Министерство науки и высшего образования Российской Федерации

Федеральное государственное автономное образовательное учреждение высшего образования «Новосибирский национальный исследовательский государственный университет»
(Новосибирский государственный университет, НГУ)

Мезенцева Мария Павловна

ОБРАБОТКА ПСИХОЛОГИЧЕСКИХ ДАННЫХ

Учебное пособие

Учебное пособие подготовлено в рамках реализации задачи развития образовательного процесса Программы развития Новосибирского государственного университета – победителя конкурса отбора программ развития университетов, в отношении которых устанавливается категория «национальный исследовательский университет»

В пособии рассмотрены базовые навыки обработки эмпирических данных психологической тематики. В качестве технологии предлагаются вопросы, призванные помочь в выборе критериев обработки. Математико-статистическое описание выводов оснащено конкретными примерами данных, алгоритмами подсчёта (согласно пунктам меню статистических программ) и результатами подсчёта для самопроверки на примере свободно распространяемых статистических программ (AtteStat, Rundom Pro, Pajek, Ministep - printscreen), а также коммерческих (SPSS и STATISTICA – интерактивные демонстрации). Уделяется особое внимание графическому представлению полученных результатов, осознанной проверке результатов подсчётов, пересечению обработки данных в экспериментальной психологии и психодиагностике.

Рекомендуется в качестве учебного пособия для бакалавров и магистров психологического направления. Книга может быть полезна также студентам, магистрантам, аспирантам гуманитарных специальностей.

Введение

Экспериментальная психология представляет собой одну из наиболее масштабных дисциплин психологической тематики. Каждый автор вводит собственное определение её границ, но главное, чего стоит избегать – это определения границ исключительно через всю совокупность эмпирических знаний психологии. Сделав содержанием дисциплины лишь результат деятельности учёных, мы никогда не сможем понять, как его продолжить. Здесь стоит вспомнить слова Фейерабенда: «Революции преобразуют не только практику тех, кто стремится к изменениям, но и сами принципы, посредством которых они намереваются осуществить изменения»[1].

В данном пособии предпринята попытка изложить простым и доступным языком одну из самых туманных для студентов-психологов сторону исследовательской работы – обработку данных. Психолог находится в достаточно неудобном положении перед лицом этой задачи, поскольку образовательная программа носит отрывочный и неполный характер. Отрывочный, потому что знания, необходимые для успешной обработки исследовательских данных, располагаются в трёх дисциплинах: математической статистике, психодиагностике и экспериментальной психологии. Этот перечень, естественно, подразумевает область методологии «как делать обработку данных», являющуюся инструментальной стороной, не касающейся сути исследования, отвечающей на вопрос «что исследовать». Неполный характер образовательной программы сказывается в подготовленности студентов к математической статистике. Строясь на теории вероятности и математическом анализе, эта дисциплина навсегда закрыта для большинства психологов в своём теоретическом великолепии. Но здесь всегда возникает вопрос – а так ли жаль? Решая вопрос в сторону «кесарю кесарево», большинство книг по прикладной статистике отказываются от теоретического изложения в пользу конкретных методов по типу «делай то – получишь это». Сколь бы ни казалась жёсткость такого подхода редукцией, в данном пособии избран тот же самый метод.

Но избрав сам метод, можно, однако, получить достаточно широкую дисперсию его воплощения. Оказывается, даже делая акцент на представлении обработки данных в стиле «кулинарной книги», можно сформировать материал таким образом, что использовать его всё же не

[1] П. Фейерабенд «Наука в свободном обществе»

представится никакой возможности. Размышляя над этим, я всегда вспоминаю одну знаменитую книгу по математической статистике для психологов, которая, при всём блеске изложения, не даёт навыков работы со статистическими программами + не даёт представления об обработке данных как технологии – только лишь как о наборе критериев, несмотря на то, что некоторые из них просто не существуют друг без друга в реальной практике использования. И поскольку представилась возможность вспомнить о непосредственном воплощении обработки данных в статистических продуктах, стоит сказать, что экономическая литература со столь практической направленностью встречается гораздо чаще. С одной стороны, эта ситуация не представляется трагичной – социально-экономические исследования во многом схожи с психологическими в плане обработки данных, но есть совершенно уникальные стороны, которые почерпнуть из такой литературы зачастую не удаётся.

Чтобы поддержать студента в самых простых, но тем не менее, обязательных аспектах обработки данных, делается не только обзор статистических программ, которые могут быть применены, но и ещё более простых с точки зрения более опытного исследователя вещей, а именно – комментариев к установке, вводу данных в тех или иных случаях и т.п.

Несмотря на всю популярность SPSS и STATISTICA, конкретные примеры этого пособия даны также в тех программах, которые распространяются свободно.

1. Выбор метода обработки данных.

1.1. Зачем нужна статистика и с чего начинается обработка данных.

Чтобы по-настоящему понять, зачем нужна статистика и математика в психологии, придётся с ней ознакомиться, как то ни странно. В общих же чертах статистика может ответить на такие вопросы психолога, как то: - отличаются ли две группы по n - параметру; - есть ли связь между психологическими проявлениями; - есть ли различие в экспериментальных воздействиях; - существует ли какой-то скрытый фактор за наблюдаемыми паттернами; - как стоит строить классификацию n – явлений и т.д.

Для начинающего психолога подобные вопросы кажутся неестественными, поскольку задумываясь над каким-либо вопросом, он формулирует совершенно «иноязычный» ряд предположений. И мне бы хотелось, чтобы этот «иноязычный», но живой взгляд любопытства навсегда остался непосредственной частью исследовательской работы. Для того чтобы это было именно так, необходимо понять один-единственный принцип, удачно сформулированный Эйнштейном: «началом каждой физической теории являются мысли и идеи, а не формулы». Чтобы понять всю глубину этого высказывания и причины исключительно такого взаимодействия между мыслью и формулой, необходимо начать изучение математического аппарата с самого начала, а именно с интуитивного понятия числа, что и будет сделано в главе о шкалах измерения. Избегать же стоит другого принципа, характерного для некоторых исследователей гуманитарного направления, а именно - нельзя верить в то, что «цифры не лгут». Переход от эмпирических отношений к системам иного языка (например, числовым) угрожает любой прикладной науке ложью допустимых преобразований. Чтобы согласиться со мной или же аргументировано опровергнуть такую точку зрения опять же стоит обратиться к главе о шкалах измерения.

Математическую статистику для психолога можно преподавать совершенно различными методами. С одной стороны, можно преподавать необходимые знания в виде «кулинарный книги» - рассказать, каким критерием когда пользоваться. С другой стороны, можно начать с теории вероятностей или даже с математического анализа, но тогда книга по психологии превратится в слабое подобие труда математика. Чтобы учесть плюсы, но избежать минусов, остановимся на «кулинарной книге», но такую книгу мы будем делать, вслед за советами Орлова А.И., не просто отрывочным списком различных математических процедур, но в виде статистических технологий, где будет учтена не проверка отдельных гипотез, но общий контекст, в котором проверка не ограничивается каким-либо единственным методом. Чтобы осуществить эту цель, необходимо начать изложение с базовых знаний, которые категорически необходимы в процессе обработки данных.

Итак, для обработки данных необходимо начать со следующего (предполагается, что исследование уже проведено по канонам признанной парадигмы, либо же в осознанном отклонении от общего стандарта):

- подготовить таблицу сырых баллов;

- выдвинуть статистические гипотезы, которые согласуются с общей исследовательской гипотезой;

- проверить статистические гипотезы и сделать соответствующие выводы.

Под выдвижением статистической гипотезы подразумевается вопрос, который сформулирован в согласии с логикой исследования и математической статистикой в целом. Вопрос, не соответствующий исследованию в целом, напоминает действие сумасшедшего в определении Ломброзо, а именно бесцелен по своей природе. Вопрос, сформулированный не на языке статистики, в процессе обработки данных столь же бесполезен, сколь бесполезен вопрос китайца к африканцу. Поэтому оба условия выдвижения статистической гипотезы необходимо учитывать.

Однако, если вопрос, выдвинутый в общей логике исследования, может быть и очевиден, то сформулировать такой вопрос на языке статистики можно только в случае знакомства с некоторым набором статистических критериев. Под статистическим критерием стоит понимать метод проверки статистической гипотезы. Не обладая должным математическим образованием, психолог может мыслить даже в обратном направлении (мне кажется, что такая ситуации встречается даже чаще) – критерий + интуитивное ощущение определяют гипотезу. Такая странная ситуация, как определение гипотезы после метода, что в рамках психологической науки не должно встречаться часто, в случае, когда мы выходим из психологии в прикладную статистику, оправдывает себя. К тому же психологу не под силу проверить такую статистическую гипотезу, которую не может проверить заранее известный статистический критерий, а значит в подавляющем количестве случаев этот диалектический процесс гипотезы и критерия решается в сторону критерия.

Итак, поскольку мы выяснили, что обработка данных исследования напрямую зависит от выбора нужного критерия, нам необходимо сформулировать несколько общих правил, которые помогут сделать этот выбор быстрее и правильнее.

Статистических критериев огромное множество и вряд ли есть человек, который долгое время в состоянии помнить абсолютно все из них, причём с областями применения. Даже в книгах по математической статистике не принято описывать все критерии. Но для правильного выбора критерия стоит двигаться не от «возможности», а от «невозможности».

Апофатический выбор состоит из следующих вопросов:

(0). Описательные статистики.

1. Нормально ли распределение данных и однородна ли дисперсия?

2. Каковы шкалы измерения?

3. Сколько выборок и переменных?

4. Каковы эти выборки и переменные?

5. Специфический вопрос.

Естественно, что эти вопросы можно задавать в любом порядке, но есть определённая логика именно в таком порядке.

Пункт (0) означает, что предшествовать выбору критерия может представление данных в виде различных показателей, отвечающих на вопросы о том, насколько велик разброс данных и т.п.

Нормальность распределения данных, что описано в соответствующей главе, не только позволяет использовать соответствующие критерии, которые называются параметрическими, но и помогает ответить на вопрос, можно ли считать шкалу измерения метрической. В соответствующем параграфе будет также освещено использование нормального распределения в психодиагностике.

Под вопросом 4 подразумевается зависимость/независимость выборок и переменных.

Под вопросом 5 подразумевается тот особенный вопрос, который возникает уже после разрешения предыдущих 4 и носит более специфический характер, свойственный самому исследованию.

Последовательно решая каждый из поставленных вопросов, мы сможем понять, какие статистические критерии применяются в тех или иных случаях, а далее нам лишь останется утвердить свои знания непосредственным применением конкретных критериев в статистических программах.

1.2. Ввод данных в Excel.

Работа над статистической обработкой прежде всего, как и следует предположить, начинается с ввода данных. Придавать сырым данным электронный вид удобнее всего в Excel, поскольку впоследствии может понадобиться, например, перейти от программы к программе или передать данные ещё кому-то. Если же сырые данные будут введены в какую-либо программу, то далеко не всегда удаётся экспортировать информацию напрямую в другую программу, а, значит, придётся снова пользоваться Excel в качестве пересадочного пункта. Что касается передачи информации другому исследователю, то стоит помнить, что статистические программы зачастую создают файлы в совершенно специфическом для себя формате, а, значит, без соответствующей программы на компьютере, прочитать такие файлы нельзя.

Исходя из приведённой логики, мы откроем именно Excel и начнём ввод данных (Пуск -> Все программы -> Microsoft Office -> Microsoft Office Excel, если пакет установлен, что крайне вероятно).

Предположим, что у нас есть опросник из 20 вопросов, измеряющий склонность человека совершать необдуманные поступки. Первые 10 вопросов, согласно идее разработчиков, составляют шкалу НП-1, т.е. необдуманные поступки в привычных условиях. Остальные 10 вопросов составляют шкалу НП-2, т.е. необдуманные поступки в непредвиденных обстоятельствах. Соответственно, НП-3 – это сумма НП-1 и НП-2, которая отражает склонность совершать необдуманные поступки в любых жизненных ситуациях. Стоит сделать уточнение: вовсе не обязательно, что шкала суммарного балла вообще присутствует в опроснике – её может и не быть. В некоторых случаях вместо суммарного балла используются обобщающие факторы, выявленные разработчиками эмпирическим путём и т.д. Таким образом, за отсутствием шкалы суммарного балла не стоит стремиться получить таковую необдуманно самостоятельными методами – стоит разобраться с причиной её отсутствия.

В зависимости от задачи исследования массив данных можно вводить, как минимум, двумя кардинально различными способами.

1. Для большинства случаев данные вводятся по шкалам.

Предположим, что в нашем вымышленном примере вопросы о необдуманных поступках имеют как прямую, так и обратную форму.

Например:

- Бывали ли в вашей жизни случаи, когда вы сначала совершали некоторое действие, а только потом думали о последствиях?

Если мы имеем ответник, в котором предполагается, что нечто нужно оценить от 1 до 5, где 1 – это абсолютное несогласие, а 5 – абсолютное согласие, то в шкале необдуманности ответ «5» означал бы максимальное проявление этого качества. Соответственно, это прямой вопрос.

- Бывают ли в вашей жизни случаи, когда вы сначала тщательно обдумываете план действий, а только потом действуете?

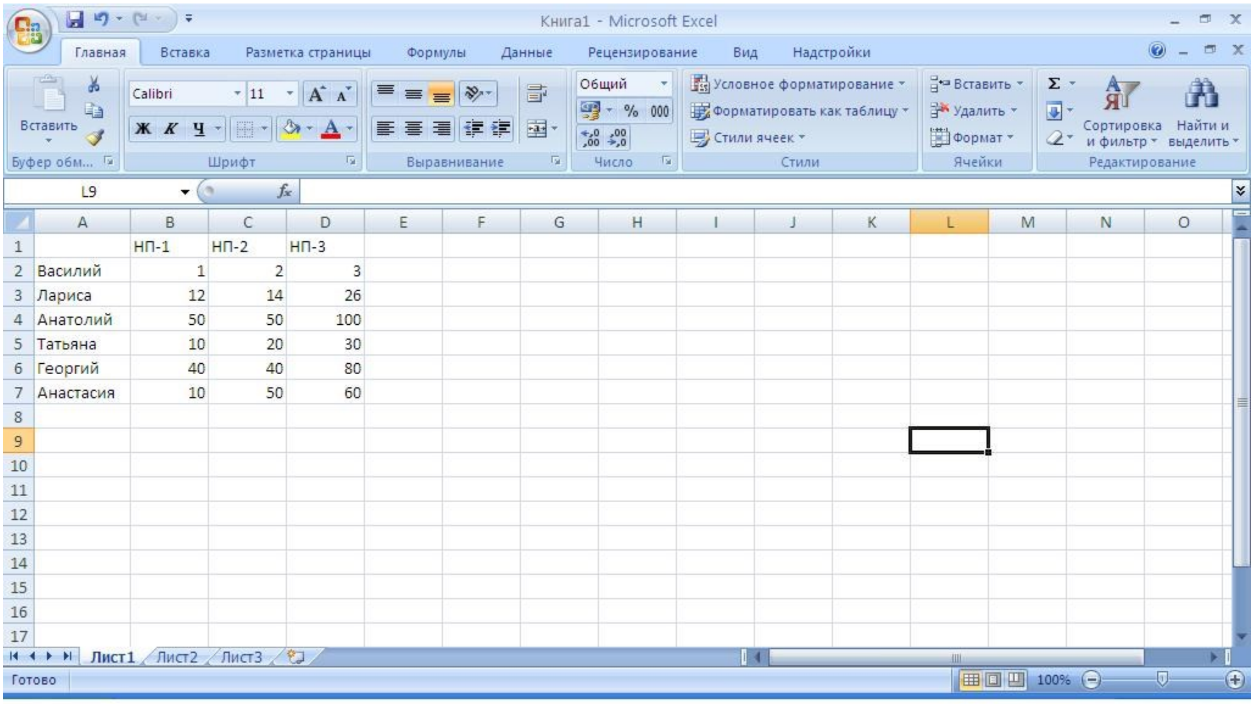
Соответственно, это обратный вопрос, потому что ответ «5» будет соответствовать максимальной обдуманности. Так как шкала у нас необдуманности, то при обработке такой первичный ответ испытуемого инвертируется. Обратные вопросы всегда помечены в ключах к тестам. Но рассмотрим общий случай, как вообще инвертируются обратные вопросы. Обычно, для этой цели при обработке обратный балл высчитывается устно, поскольку это очень легко. Если у нас была 5-балльная шкала оценки, где 1 – абсолютное несогласие, а 5 – абсолютное согласие, то легко предположить, что инвертированные баллы будут: **1** = 5; **2** = 4; **3** = 3; **4** = 2; **5** = 1. Но если возникает трудность, то можно воспользоваться формулой:

$$r = \min - s + \max,$$

где: r – отклик на прямой вопрос, min и max – минимальное и максимальное значение в шкале, s – отклик на обратный вопрос.

Соответственно, удобно бывает сразу на заполненных бланках отмечать: к какой шкале относится вопрос, а также обратные вопросы.

После этого нужно взглянуть, какие шкалы предполагаются опросником и в результате суммирования баллов вводить данные в Excel следующим образом:



Обратите внимание, что испытуемые представляют собой строки, а шкалы – столбцы. Для большинства случаев необходим именно такой способ расположения.

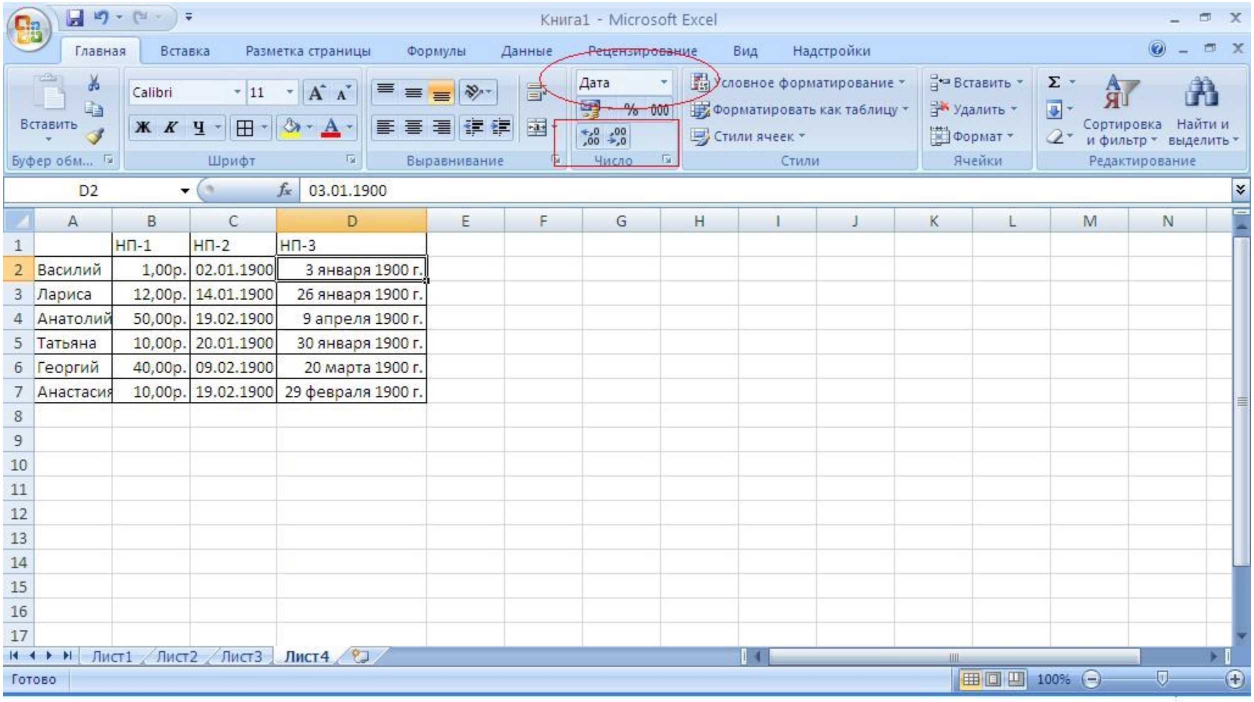
В некоторых случаях возникает несколько трудностей по непосредственному вводу конкретных значений в ячейку. Мы рассмотрим две самые распространённые трудности.

Во-первых, по неведомым причинам, данные могут приобрести следующий вид:

	НП-1	НП-2	НП-3
Василий	1,00р.	02.01.1900	3 января 1900 г.
Лариса	12,00р.	14.01.1900	26 января 1900 г.
Анатолий	50,00р.	19.02.1900	9 апреля 1900 г.
Татьяна	10,00р.	20.01.1900	30 января 1900 г.
Георгий	40,00р.	09.02.1900	20 марта 1900 г.
Анастасия	10,00р.	19.02.1900	29 февраля 1900 г.

Это означает, что стоит неверный формат ячейки. Для исправления: правой кнопкой мыши -> формат ячеек -> нужно сменить на общий. Либо из главного меню, ниже будет PrintScreen.

В случае, если вводятся дробные значения показателей, Excel может их округлять. Чтобы этого не происходило, нужно увеличить количество знаков после запятой. Для этого используется либо главное меню, либо в быстром меню, вызываемом правой кнопкой мыши, ищется следующие знаки (также приведено расположение пункта меню формата ячейки):



Далее остановимся на нескольких простых приёмах упрощения жизни исследователя в случаях, когда нужно видоизменить существующую таблицу данных для конкретной задачи.

Как мы помним, НП-3 – это суммарная шкала, получающаяся из сложения НП-1 и НП-2. Мы можем считать эту шкалу устно, вручную или на калькуляторе, а можем воспользоваться помощью Excel. Для этого достаточно задать функцию суммирования: Формулы -> Вставить функцию -> Полный алфавитный перечень -> СУММ, либо в главном меню найти значок «∑», нажать на стрелочку рядом с этим значком и выбрать «Сумму».

И если этот пункт объяснения, а именно, как посчитать сумму в Excel, можно считать зачастую излишним, то со следующим шагом трудности возникают уже чаще. Если посчитать какую-либо шкалу, задав ту или иную формулу, то при необходимости копирования этой шкалы в другой столбец Excel или на другой лист (обычно перечень листов находится внизу, добавление листов может происходить в некоторых версиях через меню вставки), с лёгкостью исполнения и ужасом воплощения можно увидеть следующее:

########
########
########
########
########
########

Либо изменение значений, поскольку копирование привело к заданию той же формулы, но других аргументов.

Чтобы такого не произошло, нужно вставить не формулу, а сами значения. Делается это следующим образом: копирование идёт по стандартному образцу (выделить необходимый массив, правой кнопкой мыши вызвать меню - выбрать копировать; либо просто выделить массив и нажать CTRL и C). А вставлять данные в необходимую ячейку не CTRL и V, а правой кнопкой вызвать меню на необходимой ячейки и нажать специальную вставку, отменив в ней опцию «значения». В этом случае мы вставим значения без формулы, но именно такой ситуации мы и добивались для перемещения шкалы НП-3 в другую область Excel.

В случае, если мы находимся всё в том же методе ввода данных через уже подсчитанные шкалы, может возникнуть потребность подготовить данные к анализу различия групп. Например, в приведённом примере, половина испытуемых – мужчины, половина – женщины. Впоследствии мы научимся искать между ними различия, но для начала этого процесса готовится таблица сырых данных по определённому принципу, а именно – каждому значению шкалы в следующем столбце соответствует код группы. Например, если мы решим сравнить две группы – мужчин (1) и женщин (2) по шкале НП-1, НП-2 или НП-3, то сама таблица данных будет выглядеть следующим образом:

	НП-1		НП-2		НП-3	
Василий	1	1	2	1	3	1
Лариса	12	2	14	2	26	2
Анатолий	50	1	50	1	100	1
Татьяна	10	2	20	2	30	2
Георгий	40	1	40	1	80	1
Анастасия	10	2	50	2	60	2

Соответственно, если групп, например, три, то коды в следующем за сырыми данными столбце, будут чередоваться от 1 до 3 и т.д.

Таким образом, вводя, например, первые два столбика (НП-1 и код) мы сможем впоследствии посчитать различие групп по этой шкале, указывая, какой столбик – значения, а какой – код группы.

В случае, когда нашим методом обработки данных является корреляционный анализ, редко, но всё же тоже встречается задача коррелирования не шкал, а испытуемых. Обычно, используя корреляционный анализ, мы размышляем над тем, связаны ли между собой психические качества. В некоторых же случаях нам необходимо узнать, является ли группа испытуемых сплочённым множеством объектов – идёт сравнение профилей испытуемых. В таком случае нам будет необходимо развернуть таблицу сырых данных прежде, чем вводить её в какую-нибудь статистическую программу. Для этого используется уже знакомая нам специальная вставка, но указывать нужно опцию транспонирования. В таком случае таблица данных приобретёт следующий вид:

	Василий	Лариса	Анатолий	Татьяна	Георгий	Анастасия
НП-1	1	12	50	10	40	10
НП-2	2	14	50	20	40	50
НП-3	3	26	100	30	80	60

В случае, когда мы разворачиваем таблицу данных для корреляционного анализа самих испытуемых, стоит помнить, что количество шкал должно быть достаточно большим – 15 и более, поскольку, в ином случае все относительно небольшие значения корреляции будут не значимыми. Это касается, естественно, и случая стандартного анализа, когда желательное количество испытуемых не менее 15, но в случае разворота таблицы можно случайно забыть о том, что теперь нужно заботиться о количестве шкал, а не испытуемых.

2. Другой способ ввода данных предполагает в качестве столбцов не шкалы, а вопросы. Этот способ требуется реже. Главное применение – анализ надёжности самого теста.

Естественно, что в случае, когда мы вводим ответы испытуемых на каждый вопрос, нам также необходимо учесть обратные баллы, как это было в предыдущем случае.

Если бы у нас был опросник из пяти вопросов, то таблица данных могла бы выглядеть следующим образом:

	1	2	3	4	5
Татьяна	1	2	1	1	5
Мария	1	4	5	2	4
Ирина	1	5	3	2	1
Лариса	4	4	5	5	4

Соответственно, в этом случае рассуждение зачастую идёт не об испытуемых, не о шкалах, а о соотношении вопросов. Использовать же такой метод, например, для сравнения испытуемых между собой, не представляется возможным, поскольку психологическое измерение носит неточный характер, а значит, делать вывод, исходя из ответа на каждый вопрос в отдельности, зачастую нельзя.

1.3. Установка программного обеспечения.

В данном параграфе мы обсудим не совокупность возможных статистических программ, которыми может пользоваться психолог-исследователь в своей работе, но только лишь те две программы, которые понадобятся нам на протяжении обучения приёмам обработки данных. О том, какие программы могут понадобиться при обработке данных вообще, в чём их отличие, официальные сайты и полезные ссылки – всё это находится во второй части пособия – перечень программ для статистической обработки и дополнительный статистический инструментарий. Естественно, что использование AtteStat и Rundom Pro в своей дальнейшей деятельности – вовсе не обязательное условие успеха, но эти программы являются легально бесплатными, в связи с чем объяснение методов подсчёта в данном пособии ведётся именно с помощью этих продуктов.

<u>AtteStat:</u>

Для установки программы AtteStat необходимо пройти по ссылке: http://attestatsoft.narod.ru/download.htm

Великолепнейшим плюсом, помимо прочих, естественно, является абсолютно русскоязычное меню программы.

Для установки достаточно выбрать 32 или 64 битную версию. В случае, если вы не знаете, какая необходима вам, то можно выяснить это чисто эмпирическим путём – скачать обе.

Установка проходит без лишних трудностей, в связи с чем переходим к рассмотрению уже установленного продукта.

У программы AtteStat есть поразительная хитрость, которая далеко не сразу становится очевидной. Есть два способа запустить программу: в уже открытом файле Excel (через меню надстройки), либо через меню пуск, но в таком случае всплывает окно разрешения запуска Excel. Окно запроса Excel всплывает из-за того, что программа AtteStat является «не самостоятельной» программой, а программой, которая «надстраивается» над Excel. Хитрость же заключается в том, что, несмотря на интуитивную аналогичность способов запуска, результат может поразить. Если мы запустим программу в заранее открытом Excel, у нас может оказаться гораздо меньшее количество модулей программы, чем есть на самом деле. В некоторых случаях, когда запуск делался, например, для дисперсионного анализа, заранее открытый Excel не будет содержать этого способа анализа. Самый простой способ избежать недостающих модулей – запускать AtteStat, разрешать запуск Excel, а уже в открытом Excel открывать интересующую таблицу данных.

Чтобы понять, какие модули анализа в AtteStat точно есть, привожу список:

1. XTAB – кросстабуляция;

2. TRFG – преобразование данных;

3. TIME – анализ временных рядов;

4. SURVIVE – анализ выживаемости;

5. SQC – контроль качества;

6. RNZ – рандомизация;

7. PS – параметрическая статистика;

8. PRT – распознавание образов;

9. OP – обработка выбросов;

10. NP – непараметрическая статистика;

11. NDC – проверка надёжности;

12. MDP – обработка пропущенных данных;

13. MDS – многомерное шкалирование;

14. KNOW – экспертные оценки;

15. IA – информационный анализ;

16. FAA – факторный анализ;

17. EXACT – точные критерии;

18. DS – описательная статистика;

19. CORA – корреляционный анализ;

20. CLA – кластерный анализ;

21. AV – дисперсионный анализ;

22. APX – аппроксимация зависимостей.

Если при запуске программы какого-либо из этих модулей нет, то что-то пошло не так – скорее всего сам запуск, как было описано выше.

Поскольку AtteStat работает непосредственно с данными, введёнными в Excel, то дополнительные объяснения по вводу данных в эту программу излишни. Конкретные рекомендации будут даны уже по использованию тех или иных модулей.

Rundom Pro (3.14):

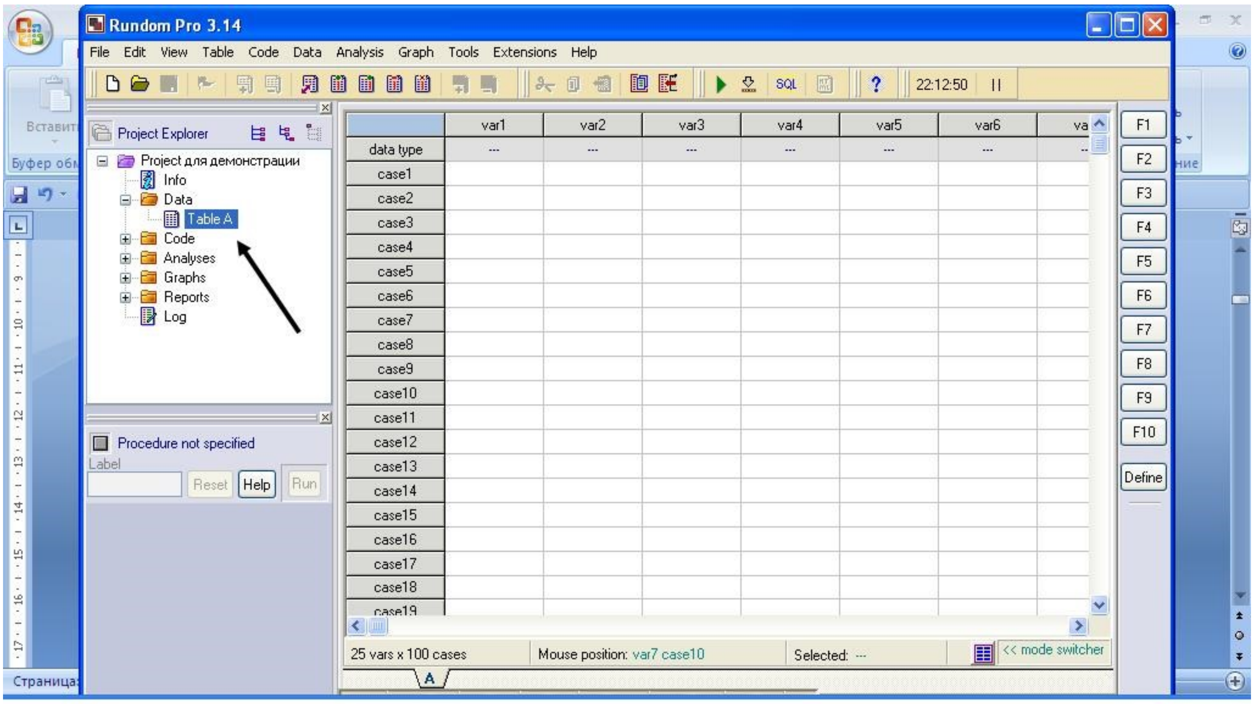
Скачать программу можно с сайта: http://pjadw.tripod.com/index.htm, только при открытии этого сайта могут открываться дополнительные ссылки – не стоит бояться – вирусов нет, просто это какие не слишком удачные нововведения, которых раньше не было, что оставляет надежду на то, что вскоре они исчезнут. К тому же, если возникает непредвиденная проблема из серии: сайт переехал, можно попытаться скачать эту программу с сайта бесплатных статистических программ: http://freestatistics.altervista.org/en/stat.php.

Установка дистрибутива обычно не вызывает никаких трудностей, поэтому переходим к рассмотрению уже установленной программы.

При запуске программы с рабочего стола или через меню пуск, первым появляется меню, где предлагается просто начать, создать новый проект и т.д.

Создадим новый проект (start new project), зададим ему имя в открывшемся окне.

Слева у нас появится меню самого проекта, нам нужно выбрать: Data -> Table A.



В данную таблицу можно перенести уже подготовленную в Excel таблицу сырых данных (без названия шкал и имён испытуемых). Для этого необходимо скопировать из Excel данные, но вставлять их через меню: Edit -> Paste Data (Active Table), в ином случае (через правую кнопку мыши или CTRL и V) все скопированные данные вставятся в одну-единственную ячейку, на которой стоит курсор.

После проведения того или иного подсчёта крайне необходимо нажимать на сохранение, поскольку при выходе из программы простым «крестиком» данные не будут сохранены.

Соответственно, если нас интересуют отчёты о результатах статистических критериев, то их можно увидеть либо в выплывшем отчёте, либо, если мы, например, делали несколько проверок и решили вновь к ним вернуться, в меню слева в папке Analyses.

Если мы делали несколько графиков и хотим просто повторно на них взглянуть, то они будут находиться в меню слева в папке Graphs.

Сам проект сохраняется, если не была выбрана иная папка для установки программы, в: Мой компьютер -> C -> RundomPro314 -> Projects.

Последним комментарием общего характера может быть следующее: Rundom Pro очень чувствителен к пунктуации, т.е. при необходимости задать название переменных, код, процентильные листы и т.д. могут возникать трудности, объясняющиеся нехваткой запятой и т.п. Поэтому в случае, когда программа выдаёт ошибки, стоит попытаться изменить пунктуацию методом перебора или при помощи меню Help, поскольку зачастую трудности ввода касаются именно этих простых вещей. Рассмотренные в пособии методы будут содержать пояснение к пунктуации.

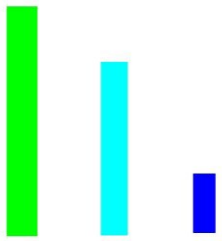
1.4. Описательные статистики, представление данных.

К вопросу о статистическом представлении данных можно подходить двумя путями: через дескриптивную и индуктивную статистику. Эти подходы дополняют друг друга примерно в том же отношении, что и система «человек – машина».

Под индуктивной статистикой понимаются выводы, которые можно получить при помощи методов, делающих обобщения на генеральную совокупность. Статистические критерии, с которыми мы познакомимся при использовании статистических программ, как то: критерий Манна-Уитни, Спирмена, Стьюдента, Пирсона и т.д., - всё это примеры индуктивной статистики.

Под дескриптивной же или описательной статистикой понимаются графики и формулы подсчёта, отвечающие на вопросы, касающиеся непосредственно результатов выборочного исследования. Дескриптивную статистику выгодно отличает наглядность, но представление данных, выраженное на «высоком языке» такой статистики требует предварительной подготовки. Однако, хотелось бы, чтобы под наглядностью данных читатель не понимал исключительно тот вариант, который разделяется противниками статистической обработки данных. Я говорю о «пирогах» и «небоскрёбах», которыми некоторые исследователи стремятся заменить аргументированный и доказательный путь. Простое внесение в график процента испытуемых, ответивших на опросник «так-то» и «так-то» не заменяет не только индуктивную статистику, но даже и описательные меры, такие, как меры центральной тенденции и рассеяния.

Чтобы окончательно и бесповоротно искоренить традиции, которые доверяют лишь своим глазам, рассмотрим пример. Предположим, что мы высчитали общий балл по опроснику тревожности, в котором было три уровня представленности признака – высокий, средний, низкий. В традиции, которая не даст нам ничего, а что ещё хуже, поведёт по ложному следу, мы должны будем наглядно представить данные, например, в виде «небоскрёбов».



Первый столбик отражает процент людей, набравших высокие баллы по шкале тревожности, второй столбик – средние и т.п.

Если подходить к обработке данных с должным уровнем скепсиса, то мы не можем с уверенностью утверждать, что первый столбик отличается от второго, потому что даже если они «на глаз» кажутся разной высоты, есть дополнительные факторы, которые могут эту разницу делать более и менее значимой. Так, например, достаточно важно, смотрим ли мы различие в 5 баллов между 75 и 70 или между 20 и 15. К тому же, полезно знать, насколько разнятся между собой данные внутри группы. Если, например, в первом столбике у нас возможны данные от 150 до 200 баллов и при этом все испытуемые этой группы набрали 165 баллов, а в другой группе у нас данные, среди которых возможны баллы от 100 до 150, но все данные в 140 баллов, то это один случай сравнения. С другой стороны, у нас могут быть представлены данные, которые представляют все возможные значения высокого уровня тревожности и среднего уровня тревожности. Сравнивая случай более-менее схожих внутри групп данных и тех данных, которые внутри группы не гомогенны, мы с лёгкостью можем понять, что гомогенные внутри групп данные гораздо более вероятно различны между собой. Никакие из этих вопросов не отражены в наших процентных столбиках. И единственная причина их популярности заключается в том, что они понятны любому прохожему «церковно-приходской школы».

Далее мы рассмотрим статистические показатели и такие диаграммы, которые без предварительного ознакомления с принципами своего построения, мало понятны пешеходу. Это, конечно, наносит удар по их «простой» наглядности, но с течением времени приносит свои плоды, поскольку даёт гораздо больше ПОЛЕЗНОЙ информации, нежели обманчивая простота. Однако, испытывая некоторые трудности с восприятием графиков, диаграмм и таблиц, я проявляю к ним некоторую враждебность, которую обосную рациональностью. На мой взгляд, использовать описательные статистики гораздо более полезно не до анализа, а после того, как доказана значимость результатов сравнения, т.е. после использования индуктивной статистики.

Второй способ использования описательных статистик пронизывает основы построения психодиагностики. Такие показатели, как процентили, квартили, квантили, дисперсия и стандартное отклонение позволяют оперировать тестовыми нормами и категорически необходимы при работе над эмпирическим исследованием, что мотивирует на изучение данного параграфа как тех, кто любит графическое представление данных, так и тех, кто его избегает. Описание стоит начать именно с этих способов подсчёта, поскольку графическое представление данных в некоторых случаях строится непосредственно на них.

1.4.1. Разнообразие описательных статистик.

Описательные статистики в самом распространённом понимании это такие показатели, подсчёт которых говорит некоторую информацию о распределении данных. В таком случае первый вопрос – что такое распределение данных?

Распределением называется частота встречаемости определённой градации признака. Например, всем известный нормальный закон распределения интеллекта, говорит о том, что в обществе встречаются в основном люди со средним интеллектом, есть небольшая прослойка очень умных и небольшая прослойка очень не умных. Конечно, частота встречаемости выраженности признака не есть истинное определение распределения, которым пользуются математики, но определения через интеграл и прочее вряд ли может помочь психологу.

Распределений огромное количество, но, как и сказано в главе о нормальном распределении, для нас гораздо более полезна удивительная типология «нормальное – ненормальное распределение». Потому что как только мы понимаем, что распределение не нормально, мы переходим в область исследований, ничего не говорящую о распределениях, т.е.е в непараметрическую статистику.

Рассматривая сколько каких данных встречается в нашем исследовании, мы выдвигаем ряд вопросов, которые и будут являться описательными статистиками. Таких вопросов огромное количество, поэтому нужно чётко осознавать, что приведённый ряд описательных статистик не является исчерпывающим.

Принято делить статистические показатели описательной статистики на две большие группы: меры центральной тенденции и меры рассеяния. К этому распространённому делению имеет смысл добавить ещё два класса: квантили и показатели формы распределения.

Меры центральной тенденции стремятся обобщить данные, чтобы был виден средний (типичный) уровень выраженности признака, в связи с чем такие меры иногда называют мерами среднего уровня.

Меры рассеяния (изменчивости, вариации) стремятся охарактеризовать разнообразие встречаемых данных.

Показатели формы распределения характеризуют особенности группировки данных на всём отрезке возможных значений. Суть этих показателей, как и квантилей, будет окончательно ясна далее.

1.4.2. Меры центральной тенденции.

1) Среднее (оценка среднего, среднее арифметическое, выборочное среднее, M) – всем знакомая величина, которая приобретает смысл или теряет таковой в зависимости от изменчивости данных. Так, если мы посчитаем среднее арифметическое для данных, которые находятся относительно близко друг к другу, то среднее арифметическое будет наполнено огромным смыслом, касающимся того значения, которое более характерно для совокупности данных. Если же мы посчитаем среднее арифметическое для весьма изменчивого ряда данных, т.е. такого, где минимальное и максимальное значение находятся на большом расстоянии друг от друга + баллы носят «рваный» характер, то среднее арифметическое свой смысл потеряет, несмотря на формальную правомерность его подсчёта. Так, известная шутка: «в больнице у одних пациентов температура зашкаливает, а другие в морге – в больнице температура 36,6» – это как раз характерный пример неправомерности использования среднего арифметического при большой изменчивости. Из трёх подруг центральной тенденции – моды, медианы и среднего арифметического, последняя является самой чувствительной к изменчивым данным, в то время как на моду изменчивость может до некоторой степени никак не повлиять.

$M = (x_1 + …. + x_n) / N$, где $(x_1 + … + x_n)$ – это сумма всех представленных значений в ряде данных, а N – это количество членов этого ряда.

Остался без рассмотрения лишь один аспект. Среднее арифметическое ещё называют выборочным средним. Что это значит? Это значит, что посчитано среднее для выборки, т.е. для тех данных, которые мы получили эмпирическим путём. Но конечная цель статистики сделать вывод не о выборочной совокупности данных, например, не о школьниках 3-го класса «В» такой-то школы, а о школьниках 3-го класса вообще. Тогда 3 «В» будет выборкой, поскольку мы именно на них проводили исследование, а 3-и классы вообще – это генеральная совокупность. Для генеральной совокупности также может быть подсчитано среднее, но уже по другой формуле и просто «среднее» его никто не называет.

Среднее арифметическое можно считать только в метрических шкалах. Это стоит запомнить, поскольку такой достаточно некорректный способ, как выведение среднего балла школьной успеваемости, например, не является редкостью, хотя и не является корректным.

2) Мода. Название этого показателя говорит само за себя – это самое частое значение в ряду данных. Так, например, если у нас есть ряд чисел: 3, 10, 10, 10, 4, то модой будет 10. К использованию моды чаще всего прибегают в случаях использования таких шкал, для которых считать среднее арифметическое или другие меры среднего попросту не имеет смысла. Например, в номинальных шкалах, о которых мы будем подробнее говорить далее, использование моды – это единственный корректный способ отражения мер центральной тенденции.

Несмотря на всю простоту подсчёта моды, могут возникнуть некоторые трудности, во избежание которых дам уточняющие примеры.

Если мы имеем ряд: 1, 2, 3, 5, то моды нет.

Если мы имеем ряд: 1, 3, 5, 5, 6, 6, 7, 8, то модой должна быть либо «5», либо «6». Одним из способов, которым с этим можно бороться, является ответ 5,5, т.е. среднее значение двух соседних мод.

Если мы имеем ряд: 1, 3, 3, 5, 6, 6, 7, 8, то модами будет «3» и «6» и находить среднее между ними уже нельзя. В таком случае распределение объявляют бимодальным и указывают обе моды. Если мод больше, то распределение становится полимодальным.

Если мы имеем дело с данными, сгруппированными в интервалы, то моду рассчитать можно исходя из модального интервала.

Рассмотрим пример. Предположим, что мы (зачем – не знаю), создали анкету, где выясняли, сколько рабочих получают тот или иной уровень заработной платы. Соответственно, мы имеем примерно следующую картину:

зар. плата	10-100	101-200	201-300	301-500
кол-во человек	3	5	10	7

«Наш завод» в Ленинграде получил такую зар. плату, что модальным интервалом являются значения «201-300». Естественно, что нам бы хотелось знать не интервал, а более узкую меру. Такой мерой может быть и мода. Для этого используется следующая формула[2]:

$$Mo = x_0 + h * \frac{f_m - f_{m-1}}{(f_m - f_{m-1}) + (f_m - f_{m+1})},$$

где:

Mo – мода,

x_0 – нижняя граница модального интервала,

h – величина интервала,

f_m – частота модального интервала,

f_{m-1} – частота интервала, предшествующего модальному,

f_{m+1} – частот интервала, следующего за модальным.

Тогда для нашего примера мода будет равна:

$$Mo = 201 + 99 * \frac{10 - 5}{(10 - 5) + (10 - 7)} = 262{,}875$$

3) Медиана. Этот показатель является наиболее адекватным для порядковых шкал. Медиана – это средний по счёту член упорядоченного вариационного ряда. Например, если мы имеем данные: 2, 10, 5, 8, то нам нужно выстроить возрастающий или убывающий ряд: 2, 5, 8, 10 и медианой для такого ряда будет число 6,5, поскольку медиана находится между 5 и 8 {(5+8)/2}. Если, соответственно, у нас был бы ряд: 2, 5, 8, 10, 15, то медианой

[2] Формула взята с сайта: http://www.grandars.ru/student/statistika/strukturnye-srednie-velichiny.html

стало бы число 8, поскольку ряд нечётный. Таким образом, медиана делит ряд данных пополам – 50% значений собранных данных находятся ниже медианы, 50% - выше.

Естественно, что медиана, как и среднее арифметическое, чувствительна к изменчивости, хотя и в меньшей мере. Так, если мы будем строить ряд: 1, 1, 1, 1000, 1000, то медианой у нас будет число 1 (опять же, потому что ряд нечётный), хотя интуитивно понятно, что мера, отвечающая за среднее (как бы это среднее не высчитывалось) в данном случае не отражает «всей правды». Но, повторюсь, в процессе исследования может случиться такая ситуация, когда без медианы обойтись будет сложно, а среднее арифметическое в меру имеющихся данных, применять не адекватно. Хотя, подолью масла в огонь, сказать, что в рассмотренном ряду из единиц и тысяч среднее арифметическое в 400,6 также мало отражает истину, как и медиана.

4) Разговор о самых распространённых мерах центральной тенденции заканчивается на вышеперечисленных трёх показателях, однако это не значит, что остальные бесполезны.

Даже средних существует огромное количество, что, впрочем, не удивительно, исходя из того, какое широкое определение было ему дано академиком О. Коши в XIX веке: «Средней величиной является любая функция f (X_1, X_2, …, X_n) такая, что при всех возможных значениях аргументов значение этой функции не меньше, чем минимальное из чисел X_1, X_2, …, X_n, и не больше, чем максимальное из этих чисел»[3].

Поэтому в этом пункте я перейду к нераспространённым среди психологов мерам среднего. Мы уже успели рассмотреть связь среднего арифметического, медианы и моды – трёх показателей, которые наиболее часто встречаются в эмпирической психологии. Но среднее арифметическое, если не обращать внимания на его популярность, имеет гораздо более тесную связь со степенными средними, нежели с медианой и модой, являясь степенным средним первой степени. Редкость использования остальных представительниц степенного среднего объясняется достаточно легко – не виной психологов, естественно, а редкостью шкал отношений в нашей науке, для которых степенные средние и применимы, в связи с чем формулы степенных средних только и находятся на страницах словаря Бурлачука Л.Ф. Но прогресс не стоит на месте, и я искренне надеюсь в ближайшие десять лет

[3] Цит, по Орлов А.И. «Эконометрика»

увидеть такие шкалы психологических качеств, в связи с чем ещё раз привожу эти формулы.

Итак, среднее геометрическое, применяемое для шкал отношений, имеет вид[4]:

$$A_0 = g = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n},$$

где $x_1 \ldots x_n$ – члены изучаемого ряда.

Среднее гармоническое:

$$A_{-1} = h = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_n}}$$

Среднее квадратическое:

$$A_2 = s = \sqrt{\frac{x_1^2 + x_2^2 + \ldots + x_n^2}{n}}$$

На этом список степенных средних не заканчивается, поскольку степеней может быть столько, сколько угодно исследователю, мы лишь рассмотрели самые известные – (-1), 0, 1 (среднее арифметическое), 2.

1.4.3. Меры изменчивости.

1) Размах – разность между максимальным и минимальным значением зафиксированного признака.

$R = x_{max} - x_{min}$

2) Дисперсия – это такой показатель, который показывает отклонение значений от среднего арифметического в квадратах единицы измерения. Под квадратом единицы измерения подразумевается та же ситуация, которая получается при измерении площади фигуры – сама фигура измерена, например, в метрах, а площадь уже в м^2. Именно потому, что это зачастую не представляется удобным, на суд общественности представляются показатели изменчивости, выраженные не дисперсией, а стандартным отклонением,

---

[4] Формулы степенных средних взяты с сайта: http://ru.wikipedia.org/wiki/%D0%A1%D1%80%D0%B5%D0%B4%D0%BD%D0%B5%D0%B5_%D1%81%D1%82%D0%B5%D0%BF%D0%B5%D0%BD%D0%BD%D0%BE%D0%B5

которое будет обсуждено далее и выражает меру, эквивалентную единице измерения.

Высчитывается по формуле:

$$\sigma^2 = \frac{\sum(x_i - \bar{x})^2}{n-1},$$

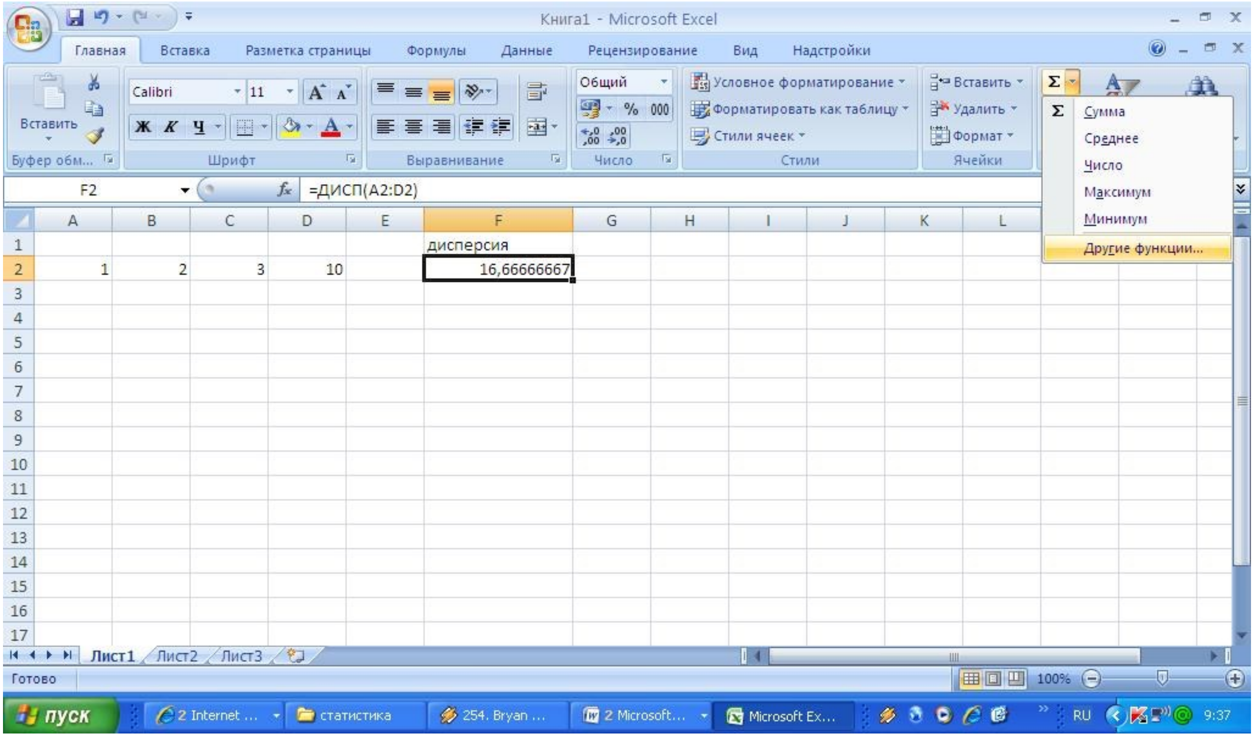
где:

x_i – i-ый элемент ряда,

$\bar{x}$ – среднее арифметическое,

n – количество элементов.

В Excel, как и в статистических программах, можно посчитать выборочную дисперсию, не затрачивая особых усилий. Для этого нужно нажать на меню различных формул, которое находится в Office 2007 здесь:



В открывшемся меню различных формул выбрать, например, полный алфавитный перечень, где найти функцию ДИСП, которая и высчитывает дисперсию для выборки.

В окне подсчёта дисперсии нужно задать весь интересующий массив данных, нажав на стрелочку в строчке «Число 1», и ОК. Стоит иметь в виду, что массив данных, для которых имеет смысл считать дисперсию, целиком и полностью зависит от целей исследователя. Например, может понадобиться дисперсия для шкалы опросника, но начинающий исследователь может сделать ошибку там, где сделать её обидно. Чтобы рассмотреть эту ошибку подробнее, рассмотрим пример. Предположим, что у нас 4 испытуемых, которые отвечали на 10 вопросов опросника «любителей яблок», где каждый возможный ответ предполагает всего два значения – 1 и 0. Итак, чем выше суммарный балл, тем больше человек любит яблоки. Но почему я позволила себе употребить высокомерное «начинающий исследователь»? Только лишь потому, что для тех, кто первый раз делает сводную таблицу данных, характерна особенность не жалеть своих сил и делать таблицу данных следующим образом:

вопрос	1	2	3	4	5	6	7	8	9	10
Василий	1	0	0	0	0	0	0	0	0	0
Пётр	1	1	0	0	0	0	0	0	0	0
Катерина	1	1	1	0	0	0	0	0	0	0
Тереза	1	1	1	1	1	1	1	1	1	1

Т.е. фиксировать ответ на каждый вопрос вместо того, чтобы вводить уже суммарные баллы по шкале. В целом, это даже хорошо, потому что при анализе качества теста зачастую требуется именно такое представление данных. Но для корреляционных исследований, например, это лишняя информация, которую всё равно придётся преобразовать.

Нам этот пример нужен затем, чтобы понять, как важно правильно задать массив при подсчёте дисперсии.

Если мы зададим массивом все эти значения на каждый вопрос, то выборочная дисперсия будет равна примерно 0,246. Если зададим массивом данных уже подсчитанный суммарный балл каждого испытуемого, а именно ряд данных: 1, 2, 3, 10, то дисперсия будет равна 16,66666...7.

Как видно из примера, дисперсия будет разительно отличаться и задать массив суммарного балла не есть эквивалентное действие массиву всех входящих в этот суммарный балл значений. Эта игра массивов, которая сводит с ума в первые попытки подсчёта, встречается при подсчёте надёжности в Excel, но это уже выходит за рамки разговора об описательных статистиках.

Если вести подсчёт в англоязычном меню Excel, то обозначением дисперсии будет VAR (от variance).

3) Стандартное отклонение (среднеквадратическое отклонение, стандартный разброс) показывает отклонение от среднего арифметического уже в тех же единицах, в которых измерено само качество.

$$\sigma = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$$

Стандартное отклонение – важнейшее понятие в психодиагностике.

Например, среднестатистическая норма и коридор нормы теста, у которого есть стандартизованные баллы (IQ, T, стэны, станайны и др.), даётся именно на языке стандартное отклонения, а именно – 2/3 такового. Нормой шкалы IQ является 100 баллов, а коридор нормы (2/3σ) ±10. Т.е. люди, чей интеллект составляет от 90 до 110 баллов, считаются среднестатистической нормой.

4) Если мы, например, решим представить наглядное сравнение интеллекта школьников двух классов дисперсией или стандартным отклонением, то никаких очевидных проблем, кроме этических, перед нами возникнуть не должно. Но что делать, если единицы измерения двух сравниваемых групп различны? Мы помним, что и дисперсия, и стандартное отклонение так или иначе напрямую зависят от единицы измерения. Эта особенность делает такие показатели абсолютными.

Для целей представления сравнения признаков с разной размерностью можно пользоваться несколькими относительными показателями рассеяния: коэффициентом вариации, линейным коэффициентом вариации, коэффициентом осцилляции, относительным квартильным расстоянием.

А. Коэффициент вариации:

$$V = \frac{\sigma}{\bar{x}} 100,$$

где:

σ - стандартное отклонение,

$\bar{x}$ - среднее арифметическое.

Б. Линейный коэффициент вариации:

$$V_d = \frac{\bar{d}}{\bar{x}} 100,$$

где:

$\bar{d}$ - среднее абсолютное отклонение, равное:

$$\bar{d} = \frac{\sum|x_i - \bar{x}|}{n}$$

$\bar{x}$ - среднее арифметическое.

В. Коэффициент осцилляции:

$$V_R = \frac{R}{\bar{x}} 100,$$

где:

R - размах вариации,

$\bar{x}$ - среднее арифметическое.

Г. Относительное квартильное расстояние:

$$d = \frac{q}{\bar{x}},$$

где:

q – среднее квартильное расстояние, равное средней величины разности между квартилями:

$$q = \frac{Q_3 - Q_1}{2}$$

$\bar{x}$ - среднее арифметическое.

О том, что такое квартиль, мы поговорим в следующем пункте – квантильных показателях, также относящихся к описательным статистикам.

1.4.4. Квантили.

1.4.4.1. Квантили и процентили.

1.4.4.1. Квантили и процентили

Квантиль – это общее название для показателей, которые делят распределение или совокупность данных исследования на n равных частей. Статистическая точка зрения рассматривает квантиль как такое число, что случайная величина (изучаемое качество) не превышает его с определённой вероятностью. Т.е., например, нас интересует, сколько нужно набрать баллов, чтобы сделать тест лучше, чем 50 % изученных людей – мы ставим вероятность 0,5 (для примера - когда мы знаем, что событие обязательно случится, мы задаём вероятность 1) и смотрим, каково же это число, такое, что вероятность равна 0,5. Подход через вероятности довольно не очевиден, поэтому гораздо более удобен другой подход, через который зачастую и идёт объяснение квантилей.

Как уже было сказано, квантили – это общее название для целого ряда показателей, отсекающих ту или иную часть совокупности данных. Синонимом квантилей являются процентили, хотя по традиции за квантилями всё же оставляют общий смысл, в то время как процентили уже считаются конкретным случаем квантилей. Но следует понимать, что от квантилей до процентилей один шаг терминологического разнообразия, в то время как остальные показатели буквально имеют определение через процентили.

Процентиль имеет очень большое сходство с процентом, но взятым наоборот. Если мы говорим, что результат теста Василия – 95-ая процентиль, то это значит, что 95 % людей из его группы (или группы стандартизации теста в зависимости от того, берём ли мы собственноручно сделанные нормы или нормы, данные в руководстве теста) сделали этот тест как он и хуже, чем он. Стоит понимать, что под процентом здесь всегда имеются в виду люди, а не сам тест. Сказать, что Василий выполнил 95 % теста – категорически не правильно. Грубейшая ошибка, которую совершать нельзя.

Итак, процентиль можно определять не только через вероятность, но и через саму выборку, что гораздо более понятно. Процентиль – это такой показатель, который делит ранжированную выборку на 100 равных интервалов и выдаёт значение границы интервала в единицах сырых баллов изучаемой совокупности. Я умышленно отступаю не только от определения через вероятность, но и от определения через среднее интервала, поскольку далее мы узнаем, что несмотря на простоту самой идеи процентиля, разные статистические программы дают совершенно разный вывод о границе таких интервалов.

Первое, что мы сделаем – это убедимся, что разные статистические пакеты считают процентили совершенно разными способами.

Предположим, что у нас в выборке 12 человек, которые делали тест на любовь к апельсинам. Максимум баллов, которые можно набрать по этому тесту – 100 баллов, минимум соответственно 0. Данные выглядят следующим образом:

	апельсины	№	процентиль
Василий	5	1	
Катерина	7	2	
Лариса	8	3	
Татьяна	9	4	25
Пётр	25	5	
Николай	29	6	
Надежда	30	7	медиана
Марина	45	8	
Фёдор	80	9	
Полина	90	10	75
Клара	91	11	
Дарья	95	12	

В первой колонке у нас имена испытуемых, во второй – сырые баллы по тесту, в третьей – простой порядковый номер испытуемых.

Этот же порядковый номер по совместительству является ранговым рядом. Т.е когда мы давали определение процентили (процентиль – это ж.р. – «она моя»), то под ранжированной выборкой имели в виду совокупность данных, выстроенную по возрастанию сырых баллов. Затем этот возрастающий ряд сырых баллов переводится в ранговый ряд путём простого присвоения номера. Из 12 получить 100 интервалов, которые нам нужны для процентилей, можно простым путём пересчёта, но мы поставим себе несколько иную цель – мы будем искать не все процентили, а только три из них – 25-ую, 50-ую и 75-ую.

Чтобы это сделать, мы разделим совокупность данных на 4 равные части, что и отмечено цветом. Ведь 12/4 получается три, в связи с чем каждые три ранга обозначены цветом. В последней колонке отмечено, между какими ячейками находится значение 25, 50 и 75 процентили.

Теперь нам нужно понять, действительно ли: 25 процентиль будет находиться между сырыми баллами 8 и 9, медиана (50 процентиль) между 29 и 30 сырыми баллами, а 75 – между 80 и 90 баллами.

Чтобы посчитать k-ую процентиль в Excel можно задавать формулу вручную, но мы воспользуемся более удобным методом - встроенной функцией: ПЕРСЕНТИЛЬ (получить к ней доступ можно либо из главного меню, нажав на стрелочку рядом со знаком «∑» -> Другие функции -> Полный алфавитный перечень -> Персентиль; либо через вкладку Формулы -> Вставить формулу -> Категория полный алфавитный перечень -> Персентиль).

После того, как перед нами открылось диалоговое окно функции Персентиль, необходимо задать всего два параметра: массив данных (которым будет являться вся совокупность данных по шкале любви к апельсинам) и необходимую процентиль. В Excel процентиль задаётся на языке вероятностей, т.е. через десятичные значения 0,25; 0,5; 0,75 и т.п. Такая ситуация может варьироваться в зависимости от программы – так, например, в SPSS (по крайней мере 17 версии) процентили нужно задавать через процентные величины, т.е. 25, 50, 75. Отчасти неудобным аспектом является тот факт, что можно посчитать при этом и 0,25 процентиль, что будет означать отнюдь не 25 процентиль. Во избежание таких недоразумений стоит осознанно подходить к обработке и при освоении новой для себя программы стремиться к проверке правильности подсчёта, сколь бы элементарным он ни казался на определённом этапе. Исходя из описанной ситуации мне более приятен вариант измерения процентилей в пределе от 0 до 1, поскольку в таком пределе программа сразу же выдаст ошибку, если задать слишком большое значение.

Результаты подсчёта в разных статистических программах разными способами таковы:

процентиль	Excel $	SPSS (1) *	SPSS (2) &	STATISTICA (1) &	STATISTICA (2) *	STATISTICA (3) $
25	8,75	8,25	8,5	8,5	8,25	8,75
50	29,5	29,5	29,5	29,5	29,5	29,5
75	82,5	87,5	85	85	87,5	82,5

На самом деле способов посчитать процентиль как минимум шесть, но давайте взглянем на приведённую таблицу. Цветом я отметила середину своих подсчётов, отходя на клетку в сторону от которой можно увидеть абсолютно одинаковые подсчёты. Таким образом, зеркальность этой таблицы показывает, что подсчёт ведётся в разных программах, но одинаковым способом.

Весьма разнообразные варианты подсчёта процентилей предлагаются в программе STATISTICA, поэтому мы возьмём именно ту классификацию, которая предложена в руководстве к этой программе. Однако русскоязычного варианта этой классификации я не встречала, в связи с чем буду приводить оригинальное название, поскольку на мой перевод надеяться ни в коем случае не следует.

1. Средневзвешенное величины X_{np} (Weighted Average at X_{np})[5].

Нам нужно определиться с тем, что же такое средневзвешенное и величина X_{np} прежде, чем переходить к самому правилу подсчёта процентилей этим методом.

Среднее арифметическое взвешенное (weighted average) – это всеми любимое среднее арифметическое, взятое с учётом веса. Например, если мы будем покупать кроликов не в одну партию, а в три, то может случиться такая ситуация, что цена кроликов будет меняться от партии к партии. Предположим, что в первую партию мы купили 20 кроликов по 30 рублей, во вторую партию мы купили 35 кроликов по 40 рублей, а в третью – 30 кроликов по 45 рублей за штуку. Тогда среднее арифметическое взвешенное цены равно:

$$\bar{x} = \frac{\sum \omega_i * x_i}{\sum \omega_i} = \frac{20 * 30 + 35 * 40 + 30 * 45}{20 + 35 + 30} \approx 39{,}41 \text{ рублей}$$

Обратите внимание, что при вычислении среднего арифметического взвешенного можно перепутать, что именно из двух параметров является весами. Мы вычисляли средневзвешенную цену, в связи с чем количество кроликов выступало весом.

Под X_{np} стоит понимать значение процентили, выраженное в единицах сырых баллов. Когда мы рассматриваем результат Василия в 35 сырых баллов, соответствующие 95-ой процентили по тесту интеллекта в рамках группы из 25 человек, то это и будет означать X (35 баллов), n – 25 человек, p – 0,95 процентиль. С тем же успехом мы можем высчитать достижения Катерины из той же группы и получить, например, 0,85 процентиль, что может быть следующим: X (28 баллов), n – 25 человек (группа ведь та же самая), p – 0,85.

---

[5] С методами подсчёта можно ознакомиться в меню программы STATISTICA – достаточно ввести в справку фразу «Percentile Calculation Options»

Теперь нам осталось понять, в чём же заключается подсчёт процентили методом средневзвешенного значения. Описание каждого из методов имеет стандартный вид, который я не буду менять в силу того, что методов становится всё больше и больше, а значит гораздо полезнее научиться читать этот стандартный образец, нежели запоминать сами методы – для их подсчёта достаточно обратиться к статистическим программам.

Стандартный вид средневзвешенного метода имеет следующие параметры:

$np = i + f$,

где n – количество наблюдений, p – k - процентиль, i – целая часть np, f – дробная часть np.

Значение процентили = $(1 - f)*x_i + f*x_{i+1}$

Пример: предположим, что мы хотим узнать, каково значение 25-ой процентили в ряду всё тех же любителей апельсинов, описанных выше:

5	X_1
7	X_2
8	X_3
9	X_4
25	X_5
29	X_6
30	X_7
45	X_8
80	X_9
90	X_{10}
91	X_{11}
95	X_{12}

Тогда $np = 12 * 0,25 = 3 = i + f = 3 + 0,00$

25-ая процентиль = $(1 - 0)*x_3 + 0*x_{3+1} = 8$

В случае, если в формуле появляется величина x_0, то она заменяется на x_1. Тогда формула подсчёта приобретает вид: $(1 - f)*\mathbf{x_1} + f*x_{(0+1)}$.

Например, высчитаем первую процентиль для нашего теста любителей апельсинов.

$np = 12*0,01 = 0,12$. Тогда 1-ая процентиль = $(1 - 0,12)*\mathbf{x_1} + 0,12*x_{0+1} = 5$. И до 8-ой процентили в нашем тесте значение будет 5.

2. Средневзвешенное величины $X_{(n+1)p}$ (Weighted Average at $X_{(n+1)p}$).

$(n+1)*p = i + f$,

где n – количество наблюдений, p – k - процентиль, i – целая часть np, f – дробная часть np.

Значение процентили: $(1 – f)*x_i + f*x_{i+1}$

Пример подсчёта 25-ой процентили из теста любителей апельсинов:

$(n+1)*p = (12 + 1)*0,25 = 3,25 = i + f = 3 + 0,25$

25-ая процентиль $= (1 – 0,25)*x_3 + 0,25*x_{3+1} = (1 – 0,25)*8 + 0,25*9 = 8,25$

В таблице подсчёта процентилей, которую мы считали разными статистическими программами, этот метод уже встречался. Им пользуется SPSS, если стоит опция подсчёта процентилей по умолчанию.

В случае, если в формуле появляется значение x_{n+1}, то вместо него подставляется x_n. Тогда формула подсчёта процентили приобретает вид: $(1 – f)*x_n +f\mathbf{x_n}$.

Например, в нашем тесте любителей апельсинов значение x_{n+1}, т.е. x_{12+1} может потребоваться для подсчёта процентилей, начиная с 93.

Тогда подсчёт выглядит следующим образом: $(n + 1)*p = (12 + 1)*0,93 = 12,09$

93 процентиль $= (1 – f)*x_i + f*x_{i+1} = (1 – 0,09)*x_{12} + 0,09*\mathbf{x_{12}} = 95$. Более того, все оставшиеся процентили нашего теста также будут равны 95.

3. Эмпирическая функция распределения (Empirical Distribution Function).

Отклоняюсь от перевода: функция эмпирического распределения, только потому что именно так на русском языке звучит этот термин.

Это самый очевидный метод, но и самый ненадёжный, на мой взгляд. Значение процентили задаётся границей интервала, выраженной в непосредственных сырых баллах. Что значит в непосредственных? Например, в предыдущем методе мы получили такое значение процентили, которое не соответствует ни одному действительно полученному в процессе исследования значению. В случае же, когда эмпирическое распределение остаётся неприкосновенным, в качестве значения процентили мы увидим сырой балл какого-нибудь испытуемого, который ближе всего стоит к искомому интервалу. С одной стороны, это, конечно, «правда в лицо». С другой, стоит прекрасно понимать, что значения, полученные психологическими тестами, носят довольно флуктуирующий характер, т.е. имеют свойство меняться как от одного тестирования до другого, так и от выборки к выборке. Это обстоятельство наводит меня на мысль, что пользоваться данным методом в психологических исследованиях рискованно.

$np = i + f$,

где n – количество наблюдений, p – k - процентиль, i – целая часть np, f – дробная часть np.

Если $f = 0$, то значение процентили = x_i

Если $f > 0$, то значение процентили = x_{i+1}

Пример подсчёта 25 процентили в тесте любителей апельсинов этим методом:

$np = 12*0{,}25 = 3 = i + f = 3 + 0{,}00$

Т.к. $f = 0$, то 25 процентиль = $x_3 = 8$.

Если бы мы получили хотя бы небольшое дробное значение в np, то нам бы пришлось использовать в качестве 25 процентили следующее число, т.е. 9.

4. Эмпирическая функция распределения с усреднением (Empirical Distribution Function with Averaging).

Этот метод является модификацией предыдущего с той лишь разницей, что сделана попытка учесть тот самый аргумент о флуктуации данных.

$np = i + f$,

где n – количество наблюдений, p – k - процентиль, i – целая часть np, f – дробная часть np.

Если $f = 0$, то значение процентили = $0{,}5\,(x_i + x_{i+1})$

Если $f > 0$, то значение процентили = $x_i + 1$.

Пример 25-ой процентили из теста апельсинов:

$np = 12*0,25 = 3 = i + f = 3 + 0,00$

Т.к. $f = 0$, то 25 процентиль = $0,5\ (x_3 + x_{3+1}) = 0,5\ (8 + 9) = 8,5$

Таким образом, мы видим, что этот метод заключается в усреднении сырых баллов, непосредственно касающихся интервала процентили.

Этот метод используется в STATISTICA (по крайней мере, в 6.00 версии), если опция подсчёта процентилей задана по умолчанию, а также в довольно часто встречающейся в наших примерах программе Rundom Pro 3.14.

5. Эмпирическая функция распределения с интерполяцией (Empirical Distribution Function with Interpolation).

Все слова, фигурирующие в названии для нас уже знакомы, кроме, пожалуй, интерполяции. Под интерполяцией стоит понимать группу процедур, позволяющих сделать вывод о недостающих промежуточных значениях. Не стоит предполагать, что интерполяция – это исключительно математический термин, вернее было бы оставить межнаучный статус. Так, например, мы с лёгкостью можем представить себе задачку для теста интеллекта, которая бы содержала интерполяцию. Например, предположите, какого значения не хватает в ряду чисел:

2, 4, …, 8, 10. Нахождение числа 6 было бы чистейшей задачей интерполяции. Однако, вернёмся к методам подсчёта процентилей.

$(n-1)*p = i + f$,

где n – количество наблюдений, p – k - процентиль, i – целая часть np, f – дробная часть np.

Если $f = 0$, то значение процентили = x_{i+1}

Если $f > 0$, то значение процентили = $x_{i+1} + f*(x_{i+2} - x_{i+1})$.

Пример подсчёта 25 процентили в тесте любителей апельсинов:

$(n-1)*p = (12 - 1)*0,25 = 2,75 = i + f = 2 + 0,75$

Т.к. $f > 0$, то 25 процентиль = $x_{2+1} + 0,75*(x_{2+2} - x_{2+1}) = 8 + 0,75*(9 - 8) = 8,75$.

Этот метод подсчёта процентилей встречается не только в STATISTICA, но и является встроенным по умолчанию алгоритмом функции

ПЕРСЕНТИЛЬ в Excel. Но всё же сделаю уточнение: в Excel может встречаться и другой алгоритм подсчёта, приближённо напоминающий гибрид эмпирического распределения без усреднения и с усреднением. Следовательно, логично предположить, что в более новых версиях, нежели Excel 2007, может встретиться отличный от описанного алгоритм, вследствие чего читатель может встретиться с расхождением собственноручно выполненного расчёта и приведённого, чего бояться не следует. Бояться стоит только ошибки, которую можно избежать простой ручной проверкой на этапе освоения программы.

6. Ближайшее наблюдение – наблюдение, наиболее близкое к величине np (Closest Observation – Observation closest to np).

Мы помним, что величина np показывает порядковый номер испытуемого, который находится на границе интервала искомой процентили. В зависимости от метода подсчёта мы выберем либо сырой балл этого испытуемого, либо какое-то значение поблизости.

«Наблюдение», которое фигурирует в названии метода, означает, что значение процентили будет выражено в эмпирически полученном сыром балле, т.е., взглянув на таблицу данных, мы обязательно найдём испытуемого, который набрал ровно то количество баллов, которое значится той или иной процентилью.

В таком случае, метод ближайшего наблюдения является сходным по задумке методом с одним из тех, который мы уже рассматривали, а именно – с методом, строящемся на эмпирическом распределении. Как мы помним, в методе эмпирического распределения процентиль обязательно имеет такое значение, которое встречалось в исследуемой выборке. Тогда нам остаётся понять лишь то условие, которое в некоторых случаях будет дифференцировать результаты, полученные этими методами. Чтобы говорить более основательно, рассмотрим суть метода ближайшего наблюдения.

$np + 0{,}5 = i + f$,

где n – количество наблюдений, p – k - процентиль, i – целая часть np, f – дробная часть np.

Значение процентили = x_i

Пример 25 процентили теста из теста любителей апельсинов:

$np + 0{,}5 = 12*0{,}25 + 0{,}5 = 3{,}5 = i + f = 3 + 0{,}50$

25 процентиль = x_3 = 8.

Поскольку мы рассмотрели все 6 методов, описанные в STATISTICA, то можем позволить себе сравнение этих методов с целью выяснения, чем же отличается метод эмпирического распределения от метода ближайшего наблюдения.

Рассчитаем массу процентилей для теста любителей апельсинов тремя методами: ближайшего наблюдения, эмпирического распределения и эмпирического распределения с интерполяцией. В свою очередь последний метод, реализованный в Excel необходим нам для контраста, чтобы видеть резко отличный от сырых баллов метод.

сырой балл	ранг	persentile	np	clos. observ.	emp. distr.	emp.distr. w. interp.
5	1	1	0,12	5	5	5,22
7	2	3	0,36	5	5	5,66
8	3	5	0,6	5	5	6,1
9	4	7	0,84	5	5	6,54
25	5	25	3	8	8	8,75
29	6	50	6	29	29	29,5
30	7	61	7,32	30	45	40,65
45	8	75	9	80	80	82,5
80	9	81	9,72	90	90	89,1
90	10	89	10,68	91	91	90,79
91	11	99	11,88	95	95	94,56
95	12					

Цветом отделены первые две колонки, чтобы не возникало соблазна сравнить их со следующими. А также цветом показана единственная процентиль из приведённых в списке, которая имеет разное значение при подсчёте двумя интересующими нас методами, а именно методом эмпирического распределения и ближайшего наблюдения. Исходя из этого различия мы уже может сделать вывод, что эти методы отличны друг от друга, но не так кардинально, как от метода интерполяции.

Ключом к разгадке является значение np. Если 0 > f < 0,50 в величине np (обратите внимание – именно в величине np, а не в величине np + 0,5), эти два метода будут давать различные выводы. Во всех остальных случаях результаты будут идентичными. Это правило различия помогает понять, как же работает метод, в котором заявлен поиск ближайшего к np значения. Вспомним школьную математику. Округление дроби идёт в сторону уменьшения значения в случае, если дробное значение формулируется через цифры, меньшие 5. В то время как дроби с цифрами от 5 до 9 при округлении дают увеличение значения числа. В таком случае, метод ближайшего наблюдения, созданный для поиска эмпирически зафиксированного значения, близкого к величине np, принимает совершенно логичные решения как в случае np = 7,32, выдавая значение 7-го члена ряда, так и в случае, когда np = 9,72, выдавая значение 10-го члена ряда.

Стоит, однако, обратить внимание на то, что при вычислении 1 и 3 процентили методом ближайшего наблюдения, STATISTICA сделала отступление от формального вида формулы, заменив X_0 на X_1. Этому не стоит удивляться, ведь мы уже встречали такое правило подстановки в методе №1 – средневзвешенном величины X_{np}.

1.4.4.2. Спорные и сложные вопросы о процентилях

1.4.4.2. Спорные и сложные вопросы о процентилях

1. Бывает ли нулевая и сотая процентиль? Это не столь тривиальный вопрос, как может показаться на первый взгляд. Давайте для начала проведём эксперимент, а именно – попробуем побудить статистическую программу эти процентили посчитать.

STATISTICA 6.0 вне зависимости от того, какой из 6 предложенных методов (переключаемых в меню: Инструменты -> Опции -> Вычисление) мы выберем, предлагает нам пересмотреть своё поведение и выдаёт вместо 0 и 100 процентили 10 и 90. Естественно, что в этой программе можно посчитать процентили от 1 до 99, так что отчаиваться не стоит.

SPSS 17 встроенным по умолчанию методом, а именно – средневзвешенным величины $X_{(n+1)p}$ (№2 в приведённой выше классификации) – откажется считать 0 процентиль, но за 100 без зазрения совести выдаст самое высокое эмпирически зафиксированное значение, а именно - любителя апельсинов с 95 баллами. Если сменить алгоритм подсчёта на эмпирическое распределение с усреднением, то картина изменится (для этого, если, например, вести подсчёт через Analyze -> Descriptive Statistics -> Frequencies, то необходимо войти в находящийся справа пункт Statistics, задать интересующие процентили и поставить галочку на Values are group midpoints). Задав другой алгоритм мы увидим, что ни 0, ни 100 процентиль SPSS считать не будет, поскольку не известны нижняя (в случае 0 процентили) и верхняя (в случае 100 процентили) границы интервалов.

Rundom Pro 3.14, пользующаяся эмпирическим распределением с усреднением откажется считать 0 и 100 процентиль, что роднит её по духу с STATISTICA и вторым способом подсчёта в SPSS.

Excel, использующий эмпирическое распределение с интерполяцией, посчитает и 0 (приняв за неё первое эмпирически обнаруженное значение), и 100 (последнее эмпирически обнаруженное значение) процентиль.

Из описания приключения по подсчётам 0 и 100 процентилей следует замечательный вывод: даже используя одни и те же методы подсчёта процентилей статистические программы могут давать разнящиеся между собой выводы, что особенно характерно для нетривиальных ситуаций из серии описанной выше.

В таком случае, отвечая на вопрос о том, имеет ли смысл считать 0 и 100 процентиль, и есть ли они вообще, стоит аргументировать свой ответ исключительно логически, не ссылаясь на ту или иную формулу подсчёта. Самый распространённый ответ на этот вопрос: нет, 0 и 100 процентили нет. Этот ответ коренится на самом понятии процентили как заданной определённым методом границы интервала. Ведь как только мы говорим об интервалах, нам нужно представлять, где эти интервалы начинаются и заканчиваются. Получить такую информацию можно в конечном множестве элементов, каждый из которых нам известен. В случае разговора о выборочном исследовании мы априорно исходим из того, что конечным множеством эмпирически мы не располагаем – только лишь экстраполируем выводы с 3 «В» класса на всех третьеклассников мира. С другой стороны, когда наше исследование носит тот редкий характер, а именно – не делает выводов на генеральную совокупность, принимая за выборку и генеральную совокупность одни и те же элементы, то может иметь смысл в случае необходимости поразмышлять на тему наличия нулевой и сотой процентили. Но даже в этом категорически редком случае поставленный вопрос может привести в тупик, поскольку есть не менее спорный вопрос, который касается строгости определения термина процентиль.

2. Стандартное определение процентили через выборку звучит примерно так: процентиль – это процент людей, набравших такое же или меньшее количество баллов. В некоторых случаях я встречала определение, которое накладывало строгость, т.е. включало в себя исключительно процент людей,

имеющих меньшее количество баллов, но это определение не является верным, если не используются дополнительные операции, отсекающие ряд одинаковых значений. Ведь недостаточно выдать желаемое за действительное, а в тех 6 методах, которые мы рассматривали выше, приёмов, отсекающих одинаковые значения, не предусмотрено, что приводит нас к нестрогому определению.

Рассмотрим абсурдный пример. Предположим, что все испытуемые набрали 100 баллов: 100, 100, 100, 100, 100.

Тогда все процентили будут иметь значение 100, вне зависимости от метода подсчёта. Это не является упущением математиков – кто же знал, что встретится сумасшедший психолог, который проведёт подобное исследование. И самое, пожалуй, парадоксальное в первый момент рассмотрения такого результата подсчёта от 1 до 99 процентилей, это тот забавный факт, что «100», которое значится как сырой балл каждой процентили, на самом деле разные «100», просто в меру того, что числа одинаковые, это не очевидно. Такой смешной пример помогает нам понять две вещи: во-первых, давать определение процентили через строгое определение не имеет смысла, поскольку не используются методы, дающие действительное право на такую строгость; во-вторых, даже если мы имеет своей целью сделать выводы исключительно о самой выборке, не распространяя выводы на другие группы (принимаем выборку и генеральную совокупность за одно и то же множество элементов), нахождение 0 и 100 процентили не представляется обоснованным шагом за счёт нестрогого определения процентили и отсутствия абсолютного нуля в большинстве психологических измерений + повторяющихся значений, которые могут нам встретиться в первом и последнем интервале.

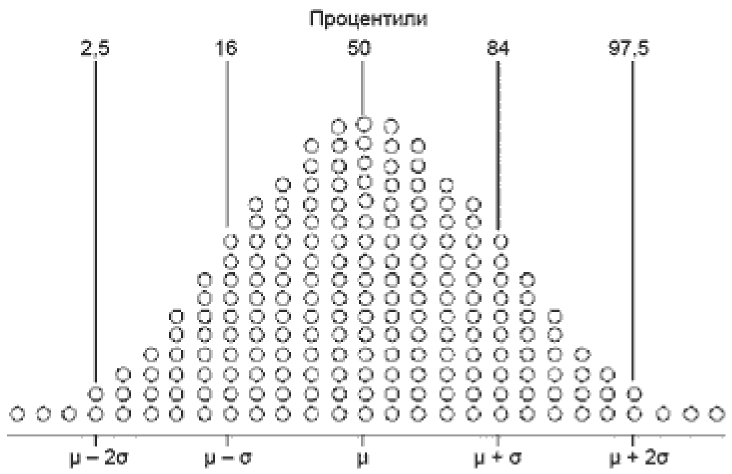
3. К вопросу о связи процентилей и стандартного отклонения.

Этот вопрос не является спорным по природе, но относится, на мой взгляд, к сложным, если столкнуться с ним в первый раз. Процентили, как мы выяснили выше, делят распределение данных на 100 равных интервалов, где засечки этих интервалов и будут искомым сырым баллом k-процентили. Но когда мы стали подробнее рассматривать эмпирическое распределение теста любителей апельсинов, то выяснили, что «равность» этих интервалов имеет свой особенный смысл – ведь и 1, и 3, и 5, и 7 процентили имеют значение «5». Где здесь равные интервалы? Исходя из нестрогости определения мы получаем ситуацию, в которой одинаковые значения сырых баллов испытуемых могут находиться на одном уровне: 5, 5, 5. Следующие за ними члены ряда могут находиться очень далеко от предыдущих: 15, 16, 18 и т.п. Но процентиль – это ранговая процедура, которая пренебрегает расстоянием между членами ряда, учитывая лишь их порядок. Тогда получается, что имея перед собой 99 испытуемых, мы сможем распределить их порядковый номер равномерно по процентилям от 1 до 99, но это вовсе не значит, что и сырые баллы будут распределяться таким же образом. Сырые баллы, в единицах которых идут значения процентилей, могут, и более того, распределяются в основном неравномерно, поскольку если мы представим, например, тест интеллекта из 100 баллов, то в среднестатистической выборке мы вряд ли встретим одинаковое количество людей, которые бы набрали 5, 10 и 50 баллов из 100. Основная часть наберёт 50 и выше баллов. Такая особенность роднит процентили со стандартным отклонением. Рассмотрим близость идей этих показателей при условии, что мы имеем интервальную и выше шкалу измерения + знаем вид распределения. Например, рассмотрим процентили стандартного нормального распределения:

Вероятность,%	99,99	99,9	99	97,72	97,5	95	90	84,13	50
Квантиль	3,715	3,09	2,326	2	1,96	1,645	1,282	1	0

В таком виде квантили стандартного нормального распределения даны в википедии[6]. Эта таблица на самом деле довольна проста, если в ней разобраться. Верхняя строка означает ту самую k-процентиль, которую мы пытались научиться считать выше. Например, значение 99 или 0,99 процентили равно 2,326. Но в таком случае возникает вопрос, откуда взялся такой сырой балл, если мы не знаем, о каком, например, тесте идёт речь – единицы измерения ведь могут самыми разными. Дело в том, что когда речь идёт о стандартном нормальном распределении, истинно сырых баллов мы уже не имеем. Значения процентили даются в стандартных баллах. Существуют специальные способы перевода сырых баллов в стандартные, вид которых может существенно отличаться. В случае же, когда речь идёт о стандартном нормальном распределении, предполагается, что среднее равно 0, а стандартное отклонение 1. В таком случае становится ясно, что 99 процентиль равна 2, 326 стандартных отклонений от среднего. Есть замечательное графическое воплощение этой идеи:

[6] http://ru.wikipedia.org/wiki/%CA%E2%E0%ED%F2%E8%EB%FC



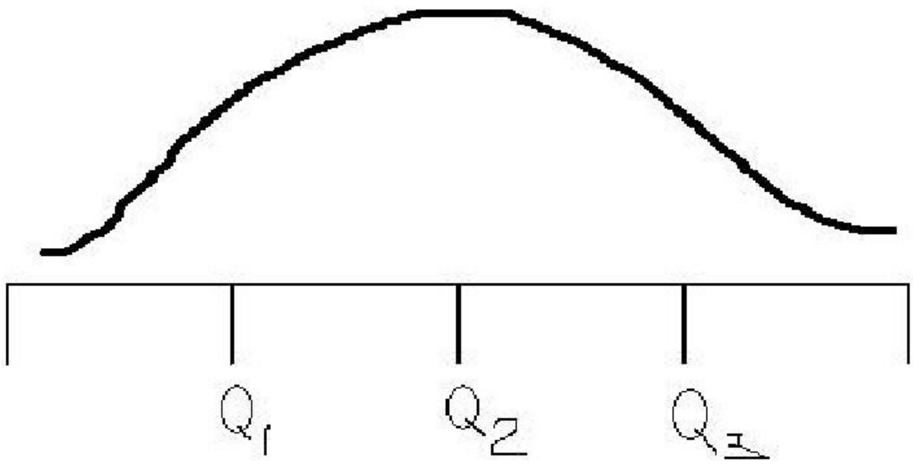
Соотношение процентилей и стандартных отклонений стандартного нормального распределения[7].

Например, мы знаем, что ±1 σ от среднего даёт коридор нормы, в который попадает 68,27 % популяции. В таком случае значения процентилей примерно с 16 до 84 дадут тот же самый коридор нормы, что и на языке стандартных отклонений, если мы имеет стандартное нормальное распределение результатов исследования. Однако же, стоит подчеркнуть, что понятие среднего и стандартного отклонения применимо далеко не всегда, а значит в случае, когда мы не уверены в законе распределения, имеет смысл использовать процентили.

1.4.4.3. Разнообразие квантилей.

Выше мы выяснили, что соотношение между квантилями и процентилями носит скорее терминологический характер традиции, нежели фактический. Однако, есть такие разновидности квантилей, которые не вызывают трудностей различения – признак, по которому идёт деление квантилей, вполне очевидно. Как мы помним, процентиль – это такая квантиль, которая делит распределение данных на 100 интервалов, но при этом стоит помнить, что процентилей всего 1-99. Если на примере процентилей было не очень удобно показать, почему это так, то на примере квартилей вполне уместно:

[7] http://www.vector-best.ru/nvb/n39/st39_3.htm



На этой картинке показано распределение данных (симуляция под функцию плотности распределения) – это не нормальный закон, хотя и похож. Слева 0 баллов, справа – например, 100. По вертикали ось, на которой откладывается количество человек, набравшее то или иное количество баллов теста. В таком случае, построив квартили этого распределения, мы можем видеть 4 равных интервала, но квартилей будет всего три. Подобно тому, как процентилей было всего 99.

Итак, квантили, которые делят распределение данных на четыре интервала, называется квартилями.

Квантили, которые делят распределение на 5 равных интервалов, называются квинтилями.

Квантили, которые делят распределение на 10 равных интервалов, называются децилями.

Сводная таблица специальных квантилей выглядит следующим образом[8]:

[8] Информация взята с сайта: http://www.machinelearning.ru/wiki/index.php?title=%D0%9A%D0%B2%D0%B0%D0%BD%D1%82%D0%B8%D0%BB%D1%8C

Процентиль	$x_{p/100}$, $p=1,...,99.$
Дециль	$x_{p/10}$, $p=1,...,9.$
Квинтиль	$x_{p/5}$, $p=1,2,3,4.$
Квартиль	$x_{p/4}$, $p=1,2,3.$

Соответственно, 1 квартиль – это 25 процентиль. 2 квартиль – это медиана или 50 процентиль. 3 квартиль – это 75 процентиль. Из этого примера хорошо видно, что квантили могут быть выражены одна через другую. Если мы, например, возьмём 1 квинтиль, то это 20 процентиль и т.д.

Квартили являются чрезвычайно полезными показателями при выведении норм теста. Если, например, нам нужно вывести уровневые нормы: плохой, средний, высокий, то достаточно использовать квартили: 1 как границу низкого уровня, от 1 до 3 – среднего, выше 3 – высокого уровня. Можно сделать и более дифференцированные нормы, даже используя те же квартили, например: ниже 1 - очень плохо, от 1 до 2 – плохо, от 2 до 3 – хорошо, выше 3 – отлично.

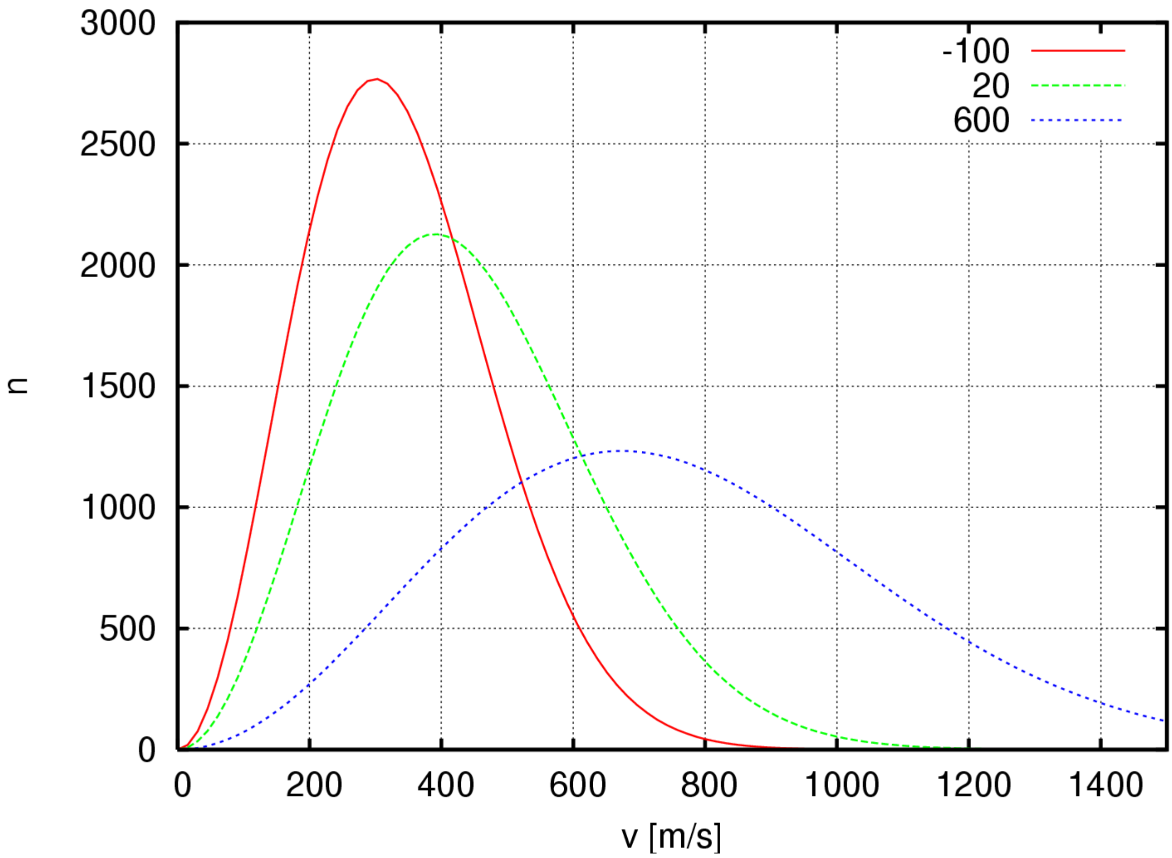
1.4.5. Показатели формы распределения данных.

Знаменитое выражение: колоколообразная кривая Гаусса, которым называют нормальное распределение данных, отражает лишь его форму, а именно – форму колокола. Когда мы графически смотрим распределение данных (подробнее с функцией плотности распределения можно познакомиться в параграфе графического представления данных – гистограмма, а также в параграфе о нормальном распределении данных), то первое, что «бросается в глаза», это тот простой факт, что кривая может быть совершенно разной формы.

Два основных вопроса, который задают по отношению к форме распределения следующие: какова асимметрия и эксцесс. Согласно распространённой традиции к описательным статистикам в основном относят меры центральной тенденции и меры рассеяния, но как квантили, так и показатели формы распределения имеет смысл относить к этой же категории статистических показателей, поскольку в конечном итоге любой показатель, который не даёт информации о генеральной совокупности, но представляет вывод о распределении выборки, представляет собой описательную статистику. И последний, абсолютно нерушимый аргумент, который роднит меры центральной тенденции, меры рассеяния, асимметрию и эксцесс – это центральные моменты распределения. Дисперсия является центральным моментом второго порядка, в то время как асимметрия высчитывается с помощью третьего, эксцесс – четвёртого. Но поскольку наша цель – не ознакомиться со статистикой как наукой, а научиться использовать её в психологических исследованиях, то мы переходим к более «ощутимым» вещам, а именно – непосредственно к подсчёту асимметрии и эксцесса.

Асимметрия. Противоположностью асимметрии является, естественно, симметрия. Когда мы говорим о симметричном распределении, то это значит, что, сложив его по линии среднего, мы получим идеально совпадающие половины. Любое отклонение одной половины от другой приводит к асимметрии. Один из самых желанных примеров симметричного распределения – это нормальное.

Рассмотрим пример асимметричного распределения, а именно – распределения Максвелла[9]:



Привожу именно этот пример, поскольку распределение Максвелла носит прикладной (правда для физики и химии), а не вымышленный

[9] Изображение взято с сайта: http://upload.wikimedia.org/wikipedia/commons/3/36/Maxwell-Boltzmann_distribution_1.png

характер. Данное распределение не является нормальным, асимметрично, но при этом нельзя сказать, что ситуации, в которых оно применимо, носят единичный случай. Отнюдь, это распределение используется для описания поведения молекул газа – оценки скорости, энергий, импульсов молекул. Данный пример призван ещё раз подчеркнуть читателю относительность понятия повсеместности нормального распределения – в психологии из-за малых выборок такое распределение скорее редкость и аномалия, нежели правило. И ещё один, пожалуй, полезный пункт. Когда мы видим на картинке распределение Максвелла и сравниваем его на память с нормальным, то мы сравниваем не функции распределения, а функции плотности распределения, потому что графически в таком виде изображены именно они. Безусловно, графически можно выразить и функцию распределения, но выглядеть она будет совершенно иначе. И даже у нормального распределения функция распределения ни в коем случае не колоколообразна, как бы мы ни привыкли не правильно думать именно в таком ключе.

Как же понять, каков коэффициент асимметрии у массива данных, состоящих из баллов шкалы тревожности n-теста? Предположим, что мы получили следующие данные:

Катерина	23
Василий	12
Татьяна	24
Регина	21
Леопольд	20
Лара	19
Николай	18
Белла	17
Венера	25
Михаил	11
Пётр	8
Вера	7
Александр	18
Кристина	14
Георгий	10

Достаточно открыть любую программу, которая предполагает подсчёт асимметрии. Англоязычная версия термина – skewness, т.е. скошенность.

Мы воспользуемся программой Rundom Pro 3.14.

Открываем программу, создаём проект, переносим числовые данные теста тревожности через функцию Edit -> Paste Date в колонку var 1.

Поскольку, как и было сказано выше, асимметрия является описательной статистикой, то запускаем меню: Analysis -> All Descriptive Statistics, и видим слева внизу меню, куда нужно задать параметры.

Это меню, естественно, позволяет высчитать и процентили, поэтому мы посчитаем: процентили, асимметрию и эксцесс, о котором речь пойдёт далее.

Variable var1

Weights пропускаем

Perc. list 0,25; 0,5.

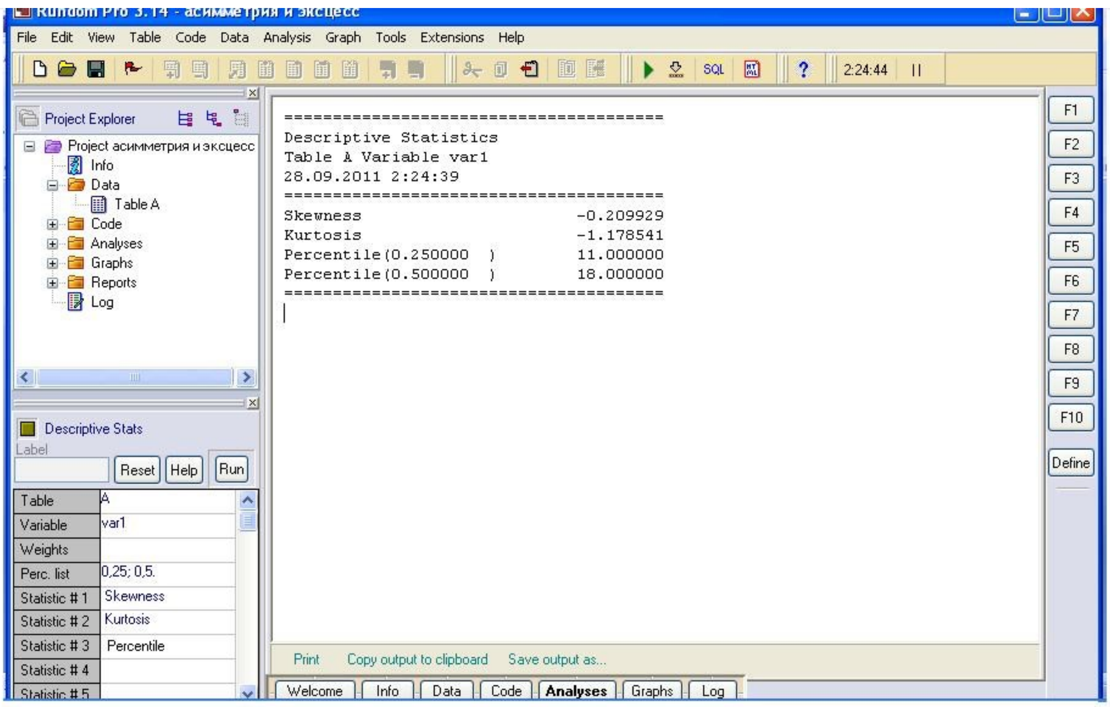
(в случае, когда мы считаем асимметрию – пропускаем; при подсчёте процентилей пишется, например, следующее: 0,25; 0,5. , причём пунктуация идёт именно через запятую дробей, точку с запятой в случае нескольких процентилей и точку в конце)

Statistic # 1 Skewness (это и есть асимметрия)

Statistic # 2 Kurtosis (посчитаем сразу и эксцесс, о котором речь пойдёт далее)

Statistic # 3 Percentile

Нажимаем Run и получаем следующую информацию:



Как же понять, хорошее ли это значение (-0,209929)?

Как мы помним, нормальное распределение симметрично, в связи с чем коэффициент асимметрии у него равен 0. Как только мы видим отклонение от этого значения, коэффициент асимметрии начинает отклоняться от 0, но в таком случае возникает вопрос – что же значит положительное и отрицательное значение? Если правый хвост распределения длиннее левого, то асимметрия будет иметь положительный знак. Если левый длиннее, то отрицательный.

Слишком асимметричные распределения могут подсказывать нам некоторую информацию, которую мы могли до этого пропустить. По крайней мере, информация об асимметрии может дать повод для размышлений. Например, мы имеем сильный скос в правую сторону в результатах контрольной работы по математике, т.е. коэффициент асимметрии имеет положительное значение. Что это может означать? Всего две альтернативы. Либо учитель математики дал слишком простые задания на контрольную, либо слишком хорошо объяснил, в связи с чем все всё поняли. Естественно, выдвигая такие альтернативы, я ни в коем случае не отрицаю возможности, что все списали у Риты с первой парты и т.п.

Размышляя об асимметрии полученных данных, полезно смотреть на графическое изображение плотности распределения (например, пользуясь гистограммой).

Эксцесс – это характеристика вершины. Если рассматривать готический собор, то его крыша будет выглядеть пиком, в то время как арабы предпочитают, видимо, купола. В программе Rundom Pro мы уже успели посчитать эксцесс (kurtosis), в связи с чем нам нужно понять, как же интерпретируется значение (-1,178541).

Для этого нужно знать, сколько же значение эксцесса при нормальном распределении. А здесь редко, но всё же возникают разночтения. Дело в том, что в формулу подсчёта эксцесса входит не только значение четвёртого момента и стандартного отклонения в четвёртой степени, но и мифическое «-3». И в зависимости от того, входит ли это «-3» в формулу подсчёта той или иной программы, эксцесс нормального распределения будет либо равен 3, либо 0. В случае Rundom Pro, AtteStat, STATISTICA и т.п. у нормального распределения 0 эксцесс.

Если хочется доподлинно понять, с какой формулой пришлось столкнуться – с той, которая высчитывает нормальному распределению эксцесс 0 или с той, которая даёт 3, необходимо ознакомиться со справкой, либо за неимением таковой, сравнить результаты подсчёта с вышеописанными программами, в которых идеальное значение – 0.

Для «нулевого идеала» формула эксцесса выглядит следующим образом:

$$kurtosis = \frac{m_4}{s^4} - 3 = \frac{m_4 - 3s^4}{s^4},$$

Где:

m_4 – центральный момент четвёртого порядка,

s^4 – стандартное отклонение в чётвёртой степени, что можно также записать как дисперсию в квадрате.

Стоит обратить внимание на то, что в разных источниках в формулу эксцесса в подсчёт центрального момента четвёртого порядка может как входить $3s^4$, так и не входить, в связи с чем у психолога совершенно точно может возникнуть недопонимание происходящего. Лично мне в этом случае понадобилась помощь Черновой Н.И., без которой я бы так и не поняла, почему формула:

$$b_2 = \frac{m_4}{s^4}$$

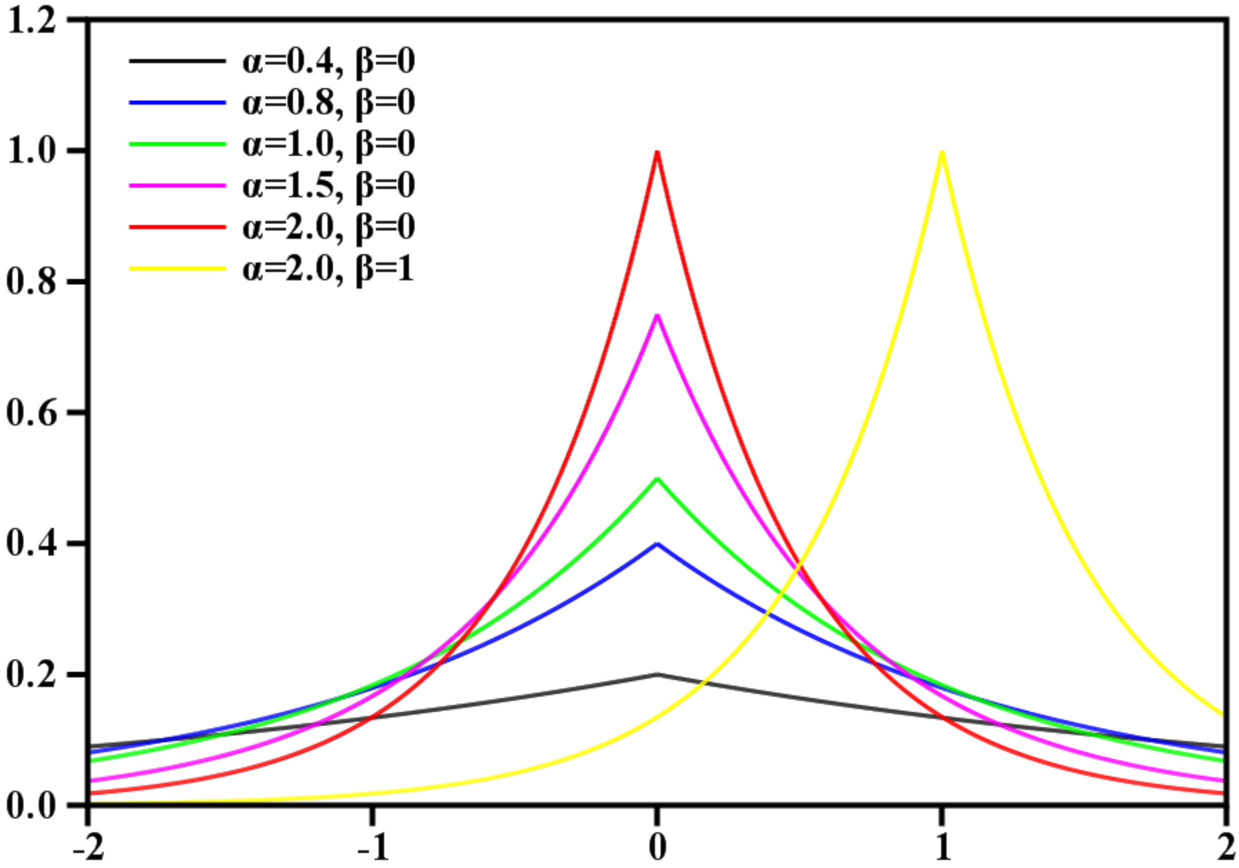
также даёт 0 для нормального распределения. Только лишь потому, что в AtteStat, где и приведена эта формула, $m_4 = m_4 - 3s^4$. Поэтому при самостоятельном выяснении сути формулы стоит смотреть не только на то, как выглядит непосредственно подсчёт эксцесса, но и на то, как высчитывается четвёртый момент.

Если значение эксцесса положительное, то пик остроконечный, если отрицательное, то гладкий.

Взглянем на функцию плотности распределения Лапласа[10], эксцесс которого равен 3:

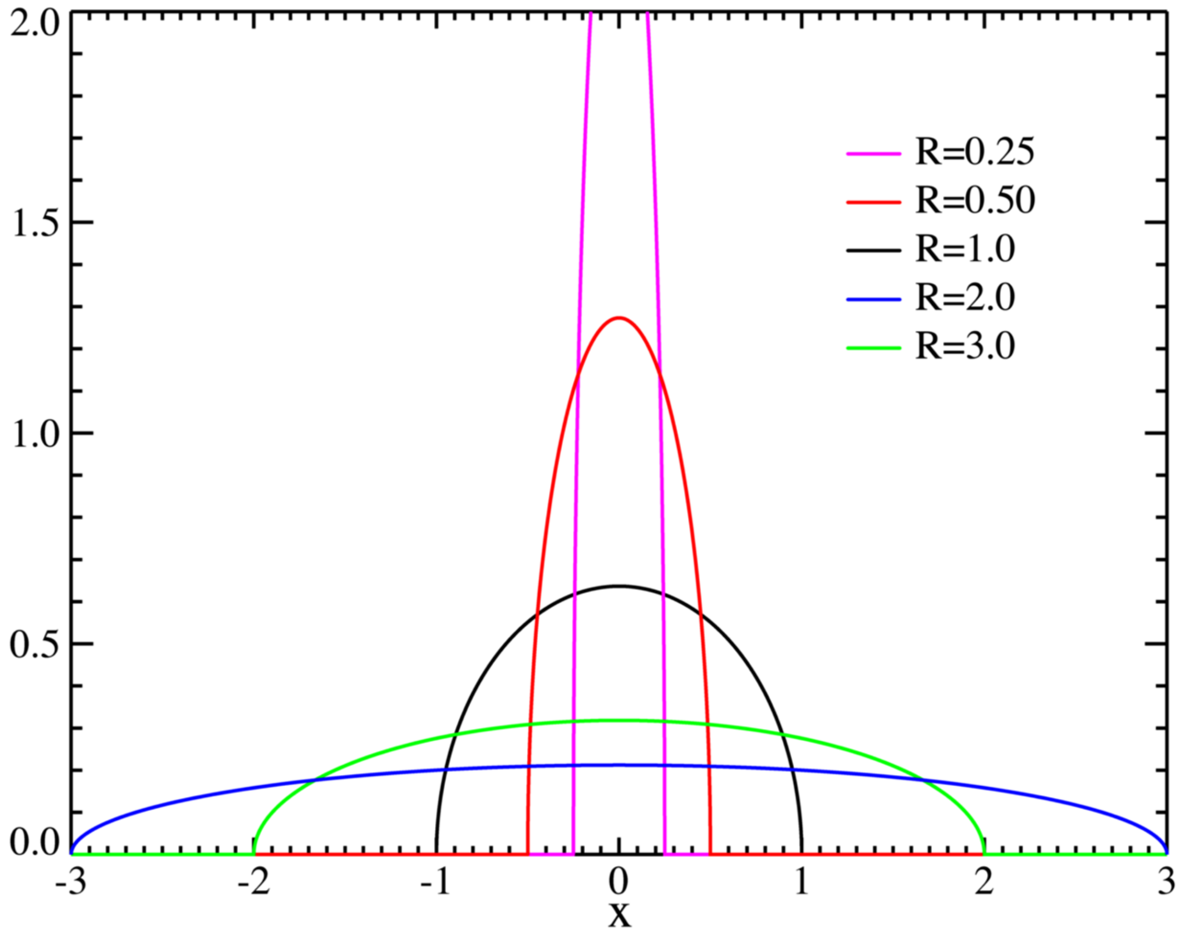
---

[10] Изображение взято с сайта: http://upload.wikimedia.org/wikipedia/ru/0/08/Laplace_distribution_%282%29_cdf.png



И сравним с полукруговым распределением Вигнера[11], эксцесс которого равен (-1):

[11] Изображение взято с сайта: http://upload.wikimedia.org/wikipedia/commons/e/ee/WignerS_distribution_PDF.png



Я думаю, после таких антиподов эксцесса, излишни дополнительные описания. К слову сказать, Юджин Вигнер вывел это распределение для своих исследовательских интересов, что ещё раз подчёркивает возможность уникальных распределений в психологических исследованиях.

Если же «пиков», на которые мы и смотрели, выясняя, чем же отрицательный эксцесс отличается от положительного, несколько, то смотреть эксцесс визуально уже не стоит. Потому что знак эксцесса согласуется с «пиком» (который на самом деле является не чем иным, как модой распределения), только тогда, когда пик у нас один, т.е. распределение унимодально. Случай нескольких мод, возвышающихся в разных местах распределения на гистограмме, означает, что стоит задуматься о вопросе: «а правильно ли набрана выборка».

Чем коэффициенты асимметрии и эксцесса отличаются от критериев асимметрии и эксцесса?

Данный вопрос не только полезен своей непосредственной сутью, но и призван раскрыть многообразие применения описанных показателей в статистическом анализе психолога.

Во-первых, асимметрия и эксцесс полезны сами по себе. Это мы видели ещё в рассуждениях об излишней скошенности распределения данных контрольной по математике.

Во-вторых, по асимметрии и эксцессу можно вести проверку нормальности. Это довольно спорный путь, поскольку наличие асимметрии и эксцесса, не отличающихся статистически значимо от 0, может и не означать нормальности, но, тем не менее, этот путь есть. Хотя более точную информацию дают специальные критерии нормальности.

Например, в программе AtteStat в модуле проверки нормальности есть критерии асимметрии и эксцесса, которые не только вычисляют, чему равны эти коэффициенты, но и указывают, значимо ли отличаются эти значения от идеала, т.е. от асимметрии и эксцесса нормального распределения.

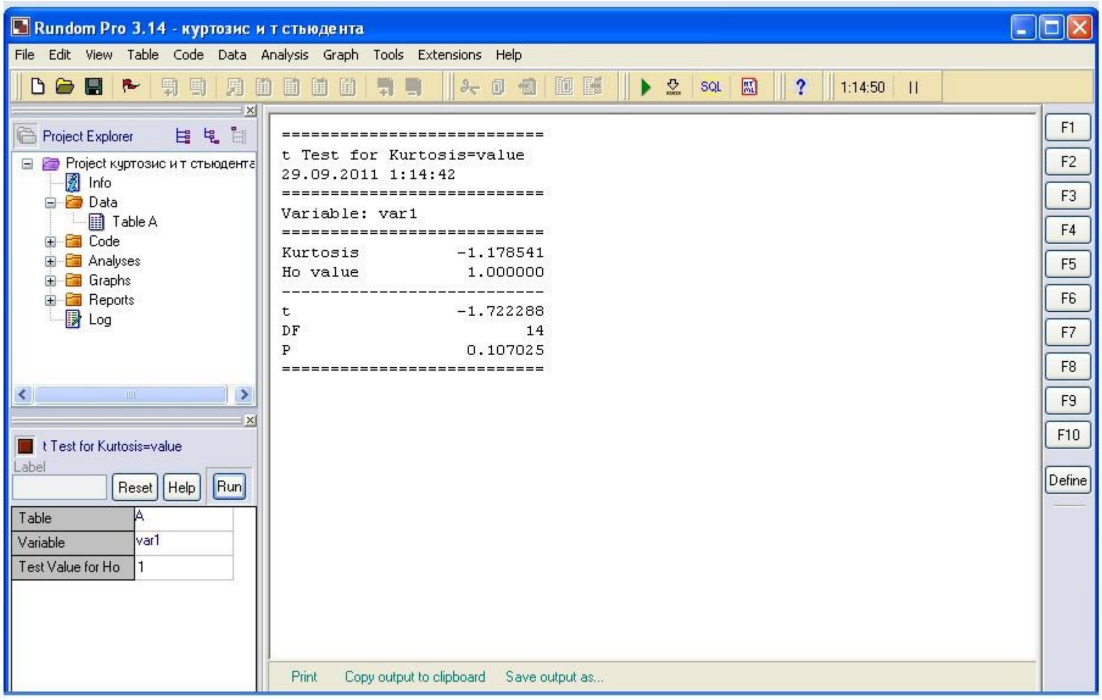
В-третьих, довольно любопытное воплощение критериев асимметрии и эксцесса дано в программе Rundom Pro. В отличие от AtteStat, здесь предлагается сравнить асимметрию и эксцесс не с нормальным распределением, а с любым, заданным собственноручно, что позволяет вести сравнение двух наборов данных по этим характеристикам распределения.

Делается это так.

Открываем программу Rundom Pro, создаём проект, вводим данные о тесте тревожности в столбик var1.

Далее нам нужно запустить сам анализ из меню: Analysis -> Classical -> Other -> t Test for Skewness (или, соответственно, Kurtosis). Посчитаем, например, эксцесс.

Для этого в графе Variable нужно задать var1 (если, конечно, вы скопировали данные именно в первый столбик). Далее нужно задать Test Value for Ho, которое по умолчанию стоит 0. Такое значение по умолчанию довольно очевидно, ведь эксцесс нормального распределения равен 0. А вообще эта графа переводится как «тестовое значение для нулевой гипотезы», т.е. с каким значением эксцесса мы собираемся сравнить то значение, которое получится при подсчёте эксцесса теста тревожности. Мы поставим значение 1, например. Нажимаем Run и получаем следующую информацию:



Чтобы понять, значимо ли отличается полученное значение эксцесса по тесту тревожности от заданного, т.е. 1, достаточно посмотреть на уровень значимости t-критерия Стьюдента, которым и велась эта проверка. Уровень значимости обозначается как Р и равен в нашем случае 0,107025. В зависимости от задачи нам будет необходимо значение Р, которое либо больше 0,05, т.е. такое, которое гласит о принятии нулевой гипотезы, либо Р, меньшее 0,05, т.е. гласящее об отвержении нулевой гипотезы и принятии альтернативной. Значение, большее 0,05 означает, что различий нет, т.е. принимается нулевая гипотеза – эксцесс теста тревожности не имеет статистически значимых отличий от 1.

Для глаза это кажется довольно странным, поскольку интуитивно кажется, что (-1) и (+1) – это очень разные величины, но статистические критерии учитывают гораздо больше условий, чем может учесть человек «глазом». А вот, например, заданное значение эксцесса 2 будет уже статистически значимо отличаться от эксцесса теста тревожности.

На этом стоит закончить разговор о формулах подсчёта описательных статистик и перейти к графическому представлению данных, которое во многом имеет право относиться к этой же группе методов.

1.4.6. Графическое представление данных.

1.4.6.1. Диаграмма рассеяния – визуализация связи параметров.

Эта диаграмма показывает как сырые баллы, так и линию наименьших квадратов. Рассмотрим пример. Предположим, что перед нами стоит абсолютно ироничная задача – узнать, связана ли сложность решения задач со временем их решения. Для этого мы задали сложность решённых каждым испытуемым задач баллами, время минутами и получили два столбика данных:

	сложность	время
Василий	23	21
Бенедикт	12	24
Карл	14	20
Георг	24	19
Татьяна	22	18
Катрин	11	16
Белла	20	25
Василиса	21	11
Лариса	43	24
Тамара	31	27

Чтобы увидеть сырые баллы в виде графика можно использовать диаграмму рассеяния. Построим её в программе Rundom Pro 3.14. Для этого нужно:

<u>А (приготовления)</u>

- запустить программу;

- создать новый проект кнопкой «Start new project» (либо закрыть стартовое диалоговое окно и из меню «File» выбрать вкладку «New project»);

- название стоит вводить аккуратно – некоторые варианты создают ошибку – например, нельзя называть проект «описательные статистики», лучше в таком случае назвать «описательныестатистики»;

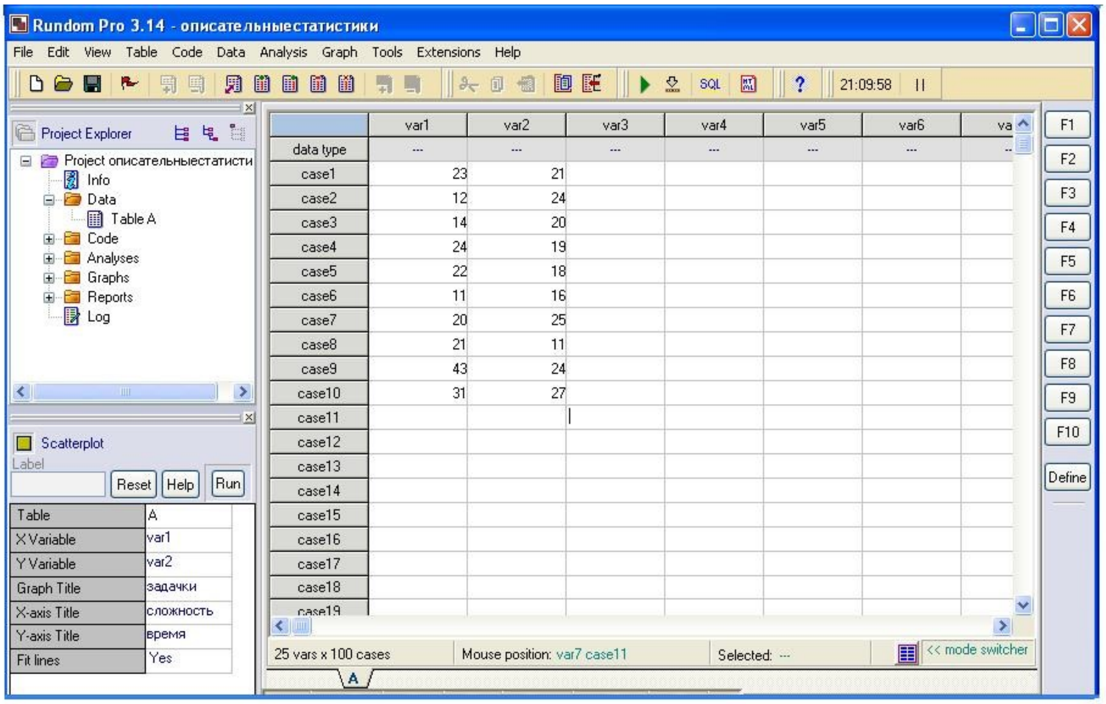
- для изменения места сохранения при введении названия в открывшемся меню создания проекта необходимо убрать галочку сохранения по умолчанию и задать собственное место (по умолчанию проекты сохраняются в C/Rundom Pro/PROJECTS);

- создав проект нам необходимо найти в меню слева строчку Table A и перенести данные, приведённые выше, в первые две колонки таблицы. Если создавать таблицу данных в Excel, то переносить данные нужно не просто процедурой «копировать – вставить», но при перенесении данных в Rundom Pro выбрать в меню «Edit» вкладку «Past Data (Active Table)», иначе все данные скопируются в первую ячейку таблицы.

Таким образом, мы создали проект и перенесли данные в таблицу программы.

<u>Б (подсчёты)</u>

Мы выбираем меню «Graph», вкладку «Scatterplot», что в переводе и означает диаграмма рассеяния. Слева снизу появляется меню построения такой диаграммы.



Мы задаём переменной X – var1, переменной Y – var2. Обратите внимание, что var1 и т.п. нужно писать именно слитно, иначе будет ошибка. Значение var1 означает, что нам нужны данные, расположенные в первом столбике таблицы. Если вас интересуют данные 5, 7 и прочих столбиков, то и порядковый номер нужно ставить соответствующий.

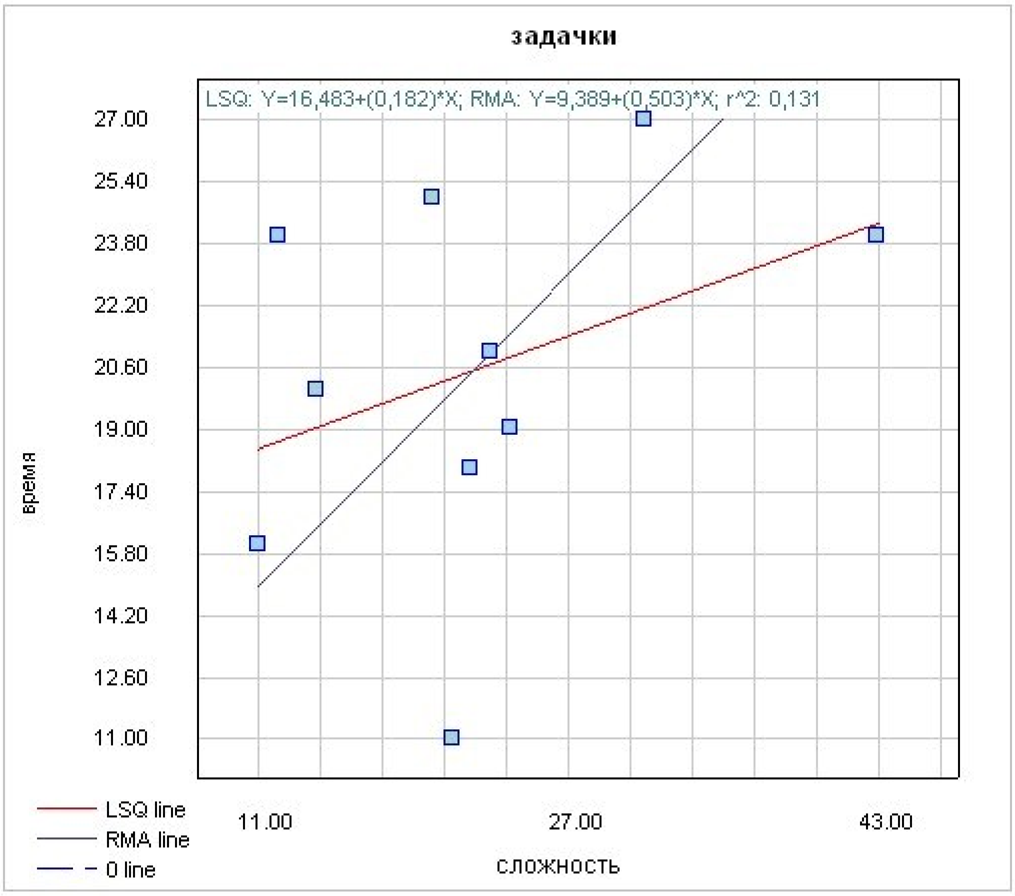
Graph Title – название диаграммы.

X-axis и Y-axis Title – названия осей. Обратите внимание, что при перенесении данных из Excel, я выделила только таблицу самих данных без названий.

Fit lines – линии приближения (приближённые линии), которые можно как показывать, так и скрывать. Мы выберем Yes, чтобы их увидеть и обсудить.

После того, как мы ввели все необходимые условия построения диаграммы рассеяния, нажимаем «Run».

Полученную диаграмму можно сохранить отдельным файлом, что мы и сделаем. Для этого нужно найти на самом поле диаграммы кнопку «Save as», выбрать место сохранения и графический формат. Полученная диаграмма выглядит следующим образом:



Анализ диаграммы можно проводить совершенно различными способами, зависящими от целей исследователя, поэтому я изложу лишь те общие принципы, которые демонстрируют принципы построения этой диаграммы.

Каждый квадратик на поле диаграммы в нашем случае – это испытуемый, который решил задачу определённой сложности за определённое время. Именно потому, что всего у нас было 10 испытуемых, квадратиков у нас тоже десять. За ось Y или ординат мы выбрали время, за ось X или абсцисс мы выбрали сложность, хотя можно было сделать и наоборот – серьёзных причин беспокоиться по этому поводу я не вижу. Хотя существует традиция в случае, когда параметры неравноправны, есть уверенность в причинно-следственной связи, представлять независимые переменные как X, а зависимые как Y. Например, если бы мы решили провести антигуманное исследование, где замеряли громкость крика от силы удара молотка по колону, то зависимой переменной выступал бы крик, а независимой – удар молотка, поскольку под независимой переменной понимается регулируемое экспериментатором воздействие, а под зависимой - отклик на это воздействие. В таком случае, X – сила удара молотка, Y – крик. Негуманно, но хорошо запоминается.

Чтобы понять, что нам представлены по-настоящему сырые баллы, стоит лишь посмотреть на исходную таблицу и полученную диаграмму. Например, квадратик, который находится на осях: (11 сложности; 16 времени) действительно имеет сырые баллы 11 и 16, это испытуемая Катрин. Таким образом, глядя на диаграмму рассеяния сказать, что она не наглядна, нельзя. Просто эта наглядность требует предварительной подготовки, которой не обладает каждый встречный пешеход.

В таком случае, узнав, что эти квадратики есть не что иное, как сырые баллы испытуемого, представленные в виде координат, можем ли мы сказать, что такая диаграмма даёт нам более полезную информацию, нежели процентные небоскрёбы, рассмотренные выше как антипример? Я думаю, что качество инструмента можно оценить только в границах его употребления. На эту тему Фейерабенд в одном из своих трудов писал замечательнейшую шутку: «… па классического балета прекрасны на сцене, но кто поверит, что с их помощью можно подняться на гору?».

Возвращаясь же с такой мерой оценки инструмента к сравнению небоскрёбов процента ответов и диаграммам рассеяния, можно сказать, что последние прекрасны хотя бы тем, что не дают почву для безосновательных выводов.

Какие же выводы мы можем сделать, глядя на диаграмму рассеяния?

Особенно важны такие диаграммы для исследователя, поставившего перед собой целью корреляционное исследование, но могут быть полезны и для общего ознакомления с массивом данных. Однако стоит понимать, что диаграмма рассеяния, где первой колонкой исходных данных будет значение группы 1, а второй колонкой значения группы 2, вряд ли имеет смысл. Диаграмма рассеяния показывает связь первой и второй координаты, поэтому не стоит пользоваться такой визуализацией, если данные двух параметров не предполагают между собой смысловой связи. Например, рисовать диаграмму рассеяния для задачи – лисы и муравьи в лесу, где первой колонкой данных будут лисы, а второй муравьи, не имеет смысла, потому что мы видим на диаграмме не автономное количество каждой группы, а связь параметров.

Первый вопрос, который мы можем задать диаграмме рассеяния, сгруппированы ли данные?

В некоторых случаях диаграмму рассеяния даже и определяют таким образом – как график, позволяющий все данные привести к единой линии.

Как мы видим, в нашем случае данные не сгруппированы. Хорошая группировка данных встречается достаточно редко в психологических исследованиях. И тем чаще, чем больше испытуемых в выборке. На графике хорошая группировка будет выглядеть как «облако», где все данные будут сконцентрированы вокруг красной линии. Причём чем больше облако похоже на круг, тем меньше связь между переменными, а чем более тонким выглядит облако, превращаясь в линию, совпадающую с красной, тем лучше. Естественно, никто не оценивает связь явлений по диаграмме рассеяния – для этого существуют специальные критерии, но так или иначе было бы странно не упомянуть о том, что диаграмма даёт информацию о связи и группировке данных.

Другой вопрос, который мы можем задать, формулируется примерно так – есть ли выбросы? Под выбросом понимается такое значение данных, которое резко отличается от остальных. Работа с выбросами – это тема для отдельной главы, поскольку слишком отклоняющееся значение вызывает сомнение. Если, например, в классе был замерен интеллект и все данные находятся в пределах 100-110 IQ, то что можно сказать об интеллекте двух девочек, одна из которых набрала 60 IQ, а другая 150? В самом кратком варианте можно рассуждать двумя путями – 1) эти девочки особенные и то, что мы получили, истина, отражающая особенное положение этих детей; 2) при проведении исследования были не учтены какие-то факторы, которые повлияли на эти результаты, т.е. измерение содержит ошибку. Соответственно, выбросы можно оставлять, можно исключать из анализа.

Поскольку сейчас мы обсуждаем диаграммы рассеяния, то нас интересует вопрос – есть ли выбросы, в свою очередь вопрос о том, что с этими выбросами делать, мы оставляем в стороне. На том графике, который мы построили для сложности задач и времени решения, принять решение о выбросах практически невозможно, поскольку характерного облака скопления данных нет. Здесь все данные достаточно разрозненно располагаются на поле диаграммы. Поэтому в данном случае мы можем принять такое решение лишь субъективно, с одной-единственной целью – показать общий принцип таких действий. Итак, выбросом, на мой взгляд, имеет смысл назвать лишь нижнее значение с координатами (21; 11), т.е. с 21 баллами по сложности и 11 баллами по времени, что соответствует испытуемой Василисе. Что же касается других значений, например, Тамары (31; 27), то я не удивлюсь, если при увеличении выборки мы обнаружим появление испытуемых, которые окажутся между значением Тамары и красной линией, поскольку есть ещё три значения, которые находятся между значением Тамары и красной линией.

Приведённое рассуждение о выбросах побуждает ещё раз подчеркнуть, что статистические методы, будь то методы хоть дескриптивной, хоть индуктивной статистики, требует большого количества наблюдений. И чем больше наблюдений, тем точнее выводы. Десять испытуемых, взятых в нашем примере, явно показывают недостаточность наблюдений уже на уровне предварительного рассмотрения данных выборочного исследования, не говоря уже о выводах, распространяемых на генеральную совокупность.

Следующий вопрос в процессе ознакомления с принципами построения диаграммы рассеяния звучит так: что такое красная линия или LSQ line?

LSQ line – Least Squares Quadratic или просто Least Squares Regression – линия наименьших квадратов. Красная линия или линия регрессии построена по методу наименьших квадратов, что в предельно простом варианте означает такое расположение линии между точками на графике, что расстояние до каждой точки от этой линии минимально. Такая линия регрессии задаётся уравнением $Y = a + b*X$, где a – константа, b – угловой (регрессионный) коэффициент.

Давайте проверим получившееся уравнение регрессии теми самыми данными, которыми располагаем. Итак, согласно логике получившегося уравнения, {время решения задачи = 16,483 + (0,182) * сложность решения задачи}. Эту информацию я взяла исключительно из построенной нами диаграммы рассеяния – вверху диаграммы есть соответствующая надпись.

Тогда нам нужно первую колонку, сложность решённых задач, умножить на регрессионный коэффициент, равный 0,182, прибавить к получившемуся значению 16, 483 и получить в идеале колонку 2, а именно, время решения задачи. Так как идеала не получится, то мы заодно посмотрим, насколько отличным от предполагаемого значения будет истинное значение времени решения задач.

	сложность	время	ур регр	отклонение от предполагаемого
Василий	23	21	20,669	0,331
Бенедикт	12	24	18,667	5,333
Карл	14	20	19,031	0,969
Георг	24	19	20,851	-1,851
Татьяна	22	18	20,487	-2,487
Катрин	11	16	18,485	-2,485
Белла	20	25	20,123	4,877
Василиса	21	11	20,305	-9,305
Лариса	43	24	24,309	-0,309
Тамара	31	27	22,125	4,875

Таким образом, мы получили таблицу, где видно, что, например, для Василия по уравнению регрессии предполагаемое время решения задач составляет 20,669, а реальное значение, полученное в ходе нашего исследования 21. Соответственно, уравнение регрессии обмануло нас на 0,331.

Считать эти показатели нет никакой нужды, поскольку на самом деле стоит присмотреться к тем значениям, которые отмечены красной линией. Особенно удобно смотреть на результаты Ларисы и Тамары, поскольку их не составляет большого труда обнаружить на диаграмме рассеяния. Оказывается, красная линия состоит как раз из тех самых координат, которые мы сейчас высчитывали, а именно – из предполагаемых значений Y по уравнению регрессии для соответствующего X.

Нам вовсе не обязательно знать, как такая линия строится, но нужно знать предельно простые признаки, как она может выглядеть.

Рассматриваемый пример показывает, что реально зафиксированные эмпирические значения достаточно далеко отстоят от красной линии регрессии, что является заочным признаком слабой связи или её отсутствия, потому что когда связь между двумя параметрами сильна, то все точки будут стремиться к этой линии.

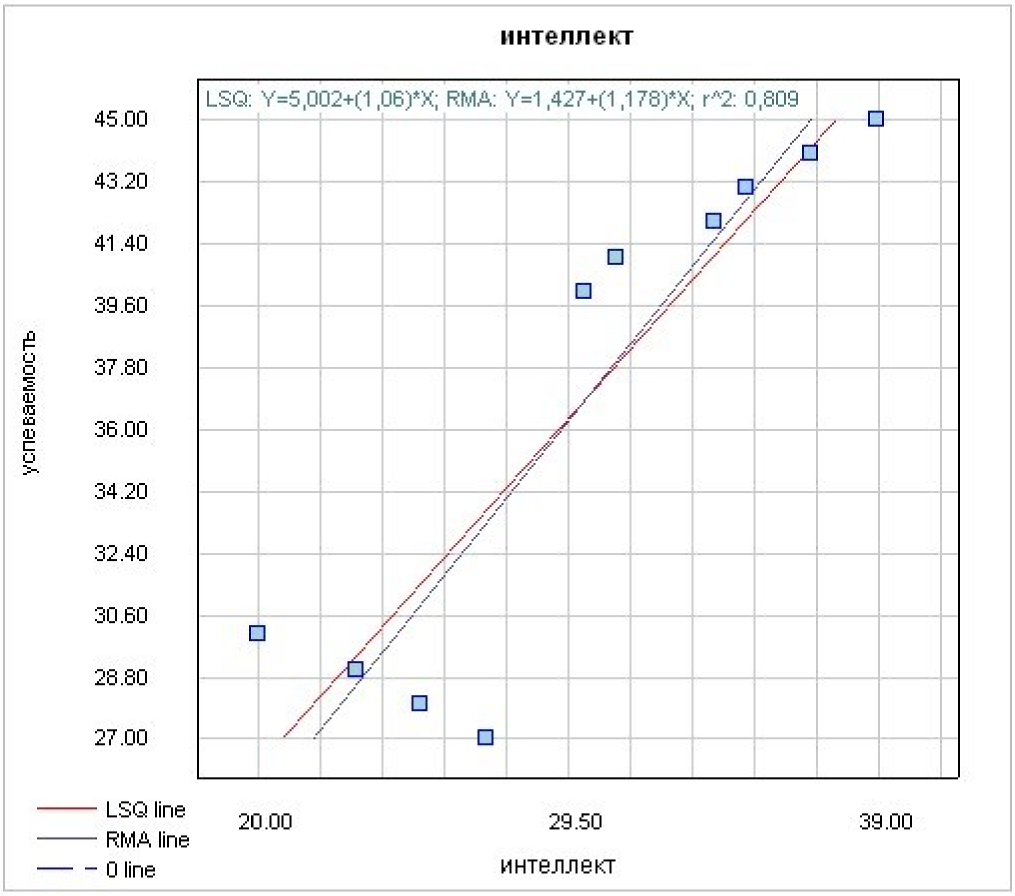
Наклон линии данных, представленных диаграммой рассеяния, может свидетельствовать о направлении корреляционной связи параметров. Соответственно, если линия регрессии выглядит как подъём в гору, т.е. левый её край ниже, чем правый (как в нашем примере, если бы корреляция была ещё и значимой), то это свидетельство положительной связи, т.е. увеличению сложности задач в нашем примере характерно увеличение времени решения задач. Если линия регрессии выглядит как спуск с горы, то связь отрицательна, т.е. увеличению сложности задач было бы характерно уменьшение затраченного времени на их решение. Соответственно, диаграмма рассеяния позволяет сделать такой замечательный вывод, который может быть и не решён даже методами индуктивной статистики. Что же может увидеть глаз? Глядя на диаграмму рассеяния можно заметить нелинейность связи, т.е. ситуацию, в которой точки/квадратики испытуемых располагаются не случайным образом, но и не линейным. Например, квадратики могли расположиться в ряд какой-нибудь кривой линии, что делало бы применение линейных методов проверки корреляции неадекватными. Применение же критерия Пирсона, о котором нам ещё предстоит узнать, к нелинейным данным, может сбить исследователя с истинного пути, потому что результат «связи нет» не отражает истинного положения дел, а лишь свидетельствует о том, что нет линейной связи.

Психологи обычно ограничиваются тем, что, заметив кривые линии, используют коэффициент корреляции Спирмена, хотя это, как обычно и бывает в психологии, «полуправда». Вторая часть этой правды звучит как немонотонность функции. Дело в том, что кривая линия может сохранять знак на всём своём пути, а может и не сохранять. В случае, если знак будет меняться, как если бы мы шли не с горы и не на гору, а по холмикам, то немонотонность функции ставила бы под вопрос не только использование коэффициента Пирсона, но и Спирмена. Поэтому единственный совет, который можно в этом случае дать, если конечно читатель не владеет респектабельным багажом знаний по аппроксимации, это делить выборку, строить для получившихся делений собственные графики рассеяния и смотреть, чтобы полученное рассеивание данных предполагало оценку простой (монотонной) связи.

Последним комментарием к рассмотрению «формы» полученных данных назову «крест». Мы выяснили, что положительная связь будет выглядеть как подъём в гору, а отрицательная как спуск с неё. Давайте создадим ситуацию, где данные будут иметь так называемую функциональную корреляционную связь, т.е. самую тесную связь из всех возможных – (1) и (-1). Но мы создадим этот пример таким образом, что половине данных будет соответствовать корреляция в 1, а другой половине данных в (-1) и сделаем вид, что мы этого не знаем.

20	30
23	29
25	28
27	27
30	26
31	40
34	41
35	42
37	43
39	44

Тогда при помощи любой программы, которая может строить графики, мы увидим следующую картину:



Если мы измерим корреляцию всего массива, то получим значение коэффициента Спирмена = 0,75. Это значение значимо, но не отражает суть получившегося явления в должной мере. Исследователь, который сделает вывод, исходя из таких данных, что высокому интеллекту характерна лучшая успеваемость, сделает серьёзную ошибку. И когда результаты исследования не удастся повторить, винить будет нужно не статистику и не методы, а самого исследователя. Я называю такую картину на графике рассеяния крестом, потому что по сути при одновременном присутствии положительной и отрицательной связи можно встретить и наложение, которое будет ничем не лучше, чем игра в «крестики-нолики», которую все мы помним из детства.

Данный пример предполагает очень простое решение – нужно разделить выборки и понять, почему они столь различны. Разделив выборки и посчитав коэффициенты корреляции отдельно для первых пяти испытуемых и остальных пяти, мы выясним, что у нас есть функциональная связь в 1 и в (-1). Стоит ли говорить, что случай двух функциональных зависимостей в (-1) при подсчёте коэффициента Спирмена может вообще не показать значимого результата без соответствующего деления выборок? Для проверки высказывания приведу данные без подсчётов:

20	30
23	29
25	28
27	27
30	26
50	60
49	61
48	62
47	63
46	64

Таким образом, нерассмотренным остался лишь один вопрос. Что же такое линия RMA?

Линия RMA расшифровывается как Reduced Major Axis – редуцирование главных осей. Этот метод является усовершенствованием и аналогом по цели методу наименьших квадратов. Считается, что такая линия может служить более адекватным предсказанием в случае, если независимая переменная характеризуется ошибками измерения. Однако, в Российской Федерации этот метод, насколько я могу судить по представленности русскоязычных источников на заданную тему, используется категорически редко. Такая ситуация означает отнюдь не низкое качество, а скорее относительную молодость создания.

Диаграмма рассеяния может выглядеть совершенно по-разному и в некоторых случаях настолько необычного, что далеко не сразу читателю придёт в голову, что такая диаграмма есть воплощение знакомых идей. Например, диаграмма рассеяния Вороного более напоминает пустыню, нежели бусинки посреди поля. К тому же современные статистические пакеты позволяют строить не только двухмерные диаграммы, но и трёхмерные, для трёх параметров. С другой стороны, можно построить такую диаграмму, которая будет сочетать в себе несколько диаграмм, построенных для разных наборов данных, одновременно. Перечисление возможного разнообразия диаграмм рассеяния может занять целое, пусть и небольшое, методическое пособие. В мои цели не входит осветить все возможные графики и диаграммы, скорее для меня важно, чтобы читатель, который предпочитает графическое восприятие информации (чем я, к своему великому сожалению, похвастаться не могу) мог позволить себе выйти за пределы «пирогов» и «небоскрёбов» пустых надежд и необоснованных выводов.

1.4.6.2. «Ящик с усами» - визуализация различия групп.

Другим способом наглядного представления данных, который мы рассмотрим, являются так называемые «ящики с усами» или «box-and-whiskers plots». Такое юмористическое название, как ни странно, совершённо чётко отражает суть того, как выглядит такой график, просто под усами стоит понимать скорее усики насекомых, например, бабочек.

Говоря об этом типе графического представления данных, нам придётся вспомнить описательные статистики, о которых речь шла выше. Приведу шутку, которая прекраснейшим образом показывает, почему график «ящик с усами» гораздо более предпочтителен, чем, например, сравнения групп в традиции «небоскрёбы» со средними баллами, на которых в худшем варианте не только заканчивается графическое представление данных, но и статистическая обработка в целом.

«Сидеть на горячей плите, надев на голову холодильник, в среднем не плохо»[12].

Другими словами, среднее арифметическое – далеко не единственный способ сравнить полученные данные. Ящик с усами не заменит индуктивную статистику, которая ответит, действительно ли стоит верить глазу, который

[12] http://libertygrant.co.uk/portal/?p=706

говорит, что данные различны. Но график прекрасно такую статистику дополняет, когда мы делаем презентацию исследования на защиту перед комиссией коллег или в случае, когда нам нужно определить, в какой именно группе значение выше (далеко не всегда индуктивная статистика выдаёт в качестве отчёта направленное различие, существуют критерии, которые показывают лишь факт различия).

Построим такой график в программе Rundom Pro. Для этого создадим новый проект.

Предположим, что перед нами данные скорости бега двух футбольных команд по 10 человек в каждой. К сожалению, я понятия не имею, какова средняя скорость бега нормального человека, а также не подозреваю, сколько человек должно быть в футбольной команде, поэтому прошу не судить меня строго за ассоциативный ряд.

	бег	команда
Василий $	45	1
Георгий $	34	1
Дмитрий $	46	1
Константин $	43	1
Пётр $	48	1
Станислав $	49	1
Геннадий $	50	1
Николай $	52	1
Прохор $	53	1
Павел $	90	1
Михаил #	23	2
Борис #	22	2
Поль #	21	2
Владимир #	20	2
Алексей #	19	2
Александр #	16	2
Роман #	25	2
Элвис #	23	2
Брюс #	21	2
Джордж #	1	2

Обратите внимание, что данные по скорости бега представлены в один столбик, а второй показывает, к какой команде относится бегун. Такое представление исходных данных весьма характерно для задач, касающихся сравнения групп, а не признаков. Я таким образом создавала данные, что знаю ответ заранее – первая группа явно бегает быстрее, но мы всё же взглянем на получившийся график, ведь экстрасенсорика далеко не всегда сопутствует исследовательскому анализу. К тому же в каждую группу я умышленно поместила два отличающихся от общего фона значения – Павел из первой группы имеет скорость бега 90, а Джорджу из второй группы вообще не повезло – он бегает со скоростью в 1 м/с.

Создав проект и перенеся исходные данные в программу, мы можем приступать к построению графика. Для этого необходимо выбрать меню «Graph», вкладку «Box Plot». В левом нижнем углу появилось меню.

Dependent Variable (зависимая переменная) в нашем случае это бег, т.е. первая колонка данных, поэтому мы пишем var1.

Grouping Variable (группирующая переменная) – это такая переменная, которая показывает, к какой из групп относится значение зависимой переменной. В нашем случае указание на соотнесение с футбольной командой находится во второй колонке, поэтому мы пишем var2.

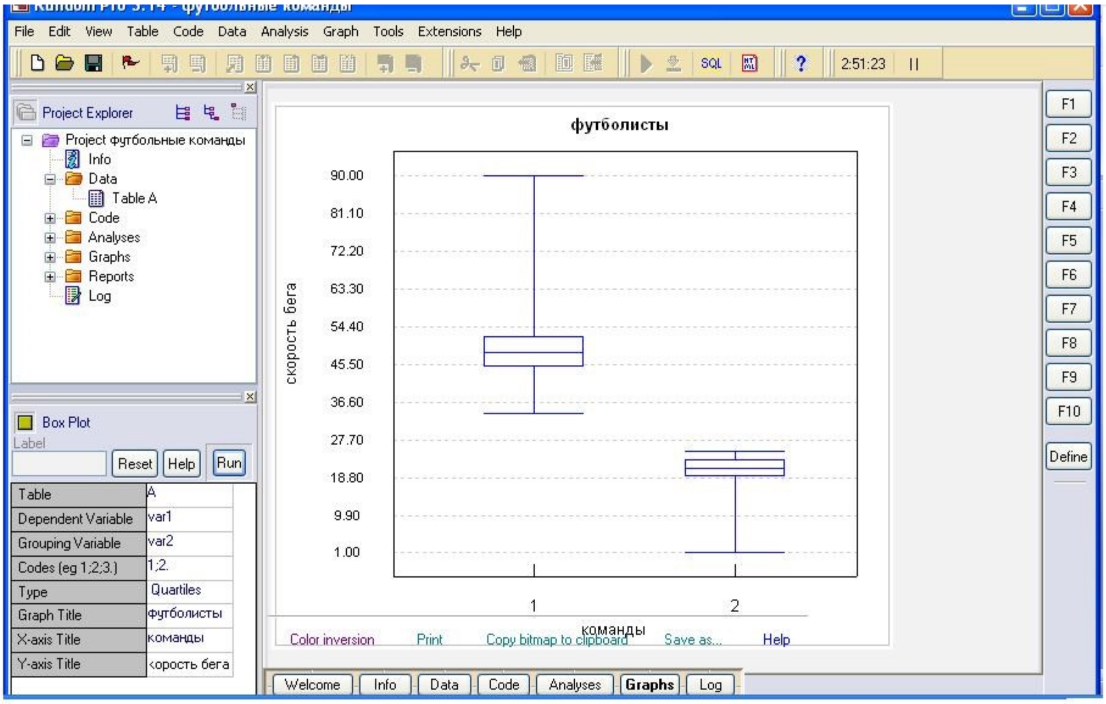
Codes – это кодировка групп. У нас было всего две команды, которые мы обозначили через 1 и 2, поэтому нам нужно написать «1;2.». Обратите внимание на то, чтобы при самостоятельном воплощении были соблюдено написание, иначе может возникнуть ошибка при выполнении алгоритма.

Graph Title – название графика, предположим «футболисты».

X-axis Title – это не название первой команды, а факт того, что вообще эти команды есть, т.е. это сущность самих ящиков. Поэтому мы пишем «команды».

Y-axis Title – это название того параметра, по которому идёт сравнение. Поэтому нам следует назвать ось Y «скорость бега».

Нажав на кнопку Run, мы увидим следующее:



У меня слегка наезжает меню сохранения результатов, инверсии цвета и прочего на саму диаграмму из-за разрешения экрана. При презентации данных этого можно с лёгкостью избежать, использовав всё ту же замечательную кнопку «Save as» для сохранения полученного графика в удобном для вас графическом формате.

Даже не зная описательных статистик, читатель может понять, что скорость бега первой команды гораздо выше, чем у второй.

Но всё же мы разберёмся с тем, что же значит «пузечко» и «усики».

Поскольку тип ящик с усами у нас был квартильный, то перечисление горизонтальных линий снизу вверх выглядит следующим образом:

- первая (нижняя) горизонтальная линия – минимальное значение;

- вторая снизу горизонтальная линия – первая квартиль;

- третья линия – вторая квартиль (медиана);

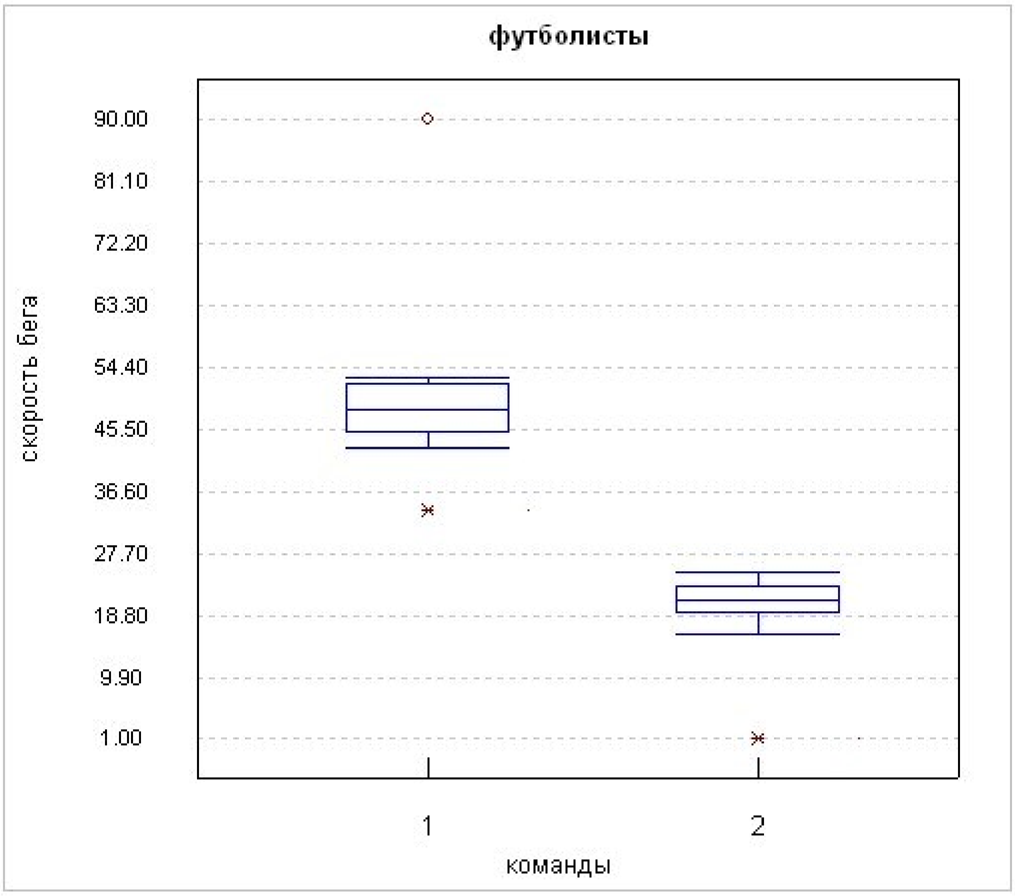
- четвёртая линия – третья квартиль;

- самая верхняя горизонтальная линия – максимальное значение.

Поскольку я уточнила, что такое перечисление горизонтальных линий свойственно квартильному ящику с усами, то логично предположить, что есть и другие.

Гораздо более любопытный (на мой субъективный взгляд) ящик с усами указан в программе Rundom Pro как тип: Quartiles2.

Чтобы сменить стиль ящика с усами, нужно навести курсор мышки на правую сторону ячейки Quartiles и щёлкнуть. Должно выплыть меню выбора типа графика, в котором будет представлено 3 разновидности. Мы выбираем Quartiles2 и нажимаем Run, что даст нам следующую картину:



В данном случае, несмотря на удивительную схожесть с предыдущим графиком, принцип построения несколько иной.

Первой (нижней) горизонтальной линией является следующая величина:

(Q_1 – 1,5*IQR), где в русскоязычном эквиваленте Q_1 означает первую квартиль, а IQR – интерквартильный размах, т.е. разность между 75 и 25 квартилью.

Второй горизонтальной линией будет первая квартиль.

Третьей – вторая квартиль (медиана).

Четвёртой – третья квартиль.

Пятой или верхней горизонтальной линией является величина: (Q_3 + 1,5*IQR).

В таком случае остался только один вопрос, касающийся второго квартильного способа представления ящика с усами. Что же это за точка и две «мухи»?

Эти точки и «мухи» - самое любопытное, что есть в таком способе построения ящика с усами. Кто-то называет их «неудачниками», кто-то бы дословно перевёл англоязычный статистический термин «outliers» как посторонних. На самом же деле русскоязычное статистическое понятие, которое и воплощают в себе эти «мухи», звучит как выбросы. Под выбросами понимают «подозрительно» высокие и низкие значения, которые не вписываются в общую картину наблюдения. Но легче сказать, чем сделать. Идентификация выбросов – это вполне достойная математическая задача, над которой бьётся целая группа специалистов по всему миру, создавая новые и, надеюсь, более точные методы. Есть крайне специализированные статистические критерии, призванные решать такие задачи, например, критерии Граббса (Груббса), Кохрена (Кокрена), Титьена-Мура и т.д. С некоторым множеством таких методов можно ознакомиться и пользоваться с помощью, например, программы AtteStat в модуле обработки выбросов соответственно. В многомерных зависимостях глаз уже никак вообще не поможет, в связи с чем есть такой экспериментальный метод как нейронная сеть персептронного типа, с помощью которой также предпринимаются попытки определить выбросы, но мы рассматривать её не будем ни в коем случае. AtteStat также советует для обработки выбросов в многомерных данных применять факторный и кластерный анализ.

Самым ужасным из всех требований «серьёзных» методов является главная печаль любого психолога, а именно – нормальность распределения. Поэтому в случаях отвержения этого требования, не остаётся никакого другого способа, кроме оценки «на глаз». Но у «глаза», помимо субъективизма, есть другая вполне объективная причина, которая к тому же объяснит, в чём же проблема определения выбросов. Выбросы – это вовсе не единая, неделимая группа явлений. Эта группа явлений в свою очередь содержит внутри себя разновидности в связи с тем, что причина возникновения выброса как факта может быть очень своенравной. Например, мы проводим тест интеллекта летом, в душном помещении, а за окном гремит трактор и экскаватор, потому что школа продала некоторую часть своей территории под постройку жилого дома. Помимо того, что в такой ситуации у нас может снизиться общий показатель интеллекта группы, но всенепременно будут и выбросы, поскольку испытуемые весьма различны в уровне устойчивости к фрустрации. Кто-то в состоянии думать под гул цивилизации, а кто-то совсем нет. Полезависимость поленезависимости рознь.

Итак, мы запомнили, что в школьном классе при проведении теста интеллекта у нас были зафиксированы выбросы, которые связаны с ошибками измерения. Конечно, нам нужно приложить титанические усилия, разгадать величайшую загадку, какой же уровень был настоящим, а какой заниженным, отсечь сор и приняться за дальнейшие подсчёты.

Совершенно другой тип выбросов будет в ситуации, когда у нас из всех мужчин мира будут выбраны мужчины, живущие в нашем дворе. Эта ошибка репрезентативности может дать нам «мух», которые «мухами» являться перестанут, как только мы сможем собраться с духом и выйти в более широкое исследование. За счёт чего? За счёт того, что слишком низкий уровень агрессивности нашего соседа дополнится относительно низким уровнем агрессивности испытуемого из соседнего двора. Т.е. исследователь сможет увидеть, что это не случайное аномальное явление, а лишь разновидность общего закона. В таком случае, по отношению к такого рода выбросам нельзя относиться как к предыдущим – их вообще нужно оставить, поскольку они отражают не сор, а истину.

Третий и последний вид выбросов, который я опишу в данном пособии, хорошо может быть представлен каким-нибудь анекдотическим примером следующего характера. Предположим, мы решили изучить отношение женщин к комнатным растениям и приготовлению пищи. Такая ситуация не представляется такой уж нереалистичной, особенно, если вспомнить шкалу MMPI на феминность – мускулинность. И вот мы видим, что первая испытуемая нашей выборки обожает комнатные растения, разговаривает с ними, потому что от голоса человека они лучше растут и т.д. Другая – вообще их ненавидит, не поливает, выживают только кактусы и т.д. На данном моменте я ещё не ввела нового критерия выброса, ведь предыдущий тоже может подойти – при увеличении выборки эта разрозненность данных может стать более ровной. Но в итоге мы видим, что у нас данные распределились весьма странным образом – одна половина женщин цветы и кулинарию обожает, а другая ненавидит. Что это за выбросы? Какова их природа? Всё дело в том, что теоретически заданная классификация далеко не всегда подтверждается эмпирическим путём. Тот факт, что мы, как исследователи, предполагаем, что люди бывают двух типов: мужчины и женщины, вовсе не означает, что так оно и есть. Я говорю сейчас, естественно, не о физиологических различиях, хотя и они не столь однозначны, как казалось бы, но говорю о культурном аспекте, который носит ещё более разрозненный характер. Феминистки, лесбиянки, эмансипе – все эти женщины вполне могут и не согласиться с нашим стереотипом о любви к цветам и кулинарным чудесам. Но увеличение выборки женщин не приведёт к тому, что выбросы исчезнут. Они всё также будут, поскольку у них совершенно отличная природа и от ошибки измерения, и от ошибки репрезентативности. Эти выбросы имеют в своей основе ошибку классификации, и деление на более гомогенные группы приведёт к тому, что исследователь «соберёт паззл». Соответственно, такого рода выбросы также нельзя выкидывать – их нужно переносить в другой подсчёт.

Таким образом, мы рассмотрели три природы выбросов, которым соответствует три стратегии поведения исследователя – исключение, дополнение и перераспределение.

Естественно, что могут быть и другие причины выбросов. Но самое, пожалуй, страшное – это невозможность определения природы выбросов без дополнительных исследований. Ведь сидя перед раз и навсегда заданными данными мы никогда не определим доподлинно, почему у нас именно такая картина. В то же время, проведя дополнительное исследование, проверяющее более вероятную гипотезу, мы можем получить хотя бы исключение того или иного предположения о природе выбросов. Но сколько бы не были радужны предположения автора об инициативности студентов, мне всё же совершенно ясно, что проверять природу выбросов получается далеко не у всех. В этом случае мне хотелось бы сказать одно – нужно, по крайней мере, знать, не много ли их. И без дополнительной проверки их природы возможно самый лучший способ – исключить их из расчёта.

После того, как мы чётко представили себе, что же такое выбросы, настало самое время разобраться с тем, почему некоторые из них обозначены точками, а некоторые звёздочками или «мухами» - право не знаю, как назвать этот знак.

В русском языке термины «outliers» и «extreme values» чаще всего переводятся одинаково – выбросами, однако это не совсем адекватно, если мы имеем дело с программой Rundom Pro. Помимо того, какова природа выбросов, их классификацию можно строить и по принципу, насколько сильно они отличаются от общего расположения данных. В AtteStat это деление переведено следующим образом: outliers – мягкие (подозрительные) выбросы, а extreme values – экстремальные выбросы. Как же они вычисляются?

Опять же рассмотрим случай Rundom Pro.

Outliers – можно исключать, можно и не исключать. Они находятся в пределах между внутренней и внешней границами, которых в общей сложности получается 4, а, соответственно, коридоров, в которых эти выбросы находятся, два.

Внутренние границы – это уже знакомые нам нижняя и верхняя горизонтальные линии самого ящика с усами.

Нижняя граница ящика с усами, являющаяся и нижней внутренней границей – величина: (Q_1 – 1,5*IQR).

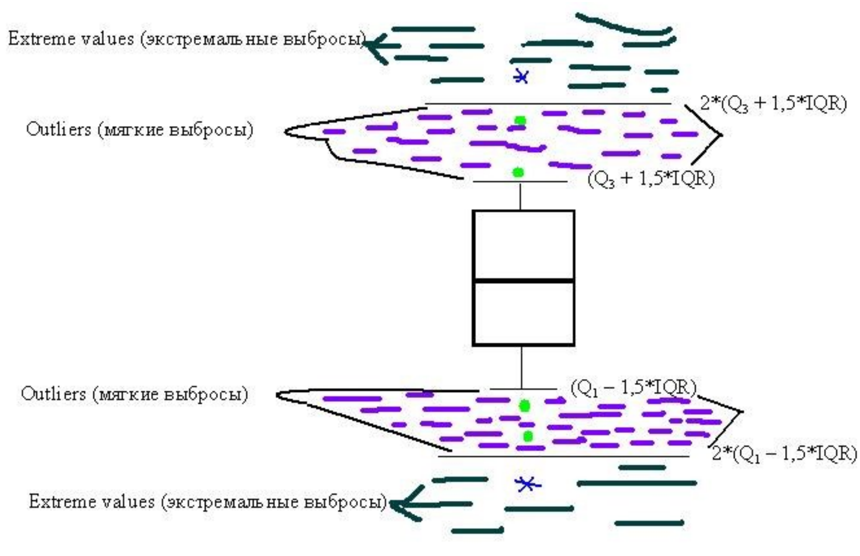
Верхняя граница ящика с усами, являющаяся и верхней внутренней границей – величина: (Q_3 + 1,5*IQR).

В свою очередь внешними границами называются удвоенные внутренние границы.

Нижняя внешняя граница – 2*(Q_1 – 1,5*IQR).

Верхняя внешняя граница – 2*(Q_3 + 1,5*IQR).

В то время как Extreme values (экстремальные выбросы) находятся за верхней и нижней внешними границами. Чтобы окончательно представить себе эту ситуацию, достаточно взглянуть на эту картинку:



В центре в меру моих эстетических сил изображён ящик с усами. Указаны в качестве нижней и верхней внутренних границ его «ножки». Дорисованы внешние границы как удвоение «ножек» и показано, в каких областях находятся оба типа выбросов.

Существует традиционный совет исключать из подсчётов хотя бы выбросы, попавшие в зону экстремальных значений, если природы выбросов не известна.

Для определения внешних границ диапазона выбросов существует и другая формула подсчёта, не реализованная в Rundom Pro, а именно: (Q_1 – 3*IQR) и (Q_3 + 3*IQR), что, судя по всему, делает распределение выбросов между двумя группами – мягких и экстремальных, несколько разными способами. Так, метод, реализованный в Rundom Pro, будет относить нижние выбросы к экстремальным чаще при общих равных, чем второй метод, поскольку его внешняя граница находится ближе к «нижней ножке» ящика с усами, но если рассматривать верхние выбросы, то Rundom Pro будет рассматривать их мягче, поскольку верхняя граница, определяющая внешнюю границу, находится дальше от «верхней ножки», чем у метода, в котором используется формула без удвоения третьей квартили. Однако, сказать, какая из этих формул более точна, если и представляется возможным, то не мне.

Таким образом, мы видим, что такой ящик с усами полезен не просто для сравнения двух групп, но и для работы с выбросами в рамкам одной-единственной группы.

Последний тип ящиков с усами может быть применён только в том случае, если мы имеем метрическую шкалу измерения признака + распределение этого признака если не нормально, то хотя бы симметрично. Но тогда возникает адекватный вопрос – действительно ли данные имеют метрическую шкалу, если распределены не нормально, учитывая тот факт, что большинство психологических шкал буквально «выбивают» себе право быть признанными метрическими исключительно за счёт «правильного» распределения. Я говорю это к тому, что стандартное отклонение и среднее арифметическое имеет смысл вычислять только для метрических шкал, поэтому для корректного использования такого типа ящиков с усами стоит осторожнее подходить к априорным требованиям.

Если же всё-таки данные имеют метрическую природу, то можно позволить себе такой график.

Нижняя горизонтальная линия – это граница стандартного отклонения.

Вторая снизу линия – это граница стандартной ошибки среднего.

Третья снизу линия – среднее арифметическое.

Четвёртая снизу или вторая сверху линия – это другая граница стандартной ошибки среднего.

И верхняя линия – это граница стандартного отклонения.

К вопросу о том, что же такое стандартная ошибка среднего. Это такая величина, которая отражает точность вычисления среднего арифметического и высчитывается по формуле: $m = \frac{\sigma}{\sqrt{n}} = \frac{\sigma^2}{n}$.

На этом мы заканчиваем рассмотрение графического представления данных. Были рассмотрены только самые основные его виды, что вовсе не означает, что на этом исчерпываются возможности исследователя. Статистические пакеты предлагают избыточное количество различных графических методов, где каждый найдёт оптимальный для своего изысканного вкуса, однако не стоит слишком злоупотреблять красочностью и сложностью графического инструментария, поскольку он может преследовать в первую очередь две цели – дать полезную информацию о распределении данных + упростить восприятие целостной картины, а не восхитить, шокировать, дать необоснованные выводы и т.д.

1.4.6.3. Гистограмма и функция плотности распределения.

Чтобы взглянуть на своё распределение данных с точки зрения количества мод, формы распределения, оценить эксцесс на «глаз» и т.п. используется гистограмма, которая строится по достаточно простым принципам. По сути гистограмма отвечает на вопрос о частоте, с которой встречаются те или иные значения (например, теста) в группе.

Не будучи достаточно математически образованным человек, думая о распределениях данных, я представляю себе их именно в терминах функции плотности, которая может быть легко построена гистограммой. Отличается ли гистограмма от графического изображения функции плотности? В принципе, конечно, отличается. Но из гистограммы можно сделать подобие кривой плотности путём увеличения объёма выборки и уменьшением длины интервала. Под интервалом стоит понимать ширину столбиков гистограммы, которая получается путём присвоения интервалам гистограммы определённого количества сырых баллов. Под высотой столбиков стоит понимать частоту, с которой встречаются значения данного интервала сырых баллов по выборке.

Рассмотрим, как построить гистограмму в программе Rundom Pro.

Предположим, что у нас есть суммарный балл по тесту скорости чтения n. Мы решили посмотреть, как выглядят полученные результаты.

Открываем программу, создаём новый проект и вставляем в первый столбик (для вставки меню: Edit -> Paste Data) результаты следующих испытуемых:

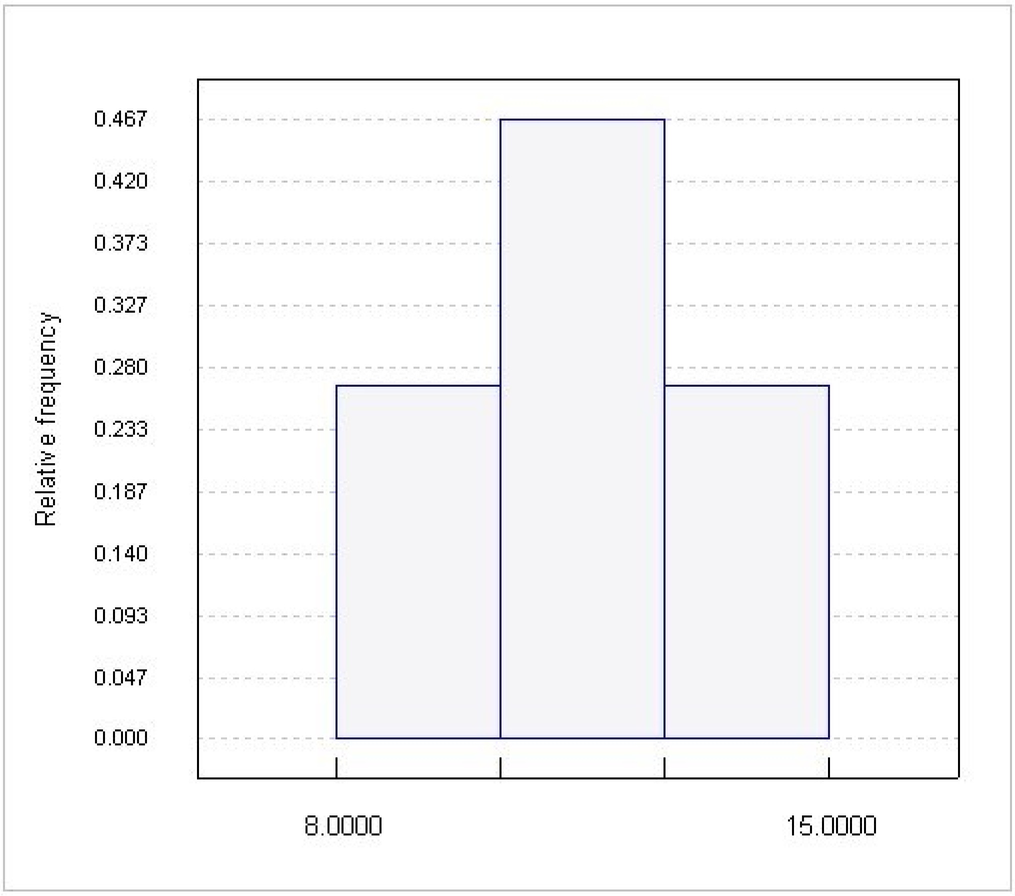
Василий	12
Татьяна	8
Лариса	9
Матвей	11
Георгий	10
Клара	12
Нора	13
Леопольд	11
Раиса	12
Ирина	12
Мария	11
Владимир	10
Михаил	15
Александр	13
Ольга	14

Далее выбираем пункты меню: Graph -> Histogram. Появляется меню, где нужно задать следующие уточнения:

Variable var1 (если данные были скопированы именно в этот столбик). Всё остальное пока что оставляем таким, как оно есть.

К вопросу, что же такое Relative frequency, которое написано в графе Y-axis по умолчанию. Это так называемая накопленная частота.

У нас появляется следующего вида гистограмма:



В некоторых статистических программах, например в STATISTICA, можно также задать опцию, чтобы кривая плотности нормального распределения тоже накладывалась на график. Это делается для визуального определения нормальности распределения, но лично мне этот метод не нравится, поскольку я категорически не доверяют своему глазу. Поэтому помимо графического наложения кривой плотности нормального распределения имеет смысл прибегать к специализированным критериям определения нормальности, например, критерию Колмогорова – Смирнова, Шапиро – Уилка и т.д. Хотя с другой стороны, графический метод как процветал, так и будет процветать, поскольку, несмотря на огромное количество критериев нормальности, «панацеи от всех бед» среди них ещё нет.

В Rundom Pro, судя по всему, такой функции нет.

И плюсом, и минусом диаграммы является относительно произвольное количество интервалов. Для подсчёта оптимального количества интервалов существует несколько методов, самым знаменитым из которых, пожалуй, является формула Стерджесса (Sturgess):

$k = 1,44 \ln n + 1$,

где: k – количество интервалов (классов) в диаграмме, ln n – натуральный логарифм числа элементов в выборке, например, испытуемых. К вопросу о том, как посчитать натуральный логарифм. Самый простой способ – это воспользоваться функцией LN в Excel, доступ к которой можно получить, либо нажав на галочку рядом со знаком «∑», либо через меню: Формулы -> Вставить функцию. Находится натуральный логарифм в категории «математические», либо, если возникают трудности, «полный алфавитный перечень».

В нашем случае было 15 испытуемых, в связи с чем по формуле Стерджесса количество классов в гистограмме должно быть равно: k = 1,44 * 2,70805 + 1 = 4,8995. Естественно, что в данном случае как значение натурального логарифма, так и само значение классов дано в приблизительном значении. Более того, такое количество классов для нас крайне неудобно – мы его округлим и построим диаграмму с 5 классами.

Чтобы построить в Rundom Pro гистограмму с 5 классами, нам необходимо изменить всего одну деталь. В строке # of bins поставить 5. Получается довольно любопытная форма гистограммы, состоящей как раз из 5 классов.

Есть и более сложные подходы к вычислению оптимального количества классов в гистограмме, актуальность их связана с проблемой – не засорить и не упустить изменения. Дело в том, что при слишком малом количестве интервалов мы не заменит важных изменений, а если сделаем количество классов в диаграмме, например, по количеству встречающихся в нашем исследовании значений, рискуем засорить картину слегка отклоняющимися данными.

1.5. Нормальное распределение и однородность дисперсии.

Для человека, хорошо знакомого с математической статистикой, названная подобным образом глава может вызвать лишь раздражение и недоумение, потому что нормальность распределения и однородность дисперсии двух выборок есть совершенно самостоятельные темы. Чтобы объяснить свою иррациональную позицию, начну описание с «горькой правды».

Введение в «горькую правду», выраженное известным происшествием анекдотичного характера:

«Господин Эддингтон, правда ли, что теорию относительности понимают три человека на земле?». После долгого и неловкого молчания: «Я понимаю, Эйнштейн понимает, а кто третий?».

Или другими словами. Кому будет полезна книга по экспериментальной психологии, если в объяснении суммы будет фигурировать определение типа: предела последовательности частичных сумм? А если я скажу, что 2 + 2 = 4 и 4 есть сумма?

Два происшествия анекдотичного характера наводят меня на мысль, что для использования какого-либо метода подсчёта вовсе не обязательно понимать его глубокую суть, гораздо важнее понимать границы его применения. Такое решение как ограничение своей компетенции никогда не даётся легко, но в рамках данной книги оно позволит овладеть предельно простыми навыками, которые впоследствии можно совершенствовать.

Итак, нам придётся отказаться от описания функции нормального распределения случайной величины, потому что для этого нам нужно будет понять, что такое случайное величина и её разновидности, что такое функция, какие распределения бывают, какие нормальные распределения бывают и т.д. Мы изберём более простой путь, который может быть полезен для психологического исследования, но который не идёт в сравнение с истинными знаниями математика.

Случайные величины – зачем они нужны? Если я скажу прохожему на остановке, что занимаюсь наукой об отношениях, то скорее всего он подумает, что я психолог. Ведь под отношениями обычно понимают отношения между людьми. Но более верной позицией было бы считать наукой об отношениях математику. Подразумевая отношения между объектами. Различные отношения и их преобразования могут быть описаны математическими законами. И если бы психология была солидной наукой с многовековой историей, то её работникам приходилось бы учить и использовать точные методы математики, но так как наука наша молода, то используются в основном такие методы, которые относят к более узкой области, а именно – математической статистике. В отличии от точных методов, статистика имеет дело со случайными величинами. В чём же их особенность?

Предположим, что перед нами брат и сестра. Разговор их носит чрезвычайно важный характер. Василий выясняет у Екатерины, когда придёт мама, потому что ему нужно к этому визиту подготовиться. Итак, если бы мы имели дело с математикой, то Екатерина точно бы знала, во сколько ждать родителя. Если же мы в области математической статистики, то Екатерина может дать только приближённый вариант, который будет более вероятным. Этот пример был бы излишним в меру своей простоты, если бы не было одного очень неприметного комментария. Из области вероятности в область точности Екатерина всегда может перейти, если выяснит точные факторы, которые задают вероятность. Другими словами, если знание будет носить минимум неполноты, то вероятность нам не нужна – мы сможем оперировать точными математическими законами. Такой переход даже в рамках психологии возможен, но очень трудоёмок, потому что понять, почему данные, которые по логике вещей должны быть постоянными, флуктуируют, зачастую не удаётся.

Непостоянство величины, природа которого исследователю не известна, делает такую величину случайной.

Мама может в принципе прийти когда угодно. И в 6 вечера, и в 7, и даже в 9 утра, причём частота, с которой мама может прийти в 9 утра следующего дня значительно меньше, чем вечером. Все эти возможные значения и их частоты будут распределением случайно величины.

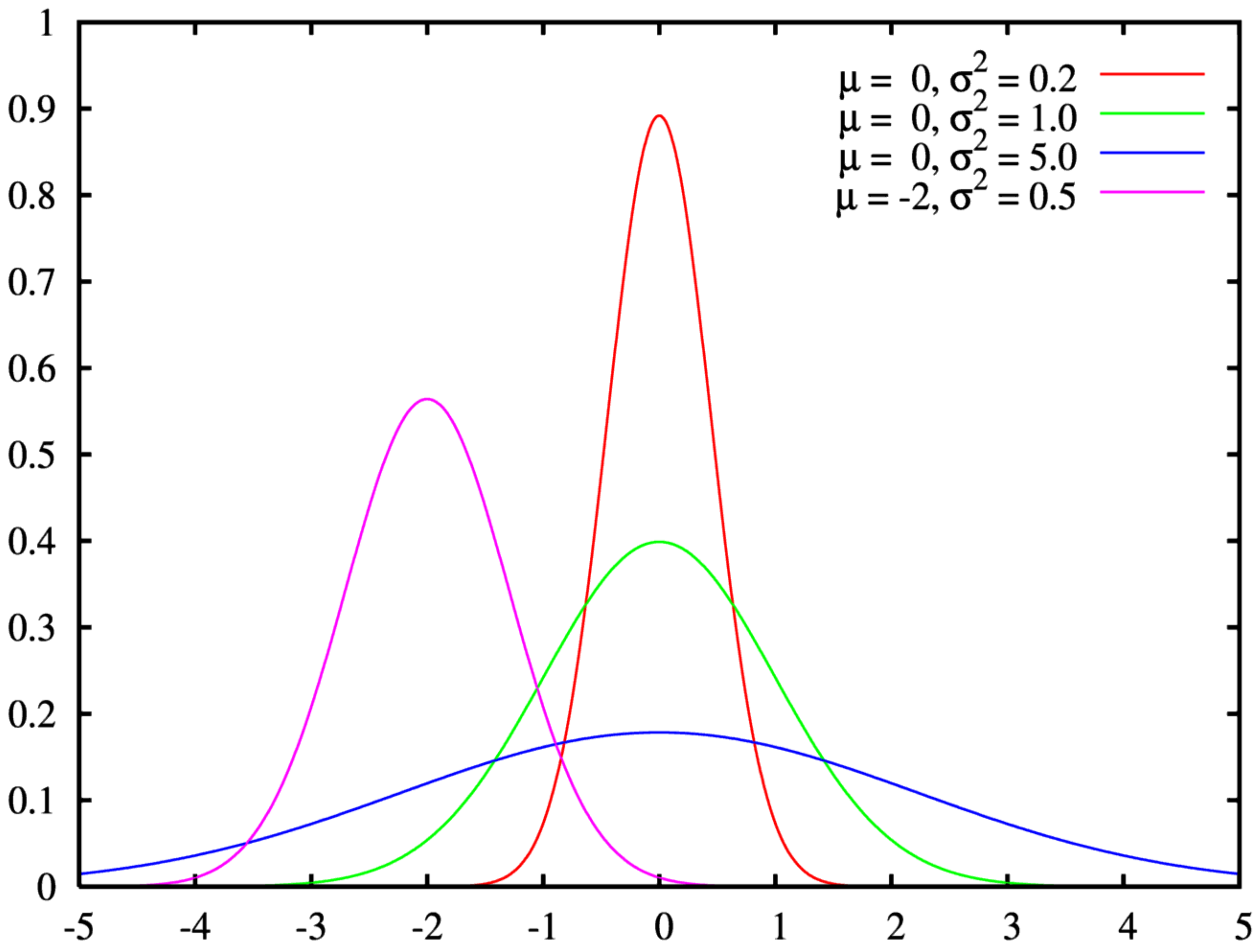
Распределений, как и самих случайных величин, есть множество. Но нас будет интересовать всего два варианта распределений (такая типология свела бы с ума любого математика): нормальное распределение и ненормальное. Потому что когда распределение будет биномиальным, пуассоновским или каким-либо другим, нам это уже будет не важно.

Нормальное распределение по общей традиции считается самым частым в жизни человека. И мало кто осмеливается сказать, что, несмотря на центральную предельную теорему, нормальное распределение данных встречается реже, чем другие. Хотя с точки зрения обывателя это вовсе не удивительно. Потому что если мы будем, например, создавать классификацию животных и скажем, что верблюды – самые часто встречающиеся животные, то даже если их действительно много, мы всё равно скорее всего обречены на неудачу, поскольку разнообразие животного мира всегда будет побуждать людей оспаривать наше высказывание о первенстве верблюдов. Примерно та же ситуация сложилась и с центральной предельной теоремой, которая гласит в самом простом виде следующее: при увеличении выборки мы в итоге получим нормальное распределение данных. Есть огромная масса сторонников этой теоремы, есть масса её доказательств различными методами, но реальность психологического исследования чаще всего касается «ненормальных» данных, что, конечно же, само себе не противоречит предельной теореме. Хотя исследования Орлова А.И. и его выводы сводятся к опровержению вездесущности нормального закона.

Оставив разрешение этой дискуссии профессионалам, я перейду к более простым вопросам.

Нормальное распределение данных можно описать следующим всем знакомым примером: если мы возьмём класс детей, то скорее всего обнаружим небольшое количество очень умных, небольшое количество очень не умных и подавляющее количество средних по успеваемости детей.

Гауссовская колоколообразная кривая или нормальное распределение выглядит следующим образом[13]:



Даже из этого изображения функции плотности нормального распределения мы видим, что оно может быть несколько различным. В психодиагностике очень часто используется стандартное нормальное

[13] Изображение взято с сайта: http://upload.wikimedia.org/wikipedia/commons/1/1b/Normal_distribution_pdf.png

распределение, а именно то, которое на рисунке обозначено зелёным цветом. Стандартное нормальное распределение имеет 0 среднее и 1 стандартное отклонение (как и 1 дисперсию, соответственно). Именно из-за особенности его рассеяния, некоторые исследователи используют термин – единичное нормальное распределение.

Например, стандартное нормальное распределение используется при стандартизации тестов путём выведения стандартизированных баллов, а не процентилей.

Основа процедуры стандартизации баллов выглядит следующим образом:

- проверяется, нормально ли распределены данные;

- в случае подтверждения предположения нормальности, делается линейное преобразование данных к виду стандартного нормального распределения путём создания z-шкалы:

$$z = \frac{x_i - \bar{x}}{\sigma},$$

где:

z – стандартизированное значение балла теста,

x_i – сырой балл теста,

$\bar{x}$ - среднее арифметическое сырых баллов,

σ – стандартное отклонение сырых баллов.

Такая процедура (стандартизации) не меняет свойств исходного распределения, именно поэтому она и называется линейным преобразованием. Но приведение распределения данных к z-шкале, где средним является 0, а стандартным отклонением 1, вовсе не имеет смысла, если исходное распределение данных не является нормальным. Именно поэтому возникает некоторая путаница – с одной стороны, процедура не меняет распределения, с другой – для данных, которые распределены не нормально, не применяется. Но разгадка очень проста – эта процедура приводит распределение данных к такому виду, что имеет строго определённое среднее и стандартное отклонение, а не делает из ненормальных данных нормальные. Поэтому приведение к такому виду данных, если они не нормальны, не даёт никакой пользы и не используется. К тому же, если у нас порядковые данные, то проведение этой процедуры не корректно, поскольку для таких данных не имеет смысла говорить о среднем и стандартном отклонении, в связи с чем в таких ситуациях используется не стандартизированный балл, а процентильные показатели.

Почему же так удобна z-шкала, если она не умеет творить чудес и делать нормальные данные из любых? Исключительно тем, что с помощью такой шкалы удобно переходить от одних стандартных баллов к другим.

Формула перевода сырых баллов в любую другую стандартизированную шкалу выглядит следующим образом:

$$y_i = a\left(\frac{x_i - \bar{x}}{\sigma}\right) + b\ ,$$

где:

y_i – балл, который мы хотим получить по новой шкале;

a – стандартное отклонение, которое мы задали новой шкале;

x_i – сырой балл;

$\bar{x}$ - среднее арифметическое сырых баллов;

σ – стандартное отклонение;

b – среднее новой шкалы.

Например, мы провели тест интеллекта, проверили данные на нормальность и решили создать стандартные баллы IQ (собственно, это-то и делается разработчиками тестов).

В таком случае, предположим, что среднее арифметическое по исследованной нами группе было 55, а стандартное отклонение по этой группе 8. В качестве шкалы, к которой мы будем приводить наши нормальные данные, была выбрана IQ, следовательно, стандартное отклонение новой шкалы 15, а среднее 100.

Для Анечки, у которой сырой балл 49, формула перевода будет выглядеть так:

$$y_i = 15\left(\frac{49 - 55}{8}\right) + 100 = 88{,}75$$

Итак, у Анечки 88,75 баллов IQ. Формулу самой z-шкалы в формуле перевода заметить очень легко – она находится в скобках, а за скобками константы.

Мы рассмотрели лишь один распространенный вариант стандартных баллов – IQ, но не только знаменитых больше, но и удобных исследователю не счесть, ведь такие переводы делаются для того, чтобы подобрать шкалу, которая будет давать более удобный вид баллов. Т.е., например, если бы в результат Анечки мы оставили в z-шкале (-0,75), то это было бы крайне неудобно сравнивать с другими результатами по группе из-за наличия отрицательных показателей и дробных значений.

Среди распространённых стандартных баллов можно также назвать:

Т-баллы: со средним 50 (b) и стандартным отклонением 10 (a);

Стэны: b = 5,5, a = 2;

Станайны: b = 5, a = 2.

В случае же, если данные не являются нормальными, но в цели исследователя входит задача повысить уровень стандартизации, применяется также нелинейное преобразование, когда одно распределение приводится к виду другого. Такая процедура называется нормализацией, хотя стоит признать, что возможна ситуация, когда исходные данные приводятся не к виду нормального распределения, а, например, к виду распределения t-Стьюдента и т.д. К тому же следует иметь в виду, что нормализация не приводит к нормальности распределения, хотя и приводит данные в большинстве случаев к «более нормальному» виду, но об этом мы поговорим после обсуждения разных методов подсчёта.

Сама эта процедура может быть сделана «вручную» путём присвоения исходной частоте встречаемости признака тех значений, которые соответствуют такой встречаемости в нормальном распределении. Есть, как минимум, три способа высчитать нормализованный балл (суть показателя от способа не меняется).

Для подсчёта нам понадобится Excel.

Предположим, что по результатам исследования нового теста интеллекта, мы получили следующий вопиющий ряд сырых баллов:

	сырой ряд
Василий	1

Татьяна	2
Клара	2
Пётр	3
Валя	3
Фёкла	3
Галя	3
Лариса	4
Ольга	4
Ирина	4
Анастасия	4
Максим	4

Из этих сырых баллов мы хотим сделать стандартные IQ баллы. Первое, что нам нужно сделать, это проверить данные на нормальность (как это делается, описано во второй части книги). Поскольку проверка нормальности критерием Шапиро-Уилка позволила нам сделать вывод о том, что данные не подчиняются нормальному распределению, то вышеописанное линейное преобразование в z-баллы для нас закрыто (хотя проверка критерием Колмогорова-Смирнова с поправкой Лиллиефорса позволит всё же принять гипотезу о нормальности, хотя и с «натяжкой» граничных значений). Предположим, что мы всё-таки решили поверить критерию Шапиро-Уилка. В таком случае для получения z-балла мы должны воспользоваться не стандартной процедурой формулы z-балла через среднее и стандартное отклонение, а 3 способами (одним из них, естественно), которые описаны ниже.

<u>1-ый способ получения нормализованных стандартных баллов</u> (описан в пособии П.Клайна «Справочное руководство по конструированию тестов»):

Из ряда сырых баллов мы делаем следующую таблицу:

шаг	Клайн				
1	сырой балл	1	2	3	4
2	частота	1	2	4	5
3	кум частота	0	1	3	7
4	к част серед интерв	0,5	2	5	9,5
5	к ч с интерв /n	0,041667	0,166667	0,416667	0,791667
6	z	-1,73166	-0,96742	-0,21043	0,812218

6 ступеней к победе выглядят следующим образом.

1. Нам нужно рассмотреть сырой ряд баллов и перечислить в первой строке все баллы, которым мы хотим получить соответствие в стандартных z-баллах.

Из этого шага, помимо его сути, виден очевидный вывод – стандартизация имеет смысл в том случае, когда выборка достаточна велика и содержит совершенно разнообразные сырые баллы, а не 2 повторяющихся на весь диапазон значения. В идеальном мире для удовлетворительной стандартизации теста или какой-либо другой исследовательской методики (которую порой и рука-то не поднимется назвать полноценным тестом), требующей наличия стандартных баллов, необходимо от 200 человек выборочной совокупности.

2. Эта ступень отвечает на вопрос, сколько раз встречается данное значение сырого балла в выборочной совокупности данных. Например, мы видим, что 2 сырых балла встречается в выборке 2 раза, потому что 2 балла набрали 2 человека – Клара и Пётр.

3 сырых балла набрали 4 человека: Пётр, Валя, Фёкла и Галя.

3. Кумулятивная (накопленная) частота – это достаточно многозначный термин. И в случае столкновения с ним в том или ином методе подсчёта требуется чрезвычайная внимательность – автор, скорее всего, пояснит, что под этим термином имеется в виду в данном конкретном случае.

В общем смысле кумулятивная частота, кумулятивный процент и прочие кумуляции математики будут означать, что значение следующего члена последовательности равно не просто своему непосредственному значению, но собственному значению с учётом значения предыдущего члена последовательности.

В случае, когда мы рассматриваем кумулятивную частоту в данном случае, это все значения, которые лежат ниже данного сырого балла.

Например, ниже 1 сырого балла значений у нас нет, поэтому кумулятивная частота равна 0. Ниже 2 сырого балла у нас есть всего одно значение, а именно – значение 1-го сырого балла. Заметьте, что кумулятивная частота в данном случае не включённая, то есть значение самого рассматриваемого балла не учитывается.

4. Кумулятивная частота средней точки интервала представляет собой учёт предыдущей невключённости. Т.е. к кумулятивной частоте, которую мы

находили на 3 шаге, прибавляется половина частоты рассматриваемого сырого балла.

Например, для второго сырого балла на четвёртом шаге мы получим: (третий шаг + ½ второго шага = 1 + ½ 2 = 1 + 1 = 2. Для третьего сырого балла на четвёртом шаге мы получим: (третий шаг + ½ второго шага = 3 + ½ 4 = 3 + 2 = 5.

5. Кумулятивная пропорция получается путём деления кумулятивной частоты средних точек каждого сырого балла на общее количество человек n.

Например, для первого сырого балла мы получили значение: 0,5/12 = 0,041667, поскольку в нашей вымышленной выборке было 12 человек.

6. Далее нам остаётся только получить z-балл. Для этого в качестве вероятности мы рассматриваем значение, полученное на шаге 5, и высчитываем односторонний z-критерий.

Благодаря Excel сделать это очень просто: мы запускаем функцию НОРМСТОБР (Формулы -> Вставить функцию -> Полный алфавитный перечень -> НОРМСТОБР), а в диалоговом окне вводим вместо вероятности ячейку, где высчитан шаг 5.

В качестве элемента истории приведу объяснение того, как пользоваться ныне не слишком востребованными (в меру встроенности в Excel) таблицами площади распределения. К слову сказать, именно эта таблица имеется в виду в качестве рекомендации для подсчёта шестого шага в руководстве Клайна.

Выглядеть она может, например, так[14]:

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549

[14] Таблица взята с сайта: http://www.risktheory.ru/distr_tab_normal.htm

0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995

3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4.0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Первый столбец – это целое значение и первый знак дроби стандартного отклонения единичного нормального распределения, а первая строка – это второй знак после запятой. Следующие строки и столбы – это значения, которые соответствуют этим стандартным отклонениям. Так, например, мы можем найти значение 2,59σ – оно будет находиться на пересечении строки 2,5 и столбца 0,09, а значит будет равно 0,9952. Соответственно, таблица не содержит двух необходимых вещей: - отрицательных значений сигмы; - некоторых промежуточных значений так необходимых нам вероятностей. Для этого использовались формулы, которые мы рассматривать не будем, поскольку нет рациональности в том, чтобы пользоваться этими таблицами при наличии Excel.

7. Последний шаг, который нам нужно сделать, чтобы нормализовать данные – это перевести найденный z-балл в стандартную шкалу, которая более приемлема для нас. Мы решили, что это будет IQ, хотя, естественно, что мы могли бы задать любую шкалу.

У IQ, как мы помним, среднее = 100, а стандартное отклонение = 15, следовательно, как и в случае получения z-баллов линейным преобразованием, мы пользуемся формулой:

$$y_i = a\left(\frac{x_i - \bar{x}}{\sigma}\right) + b\ ,$$

получая следующую таблицу соответствия сырых, z и IQ баллов для нашего вымышленного теста интеллекта:

сырой балл	1	2	3	4

z нормализ.	-1,73166	-0,96742	-0,21043	0,812218
IQ	74,02503	85,48868	96,84357	112,1833

2 – ой способ получения нормализованных стандартных баллов (описан в учебнике «Общая психодиагностика» Бодалёва, Столина, Аванесова):

шаг	Бодалёв, Столин, Аванесов				
1	сырой балл	1	2	3	4
2	Частота (P_i)	1	2	4	5
3	кум частота (F_i)	1	3	7	12
4	F_i*	0,5	2	5	9,5
5	PR	4,166667	16,66667	41,66667	79,16667
6	σ	-1,73166	-0,96742	-0,21043	0,812218

В данном случае кумулятивная частота (3-ий шаг) получается с учётом частоты рассматриваемого сырого балла: $F_i = \sum_{j=1}^{i} P_{ji}$.

Например, для 4-го сырого балла частота будет высчитываться не как «7» (предыдущий метод), а как «12».

На четвёртом шаге высчитывается F* - кумулятивный балл:

$$F_i^* = F_i - \frac{1}{2} P_i$$

Так, например, мы видим, что для второго сырого балла $P_i = 2$, а $F_i = 3$, следовательно, $F_i^* = 3 - ½\ 2 = 2$.

Процентильный ранг (PR) равен:

$$PR_i = F_i^* \cdot 100 / n \text{ .}$$

N – это количество человек.

И, наконец, в строке σ даётся значение нормализованного или z-балла, которое можно находить либо по таблицам, либо с помощью функции НОРМСТОБР. В случае, если мы находим значение сигмы функцией Excel, необходимо на шаге 5 изменить формулу, а именно – разделить F_i* на количество человек, не умножая на сто. В таком случае вероятность будет дана в пределе от 0 до 1 и нам не придётся в уме делить значение PR снова на сто, чтобы задать нужную вероятность в Excel – достаточно просто указать ячейки PR и нажать НОРМСТОБР.

3-ий способ (самый простой):

Я привела предыдущие два способа подсчёта нормализованных баллов по двум причинам: во-первых, поскольку они описаны в известных руководствах по психодиагностике, исследователь может встретиться с подобными же описаниями в любых других книгах и отчётах о процедуре проведения какого-нибудь исследования; во-вторых, эти два метода хорошо отражают суть того, как получаются нормализованные баллы.

Третий способ я лично не встречала в описаниях (хотя, естественно, что он где-то написан), но придумала его, поскольку он более лёгок для запоминания в плане включённых в него шагов подсчёта. По-моему, достаточно всего один раз посчитать данные таким образом, чтобы запомнить все необходимые действия.

сырой ряд	R	(R-0,5)/n	z
1	1	0,041667	-1,73166
2	2,5	0,166667	-0,96742
2	2,5	0,166667	-0,96742
3	5,5	0,416667	-0,21043
3	5,5	0,416667	-0,21043
3	5,5	0,416667	-0,21043
3	5,5	0,416667	-0,21043
4	10	0,791667	0,812218
4	10	0,791667	0,812218
4	10	0,791667	0,812218
4	10	0,791667	0,812218
4	10	0,791667	0,812218

Шаг первый. Ввести просто сырой ряд данных.

Шаг второй. Присвоить этому ряду ранги с учётом повторений. Ранги с учётом повторений – это довольно часто встречающаяся процедура, которая получается следующим образом: если два или более сырых значения повторяются, то берётся среднее арифметическое их рангов и присваивается каждому, поскольку у нас нет достаточной информации, чтобы выстроить между ними приоритеты.

Так, например, сырой балл «3» встречается у нас 4 раза, занимая по возрастанию позицию с 4 по 7. В таком случае ранг с учётом связки будет равен среднему рангу: (4 + 5 + 6 + 7)/4 = 5,5.

Третий шаг хорошо отражён в собственном названии – нужно из полученного ранга отнять 0,5 и разделить получившееся на количество человек в выборке, т.е. на 12 в нашем случае.

Наконец, последний шаг заключается в том, что полученное значение нужно использовать в качестве вероятности в функции НОРМСТОБР.

Поскольку мы научились делать нормализованные стандартные баллы или нормализацию по составу, то естественно перед нами возникает вопрос, является ли преобразованное распределение нормальным? Чтобы увидеть различие между линейным и нелинейным преобразованием во всей красе, я приведу несколько показателей (задача остаётся прежней – новый тест интеллекта), а именно:

	сырой ряд	z-балл	IQ из z	норм z-балл	IQ из норм z
Василий	1	-2,09127	68,63095	-1,73166	74,0251
Татьяна	2	-1,08746	83,6881	-0,96742	85,4887
Клара	2	-1,08746	83,6881	-0,96742	85,4887
Пётр	3	-0,08365	98,74525	-0,21043	96,84355
Валя	3	-0,08365	98,74525	-0,21043	96,84355
Фёкла	3	-0,08365	98,74525	-0,21043	96,84355
Галя	3	-0,08365	98,74525	-0,21043	96,84355
Лариса	4	0,92016	113,8024	0,812218	112,18327
Ольга	4	0,92016	113,8024	0,812218	112,18327
Ирина	4	0,92016	113,8024	0,812218	112,18327
Анастасия	4	0,92016	113,8024	0,812218	112,18327
Максим	4	0,92016	113,8024	0,812218	112,18327
Колмогоров-Смирнов (d)	0,238	0,238	0,238	0,252	0,252
Колмогоров-Смирнов (p)	0,059	0,059	0,059	0,034	0,034
Шапиро-Уилк (W)	0,84	0,84	0,84	0,849	0,849
Шапиро-Уилк (p)	0,028	0,028	0,028	0,035	0,035

Таблица устроена следующим образом. Столбик «сырой ряд» представляет собой баллы испытуемых, которые они набрали «как есть». Соответственно, четыре нижние строчки – это эмпирические значения критерия и уровни значимости (p) критериев нормальности Колмогорова-Смирнова (d) и Шапиро-Уилка (W). Второй столбик – z-балл – представляет собой стандартный балл, получаемый через формулу среднего и стандартного отклонения. IQ из z – это столбик, который показывает, как будут выглядеть IQ баллы, если высчитывать их из стандартного z. «Норм. Z-балл» - это столбик, в котором посчитано значение каждого z балла, исходя из распространённости этой частоты в нормальном распределении. «IQ из норм. z» есть IQ балл, полученный тем же способом, что и в третьем столбике, но с той лишь разницей, что в формуле перевода используется не стандартный z-балл, а нормализованный.

Какие выводы мы можем сделать из этой таблицы?

1) ни перевод в стандартный z балл, ни последующий перевод в IQ не меняет форму распределения, в связи с чем значимость критериев нормальности для первых трёх столбиков данных одинаковая;

2) движение по пути к нормальности носит в данном случае странную траекторию. С одной стороны, критерий Шапиро-Уилка ведёт себе точно так, как мы и ожидали. Наши нормализованные баллы становятся ближе к нулевой гипотезе о нормальности, нежели исходный сырой ряд данных. С другой стороны, критерий Колмогорова-Смирнова ведёт себя на первый взгляд совершенно странно – при нормализации данные, по его мнению, стали более ненормальными. Чтобы понять причины этого явления можно, конечно, создавать эксперименты Монте-Карло; внимательно изучать исходные распределения, ведь в зависимости от формы ненормального распределения во многом зависит, как поведут себя эти два критерия нормальности, т.е. против какой альтернативы они выступают. Но мы не будем этого делать, мы сделаем более простой шаг. Создадим более нормальное распределение сырых баллов, нежели наш новый тест интеллекта, и менее нормальное. Посмотрим, к чему же это приведёт.

Более нормальное распределение будет отражено следующими результатами:

	сырой ряд	z-балл	IQ из z	норм z-балл	IQ из норм z
Василий	1	-1,63707	75,44395	-1,7316644	74,02503
Татьяна	2	-1,07579	83,86315	-0,96742157	85,48868
Клара	2	-1,07579	83,86315	-0,96742157	85,48868
Пётр	3	-0,51451	92,28235	-0,4307273	93,53909
Валя	3	-0,51451	92,28235	-0,4307273	93,53909
Фёкла	4	0,04677	100,7016	-1,3921E-16	100
Галя	4	0,04677	100,7016	-1,3921E-16	100
Лариса	5	0,60805	109,1208	0,54852228	108,2278
Ольга	5	0,60805	109,1208	0,54852228	108,2278

Ирина	5	0,60805	109,1208	0,54852228	108,2278
Анастасия	6	1,16933	117,54	1,15034938	117,2552
Максим	7	1,73061	125,9592	1,7316644	125,975
Колмогоров-Смирнов (d)	0,145	0,145	0,145	0,13	0,13
Колмогоров-Смирнов (p)	0,2	0,2	0,2	0,2	0,2
Шапиро-Уилк (W)	0,97	0,97	0,97	0,98	0,98
Шапиро-Уилк (p)	0,916	0,916	0,916	0,982	0,982

Поскольку мы используем критерий Колмогорова-Смирнова с поправкой Лилиефорса, то ориентироваться следует не по уровню значимости (поскольку в случае принятия нулевой гипотезы SPSS, в котором и вёлся подсчёт, выдаёт значение нижней границы истинной значимости – (0,2)), а по самому эмпирическому значению критерия (d). Чем выше эмпирическое значение критерия Колмогорова-Смирнова, тем менее вероятна нулевая гипотеза, а значит для нашей задачи нужно, чтобы движение эмпирического значения критерия от сырого балла (и стандартного z-балла) к нормализованному z-баллу было в сторону уменьшения (в то время как уровень значимости критерия Колмогорова-Смирнова должен наоборот уменьшаться от стандартного z к нормализованному). К вопросу о том, почему бы не посчитать критерий Колмогорова-Смирнова без этой поправки, то это не корректный путь для нашей задачи, поскольку генеральное среднее и стандартное отклонение мы не задаём, а высчитываем из выборочных данных, которые собрали.

Как мы видим из этой таблицы, оба критерия нормальности ведут себя запланированным способом – нормализация сделала данные более нормальными.

Рассмотрим результат нормализации данных, которые изначально были ненормальными.

	сырой ряд	z-балл	IQ из z	норм z-балл	IQ из норм z
Василий	1	-1,04536	84,3196	-1,73166	74,02503
Татьяна	2	-0,99784	85,0324	-1,15035	82,74476
Клара	3	-0,95033	85,74505	-0,81222	87,81673
Пётр	5	-0,85529	87,17065	-0,54852	91,77217
Валя	6	-0,80778	87,8833	-0,31864	95,22041
Фёкла	7	-0,76026	88,5961	-0,10463	98,4305
Галя	25	0,09503	101,4255	0,104633	101,5695
Лариса	37	0,66523	109,9785	0,318639	104,7796
Ольга	40	0,80778	112,1167	0,548522	108,2278

Ирина	49	1,23542	118,5313	0,812218	112,1833
Анастасия	50	1,28294	119,2441	1,150349	117,2552
Максим	51	1,33046	119,9569	1,731664	125,975
Колмогоров-Смирнов (d)	0,276	0,276	0,276	0,044	0,044
Колмогоров-Смирнов (p)	0,012	0,012	0,012	0,2	0,2
Шапиро-Уилк (W)	0,817	0,817	0,817	0,998	0,998
Шапиро-Уилк (p)	0,015	0,015	0,015	1	1

Из этой таблицы мы опять же видим, что и Шапиро-Уилк, и Колмогоров-Смирнов ведут себя ожидаемым способом – нормализация делает данные нормальнее. Что мы получили? Чаще всего нормализация приводит распределение к более нормальному виду, хотя это не значит, что данные стали нормальными, они стали нормальнее. Более того, мы выяснили, что такая радужная картина возникает далеко не всегда – первый пример, на котором мы и учились считать нормализованный z-балл, отлично показал невсеобщность этого правила.

На этом я заканчиваю рассмотрение нормализованных стандартных баллов. Однако, стоит понимать, помимо прочего, что это жёсткая процедура, которая лично у меня вызывает больше вопросов, нежели ответов, поскольку, «ломая» распределение, мы можём впасть в ошибку – если исходное распределение не является нормальным в связи с тем, что его природа именно такова, то попытка сделать данные более нормальными есть варварство над естеством феномена. Если же отклонение от нормальности вызвано неточностью измерения или недостаточно большой выборкой, то нормализация – это вполне достойный путь повысить уровень стандартизации. Но вопрос о природе распределения и причинах отклонения может быть настолько сложен, что быть может имеет смысл без точной уверенности остановиться на процентильной стандартизации и не ломать распределение.

Другой областью применения нормального распределения в психодиагностике является определение уровневых норм теста. Мы уже рассматривали, как задать уровневые нормы, используя процентили. Для задания таких норм через стандартные отклонения, если наши данные подчиняются нормальному закону, также не составит труда. Например, мы предположим, что «средний уровень» наших норм должен охватывать 50 % испытуемых, ниже этого уровня будет находиться «низкий уровень», а выше – «высокий». В таком случае при нормальном распределении данных «средний уровень» или коридор среднестатистической нормы будет составлять 2/3σ в обе стороны от математического ожидания. Например, IQ баллы имеют коридор нормы от 90 до 110 IQ (напоминаю – среднее – 100, а стандартное отклонение 15, в связи с чем 2/3σ это 10).

Чтобы задать нормы таким способом, отсекая те коридоры, которые нужны исследователю, используется таблица стандартных отклонений единичного нормального распределения с указанием процента выборки, попадающего в эту область.

Самые используемые значения следующие:

±1σ – около 68,27 % случаев;

±1,65σ – около 90,10% случаев;

±1,96σ – около 95% случаев;

±2,58σ – около 99% случаев;

±3σ – 99,73% случаев.

Из приведённых значений мы, сами того не ожидая, уже видим, в чём же заключается знаменитое правило трёх сигм. Три сигмы отсекают 99,73 % случаев, что и придаёт этому значению стандартного отклонения единичного нормального распределения такую славу. Однако, если задуматься, то 3 сигмы – это довольно необычное правило, потому что несмотря на «намёк», что это практически все испытуемые выборки, это всё же не все испытуемые. А тогда получается, что с тем же успехом мы могли бы сделать знаменитым правило «три и три (3,3) сигмы», подразумевая, что в диапазон 3,3σ попадает около 99,9% случаев. Но как бы там ни было, известны лишь три сигмы, вследствие чего не стоит удивляться при встрече с ними. Хотя, нужно отдать должное этому правилу (3σ) – это единственный случай, когда я не говорю, что значение примерно такое – оно точно такое.

Что же делать, если необходимо узнать, сколько стандартных отклонений от среднего отсекает тот или иной процент случаев, а значения в данном пособии такого нет.

Исследователи, которые возможно и не ценят своего счастья, но, тем не менее, обладают прекраснейшей программой – STATISTICA, могут позволить себе воспользоваться истинным шедевром современных технологий – вероятностным калькулятором. Он может находиться в меню основных статистик и называться Probability Calculator. И, несмотря на то, что данное пособие построено по принципу эконом - класса и описывает лишь бесплатное программное обеспечение, я всё же уделю пару слов этой функции, потому что если не уметь ей пользоваться, то она точно будет пылиться в закоулках. Чтобы решить обозначенную выше задачу, а именно – определить количество единиц отклонений от среднего для того или иного процента случаев, необходимо открыть в вероятностном калькуляторе пункт z (normal) и заполнить всего несколько ячеек. Нужно поставить галочку на «двойном критерии», почему, я объясню чуть позже. Отметить в качестве вероятности (p) необходимое количество случаев, например, для проверки, можно взять значение 0,95, т.е. под этим мы понимаем, что хотим узнать, сколько стандартных отклонений в обе стороны отсекут 95 % случаев. Нажать Enter или Подсчёт и посмотреть в графу X. Для 95% случаев необходимо 1,959964 отклонения в обе стороны, можно округлить до 1,96, что мы и запоминали из приведённых распространённых значений.

Если по каким-либо причинам нужно задать не единичное, а какое-нибудь другое нормальное распределение, то достаточно просто сменить заданное по умолчанию значение стандартного отклонения с 1 на какое-нибудь другое, а также сменить значение со среднего (0 по умолчанию) на другое.

Если требуются другие распределения, то и это есть в вероятностном калькуляторе. Например, при подсчёте корреляции Спирмена используются критические значения критерия t-Стьюдента. Можно сравнить те данные, которые выводятся по t-Стьюденту во время подсчёта корреляции с теми вероятностями, которые выдаст калькулятор. Совершенно бесполезно, поскольку информация будет одной и той же, однако приятно делать хоть что-то своими руками. Комментарий: подсчёт вероятностей Стьюдента в этом калькуляторе носит уже несколько более сложный характер, нежели подсчёт вероятностей для нормального распределения, ведь нужно задать степени свободы (которые вычисляются как n-2), задать двухсторонность, а также указать галочкой величину (1-p).

Но мы несколько отклонились от темы. Для вычисления количества σ нормального распределения мы воспользуемся Excel. Однако, нам нужно понять, чем односторонний критерий отличается от двухстороннего. Односторонний критерий проверяет только одну сторону распределения относительно среднего, поскольку гипотеза исследователя задаётся умышленно точно. Например, если мы сравниваем две группы на предмет различий и решаем, что мы уверены, что первая группа получила результаты более высокие, чем вторая, но не знаем, значимо ли это статистически, то можно воспользоваться односторонним критерием, поскольку он более «мягкий», а значит мы не отвергнем гипотезу «зазря». Но мне кажется, что такой «мягкостью» стоит пользоваться тогда, когда уверенность исследователя обоснованна настолько, что одна лишь мысль о сомнении вызывает оскорбление. Часто ли случаются такие уверенные исследования? Нет. Поэтому зачастую стоит более критично подходить к данным и использовать ненаправленные гипотезы, которые проверяют распределение с обеих сторон.

Совсем другой случай использования одностороннего z-критерия – это получение нормализованных стандартных баллов, которые мы рассматривали чуть выше. В этом случае совершенно логично было рассуждать не от среднего, а слева на право.

Что же касается конкретной задачи, то нам вообще не нужен односторонний критерий, поскольку наша цель нуждается в «коридоре». Чтобы научиться задавать тот или иной процент случаев, необходимо научиться преобразовывать односторонний z-критерий в двухсторонний.

Итак, односторонний z-критерий высчитывает вероятность как: $(1 – \alpha/2)$.

Двухсторонний как: $(1 – \alpha)$.

Excel, которым мы и будем для этого пользоваться, высчитываем только односторонний критерий. По крайней мере в той версии, которой располагаю я. Возможно, что более поздние версии учтут эту обидную деталь.

В таком случае, нам необходимо вручную вычислить, чему будет равна вероятность одностороннего критерия для той вероятности двухстороннего критерия, которую мы хотим получить.

Если мы, например, хотим, вычислить, чему равно z для вероятности p = 0,95 у двухстороннего критерия путём подсчёта одностороннего, то предпринять необходимо следующие шаги:

1. Вычислить $\alpha = (1 – p) = (1 – 0{,}95) = 0{,}05$.

2. Разделить α, чтобы получить $\alpha/2$, т.е. $0{,}05/2 = 0{,}025$.

3. Вычислить p для одностороннего критерия, а именно $p = (1 – \alpha/2) = 1 – 0{,}025 = 0{,}975$.

Итак, чтобы вычислить z для 95% случаев двухстороннего критерия, необходимо вычислить z для 97,5% случаев одностороннего. Z будет одним и тем же при этих двух значениях вероятности, в связи с чем мы получим желаемый результат.

Для того, чтобы вычислить z для одностороннего критерия при заданной вероятности, используется функция Excel, которая называется НОРМСТОБР (мы ею уже пользовались для присвоения сырому баллу значение нормализованного стандартного). Мы просто ставим курсор на любой понравившейся пустой ячейке, запускаем меню функций, находим необходимую нам и задаём в графе Вероятность значение 0,975. Получаем: 1,959964, что примерно равно 1,96.

Таким образом, отложив 1,96 стандартных отклонений в обе стороны, мы получим коридор, в который попадут 95% испытуемых.

Та же самая идея о нормальном распределении используется при нахождении доверительного интервала измерения n-теста. Совершенно очевидной представляется ситуация неточности психологического измерения, что побуждает исследователей искать пути выхода из этой ситуации. Заложив в основу измерения понятие об истинной дисперсии и дисперсии ошибки, исследователи вывели не только представления о надёжности теста как точности измерения, но о стандартной ошибке измерения как эквиваленте надёжности измерения, воплощённой в конкретном психодиагностическом заключении для одного-единственного испытуемого (тестируемого). И если о надёжности теста (как неотъемлемой части психометрической оценки теста) в данном пособии мы говорить не будем, поскольку это уже область дисциплины психологической диагностики, то ошибку измерения затронем, поскольку она может быть использована в абсолютно любом исследовании, использующем тестовый подход, а не только в исследовании, направленном на конкретную задачу проверки психометрических характеристик теста.

Предположим, что мы решили понять, каков интеллект Анечки на самом деле. Совершенно очевидным является факт, что эмпирически зафиксированное значение 112 IQ не является окончательным вердиктом, который никогда не изменится. Даже если не брать в расчёт такое явление как прогресс/регресс, другие, более случайные, факторы также повлияли на точность нашего измерения. Если бы мы представили себе ситуацию, что проводим тест интеллекта на этой бедной девочке несколько десятков раз в относительно небольшой срок (и пусть даже не учитываем эффект тестирования), то эмпирическое значение показателя всё равно будет флуктуировать. Чтобы избежать этой неприятной действительности заранее, не проводя десятка тестирований, но при этом учесть неточность полученного результата, достаточно высчитать стандартную ошибку измерения и задать интервал.

Стандартная ошибка (SEM) обычно даётся в руководствах тестов и вычисляется по формуле[15]:

$$SEM = SD_t\sqrt{1 - r_{tt}},$$

где:

SD_t – стандартное отклонение показателей теста,

r_{tt} - коэффициент надёжности теста.

Например, если надёжность теста n в руководстве разработчиков равна 0,95, а показатели теста даны в Т-баллах (у которых стандартное отклонение всегда равно 10 баллам), то SEM будет равна: $10\sqrt{1-0{,}95} = 2{,}23$.

Соответственно, если нет данных о надёжности и стандартном отклонении; нет вообще никаких данных о тесте, поскольку вы создали его сами, то настоятельно рекомендуется обратиться к пособиям, раскрывающим вопросы психометрических характеристик.

Чтобы задать доверительный интервал, в связи с которым рассказ о стандартной ошибке измерения и помещён в эту главу, используются идеи о заданной вероятности. Насколько вы хотите быть уверены в том, что «уловили» истинный уровень интеллекта нашей бедной Анечки? На 50 %? Я или уверен, или не уверен. Может быть на 70? С вами всё будет хорошо в 7 случаев из 10. Вообще, для этой задачи, как и для большинства других, используются общепринятые нормы 0,95 и 0,99 вероятности. Причём, 0,95 в этом случае даже более приемлема, поскольку помимо ложного отвержения гипотез может быть ложное принятие. Чтобы задать 95-ый доверительный интервал измерения необходимо умножить SEM на 1,96. Таким образом, интеллект Анечки находится в пределах: от (112 – 1,96*2,23) до (112+1,96*2,23), т.е. от 107,6292 до 116,3708. Естественно, что такая точность до десятитысячных не требуется, я просто привожу точное вычисление, чтобы читатель не усомнился в правильности своего.

[15] Формула взята из книги: А.Анастази, С.Урбина «Психологическое тестирование» - Спб.: Питер, 2009

Если по каким-либо причинам необходимо задать 99-ую вероятность, то умножать значение стандартной ошибки измерения нужно на 2,58.

Бывают случаи, когда требуется сравнить два доверительных интервала. Например, в случае, если мы имеет несколько субтестов интеллекта у одного и того же испытуемого. Естественно, что в таком случае нам требуется выполнить два условия:

- шкалы измерения должны быть сопоставимыми – IQ с IQ, T с T, чего, как мы помним, легко достичь, используя методы преобразования шкал, описанные выше;

- ошибка измерения должна учитывать случайные факторы, повлиявшие как на эмпирическое значение по первой шкале, так и по второй.

Для учёта факторов обеих шкал вместо стандартной ошибки измерения используется формула стандартной ошибки различий:

$$SE_{diff.} = \sqrt{(SEM_1)^2 + (SEM_2)^2} = SD\sqrt{2 - r_1 - r_2},$$

где:

$SE_{diff.}$ – стандартная ошибка разности показателей,

SEM_1 и SEM_2 – стандартные ошибки измерения первой и второй шкалы,

SD – стандартное отклонение, которое одинаково для обеих шкал из-за предварительной подготовки к сопоставимости,

r_1 и r_2 – надёжность первой и второй шкалы.

Соответственно, эти две формулы подсчёта стандартной ошибки разности эквивалентны. После подсчёта этого показателя используется тот же принцип подсчёта доверительного интервала, что и просто для стандартной ошибки измерения (через константы z).

Два любопытных вопроса о SEM:

- Во-первых, не стоит путать SEM и SEM. Один вариант расшифровки аббревиатуры мы сейчас рассмотрели: Standard Error of Measurement. Совершенно другой вариант расшифровки: Standard Error of Mean. И англоязычные жители планеты поначалу очень мучаются с этим недоразумением.

Стандартная ошибка среднего высчитывается совершенно по-другому, отражая точность вычисления среднего:

$$SEM = \frac{\sigma}{\sqrt{n}},$$

где:

SEM – стандартная ошибка среднего,

σ – стандартное отклонение выборочной совокупности,

n – количество человек.

- Во-вторых, несмотря на то, что в очень многих учебных пособиях дана формула подсчёта стандартной ошибки измерения, редко встречается уточнение, о какой именно надёжности идёт речь. В целом, в качестве коэффициента надёжности теста может быть использована как надёжность частей, так и ретестовая, параллельных форм и т.д. В зависимости от того, какая задача стоит перед исследователем – например, использовать тест дважды (в таком случае при подсчёте SEM под надёжностью понимать ретестовую), использовать два разных теста (использовать надёжность параллельных форм) или какая-либо другая задача, может быть использован более приемлемый и логичный для данного конкретного случая вид надёжности[16].

Естественно, что перечислить все методы, основанные на идеях нормальности распределения, не входит в задачи данного пособия. Однако, есть ещё один крайне важный метод, которым мы и завершим данное рассмотрение. В более широком контексте, а именно в контексте экспериментальной психологии, нормальное распределение полезно и по другим причинам. При обработке данных исследования крайне важно знать, является ли распределение нормальным, поскольку ответ на этот вопрос позволяет выбрать между двумя группами статистических критериев – параметрическими и непараметрическими.

Параметрическая статистика достаточно часто предъявляет требование не только нормальности распределения, но и однородности дисперсий, что и побудило меня подчеркнуть эти требования названием главы. Например,

---

[16] Ознакомиться с довольно подробным описанием различных аспектов SEM можно в статье Лео Харвилла, размещённой по адресу: http://www.ncme.org/pubs/items/16.pdf

критерий t-Стьюдента может быть применён только после проверки этих двух условий.

Для проверки нормальности может быть использовано большое количество специализированных критериев, среди которых особенно стоит отметить критерий Шапиро - Уилка. Несмотря на различность условий, в которых тот или иной критерий нормальности лучше справляется со своей задачей, а именно точнее принимает решение о том, нормальны ли данные, критерий Шапиро – Уилка занимает одно из лидирующих мест как в отечественной, так и в зарубежной литературе.

Однородность дисперсии была бы уместна скорее в главе об описательных статистиках, поскольку именно там стоит давать определение дисперсии, но всё же хотелось бы сказать, что ответ на вопрос – однородны ли дисперсии групп – решается по-разному. В самом простом варианте считается, что изменчивость данных однородна в случае равного количества людей в двух сравниваемых группах. Но такой подход на мой взгляд не слишком оправдан, учитывая тот факт, что проверка однородности в современных статистических программах занимает около двух минут. Поэтому всяческих априорных предположений, неоправданных ничем кроме надежды и общей логики, нужно избегать. Для этого случая рассматривается два критерия: Левена и Брауна-Форсайта.

Описание, данное в этой главе, носит, естественно, ультракраткий характер, но главное, что должен уяснить читатель – это лишь то, что два вопроса, а именно:

- нормальность распределения может быть использована в психодиагностических целях;

- нормальность распределения данных и однородность дисперсий, необходимо проверять прежде, чем использовать параметрические критерии проверки статистических гипотез.

1.6. Шкалы измерения и отношение.

Прежде чем непосредственно перейти к заявленной теме, нам необходимо рассмотреть более широкий контекст использования математического аппарата в психологической науке.

Многие студенты, да, пожалуй, и не только они, впадают в недоумение, когда сталкиваются с огромным количеством совершенно разнообразных измерений и подсчётов, используемых в ГУМАНИТАРНЫХ науках. И если после нескольких минут раздумий становится совершенно очевидно, что измерением мы обязаны научности подхода, то как же быть с числами и статистическими процедурами? Поэтому длинный разговор, завершающийся откровенно пугающими описаниями программ и алгоритмов подсчёта, мы начнём с предельно очевидных уроков математики для первоклассника психологической школы.

Что же такое число в обычном его понимании? Поскольку класс, скептически ухмыляясь, молчит, предлагаю рассмотреть рисунок.

Искренне надеюсь, что каждый из нас, глядя на эту картинку, видит две вишни. Можно спорить скорее о том, является ли данное изображение воплощением вишни. Может быть черешни? Но о том, что этих предметов два… сомнительная софистика. В этом есть неоспоримое преимущество математического аппарата.

Сколько же вишенок на следующей картинке?

Насколько я понимаю, на этой картинке вишенок уже 4. Если мы будем рассуждать как истинные дегустаторы прекрасного дара, то будет ли отличаться вкус каждой из вишен, если за условие мы возьмём, что в корзине их было 4? А если в корзине изначально было 5 вишенок, то будет ли отличаться вкус каждой из них на этот раз? Очевидно, что количество не накладывает отпечаток на вкус каждой отдельно взятой вишенки. Если вкус и отличается, то в силу других факторов генетики и воспитания, которыми мы будем пренебрегать в условиях данной задачи.

Ни вкус, ни цвет, ни форма никак не зависят от того, рассматриваем ли мы корзинку, где лежат 4 вишенки, или же 5. Прибавление и вычитание количества также не даст эффекта. Тогда получается, что количество не выражает непосредственного свойства предмета. Что же в таком случае изменяется, когда мы говорим о том, что у нас было 4 вишни, а стало 5? Количество есть не атрибутивное свойство, «прописанное» в материале вишни, а свойство системное, выражающее отношение элементов внутри некоего единого целого. Для начала стоит отметить, что целое, накладывающее границы на наш подсчёт, носит довольно произвольный характер, напрямую отвечающий задаче исследователя. В большинстве случаев мы ничего не знаем о поле опыта целиком, например, мы не знаем, сколько вишен вообще на земле было или будет. Но нам зачастую это и не нужно. Нам достаточно задать систему отсчёта, которой и будет ограничиваться всё рассуждение. Будь то хоть корзинка, хоть выборка, хоть капелька воды. Какой смысл изучать одну-единственную корзинку? Подспудно предполагая, что все корзинки одинаковые, мы с лёгкостью можем рассуждать о поле опыта целиком, изучая лишь капельку океана.

Итак, когда мы говорим, что в корзинке 5 вишенок, мы имеем в виду лишь такое целое, что каждая из этих вишенок имеет конкретно специфическое отношение к своим соседкам. Исходя из того, что число есть отношение между предметами, уже становится легко предположить, что эти самые отношения могут разнообразными. Под разнообразием можно понимать как однотипное разнообразие – 4, 5, 1000… вишенок, так и разнообразие, полученное путём смыслового уточнения отношений.

Разные по смыслу отношения между предметами закладывают классификацию шкал измерения. Ведь шкала измерения есть не что иное, как попытка упорядочить изучаемые предметы между собой. Это построение лестницы, где ступеньками становятся наши любимые испытуемые, вишенки или другие явления.

Прежде чем назвать отношения и шкалы, зададимся вопросом, откуда эти отношения берутся? Как ни странно, дав нам право жить и пользоваться благами мира, Бог ничего не сказал о том, как же этим миром пользоваться. Именно поэтому рациональность взяла верх в жизнедеятельности человека и объяснение окружающего мира с помощью каких бы то ни было средств есть любимое хобби практически каждого человека. Мы изучаем мир, всё больше (по крайней мере, нам так кажется) постигая его глубинные механизмы. Знание об n-предметах, которое с течением времени имеет тенденцию увеличиваться и уточняться, неизбежно начинает «путешествовать» по шкалам измерения.

Рассмотрим вымышленную и абсурдную на данный момент ситуацию. Итак, мужчин и женщин мы можем дифференцировать между собой исключительно в номинальной шкале (несколько не понятно, если термин не знаком, но далее размышление окончательно прояснится). Т.е. более чем сказать, что это – мужчина, а это – женщина, мы и не можем. Предположим, что гениальный и немного странный учёный обнаружил, что мужчина и женщина – это довольно грубое деление. На самом деле каждый мужчина – это более мужчина, но ещё и женщина; а каждая женщина – это более женщина, но чуть-чуть и мужчина. Таким образом, мы получили континуум, где пол является лишь полюсами. Предположим, что этот странный учёный изобрёл прибор, который позволяет по биохимии крови определять пол более дифференцированно, т.е. если загрузить 20 данных крови, то прибор сможет определить, что 1 испытуемый более мужчина, чем второй и т.д. Мы получили порядковую шкалу. Ещё через 10 лет этот странный учёный смог усовершенствовать свой прибор таким образом, что ввёл эталоны пола и единицы измерения, что позволяет измерять пол в G-баллах. Теперь довольный собой молодой человек сможет сказать любимой девушке, что у него 150 G-баллов, а девушка в ответ, улыбнувшись, скажет, что у неё 200 G[1]-баллов. Шкала получилась метрической.

Такой вымышленный пример показывает нам, что знание о чём-либо, уточняясь, позволяет совершать всё более сложные операции над отношениями. Не стоит думать, что это редкое явление – одна классификация, наложенная на другую, позволяет сделать изначально номинальные переменные, которые есть возможность только назвать разными именами, переменными более высокого порядка. Именно так получилась измеримая в подлинном смысле классификация цветов, когда каждый цвет имеет не только своё уникальное название, но и длину волны.

Итак, шкала измерения говорит нам какую-то информацию об отношениях между предметами. Можно придумать большое разнообразие различных отношений, но условно групп таких шкал будет всего две – группа качественных отношений между предметами и группа количественных отношений. К первой из них будут относиться всего две шкалы нашего списка – номинальная и порядковая, к количественным – остальные.

Рассматривая вишню мы имели в виду то понятие числа, которое лежит на поверхности и общеупотребимо. Но сама по себе цифра не всегда предполагает те операции (или так называемые допустимые преобразования), к которым каждый из нас привык в школе. Вовсе не всегда 2 + 2 = 4, а 4 * 4 = 16. Так получается именно из-за отношений. Указанные законы сложения и умножения правомерны лишь тогда, когда цифрам присвоены строго определённые отношения, но, узнав, что отношения могут быть разными, мы выясним, что и операции над этими отношениями не всегда столь предсказуемы, как мы ожидаем. Ведь если посадить в клетку 2 кроликов, никто из вас не удивится, увидев вместо 2 кроликов или, на крайний случай, 3-х - целых 24, например.

Качественные шкалы, как ни странно для логики психолога, отражают меньшее количество информации о предмете изучения, чем количественные шкалы. Название отражает лишь тот предупредительный факт, что номинальная и порядковая шкала не предполагает тех действий, которые применимы для «настоящих» чисел, к которым мы привыкли.

Номинальная шкала.

Исходя из названия, это шкала наименований. Чтобы быстро ввести в курс дела, рассмотрим анекдотический пример:

- Петя, сколько человек было в комнате, когда ты вошёл?

- Ну… Таня, Катя и Денис.

Т.е. номинальная шкала представляет собой предельно простую классификацию, где подклассы не пересекаются (невозможно спутать, к какому подклассу относится тот или иной предмет, потому что критерии деления чётко оговорены). Тогда Таня – 0, Катя – 1, Денис – 2. Единственное, что мы можем сделать, если оставим такую шкалу в виде или имён, или даже присвоим каждому подклассу какую-нибудь цифру – это констатировать тождество или различие.

$А = Б$, $А \neq Б$.

Катя – это не Денис – вот то единственное, что мы знаем об этом явлении.

Обратите внимание, что номинальная шкала может очень удивить. Например, ИНН налогоплательщика (что-нибудь из серии: 540811794…) представляет собой ничто иное как код, присваиваемый вместо имени, что относит такой длинный числовой ряд к номинальной шкале. Встречаются и спорные шкалы, но об этом далее.

Находить среднее арифметическое, умножать, делить и т.п. находясь с такими данными на руках попросту нельзя. Однако небольшая группа статистических процедур для такой шкалы всё же есть. Но стоит учитывать, что операции всё же идут не столько над самими подклассами, сколько над количеством объектов, отнесённым к этим подклассам. Поскольку это так, то можно сделать вывод, что математический аппарат по большей части не предназначен для обработки такого рода данных.

<u>Порядковая шкала.</u>

Порядковая шкала также говорит сама за себя. Объекты здесь упорядочены, что даёт исследователю больше свободы действий, чем в случае с номинальными шкалами, но всё же свобода эта не безгранична.

Опять же рассмотрим юмористический пример.

За первой улиткой ползла вторая, а за второй третья. Подружки ли они?

Согласно порядковой шкале они подружки, потому что такой шкале в принципе всё равно, каково между улитками расстояние. Если бы мы были, конечно, в более высоком уровне измерения, то можно было бы подискутировать о расстоянии, но поскольку сейчас говорим только лишь о порядке, в котором одна улитка следует за другой, то и спор исчерпан.

Допустимые операции:

А = Б, А ≠ Б, что было свойственно номинальной шкале, а также

А > Б, А < Б.

Порядковый ряд в статистических процедурах будет называться ранговым или вариационным рядом. Обработка таких данных осуществляется с помощью достаточно большой группы методов непараметрической статистики, на чём мы опять же остановимся в следующих параграфах. Но прежде чем закончить разговор, стоит сделать несколько крайне важных для психолога комментариев.

Большинство психологических тестов имеют на выходе именно порядковые баллы. На самом деле даже стандартизированные баллы по типу: IQ, T, стенов и станайнов не всегда относятся к шкалам более высокого порядка. Здесь имеет смысл помнить правило: лучше шкалу понизить, нежели повысить. Почему? Как заметил внимательный читатель, все те операции, которые допустимы в шкалах более низкого порядка, допустимы и в высоких шкалах, но не наоборот.

Например, оценка группой экспертов некоего явления, как и самооценка во многих других тестах, всенепременно относится к шкале порядка.

Следующее замечание будет ещё более любопытным, поскольку многие исследователи делают эту ошибку совершенно неосознанно, но при том решительно повсеместно.

Чтобы эту ошибку понять и не совершать, давайте строить рассуждение вместе. Итак, предположим, что у нас есть психологический тест, в котором испытуемому предлагается ответить на ряд вопросов о своей тревожности. Оценка идёт по 5-балльной шкале (что иногда называется шкалой Лайкерта).

1. Тревожны ли Вы после обеда?

-2 … -1 … 0 … 1 … 2

2. Тревожны ли Вы по понедельникам?

-2 … -1 … 0 … 1 … 2

3. Тревожны ли по вторникам?

-2 … -1 … 0 … 1 … 2

И так далее, 60 вопросов. Один из показателей теста (30 вопросов) измеряет тревожность как состоянии внутреннее, а остальные 30 вопросов объединяются в тревожность как состояние, обусловленное внешними причинами. Далее мы обычно что делаем? Выводим общий показатель, который получается либо суммированием двух предыдущих показателей (или всех вопросов), либо усреднением показателей (чему зачастую соответствует несколько иной тип задач – например, средняя успеваемость школьников по разным предметам). Начнём с суммирования. Строго говоря, мы не имеем на это права. Ведь сумма баллов или суммарный показатель имеет смысл в том случае, когда у нас есть основание предположить, что интервалы равны. Почему же мы в таком случае позволяем себе строить суммарные показатели, не имея на них права? Потому что у нас выхода нет – запрос практики для психолога, как это ни прискорбно, выше абстрактной математической теории. В случае же, если бы мы решили вывести общий показатель тревожности или же показатель успеваемости по предметам как средний, то у нас бы появилась возможность поступить более корректно. Как же вывести хотя бы средний показатель корректно? *Для порядковых шкал* *имеет смысл считать средним показателем не среднее арифметическое, а медиану.* Это никак не противоречит общим идеям о вариационных рядах.

И, наконец, последнее замечание о порядковых шкалах носит частный характер, но показывает некоторую относительность любого понятия. Является ли порядок порядком? Например, является ли номер дома порядковой шкалой? Может быть номинальной? На этот смешной вопрос ответ бывает самый разный. Чтобы решить, является ли порядок порядком необходимо договориться о том, присвоен ли этот порядок согласно какой-либо логике. Т.е., например, если дома расположены строго определённым образом, таким, что более новый дом носит больший порядковый номер, то почему бы нам и не признать адрес порядковой шкалой? Если дети в шеренге стоят строго по росту, то порядковый номер каждого из них будет означать, что мы имеем порядковую шкалу роста. Однако же, если встали они как попало, то порядковый номер ничего значить не будет, кроме номинального разграничения. Таким образом, на вопрос – является ли порядок порядком всегда стоит привлекать дополнительные уточняющие сведения о способе построения такой классификации.

С качественными (неметрическими) шкалами измерения на этом мы заканчиваем. Какие же количественные (метрические) шкалы бывают?

Шкала интервалов.

Рассмотрим опять же пример. Давний (так и не решённый, между прочим) вопрос формулируется очень просто – можно ли считать интеллект в IQ баллах метрической шкалой? Вступим в этот спор поступательным размышлением.

Если мы находимся в интервальной шкале при разговоре об IQ-баллах, то очередной юмористический пример должен быть более правдой, чем шуткой. У нас есть четыре мальчика, чей интеллект мы очень хотим сравнить попарно. У Пети интеллект - 100, у Васи – 130, у Грини – 80, у Феликса – 50. Разница между интеллектом первых двух героев – 30 единиц, вторых – тоже 30. Можно ли считать, что 30 единиц разницы Пети и Васи равны 30 единицам Грини и Феликса? Если мы в интервальной шкале, а в данном примере мы якобы именно там, то ответ – да. Однако же такой ответ, на мой скромный взгляд, является слишком преждевременным и опрометчивым. И если читатель ещё не проникся моей логикой (хотя вовсе не обязательно, что это вообще произойдёт), то заострю иронию ситуации. 30 баллов Пети и Васи означало, что Петя находится в границах среднестатистической нормы, а Вася имеет интеллект выше среднего. В то время как 30 баллов Грини и Феликса означало, что Гриня находится чуть ниже среднего, а Феликс скорее всего имеет патологию развития интеллекта. Таким образом, разница в 30 баллов IQ, взятая из разных отрезков распределения, может сопровождаться совершенно разными качественными проявлениями – в определённом соотношении 30 баллов могут быть и не слишком заметны, в другом же – определять совершенно разные возможности в жизнедеятельности. Такая картина, если принять её истинность, означает два заключения: на данный момент мы не можем признать за IQ-баллами правомерность интервальности шкалы, поскольку равность интервалов вызывает большие сомнения; в построении IQ-шкалы есть масса неартикулированных качественных признаков, которые никак не отражены в количественных эквивалентах её измерения. Завершая разговор об IQ-баллах стоит назвать ещё два распространённых мнения исследователей на этот вопрос: - можно во что бы то ни стало признавать интервальность шкалы; - либо ввести новое понятие - квазиметрической шкалы, чем и отметить отличие подобного рода шкал от порядковых и интервальных. Относительно первого предложения было приведено размышление о различии интеллекта на разных отрезках распределения, что же касается второго, то очевидных причин отрицать такое предложение нет, поскольку очевидно, что IQ-баллы представляет весьма доработанную шкалу, которая уже не является «просто» порядковой. Однако же признание этого нового статуса в практическом смысле не имеет на данный момент больших преимуществ, поскольку отдельного статистического аппарата, который бы позволял обрабатывать именно такие шкалы, пока что не создан, а повышать уровень шкалы без серьёзных на то оснований у нас нет права. А значит мы зачастую возвращаемся к обработке непараметрическими методами для порядковых шкал.

Допустимые преобразования: $x' = kx + b$.

И, начиная с этой шкалы, мы вводим ещё два новых понятия – точка отсчёта и единица измерения. Под точкой отсчёта стоит понимать абсолютный ноль, т.е. «ноль», в котором измеряемое данной шкалой свойство равно нулю. Под единицей измерения стоит понимать распространённую меру, которая используется для сравнения при измерении. Единицами измерения являются: сантиметры, килограммы, фунты, футы, амперы и т.д.

Для шкалы интервалов характерны произвольная точка отсчёта и единица измерения, которые присваиваются самим исследователем.

Например, шкалы температур Фаренгейта и Цельсия относятся к равным интервалам, поскольку нет ни естественной точки отсчёта, ни естественной единицы измерения. В то же время шкала температур Кельвина имеет естественный абсолютный ноль, что позволяет использовать большее количество преобразований и переводит её на уровень шкалы отношений.

Шкала отношений.

Если бы мы модифицировали задачу с интеллектом мальчиков для шкалы отношений, то выясняли бы равенство не вычитанием, а делением. Действительно ли можно составить пропорции, где отношение первого мальчика ко второму равно отношению третьего к четвёртому? Это вопрос как раз для шкалы отношений.

На этом уровне измерения уже имеется естественная точка отсчёта (которая и позволяет сравнивать отношения одного члена ряда к другому), но единица измерения всё ещё произвольна.

Допустимые преобразования: $x' = kx$.

Шкала разностей.

Примером шкалы разностей является время, измеренное в сутках, годах и т.п., несмотря на то, что время, измеренное в секундах или часах, является шкалой интервалов. Разница здесь в том, что в шкале разностей присутствует естественная единица измерения, но начало отсчёта произвольно.

Допустимые преобразования: $x' = x + b$.

Абсолютная шкала.

У абсолютной шкалы естественная точка отсчёта и естественная единица измерения. Примером может служить количество людей в комнате – вполне очевиден как абсолютный ноль, так и приращение величины.

Допустимые преобразования: $x' = x$.

Таким образом, мы видим, что деление шкал получается по трём признакам: - упорядоченность позиций шкалы относительно друг друга; - интервальность, т.е. равные интервалы между позициями; - нулевая точка – наличие абсолютного нуля отсчёта. В зависимости от того, присутствуют ли эти три признака, мы можем судить о типе шкалы.

1.7. Независимость/зависимость выборок и переменных.

Часто, слишком часто в качестве определения зависимости/независимости выборок я слышу нечто вроде: независимые выборки – это две выборки, а зависимые – это одна. Такое определение не является неправильным, но совершенно точно не достаточно. Гораздо более адекватно определять зависимые выборки как выборки, в которых есть строгое соответствие величины X и Y, в то время как независимые выборки такого соответствия не предполагают.

Рассмотрим пример.

Мы измерили интеллект и креативность детей 1 «Б» класса. Таким образом, интеллект будет представлять первый столбик нашего массива данных, а креативность – второй. Совершенно очевидно, что мы не можем спутать порядок, в котором будем вводить набранные баллы – перепутать местами, например, данные креативности первой и десятой строки. Почему мы не можем это сделать? Потому что каждая строка представляет соответствие интеллекта и креативности совершенно конкретного испытуемого. Иначе же интеллект у него останется свой, а креативность – чужая.

Рассмотрим другой пример.

Мы делаем массив данных, например, по интеллекту 1 «Б» и 1 «В». Опять же очевидно, что вводить порядок данных по интеллекту этих двух классов можно разными способами – главное – не внести данные из одной группы в другую группу. Такая свобода является явным признаком независимости выборок.

Это правило не вызывает никакой сложности, но вот ситуации более сложного характера – вполне могут.

Предположим, что у нас есть данные интеллекта по матерям и дочерям.

Первый столбик в таком случае – это интеллект матерей, а второй – дочерей. Здесь мы, несмотря на то, что это две выборки, заметим, что помещать в первую строку интеллект мамы X_1, а во вторую строку – дочки Y_{20}, не имеет смысла. Даже интуитивно чувствуется, что необходимо делать соответствие между этими столбцами. Стоит внимательно относиться к такой интуиции и как только возникает желание выстроить соответствие между X и Y, это явный признак зависимости выборок. Это правило является достаточно важным помощником выбора статистического критерия, потому что многие из них существуют в двух модификациях: для зависимых и независимых выборок. Поскольку от одной модификации к другой меняется логика, подсчёт и вероятности, перепутать их между собой – это сделать достаточно серьёзную ошибку в смысле корректности, да и результата.

Исходя из этого правила, становится совершенно очевидно, что считать, например, корреляцию не всегда имеет смысл. Если мы имеем интеллект 1 «Б» и 1 «В», то посчитать корреляцию в принципе невозможно, поскольку рассуждение в таких критериях идёт об изменении значений от X к Y, а в этом случае мы не сможем вытроить такую таблицу, чтобы было чёткое соответствие X и Y.

Переходим к переменным. Наличие контролируемого воздействия, как следует из теоретических основ экспериментальной психологии, всегда верный признак эксперимента, в то время как его отсутствие характерно, например, корреляционному исследованию. Именно экспериментальное воздействие называется влияющим фактором или независимой переменной. Например, если мы предположим, что ставим неэтичный пример удара по ноге молотком, то независимой переменной будет сила удара молотка, а зависимой будет реакция ноги.

Если при попытке использования того или иного статистического критерия мы встретились с необходимостью различать зависимые и независимые переменные, то это достаточно неплохой признак попытки математического обоснования именно причинно-следственной связи между явлениями, а не такой, как, например, корреляционная, где нет указания на то, что есть влияние, а что – реакция на это влияние. Например, ANOVA построен по принципу чёткого разделения независимых и зависимых переменных, что даёт свой неоспоримый плюс – мы можем вполне обоснованно, пользуясь факторными планами (не путать с факторным анализом как статистическим методом), доказывать причинность явлений. К независимым переменным при использовании того или иного метода может быть предъявлено требование совершенно особенного рода, а именно – градация факторов. Например, при использовании ANOVA желательно иметь не менее трёх градаций независимой переменной, что означает следующее. Предположим, что мы решили изучить влияние удовлетворённости жизнью и семейных традиций на результативность профессиональной деятельности человека. Не представляю, зачем психологу такое исследование, но почему бы и нет. В таком случае под градациями этих двух факторов будет пониматься, например, высокая, средняя и низкая удовлетворённость жизнью, а также какие-нибудь градации семейных традиций, например, ярко выраженные семейный «мифы», средняя выраженность, низкая или отсутствие «мифов» семьи.

С другой стороны, независимая переменная в статистическом критерии не всегда берётся как варьируемое экспериментальное воздействие – это может быть и фиксированный фактор разделения групп. Зачастую такая независимая переменная называется группирующим фактором и отражает лишь принадлежность испытуемого к той или иной группе сравнения. Критерий Манна-Уитни, Крускала-Уоллиса и многие другие могут быть посчитаны в некоторых программных продуктах на массиве, где будет всего два столбика данных: первый – сырые баллы, а второй – указание, к какой группе данных относится этот сырой балл.

1.8. Специфический вопрос исследователя и многообразие статистических критериев.

Мы рассмотрели вопросы, являющиеся подготовкой к решительному шагу – выбору статистического критерия.

Напомню, каков список этих вопросов:

1. Нормально ли распределение (однородна ли дисперсия в необходимых случаях).

2. Каковы шкалы измерения.

3. Сколько выборок и переменных.

4. Каковы эти выборки и переменные (зависимы, независимы).

5. Специфический вопрос.

Поскольку мы уже знаем, как искать ответ на первые четыре вопроса, то нам остаётся понять только то, какую статистическую гипотезу мы можем задать, чтобы получить ответ на тот исследовательский вопрос, который поставили.

Классификация статистических критериев будет представлять собой форму «вопрос – ответ». Естественно, что авторство предлагаемой таблицы-списка ни в коем случае не принадлежит автору данного пособия – это «совместный», хотя и «независимый» труд других авторов[17]. Я умышленно отойду от стандартного списка таких вопросов, переименовав (в скобках будет указана распространённая формулировка) и добавив некоторые пункты, чтобы читателю было проще ориентироваться в многообразии статистических критериев. Естественно, что переименованные и добавленные пункты являются алогичными для математика, поскольку неточно и дублирующее передают ту же самую суть, что и распространённые деления, но для начинающего исследователя такие корректировки могут быть очень полезны. Однако, сколь бы полным не казался список критериев, приведённых ниже, это далеко от истины – статистика располагает гораздо большим количеством методов.

Классификация методов по задачам исследования:

1. Проверка различия независимых групп.

Распространённое название такого пункта классификации - выявление различий в уровне исследуемого признака, но оно не отражает того простого факта, говорим ли мы при этом о зависимых или независимых выборках, в связи с чем у разных авторов встречаются разногласия, которые в меру размытости формулировки трудно критиковать – правы все диалектически. Чтобы не мучить читателя, я задаю область достаточно строго.

А) 2 выборки испытуемых:

- Q – критерий Розенбаума;

- U – критерий Манна – Уитни;

- φ* - критерий (угловое преобразование Фишера);

[17] Сидоренко Е.В. «Методы математической обработки в психологии» (http://www.koob.ru/sidorenko/)

Нижегородцева Н.В., Мишина Т.В. «Методические рекомендации по написанию и оформлению курсовой и выпускной квалификационной работы по психологии и конфликтологии» (http://cito-web.yspu.org/link1/metod/met125/met125.html)

Афанасьев В.В. «Теория вероятностей в вопросах и задачах» (http://cito-web.yspu.org/link1/metod/theory/theory.html)

Кондаков И.М. «Базовый курс SPSS for Windows для обработки психологических данных» (http://www.matlab.mgppu.ru/work/0024.htm)

Шишлянникова Л.М. «Математическое сопровождение научной работы с помощью статистического пакета SPSS for Windows 11.5.0» (http://www.matlab.mgppu.ru/work/0022.htm) и др.

- t - критерий Стьюдента для независимых выборок;

- критерий Крамера-Уэлча.

Б) 3 и более выборок испытуемых:

- S – критерий Джонкира;

- H – критерий Крускала – Уоллиса.

Сюда же включают иногда и ANOVA для независимых выборок, но нужно понимать, что при всей правомерности такого подхода читателю придётся ответить на вопрос, есть ли в его распоряжении независимые переменные, содержащие не менее трёх градаций каждая.

2. Проверка различия зависимых групп.

Распространённое название данного пункта – оценка сдвига значений исследуемого признака, но, поскольку зависимыми выборками могут быть и не одни и те же люди, как это принято давать в определении довольно большого количества пособий и статей, то я буду избегать такой формулировки во избежание путаницы.

А) 2 замера:

- T – критерий Вилкоксона;

- G – критерий знаков;

- φ* - критерий (угловое преобразование Фишера);

- t –критерий Стьюдента для зависимых групп.

Б) 3 и более замеров:

- χ^2 – Фридмана;

- L – критерий тенденций Пейджа.

Некоторые авторы опять же добавляют в список ANOVA, имея в виду случай критерия для зависимых выборок.

3. Выявление различий в распределении.

Прекрасная группа критериев, которая сравнивает либо функции распределения между собой, либо с теоретически заданным эталоном – нормальным, экспоненциальным, пуассоновским, равномерным или каким-то другим. Для психолога зачастую такие слова звучат как страшный шум, но в том же SPSS реализованы алгоритмы, которые позволяют сравнить эмпирическое распределение с теоретически заданным, например, одновыборочным критерием Колмогорова-Смирнова; существует множество опций для проверки нормальности и т.д.

А) при сопоставлении эмпирического распределения с теоретическим:

- χ^2 – критерий Пирсона (имеется в виду вариант, который называется критерием согласия, а не независимости);

- λ – критерий Колмогорова – Смирнова (имеется в виду критерий Колмогорова в модификации Смирнова, хотя, естественно, что может быть использован и какой-либо другой вариант этого семейства; одновыборочный для сравнения с такими теоретическими распределениями как: нормальное, экспоненциальное, равномерное и др.);

- m – биномиальный критерий.

Б) при сопоставлении двух эмпирических распределений:

- χ^2 – критерий Пирсона (независимости);

- λ – критерий Колмогорова-Смирнова (естественно, что имеется в виду уже не одновыборочный);

- φ^* - критерий (угловое преобразование Фишера).

Нужно сказать, что пункт Б, с точки зрения психолога, может потерять свою красоту, поскольку если мы не говорим о сравнении с теоретическим распределением, то углубление в сравнение эмпирических распределений между собой представляет собой на первый взгляд не слишком отличную задачу от сравнения двух дисперсий или средних другими группами методов, в связи с чем стоит остановиться подробнее на случае, когда именно нам интересно сравнивать эмпирические распределения, а когда, например, математические ожидания.

H_0 (обозначим эту гипотезу H_a) для одних критериев может быть сформулирована так: F(x) = G(x) при всех x, где F и G – функции распределения случайных величин X и Y.

Для других же критериев H_0 (обозначим эту гипотезу H_b) может быть сформулирована так (например, t-критерий Стьюдента): M(x) = M (y), где M(x) и (y) – математические ожидания случайных величин X и Y.

H_a предполагает H_b, но принятие гипотезы H_b не обязательно должно означать принятие гипотезы H_a.

Предположим, что перед нами стоит задача понять, можно ли считать школьников 1 «А» и 1 «Б» однородными группами, т.е. группами, между которыми нет различий по какому-либо признаку. Такая ситуация весьма часто встречается в случае, когда мы собираемся провести эксперимент с использованием экспериментальной и контрольной группы – замер «до» выглядит именно таким образом и призван ответить на вопрос, сможем ли мы впоследствии обоснованно делать выводы о динамике в группе «с воздействием» и группе «без воздействия» (контрольной). На мой взгляд, такой случай желательно было бы обработать именно критериями, сравнивающими распределения. Никто, конечно, не запрещает сравнить группы «до» и методами, сравнивающими средние или дисперсии, но это не так красиво и доказательно. Ведь наша задача была подобрать как можно более схожие группы, чтобы на одной смотреть естественное развитие, а с другой экспериментировать. В таком случае, логично сравнивать именно распределения, а не средние двух групп.

Однако, есть случаи, где нам не требуется такой однородности, как в первом примере. Нам вполне достаточно равенства средних. Например, мы решили понять, существует ли различие между интересами к учёбе у первоклассников и пятиклассников. Доказывать, что распределения интересов у этих двух групп схожи, у нас нет необходимости. Нам будет достаточно понять, можно ли по групповым показателям зафиксировать статистически значимое различие. В таком случае мы можем использовать t-Стьюдент или какой-либо другой критерий.

4. Выявление связи.

Конечно, распространённое название данного пункта – выявление степени согласованности изменений – весьма чётко отражает суть критериев этой группы, однако для читателя, который не слишком хорошо знаком со статистическими методами, «связь» «роднее».

А) двух признаков:

- r_s –коэффициент ранговой корреляции Спирмена;

- τ – коэффициент корреляции Кендалла;

- r – коэффициент линейной корреляции Пирсона;

- η – корреляционное отношение;

- r_{bis} – бисериальный коэффициент корреляции Пирсона (а также точечно-бисериальный и рангово-бисериальный);

- r_{tetr} – тетрахорический коэффициент корреляции Пирсона;

- линейная и криволинейная регрессии.

Поскольку коэффициентов корреляции довольно много, то приведу таблицу, которая имеет то преимущество, что указывает, какой метод должен быть выбран в случае, если одна переменная измерена в одной шкале, а другая – в другой:

	дихотомия	порядковая	метрическая
дихотомия	тетрахорический	рангово-бисериальный	точечно-бисериальный; бисериальный
порядковая		критерий Спирмена ; критерий Кендалла	критерий Спирмена; критерий Кендалла
метрическая			критерий Пирсона

Б) двух иерархий или профилей:

- r_s – коэффициент ранговой корреляции Спирмена.

Признаться честно, я не имею ни малейшего понятия, почему мне так и не удалось встретить книгу, в которой бы в этот список добавили, например, линейный коэффициент корреляции Пирсона. Исходя из общих соображений, поскольку коэффициент Спирмена есть частный случай именно коэффициента Пирсона для вырожденных значений – рангов, то нет особых возражений за исключением нормальности, конечно, в применении для анализа двух иерархий и коэффициента Пирсона. Однако, поскольку в данном случае Пирсону объявлено «эфирное молчание», подобно тому, как Спирмену объявлено без всяческих причин «эфирное молчание» в подсчёте психометрических характеристик тестов, я не включаю этот коэффициент в официальный список для такой исследовательской задачи, хотя категорически намекаю на такую возможность.

В) трёх и более признаков:

- линейная, криволинейная и множественная регрессия;

- множественная и частная корреляция;

- r_s – коэффициент ранговой корреляции Спирмена;

- r – коэффициент линейной корреляции Пирсона.

Добавлю W-коэффициент конкордации Кендалла, который позволяет оценить, например, согласованность мнений нескольких экспертов.

Опять же позволю себе несколько комментариев.

Несмотря на то, что в этом списке перечислены коэффициенты Пирсона и Спирмена, это вовсе не значит, что мы сможем «убедить» этих коэффициентов проверить наличие связи одновременно для более чем двух признаков. Нет, нет и ещё раз нет. Мы просто можем сделать несколько отдельных проверок, что обычно и делается.

Что же касается звонкого названия «множественная регрессия», стоит понимать, что связь в таком анализе идёт не между несколькими зависимыми переменными, а между несколькими зависимыми.

<u>5. Анализ изменений признака под влиянием контролируемых условий.</u>

Альтернативным названием, если распространённое вызывает сомнения, могу предложить следующее: изучение влияния независимых факторов на зависимые.

А) под влиянием одного фактора:

- S – критерий тенденций Джонкира;

- L – критерий Пейджа;

- ANOVA;

- T - критерий Бартлетта;

- G - критерий Кохрена

- критерий Линка и Уоллеса;

- критерий Немени.

Б) под влиянием двух и более факторов:

- ANOVA.

<u>6. Вспомогательные методы.</u>

Под вспомогательными методами я хочу ещё раз напомнить читателю те статистические критерии, которые чаще всего используются не как самоцель исследования, но как средства для проверки общих требований к другим критериям. Соответственно, получив ответ от этих критериев мы либо получаем право использовать следующие, либо не получаем.

А) нормальность распределения:

Критерии согласия как раз используются для сравнения распределений (мы уже встречались с ними в пункте 3). В случае, когда мы задумываемся об их вспомогательной функции, то на ум, естественно, приходит в первую очередь нормальность распределения, которую можно проверить с помощью достаточно большого количества критериев данной группы. Распространёнными критериями проверки нормальности, как мы уже помним, являются: критерий Колмогорова-Смирнова; критерий Шапиро-Уилка. В случае подтверждения нормальности распределений мы можем использовать параметрические критерии для проверки своих амбициозных планов сравнения групп, поиска связей признаков и т.д. В случае, если данные не подчиняются нормальному закону распределения стоит использовать непараметрические методы, а повторяю я эти правила лишь затем, чтобы выбор того или иного критерия в исследовательской работе читателя носил корректный и обоснованный характер.

Б) однородность дисперсий групп – ещё одна задача вспомогательного характера, которая может быть указана требованием к использованию статистического критерия. Например, перед использованием t-критерия Стьюдента для независимых групп требуется, чтобы обе группы имели одинаковую дисперсию, что может быть проверено с помощью, например, следующих трёх критериев: - Левена, - Брауна-Форсайта; - F-критерия Фишера.

В) сравнение распределений с целью объединения групп данных для корреляционного анализа.

Этот случай, а именно – необходимость применения критериев сравнения двух эмпирических распределений с целью объединения двух групп в общую третью группу, может быть описан следующим примером, который, как то ни странно, встречается довольно часто. Предположим, что наше исследование посвящено алкоголикам и учителям литературы. Мы считаем, что это замечательнейшие группы для сравнения – одни - «хорошие», другие - «плохие».

Сделаю комментарий о методе контрастных групп вообще. Я шутливо рассказываю о такого рода исследованиях по двум причинам:

- во-первых, к контрастным группам из-за угрозы статистической регрессии к среднему (с этим видом угрозы внутренней валидности эксперимента можно ознакомиться, наряду с другими угрозами, в книге Д. Кэмпбелла «Модели экспериментов в социальной психологии и прикладных исследованиях») стоит относиться с должным вниманием и не использовать без необходимых для этого причин, когда речь идёт о лонгитюдных исследованиях;

- во-вторых, мой скепсис объясняется удивительно идеальными результатами подобных исследований на контрастных группах, где герои всегда оказываются на белых конях, а злодеи повержены, хотя другое подобное исследование обнаружит ещё героев и ещё злодеев, что в гротеске, возможно, может означать, что злодеи по сравнению с другими злодеями станут даже и героями. Психологии лишь предстоит осознать природу этой бессмысленности и неоправданности контрастности групп «плохих» и «хороших» большинства эмпирических исследований (эта ситуация встречается даже в тех случаях, когда о достоверности полученных студентом данных не стоит волноваться), что я, в свою очередь, на данный момент не могу рационально объяснить. Я могу лишь выдвинуть несколько возможных причин некорректности параллельных срезов контрастных групп для некоторых исследовательских случаев (угрозы продольных срезов уже описаны Кэмпбеллом):

- статистическая регрессия может проявляться не только в лонгитюде, но иметь своеобразный аналог и в параллельном срезе, что, согласитесь, не слишком великая экстраполяция от Кэмпбелла, хотя, безусловно, требует дополнительных размышлений и проверок;

- неполнота выделения системы, как «методологическая печаль», в данном случае обращается оружием против любого исследователя, поскольку адекватность присвоения именно этим группам статуса контрастности НИКОГДА не может быть окончательно обоснована теоретическим путём, что невольно вызывает предположение из серии: а другая «хорошая» выборка будет такой же «хорошей», как эта; а другая «плохая» выборка – например, не алкоголики, как в нашем вымышленном примере, а наркоманы не будут ли демонстрировать ещё низкие или такие же низкие результаты, как и выборка алкоголиков в сравнении с потенциально нормальной выборкой. Здесь я пытаюсь заострить вопрос классификации индивидов – мы всегда наивно полагаем, что мир состоит из чёрных и белых тонов, хотя, сколь бы удивительно это не было, мир может оказаться многомерными пространствами, где одна ось логически связана с другой, где полюс имеет свойство трансформироваться в абсолютно противоположный, подобно тому как идеальные стоики, которые никогда не употребляют вредных веществ, при неудачном стечении обстоятельств превращаются в заядлых алкоголиков. В этом смысле мне весьма и весьма импонирует квадриполярный подход к когнитивным стилям М.А. Холодной, потому что я не хочу всерьёз думать, что мир также убог, как и мои исследовательские представления о нём. Гораздо лучше на определённом этапе сказать «я не знаю» или «у меня нет научного объяснения», чем привести представление о мире к какой-то жуткой карикатуре истинного положения вещей. Поэтому вопрос контрастности групп зиждется на строгом классификационном представлении, которое зачастую отсутствует.

- и, пожалуй, апофеозом проблем контрастных групп как исследовательского метода является патологическая неправомерность экстраполяции вывода. Если даже предположить, что полученный факт есть именно факт, а не артефакт предыдущих причин, то мы не можем сделать вывод о том, в связи с чем столкнулись с такими особенностями контрастных групп. Наше психологическое замалчивание причинности в корреляционном исследовании носит скорее дань моде статистической корректности, нежели истинное отражение наших «тайных» мыслей. Делая любой корреляционный план, будь то хоть план, использующий в качестве статистических методов непосредственно корреляционные статистические методы (критерий Спирмена и т.д.) или критерии различий (типа: Манна-Уитни, t-Стьюдента и т.п.), чаще всего исследователь надеется понять, что за причина стоит за эмпирическими фактами. Безусловно, статистическая корректность, которая гласит о невозможности причинных выводов, более чем правомерна. Но, в таком случае как интерпретировать выводы о том, что у алкоголиков интеллект ниже, чем у нормальных испытуемых (т.е. у наркоманов, зоофилов, любителей контакта, домохозяек, инженеров, рабочих и архитекторов). Мы поневоле должны рассматривать алкоголизм как следствие влияния независимого фактора, с той лишь разницей, что в меру невозможности контроля над ним, его влияние выражается в нашем случае через дихотомию эффекта воздействия фактора – алкоголики и нормальные. Совершенно очевидно, что без полной картины события, без знания континуума выраженности, мы даже никому не нужный вывод из серии: у алкоголиков такой-то интеллект, а у нормальной группы такой-то, не можем сделать корректно. Потому что мы не знаем, дело здесь в алкоголизме или наш результат случаен. Единственный вывод, который будет корректен – это у Василия такой-то интеллект (хоть он и алкоголик), у Катерины (хоть она и находится в группе нормальных) – такой-то и т.д. Единственный случай, когда мы можем идеально корректно сделать вывод о том, что такая-то контрастная группа имеет такой-то уровень выраженности признака, а другая – другой, это когда мы имеем дело с полной группой несовместных событий, например, с полом. Люди бывают двух типов: мужчины и женщины. Если предположить, что другого не может быть (хотя это, естественно, алогично, поскольку может – начиная от нетрадиционной ориентации и заканчивая ошибками хромосом), но если всё-таки поверить в правомерность именно такой классификации, то вывод о различии двух выборок имеет свою корректную основу. Но, несмотря на то, что мы получили правомерное основание классификации и можем без всяческих сомнений рассматривать только мужчин и женщин, о причинах мы всё равно говорить не можем, поскольку на классификацию пола может быть наложена другая классификация, например, место проживания, что и даёт либо совместный, либо к ужасу исследователя, единственный эффект. Такое явление названо влиянием через третью переменную.

Естественно, что моя ужасающая зарисовка на тему контрастных групп не является причиной никогда не проводить исследования с помощью контрастных групп, ведь одна десятикилограммовая кошка не приводит к отвержению общей теории о среднем двухкилограммовом весе кошек, но эти размышления стоит учитывать, когда охватывает уверенность праведности пути. Метод контрастных групп ничем не хуже другого способа составления корреляционных выборок, но и не лучше, несмотря на свою псевдопривилегию.

Однако, «высшим пилотажем» многих исследований с использованием контрастных групп является отнюдь не факт избрания «плохих» и «хороших», а сам ход статистической проверки. Предположим, что у алкоголиков и учителей литературы мы изучали креативность (творческую способность). Данные мы проверили на нормальность, предположим, что она подтвердилась. В таком случае, чтобы понять, различен ли уровень креативности в наших группах, мы выбрали опять же критерий t-Стьюдента для независимых выборок. Обнаружили, что существуют различия. А дальше… и собственно ради этого номера программы было всё повествование… нас укусила какая-то неведомо коварная муха и мы объединили группы, чтобы посмотреть связь креативности с каким-нибудь другим параметром, например, интеллектом. Ради чего бы нам это делать? Единственным аргументом, который говорит «за» такое действие – это увеличение количества человек. Чем больше людей в группе для корреляционного анализа, тем более точным и чувствительным становится корреляционный анализ при общих равных. Но. И это «но» перечёркивает предыдущий плюс полностью и без остатка. Если мы опустим в «тазик» корреляционного анализа разрозненных испытуемых как, например, наших алкоголиков и учителей литературы, то можем «затереть» силу связи или потерять её вовсе только за счёт того, что одной части нашей выборки будет свойственно одно направление связи, а другой другое. Поэтому для корреляционного анализа, проводимого на объединённых группах, как то: контрастные, параллельные срезы лонгитюда (1 «А», 5 «Б», 5 «Г», 7 «Б» классы) и т.п., имеет смысл проверять однородность групп критериями, нулевые гипотезы которых исходят из равенства распределений. Причём объединение двух групп в одну общую дают право именно критерии сравнения распределений, а не, например, средних или дисперсий. А, уже получив сведения о том, что у нас есть право относить выборки к одной генеральной совокупности, внимательно рассмотрев графики рассеяния и не обнаружив подозрительных противоречий задумке можно провести корреляционный анализ на объединённой выборке. Хотя, если подтверждаются условия для объединения выборок, приходится признать неправомерность рассмотрения этих двух групп как контрастных. Но, к сожалению, либо одно, либо другое. Если же хочется посмотреть как различие, так и связь, то при различии распределений это можно сделать отдельно для одной группы, отдельно для другой, чему нет каких-либо очевидных препятствий.

7. Многомерные статистические методы.

Самой, пожалуй, сложной в понимании группой методов являются многомерные методы. И сложность начинается буквально с того, что у студента возникают трудности с определением списка таких критериев. Попытаемся в этом пункте ответить на вопрос, какие же критерии можно отнести к этой мифической группе.

Я приведу таблицу, которая приведена в руководстве Шишлянниковой Л.М. «Математическое сопровождение научной работы с помощью статистического пакета SPSS for Windows 11.5.0» (http://www.matlab.mgppu.ru/work/0022.htm):

Зависимая переменная	Независимые переменные	Многомерный метод
Дихотомическая	Любые	Двоичная логистическая регрессия; дискриминантный анализ
Дихотомическая	С номинальной или порядковой шкалой	Логит-логарифмические линейные модели
С номинальной шкалой	С номинальной или порядковой шкалой	Мультиномиальная логистическая регрессия
С порядковой шкалой	С номинальной или порядковой шкалой	Порядковая регрессия
С интервальной шкалой	С номинальной или порядковой шкалой	Дисперсионный анализ
С интервальной шкалой	Любые	Ковариационный анализ; множественный регрессионный анализ

К списку многомерных статистических методов, перечисленных в таблице, следует также относить весьма популярные методы факторного, кластерного анализа и многомерного шкалирования.

На этом мы заканчиваем наш список критериев, что отчасти должно даже порадовать читателя.

1.9. Как пользоваться статистическим критерием.

Как мы уже говорили, статистический критерий есть способ проверки статистических гипотез. Такая идея в целом ничем не отличается от любой научной области, в том числе и психологии. Ведь мы также располагаем массой, например, тестов, с помощью которых ищем ответы на вопросы о выраженности тех или иных психических качеств человека. Разница в данном случае заключается лишь в том, что выраженность у нас уже есть, а статистика ищет тенденции, закономерности, с тем, чтобы предсказание о всей генеральной совокупности было правомерным.

Общенаучность в статистическом подходе видна буквально с самых первых шагов ознакомления. А именно: метод проверки любой гипотезы строится так, что гипотезу невозможно подтвердить – её можно лишь опровергнуть, либо не опровергнуть. Не это ли торжество познавательной асимметричности знаменитого принципа фальсифицируемости Карла Поппера?

Так как принцип фальсифицируемости важен не только для верного понимания строения статистических критериев, но и для психологического знания в целом, рассмотрим его подробнее на примере. Договоримся о том, что у нас есть некое поле опыта, предположим, - кошки. Мы решили изучить их вес, измерили 20 кошек и обнаружили, что их средний вес равен 2 кг. В таком случае, мы можем сказать, что кошка – это животное, чей примерный вес равен 2 кг. Мы напечатали 20 статей на эту тему, выступили на конференциях, решаем прикладные задачи с помощью своей революционной теории и до того всех замучили своим открытием, что нас решили опровергнуть. Критики взялись измерять кошек своей независимой экспертной комиссией. День измеряют, два измеряют, три. А кошки-то все по 2 кг. Председатель экспертной комиссии уже было расстроился, но тут попалась кошка, чей вес 10 кг. Это никак не 2. И всё. Наша теория опровергнута. Естественно, что никто не отказывается от теории только потому, что попалась одна-единственная десятикилограммовая кошка. Собственно, почему? Только лишь потому, что нет идеальных теорий, которые бы могли со 100-% точностью предсказать условия для тех явлений, о которых повествуют. Даже теория Ньютона пала под натиском теории Эйнштейна, даже евклидова геометрия не выдержала присутствия Лобачевского. Но эта десятикилограммовая кошка означает, что для подтверждения чего-либо нужно поле опыта целиком (что зачастую невозможно), хотя при этом достаточно одного - единственного противоречащего факта для опровержения. Именно эта ситуация «несправедливости» подтверждения (верификации) и опровержения (фальсификации) теории легла в основу крылатого термина Поппера «познавательная асимметрия».

Именно по принципу фальсификации созданы и статистические критерии (хотя, не стоит думать, что математическая статистика обязана рождением Карлу Попперу). Если мы сформулируем две гипотезы, которые будут как несовместные события отражать явление целиком, то отвержение одной из них будет автоматически означать принятие другой. Например, если наша кошка может быть только чёрной или белой, то, выяснив, что она не белая, мы сможем вполне логично заключить, что она – чёрная.

Таким образом, статистический критерий проверяет по сути лишь одну гипотезу – нулевую (H_0), а в случае её отвержения принимает альтернативную (H_1).

Естественно, что при плохом освещении и без своих любимых очков мы будем принимать решение неуверенно. Степень уверенности характерна и статистическим критериям, что отражено в термине «уровень значимости». Взглянув на получившийся уровень значимости (p-level, signature), мы можем сказать, какая из гипотез не опроверглась. Уровень значимости есть вероятность, измеряемая в пределе от 0 до 1. Рубежным значением принято считать значение 0,05 (либо 0,01 при более строгой проверке). Если $p>0{,}05$, то принимается нулевая гипотеза, если $p<0{,}05$, то нулевая гипотеза отвергается и принимается альтернативная. Самое простое объяснение $p<0{,}05$ следующее: это 95% вероятность того, что нулевая гипотеза не верна. Правило, в котором уровень значимости соотносится с нулевой и альтернативной гипотезой, весьма полезно помнить, поскольку при самостоятельном ознакомлении с каким-либо новым критерием будет достаточно взглянуть на нулевую и альтернативную гипотезы данного критерия, как уже станет ясно, какой уровень значимости будет подтверждать предположение исследователя, а какой – отвергать.

Рассмотрим процесс принятия и отвержения гипотез на примере двух критериев.

Возьмём критерий Шапиро-Уилка. Этот критерий необходим для того, чтобы проверить эмпирические данные на нормальность.

H_0: случайная величина X распределена нормально.

H_1: для психолога альтернативная гипотеза такого критерия может быть сформулирована просто: случайная величина X распределена не нормально, хотя, конечно, в зависимости от того, каким именно ненормальным распределением является изучаемое, мы можем получить различную оценку критериев нормальности – они очень чувствительны к тому, против какой альтернативы смотрится нормальность. Так, например, логнормальное распределение – это достаточная трудность проверки, которая при использовании некоторых критериев может закончиться признанием нормальности. Однако для поставленной здесь задачи мы можем позволить себе ограничиться ненормальностью как альтернативой.

Итак, мы выяснили, что нулевая гипотеза Шапиро-Уилка – нормальность распределения. Предположим, что у нас есть исследование, в котором нужно посмотреть различие двух групп. Чтобы решить, можно ли использовать t-критерий Стьюдента для этой цели, мы должны проверить данные на нормальность. Это мы и делаем критерием Шапиро-Уилка. В таком случае, если уровень значимости этого критерия больше 0,05, то наши надежды оправдались и распределение нормально. Если уровень значимости меньше 0,05, то распределение не нормально и нам придётся использовать непараметрические методы проверки различия групп.

Вторым примером будет, например, t-критерий Стьюдента для независимых групп. Предположим, что предшествующая проверка нормальности распределения в обеих группах дала право использовать этот критерий (также мы проверили гипотезу об однородности дисперсий; удостоверились, что группы действительно независимы). В таком случае мы продолжаем свой анализ параметрическим критерием. Естественно, что нулевая исследовательская гипотеза, к которой мы стремимся своим исследованием – это, скорее всего, обнаружить различие, т.е. существует ли различие между группой 1 и 2. Однако, при переходе к статистической проверке, нам необходимо помнить, что статистическая гипотеза должна быть сформулирована с точностью до наоборот.

H_0 критерия Стьюдента (например, для независимых групп): $M_1 = M_2$.

H_1: $M_1 \neq M_2$.

M – это математические ожидания 1 и 2 случайной величины.

Это означает, что в нулевой гипотезе проверяется предположение о том, что средние значения двух генеральных совокупностей, из которых извлечены сравниваемые независимые выборки, равны.

В таком случае, мы видим, что применение t-Стьюдента будет интерпретироваться следующим образом: если $p>0,05$, то принимается нулевая гипотеза, если $p<0,05$, то нулевая гипотеза отвергается и принимается альтернативная. И в этом моменте никаких противоречий с предыдущим примером у нас нет. Но. В случае поиска различий исследователь надеется увидеть не значение >0,05, а меньше 0,05, потому что согласно исследовательской гипотезе он ищет не однородность групп, а их различие, в связи с чем уповает на альтернативную статистическую гипотезу.

Этот комментарий не является сложным, однако может всё же вызвать трудность у начинающих исследователей, которые чаще всего встречались только с теми критериями, применение которых ознаменовывалось надеждой на альтернативную гипотезу. Итого: значение оптимального уровня значимости не всегда должно быть меньше 0,05, скорее оно зависит от того, согласуется ли исследовательская гипотеза с нулевой или альтернативной статистической гипотезой.

Та же самая неточность встречается при защите курсовых и дипломных работ. Студент выходит и говорит, например: «Гипотеза подтвердилась, связь есть». Соответственно, психолог психолога поймёт, и мы начинаем отважно качать головами в знак согласия, но такая формулировка правомерна лишь в том, что исследовательская гипотеза подтвердилась. Статистическая, а именно – нулевая статистическая гипотеза – была отвергнута. Более того, альтернативная не подтвердилась, а была принята, потому как мы помним правило Поппера, с которым вряд ли имеет смысл спорить в столь сложных вопросах как законы случайных величин на малых выборках.

Существует два разных способа формулировать альтернативную гипотезу – направленный и ненаправленный. Рассмотрим пример. Предположим, что у нас есть две выборки, на которых мы получили данные об агрессивности. Проверив нормальность данных, однородность дисперсий, мы взялись за проверку статистической значимости различий критерием t-Стьюдента, который так часто попадает в примеры этого учебного пособия.

Направленная гипотеза всегда уточняет предполагаемый результат, не просто указывая эффект, но и его направление.

H_0 критерия Стьюдента (и в случае направленной, и в случае ненаправленной альтернативной гипотезы): $M_1 = M_2$, где M_1 и M_2 – это математические ожидания первой и второй группы.

В случае, если мы рассматриваем ненаправленную альтернативную гипотезу критерия Стьюдента, то она будет формулироваться следующим образом: $M_1 \neq M_2$.

В случае, если мы рассматриваем направленную альтернативную гипотезу критерия Стьюдента, то она может быть сформулирована двумя способами:

1) $M_1 > M_2$;

2) $M_2 > M_1$.

Рассмотрим пример, как воплощается на практике идея о ненаправленных и направленных альтернативных гипотезах.

Предположим, что у нас есть данные 1 «Г» и 1 «Е» класса по симпатии к кошкам. Мы решили узнать, можно ли считать симпатию разных классов статистически значимо различной.

2	1
3	1
5	1
6	1
7	1
8	1
12	1
14	1
15	1
3	2
4	2
6	2
7	2
10	2
12	2
13	2
14	2
16	2

В случае, если мы будем формулировать ненаправленную гипотезу, то вывод будет следующий:

t	-0.656809
DF	16
P	0.520640

Если же мы формулируем направленную альтернативную гипотезу, например, следующего вида: $M_1<M_2$, то нам нужно взять отрицательное значение t (в данном случае – (-0,656809)), уровень значимости будет другим: он будет равен половине уровня значимости двустороннего критерия, т.е. при направленной гипотезе p в данном случае будет равен 0,26032. Если мы формулируем направленную альтернативную гипотезу вида: $M_2 > M_1$, то в нашем случае нужно взять положительное значение t (0,656809), уровень значимости 0,26032. Чтобы разобраться с тем, отрицательное или положительное значение имеет критерий, достаточно один раз «сыграть» с данными. Если, например, обратиться к только что приведённому массиву данных, можно посчитать средние (точнее, просто на них взглянуть в статистической программе) и обнаружить, что $M_1 = 8$, а $M_2 = 9,44$. Если мы зададим группой 1 ту группу, где среднее равно 8, то эмпирическое значение критерия будет отрицательным, а если первой группой мы зададим ту группу, в которой среднее равно 9,44, то эмпирическое значение критерия будет положительным. Таким образом, видно, что в случае t-критерия Стьюдента знак эмпирическое значения не имеет исследовательского значения (он лишь говорит о том, о левой или правой части распределения мы говорим), в отличие, например, от коэффициента корреляции, где отрицательное значение строго отличается по смыслу от положительного.

Направленная гипотеза всегда имеет более высокий шанс отвергнуть нулевую гипотезу, чем ненаправленная.

Что касается использования двусторонних (ненаправленная гипотеза) и односторонних (направленная гипотеза) критериев, то правило достаточно размыто. Прикладники спорят и опасаются односторонних критериев, придумывают различные причины того, почему ими нельзя пользоваться. Самой неверной причиной неиспользования односторонних критериев является следующая: если исследователь ожидает увидеть, что первая группа меньше второй, а она на самом деле больше, то исследователь впадёт в заблуждение. Это маловероятно, поскольку у исследователя есть глаза, а односторонний критерий он всё равно будет задавать «вручную», исходя из данных об эмпирическом значении критерия и одностороннем уровне значимости. А вот другая причина неиспользования одностороннего критерия выглядит вполне разумной – без рациональных причин априорного обоснования направленности гипотезы мы можем совершить ошибку первого рода. С другой стороны, использование односторонних критериев призвано не затмить исследовательский мир ошибками первого рода, а справиться с ошибкой второго рода. Поэтому окончательный вердикт по использованию двусторонних и односторонних критериев мы вынесем после рассмотрения этих двух типов статистических ошибок.

Как мы помним, статистический критерий имеет две гипотезы – нулевую и альтернативную. Соответственно, нулевой гипотезой может быть, например, отсутствие эффекта, а альтернативной – присутствие эффекта.

Поскольку мы имеем дело со случайными величинами, то неуверенность математической статистики настолько велика, что некоторые математики вообще не считают эту отрасль своего знания научной. Если мы допускаем, что в наши вычисления может закрасться ошибка, то как она будет выглядеть? Совершенно очевидно, что она может быть двух видов – мы либо понапрасну решим, что эффект есть, либо что его нет.

Ошибка первого рода – это ошибочное отвержение нулевой гипотезы, а ошибка второго рода – это ошибочное принятие нулевой гипотезы. Зачастую это выражено следующей небольшой табличкой[18]:

		Верная гипотеза	
		H_0	***H_1***
Результат применения критерия	***H_0***	H_0 верно принята	H_0 неверно принята (Ошибка *второго* рода)
	H_1	H_0 неверно отвергнута (Ошибка *первого* рода)	H_0 верно отвергнута

Соответственно, возвращаясь к вопросу о двустороннем и одностороннем критерии мы видим попытки справиться именно с этой проблемой.

Двусторонний критерий помогает бороться с тем, что не отвергнуть H_0, когда на самом деле эффекта нет и т.п. А односторонний критерий помогает не пропустить иголку в стогу сена и зафиксировать, например, различие в тот момент, когда двусторонний критерий эту гипотезу не принял.

Как мы помним, односторонний критерий полезен задачам психодиагностики, что описано в главе о нормальном распределении. Что же касается исследования, то всё остаётся на совести исследователя. Вопиющее прикладное решение порой озвучивается следующим образом – каковы последствия той и другой ошибки в данной конкретной ситуации? Менее вопиющем и даже, пожалуй, обоснованным является только лишь распространённый совет априорного выдвижения обоснования гипотезы, но здесь я поясню, что не опора на теорию, а не предыдущие схожие исследования может быть достойным обоснованием использования одностороннего критерия.

В случае же, когда и двусторонний, и односторонний критерий выбирают одну и ту же гипотезу, то при защите работ или в публикациях

[18] Таблица взята с сайта: http://ru.wikipedia.org/wiki/%CE%F8%E8%E1%EA%E8_%EF%E5%F0%E2%EE%E3%EE_%E8_%E2%F2%EE%F0%EE%E3%EE_%F0%EE%E4%E0

стоит указывать значение двустороннего критерия, поскольку научное сообщество проявляет к нему большее доверие.

2. Программное обеспечение для обработки данных.

Первое столкновение с исследованием зачастую несёт в себе массу слёз, ужаса и ошибок по той причине, что студент просто не знает, с чего начать свой поиск. Моё первое знакомство с обработкой данных исследования началось с ночи мучений над тем, как посчитать различие 58 переменных. Программа почему-то активно сопротивлялась, а я не знала, что вообще туда нужно вводить. Так я познакомилась с критерием Манна-Уитни. Следующий ужас был не менее запоминающимся и сопровождался слезами над результатами статистической обработки критерием Спирмена – и куда же там смотреть? Ещё один стрессовый случай случился со мной, когда мой научный руководитель выслал мне, как было сказано в письме, результаты подсчёта (и, надо сказать, что к тому времени я уже ожидала увидеть совершённо определённого вида таблицу), но увидела лишь собственные сырые баллы. Каково же было моё удивление минут через 15 разглядывания собственной таблицы, что Excel столь удобен, что содержит листы, которые можно перелистывать. Такие ошибки могут постичь каждого новобранца, а моя задача – не забыть того, как это было в первый раз тяжело и страшно. Поэтому я начну с обзора программного обеспечения, которое для простых статистических задач является в сущности альтернативным. Следовательно, оценивать преимущество этих программ в начале своего пути стоит совсем по другим критериям, нежели приводятся в технической документации к этим продуктам.

2.1. Перечень программ для статистической обработки.

Статистическая программа для психолога-исследователя становится с течением времени настолько родной, что сменить одну на другую будет уже нелегко. Те, кто любят SPSS с неприязнью смотрят на любителей STATISTICA и наоборот. Такая ситуация объясняется довольно просто. Дело не в том, что 250-ый критерий, который умеет считать одна программа, не умеет считать другая. Вполне возможно, что вам никогда не понадобится этот 250-ый критерий. Здесь дело в первом знакомстве по типу того, которое описано выше. Тем, кто никогда не встречался со статистическими программами, эта ситуация будет хорошо знакома по борьбе Windows и

Linux, по трудности перехода от Windows XP к 7, от Microsoft Office 2003 к 2007 и т.д. Чтобы имело смысл переучиваться, должна быть более сильная мотивация, чем 250-ый критерий или какая-то там возможность, которая обычному пользователю ничего не говорит. В отношении статистических критериев единственным мотивом, который побудил меня открыть что-то, кроме STATISTICA, стала крайняя необходимость посчитать точечно-бисериальный коэффициент, который может считать только AtteStat, насколько мне известно. Поэтому мой совет прост: прежде чем начать всерьёз использовать какую-либо из приведённых ниже программ, стоит открыть каждую из них и посмотреть, удобен ли вам интерфейс, нравится ли расположение пунктов меню и т.п. Потому что только сейчас вы по-настоящему выбираете. Далее наступит инерция, которая может даже противоречить здравому смыслу. К столь же простым критериям, как интерфейс, относится логика создателей. Как бы вы делали статистическую программу, если бы представилась возможность? Если вы чувствуете себя в программе так, как в доме, где убирался кто-то другой, вам имеет смысл поискать ещё какой-нибудь продукт – возможно, что с ним будет более удобно.

Лидерами программного обеспечения в данной области на отечественном рынке без преувеличения можно назвать программы SPSS и STATISITICA.

<u>SPSS (Statistical Package for Social Science):</u> первая версия программы была создана ещё в 1968 году студентами-политологами Норманом Наем и Дейлом Бентом (далее к ним примкнул Хэдлай Халл), которые не могли довольствоваться ситуацией отсутствия электронной обработки статистической информации. Столь длинная история включает в себя версии программы, созданные для совершенно разных вычислительных машин, включая даже такие мифические ныне инструменты, как перфокарты. Признаком истории и динамики также может служить изменение названий продуктов – выпуск 2009 года носит название PASW (Predictive Analytics SoftWare), а с 2010 года, в связи с покупкой фирмой IBM, все продукты носят название: IBM SPSS.

В настоящее время существуют версии, созданные как для Windows (кроме 2000 версии), Linux (86), так и для довольно редко встречающихся у психологов операционных систем - леопардов Mac OS x10.5 и x10.6. Последняя на данный момент версия статистического пакета носит порядковый номер 19 и имеет русскоязычный интерфейс.

Учебники по SPSS: без преувеличения можно сказать, что печатные и электронные издания на русском языке по использованию данного программного продукта лидируют на книжном рынке. Стоит хотя бы даже обратить внимание на долю публикаций знаменитого издательского дома «Питер», относящуюся к SPSS, и сравнить с долей других опубликованных руководств по статистическим программам, чтобы понять это. Ниже приведены далеко не все печатные издания соответствующей тематики.

- Алексеева А.Ю., Ечевская Е.Г., Ковалёва Г.Д., Ростовцев П.С. «Анализ социологических данных с применением статистического пакета SPSS», Новосибирск: Редакционно-издательскй центр НГУ, 2003.

- Наследов А.Д. «SPSS: компьютерный анализ данных в психологии и социальных науках», Спб.: Питер, 2005, 2007.

- Наследов А.Д. «SPSS 15: профессиональный статистический анализ данных», Спб.: Питер, 2009.

- Наследов А.Д. «SPSS 19: профессиональный статистический анализ данных», Спб.: Питер, 2011.

- Таганов Д.Н. «SPSS: статистический анализ в маркетинговых исследованиях», Спб.: Питер, 2005.

- Хили Д. «Статистика. Социологические и маркетинговые исследования», Спб.: Питер, 2005.

Электронные ресурсы:

- Особенно хочется отметить труд Рейналя Левека и Антона Балабанова, где есть всё, что необходимо студенту, избравшему для себя SPSS. Присутствуют как иностранные статьи и интерактивные обучающие файлы, так и русские источники, а для любителей программирования имеются пояснения к синтаксису и макросам: http://www.spsstools.ru/spss.htm.

- Для нелюбителей читать более одной страницы пояснения, имеется электронный учебник с примерами из SPSS 11: http://www.learnspss.ru/.

Официальные сайты компании: англоязычная версия – http://www.spss.com (присутствует настройка русского языка для некоторых вкладок сайта, лучше начинать с неё); русскоязычная версия - http://www.spss.ru.

STATISTICA: разработана компанией StatSoft Inc., основанной в 1984 году. STATISTICA 6 была подготовлена уже специально для Windows NT/2000/XP и переведена на русский язык. Самый современный продукт этой фирмы – STATISTICA 10, но русского аналога ещё нет. Насколько мне известно, версии для Linux и других операционных систем отсутствуют, а также не стоит заблуждаться – STATISTICA 6 без дополнительных «ухищрений» не может быть установлена на Windows 7. Единственный продукт, который может быть использован с операционной системой Linux – это Statistica Enterprise Server Applications, что столь же прекрасно для простого студента, как и бесполезно. Что касается надстроек, которые могут быть освоены лишь любителями программирования, то во вкладке главного меню «Statistica» присутствует Statistica Visual Basic.

Учебные пособия: найти какой бы то ни было учебник по STATISTICA в книжных магазинах реально настолько же, насколько встретить пегаса. Кто сказал, что нет пегасов? Вот и книгу Боровикова В.П. наверняка можно будет найти. Особенно, когда будет переиздание. А другими словами, на данный момент книг по STATISTICA в продаже не осталось, по сему перехожу к электронным изданиям.

Электронные ресурсы:

- Стоит отметить, что официальный русскоязычный сайт содержит довольно исчерпывающий для начинающего пользователя интерактивный учебник, что практически полностью нормализует ситуацию с печатными изданиями.

- Портал Статосфера целиком и полностью посвящён программе STATISTICA. http://statosphere.ru/. Конечно, юмористическим наброском для отечественного читателя представляется статья о сравнении статистических программ, где против STATISTICA ставится Eviews, которая большому кругу исследователей социальных наук в принципе не нужна, поскольку является специализированной программой эконометрического анализа. Однако, в остальном, это очень полезный портал, где можно встретить такую уникальную попытку, как перевод видео курсов, посвящённых работе с программой. Здесь же представлены публикации о программе и статистике в целом.

- Отдельным пунктом стоит назвать статистические советники, представленные на официальном русскоязычном сайте во вкладке: http://www.statsoft.ru/statportal/tabID__43/DesktopDefault.aspx . Таковых компания предлагает пять: - советник по прогнозированию; - советник по многомерному анализу; - советник по нейронным сетям; - советник по STATISTICA Visual Basic; - медицинский советник. Первые четыре представляют собой краткие учебные руководства и не более, тем, кто столкнётся с такого рода задачами было бы полезно на них взглянуть. На медицинском советнике остановимся подробнее. Изначально идея статистического советника проста. Сталкиваясь с извечной путаницей в выборе необходимого критерия, мне приходится освещать правила, которые помогают студентам выбирать нужный вид анализа, постепенно отбрасывая те варианты, которые для данного конкретного случая теряют всяческий смысл. Ведь выбор критерия – это набор вопросов, поэтапно выдвигаемый к имеющимся данным и не более того. Конечно, рисуя столь радужную перспективу, я не имею в виду слишком сложных случаев, но в большинстве своём это достаточно несложная задача. Необходимость такого советника очевидна. Возвращаясь к медицинскому советнику, реализованному компанией Statsoft, я не рекомендую исследователю-психологу его использовать, если, конечно, не появится новой версии. Причины, по которым не стоит его использовать, помогут также выяснить, об одной ли версии советника мы говорим.

Во-первых, данный советник строится на предположении о том, что обе переменные измерены в одной шкале. За это не стоит винить именно статистического советника – на таком предположении строится большинство неспециализированных учебников по статистике, которые абсолютно игнорируют далеко не редко встречающуюся в психологии ситуацию, когда одна переменная измерена, например, в порядковой шкале, а другая – дихотомия и т.п.

Во-вторых, статистический советник предполагает, что вы уже знакомы с некоторыми, пусть и простыми с точки зрения компании, понятиями статистики, которыми обычные студенты, начинающие свои исследования, вполне возможно и не владеют в достаточной мере. Конечно, есть специальные разделы, поясняющие, например, значения терминов «независимая» и «зависимая» выборки и т.п., но данная ситуация свидетельствует, что пока нет минимальной базы, даже советник вам не поможет. И тип выборки не самое опасное место в данном случае, потому что нехватка этого знания будет очевидна при первом же использовании советника. Есть куда более опасные «подводные камни». Например, если мы выберем вопрос «Вам нужно исследовать две независимые группы пациентов», то следующим решающим вопросом будет шкала измерения.

Предположим, что обе наши переменные имеют метрическую шкалу измерения. Тогда ответом будет t-критерий Стьюдента для независимых групп, что само по себе не верно, если не соблюдено правило нормальности, о котором, конечно, написано в самом критерии Стьюдента, но если студент будет действовать безотчётно, не прочитав описание самого критерия, то совершит ошибку. К тому же t-критерий Стьюдента предполагает равенство дисперсий, которое также желательно проверять специализированным критерием.

Другой пример угрозы использования данного советника кроется недалеко. Предположим, что нас всё так же интересует различие независимых групп, которых теперь более двух, измерение порядковое. Нам посоветуют критерий Краскала-Уоллиса, что само по себе есть чистая истина, но не достаточная. После использования данного критерия психологам настоятельно рекомендуется использовать критерий Манна-Уитни с поправкой Бонферрони, без которого полученный результат является относительно бесполезным, потому что не отвечает на поставленный вопрос целиком.

Таким образом, использование данного советника может помочь в обработке данных исследования, а может и серьёзно навредить в зависимости от того, в чьи руки попадёт. Такая ситуация ничем не лучше самостоятельного выбора, а скорее хуже, поскольку в данном случае процесс принятия решения автоматизирован и иллюзорно выглядит куда более надёжным, чем самостоятельное действие.

Официальные сайты компании: англоязычная версия http://www.statsoft.com/#; русскоязычная версия - http://www.statsoft.ru/.

AtteStat: не столь известная программа, как SPSS или STATISTICA, но бесспорно имеющая свои преимущества. Во-первых, она имеет изначально русскоязычное меню. Во-вторых, представляет собой надстройку Excel, что позволяет не только быстро её устанавливать, но и творить такие чудеса, как работать с Windows 7, не будучи под него написанной. И, в-третьих, Игорь Гайдышев, разработчик программы, распространяет её бесплатно.

Дополнительные сайты и учебные пособия по данной программе мне не встречались, но руководство, встроенное в саму программу, носит достаточно исчерпывающий характер.

Официальный сайт (где, в том числе, доступны ссылки для бесплатного скачивания самой программы): http://attestatsoft.narod.ru/index.htm.

Rundom Pro: совершенно бесплатная программа, которая позволяет вычислять массу полезных и относительно редких показателей. Так как с ней мы будем знакомиться при подсчёте различных показателей, то комментарии излишни.

Официальный сайт: http://pjadw.tripod.com/index.htm

К тому же скачать программу можно, перейдя по ссылкам сайта: http://freestatistics.altervista.org/en/stat.php

OpenStat: разработчик программы – Уильям Миллер – в настоящий момент заслуженный профессор института Айовы. Хочется отметить доброту и искренность этого человека, которая просто восхищает. В предыдущих описаниях я давала в качестве обучающих ресурсов электронные порталы, здесь же можно просто написать почту разработчика, потому что обязательно получишь ответ на любой вопрос о статистике как науке или самом авторе. Работа над программой OpenStat началась в 1970-х годах и призвана была помочь студентам использовать электронные ресурсы в своих исследованиях. Позже был написан аналог LazStats для Linux. Если говорить о плюсах программы, то результаты подсчёта носят в некоторых случаях более подробный характер, чем у других статистических продуктов. Однако, бывают, конечно, и курьёзы. Так, например, если посчитать корреляцию Спирмена для 7 человек, то вы увидите значение самого коэффициента, но уровня значимости не будет, потому что таблица значимости t-критерия для такой выборки почему-то в программу не помещена. И, единственный совет, который даст программа – это посмотреть на 284 страницу учебника Сигэла, что, естественно, менее вероятное событие для студента, чем пересчитать другой статистической программой. Подводя итог, можно сказать, что программа носит самобытный характер, отражающий душу и логику создателя, но, если с этой логикой не совпасть, то лучше не использовать эту программу. К тому же, как я думаю, ясно, что все пояснения идут на английском языке. До июня 2011 года у этой программы было то преимущество, что она была абсолютно бесплатна для всех. В июне Уильям решил изменить своё решение и предлагает купить программу на диске почтой, что на данный момент не носит коммерческого характера, а возмещает расходы на пересылку и покупку диска. По словам самого автора, если до августа 2011 года не будут очевидны преимущества данного подхода, то программа вновь станет бесплатной. Ждём решения.

Официальный сайт: (существует только англоязычная версия) http://statprograms4u.com/. Стоит особенно отметить вкладку:

http://www.statprograms4u.com/Sources.htm, поскольку в ней содержатся ссылки на весьма полезные англоязычные сайты, в том числе содержащие другие бесплатные статистические продукты.

В целом список альтернативных в большинстве вопросов программ можно продолжать долго. Поэтому я ограничусь лишь тем, что приведу названия некоторых из них, чтобы у читателя сохранялся оптимизм найти наиболее подходящее себе программное обеспечение: Analyse-it, GenStat, JMR, MatLab, MINITAB, NCSS, Octave, PRISM, STADIA, STATA, STATGRAPHICS PLUS, StatPlus, SYSTAT и т.д. И, естественно, прежде чем искать что-то «на стороне», стоит попробовать выполнить это в MS Excel – возможно, что дальнейшие поиски будут излишни.

2.2. Дополнительный статистический инструментарий.

Как было сказано выше, для нужд студента-психолога, а порой даже и психолога-исследователя, бывает вполне достаточно выбрать какую-нибудь из вышеописанных программ, например, и обрабатывать исследования таким образом на протяжении всего поля опыта. Причина тому проста: любая корпорация, производящая статистические продукты ставит своей целью быть конкурентоспособной, а значит, «простые запросы» учитывает. Но есть запросы более специализированные, что побуждает меня представить некоторые продукты, раскрывая их возможности именно с точки зрения узких задач.

Вряд ли стоит проводить отдельный анализ, чтобы выяснить очевидный факт: корреляционные исследования занимают «самую большую долю рынка» психологических исследований. Углубляться в причины подобной картины я сейчас не буду, но лишь напомню, что неотъемлемой частью таких исследований являются корреляционные плеяды. Если Ваше исследование насчитывает 4 связи, то никаких проблем не возникает, но если хотя бы 20, то Вам понадобится помощь в их визуализации. Поэтому ниже приведено описание программы, которая помогает сделать процесс построения корреляционной плеяды весьма увлекательным занятием.

<u>Pajek:</u>

Поскольку программа распространяется свободно, то можно без каких-либо трудностей скачать её с сайта разработчиков:

http://vlado.fmf.uni-lj.si/pub/networks/pajek/

Либо с сайта Википедии: http://vlado.fmf.uni-lj.si/pub/networks/pajek/

Использовать Pajek для построения корреляционных плеяд легко, но необходимо подготовить файл соответствующего формата. Поэтому давайте внимательнее рассмотрим всю процедуру подготовки.

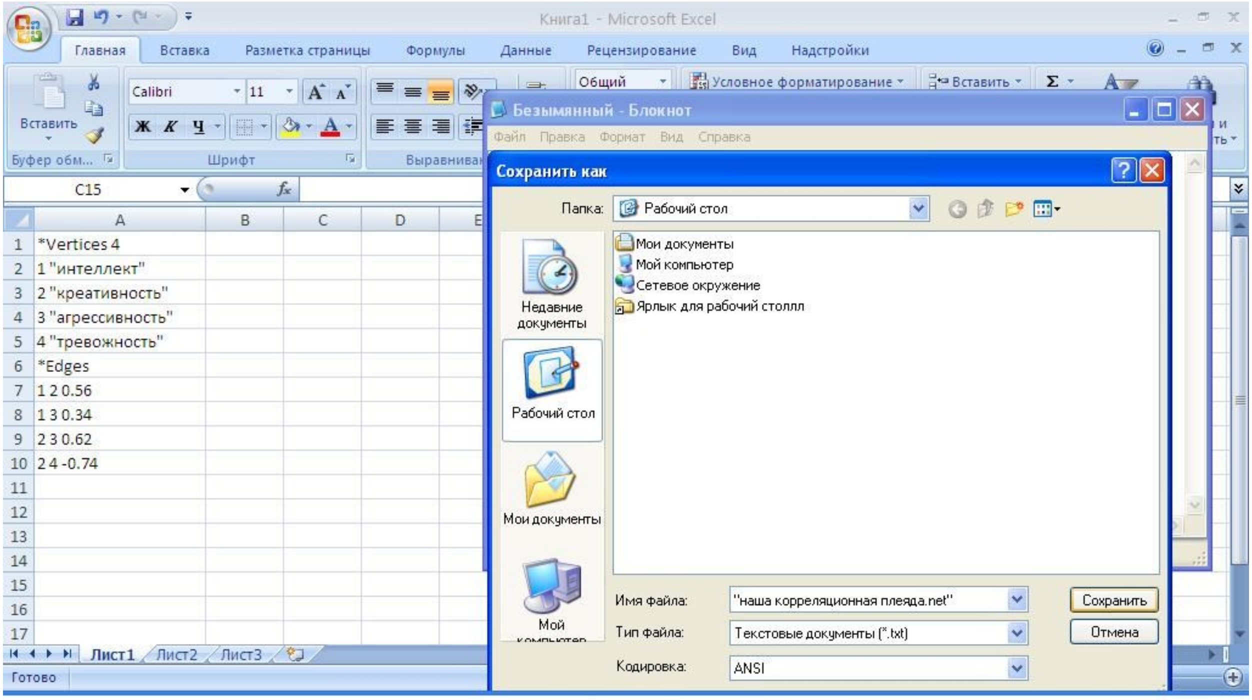
Для начала нам следует приготовить в Excel таблицу с уже посчитанными корреляциями. Располагать данные придётся в несколько необычной форме: нужно задать имена переменных (*Vertices) и значения корреляций (*Edges). Например:

*Vertices 4
1 "интеллект"
2 "креативность"
3 "агрессивность"
4 "тревожность"
*Edges
1 2 0.56
1 3 0.34
2 3 0.62
2 4 -0.74

Так, корреляция интеллекта и креативности в нашем вымышленном примере равна 0.56 и т.д. Обратите внимание, что вводить значение корреляций нужно через точку, а не через запятую. К тому же, если вам удобнее сначала вводить значение корреляций, а затем вписывать номера коррелирующих переменных, то нужно поставить апостроф перед отрицательным значением корреляции, иначе Excel будет выдавать ошибку. Т.е. вот так: ‘-0.74, но в последующих шагах, когда вводятся номера переменных, нужно внимательно проследить, чтобы апострофов больше не было, потому что Pajek не распознает апостроф, что в итоге приведёт к исчезновению отрицательных корреляций из корреляционной плеяды.

Далее нам нужно перенести созданную таблицу в блокнот (в Windows XP эта встроенная программа находится по пути: Пуск -> Все программы -> Стандартные -> Блокнот).

В открывшееся окно блокнота нужно просто скопировать таблицу из Excel и сохранить определённым образом: нужно задать расширение .net. Делается это просто. Нужно нажать на кнопку «сохранить» или «сохранить как» во вкладке меню «файл». И ввести в само название необходимое расширение, но для того, чтобы блокнот не превратил это расширение обратно в текстовый формат, нужно название закавычить целиком, например “наша корреляционная плеяда.net”:

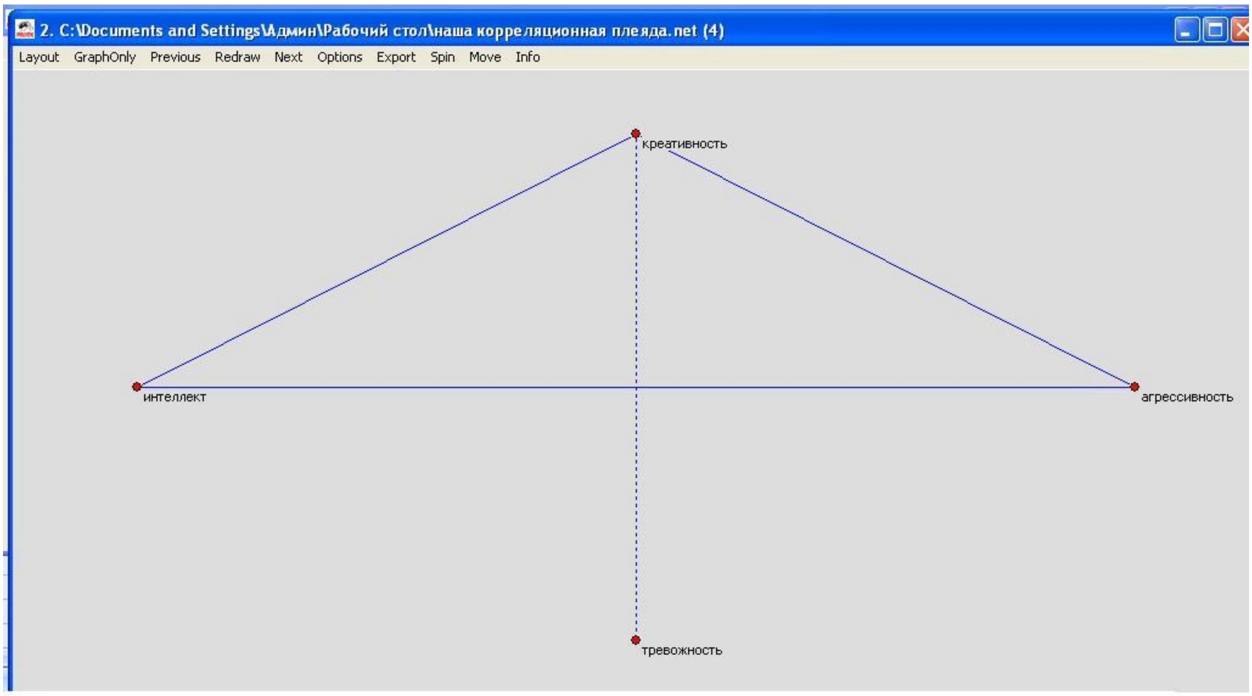


После того, как мы получили совершенно бесполезный файл, который даже сам блокнот теперь прочитать не может, можно приступать к завершающему шагу – построению самой плеяды.

Для этого открываем программу Pajek, которую вы благополучно установили, скачав с сайта создателей. Ярлык этой программы находится либо на рабочем столе, если вы его перенесли туда, либо в папке c/pajek, либо там, куда вы сохранили программу, изменив директорию при установке.

Я привожу примером версии 200, но этого не стоит бояться, потому что основные шаги от версии к версии не меняются.

Нам нужно нажать следующие кнопки меню программы: File -> Network -> Read. Выбрать путь к своему файлу и нажать «открыть». На белом листе появится отчёт о прочтении файла, мы его закрываем. Если при чтении возникает ошибка, то программа весьма прозрачно на это намекнёт – она начнёт кричать, выдавать ошибки и вообще может не дать себя закрыть, что придётся делать через знаменитый Ctrl+Alt+Delete. Но в случае нашего показательного примера всё в полном порядке, поэтому мы идём дальше. Находим следующие вкладки меню: Draw -> Draw. И появляется долгожданное чудо:



Как видно из плеяды, если присутствуют отрицательные корреляции, то они обозначаются пунктирной линией, в то время как положительные значения обозначаются сплошной линией.

Получившийся миловидный домик на самом деле можно сделать практически любого цвета и разместить точки переменных в любые доступные места экрана, просто передвигая их мышью.

Для изменения цвета тех или иных составляющих плеяды: Options -> Colors -> …

Для того, чтобы рядом с вершинами графа, которые представлены коррелирующими переменными, появились или исчезли сами названия шкал: Options -> Mark Vertices Using -> Labels (или No Labels).

Для того, чтобы рядом с рёбрами графа, обозначающими корреляции, появились надписи о числовом значении: Options -> Lines -> Mark Lines -> With Values.

Экспортирование корреляционной плеяды в виде картинки расширения .bmp, которое является хорошо всем знакомой альтернативой .jpg, делается следующим образом: Export -> 2D -> Bitmap.

<u>MiniStep:</u>

MiniStep (демо-версия Winsteps) – является полезным инструментом для конвертации сырых баллов в шкалу Раша.

Скачать программу можно с сайта разработчика: http://www.winsteps.com/ministep.htm

Стоит отметить, что демо-версия (MiniStep) работает аналогично коммерческой версии, с той лишь разницей, что предполагает введение до 25 вопросов и 75 испытуемых. Ограничения срока использования демо-версии нет.

Достаточно удобно переносить сырые данные как из SPSS, так и из Excel, хотя этими программами, как хорошо видно из диалогового окна, способности MiniStep по импорту данных не ограничиваются. Мы будем рассматривать вариант Excel, потому как представить себе компьютер без этой программы сложно.

Для начала приготовим какой-нибудь вымышленный пример в Excel. Предположим, что у нас есть тест для любителей овощей, оцениваемый каждым испытуемым на предмет симпатии к ним шкалой с баллами от 1 до 5 баллов.

В таком случае в Excel необходимо сделать таблицу, где первой строкой ОБЯЗАТЕЛЬНО будут значиться названия вопросов опросника (можно использовать просто цифры), иначе при конвертации в MiniStep мы потеряем первого испытуемого. Если при этом мы хотим, чтобы имена испытуемых в последующих вариантах анализа также были видны, а не представлены номерами, то их также нужно внести в таблицу.

имена	морковь	свёкла	картофель	лук	брюква
Василий	5	1	2	4	4
Георгий	2	5	4	2	3
Татьяна	3	4	2	1	3
Нина	5	1	2	3	4
Раиса	3	2	3	4	3
Николай	2	1	5	3	3
Мария	1	1	1	2	2
Пётр	3	2	4	2	1
Лариса	2	3	4	2	3
Евгений	4	4	3	2	1
Анатолий	5	5	5	2	2
Ирина	3	4	4	2	1
Ольга	3	5	5	4	2
Константин	2	4	4	1	1
Александр	4	3	5	4	3

Елена	3	4	4	2	1
Анастасия	3	2	4	5	3
Эдит	2	4	4	2	2
Клод	5	4	5	5	4
Брюс	2	1	2	2	1
Моника	3	5	3	3	4
Жюль	3	2	4	4	1
Билл	5	3	3	4	2
Борис	2	3	1	4	5
Белла	4	4	2	3	3
Поль	5	5	5	1	1
Полина	3	3	4	3	2
Зигмунд	4	5	5	3	2
Эмма	2	4	1	3	3
Ингеборга	3	3	2	1	1

В случае, если конвертация проводится из SPSS, то первую строку и столбец внутри программы SPSS обозначать названиями пунктов и именами респондентов не обязательно, хотя если этого не сделать, то названия пунктов будут обозначаться привычным для SPSS способ – Var1 и т.д, а имена респондентов будут выражаться номерами. Остальные шаги по конвертации из Excel и SPSS совпадают.

Затем мы сохраняем этот файл под каким-нибудь узнаваемым для нас названием и переходим к программе MiniStep.

При запуске программы появляется диалоговое меню, в котором, помимо приветствия, даётся возможность, например, импортировать данные, что нам и нужно сделать. Итак, мы нажимаем кнопку «Import from Excel, …».

Далее снова выбираем вариант Excel, после чего появляется пустое окно.

Мы нажимаем кнопку «Select Excel file», находим созданный Excel-файл, нажимаем «открыть» и ждём.

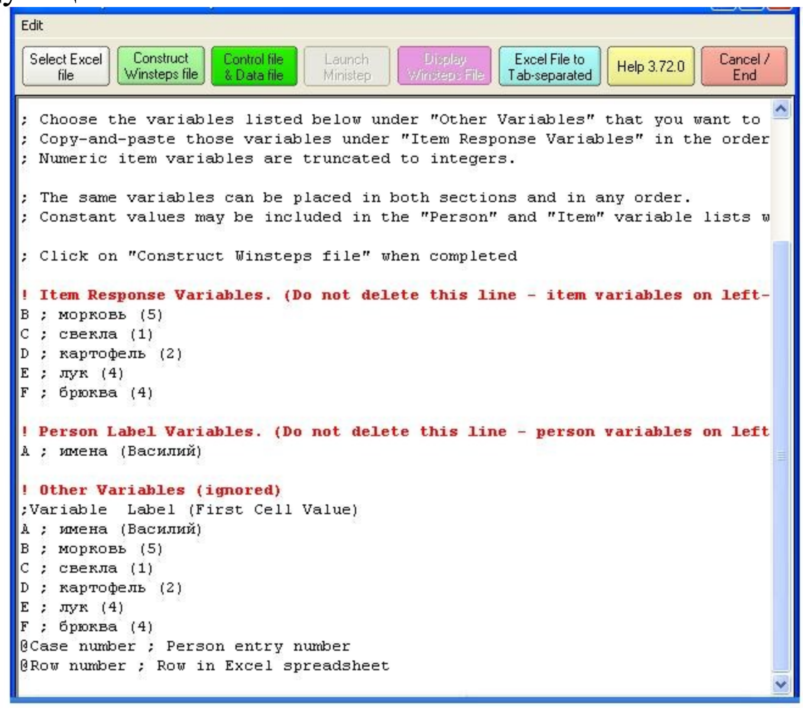
В итоге в нашем до этого момента пустом листе появляется большое количество текста. В конце этого текста есть красная строка «Other Variables (ignored)». Нам понадобится всё, что после строки «;Variable Label (First Cell Value)» и до строки «@Case number ; Person entry number» скопировать. Это наши переменные:

```
A ; имена (Василий)
B ; морковь (5)
C ; свекла (1)
D ; картофель (2)
E ; лук (4)
F ; брюква (4)
```

Далее нам нужно скопировать и вставить строку «A; имена (Василий)» после строки «! Person Label Variables. (Do not delete this line - person variables on left…».

Строки «морковь», «свекла» и т.д. нужно скопировать и вставить после красной надписи «! Item Response Variables. (Do not delete this line - item variables on left…».

Получается следующее:



Далее нажимаем «Construct Winsteps file», сохраняем под каким-нибудь именем и можем закрыть окно появившегося блокнота, а также окно, где копировали и вставляли переменные.

Переходим в оставшееся окно, где нас спрашивают «Control file name…» и нажимаем клавишу Enter. Указываем место, где хранится только что сделанная усилиями самого MiniStep таблица.

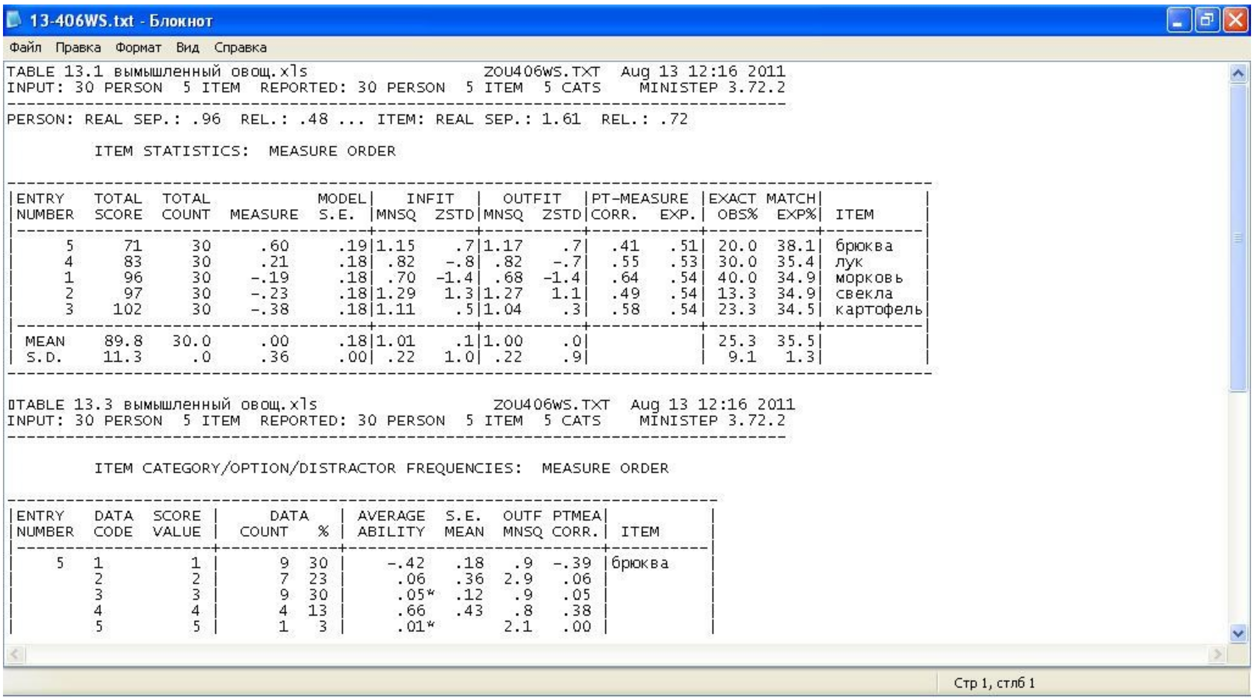
На вопрос «Report output file name (or press Enter for temporary file, Ctrl+O for Dialog Box)» нажимаем Enter в случае, если не хотим сохранять полученные результаты, Ctrl+O, если хотим сохранить. Но мне кажется, что особой необходимости сохранять эти данные нет. Гораздо полезнее сохранить результаты, о которых пойдёт речь ниже.

На вопрос «Extra specifications (if any). Press Enter to analyze» просто нажимаем Enter. Появился подсчёт.

Далее нажимаем тот вариант представления данных, который необходим.

Например, выберем следующий: Output Tables -> 13. Item: measure.

Появится следующая таблица:



```
13-406WS.txt - Блокнот
Файл  Правка  Формат  Вид  Справка
TABLE 13.1 вымышленный овощ.xls                     ZOU406WS.TXT  Aug 13 12:16 2011
INPUT: 30 PERSON  5 ITEM  REPORTED: 30 PERSON  5 ITEM  5 CATS    MINISTEP 3.72.2
--------------------------------------------------------------------------------
PERSON: REAL SEP.: .96  REL.: .48 ... ITEM: REAL SEP.: 1.61  REL.: .72

         ITEM STATISTICS:  MEASURE ORDER

-----------------------------------------------------------------------------------------------
|ENTRY   TOTAL  TOTAL           MODEL|   INFIT    |  OUTFIT   |PT-MEASURE |EXACT MATCH|          |
|NUMBER  SCORE  COUNT  MEASURE  S.E. |MNSQ  ZSTD|MNSQ  ZSTD|CORR.  EXP.| OBS%  EXP%| ITEM     |
|------------------------------------+----------+----------+-----------+-----------+----------|
|     5     71     30      .60    .19|1.15    .7|1.17    .7|  .41   .51| 20.0  38.1| брюква   |
|     4     83     30      .21    .18| .82   -.8| .82   -.7|  .55   .53| 30.0  35.4| лук      |
|     1     96     30     -.19    .18| .70  -1.4| .68  -1.4|  .64   .54| 40.0  34.9| морковь  |
|     2     97     30     -.23    .18|1.29   1.3|1.27   1.1|  .49   .54| 13.3  34.9| свекла   |
|     3    102     30     -.38    .18|1.11    .5|1.04    .3|  .58   .54| 23.3  34.5| картофель|
|------------------------------------+----------+----------+-----------+-----------+----------|
| MEAN    89.8   30.0      .00    .18|1.01    .1|1.00    .0|           | 25.3  35.5|          |
| S.D.    11.3     .0      .36    .00| .22   1.0| .22    .9|           |  9.1   1.3|          |
-----------------------------------------------------------------------------------------------

TABLE 13.3 вымышленный овощ.xls                      ZOU406WS.TXT  Aug 13 12:16 2011
INPUT: 30 PERSON  5 ITEM  REPORTED: 30 PERSON  5 ITEM  5 CATS    MINISTEP 3.72.2
--------------------------------------------------------------------------------

         ITEM CATEGORY/OPTION/DISTRACTOR FREQUENCIES:  MEASURE ORDER

-------------------------------------------------------------------------
|ENTRY   DATA  SCORE |     DATA    | AVERAGE  S.E.  OUTF PTMEA|          |
|NUMBER  CODE  VALUE |  COUNT    % | ABILITY  MEAN  MNSQ CORR.| ITEM     |
|--------------------+-------------+--------------------------+----------|
|    5   1         1 |      9   30 |    -.42   .18    .9  -.39 |брюква    |
|        2         2 |      7   23 |     .06   .36   2.9   .06 |          |
|        3         3 |      9   30 |     .05*  .12    .9   .05 |          |
|        4         4 |      4   13 |     .66   .43    .8   .38 |          |
|        5         5 |      1    3 |     .01*        2.1   .00 |          |
Стр 1, стлб 1
```

Чтобы сохранить этот документ достаточно нажать на вкладку меню «файл» - > «сохранить как».

Возникает коронный вопрос – куда смотреть?

Entry Number: это номера пунктов нашего вымышленного опросника. Хорошей проверкой является колонка Item – если в ней оказались названия вопросов (или номера, если пункты были обозначены номерами), то всё сделано правильно. Если же там написано что-нибудь типа: 1 1 2 2 4, то вместо названий пунктов опросника программа подставила данные первого испытуемого, поскольку при создании таблицы Excel вы забыли ввести в первую строку названия пунктов. Итак, Entry Number и Item – это названия пунктов.

Total score: сумма сырых баллов.

Total count: количество испытуемых. Если, опять же, там неожиданное количество, то, скорее всего, напутано с названием пунктов при конвертации.

Measure: логиты, т.е. единицы шкалы Раша, подобно тому, как в самом начале у нас были сырые баллы 5-балльной шкалы Лайкерта.

Model S.E.: ошибка измерения модели Раша.

MNSQ: соответствие/несоответствие результатов самой модели.

ZSTD: стандартизованные значения MNSQ (где среднее равно 0, а стандартное отклонение 1).

PT-measure corr. – бисериальная корреляция, отражающая согласованность пунктов опросника.

<u>Анализ результатов по пунктам опросника:</u>

Всё зависит от задачи. Например, мы хотим узнать, какой овощ является самым любимым. Здравый смысл, направленный на получившуюся таблицу, сказал бы нам, что самый любимый овощ людей – это брюква. На самом деле это не так, ведь даже для вымышленного примера такой результат был бы ужасной ложью. Конечно же, это картофель.

<u>Внимание:</u> согласно логике создателя этой модели, наибольшее количество баллов получает наименьшее количество логитов. Это достаточно неудобно для обыденных задач, поэтому есть два выхода из сложившейся ситуации.

1) Можно конвертировать исходные данные подобно тому, как конвертируются пункты с обратной формулировкой вопросов. Если мы имеем шкалу от 1 до 5, то 5 превращается в 1, 4 в 2, 3 в 3, 2 в 4, 1 в 5. Прошу, опять же, обратить внимание на то, что в нашем примере шкала задумывалась именно 5-балльной, т.е. без 0 как возможности ответа. Если же шкала подразумевает такое значение, то она получается 6-балльной, а значит, 0 будет превращаться в 1, 1 в 4, 2 в 3 и т.д.

После конвертации у нас получается следующая таблица исходных данных:

имена	морковь	свёкла	картофель	лук	брюква
Василий	1	5	4	2	2
Георгий	4	1	2	4	3
Татьяна	3	2	4	5	3
Нина	1	5	4	3	2
Раиса	3	4	3	2	3
Николай	4	5	1	3	3
Мария	5	5	5	4	4

Пётр	3	4	2	4	5
Лариса	4	3	2	4	3
Евгений	2	2	3	4	5
Анатолий	1	1	1	4	4
Ирина	3	2	2	4	5
Ольга	3	1	1	2	4
Константин	4	2	2	5	5
Александр	2	3	1	2	3
Елена	3	2	2	4	5
Анастасия	3	4	2	1	3
Эдит	4	2	2	4	4
Клод	1	2	1	1	2
Брюс	4	5	4	4	5
Моника	3	1	3	3	2
Жюль	3	4	2	2	5
Билл	1	3	3	2	4
Борис	4	3	5	2	1
Белла	2	2	4	3	3
Поль	1	1	1	5	5
Полина	3	3	2	3	4
Зигмунд	2	1	1	3	4
Эмма	4	2	5	3	3
Ингеборга	3	3	4	5	5

Для неё мы создаём другой файл, конвертированный в MiniStep и выбираем нужные опции представления результатов, как делали это выше.

Если мы сравним данные, полученные при анализе прямых и обратных баллов, то получим следующее:

item measure для овоща.txt - Блокнот

Файл Правка Формат Вид Справка

```
PERSON: REAL SEP.: .96  REL.: .48 ... ITEM: REAL SEP.: 1.61  REL.: .72

        ITEM STATISTICS:  MEASURE ORDER

-------------------------------------------------------------------------------------------
|ENTRY   TOTAL  TOTAL           MODEL|   INFIT    |  OUTFIT   |PT-MEASURE |EXACT  MATCH|         |
|NUMBER  SCORE  COUNT  MEASURE  S.E. |MNSQ  ZSTD|MNSQ  ZSTD|CORR.  EXP.| OBS%  EXP%| ITEM    |
|------------------------------------+----------+----------+-----------+-----------+---------|
|     5     71     30      .60     .19|1.15    .7|1.17    .7|  .41   .51| 20.0  38.1| брюква  |
|     4     83     30      .21     .18| .82   -.8| .82   -.7|  .55   .53| 30.0  35.4| лук     |
|     1     96     30     -.19     .18| .70  -1.4| .68  -1.4|  .64   .54| 40.0  34.9| морковь |
|     2     97     30     -.23     .18|1.29   1.3|1.27   1.1|  .49   .54| 13.3  34.9| свекла  |
|     3    102     30     -.38     .18|1.11    .5|1.04    .3|  .58   .54| 23.3  34.5| картофель|
|------------------------------------+----------+----------+-----------+-----------+---------|
| MEAN    89.8   30.0      .00     .18|1.01    .1|1.00    .0|           | 25.3  35.5|         |
| S.D.    11.3     .0      .36     .00| .22   1.0| .22    .9|           |  9.1   1.3|         |
-------------------------------------------------------------------------------------------
```

item measure для инвертированного овоща.txt - Блокнот

Файл Правка Формат Вид Справка

```
TABLE 13.1 инвертированный овощ.xls              ZOU002WS.TXT  Aug 12 15:16 2011
INPUT: 30 PERSON  5 ITEM  REPORTED: 30 PERSON  5 ITEM  5 CATS    MINISTEP 3.72.2
--------------------------------------------------------------------------------
PERSON: REAL SEP.: .96  REL.: .48 ... ITEM: REAL SEP.: 1.61  REL.: .72

        ITEM STATISTICS:  MEASURE ORDER

-------------------------------------------------------------------------------------------
|ENTRY   TOTAL  TOTAL           MODEL|   INFIT    |  OUTFIT   |PT-MEASURE |EXACT  MATCH|         |
|NUMBER  SCORE  COUNT  MEASURE  S.E. |MNSQ  ZSTD|MNSQ  ZSTD|CORR.  EXP.| OBS%  EXP%| ITEM    |
|------------------------------------+----------+----------+-----------+-----------+---------|
|     3     78     30      .38     .18|1.11    .5|1.04    .3|  .58   .54| 23.3  34.5| картофель|
|     2     83     30      .23     .18|1.29   1.3|1.27   1.1|  .49   .54| 13.3  34.9| свекла  |
|     1     84     30      .19     .18| .70  -1.4| .68  -1.4|  .64   .54| 40.0  34.9| морковь |
|     4     97     30     -.21     .18| .82   -.8| .82   -.7|  .55   .53| 30.0  35.4| лук     |
|     5    109     30     -.60     .19|1.15    .7|1.17    .7|  .41   .51| 20.0  38.1| брюква  |
|------------------------------------+----------+----------+-----------+-----------+---------|
| MEAN    90.2   30.0      .00     .18|1.01    .1|1.00    .0|           | 25.3  35.5|         |
| S.D.    11.3     .0      .36     .00| .22   1.0| .22    .9|           |  9.1   1.3|         |
-------------------------------------------------------------------------------------------

TABLE 13.3 инвертированный овощ.xls              ZOU002WS.TXT  Aug 12 15:16 2011
```

Как видно из этого сравнения, полностью меняется колонка Total score, поскольку сумма сырых баллов вполне естественно меняется при инвертировании сырых баллов. Также меняется и «Mean S.D.» для колонки Total Score, поскольку среднее и стандартное отклонение неотъемлемо зависят от тех данных, на которых идёт их подсчёт.

Баллы шкалы Раша меняют знак – то, что в колонке Measure имеет положительный знак, в случае инвертирования приобретёт отрицательный знак и наоборот.

Всё остальное в рассматриваемой таблице просто изменяет своё место, поскольку меняется место пункта, которому этот подсчёт соответствует.

По инвертированным данным мы можем дать прямую интерпретацию – самым любимым овощей нашей выборки является картофель, затем свекла и т.д.

Соответственно, подробно рассмотрев получившиеся изменения от инвертирования сырых баллов, мы подходим ко второму способу видоизменения подсчётов для более удобной интерпретации.

2) Если таблица исходных данных слишком велика и исследователю представляется слишком трудоёмким изменять каждое значение сырых баллов, то можно просто помнить о том, что отрицательные значения колонки Measure представляют собой положительное значение и наоборот.

Или просто читать колонку Item «снизу», если интересует вопрос самых «симпатичных» пунктов опросника.

<u>Анализ результатов опросника по испытуемым:</u>

Чтобы сделать анализ по испытуемым, необходимо обратиться, например, к пункту: Output Tables -> 17. Person: measure.

Мы сделаем это для сырых баллов, как они были даны в первой таблице (т.е. без инвертирования), логичность чего будет ясна чуть позже.

Итак, таблица анализа будет выглядеть следующим образом:

17-406WS.txt - Блокнот

Файл Правка Формат Вид Справка

```
TABLE 17.1 вымышленный овощ.xls                  ZOU406WS.TXT  Aug 13 12:16 2011
INPUT: 30 PERSON  5 ITEM  REPORTED: 30 PERSON  5 ITEM  5 CATS    MINISTEP 3.72.2
--------------------------------------------------------------------------------
PERSON: REAL SEP.: .96  REL.: .48 ... ITEM: REAL SEP.: 1.61  REL.: .72

         PERSON STATISTICS:  MEASURE ORDER

-------------------------------------------------------------------------------------------------
|ENTRY   TOTAL  TOTAL           MODEL|   INFIT    |  OUTFIT   |PT-MEASURE |EXACT  MATCH|         |
|NUMBER  SCORE  COUNT  MEASURE  S.E. |MNSQ  ZSTD|MNSQ  ZSTD|CORR.  EXP.| OBS%  EXP%| PERSON   |
|------------------------------------+----------+----------+-----------+-----------+----------|
|    19     23      5    1.92    .71| .50   -.4| .57   -.2|  .43   .24| 80.0  64.6| -лод     |
|    11     19      5     .73    .45|1.32    .7|1.27    .6|  .92   .36|   .0  35.2| -натолий |
|    13     19      5     .73    .45| .83   -.1| .85   -.1|  .77   .36| 20.0  35.2| +льга    |
|    15     19      5     .73    .45| .40  -1.2| .46   -.9|  .57   .36| 60.0  35.2| -лександр|
|    28     19      5     .73    .45| .67   -.4| .66   -.4|  .97   .36| 20.0  35.2| Зигмунд  |
|    21     18      5     .53    .43| .82   -.2| .84   -.1| -.10   .37| 20.0  34.2| |оника   |
|    17     17      5     .35    .42|1.18    .5|1.15    .4| -.14   .37| 40.0  33.5| -настасия|
|    23     17      5     .35    .42| .73   -.4| .74   -.3|  .43   .37|   .0  33.5| -илл     |
|    26     17      5     .35    .42|2.19   1.9|2.13   1.8|  .92   .37|   .0  33.5| Поль     |
|     1     16      5     .18    .41|2.51   2.2|2.53   2.3| -.48   .37|   .0  31.7| -асилий  |
|     2     16      5     .18    .41| .97    .1| .95    .1|  .41   .37| 20.0  31.7| |еоргий  |
|    25     16      5     .18    .41| .62   -.7| .64   -.6|  .03   .37| 20.0  31.7| -елла    |
|     4     15      5     .01    .41|2.26   2.0|2.27   2.0| -.41   .37| 20.0  31.4| -ина     |
|     5     15      5     .01    .41| .67   -.5| .67   -.6| -.38   .37| 40.0  31.4| Раиса    |
|    24     15      5     .01    .41|2.80   2.5|2.94   2.7| -.94   .37| 20.0  31.4| -орис    |
|    27     15      5     .01    .41| .10  -3.0| .10  -2.9|  .87   .37| 80.0  31.4| Полина   |
|     6     14      5    -.16    .41|1.59   1.1|1.53   1.1|  .05   .36| 20.0  32.0| -иколай  |
|     9     14      5    -.16    .41| .45  -1.1| .47  -1.1|  .30   .36| 20.0  32.0| -ариса   |
|    10     14      5    -.16    .41| .59   -.8| .62   -.7|  .88   .36| 20.0  32.0| +вгений  |
|    12     14      5    -.16    .41| .51  -1.0| .54   -.9|  .97   .36| 20.0  32.0| -рина    |
|    16     14      5    -.16    .41| .51  -1.0| .54   -.9|  .97   .36| 20.0  32.0| +лена    |
|    18     14      5    -.16    .41| .47  -1.1| .45  -1.1|  .70   .36| 20.0  32.0| Эдит     |
|    22     14      5    -.16    .41| .87   -.1| .90    .0|  .55   .36| 20.0  32.0| Жюль     |
|     3     13      5    -.34    .42| .94    .1| .98    .2|  .14   .35| 20.0  33.7| Татьяна  |
|    29     13      5    -.34    .42|1.35    .8|1.32    .7| -.41   .35|   .0  33.7| Эмма     |
```

Стр 1, стлб 1

Хорошо видны несоответствия кодирования кириллицы, но это никак не влияет на способность исследования понимать, о каком испытуемом идёт речь. Если, конечно, в таблице нет одинаковых имён, хотя этому недоразумению можно найти простое решение – посмотреть в колонку Entry number и сравнить номер одноимённых испытуемых с исходной Excel-таблицей.

Мы взяли для анализа именно исходную таблицу данных, поскольку количество логитов и последующая интерпретация полученных результатов должна идти в случае анализа испытуемых именно в прямом порядке.

Внимание: наибольшее количество баллов в данном случае получает наибольшее количество логитов. Ошибка будет стоить вам обратной интерпретации.

В нашем случае Клод действительно любит овощи, приведённые в опроснике, больше, чем Анатолий.

Чтобы понять, прямая или обратная интерпретация логитов предполагается в каждом виде анализа, достаточно посмотреть в колонку Total Score. В случае, когда мы анализировали сами овощи, сумма сырых баллов возрастала. В случае же, когда мы перешли к анализу испытуемых, сумма баллов выстроена в убывающем порядке.

Total Count: количество пунктов опросника, заполненное каждым испытуемым, на основе которых и выводятся последующие результаты.

Таким образом, видно, что колонки имеют тот же смысл, что и при анализе пунктов опросника.

2.3. Проверка нормальности распределения (AtteStat).

Для того, чтобы принять столь важное для последующего анализа решение, достаточно проделать следующие шаги.

Как известно, проверка нормальности чаще всего идёт для каждой шкалы отдельно. Представим, что в школе проводили тест интеллекта и получили следующие данные общего IQ:

25
105
125
39
60
70
98
75
131
125
111
110
124
143

132
119
93
81
59
119

Итого: в нашем классе 20 ребят. К смысловому значению полученных данных относиться стоит исключительно с юмором.

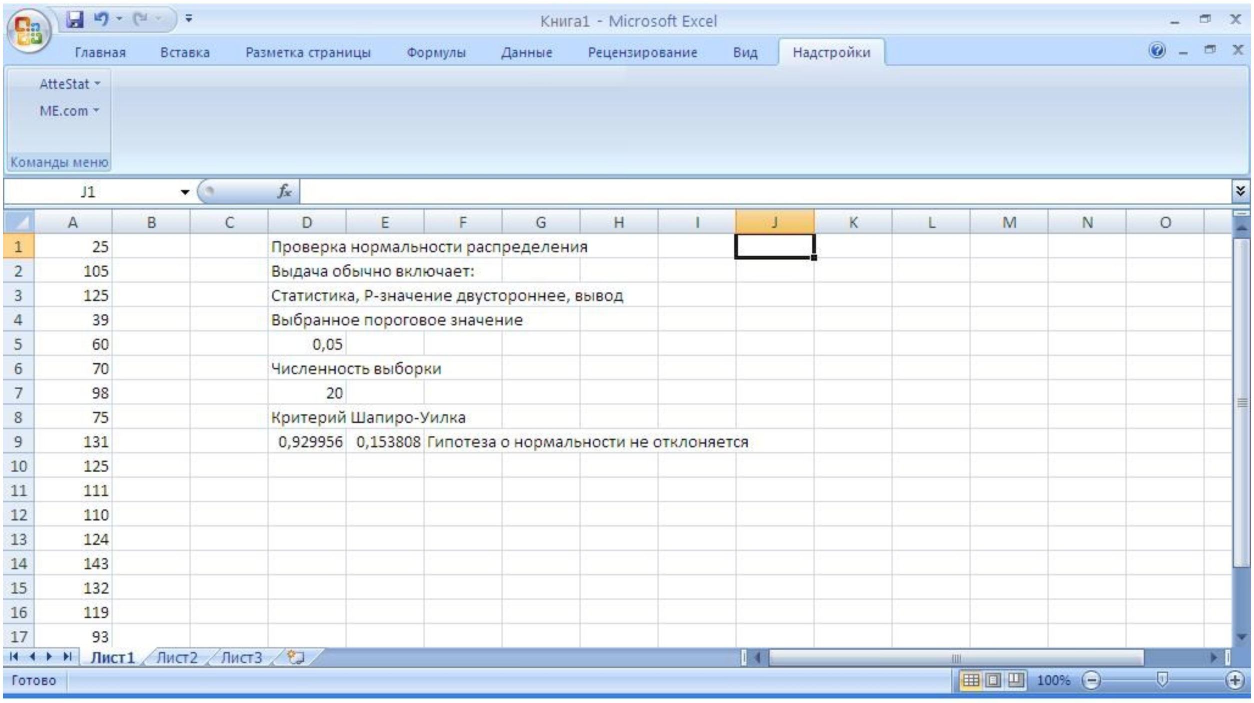
В моей версии MS Office программа AtteStat находится во вкладке «Надстройки». В открывшемся меню программы нужно выбрать модуль «Проверка нормальности» -> Проверка нормальности.

Интервал выборки – выделяем все имеющиеся данные теста интеллекта.

Интервал вывода – ту ячейку Excel, где мы хотели бы видеть вывод анализа.

Выбираем тест Шапиро-Уилка (по умолчанию стоит уровень значимости 0,05, но если по каким-то причинам необходимо сделать критерий более жёстким, значение можно изменить), «Выполнить расчёт».

Появляется примерно такая картина:



Согласно расчётам, эмпирическое значение критерия Шапиро-Уилка равно 0,929956. Уровень значимости равен 0,153808. Удобство программы заключается в том, что она совершенно русскими буквами пишет, что гипотеза о нормальности не отклоняется, т.е. распределение является нормальным. И такая надпись будет появляться каждый раз, когда уровень значимости будет больше 0,05 (или того порога, который вы задаёте вручную). Как мы увидим в других критериях, это достаточно редкая ситуация – когда нам нужно, исходя из целей исследования, чтобы значение p-уровня было больше 0,05. Чаще же нам нужна как раз обратная ситуация, но об этом далее.

2.4. Проверка равенства дисперсий (Rundom Pro 3.14).

Предположим, что у нас есть два класса школьников, на которых мы провели тест интеллекта. Полученные результаты выглядят следующим образом:

25	23
105	110
125	121
39	145
60	121
70	142
98	110
75	101
131	34
125	26
111	49
110	70
124	86
143	58
132	120
119	105
93	29
81	45
59	69
119	130

Нам нужно создать новый документ в программе Rundom Pro и вставить туда сырые баллы, но не спешите это делать, прежде чем ознакомитесь с двумя пунктами.

1. Ещё раз напоминаю, что для того, чтобы это сделать правильно, нельзя пользоваться Ctrl+V, правой кнопкой мыши Вставить или Paste. Иначе все данные окажутся в одной ячейке. Для правильного импорта данных необходимо использовать вкладку меню Edit -> Paste Data (Active Table).

2. Более того, необходимо скопировать не приведённую выше таблицу, а нижеследующую. Небольшое пояснение: есть два варианта, которым статистические программы рассматривают данные двух выборок. В первом варианте (как это в таблице выше) данные представлены «по переменным», т.е. каждый новый столбик представляет собой каждую новую выборку испытуемых. Во втором варианте программы желают видеть данные, сгруппированные «по группам», что означает расположение всех выборок в одной колонке, а в другой указывается, к какой именно выборке относятся полученные данные. Это особенность не только программы Rundom Pro, так что праведный гнев здесь не уместен.

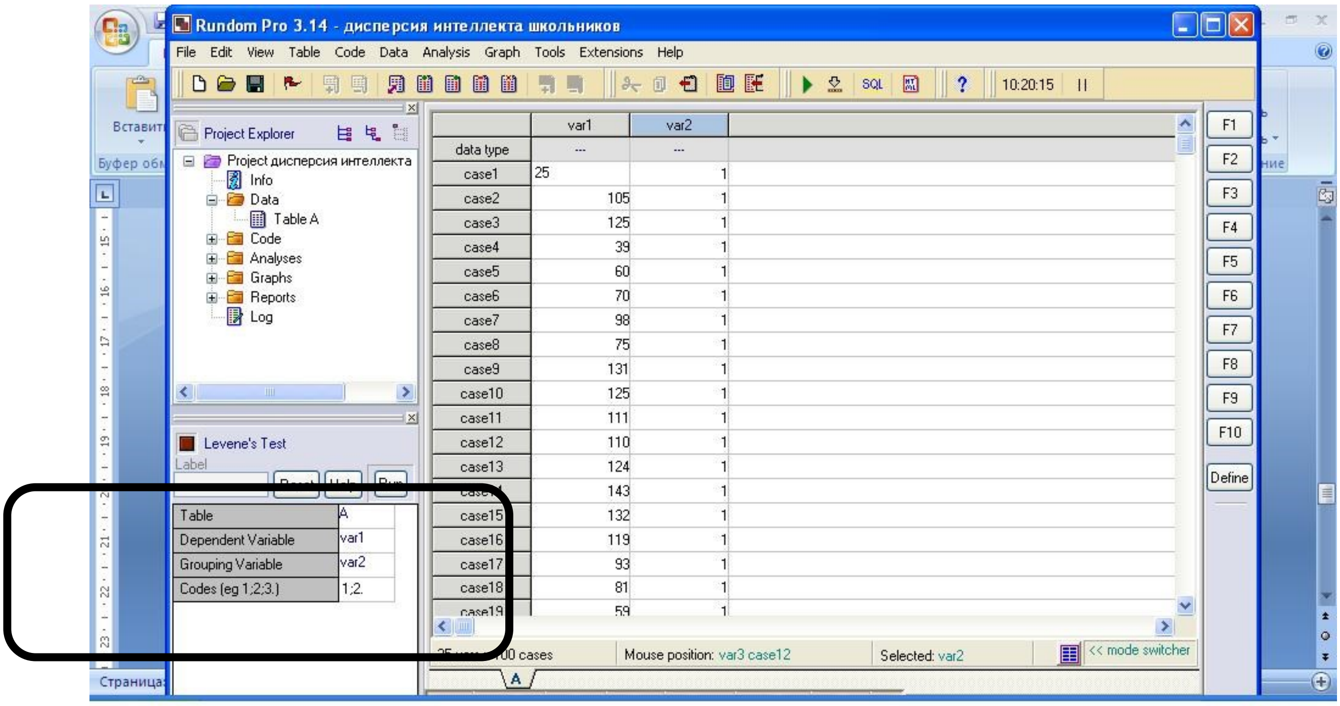
Для реализации актуальных задач нам нужно сделать таблицу таким образом, который соответствует второму варианту, а именно:

25	1
105	1
125	1
39	1
60	1
70	1
98	1
75	1
131	1
125	1
111	1
110	1
124	1
143	1
132	1
119	1
93	1
81	1
59	1
119	1
23	2
110	2
121	2
145	2
121	2
142	2
110	2
101	2
34	2
26	2
49	2

70	2
86	2
58	2
120	2
105	2
29	2
45	2
69	2
130	2

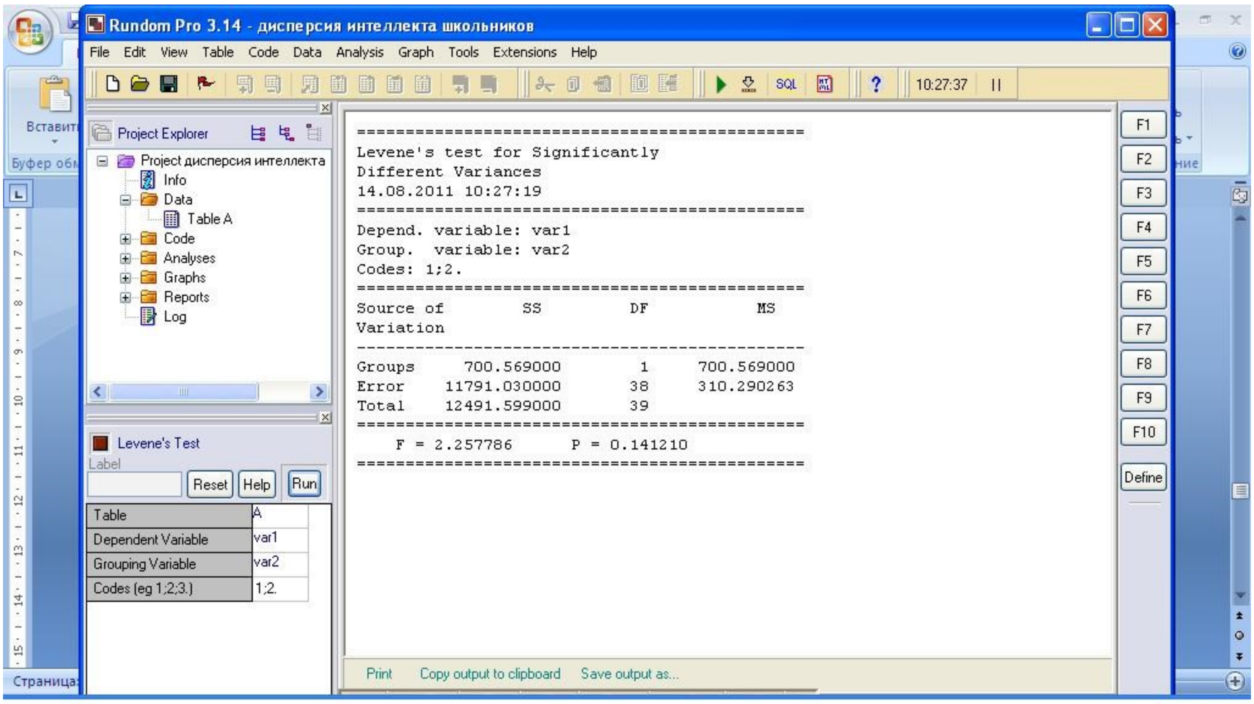
И теперь, с чувством выполненного долга, правильным образом копируем данные в программу. Нас интересуют следующие вкладки меню: Analysis -> Classical -> Other -> Levene's test.

В левом нижнем углу появляется диалоговой окно критерия Левена. Нам необходимо задать переменные и коды выборок. Зависимая переменная представлена самими сырыми баллами, т.е. в Dependent Variable мы пишем var1 (обратите внимание, что между var и 1 нет пробела – так необходимо писать, поскольку в ином случае программа выдаст ошибку), поскольку скопировали сырые баллы в колонку var1. Как и в других статистических программах, та колонка данных, которая представляет собой категории, т.е. 1 и 2 выборку в нашем случае, называется группирующей переменной, поэтому Grouping Variable – var2. Коды нужно указывать строго через точку с запятой и точку в конце перечисления. Т.е. «1;2.». В ином случае программа выдаст ошибку. В итоге диалоговое окно должно быть заполнено следующим образом:



Нужно нажать «Run» и начнётся анализ.

Будут получены следующие результаты:



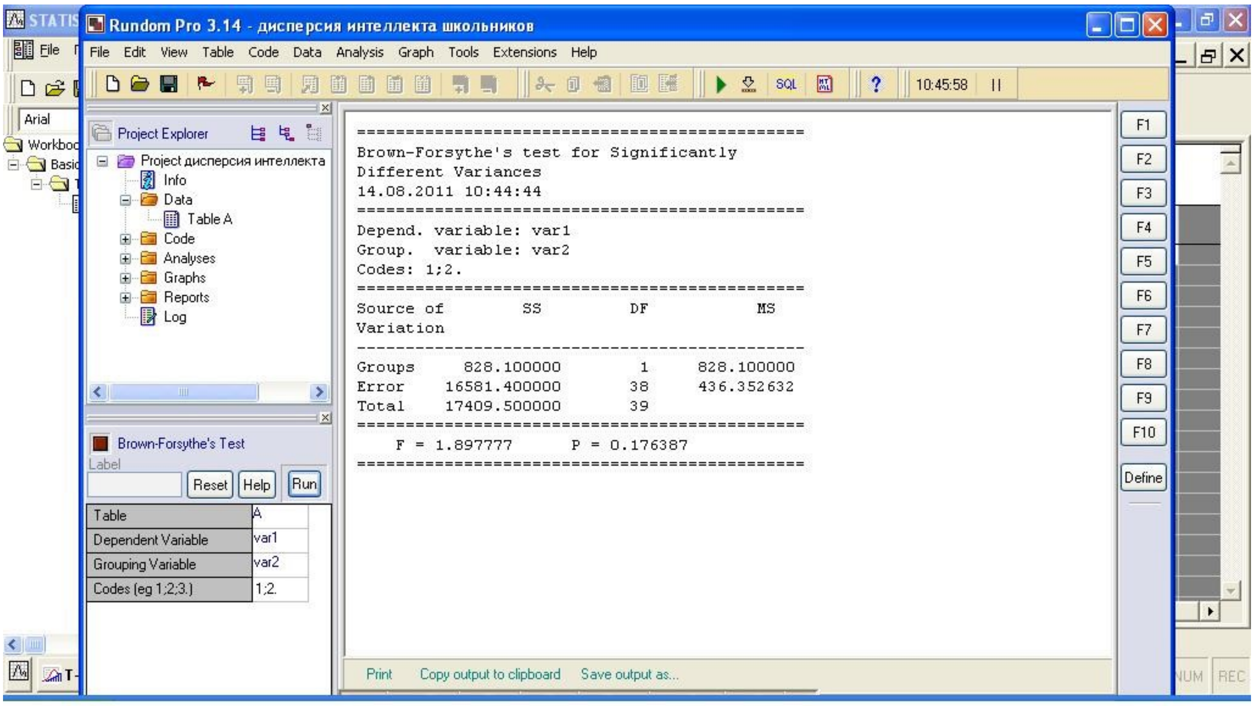
Итак, значение самого критерий F=2,257786, уровень значимости p=0,141210. Это означает, что у наших групп равная дисперсия. Если же p будет меньше 0,05, то равенство дисперсий придётся отклонить, что закрывает для исследователя возможность использовать, например, такой параметрический критерий как t-Стьюдент и придётся перейти к непараметрическим аналогам.

Но поскольку критерием Левена не ограничивается возможность сравнения дисперсий двух выборок, то рассмотрим и критерий Брауна-Форсайта. Считается, что это критерий более мощный, чем критерий Левена, хотя оба эти критерия вызывают массу недовольств статистиков.

Для этого нам можно продолжать свой анализ ничего не закрывая (хотя вернуться к проверке равенства дисперсий на следующий день вам также никто не помешает). Нужно выбрать следующие вкладки меню: Analysis -> Classical -> Other -> Brown-Forsythe’s Test.

В нижнем левом углу появится диалоговое окно критерия Брауна-Форсайта.

Заполнить условия необходимо также, как и для критерия Левена. Т.е. зависимая переменная – var 1; группирующая переменная – var 2; коды обозначения групп – «1;2.». Нажимаем «Run» и видим следующий результат:



Результаты несколько отличаются от критерия Левена, но смысловая нагрузка не изменилась. Мы всё также можем признать равенство за дисперсиями групп, что можно судить по значению p-уровня. Однако, если бы значение p-уровня было меньше 0,05, то также, как и в критерии Левена, нам пришлось бы отклонить гипотезу о равенстве дисперсий и перейти к использованию непараметрических критериев.

В отечественной традиции, которая выходит из употребления, равенство дисперсий определялось F-критерием Фишера. Мы не будем рассматривать его электронный подсчёт, поскольку для задач, описанных выше, в сфере современных компьютерных программ рекомендует использовать критерии Левена и Брауна-Форсайта. Однако, читатель может ради интереса подсчитать критерий Фишера вручную, чтобы понять общий смысл, используемый, однако в других критериях, а также для того, чтобы понять, каким образом раньше обрабатывались эмпирические данные.

Алгоритм подсчёта «вручную» выглядит следующим образом:

$$F = \frac{\sigma_1^2}{\sigma_2^2},$$

где:

σ_1^2 - большая дисперсия,

σ_2^2 - меньшая дисперсия.

Если вычисленное значение критерия F больше критического для определенного уровня значимости и соответствующих чисел степеней свободы для числителя и знаменателя, то дисперсии считаются различными.

Число степеней свободы числителя определяется по формуле:

$\upsilon_1 = n_1 - 1$,

где n_1- число вариант для большей дисперсии.

Число степеней свободы знаменателя определяется по формуле:

$\upsilon_2 = n_2 - 1$,

где n_2- число вариант для меньшей дисперсии.

Распределение F-Фишера полезно, в частотности для подсчёта критерия Левена.

Алгоритм следующий:

Проверяемая гипотеза о постоянстве дисперсии m выборок имеет вид:

$H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_m^2$.

а конкурирующая с ней гипотеза:

$H_1 : \sigma_{i_1}^2 \neq \sigma_{i_2}^2$,

где неравенство выполняется, по крайней мере, для одной пары индексов i_1, i_2.

Пусть n_i – объем i-й выборки, $N = \sum_{i=1}^{m} n_i$, X_{ij} – j-е наблюдение в i-й выборке. Статистика критерия Левена имеет вид:

$$W = \frac{N-m}{m-1} \frac{\sum_{i=1}^{m} n_i \left(\overline{Z}_{i\bullet} - \overline{Z}_{\bullet\bullet}\right)^2}{\sum_{i=1}^{m} \sum_{j=1}^{n_i} \left(Z_{ij} - \overline{Z}_{i\bullet}\right)^2},$$

где Z_{ij} может определяться одним из следующих способов:

1. $Z_{ij} = \left| X_{ij} - \overline{X}_{i\bullet} \right|$

где $\overline{X}_{i\bullet}$ – есть среднее в i-й выборке.

2. $Z_{ij} = \left| X_{ij} - \tilde{X}_{i\bullet} \right|$

где $\tilde{X}_{i\bullet}$ – есть медиана в i -й выборке.

3. $Z_{ij} = \left| X_{ij} - \overline{X}'_{i\bullet} \right|$

где $\overline{X}'_{i\bullet}$ – есть усеченное среднее в i -й выборке.

$\overline{Z}_{i\bullet}$ – есть среднее Z_{ij} по i -й выборке, $\overline{Z}_{\bullet\bullet}$ – есть среднее Z_{ij} по всем выборкам.

В критерии Левене проверяемая гипотеза о равенстве дисперсий отклоняется, если

$$W > F_{\alpha, m-1, N-m},$$

где $F_{\alpha, m-1, N-m}$ – верхнее критическое значение F -распределения с $m-1$ и $N-m$ степенями свободы и уровнем значимости α.

2.5. Различие выборок.

2.5.1. Параметрические критерии различия выборок.

2.5.1.1. t-критерий Стьюдента для независимых выборок.

Вернёмся к примеру с интеллектом двух классов.

25	23
105	110
125	121
39	145
60	121
70	142
98	110
75	101
131	34
125	26
111	49
110	70
124	86
143	58
132	120
119	105
93	29
81	45
59	69
119	130

Если нам нужно понять, можно ли считать интеллект ребят из одного класса значимо отличным от интеллекта детей из другого класса, то этот критерий нам подойдёт.

В таком случае нам необходимо проверить нормальность распределения каждого столбца данных, равенство дисперсий этих двух выборок и приступать непосредственно к t-критерию Стьюдента.

Если мы считаем этот критерий в AtteStat, то нам понадобится таблица, в которой сырые данные представлены без группирующих переменных.

Для подсчёта нужно выбрать вкладку меню: Параметрическая статистика. Задать диапазоны выборок и вывода. Поставить галочку на критерии Стьюдента для независимых выборок.

Получается следующий результат:

Параметрическая статистика			
Результаты обычно включают:			
Статистика, Р-значение одностороннее, Р-значение двустороннее			
Критерий Стьюдента для независимых выборок			
1,060423	0,147823	0,295646	

Эмпирическое значение критерия t=1,060423…; p = 0,295646… (Обычно нас интересует именно двустороннее значение уровня значимости).

Если вести подсчёт в программе Rundom Pro, то таблицу данных необходимо представлять в виде зависимых и группирующих переменных. Чтобы посчитать критерий, необходимо пройти следующим путём: Analysis -> Classical -> Two-Sample t Test.

Получается следующий результат:

```
============================
Two-Sample t Test
(equal variances assumed)
14.08.2011 15:45:43
============================
Depend. variable: var1
Group.  variable: var2
Code, 1st sample: 1
Code, 2nd sample: 2
============================
Mean1              97.200000
```

```
Mean2             84.700000
Diff.             12.500000
Ho value           0.000000
----------------------------
t                  1.060423
DF                       38
P                  0.295646
============================
```

Если нужно задать условие, что равенство дисперсий не подразумевается (хотя с чего бы нам это делать, если мы это условие уже проверили критериями Левена и Брауна-Форсайта), то необходимо в диалоговом меню критерия помимо зависимой и группирующей переменной, а также кодов, задать ещё и это условие (Equal Variance Assumed – No). Тогда при анализе эмпирическое значение критерия не изменится, но изменится уровень значимости:

```
============================
Two-Sample t Test
(Welch's approach)
14.08.2011 15:58:34
============================
Depend. variable: var1
Group.  variable: var2
Code, 1st sample: 1
Code, 2nd sample: 2
============================
Mean1             97.200000
Mean2             84.700000
Diff.             12.500000
Ho value           0.000000
----------------------------
t                  1.060423
DF                36.493793
P                  0.295920
============================
```

Как видно, уровень значимости при таком условии становится более благосклонном к гипотезе, что различий нет.

Итак, если p-уровень <0,05, то мы принимаем альтернативную гипотезу о том, что различия есть. Если >0,05, то, увы, но говорить о различиях выборок повода нет.

Что же такое пункт меню «two-sample t Test (raw data not known)»?

Это тот же самый t-критерий Стьюдента для независимых групп, но использующий в качестве основы анализа не сырые данные, а такие признаки сырых баллов, как: число испытуемых (N), среднее по выборке (Sample Mean), стандартное отклонение по выборке (Sample Std. Dev.). Кому нужен такой анализ? Предположим мифическую ситуацию, когда вы делаете подробнейший анализ литературы, в которой на ваши глаза попались интереснейшие результаты двух исследователей. Предположим, что они как раз и изучали интеллект в двух школах разных районов города Новосибирска. В принципе, если бы вы были знакомы с этими исследователями, то попросили бы у них сырые баллы и посчитали t-критерий, но если у вас в распоряжении только данные о стандартных отклонениях и средних выборок (или хотя бы дисперсиях – ведь получить из них стандартное отклонение можно простым извлечением корня), то вы также сможете провести достойный анализ. На мой взгляд, такой модуль очень полезен и побуждает писать в академических статьях необходимую информацию, ведь приводить сырые баллы зачастую не представляется возможным.

Никогда не стоит принимать слова другого человека на веру, если есть возможность их проверить. В данном случае вам понадобится всего лишь сделать новый документ в Rundom Pro (можно было бы и в старом, но где гарантия, что критерий будет посчитан не по сырым баллам, которые уже введены?). Выбрать: Analysis -> Classical -> Two Sample t Test (raw data not known) и ввести необходимые описательные статистики.

А вот чтобы эти описательные статистики получить, придётся приложить усилия. С тем, как подсчитать описательные статистики, можно ознакомиться в соответствующей главе. Из собственноручно выполненных подсчётов (если подсчёт программой можно так назвать) должен был обнаружиться уточняющий факт – когда мы имеем дело с подсчётом t-критерия Стьюдента по описательным статистикам, а не исходным данным, имеется в виду тот уровень значимости, который не предполагает равенства дисперсий.

И последнее замечание о t-критерии для независимых выборок в разных программных продуктах. Подсчёты эмпирического значения совпали в обеих программах лишь потому, что среднее по первой выборке было больше среднего по второй выборке. Если бы мы, к примеру, поменяли их местами, обозначив в программе Rundom Pro за код первой выборки категорию 2, а за код второй выборки категорию 1, то в Rundom Pro значение t-критерия было бы точно таким же, как и было, но с обратным знаком, т.е. (- 1.060423). Уровень значимости, однако, не изменился бы. Если же мы проделаем подобную процедуру в программе AtteStat (т.е. зададим за диапазон первой выборки вторую колонку, а за диапазон второй выборки первую колонку), то значение t-критерия как было положительным, так и останется. Психологический разум, глядя на такую «какофонию», впадает в беспамятство и теряет связь с реальностью, однако секрет кроется в числителе формулы для подсчёта t-критерия. В программе AtteStat используется модуль, что означает всегда положительные значения, в то время как в Rundom Pro, SPSS и STATISTICA, например, модуль не используется, что и позволяет эмпирическому значению принимать отрицательные значения. Симметричность распределения t-Стьюдента даёт шанс обеим формулам. С другой стороны, те простые шаги, как изменить порядок выборок, приводит к изменению знака, что, хоть и безосновательно, но всё же располагает меня в сторону формулы с модулем в числителе. Что же касается нашего метода менять выборки местами, обозначая 1 за 2, 2 за 1, то лучше в реальном исследовании так не делать, поскольку это лишь засоряет и без того загруженное сознание исследователя. Нужно просто знать, что значение может быть как положительным, так и отрицательным в зависимости от того, какая группа (большая или меньшая) обозначена за первую. В случае же, когда мы говорим о направленной альтернативной гипотезе, знак может пригодиться. О том, как именно он используется, можно прочитать в главе «Как пользоваться статистическим критерием».

2.5.1.2. t-критерий Стьюдента для зависимых выборок.

Предположим, что те данные, которые мы столь часто рассматриваем, а именно – интеллект школьников, представляет сырые баллы, которые были собраны по другому принципу. Первая колонка – это интеллект в начале года, а вторая колонка – этот же тест в конце года. Т.е. это план «тест-ретест», проведённый на одном-единственном классе. Если нас интересует различие этих двух показателей, то нам нужен какой-нибудь критерий, который заканчивается словами «для зависимых (или парных) выборок». А поскольку мы проверили наши данные и обнаружили, что они подчиняются нормальному закону распределения, то можем позволить себе использовать t-критерий.

Итак, для программы AtteStat нам по-прежнему необходима вышеописанная таблица, в которой каждая колонка отвечает за свою выборку, а для Rundom Pro таблица, в которой первая колонка содержит сырые баллы обеих выборок, а вторая – группирующая переменная.

Для AtteStat: Параметрическая статистика -> Критерий Стьюдента (парные).

Получаются следующие результаты:

Параметрическая статистика			
Результаты обычно включают:			
Статистика, Р-значение одностороннее, Р-значение двустороннее			
Критерий Стьюдента для связанных выборок			
1,019218	0,160451	0,320903	

Как мы видим, t = 1,019218, p = 0,320903 (поскольку мы выдвигаем ненаправленную гипотезу).

Для Rundom Pro: Analysis -> Classical -> Paired Comparison t Test.

Задаём зависимой переменной var1, группирующей var2, коды 1 и 2. Нажимаем «Run».

Полученные результаты выглядят следующим образом:

```
========================
Paired Comparison t Test
14.08.2011 17:48:13
========================
Depend. variable: var1
Group.  variable: var2
Code, 1st sample: 1
Code, 2nd sample: 2
========================
Mean diff.     12.500000
------------------------
t               1.019218
DF                    19
P               0.320903
========================
```

Как мы видим, результаты подсчёта полностью согласуются. Если поменять группы местами, то мы опять получим отрицательное значение в Rundom Pro, SPSS и STATISTICA. Но бояться этого не следует, потому что в конечном счёте нас интересует уровень значимости.

Итак, мы выяснили, что различий между первым и вторым полугодием по признаку интеллекта нет. Если бы p-уровень был меньше 0,05, то мы бы посчитали, что различия в выборках статистически значимы.

2.5.1.3. t-критерий Стьюдента для сравнения со средним значением.

Предположим, что у нас есть (вы, конечно, удивитесь) группа школьников, у которых мы решили измерить интеллект. Наша исследовательская задача относительно проста: ходит слух, что средний интеллект по школе равен 50, и мы решили сравнить интеллектуальный потенциал нашего класса с общим потенциалом школы. Естественно, людям, которые помнят нормы интеллекта, мой пример покажется странным, но помимо чайной ложечки юмора стоит обязательно посчитать приведённые ниже данные, потому что нам нужно обсудить любопытную особенность выдачи данных.

Итак:

25	50
105	50
125	50
39	50
60	50
70	50
98	50
75	50
131	50
125	50
111	50
110	50
124	50
143	50
132	50
119	50
93	50
81	50
59	50
119	50

Для AtteStat данные нужно представлять именно таким образом и никак иначе. Первая колонка данных представляет результаты тестирования интеллекта в классе, а вечно повторяющееся «50» есть ничто иное как средний интеллект по школе. Нас интересует вкладка меню «Параметрическая статистика». Диапазоном выборки 1 мы задаём все 20 значений первой колонки, а диапазоном выборки 2 задаём все 20 пятидесяток. Интервал вывода оставляю на ваше усмотрение. Затем мы галочкой отмечаем «Критерий Стьюдента» - сравнение со средним, который находится в правой части меню.

И вполне может получиться вот такой результат:

Параметрическая статистика

Результаты обычно включают:						
Статистика, Р-значение одностороннее, Р-значение двустороннее						
Критерий Стьюдента						
6,343659	2,18E-06	4,36E-06				

Достаточно приемлемым на интуитивном уровне кажется эмпирическое значение t = 6,343659, но что же это за уровень значимости такой? Даже не представляю, каковы будут глаза читателя какой-нибудь психологической статьи, если он увидит вот такой уровень значимости в результатах исследования. Всё дело в том, что это очень маленький уровень значимости (исходя из наших задач нам это и нужно). Но чтобы никого не напугать, нужно нажать на интересующую нас ячейку, т.е. на ячейку двухстороннего уровня значимости и изменить её формат. Делается это в 2007 Office, например, через правую кнопку мыши – формат ячеек. Нам нужно получить числовой формат этой ячейки. Осуществив такую подмену мы увидим скорее всего только нули. И это правда, но как обычно – полуправда. На самом деле нам достаточно увеличить число десятичных знаков (что делается в том же меню, где и ставится числовой формат – если название формата слева, то справа скорее всего можно выбрать количество знаков), чтобы увидеть, что уровень значимости в данном случае равен примерно 0,0000043620…

Естественно, что такой уровень значимости может быть и не нужен. Можно просто в случае, когда программа AtteStat выдаёт странный уровень значимости с обозначением «Е», указывать p = 0. Это достаточно корректно и зачастую труднодостижимо в реальном исследовании.

В Rundom Pro также есть возможность посчитать одновыборочный t-критерий Стьюдента. Для этого достаточно и одной колонки, содержащей сырые баллы теста интеллекта в классе. Меню: Analysis -> Classical -> One Sample t Test. В диалоговом меню нужно будет просто указать номер переменной и среднее значение, с которым нужно сравнить эту переменную (Reference Mean – в нашем случае 50).

Получаются следующие результаты:

```
=========================
One-Sample t Test
14.08.2011 19:27:16
=========================
Variable: var1
=========================
Mean              97.200000
Ref. Mean         50.000000
Diff.             47.200000
-------------------------
```

```
t                  6.343659
DF                       19
P                  0.000004
===========================
```

Значимое различие налицо, как говорится.

Если вы внимательно читали моё пояснение о возможных различиях в знаке эмпирического значения t в случае независимых и зависимых выборок, то сможете с огромной лёгкостью придумать такое значение среднего интеллекта по школе, чтобы Rundom Pro вывела отрицательное значение, а AtteStat – положительное. А говорю это я к чему? К тому, что результаты, получаемые данным видом t-критерия абсолютно такие же, как если бы мы считали парный t-критерий, т.е., например, сравнивали два полугодия, во втором из которых интеллект у всех испытуемых вдруг стал равен 50. Опять же, резонно, не доверяя мне априорно, посчитать самим.

Если же в нашей выборке очень малое количество испытуемых или же не выполняются требования t-Стьюдента, то можно воспользоваться биномиальным критерием. Этот критерий также предполагает, что исследование было проведено всего на одной выборке, чей результат сравнивается с заданной теоретической частотой.

Гипотезы

H_0: Частота встречаемости данного эффекта в обследованной выборке не превышает теоретической (заданной, ожидаемой, предполагаемой).

H_1: Частота встречаемости данного эффекта в обследованной выборке превышает теоретическую (заданную, ожидаемую, предполагаемую).

Биномиальный критерий имеет следующие ограничения[19]:

1. В выборке должно быть не менее 5 наблюдений. В принципе возможно применение критерия и при $2 \leq n < 5$, но лишь в отношении определенного типа задач, которые приведены в таблице ниже.

2. Верхний предел численности выборки зависит от ограничений, определяемых пп.3-8 и варьирует в диапазоне от 50 до 300 наблюдений, что определяется имеющимися таблицами критических значений.

3. Биномиальный критерий m позволяет проверить лишь гипотезу о том, что частота встречаемости интересующего нас эффекта в обследованной

[19] Сидоренко Е.В. «Математические методы обработки в психологии»

выборке *превышает* заданную вероятность Р. Заданная вероятность при этом должна быть: Р ≤0,50.

4. Если мы хотим проверить гипотезу о том, что частота встречаемости интересующего нас эффекта достоверно *ниже* заданной вероятности, то при Р=0,50 мы можем сделать это с помощью уже известного критерия знаков G, при Р>0,50 мы должны преобразовать гипотезы в противоположные, а при Р<0,50 придется использовать критерий χ^2.

Выбор критерия для сопоставлений эмпирической частоты с теоретической при разных вероятностях исследуемого эффекта Р и разных гипотезах.

Заданные вероятности	H_1: $f_{эмп}$ достоверно выше $f_{теор}$	H_1: $f_{эмп}$ достоверно ниже $f_{теор}$
P<0,50	для 2 ≤n ≤50	для n ≥30
P=0,50	для 5 ≤n ≤300	для 5 ≤n ≤300
P>0,50	для n ≤30	для 2 ≤n ≤50

Рассмотрим один из примеров, приведённых в учебном пособии «Методы математической обработки в психологии» Сидоренко Е.В.:

«В тренинге профессиональных наблюдателей допускается, чтобы наблюдатель ошибался в оценке возраста ребенка не более чем на 1 год в ту или иную сторону. Наблюдатель допускается к работе, если он совершает не более 15% ошибок, превышающих отклонение на 1 год. Наблюдатель Н допустил 1 ошибку в 50-ти попытках, а наблюдатель К - 15 ошибок в 50-ти попытках. Достоверно ли отличаются эти результаты от контрольной величины?

Определим частоту допустимых ошибок при n = 50:

$f_{теор}=n\cdot P=50\cdot 0,15=7,5$

Для наблюдателя Н $f_{эмп}<f_{теор}$. Для наблюдателя К $f_{эмп}>f_{теор}$

Сформулируем гипотезы для наблюдателя Н.

H_0: Количество ошибок у наблюдателя Н не меньше, чем это предусмотрено заданной величиной.

H_1: Количество ошибок у наблюдателя Н меньше, чем это предусмотрено заданной величиной.

В данном случае Р=0,15<0,50; $f_{эмп}>f_{теор}$.

Этот случай попадает под вариант Б Табл. 5. 12. Нам придется применить критерий *у},* сопоставляя полученные эмпирические частоты ошибочных и правильных ответов с теоретическими частотами, составляющими, соответственно, 7,5 для ошибочного ответа и (50-7,5)=42,5 для правильного ответа. Подсчитаем χ^2 по формуле, включающей поправку на непрерывность[20]:

$$\chi^2 = \sum \frac{\left(\left|f_j^э - f_j^т\right| - 0,5\right)^2}{f_j^т} = \frac{\left(\left|1-7,5\right|-0,5\right)^2}{7,5} + \frac{\left(\left|49-42,5\right|-0,5\right)^2}{42,5} =$$
$$= 4,800 + 0,85 = 5,65$$

Определяем критические значения χ^2 при ν=l:

$$\chi^2_{кр.} = \begin{cases} 3,841\ (\rho \le 0,05) \\ 6,635\ (\rho \le 0,01) \end{cases}$$
$$\chi^2_{эмп} > \chi^2_{кр.}\ (\rho \le 0,05)$$

Ответ: H_0 отвергается. Количество ошибок у наблюдателя Н меньше, чем это предусмотрено заданной величиной (р≤0,05)

Сформулируем гипотезы для наблюдателя К.

H_0: Количество ошибок у наблюдателя К не больше, чем это предусмотрено заданной величиной.

H_1: Количество ошибок у наблюдателя К больше, чем это предусмотрено заданной величиной.

В данном случае Р=0,15<0,5; $f_{эмп}>f_{теор}$. Этот случай подпадает под вариант А Табл. 5.12. Мы можем применить биномиальный критерий, поскольку n=50.

Определяем критические значения при п=50, Р=0,15, Q=0,85:

$$m_{кр} = \begin{cases} 13\ (p \le 0,05) \\ 16\ (p \le 0,01) \end{cases}$$
$$m_{эмп} = f_{эмп} = 15$$
$$m_{эмп} > m_{кр}\ (\rho \le 0,05)$$

[20] Поправка на непрерывность вносится во всех случаях, когда признак принимает всего два значения и число степеней свободы поэтому равно 1 (см. параграф 4.2)

Ответ: H_0 отвергается. Количество ошибок у наблюдателя Н меньше, чем это предусмотрено заданной величиной ($p<0,05$)».

2.5.2. Непараметрические критерии различия выборок.

2.5.2.1. U-критерий Манна-Уитни (для независимых выборок).

В большинстве случаев психологические данные не подчиняются нормальному закону распределения. Сложно сказать, что именно является причиной такой картины – сама ли природа данных или же небольшие объёмы выборок, но факт остаётся фактом. Непараметрические критерии, которые не накладывают предположений о параметрах распределения данных, бывают весьма полезны.

Критерий Манна-Уитни – это непараметрический критерий, который позволяет ответить на вопрос, есть ли различие между двумя независимыми выборками по заданному признаку, измеренному, по крайней мере, в порядковой шкале. Совершенно очевидно, что критерий Манна-Уитни является непараметрическим аналогом (по исследовательской задаче) критерию t-Стьюдента для независимых выборок. Соответственно этот критерий не требует нормальности данных.

Безусловно, как и для многих других критериев, в случае Манна-Уитни допустимое количество человек 7, но не стоит предполагать, что проведя исследование на такой маленькой выборке, имеет смысл делать какие-либо уверенные выводы.

H_0: Уровень признака в группе 2 не ниже уровня признака в группе 1.

H_1: Уровень признака в группе 2 ниже уровня признака в группе 1.

Поскольку это ранговый критерий, т.е. такой критерий, который переводит исходные сырые баллы в упорядоченный ряд, состоящий из рангов, то повторяющиеся значения должны быть учтены в формуле подсчёта. Также стоит понимать, что качество подсчёта падает, если повторяющихся значений слишком много. Т.е., например, пользоваться критерием Манна-Уитни, когда значения шкалы могут принимать только два значения - 1 и 2, не имеет смысла.

Предположим, что мы располагаем данными по музыкальной аддикции подростков 10 «Б» и 10 «В» класса (результаты опросника L, проведённого на двух независимых группах). Исходя из той ситуации, которая складывается на данный момент, а именно – усиленному разрастанию списка аддикций в различных исследованиях, мы наверняка скоро встретим такое исследование.

Сырой ряд данных будет выглядеть следующим образом:

10"Б"	10"В"
1	12
2	13
3	14
4	15
5	16
6	17
7	18
8	19
9	20
10	21

Посчитаем критерий Манна-Уитни в программе AtteStat. Для этого нам нужно открыть модуль «Непараметрическая статистика» задать в качестве первой выборки данные 10 «Б», например, а в качестве второй выборки значения 10 «В», выбрать ячейку для вывода результатов подсчёта, а также поставить галочку на подсчёте критерия Манна-Уитни. В данном случае нам не требуются группирующие переменные, поскольку программа позволяет ввести данные в сыром виде. Более того, вводить данные с группирующей переменной (классификатором) не рекомендуется, потому что даже если поставить опцию представления данных в качестве выборки-классификатора, получатся ошибки, поскольку эта опция предназначена для других методов. Поэтому мы вводим просто две колонки данных и видим следующие результаты подсчёта критерия Манна-Уитни:

Непараметрическая статистика и анализ качественных данных	
Учет связок	
Учет поправки на непрерывность	
Критерий Манна-Уитни	
Статистика, Р-значение (двустороннее)	
100	0,000157052

Если мы сравним результаты подсчёта критерия Манна-Уитни в программе AtteStat с программой SPSS, например, то очень удивимся огромной разнице эмпирического значения. Для этого же массива данных в SPSS эмпирическое значение U = 0. Как объяснить такую странность?

Нам ничего не остаётся, как посчитать данные вручную, чтобы понять, отчего это зависит. Правда, можно было бы конечно воспользоваться и справкой программы AtteStat, но тогда мы упустим возможность понять чёткий алгоритм подсчёта этого критерия.

Чтобы посчитать критерий Манна-Уитни используется следующая процедура:

1. Ряды данных обеих групп объединяются в единую упорядоченную по возрастанию значения строку, но таким образом, чтобы всегда помнить, из какой группы было данное значение. При ручном подсчёте используются цветные карандаши или рядом с каждым значением рисуется символ, отвечающий за его группу. Мы сделаем проще – мы отметит ячейки цветом в Excel, т.е. наши данные приобрели следующий вид:

1
2
3
4
5
6
7
8
9
10
12
13
14
15
16
17
18
19
20
21

Обратите внимание на тот факт, что может встречаться ситуация, когда некоторые значения из первой группы будут меньше значений из другой группы и наоборот, что приведёт к «перемешиванию» цветов – этого не стоит бояться.

2. Далее упорядоченному ряду присваиваются ранги. Выглядит это следующим образом:

1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
12	11
13	12
14	13
15	14
16	15
17	16
18	17
19	18
20	19
21	20

В нашем случае нет повторяющихся рангов, в связи с чем ранг, по сути, совпадает с порядковым номером значения. Если бы таковы имелись, то нужно было бы подсчитать среднее арифметическое повторяющихся рангов и присвоить соответствующим значениям усреднённый ранг. Например, если бы первый и второй испытуемый первой группы набрали бы не «1» и «2» сырых балла соответственно, а по «2», то наша ранговая таблица начиналась бы не с рангов «1» и «2», а с «1,5» и «1,5». Хотя третий испытуемый, который набрал «3» сырых балла, получил бы ранг «3».

3. Далее подсчитывается сумма РАНГОВ для каждой группы отдельно. В нашем случае она такова:

$R_1 = 1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 + 10 = 55$

$R_2 = 11 + 12 + 13 + 14 + 15 + 16 + 17 + 18 + 19 + 20 = 155$

4. На следующем шаге подсчитывается эмпирическое значение величины U.

$U_k = n_1 * n_2 + n_k * (n_k + 1)/2 – R_k$,

где: n – количество человек в группе, k – номер группы, у которой была наибольшая сумма рангов, R – сумма рангов.

Соответственно, для подсчёта эмпирического значения нам нужно посчитать U второй группы, поскольку именно для неё сумма рангов была больше, но мы посчитаем обе величины U, поскольку наша цель в том числе – понять, почему эмпирическое значение в AtteStat и SPSS различно.

U_1 = 10*10 + 10 {имеется в виду количество человек первой группы}* (10+1)/2 – 55 = 100

U_2 = 10*10 + 10 {имеется в виду количество человек во второй группе}* (10+1)/2 – 155 = 0

5. Далее нам нужно сравнить полученное эмпирические значение критерия с таблицей значимости.

Как мы помним, нам нужно U_2. Таблица значимости этого критерия устроена довольно просто – критические значения находятся на пересечении количества человек первой и второй группы. В нашем случае было 10 человек в той и другой группе, в связи с чем мы легко можем увидеть, что $U_{крит}$ = 23. В нашем случае $U_{эмп} = 0 \Rightarrow U_{эмп} < U_{крит}$, что означает для критерия Манна-Уитни отвержение нулевой гипотезы и принятие альтернативной о различии групп. Соответственно, если бы $U_{змп}$ было больше $U_{крит}$, то была бы принята нулевая гипотеза.

Критические значения критерия U Манна-Уитни[21]

(для проверки ненаправленных альтернатив)

P=0,05

N_1

N_2

0 1 2 3 4 5 6 7 8 9 0

[21] Таблица значимости взята с сайта: http://statexpert.org/articles/35/#5

									0	1	1	2	3	3
						1	2	3	4	5	7	8	9	0
			0	1	3	4	6	7	9	1	2	4	5	7
		0	2	4	6	8	0	2	4	6	8	0	2	4
	0	3	5	7	9	2	4	6	9	1	4	6	8	1
	2	5	7	0	3	6	8	1	4	7	9	2	5	8
0	4	7	0	3	6	9	3	6	9	2	5	8	2	5
1	6	9	3	6	0	3	7	0	4	7	1	5	8	2
2	8	2	6	9	3	7	1	5	9	3	7	1	5	9
3	0	4	8	3	7	1	5	0	4	9	3	7	2	6
4	2	6	1	6	0	5	0	5	9	4	7	4	8	3
5	4	9	4	9	4	9	4	9	4	0	5	0	5	0
6	6	1	7	2	7	3	9	4	0	5	1	6	2	8
7	8	4	9	5	1	7	3	7	5	1	7	3	9	05
8	0	6	2	8	5	1	7	4	0	6	3	9	06	12

9	2	8	5	2	8	5	2	8	5	2	9	06	13	19
0	4	1	8	5	2	9	6	3	0	8	05	12	19	27

Во-первых, мы поняли, что «0» и «100» - это $U_{эмп}$, подсчитанное для той или другой группы. Соответственно, в SPSS было представлено U_2, а в AtteStat – U_1. Этого не стоит бояться, поскольку уровень значимости, что и важно для нас, от этого не меняется, в связи с чем мы можем совершенно не беспокоиться о таких различиях статистических программ. Соответственно, в нашем случае $p<0,05$, а значит музыкальная аддикция этих двух групп статистически значимо различна.

Но достаточен ли вывод о том, что две группы данных различны? Зачастую такой ответ вовсе не устраивает исследователя, поскольку ему очень интересно знать, о каком направлении различий идёт речь.

И в этом пункте начинается наше новое путешествие к ответу. Дело в том, что есть, как минимум, четыре способа, как можно посмотреть, в какой группе значение признака больше.

1. В таких программах как SPSS и STATISTICA при подсчёте критерия Манна-Уитни приводится сумма рангов каждой из групп. Да, та самая сумма рангов R_1, которая в нашем случае была равна 55, и R_2, которая в нашем случае была равна 155.

Соответственно, пользователи STATISTICA любят смотреть на сумму рангов, чтобы понять, в какой группе значение признака больше. Так, мы бы решили, что во второй группе значение музыкальной аддикции выше, чем в первой.

Но этот способ имеет существенный минус – при разном количестве испытуемых в первой и второй группе сумма рангов не будет отражает реального направления – сумма рангов скорее всего окажется искусственно завышенной в той группе, где больше испытуемых. Это очень серьёзное искажение, поэтому методом определения направления различий по сумме рангов лучше не пользоваться, а точнее – вообще нельзя пользоваться при разном количестве испытуемых в первой и второй группе.

2. Второй способ, который я встречала, это определение среднего арифметического по каждой группе. А почему бы в самом деле и нет? Ведь

среднее арифметическое – это такая величина, которая очень чутко реагирует на количество человек в группе.

Я бы и против такого момента возразила по двум причинам.

Во-первых, в данном случае мы имеем дело с ранговым критерием для порядковых шкалах. Что это значит? С одной стороны, этот критерий, совершенно оправдывая своё название, имеет дело исключительно с рангами величин, а не с самими значениями, что делает подсчёт среднего логически некорректным. Мы хорошо это увидели при «ручном» подсчёте – присвоив на самом первом шаге ранги, мы ни разу не вернулись к этим данным – все операции касались уже именно рангов. С другой, для порядковых шкал считать среднее не этично. В этом нет такого уж строго математического смысла, какой мы привыкли навязывать среднему арифметическому.

Но первый аргумент касается исключительно корректности математической операции, что, естественно далеко не всегда соблюдается в прикладных задачах. В таком случае, чтобы окончательно отвергнуть метод определения направлений различий нам необходимо понять, может ли быть такая ситуация, когда некая ранговая величина будет расходиться со средним арифметическим? Это и будет нашим следующим аргументом.

Во-вторых, что, пожалуй, окончательно низвергает этот метод с учётом предыдущего аргумента, среднее арифметическое чувствительно к выбросам, а, значит, есть опасность, что операции на ранжированном ряду будут выглядеть совершенно иначе, нежели среднее арифметическое на сыром ряду данных. Раз нам нужно найти хорошую ранговую альтернативу среднему арифметическому, но использовать сумму рангов при этом нельзя по вышеописанным причинам, то наиболее оптимальной альтернативой будет средний ранг.

Средний ранг – это среднее арифметическое рангов.

И есть случай, когда среднее арифметическое и средний ранг будут расходиться. Давайте его рассмотрим.

Предположим, что у нас есть следующие данные:

1	12
2	13
3	10
4	14
5	15

	50	16
	49	17
	47	18
	6	19
	7	20
	8	21
	9	22
	43	23
	11	24
	45	26
среднее	19,33333	18
ср ранг	13,06667	17,93333

Как мы видим из этой таблицы, среднее арифметическое больше в первой группе, а средний ранг во второй. Так происходит из-за того, что среднее арифметическое очень чувствительно к выбросам, в то время как ранги от этой беды практически избавлены. Если в исследовании не идёт работа с выбросами или выбросы приняты за истинные значения, то среднее арифметическое даст попросту неверное представление о различиях.

И в SPSS по этой, насколько я понимаю, причине в подсчёте Манна-Уитни даётся информация о среднем ранге каждой группы.

3. Средний ранг – это третий и наиболее оптимальный из вышеперечисленных способ определения направления различий.

4. Однако стоит понимать, какую задачу мы решаем. Для определения различий при подсчёте критерия Манна-Уитни стоит использовать средний ранг. Если же исследователь далее пускается в подробнейшие размышления о различиях и их природе, то средний ранг уже не является панацеей – стоит привлекать описательную статистику, ведь средний ранг – это разновидность мер центральной тенденции, в то время как есть меры рассеяния. Поэтому качественные размышления о различиях стоит строить не только на критерии Манна-Уитни и среднем ранге, но и при помощи, например, ящиков с усами и других графических средств и описательных статистик.

2.5.2.2. Критерий Т-Вилкоксона для зависимых выборок.

Критерий Т-Вилкоксона используется для проверки различий двух зависимых выборок. Нормальности распределения, как и все непараметрические критерии, не требует.

H_0: Интенсивность сдвигов в типичном направлении не превосходит интенсивности сдвигов в нетипичном направлении.

H_1: Интенсивность сдвигов в типичном направлении превышает интенсивность сдвигов в нетипичном направлении.

Этот критерий является непараметрическим аналогом t-критерия Стьюдента для зависимых выборок и зачастую показывает достаточно большую мощность. Более того, если верить справочной системе программы STATISTICA, то при выполнении требований к t-Стьюденту, т.е. нормальности данных, эти два критерия покажут схожую мощность.

Но в России, на мой взгляд, с этим критерием наблюдается некоторая напряжённость. Всё дело в том, что критерий Вилкоксона применяется в схожих исследовательских задачах, что и критерий знаков. И в попытке понять, в чём же отличие этих двух критериев, помимо того что критерий Вилкоксона – «хороший», а критерий знаков – «плохой», ищущий истины может встретить массу трудностей, умалчиваний и неправильных трактовок.

Как то ни странно, англоязычные источники дают статистическую информацию «более русским языком», чем русские учебники, ориентированные, видимо, зачастую на профессоров и академиков, которые поймут автора с полуслова.

Критерий знаков лишь потому встречается в подсчётах до нынешнего времени, что:

1) некоторые исследования просто нерационально проводить с тем уровнем точности, который необходим для обработки критерием Вилкоксона – гораздо дешевле делать некоторые измерения без точности, просто указывая направление изменений;

2) у критерия знаков нет таких ограничений, которые есть у критерия Вилкоксона.

Поэтому своё рассмотрение этого критерия мы начнём с ограничений, которые должны быть соблюдены для корректного использования.

Предположим, что $Z_i = X_i - Y_i$, где $i = 1, \ldots, n$.

В таком случае условия использования критерий Вилкоксона будут следующими[22]:

1. Z_i должны быть независимы (не путать с независимость X_i и Y_i, которые должны быть как раз зависимы).

2. Каждое Z_i непрерывная величина, симметричная относительно общей медианы θ.

3. Шкалы измерения X_i и Y_i как минимум порядковые, причём такие, что «больше чем», «меньше чем» и «равно» имеют смысл.

С независимостью величин Z_i проблем возникнуть не должно, хотя, строго говоря, большинство исследований, которые сделаны НЕ ПРАВИЛЬНО, этой проблемой страдают «молча», а именно – испытуемые – это друзья экспериментатора, что, конечно, делает их зависимыми между собой.

А вот, например, третье требование – шкалы измерения такие, что градация величины имеет свой строгий смысл, делает критерий Вилкоксона применимым далеко не всегда. Как сказано в справочной системе, такую порядковую шкалу имеет смысл считать ранжированной метрической шкалой, что, безусловно, имеет свой смысл.

Например, в исследованиях, построенных по ретестовому плану с большим промежутком такое строгое соответствие выраженности оценок испытуемого имеет тем меньший смысл, чем более изменчивым и неопределённым является оцениваемое им качество/признак, а также тем меньше смысла в применении обсуждаемого критерии при большом количестве градаций самой шкалы. Собственно, почему я вывожу именно такие признаки исследований в качестве ограничения на применение критерия Вилкоксона? Представим себе, что мы создали шкалу оценки удовлетворённости жизнью, в которой предлагается оценить это качество своей жизни в пределах от 0 до 100. Мы провели её на некотором количестве испытуемых, а затем, месяца через три решили провести эту анкету ещё раз. Почему критерий Вилкоксона в данном случае лучше не использовать? Во-первых, мы никогда не можем точно сформулировать свою оценку удовлетворённости жизнью, если, конечно, не переживаем какое-то критическое состояние, когда нам кажется, что жизнь отвратительна или

---

[22] http://www.answers.com/topic/wilcoxon-signed-rank-test

удивительно прекрасна. На эту неочевидность оценки объекта исследования накладывается достаточно большой ретест, который характерен не только изменением уровня жизни, но и изменением самого измерительного инструмента, а именно испытуемого. И что самое страшное – это изменение инструмента вовсе не изменение уровня жизни, а попросту плохая память и интуитивное случайное размытие интервалов. И вот на эту какофонию накладывается ещё и тот факт, что у нас было целых 100 возможных единиц оценки, а значит, человек вряд ли в состоянии точно оценить изменение с 95 до 96 или 97 баллов. Если он решил во время первого тестирования пошутить и написать 94, а во время второго – 96, то это настоящая катастрофа для нас, поскольку мы не просто смотрим изменение его удовлетворённости жизнью, а выраженность этих изменений.

Поэтому для описанной исследовательской ситуации не стоит по традиции закрывать глаза и наивно предполагать ранжированную метрическую шкалу, когда на самом деле она не более чем порядковая в подлинном смысле этого слова.

Серьёзность этого допущения критерия Вилкоксона очень опасно для исследователя-прикладника, но именно оно, на мой взгляд, позволяет этому критерию добиваться сопоставимой с t-критерием мощности. В сущности, критерий Вилкоксона больше подходит для метрических данных, которые не распределены нормально, нежели для порядковых данных психологического исследования.

Что же касается 2-го ограничения, которое было указано выше, так это ограничение – мифический герой, который от учебника к учебнику возникает и исчезает. Например, в интерактивном учебнике STATISTICA и знаменитейшем руководстве Сидоренко Е.В. «Методы математической обработки в психологии» нет упоминаний об ограничении, касающемся симметричности. В это же самое время в оксфордском словаре есть это требование к критерию Вилкоксона. На сайте англоязычной Википедии такое ограничение есть. Мы же возьмём для цитаты книгу «International Encyclopedia of Statistical Science». Я даже приведу эту цитату на английском языке, чтобы не возникало соблазна свести эту спорную симметричность исключительно к моему неправильному переводу: «The classical Wilcoxon-signed-signed-rank test assumes that the differences D_i are mutually independent and D_i, i = 1, …, N comes from continuous distribution F that is symmetric about a median θ…

The null hyposesis states that H_0: $\theta = 0$, i.e., the distribution of the differences is symmetric about zero corresponding to no differences in location between the two samples. The two-sided alternative is H_1: $\theta \neq 0$…

Another well-known test for the one-sample location problem is the sign test. Compared to the sign test, the Wilcoxon-signed-rank test has the additional assumption of the symmetry of the distribution but uses the ordering of the differences as additional information…».

Таким образом, мы видим, что у нас возникает некоторое недопонимание, заключающееся примерно следующей альтернативой: «критерий Вилкоксона требует предварительную проверку симметричности различий относительно общей медианы или сам является такой проверкой». Это тот аспект, который можно недопонять, особенно, учитывая, что где-то я даже видела ссылку на то, что при несимметричности данных стоит использовать критерий знаков. К вопросу о том, какая в этом разница. Мы помним, что t-Стьюдент выдвигает требование нормальности данных, но никому не придёт в голову сказать, что с помощью t-Стьюдента можно проверить нормальность данных.

Чтобы провести своё собственное расследование этого вопроса мы для начала посчитаем критерий Вилкоксона вручную.

Предположим, что располагаем данными теста толерантности к неопределённости n-теста, замеренными «до» и «после» просмотра серии постмодернистских фильмов. Очередное увлекательное исследование, которое никому не спасёт жизнь, но мы будем усиленно считать результаты, ведь цель наша сейчас совершенно другая.

до	после
23	24
21	18
24	20
25	29
27	23
28	35
29	36
30	38

1. Нам нужно выстроить таблицу, где мы посчитаем разность замера «до» и «после». Затем выстроим ранговый ряд этих разностей (всех, безотносительно знака). В случае, если изменение от замера к замеру не происходит, то эти данные выбрасываются из подсчёта. Если ранги повторяются, то высчитывается средний.

до	после	разнос ть	ранг
23	24	-1	1
21	18	3	2
24	20	4	4
25	29	-4	4
27	23	4	4

28	35	-7	6,5
29	36	-7	6,5
30	38	-8	8

2. Далее нам нужно посчитать сумму рангов отрицательных и положительных сдвигов.

$T^{+} = 10$

$T^{-} = 26$.

За эмпирическое значение Т берётся величина «редких» сдвигов. В нашем случае – это 10.

По таблице значимости находится критическое значение критерия Т-Вилкоксона. Поскольку в нашем случае 8 человек, то критическое значение равно 3.

В случае, если $T_{эмп} < T_{крит}$, нулевая гипотеза отклоняется и принимается альтернативная о нетипичности сдвига.

В нашем случае $T_{эмп}$ (10) < $T_{крит}$ (3), следовательно мы принимаем нулевую гипотезу о случайности сдвига.

Посчитать критерий Вилкоксона в программе AtteStat не составляет большого труда. Для этого достаточно открыть модуль непараметрической статистики, задать массивы сырых данных точно также, как они даны в вышеописанном примере, задать ячейку выхода данных и выбрать критерий Вилкоксона (парный).

Для наших данных результат подсчёта будет следующим:

Непараметрическая статистика и анализ качественных данных	
Учет связок	
Учет поправки на непрерывность	
Критерий Вилкоксона (парный)	
Статистика, Р-значение (двустороннее)	
10	0,262374488

Как мы видим, p>0,05, следовательно принимается нулевая гипотеза о случайности сдвигов.

2.6. Связь признаков.

2.6.1. Критерии Пирсона и Спирмена.

Для выяснения вопроса о том, согласованно ли изменяются, например, результаты испытуемого по двум тестам, можно использовать критерий Пирсона. Но прежде следует проверить данные на нормальность, потому что если такое предположение не подтверждается, используется критерий Спирмена для порядковых шкал.

Предположим, что мы провели исследование на тему связи интеллекта и креативности. В таком случае мы будем располагать следующими данными:

23	24
21	23
11	12
10	13
9	12
8	10
7	9
13	14
14	16
15	18
16	19
17	12
16	19
23	20
11	13

Обращать внимание на то, что у нас размах данных вряд ли соответствует адекватному для тестов креативности и интеллекта, не стоит.

Критерий Шапиро-Уилка позволяет сказать, что распределение обоих наборов данных нормально.

В таком случае нам стоит перейти к подсчёту корреляции.

Воспользуемся программой AtteStat. Нас интересует модуль корреляционного анализа.

Под интервалом выборки 1 мы задаём первую колонку данных, под интервалом выборки 2 – вторую. Задаём в любом удобном для нас свободном пространстве таблицы выходной интервал (достаточно указать 1 ячейку). Выбираем в качестве метода анализа коэффициент корреляции Пирсона и нажимаем расчёт.

Перед нами появится следующая информация:

Коэффициент Пирсона
0,893351747
Р-значение
7,28164E-06
Доверительный интервал 95%
0,702637626

0,964314474

Соответственно, бояться странного значения уровня значимости не стоит – это совершенно обычное число. Достаточно изменить формат ячейки на числовой и добавить количество знаков после запятой, как станет ясно, что это просто очень маленькое число, которое в других статистических программах было бы не столь точно - 0,0000072816.

Чтобы понять, что за корреляцию мы получили, достаточно посмотреть в две ячейки. Первая ячейка, в которую стоит нацелить взгляд, это уровень значимости или Р-значение. Если это значение меньше 0,05, то связь между явлениями статистически значима.

Как только мы увидели, что связь значима, нам нужно определить характер этой связи.

Корреляция может измеряться в пределах [-1;+1], причём значение 0 означает отсутствие связи, (-1) – отрицательная связь или обратная; +1 – положительная связь.

Отрицательная связь означает, что увеличение одной переменной согласуется с уменьшением другой. Например, возрастание выраженности интеллекта связано со снижением уровня креативности или наоборот.

Положительная связь означает сонаправленные изменения. У людей с высоким интеллектом наблюдается высокий уровень креативности, у людей с низким – низкий.

В случае, когда коэффициент корреляции используется для психодиагностических целей, например, при подсчёте ретестовой надёжности или надёжности параллельных форм, то в качестве удовлетворительного результата принимается значение корреляции, равное 0,7. Такая критическая точка выдвинута, исходя из соображений коэффициента детерминации, который равен квадрату коэффициента корреляции и означает, какова доля объяснённой дисперсии. Конечно, более логичной точкой отсчёта в таком случае был бы коэффициент корреляции, соответствующий коэффициенту детерминации в 0,51, но это значение является дробным, что, скорее всего, и послужило аргументом сделать началом отсчёта 0,49.

Для некоторых психодиагностических подсчётов, как и вообще подсчётов в экспериментальной психологии, может понадобиться усреднённый коэффициент корреляции, т.е. среднее арифметическое нескольких коэффициентов корреляции. Например, если мы будем вычислять ретестовую надёжность теста, состоящего из нескольких шкал, у нас появится несколько коэффициентов корреляции. Далее возникает вопрос о том, как презентовать в научной статьи такие результаты. Самый выгодный, если не единственный способ – это задать предел, в котором были получены коэффициенты корреляции, например, ретестовая надёжность n-теста: (0,75 – 0,86). Находить же среднее арифметическое между коэффициентами корреляции нельзя, поскольку эти величины не аддитивны. Если всё же по каким-то причинам нужно именно среднее арифметическое, то необходимо пользоваться в этом случае коэффициентами детерминации.

Критерий Пирсона не только налагает требование нормальности данных, но и исходит из линейности зависимости. Линейность зависимости можно и нужно проверять глазомерным способом с помощью диаграмм рассеяния, о которых говорилось выше. Если связь есть, но она не линейна, то коэффициент Пирсона будет не значим и близок к нулю.

Чтобы оценить связь нелинейного характера можно пользоваться критерием ранговой корреляции Спирмена, даже при том, что данные относятся к метрической шкале. Такой анализ встречается достаточно часто и основывается на том простом эмпирическом факте, что оценить нелинейную зависимость другими методами для психолога зачастую не представляется возможным. Критерий Спирмена может определить далеко не любую нелинейную связь, но за неимением навыка пользоваться более сложными процедурами подсчёта, нам остаётся только перейти к его описанию.

Критерий Спирмена мы посчитаем на том же самом массиве данных, на котором считали критерий Пирсона.

Для этого мы открываем меню корреляционного анализа в AtteStat и выбираем критерий Спирмена.

Результаты подсчёта следующие:

Показатель Спирмэна
0,84375
Р-значение
0,00007694944
Доверительный 95% интервал
0,583832174
0,946789703

Если мы, например, посчитали бы критерий Спирмена в STATISTICA, то получили бы несколько иные результаты, в связи с тем, что AtteStat делает поправки на повторяющиеся ранги.

Само значение интерпретируется тем же способом, что и при подсчёте критерия Пирсона.

Алгоритм подсчёта критерия Пирсона вручную выглядит следующим образом:

1. Коэффициент корреляции Пирсона подсчитывается по формуле:

$$r_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{SS_x SS_y}},$$

где $SS_x = \sum(x_i - \bar{x})^2$, $SS_y = \sum(y_i - \bar{x})^2$.

Эмпирическое значение данного коэффициента находится путём подсчёта эмпирического значения t по следующей формуле:

$$t = r\sqrt{\frac{n-2}{1-r^2}}$$

Полученное эмпирическое значение сравнивается с критическим значением t-критерия Стьюдента со степенями свободы df = n – 2. Если эмпирическое значение больше критического, то принимается нулевая гипотеза.

Агоритм подсчёта коэффициента корреляции Спирмена вручную выглядит следующим образом:

1. Коэффициент Спирмена вычисляется по формуле:

$$r_s = \frac{6\sum d_i^2}{n\,(n^2 - 1)},$$

где:

d_i – разность рангов двух признаков i-го элемента выборки.

Для проверки используются таблицы критерия t-Стьюдента.

Список литературы

Основная литература:

1. Готтсданкер Р. Основы психологического эксперимента. - М.: Изд-во МГУ, 1982.
2. Дружинин В.Н. Экспериментальная психология. М.: Инфра-М, 1997 (или СПб, Питер, 2000 2-е издание.)
3. Корнилова Т.В. Введение в психологический эксперимент. - М.,1997
4. Кэмпбелл Д. Модели экспериментов в социальной психологии и прикладных исследованиях. - М.: Прогресс, 1980
5. Методы исследования в психологии: квазиэксперимент./Под ред. Т.В. Корниловой. - М.: Форум, 1998.
6. Милграм С. Эксперимент в социальной психологии. - СПб.: Питер, 2000.
7. Пайнс Э., Маслач К. Практикум по социальной психологии. - СПб.: Питер, 2000.
8. Практикум по общей, экспериментальной и прикладной психологии. /Под ред. А.А. Крылова, С.А. Маничева. - СПб.: «Питер», 2000.
9. Солсо Р. Когнитивная психология. - М.: Тривола,1996.
10. Солсо Р., Джонсон Х., Бил К. Экспериментальная психология: практический курс. -СПб.: прайм-ЕВРОЗНАК, 2001.
11. Хьелл Л., Зиглер Д. Теории личности. – СПб.: Питер,1997.
12. Мартин Д. Психологические эксперименты. - СПб .: прайм-ЕВРОЗНАК, 2002.

Дополнительная литература:

1. Атватер И. Я вас слушаю... - М., 1988.
2. Ананьев Б.Г. О проблемах современного человекознания. - М., 1977.
3. Бардин К.В., Забродин Ю.М. Проблемы сенсорной психофизики.//Познавательные процессы: ощущения, восприятие. Под ред. А.В. Запорожца, Б.Ф. Ломова, В.П. Зинченко. М.: Педагогика, 1982.
4. Зароченцев К. Д., Худяков А. И. Экспериментальная психология. - М.: Проспект, 2005. С. 74
5. Дружинин В.Н. Структура и логика психологического исследования. - Изд.2-е, М., 1994.
6. Дубина И.Н. Математические основы эмпирических социально-экономических исследований. – Барнаул: Изд-во Алт. ун-та, 2006.
7. Зинченко В.П., Смирнов С.Д. Методологические вопросы психологии. - М., 1982.
8. Куликов Л.В. Психологическое исследование. - СПб.: Наука, 1994.
9. Мельников В.М., Ямпольский Л.Т. Введение в экспериментальную психологию личности. - М.: Просвещение, 1985.
10. Методы исследования в психологии: квазиэксперимент./Под ред. Т.В. Корниловой. - М.: Форум, ИНФРА-М, 1998.

11. Никандров В. В. Наблюдение и эксперимент в психологии. - СПб.: Речь, 2002.

12. Орлов А.И. Эконометрика. – М.: Экзамен, 2002.

13. Петренко В. Ф., Нистратов А. А. Построение вербального семантического дифференциала на базе русской лексики. - В кн: Исследование проблем речевого общения. - М.: Институт языкознания АН СССР, 1979.

14. Практикум по психологии. /Под ред. А.Н. Леонтьева, Ю.Б. Гиппенрейтер. - М.: Изд-во МГУ, 1972.

15. Практикум по экспериментальной и прикладной психологии. /Под ред. А.А. Крылова. СПб.: Изд-во СПб университета, 1997.

16. Психологический словарь / Под ред. В. П. Зинченко, Б. Г. Мещерякова. – 2-е изд., перераб. и доп. – М.: ООО «Издательство Астрель», 2004/.

17. Фейерабенд П. Избранные труды по методологии науки. М., 1986.

18. Франселла Ф., Баннистер Д. Новый метод исследования личности. М., 1987.

19. Шапкин С.А. Экспериментальное изучение волевых процессов. - М.: Смысл, 1997.

20. Экспериментальная психология. /Под ред. С. Стивенса. Т1. - М., 1960.

21. Экспериментальная психология. /Под ред. С. Стивенса. Т2. - М., 1963.

22. Экспериментальная психология. /Под ред. П. Фресса и Ж. Пиаже. Вып.1,2. - М.: Прогресс, 1966.

23. Экспериментальная психология. /Под ред. П. Фресса и Ж. Пиаже. Вып.3. - М.: Прогресс, 1970.

24. Экспериментальная психология. /Под ред. П. Фресса и Ж. Пиаже. Вып.4. - М.: Прогресс, 1973.

25. Экспериментальная психология. /Под ред. П. Фресса и Ж. Пиаже. Вып.5. - М.: Прогресс, 1975.

26. Экспериментальная психология. Под ред. П. Фресса и Ж. Пиаже. Вып.6. - М.: Прогресс, 1978.

Содержание