определять, какие именно описательные статистики и статистические критерии использовать далее.

Приведем пример построения гистограммы и проверки данных на нормальность распределения по реальным данным, полученным в результате измерения случайной выборки 200 взрослых людей коэффициента интеллекта (в программе Statistica).

На рис. 4.5 представлены первые 11 значений признака, всего их 200.

В меню Statistics выбираем Basic Statistics and Tables, Descriptive Statistics, открываем вкладку Normality. Далее производим действия, помеченные стрелочками, и нажимаем на Summary (см. рис. 4.6).

В результате получаем гистограмму (см. рис. 4.7) распределения переменной с количеством интервалов в данном случае равным 10, ожидаемая кривая нормального распределения показана кривой, вверху графика результаты проверки на нормальность с помощью критериев Колмогорова–Смирнова, Лилифорса и Шапиро–Уилка. По всем критериям видим, что $p > 0,05$, это означает, что нулевая гипотеза (о соответствии полученного распределения нормальному) не может быть отвергнута.

| STATISTICA - [Data: пример с... | |
|--------------------------------------|-----|
| File Edit View Insert F... | |
| | |
| <div> <div></div> <div></div> </div> | |
| | 1 |
| | IQ |
| 1 | 91 |
| 2 | 122 |
| 3 | 109 |
| 4 | 102 |
| 5 | 104 |
| 6 | 131 |
| 7 | 107 |
| 8 | 119 |
| 9 | 110 |
| 10 | 93 |
| 11 | 129 |

Рис. 4.5

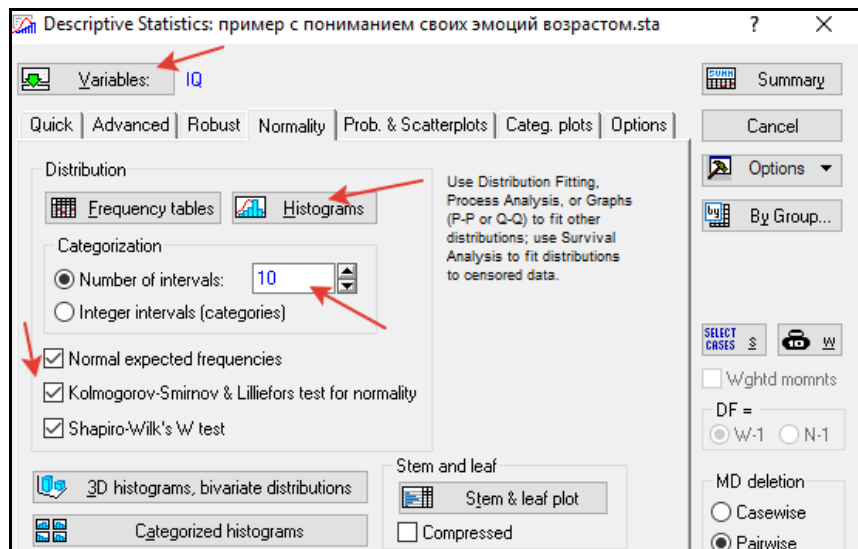


Рис. 4.6

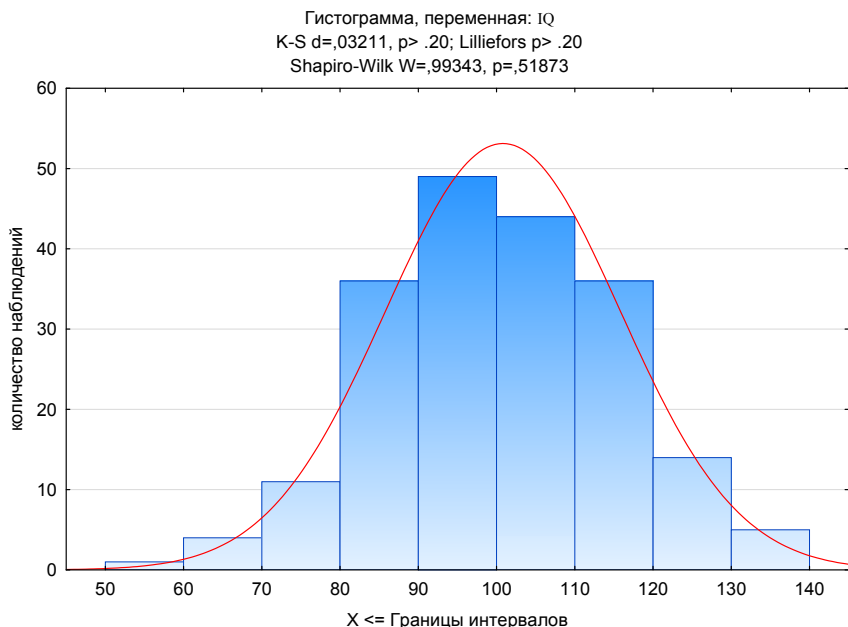


Рис. 4.7

Отметим, что при малых выборках (менее 50 вариантов) для проверки на нормальности лучше всего работает критерий Шапиро–Уилка как наиболее чувствительный.

В программе «Медстат» проверка на нормальность организована в автоматическом режиме, сразу с учетом того, какой критерий лучше подходит для данного количества вариантов в выборке.

Приведем следующий пример: для психотерапии эмоционального выгорания были применены три различные методики, представлены результаты терапии в некоторых экспертных оценках (см. рис. 4.8).







| | | | | |
|---|---------|---------|---------|----------|
| <div>  Данные Окна Help </div> <div>   N=?    </div> | | | | |
| | вход. | вход. | вход. | вход. |
| | Метод 1 | Метод 2 | Метод 3 | Здоровые |
| 1 | 3,6 | 4,8 | 4,2 | 4 |
| 2 | 2,6 | 3,6 | 6,4 | 3,3 |
| 3 | 6,4 | 8,1 | 5,9 | 4,4 |
| 4 | 4,1 | 7,1 | 4,3 | 5,1 |
| 5 | 4,1 | 6,5 | 4,7 | 5 |
| 6 | 4 | 4,3 | 5 | 5,4 |
| 7 | 5,9 | 4,9 | 6,2 | 2,7 |
| 8 | 5,2 | 4,2 | 5,9 | 4 |
| 9 | 3,6 | 6,9 | 6,6 | 5 |
| 10 | 3,4 | 6,8 | 5,7 | 3,4 |
| 11 | 4,4 | 7,8 | 5,6 | 3,7 |
| 12 | 3,5 | ? | 3 | 3 |
| 13 | 3,5 | ? | 4,5 | 2,9 |
| 14 | ? | ? | 3,7 | 3,5 |
| 15 | ? | ? | 3,5 | 3,7 |
| 16 | ? | ? | 4,2 | 2,7 |
| 17 | ? | ? | ? | 3,8 |
| 18 | ? | ? | ? | 3,9 |
| 19 | ? | ? | ? | 4,3 |
| 20 | ? | ? | ? | 3,9 |

Рис. 4.8

Заметим, что контрольная выборка (испытуемые без признаков эмоционального выгорания) составляет 50 человек (показано на рисунке лишь 20), а группы сравнения (которым проводилась психотерапия разными методами) — соответственно 12, 11 и 16 человек. При нажатии на указанную кнопку ($N = ?$) осуществляется процедура проверки переменных на нормальность. Выводится соответствующий результат (рис. 4.9).

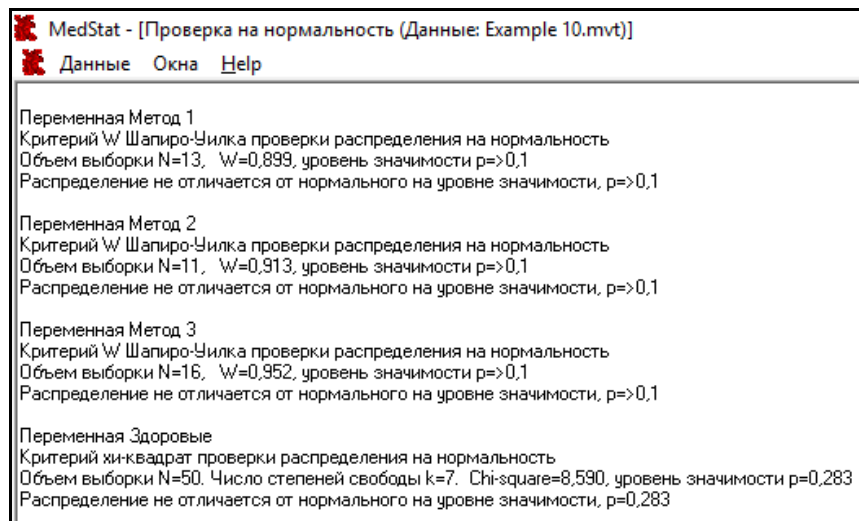


Рис. 4.9

В данном примере: если значение $p \geq 0,1$, то это означает, что данное распределение не отличается от нормального на уровне статистической значимости ($p \leq 0,05$), а следовательно, в данном случае рассматриваем его как нормальное.

Обратим внимание, что от нормального распределения могут быть отклонения разного рода — как в виде выбросов (когда наблюдаются единичные аномально большие или малые значения, и исследователь должен дополнительно разобраться с причинами их появления — от ошибок набора данных в таблицу до артефак-

тов любого рода), так и различные значения асимметрии и эксцесса, при этом само нормальное распределение имеет нулевые асимметрию и эксцесс.

Асимметрия — мера отклонения графика распределения от симметричного вида относительно среднего значения. Если данные представлены в зет-значениях, то асимметрия вычисляется как:

$$As = \frac{\sum_{i=1}^n z_i^3}{n}.$$

Асимметричные распределения представлены на рис. 4.10.



Рис. 4.10

Психологическое значение левосторонней асимметрии состоит в том, что в изучаемой группе наблюдается большее количество людей с высоким уровнем изучаемой характеристики, чем это ожидается при симметричном распределении.

Например, среди осужденных за насильственные преступления уровень агрессивности будет высоким в большинстве случаев, и графически это будет выглядеть как левосторонняя асимметрия.

Правосторонняя асимметрия указывает на то, что в изучаемой группе наблюдается большое количество людей с низким уровнем изучаемого признака. Например, среди учащихся коррекционных классов уровень интеллекта будет сдвинут в сторону низких значений, соответственно, в правую сторону на графике (рис. 4.10).

Эксцесс — степень плосковершинности или остроконечности графика распределения измеренного признака. Вычисляется по формуле:

$$Ex = \frac{\sum_{i=1}^n z_i^4}{n} - 3.$$

Различные значения эксцесса представлены на рис. 4.11.

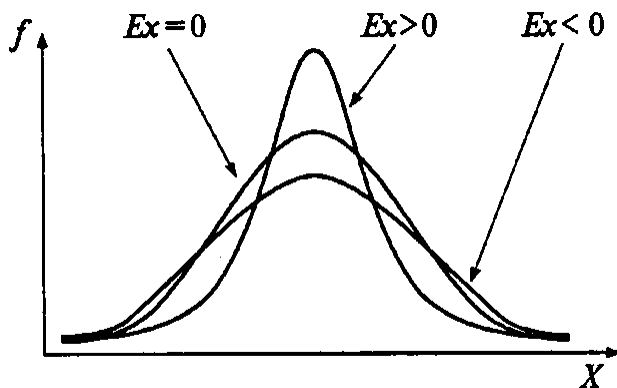


Рис. 4.11

Психологическое значение плосковершинного эксцесса состоит в том, что в группе относительно равномерно представлены все возможные показатели изучаемого признака, средних несколько больше. Например, учащиеся в классе распределены по успеваемости так, что двоечников и отличников примерно одинаковое количество, а хорошистов чуть больше.

Острровершинный эксцесс указывает на то, что средних значений изучаемого признака в исследуемой группе очень много, а высоких и низких почти нет. Например, социально-психологический климат группы благоприятен, нет отверженных и изолированных и минимальное количество «звезд».

При разработке тестовых шкал и проведении стандартизации (переводу сырых баллов в стандартные) существует ряд приемов приведения к нормальному виду таких распределений [9].

Так, лишь при условии измерения данных в шкале интервалов или отношений и соответствия их распределения нормальному, мы имеем полное право в качестве описательных статистик применять *среднее арифметическое, стандартное (среднеквадратичное) отклонение, стандартную ошибку, доверительные интервалы*. В случае интервальных данных или данных в шкале отношений, не распределенных нормально, мы уже не имеем права использовать стандартное отклонение, стандартную ошибку и доверительные интервалы, а помимо среднего обязательно анализировать моду и медиану.

Если же данные не соответствуют нормальному распределению или даже формально соответствуют, но измерены в порядковой шкале, то в качестве описательных статистик используются медиана, мода, квартили, процентиля, а также специальным образом рассчитываемые с помощью других формул ошибка медианы и доверительные интервалы, что реализовано, например, в программе «Медстат» [9].

И, что самое важное, именно соответствие конкретного распределения данных нормальному типу позволяет нам осуществить обоснованный выбор того или иного статистического критерия, а также его типа.

Можно разделить статистические критерии на параметрические и непараметрические.

Если данные количественные (неважно, сколько выборок в исследовании) и их распределение не отличается от нормального, то мы имеем право применять *параметрические статистические критерии*, ценность которых состоит в их мощности и чувствительности, учете в средних значениях каждой варианты и, в конечном счете — в более точном и полном отражении характеристик выборки.

В противном случае, если данные не проходят проверку на нормальность распределения, мы вынуждены пользоваться *непараметрическими критериями* — более устойчивыми к виду распределения, однако в целом менее информативными из-за огрубления шкалы — из интервалов или отношений — в порядковую.

4.3. ОПРЕДЕЛЕНИЕ РАЗЛИЧИЙ И СВЯЗЕЙ В КОЛИЧЕСТВЕННЫХ И ПОРЯДКОВЫХ ДАННЫХ

Наиболее привычная для психолога-исследователя ситуация — это когда каждому объекту наблюдения приписываются определенные признаки, выраженные в числах (баллы по опроснику, возраст и т. д.) или какие-либо качества (профессия, вероисповедание и проч.). Первые относятся к шкалам отношений, интервалов или порядка, вторые — к шкале наименований или качественным данным. Мы их можем только подсчитать и разбить испытуемых по соответствующим категориям. Однако именно количественные данные, когда каждому объекту (испытуемому) сопоставляется некоторое число (числа) или ранг (ранги) в определенных признаках и по некоторым правилам, имеют особую ценность для психолога-исследователя, поскольку позволяют в принципе получить более обширную и точную информацию об объекте или выборке. Всегда имеется возможность получить из количественных данных качественные, точнее, категориальные или дихотомические (например,

разбив их по категориям), но обратное — невозможно. Следовательно, количественные данные по умолчанию содержат большую информацию об объекте.

Например, в результате обследования испытуемых методикой М. Рокича «Ценностные ориентации» будет получен ряд рангов в отношении выбора испытуемыми терминальных и инструментальных ценностей, какие-то ценности окажутся на первых местах, какие-то — на последних. Примером перевода количественных данных (в данном случае рангов ценностей) в качественные будет присвоение первым в ранговом ряду ценностям категории «предпочитаемые ценности», а последним — «отвергаемые ценности». Обратный вариант, например, назначить терминальной ценности «хорошие, верные друзья» некое числовое обозначение, невозможен.

4.3.1. Сравнение выборочного среднего с константой. Параметрический одновыборочный критерий Стьюдента

Пожалуй, наиболее простым примером критерия, в данном случае параметрического, когда распределение признака не отличается от нормального, будет сравнение выборочного среднего с конкретным числом. При всей кажущейся простоте возможен довольно широкий класс задач, когда применение именно этого критерия (одновыборочного критерия Стьюдента) окажется более чем уместным. Например, исследователя может заинтересовать вопрос: значительно ли отличается значение показателя, полученного в выборочной совокупности, от популяционного среднего.

Например, исследователя может заинтересовать, отличается ли тревожность (фактор О) у врачей от популяционного среднего по опроснику Р. Кеттелла.

Рассмотрим данный пример в программе Statistica. Выборка врачей разного возраста и профиля составила 34 человека.

Вносим данные в программу (рис. 4.12).

| | 1 A | 2 B | 3 C | 4 E | 5 F | 6 G | 7 H | 8 I | 9 L | 10 M | 11 N | 12 O | 13 Q1 | 14 Q2 | 15 Q3 | 16 Q4 |
|----|--------|--------|--------|--------|--------|--------|--------|--------|--------|---------|---------|---------|----------|----------|----------|----------|
| 1 | 4 | 9 | 3 | 5 | 3 | 9 | 7 | 7 | 8 | 4 | 4 | 8 | 5 | 6 | 6 | 8 |
| 2 | 3 | 10 | 5 | 9 | 3 | 9 | 5 | 7 | 6 | 7 | 3 | 10 | 4 | 3 | 5 | 9 |
| 3 | 4 | 9 | 2 | 4 | 6 | 9 | 3 | 8 | 5 | 5 | 2 | 8 | 4 | 2 | 9 | 4 |
| 4 | 3 | 9 | 6 | 6 | 2 | 7 | 8 | 9 | 8 | 6 | 5 | 10 | 4 | 2 | 4 | 7 |
| 5 | 9 | 10 | 3 | 4 | 7 | 8 | 6 | 7 | 7 | 9 | 4 | 10 | 5 | 5 | 8 | 9 |
| 6 | 5 | 10 | 7 | 8 | 3 | 8 | 9 | 9 | 6 | 6 | 2 | 9 | 2 | 2 | 7 | 5 |
| 7 | 10 | 9 | 2 | 7 | 8 | 9 | 3 | 7 | 6 | 7 | 3 | 8 | 6 | 2 | 9 | 7 |
| 8 | 4 | 10 | 8 | 9 | 2 | 9 | 9 | 9 | 9 | 4 | 1 | 9 | 3 | 3 | 5 | 9 |
| 9 | 8 | 10 | 6 | 5 | 9 | 4 | 8 | 7 | 4 | 7 | 7 | 6 | 7 | 6 | 4 | 4 |
| 10 | 10 | 8 | 7 | 8 | 7 | 4 | 7 | 5 | 6 | 3 | 4 | 4 | 9 | 7 | 6 | 5 |

Рис. 4.12

Для экономии места мы показали лишь 10 первых значений из 34. Итак, нас интересует фактор О — тревожность. Для работы с ним в меню Statistics (Статистика) выбираем Basic Statistics/Tables (Описательные статистики / таблицы) (рис. 4.13).

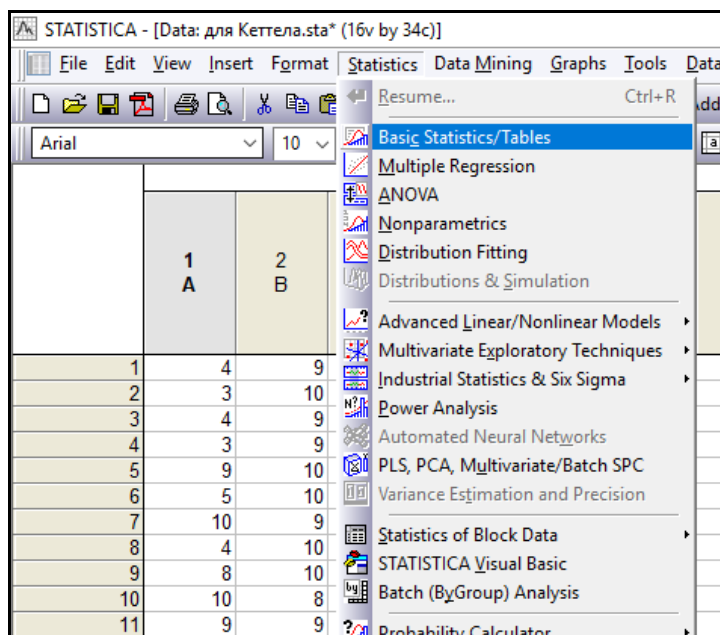


Рис. 4.13

Затем выбираем пункт в открывшемся следующем меню t-test, single sample (одновыборочный критерий Стьюдента) (рис. 4.14).

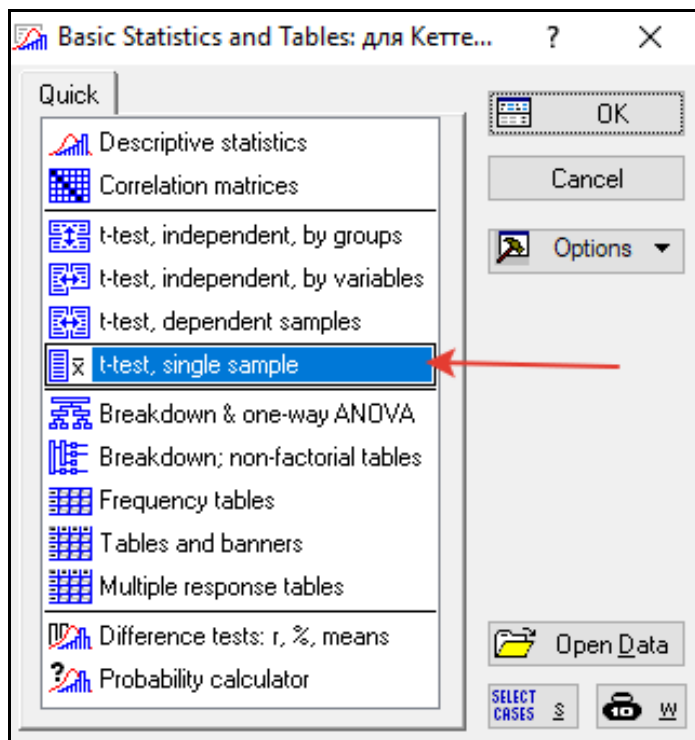


Рис. 4.14

Далее в раскрывшемся меню (рис. 4.15) выбираем переменную для анализа (в данном случае О), в качестве среднего выбираем 5,5, поскольку именно таковым является популяционное среднее для опросника Р. Кеттелла (выраженное в шкале стенов).

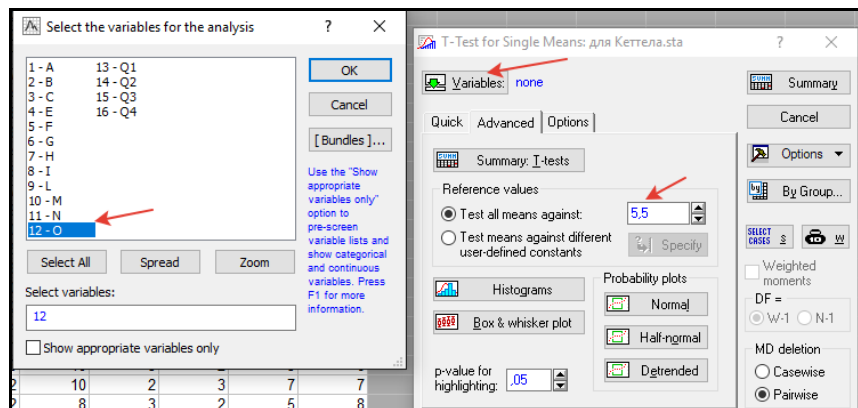


Рис. 4.15

В итоге получаем результат (рис. 4.16).

| Test of means against reference constant (value) (для Кеттлена.sta) | | | | | | | | |
|---|----------|----------|----|----------|--------------------|----------|----|----------|
| Variable | Mean | Std.Dv. | N | Std.Err. | Reference Constant | t-value | df | p |
| O | 7,470588 | 2,584749 | 34 | 0,443281 | 5,500000 | 4,445463 | 33 | 0,000094 |

Рис. 4.16

Наблюдаются высоко значимые различия в сторону большей тревожности от среднего популяционного уровня. Таким образом, для врачей характерен завышенный уровень тревожности, в сравнении с общей популяцией, что может проявляться в постоянном эмоциональном и психофизиологическом напряжении, в нарушениях сна, снижении настроения.

4.3.2. Параметрические критерии для оценки различий в уровне и разбросе признака в двух независимых и зависимых выборках

Хотя существует множество непрерывных и дискретных распределений, в психологии мы оперируем в основном признаками, чье распределение соответствует или не соответствует нормальному.

Так, в случае нормальности распределения, достаточно всего двух параметров (среднего и дисперсии), чтобы полностью охарактеризовать вероятность нахождения того или иного числа (признака) на соответствующей кривой нормального распределения. Это означает, применительно к критериям, что мы можем с вполне определенной точностью судить о том, с какой вероятностью будут отличаться между собой числа, принадлежащие к разным выборкам, в частности, средние или дисперсии как параметры нормального распределения. Классические формулы для расчетов критериев Стьюдента включают требования равенства дисперсий, однако если дисперсии у двух распределений отличаются, это учитывается в формулах расчета в статистических пакетах.

Очевидно, что количественные оценки различий всегда более точны, чем ранговые, поэтому, если распределение не отличается от нормального, как правило, мы применяем именно параметрические критерии. Итак, для порядковых данных, чье распределение не отличается от нормального, принято применять параметрические критерии.

Напомним, что следует различать независимые и зависимые выборки. *Независимые выборки*, как правило, состоят или из испытуемых, которые отличаются в отношении свойства, влияние которого исследуется на определенный психологический показатель (например, изучаются различия когнитивных стилей у бухгалтеров и художников и т. д.), или из испытуемых, случайным образом отобранных в две и более группы из генеральной совокупности (одинаковых по определенным свойствам субъектов) для того, чтобы изучить эффект от воздействия того или иного фактора — например, разные способы психотерапии, разные социальные ситуации и проч.

Зависимые выборки состоят, как правило, из одних и тех же испытуемых, у которых два или более раз в разных ситуациях, или при разных воздействиях, либо до и после определенных воздействий измеряются определенные признаки.

Напомним, что существует много статистических критериев для одной и той же цели, мы демонстрируем лишь некоторые из них, наиболее часто употребляемые, не вдаваясь глубоко в их специфику.

Критерий Стьюдента для оценки уровня различий в уровне признака в двух независимых выборках

Один из самых «узнаваемых» параметрических критериев — критерий Стьюдента, а также его разновидности.

Требованием к применению t -критерия Стьюдента является, прежде всего, нормальность статистического распределения в выборках.

Применять параметрические критерии рискованно, если в каждой из выборок меньше 10 испытуемых, поскольку в таком случае сложно оценить, действительно ли распределение носит характер нормального. Однако на практике все зависит от верности оценки исследователем распределения признака в генеральной совокупности (содержательных психологических знаний о ее свойствах) при условии, в том числе, и визуального анализа распределения, выводов о причинах возможных выбросов и т. д.

Статистическая гипотеза при применении критерия Стьюдента состоит в проверке равенства (нулевая гипотеза) или ненулевой разности (альтернативная гипотеза) средних значений двух выборок.

Покажем на примере, как применяется t -критерий Стьюдента в программе «Медстат». Представим, что задачей исследования было сравнение продавцов-консультантов в компании. В качестве одной группы были отобраны те, кто постоянно перевыполняет план продаж, а другой — наоборот, постоянно недовыполняет. Обозначим их как «успешные продавцы» и «неуспешные продавцы». Данные выборки состоят из разных сотрудников, отличающихся по определенному фактору — «успешности». В данном контексте, несмотря на то, что они могут взаимодействовать между собой и обмениваться информацией, опытом, — выборки независимые в том отношении, что у различных испытуемых сохраняется различным фактор «успешности». Было выдвинуто предположение, что успешность продавца-консультанта связана с его коммуникативными характеристиками. Для оценки коммуникативных характеристик использовалась методика диагностики межличностных отношений

Т. Лири. В данном примере рассматривается шкала «агрессивность». Максимальная оценка уровня агрессивности — 16 баллов, но она разделена на четыре степени выраженности (табл. 4.3).

Таблица 4.3

**Уровни шкалы «агрессивность»
по степени выраженности**

| | |
|------------------------------|--|
| 0-4 балла — низкая | Адаптивное поведение. По шкале агрессивности — это упрямый, упорный, настойчивый и энергичный тип |
| 5-8 баллов — умеренная | |
| 9-12 баллов — высокая | Экстремальное поведение: требовательный, прямолинейный, откровенный, строгий и резкий в оценке других, непримиримый, склонный во всем обвинять окружающих, насмешливый, ироничный, раздражительный |
| 13-16 баллов — экстремальная | Доходящее до патологии поведение: жесткий и враждебный по отношению к окружающим, резкий, жесткий, агрессивность может доходить до асоциального поведения |

В данном случае, поскольку методика является стандартизированной, мы можем считать шкалу интервальной и применять параметрические критерии при условии соответствия статистического распределения показателя в каждой выборке нормальному.

Итак, вносим данные в программу (рис. 4.17).

Выборка успешных продавцов составила 20 человек, неуспешных — 19. Рекомендация примерного равенства по объему выборок сохранена. Проверяем данные на нормальность (см. рис. 4.18).






| | | |
|---|------------|------------|
|   N=?    | | |
| | вход. | вход. |
| | успешные г | неуспешный |
| 1 | 3 | 11 |
| 2 | 5 | 8 |
| 3 | 0 | 3 |
| 4 | 6 | 8 |
| 5 | 4 | 8 |
| 6 | 4 | 9 |
| 7 | 4 | 5 |
| 8 | 5 | 13 |
| 9 | 4 | 4 |
| 10 | 2 | 9 |
| 11 | 3 | 7 |
| 12 | 4 | 6 |
| 13 | 3 | 11 |
| 14 | 9 | 9 |
| 15 | 2 | 6 |
| 16 | 4 | 4 |
| 17 | 4 | 12 |
| 18 | 1 | 8 |
| 19 | 7 | 7 |
| 20 | 2 | ? |

Рис. 4.17

Проверка на нормальность (Данные: Noname.mvt)

Переменная успешные продавцы

Критерий W Шапиро-Уилка проверки распределения на нормальность

Объем выборки N=20, W=0,945, уровень значимости $p \geq 0,1$

Распределение не отличается от нормального на уровне значимости, $p \geq 0,1$

Переменная неуспешные продавцы

Критерий W Шапиро-Уилка проверки распределения на нормальность

Объем выборки N=19, W=0,971, уровень значимости $p \geq 0,1$

Распределение не отличается от нормального на уровне значимости, $p \geq 0,1$

Рис. 4.18

После того как мы убедились, что данные в выборках соответствуют нормальному распределению, нажимаем (рис. 4.19) на кнопку обработки данных параметрическими методами и затем, в выпадающем списке, на пункт «Сравнение средних для двух независимых выборок».

Данные: Noname.mvt

Нормальный закон распределения

| | вход. | вход. | |
|----|------------|------------|---|
| | успешные г | неуспешный | |
| 1 | 3 | 11 | Описательная статистика - Alt 1 |
| 2 | 5 | 8 | $\bar{X} > \bar{Y}$ Сравнение средних для двух независимых выборок - Alt 2 |
| 3 | 0 | 3 | $\bar{X}_1 \bar{X}_2$ Сравнение для двух связанных выборок - Alt 3 |
| 4 | 6 | 8 | $S^2_1 S^2_2$ Сравнение дисперсий для независимых выборок - Alt 4 |
| 5 | 4 | 8 | Множественные сравнения - Alt 5 |
| 6 | 4 | 9 | Сравнения с контрольной группой - Alt 6 |
| 7 | 4 | 5 | Линейная корреляция (коэффициент Пирсона) - Alt 7 |
| 8 | 5 | 13 | Линейная регрессия (однофакторная модель) - Alt 8 |
| 9 | 4 | 4 | Критическая область |
| 10 | 2 | 9 | <input type="radio"/> Односторонняя <input checked="" type="radio"/> Двусторонняя |

Рис. 4.19

Далее (рис. 4.20) получаем представление результатов применения критерия Стьюдента для независимых выборок с двусторонней критической областью и учетом возможного неравенства дисперсий в выборках:

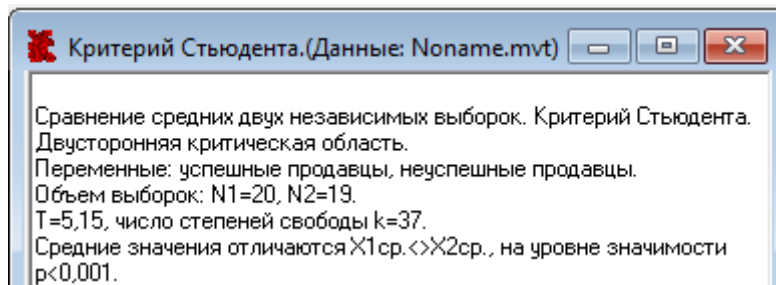


Рис. 4.20

Нами рассматривается двусторонняя альтернативная гипотеза (то есть мы предполагаем, что данные могут меняться как в сторону увеличения, так и в сторону уменьшения по каждой группе). Если бы мы имели теоретические основания полагать, что успешные продавцы-консультанты более конструктивны во взаимодействии с другими людьми, а неуспешные — более конфликтны и раздражительны, тогда мы могли бы использовать одностороннюю критическую область и получить более высокие показатели статической значимости.

В данном примере двусторонняя альтернативная гипотеза о различии средних подтверждается. Если нам нужно посмотреть и обсудить конкретные значения, то выбираем пункт «Описательная статистика» (рис. 4.21).

| Нормальный закон распределения (Данные: Noname.mvt) | | | | |
|---|--------|---------|--------|--------------|
| Переменная | Кол-во | Среднее | С.к.о. | Ош. среднего |
| успешные продавц | 20 | 3,8 | 2,042 | 0,4565 |
| неуспешные прода | 19 | 7,789 | 2,76 | 0,6333 |

Рис. 4.21

Таким образом, полученные результаты позволяют нам говорить о том, что успешные продавцы конструктивно-настойчивы, и это обеспечивает их адекватный контакт с покупателями и позволяет довести сделку до логического завершения. Тогда как неуспешные продавцы-консультанты избыточно резки и раздражительны, что, вероятно, отпугивает клиентов от взаимодействия с ними.

Критерий Стьюдента для оценки сдвига в значении признака в двух зависимых выборках

Если в случае независимых выборок мы можем перемешать одну и другую выборки и рассматривать статистические гипотезы как значимость или незначимость в величине разностей между средними арифметическими двух выборок, то одну и другую выборки в зависимых выборках (например, испытуемые до и после тренинга) мы не можем перемешать, каждое значение в одной выборке как бы намертво привязано к своему двойнику. Например: измерения у одного и того же субъекта одинакового признака в разных ситуациях, в частности, «до» и «после» воздействия экспериментатора; по прошествии какого-то времени; «до» и «после» появления специфического опыта (травма, рождение тройни, употребление психоактивных веществ и т. д.). Поэтому мы можем говорить о задаче оценки статистической значимости (альтернативная гипотеза) или незначимости (нулевая гипотеза) сдвига в значении признака.

Приведем пример конкретной задачи и ее расчета в программе «Медстат». Например, проводится обучение группы пенсионеров в области экономической безопасности при использовании ими банковских карт на предмет распознавания мобильного мошенничества. Оценивается по одной и той же тестовой методике их компетенция ДО (входной контроль уровня знаний) и ПОСЛЕ (экзамен) проведения обучения по принятой исследователями шкале оценок от 0 до 120. Существенно, что при проведении первичного тестирования испытуемым не дается обратная связь о правильности или неправильности ответа на тот или иной вопрос (в данном случае — чтобы исключить возможность механического запомина-

ния правильного варианта ответа и его воспроизведения при повторном тестировании). Поскольку шкала оценок хорошо дифференцирована, мы имеем право воспользоваться параметрическими критериями при условии нормальности распределения результатов тестирования в обоих случаях (до и после обучения). Внесем данные в таблицу (рис. 4.22).

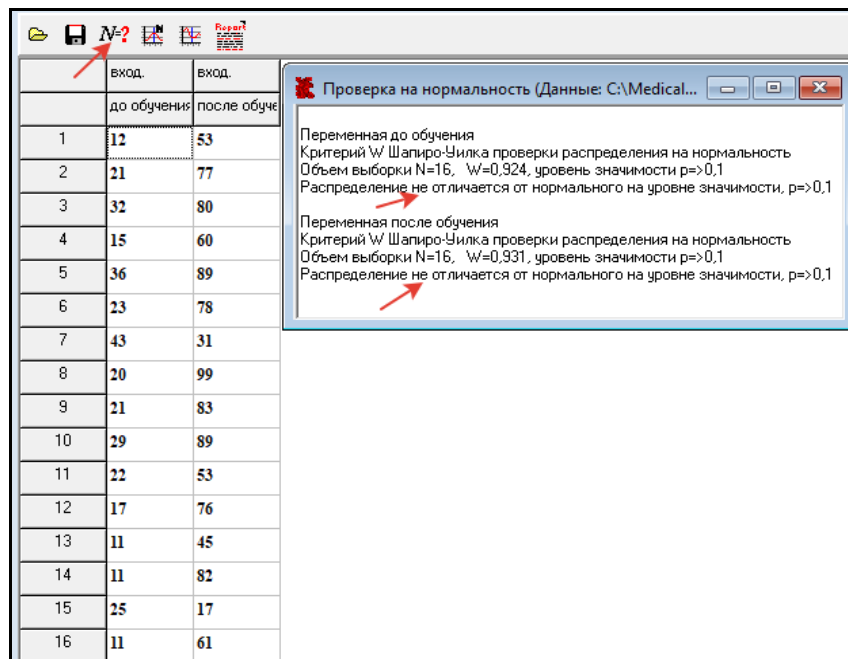


Рис. 4.22

Далее мы можем проанализировать описательную статистику в меню для параметрических методов, что нам может понадобиться при описании результатов, и, собственно, приступить к статистическому сравнению для двух связанных выборок (см. рис. 4.23), то есть в данном случае — оценить значимость различий компетентности пенсионеров до и после проведения обучения.

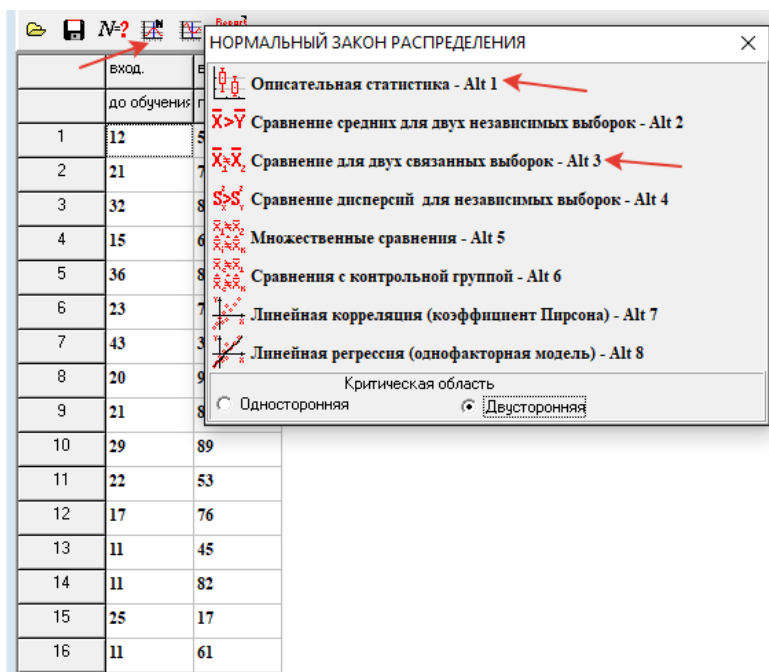


Рис. 4.23

В данном случае сдвиг в значениях среднего (до и после) оказывается статистически высоко значим на уровне $p < 0,001$ (рис. 4.24) — подтверждается альтернативная гипотеза о том, что оценки, полученные по тестам, а значит, и компетентность, значительно возросли.

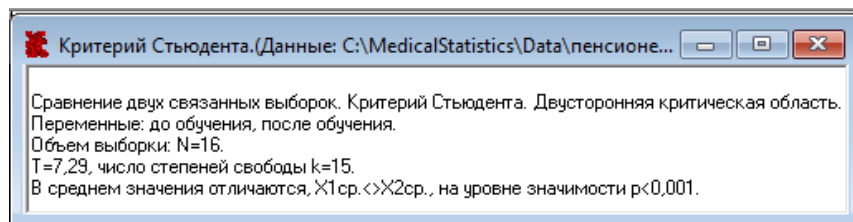


Рис. 4.24

С помощью подсчета описательной статистики можно установить, что если до обучения оценка за тест составила $21,81 \pm 9,45$, то после — $67,06 \pm 22,66$, где в качестве меры разброса мы использовали среднеквадратичное отклонение. Как мы уже знаем, это означает, что примерно 68% значений признака находятся в этих пределах. Однако, с учетом небольшой величины выборки, эта оценка разброса носит лишь крайне приблизительный характер. Приблизительно можно сказать: если перед обучением пенсионеры в среднем были «почти одинаково некомпетентны», то после обучения, компетентность одних существенно возросла, а других — несущественно, или даже уменьшилась. Поскольку величина среднеквадратичного отклонения в начале обучения существенно меньше, чем после обучения), что можно объяснить возрастными особенностями выборки, разной мотивацией и способностью к обучению в данной сфере, различными способностями и т. п.

**Параметрический критерий Фишера
для оценки различий в дисперсиях (разбросе)
в значениях признака в двух независимых выборках**

Мы уже упоминали о том, что не всегда наша задача — сравнение средних значений. Иной раз требуется определить степень выраженности различий в разбросе (дисперсии) данных двух выборок — иными словами, насколько сильно или слабо основной массив значений сгруппирован относительно среднего значения, при этом определив статистическую значимость различий степени этой группировки в двух различных выборках.

Рассмотрим следующий пример. Исследователь предполагает, что ревнивые мужчины имеют или завышенную, или заниженную самооценку. Для оценки данного фактора исследователь выбрал шкалу № 6 (шкала «ригидности») стандартизированного многофакторного метода исследования личности СМЛ (модифицированный тест ММРІ). Выдвинута гипотеза, что ревнивцы будут иметь либо заниженные, либо завышенные показатели по данной шкале. Супруги мужчин были изначально опрошены на предмет патологической ревности своих партнеров, таким образом мужчин разделили на две подвыборки.

Внесем данные по шкале № 6 в программу «Медстат» (рис. 4.25).

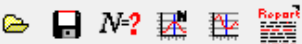
| | | |
|---|----------|------------|
|  | | |
| | вход. | вход. |
| | ровнивые | не ровнивы |
| 5 | 70 | 48 |
| 6 | 61 | 62 |
| 7 | 50 | 71 |
| 8 | 45 | 48 |
| 9 | 48 | 61 |
| 10 | 31 | 65 |
| 11 | 28 | 71 |
| 12 | 50 | 63 |
| 13 | 27 | 51 |
| 14 | 21 | 62 |
| 15 | 45 | 64 |
| 16 | 50 | 55 |
| 17 | 81 | 49 |
| 18 | 19 | |
| 19 | 28 | |
| 20 | 83 | |

Рис. 4.25

Проверка данных на нормальность (ее мы опускаем) показала, что статистическое распределение в каждой из групп не отличается от нормального. Далее применяем критерий Фишера (параметрический) для сравнения дисперсий (рис. 4.26).

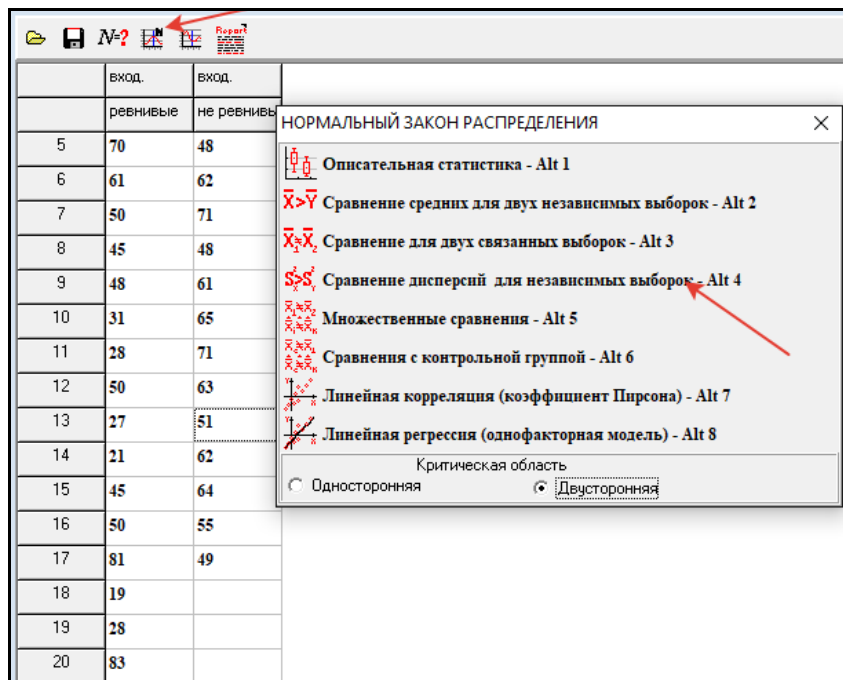


Рис. 4.26

В результате получаем, что дисперсии отличаются на высоком уровне статистической значимости $p < 0,01$ (рис. 4.27).

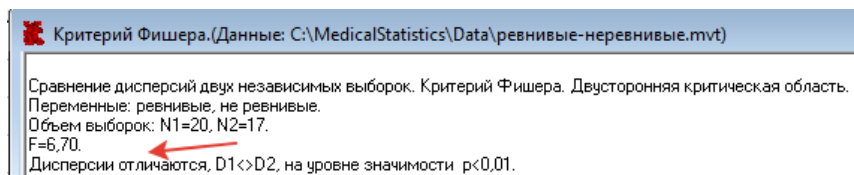


Рис. 4.27

Для того чтобы оценить эти различия количественно, следует выбрать пункт «Описательная статистика» (см. рис. 4.28).

| Нормальный закон распределения (Данные: C:\MedicalStatistics\Data\ревнивые-неревни | | | | |
|--|--------|---------|--------|--------------|
| Переменная | Кол-во | Среднее | С.к.о. | Ош. среднего |
| ревнивые | 20 | 53,4 | 23,49 | 5,253 |
| не ревнивые | 17 | 58,41 | 9,076 | 2,201 |

Рис. 4.28

Действительно, при незначительных различиях в средних обнаруживаются весомые отличия в среднеквадратичных отклонениях (корень из дисперсии). См. также график ошибки среднего и доверительного интервала (рис. 4.29).

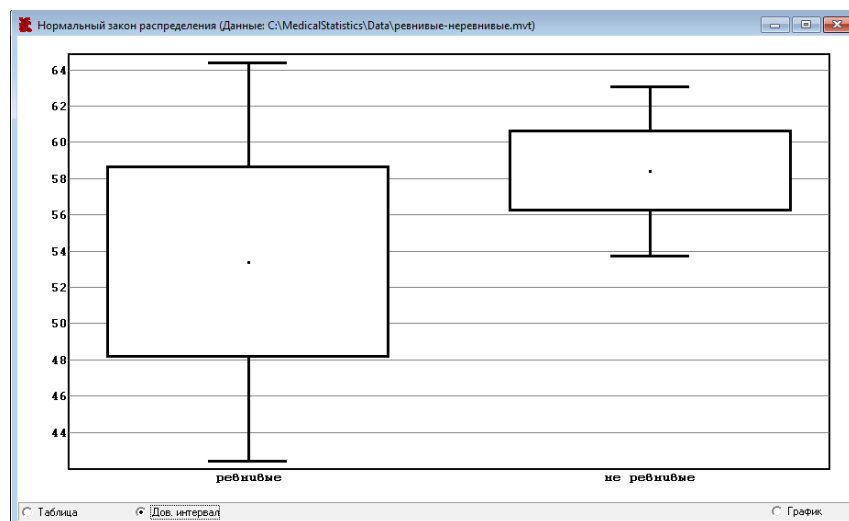


Рис. 4.29

Особо отметим, что при применении критерия Стьюдента для независимых выборок в данной задаче мы не получаем статистически значимых различий по данной шкале (рис. 4.30).



Сравнение средних двух независимых выборок. Критерий Стьюдента. Двусторонняя критическая область.
Переменные: ревнивые, не ревнивые.
Объем выборок: $N_1=20$, $N_2=17$.
 $T=0,88$, число степеней свободы $k=25$.
Различие средних не является статистически значимым, $p=0,387$.

Рис. 4.30

Иными словами, при том, что средние значения по шкале № 6 СМИЛ у ревнивых и не ревнивых не обнаруживаются, они статистически значимо отличаются по показателям разброса. Исследователь может сделать вывод, что ревнивцы чаще имеют либо завышенную, либо заниженную самооценку, нежели не ревнивые мужчины.

4.3.3. Непараметрические критерии для оценки положения и сдвигов в уровне признака в двух независимых и зависимых выборках

Ранговые (непараметрические) критерии потому и ранговые, что пересчитывают изначальные числа в ранги, тем самым переводя значения признаков в любом случае в порядковую шкалу из шкал интервалов или отношений. После перевода чисел в ранги мы лишаемся возможности оценить средние, дисперсии, стандартное отклонение, но все еще можем оценить медиану, квартили и т. д. Так, медиана учитывает только одно — серединное значение в выборке, что имеет свои плюсы и минусы. Как оценка среднего, очевидно, медиана никак не зависит от выбросов и артефактов и является более устойчивой к выбросам, отклонениям от нормальности распределения признака и т. д. Однако и информации мы получаем также меньше, потому что закругляем шкалу до порядковой. В большинстве случаев расчеты «вручную» и статистические пакеты присваивают большим значениям большие, а не меньшие ранги. Так, олимпийские чемпион в забеге из ста бегунов получит не 1-е место (ранг), а наибольшее по порядку значение (например, 100).

У параметрических критериев для сравнения средних арифметических в независимых выборках и для определения статистической значимости сдвига среднего арифметического есть свои непараметрические аналоги. В частности, наиболее применимые из них: критерий Манна–Уитни для независимых выборок и критерий Вилкоксона — для зависимых выборок.

Критерий Манна–Уитни для оценки различий в выраженности центральных тенденций в двух независимых выборках

Прежде всего, как было сказано выше, необходимо различать шкалы, в которых измерены переменные. Если шкала изначально порядковая (например, шкала рейтинга), то нет ни малейшего смысла пытаться использовать параметрические критерии. Также, если исходные данные являются количественными, но при этом статистическое распределение хотя бы одной из выборок отличается от нормального, то следует использовать, в частности, непараметрический ранговый критерий Манна–Уитни, который показывает, насколько интенсивны значения между рангами двух несвязанных выборок.

Приведем пример использования данного критерия с помощью программы «Медстат».

Например, исследователь сравнивает рейтинги показателей научной активности аналогичных по профилю кафедр в двух разных вузах. При этом можно использовать уже подсчитанные ранги (места), которые заняли преподаватели в общем рейтинге, либо указать их значения (которые были подсчитаны по одной и той же некоторой методике), и критерий Манна–Уитни по формуле проведет их общее ранжирование и будет выявлять различия между рангами, а не изначальными значениями. Введем данные в программу «Медстат» (рис. 4.31). В данном случае оказалось, что на двух кафедрах одинаковое количество преподавателей, однако, поскольку выборки независимые, напомним, что это условие не является строго обязательным.

Проверяем данные на нормальность статистического распределения (рис. 4.32).

| | | |
|----|-------|-------|
| | | |
| | вход. | вход. |
| | вуз 1 | вуз 2 |
| 1 | 600 | 500 |
| 2 | 10 | 110 |
| 3 | 5 | 400 |
| 4 | 0 | 0 |
| 5 | 16 | 280 |
| 6 | 125 | 570 |
| 7 | 750 | 1010 |
| 8 | 20 | 750 |
| 9 | 35 | 10 |
| 10 | 450 | 14 |
| 11 | 2 | 11 |
| 12 | 28 | 80 |
| 13 | 13 | 200 |
| 14 | 1200 | 590 |
| 15 | 17 | 850 |
| 16 | 0 | 126 |

Рис. 4.31

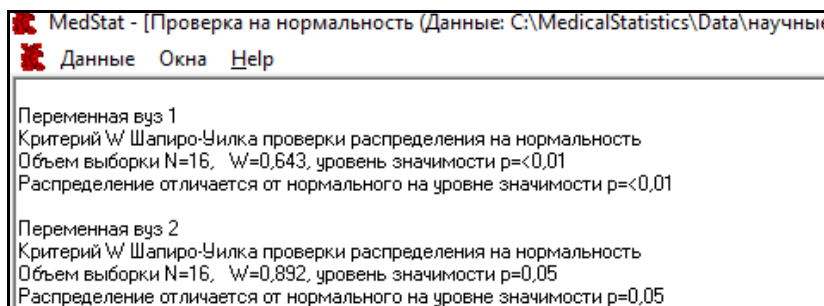


Рис. 4.32

Оказывается, что в обеих выборках не соблюдается условие нормальности распределения, поэтому применяем непараметрический критерий Манна–Уитни (рис. 4.33).

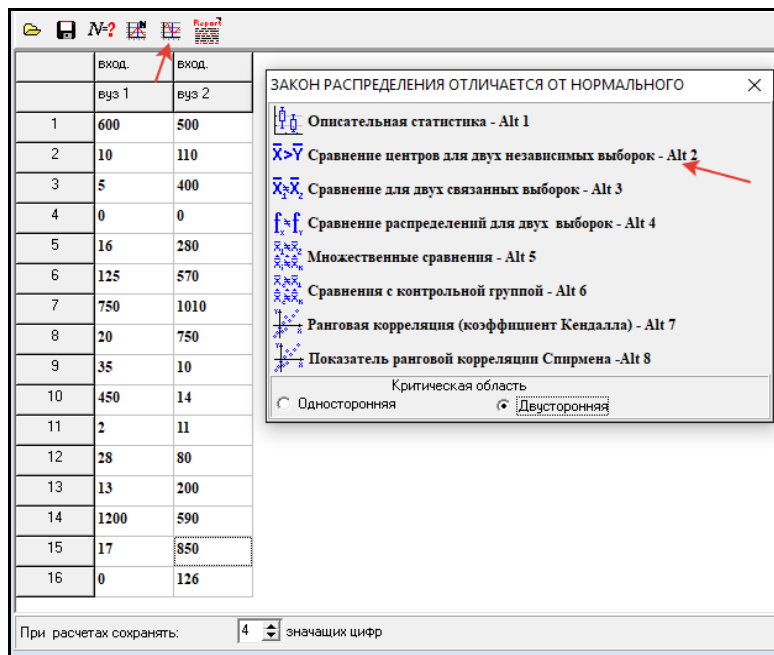


Рис. 4.33

В результате оказалось, что различия не являются статистически значимыми, поскольку в данном случае уровень статистической значимости $p = 0,113$, что больше граничного уровня $p = 0,05$ (рис. 4.34).

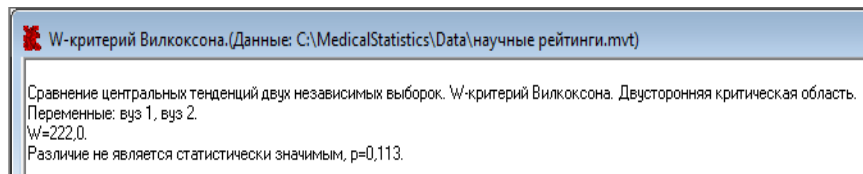


Рис. 4.34

Нулевая гипотеза о значимых различиях в рейтингах преподавателей кафедр разных вузов не может быть опровергнута. Иными словами, нельзя однозначно судить о том, что две кафедры аналогичной направленности в разных вузах характеризуются разной научной активностью по конкретной методике расчета рейтинга.

Критерий Вилкоксона для оценки интенсивности сдвигов в двух зависимых выборках

Критерий Вилкоксона — это некий аналог критерия Манна–Уитни за тем исключением, что критерий Вилкоксона оперирует с зависимыми выборками и оценивает статистическую значимость интенсивности сдвига признака.

Приведем пример использования данного критерия с помощью программы «Медстат».

Исследователя интересовали различия в тревожности одной и той же выборки студентов-психологов в начале учебного года и непосредственно перед зачетно-экзаменационной сессией. Использовалась в качестве диагностического инструмента шкала Спилбергера–Ханина для оценки ситуативной тревожности.

Как мы уже упоминали, в одной и той же строке исходной таблицы — данные одного и того же испытуемого для разных ситуаций (начало года и перед сессией).

Проверка на нормальность распределения показала, что в ситуации перед сессией распределение показателей тревожности отличается от нормального распределения (см. рис. 4.35).

Поэтому, хотя шкала опросника не ранговая, а балльная, мы все же применяем критерий Вилкоксона — на основании того, что распределение признака хотя бы в одной из выборок (в данном случае, «перед сессией») отличается от нормального (см. рис. 4.36).

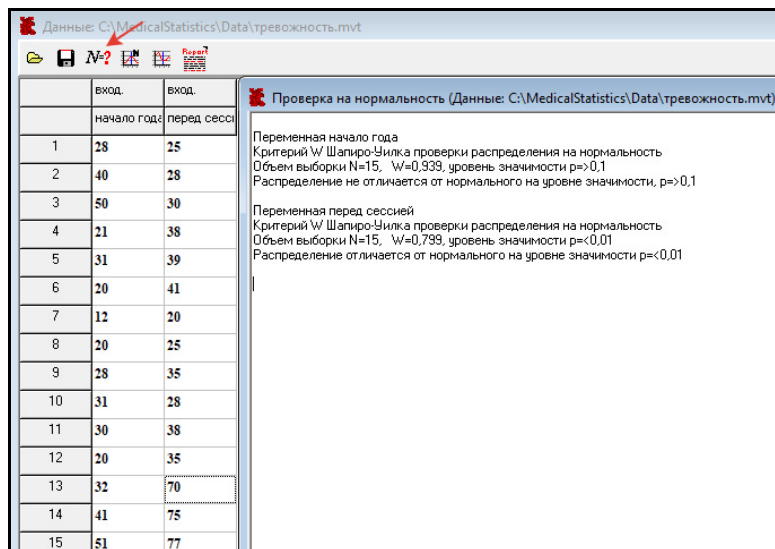


Рис. 4.35

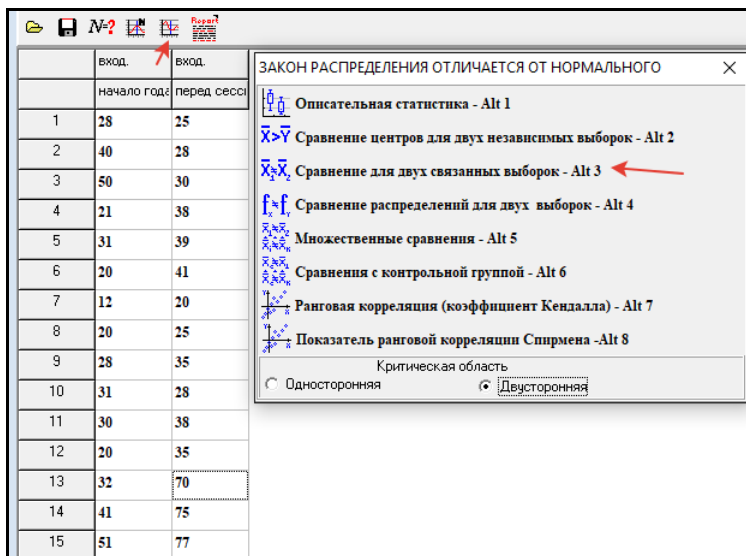


Рис. 4.36

Итог проверки нулевой гипотезы о том, что тревожность в начале учебного года не проявляет сдвига в значениях (в данном случае рангах), таков (рис. 4.37).

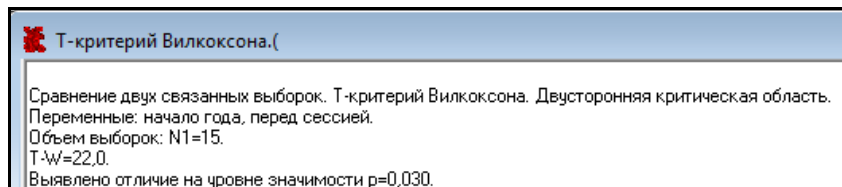


Рис. 4.37

Нулевая гипотеза отвергается и подтверждается альтернативная о сдвиге в значениях ситуативной тревожности. С помощью описательной статистики можно уточнить детали (рис. 4.38).

| Переменная | Кол-во | Медиана | I квартиль | III квартиль | Минимум | Максимум | Ош. медианы | Лев.(95% ДИ) | Прав.(95% ДИ) |
|---------------|--------|---------|------------|--------------|---------|----------|-------------|--------------|---------------|
| начало года | 15 | 30 | 20 | 40 | 12 | 51 | 3,648 | 20 | 40 |
| перед сессией | 15 | 35 | 28 | 41 | 20 | 77 | 5,986 | 28 | 41 |

Рис. 4.38

Несмотря на то, что показатели средних и квартилей и доверительных интервалов близки, в данном случае крайние значения (выбросы) определяют значимость различий. Понятно, что поскольку выбросы несимметричны относительно среднего, то и распределение отличалось от нормального (см. рис. 4.39).

Именно те несколько испытуемых, у которых тревожность резко возросла, определяют в данном случае значимость различий. Далее для исследователя может иметь практическое значение организация психокоррекционной работы именно с ними, при условии анализа причин столь высокой тревожности.

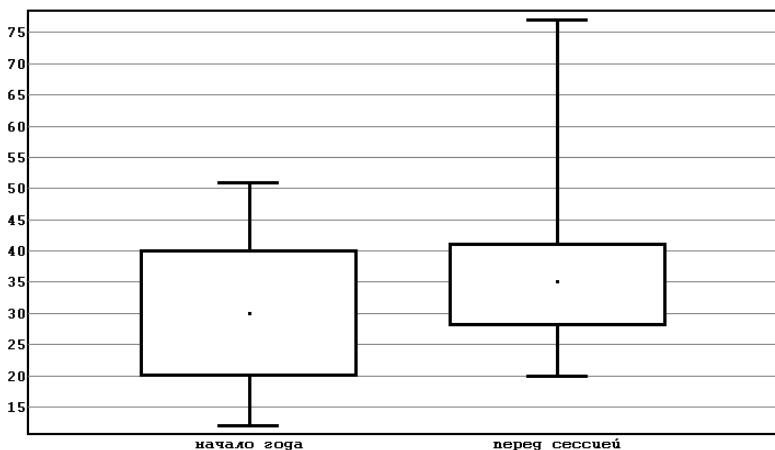


Рис. 4.39

4.3.4. Параметрические и непараметрические коэффициенты корреляции для оценки уровня статистической связи между переменными

Не менее важной задачей, чем определение статистической и фактической значимости различий, разброса или сдвигов в показателях, выступает задача определения интенсивности значимости статистической связи различных показателей, измеренных, чаще всего, на одной выборке. Речь идет о корреляции — статистической (не обязательно причинной) связи между двумя признаками, измеренными чаще всего на одних и тех же испытуемых.

Например, нас интересует связь между самооценкой и степенью избыточного веса у выборки подростков. У каждого подростка определенным образом измеряется признак, отвечающий за свойство самооценки своей внешности, и признак, принимающий различные значения в интервале от нормы до крайней степени ожирения. Очевидно, что найдутся испытуемые, которые даже при нормальном весе будут полагать его избыточным, и их самооценка будет неадекватно снижена, и, наоборот, будут испытуемые, само-

оценка которых будет высока, несмотря на избыточный вес. Кроме того, на самооценку будут оказывать влияние множество иных факторов, помимо веса, как бы мы ни старались сделать выборку гомогенной. Следовательно, на этом примере видно, что связь, хотя и будет, вероятно, наблюдаться — чем выше степень ожирения, тем ниже уровень самооценки внешности, — не будет столь однозначной и предсказуемой для каждого конкретного индивида, а значит, будет носить статистический характер. Поэтому для оценки степени связи двух признаков используется статистическая мера — *коэффициент корреляции*.

Рассмотрим и вкратце поясним свойства коэффициента корреляции.

Коэффициент корреляции (чаще всего обозначается латинской буквой R) может принимать значения от -1 до $+1$.

Чем ближе по модулю значения коэффициента корреляции к единице, тем сильнее статистическая связь между признаками. При максимальных по абсолютной величине значениях (-1 или $+1$), статистическая связь превращается в функциональную, то есть любому значению одного признака однозначно соответствует одно и только одно значение другого признака, что, конечно же, не так, если связь носит статистический, случайный характер.

На деле же мы получаем некоторые промежуточные значения. При нулевом или близком к нулю коэффициенте корреляции связь отсутствует или близка к нулю. Если коэффициент корреляции положителен, то большим значениям одного признака чаще (но не всегда!) соответствуют большие значения второго признака, а меньшим значениям одного признака чаще соответствует меньшие значения второго признака. Если коэффициент корреляции отрицателен, то, наоборот, большим значениям одного признака чаще соответствуют меньшие значения другого признака.

Следует отличать общую и частную классификацию коэффициентов корреляции.

Частная классификация делит их на *статистически значимые* и *незначимые* применительно к каждому конкретному случаю расчетов ($p > 0,05$ — незначимая, $0,01 \leq p < 0,05$ — значимая,

$0,01 \leq p < 0,001$ — высоко значимая, $p < 0,001$ — очень высоко значимая). Нулевой статистической гипотезой при использовании корреляционного анализа в данном случае является отсутствие статистически значимой связи между признаками, а альтернативной — наличие статистически значимой связи.

Наряду с частной классификацией существует и несколько вариантов общей классификации. Например, классификация по Э. В. Ивантер и А. В. Коросову: *тесная* при $|R| \geq 0,70$; *средняя* при $0,50 \leq |R| < 0,70$; *умеренная* при $0,30 \leq |R| < 0,50$; *слабая* при $0,20 \leq |R| < 0,30$; *очень слабая* при $|R| < 0,20$. Здесь речь идет о модуле коэффициента корреляции, то есть о его абсолютном значении, неважно — отрицательный или положительный коэффициент корреляции.

Однако главное — это какой коэффициент корреляции и на каких основаниях исследователь считает достаточным по величине для конкретного исследования и тем самым определяет для себя объем выборок путем специальных процедур в программных пакетах, которые обеспечат исследователю статистически значимую связь при вполне определенном коэффициенте корреляции. Об этом более подробно будет объяснено ниже, в главе о планировании проспективного (предстоящего) исследования и определении объема выборок для него.

Крайне важно, чтобы исследователь или интерпретатор результатов понимал, что корреляционная связь никоим образом не является основанием для суждения о причинной зависимости между переменными. Какая из переменных зависит от какой и зависит ли вообще — отдельная задача психологического исследования, как правило, выходящая за рамки математико-статистической обработки данных.

Так, в выборке младших школьников могут обратно коррелировать ответственность и рост: коррелировать могут, но интерпретации это не поддается. Либо же нужен поиск фактора (или факторов), который объяснит наблюдаемую (но причинно не связанную корреляцию). В частности, может оказаться, что чаще более ответственны девочки, которые, как правило, в предпубертатном возрасте имеют меньший рост.

Пример случая, когда зависимости как бы перепутаны, — тревожность и ответственность. Люди думают: чем человек ответственнее, тем больше он тревожится. На самом же деле, чем человек тревожнее, тем с большей вероятностью он будет демонстрировать более высокую ответственность при выполнении задания.

Могут находиться и просто лишённые смысла корреляции в данной выборке. Поэтому главная задача исследователя — не нахождение высокого и статистически значимого коэффициента корреляции, а объяснение и доказательство наличия причинной связи между переменными, причем с указанием, какая из переменных является фактором, а какая — результатом действия данного фактора («что на что влияет?»).

Существует несколько коэффициентов корреляции параметрических и непараметрических для количественных данных. Рассмотрим лишь два из них, наиболее часто используемых в психологических исследованиях.

Коэффициент линейной корреляции Пирсона

Как мы уже упоминали, могут быть самые разные формы зависимости между переменными, но в психологических задачах при наличии количественных признаков, действительно, чаще всего наблюдаются связи, близкие к линейным. Поэтому коэффициент линейной корреляции Пирсона (пригодный для выборок с нормальным распределением) применяется при соблюдении данного условия чаще всего.

Рассмотрим его применение на примере в программе «Мед-стат»: 40 испытуемых-педагогов проходили тест эмоционального интеллекта Д. В. Люсина и опросник толерантности к неопределённости Д. МакЛейна в адаптации Е. Н. Осина. Допустим, нас интересует связь между этими переменными.

Для экономии места покажем результат, выводимый программой (см. рис. 4.40).

В данном случае величина корреляционной связи составляет $T = +0,571$, что означает наличие положительной средней величины и, следовательно, очень высоко значимую статистическую связь между данными параметрами. Это означает, что хорошая способ-

ность к распознаванию и управлению эмоциями сопровождается терпимым отношением к ситуациям с неопределенным исходом, такие ситуации не вызывают эмоционального напряжения у исследуемых педагогов.

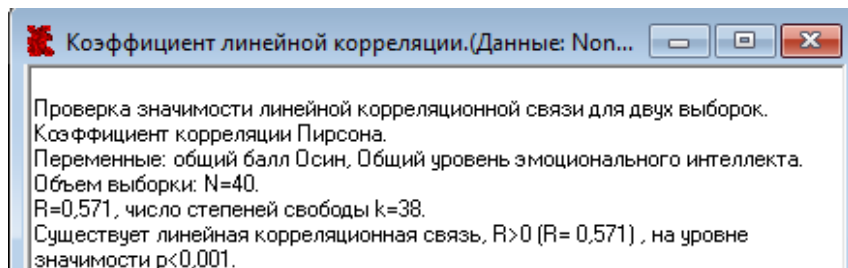


Рис. 4.40

Довольно часто в психологических задачах следует найти корреляции не между двумя, а несколькими переменными, например показателями различных опросников, измеренных у различных испытуемых. Рассмотрим лишь один из типичных примеров в программе Statistica, когда нас интересует, как различные переменные в двух опросниках коррелируют между собой. В частности, такая задача может встречаться на некоторых этапах разработки новых опросников, и не только. Так, 91 испытуемый — сотрудники силовых ведомств — заполняли опросник «Параметры интуиции в структуре саморегуляции деятельности» И. В. Васильевой, П. Е. Григорьевой и опросник «Стиль саморегуляции поведения» В. И. Моросановой. Задача исследователя — оценить конвергентную валидность шкал нового опросника интуиции в структуре саморегуляции. Нас будет интересовать, существуют ли и какие именно связи между переменными данных опросников (в данном случае, между шкалами первого и второго опросников, оценивающих с разных сторон параметры саморегуляции).

Загружаем данные в статистический пакет Statistica данные, включая именование переменных. Кстати, как мы знаем, Statis-

tica поддерживает полный импорт данных из Excel, включая заголовки.

Допустим, что наши данные прошли проверку на нормальность распределения, и мы приняли решение применить коэффициент линейной корреляции Пирсона. Для этого в меню программы выбираем Statistics → Basic Statistics and Tables → Correlation Matrices (Статистики, Базовые статистики и таблицы, Корреляционные матрицы) и попадаем в окно работы с корреляциями (рис. 4.41).

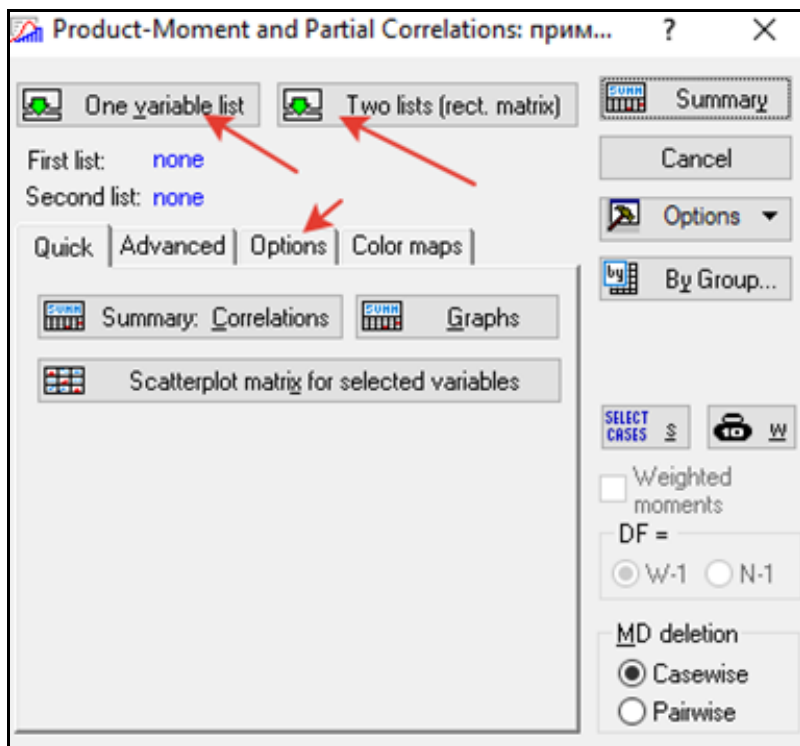


Рис. 4.41

Если бы нас интересовали все корреляции между всеми переменными, то нам бы следовало нажать на One variable list (Общий

список переменных), но нас будут интересовать два списка, поскольку будет проверяться наличие корреляций только между переменными двух опросников, а не переменных внутри каждого в том числе. Поэтому нам следует нажать на кнопку Two lists. В каждый список выберем интересующие нас переменные (рис. 4.42).

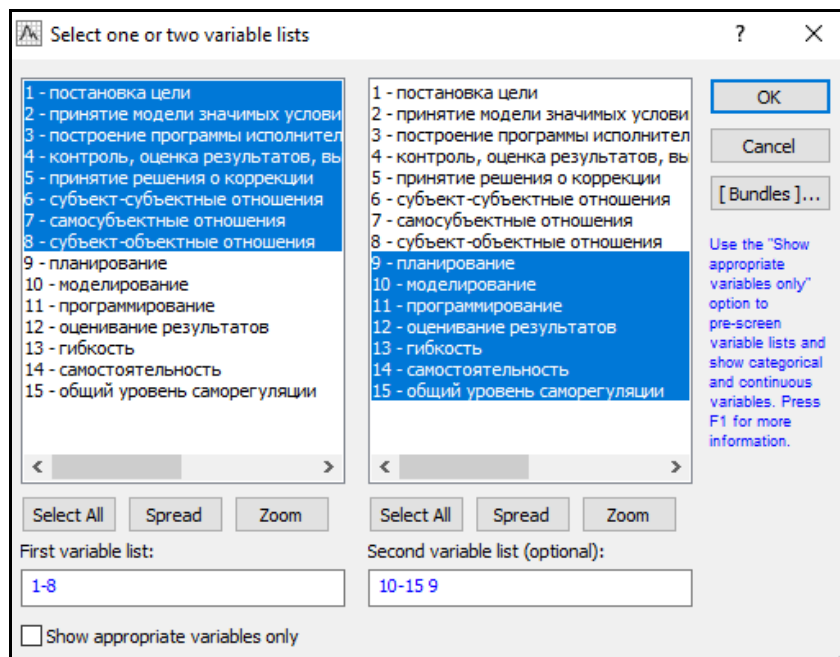


Рис. 4.42

Далее рекомендуется открыть вкладку Options (Опции) и выбрать отображение в результатах не только самих коэффициентов корреляций, но и значений статистической значимости для каждого из них (рис. 4.43), после чего нажать на вывод результатов Summary. Итог можно увидеть на рис. 4.44.

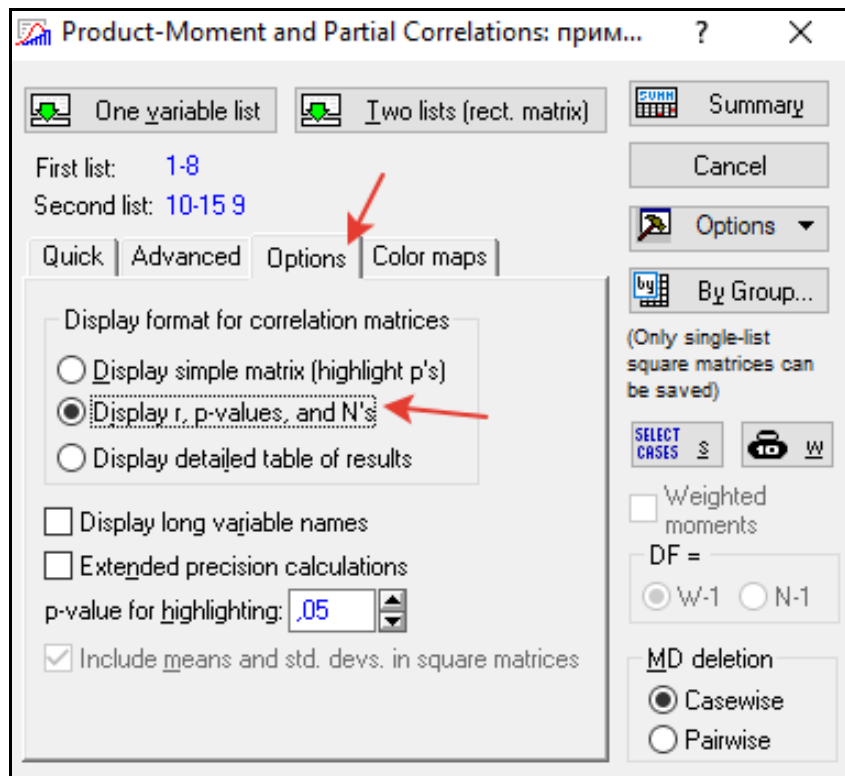


Рис. 4.43

| Variable | моделирование | программирование | оценивание результатов | гибкость | самостоятельность | общий уровень саморегуляции | планирование |
|---|---------------|------------------|------------------------|----------|-------------------|-----------------------------|--------------|
| постановка цели | -.0270 | .0972 | .1910 | .1948 | .3116 | .2696 | .2098 |
| | p= .799 | p= .359 | p= .070 | p= .064 | p= .003 | p= .010 | p= .046 |
| принятие модели значимых условий | .2003 | .1675 | .3192 | .2961 | .2990 | .3721 | .2346 |
| | p= .057 | p= .113 | p= .002 | p= .004 | p= .004 | p= .000 | p= .025 |
| построение программы исполнительских действий | .0160 | .1384 | .1427 | .1846 | .2896 | .2606 | .2485 |
| | p= .881 | p= .191 | p= .177 | p= .080 | p= .005 | p= .013 | p= .018 |
| контроль, оценка результатов, выделение критериев достижения цели | .0252 | .1654 | .2045 | .2909 | .2618 | .3115 | .2663 |
| | p= .812 | p= .117 | p= .052 | p= .005 | p= .012 | p= .003 | p= .011 |
| принятие решения о коррекции | .0579 | .0993 | .1133 | .2100 | .3490 | .2665 | .2162 |

Рис. 4.44

Проинтерпретируем наличие положительной значимой статистической связи между шкалами «гибкость» (опросник В. И. Моросановой) и «принятие решения о коррекции» (опросник И. В. Васильевой, П. Е. Григорьева). Поскольку шкала «гибкость» диагностирует уровень способности перестраивать, вносить коррекции в систему саморегуляции при изменении внешних и внутренних характеристик ситуации, а в шкалу «принятие решения о коррекции» закладывался диагностический конструкт способности принимать решение о коррекции на основе интуитивных предчувствий, неосознаваемой информации, то можно говорить о том, что конвергентная валидность шкалы «принятие решения о коррекции» (опросник И. В. Васильевой, П. Е. Григорьева) подтверждена.

При анализе матриц, где речь идет о статистической связи не двух, а целой совокупности коэффициентов корреляций, существует риск получить завышенную вероятность получения статистически значимых результатов, поскольку чем с большим числом выборок (рядов значений) мы имеем дело, тем, соответственно, чаще они будут давать значимые связи (либо различия) по чистой случайности. Вследствие этого исследователь должен крайне осторожно оперировать полученными данными. В частности, уметь содержательно показать и доказать их психологический смысл, а также, при возможности, подкрепить их аналогичными данными, полученными с применением других методов и методик (беседа, анамнез, экспертный анализ, метод тестов и др.).

4.3.5. Простая и множественная линейная регрессия

Суть линейного корреляционного анализа состоит не только в нахождении коэффициента корреляции, но и в нахождении уравнения регрессии, то есть формулы (в данном случае, прямой линии), вокруг которой ближайшим образом группируются отдельные точки на плоскости, где абсциссой является один параметр, а ординатой — другой.

В этом пособии мы рассматриваем лишь линейную модель регрессии, хотя не всегда переменные связаны между собой линейно. Например, эффективность логического мышления сначала возрас-

тает с возрастом, а затем — снова падает при старении. Количество употребляемых слов резко возрастает от 2 до 20 лет, затем остается почти на том же уровне и т. д.

Очень часто между двумя переменными наблюдается линейная или почти линейная связь, которая математически выражается как $y = ax + b$, где y — результирующая (зависимая) переменная; x — факторная (независимая) переменная, в том смысле, что исследователь предполагает, что именно от нее зависит y , а не наоборот; a — коэффициент пропорциональности и b — свободный член, согласующий размерности и значения результирующей и факторной переменной.

Иногда легко понять, где результирующая, а где — факторная переменная. Например, именно словарный запас, вероятно, будет скорее зависеть от количества прочитанных книг, а не наоборот; риск заболеваний органов дыхания будет зависеть от количества выкуриваемых сигарет, а не наоборот. Однако не всегда это столь очевидно и требует, как уже было указано для коэффициента корреляции, отдельного содержательного анализа. Тем не менее, по уравнению регрессии мы можем количественно судить о связи между переменными с определенной вероятностью. В данном случае, такого рода показатель — коэффициент детерминации (T^2), то есть квадрат коэффициента корреляции, о котором более подробно будет сказано ниже, показывает ту долю разброса результирующего признака, которая может объясняться (если существует причинная связь) факторным признаком.

Простая линейная регрессия

Простая линейная регрессия характерна тем, что мы ищем формулу прямой линии, которая бы лучше всего приближала так называемое облако корреляции или эллипс значений. Чем ближе точки на плоскости располагаются к линии регрессии, тем лучше изменчивость одной переменной объясняет изменчивость другой переменной. Исследователь произвольно оценивает и выбирает, какая переменная будет зависимой (результирующей), а какая независимой, факторной (влияющей на результирующую).

Вернемся к примеру из предыдущего пункта, где исследовалась корреляция между общим баллом толерантности к неопределенности опросника Д. МакЛейна в адаптации Е. Н. Осина и общим уровнем эмоционального интеллекта (Д. В. Люсин). Допустим, что теперь мы хотим получить уравнение регрессии — постулируемой зависимости толерантности к неопределенности от эмоционального интеллекта. Допустим, мы выдвинули альтернативную гипотезу именно о зависимости толерантности к неопределенности от эмоционального интеллекта (хотя имели полное право поменять местам переменные в процессе выдвижения и обоснования гипотез). Итак, воспользуемся программой «Статмед». После проверки на нормальность и выбора меню для обработки данных параметрическими методами, выбираем пункт «Простая регрессия (линейная модель)» (рис. 4.45).

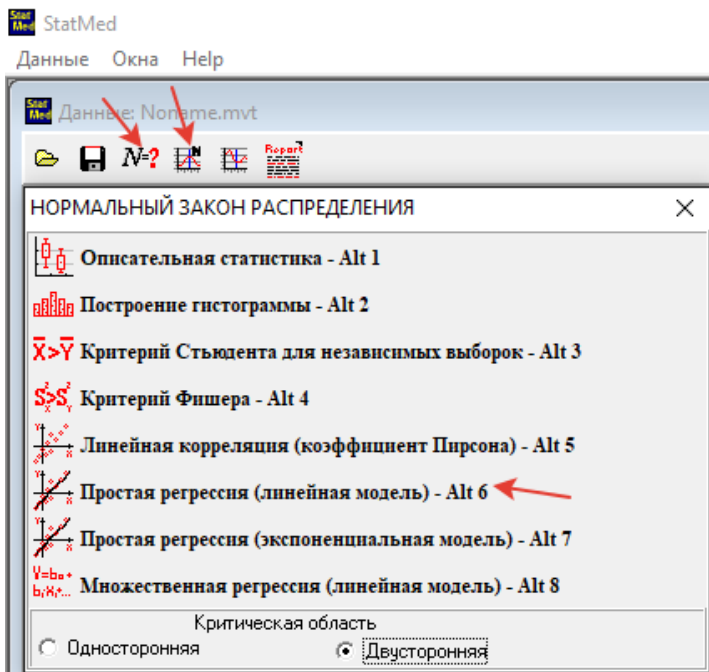


Рис. 4.45

Текст и рисунок, представленные ниже, сгенерированы программой.

Уравнение линейной регрессии имеет вид:

общий балл Осин = 0,6412*Общий уровень эмоционального интеллекта + 25,55

Объем выборки: $N = 40$. Коэффициент линейной корреляции $R = 0,571$, число степеней свободы $m, k = 1,38$.

Модель адекватна, $F = 18,340$, уровень значимости $p < 0,001$.

Коэффициент детерминации = 0,308

В данном случае самое главное в результатах — это полученная формула (во второй строке таблицы с результатами, приведенной выше), которая описывает искомую зависимость. Визуально «облако» корреляции (расположение точек, где по оси ординат отложен общий балл толерантности к неопределенности, а по оси абсцисс — общий уровень эмоционального интеллекта) и прямая линия регрессии, приблизительно описывающая формулой данную статистическую связь, представлены на рис. 4.46.

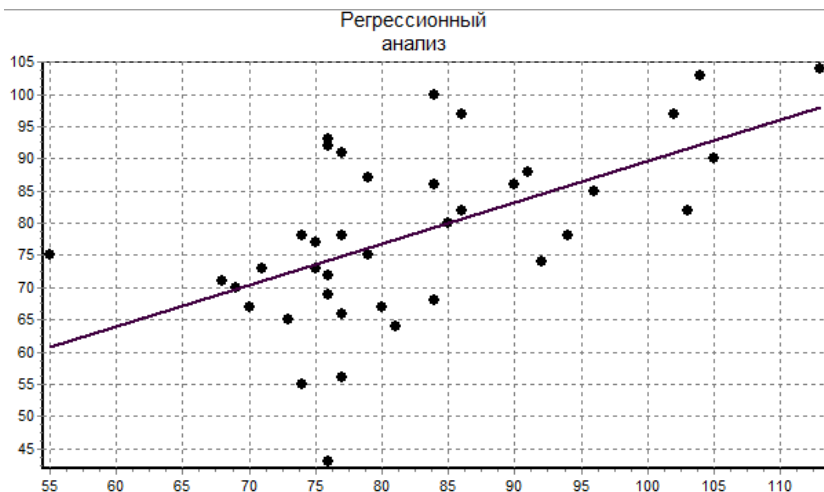


Рис. 4.46

Статистически доказана адекватность модели (значимость на очень высоком уровне $p < 0,001$, подтверждение альтернативной гипотезы). Коэффициент детерминации составляет, однако, лишь 0,308. Это означает, что 30,8% изменчивости одной переменной может быть объяснено изменчивостью второй переменной. Иными словами, для данной выборки на 30,8% изменчивость общего балла толерантности к неопределенности объясняется уровнем эмоционального интеллекта, если полагать установленной между ними причинную связь.

В некотором смысле регрессия позволяет не только представить статистическую связь между переменными в виде простейшей формулы (хотя и не полностью объясняющей данную связь), но и данная формула имеет хоть и ограниченную, но определенную предсказательную ценность. По значению факторной переменной мы можем представить, каким может быть в идеальном случае значение результирующей переменной, если бы на нее не действовали иные факторы и зависимость имела бы в точности линейный характер.

Множественная линейная регрессия

Метод множественной регрессии отличается от простой линейной регрессии наличием не одного, а нескольких факторных признаков. Таким образом, искомое уравнение:

$$y = a_1x_1 + a_2x_2 + \dots + a_nx_n + b,$$

где x — факторные признаки со своими коэффициентами a ; b — свободный член.

Проверка гипотезы об адекватности регрессии проводится с помощью F-статистики Фишера с числом степеней свободы ($k, n - k$).

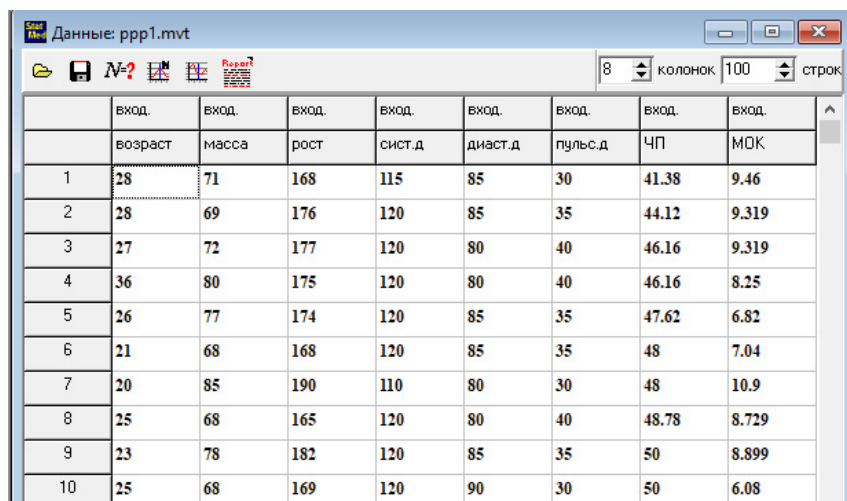
Показателем качества подгонки регрессии является коэффициент множественной детерминации R^2 , фактически представляющий собой квадрат коэффициента корреляции, изменяющийся соответственно в пределах от 0 до 1 и призванный численно характеризовать долю выборочной дисперсии результирующего признака в зависимости от факторных признаков, которые объясняются (выражаются) с помощью регрессии.

Чем ближе R^2 к 1, тем лучше качество подгонки регрессии. Однако обычно используют коэффициент множественной детерминации, скорректированный на число степеней свободы, поскольку добавление каждой независимой (факторной) переменной формально привело бы к увеличению R^2 , при этом ничуть не улучшив качество самой модели.

Как и в случае простой регрессии, исследователь решает, какая именно переменная будет результирующей (зависимой), а какие переменные — факторными (независимыми), то есть теми, которые, собственно, должны носить вклад в единственную зависимую переменную.

Рассмотрим пример применения данного метода в статистическом пакете «Статмед».

Итак, 100 испытуемым-мужчинам (на рис. 4.47 показано для экономии места только первые 10) с известным паспортным возрастом (от 20 до 40 лет) измерили 7 параметров (рост, вес и легко измеряемые и рассчитываемые в домашних условиях показатели сердечно-сосудистой системы — систолическое и диастолическое давление крови, пульсовое давление, частота пульса, минутный объем крови).



| | Вход. | Вход. | Вход. | Вход. | Вход. | Вход. | Вход. | Вход. |
|----|---------|-------|-------|--------|---------|---------|-------|-------|
| | возраст | масса | рост | сист.д | диаст.д | пульс.д | ЧП | МОК |
| 1 | 28 | 71 | 168 | 115 | 85 | 30 | 41.38 | 9.46 |
| 2 | 28 | 69 | 176 | 120 | 85 | 35 | 44.12 | 9.319 |
| 3 | 27 | 72 | 177 | 120 | 80 | 40 | 46.16 | 9.319 |
| 4 | 36 | 80 | 175 | 120 | 80 | 40 | 46.16 | 8.25 |
| 5 | 26 | 77 | 174 | 120 | 85 | 35 | 47.62 | 6.82 |
| 6 | 21 | 68 | 168 | 120 | 85 | 35 | 48 | 7.04 |
| 7 | 20 | 85 | 190 | 110 | 80 | 30 | 48 | 10.9 |
| 8 | 25 | 68 | 165 | 120 | 80 | 40 | 48.78 | 8.729 |
| 9 | 23 | 78 | 182 | 120 | 85 | 35 | 50 | 8.899 |
| 10 | 25 | 68 | 169 | 120 | 90 | 30 | 50 | 6.08 |

Рис. 4.47

Проверку на нормальность распределения опускаем также ради экономии места, но мы удостоверились, что распределение ни одной из переменных значимо не отличается от нормального. Выбираем пункт «Множественная регрессия (линейная модель)» (рис. 4.48).

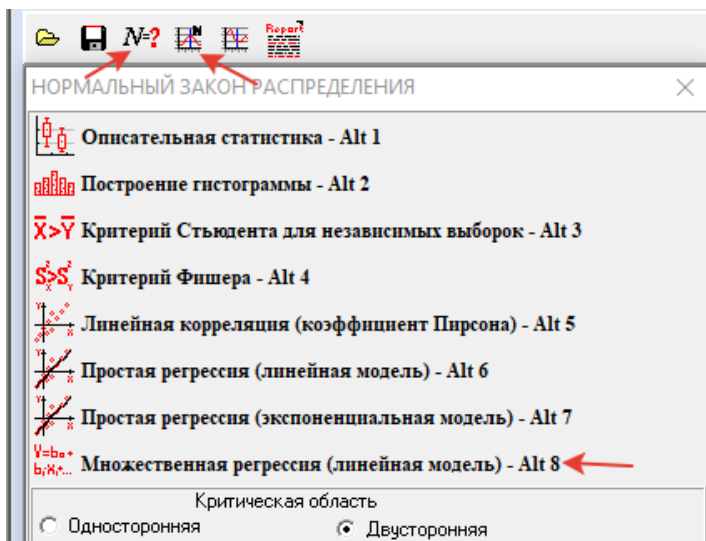


Рис. 4.48

Результаты множественной регрессии представлены на рис. 4.49.

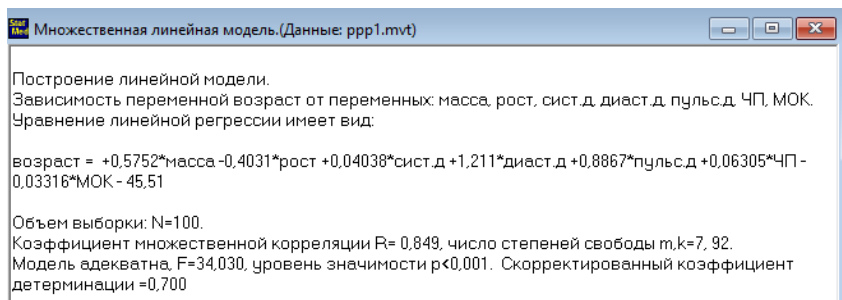


Рис. 4.49

Данный результат может быть интерпретирован следующим образом.

Возраст мужчин от 20 до 40 лет (с вполне определенными коэффициентами при каждой из переменных) тем больше, чем больше масса тела, меньше рост, больше артериальное давление, больше частота пульса и меньше минутный объем крови). Иными словами, с возрастом в среднем определенным образом изменяются эти показатели. Теперь, если мы, допустим, измерим у испытуемого с неизвестным возрастом эти показатели, то можем приблизительно определить его возраст.

Ценность регрессионных моделей может заключаться не только в том, что мы знаем, как в совокупности и в частности связана результирующая переменная с факторными, но и в прогностических (предсказательных возможностях) подобного рода вероятностных моделей. Так, в данном примере, мы, конечно, не определим на 100% возраст человека, зная остальные параметры, но, что интересно, можем получить некоторое представление о биологическом его возрасте, который, вообще говоря, зависит от биометрических показателей и состояния сердечно-сосудистой системы. Конечно же, следует с осторожностью использовать предсказательный аспект моделей, так как в таком случае нам необходимо иметь репрезентативную и хорошо рандомизированную выборку относительно релевантной генеральной совокупности. Более того, следует также с осторожностью относиться к предсказательной функции любой модели, особенно, как в данном случае, статистической. Помимо того, что зависимости могут носить нелинейный характер, каждый человек уникален и может просто не вписываться в характеристики модели. Например, в нашем случае получилось, что чем больше рост и меньше вес — тем более молод мужчина соответствующей возрастной категории. Однако здесь не учитывается конституция тела, наличие особых условий (занятия силовыми видами спорта, наличие заболеваний, ведущих к исхуданию, и т. п.). Именно поэтому нельзя «в лоб» пользоваться

любыми моделями, а использовать их лишь в ограниченном смысле — либо как скрининговые, либо как эксплораторные. Если же мы хотим трактовать модель как подтверждающую, нужны специальные оговорки, наряду с тщательным отбором испытуемых в выборку, тем самым описывая генеральную совокупность, которой они соответствуют: например, здоровые мужчины молодого и среднего возраста, и при этом не занятые силовыми видами спорта.

Коэффициент ранговой корреляции Спирмена

Для непараметрических данных, а также для сочетаний порядковых и номинативных данных существует довольно большой арсенал коэффициентов корреляции, которые освещены в соответствующих пособиях (подробнее об этом можно прочитать в заключении). Мы же остановимся лишь на одном из ранговых коэффициентов корреляции — Спирмена, который отличается мощностью, то есть выдает высокие значения, и рассчитывается в любом из статистических пакетов. При использовании корреляционного анализа по Спирмену предполагается, что в каждом из признаков (переменных) имеется не слишком мало различающихся между собой значений (во всяком случае, их должно быть больше двух).

Приведем пример расчета коэффициента корреляции Спирмена в программе «Статмед». Допустим, мы измерили у 30 социальных работников показатели толерантности по вопроснику В. С. Магуна (рис. 4.50), и нас интересует, существует ли связь между этнической и социальной толерантностью.

На рис. 4.50 представлены первые 10 значений переменных. Теперь проверяем переменные на нормальность распределения (рис. 4.51).







| | | |
|---|------------|------------|
| StatMed | | |
| Данные Окна Help | | |
| Данные: Noname.mvt | | |
|       | | |
| | вход. | вход. |
| | этническая | социальная |
| 1 | 76 | 44 |
| 2 | 43 | 22 |
| 3 | 84 | 61 |
| 4 | 72 | 60 |
| 5 | 80 | 57 |
| 6 | 60 | 98 |
| 7 | 100 | 53 |
| 8 | 46 | 31 |
| 9 | 84 | 50 |
| 10 | 99 | 45 |

Рис. 4.50

| | |
|---|---|
| StatMed | Проверка на нормальность (Данные: Noname.mvt) |
| <p>Переменная этническая толерантность</p> <p>Критерий W Шапиро-Уилка проверки распределения на нормальность</p> <p>Объем выборки N=30, W=0,906, уровень значимости p=0,02</p> <p>Распределение отличается от нормального на уровне значимости p=0,02</p> | |
| <p>Переменная социальная толерантность</p> <p>Критерий W Шапиро-Уилка проверки распределения на нормальность</p> <p>Объем выборки N=30, W=0,834, уровень значимости p<0,01</p> <p>Распределение отличается от нормального на уровне значимости p<0,01</p> | |

Рис. 4.51

Распределение как минимум одной переменной (а здесь даже и двух) отличается от нормального. В то же время данные количественные и имеют много разных значений, значит, применяем коэффициент корреляции Спирмена (рис. 4.52).

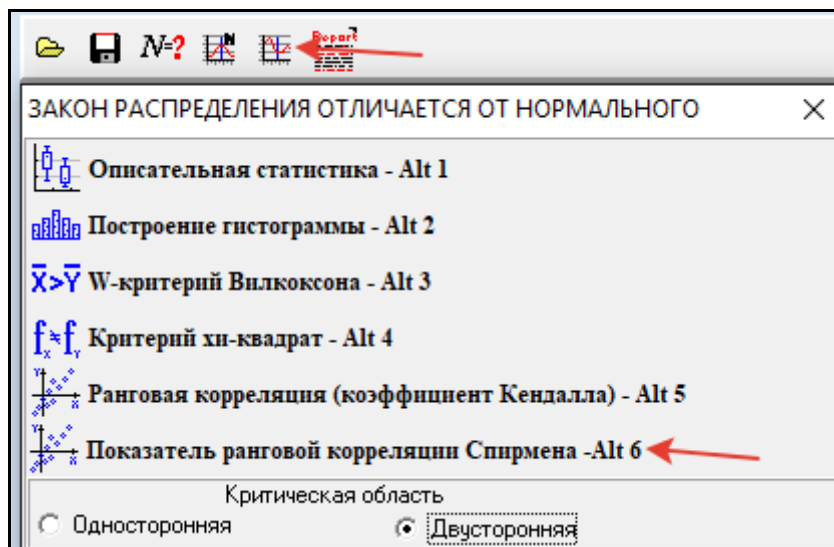


Рис. 4.52

После чего получаем итоговые результаты (рис. 4.53).

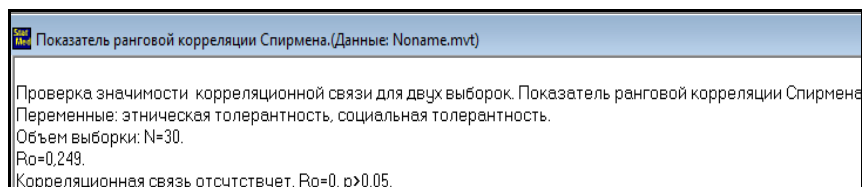


Рис. 4.53

В данном случае отсутствует значимая корреляционная связь. Коэффициент корреляции равен 0,249. Однако такое значение не является значимым, следовательно, принимаем нулевую гипотезу о том, что в группе социальных работников между социальной и этнической толерантностью отсутствует связь.

Аналогичным образом, как в примере с коэффициентом корреляции Пирсона, мы можем анализировать статистические связи сразу нескольких переменных. Для этого, например, в программе Statistica, следует проделать последовательность операций Statistics → Nonparametrics → Correlations (Spearman, Kendall tau, gamma) → Spearman Rank R (Статистики, Непараметрические статистики, Корреляции (Спирмен, тау Кендалла, гамма) → Ранговая корреляция Спирмена)», а дальше действовать аналогично, выбрав тип матрицы (единая или состоящая из двух списков), а также определив детальность вывода результатов.

В целом, мы рекомендуем пользоваться корреляцией Спирмена вследствие большей мощности и аналогией с коэффициентом корреляции Пирсона для непараметрических данных или коэффициентом корреляции гамма, когда имеется много совпадающих значений.

4.3.6. Анализ различий уровня признака в трех и более независимых выборках с нормальным распределением

Однофакторный дисперсионный анализ для независимых выборок

Довольно часто в психологических задачах приходится иметь дело с сравнением более чем двух выборок. Пожалуй, наиболее часто встречается случай с количественными данными и в случае простого сопоставления выборок, различных по некоторому фактору, поэтому данный метод называется однофакторным дисперсионным анализом (ANOVA — общепринятое англоязычное сокращение — analysis of variance). Алгоритм применения статистического анализа в данной задаче примерно следующий:

1. Проверка на нормальность всех выборок. Условие равенства объемов выборок не является обязательным, но существенные различия в них могут повлиять на параметры разброса данных, такие,

как доверительный интервал и стандартная ошибка, тем самым в ряде случаев сбив с толку исследователя. Поэтому различие объемов выборок более чем в два раза при условии, что объем наименьшей составляет не более 30 испытуемых, не является допустимым. Если самое главное условие нормальности распределения каждой выборки соблюдено, переходим к следующему пункту.

2. Рассмотрение описательной статистики (меры центральной тенденции: мода, медиана, среднее арифметическое; меры положения: максимум, минимум, квартили, процентиля; меры рассеяния: размах, дисперсия, стандартное отклонение, междуквартильный размах и др.), в том числе визуально. Часто и обоснованно эти данные приводятся в работах.

3. Применение F-критерия Фишера для оценки соотношения межгрупповой и внутригрупповой изменчивости.

4. Однако, как правило, исследователя интересует не только знание о том, что средние значения в выборках в принципе отличаются, но и сведения о конкретных различиях между каждой из пар выборок.

5. Поэтому далее может быть применено целое семейство так называемых апостериорных критериев, которые учитывают возможность появления случайных различий значений в разных выборках, поскольку их больше двух. Одним из наиболее предпочтительных критериев в данном случае является критерий Шеффе, так как он дает более широкий доверительный интервал, а также особенно пригоден в тех случаях, когда имеется подозрение о неравенстве дисперсий выборок между собой, что зачастую и происходит в психологических задачах.

6. По результатам применения апостериорного критерия исследователь имеет информацию обо всех попарных различиях между группами.

Рассмотрим самый простой вариант применения ANOVA, который реализован в программе «Медстат». Зачастую также используется программа Statistica, применение в которой данного метода в целом аналогично, но усложнено для большинства задач. Поэтому в данном случае мы советуем обратиться к руководству по работе с этой программой [16], а здесь для простоты и одновременно

с сохранением необходимой статистической строгости используем именно программу «Медстат».

Допустим, мы имеем дело с 4 выборками детей, самооценка которых была оценена методикой Дембо–Рубинштейн по параметру «ум». Выборку составили группа интеллектуально одаренных детей, группа средненормативных детей, группа детей с нарушениями интеллектуального развития и группа детей с нарушениями социальных отношений. Возможные показатели самооценки варьируются по шкале в диапазоне от 0 до 10 (рис. 4.54).






|   N=?    | | | | |
|---|------------|------------|-----------|-------------|
| | вход. | вход. | вход. | вход. |
| | дети-изгой | обычные де | одаренные | дети с инте |
| 1 | 3,6 | 4,8 | 4,2 | 4 |
| 2 | 2,6 | 3,6 | 6,4 | 3,3 |
| 3 | 6,4 | 8,1 | 5,9 | 4,4 |
| 4 | 4,1 | 7,1 | 4,3 | 5,1 |
| 5 | 4,1 | 6,5 | 4,7 | 5 |
| 6 | 4 | 4,3 | 5 | 5,4 |
| 7 | 5,9 | 4,9 | 6,2 | 2,7 |
| 8 | 5,2 | 4,2 | 5,9 | 4 |
| 9 | 3,6 | 6,9 | 6,6 | 5 |
| 10 | 3,4 | 6,8 | 5,7 | 3,4 |
| 11 | 4,4 | 7,8 | 5,6 | 3,7 |
| 12 | 3,5 | ? | 3 | 3 |
| 13 | 3,5 | ? | 4,5 | 2,9 |
| 14 | ? | ? | 3,7 | 3,5 |
| 15 | ? | ? | 3,5 | 3,7 |
| 16 | ? | ? | 4,2 | 2,7 |

Рис. 4.54

Проверяем данные на нормальность (рис. 4.55).

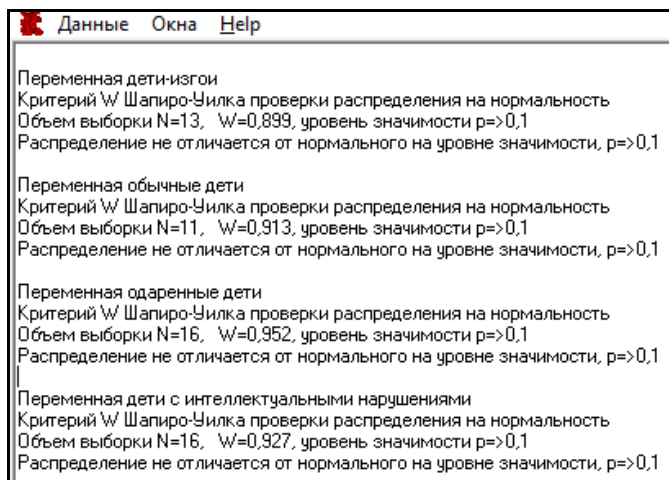


Рис. 4.55

Убеждаемся в том, что все переменные имеют нормальное распределение (рис. 4.55).

Далее полезно ознакомиться с описательной статистикой для представления и количественного описания данных. Опустим числовые данные и рассмотрим лишь график (рис. 4.56).

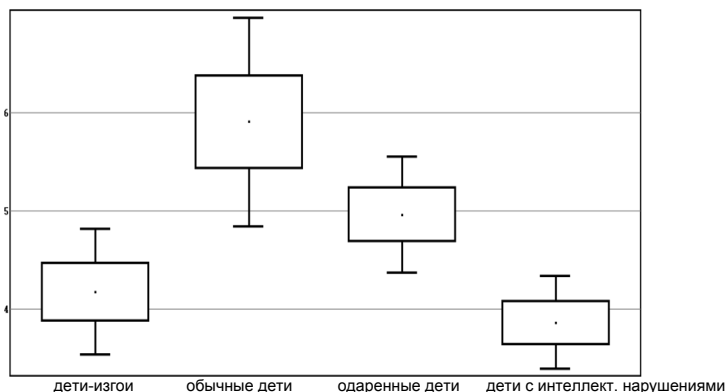


Рис. 4.56

Как мы помним, точки — это средние арифметические, «коробки» — стандартные ошибки, а «усы» — доверительные интервалы.

Далее применяем множественные сравнения в меню для параметрических выборок (рис. 4.57).

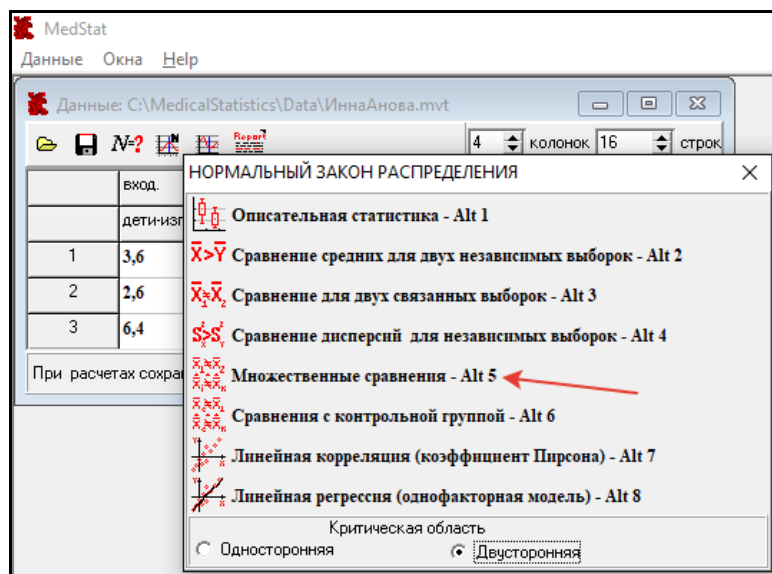


Рис. 4.57

После этого получаем результаты анализа (см. рис. 4.58).

Видно, что сначала выявлено различие в средних у выборок с высокой степенью значимости $p < 0,001$. Затем применяется критерий Шеффе для определения попарных различий. Различия значимы в самооценке у детей-изгоев и обычных детей, а также у обычных детей и детей с интеллектуальными нарушениями. Самооценка детей из средненормативной выборки выше, это типичная ситуация. Самооценка одаренных детей находится на среднем уровне, как правило, такие дети более чувствительны к внешним событиям и оценкам значимых людей, что может снижать их самооценку по шкале «ум». Дети с интеллектуальными нарушениями, как правило,

оценивают себя невысоко, если они не изолированы от других детей. Дети с нарушениями социальных отношений оценивают себя ниже среднего, но не предельно низко.

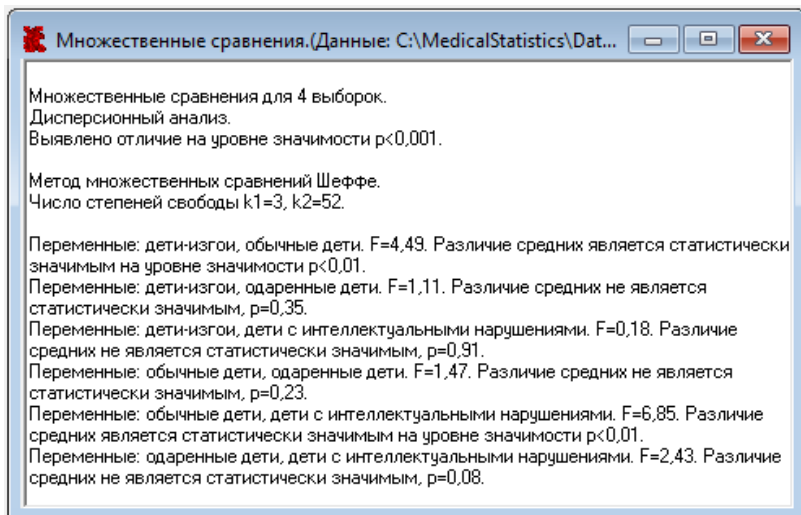


Рис. 4.58

Многофакторный дисперсионный анализ для выборок с нормальным распределением

Помимо однофакторного дисперсионного анализа довольно часто перед исследователями стоит задача выявить зависимости средних арифметических в выборках более чем от одного фактора (двух и более). Наиболее часто встречается двухфакторный анализ. Если факторов больше, то порядок действий и рассуждений никак не меняется.

Например, рассмотрим задачу зависимости эмоционального выгорания у силовиков от стажа службы (до и более 5 лет) и пола. Предположим, что оба фактора могут играть роль в особенностях выгорания, поскольку имеются литературные данные о том, что стаж 5-6 лет является критическим для формирования эмоциональ-

ного выгорания сотрудников силовых структур. При этом, если не учесть какой-либо из факторов, то есть учитывать лишь один из них, можно упустить важную зависимость как от пола, так и от боевого опыта.

Данный пример рассмотрим в программе Statistica. Допустим, что мы убедились в нормальности распределения интересующей нас зависимой переменной, в то время как — пол и стаж — это факторные переменные.

Исследование прошли 264 испытуемых, из них 128 женщин (48 со стажем до 6 лет и 80 — от 6 лет) и 136 мужчин (64 со стажем до 6 лет и 72 — от 6 лет). Для того чтобы выбрать интересующий нас тип анализа после ввода данных, следует нажать последовательно Statistics → ANOVA → Factorial ANOVA. Затем в открывшемся меню выбрать зависимую (dependent) переменную (variable) и категориальные факторы-предикторы (categorical factors-predictors), после чего будет получено общее окно с выводом разнообразных результатов (рис. 4.59).

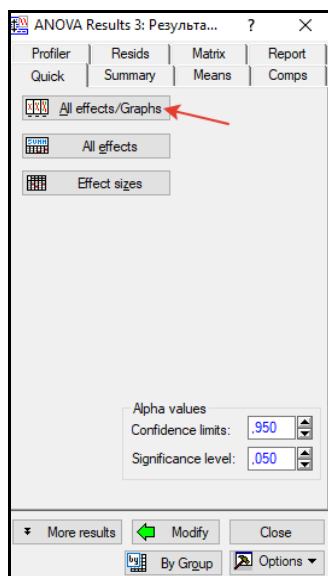


Рис. 4.59

Далее (как вариант обработки) нажимаем на кнопку All effects/Graphs (Все эффекты/графики) (рис. 4.60).

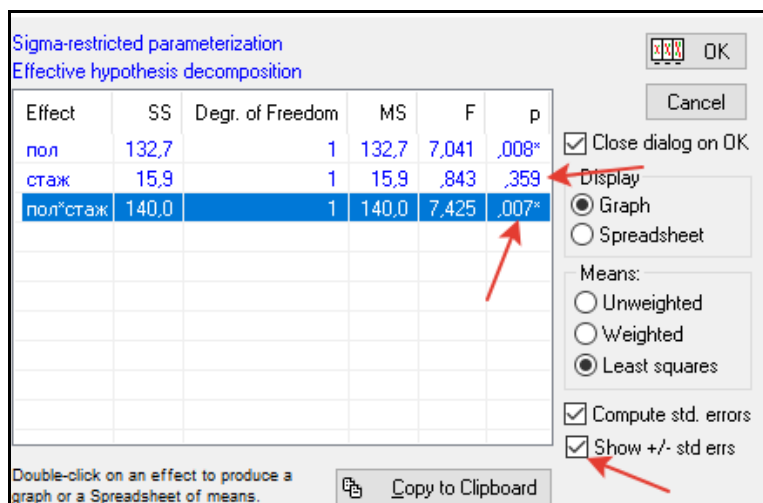


Рис. 4.60

Сначала обратим внимание, что если бы мы рассматривали только переменную «стаж» как фактора без учета влияния фактора пола на зависимую переменную, то получили бы превратное представление о явлении (рис. 4.61).

То есть различия не были бы существенными и значимыми, следовательно, мы бы полагали, что стаж, в сущности, не влияет на показатель «тревога–депрессия». На самом же деле при учете второго фактора (пол) перед исследователем открывается совсем иная, более адекватная картина явления: у мужчин со стажем этот показатель растет, в то время как у женщин, наоборот, имеет некоторую тенденцию к уменьшению (рис. 4.62). И это при том, что у женщин с меньшим стажем профессиональной деятельности в силовых структурах показатель «тревога–депрессия» намного выше, чем у мужчин, а у женщин с большим стажем он полностью выравнивается и практически не отличается от показателя мужчин.

Current effect: $F(1, 260)=.84287, p=.35943$

Effective hypothesis decomposition

Vertical bars denote \pm standard errors

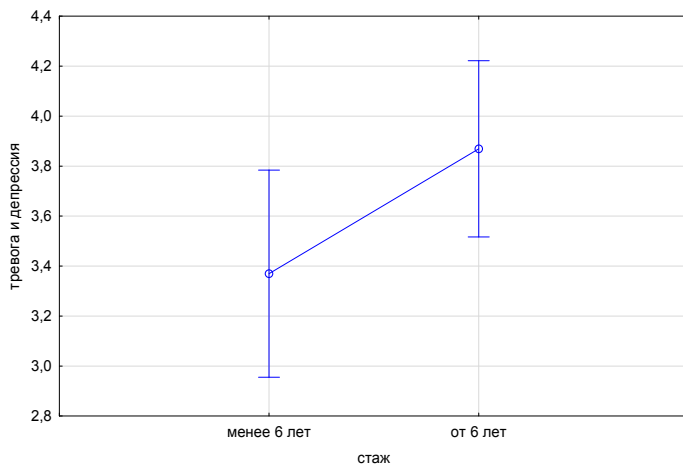


Рис. 4.61

пол*стаж; LS Means

Current effect: $F(1, 260)=7.4251, p=.00687$

Vertical bars denote \pm standard errors

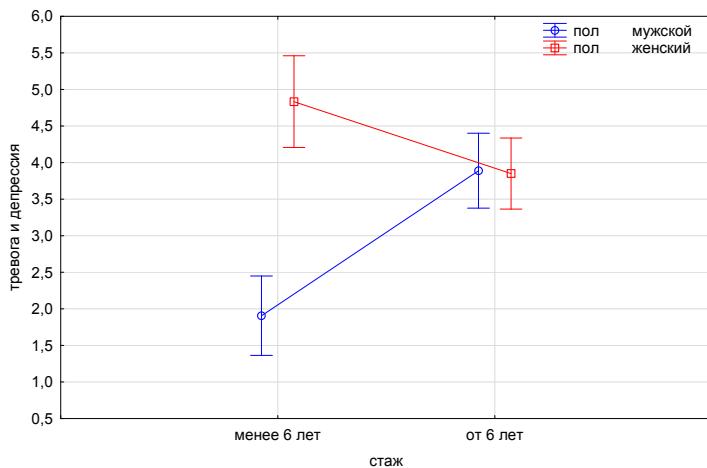


Рис. 4.62

Естественно, исследователя интересует не только графическое, но и численное представление результатов. Для этого необходимо открыть вкладку Means (Средние) и нажать на кнопку Observed, unweighted (Наблюдаемые, невзвешенные), поскольку в данном случае мы не придавали особого статистического веса какой-либо части наблюдений (чаще всего так оно и бывает) (рис. 4.63).

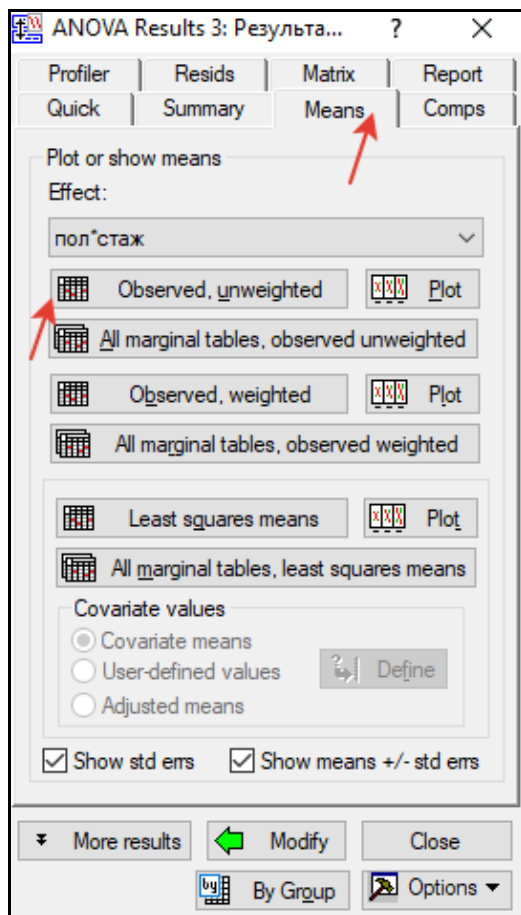


Рис. 4.63

В результате мы получаем статистику, представленную в численном виде, для каждой из подвыборок (рис. 4.64).

| Cell No. | пол | стаж | тревога и депрессия Mean | тревога и депрессия Std.Err. | тревога и депрессия -Std.Err | тревога и депрессия +Std.Err | N |
|----------|---------|-------------|-----------------------------|---------------------------------|---------------------------------|---------------------------------|----|
| 1 | мужской | менее 6 лет | 1,906250 | 0,542730 | 1,363520 | 2,448980 | 64 |
| 2 | мужской | от 6 лет | 3,888889 | 0,511691 | 3,377198 | 4,400580 | 72 |
| 3 | женский | менее 6 лет | 4,833333 | 0,626691 | 4,206642 | 5,460024 | 48 |
| 4 | женский | от 6 лет | 3,850000 | 0,485433 | 3,364567 | 4,335433 | 80 |

Рис. 4.64

Однако нас интересуют также апостериорные сравнения, которые позволяют провести анализ статистической значимости различий между парами групп. Для этого нажимаем на кнопку More results (Больше результатов) (рис. 4.65) и на вкладку Post-hoc (Апостериорные сравнения) (см. рис. 4.66).

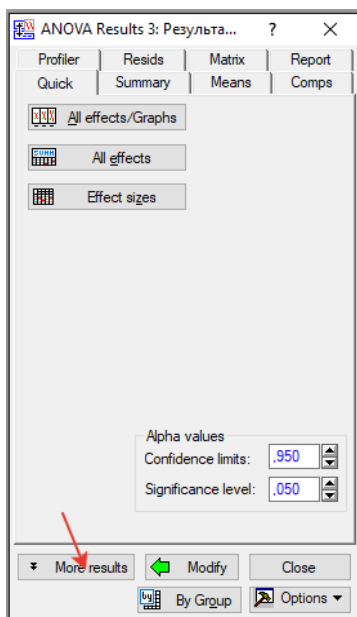


Рис. 4.65

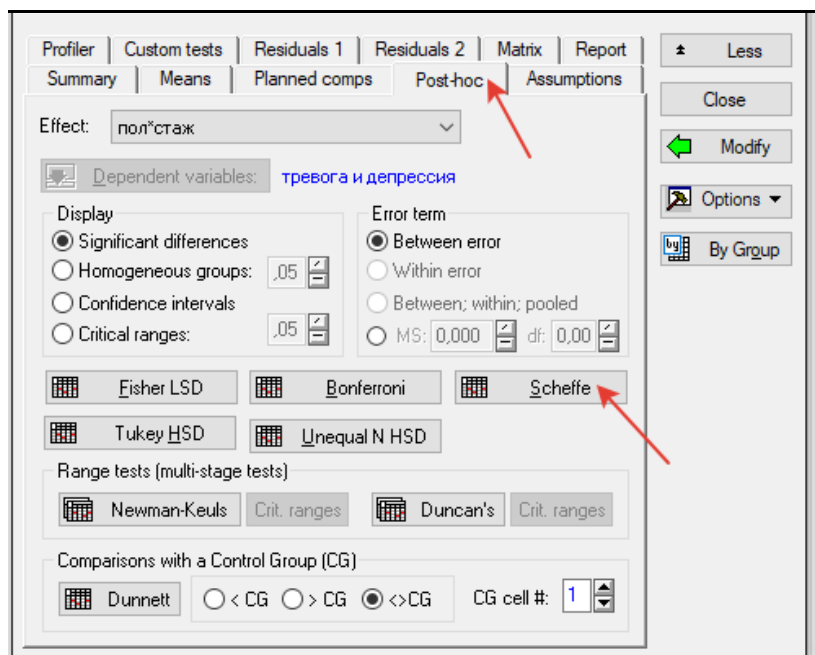


Рис. 4.66

Выбираем критерий Шефе (Scheffe) и получаем матрицу различий между всеми выборками попарно (рис. 4.67).

| Scheffe test; variable тревога и депрессия | | | | | | |
|--|---------|-------------|---------------|---------------|---------------|---------------|
| Probabilities for Post Hoc Tests | | | | | | |
| Error: Between MS = 18,852, df = 260,00 | | | | | | |
| Cell No. | пол | стаж | {1} 1,9063 | {2} 3,8889 | {3} 4,8333 | {4} 3,8500 |
| 1 | мужской | менее 6 лет | | 0,008348 | 0,000490 | 0,008076 |
| 2 | мужской | от 6 лет | 0,008348 | | 0,244141 | 0,956072 |
| 3 | женский | менее 6 лет | 0,000490 | 0,244141 | | 0,215919 |
| 4 | женский | от 6 лет | 0,008076 | 0,956072 | 0,215919 | |

Рис. 4.67

Из анализа таблицы можно сделать вывод, что высоко значимые попарные различия наблюдаются между средними групп «мужчины со стажем менее 6 лет» — «мужчины со стажем от 6 лет» ($p = 0,008348$), «мужчины со стажем менее 6 лет» — «женщины со стажем менее 6 лет» ($p = 0,000490$), «мужчины со стажем менее 6 лет» — «женщины со стажем от 6 лет» ($p = 0,008076$). Различия же между средними групп «женщины со стажем менее 6 лет» — «женщины со стажем от 6 лет» не значимы ($p = 0,215919$). Однако эту тенденцию также можно обсудить в результатах как предположение для дальнейших исследований о том, что если мужчины со стажем склонны становиться более депрессивно-тревожными, то у женщин может иметься иная тенденция. Еще один заслуживающий внимания результат — это выравнивание абсолютных показателей величин показателя «тревога–депрессия» при увеличении стажа (см. рис. 4.64), что позволяет говорить о гомогенности группы представителей обеих полов в отношении показателя «тревога–депрессия» с увеличением срока службы в силовых структурах.

4.3.7. Анализ различий уровня признака в трех и более независимых выборках с отличным от нормального законом распределения

Мы имеем ограниченное право применять мощные параметрические методы лишь в случае очень больших выборок, распределение которых существенно отлично от нормального (очень грубо — от нескольких сотен значений в выборке, в силу центральной предельной теоремы), и полное право — при неотличимости наличного распределения от нормального, исходя из соответствующего статистического критерия. Однако на практике часто встречаются небольшие выборки, распределение которых существенно отличается от нормального, в том числе, и в случае сравнения более чем двух выборок. Поэтому в этом пункте мы рассмотрим непараметрический аналог однофакторного дисперсионного анализа. Следует отметить, что часто мы *изначально* сталкиваемся с рангами (порядковыми номерами) в психологических задачах. Такую задачу мы и рассмотрим в качестве примера в программе «Медстат».

В олимпиаде по физике участвовали одаренные представители трех школ (40 человек). Нас интересует наличие или отсутствие различий в уровне мастерства учителей (полагаем его в качестве фактора) исходя из успешности участников олимпиады, судя по рангу (чем ранг выше, тем выше успешность ученика — в данном случае мы произвольно выбрали такой порядок ранжирования). В качестве данных представлено количество баллов по олимпиаде, набранное учениками (рис. 4.68).

Данные: C:\MedicalStatistics\Data\anovano

Иконки: $N=?$

| | вход. | вход. | вход. |
|----|---------|---------|---------|
| | школа 1 | школа 2 | школа 3 |
| 1 | 40 | 1 | 4 |
| 2 | 7 | 17 | 6 |
| 3 | 2 | 10 | 8 |
| 4 | 19 | 14 | 33 |
| 5 | 12 | 9 | 32 |
| 6 | 27 | 15 | 16 |
| 7 | 36 | 11 | 13 |
| 8 | 31 | 25 | 8 |
| 9 | 35 | 20 | 18 |
| 10 | 39 | 3 | 30 |
| 11 | 26 | 5 | 24 |
| 12 | 34 | 24 | 21 |
| 13 | 37 | 23 | 20 |
| 14 | 15 | ? | 22 |
| 15 | 25 | ? | 28 |
| 16 | 29 | ? | ? |

Рис. 4.68

Данная шкала — изначально ранговая, проверка на нормальность также показала отсутствие соответствия нормальности. Поэтому мы применяем непараметрический аналог дисперсионного анализа (рис. 4.69).

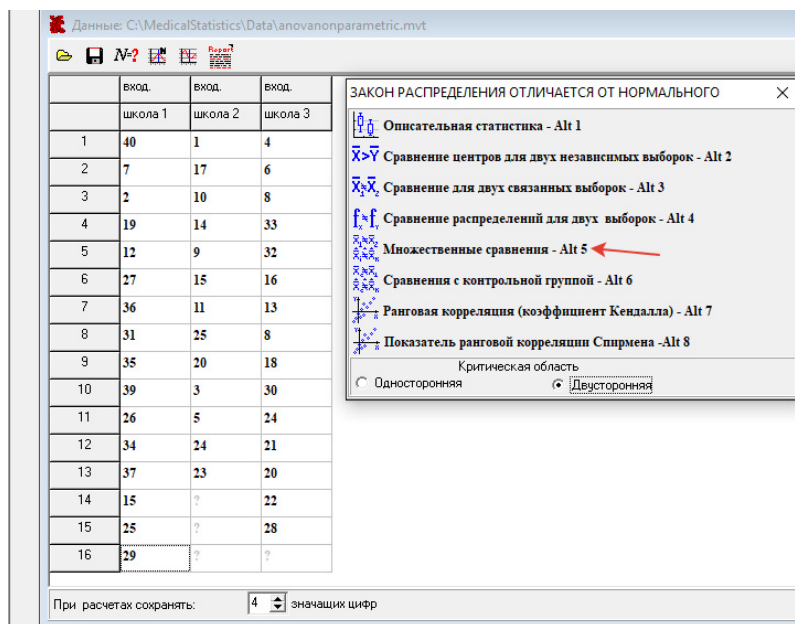


Рис. 4.69

Как известно, описательную статистику мы можем представить не только в виде чисел (рис. 4.70), но и графика (см. рис. 4.71).

| Переменная | Кол-во | Медиана | I квартиль | III квартиль | Минимум | Максимум | Ош. медианы | Лев.(95% ДИ) | Прав.(95% ДИ) |
|------------|--------|---------|------------|--------------|---------|----------|-------------|--------------|---------------|
| школа 1 | 16 | 28 | 17 | 35,5 | 2 | 40 | 3,681 | 19 | 35 |
| школа 2 | 13 | 14 | 9 | 20 | 1 | 25 | 2,779 | 5 | 23 |
| школа 3 | 15 | 20 | 8 | 28 | 4 | 33 | 3,104 | 8 | 28 |

Рис. 4.70

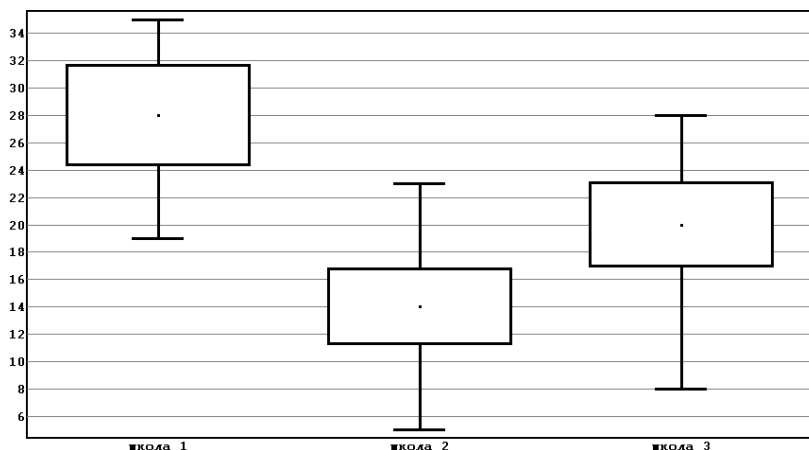


Рис. 4.71

На рисунке в виде «коробок» представлена ошибка медианы, а в виде «усов» — доверительный интервал для непараметрических данных [7].

Непараметрический аналог ANOVA, а именно — критерий Крускала–Уоллиса, определил наличие значимых различий между успешностью представителей разных школ на уровне значимости $p = 0,011$ (рис. 4.72).

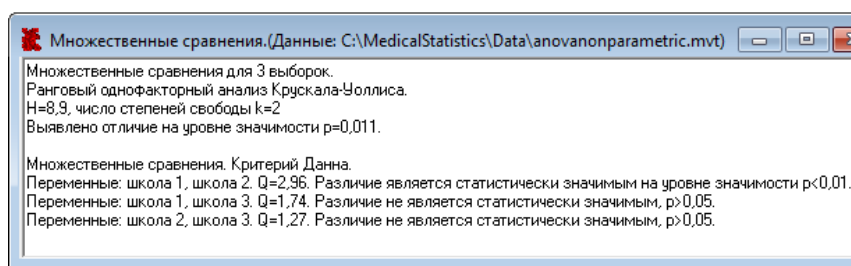


Рис. 4.72

Апостериорный критерий Данна далее автоматически применяется программой для попарных сравнений. С его помощью устанавливается, что высоко значимые различия имеются лишь между успеваемостью учеников из 1-й и 2-й школ, при этом ученики 1-й школы лучше написали олимпиадные задания, что следует из описательной статистики. В соответствии с гипотезой исследования делаем вывод о лучшей подготовке учеников лишь 1-й школы, при этом остальные попарные различия не являются статистически значимыми. В целом подтвердились следующие альтернативные гипотезы: о различном уровне успешности учеников и о том, что учитель первой школы более эффективен при подготовке к олимпиаде учеников, чем второй.

4.3.8. Анализ различий уровня признака при повторных измерениях у одной и той же выборки

Как и в случае двух выборок, мы можем иметь дело с повторными измерениями в нескольких зависимых выборках, а именно, в простейшем и наиболее частом на практике случае — повторными измерениями, проводимыми на одной и той же выборке.

Дисперсионный анализ с повторными измерениями

Если данные имеют нормальное распределение и измерены в интервальной шкале или шкале отношений, следует использовать дисперсионный одно- или многофакторный анализ, в зависимости от количества факторных переменных. Рассмотрим в качестве примера однофакторную модель, поскольку в случае нескольких факторов особенности применения аналогичны тем, что уже были рассмотрены выше в примере с многофакторным дисперсионным анализом для выборок с нормальным распределением.

Будем так же рассматривать ниже приведенный пример в программе Statistica.

Допустим, директора сети магазинов заинтересовало, какой из тренинг-менеджеров более эффективно и устойчиво обучает

менеджеров по работе с клиентами в торговых залах. В качестве одного из показателей взяли показатель «активность» (по шкале от 1 до 7) опросника САН (самочувствие–активность–настроение), который заполняли испытуемые до тренинга, после каждой из сессий и через три месяца после его завершения, поскольку нас интересует, насколько устойчивым будет эффект от тренинга.

Факторов может быть больше, мы могли бы еще учесть пол, возраст испытуемого, пол тренера, метод тренинга и т. д. Впрочем, мы можем рассматривать каждый фактор и отдельно, без учета остальных.

Вначале были случайным образом сформированы выборки, уравновешенные по полу, возрасту, опыту работы менеджеров, а также по среднему значению и разбросу изучаемой переменной (измерение «до тренинга»). Это было проверено с помощью критериев Стьюдента и Фишера. Каждому из испытуемых был присвоен в качестве факторной переменной свой тренинг-менеджер.

Как мы увидим ниже, на самом деле в данном анализе мы получим гораздо больше результатов, чем изначально требуется.

После формирования таблицы, где каждая строка соответствует испытуемому, выбираем необходимый тип анализа (рис. 4.73).

Итак, в главном меню выбираем ANOVA, то есть дисперсионный анализ.

Далее выбираем Repeated measures ANOVA, то есть дисперсионный анализ с повторными измерениями (см. рис. 4.74).

Далее нам предстоит выбрать зависимые переменные (то есть переменные, содержащие значения признаков) и категориальную переменную — предиктор. В данном случае 1 и 2 — это, соответственно, условные обозначения, кодирующие разных тренеров (см. рис. 4.75).

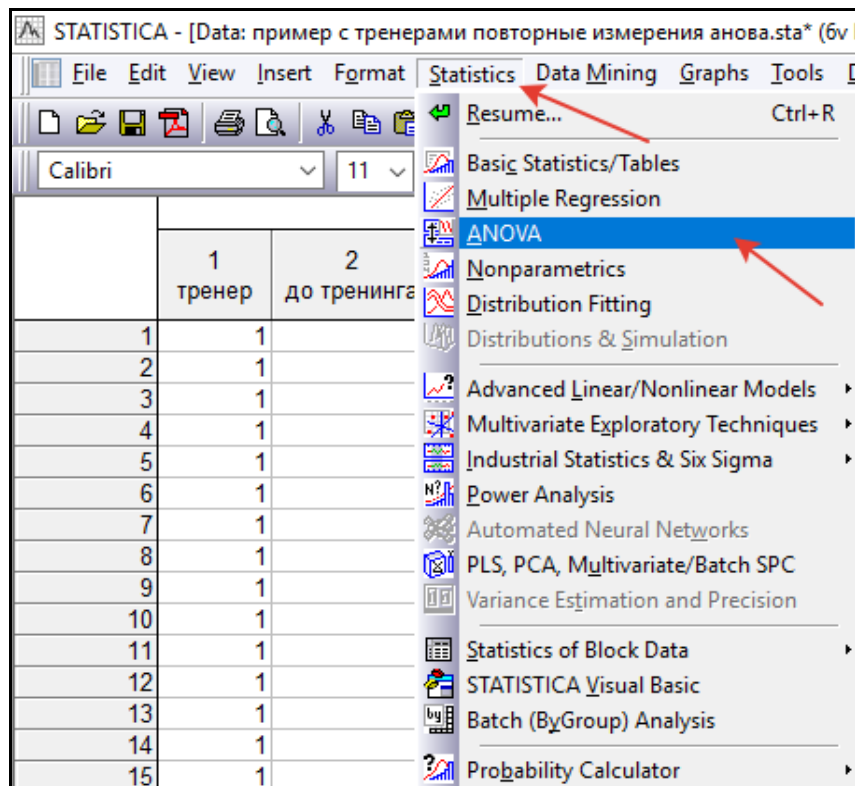


Рис. 4.73

| | 1 тренинг | 2 до тренинга | 3 после 1 сессии | 4 после 2 сессии | 5 после 3 сессии | 6 через 3 месяца | | |
|----|--------------|------------------|------------------------|------------------------|------------------------|------------------------|--|--|
| 1 | 1 | 3 | 3 | 4 | 5 | 5 | | |
| 2 | 1 | 1 | 3 | 3 | 4 | 4 | | |
| 3 | 1 | 1 | 2 | 3 | 4 | 3 | | |
| 4 | 1 | 3 | 3 | 2 | 5 | 5 | | |
| 5 | | | | | | | | |
| 6 | | | | | | | | |
| 7 | | | | | | | | |
| 8 | | | | | | | | |
| 9 | | | | | | | | |
| 10 | | | | | | | | |
| 11 | | | | | | | | |
| 12 | | | | | | | | |
| 13 | | | | | | | | |
| 14 | | | | | | | | |
| 15 | | | | | | | | |
| 16 | | | | | | | | |
| 17 | | | | | | | | |
| 18 | | | | | | | | |
| 19 | | | | | | | | |
| 20 | | | | | | | | |
| 21 | 2 | 3 | 2 | 5 | 3 | 5 | | |
| 22 | 2 | 4 | 4 | 3 | 4 | 5 | | |
| 23 | 2 | 1 | 0 | 2 | 6 | 5 | | |
| 24 | 2 | 1 | 3 | 3 | 5 | 5 | | |
| 25 | 2 | 1 | 3 | 2 | 4 | 4 | | |
| 26 | 2 | 2 | 2 | 3 | 4 | 4 | | |
| 27 | 2 | 4 | 2 | 3 | 4 | 6 | | |
| 28 | 2 | 1 | 5 | 4 | 3 | 6 | | |
| 29 | 2 | 2 | 3 | 2 | 4 | 3 | | |
| 30 | 2 | 4 | 2 | 3 | 4 | 5 | | |

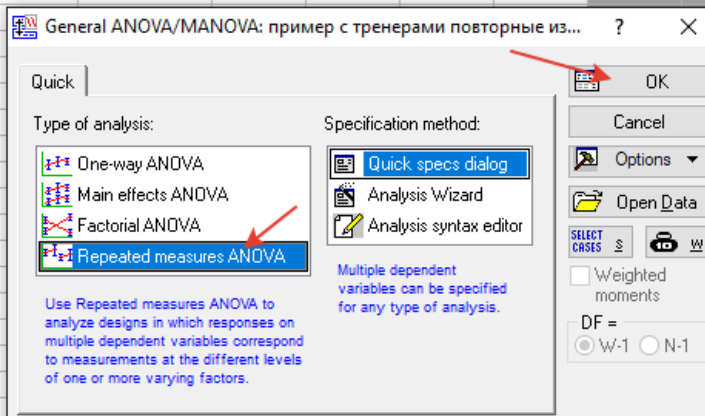


Рис. 4.74

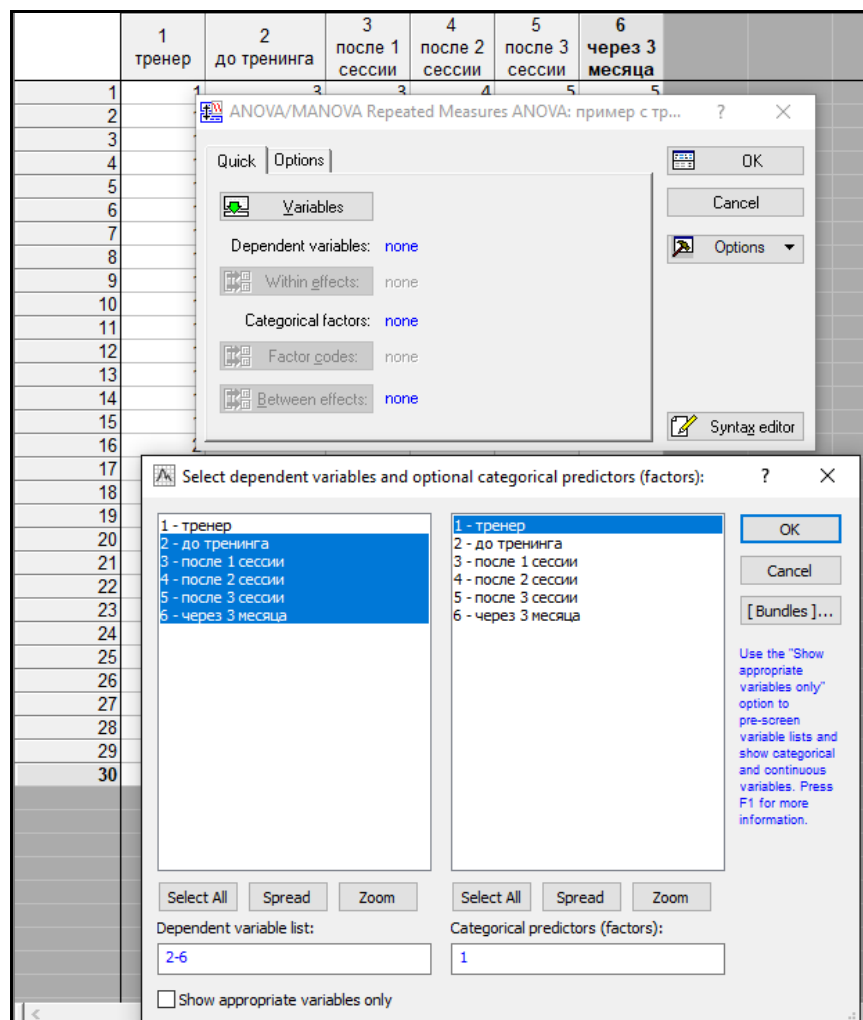


Рис. 4.75

Далее предстоит выбрать переменные, ответственные за внутригрупповые эффекты Within effects. В выпадающем меню после нажатия на «ОК», выбираем количество уровней (в данном случае

повторностей, например, пять: до тренинга, после 1, 2 и 3 сессий, а также по истечении трех месяцев после окончания тренинга) (рис. 4.76).

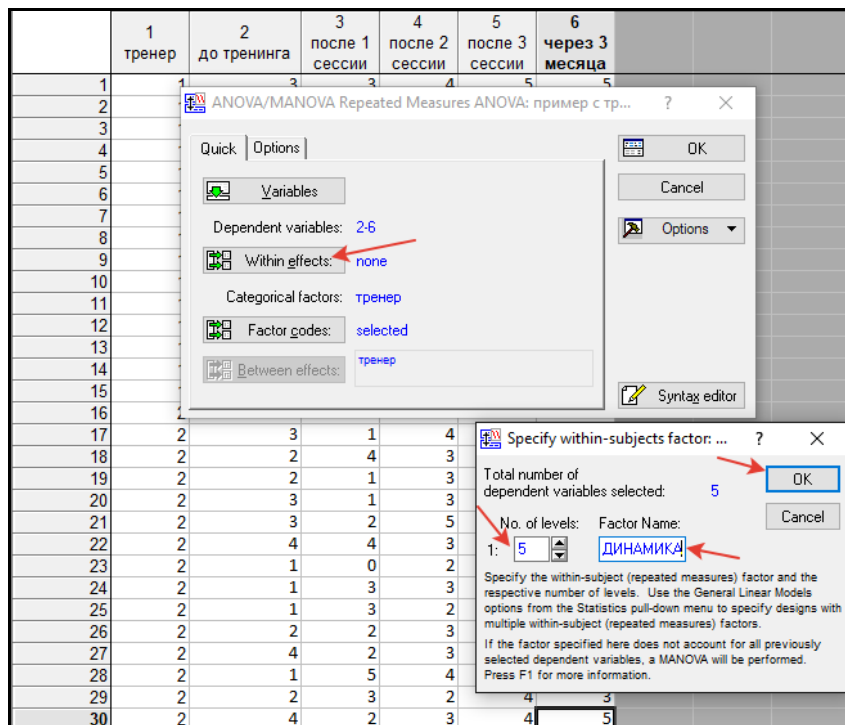


Рис. 4.76

Присвоим данному фактору имя «Динамика», поскольку речь идет об изменении состояния испытуемого.

Далее после нажатия на «ОК», можно выбрать множество вариантов представления результатов, из которых для данной задачи наиболее интересны и наглядны графические результаты, совмещенные со статистическими эффектами All effects/Graphs (рис. 4.77).

| | 1 тренер | 2 до тренинга | 3 после 1 сессии | 4 после 2 сессии | 5 после 3 сессии | 6 через 3 месяца | |
|----|-------------|------------------|------------------------|------------------------|------------------------|------------------------|--|
| 1 | 1 | 3 | | | | | |
| 2 | 1 | 1 | | | | | |
| 3 | 1 | 1 | | | | | |
| 4 | 1 | 3 | | | | | |
| 5 | 1 | 2 | | | | | |
| 6 | 1 | 2 | | | | | |
| 7 | 1 | 5 | | | | | |
| 8 | 1 | 1 | | | | | |
| 9 | 1 | 3 | | | | | |
| 10 | 1 | 1 | | | | | |
| 11 | 1 | 1 | | | | | |
| 12 | 1 | 5 | | | | | |
| 13 | 1 | 2 | | | | | |
| 14 | 1 | 1 | | | | | |
| 15 | 1 | 5 | | | | | |
| 16 | 2 | 5 | | | | | |
| 17 | 2 | 3 | | | | | |
| 18 | 2 | 2 | | | | | |
| 19 | 2 | 2 | | | | | |
| 20 | 2 | 3 | | | | | |
| 21 | 2 | 3 | | | | | |
| 22 | 2 | 4 | | | | | |
| 23 | 2 | 1 | | | | | |
| 24 | 2 | 1 | | | | | |
| 25 | 2 | 1 | | | | | |
| 26 | 2 | 2 | | | | | |
| 27 | 2 | 4 | | | | | |
| 28 | 2 | 1 | | | | | |
| 29 | 2 | 2 | | | | | |
| 30 | 2 | 4 | | | | | |

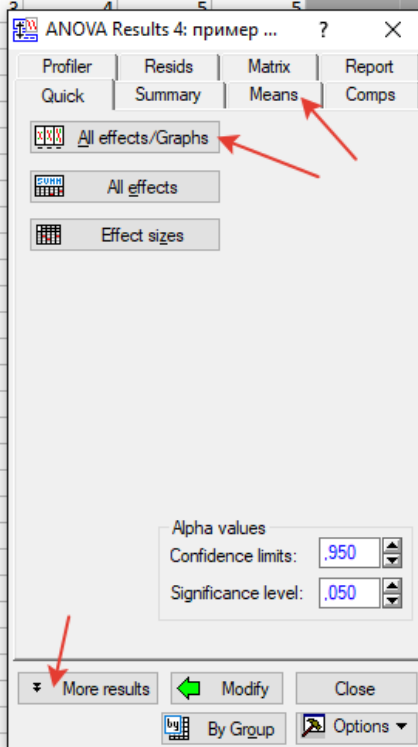


Рис. 4.77

В выпадающем меню (см. рис. 4.78) можно поставить галочку, чтобы на графиках вместо доверительных интервалов были представлены стандартные ошибки. В ряде случаев это удобно, так как довольно часто наглядно дает представление о статистической значимости различий: зачастую (но не всегда), если визуально «усы» стандартных ошибок не пересекаются между собой, то по крите-

рию Стьюдента имеются значимые различия между значением признака в соответствующих группах.

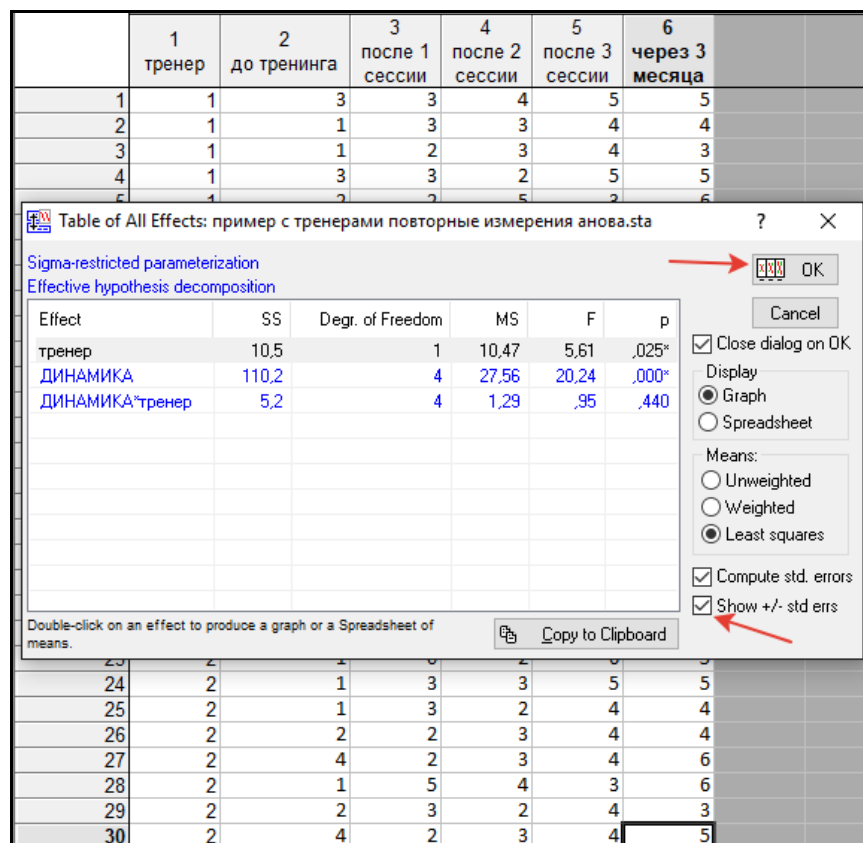


Рис. 4.78

Можно увидеть (рис. 4.79), что первый тренинг-менеджер в целом эффективнее второго (по параметру «активность»), но здесь не видна динамика.

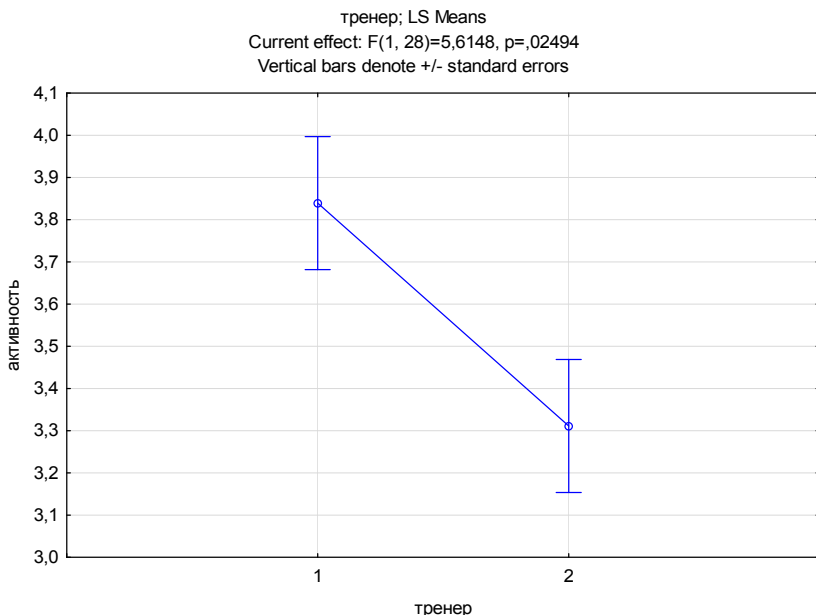


Рис. 4.79

С другой стороны, наблюдается очень высоко значимый эффект со статистической значимостью $p < 10^{-5}$, от работы каждого из тренинг-менеджеров (см. рис. 4.80), но не ясно, насколько работа одного из тренинг-менеджеров эффективнее другого на каждом из этапов.

Теперь (см. рис. 4.81) при объединении двух факторов «Динамика*Тренер» можно видеть всю картину полностью: результат работы каждого из тренинг-менеджеров на каждом этапе, а также значения активности перед тренингами и по истечении трех месяцев после их окончания (оценка посттренингового эффекта для задач консультативной психологии — это проверка экологической валидности).

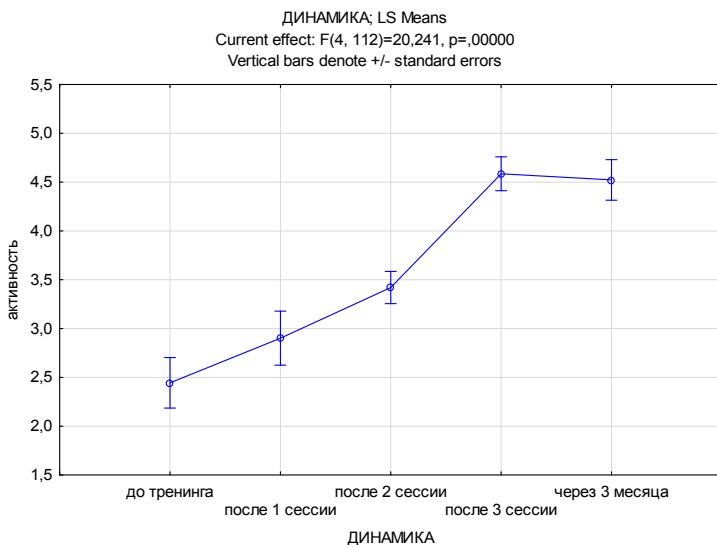


Рис. 4.80

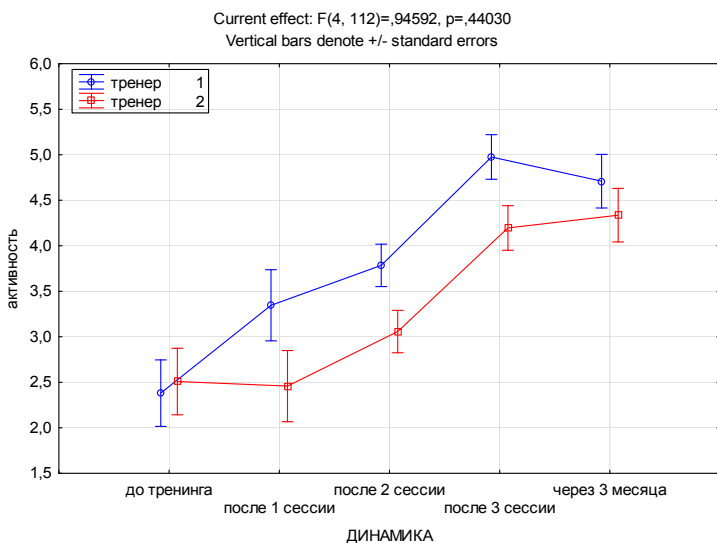


Рис. 4.81

Однако на основании данного объединения факторов мы не можем сказать, что с учетом этих этапов работа одного тренера эффективнее другого, поскольку уровень статистической значимости различий совокупного фактора «Динамика*Тренер» равен $p = 0,44$. Впрочем, это и не требовалось по условию исследования.

Тем не менее, очевидно, что первый тренер результативнее непосредственно после каждой из сессий, в то время как результирующий эффект от воздействия практически не отличается у обоих тренинг-менеджеров. Более того, если активность несколько уменьшилась через три месяца после работы первого тренинг-менеджера, то она несколько увеличилась через то же время после работы второго тренинг-менеджера. Таким образом, эффективность первого тренинг-менеджера сильнее выражена на этапах самого тренинга, но затем активность менеджеров несколько уменьшается, а эффективность второго меньше на этапах тренинга, но дает некоторый рост активности менеджеров уже после его окончания.

Столь подробный разбор этих деталей необходим для того, чтобы, принимая решение о выборе тренинг-менеджера, директор (заказчик тренингов) мог бы точнее оценивать потребности компании. Если компанию интересует кратковременный эффект, то, безусловно, следует нанимать первого тренинг-менеджера. Если же важен продолжительный эффект — то нет никакой разницы между эффективностью от работы тренинг-менеджеров.

Непараметрический аналог дисперсионного анализа с повторными измерениями. Коэффициент конкордации как мера согласия экспертных оценок

Дисперсионный анализ Фридмана является альтернативой для данных с отличным от нормального распределением или для порядковых данных с одним лишь внутригрупповым фактором с повторными измерениями (в дисперсионном анализе с повторными измерениями можно еще вводить межгрупповые факторы, в предыдущем примере таким межгрупповым фактором был тренинг-менеджер).

Примечательно, что в этом же блоке анализа в программе Statistica выдается результат вычисления коэффициента конкордации Кендалла, который является, по сути, усредненной ранговой корреляцией и позволяет, в частности, определить уровень согласованности экспертных оценок относительно друг друга.

Рассмотрим следующий пример в программе Statistica.

Участники олимпиады по психологии в университете, 15 школьников 10-11 классов, состязаются в конкурсах, которые требуют предметной компетентности и коммуникативных навыков. Преподаватели психологии выступают в качестве экспертов (12 человек) и оценивают способность и потенциальную готовность поступить на направление «Психология». Школьники ранжируются от минимального до максимального значения «способности и готовности», в данном случае, от 1 до 15 места/ранга. Причем, чем успешнее участник, тем больше его ранг. Таким образом, самый успешный участник получает «оценку» 15, а неуспешный — «оценку» 1. Очевидно, что раз речь идет изначально о порядковых данных, мы должны использовать непараметрические методы.

Статистически задача состоит в том, чтобы оценить, различаются ли участники по успешности, а также определить, насколько согласованы мнения экспертов относительно успешности участников олимпиады.

Для оценки согласованности мнений экспертов используют коэффициент конкордации (Кендалла). Если мнения экспертов хорошо согласованы, то его значение должно быть в пределах 0,6-1,0. Если мнения экспертов плохо согласованы, то его значение в пределах 0,0-0,4. Соответственно, граничные значения коэффициента конкордации — от нуля до единицы. Впрочем, специфика задачи и толкование исследователя также играют важную роль, ведь сами по себе эксперты могут быть значительно неоднородны по профессиональному или другому признаку, тогда требования к величине коэффициента конкордации могут быть снижены.

Итак, формируем исходную таблицу по результатам оценок экспертов (рис. 4.82).

В данном случае мы имеем дело с зависимыми выборками в том смысле, что один и тот же субъект многократно оценивается

различными экспертами. Следует обратить внимание, что данная матрица, по сравнению с дисперсионным анализом с повторениями, повернута набок (транспонирована), то есть преподаватели-эксперты выступают в качестве зависимых переменных, а участники олимпиады — независимых, факторных.

| | 1 школьн ик1 | 2 школьни к2 | 3 школьни к3 | 4 школьни к4 | 5 школьни к5 | 6 школьни к6 | 7 школьни к7 | 8 школьни к8 | 9 школьни к9 | 10 школьни к10 | 11 школьни к11 | 12 школьни к12 | 13 школьни к13 | 14 школьни к14 | 15 школьни к15 |
|-----------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| эксперт1 | 5 | 15 | 6 | 14 | 1 | 2 | 3 | 13 | 8 | 11 | 4 | 7 | 9 | 12 | 10 |
| эксперт2 | 3 | 13 | 7 | 15 | 4 | 2 | 1 | 14 | 8 | 10 | 5 | 6 | 9 | 11 | 12 |
| эксперт3 | 4 | 10 | 6 | 11 | 5 | 3 | 1 | 12 | 7 | 8 | 2 | 9 | 13 | 15 | 14 |
| эксперт4 | 3 | 8 | 4 | 10 | 1 | 2 | 5 | 11 | 7 | 6 | 9 | 12 | 13 | 14 | 15 |
| эксперт5 | 2 | 14 | 5 | 15 | 1 | 3 | 4 | 13 | 6 | 7 | 8 | 9 | 10 | 12 | 11 |
| эксперт6 | 6 | 7 | 5 | 12 | 3 | 1 | 2 | 11 | 8 | 10 | 4 | 9 | 15 | 14 | 13 |
| эксперт7 | 7 | 8 | 6 | 10 | 2 | 1 | 3 | 11 | 5 | 9 | 4 | 12 | 15 | 13 | 14 |
| эксперт8 | 5 | 10 | 7 | 11 | 4 | 3 | 1 | 13 | 6 | 15 | 2 | 8 | 14 | 12 | 9 |
| эксперт9 | 3 | 11 | 12 | 10 | 2 | 1 | 4 | 5 | 6 | 13 | 7 | 8 | 15 | 14 | 9 |
| эксперт10 | 3 | 12 | 8 | 15 | 2 | 1 | 4 | 9 | 7 | 14 | 5 | 6 | 11 | 10 | 13 |
| эксперт11 | 4 | 7 | 8 | 9 | 5 | 3 | 2 | 10 | 6 | 11 | 1 | 12 | 14 | 15 | 13 |
| эксперт12 | 2 | 8 | 7 | 14 | 1 | 3 | 4 | 11 | 15 | 10 | 5 | 6 | 13 | 9 | 12 |

Рис. 4.82

Для начала анализа выбираем в основном меню Statistics (Статистика), затем Nonparametrics (Непараметрические методы) (рис. 4.83).

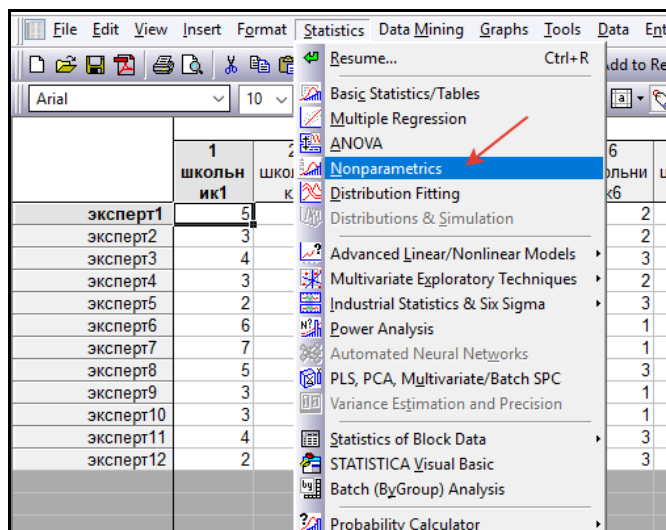


Рис. 4.83

И затем в открывшемся меню — Comparing multiple dependent samples (variables) (Сравнения многих зависимых выборок (переменных)) (рис. 4.84).

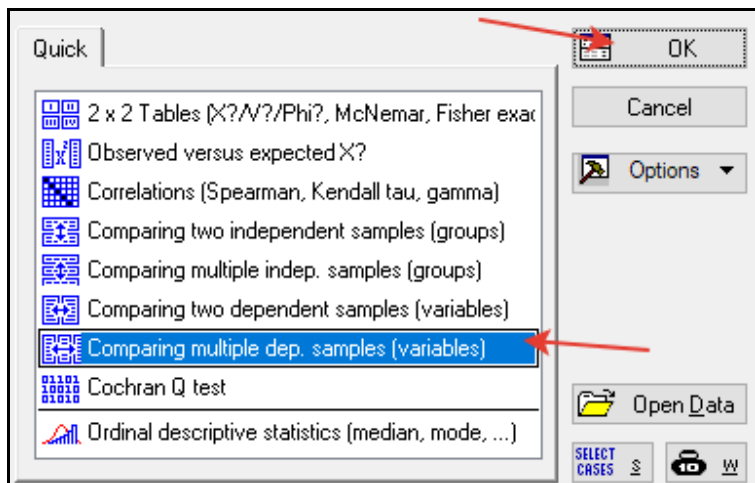


Рис. 4.84

После этого выбираем переменные для анализа. В данном случае — это все школьники, которые участвовали в олимпиаде (рис. 4.85).

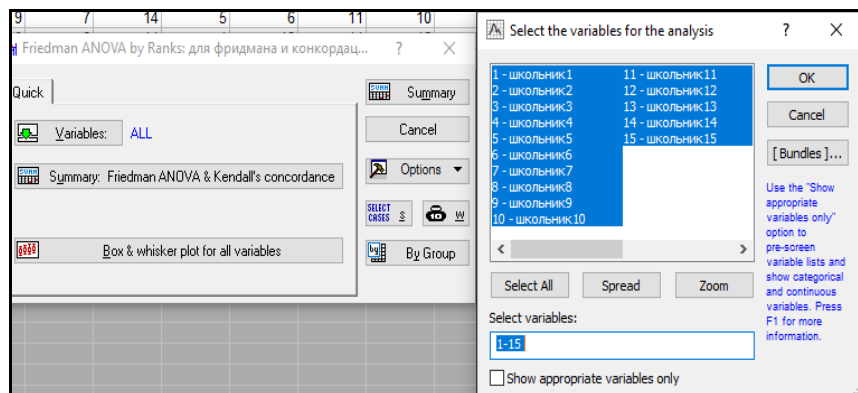


Рис. 4.85

После нажатия на кнопку Summary (Итог) получаем итоговую статистику для критерия Фридмана и коэффициента конкордации (надпись вверху скриншота) (рис. 4.86).

| Variable | Friedman ANOVA and Kendall Coeff. of Concordance ANOVA Chi Sqr. (N = 12, df = 14) = 130,3417 p = 0,00000 Coeff. of Concordance = ,77584 Aver. rank r = ,75547 | | | | |
|------------|---|--------------|----------|----------|--|
| | Average Rank | Sum of Ranks | Mean | Std.Dev. | |
| школьник1 | 3,91667 | 47,0000 | 3,91667 | 1,564279 | |
| школьник2 | 10,25000 | 123,0000 | 10,25000 | 2,767506 | |
| школьник3 | 6,75000 | 81,0000 | 6,75000 | 2,050499 | |
| школьник4 | 12,16667 | 146,0000 | 12,16667 | 2,289634 | |
| школьник5 | 2,58333 | 31,0000 | 2,58333 | 1,564279 | |
| школьник6 | 2,08333 | 25,0000 | 2,08333 | 0,900337 | |
| школьник7 | 2,83333 | 34,0000 | 2,83333 | 1,403459 | |
| школьник8 | 11,08333 | 133,0000 | 11,08333 | 2,391589 | |
| школьник9 | 7,41667 | 89,0000 | 7,41667 | 2,574643 | |
| школьник10 | 10,33333 | 124,0000 | 10,33333 | 2,708013 | |
| школьник11 | 4,66667 | 56,0000 | 4,66667 | 2,424621 | |
| школьник12 | 8,66667 | 104,0000 | 8,66667 | 2,309401 | |
| школьник13 | 12,58333 | 151,0000 | 12,58333 | 2,274696 | |
| школьник14 | 12,58333 | 151,0000 | 12,58333 | 1,928652 | |
| школьник15 | 12,08333 | 145,0000 | 12,08333 | 1,975225 | |

Рис. 4.86

Как видно из представленного скриншота, в данном случае статистическая значимость различий в успешности участников очень высока: $p = 0,00000$, то есть фактически $p < 10^{-5}$, что означает достоверные различия в уровне подготовки участников, по мнению экспертов. Коэффициент конкордации (Coeff. of Concordance), отражает уровень согласия экспертов относительно успешности различных испытуемых. Его значение 0,77584 означает в данном случае высокий уровень согласия экспертов.

Для большей наглядности различий в успешности можно построить график, вернувшись в главное меню анализа. Нажав на Box and whisker plot for all variables (график по типу «ящик с усами» для всех переменных) и выбрав затем тип графика «исключительно» («Медиана/квартиль/размах») (рис. 4.87), поскольку мы имеем дело с порядковыми данными, можно визуальнo оценить успешность каждого испытуемого, исходя из мнений всех экспертов.

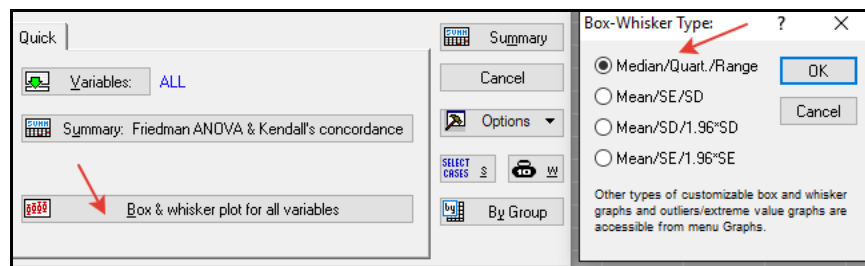


Рис. 4.87

В данном случае не все подрисуночные подписи по оси абсцисс поместились (в ней — но не в графике! — пропущены четные «участники»). Тем не менее, на графике представлены все 15 участников, и мы можем ясно видеть, как распределились мнения экспертов относительно участников олимпиады (рис. 4.88).

Из графика может быть визуальнo получена дополнительная информация о распределении в рейтинге каждого из участников.

Итак, подтверждается альтернативная гипотеза о том, что успешность участников различна. Можно утверждать, что эта гипотеза объективно подтверждается еще и потому, что мнения экспертов хорошо согласованы между собой.

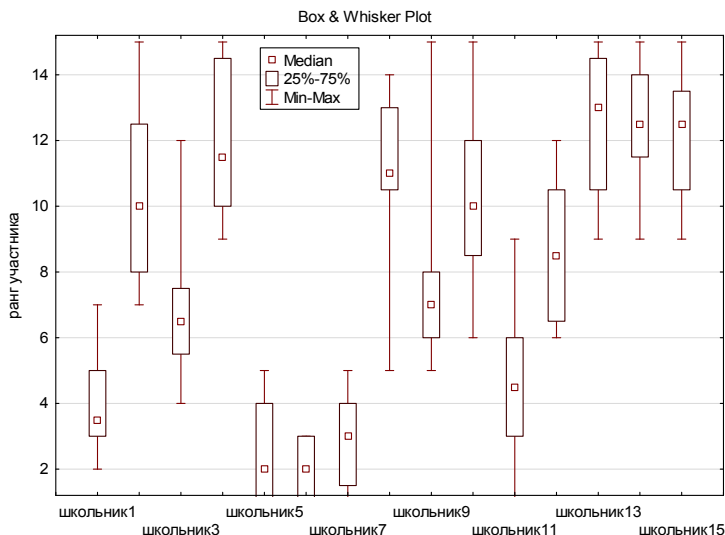


Рис. 4.88

4.4. СТАТИСТИЧЕСКИЕ КРИТЕРИИ ДЛЯ ОПРЕДЕЛЕНИЯ РАЗЛИЧИЙ И СВЯЗЕЙ В КАЧЕСТВЕННЫХ И КАТЕГОРИАЛЬНЫХ ДАННЫХ

Прежде всего, следует четко понимать различие между количественными и порядковыми данными, с одной стороны, и качественными (и категориальными) — с другой стороны. Если мы имеем дело с количественными или порядковыми данными, то каждому объекту совокупности приписываются одно или несколько чисел (значение IQ, уровень тревожности и проч.) или рангов. Если же данные качественные, то каждому объекту совокупности приписывается одно или несколько качеств («невротик, не невротик», «пенсионер, работающий, безработный», «гетеросексуал, гомосексуал, бисексуал» и т. п.). В некоторых задачах объект может принадлежать к сразу нескольким категориям. Например, «гетеросексуал, невротик, безработный». Затем проводится подсчет количества объектов (испытуемых), принадлежащих к определенной

категории. И далее мы оперируем уже не с числами, приписанными каждому объекту, как в случае количественных данных, а количеством объектов (испытываемых), относящихся к той или иной категории или их пересечению.

Еще раз подчеркнем, что важным и продуктивным в психологических исследованиях является использование критериев для качественных данных, когда каждому объекту из выборки приписывается определенное качество или принадлежность к той или иной категории. Собственно говоря, данные, принадлежащие к шкале наименований (номинативной шкале), уже являются качественными. Все критерии, работающие с качественными данными, работают по одной и той же схеме. Каждому объекту из выборки приписано или по определенному правилу приписывается одно из качеств, которые в совокупности составляют систему событий: «ребенок, взрослый, старик», «мужчина, женщина», «больной, здоровый», «поступил в вуз, не поступил в вуз», «психолог, физик-теоретик, гастроэнтеролог...» и т. д. Иногда нам удобно загрузить шкалу и перейти от более мощной к менее мощной, например от шкалы интервалов к шкале наименований, и затем использовать статистические критерии для качественных данных.

Например, мы сравниваем личностные характеристики людей, занимающихся медитацией разное время. В шкале интервалов мы можем рассматривать всю выборку в соответствии со значениями стажа занятий медитацией в показателях лет, месяцев, количестве пройденных «ретритов» (курсовых тренингов). При переходе из шкалы интервалов в шкалу наименований мы как исследователи можем решить, что люди, занимающиеся медитацией до трех лет, условно будут названы «начинающие», более трех лет — «опытные». И после такой категоризации можно оперировать этими данными как с независимыми выборками.

Далее мы перечислим критерии, которые наиболее универсальны и нетребовательны к распределению данных и решают определенные задачи сравнения или определения связи между качественными данными. Ниже будут рассмотрены лишь некоторые из возможных способов (критериев) решения этих задач.

4.4.1. Сравнение доли выраженности признака в двух независимых выборках

Довольно часто исследователь задается вопросом о том, в какой из двух независимых выборок доля признака значимо отличается и в какую сторону, когда речь идет именно о качественных данных. Довольно часто в таких случаях используются, например, точный критерий Фишера, критерий хи-квадрат с поправкой Йейтса, критерий «угловое преобразование Фишера».

Именно этот критерий, «угловое преобразование Фишера», мы и рассмотрим более подробно, поскольку он является довольно мощным и позволяет эффективно находить различия. Единственное относительное ограничение критерия — его значения являются несколько завышенными в смысле нахождения значимых различий, когда доля признака в выборке равна или очень близка к нулю. Но в реальных задачах такое бывает крайне редко.

Данный критерий является многофункциональным во многих смыслах. Например, иногда бывает удобно перевести количественные данные в качественные. В частности, мы можем разделить выборку на две части исходя из граничного значения, которое принимает признак. Так, возможно измерить уровень выраженности депрессии количественно (получив конкретные числа, отражающие уровень выраженности признака), а затем разделить выборку на две части — с нормальными и повышенными значениями. В некоторых случаях нам просто не обойтись без применения данного или сходных критериев, когда мы обрабатываем данные, где предполагается лишь два варианта ответа или значения — да или нет, или, в более общих терминах, «обладает одним свойством» — «обладает другим свойством». Важно, чтобы в конкретной когорте конкретная дихотомия охватывала бы всю выборку и каждому из ее объектов можно было бы приписать или одно, или другое свойство. Таким образом, в каждой из подвыборок в сумме мы должны иметь полную систему событий, выражаясь языком теории вероятностей.

Рассмотрим два типичных примера, когда применяется критерий «угловое преобразование Фишера» для качественных данных.

В первом примере предполагается, что выборки гомогенны (по интеллекту, полу, возрасту и т. п.), но на результат (то есть различия выраженности долей признака) может повлиять фактор обстановки, внешняя среда (неважно, что это — преподаватель, консультант, место, время и т. д.). Итак, исследователь задался вопросом: отличаются ли доли студентов-психологов, выпускившихся в Москве и в Тюмени, в равнозначных по рейтингу вузах и не более чем в течение года, трудоустроившихся по профессии?

Рассмотрим данный пример в программе «Медстат».

Выборка тюменских выпускников составила 100 человек (из которых 68 трудоустроились), московских — 120 (из которых 105 трудоустроились) (рис. 4.89).

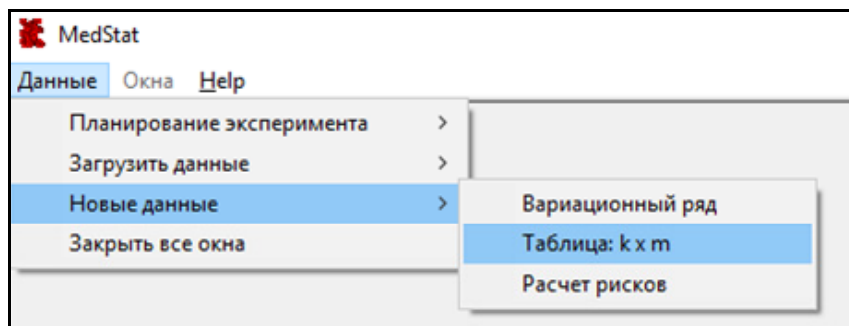


Рис. 4.89

Выбираем в меню Данные → Новые данные → Таблица: k x m.

Вносим данные (рис. 4.87). Следует обратить особое внимание, что Level1 означает долю одного рода событий (в данном случае — сколько выпускников трудоустроились), а Level2 — другого (не трудоустроились). Очевидно, что в сумме одни должны дать в каждом из столбцов совокупные значения по каждой выборке (то есть $68 + 32 = 100$, а $105 + 15 = 120$).

Далее, после выбора в меню «k x m» пункта «Сравнение доли (таблицы 2×2)» (рис. 4.90) получаем результат статистического сравнения (рис. 4.91).

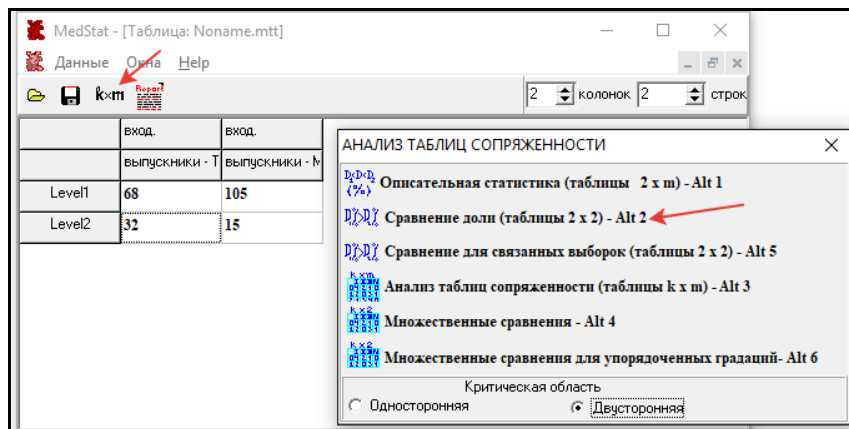


Рис. 4.90

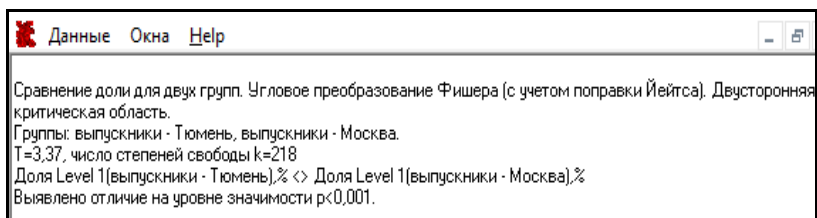


Рис. 4.91

В данном случае оказалось, что с высокой степенью статистической значимости ($p < 0,001$) выпускники московских вузов трудоустраиваются по специальности чаще тюменских. Однако следует заметить, что по условиям задачи исследователь не уточнял, где именно трудоустраиваются выпускники. Возможно, что в Москве, где в основном и трудоустраиваются выпускники-психологи московского вуза, больше вакансий для психологов; а может быть, уровень подготовки в столичном вузе выше, а может быть — в Москве обучаются наиболее талантливые и целеустремленные студенты из России и других стран, в частности, бывшего СССР, в отличие от регионального контингента, обучающегося в основном в Тюмени, и т. д. Таким образом, мы еще раз видим, как важна

не столько математическая, сколько содержательная обработка и понимание результатов исследования; не говоря уже о дизайне, подготовке исследования и формировании выборов.

Тем не менее, напомним, как можно представить результаты:

- в виде долей и их доверительных интервалов (рис. 4.92);

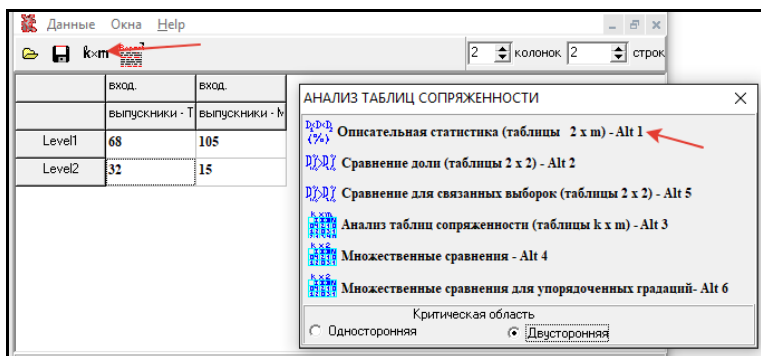


Рис. 4.92

- с получением текстовых (числовых) (рис. 4.93);

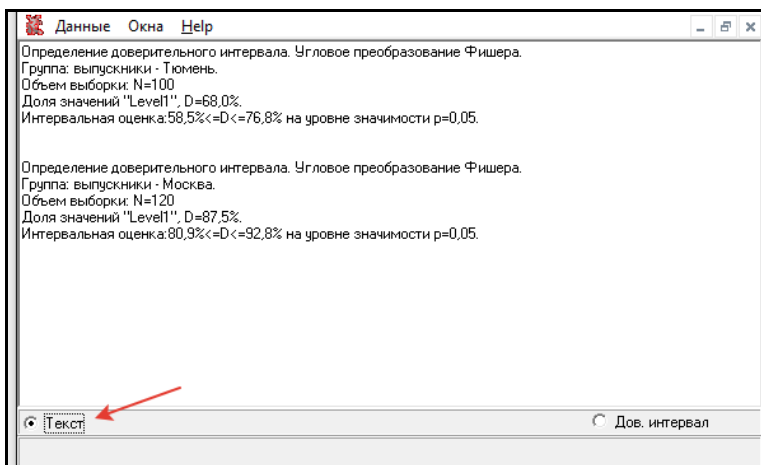


Рис. 4.93

- графически представленных результатов (рис. 4.94).

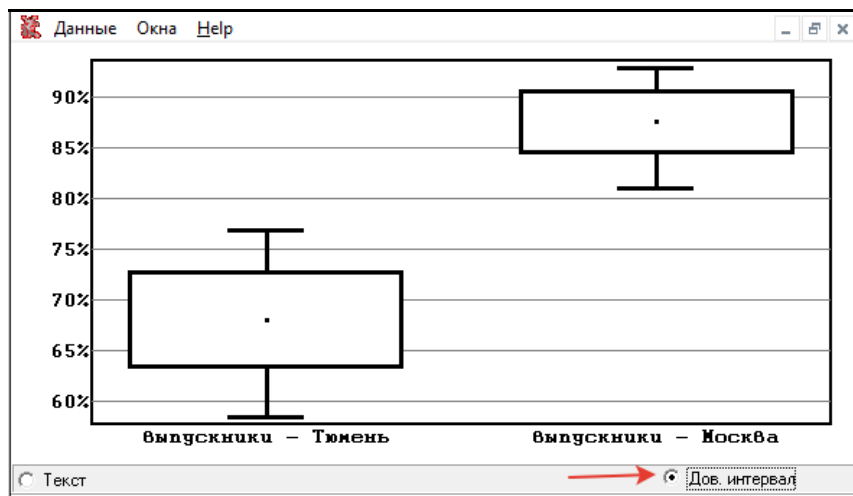


Рис. 4.94

Рассмотрим далее, в некотором смысле, иного рода пример, где фактором является некий признак, по которому выборки существенно отличаются.

Например, исследователя интересует: значимо ли отличаются доли совершивших правонарушения хотя бы один раз подростков 15 лет из выборок полных семей (5 из 140) и неполных семей (11 из 120) из одних и тех же социальных слоев пропорционально. В данном случае выборки явно негомогенны. Однако применение критерия так же уместно. В сущности, если бы мы имели дело с количественными признаками, мы бы могли поступить точно так же: полагать выборки независимыми, выбирая их как бы из общего пула, и затем помещать их в разные условия или совершать над ними разные экспериментальные действия. Либо изначально формировать выборки существенно различными по характерному признаку, сопоставляя у них определенную зависимую переменную.

Данный пример так же рассмотрим в программе «Медстат».

Внесем данные в таблицу (рис. 4.95).

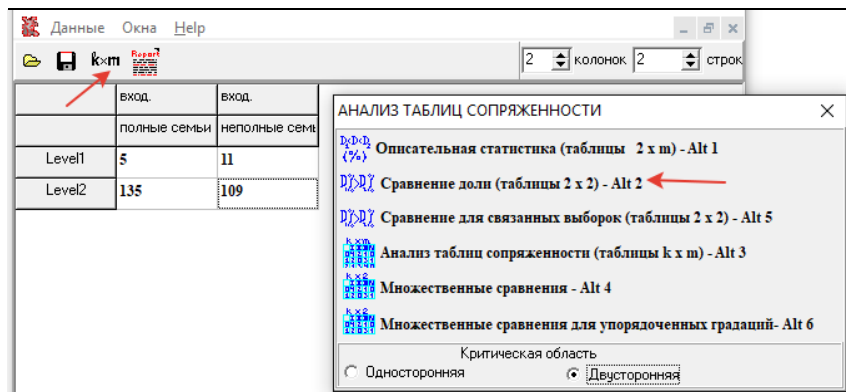


Рис. 4.95

И получим результаты статистического сравнения (рис. 4.96).

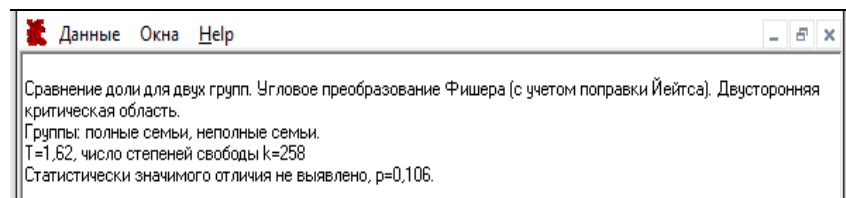


Рис. 4.96

В данном случае различия не являются статистически значимыми (уровень значимости $p = 0,106$), поэтому для выборок данного размера мы не можем сделать вывода о том, что правонарушителей в неполных семьях больше, чем в полных, хотя, судя по доле, их больше (9,2% против 3,6%). Однако поскольку мы не можем отвергнуть нулевую гипотезу на должном уровне значимости ($p < 0,05$), полагаем различия случайными.

4.4.2. Сравнение долей выраженности признака в двух связанных выборках

Как мы знаем, очень важно уметь сравнивать выраженность признака также и в зависимых выборках. Когда речь идет о качественных данных, то подобный тип задач так же актуален. Критерий Мак-Немара в некотором роде является аналогом параметрического критерия Стьюдента и непараметрического критерия Вилкоксона и применяется для анализа связанных измерений в случае изменения реакции с помощью дихотомической переменной.

Рассмотрим некоторые из возможных дизайнов, которые могут встретиться в психологических исследованиях с использованием дихотомических переменных в зависимых выборках:

1. Измеряют дихотомическую переменную (условно — «да»/«нет») на одной и той же выборке в разных условиях или до и после некоторого воздействия. Задача: установить, значимо ли повлиял некоторый фактор (изменение обстановки, экспериментальное воздействие и проч.) на соотношение (долю) «да» к «нет».

2. На одной и той же выборке с заранее известным свойством проводят измерения двумя разными методами. Задача: установить, одинаково ли хорошо разные измерения выявляют наличие данного (априори имеющегося) свойства у данной выборки.

Рассмотрим пример, основанный на первом из перечисленных типов дизайна. Например, можно сравнить удовлетворенность своими отношениями с партнером у женщин до и после проведения тренинга женской привлекательности (вумбилдинг), который прошли 50 женщин. Поскольку мы имеем дело с качественными данными, женщины могли ответить только «да» или «нет» на вопрос, который им среди прочих задавали до начала и после окончания тренинга: «Удовлетворены ли Вы отношениями с Вашим партнером?»

Допустим, что по результатам опроса мы имели следующие результаты (см. табл. 4.4).

Решим данную задачу с помощью пакета «Медстат».

В меню «Данные» выбираем пункты «Новые данные» и «Таблица k x m» (см. рис. 4.97).

**Удовлетворенность отношениями до и после тренинга
(стандартный вариант)**

| Удовлетворенность отношениями (да/нет) | До: «Да» | До: «Нет» | После: «Да» | После: «Нет» |
|---|----------|-----------|-------------|--------------|
| Количество участниц с соответствующими ответами | 30 из 50 | 20 из 50 | 10 из 50 | 40 из 50 |

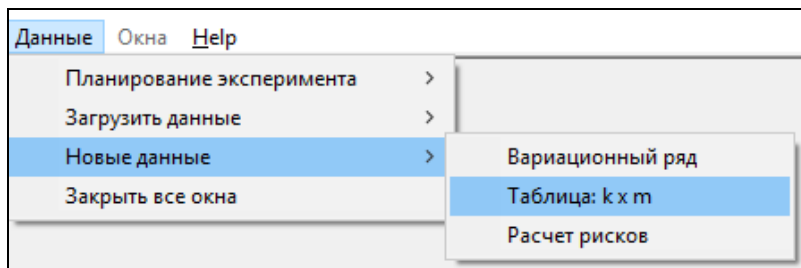


Рис. 4.97

После этого формируем таблицу размерностью две строки на два столбца (колонки), выставляя соответствующие значения в указанных стрелками пунктах меню (рис. 4.98).

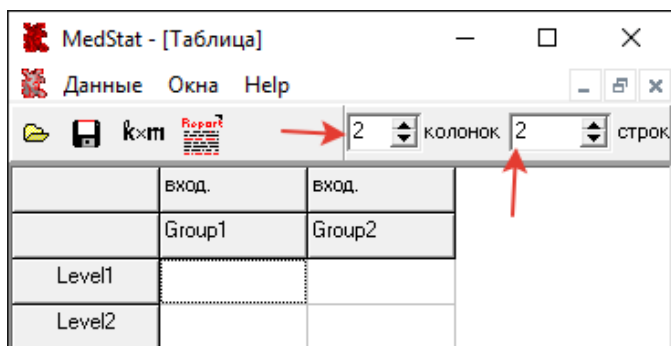


Рис. 4.98

Теперь рассмотрим, как следует вводить расчетные данные в соответствующую таблицу в программе «Медстат», невзирая на стандартные обозначения в программе.

Составим таблицу (табл. 4.5) в том виде, как данные будут внесены в программу.

Таблица 4.5

**Удовлетворенность отношениями до и после тренинга
(вариант для программы)**

| Количество ответов | До тренинга: НЕТ | После тренинга: ДА |
|--------------------|------------------|--------------------|
| До тренинга: НЕТ | $20 + 40 = 60$ | $20 + 10 = 30$ |
| После тренинга: ДА | $40 + 10 = 50$ | $30 + 10 = 40$ |

Теперь вносим данные в таблицу уже программы «Медстат» в том же порядке, не обращая внимания на отличия в обозначениях строк и столбцов таблицы (рис. 4.99).

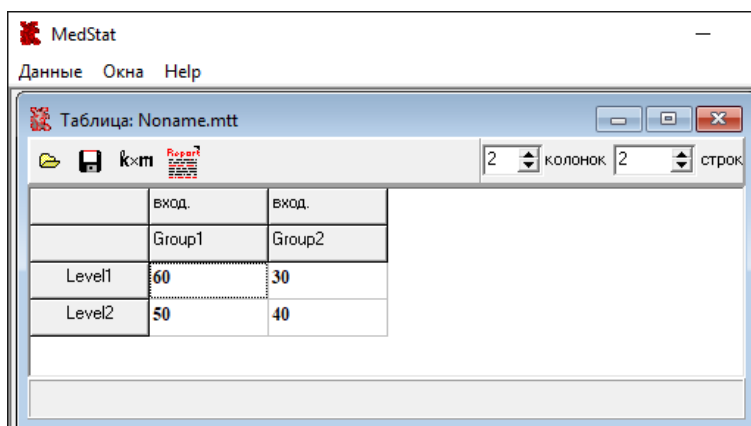


Рис. 4.99

Затем выбираем в расчетном меню «Сравнение для связанных выборок» (см. рис. 4.100).

Ниже представлен результат применения критерия Мак-Немара к изучаемой выборке (см. рис. 4.101).

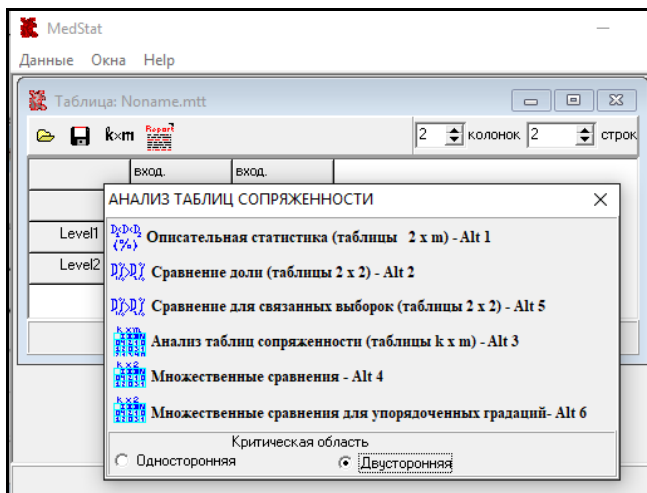


Рис. 4.100

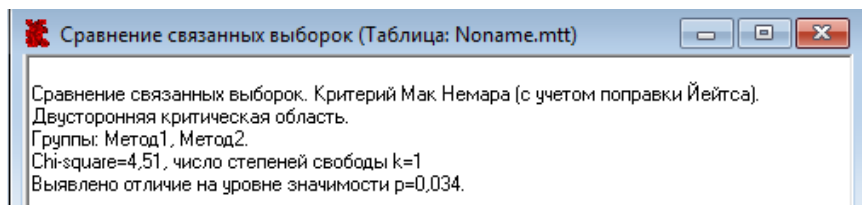


Рис. 4.101

В данном случае можно сделать вывод о том, что в результате прохождения конкретного тренинга по вумбилдингу доля женщин, недовольных своими партнерами, значимо увеличилась ($p = 0,034$). Какие выводы может сделать из этого результата исследователь, зависит от особенностей методики его проведения:

1) возможно, мы имеем дело с тренингом, который имеет неконструктивную методологию (например, женщин учат быть «стервами», а не искать способы понять партнера);

2) возможно, это недовольство носит конструктивный характер и будет направлено на улучшение и укрепление отношений, а не на их разрушение.

Рассмотрим еще один пример, рассчитываемый с помощью критерия Мак-Немара, построенный по второму из представленных дизайнов.

К психологу за определенный промежуток времени обратились родители 200 подростков с интернет-зависимостью и зависимостью от гаджетов. Психолог — эксперт в подобных вопросах — подтвердил наличие данной зависимости. Исследовательская задача — протестировать две разные методики (опросники, направленные на выявление данной зависимости) на предмет того, насколько они чувствительны к распознаванию зависимости (то есть у какой доли подростков они выявят наличие зависимости), а также является ли одна методика значимо чувствительнее (лучше) другой для выявления данной зависимости. Ценность данного исследования ясна, поскольку чувствительная методика может стать скрининговой для отбора подростков с риском или наличием данной зависимости. Если же методика не чувствительна, то есть неполно, плохо выявляет данную зависимость, то, очевидно, ценность ее как скрининговой, мала. Если же у нас есть две различных, в некотором смысле конкурирующих между собой методики, то мы можем выбрать лучшую в отношении чувствительности при наличии значимых различий.

Итак, составляем исходную таблицу (табл. 4.6).

Таблица 4.6

Данные выявляемости признака (стандартный вариант)

| Выявляемость зависимости у подростков, у которых она априори имеется | Методика 1: «Да» | Методика 1: «Нет» | Методика 2: «Да» | Методика 2: «Нет» |
|--|------------------|-------------------|------------------|-------------------|
| Количество подростков, у которых выявлена зависимость двумя разными методами | 150 из 200 | 50 из 200 | 180 из 200 | 20 из 200 |

Аналогично первому примеру применения данного критерия составим таблицу в том виде, как данные будут внесены в программу (табл. 4.7).

Таблица 4.7

Данные выявляемости признака (вариант для программы)

| Количество ответов | Методика 2: НЕТ | Методика 2: ДА |
|--------------------|------------------|-------------------|
| Методика 1: НЕТ | $50 + 20 = 70$ | $50 + 180 = 230$ |
| Методика 1: ДА | $150 + 20 = 170$ | $150 + 180 = 330$ |

Так же вносим данные уже в таблицу программы, в том же порядке, не обращая внимания на отличия в обозначениях строк и столбцов таблицы (рис. 4.102).

| | Вход. | Вход. |
|--------|-------|-------|
| Group1 | 70 | 230 |
| Group2 | 170 | 330 |

Рис. 4.102

Аналогично в меню выбираем пункт «Сравнение для связанных выборок, и получаем результат (рис. 4.103).

| |
|--|
| <p>Сравнение связанных выборок. Критерий Мак Немара (с учетом поправки Йейтса). Двусторонняя критическая область. Группы: Метод1, Метод2. Chi-square=8,70, число степеней свободы k=1 Выявлено отличие на уровне значимости p=0,003.</p> |
|--|

Рис. 4.103

В результате получили, что вторая методика выигрывает в выявляемости (90% против 75%), а также намного эффективнее первой с точки зрения статистической значимости различий ($p = 0,003$) и, следовательно, может быть рекомендована к применению в качестве скрининговой.

4.4.3. Сравнение доли выраженности признака в более чем двух независимых выборках

Как и в случае, когда мы имели дело с количественными данными, при работе с качественными данными также очень важно иметь возможность проводить сравнение выраженности признака в нескольких выборках, с последующими попарными межгрупповыми сравнениями. Если при нормальном распределении признака этой задаче соответствовал однофакторный дисперсионный анализ (ANOVA), а для порядковых данных — критерий Краскела–Уоллиса, то в этом пункте мы разберем для качественных данных аналогичную возможность сравнивать доли дихотомического признака с помощью процедуры Мараскуило в модификации Ляха-Гурьянова [9], при условии выявления значимых различий в распределении признака критерием хи-квадрат.

В некотором смысле речь идет об обобщении уже рассмотренного выше критерия «угловое преобразование Фишера» или подобных ему на случай сопоставления выраженности признака более чем в двух выборках.

Поясним и проиллюстрируем процедуру множественных сравнений дихотомических данных в программе «Медстат» на следующем примере.

Исследователь поставил цель сравнить удовлетворенность своей профессией у представителей разных типов профессий (по Е. А. Климову). Исследование планировалось среди тех, кто проработал не менее пяти лет по данной профессии, среди жителей Тюмени в возрасте от 25 до 35 лет. Как известно, Е. А. Климов выделил пять основных объектов труда: человек, техника, художественный образ, знак, природа. В первой части названия типа профессии обозначен субъект труда, которым всегда является человек,

например, «человек–человек», «человек–техника» и др. Альтернативная гипотеза состояла в том, что в современных условиях мегаполиса удовлетворенность профессиями, относящимся к разным типам объектов труда, будет различаться между собой. В случае же выявления значимых различий исследователь сможет провести попарное статистическое сопоставление степени удовлетворенности профессиями представителей данных профессиональных групп.

Составляем таблицу в программе «Медстат», выбрав необходимое количество строк (2) и столбцов (5) (рис. 4.104) и присвоив переменным соответствующие названия (рис. 4.105).

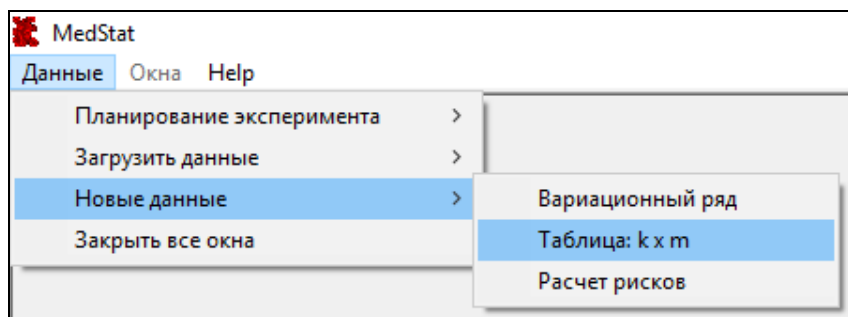


Рис. 4.104

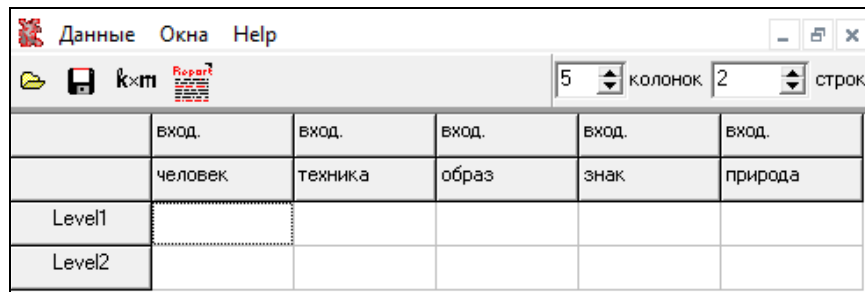


Рис. 4.105

В строку Level1 внесем количество испытуемых данной группы, удовлетворенных своей профессией, а в строку Level2 — неудовлетворенных. Затем нажимаем меню «k x m», далее — «Множественные сравнения» (рис. 4.106).

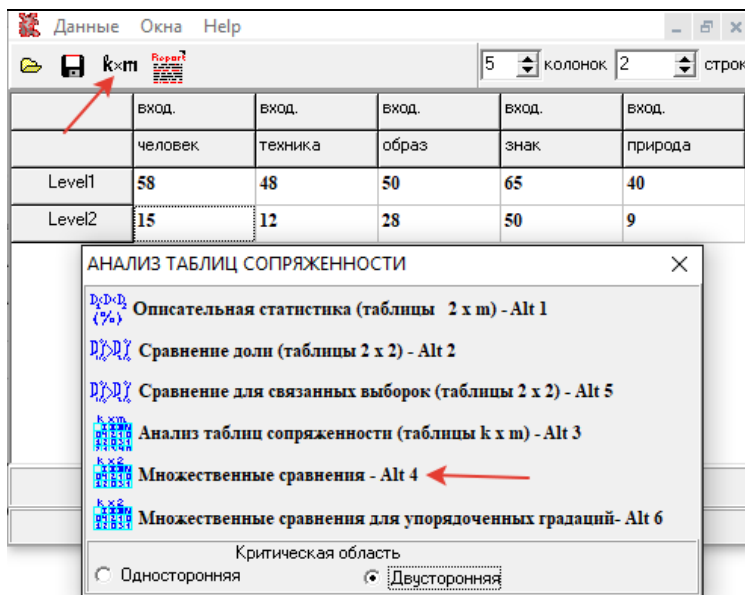


Рис. 4.106

Итак, ниже представлены результаты множественных сравнений (см. рис. 4.107).

Из представленного примера видно, что, во-первых, доли признака (удовлетворенных профессиями) значительно различаются, причем на высоком уровне значимости $p < 0,001$. Однако при рассмотрении попарных различий лишь доли удовлетворенных профессиями типа «человек–человек» и «человек–знак» значительно различаются ($p = 0,040$). Можно представить результаты в виде описательной статистики или графически (см. рис. 4.108).

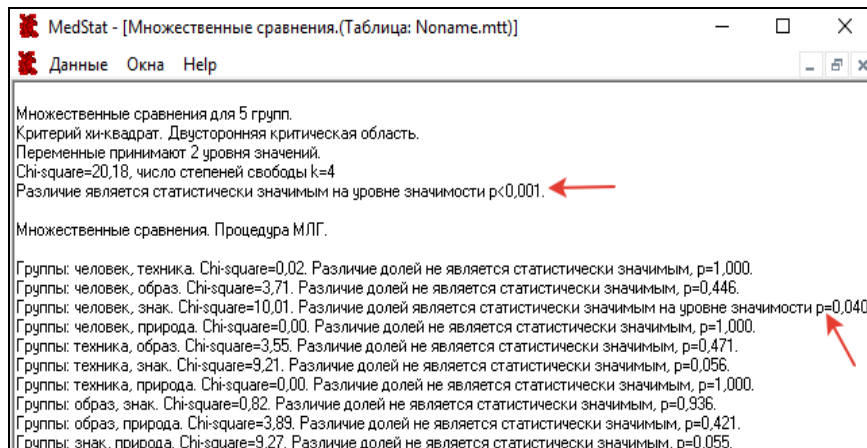


Рис. 4.107

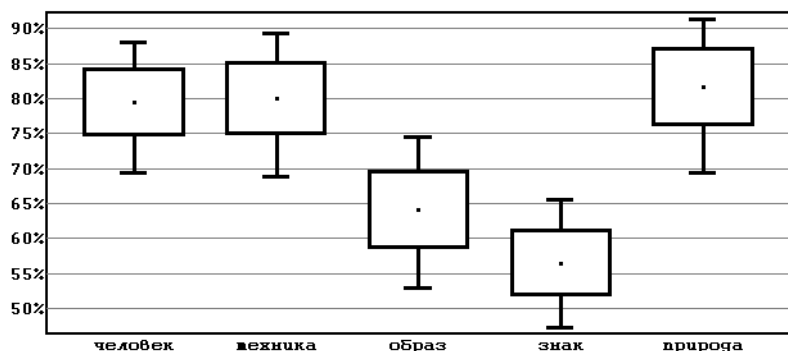


Рис. 4.108

Из графика видно, что удовлетворенность профессиями типа «человек–знак» действительно ниже, чем у представителей профессий «человек–человек». Исходя из этого, исследователь мог бы сделать вывод, что взаимодействие с другими людьми дает дополнительное ощущение благополучия, нужности своего труда, тогда как в системе «человек–знак» фрустрированной оказывается об-

ратная связь — ощущение востребованности продуктов труда, что и обеспечивает ощущение неудовлетворенности.

4.4.4. Выявление отличий наблюдаемой доли признака от эталонной (теоретически ожидаемой)

Довольно распространенной задачей психологического исследования является изучение статистической значимости отклонения доли признака от ожидаемой величины или, что аналогично с точки зрения математики, обнаружение преобладания одной из двух альтернатив при заранее известном или предполагаемом процентном соотношении между альтернативами. Для таких случаев, как правило, следует использовать биномиальный критерий, реализованный, в частности, в статистическом пакете SPSS.

Поясним это примерами:

1. Допустим, исследователь задался вопросом, каких девушек предпочитают парни — блондинок или брюнеток. Были отобраны фотографии брюнеток и блондинок «идеальной» внешности. На экране монитора располагалось достаточное количество фотографий лиц, где было поровну блондинок и брюнеток, и каждый респондент должен был выбрать одну из фотографий как наиболее понравившуюся. Понятно, что единственный выбор испытуемого так или иначе падал на блондинку или брюнетку. Будем исходить из априорного предположения, что респонденты — совершеннолетние студенты вузов Тюмени (пусть выборка насчитывает 250 человек) — будут поровну выбирать брюнеток и блондинок, то есть теоретически ожидаемая частота составит 50%.

Решим данную задачу в программе SPSS. Выбору блондинки сопоставим условное обозначение «0», а выбору брюнетки — условное обозначение «1». В нашем примере, поскольку данные качественные, мы могли бы присвоить выборам блондинок или брюнеток любые другие значения, главное — чтобы они отличались.

Каждой строке соответствует выбор одного испытуемого. Таким образом, наша переменная насчитывает 250 строк, каждая из которых может иметь только два значения: «блондинка» — 0 или «брюнетка» — 1.

Для экономии места покажем лишь первые пять значений, которые мы вводим в новый файл или копируем их в новую переменную из табличного процессора (рис. 4.109).

| | VAR00001 | пер | пер | пер | пер | пер | пер | пер |
|---|----------|-----|-----|-----|-----|-----|-----|-----|
| 1 | .00 | | | | | | | |
| 2 | .00 | | | | | | | |
| 3 | .00 | | | | | | | |
| 4 | .00 | | | | | | | |
| 5 | 1.00 | | | | | | | |

Рис. 4.109

В меню «Анализ» сначала выбираем «Непараметрические критерии», затем «Устаревшие диалоговые окна» и «Биномиальный...» (рис. 4.110).

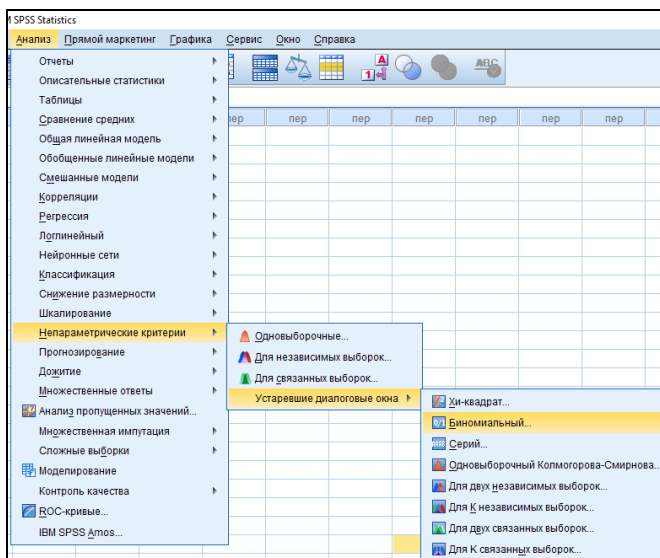


Рис. 4.110

После чего выставляем проверяемую долю (теоретически ожидаемое значение пропорции), в данном случае — 0,50, то есть 50%. Далее с помощью стрелки-указателя при предварительном однократном нажатии на интересующей переменной (у нас всего одна переменная — VAR0001) перемещаем переменную в список проверяемых (рис. 4.111) и нажимаем на «ОК» (рис. 4.112).

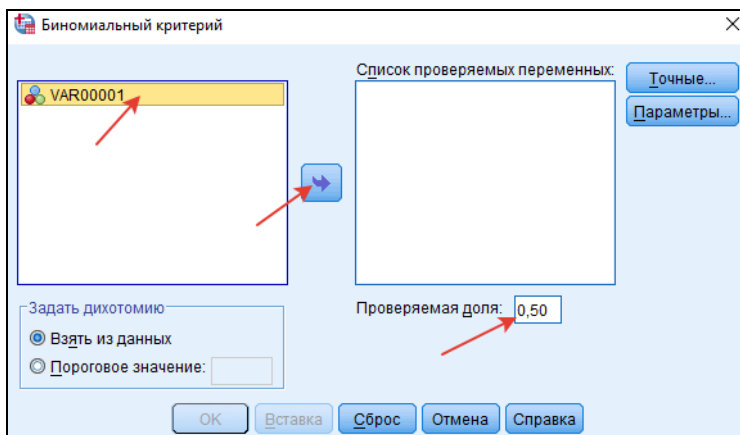


Рис. 4.111

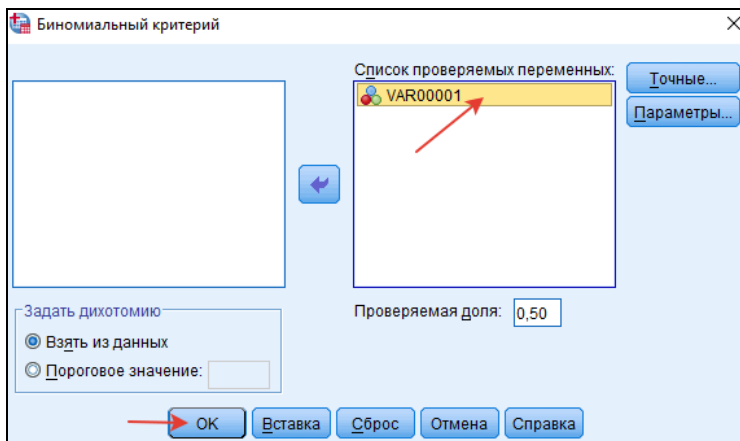


Рис. 4.112

В окне вывода результатов программы SPSS (рис. 4.113) получаем итоговый результат: примерно 56% испытуемых выбрали блондинок, а 44% — брюнеток, однако этого оказалось недостаточно, чтобы отвергнуть нулевую гипотезу об отсутствии различий в выборах при ожидаемых выборах «поровну» блондинок и брюнеток ($p = 0,066$).

Биномиальный критерий

| | Категория | N | Наблюдаемая доля | Проверяемая доля | Точная з.нч. (2-сторонняя) |
|----------|-----------|------|------------------|------------------|----------------------------|
| VAR00001 | Группа 1 | ,00 | 140 | ,56 | ,066 |
| | Группа 2 | 1,00 | 110 | ,44 | |
| | Всего | | 250 | 1,00 | |

Рис. 4.113

В данном случае использовалась, как мы видим, двусторонняя критическая область, то есть мы не имели априорных предположений о преобладании тех или иных типов выборов.

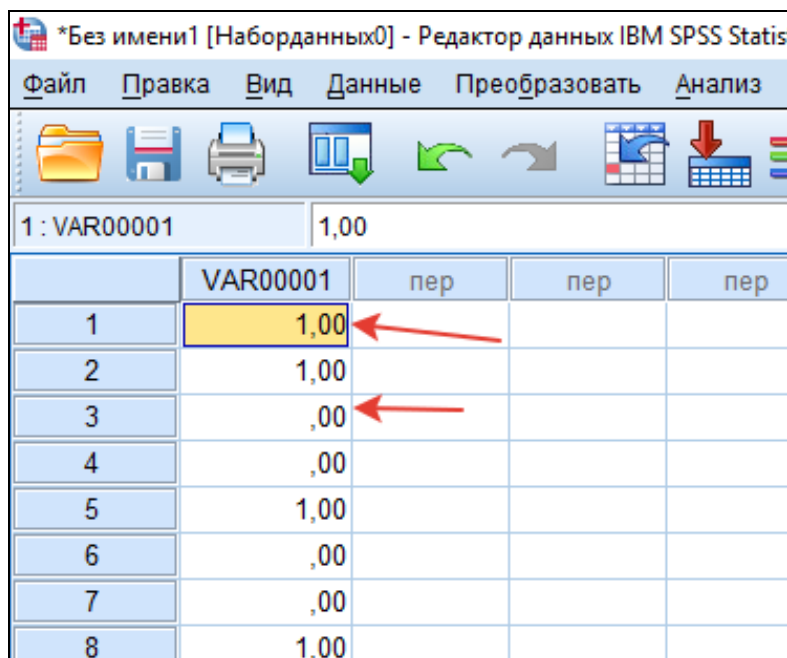
2. Допустим, исследователь проводит эксперимент, который призван выявлять «экстраординарные» способности к интуитивным предсказаниям, с помощью специальной программы ESP Trainer (<http://www.espresearch.com/iphone>). Исследователь желает проверить, будут ли испытуемые значимо чаще предугадывать стимулы, чем это теоретически ожидается.

Каждый испытуемый в сессии, состоящей из 24 последовательных попыток, должен предугадывать, нажатие на какой именно квадрат (зеленый, синий, красный или желтый) приведет к открытию фотографии. Следует отметить, что мы можем выявить здесь как общие, так и групповые, а также индивидуальные особенности предугадывания.

Пусть мы имеем дело с данными 166 испытуемых-женщин разного возраста, которые прошли по одной сессии тестирования. То есть всего в качестве исходных данных исследователь получил

24 попытки, помноженные на 166 испытуемых, то есть 3984 результатов выборов. Теоретически ожидаемая частота правильного выбора составляет, естественно, 25%, то есть один к четырем. Присвоим символу «1» результат «угадала», а символу «0» — «не угадала».

В программе SPSS в качестве проверяемой доли будем использовать 0,25, но при этом важно, чтобы первым по счету символом стоял бы «1» (в данном случае этого легко добиться, просто поменяв местами ноль и единицу в любом месте, если первым символом стоял ноль. Так, допустим, что в первой строке у нас стоял ноль, а в третьей — единица. Нам следует, как вариант, поменять их местами, чтобы программа затем правильно сработала (рис. 4.114).



| 1 : VAR00001 | | 1,00 | | |
|--------------|----------|------|-----|-----|
| | VAR00001 | пер | пер | пер |
| 1 | 1,00 | | | |
| 2 | 1,00 | | | |
| 3 | ,00 | | | |
| 4 | ,00 | | | |
| 5 | 1,00 | | | |
| 6 | ,00 | | | |
| 7 | ,00 | | | |
| 8 | 1,00 | | | |

Рис. 4.114

Далее в качестве проверяемой доли в критерии указываем 0,25 (рис. 4.115).

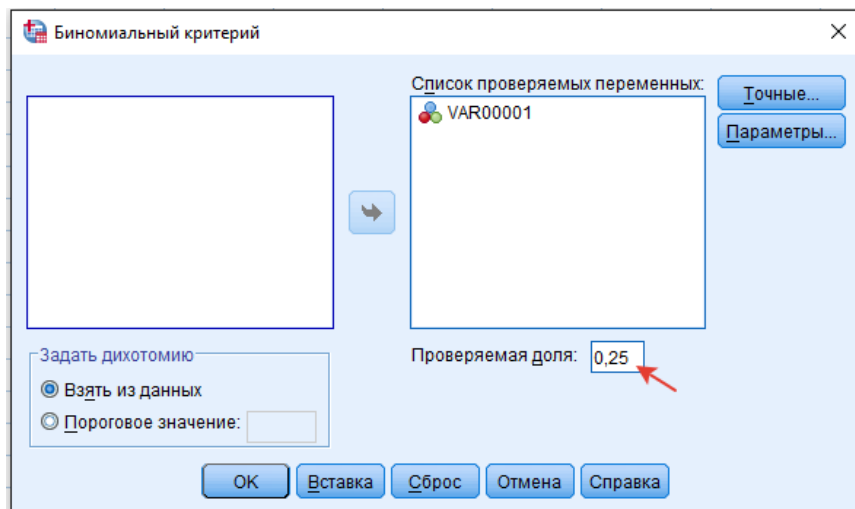


Рис. 4.115

И получаем результат (рис. 4.116).

Биномиальный критерий

| | Категория | N | Наблюденная доля | Проверяемая доля | Точная знач. (1-сторонняя) |
|----------|-----------|------|------------------|------------------|----------------------------|
| VAR00001 | Группа 1 | 1,00 | 1046 | ,26 | ,036 |
| | Группа 2 | ,00 | 2938 | ,74 | |
| | Всего | | 3984 | 1,00 | |

Рис. 4.116

Отметим, что для любых других долей, кроме 0,5 (50%) в пакете SPSS рассчитывается биномиальный критерий с односторонней критической областью, то есть проверяется альтернативная гипоте-

за о том, что наблюдаемая доля значений признака превышает теоретически ожидаемую. В данном случае мы опровергли нулевую гипотезу на уровне статистической значимости $p = 0,036$, но в предположении о том, что доля предугаданных значений больше, чем 25%. Если бы мы проверяли двустороннюю гипотезу (то есть априори не предполагали, что доля предугаданных значений может быть и значимо меньше, чем 25%, а не только больше), то мы бы не смогли опровергнуть такую двустороннюю нулевую гипотезу, поскольку приблизительно уровень значимости был бы равен $0,036 \times 2 = 0,072$, что, очевидно, не позволило бы отвергнуть нулевую гипотезу с двусторонней критической областью.

4.5. АЛГОРИТМ ВЫБОРА НЕКОТОРЫХ ИЗ ВОЗМОЖНЫХ К ПРИМЕНЕНИЮ СТАТИСТИЧЕСКИХ КРИТЕРИЕВ ДЛЯ БАЗОВЫХ ЗАДАЧ ПСИХОЛОГИЧЕСКИХ ИССЛЕДОВАНИЙ

Перейдем к описанию некоторых «рецептов», исходя из которых выбираются те или иные критерии для обработки результатов. По отдельности мы уже разбирали почти все эти критерии и приводили примеры их применения. Однако для удобства целесообразно подытожить эти сведения в таблицу как некий путеводитель в отношении применения того или иного критерия в зависимости от экспериментального дизайна.

Рассмотрим примерный алгоритм выбора направления статистического анализа в случае определенного вида экспериментальных дизайнов, когда мы уже разобрались с тем, сколько у нас выборов, какие они (независимые, зависимые), в какой именно шкале выражены признаки и имеем ли мы дело с данными с нормальным или альтернативным статистическими распределениями. А главное, когда сформулированы статистические гипотезы на основе содержательных, то есть реализован этап психолого-математической интерпретации данных. Сразу следует оговориться, что на самом деле для каждого случая существуют и другие критерии со своими особенностями. Приведенная ниже табл. 4.8 не охватывает некоторые дизайны, с которыми мы можем встретиться в исследованиях,

например, анализ выживаемости, тенденций, множественный анализ для упорядоченных градаций, отношение шансов и др. В то же самое время некоторые из методов, приведенных в таблице, не были рассмотрены в примерах данного пособия по причине того, что они встречаются довольно редко в реальных психологических исследованиях. Однако их легко найти в руководствах из приложенного библиографического списка или других (в том числе открытых) источников, которыми изобилует современное информационное пространство.

Таблица 4.8

**Выбор критерия для проведения статистического анализа
в зависимости от перевода экспериментального дизайна
в статистическую гипотезу**

| <i>Тип экспериментального дизайна</i> | <i>Шкала измерений</i> | | |
|---------------------------------------|---|--|---|
| | <i>Наименование</i> | <i>Порядок</i> | <i>Интервалы и отношения</i> |
| <i>1</i> | <i>2</i> | <i>3</i> | <i>4</i> |
| Одна выборка | Биномиальный критерий, одновыборочный критерий хи-квадрат, критерий серий | Одновыборочный критерий хи-квадрат, критерий серий | Одновыборочный критерий Стьюдента, критерий серий, основанный на среднем; критерии Колмогорова–Смирнова, Шапиро–Уилка, хи-квадрат и др. — для оценки типа распределения |
| Две независимые выборки | Критерий хи-квадрат, точный критерий Фишера, угловое преобразование Фишера, биномиальный критерий | Критерий Манна–Уитни, критерий знаков | Критерий Стьюдента для независимых выборок, F-критерий Фишера |

| 1 | 2 | 3 | 4 |
|---------------------------------|---|---|--|
| Три и более независимые выборки | Анализ таблиц сопряженности критерием хи-квадрат | Критерий Краскела–Уоллиса | Одно- и многофакторный дисперсионный анализ |
| Две зависимые выборки | Критерий Мак-Немара | Критерий Вилкоксона | Критерий Стьюдента для зависимых выборок |
| Три и более зависимые выборки | Критерий Кокрана | Критерий Фридмана, коэффициент конкордации | Дисперсионный анализ с повторными измерениями |
| Связь между переменными | Коэффициенты сопряженности, контингенции Пирсона, коэффициент Юла | Ранговая корреляция Спирмена, Кендалла, гамма | Линейная регрессия, множественная линейная регрессия, корреляция Пирсона |

5. ПРОСПЕКТИВНЫЕ ИССЛЕДОВАНИЯ. ПИЛОТАЖНЫЕ И ПОДТВЕРЖДАЮЩИЕ ИССЛЕДОВАНИЯ. РЕТРОСПЕКТИВНЫЕ ИССЛЕДОВАНИЯ

Если не существует особых ограничений или специфики, то исследование планируется заранее. В соответствии с методологией (в более узком смысле — с теорией), которой придерживается исследователь, выдвигается теоретическая и/или эмпирическая гипотеза, затем формируется план исследования, в том числе эксперимента или экспериментов, после чего составляются их дизайны, и наконец, эти дизайны нуждаются в переводе с языка психологии на язык статистики.

Когда же мы сталкиваемся с языком статистики, прежде чем сформулировать статистические гипотезы и выбрать методы обработки данных, мы с необходимостью должны вспомнить целый ряд понятий и разобраться с ними применительно к каждому из психологических экспериментов:

I. Какие именно выборки отобраны для исследования?

1. Это единственная выборка, и целью исследования является выявление соответствия или несоответствия некоему эталонному распределению признака (пропорции, нормальному распределению и т. д.). Например:

А. 30% сотрудников организации обычно приходят на работу за 20 минут до начала работы, 50% — за 10 минут до начала работы, 20% опаздывают, приходя через 10 минут после начала рабочего дня (то есть имеется некое эталонное распределение категорий сотрудников). В дни, когда компания представляет новую продукцию, по мнению начальника отдела по персоналу, количество опозданий увеличивается. Так ли это на самом деле или причина в других факторах (тревожность самого руководителя, погодные и дорожные условия и проч.)?

Б. По мнению завуча по учебной работе, в дни контрольных работ количество учащихся, оставшихся дома и не пришедших в школу по болезни, увеличивается. Так ли это на самом деле или

причина в других факторах (в предвзятом отношении завуча к учащимся, его встревоженности о результатах контрольных работ, нежелании дополнительно организовывать передачи и т. д.)?

2. Это независимые выборки, отобранные из разных генеральных совокупностей; у них будут измеряться одни и те же параметры, при этом будет изучаться влияние факторов, которые определяют принадлежность к разным генеральным совокупностям и полагаются причинами тех или иных различий изучаемых психологических свойств. Возможно также, что далее эти группы станут подвергаться различным воздействиям, при этом исследователь будет иметь в виду эти изначальные различия существенных свойств выборок. Например:

А. Жители Африки и жители Европы. Оценка невербального интеллекта (как наиболее независимого от влияния культуры). Исследование проводится одними и теми же инструментами (тест «Прогрессивные матрицы» Дж. Равена, тест интеллекта, свободного от влияния культуры, Р. Кеттелла).

Б. Ветераны боевых действий Афганистана и Вьетнама. Различие выборок состоит в том, что это комбаттанты, принадлежащие к разным странам и разным культурам, участвовавшие в разных войнах. Общее, что позволяет исследовать их по одним и тем же параметрам, — опыт участия в боевых действиях на территории не своей страны. Потенциально сравниваемые параметры: выраженность, особенности симптомов и синдромов посттравматического стрессового расстройства (ПТСР — PTSD), особенности протекания их коррекции.

3. Это независимые выборки, отобранные случайным образом из одной и той же генеральной совокупности, над которыми затем проводят те или иные измерения, испытания. Например:

А. Сотрудники двух разных компаний, работающие на должности менеджера по продажам. Оцениваем особенности корпоративной культуры в представлениях менеджеров по продажам (в каждой из компаний) и сравниваем успешность, продуктивность деятельности менеджеров.

Б. Учащиеся общеобразовательных учреждений, среднего звена (5-6 класс) и старшего звена (10-11 класс). Оцениваем зависимость от

электронных гаджетов (частота пользования смартфоном, опасение остаться без смартфона, доступа к социальным сетям и электронной почте) и общий уровень тревожности и невротизации учащихся.

4. Это одна и та же выборка (хоть это и одна и та же выборка — мы все равно называем их зависимыми), у которой в различных ситуациях двукратно или многократно измеряются одни и те же признаки: как в относительно краткосрочной, так и долгосрочной перспективе. Например:

А. Учащиеся общеобразовательных учреждений. Оценка когнитивных функций одних и тех же учащихся в 1, 2, 3, 6, 9, 11 классах. В 1-3 классах оценка когнитивных функций проводится для выявления слабых и сильных мест, соответствующей коррекции и учета этого в обучении. В 6 и 9 классах для учета в профилизации обучения. В 9-11 классах — с целью профессиональной ориентации.

Б. В рамках семейной терапии возможно оценивать состояние ребенка и каждого из родителей на разных этапах терапевтического процесса.

В. В исправительном учреждении осужденные проходят несколько периодов адаптации к условиям заключения. Возможно оценивать их мотивы к ресоциализации, психоэмоциональное состояние на этапе адаптации к условиям заключения (примерно первая четверть срока отбывания наказания), на основном этапе отбывания наказания, на завершающем этапе, когда необходимо готовиться к выходу на свободу и осваивать появившиеся социальные практики (примерно один-два месяца до освобождения).

5. Это одна и та же выборка (но мы все равно называем их зависимыми) или несколько подвыборок, у каждой из которых мы измеряем более одного параметра и затем ищем связи между параметрами.

Например, сотрудники разных организаций, офисных и производственных. Обобщающий признак для выборки — сотрудники организации. Если мы зададимся целью узнать, с какими параметрами трудовой деятельности связаны симптомы эмоционального выгорания сотрудников, то мы должны будем оценить каждый из симптомов эмоционального выгорания (например, методикой В. В. Бойко) и параметры трудовой деятельности: принадлежность к системе «человек–человек», «человек–знак», «человек–техника»,

«человек—художественный образ», удовлетворенность трудом, наличие/отсутствие опредмеченного продукта труда (сшитое платье, выпеченная булочка, выпущенный в 11 классе ученик, годовой отчет) и т. д.

6. Это разные выборки, но при этом каждому субъекту одной выборки мы можем сопоставить одного и только одного субъекта (например, муж и жена, монозиготные сиблинги и т. п.). Например:

А. Особенности конфликтов в семейных парах с разным доходом. Мы хотим выяснить, в каких семьях чаще конфликтуют: в тех в которых муж зарабатывает больше, чем жена, или наоборот?

Б. Метод разлученных близнецов. Близнецы, воспитывавшиеся в разных семьях, обладают сходными чертами личности или различными?

II. К какой шкале принадлежат измеренные в выборках признаки?

Мы уже довольно подробно и неоднократно останавливались на вопросе о принадлежности данных к той или иной шкале, но подчеркнем лишь несколько деталей:

1. Принадлежность данных к той или иной шкале будет ограничивать нас применением определенных описательных статистик и статистических критериев.

2. Каждому субъекту выборки, можно сопоставить несколько признаков, выраженных в разных шкалах — от наименований до отношений. Арсенал математико-статистических методов невероятно широк, что находит свое описание в соответствующих руководствах и статистических пакетах, но, как правило, если в выборке каждому субъекту приписаны количественные признаки (значение баллов по опроснику, баллы за академические достижения, размер финансового вознаграждения как показатель производственной успешности), помимо качественных (пол, статус, группа), то, скорее всего, исследователь вправе рассматривать эти качественные признаки как группирующие данные в подвыборки. Например:

А. В классах школы есть учащиеся, принадлежащие к разным социальным стратам: изолированные, предпочитаемые, лидеры, отвергаемые. Задача, поставленная перед психологом, состоит в том, чтобы понять, учащиеся какой страты привержены каким

ценностным ориентациям (терминальным и инструментальным, методика М. Рокича).

Б. В процессе профессионального отбора в правоохранительные органы, специалисты выделяют четыре группы претендентов: не рекомендуемые ни при каких условиях, условно рекомендуемые (могут быть приняты при условии большого количества вакансий и постоянного контроля за деятельностью и динамикой адаптации), рекомендуемые и рекомендуемые в первую очередь. Задача психолога может состоять в том, чтобы выяснить особенности, присущие представителям условно рекомендуемой и рекомендуемой группы, и критическое значение психологических показателей, при которых происходит переход из одной группы в другую.

3. Если наши данные исключительно качественные или категориальные (что мы разобрали в пункте 4.4), то тогда и статистические гипотезы следует формулировать в терминах долей.

III. Далее мы переходим к формулировке одной или нескольких частных и (желательно) независимых друг от друга статистических гипотез в том смысле, что выполнение одной гипотезы не должно увеличивать шансы другой гипотезы в данном исследовании.

Например, в двух классах общеобразовательной школы, учащиеся обладают сходным уровнем когнитивных способностей, однако академическая успеваемость в одном классе по средним оценкам 4,5, а в другом 3,7. Содержательная гипотеза может быть такая: академическая успеваемость связана с социально-психологическим климатом в классе. Статистические гипотезы:

1) особенности социально-психологического климата в классе связаны (влияют) с академической успеваемостью учащихся (двунаправленная гипотеза);

2) негативный социально-психологический климат отрицательно влияет на академическую успеваемость учащихся (однонаправленная гипотеза);

3) позитивный социально-психологический климат положительно влияет на академическую успеваемость учащихся (однонаправленная гипотеза);

4) особенности социально-психологического климата в классе не связаны (не влияют) на академическую успеваемость учащихся.

Если у нас нет априорных (заранее известных, хорошо обоснованных) предположений о результатах исследования, то нам следует провести перед *подтверждающим* нашу гипотезу исследованием *пилотажное исследование*, на выборке из той же генеральной совокупности. Его цель состоит отнюдь не в доказательстве гипотезы исследования, а в том, чтобы в конечном счете определить объем выборки для подтверждающего исследования таким образом, чтобы мы смогли с должной точностью проверить нашу гипотезу (выборки не должны быть слишком малы) и в то же время не злоупотребить ресурсами (выборки не должны быть слишком велики).

Итак, классическое проспективное исследование (заранее планируемое — от методологии и теории до гипотез и, конечно, методик) часто начинают с пилотажного исследования. Для пилотажного исследования отбирается довольно малая часть генеральной совокупности и подвергается тем же экспериментальным методам, что и основная часть — для подтверждающего исследования. Возможны нюансы. На основе особенностей процесса или результатов пилотажного исследования методика основного (подтверждающего гипотезу) исследования может быть скорректирована. Представим наиболее простой случай, когда корректировка не потребовалась. Прежде всего, мы получим некий результат, который, однако, за редким исключением не будет обладать статистической значимостью, поэтому не может считаться подтвердившим гипотезу. Почему?

Потому что пилотажное исследование не преследует изначально цели доказать гипотезу, а, скорее, необходимо для получения дополнительной информации о предмете и объекте, для уточнения гипотез и задач, для проверки качества инструментария (тесты, опросники, экспериментальные процедуры и т. п.), разработанного для подтверждающего исследования, для выявления потенциальных неочевидных трудностей проведения эксперимента. Объем выборки для пилотажного исследования существенно варьируются — от 10 до 100 человек — в зависимости от проработанности инструментария, количества выборок и конечных целей и задач подтверждающего исследования.

В практическом отношении важной функцией пилотажного исследования может являться примерное определение важнейших параметров выборок (средние для каждой из выборок, разность между ними, стандартные отклонения и т. д. — в зависимости от типа математико-статистической процедуры). Так в пилотажном исследовании мы примерно определяем основные описательные статистики, которые в случае нормального распределения в подтверждающем исследовании будут не слишком отличаться от полученных в пилотажном. Если исследователь не вправе ожидать нормальности распределения или уверен в отклонении распределения выборок от нормального, ему следует увеличить объем выборок приблизительно в 1,8 раз (квадратный корень из числа пи) [9] относительно тех, что были получены для случая нормального распределения.

5.1. СОДЕРЖАТЕЛЬНАЯ И СТАТИСТИЧЕСКАЯ ЗНАЧИМОСТЬ РЕЗУЛЬТАТОВ. РАСЧЕТ ОБЪЕМА ВЫБОРОК ДЛЯ ПОДТВЕРЖДАЮЩЕГО ИССЛЕДОВАНИЯ

Чтобы прояснить разницу между функциями пилотажного и подтверждающего исследования, следует основательно разобраться в различиях между понятиями статистической и содержательной значимости результатов. Различия могут быть статистически значимыми (например, мы будем сравнивать по параметру IQ две группы довольно большие по объему, по 1100 человек каждая, и получим разницу в показателе примерно 1,8 единиц). Статистическая значимость, посчитанная с помощью критерия Стьюдента, будет равна 0,05 с мощностью критерия 0,8. Можем ли мы с позиций психологии обоснованно утверждать, что средней разницы в 1,0 единицу (при известном стандартном отклонении для данной шкалы в 15) достаточно для того, чтобы утверждать, что представители одной группы действительно «умнее» другой? На это может дать ответ только психолог-исследователь исходя из своей компетентности и цели исследования.

Представим, наоборот, что различия в показателе велики (10 единиц в IQ), но выборки содержат лишь по 11 человек. В таком случае различия оказываются незначимыми статистически, хотя легко представить, что фактически (содержательно, с точки зрения психологического исследования и практики) в одной из выборок интеллект реально (с точки зрения психологического исследования) выше, чем в другой.

Для того и существует планирование объема выборок на основании дизайна и метода исследования, чтобы наверняка подтвердить или опровергнуть статистическую гипотезу исходя из содержательной (психологической) функции исследования, рассчитав необходимый и достаточный объем выборки (выборок) для конкретных показателей. В последнем примере, например, нас бы интересовало — какое количество испытуемых в выборках достаточно для того, чтобы установить статистически значимо интересующие нас различия коэффициента интеллекта в выборках.

Мы можем планировать подтверждающее исследование и без пилотажного, но для этого нам нужно наверняка знать параметры распределения измеряемой величины

Рассмотрим этот пример в программе Primer of Biostatistics. Допустим, психолог в своем исследовании двух независимых выборок (к примеру, учащихся физико-математической школы и обычной общеобразовательной школы) по признаку IQ считает минимально значимым с точки зрения содержательной (наполненной психологическим смыслом для данного исследования) различием в 10 единиц. Такое значение разницы в IQ он получил в пилотажном исследовании. Теперь пошагово рассмотрим минимальные объемы выборок, чтобы заведомо получить статистически значимые различия при данных разностях и стандартных отклонениях, полученных в пилотажном исследовании. Для этого в программе во вкладке Sample Size (Размер выборки) выберем t-test (Критерий Стьюдента) (см. рис. 5.1), поскольку в пилотажном исследовании, положим, распределение переменных обеих выборок не отличались от нормального и сам параметр IQ стандартизирован так, чтобы

в большинстве случаев давать именно нормальное распределение. Правда, лишь постфактум, проведя подтверждающее исследование, мы увидим, действительно ли распределение явилось нормальным. Но тогда у нас всегда есть альтернатива воспользоваться аналогичными непараметрическими критериями. В любом случае в реальное исследование следует отбирать несколько большее количество испытуемых (не менее чем на 10-30% от рекомендуемого программой — чем меньшее количество испытуемых планируется включить в исследование, тем больше в процентах должен быть запас испытуемых) по многим причинам — от возможного вынужденного использования непараметрических, менее мощных критериев до отказа участвовать или выбытия испытуемых из исследования по самым разным причинам.

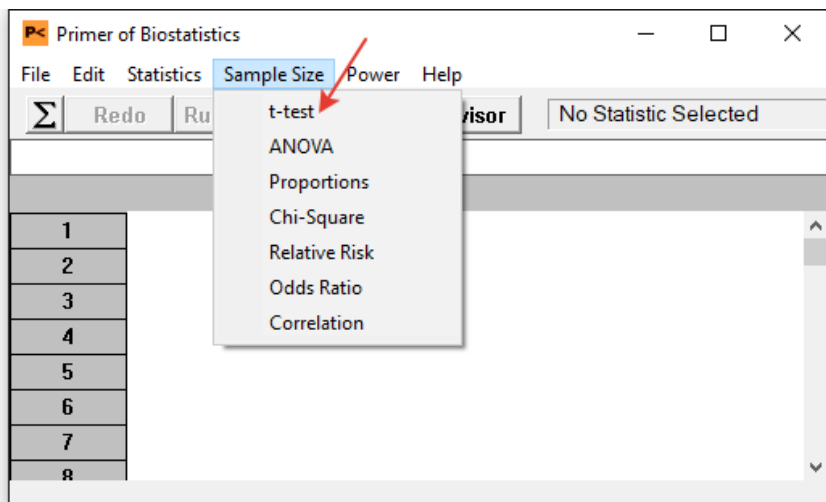


Рис. 5.1

Далее в открывшемся меню выберем тип критерия Unpaired t-test (two separate groups) (Критерий Стьюдента для независимых (несвязанных) выборок) (рис. 5.2).

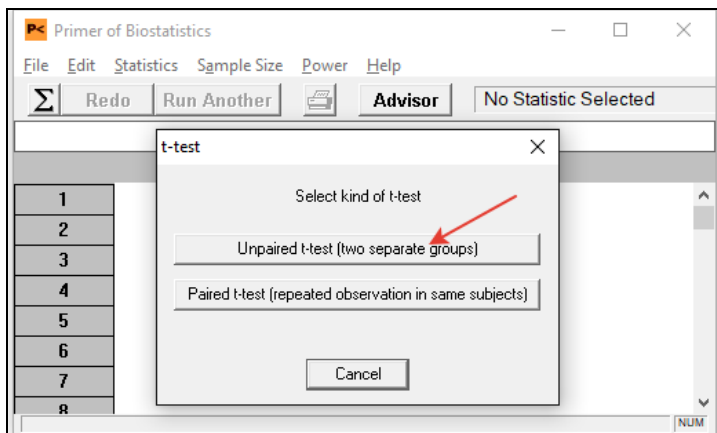


Рис. 5.2

После этого откроется вкладка, где должны быть введены некоторые параметры. Так, пусть граничным уровнем величины ошибки первого рода или, что в данном случае то же самое, статистической значимости (Alpha) будет 0,05 (стандартная величина), различие в средних (Difference of Means) 10 и ожидаемое внутригрупповое стандартное отклонение 15 (рис. 5.3).

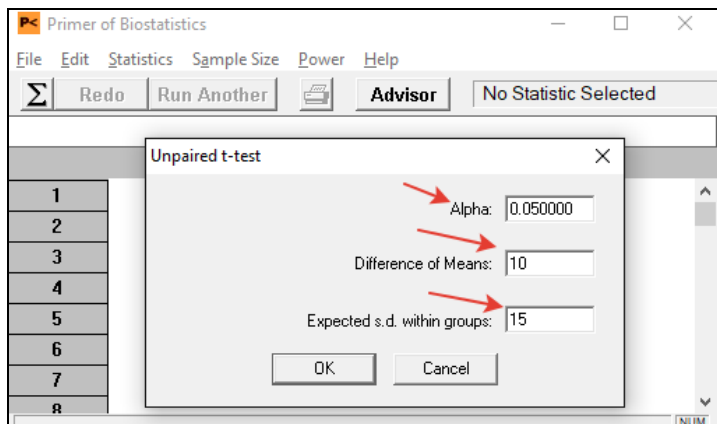


Рис. 5.3

Следует отметить, что здесь подразумевается популяционное стандартное отклонение (для данной генеральной совокупности). Авторы посоветовали бы для уверенности использовать рассчитанное нами объединенное стандартное отклонение двух выборок, полученное в пилотажном исследовании. Или, если пилотажное исследование не было проведено, но есть доказательные основания полагать, что действительно для данной генеральной совокупности, из которой взяты выборки, параметры (разность средних и стандартное отклонение) именно таковы, то вносим их. Не стоит забывать, что мы можем произвольно менять желаемый минимально допустимый уровень значимости, и от этого будет зависеть расчетное количество испытуемых в выборках.

В следующем меню мы должны указать желаемую мощность (Desired Power) (рис. 5.4). По умолчанию она составляет 0,8, однако мы можем выбрать любое число от 0 до 1 — в частности, можно встретить 0,9 или 0,95. И снова — чем больше мощность (то есть разность единицы и ошибки второго рода), тем больше потребуется испытуемых в выборках.

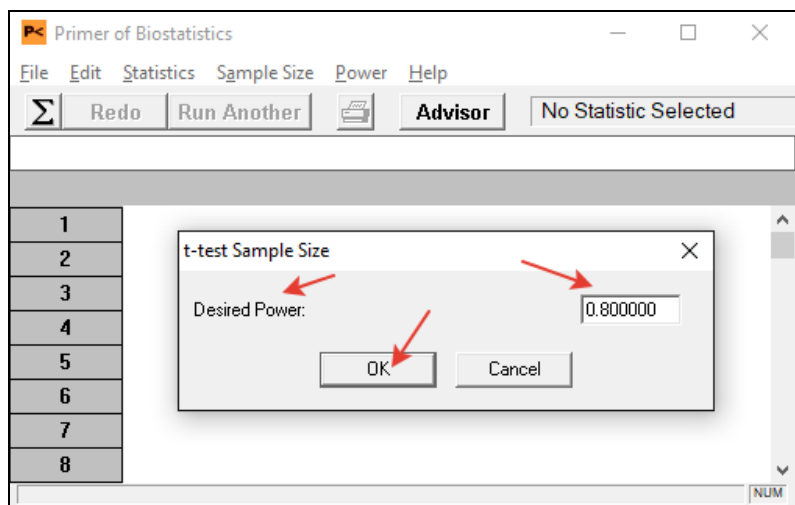


Рис. 5.4

Для нашего же выбора, как видно, необходимо, чтобы в каждой из выборок было не менее 37 человек (рис. 5.5).

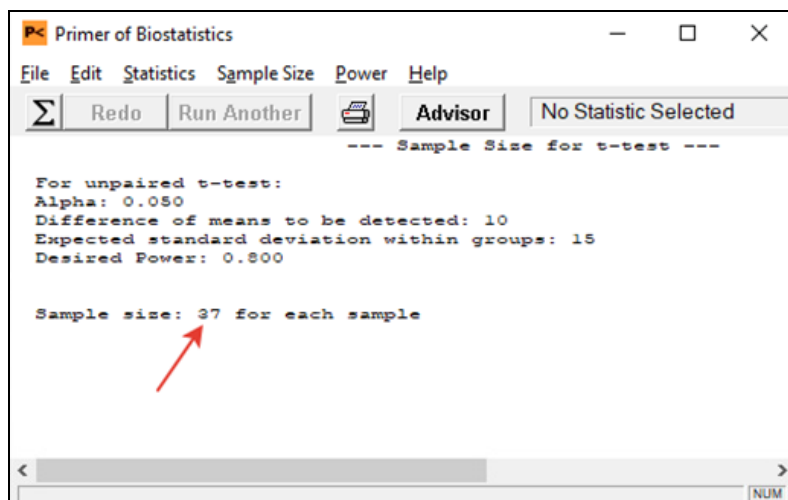


Рис. 5.5

Обратим внимание также, что в случаях с независимыми выборками расчет выдает равное количество испытуемых в них. Собственно, этой рекомендации следует придерживаться в любом случае — насколько это возможно.

На приведенном примере видно, что бытующие и зачастую верные представления о том, что в выборке исследования должно быть не менее 30 человек, не всегда корректны, а необходимо ориентироваться на конкретные показатели статистической значимости, мощности и описательные статистики.

В следующем примере рассмотрим случай с двумя зависимыми выборками, когда сравниваются сдвиги значений в повторных измерениях на одних и тех же субъектах, что предлагается сделать с помощью критерия Стьюдента для зависимых выборок (см. рис. 5.6). Например, сравнивают коммуникативные умения студентов до и после их участия в социально-психологическом тренинге.

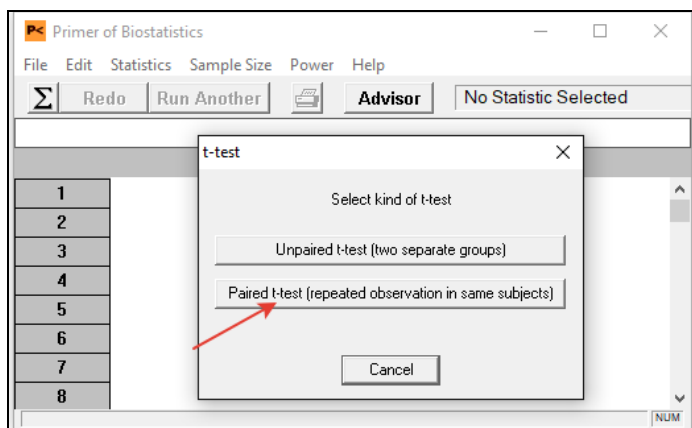


Рис. 5.6

Вводим показатели. Допустим, у нас так же значимым психологически будет считаться сдвиг в значении в 10 единиц. При этом мы получили в пилотажном исследовании или у нас есть основания полагать, что стандартная ошибка разности средних также составила 15 единиц. Выберем величину ошибки первого рода так же 0,05 (рис. 5.7), а мощность — 0,8 (рис. 5.8).

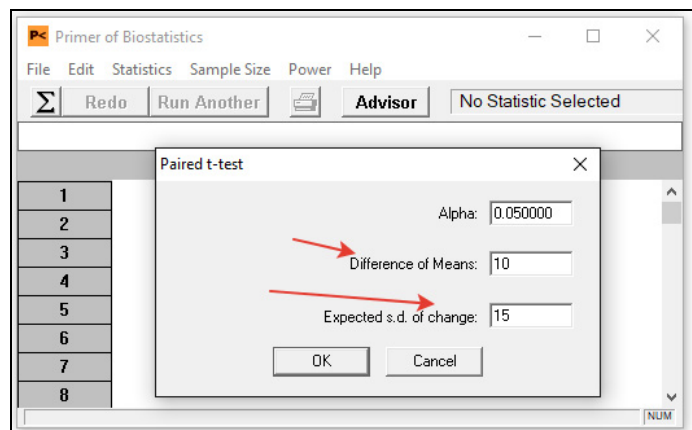


Рис. 5.7

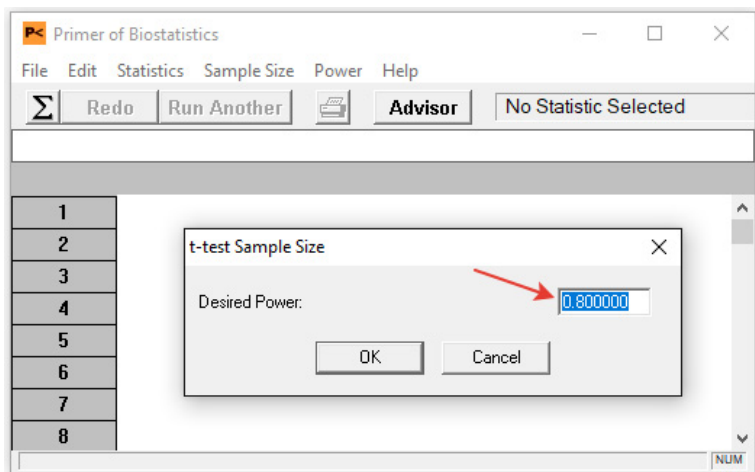


Рис. 5.8

В итоге получаем, что величина выборки должна составить не менее 20 испытуемых (рис. 5.9).

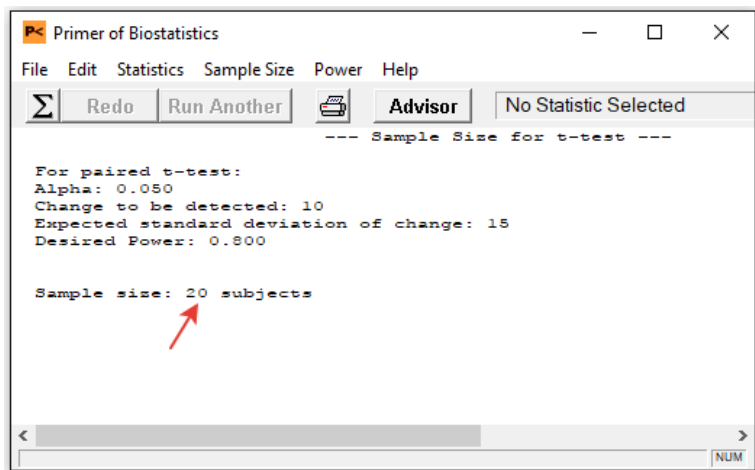


Рис. 5.9

Далее рассмотрим типичный и важный пример — расчет выборки для заданного коэффициента корреляции, который считается минимально приемлемым для исследователя как содержательно значимый или был получен в пилотажном исследовании и устроил исследователя как содержательно значимый.

К примеру, предполагается установить связь между стилем взаимодействия преподавателя и благоприятностью социально-психологического климата в классе.

Выбираем в меню соответствующий пункт Correlation (Корреляция) (рис. 5.10).

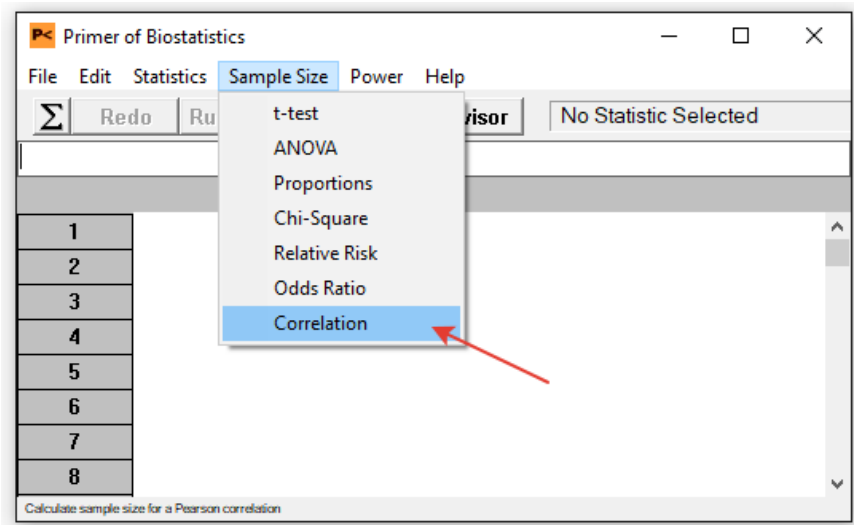


Рис. 5.10

В следующем меню выбираем желаемый уровень статистической значимости. Пусть он снова будет равен 0,05. Затем приемлемый минимальный содержательно значимый (обоснованный для данного исследования и с точки зрения самого экспериментатора) коэффициент корреляции — например, 0,5 (рис. 5.11).

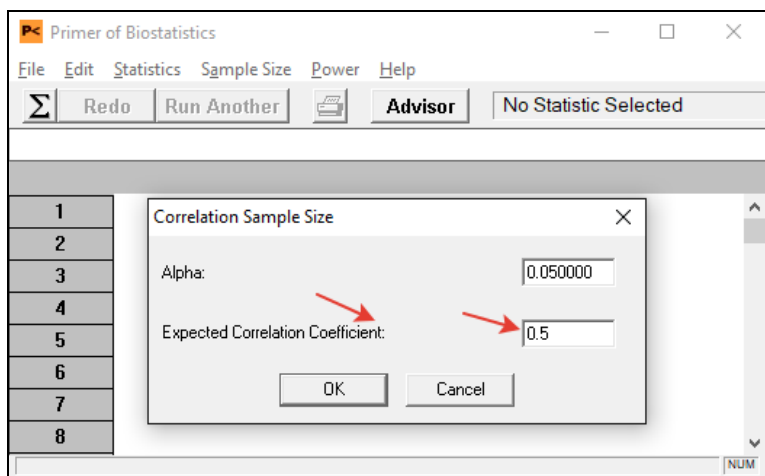


Рис. 5.11

Аналогично предыдущим примерам, пусть мощность будет стандартно равна 0,8 (рис. 5.12).

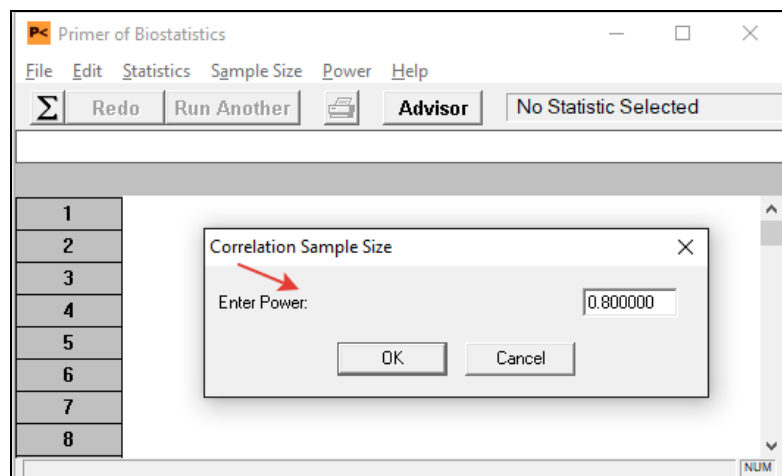


Рис. 5.12

В итоге получаем, что размер выборки должен составить 30 испытуемых (рис. 5.13).

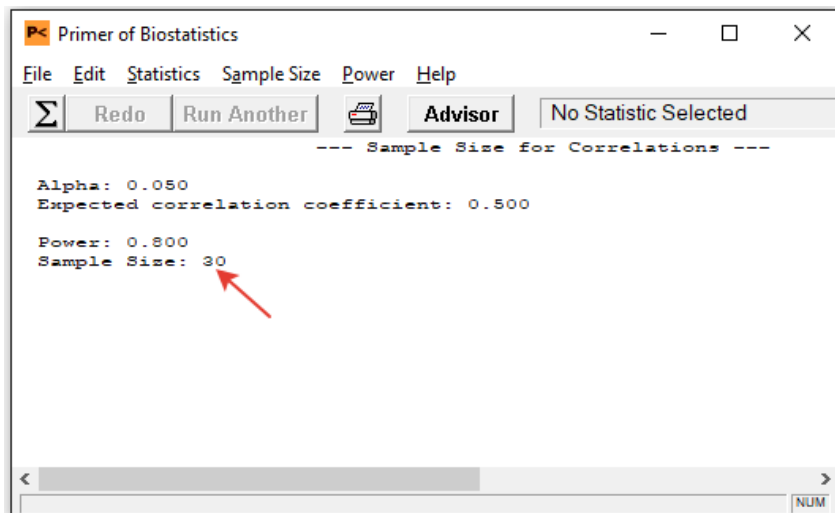


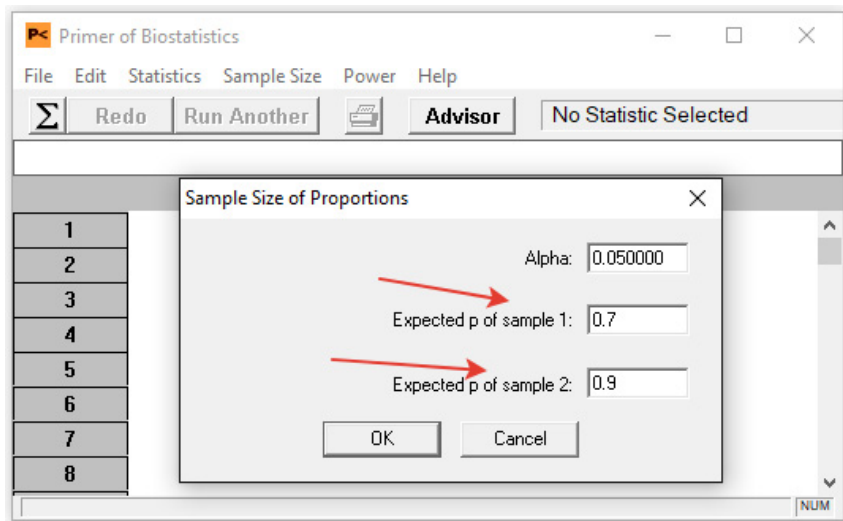
Рис. 5.13

Рассмотрим также типичный в исследованиях пример, когда сопоставляются две доли признака в независимых выборках (речь идет о качественных признаках).

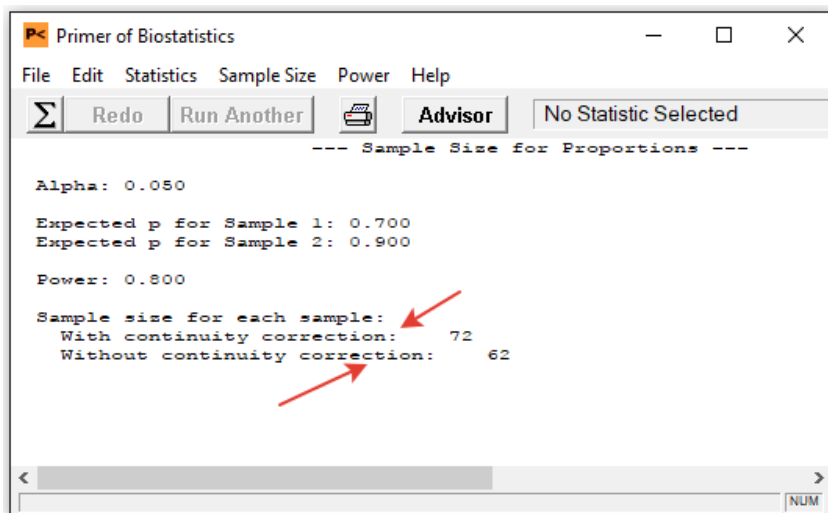
Например, в одной группе участников исследования 70% экстравертов, а в другой — 90%.

В меню Sample Size выбираем Proportions (Пропорции) и вносим соответствующие данные в единицах вероятности (рис. 5.14).

И получаем, что для данной ситуации следует выбрать не менее чем по 72 испытуемых, с условием поправки на непрерывность (корректней реализовать именно этот вариант с поправкой) (рис. 5.15).



Puc. 5.14



Puc. 5.15

Как можно увидеть в меню данной программы, существуют и другие дизайны, при которых рассчитывается объем выборок (для дисперсионного анализа ANOVA, отношения шансов Odds Ratio, отношения рисков Relative Risk, различий в распределениях Chi-Square). Если у читателя возникнет необходимость разобраться, как действовать в этих случаях, существует богатый справочный материал (в частности, источники из библиографии), который определяет эти понятия и параметры.

Еще более богатый арсенал дизайнов исследования, для которых можно вычислить размеры выборок исходя из предпочтения исследователя, предоставляют более сложные multifunctional статистические пакеты, такие как Statistica. В раздел Power Analysis (Анализ мощности) мы попадаем непосредственно из вкладки Statistics (рис. 5.16).

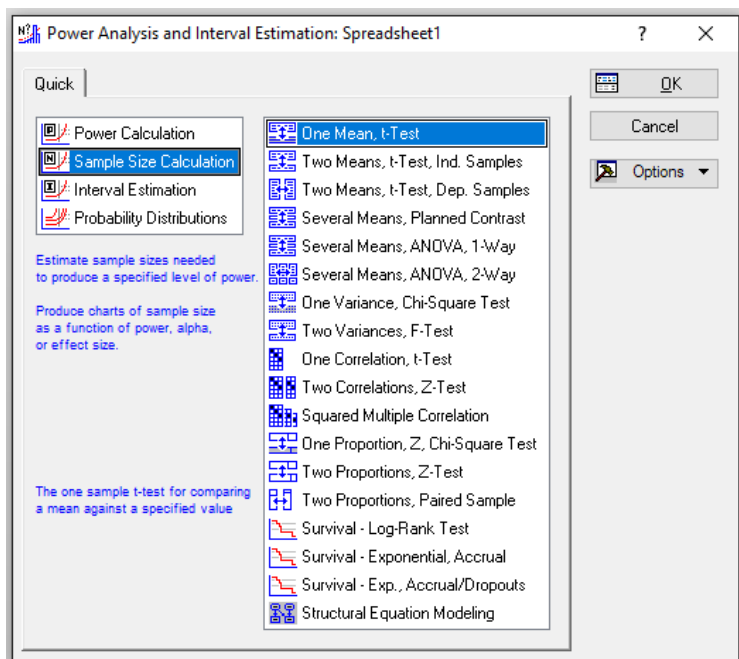


Рис. 5.16

Информацию о каждом типе дизайна (список справа) можно почерпнуть из справочных материалов системы Statistica, на специализированных официальных сайтах (например, statistica.ru) и т. д.

Однако изложение не было бы полным, если бы мы не остановились более подробно на задачах с одной выборкой (а не двумя или более, как рассматривали выше):

1) расчет объема выборки для массового опроса — задача, актуальная не только для социологии, но и собственно для психологии;

2) расчет объема выборки для того, чтобы (с заданной вероятностью) наблюдать определенное количество событий, распространенность которых в популяции нам известна (или неизвестна).

Данные виды расчета объема выборки реализованы в программе Biostat.

Сначала научимся рассчитывать объем выборки для массового опроса. После запуска программы выбираем последовательность «Статистика → Анализ мощности / Оценка объема выборки → Объем выборки для массового опроса (конечная ГС)...» (рис. 5.17).

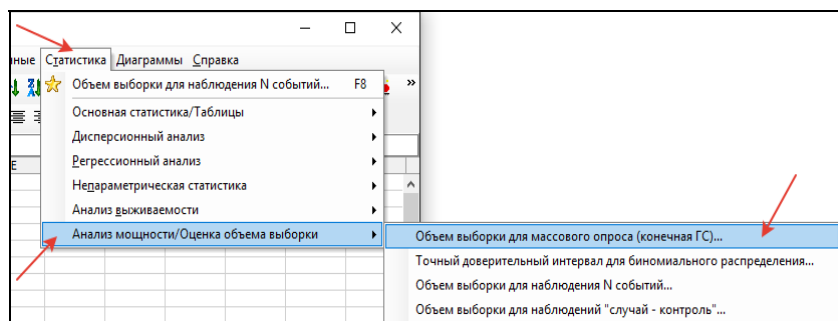


Рис. 5.17

Далее в открывшемся меню нам необходимо внести интересные значения (см. рис. 5.18).













| | |
|---|---|
| Точность (%) | |
|  | Введите число. [Необходимое поле]   |
| Распространенность явления (болезненность) в процентах % (50% если неизвестна) | |
|  | Введите число. [Необходимое поле]   |
| Генеральная совокупность (0 - если объем неизвестен) | |
|  | Введите число. [Необходимое поле]   |
| Уровень значимости для дов. интервала | |
|  | Введите число. [Необходимое поле]   |

Рис. 5.18

Рассчитывается объем выборки для заданного уровня значимости доверительного интервала (например, чаще всего 0,95, то есть 95%) с нижней границей «Распространенность явления – Точность» и верхней границей «Распространенность явления + Точность». При вводе значения «Генеральная совокупность» равным нулю она предполагается равной бесконечности. Для генеральных совокупностей конечного и малого размера применяется поправка, что приводит к меньшим объемам выборки.

Если мы представляем ожидаемую распространенность явления в процентах, то необходимо ее ввести. Например, сколько респондентов ответят положительно или отрицательно на интересующий нас вопрос. Однако зачастую это неизвестно. Поэтому, если это так, следует вводить 50%. Ввод именно этого значения имеет и математический смысл, поскольку объем совокупности максимален именно при 50%, а это значит, что в любом случае исследователь не ухудшит требуемую точность. Точно так же относительно генеральной совокупности. Если она ограниченного размера (например, все студенты-психологи, обучающиеся в Тюменском государственном университете), то есть нас интересуют только эти и никакие другие студенты, то следует указать ее объем (количество человек), так как в данном случае это позволит снизить требуемое число респондентов. Однако, если нам неизвестен объем генеральной совокупности или она крайне велика, то следует оставить «0», то есть ее объем будет полагаться равным бесконечности, и это опять же позволит не ухудшить точность исследования.

Рассмотрим пример, когда нам известен объем генеральной совокупности, скажем, это трудоспособное население России — примерно 83 млн человек по состоянию на 2017 год.

Точность определяет сам исследователь сообразно своим задачам, в соцопросах она чаще всего варьируется в пределах $\pm 2-4\%$. Допустим, нас устроит точность $\pm 3\%$. Далее, пусть доверительный интервал составляет 95% (0,95 в терминах вероятности), то есть 95% значений попадет в интервал $\pm 3\%$ от распространенности. Допустим, мы не знаем, как много людей из генеральной совокупности ответят «да» или «нет». Поэтому распространенность положим равной 50% (рис. 5.19).

| | |
|---|---|
| Точность (%) | |
| 3 | X |
| Распространенность явления (болезненность) в процентах % (50% если неизвестна) | |
| 50 | X |
| Генеральная совокупность (0 - если объем неизвестен) | |
| 83000000 | X |
| Уровень значимости для дов. интервала | |
| 0,95 | X |

Рис. 5.19

Далее получаем искомый результат (рис. 5.20).

| | | |
|----|---|------------|
| 1 | Объем выборки для массового опроса | |
| 2 | | |
| 3 | Допущения | |
| 4 | Точность (%) | 3 |
| 5 | неизвестна | 50 |
| 6 | Объем генеральной совокупности | 83 000 000 |
| 7 | 95% Определенные границы доверительного интервала | 53,0000% |
| 8 | | |
| 9 | Рассчитанный объем выборки | |
| 10 | <i>n</i> | 1068 |

Рис. 5.20

Теперь рассмотрим пример 2: как можно рассчитать необходимое количество участников выборки для того, чтобы среди них было минимум столько-то с заданным свойством. Например, мы хотим отобрать 10% самых способных учеников для участия в олимпиаде так, чтобы размер выборки составил 25 человек. Сколько всего человек нам надо отобрать в исходную совокупность, которая затем подвергнется исследованию на предмет интересующих нас способностей?

Запускаем меню «Статистика → Анализ мощности / Оценка объема выборки → Объем выборки для наблюдения N событий».

Вносим необходимые данные (рис. 5.21).

| Распространенность явления (болезненность) в процентах % (50% если неизвестна) | |
|--|---|
| 10 | X |
| Минимальное кол-во успешных случаев | |
| 25 | X |
| Уровень значимости | |
| 0,95 | X |

Рис. 5.21

Получаем результат (рис. 5.22).

| | | |
|---|--|-----|
| 1 | Объем выборки для наблюдения N событий | |
| 2 | | |
| 3 | Допущения | |
| 4 | Распространенность явления (болезненность) в процентах % (50% если неизвестна) | 10 |
| 5 | Минимальное кол-во успешных случаев | 25 |
| 6 | Уровень значимости | 95 |
| 7 | | |
| 8 | Рассчитанный объем выборки | |
| 9 | <i>n</i> | 333 |

Рис. 5.22

Таким образом, чтобы отобрать для участия в олимпиаде 25 человек, которые будут составлять 10% самых способных в определенном отношении испытуемых, нам следует изначально составить выборку из 333 человек.

Данный дизайн мог бы быть построен иначе. Например, мы бы изначально могли знать, сколько процентов составляет искомая часть популяции с заданным свойством (например, исходя из популяционных значений). Например, известно, что по состоянию на 2017 год 1% населения Тюменской области является носителем ВИЧ-инфекции. Исследователь может задать вопрос: сколько жителей Тюмени нужно отобрать (допустим, что все условия стратифицированного случайного отбора соблюдены), чтобы среди них было как минимум 25 ВИЧ-инфицированных? Аналогично, подставив значения в форму (вместо 10% в предыдущем примере — 1%), получим значение в 3372 человека.

Следует отметить в заключение данной части, что в психологических исследованиях часто требуется сравнить две выборки или реализовать другие дизайны исследования, но не по одному, а по множеству признаков одновременно — совокупности показателей нескольких опросников, тестов, экспериментальных процедур и т. п. Какой же в таком случае объем выборок следует выбрать? Наиболее общей могла бы стать рекомендация провести расчет размера выборки по каждому из этих признаков в отдельности и в итоге выбрать те объемы выборок или выборки, которые определяют самые «требуемые» или важные в отношении гипотез параметры, интересующие исследователя.

5.2. РЕТРОСПЕКТИВНЫЕ ИССЛЕДОВАНИЯ.

Отличия от проспективных исследований.

ОЦЕНКА МОЩНОСТИ ИМЕЮЩИХСЯ ВЫБОРОК

Помимо расчета размера выборок, нас может интересовать расчет мощности, который необходим в случае ретроспективных исследований.

В отличие от проспективных исследований, которые сначала планируются, потом выполняются по определенным правилам, ретроспективные исследования используют уже полученные ранее данные. В некоторых случаях ретроспективные исследования — единственно возможные, например, когда анализируется материал,

который был получен в прошлом другими исследователями, или в других местах, или требует перепроверки и т. д. Хорошим тоном в науке считается проведение проспективного исследования. Тем не менее, если перед нами, например, архивные данные, относительно которых известна методика проведения исследования, и мы надеемся, что оно было проведено согласно декларируемым правилам, то актуальной задачей становится анализ мощности. Мы должны понимать, какова будет вероятность безошибочно отвергнуть нулевую гипотезу при данных условиях (уровень статистической значимости, объемы выборок, наблюдаемые эффекты, как то: различие средних и соответствующее стандартное отклонение, коэффициент корреляции, доли признака и т. д.).

Рассмотрим теперь расчет мощности в тех же типах исследований, которые ранее рассматривали для расчета размера выборок. Будем придерживаться того же стандартного уровня статистической значимости 0,05.

В первом примере, допустим, мы наблюдали различия средних в 10 единиц при стандартном отклонении в 15 единиц (рис. 5.23).

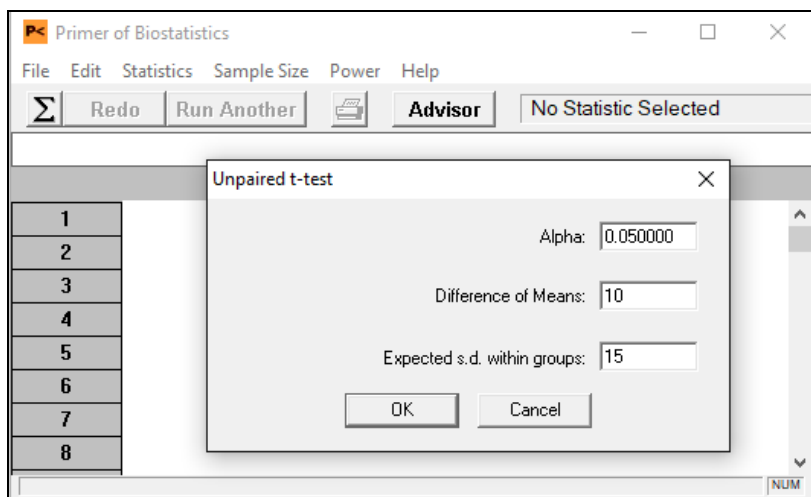


Рис. 5.23

Допустим, что в данном двухвыборочном исследовании участвовали соответственно 22 и 18 испытуемых в каждой из выборок (рис. 5.24).

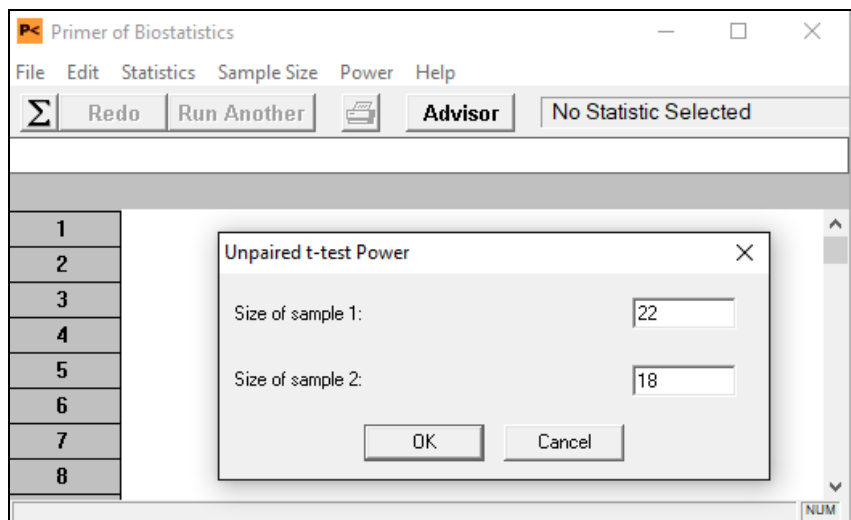


Рис. 5.24

Оказывается, что мощность данного исследования составила всего 0,534 (см. рис. 5.25). Иными словами, вероятность обнаружить верную альтернативную гипотезу о наличии различий составила в данном исследовании всего 53,4%, что недопустимо.

Как мы рассчитали в предыдущем пункте, рассматривая сходный пример в подобном исследовании для обеспечения адекватной мощности в 80% (0,8) необходимо не менее 37 человек в каждой выборке (см. рис. 5.5).

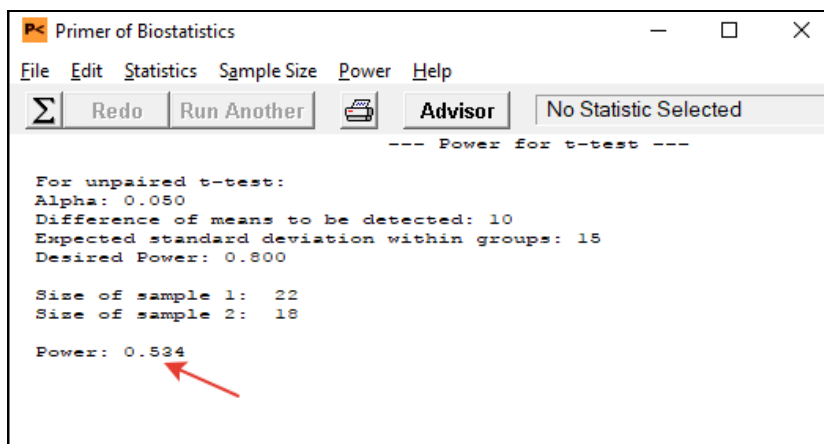


Рис. 5.25

Рассмотрим следующий пример, где участвуют первое и повторное измерения на одной и той же выборке (рис. 5.26). В числовом выражении оставим те же самые значения параметров, что и в примере, рассмотренном на проспективном исследовании.

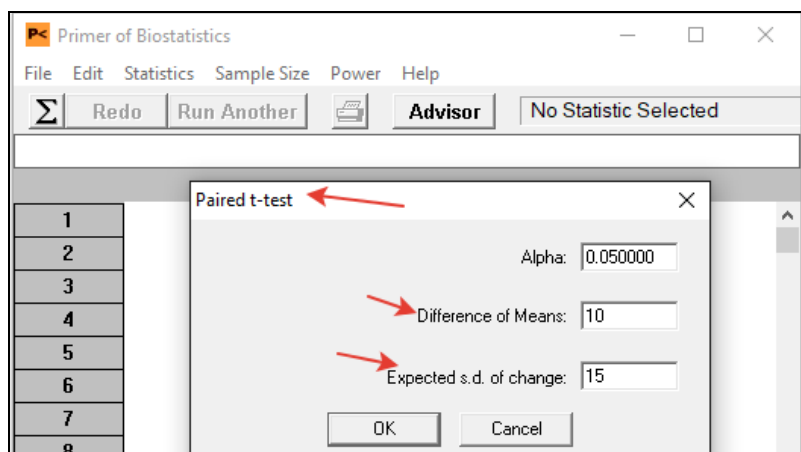


Рис. 5.26

Допустим, что в нем приняли участие 40 испытуемых (рис. 5.27).

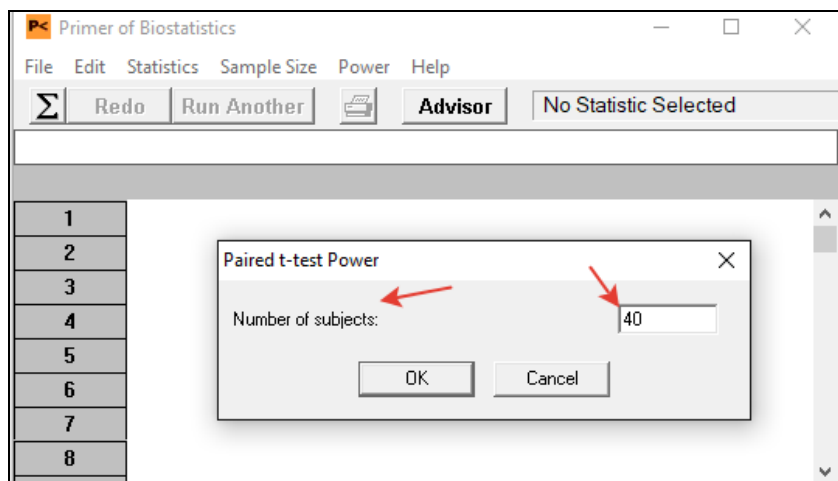


Рис. 5.27

Для данного исследования мощность уже составила 98,4%, что является очень хорошим результатом (рис. 5.28).

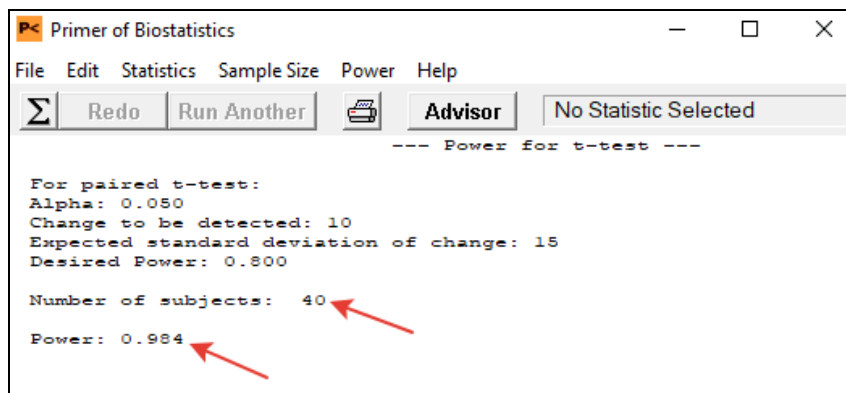


Рис. 5.28

Рассмотрим далее пример с линейной корреляцией (рис. 5.29).

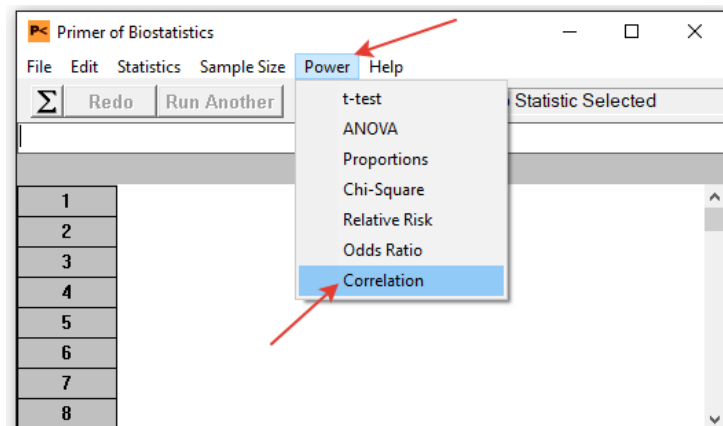


Рис. 5.29

Допустим, мы получили коэффициент корреляции, равный 0,5, и уровень значимости 0,05 (рис. 5.30) (на самом деле, «в реальной жизни», конечно, следует указывать именно тот уровень значимости, который был реально получен в исследовании, например, 0,023 или любой другой, полученный в конкретном исследовании).

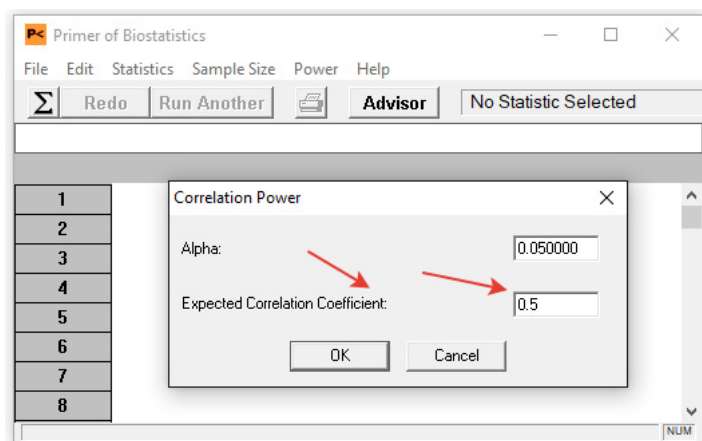


Рис. 5.30

Пусть объем выборки составил 30 человек (рис. 5.31).

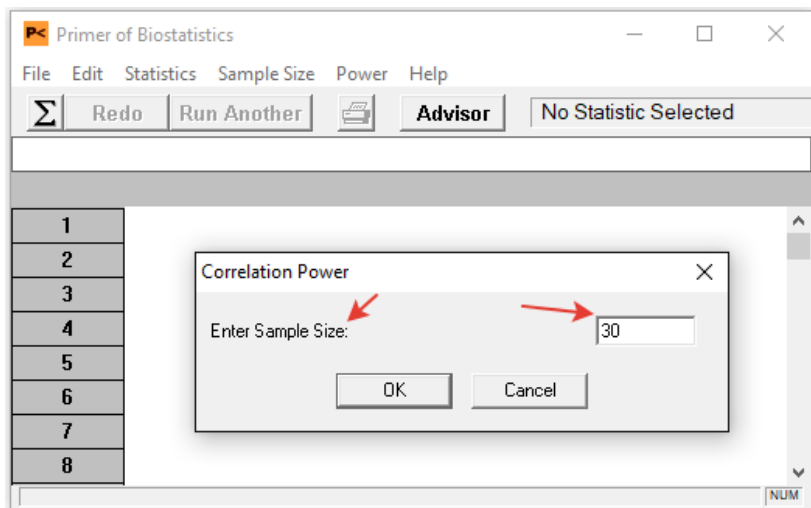


Рис. 5.31

В данном случае уровень мощности составил 81,4% (0,814) (рис. 5.32), что является вполне приемлемым.

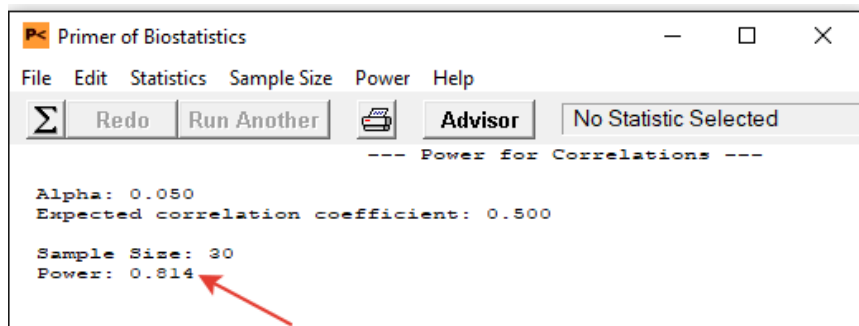


Рис. 5.32

Рассмотрим далее случай независимых выборок для качественных данных (см. рис. 5.33).

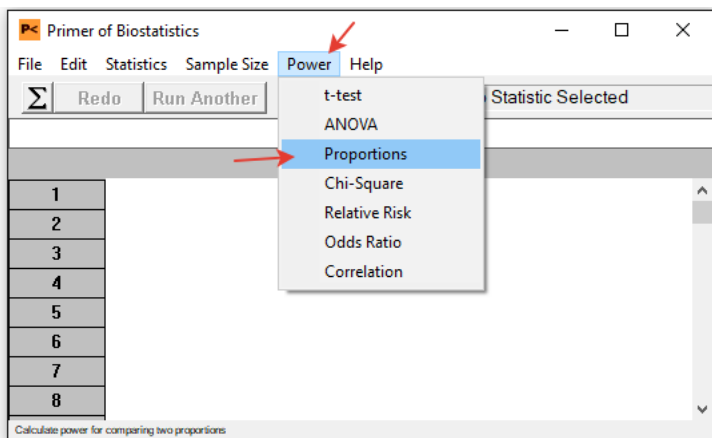


Рис. 5.33

Пусть в одной выборке доля интересующих нас исходов составила 30%, а в другой — 45%, например, испытуемых с высоким уровнем тревожности (рис. 5.34).

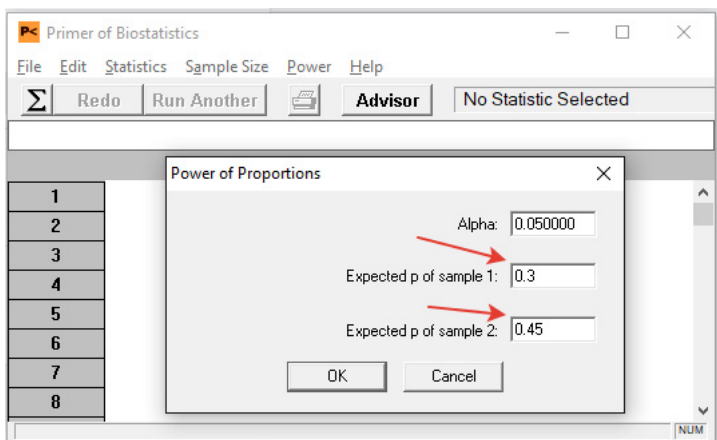


Рис. 5.34

Пусть объемы выборок составили, соответственно, 30 и 32 испытуемых (рис. 5.35).

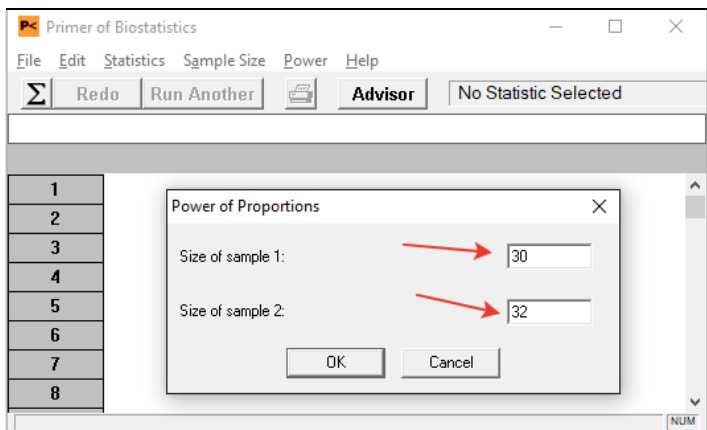


Рис. 5.35

Оказывается, что мощность в данном примере составила всего 15,4% (0,154) с учетом поправки на непрерывность (рис. 5.36), что, безусловно, крайне мало для суждения о том, имеются ли значимые различия в выборках или нет.

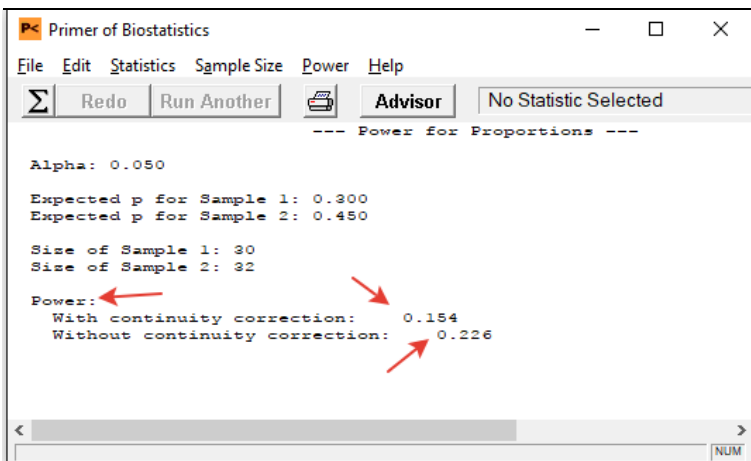


Рис. 5.36

Следовательно, без анализа ретроспективных исследований — полученных ранее нами без расчета объема выборок или другими исследователями — следует обязательно проводить анализ мощности, чтобы понять, насколько можно доверять нулевой или альтернативной гипотезе, полученной в данном исследовании.

6. ВВЕДЕНИЕ В ПРОБЛЕМАТИКУ ВОСПРОИЗВОДИМОСТИ РЕЗУЛЬТАТОВ ИССЛЕДОВАНИЙ

Представим ситуацию: исследователь-психолог подтвердил альтернативную гипотезу с высоким уровнем статистической значимости. Но зададимся вопросом: какова вероятность получить в другом исследовании при прочих равных условиях результат такой же направленности? Какова вероятность получить результат такой же, как в первом, или более высокой статистической значимости? А что, если существует много публикаций, сходных по методам, подходам и проблематике, подтверждают ли они или опровергают нашу гипотезу? Или мы получили новый результат и хотим знать, действительно ли он воспроизведется в других исследованиях, максимально сходных по дизайну и методам, и найдутся коллеги с безупречной научной репутацией, но скептически настроенные в отношении полученных нами результатов, которые возьмутся повторить наше исследование?

Если кто-то из наших читателей задавался хотя бы одним из этих или похожих вопросов, то он вступил на тернистый путь проблематики воспроизводимости результатов исследования. Вообще говоря, воспроизводимость исследований является одним из ключевых требований фундаментальных и отчасти естественных наук, медицины, зарубежной психологии и т. д.

Воспроизводимость исследований — это точное (для фундаментальных наук) или приблизительное сходство результатов, где бы и кем бы они ни были получены, при повторении всех условий эксперимента (в психологии — ключевые особенности испытуемых, одинаковые условия, методы и инструментарий эксперимента, размер выборки не меньший, чем в исходном эксперименте, много независимых лабораторий в разных городах и странах, если инструментарий одинаково соответствует особенностям восприятия и понимания испытуемыми, и т. д.).

Покажем на примере двухвыборочного критерия Стьюдента с одинаковым количеством испытуемых в каждой из выборок, как это работает.

Число степеней свободы для данного критерия рассчитывается как $f = (n_1 + n_2) - 2$.

Используем для этой иллюстративной задачи лишь некоторые из возможностей программы LePrep (рис. 6.1).

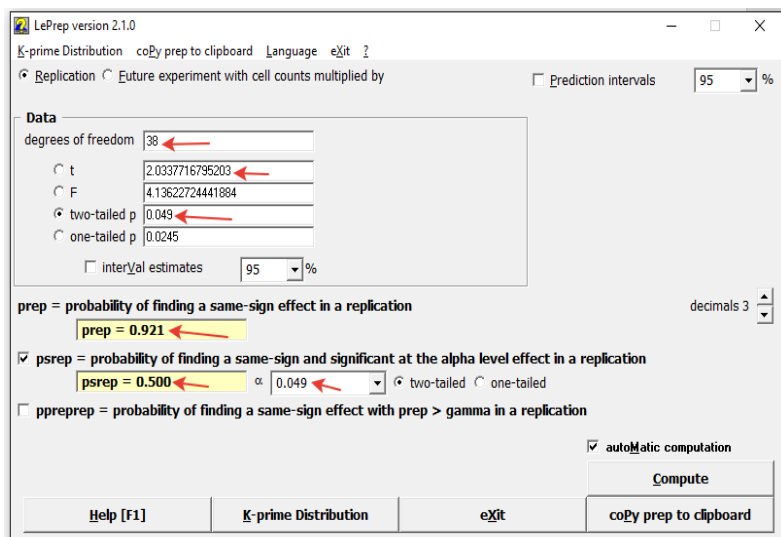


Рис. 6.1

Представим, что у нас в выборках по 20 испытуемых, по каждому из которых одно измерение. Распределение выборок не отличается от нормального, дисперсии двух выборок примерно равны, то есть соблюдены все необходимые условия для использования параметрического критерия. Представим, что мы имели дело с двухсторонней критической областью. Итак, количество степеней свободы равно 38 ($20 + 20 - 2$). Представим, что статистическая значимость составила 0,049, то есть почти на границе допустимого, пока еще верна альтернативная гипотеза, значение распределения

Стьюдента примерно равно 2,03 (см. указательные стрелки вверху скриншота программы).

Мы хотим узнать, во-первых, какова вероятность обнаружить эффект того же знака (разность средних в выборках будет той же, что и в исходном эксперименте); во-вторых, какова вероятность воспроизвести уровень статистической значимости не хуже, чем в исходном эксперименте. Увы, мы получаем не столь уж утешительные результаты: соответствующие вероятности будут равны 0,921 и 0,5. Последнее число особенно обескураживает: оказывается, что воспроизвести этот результат с уровнем значимости, не худшим, чем в исходном, удастся только в половине исследований!

Попробуем теперь другой вариант (рис. 6.2). Допустим, что мы получили на тех же выборках статистически очень высоко значимый результат с $p = 0,00049$, и теперь хотим узнать, как часто нам удастся воспроизвести в повторном эксперименте результат со статистической значимостью не хуже, чем с $p = 0,049$, то есть фактически получить повторно тоже статистически значимый результат, пусть и на наименьшем из возможных уровней значимости, для отвержения нулевой гипотезы.

LePrep version 2.1.0

K-prime Distribution coPy prep to clipboard Language eXit ?

☒ Replication ☐ Future experiment with cell counts multiplied by

☐ Prediction intervals 95 %

Data

degrees of freedom 38

☐ t 3.81282717548647

☐ F 14.5376510701281

☒ two-tailed p 0.00049

☐ one-tailed p 0.000245

☐ interval estimates 95 %

prep = probability of finding a same-sign effect in a replication decimals 3

prep = 0.995

☒ psprep = probability of finding a same-sign and significant at the alpha level effect in a replication

psprep = 0.881 α 0.049 ☒ two-tailed ☐ one-tailed

☐ pprep = probability of finding a same-sign effect with prep > gamma in a replication

☒ autoMatic computation

Compute

Help [F1] K-prime Distribution eXit coPy prep to clipboard

Рис. 6.2

Как видно из скриншота, ситуация явно изменилась к лучшему: вероятность получить эффект того же знака уже близка к единице ($prep = 0,995$), но вероятность получить статистически значимый результат, как и в предыдущем примере, на уровне 0,049, то есть, в принципе, отвергнуть нулевую гипотезу на близком к критическому уровне значимости, составляет лишь ($psprep = 0,881$), то есть чуть менее чем в девяти из десяти случаев. И это при условии, что в начальном эксперименте наблюдалась очень высокая статистическая значимость! В данном случае ситуация лишь несколько улучшится, если применять направленные статистические гипотезы (с односторонней критической областью).

В психологии (как и в некоторых других науках) к воспроизводимости результатов исследований приковано самое пристальное внимание, и ведутся острые дебаты. Так, в 2015 году В. А. Nosek [20] реализовал крупномасштабный проект по воспроизводимости психологических исследований на одну и ту же тему (порядка сотни работ). Экспертами были отобраны исследования, опубликованные в журналах “Psychological Science”, “Journal of Personality and Social Psychology”, “Journal of Experimental Psychology: Learning, Memory and Cognition”, введены критерии оценки соответствия оригинального исследования и воспроизводимости. В итоге, даже при условии сотрудничества исследователей, проводящих репликацию, и авторов оригинальной работы, «средний размер эффекта воспроизводимости ($M_r = 0,197$, $\sigma = 0,257$) оказался в два раза меньше величины среднего эффекта размера исходных эффектов ($M_r = 0,403$, $\sigma = 0,188$), что представляет собой существенное снижение». Несмотря на неутешительные показатели воспроизводимости, группа исследователей из организации “Center for Open Science” (США) сделала вывод о необходимости проведения дальнейших репликаций, поскольку они указывают на потенциальные пути улучшения научных исследований. В частности, сделан вывод о том, что наиболее значимым параметром воспроизводимости является уровень статистической значимости, в сравнении с параметрами опыта и квалификации авторов эксперимента. На данную работу поступили комментарии, защищающие психологию от

обвинений в «кризисе воспроизводимости» [18]. Эти авторы указывают на то, что В. А. Nosek при проведении анализа использовал критерий, не учитывавший ошибки в данных, был выбран недостаточно мощный дизайн для оценки истинного размера эффекта воспроизводимости, допускались нарушения в организации экспериментальных процедур, приведших к невоспроизводимости данных. Stroebe W. и Strack F. [22] полагают необоснованными заявления о «кризисе воспроизводимости», поскольку репликации в большей части ориентируются на воспроизведение собственно феномена, а не его механизмов. Главный редактор журнала “Science” высказалась о мерах в редакционной политике, которые позволят убедить читателей в достоверности представляемой в журнале информации: привлечение экспертов в области статистики для оценки качества статей, запрос к авторам на предоставление материалов, удостоверяющих полученные в статье данные [19].

Исследователи предлагают следующие меры по улучшению воспроизводимости: отслеживание репликаций уже опубликованных исследований посредством создания базы открытого доступа к репликациям; контроль над реализацией условий, отражающих измеряемый концепт; использование подхода в метаанализе, подразумевающего его постоянное пополнение данными (the continuously cumulating meta-analytic approach).

Есть указания на то, что для разных областей наук о человеке должны быть разные требования к воспроизводимости: например, в психологии развития и в поведенческих науках (например, в прикладной экономике). При этом речь идет не о снижении достоверности предоставляемых результатов, а о вариациях в размерах эффектов [17]. Можно встретить мнение, что воспроизводимость в психологических исследованиях не является тем критерием научности, на который необходимо ориентироваться в первую очередь: «с методологических позиций постнеклассической науки можно говорить о новых нормах научности в психологии, прежде всего таких, как диалогичность, плюралистичность, множественность истин, нестабильность как свойство систем» [3]. В. Н. Бинги вводит представление о методологии невоспроизводимости как в фи-

зике, так и в психологии, на основе «неполной реализации методологического принципа объективности — разделения объекта и субъекта исследования — <...> что означает допустимость влияния сознания экспериментатора на результат измерений» [1, с. 508]. В психологических исследованиях данная методология заслуживает особого внимания в силу самых разных эффектов: скептицизма или веры испытуемого или экспериментатора в изучаемый феномен, множества ситуативных факторов, априорной невозможности не только отбора совершенно гомогенных испытуемых в аналогичные по дизайну исследования, но и принципиально не учитываемые изменения в состоянии не только участников эксперимента, но и самой психологической атмосферы, равно как и погодных и прочих условий. Кроме того, у репликаторов может быть искушение привести в эксперимент, подвергающийся репликации, какие-то пусть незначительные детали, не совпадающие с исходным экспериментом, и т. д.

Таким образом, несмотря на множество мнений о необходимости соблюдения воспроизводимости в психологических исследованиях, следует, скорее, ориентироваться на сходство устанавливаемых психологических механизмов и паттернов (направленности эффектов), нежели единственно на уровни значимостей. Поскольку на примере даже строгих математических расчетов, которые можно провести, например, как было показано выше, в программе LePter или аналогичных, не всегда может быть достигнут принятый уровень статистической значимости, равный 0,05.

Наиболее эффективными на сегодня способами заранее застраховаться от обвинений в осознанных или неосознанных подтасовках и артефактах (experimenter fraud) являются:

1. Использование информационных технологий для автоматической фиксации плана эксперимента, гипотезы, результатов, их «слепая» обработка, предварительная публикация плана эксперимента в многочисленных специальных, доступных научному сообществу регистрах, например: <http://centerforopenscience.org>, <https://clinicaltrials.gov> (Служба национальных институтов здравоохранения в США), <https://www.socialscienceregistry.org> (Регистр

Американской экономической ассоциации для рандомизированных контролируемых испытаний), <http://egar.org> (Регистр доказательств в области управления и политики), <https://aspredicted.org> (база пререгистрации экспериментов, созданная отдельной группой ученых), <http://www.who.int/ictip/network/en> (Международная организация здравоохранения, платформа регистрации клинических испытаний). Редакторы психологических журналов все чаще требуют предварительной регистрации экспериментов, результаты которых предлагаются к опубликованию. Более того, пререгистрация становится или обязательной в ряде журналов, или фактором, повышающим вероятность публикации статьи, а не ее отклонения (пример такого рода требований приводится на сайте Association for psychological science (http://www.psychologicalscience.org/publications/psychological_science/preregistration) и на сайте American Psychological Association (<http://www.apa.org/science/about/psa/2015/08/pre-registration.aspx>)).

2. По примеру доказательной медицины проведение проспективных двойных или лучше тройных слепых многоцентровых контролируемых рандомизированных исследований, строящихся по идентичному плану, и затем совокупная их обработка уже, в том числе, и по правилам метаанализа литературы. См., например, классическую статью по метаанализу в области психологии *Meta-analysis of psychotherapy outcome studies* [21], а также другие ресурсы и источники, в частности: <https://www.meta-analysis.com>.

Разъясним понятия, которые могут часто встречаться в описании методов исследования и обработки его результатов, независимо от того, какой именно дизайн эксперимента используется исследователем.

В *открытом* исследовании и испытуемый, и экспериментатор, и статистик в общем знают цель и смысл эксперимента. В *слепом* эксперименте испытуемые не знают истинной цели исследования. В *двойном слепом* — экспериментатор так же не знает, что на самом деле составляет предмет и цель исследования. В *тройном слепом* — статистик тоже не знает, какие именно данные и для чего он обрабатывает. Ясно, что чем более полной является «слепота», тем

лучше мы исключаем разного рода эффекты — веры, скептицизма, выгоды испытуемого, исследователя, статистика и т. д.

Контролируемое исследование подразумевает наличие одной или нескольких контрольных групп, чаще из одной генеральной совокупности, должным образом рандомизированных. Контрольной группой чаще всего является не группа, с которой «ничего не происходит», а та или те, с которыми проводится определенная экспериментальная работа или организуется наблюдение и по отношению к которой или которым нам требуется установить или опровергнуть наличие различий или зависимостей разного рода.

Наконец, проведение исследований разными экспериментаторами с разными испытуемыми в разное время и/или в разных местах, но с использованием идентичной методики исключает эффекты веры/скептицизма, недостатков рандомизации, позволяет отсеять региональные или иные эффекты, которыми могли бы быть обусловлены результаты эксперимента. Другие исследователи не должны знать истинных целей и задач эксперимента, а также приветствуется, чтобы среди них были не только симпатизирующие конкретному заказчику (организатору исследования, теоретику), но и скептически настроенные.

Естественно, соблюдение этих правил требует высокой профессиональной этики и, в идеале, немедленной фиксации результатов наблюдений в Интернете, на соответствующих порталах, а также на электронных носителях, желательно в удаленном режиме, в единой системе (для многоцентровых исследований), во избежание подтасовок результатов. Без соблюдения таких правил, в среднесрочной перспективе, ведущие мировые научные издания, вероятно, уже не будут оценивать научные исследования в областях психологии, социологии, политологии и т. п. как проведенные на высоком уровне доказательности.

ЗАКЛЮЧЕНИЕ

Итак, мы рассмотрели основы математико-статистических методов в психологии, дали некоторые рецепты по выбору критерия или технологии обработки данных в некоторых базовых дизайнах психологических исследований. Несомненно, на практике исследователь, в частности, обучающийся в аспирантуре, может столкнуться с необходимостью применять другие методы, не рассмотренные в данном учебном пособии. Для этого мы отсылаем читателя к доступным учебникам и пособиям, где на высоком методическом уровне описаны многие из нерассмотренных здесь методов.

Для углубленного понимания основ статистики и, в частности, множества вариантов распределений данных (не только нормального, не исчерпывающего всего многообразия случаев возможных распределений) рекомендуем изучить классический учебник Г. Ф. Лакина [8], разделы: III.5 «Биномиальное распределение»; III.6 «Распределение Пуассона»; III.7 «Параметры дискретных распределений»; III.9 «Распределение Максвелла»; III.11 «Распределение Шарлье»; VIII.3 «Множественная и частная корреляция».

Учебник Е. В. Сидоренко [15] написан специально для психологов и иллюстрирует решение множества психологических примеров. Рекомендуем к изучению разделы 2.2 «Критерий Розенбаума»; 2.5 «Критерий тенденций Джонкира»; 3.2 «Критерий знаков»; 3.5 «Критерий тенденций Пейджа».

Пособие Е. Л. Доценко и З. З. Вахитовой [7] будет полезно разделами, посвященными факторному и кластерному анализу, а также многомерному шкалированию данных.

В пособии Р. И. Остапенко [12] в разделе 2.1 дано понятное описание метода непараметрической корреляции «Коэффициент корреляции τ -Кендалла».

В практикуме О. В. Митиной [10] следует обратить внимание на часть 4, главу 11 «Проведение статистического анализа использования психодиагностического опросника».

Содержательная основа книги С. Гланца [4] — медицинская статистика. Однако большая часть примеров с легкостью экстраполируется на предметное поле психологии — на строгом и в то же время понятном языке. Особый интерес представляют разделы 5 «Анализ качественных признаков»; 7 «Доверительные интервалы»; 11 «Анализ выживаемости».

В пособии Г. В. Рублевой [13] рассмотрены не только линейные, но и нелинейные корреляционные и регрессионные меры связи между признаками, даются доступные формулы для расчетов показателей. Пособие насыщено понятными примерами, пусть и не из психологии. Методы доступны для самостоятельного, без применения пакетов статистики, расчета. Тем более что не во всех доступных и популярных пакетах статистики они присутствуют.

Книга Дж. Гудвина [5] может быть рекомендована для изучения методологии планирования и организации психологических исследований, что имеет первостепенное значение для аспирантов и магистрантов.

В книге Б. Л. Вардена [2] основы статистики излагаются с определенной математической строгостью. Она может быть полезна пытливым читателям, которым действительно понадобится вникнуть в предмет несколько глубже, чем в большинстве случаев.

Мы показали, как те же самые или различные задачи математико-статистической обработки данных могут быть успешно решаемы в различных статистических пакетах (их много десятков, если не сотен, заслуживающих внимания), так что исследователю вовсе не обязательно «привязываться» к одному из них. Более детальное рассмотрение принципов работы с некоторыми программами можно найти в соответствующих руководствах пользователя [14, 16] и пособиях.

Учебное пособие Ю. Е. Ляха и др. [9] содержит почти исчерпывающие данные по большинству базовых методов, которые могут понадобиться в психологическом исследовании. Переформулировка примеров с языка биологии и медицины не составит труда. Особенно имеет смысл обратить внимание на следующие части книги: пункт 7.1 «Критерий Шапиро–Уилка» (проверка на нормальность);

глава 14 «Количественная оценка эффекта лечения (воздействия)» — по сути дела, оценка эффекта любого воздействия, в том числе психокоррекционного.

В книге А. Наследова [11] наиболее понятно и пошагово описаны многомерные эмпирические математико-статистические методы (главы: 19 «Анализ надежности»; 22 «Дискриминантный анализ»; 23 «Многомерное шкалирование»; 24 «Логистическая регрессия»; 25 «Логлинейный анализ таблиц сопряженности»; 26 «AMOS: Моделирование структурными уравнениями»).

В случае необходимости стоит обратиться к теме 17 «Компьютерные программы психолингвистического анализа текста» в учебном пособии В. Н. Денисенко и Е. Ю. Чеботаревой [6].

В электронном учебнике по программе Statistica [16] следует обратить внимание на пункт «Анализ выживаемости». Данная группа методов применяется не только в медицине и биологии, но также, например, в исследованиях отдаленных эффектов психокоррекции, изменений установок и т. п.

Авторы надеются, что благодаря избранному стилю изложения — почти без математических формул, но с разъяснением смысла каждого показателя — читатель смог прочитать эту книгу до конца, уяснить некоторые навыки работы с типовыми статистическими пакетами.

Наконец, следует помнить, что в психологических исследованиях математика и статистика — это не самоцель, а лишь способ обработки данных и формальный инструмент проверки на «прочность» той или иной гипотезы исследователя из безграничного поля психологической науки.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Бинги, В. Н. Принципы электромагнитной биофизики / В. Н. Бинги. — М.: Физматлит, 2011. — 593 с.
2. Варден, Б. Л. Математическая статистика / Б. Л. Варден; пер. с нем. — М.: Изд-во иностранной литературы, 1960. — 435 с.
3. Вачков, И. В. Воспроизводимость психологических экспериментов как проблема постнеклассической науки / И. В. Вачков, С. Н. Вачкова // Культурно-историческая психология. — 2016. — Т. 12, № 1. — С. 97-101. — DOI: 10.17759/chp.2016120110
4. Гланц, С. Медико-биологическая статистика / С. Гланц; пер. с англ. — М.: Практика, 1998. — 459 с.
5. Гудвин, Дж. Исследование в психологии: методы и планирование / Дж. Гудвин. — 3-е изд. — СПб.: Питер, 2004. — 558 с.
6. Денисенко, В. Н. Современные психолингвистические методы анализа речевой коммуникации: учеб. пособие / В. Н. Денисенко, Е. Ю. Чеботарева. — М.: РУДН, 2008. — 258 с.
7. Доценко, Е. Л. Психосемантика: учеб. пособие / Е. Л. Доценко, З. З. Вахитова. — Тюмень: Изд-во Тюменского гос. ун-та, 2014. — 292 с.
8. Лакин, Г. Ф. Биометрия: учеб. пособие для биол. спец. вузов / Г. Ф. Лакин. — 4-е изд., перераб. и доп. — М.: Высшая школа, 1990. — 350 с.
9. Лях, Ю. Е. Основы компьютерной биостатистики: анализ информации в биологии, медицине и фармации статистическим пакетом MedStat: учеб. пособие / Ю. Е. Лях [и др.]. — Донецк: Папакица Е. К., 2006. — 214 с.
10. Митина, О. В. Математические методы в психологии: практикум / О. В. Митина. — М.: Аспект Пресс, 2008. — 238 с.
11. Наследов, А. IBM SPSS Statistics 20 и AMOS: профессиональный статистический анализ данных / А. Наследов. — СПб.: Питер, 2013. — 416 с.
12. Остапенко, Р. И. Многомерный анализ данных для психологов: учеб.-метод. пособие / Р. И. Остапенко. — Воронеж: ВГПУ, 2012. — 72 с.
13. Рублева Г. В. Математическая статистика: изучение взаимосвязей между признаками: учеб.-метод. пособие / Г. В. Рублева. — Тюмень: Изд-во Тюменского гос. ун-та, 2014. — 30 с.

14. Руководство пользователя по базовой системе IBM SPSS Statistics 20. — 475 с. [Электронный ресурс]. — Режим доступа: ftp://public.dhe.ibm.com/software/analytics/spss/documentation/statistics/20.0/ru/client/Manuals/IBM_SPSS_Statistics_Core_System_Users_Guide.pdf (дата обращения: 31.12.2017).
15. Сидоренко, Е. В. Методы математической обработки в психологии / Е. В. Сидоренко. — СПб.: Речь, 2006. — 350 с.
16. Электронный учебник по статистическому пакету Statistica StatSoft Inc. [Электронный ресурс]. — Режим доступа: <http://statsoft.ru/home/textbook/default.htm> (дата обращения: 31.12.2017).
17. Duncan, G. J. Replication and robustness in developmental research / G. J. Duncan [et al.] // *Developmental Psychology*. — 2014. — Nov. — Vol. 50(11). — P. 2417-2425. — DOI: 10.1037/a0037996
18. Gilbert, D. T. Comment on “Estimating the reproducibility of psychological science” / D. T. Gilbert [et al.] // *Science*. — 2016. — Mar 4. — Vol. 351(6277). — P. 1037. — DOI: 10.1126/science.aad7243
19. McNutt, M. Reproducibility / M. McNutt // *Science*. — 2014. — Jan 17. — Vol. 343(6168). — P. 229. — DOI: 10.1126/science.1250475
20. Open Science Collaboration. PSYCHOLOGY. Estimating the reproducibility of psychological science // *Science*. — 2015. — Aug 28. — Vol. 349(6251). — aac4716. — DOI: 10.1126/science.aac4716
21. Smith, M. L. Meta-analysis of psychotherapy outcome studies / M. L. Smith, G. V. Glass // *American Psychologist*. — 1977. — Sep. — Vol. 32(9). — P. 752-60.
22. Stroebe, W. The alleged crisis and the illusion of exact replication / W. Stroebe, F. Strack // *Perspectives on Psychological Science*. — 2014. — Jan. — Vol. 9(1). — P. 59-71. — DOI: 10.1177/1745691613514450

Учебное издание

ГРИГОРЬЕВ Павел Евгеньевич
ВАСИЛЬЕВА Инна Витальевна

СТАТИСТИЧЕСКИЕ МЕТОДЫ В ПСИХОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ

Учебное пособие

Верстка и редактирование издания выполнены в соответствии
с Гражданско-правовым договором автономного учреждения
№ 2Т/01068-18 от 18 июня 2018 года

Печать электрографическая *А. В. Баширов*

Печать офсетная *В. В. Торопов, С. Г. Наумов*



Подписано в печать 17.07.2018. Тираж 50 экз.
Объем 12,56 усл. печ. л. Формат 60×84/16. Заказ 644.

Издательство Тюменского государственного университета
625003, г. Тюмень, ул. Семакова, 10
Тел./факс: (3452) 59-74-68, 59-74-81
E-mail: izdatelstvo@utmn.ru