Министерство образования и науки Российской Федерации
КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

С.Д. НЕКРАСОВ

МАТЕМАТИЧЕСКИЕ МЕТОДЫ В ПСИХОЛОГИИ (MS EXCEL)

Рекомендовано отделением Научно-методического совета по математике Министерства образования и науки РФ в ЮФО в качестве учебного пособия для студентов, изучающих методы математической обработки эмпирических данных

Краснодар
2014

УДК 303.2:159.9
ББК 22.311:в7
Н 48

Рецензенты:
Доктор психологических наук, профессор
З.И. Рябикина
Доктор физико-математических наук, профессор
Е.А. Семенчин
Кандидат физико-математических наук, доцент
А.П. Савченко

Некрасов, С.Д.
Н 48 Математические методы в психологии (MS Excel): учеб. пособие. 3-е изд., испр. и доп. / С.Д. Некрасов. Краснодар: Кубанский гос. ун-т, 2014. 147 с. 500 экз.
ISBN 978-5-8209-0952-8

В пособии представлены основные математические методы, необходимые для получения статистических обоснований психологических гипотез, а также для проведения простейших расчетов с помощью Microsoft Office Excel.

Адресуется студентам, изучающим методы математической обработки эмпирических данных, а также может быть полезно исследователям психологических и социальных явлений.

УДК 303.2:159.9
ББК 22.311:в7

ISBN 978-5-8209-0952-8

ВВЕДЕНИЕ

Степень использования математических методов в практической и исследовательской деятельности является показателем зрелости различных наук. Психология не является исключением, психологами используются методы различных математических дисциплин, в том числе математической статистики, теории множеств, теории вероятностей, математической логики.

Использование математических методов позволяет исследователю осуществлять измерение свойств психической реальности, описывать и сравнивать качественные и количественные характеристики психологических феноменов, моделировать отдельные психические процессы, свойства и состояния явлений, делать обоснованные выводы и прогнозы.

Для психолога-исследователя, психолога-практика, студента, осваивающего профессию психолога, написаны пособия, позволяющие освоить математические методы обработки и анализа эмпирических данных.

Можно выделить основные пособия, выдержавшие проверку практикой, пользующиеся большим спросом у студентов, которые были использованы при подготовке настоящего пособия: «Методы математической обработки в психологии» Е.В. Сидоренко, «Математические методы психологического исследования: анализ и интерпретация данных» Д.А. Наследова. Кроме того, были использованы теоретические источники авторов: Е.Ю. Артемьевой и Е.М. Мартынова, Дж. Гласса и Дж. Стэнли, Дж. Гудвина, О.Ю. Ермолаева, В.С. Иванова, Г.Ф. Лакина, Д.Б. Мангейма и Р.К. Рича, Г.В. Осипова и Э.П. Андреева, В.И. Паниотто и В.С. Максименко, Г.В. Суходольского, Ю.Н. Толстовой, Ю.Н. Тюрина и А.А. Макарова и др.

При подготовке пособия решались следующие задачи:

1) обеспечение принципов последовательности, логичности и доступности изложения использования математических методов в психологии;

2) ориентация на оснащение учебного процесса студентов направления (специальности) «Психология» как очной, так и заочной формы обучения;

3) оснащение психолога-практика минимальным набором математических методов, необходимых для оперативной обработки собственного эмпирического материала;

4) демонстрация возможностей обработки полученных эмпирических данных и проведения простейших расчетов с помощью Microsoft Excel для начинающего пользователя.

Пособие состоит из десяти глав.

Начинается с основных понятий (число, величина, измерение, шкала) и технологий протоколирования эмпирических данных.

Две главы посвящены построению распределения частот выборок (абсолютных, относительных, кумулятивных, процентильных) и описательным статистикам (моде, среднему, стандартному отклонению).

Отдельные главы знакомят с непараметрическими (φ-критерий Фишера, λ-критерий Колмогорова-Смирнова, G-критерий знаков, U-критерий Манна-Уитни) и параметрическими (T-критерий Стьюдента, F-критерий Фишера) критериями сравнения свойств выборок и генеральных совокупностей.

В двух завершающих пособие главах приводится описание критериев выявления и сравнения силы связей свойств психологических явлений (r-критерий Спирмена, r-критерий Пирсона, Z-критерий Фишера), однофакторного дисперсионного анализа (ANOVA).

Пособие предназначено в первую очередь для студентов, может быть полезно различным исследователям психологических свойств и социальных процессов.

1. ОСНОВНЫЕ ПОНЯТИЯ

1.1. Число

Потребность в решении практических задач на заре человечества привела к возникновению понятия «число». Зародилось оно из интуитивных представлений и практической потребности людей, поэтому понимается всеми одинаково и с достаточной для общения ясностью.

Попытки разъяснить смысл понятия *число* предпринимались многими учеными. Пифагореец Филолай (V век до н.э.) утверждал, что «мощь числа проявляется, как нетрудно заметить, не только в деяниях демонов или богов, но и во всех поступках и помыслах людей, во всех ремеслах и музыке». Немецкий математик Леопольд Кронекер (1823–1891) подчеркнул эту мысль следующей фразой: «Господь Бог создал натуральные числа. Все остальное дело рук человеческих». Фридрих Энгельс (1820–1895) писал: «Число есть чистейшее количественное определение, какое мы только знаем. Но оно полно качественных различий». В этих размышлениях выделена особая роль понятия *число* как первичного базового понятия. Операции с *числами* стали называть *вычислениями*.

Числа могут быть точным и приближенным значением свойства некоторого явления. При выполнении вычислений психологам приходится иметь дело, как правило, не с точными числами, а с их приближенными значениями.

Вычислить точное значение исследуемого свойства сложнее, чем приближенное значение. Например, представить значение числа **sin2 + ln2** (точное значение) сложнее, чем числа 0,728, равного 0,035 + 0,693, где 0,035 – приближенное значение числа **sin2,** а 0,693 – приближенное значение числа **ln2**.

При выполнении вычислений, в том числе с использованием различной вычислительной техники (от калькулятора до компьютера), часто приходится иметь дело с числами, которые содержат разное количество знаков после запятой. Количество десятичных знаков, которых достаточно для формулирования психологических предположений, гипотез, выводов, зависит от цели измерения, но не превышает, как правило, одного-двух знаков после запятой. Остальные десятичные знаки целесообразно отбрасывать, если они расположены после запятой в десятичной записи числа, или заменять их нулями, если они расположены до запятой.

Правило округления. Чтобы округлить число до какого-нибудь разряда, отбрасывают все цифры, расположенные правее этого разряда, или заменяют их нулями.

Если первая за сохраняемым разрядом цифра 0, 1, 2, 3, 4, то значение сохраняемого разряда не изменяют.

Если первая за сохраняемым разрядом цифра 5, 6, 7, 8, 9, то значение сохраняемого разряда увеличивают на единицу.

Пример округления чисел

16,80 = 16,8	16,85 = 16,9
16,81 = 16,8	16,86 = 16,9
16,82 = 16,8	16,87 = 16,9
16,83 = 16,8	16,88 = 16,9
16,84 = 16,8	16,89 = 16,9

1.2. Качество и количество

Исследуемая сущность психического явления, как правило, имеет наименование и ограниченное число свойств.

Термины, обозначающие явления и их свойства, принято называть *качественными* характеристиками явления.

Если описание свойства явления содержит числа, то эти числа принято называть *количественными* (численными) характеристиками явления.

Пример количественных и качественных характеристик

Исследуемое явление – свойства памяти человека, различающиеся по времени сохранения информации.

Качественные характеристики свойств памяти (наименования видов памяти): мгновенная, кратковременная, оперативная и т.д.

Количественные характеристики свойств:

– до 0,5 с – мгновенная память,

– от 0,5 до 20 с – кратковременная память и т.д.

Значения свойств явления бывают: дискретные или непрерывные; постоянные (неизменяемые) или переменные (изменяемые); связанные (взаимозависимые) или несвязанные (независимые).

Численные значения явления называются *дискретными*, если эти значения можно пронумеровать.

Численные значения явления называются *непрерывными*, если эти значения невозможно пронумеровать.

Пример дискретных и непрерывных значений

Исследуемое явление – внимание учеников во время нескольких уроков.

Дискретное значение – количество учеников на каждом уроке.

Непрерывное значение – изменение уровня внимания одного ученика на уроке.

Численное значение явления называется *постоянным*, если оно не изменяется на протяжении какого-нибудь процесса.

Численное значение явления называется *переменным*, если оно принимает разные значения на протяжении какого-нибудь процесса.

Пример постоянных и переменных величин

Исследуемое явление – внимание учеников во время урока.

Постоянными величинами являются, как правило, число учеников на уроке, продолжительность урока и др.

Переменные величины – уровни внимания на уроке различных учеников.

Две переменные называются *независимыми* на протяжении какого-нибудь процесса, если изменение одной переменной не связано с изменением другой переменной на протяжении этого процесса.

Две переменные называются *взаимозависимыми* на протяжении какого-нибудь процесса, если изменение одной переменной связано с изменением другой переменной на протяжении этого процесса.

Пример зависимых и независимых переменных

Исследуемое явление – связь учебной мотивации и уровня внимания учеников во время урока.

Переменные: количество учеников на уроке, учебная мотивация ученика, уровень внимания ученика, возраст ученика.

Зависимые переменные: учебная мотивация и уровень внимания ученика во время урока.

Независимые переменные: количество учеников на уроке, возраст учеников.

1.3. Измерение и шкала

Нахождение численных значений свойств явлений по установленным правилам (инструментам измерения) принято называть *измерением*.

Измерение – один из основных методов исследований свойств явлений. Если рассматривать человека как сложное явление, то его свойства можно классифицировать на физические и психические.

Измерение свойств физических явлений имеет продолжительную историю. Единицы измерения свойств физических явлений установлены естественно-научным путем и определены для нужд практики. Для измерения свойств физических явлений введены начало отсчета и мера измерения. Мера (единица) задает название шкалы измерения: шкала роста человека, шкала веса человека, шкала температуры тела и т.п.).

Инструментом измерения свойств физических явлений служит непрерывная шкала, которая содержит начало и меру измерения. Начало шкалы принято обозначать числом 0 (нуль), меру измерения – числом 1 (единица).

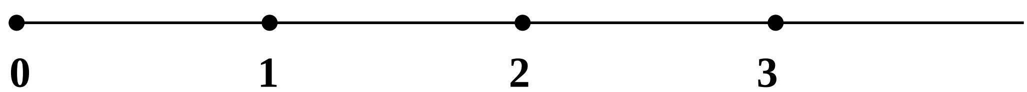


Рис. 1. Модель шкалы измерения свойств физических явлений

В психологии шкала измерения свойств физических явлений используется, если возможно установить начальную точку и единицу измерения свойства психического явления. Данную шкалу называют шкалой отношений.

Шкала отношений – это правило, устанавливающее взаимно-однозначное соответствие множества возможных значений свойства явления с числовым отрезком. Особенностью шкалы отношений является существование:

– начальной точки отсчета, которая обозначается нулём;

– меры измерения, которая называется единицей измерения.

Пример шкалы отношений

Измеряемое явление – стаж работы служащего в организации.

Инструментом измерения стажа работы служит анализ биографических данных респондента.

Нулевая точка шкалы отношений соответствует началу работы в организации.

Единица шкалы – один год (365 дней).

Шкалу отношений называют также измерительной шкалой, иногда абсолютной шкалой.

Если сложно установить начальную точку измерения свойства психического явления, но можно задать единицу измерения, то используют шкалу интервалов.

Шкала интервалов – это правило, устанавливающее классификацию множества возможных значений свойства психического явления по упорядоченным интервалам, содержащим установленное количество заданных единиц измерения.

Единица измерения, как правило, устанавливается эмпирическим путем.

Шкала интервалов имеет градации – интервалы, длина которых равна установленному количеству единиц измерения. Для вычисления границ градаций шкалы интервалов используют статистические методы: процентильное распределение, нормальное распределение и др.

Шкала интервалов состоит из упорядоченной последовательности градаций. Крайние градации являются, как правило, открытыми интервалами. Шкалу интервалов также называют интервальной шкалой, метрической шкалой.

Пример шкалы интервалов

Психическое свойство – уровень интеллектуального развития человека.

Инструментом измерения уровня интеллектуального развития респондента служит тест IQ.

Единица измерения – один балл теста IQ. Для вычисления границ градаций используют: среднее – 100 баллов, стандартное отклонение – 15 баллов.

Градации шкалы интервалов:

– до 54 баллов;
– от 55 до 69 баллов;
– от 70 до 84 баллов;
– от 85 до 99 баллов;
– от 100 до 114 баллов;
– от 115 до 129 баллов;
– от 130 до 144 баллов;
– свыше 145 баллов.

Если сложно задать единицу измерения, но можно установить уровни выраженности свойства психического явления, то используют шкалу порядка.

Шкала порядка – правило, устанавливающее классификацию множества возможных значений свойства психического явления по разным уровням выраженности этого свойства.

Уровни выраженности свойства психического явления, как правило, устанавливаются эмпирическим путем. Градации шкалы порядка – это качественные описания каждого уровня свойства психического явления. Разные градации отличаются по признаку «меньше – больше», «выше – ниже», «лучше – хуже» и т.п.

Шкала порядка состоит из возрастающей последовательности градаций.

Пример шкалы порядка

Психическое свойство – академическая успеваемость студента.

Инструментом измерения академической успеваемости студента служит экзамен.

Градации шкалы порядка: плохо, удовлетворительно, хорошо, отлично.

Описание градаций шкалы порядка:

– плохо (не смог ответить на оба вопроса билета);

– удовлетворительно (ответил на оба вопроса билета, но допустил более двух ошибок);

– хорошо (ответил на оба вопроса билета, но допустил одну-две ошибки);

– отлично (ответил на оба вопроса билета без ошибок).

Шкалу порядка также называют порядковой шкалой, ранговой шкалой.

Если сложно установить уровни выраженности свойства психического явления, но можно выделить качественное основание классификации свойства, то используют шкалу наименований.

Шкала наименований – правило, устанавливающее качественную классификацию множества различных значений свойства психического явления, которые нельзя упорядочить.

Отдельные градации являются описанием однородных качеств психического явления, как правило, найденных эмпирическим путем.

Градации шкалы наименований задаются описанием однородных значений свойства явления. Каждой градации присваивается наименование.

Пример шкалы наименований

Измеряемое явление – тип темперамента человека.

Инструментом измерения типа темперамента респондента являются особенности нервных процессов.

Градации шкалы наименований:

– меланхолик (слабый, неуравновешенный, инертный);

– сангвиник (сильный, уравновешенный, подвижный);

– флегматик (сильный, уравновешенный, инертный);

– холерик (сильный, неуравновешенный, подвижный).

Шкалу наименований также называют номинальной шкалой, номинативной шкалой, неметрической шкалой.

Резюме

Одним из основных методов исследования сущности психических явлений является измерение их свойств. Измерительная шкала – это правило проведения измерения. Выбор измерительной шкалы – важная часть любого психологического исследования.

Можно рекомендовать следующую последовательность использования различных шкал для измерения свойств явлений.

Полезно начинать с выделения различающихся однородных значений свойства исследуемого явления и описания основания для их различения. Название и описание сущности однородных значений свойства задают шкалу наименований.

Далее, если возможно, устанавливается уровень выраженности отдельного свойства исследуемого явления, например: низкий, средний, высокий. Правило установления уровня выраженности отдельного свойства задает шкалу порядка.

Если можно выделить меру выраженности отдельного свойства исследуемого явления, например, в баллах, то можно построить шкалу интервалов.

Если можно установить полное отсутствие исследуемого свойства явления и выделить меру его выраженности, то строится шкала отношений.

Упражнения

1.1. Приведите пример шкалы:
- а) наименований;
- б) порядка;
- в) интервалов;
- г) отношений.

1.2. Запишите значение собственного роста и укажите:
- а) единицу измерения;
- б) инструмент измерения;
- в) точность измерения.

1.3. Запишите значение собственного возраста и укажите:
- а) единицу измерения;
- б) инструмент измерения;
- в) точность измерения.

1.4. Определите вид шкалы:

а) шкала исчисления тысячелетий: …, второе тысячелетие до н.э., первое тысячелетие до н.э., первое тысячелетие н.э., второе тысячелетие н.э., …;

б) шкала «Человекоподобные приматы»:

– семейство Гибоновые;
– семейство Люди;
– семейство Человекообразные обезьяны;

в) шкала удовлетворенности:

– «вполне удовлетворен»;
– «удовлетворен»;
– «скорее удовлетворен, чем не удовлетворен»;
– «затрудняюсь сказать»;
– «скорее не удовлетворен, чем удовлетворен»;
– «не удовлетворен»;
– «совершенно не удовлетворен»;

г) шкала ощущения громкости звука в децибелах;

д) шкала аффилиации при ожидании поезда:

– желание находиться вместе с другими;
– желание находиться в одиночестве;
– без предпочтений;

е) шкала популярности психологических методик (в порядке убывания):

– личностный опросник Р. Кеттела;
– цветовой тест М. Люшера;
– опросник MMPI, тест IQ;
– тест Д. Векслера;
– проективные методики «Рисунок несуществующего животного» и «Дом, дерево, человек»;
– другие методики.

2. ПРОТОКОЛИРОВАНИЕ ЭМПИРИЧЕСКИХ ДАННЫХ

Исследуя сущность психических явлений, психолог формулирует предположения, теоретически их обосновывает и ищет им эмпирические подтверждения.

Поиск эмпирических подтверждений гипотез исследования начинается со сбора количественных и качественных характеристик сущности явления, которые принято называть эмпирическими данными.

Полученные эмпирические данные заносятся в документы, основные из которых:

– индивидуальные данные респондента (опросные листы, анкеты, листы ответов, бланки исследования и др.);

– протокол сводных данных (таблица результатов исследования).

2.1. Бланк исследования

Индивидуальные данные на каждого респондента принято заносить в документ, который будем называть бланком исследования. Бланк исследования заполняется, как правило, индивидуально на каждого респондента.

Бланк исследования состоит из следующих частей:

– заголовок исследования,

– краткие биографические данные респондента,

– регистрируемые эмпирические данные.

В заголовке, как минимум, отражаются краткое название, время и место проведения исследования.

Краткие биографические данные респондента содержат фамилию, имя, возраст респондента, а также другие интересующие исследователя сведения о респонденте (например, образование, стаж работы, семейное положение и т.п.).

Содержание регистрируемых эмпирических данных определяется задачами и шкалой методики исследования.

Пример бланка исследования самооценки и оценки учебной компетентности студентов-бакалавров направления «Психология»

Дата проведения «20» февраля 2013 г.

Анисимов Петр, число полных лет – 20 лет.

1. Оценка собственной компетентности (десятибалльная шкала):

а) по общей психологии – 8 баллов;

б) по математике – 6 баллов.

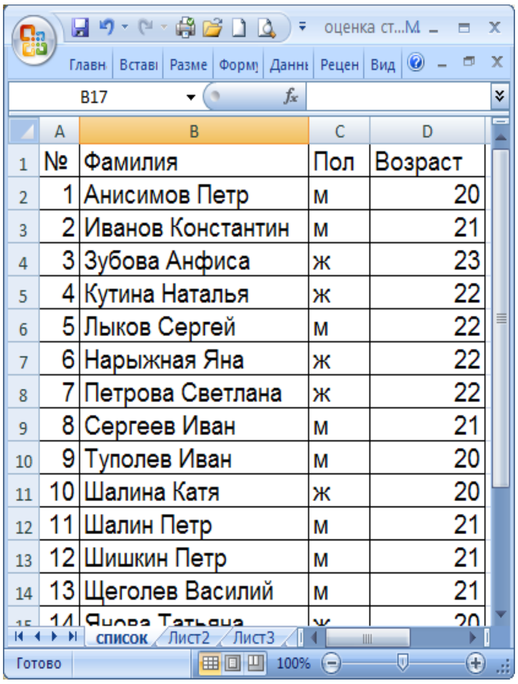
2. Сумма оценок, полученных в зимнюю сессию – 23 балла.

Собранные бланки исследования нумеруются.

В Microsoft Office Excel составляется таблица, в которую заносятся номер бланка исследования, фамилия и имя, пол, возраст респондентов.

Файлу, как правило, присваивают краткое название предмета исследования, а первому листу в файле – название «Список».

Пример списка респондентов (Excel)



№	Фамилия	Пол	Возраст
1	Анисимов Петр	м	20
2	Иванов Константин	м	21
3	Зубова Анфиса	ж	23
4	Кутина Наталья	ж	22
5	Лыков Сергей	м	22
6	Нарыжная Яна	ж	22
7	Петрова Светлана	ж	22
8	Сергеев Иван	м	21
9	Туполев Иван	м	20
10	Шалина Катя	ж	20
11	Шалин Петр	м	21
12	Шишкин Петр	м	21
13	Щеголев Василий	м	21
14	Янова Татьяна	ж	20

Список респондентов необходим исследователю для идентификации в случае надобности данных о респонденте и замены его фамилии и имени на присвоенный номер.

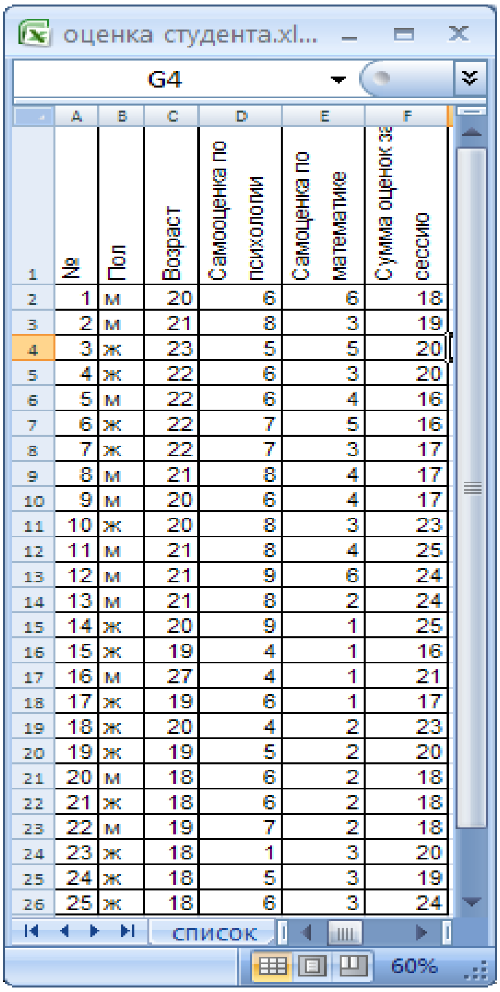
2.2. Протокол эмпирических данных

Протокол результатов исследования представляет собой таблицу, в которую заносятся эмпирические данные из собранных опросных листов.

Протокол рекомендуется заполнять в Excel на новом *листе* в том же файле, в котором содержится *лист* со списком. *Листу* рекомендуется присваивать название «Протокол».

В *лист* «Протокол» переносятся из *листа* «Список» номер регистрационного бланка, пол респондента, возраст респондента. Столбцам протокола, следующим за столбцом «Возраст», присваивают наименования, соответствующие каждому виду эмпирических данных. В соответствующих строках таблицы размещают полученные эмпирические данные каждого отдельного респондента из их бланков исследования.

Пример протокола исследования «Оценка – самооценка» (Excel)



	A	B	C	D	E	F
1	№	Пол	Возраст	Самооценка по психологии	Самооценка по математике	Сумма оценок за сессию
2	1	м	20	6	6	18
3	2	м	21	8	3	19
4	3	ж	23	5	5	20
5	4	ж	22	6	3	20
6	5	м	22	6	4	16
7	6	ж	22	7	5	16
8	7	ж	22	7	3	17
9	8	м	21	8	4	17
10	9	м	20	6	4	17
11	10	ж	20	8	3	23
12	11	м	21	8	4	25
13	12	м	21	9	6	24
14	13	м	21	8	2	24
15	14	ж	20	9	1	25
16	15	ж	19	4	1	16
17	16	м	27	4	1	21
18	17	ж	19	6	1	17
19	18	ж	20	4	2	23
20	19	ж	19	5	2	20
21	20	м	18	6	2	18
22	21	ж	18	6	2	18
23	22	м	19	7	2	18
24	23	ж	18	1	3	20
25	24	ж	18	5	3	19
26	25	ж	18	6	3	24

В протоколах содержатся основные данные о результатах исследования, которые называют *сырыми данными*, подчеркивая их первичный характер.

2.3. Выборка

Выборкой (выборочной совокупностью) называется множество измеренных значений свойства явления. Выборкой также называют множество испытуемых.

Выборки рекомендуется называть кратким наименованием измеренного свойства явления или обозначать символами.

Пример выборки

Выборка оценок собственной компетентности по общей психологии обозначена «Самооценка по психологии».

Значения выборки «Самооценка по психологии»:

6, 8, 5, 6, 6, 7, 7, 8, 6, 8, 8, 9, 8, 9, 4, 4, 6, 4, 5, 6, 6, 7, 1, 5, 6.

Объемом выборки называется количество измеренных значений свойства явления.

Объем выборки будем обозначать буквой n.

Пример объема выборки

Объем выборки «Самооценка по психологии» равен 25.

Обозначение: $n = 25$.

Вариантой выборки называется отдельное значение свойства явления (элемент множества измеренных значений).

Пример варианты выборки

6 – одна из вариант выборки «Самооценка по психологии».

Вариационным рядом называется последовательность значений вариант, расположенных в порядке возрастания (если измерительная шкала неноминальная).

Пример вариационного ряда

Вариационный ряд выборки «Самооценка по психологии»:

1, 4, 5, 6, 7, 8, 9.

Резюме

Протоколирование эмпирических данных – важный этап исследования.

Рекомендации для составления протокола.

1. В одну ячейку протокола заносят только одну информационную единицу.
2. Не рекомендуется объединение ячеек протокола.
3. На листе протокола расчеты не проводят.
4. Для расчетов используются копии протокола.

Упражнения

2.1. Проведите опрос «Оценка – самооценка» студентов-психологов (не менее 20 респондентов).

Бланк исследования

Фамилия, имя ___________________ Число полных лет ______

Оценка собственной компетентности (десятибалльная шкала):
а) по психологии – ____ баллов;
б) по математике – _____ баллов.
Сумма баллов оценок за зимнюю сессию – ___ балла.

2.2. Составьте протокол результатов исследования «Оценка – самооценка».

2.3. Сколько выборок получено в исследовании «Оценка – самооценка»?

2.4. Каковы объем и варианты выборки:

а) «Пол»;
б) «Возраст»;
в) «Самооценка по психологии»;
г) «Самооценка по математике»;
д) «Сумма баллов за сессию».

3. РАСПРЕДЕЛЕНИЯ ЧАСТОТ ВЫБОРКИ

3.1. Частота варианты выборки

Частотой варианты называется число одинаковых значений этой варианты в выборке.

Частоту варианты из вариационного ряда принято обозначать символом n_i, где i – место варианты в вариационном ряду.

Пример частоты варианты

Частота варианты «6» выборки «Самооценка по психологии» равна 8, то есть $n_2 = 8$.

3.2. Распределение частот выборки

Распределением частот вариант выборки называется таблица, содержащая две строки:

– перечень вариант выборки;

– перечень частот, соответствующих вариантам.

Пример распределения частот выборки (x_i)

Варианта (x_i)	x_1	x_2	x_3	...	x_k
Частота (n_i)	n_1	n_2	n_3	…	n_k

Заметим, что сумма частот всех вариант выборки равна объему выборки.

Пример распределения частот выборки «Самооценка по математике»

Самооценка по математике	1	2	3	4	5	6
Частота	4	6	7	4	2	2

Заметим, 4 + 6 + 7 + 4 + 2 + 2 = 25, что равно объему выборки «Самооценка по математике».

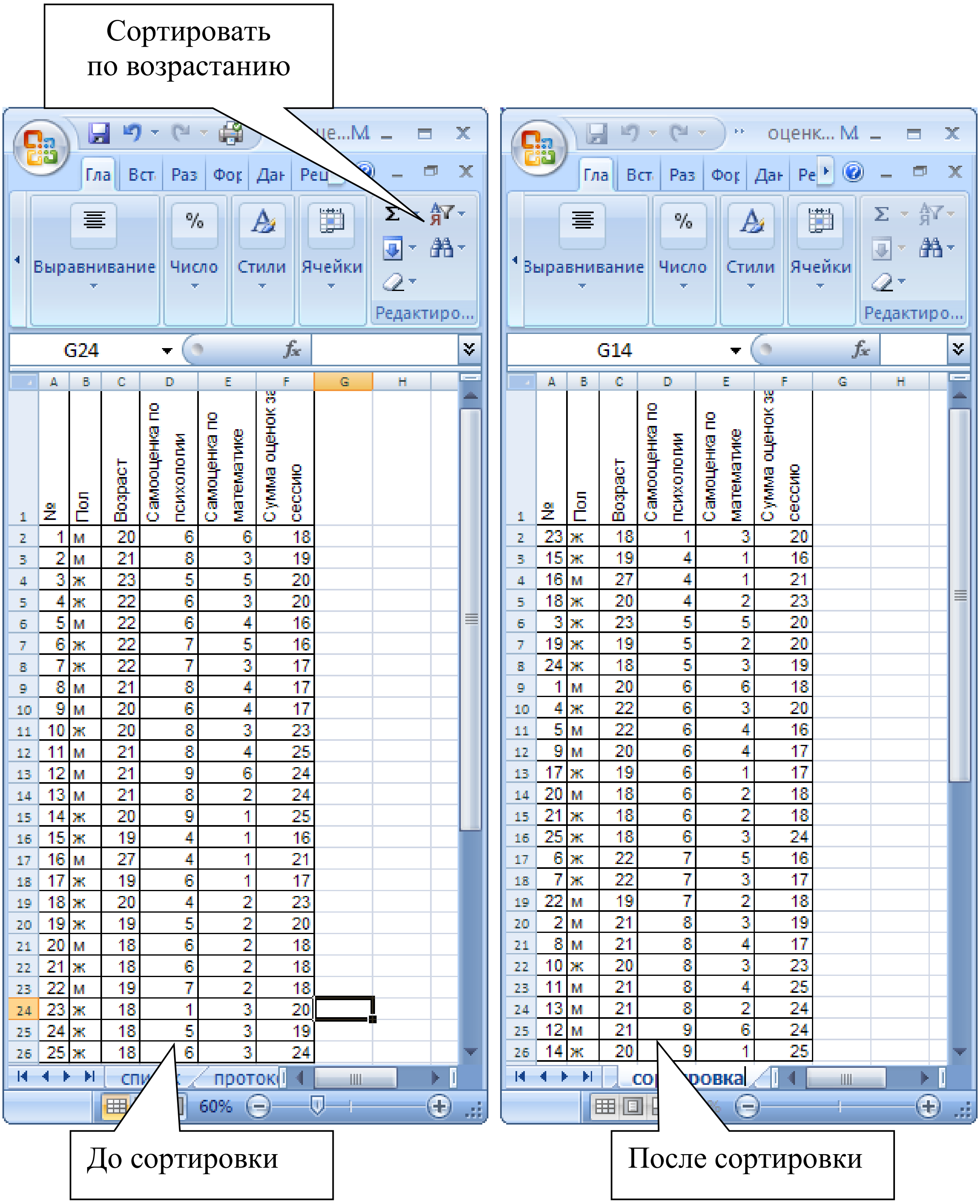
Алгоритм построения распределения частот выборки (Excel)

1. В файле с протоколом эмпирических данных создайте новый лист с названием «Сортировка».

2. Скопируйте «Протокол» и вставьте на лист «Сортировка».

3. Курсор установите на название сортируемого столбца.
4. Нажмите кнопку «Сортировать по возрастанию».
5. Составьте таблицу искомого распределения частот.

Пример сортировки выборки «Самооценка по психологии»



До сортировки:

№	Пол	Возраст	Самооценка по психологии	Самооценка по математике	Сумма оценок за сессию
1	м	20	6	6	18
2	м	21	8	3	19
3	ж	23	5	5	20
4	ж	22	6	3	20
5	м	22	6	4	16
6	ж	22	7	5	16
7	ж	22	7	3	17
8	м	21	8	4	17
9	м	20	6	4	17
10	ж	20	8	3	23
11	м	21	8	4	25
12	м	21	9	6	24
13	м	21	8	2	24
14	ж	20	9	1	25
15	ж	19	4	1	16
16	м	27	4	1	21
17	ж	19	6	1	17
18	ж	20	4	2	23
19	ж	19	5	2	20
20	м	18	6	2	18
21	ж	18	6	2	18
22	м	19	7	2	18
23	ж	18	1	3	20
24	ж	18	5	3	19
25	ж	18	6	3	24

После сортировки:

№	Пол	Возраст	Самооценка по психологии	Самооценка по математике	Сумма оценок за сессию
23	ж	18	1	3	20
15	ж	19	4	1	16
16	м	27	4	1	21
18	ж	20	4	2	23
3	ж	23	5	5	20
19	ж	19	5	2	20
24	ж	18	5	3	19
1	м	20	6	6	18
4	ж	22	6	3	20
5	м	22	6	4	16
9	м	20	6	4	17
17	ж	19	6	1	17
20	м	18	6	2	18
21	ж	18	6	2	18
25	ж	18	6	3	24
6	ж	22	7	5	16
7	ж	22	7	3	17
22	м	19	7	2	18
2	м	21	8	3	19
8	м	21	8	4	17
10	ж	20	8	3	23
11	м	21	8	4	25
13	м	21	8	2	24
12	м	21	9	6	24
14	ж	20	9	1	25

Столбец «После сортировки» позволяет непосредственно сосчитать частоты выборки «Самооценка по психологии» и составить искомое распределение частот.

Таблица распределения частот выборки «Самооценка по психологии»

Самооценка по психологии	1	4	5	6	7	8	9
Частота	1	3	3	8	3	6	1

3.3. ОТНОСИТЕЛЬНАЯ ЧАСТОТА ВАРИАНТЫ

Относительной частотой варианты называется число, равное отношению частоты варианты к объему выборки (n_i / n).

Относительную частоту можно представить в виде процентов ($(n_i / n) \cdot 100\%$).

Пример распределения относительных частот выборки «Самооценка по математике»

Самооценка по математике	1	2	3	4	5	6
Относительная частота	0,16	0,24	0,28	0,24	0,08	0,08
Относительная частота, %	16%	24%	28%	24%	8%	8%

Алгоритм построения распределения относительных частот выборки (Excel)

1. В файле с протоколом эмпирических данных новый лист назовите «Частота». На листе разместите распределение частот выборки.

2. К таблице распределения частот выборки добавьте строку «Относительная частота» и установите курсор на первую слева ячейку этой строки.

3. Последовательно выполните операции:

– нажмите клавишу со знаком «=»;

– курсор поставьте в ячейку, содержащую значение частоты первой варианты;

– нажмите клавишу со знаком «/»;

– наберите значение объема выборки;

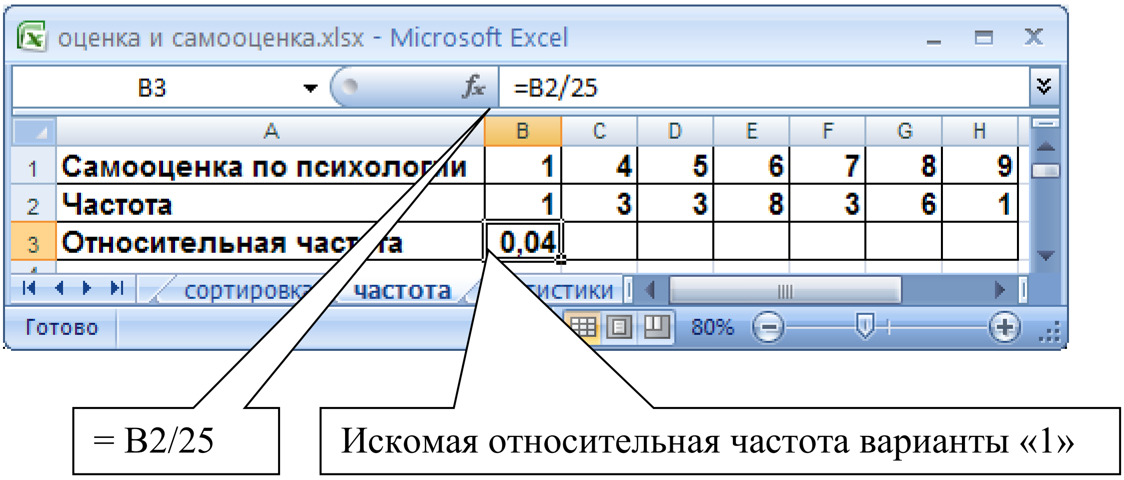
– в строке «f_x» появится запись: = (код ячейки)/(значение объема выборки);

– нажмите клавишу со знаком «Enter».

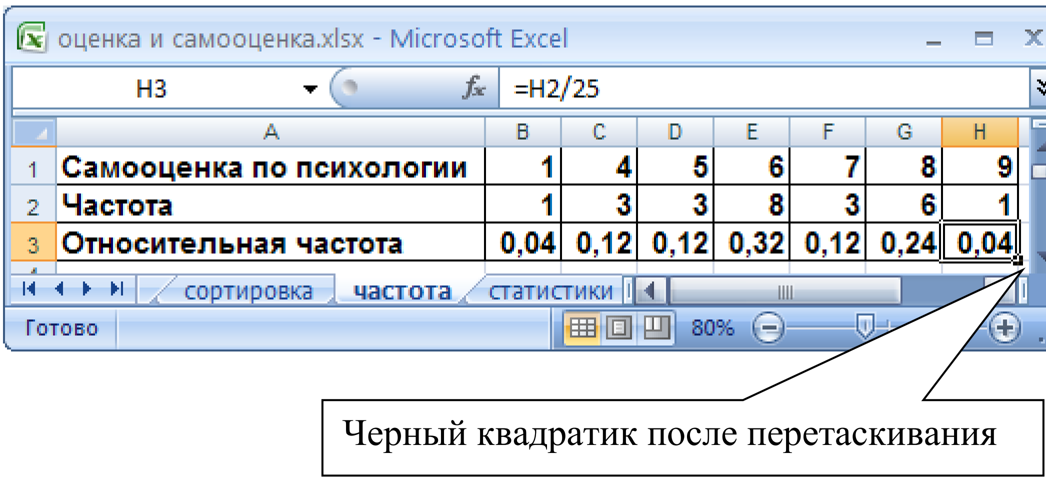
4. Найдите в правом нижнем углу ячейки черный квадратик и перетащите его в остальные пустые ячейки строки «Относительная частота».

Пример построения распределения относительных частот выборки «Самооценка по психологии» (Excel)

К распределению частот выборки «Самооценка по психологии» добавим строку «Относительная частота» и составим формулу.



Найдем черный квадратик в правом нижнем углу ячейки и перетащим его курсором в остальные пустые ячейки строки «Относительная частота».



3.4. Кумулятивная частота варианты

Кумулятивной частотой варианты ($\boldsymbol{x}_i$) называется число, равное сумме частот всех вариант, не превышающих x_i: $n_1 + n_2 + \ldots + n_i$.

Кумулятивную частоту называют также накопленной частотой варианты. Обозначение: $CUMi = n_1 + n_2 + \ldots + n_i$.

Заметим, что кумулятивная частота последней варианты вариационного ряда равна объему выборки.

Пример распределения кумулятивных частот выборки «Самооценка по математике»

Самооценка по математике	1	2	3	4	5	6
CUM	4	10	17	21	23	25

Заметим, что кумулятивная частота *последней варианты выборки* (8) равна 25, то есть *CUM*(6) = 25, что равно объему выборки «Самооценка по математике».

Алгоритм построения распределения кумулятивных частот выборки (Excel)

1. На листе «Частота» постройте распределение частот выборки.

2. К таблице распределения частот выборки добавьте строку «Кумулятивная частота».

3. В первую ячейку строки «Кумулятивная частота» запишите значение первой ячейки строки «Частота».

4. Курсор установите на вторую ячейку строки «Кумулятивная частота».

5. Последовательно выполните операции:

– нажмите клавишу со знаком «=»;

– курсор поставьте в первую ячейку строки «Кумулятивная частота»;

– нажмите клавишу со знаком «+»;

– курсор поставьте во вторую ячейку строки «Частота»;

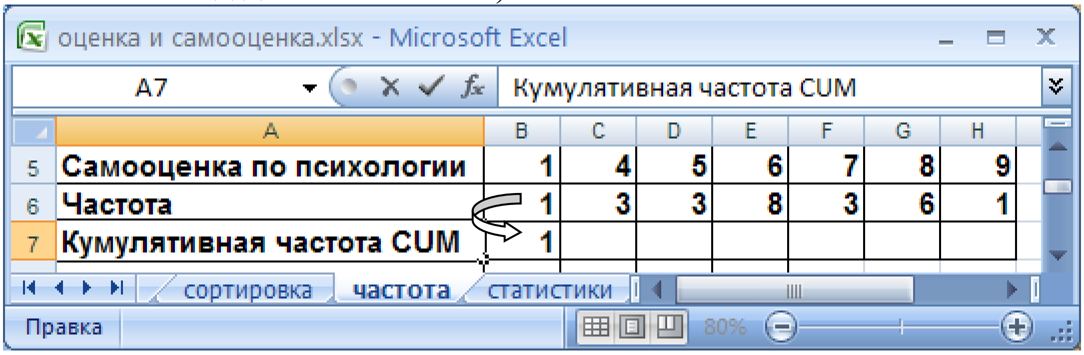
– в строке «f_x» появится запись: = (код первой ячейки строки «Кумулятивная частота») + (код второй ячейки строки «Частота»);

– нажмите клавишу со знаком «Enter».

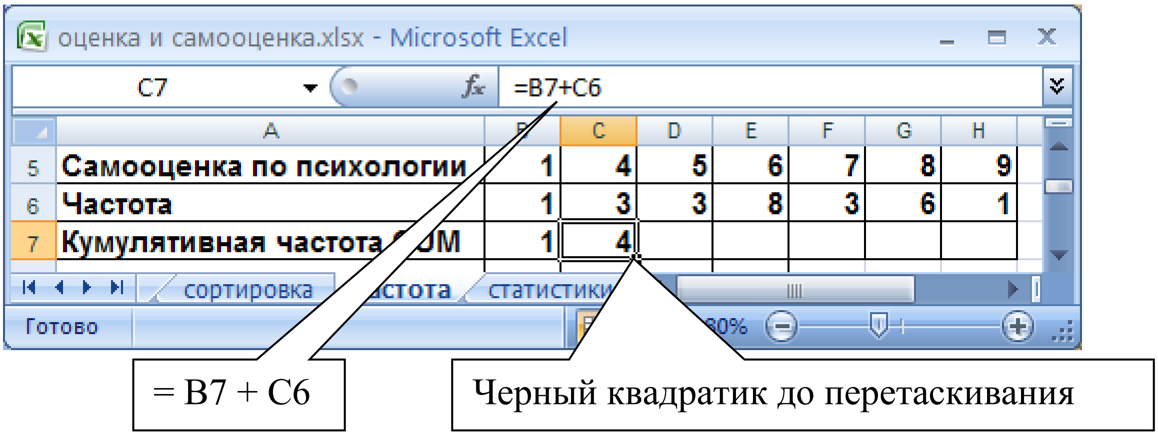
6. Найдите черный квадратик и перетащите его курсором в остальные пустые ячейки строки «Кумулятивная частота».

Пример построения распределения кумулятивных частот выборки «Самооценка по психологии» (Excel)

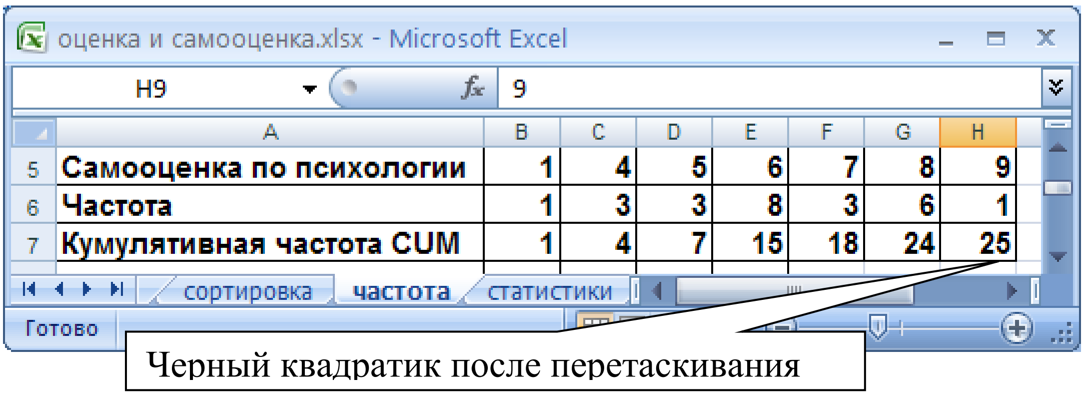
В первую слева ячейку строки «Кумулятивная частота» запишем значение первой ячейки строки «Частота» (из ячейки, расположенной над данной ячейкой).



Выполним операции.



Найдем черный квадратик в правом нижнем углу ячейки и перетащим его курсором в остальные пустые ячейки строки «Кумулятивная частота».



3.5. Процентильная частота варианты

Процентилем *Pk* называется такое значение варианты, левее (меньше) которой находится *k* процентов вариант выборки.

Процентильной частотой варианты (x_i) называется число, которое равно отношению кумулятивной частоты к объему выборки (*CUMi* / *n*).

Процентильную частоту можно представить в виде процентов ((*CUMi* / *n*) · 100%).

Обозначения: *PCUM* и *PCUM, %*.

Пример распределения процентильных частот выборки «Самооценка по математике»

Самооценка по математике	1	2	3	4	5	6
PCUM	0,16	0,40	0,68	0,84	0,92	1
PCUM, %	16%	40%	68%	84%	92%	100%

Заметим, что процентильная частота последней варианты вариационного ряда (6) равна 1 (100%).

Алгоритм построения распределения процентильных частот выборки (Excel)

1. На листе «Частота» постройте распределение кумулятивных частот выборки.

2. К таблице распределения кумулятивных частот выборки добавьте строку «*PCUM*» (процентильная частота).

3. Курсор установите на первую слева ячейку строки «*PCUM*».

4. Последовательно выполните операции:

– нажмите клавишу со знаком «=». Курсор поставьте в ячейку, содержащую значение частоты первой варианты;

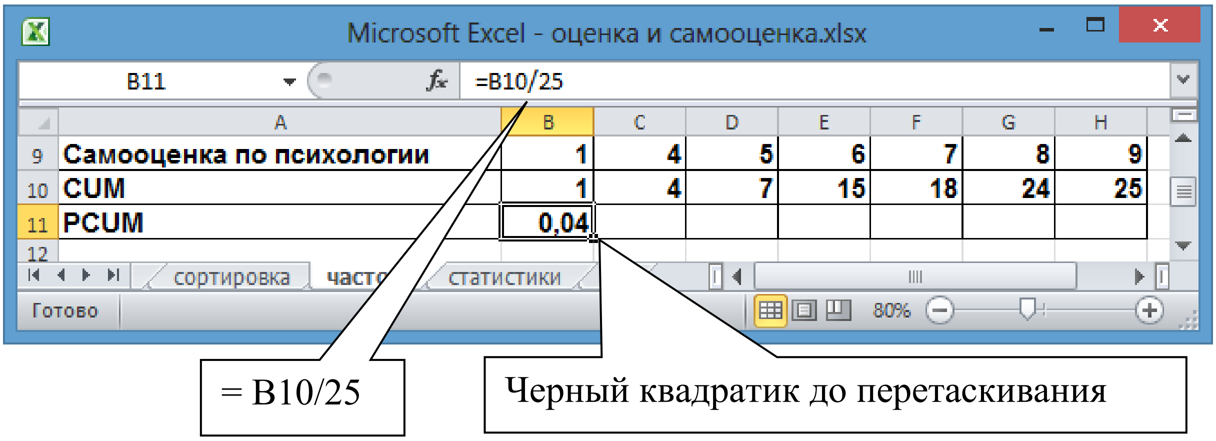
– нажмите клавишу со знаком «/». Наберите значение объема выборки. В строке «f_x» появится запись: = (код ячейки)/(значение объема выборки);

– нажмите клавишу со знаком «Enter».

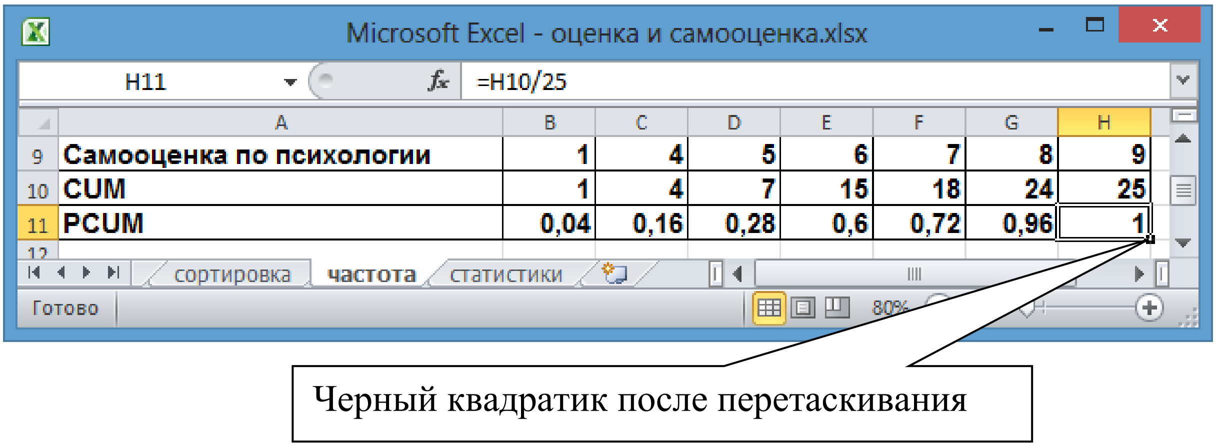
4. Найдите в правом нижнем углу ячейки черный квадратик и перетащите его курсором в остальные пустые ячейки строки «*PCUM*».

Пример построения распределения относительных частот выборки «Самооценка по психологии» (Excel)

К распределению *кумулятивных* частот добавим строку «*PCUM*» и составим формулу.



Найдем черный квадратик в правом нижнем углу ячейки и перетащим его курсором в остальные пустые ячейки строки «Процентильная частота».



Резюме

С построения распределения частот выборок, как правило, начинается анализ и обобщение эмпирических данных. Распределения частот выборок позволяют:

– классифицировать выборку по уровням выраженности измеренного свойства с использованием процентильного распределения частот;

– выявлять уровень выраженности исследуемого свойства у отдельных респондентов;

– проводить сравнение распределений частот с использованием диаграмм.

Рекомендации для построения диаграмм

1. Диаграммы строятся в прямоугольной системе координат с помощью Excel.

2. На горизонтальной оси отмечают значения вариант выборок.

3. Название оси и градаций шкалы подписывают снизу.

4. На вертикальной оси отмечают относительную (в процентах) частоту вариант выборок.

5. Название оси и градации шкалы подписывают слева.

6. Рекомендуется в одной диаграмме размещать не более пяти распределений частот выборок.

Пример диаграммы распределения частот выборок, измеренных в шкале наименований

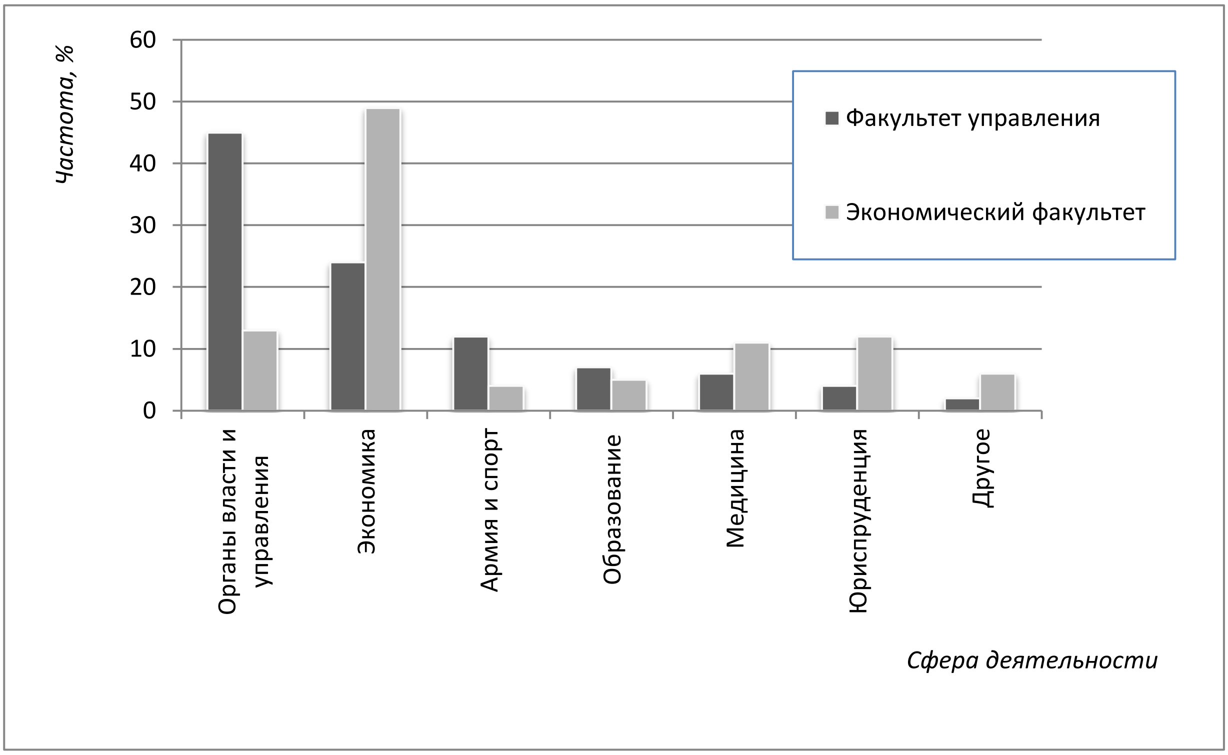


Рис. 2. Диаграмма представлений абитуриентов о сфере будущей профессиональной деятельности

Упражнения

3.1. Постройте распределение частот исследования «Оценка – самооценка» выборки:

а) «Пол»;

б) «Возраст»;

в) «Сумма баллов за сессию».

3.2. Постройте распределение относительных частот исследования «Оценка – самооценка» выборки:

а) «Возраст»;

б) «Сумма баллов за сессию».

3.3. Постройте распределение кумулятивных частот исследования «Оценка – самооценка» выборки:

а) «Возраст»;

б) «Сумма баллов за сессию».

3.4. Определите для выборки «Оценка – самооценка» уровень выраженности самооценок по шкале (низкий, средний, высокий):

а) по психологии; б) по математике.

3.5. В таблице представлены распределения частот представлений студентов экономического факультета и факультета управления о своей будущей профессиональной деятельности. Изобразите на диаграмме результаты исследования.

Факультет	Сфера будущей профессиональной деятельности										
	Государственная служба	Предпринимательство	Органы власти	Образование	Медицина	Экономика	Армия	Культура	Юриспруденция	Спорт	Другое
управления	81	45	27	19	18	15	12	10	9	8	6
экономический	23	51	2	5	12	61	23	4	27	2	13

3.6. В таблице представлены результаты тестирования школьников по физике в баллах.

Имя	Балл	Имя	Балл	Имя	Балл	Имя	Балл
Алексей	60	Дима	55	Леонид	46	Роман	75
Алена	55	Елена	61	Марина	86	Света	58
Андрей	55	Жанна	51	Мария	64	Сергей	50
Белла	31	Зина	48	Михаил	85	Стас	24
Борис	89	Игорь	92	Настя	63	Тарас	27
Вадим	69	Ирина	42	Ник	84	Татьяна	75
Вера	38	Катя	71	Олег	62	Ульяна	80
Галина	39	Клава	73	Ольга	77	Федор	49
Гриша	52	Костя	64	Петр	77	Юрий	24
Дина	54	Лариса	70	Рита	39	Яна	90

Выполните следующие задания:

а) постройте интервальное распределение частот для выборки девочек и выборки мальчиков (длина интервала десять баллов);

б) постройте процентильное распределение частот всей выборки;

в) установите уровень академической успеваемости школьников по шкале с четырьмя градациями: неудовлетворительно, удовлетворительно, хорошо, отлично;

г) составьте список школьников, получивших оценку «отлично».

4. ОПИСАТЕЛЬНЫЕ СТАТИСТИКИ

Одной из задач психолога является составление психологических портретов как одного человека, так и группы людей. Решение этой задачи можно осуществлять с использованием выборочных данных путем выделения количественных характеристик выборки.

Числа, которые являются характеристиками выборки, называются *описательными статистиками выборки*.

К основным описательным статистикам выборки относятся: мода, среднее, стандартное отклонение.

4.1. Мода

Мода выборки – статистика, равная варианте выборки, частота которой в данной выборке наибольшая.

Выборку, в которой только одна варианта имеет наибольшую частоту, называют *унимодальной выборкой*. Выборка, в которой только две смежные варианты имеют наибольшую частоту, также является *унимодальной выборкой*.

Пример моды выборки

Самооценка по психологии	1	4	5	6	7	8	9
Частота	1	3	3	8	3	6	1

Мода выборки равна 6 баллов, так как частота соответствующей варианты равна 8 – наибольшая частота в данной выборке.

Пример моды выборки (шкала наименований)

Тип темперамента	Холерик	Флегматик	Меланхолик	Сангвиник
Частота	6	3	5	9

Мода выборки – «Сангвиник», так как частота соответствующей варианты равна 9 – наибольшая частота в данной выборке.

Пример моды выборки (смежные варианты)

Самооценка по логике	1	2	3	4	5	6	7
Частота	1	3	5	5	3	2	1

Мода выборки равна 3,5 балла, так как частоты вариант «3» и «4» – наибольшие частоты смежных значений вариант в данной выборке. Мода равна среднему этих значений вариант.

Выборку, в которой только две несмежные варианты имеют наибольшую частоту, называют *бимодальной выборкой*.

Пример бимодальной выборки

Рост, см	165	166	167	168	169	170	17	172	173	174	175
Частота	3	3	4	8	2	5	3	8	2	1	1

Выборка имеет две моды – 168 см и 172 см – равные частоты.

В остальных случаях принято считать, что выборка не имеет моды.

4.2. Среднее

Среднее выборки – статистика, равная отношению суммы всех значений варианты к объему выборки.

Если варианты выборки представить точками некоторого числового отрезка, то среднее будет условным центром отрезка, а все варианты выборки располагаются слева или справа от среднего. *Среднее* – статистика, обозначающая условный центр выборки.

Среднее принято обозначать латинской буквой m (первая буква английского слова *mean*).

Средние выборок, объем которых небольшой, вычисляют по определению. В остальных случаях – с помощью Excel.

Пример среднего небольшой выборки

Вычислим среднее выборки «Самооценки по психологии», объем которой 9 студентов: 6, 8, 5, 6, 6, 7, 7, 8, 6.

m = (6+8+5+6+6+7+7+8+6)/9 = 59/9 = 6, 56 баллов.

Алгоритм вычисления среднего (Excel)

1. В файле с протоколом эмпирических данных создайте новый лист с названием «Статистики».

2. Скопируйте лист «Протокол» и вставьте на лист «Статистики».

3. К протоколу добавьте строку «Среднее».

4. Курсор установите на первую пустую ячейку строки «Среднее».

5. Последовательно выполните операции:

– нажмите клавишу со знаком «f_x»;

– в появившемся окне «Мастер функций» в ячейке «Поиск функции»: наберите СРЗНАЧ. Нажмите кнопку «Найти»;

– в окне «Мастер функций» нажмите «ОК»;

– в появившемся окне «Аргументы функции» нажмите «ОК»;

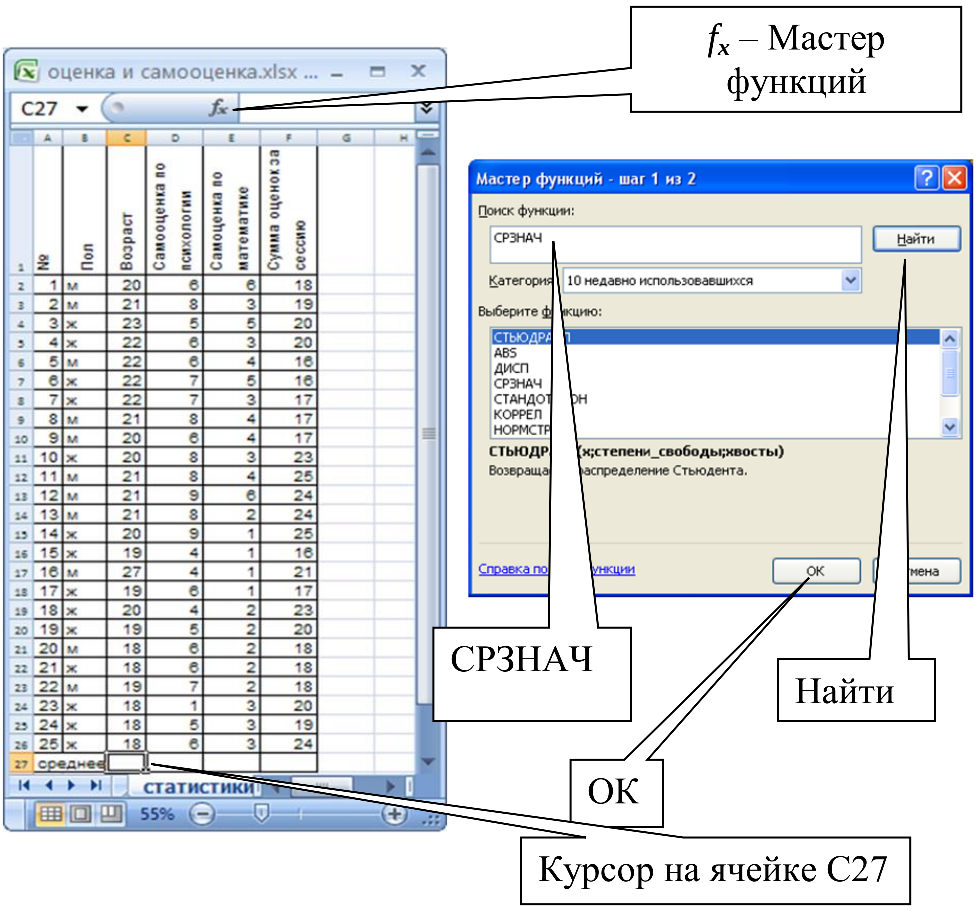
– в ячейке появится значение среднего выборки;

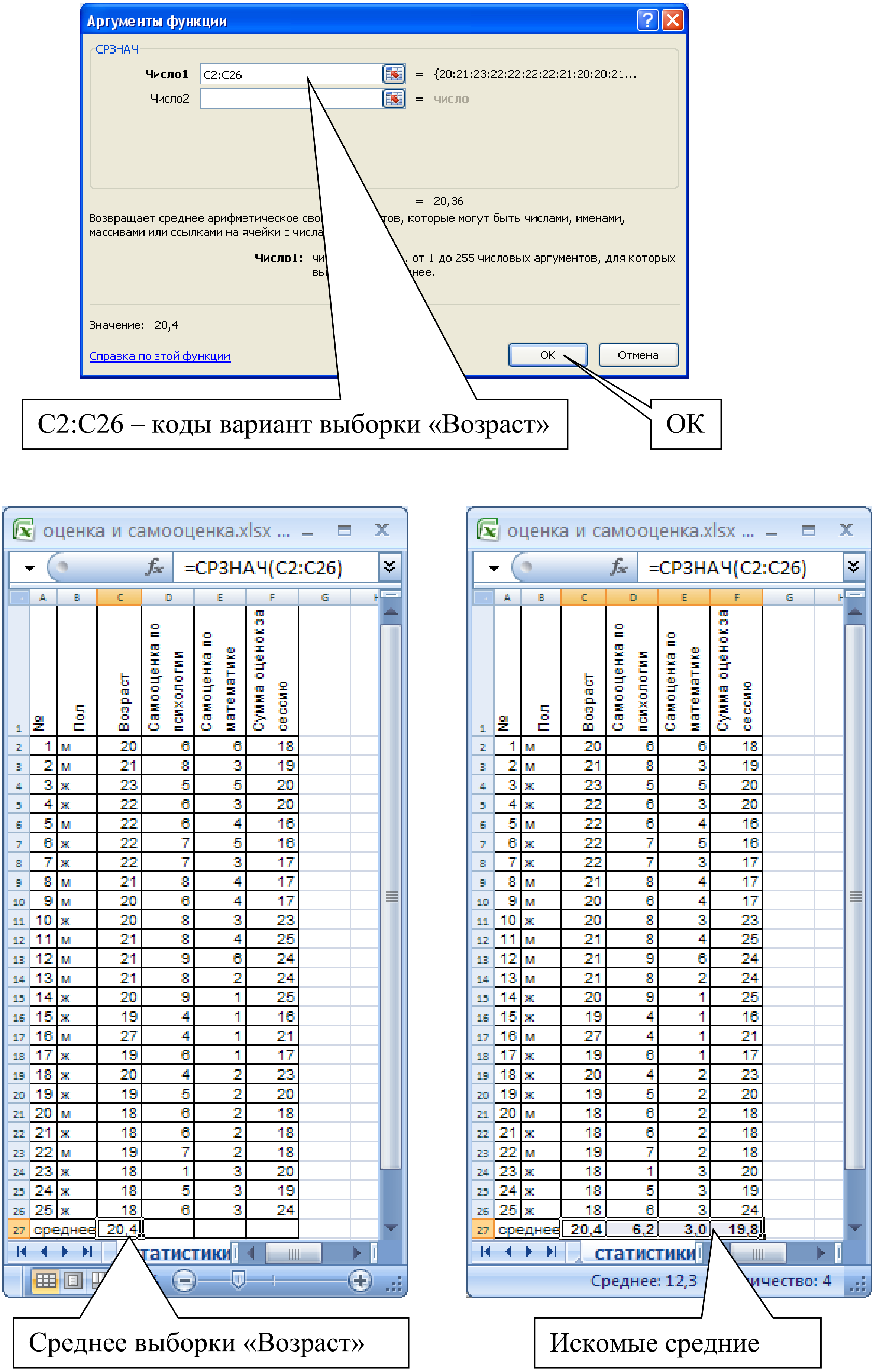
– в строке «f_x» появится запись: = СРЗНАЧ (код первой ячейки выборки : код последней ячейки выборки);

6. Найдите черный квадратик и перетащите его курсором в остальные пустые ячейки строки «Среднее».

Пример вычисления среднего выборок исследования «Оценка – самооценка» (Excel)

Вычислим среднее выборки «Возраст».




Аргументы функции
СРЗНАЧ
Число1 C2:C26 = {20:21:23:22:22:22:22:21:20:20:21…
Число2 = число
= 20,36
Возвращает среднее арифметическое сво… …тов, которые могут быть числами, именами,
массивами или ссылками на ячейки с числа…
Число1: чи… …, от 1 до 255 числовых аргументов, для которых
вы… …нее.
Значение: 20,4
Справка по этой функции
OK
Отмена
C2:C26 – коды вариант выборки «Возраст»
OK
оценка и самооценка.xlsx …
=СРЗНАЧ(C2:C26)
№
Пол
Возраст
Самооценка по психологии
Самооценка по математике
Сумма оценок за сессию
среднее
20,4
татистики
Среднее выборки «Возраст»
оценка и самооценка.xlsx …
=СРЗНАЧ(C2:C26)
№
Пол
Возраст
Самооценка по психологии
Самооценка по математике
Сумма оценок за сессию
среднее
20,4
6,2
3,0
19,8
статистики
Среднее: 12,3
…ичество: 4
Искомые средние

Если составлено распределение частот выборки, то для вычисления выборочного среднего используется формула

$$m = \frac{x_1 n_1 + x_2 n_2 + \ldots + x_k n_k}{n},$$

где m – среднее выборки;

$x_1, x_2, \ldots, x_k$ – варианты выборки;

$n_1, n_2, \ldots, n_k$ – частоты вариант выборки;

n – объем выборки.

Алгоритм вычисления среднего, если составлено распределение частот выборки (Excel)

1. Таблицу распределения частот выборки наберите в Excel и добавьте столбец «Сумма» и строки «$x_i * n_i$», «Среднее».

2. Курсор установите на первую ячейку строки «$x_i * n_i$». Последовательно выполните операции:

– нажмите клавишу со знаком «=»;

– курсор установите на первую ячейку строки «Варианта x_i»;

– нажмите клавишу со знаком «*»;

– курсор установите на первую ячейку строки «Частота n_i»;

– в строке «f_x» появится запись: = (код первой варианты выборки)* (код частоты первой варианты выборки);

– нажмите клавишу со знаком «Enter».

3. Перетащите черный квадратик курсором в остальные пустые ячейки строки «$x_i * n_i$».

4. Курсор установите на первую ячейку столбца «Сумма». Последовательно выполните операции:

– нажмите клавишу со знаком «$\sum$». Автоматически выделятся все ячейки строки «Частота n_i»;

– в строке «f_x» появится запись: = СУММ (код частоты первой варианты выборки : код частоты последней варианты выборки);

– нажмите клавишу со знаком «Enter».

5. Найдите черный квадратик в правом нижнем углу первой ячейки столбца «Сумма» и перетащите его курсором во вторую ячейку столбца «Сумма».

6. Курсор установите на первую ячейку строки «Среднее». Последовательно выполните операции:

– нажмите клавишу со знаком «=»;

– курсор установите на вторую ячейку столбца «Сумма»;

– нажмите клавишу со знаком «/»;

– курсор установите на первую ячейку столбца «Сумма»;

– в строке «f_x» появится запись: = (код второй ячейки столбца «Сумма»)/(код первой ячейки столбца «Сумма»);

– нажмите клавишу со знаком «Enter».

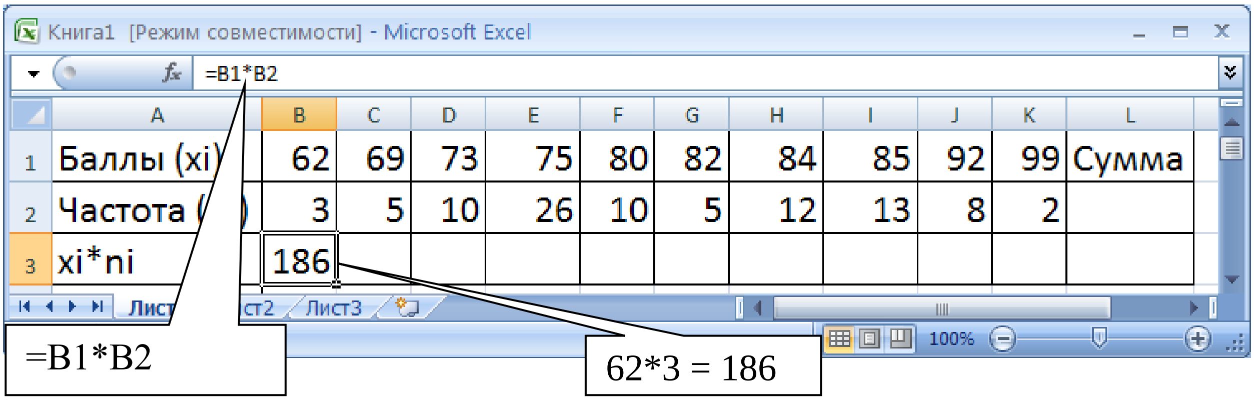
7. В ячейке строки «Среднее» появится искомое значение среднего выборки.

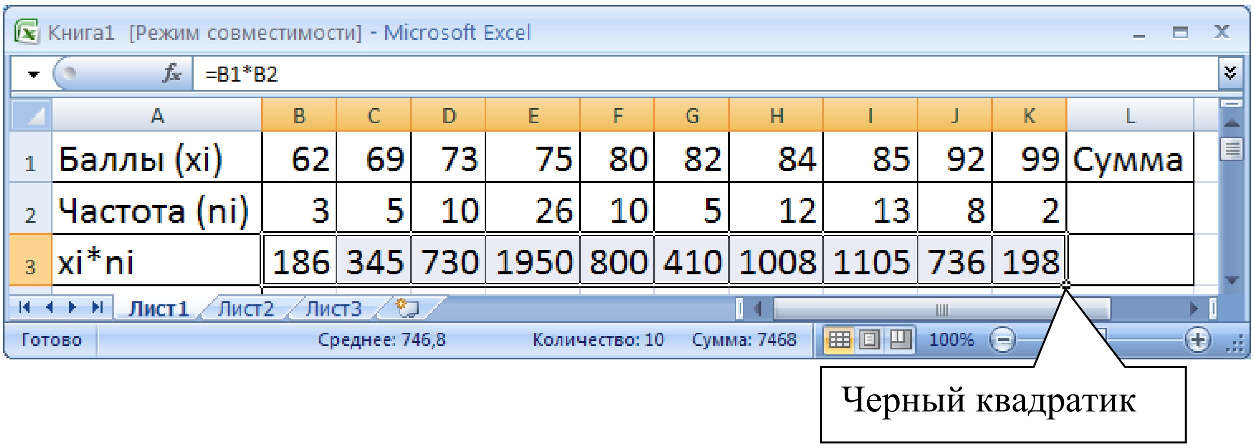
Пример вычисления среднего, если составлено распределение частот выборки «Тест по химии» (Excel)

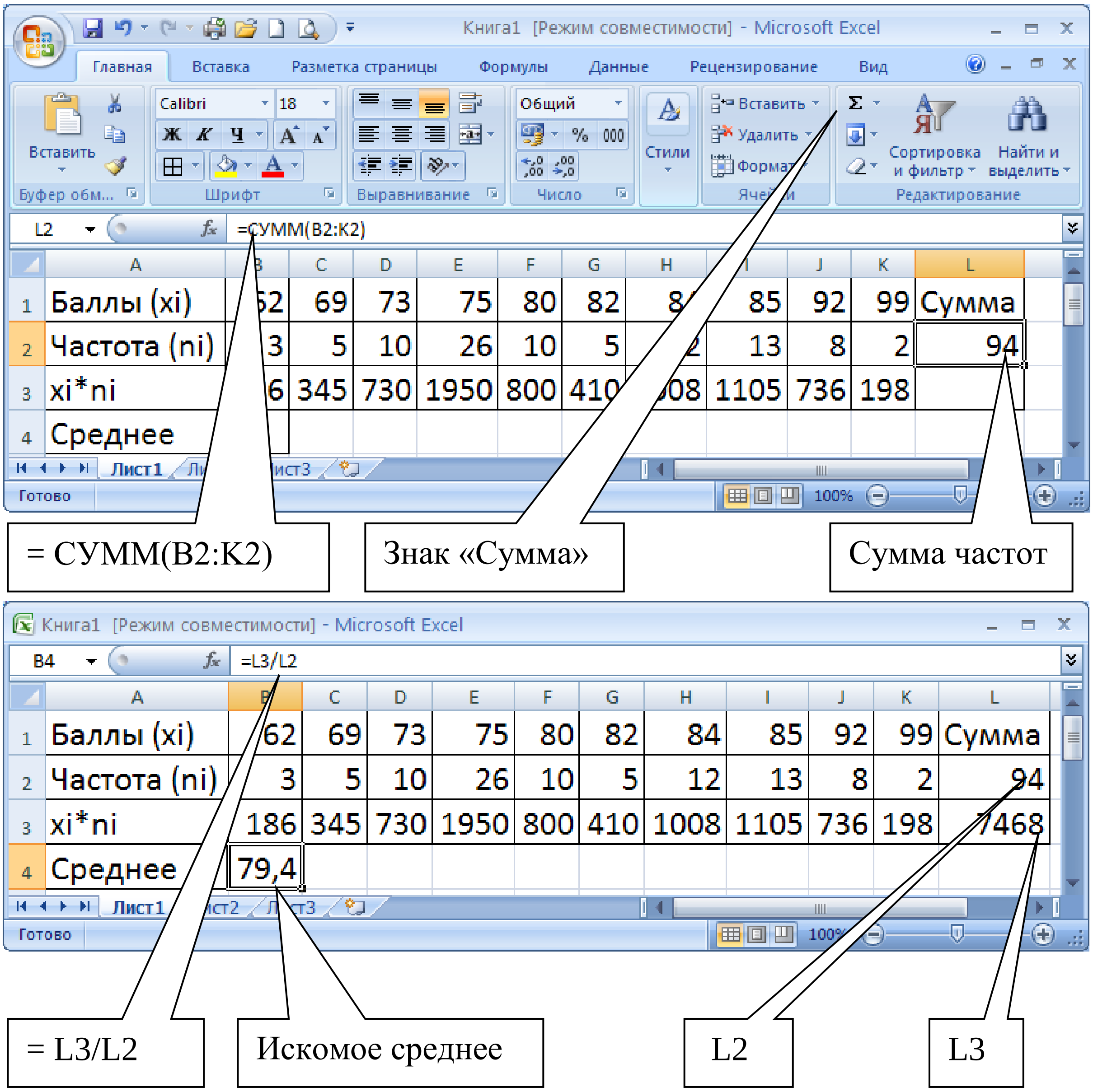
Баллы (x_i)	62	69	73	75	80	82	84	85	92	99
Частота (n_i)	3	5	10	26	10	5	12	13	8	2

Разместим таблицу в Excel.

Добавим столбец «Сумма» и строки «x_i*n_i», «Среднее».







4.3. Стандартное отклонение

Среднее – статистика, обозначающая условный центр выборки. Но одной статистики недостаточно для описания выборки, так как варианты выборки могут находиться на разных расстояниях от центра выборки.

Пример описания выборки с помощью одной статистики (средней)

Рассмотрим выборки:

A: –10, 1, 2, 3, 4, 5, 10.

B: –10, –5, –1, 5, 6, 10, 10.

C: –10, 2,80; 2,99; 3,01; 3,03; 3,17; 10.

Выборки *A, B* и *C* имеют одинаковый объем (7), у них совпадает значение среднего (2,14).

Если варианты выборок *A, B* и *C* представить точками числового отрезка, то они разместятся в отрезок от –10 до 10.

Заметим, что отличительной особенностью выборок *A, B* и *C* являются расстояния вариант от центра выборки (среднего).

Статистика, характеризующая отклонение вариант от среднего, называется *стандартным отклонением* выборки.

Стандартное отклонение – статистика, обозначающая стандартный диапазон изменчивости (рассеяния) вариант выборки от среднего (центра выборки).

Если варианты выборки представить точками некоторого числового отрезка, среднее – условным центром отрезка, то большинство вариант выборки располагаются слева или справа от среднего на величину стандартного отклонения.

Стандартным отклонением выборки (x_i) объемом n со средним m называют число s, равное квадратному корню отношения суммы квадратов отклонений всех значений варианты от выборочного среднего к $n – 1$.

Значение стандартного отклонения находят по формуле

$$s = \sqrt{\frac{(x_1 - m)^2 + (x_2 - m)^2 + \ldots + (x_k - m)^2}{n-1}},$$

где $x_1, x_2, \ldots, x_k$ – варианты выборки;

m – среднее выборки;

n – объем выборки.

Статистика s^2 называется дисперсией выборки.

Пример стандартных отклонений выборок A, B и C

$$sA = \sqrt{\frac{(-10-2{,}1)^2 + (1-2{,}1)^2 + (2-2{,}1)^2 + (3-2{,}1)^2 + (4-2{,}1)^2 + (5-2{,}1)^2 + (10-2{,}1)^2}{7-1}} = 6{,}09.$$

$$sB = \sqrt{\frac{(-10-2{,}1)^2 + (-5-2{,}1)^2 + (-1-2{,}1)^2 + (5-2{,}1)^2 + (6-2{,}1)^2 + (10-2{,}1)^2 + (10-2{,}1)^2}{7-1}} = 7{,}69.$$

$$sC = \sqrt{\frac{(-10-2{,}1)^2 + (2{,}8-2{,}1)^2 + (2{,}99-2{,}1)^2 + (3{,}01-2{,}1)^2 + (3{,}03-2{,}1)^2 + (3{,}17-2{,}1)^2 + (10-2{,}1)^2}{7-1}} = 5{,}96.$$

Видно, что у выборок *A, B* и *C* значения стандартных отклонений различаются.

Заметим, что у выборки *C* большая часть вариант расположена ближе к среднему, чем у выборок *A* и *B*. А у выборки *B* большая часть вариант расположена дальше от среднего, чем у выборок *A* и *C*.

Алгоритм вычисления стандартного отклонения выборки (Excel)

1. В файле с протоколом эмпирических данных на листе с названием «Статистики» добавьте строку «Стандотклон».

2. Курсор установите на первую ячейку строки «Стандотклон».

5. Последовательно выполните операции:

– нажмите клавишу со знаком «f_x»;

– в появившемся окне «Мастер функций» в ячейке «Поиск функции» наберите СТАНДОТКЛОН;

– нажмите кнопку «Найти»;

– в окне «Мастер функций» нажмите «ОК»;

– в появившемся окне «Аргументы функции» выделите курсором все варианты выборки;

– нажмите «ОК»;

– в ячейке появится значение стандартного отклонения выборки;

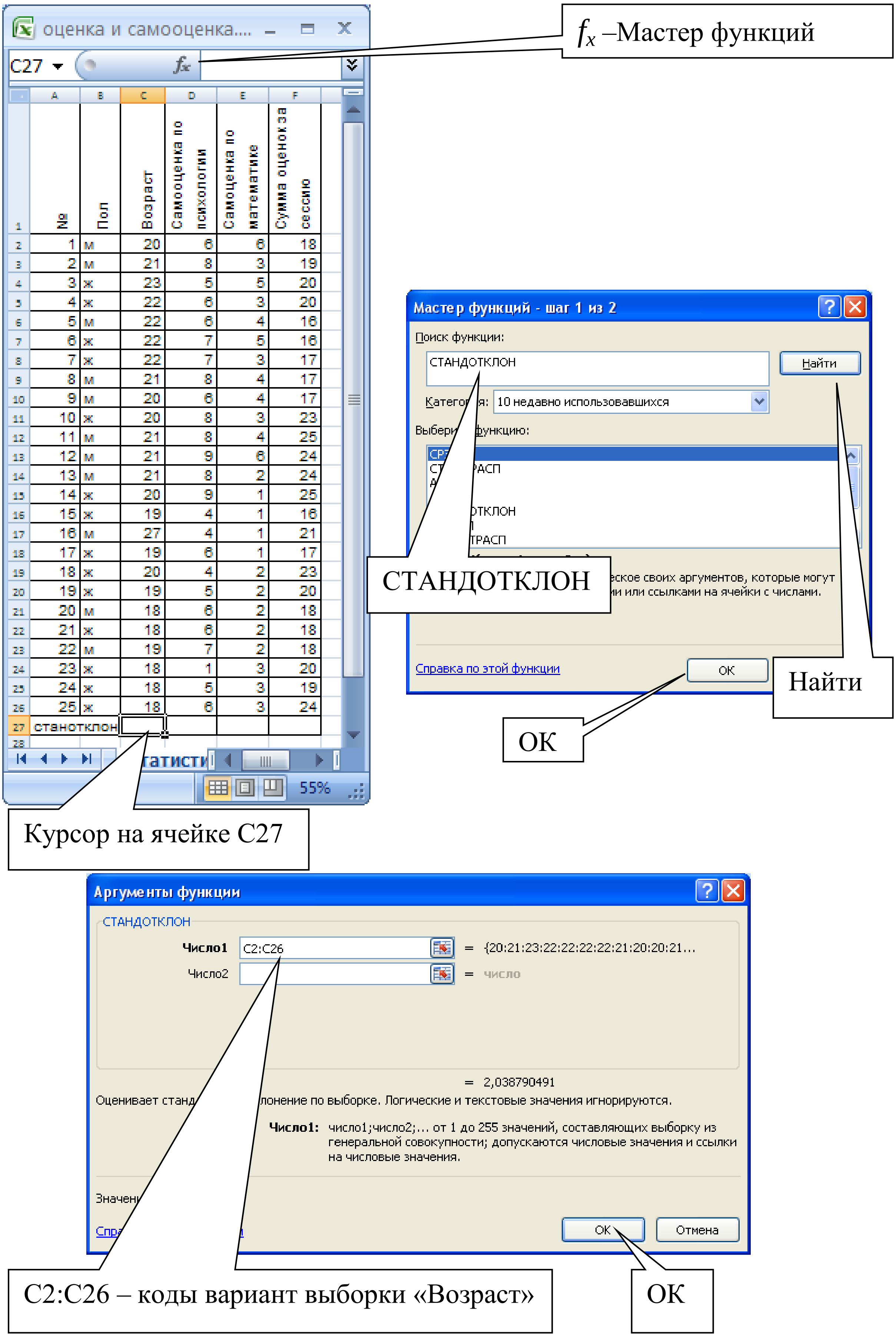
– в строке «f_x» появится запись: = СТАНДОТКЛОН (код первой ячейки выборки : код последней ячейки выборки).

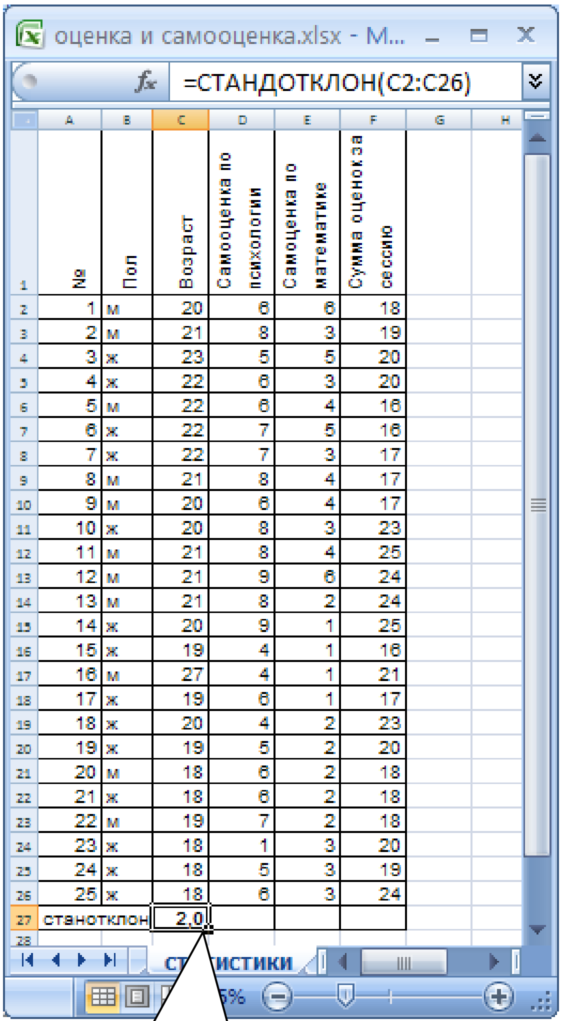
6. Найдите черный квадратик в правом нижнем углу ячейки и перетащите его курсором в остальные пустые ячейки строки «Стандотклон».

7. В ячейке строки «Стандотклон» появится искомое значение стандартного отклонения выборки.

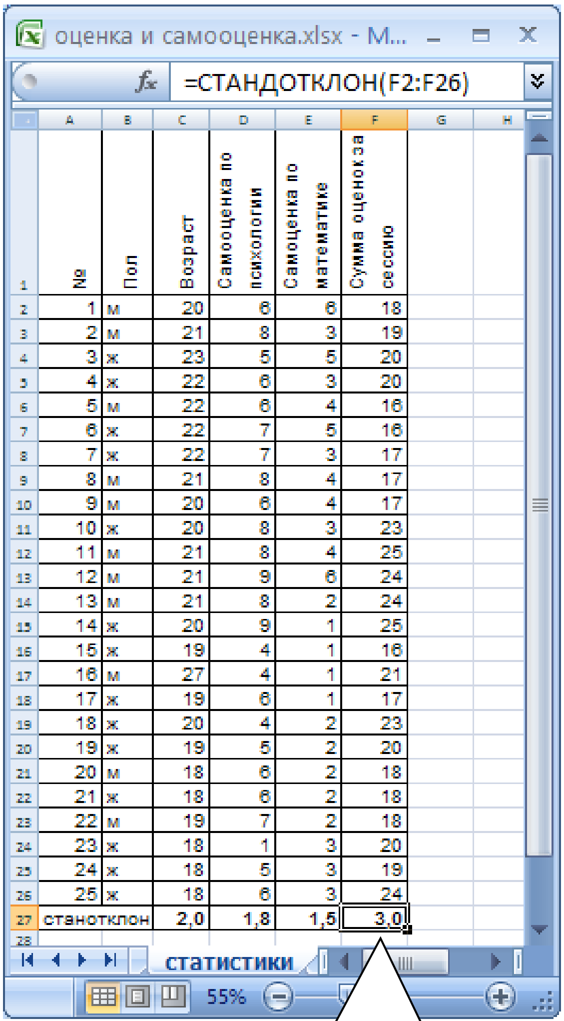
Пример вычисления стандартного отклонения выборок исследования «Оценка – самооценка» (Excel)

Вычислим вначале стандартное отклонение выборки «Возраст».


f_x –Мастер функций
оценка и самооценка....
C27
Мастер функций - шаг 1 из 2
Поиск функции:
СТАНДОТКЛОН
Найти
10 недавно использовавшихся
Справка по этой функции
OK
СТАНДОТКЛОН
Найти
OK
Курсор на ячейке С27
Аргументы функции
СТАНДОТКЛОН
Число1
C2:C26
= {20:21:23:22:22:22:22:21:20:20:21...
Число2
= число
= 2,038790491
Число1: число1;число2;... от 1 до 255 значений, составляющих выборку из генеральной совокупности; допускаются числовые значения и ссылки на числовые значения.
OK
Отмена
С2:С26 – коды вариант выборки «Возраст»
ОК



	A	B	C	D	E	F
1	№	Пол	Возраст	Самооценка по психологии	Самоценка по математике	Сумма оценок за сессию
2	1	м	20	6	6	18
3	2	м	21	8	3	19
4	3	ж	23	5	5	20
5	4	ж	22	6	3	20
6	5	м	22	6	4	16
7	6	ж	22	7	5	16
8	7	ж	22	7	3	17
9	8	м	21	8	4	17
10	9	м	20	6	4	17
11	10	ж	20	8	3	23
12	11	м	21	8	4	25
13	12	м	21	9	6	24
14	13	м	21	8	2	24
15	14	ж	20	9	1	25
16	15	ж	19	4	1	16
17	16	м	27	4	1	21
18	17	ж	19	6	1	17
19	18	ж	20	4	2	23
20	19	ж	19	5	2	20
21	20	м	18	6	2	18
22	21	ж	18	6	2	18
23	22	м	19	7	2	18
24	23	ж	18	1	3	20
25	24	ж	18	5	3	19
26	25	ж	18	6	3	24
27	станотклон		2,0			



	A	B	C	D	E	F
1	№	Пол	Возраст	Самооценка по психологии	Самоценка по математике	Сумма оценок за сессию
2	1	м	20	6	6	18
3	2	м	21	8	3	19
4	3	ж	23	5	5	20
5	4	ж	22	6	3	20
6	5	м	22	6	4	16
7	6	ж	22	7	5	16
8	7	ж	22	7	3	17
9	8	м	21	8	4	17
10	9	м	20	6	4	17
11	10	ж	20	8	3	23
12	11	м	21	8	4	25
13	12	м	21	9	6	24
14	13	м	21	8	2	24
15	14	ж	20	9	1	25
16	15	ж	19	4	1	16
17	16	м	27	4	1	21
18	17	ж	19	6	1	17
19	18	ж	20	4	2	23
20	19	ж	19	5	2	20
21	20	м	18	6	2	18
22	21	ж	18	6	2	18
23	22	м	19	7	2	18
24	23	ж	18	1	3	20
25	24	ж	18	5	3	19
26	25	ж	18	6	3	24
27	станотклон		2,0	1,8	1,5	3,0

Стандартное отклонение выборки «Возраст»

Стандартное отклонение выборки «Сумма оценок за сессию»

Если составлено распределение частот выборки, то для вычисления выборочного *стандартного отклонения* используется формула

$$s = \sqrt{\frac{n \cdot B - A^2}{n(n-1)}},$$

где n – объем выборки;

$B = x_1^2 + x_2^2 + \ldots + x_k^2$;

$A = x_1 + x_2 + \ldots + x_k$.

Алгоритм вычисления стандартного отклонения, если составлено распределение частот выборки (Excel)

1. Таблицу распределения частот выборки наберите в Excel и добавьте столбец «Сумма» и строки «x_i*n_i», «x_i*x_i», «$x_i*x_i*n_i$», «СТАНДОТКЛОН».

2. Курсор установите на первую пустую ячейку строки «x_i*n_i». Последовательно выполните операции:

– нажмите клавишу со знаком «=»;

– курсор установите на первую ячейку строки «Варианта x_i»;

– нажмите клавишу со знаком «*»;

– курсор установите на первую ячейку строки «Частота n_i»;

– в строке «f_x» появится запись: = (код первой варианты выборки)* (код частоты первой варианты выборки);

– нажмите клавишу со знаком «Enter».

3. Перетащите черный квадратик курсором в остальные пустые ячейки строки «x_i*n_i».

4. Курсор установите на первую ячейку строки «x_i*x_i». Последовательно выполните операции:

– нажмите клавишу со знаком «=»;

– курсор установите на первую ячейку строки «Варианта x_i»;

– нажмите клавишу со знаком «*»;

– курсор установите на первую ячейку строки «Варианта x_i»;

– в строке «f_x» появится запись: = (код первой варианты выборки)* (код частоты первой варианты выборки);

– нажмите клавишу со знаком «Enter».

5. Перетащите черный квадратик курсором в остальные пустые ячейки строки «x_i*x_i».

6. Курсор установите на пустую ячейку строки «$x_i*x_i*n_i$». Последовательно выполните операции:

– нажмите клавишу со знаком «=»;

– курсор установите на первую ячейку строки «$x_i* x_i$»;

– нажмите клавишу со знаком «*»;

– курсор установите на первую ячейку строки «Частота n_i»;

– в строке «f_x» появится запись: = (код ячейки «$x_i* x_i$»)* (код частоты первой варианты выборки);

– нажмите клавишу со знаком «Enter».

7. Перетащите черный квадратик курсором в остальные ячейки строки «x_i* x_i*n_i».

8. Курсор установите на первую ячейку столбца «Сумма». Последовательно выполните операции:

– нажмите клавишу со знаком «∑»; Автоматически выделятся все ячейки строки «Частота n_i»;

– в строке «f_x» появится запись: = СУММ (код частоты первой варианты выборки: код частоты последней варианты выборки);

– нажмите клавишу со знаком «Enter».

9. Найдите черный квадратик в правом нижнем углу первой ячейки столбца «Сумма» и перетащите его курсором в пустые ячейки столбца «Сумма».

10. Курсор установите на первую ячейку строки «СТАНДОТКЛОН». Последовательно выполните операции:

– в строке «f_x»: = КОРЕНЬ((L5*L2-L3*L3)/(L2*L2-L2));

– нажмите клавишу со знаком «Enter».

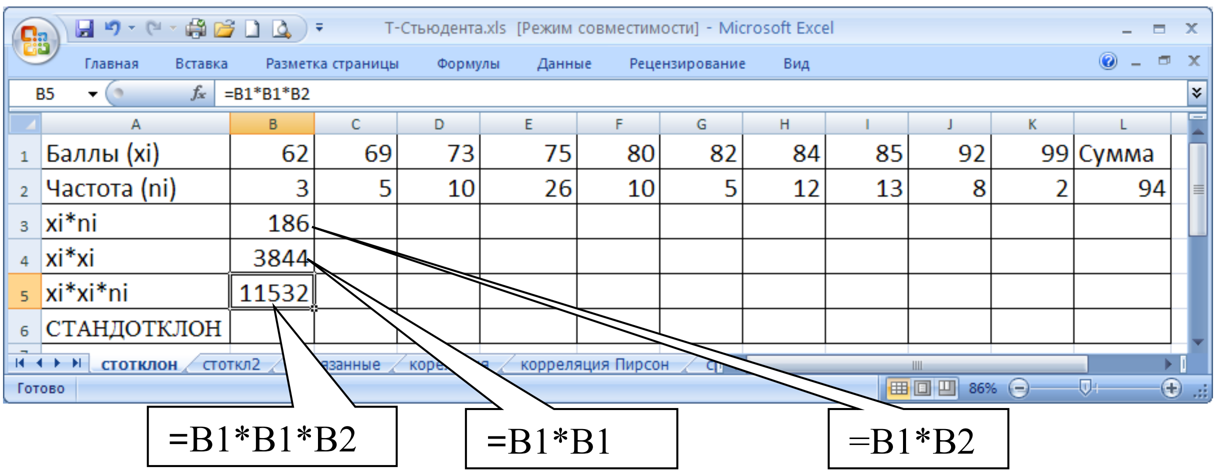
11. В ячейке строки «СТАНДОТКЛОН» появится искомое значение стандартного отклонения выборки.

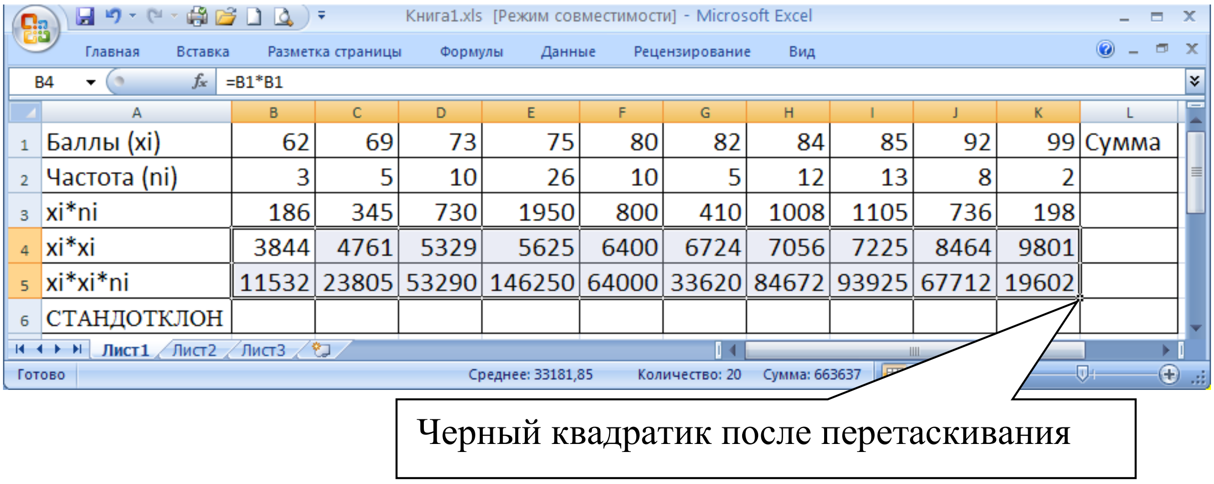
Пример вычисления стандартного отклонения, если составлено распределение частот выборки «Тест по химии» (Excel)

Баллы (x_i)	62	69	73	75	80	82	84	85	92	99
Частота (n_i)	3	5	10	26	10	5	12	13	8	2

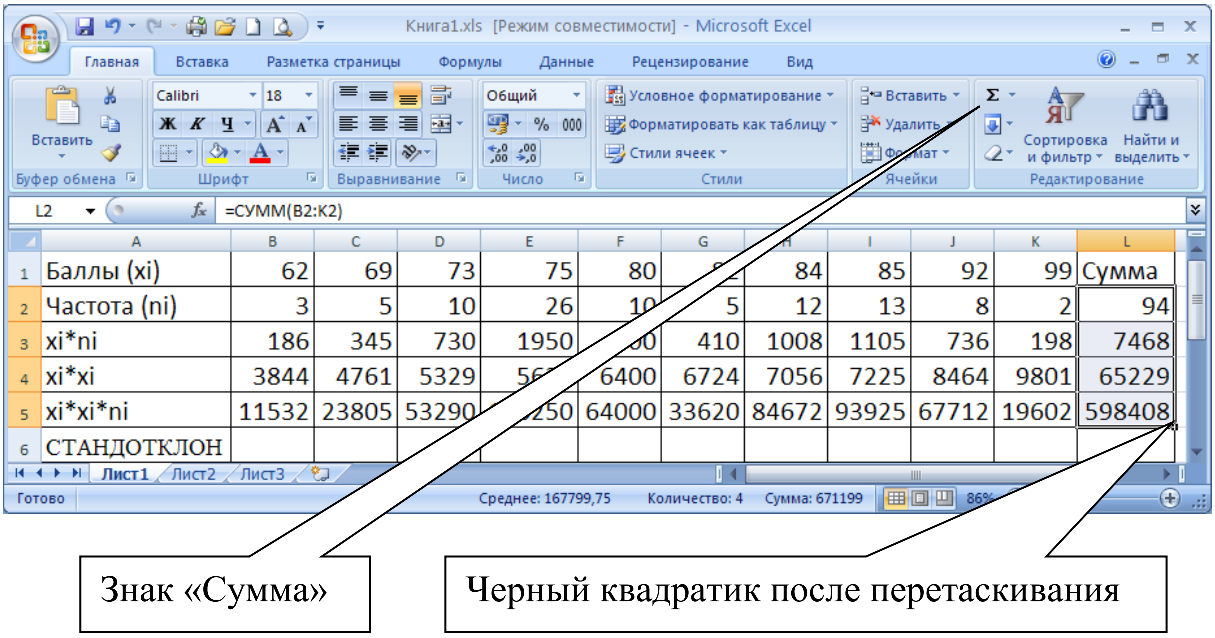
Разместим таблицу в Excel.

Добавим столбец «Сумма» и строки «$x_{i*}n_i$», «x_i^2»; «x_i^2 n_i», «Стандартное отклонение».

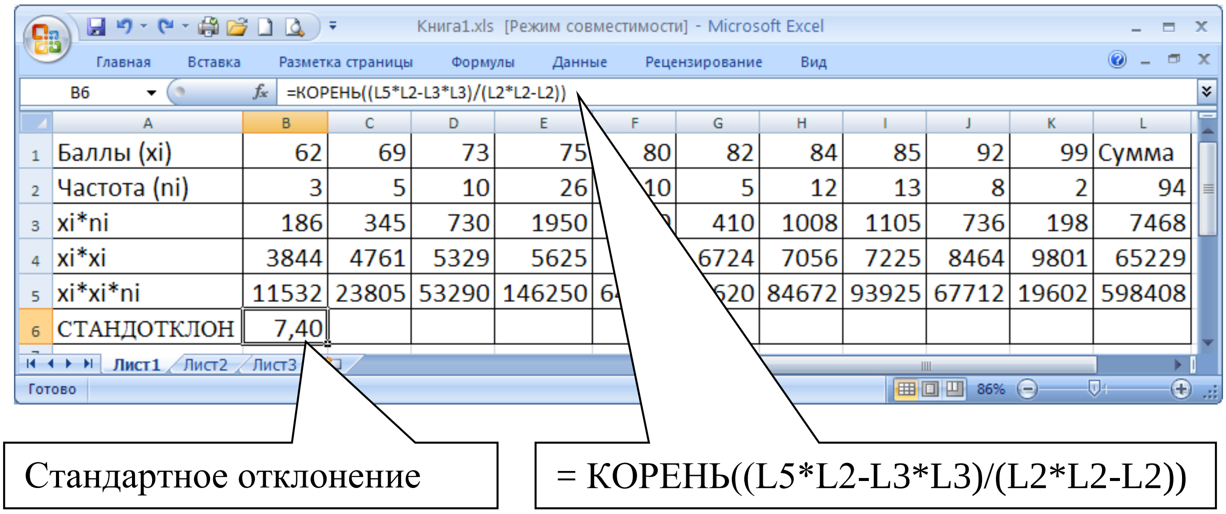




	A	B	C	D	E	F	G	H	I	J	K	L
1	Баллы (xi)	62	69	73	75	80	82	84	85	92	99	Сумма
2	Частота (ni)	3	5	10	26	10	5	12	13	8	2	
3	xi*ni	186	345	730	1950	800	410	1008	1105	736	198	
4	xi*xi	3844	4761	5329	5625	6400	6724	7056	7225	8464	9801	
5	xi*xi*ni	11532	23805	53290	146250	64000	33620	84672	93925	67712	19602	
6	СТАНДОТКЛОН											



	A	B	C	D	E	F	G	H	I	J	K	L
1	Баллы (xi)	62	69	73	75	80	[illegible]	84	85	92	99	Сумма
2	Частота (ni)	3	5	10	26	10	5	12	13	8	2	94
3	xi*ni	186	345	730	1950	[illegible]	410	1008	1105	736	198	7468
4	xi*xi	3844	4761	5329	[illegible]	6400	6724	7056	7225	8464	9801	65229
5	xi*xi*ni	11532	23805	53290	[illegible]	64000	33620	84672	93925	67712	19602	598408
6	СТАНДОТКЛОН											



	A	B	C	D	E	F	G	H	I	J	K	L
1	Баллы (xi)	62	69	73	75	80	82	84	85	92	99	Сумма
2	Частота (ni)	3	5	10	26	10	5	12	13	8	2	94
3	xi*ni	186	345	730	1950	[illegible]	410	1008	1105	736	198	7468
4	xi*xi	3844	4761	5329	5625	[illegible]	6724	7056	7225	8464	9801	65229
5	xi*xi*ni	11532	23805	53290	146250	[illegible]	[illegible]	84672	93925	67712	19602	598408
6	СТАНДОТКЛОН	7,40										

Резюме

К основным описательным статистикам выборки относятся: мода, среднее, стандартное отклонение. Существуют и другие описательные статистики: процентиль, различные квантили распределения, медиана, квартиль, размах выборки и др.

Квантиль – возможное значение варианты выборки, разбивающее выборку в заданном отношении.

Процентилем Pk называется такое значение варианты выборки, левее (меньше) которой находится *k* процентов вариант.

Медианой Md называют 50-й процентиль выборки (*P*50).

Квартилем Kv1 называют 25-й процентиль выборки (*P*25).

Квартилем Kv2 называют 50-й процентиль выборки (*P*50).

Квартилем Kv3 называют 75-й процентиль выборки (*P*75)

Размахом выборки называется разность между максимальным и минимальным ее значениями: $l = x_{max} - x_{min}$.

Возникает вопрос: «Какие описательные статистики пригодны (валидны) для описания выборки?».

Валидность (англ. valid) – действительный, имеющий силу, пригодный.

Валидность статистики – мера (силы) пригодности статистики для описания выборки.

Шкала валидности: отсутствует, низкая, средняя, высокая.

Валидность статистики зависит от объема выборки.

Мода, квантили, размах – статистики с низкой валидностью. Из этих статистик в описании психологических портретов преимущественно используется статистика *Мода* как указание на значение чаще всего встречающегося исследуемого свойства психического явления.

Среднее (m) и *Стандартное отклонение* (s) – статистики со средней валидностью. Для описания исследуемого свойства психического явления принято использовать обе статистики m и s.

Представим все варианты некоторой выборки точками числового отрезка, где *среднее* (m) – условный центр выборки, *стандартное отклонение* (s) – мера рассеяния вариант от центра выборки (мера изменчивости выборки).

Получим графическую модель выборки с основными описательными статистиками.

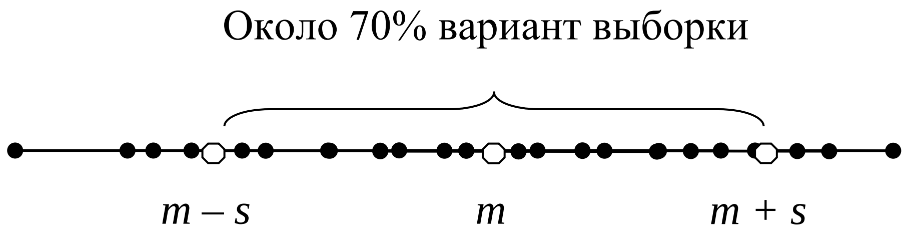


Рис. 3. Модель выборки со средним и стандартным отклонением

Как правило, около 70% вариант выборки попадает в промежуток от $m-s$ до $m+s$.

Стандартное отклонение и среднее часто используют для построения шкал, содержащих различное число градаций.

Упражнения

4.1. Найдите моду выборок по таблицам распределения частот:

Таблица А

Распределение частот исследования темперамента

Тип темперамента	Меланхолик	Сангвиник	Флегматик	Холерик
Частота	11	23	6	21

Таблица Б

Распределение частот исследования «Рост студентов»

Рост, см	162	164	165	168	169	170	176	177	181
Частота	3	6	11	3	4	9	11	3	1

Таблица В

Распределение оценок по политологии

Оценка	Неудовлетворительно	Удовлетворительно	Хорошо	Отлично
Частота	5	5	5	5

4.2. В таблице представлено распределение частот оценок школьников по химии.

Оценка, в баллах	51	58	63	66	68	69	70	71	75	80
Частота	1	3	5	9	10	11	8	7	4	2

Выполните следующие задания:

а) вычислите среднее выборки;

б) вычислите стандартное отклонение выборки;

в) найдите значения квартилей выборки.

4.3. В протоколе представлены результаты тестирования студентов по психологии в баллах.

Имя	Балл	Имя	Балл	Имя	Балл	Имя	Балл	Имя	Балл	Имя	Балл
АА	51	ДА	66	КА	68	ОА	69	РА	70	ТА	71
АВ	58	ДВ	66	КВ	68	ОВ	69	РВ	70	ТВ	71
БВ	58	ЕВ	66	КВ	68	ПВ	69	СВ	70	ФВ	71
БН	58	ЕН	66	КН	68	ПН	69	СН	70	ФН	71
ВИ	63	ЖИ	66	ЛИ	68	ПИ	69	СИ	70	ХИ	75
ВР	63	ЖР	66	МР	68	ПР	69	СР	70	ХР	75
ВТ	63	ЗТ	66	МТ	68	ПТ	69	СТ	70	ЦТ	75
ГП	63	ЗП	66	МП	68	ПП	69	СП	71	ЧП	75
ГН	63	ИН	68	НН	69	РН	69	ТН	71	ШН	80
ГЛ	66	КЛ	68	НЛ	69	РЛ	70	ТЛ	71	ЯЛ	80

Выполните следующие задания:

а) найдите моду выборки;

б) найдите медиану выборки;

в) вычислите среднее выборки;

г) найдите стандартное отклонение выборки.

5. СТАТИСТИЧЕСКИЙ ВЫВОД

Задача психолога – выявить отличия одного человека от другого, одной выборки от другой.

Методом решения этой задачи является сбор эмпирических данных не менее чем в двух выборках, а затем сравнение полученных показателей. Сравнение позволяет сделать вывод о сходстве или различии выборок.

Для сравнения выборок используются статистические критерии, позволяющие получить вывод о статистической значимости различия выборок. Вывод имеет вероятностный характер.

5.1. Глоссарий статистического вывода

Связанные и несвязанные выборки (зависимые и независимые)

Связанные выборки – множества значений *двух свойств*, полученные в одной группе респондентов.

Несвязанные выборки – множества значений одного свойства, полученные в *двух группах* респондентов.

Пример связанных и несвязанных выборок

№	Форма обучения	«Мотивация»	«Успешность»	«Общительность»
1	Очная	7	3	3
2	Очная	4	12	3
3	Очная	1	14	2
4	Очная	4	9	1
5	Очная	5	13	6
6	Очная	3	34	6
7	Заочная	9	23	7
8	Заочная	4	19	8
9	Заочная	4	18	9
10	Заочная	3	11	3
11	Заочная	5	23	4
12	Заочная	6	23	3
13	Заочная	6	23	3
14	Заочная	2	11	5

1. Выборки «Мотивация» и «Успешность» для всех студентов являются связанными.

2. Выборка «Общительность» студентов очной формы обучения и выборка «Общительность» студентов заочной формы обучения являются несвязанными.

Статистические гипотезы

Статистическими гипотезами называют *предположения* о статистически значимых различиях выборок.

На вопрос «Значимо или не значимо отличаются выборки?» может быть два варианта ответа:

– выборки отличаются не значимо;

– выборки отличаются значимо.

Следовательно, всегда можно сформулировать две гипотезы:

1) выборки статистически значимо не различаются;

2) выборки статистически значимо различаются.

Первую гипотезу называют *нулевой* гипотезой.

Обозначение – H_0 (различий нет).

Вторую гипотезу называют *альтернативной* гипотезой.

Обозначение – H_1 (различия есть).

Пример статистических гипотез

H_0: выборки «Мотивация» у студентов очной и заочной форм обучения статистически значимо не различаются.

H_1: выборки «Мотивация» у студентов очной и заочной форм обучения статистически значимо различаются.

Алгоритм проверки статистических гипотез

1. Проверяется гипотеза H_0.
2. Если H_0 принимается, то H_1 не рассматривается.
3. Если H_0 не принимается, тогда принимается H_1.

Принятие решения о H_1 имеет вероятностный характер, потому указывается уровень значимости принятия правильного решения о H_1 (вероятность вывода).

Пример вероятности подтверждения гипотезы H_1

H_1: «Самооценки студентов по психологии и самооценки по математике статистически значимо ($p \leq \alpha$) различаются»,

где p – вероятность;
α – уровень статистической значимости.

Вероятность вывода

Гипотеза H_0 принимается, когда вероятность различия выборок не высока (менее 0,90).

Гипотеза H_0 не принимается тогда, когда вероятность отличия выборок высока (более 0,90; 0,95, 0,99 и т.д.).

Пусть p – вероятность принятия правильного решения о гипотезе H_0, тогда вероятность ошибки равна $1 - p$.

Вероятность ошибки принятия гипотезы H_0 называется *уровнем статистической значимости*.

Уровень статистической значимости принято обозначать α.

В психологии различают следующую шкалу уровней статистической значимости:

– $p > 0{,}10$ (статистически незначимый);
– $0{,}05 < p \leq 0{,}10$ (невысокий – тенденция);
– $0{,}01 < p \leq 0{,}05$ (нормальный);
– $p \leq 0{,}01$(высокий).

Запись $p \leq \alpha$ означает, что вероятность ошибки непринятия H_0 меньше α. То есть уровень статистической значимости принятия H_1 равен α.

Пример записи подтверждения гипотезы H_1

H_1: «Самооценки студентов по психологии и самооценки по математике статистически значимо ($p \leq 0{,}05$) различаются».

Для проверки статистических гипотез используют статистические критерии.

Критерии проверки статистических гипотез

Статистические критерии представлены в форме алгоритма проверки статистических гипотез и содержат таблицы критических значений случайной величины. Критерии имеют названия, как правило, связанные с именами авторов.

Пример названия статистических критериев

– φ-критерий Фишера;
– λ-критерий Колмогорова-Смирнова;

– G-критерий знаков;
– U-критерий Манна-Уитни.

Статистические критерии, зависящие от объема выборок и уровня статистической значимости, опубликованы в пособиях или размещены в компьютерных программах.

Алгоритм проверки статистических гипотез

1. Выбирается критерий сравнения.
2. Вычисляется статистика для сравниваемых выборок по правилу, соответствующему критерию (C).
3. Находится предельное значение статистики ($C\alpha$) для установленного исследователем уровня значимости α.
4. Сравниваются значения C и $C\alpha$. Исходя из того, какое значение больше, делается вывод о том, принимается H_0 или принимается H_1.

5.2. φ-КРИТЕРИЙ ФИШЕРА

С помощью φ-критерия Фишера устанавливается значимость различия долей выраженности одинакового свойства в двух выборках или двух разных свойств в одной выборке.

Под долей выраженности свойства понимается отношение числа респондентов, имеющих это психическое свойство, к объему выборки.

Вопросы, ответы на которые можно найти с помощью φ-критерия Фишера

1. Есть ли статистически значимые различия долей студентов одной группы от студентов другой группы, имеющих высокий уровень общительности, если доля студентов с высоким уровнем общительности в одной группе равна 0,25, а в другой – 0,44?

2. Можно ли считать статистически значимыми различия высоких самооценок по психологии и высоких самооценок по математике у студентов, если высокие самооценки по психологии у 80% студентов, а высокие самооценки по математике у 55% студентов?

Особенности применения φ-критерия Фишера

1. В каждой из сравниваемых выборок должно быть не менее пяти респондентов.

2. Выборки могут быть связанными или несвязанными.

3. Используется таблица φ-критерия Фишера (замены долей выраженности исследуемого свойства на φ_1 и φ_2).

Таблица φ-критерия Фишера

0,а	*,00*	*,01*	*,02*	*,03*	*,04*	*,05*	*,06*	*,07*	*,08*	*,09*
,00	0,000	0,200	0,284	0,348	0,403	0,451	0,495	0,536	0,574	0,609
,10	0,644	0,676	0,707	0,738	0,767	0,795	0,823	0,850	0,876	0,902
,20	0,927	0,952	0,976	1,000	1,024	1,047	1,070	1,093	1,115	1,137
,30	1,159	1,182	1,203	1,224	1,245	1,266	1,287	1,308	1,328	1,349
,40	1,369	1,390	1,410	1,430	1,451	1,471	1,491	1,511	1,531	1,551
,50	1,571	1,591	1,611	1,631	1,651	1,671	1,691	1,711	1,731	1,752
,60	1,772	1,793	1,813	1,834	1,855	1,875	1,897	1,918	1,939	1,961
,70	1,982	2,004	2,026	2,049	2,071	2,094	2,118	2,141	2,165	2,190
,80	2,214	2,240	2,265	2,292	2,319	2,246	2,375	2,404	2,434	2,465
,90	2,498	2,532	2,568	2,606	2,647	2,691	2,739	2,793	2,858	2,941

Пример замены долей выраженности свойства на φ_1 и φ_2

Допустим, что студентов с высоким уровнем общительности в одной группе 0,25 от всей выборки (25%), а в другой группе – 0,44 (44%).

Найдем соответствующие значения φ_1 и φ_2.

0,а	*,00*	*,01*	*,02*	*,03*	*,04*	*,05*	*,06*	*,07*	*,08*	*,09*
,00	0,000	0,200	0,284	0,348	0,403	0,451	0,495	0,536	0,574	0,609
,10	0,644	0,676	0,707	0,738	0,767	0,795	0,823	0,850	0,876	0,902
,20	0,927	0,952	0,976	1,000	1,024	1,047	1,070	1,093	1,115	1,137
,30	1,159	1,182	1,203	1,224	1,245	1,266	1,287	1,308	1,328	1,349
,40	1,369	1,390	1,410	1,430	1,451	1,471	1,491	1,511	1,531	1,551
,50	1,571	1,591	1,611	1,631	1,651	1,671	1,691	1,711	1,731	1,752
,60	1,772	1,793	1,813	1,834	1,855	1,875	1,897	1,918	1,939	1,961
,70	1,982	2,004	2,026	2,049	2,071	2,094	2,118	2,141	2,165	2,190
,80	2,214	2,240	2,265	2,292	2,319	2,246	2,375	2,404	2,434	2,465
,90	2,498	2,532	2,568	2,606	2,647	2,691	2,739	2,793	2,858	2,941

$\varphi_2 = 1,451$

$\varphi_1 = 1,047$

Искомые $\varphi_1 = 1{,}047$, $\varphi_2 = 1{,}451$.

Алгоритм φ-критерия Фишера

1. Вычисленные доли (проценты) выраженности одинакового свойства в I и II выборках заменяют на соответствующие им значения φ_1 и φ_2 с помощью таблицы φ-критерия Фишера.

2. Вычисляют значение φ по формуле

$$\varphi = |\varphi_1 - \varphi_2|\sqrt{\frac{n_1 n_2}{n_1 + n_2}},$$

где φ_1 – табличное значение доли выраженности свойства в выборке I;

φ_2 – табличное значение доли выраженности свойства в выборке II;

n_1 – объем выборки I;

n_2 – объем выборки II.

3. Статистический вывод.

Если $\varphi < 1{,}29$, то принимается гипотеза H_0.

Если $1{,}29 \le \varphi < 1{,}64$, то принимается гипотеза H_1 ($p \le 0{,}10$).

Если $1{,}64 \le \varphi < 2{,}31$, то принимается гипотеза H_1 ($p \le 0{,}05$).

Если $2{,}31 \le \varphi$, то принимается гипотеза H_1 ($p \le 0{,}01$).

Пример использования φ-критерия Фишера

В таблице приведены распределения частот самооценок студентов по психологии и по математике.

Самооценка в баллах	1	2	3	4	5	6	7	8	9
Частота по психологии	1	0	0	3	5	5	7	6	0
Частота по математике	3	5	4	1	3	5	3	2	1

Требуется установить уровень статистической значимости различий высоких самооценок по психологии и высоких самооценок по математике у студентов.

Высокими считаются самооценки не менее семи баллов.

Вычислим объем выборки – сумму частот самооценок по одной дисциплине, например, по математике.

$3 + 5 + 4 + 1 + 3 + 5 + 3 + 2 + 1 = 27$.

Вычислим частоту высоких самооценок (не менее 7 баллов):

– по психологии – 13;

– по математике – 6.

Вычислим доли (проценты) высоких самооценок:

– по психологии – 13:27 = 0,48 (48%);

– по математике – 6:27 = 0,22 (22%).

Заменим вычисленные доли выраженности свойства в первой и во второй выборках на соответствующие им значения φ_1 и φ_2 с помощью таблицы φ-критерия Фишера.

0,a	*,00*	*,01*	*,02*	*,03*	*,04*	*,05*	*,06*	*,07*	*,08*	*,09*
,00	0,000	0,200	0,284	0,348	0,403	0,451	0,495	0,536	0,574	0,609
,10	0,644	0,676	0,707	0,738	0,767	0,795	0,823	0,850	0,876	0,902
,20	0,927	0,952	0,976	1,000	1,024	1,047	1,070	1,093	1,115	1,137
,30	1,159	1,182	1,203	1,224	1,245	1,266	1,287	1,308	1,328	1,349
,40	1,369	1,390	1,410	1,430	1,451	1,471	1,491	1,511	1,531	1,551
,50	1,571	1,591	1,611	1,631	1,651	1,671	1,691	1,711	1,731	1,752
,60	1,772	1,793	1,813	1,834	1,855	1,875	1,897	1,918	1,939	1,961
,70	1,982	2,004	2,026	2,049	2,071	2,094	2,118	2,141	2,165	2,190
,80	2,214	2,240	2,265	2,292	2,319	2,246	2,375	2,404	2,434	2,465
,90	2,498	2,53[illegible]	2,568	2,606	2,647	2,691	2,739	2,79[illegible]	2,858	2,941

φ_2 = 0,976

φ_1 = 1,531

Искомые φ_1 = 1,531, φ_2 = 0,976.

Вычислим $\varphi = \left|1{,}531 - 0{,}976\right| \cdot \sqrt{(27 \cdot 27)/(27 + 27)} = 2{,}04$.

Статистический вывод.

Так как $1{,}64 \le \varphi < 2{,}31$, то принимается гипотеза H_1 ($p \le 0{,}05$).

Содержательный вывод.

У студентов процент высоких самооценок по психологии (48%) статистически значимо ($p \le 0{,}05$) *больше* процента высоких самооценок по математике (22%).

Алгоритм φ-критерия Фишера (*Excel*)

1. Пусть исследуемое свойство в первой выборке отмечено у m_1 респондентов, во второй выборке – у m_2 респондентов. Обозначим n_1 – объем первой выборки, n_2 – объем второй выборки.

2. Строкам столбца **A** таблицы *Excel* присваиваются названия: «*m*1», «*m*2», «*n*1», «*n*2», «*a*1», «*a*2», «фи1», «фи2», «ФИ», «Н».

3. В строки столбца **B** таблицы *Excel* заносятся численные значения «*m*1», «*m*2», «*n*1», «*n*2», соответствующие строкам столбца A1 таблицы *Excel*.

4. В строках столбца **B** таблицы *Excel* («a1», «a2», «фи1», «фи2», «ФИ») проводятся вычисления по формулам:

$a1 = m_1/n_1$;
$a2 = m_2/n_2$;
фи1 =2*ASIN(КОРЕНЬ(*a*1));
фи2 =2*ASIN(КОРЕНЬ(*a*2));
ФИ =ABS(фи1 – фи2)*КОРЕНЬ(n_1*n_2 /(n_1 + n_2)).

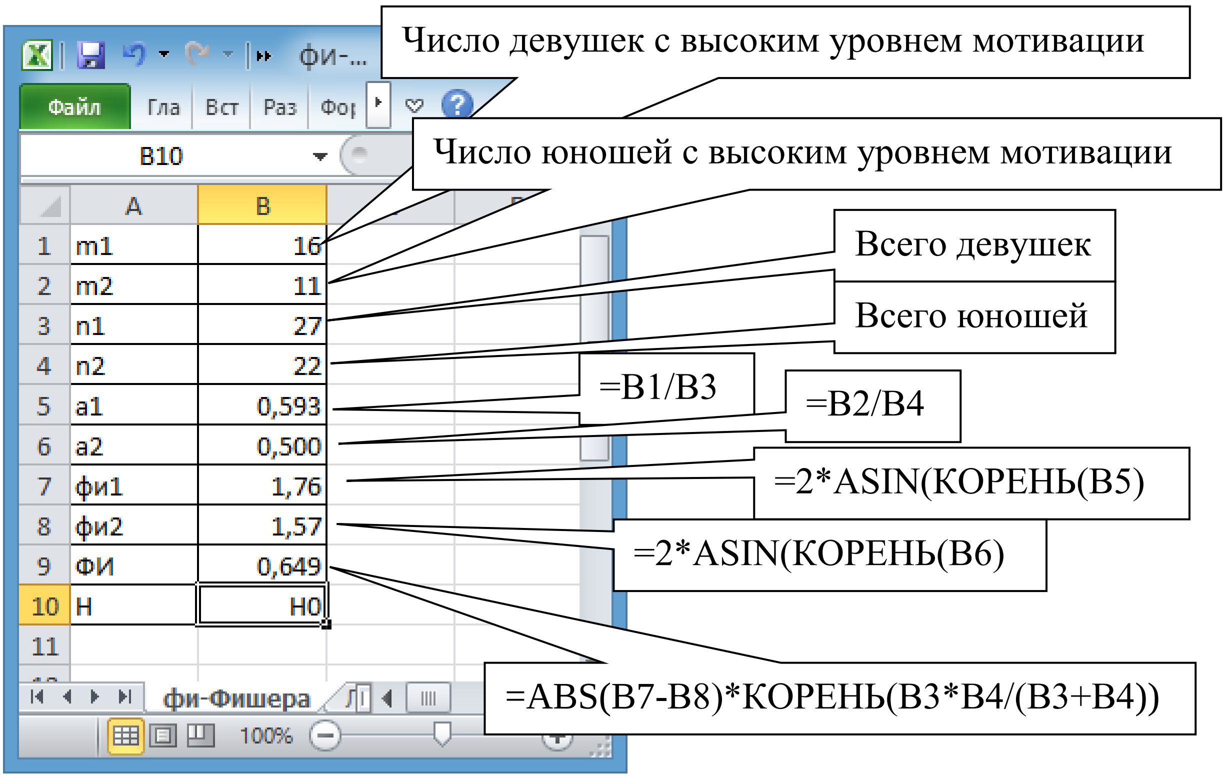
5. Вычисленное значение **ФИ** является основанием для статистического вывода.

Если ФИ < 1,29, то принимается гипотеза H_0.
Если 1,29 ≤ ФИ < 1,64, то принимается гипотеза H_1 ($p \leq 0{,}10$).
Если 1,64 ≤ ФИ < 2,31, то принимается гипотеза H_1 ($p \leq 0{,}05$).
Если 2,31 ≤ ФИ, то принимается гипотеза H_1 ($p \leq 0{,}01$).

Пример использования φ-критерия Фишера (Excel)

У 27 девушек и 22 юношей измерили уровень мотивации к избеганию неудач. Высокий уровень выявлен у 16 девушек и у 11 юношей.

Есть ли статистически значимые различия долей девушек от юношей, имеющих высокий уровень мотивации к избеганию неудач?



Статистический вывод.

Так как ФИ < 1,29, то принимается гипотеза H_0.

Содержательный вывод.

Нет статистически значимых отличий процентов девушек (59%) и юношей (50%) с высоким уровнем мотивации к избеганию неудач.

5.3. λ-КРИТЕРИЙ КОЛМОГОРОВА-СМИРНОВА

С помощью λ-критерия Колмогорова-Смирнова устанавливается уровень статистической значимости различий *распределений частот* одинакового свойства в двух выборках или двух разных свойств в одной выборке.

Вопросы, ответы на которые можно найти с помощью λ-критерия Колмогорова-Смирнова

1. Есть ли статистически значимые различия распределений частот показателей общительности студентов одной группы от студентов другой группы?
2. Можно ли считать статистически значимыми различия распределений частот самооценок по психологии и самооценок по математике у студентов?

Особенности применения λ-критерия Колмогорова-Смирнова

1. В каждой из сравниваемых выборок должно быть не менее пятидесяти респондентов.
2. Выборки могут быть связанными или несвязанными.

Алгоритм λ-критерия Колмогорова-Смирнова

1. Составляют процентильные распределения частот исследуемого свойства для I и II выборок в общей таблице.
2. К таблице добавляют строку «|PCUM I – PCUM II|» и заполняют ее.
3. Вычисляют значение λ по формуле

$$\lambda = d\sqrt{\frac{n_1 n_2}{n_1 + n_2}},$$

где d – наибольшее число из строки «|PCUM I – PCUM II|»;

n_1 – число респондентов в выборке I;

n_2 – число респондентов в выборке II.

4. Статистический вывод.

Если $\lambda < 1{,}22$, то принимается гипотеза H_0.

Если $1{,}22 \leq \lambda < 1{,}36$, то принимается гипотеза H_1 ($p \leq 0{,}10$).

Если $1{,}36 \leq \lambda < 1{,}63$, то принимается гипотеза H_1 ($p \leq 0{,}05$).

Если $1{,}63 \leq \lambda$, то принимается гипотеза H_1 ($p \leq 0{,}01$).

Пример использования λ-критерия Колмогорова-Смирнова

В таблице приведены распределения частот рангов ценности «Здоровье» у студентов экспериментальной (I) и контрольной (II) групп.

Ранг ценности «Здоровье»	1	2	3	4	5
Частота I выборки	81	45	27	19	18
Частота II выборки	23	51	2	5	12

Требуется установить уровень статистической значимости различий распределений частот рангов ценности «Здоровье» у студентов экспериментальной и контрольной групп.

Последовательно составим процентильные распределения частот исследуемого свойства для I и II выборок в общей таблице.

Добавим строку «модуль PCUM I – PCUM II» и заполним ее.

Ранг ценности «Здоровье»	1	2	3	4	5
Частота I выборки	81	45	27	19	18
Частота II выборки	23	51	2	5	12
CUM I	81	126	153	172	190
CUM II	23	74	76	81	93
PCUM I	0,426	0,663	0,805	0,905	1,000
PCUM II	0,247	0,796	0,817	0,871	1,000
|PCUM I – PCUM II|	**0,179**	0,133	0,012	0,034	0,000

$d = 0{,}179$

$n_1 = 190$

$n_2 = 93$

Вычислим значение λ:

$$\lambda = d\sqrt{\frac{n_1 n_2}{n_1 + n_2}} = 0{,}179\sqrt{\frac{190 \cdot 93}{190 + 93}} = 0{,}179 \cdot 7{,}90 = 1{,}41.$$

Статистический вывод.

Так как $1{,}36 \leq \lambda < 1{,}63$, то принимается гипотеза H_1 ($p \leq 0{,}05$).

Содержательный вывод.

Выявлены статистически значимые различия ($p \leq 0{,}05$) распределений рангов ценности «Здоровье» у студентов экспериментальной и контрольной групп.

5.4. *G*-КРИТЕРИЙ ЗНАКОВ

С помощью *G*-критерия знаков устанавливается уровень статистической значимости различий свойства *A* и свойства *B* у респондентов одной выборки.

Вопросы, ответы на которые можно найти с помощью G-критерия знаков

1. Есть ли статистически значимые различия показателей общительности (*A*) и академической успеваемости (*B*) у студентов исследуемой группы?
2. Есть ли статистически значимые различия показателей агрессивности у студентов до экспериментального воздействия (*A*) и после него (*B*)?

Особенности применения G-критерия знаков

1. В выборке должно быть не менее пяти респондентов.
2. Выборки свойств *A* и *B* должны быть связанными.
3. Свойства *A* и *B* должны быть измерены в одной шкале или ранжированы.
4. Используется таблица *G*-критерия знаков (значения рассчитаны в *Excel*).

Алгоритм *G*-критерия знаков

1. К протоколу свойств *A* и *B*, *измеренных в одной шкале или ранжированных*, добавляют столбец «Знак A – B» и заполняют его.

2. Вводят обозначения:

a – число «плюсов» в столбце «Знак A – B»;

b – число «минусов» в столбце «Знак A – B»;

n – сумма *a* и *b*;

G – число, равное меньшему из чисел *a* и *b*.

3. Находят в таблице G-критерия знаков, в строке n соответствующие значения $G0{,}10$, $G0{,}05$ и $G0{,}01$.

4. Статистический вывод.

Если $G > G0{,}10$, то принимается гипотеза H_0.

Если $G0{,}05 < G \le G0{,}10$, то принимается гипотеза H_1 ($p \le 0{,}10$).

Если $G0{,}01 < G \le G0{,}05$, то принимается гипотеза H_1($p \le 0{,}05$).

Если $G \le G0{,}01$, то принимается гипотеза H_1 ($p \le 0{,}01$).

В случае принятия гипотезы H_1 можно установить направленность изменений измеренного свойства по количеству большего из чисел a или b.

Таблица *G*-критерия знаков

n	α			**n**	α			**n**	α			**n**	α		
	0,10	0,05	0,01		0,10	0,05	0,01		0,10	0,05	0,01		0,10	0,05	0,01
5	1	0	0	**27**	9	8	7	**49**	19	18	16	**92**	38	37	34
6	1	1	0	**28**	10	9	7	**50**	19	18	16	**94**	39	38	35
7	1	1	0	**29**	10	9	8	**52**	20	19	17	**96**	40	38	35
8	2	1	1	**30**	11	10	8	**54**	21	20	18	**98**	41	39	36
9	2	2	1	**31**	11	10	8	**56**	22	21	18	**100**	42	40	37
10	2	2	1	**32**	11	10	9	**58**	23	22	19	**110**	46	45	42
11	3	2	1	**33**	12	11	9	**60**	24	22	20	**120**	51	49	46
12	3	3	2	**34**	12	11	10	**62**	25	23	21	**130**	56	54	50
13	4	3	2	**35**	13	12	10	**64**	25	24	22	**140**	60	58	55
14	4	3	2	**36**	13	12	10	**66**	26	25	23	**150**	65	63	59
15	4	4	3	**37**	14	13	11	**68**	27	26	23	**160**	70	68	64
16	5	4	3	**38**	14	13	11	**70**	28	27	24	**170**	74	72	68
17	5	5	3	**39**	14	13	12	**72**	29	28	25	**180**	79	77	73
18	6	5	4	**40**	15	14	12	**74**	30	29	26	**190**	84	82	77
19	6	5	4	**41**	15	14	12	**76**	31	29	27	**200**	88	86	82
20	6	6	4	**42**	16	15	13	**78**	32	30	28	**220**	98	95	91
21	7	6	5	**43**	16	15	13	**80**	33	31	29	**240**	107	105	100
22	7	6	5	**44**	17	16	14	**82**	34	32	29	**260**	117	114	109
23	8	7	5	**45**	17	16	14	**84**	34	33	30	**280**	126	124	118
24	8	7	6	**46**	17	16	14	**86**	35	34	31	**300**	136	133	128
25	8	8	6	**47**	18	17	15	**88**	36	35	32				
26	9	8	7	**48**	18	17	15	**90**	37	36	33				

Примечание: α – уровень значимости.

Пример использования G-критерия знаков

В таблице приведены результаты измерения уровней агрессивности школьников до (*A*) и после (*B*) просмотра боевика.

Имя	A	B
Алексей	3	4
Борис	4	5
Леонид	3	5
Марина	3	4
Мария	3	4
Михаил	2	3
Настя	3	2
Николай	4	5
Олег	4	5
Ольга	5	5
Петр	2	3
Рита	5	4

Требуется установить уровень статистической значимости различий уровней агрессивности школьников до (A) и после (B) просмотра боевика.

К протоколу свойств *A* и *B* добавим и заполним столбец «Знак A – B».

Имя	A	B	Знак A – B
Алексей	3	4	–
Борис	4	5	–
Леонид	3	5	–
Марина	3	4	–
Мария	3	4	–
Михаил	2	3	–
Настя	3	2	+
Николай	4	5	–
Олег	4	5	–
Ольга	5	5	0
Петр	2	3	–
Рита	5	4	+

Введем обозначения:

$2 = a$ – число «плюсов» в столбце «Знак A – B»;

$9 = b$ – число «минусов» в столбце «Знак A – B»;

11 = n – сумма a и b;

2 = G – число, равное меньшему из чисел a и b.

Найдем в таблице G-критерия знаков в строке 11 = n значения $G0{,}10 = 3$, $G0{,}05 = 2$ и $G0{,}01 = 1$.

Статистический вывод.

Так как $1 < G \leq 2$, то есть $G0{,}01 < G \leq G0{,}05$, значит, принимается гипотеза H_1 ($p \leq 0{,}05$).

Содержательный вывод.

Выявлены статистически значимые различия ($p \leq 0{,}05$) уровней агрессивности школьников до и после просмотра боевика, который повысился после просмотра.

5.5. Ранжирование выборки

Рангом варианты выборки называют число, обозначающее место варианты в упорядоченной последовательности всех вариант выборки.

Ранжирование выборки – замена каждой варианты выборки соответствующими им рангами.

Алгоритм ранжирования выборки, размещенной в протоколе (Excel)

1. В файле с протоколом эмпирических данных создайте новый лист с названием «Ранжирование».

2. Лист «Протокол» скопируйте и вставьте на лист «Ранжирование».

3. Проведите сортировку вариант выборки, значения которой предстоит ранжировать.

4. К таблице добавьте столбцы «Номер» и «Ранг».

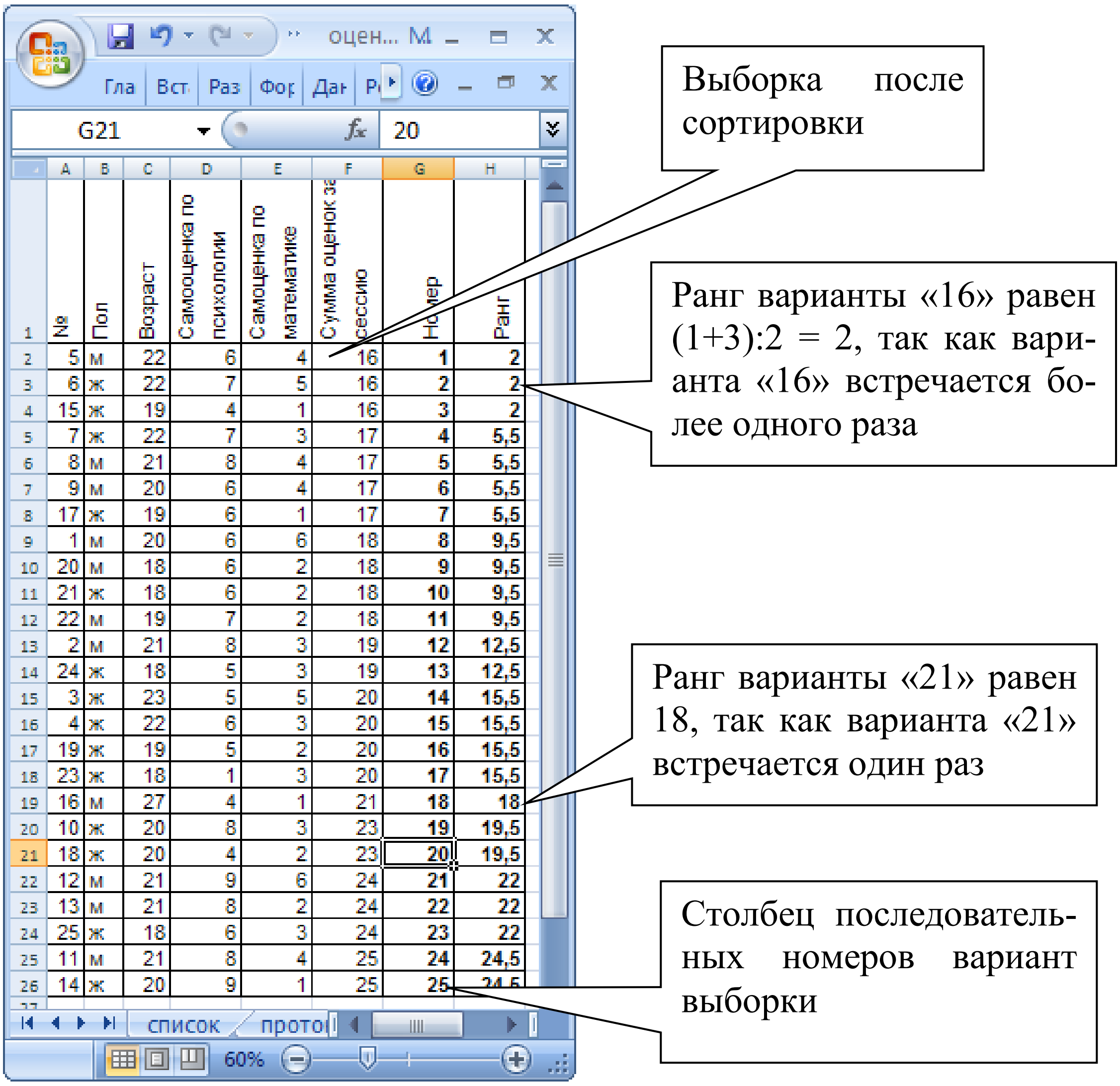
5. В столбце «Номер» последовательно пронумеруйте отсортированные значения вариант выборки.

6. В столбце «Ранг» присвойте ранг каждой варианте по правилу:

– если номер варианты встречается один раз, то *ранг* равен этому *номеру*;

– если номер варианты встречается более одного раза, то *ранг* равен *полусумме крайних номеров* варианты.

Пример ранжирования выборки «Сумма оценок за сессию»



№	Пол	Возраст	Самооценка по психологии	Самооценка по математике	Сумма оценок за сессию	Номер	Ранг
5	м	22	6	4	16	1	2
6	ж	22	7	5	16	2	2
15	ж	19	4	1	16	3	2
7	ж	22	7	3	17	4	5,5
8	м	21	8	4	17	5	5,5
9	м	20	6	4	17	6	5,5
17	ж	19	6	1	17	7	5,5
1	м	20	6	6	18	8	9,5
20	м	18	6	2	18	9	9,5
21	ж	18	6	2	18	10	9,5
22	м	19	7	2	18	11	9,5
2	м	21	8	3	19	12	12,5
24	ж	18	5	3	19	13	12,5
3	ж	23	5	5	20	14	15,5
4	ж	22	6	3	20	15	15,5
19	ж	19	5	2	20	16	15,5
23	ж	18	1	3	20	17	15,5
16	м	27	4	1	21	18	18
10	ж	20	8	3	23	19	19,5
18	ж	20	4	2	23	20	19,5
12	м	21	9	6	24	21	22
13	м	21	8	2	24	22	22
25	ж	18	6	3	24	23	22
11	м	21	8	4	25	24	24,5
14	ж	20	9	1	25	25	24,5

Алгоритм ранжирования, если составлено распределение частот выборки

1. К таблице распределения частот выборки добавьте строки «Номер» и «Ранг».

2. В строке «Номер» последовательно пронумеруйте значения вариант с учетом частоты варианты.

3. В строке «Ранг» присвойте ранг каждой варианте по правилу:

– если номер варианты встречается один раз, то *ранг* равен этому *номеру*;

– если номер варианты встречается более одного раза, то *ранг* равен *полусумме крайних номеров* варианты.

Пример ранжирования выборки «Оценки по химии»

В таблице приведены результаты аттестации школьников по химии.

Оценка в баллах	17	23	36	42	58	61	70
Частота	1	3	1	3	8	5	2

Проведем ранжирование вариант выборки.

Дополним таблицу строкой «Номер».

Заполним строку «Номер» начиная с 1 и заканчивая 23 (объем выборки 23 респондента).

Варианта «17» имеет номер 1-й, так как ее частота равна 1.

Варианта «23» имеет номера со 2-го по 4-й (2, 3, 4), так как ее частота равна 3.

Варианта «36» имеет номер 5-й (следующий за 4-м), так как ее частота равна 1.

И т.д.

Оценка в баллах	17	23	36	42	58	61	70
Частота	1	3	1	3	8	5	2
Номер	1	2–4	5	6–8	9–16	17–21	22–23

Дополним таблицу строкой «Ранг».

Заполним строку «Ранг», следуя правилу алгоритма.

Варианта «17» имеет ранг 1, так как ее номер встречается один раз.

Варианта «23» имеет ранг 3 = (2 + 4):2, так как ее номер встречается более одного раза.

Варианта «36» имеет ранг 5, так как ее номер встречается один раз.

И т.д.

Оценка в баллах	17	23	36	42	58	61	70
Частота	1	3	1	3	8	5	2
Номер	1	2–4	5	6–8	9–16	17–21	22–23
Ранг	1	3	5	7	12,5	19	22,5

5.6. *U*-КРИТЕРИЙ МАННА-УИТНИ

С помощью *U*-критерия Манна-Уитни устанавливается уровень статистической значимости различий одного свойства у респондентов двух выборок (I и II).

Вопросы, ответы на которые можно найти с помощью U-критерия Манна-Уитни

1. Есть ли статистически значимые различия показателей общительности у студентов, обучающихся на первом курсе (I), и студентов, обучающихся на втором курсе (II)?
2. Есть ли статистически значимые различия показателей респондентов экспериментальной (I) и контрольной групп (II)?

Особенности применения *U*-критерия Манна-Уитни

1. В выборке должно быть не менее четырех-пяти респондентов.

2. Выборки одного свойства (I и II) должны быть несвязанными.

3. Используется таблица *U*-критерия Манна-Уитни (по Е.В. Сидоренко).

Предельное значение *U*-критерия Манна-Уитни находится на пересечении строки n_1 (объем первой выборки) и столбца n_2 (объем второй выборки).

Если пересечение является пустой ячейкой, тогда предельное значение *U*-критерия Манна-Уитни находится на пересечении строки n_2 (объем второй выборки) и столбца n_1 (объем первой выборки).

Таблица критерия Манна-Уитни ($\alpha = 0{,}05$)

n_1 / n_2	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
5	2	4																									
6	3	5	7																								
7	4	6	8	11																							
8	5	8	10	13	15																						
9	6	9	12	15	18	21																					
10	7	11	14	17	20	24	27																				
11	8	12	16	19	23	27	31	34																			
12	9	13	17	21	26	30	34	38	42																		
13	10	15	19	24	28	33	37	42	47	51																	
14	11	16	21	26	31	36	41	46	51	56	61																
15	12	18	23	28	33	39	44	50	55	61	66	72															
16	14	19	25	30	36	42	48	54	60	65	71	77	83														
17	15	20	26	33	39	45	51	57	64	70	77	83	89	96													
18	16	22	28	35	41	48	55	61	38	75	82	88	95	102	109												
19	17	23	30	37	44	51	58	65	72	80	87	94	101	109	116	123											
20	18	25	32	39	47	54	62	69	77	84	92	100	107	115	123	130	138										
21	19	26	34	41	49	57	65	73	81	89	97	105	113	121	130	138	146	154									
22	20	28	36	44	52	60	69	77	85	94	102	111	119	128	136	145	154	162	171								
23	21	29	37	46	55	63	72	81	90	99	107	116	125	134	143	152	161	170	180	189							
24	22	31	39	48	57	66	75	85	94	103	113	122	131	141	150	160	169	179	188	198	207						
25	23	32	41	50	60	69	79	89	98	108	118	128	137	174	157	167	177	187	197	207	217	227					
26	24	33	43	53	62	72	82	93	103	113	123	133	143	154	164	174	185	195	206	216	226	237	247				
27	25	35	45	55	65	75	86	96	107	118	128	139	150	160	171	182	193	203	214	225	236	247	258	268			
28	26	36	47	57	68	79	89	100	111	122	133	144	156	167	178	189	200	212	223	234	245	257	268	279	291		
29	27	38	48	59	70	82	93	104	116	127	139	150	162	173	185	196	208	220	232	243	255	267	278	290	302	314	
30	28	39	50	62	73	85	96	108	120	132	144	156	168	180	192	204	216	228	240	252	265	277	289	301	313	326	338

Таблица критерия Манна-Уитни ($\alpha = 0,01$)

n_1 / n_2	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
5	0	1																									
6	1	2	3																								
7	1	3	4	6																							
8	2	4	6	7	9																						
9	3	5	7	9	11	14																					
10	3	6	8	11	13	16	19																				
11	4	7	9	12	15	18	22	25																			
12	5	8	11	14	17	21	24	28	31																		
13	5	9	12	16	20	23	27	31	35	39																	
14	6	10	13	17	22	26	30	34	38	43	47																
15	7	11	15	19	24	28	33	37	42	47	51	56															
16	7	12	16	21	26	31	36	41	46	51	56	61	66														
17	8	13	18	23	28	33	38	44	49	55	60	66	71	77													
18	9	14	19	24	30	36	41	47	53	59	65	70	76	82	88												
19	9	15	20	26	32	38	44	50	56	63	69	75	82	88	94	101											
20	10	16	22	28	34	40	47	53	60	67	73	81	87	93	100	107	114										
21	10	16	22	29	35	42	49	56	63	70	77	84	91	98	105	113	120	127									
22	10	17	23	30	37	45	52	59	66	74	81	89	96	104	11	119	127	134	142								
23	11	18	25	32	39	47	55	62	70	78	86	94	102	109	117	125	133	141	150	158							
24	12	19	26	34	42	49	57	66	74	82	90	98	107	115	123	132	140	149	154	166	174						
25	12	20	27	35	44	52	60	69	77	86	95	103	112	121	130	138	147	156	165	174	183	192					
26	13	21	29	37	46	54	63	72	81	90	99	108	117	126	136	145	154	163	173	182	191	201	210				
27	14	22	30	39	48	57	66	75	85	94	103	113	122	132	142	151	161	171	180	190	200	209	219	229			
28	14	23	32	41	50	59	69	78	88	98	108	118	128	138	148	158	168	178	188	198	208	218	229	239	249		
29	15	24	33	42	52	62	72	82	92	102	112	123	133	143	154	164	175	185	196	206	217	227	238	249	259	270	
30	15	25	34	44	54	64	75	85	95	106	117	127	138	149	160	171	182	192	203	214	225	236	247	258	270	281	292

Алгоритм *U*-критерия Манна-Уитни

1. Проведите ранжирование общей выборки, в которую входят сравниваемые выборки (I и II).

2. Найдите сумму рангов вариант выборки I и сумму рангов вариант выборки II.

3. Вычислите значение U по формуле

$$U = n_1 * n_2 + 0{,}5 * n_R * (n_R + 1) - R.$$

где n_1 – объем выборки I;

n_2 – объем выборки II;

n_R – объем выборки, имеющей большую сумму рангов;

R – значение большей суммы рангов.

3. Найдите в таблице U-критерия Манна-Уитни на пересечении строки и столбца n_1 и n_2 значения $U0{,}05$ и $U0{,}01$.

4. Статистический вывод.

Если $U > U0{,}05$, то принимается гипотеза H_0.

Если $U0{,}01 < U \leq U0{,}05$, то принимается гипотеза H_1 ($p \leq 0{,}05$).

Если $U \leq U0{,}01$, то принимается гипотеза H_1 ($p \leq 0{,}01$).

Пример использования U-критерия Манна-Уитни

В протоколе приведены результаты измерения самооценок по математике студентов очной (I) и заочной (II) форм обучения.

Инициалы	Обучение	Оценка	Инициалы	Обучение	Оценка
АК	очное	6	АВ	заочное	1
БМ	очное	3	ВД	заочное	2
ВГ	очное	3	ГЕ	заочное	5
ЖА	очное	5	ЖФ	заочное	1
НМ	очное	2	ОВ	заочное	3
ПА	очное	2	ПВ	заочное	2
ПВ	очное	3	ПФ	заочное	5
РС	очное	6	ПР	заочное	6
РК	очное	8	РА	заочное	9
РЛ	очное	6	РБ	заочное	2
ТФ	очное	6	РС	заочное	7
ТЛ	очное	7	ФА	заочное	4
УВ	очное	7	ЦД	заочное	1
ЧС	очное	8			

Требуется установить уровень статистической значимости различий самооценок по математике студентов очной (I) и заочной (II) форм обучения.

Составим распределение частот самооценок по математике студентов всей выборки.

Самооценка	1	2	3	4	5	6	7	8	9
Частота	3	5	4	1	3	5	3	2	1

Осуществим ранжирование выборки самооценок по математике.

Самооценка	1	2	3	4	5	6	7	8	9
Частота	3	5	4	1	3	5	3	2	1
Номер	1–3	4–8	9–12	13	14–16	17–21	22–24	25–26	27
Ранг	2	6	10,5	13	15	19	23	25,5	27

Заменим варианты выборки соответствующими им рангами и вычислим сумму рангов в выборках очной (I) и заочной (II) форм обучения.

Инициалы	Обучение	Оценка	Ранг		Инициалы	Обучение	Оценка	Ранг
АК	очное	6	19		АВ	заочное	1	2
БМ	очное	3	10,5		ВД	заочное	2	6
ВГ	очное	3	10,5		ГЕ	заочное	5	15
ЖА	очное	5	15		ЖФ	заочное	1	2
НМ	очное	2	6		ОВ	заочное	3	10,5
ПА	очное	2	6		ПВ	заочное	2	6
ПВ	очное	3	10,5		ПФ	заочное	5	15
РС	очное	6	19		ПР	заочное	6	19
РК	очное	8	25,5		РА	заочное	9	27
РЛ	очное	6	19		РБ	заочное	2	6
ТФ	очное	6	19		РС	заочное	7	23
ТЛ	очное	7	23		ФА	заочное	4	13
УВ	очное	7	23		ЦД	заочное	1	2
ЧС	очное	8	25,5		Сумма рангов выборки II			146,5
Сумма рангов выборки I			231,5					

Заметим, что выборка I составила 14 студентов, выборка II – 13. Большая сумма рангов в выборке II – 231,5.

Вычислим значение U по формуле алгоритма U-критерия Манна-Уитни:

$$U = 14*13 + 0{,}5*14*(14+1) - 231{,}5 = 55{,}5,$$

где 14 (n_1) – объем выборки I;
13 (n_2) – объем выборки II;
14 (n_R) – объем выборки, имеющей большую сумму рангов;
231, 5 (R) – значение большей суммы рангов.

Найдем в таблице U-критерия Манна-Уитни на пересечении строки n_1 и столбца n_2 значения $U0{,}05 = 56$ и $U0{,}01 = 43$.

Статистический вывод.

Так как $43 < U \leq 56$, то принимается гипотеза H_1 ($p \leq 0{,}05$).

Содержательный вывод.

Выявлены статистически значимые различия ($p \leq 0{,}05$) самооценок по математике студентов очной и заочной форм обучения.

Резюме

Рассмотрены критерии сравнения выборок (φ-критерий Фишера, λ-критерий Колмогорова-Смирнова, G-критерий знаков, U-критерий Манна-Уитни), а также алгоритм ранжирования выборки, замены выборочных данных соответствующими им рангами. Эти критерии относятся к непараметрическим методам сравнения выборок.

Отличия непараметрических и параметрических методов будет описано в следующих главах.

Существуют и другие критерии сравнения выборок, но, по нашему мнению, для психологических исследований данный перечень непараметрических критериев почти полный.

Упражнения

5.1. По данным Всемирной метеорологической организации вероятность прогноза погоды на один день равна 95%, на три дня – 90%, на десять дней – 85%.

Определите вероятность ошибки прогноза погоды:

а) на один день;

б) на три дня;

в) на десять дней.

5.2. Приведите примеры:
а) связанных выборок;
б) несвязанных выборок.

5.3. Сформулируйте статистические гипотезы о различиях выборок студентов:
а) самооценок по математике юношей и девушек;
б) самооценок студентов по психологии и по математике.

5.4. В таблице приведены распределения частот рангов ценности «Работа» у студентов до (I выборка) и после (II выборка) производственной практики.

Ранг ценности «Работа»	1	2	3	4	5	6
Частота I выборки	4	8	9	6	12	16
Частота II выборки	13	6	14	9	3	10

Установите уровень статистической значимости различий распределений частот рангов ценности «Работа» у студентов до и после производственной практики.

5.5. В протоколе приведены самооценки студентов по психологии (А) и самооценки по математике (В).

Имя	А	В	Имя	А	В	Имя	А	В
Алексей	4	1	Рая	7	4	Николай	7	2
Борис	4	1	Роза	5	5	Олег	1	3
Вова	5	7	Тарас	8	6	Юля	8	7
Леонид	6	1	Света	5	5	Ольга	5	3
Марина	4	2	Сергей	7	5	Петр	6	3
Мария	5	2	Стас	7	6	Рита	6	3
Михаил	6	2	Толя	7	6	Роман	8	9
Настя	6	6	Таня	7	6	Таня	7	6
Юра	7	8	Федя	8	6	Вася	7	8

Установите уровень статистической значимости различий самооценок по психологии и самооценок по математике у студентов.

5.6. В первой группе, состоящей из 24 студентов, у шести из них выявлен высокий уровень общительности. Во второй группе, состоящей из 28 студентов, у двенадцати выявлен высокий уровень общительности.

Есть ли статистически значимые различия долей студентов одной группы от студентов другой группы, имеющих высокий уровень общительности?

5.7. В таблице приведены результаты теста «Логические способности», проведенного среди школьников.

Оценка в баллах	17	19	20	21	22	24	25
Частота	6	3	8	1	5	3	2

Проведите ранжирование вариант выборки.

5.8. В протоколе представлены результаты тестирования школьников по физике в баллах.

Имя	Балл	Имя	Балл	Имя	Балл	Имя	Балл
Алексей	60	Дмитрий	55	Леонид	46	Роман	75
Алена	55	Елена	61	Марина	86	Светлана	58
Андрей	55	Жанна	51	Мария	64	Сергей	50
Белла	31	Зина	48	Михаил	85	Стас	24
Борис	89	Игорь	92	Настя	63	Тарас	27
Вадим	69	Ирина	42	Николай	84	Татьяна	75
Вера	38	Катя	71	Олег	62	Ульяна	80
Галина	39	Клава	73	Ольга	77	Федор	49
Григорий	52	Костя	64	Петр	77	Юрий	24
Дина	54	Лариса	70	Рита	39	Яна	90

Установите уровень статистической значимости различий результатов тестирования школьников-мальчиков и школьников-девочек.

6. ВЫБОРОЧНЫЙ МЕТОД

Задача психолога – выявить психологические особенности, общие для группы людей.

Если группа людей относительно небольшая (например, студенческая группа), то для установления общепсихологических особенностей группы проводят опросные процедуры с каждым членом группы.

Если группа людей большая (например, студенты университета), тогда для опроса формируется подгруппа (выборка), отражающая особенности всей группы.

В психологии основным методом установления общепсихологических закономерностей поведения большой группы людей является *выборочный* метод.

6.1. ГЛОССАРИЙ ВЫБОРОЧНОГО МЕТОДА

Генеральная совокупность и выборка

Генеральной совокупностью принято называть группу людей или множество всех возможных значений измеряемого у группы свойства.

Объем *генеральной совокупности* принято обозначать η (эта).

Выборкой принято называть любую подгруппу *генеральной совокупности*.

Объем выборки принято обозначать n.

Пример генеральной совокупности и выборки

Генеральная совокупность – студенты второго курса университета. Объем генеральной совокупности – около 3500 студентов ($\eta = 3500$).

Выборка – студенты второго курса, осваивающие специальность «Психология». Объем выборки – 27 студентов ($n = 27$).

Параметры и статистики

Численные характеристики генеральной совокупности принято называть *параметрами генеральной совокупности*: среднее, дисперсия, стандартное отклонение.

Численные характеристики выборки принято называть *статистиками выборки*.

Принятые обозначения

m (эм) – среднее выборки;

s (эс) – стандартное отклонение выборки;

n (эн) – объем выборки;

μ (мю) – среднее генеральной совокупности;

σ (сигма) – стандартное отклонение генеральной совокупности;

η (эта) – объем генеральной совокупности.

Репрезентативность выборки

Выборку, *статистики* которой с высокой вероятностью равны *параметрам* генеральной совокупности, принято называть *репрезентативной выборкой* для данной генеральной совокупности.

Репрезентативность выборки имеет вероятностный характер.

Репрезентативность является двумерной характеристикой выборки и зависит:

– от объема выборки (доверительная вероятность репрезентативности объема);

– от способа формирования выборки (вероятность репрезентативности отбора).

По объему выборки классифицируются на малые, средние и большие, которые имеют объем:

– менее 30 респондентов – малая выборка;

– от 30 до 200 респондентов – средняя выборка;

– более 200 респондентов – большая выборка.

Отбор вариант выборок классифицируют на случайный и неслучайный отбор.

Случайный отбор имеет случайную ошибку репрезентативности, которую можно оценить, используя математические методы.

Неслучайный (смещенный) отбор имеет систематическую ошибку репрезентативности, которая не поддается оценке. Систематическую ошибку необходимо устранять с помощью процедуры отбора.

Доверительная вероятность репрезентативности выборки

Репрезентативность выборки в первую очередь зависит от ее объема. Чем меньше объем выборки отличается от объема генеральной совокупности, тем более выборка репрезентативна.

Вероятность ошибки отличия объема выборки от объема генеральной совокупности называется доверительной вероятностью. Если доверительную вероятность обозначим *p*, тогда ошибка, допустимая при установлении объема выборки, будет равна 1 – *p*, которую будем обозначать α.

В психологических исследованиях уровень ошибки репрезентативности при установлении объема выборки принимается:

– α = 0,20 (тенденция),
– α = 0,10 (допустимая),
– α = 0,05 (нормальная),
– α = 0,01 (высокая).

Объем выборки с установленной ошибкой репрезентативности вычисляется по формуле

$$n = \frac{1}{\alpha^2 + \frac{1}{\eta}},$$

где n – объем выборки;
α – уровень ошибки;
η – объем генеральной совокупности.

Значение α вычисляется по формуле

$$\alpha = \sqrt{\frac{1}{n} - \frac{1}{\eta}}$$

Таблица предельных значений объема выборки (*n*), зависящих от объема генеральной совокупности (η) и ошибки репрезентативности объема выборки (*α*)

Ошибка	η=100	η=500	η=1000	η=10000	η=100000	η =1000000
α = 0,20	23	24	24	25	25	25
α = 0,10	50	83	91	99	100	100
α = 0,05	80	222	286	385	398	400

Случайные и систематические ошибки репрезентативности

Основные виды ошибок репрезентативности – *систематические и случайные*.

Систематические ошибки возникают тогда, когда статистики выборки отклоняются от параметра генеральной совокупности только в одну сторону. В этих случаях статистики выборки отличаются от параметров генеральной совокупности (имеют только большие или только меньшие значения). Выборку в этом случае называют *смещенной выборкой*.

Случайными ошибками называют ошибки, при которых статистики выборки отклоняются от параметра генеральной совокупности приблизительно одинаково как в большую, так и в меньшую сторону. Случайные ошибки имеют вероятностный характер, их можно оценить с помощью математических методов. Случайная ошибка присутствует всегда. *Вероятность случайных ошибок* – одна из важных характеристик исследования.

Пример систематической ошибки

В США регулярно проводятся опросы, в ходе которых избирателям рассылаются опросные листы, с просьбой отметить, чья кандидатура на предстоящих президентских выборах более предпочтительна.

В 1936 г. прогноз популярного в США журнала «Literary Digest», что победит Ландон с результатом 60% голосов, оказался неверным. Альф Ландон проиграл Франклину Рузвельту, который набрал около 60% голосов. После этого журнал перестали покупать и его закрыли.

Причиной неточного прогноза была *систематическая ошибка* репрезентативности выборки. Выборку, объемом *два миллиона* респондентов, сотрудники журнала образовали по телефонным книгам и спискам регистрации автомобилей. В 1936 г. телефоны и автомобили были преимущественно у обеспеченных граждан. В выборке отсутствовали мнения малообеспеченных граждан, поэтому прогноз оказался смещенным в пользу мнений состоятельных избирателей, доля которых к тому же меньше.

Выборка оказалась репрезентативной для генеральной совокупности, состоящей из обеспеченных избирателей.

Пример случайной ошибки

В то же время (1936 г.) прогноз исследователей общественного мнения Дж. Гелапа и Эл. Роупера, что победит Рузвельт с результатом 60% голосов, оказался верным.

Исследователи правильно предсказали победу Рузвельта, основываясь на анализе опросов всего *четырех тысяч* избирателей. Они сформировали выборку с одинаковым представительством всех слоев общества с учетом данных о составе электората (пол, возраст, уровень благосостояния, семейное положение и др.).

Ошибка прогноза в этом случае имела *случайный* (вероятностный) характер.

Выборка оказалась репрезентативной для генеральной совокупности, пропорционально представляющей структуру электората США.

6.2. Случайный отбор респондентов

Процедура случайного отбора респондентов называется *рандомизацией* (*англ. r*andom – случайно). Выборку в этом случае называют *рандомизированной* выборкой (отобранной случайно).

Выбор одного респондента из генеральной совокупности объема η считается случайным, если вероятность его выбора равна вероятности выбора любого другого респондента – $1/\eta$.

Рандомизированную выборку можно получить, если каждый респондент отобран случайно из генеральной совокупности. Опишем основные способы формирования рандомизированной выборки.

Простой случайный отбор

Простой случайный отбор применяется тогда, когда генеральная совокупность имеет небольшой объем. Для этого способа необходима таблица равномерно распределенных случайных чисел.

Алгоритм простого случайного отбора

1. Каждому респонденту генеральной совокупности присваивают номер от 1 до η.

2. Определяется число цифр случайных чисел.

Если $\eta < 10$, то однозначное случайное число.

Если $9 < \eta < 100$, то двухзначное случайное число.

Если $99 < \eta < 1000$, то трехзначное случайное число.

3. Из таблицы случайных чисел выписывают подряд все числа с установленным числом цифр. При этом руководствуются следующими правилами:

– если случайное число из таблицы превышает число, равное объему генеральной совокупности, то его пропускают;

– если случайное число встречается более одного раза, то его пропускают.

4. Отбор завершают тогда, когда получена выборка требуемого объема.

Пример простого случайного отбора

Из 36 курсантов необходимо отобрать случайным образом 9 курсантов. Объем генеральной совокупности (η) равен 36, объем выборки (n) равен 8.

Каждому курсанту присвоим номер от 1 до 36.

Так как $\eta = 36$ – двухзначное число, значит, будем отбирать из таблицы случайных чисел двухзначные числа.

Таблица случайных чисел

03 47 43 73 86 97 74 24 67 62 16 76 62 27 66 12 56 85 99 26 55 59
56 35 64 16 22 77 94 39 84 42 17 53 31 63 01 63 78 59 33 21 12 34

Подчеркиваем последовательно двухзначные числа, начиная с числа 03, пропуская числа, большие 36 и ранее подчеркнутые.

Таблица подчеркнутых случайных чисел

03 47 43 73 86 97 74 24 67 62 16 76 62 27 66 12 56 85 99 26 55 59
56 35 64 16 22 77 94 39 84 42 17 53 31 63 01 63 78 59 33 21 12 34

Выпишем подчеркнутые числа. В искомую выборку попадут курсанты, имеющие следующие номера: 03, 24, 16, 27, 12, 26, 35, 22, 17.

Механический случайный отбор

Механический случайный отбор применяется тогда, когда генеральная совокупность имеет небольшой объем. Для этого способа составляется список респондентов.

Алгоритм механического случайного отбора

1. Каждому респонденту списка присвойте номер от 1 до η.
2. Найдите число k, равное отношению объема выборки к объему генеральной совокупности: $k = \eta/n$.
3. Случайным образом выберите номер первого респондента из первых k респондентов.
4. В списке, начиная с выбранного номера, последовательно отсчитайте и подчеркните каждый k-й номер.
5. Респонденты с отмеченными номерами образуют искомую выборку.

Пример механического случайного отбора

Из 36 курсантов необходимо отобрать случайным образом 9 курсантов. Объем генеральной совокупности (η) равен 36, объем выборки (n) равен 9.

Вычислим k: k = 36:9 = 4.

Каждому курсанту присвоим номер от 1 до 36. Просим кого-нибудь назвать число от 1 до 4-х. Например, названо число 3.

В списке респондентов, начиная с номера три, последовательно отсчитываем и подчеркиваем каждый четвертый номер.

Таблица подчеркнутых четвертых номеров

1 2 <u>3</u> 4 5 6 <u>7</u> 8 9 10 <u>11</u> 12 13 14 <u>15</u> 16 17 18 <u>19</u> 20 21 22 <u>23</u> 24 25 26 <u>27</u> 28 29 30 <u>31</u> 32 33 34 <u>35</u> 36

Выпишем подчеркнутые числа. В искомую выборку попадут курсанты, имеющие номера 3, 7, 11, 15, 19, 23, 27, 31, 35.

Типический случайный отбор

Типический случайный отбор применяется тогда, когда генеральная совокупность имеет большой объем. Этот способ применяется тогда, когда генеральную совокупность можно классифицировать по группам (типам).

Алгоритм типического случайного отбора

1. Найдите число k, равное η/n, где n – объем выборки; η – объем генеральной совокупности.
2. Выделите группы генеральной совокупности и их объемы: $\eta_1, \eta_2, \dots, \eta_m$.

4. Найдите число респондентов, отбираемых в каждой группе генеральной совокупности по формуле $n_i = \eta_i/k$.

5. Способом механического случайного отбора в каждой группе генеральной совокупности отберите по n_i респондентов.

Пример типического случайного отбора

Из 300 студентов, в том числе 200 женского и 100 мужского полов, необходимо отобрать случайным образом 50 студентов.

Объем генеральной совокупности (η) равен 300, объем выборки (n) равен 50.

Вычислим k: $k = 300:50 = 6$.

Объем групп генеральной совокупности: 200 женского (η_1) и 100 мужского (η_2) полов.

Вычислим число респондентов, выбираемых:

– из группы женского пола: $n_1 = 200:6 = 33$;

– из группы мужского пола: $n_2 = 100:6 = 17$.

Составляем список девушек и список юношей.

В списке девушек последовательно отсчитываем и подчеркиваем каждый тридцать третий номер. В списке юношей – каждый семнадцатый номер.

В выборку попадут респонденты из двух списков, чьи номера подчеркнуты.

Ступенчатый случайный отбор

Ступенчатый случайный отбор применяется тогда, когда генеральная совокупность имеет большой объем. Этот способ применяется тогда, когда генеральная совокупность имеет сложную структуру, ее можно классифицировать по группам и подгруппам.

Алгоритм ступенчатого случайного отбора

1. Выделите структуру генеральной совокупности.

2. Пошагово примените процедуру типического случайного отбора.

Пример ступенчатого случайного отбора

Из 300 студентов необходимо отобрать случайным образом 50 студентов пропорционально форме обучения и полу студентов. На очной форме обучения 150 студентов женского и 30 студентов мужского пола. На заочной форме обучения 50 студентов женского и 70 студентов мужского пола.

Объем генеральной совокупности (η) равен 300, объем выборки (n) равен 50.

Шаг 1

Классифицируем генеральную совокупность на группы по форме обучения.

Получим, что 180 студентов обучаются по дневной, а 120 – по заочной форме обучения.

Классифицируем группы на подгруппы по полу.

На очной форме обучения 150 студентов женского и 30 студентов мужского пола.

На заочной форме обучения 50 студентов женского и 70 студентов мужского пола.

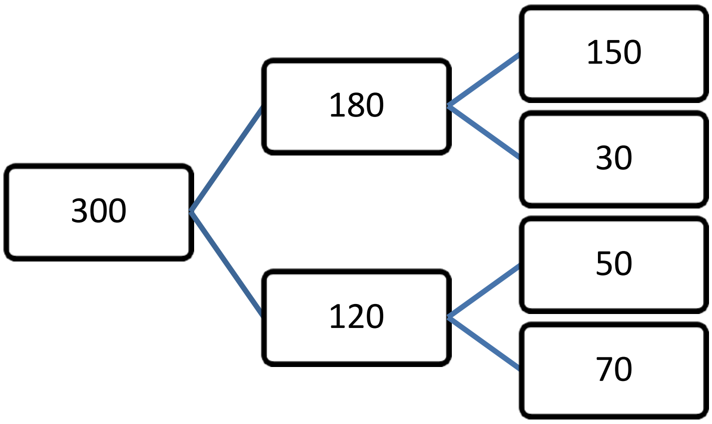


Рис. 4. Структура генеральной совокупности

Шаг 2

Вычислим k: k = 300:50 = 6.

Вычислим число респондентов, выбираемых из подгруппы очной формы обучения: 180:6 = 30.

Вычислим число респондентов, выбираемых из подгруппы заочной формы обучения: 120:6 = 20.

Вычислим число респондентов, выбираемых из подгруппы очной формы обучения:

– женского пола: 150:6 = 25;

– мужского пола: 30:6 = 5.

Вычислим число респондентов, выбираемых из подгруппы заочной формы обучения:

– женского пола: 50:6 = 8;

– мужского пола: 70:6 = 12.

Шаг 3

Составляем списки студентов:
– женского пола очной формы обучения;
– мужского пола очной формы обучения;
– женского пола заочной формы обучения;
– мужского пола заочной формы обучения.

Шаг 4

В списке студентов женского пола очной формы обучения отсчитываем и подчеркиваем каждый двадцать пятый номер.

В списке студентов мужского пола очной формы обучения отсчитываем и подчеркиваем каждый пятый номер.

В списке студентов женского пола заочной формы обучения отсчитываем и подчеркиваем каждый восьмой номер.

В списке студентов мужского пола заочной формы обучения отсчитываем и подчеркиваем каждый двенадцатый номер.

Шаг 5

Составляем выборку, в которую попадут респонденты из четырех списков, чьи номера подчеркнуты.

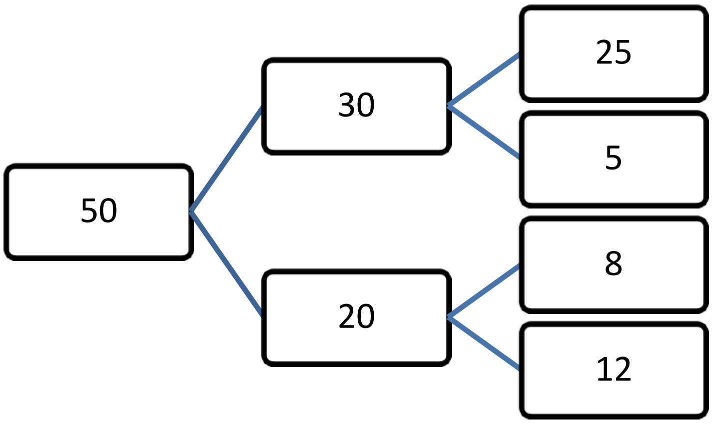


Рис. 5. Структура выборки

6.3. Виды распределения частот генеральной совокупности

Можно построить распределение частот измеренного свойства респондентов выборки из генеральной совокупности.

Если выборка является большой, то распределение частот выборки можно рассматривать как модель распределения частот генеральной совокупности интересующего исследователя психологического свойства, имеющегося у разных людей.

Нормальное распределение

Классическим примером моделирования распределения частот генеральной совокупности является исследование французского математика Абрахама де Муавра. Он измерил рост у 1375 случайно выбранных женщин и построил диаграмму распределения частот.

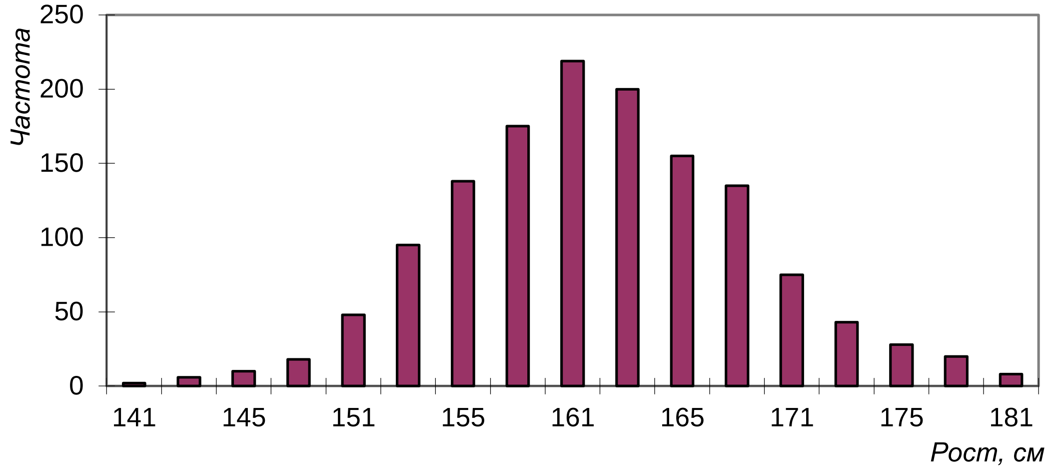


Рис. 6. Распределение частот роста женщин (выборка)

Если рост всех женщин является генеральной совокупностью, которая содержит несколько миллиардов респондентов, то распределение частот их роста можно представить в виде сглаженной колоколообразной кривой (аппроксимацией выборочного распределения частот).

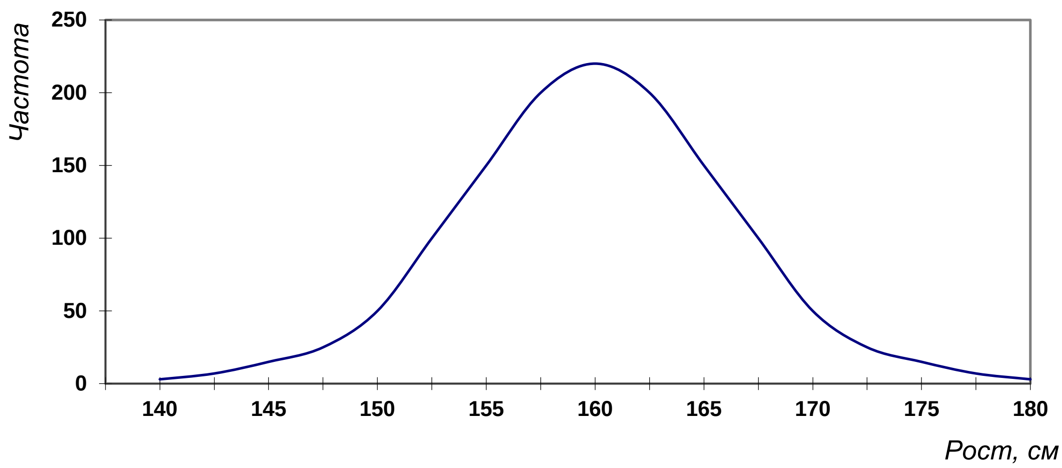


Рис. 7. Распределение частот роста женщин
(генеральная совокупность)

Оказалось, что аналогичное распределение частот можно наблюдать у многих случайных событий (скорость газовых молекул, показатели качества товаров, скорость автомобилей и др.), а также у большинства общепсихологических свойств людей.

Данное распределение получило название *нормальное распределение*.

Свойства нормального распределения

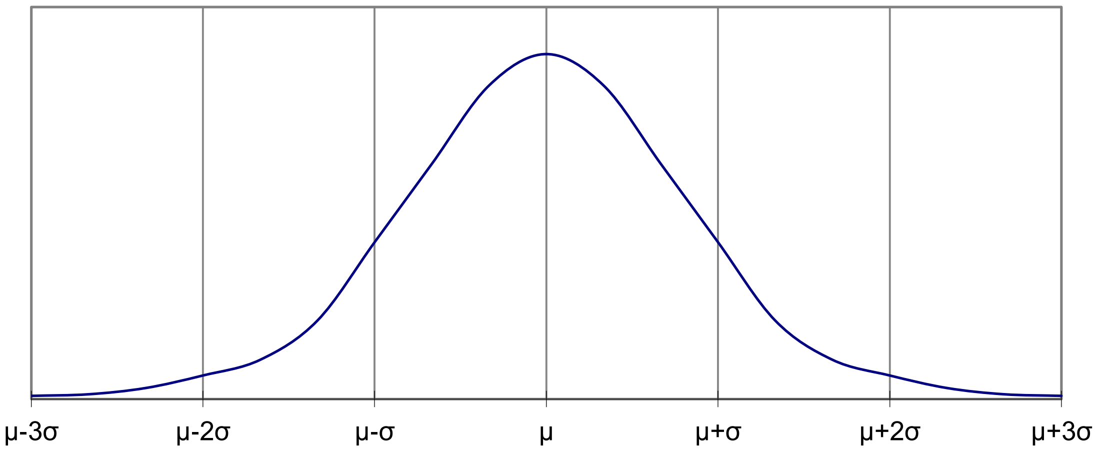


Рис. 8. График нормального распределения

Закон нормального распределения описывается формулой

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

где μ – среднее генеральной совокупности;

σ – стандартное отклонение генеральной совокупности.

Выделим основные свойства нормального распределения.

1. Нормальное распределение симметрично относительно среднего генеральной совокупности (μ).

2. До среднего генеральной совокупности (μ) нормальное распределение возрастает, после среднего – убывает.

3. От $\mu - \sigma$ до $\mu + \sigma$ находится 68,3% респондентов генеральной совокупности.

От $\mu - 2\sigma$ до $\mu + 2\sigma$ находится 95,4% респондентов генеральной совокупности.

От $\mu - 3\sigma$ до $\mu + 3\sigma$ находится 99,7% респондентов генеральной совокупности.

Равномерное распределение

Классическим примером равномерного распределения является распределение частот новорожденных по полу в конкретной стране.

Рассмотрим распределение частоты рождения девочек на каждые 1000 новорожденных детей (Швеция, 1935 г.).

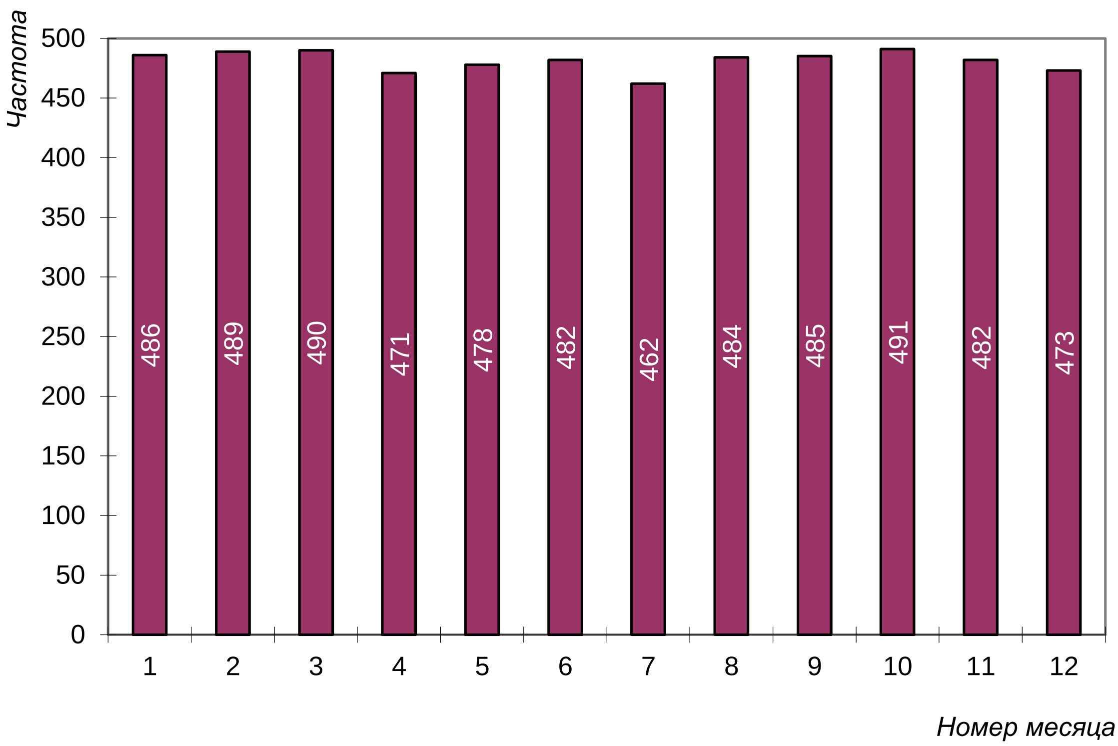


Рис. 9. Распределение частот рождения девочек (выборка)

Если число всех родившихся девочек является генеральной совокупностью, которая содержит несколько миллиардов респондентов, то распределение частот их рождения можно представить в виде прямой линии (аппроксимацией выборочного распределения частот).

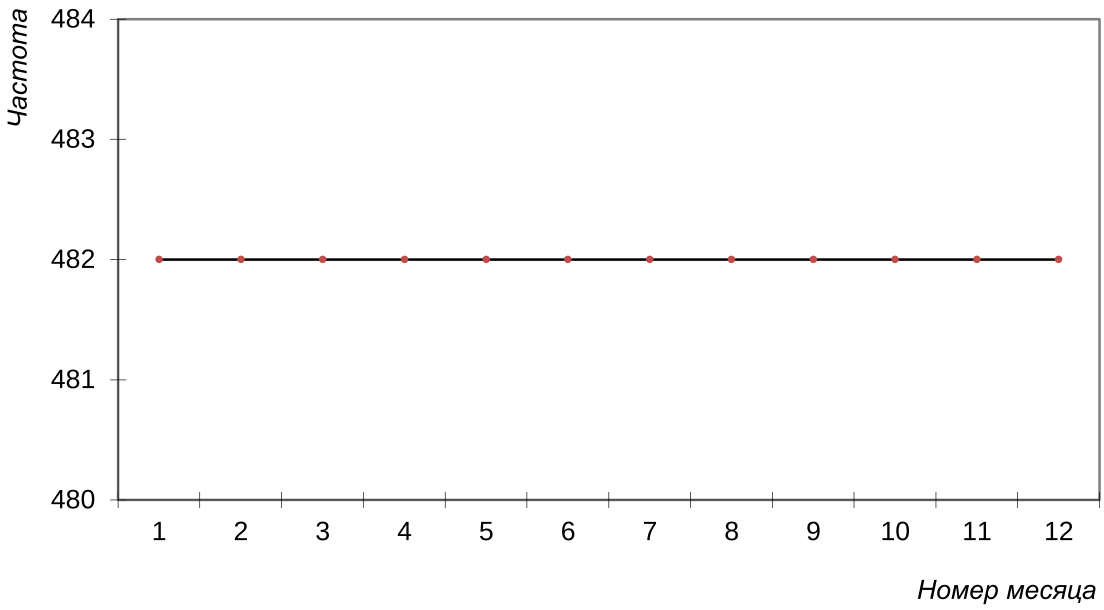


Рис. 10. Распределение частот рождения девочек на каждую тысячу новорожденных (генеральная совокупность)

Таким образом, из каждой 1000 новорожденных детей в среднем рождаются 482 девочки.

Равномерное распределение частот можно наблюдать у ряда психологических свойств людей: продолжительность работы в отдельных учреждениях, время сна новорожденного в сутки, посещаемость лекций студентами и др.

Выделим основные свойства равномерного распределения.

1. Значения распределения постоянны в течение наблюдаемого события.

2. Среднее генеральной совокупности приблизительно равно варианте выборки.

3. Стандартное отклонение генеральной совокупности равно нулю.

Монотонное распределение

Классический пример монотонного распределения можно увидеть в результатах исследования памяти, проведенных Германом Эббингаузом.

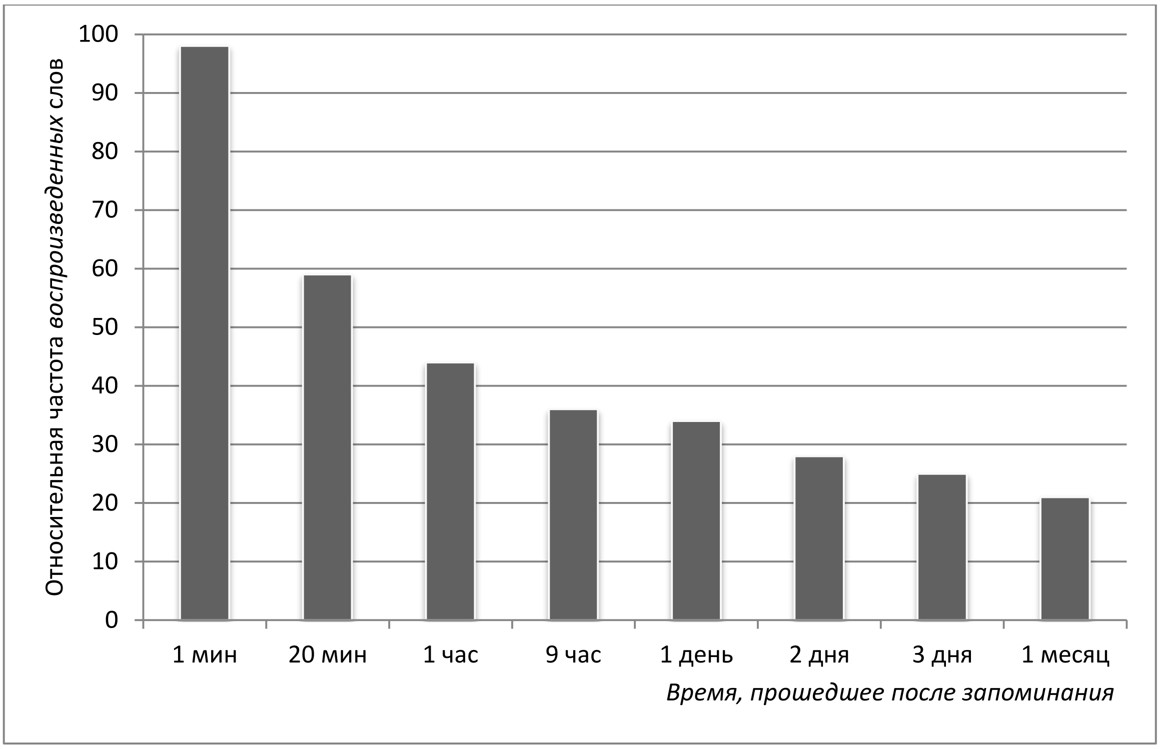


Рис. 11. Распределение частот бессмысленных слов, которые воспроизвел респондент (выборка)

Аппроксимацию выборочного распределения частот можно представить в виде сглаженной кривой.

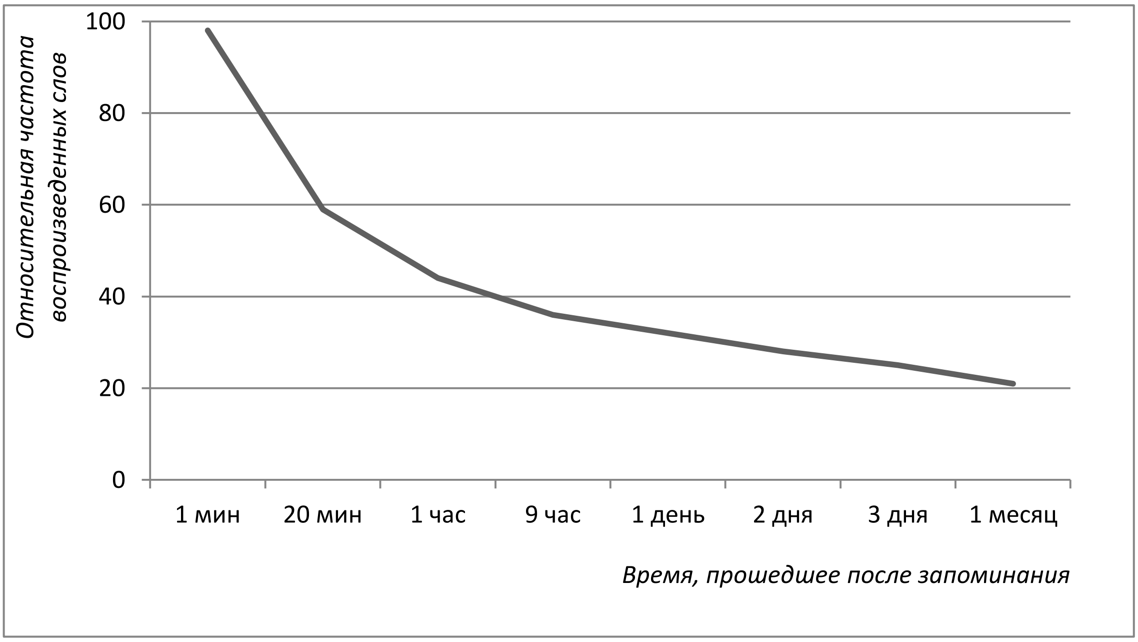


Рис. 12. Распределение частот бессмысленных слов, которые может воспроизвести человек (генеральная совокупность)

Монотонное распределение частот можно наблюдать у ряда психологических свойств людей: усталость в процессе учебных занятий, сосредоточенность внимания, продолжительность жизни и др.

Выделим основные свойства равномерного распределения.

1. Значения распределения монотонно уменьшаются или монотонно увеличиваются во время наблюдаемого события.

2. Монотонные распределения имеют, как правило, временной характер.

6.4. Проверка нормальности распределения частот выборки

Большинство общепсихологических свойств людей (рост, вес, кровяное давление, стаж работы, коммуникативные способности, академическая успеваемость, тревожность и др.) имеет нормальное распределение.

Если распределение частот выборки является близким к нормальному распределению, то это будет необходимым условием для сопоставления статистик выборки с общепсихологическими параметрами генеральной совокупности.

Одним из основных математических методов проверки нормальности распределения частот выборки является метод, основанный на вычислении коэффициентов асимметрии (А) и эксцесса (Е).

Коэффициент асимметрии (А) показывает степень отклонения распределения частот выборки от нормального распределения вправо или влево.

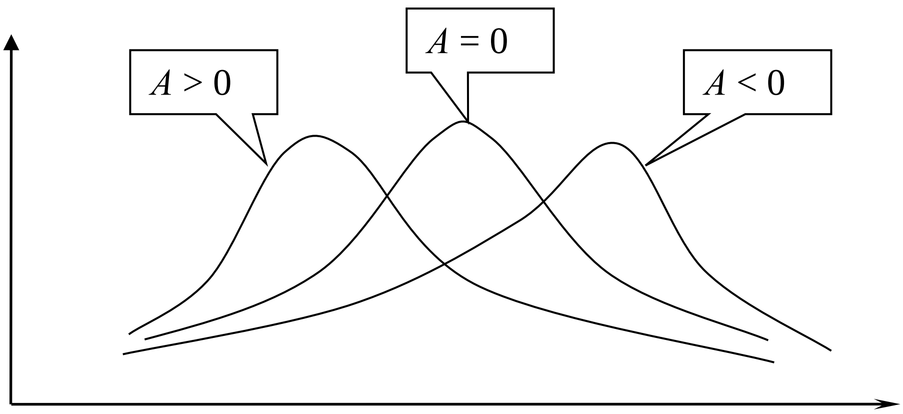


Рис. 13. Вид распределений частот с разными значениями асимметрии ($A = 0$ – нормальное распределение)

Коэффициент эксцесса (E) показывает степень отклонения распределения частот выборки от нормального распределения вверх или вниз.

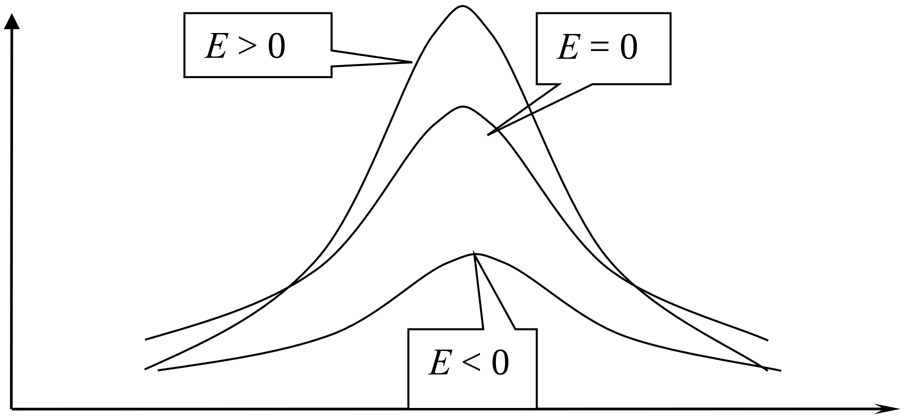


Рис. 14. Вид распределений частот с разными значениями эксцесса (E = 0 – нормальное распределение)

Алгоритм вычисления коэффициентов асимметрии и эксцесса (Excel)

1. К таблице вариант выборки добавьте строки «Асимметрия» и «Эксцесс».

2. Курсор установите на первую ячейку строки «Асимметрия» и последовательно выполните операции:

– нажмите клавишу со знаком «f_x»;

– в появившемся окне «Мастер функций» в ячейке «Поиск функции» наберите СКОС. Нажмите кнопку «Найти»;

– в окне «Мастер функций» нажмите «ОК»;

– в появившемся окне «Аргументы функции» введите коды ячеек вариант выборки (код первой ячейки выборки : код последней ячейки выборки). Нажмите «ОК»;

– в ячейке строки «Асимметрия» появится значение коэффициента асимметрии выборки;

– в строке « f_x» появится запись: = СКОС (код первой ячейки выборки : код последней ячейки выборки).

3. Перетащите курсором черный квадратик в остальные пустые ячейки строки «Асимметрия».

4. Курсор установите на первую ячейку строки «Эксцесс» и последовательно выполните операции:

– нажмите клавишу со знаком «f_x»;

– в появившемся окне «Мастер функций» в ячейке «Поиск функции» наберите ЭКСЦЕСС. Нажмите кнопку «Найти»;

– в окне «Мастер функций» нажмите «ОК»;

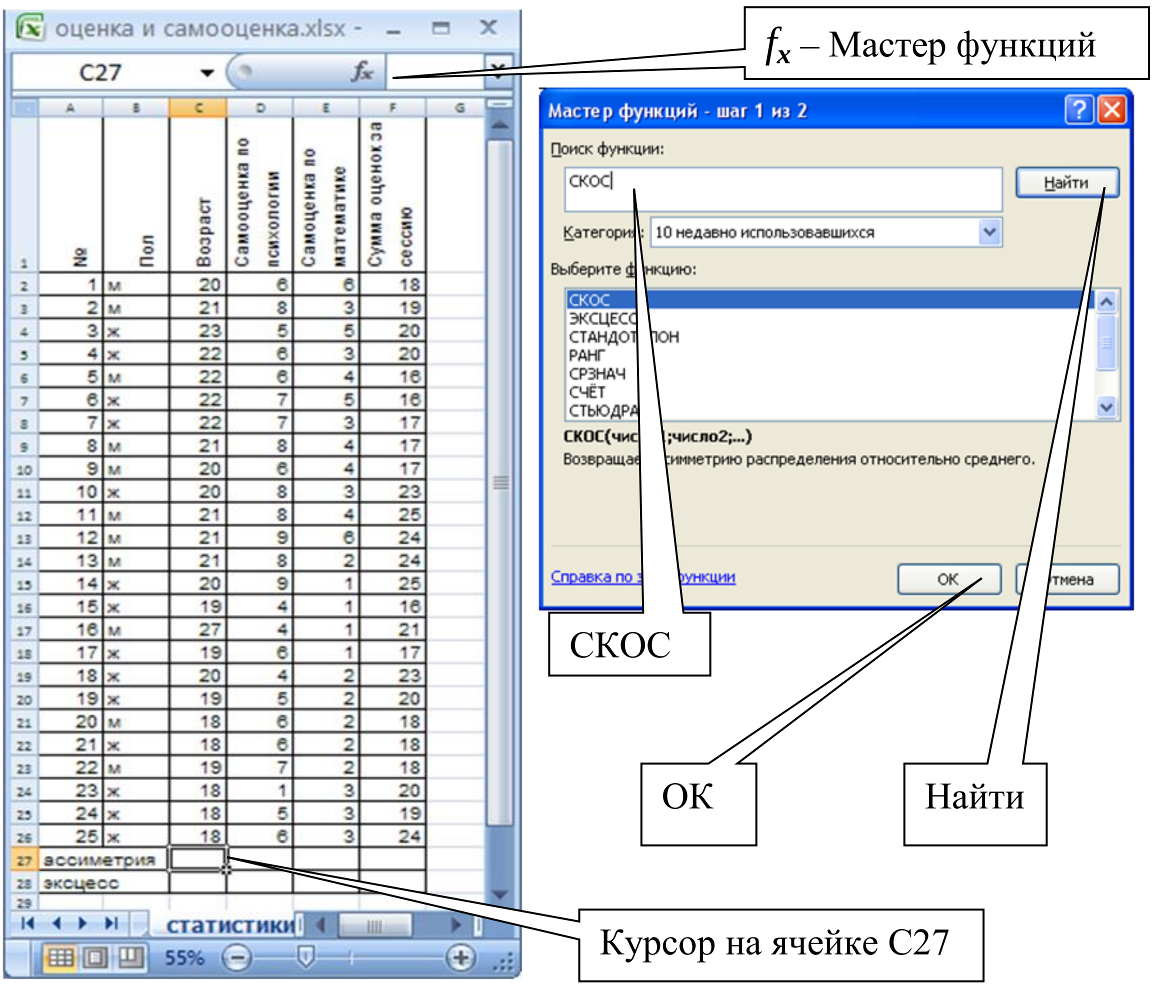
– в появившемся окне «Аргументы функции» введите коды ячеек вариант выборки (код первой ячейки выборки : код последней ячейки выборки). Нажмите «ОК»;

– в ячейке строки «Эксцесс» появится значение коэффициента эксцесса выборки;

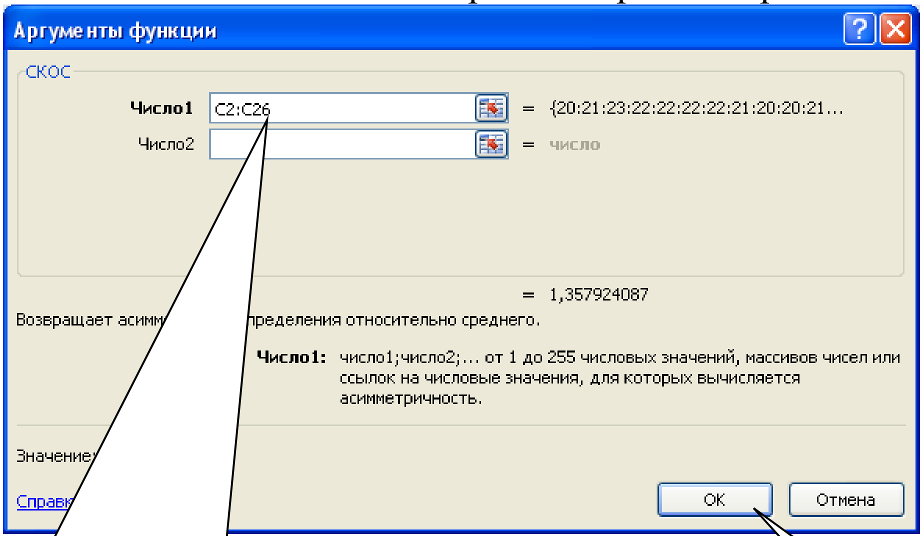
– в строке «f_x» появится запись: = ЭКСЦЕСС (код первой ячейки выборки : код последней ячейки выборки).

5. Перетащите курсором черный квадратик в остальные пустые ячейки строки «Эксцесс».

Пример вычисления коэффициентов асимметрии и эксцесса выборок исследования «Оценка – самооценка» (Excel)

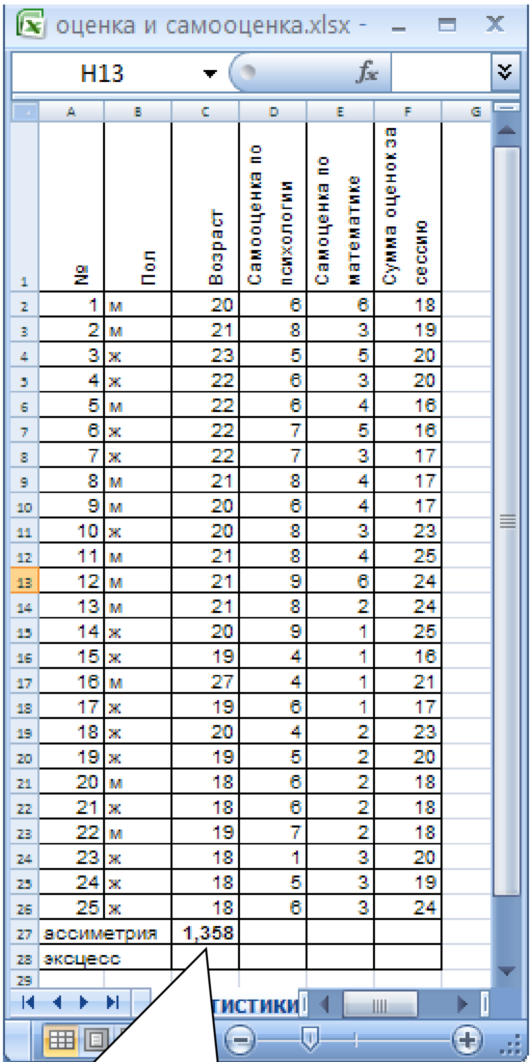


Начнем с вычисления асимметрии выборки «Возраст».

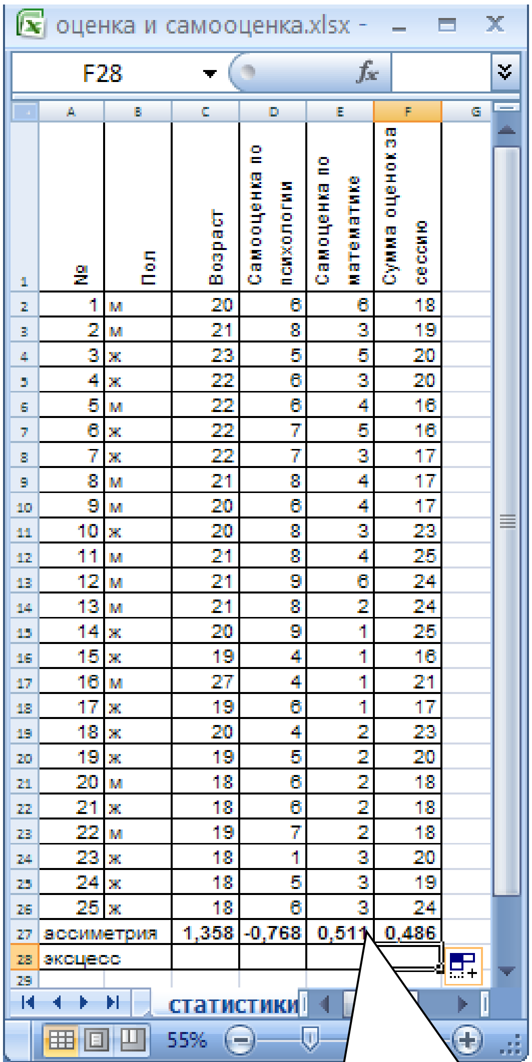


C2:C26 – коды вариант выборки «Возраст»

ОК



	A	B	C	D	E	F
1	№	Пол	Возраст	Самооценка по психологии	Самооценка по математике	Сумма оценок за сессию
2	1	м	20	6	6	18
3	2	м	21	8	3	19
4	3	ж	23	5	5	20
5	4	ж	22	6	3	20
6	5	м	22	6	4	16
7	6	ж	22	7	5	16
8	7	ж	22	7	3	17
9	8	м	21	8	4	17
10	9	м	20	6	4	17
11	10	ж	20	8	3	23
12	11	м	21	8	4	25
13	12	м	21	9	6	24
14	13	м	21	8	2	24
15	14	ж	20	9	1	25
16	15	ж	19	4	1	16
17	16	м	27	4	1	21
18	17	ж	19	6	1	17
19	18	ж	20	4	2	23
20	19	ж	19	5	2	20
21	20	м	18	6	2	18
22	21	ж	18	6	2	18
23	22	м	19	7	2	18
24	23	ж	18	1	3	20
25	24	ж	18	5	3	19
26	25	ж	18	6	3	24
27	ассиметрия		1,358			
28	эксцесс					

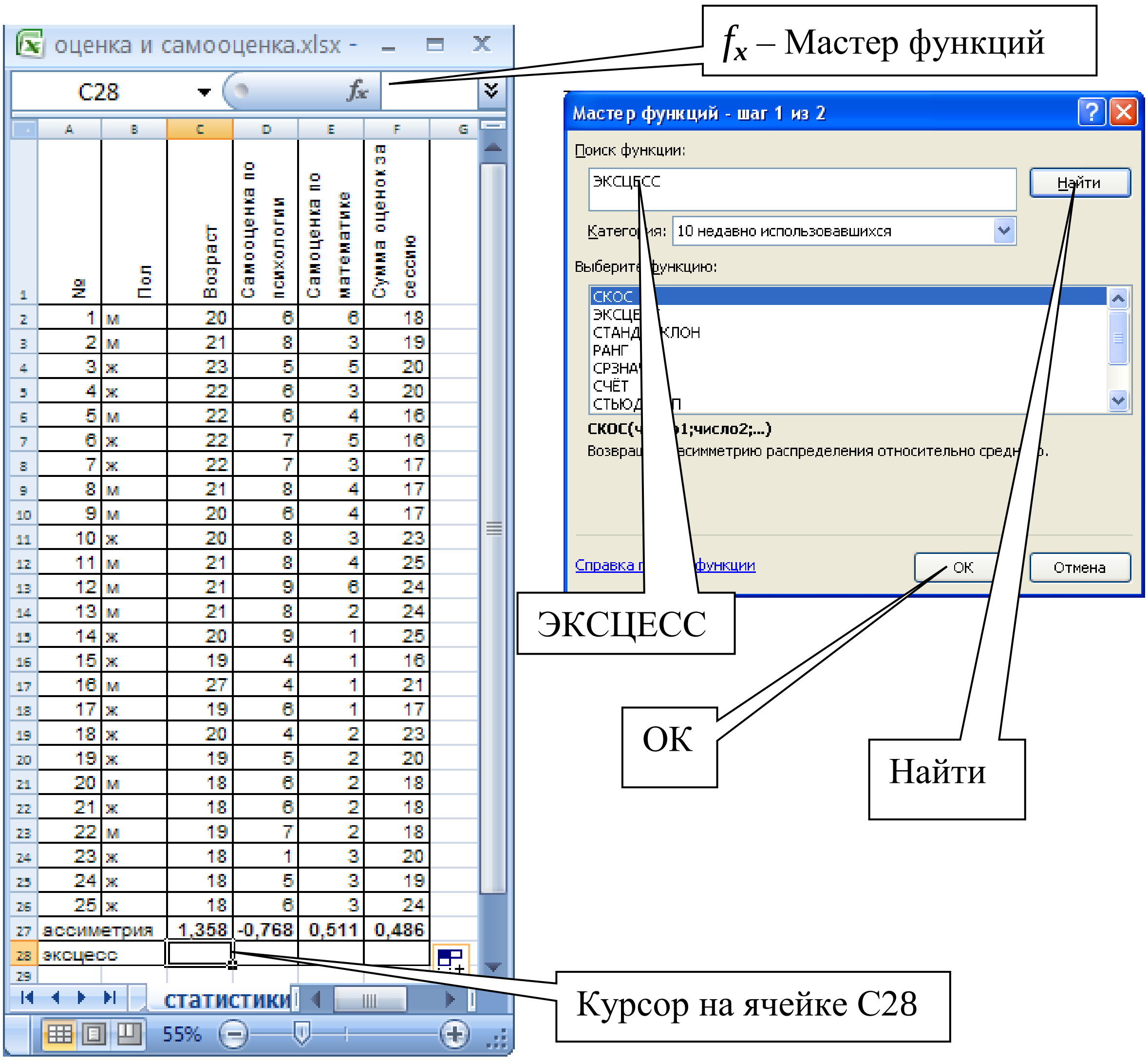


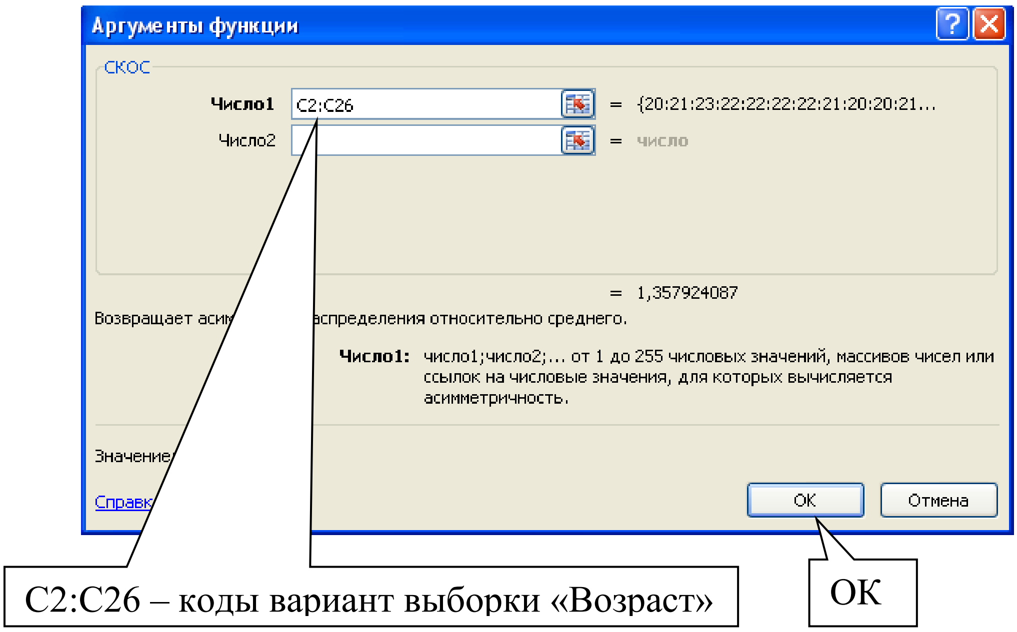
	A	B	C	D	E	F
1	№	Пол	Возраст	Самооценка по психологии	Самооценка по математике	Сумма оценок за сессию
2	1	м	20	6	6	18
3	2	м	21	8	3	19
4	3	ж	23	5	5	20
5	4	ж	22	6	3	20
6	5	м	22	6	4	16
7	6	ж	22	7	5	16
8	7	ж	22	7	3	17
9	8	м	21	8	4	17
10	9	м	20	6	4	17
11	10	ж	20	8	3	23
12	11	м	21	8	4	25
13	12	м	21	9	6	24
14	13	м	21	8	2	24
15	14	ж	20	9	1	25
16	15	ж	19	4	1	16
17	16	м	27	4	1	21
18	17	ж	19	6	1	17
19	18	ж	20	4	2	23
20	19	ж	19	5	2	20
21	20	м	18	6	2	18
22	21	ж	18	6	2	18
23	22	м	19	7	2	18
24	23	ж	18	1	3	20
25	24	ж	18	5	3	19
26	25	ж	18	6	3	24
27	ассиметрия		1,358	-0,768	0,511	0,486
28	эксцесс					

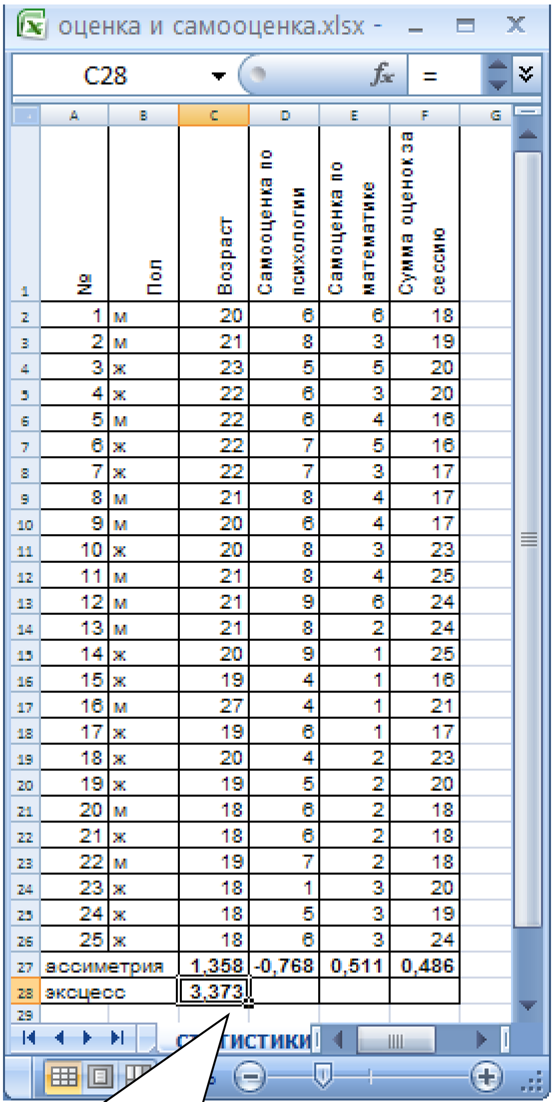
Асимметрия выборки «Возраст»

Коэффициенты асимметрии

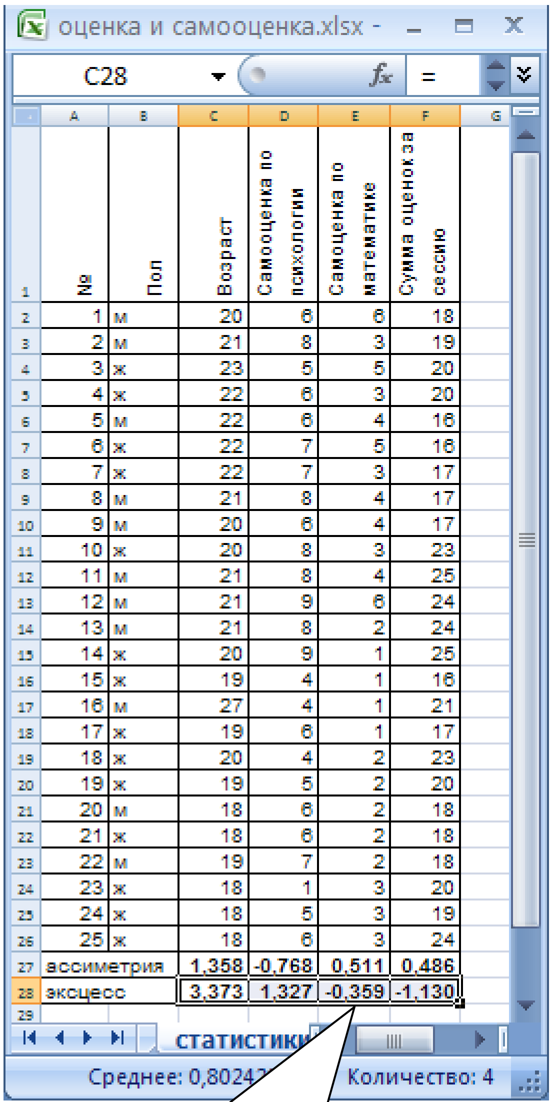
Вычислим эксцесс выборки «Возраст».







	A	B	C	D	E	F
1	№	Пол	Возраст	Самооценка по психологии	Самооценка по математике	Сумма оценок за сессию
2	1	м	20	6	6	18
3	2	м	21	8	3	19
4	3	ж	23	5	5	20
5	4	ж	22	6	3	20
6	5	м	22	6	4	16
7	6	ж	22	7	5	16
8	7	ж	22	7	3	17
9	8	м	21	8	4	17
10	9	м	20	6	4	17
11	10	ж	20	8	3	23
12	11	м	21	8	4	25
13	12	м	21	9	6	24
14	13	м	21	8	2	24
15	14	ж	20	9	1	25
16	15	ж	19	4	1	16
17	16	м	27	4	1	21
18	17	ж	19	6	1	17
19	18	ж	20	4	2	23
20	19	ж	19	5	2	20
21	20	м	18	6	2	18
22	21	ж	18	6	2	18
23	22	м	19	7	2	18
24	23	ж	18	1	3	20
25	24	ж	18	5	3	19
26	25	ж	18	6	3	24
27	ассиметрия		1,358	-0,768	0,511	0,486
28	эксцесс		3,373			



	A	B	C	D	E	F
27	ассиметрия		1,358	-0,768	0,511	0,486
28	эксцесс		3,373	1,327	-0,359	-1,130

Эксцесс выборки «Возраст»

Коэффициенты эксцесса

Проверку отклонения распределения частот выборки от нормального распределения (нормальности выборки) завершают сравнением вычисленных значений коэффициентов асимметрии и эксцесса выборок с критическими значениями этих коэффициентов. Используется таблица критических значений коэффициентов асимметрии и эксцесса (по Г.Ф. Лакину).

Если модули вычисленных значений коэффициентов асимметрии и эксцесса выборок одновременно не превосходят соответствующие критические значения, то принимается гипотеза H_0.

Если хотя бы один модуль вычисленных значений коэффициентов асимметрии и эксцесса выборок превосходит соответствующие критические значения, то принимается гипотеза H_1:

«Выборочное распределение статистически значимо отличается от нормального распределения ($p < 0,05$)».

Таблица критических значений коэффициентов асимметрии и эксцесса ($p < 0,05$)

Объем выборки	Асимметрия	Эксцесс
25	0,711	0,871
30	0,661	0,864
35	0,621	0,859
40	0,587	0,855
45	0,588	0,852
50	0,533	0,849
60	0,492	0,844
70	0,459	0,840
80	0,432	0,838
90	0,409	0,835
100	0,389	0,834
125	0,350	0,823
150	0,321	0,818
175	0,298	0,816
200	0,280	0,814
300	0,230	0,871
400	0,200	0,864
500	0,179	0,859

Примеры проверки гипотез о нормальности выборок

1. Проверка нормальности выборки «Возраст».

Коэффициент асимметрии равен 1,358.

Критическое значение коэффициента асимметрии равно 0,711.

Сравним коэффициенты асимметрии и эксцесса выборки «Возраст» с критическими значениями этих коэффициентов.

Статистический вывод.

Так как 0,711 < 1,358, то принимается гипотеза H_1 ($p \leq 0,05$).

Общий вывод.

Распределение частот выборки «Возраст» статистически значимо отличается от нормального распределения ($p \leq 0,05$).

2. Проверка нормальности выборки «Самооценка по математике».

Коэффициент асимметрии выборки «Самооценка по математике» равен 0,511.

Критическое значение коэффициента асимметрии выборки «Самооценка по математике» равно 0,711.

Сравним коэффициенты асимметрии выборки «Самооценка по математике» с критическими значениями этого коэффициента.

Статистический вывод.

Так как 0,511 < 0,711, то принимается гипотеза H_0.

Вывод об *асимметрии*.

Статистически значимых отличий коэффициентов асимметрии распределения частот выборки «Самооценка по математике» и нормального распределения нет.

Модуль коэффициента эксцесса выборки «Самооценка по математике» равен 0,359.

Критическое значение коэффициента эксцесса выборки «Самооценка по математике» равно 0,871.

Сравним коэффициенты эксцесса выборки «Самооценка по математике» с критическими значениями этого коэффициента.

Статистический вывод.

Так как 0,359 < 0,871, то принимается гипотеза H_0.

Вывод об *эксцессе*.

Статистически значимых отличий коэффициентов эксцесса распределения частот выборки «Самооценка по математике» и нормального распределения нет.

Общий вывод. Распределение частот выборки «Самооценка по математике» соответствует нормальному распределению.

6.5. Надежность оценки среднего генеральной совокупности

Чтобы определить значения параметров генеральной совокупности, среднего (μ) и стандартного отклонения (σ), находят статистики выборки, ее среднее (m) и стандартное отклонение (s).

Оценка параметров состоит в получении наиболее точных статистик. Оценки параметров бывают точечные и интервальные.

Точечная оценка – число, полученное по выборочным данным, значение которого очень близко к значению искомого параметра генеральной совокупности. Точечные оценки получают, как правило, на выборках, имеющих большой объем ($n > 200$).

Интервальная оценка – числовой интервал, которому с большой вероятностью, принадлежит значение искомого параметра генеральной совокупности. При выборке малого объема ($n < 30$), как правило, используют *интервальные* оценки, так как точечная оценка может значительно отличаться от значения параметра.

Точностью оценки параметра генеральной совокупности называют модуль разности статистики выборки и соответствующего параметра. Обозначим точность оценки символом δ (дельта).

Пример точности оценки среднего генеральной совокупности

Точность оценки среднего генеральной совокупности по среднему выборки: $\delta = | \mu - m |$.

То есть среднее генеральной совокупности наиболее вероятно принадлежит интервалу от $m - \delta$ до $m + \delta$.

Надежностью интервальной оценки среднего генеральной совокупности называют вероятность того, что значение параметра попадет в найденный интервал.

В психологии надежность *оценки* среднего генеральной совокупности принимают не менее 0,95. Найденный интервал называют доверительным интервалом для среднего генеральной совокупности со статистической надежностью ($p \geq 0{,}95$).

Алгоритм нахождения доверительного интервала для среднего генеральной совокупности

1. Вычисляют значения среднего (m) и стандартного отклонения (s) выборки.

2. Вычисляют значение δ для среднего генеральной совокупности по формуле

$$\delta = \frac{t_{\gamma} * s}{\sqrt{n}},$$

где t – случайная величина, значения которой зависят от n и находят в таблице значений t для нахождения доверительного интервала с надежностью (по В.Е. Гмурману);

s – стандартное отклонение выборки;

n – объем выборки.

Таблица значений *t* для нахождения доверительного интервала ($p \geq 0{,}95$)

n	t	n	t	n	t	n	t	n	t
5	2,78	11	2,23	17	2,12	35	2,032	80	1,991
6	2,57	12	2,20	18	2,11	40	2,023	90	1,987
7	2,45	13	2,18	19	2,10	45	2,016	100	1,984
8	2,37	14	2,16	20	2,093	50	2,009	120	1,980
9	2,31	15	2,15	25	2,064	60	2,001	∞	1,960
10	2,26	16	2,13	30	2,045	70	1,996		

Примечание: n – объем выборки, t – случайная величина.

Пример оценки среднего генеральной совокупности

По результатам теста по физике в группе из 25 учащихся среднее выборки равно 45 баллов (m), стандартное отклонение равно 3,1 балла (s).

Требуется найти доверительный интервал среднего генеральной совокупности со статистической надежностью ($p \geq 0{,}95$).

Найдем для n = 25 в таблице значение t = 2,064.

Известно, что s = 3.

Вычислим δ по формуле

$$\delta = \frac{t\ * s}{\sqrt{n}} = \frac{2{,}064 * 3{,}1}{\sqrt{25}} = 1{,}28.$$

Искомый доверительный интервал 45 ± 1,24 балла.

Значит, со статистической надежностью ($p \geq 0{,}95$) можно заключить, что среднее генеральной совокупности попадет в доверительный интервал от 43,72 до 46,28 балла.

Резюме

Выявление общих психологических особенностей для группы людей, образующих некоторую генеральную совокупность, осуществляется с помощью выборочного метода. Использование выборочного метода основано на выполнении условий, необходимых для его применения.

Во-первых, измерения должны быть проведены в шкалах интервалов или шкале отношений, то есть необходимо существование единицы измерения свойства исследуемого явления. Другими словами, значения исследуемого свойства должны быть непрерывной переменной.

Во-вторых, должны быть обеспечены оптимальная репрезентативность выборки, случайный отбор респондентов. Случайный отбор осуществляется с помощью различных методов, основными из которых являются: простой, ступенчатый, механический, типический ступенчатый случайный отбор. Заметим, что психологи комбинируют способы отбора с учетом целей и содержания исследований.

В-третьих, должна быть проведена проверка нормальности распределения частот выборки. Для проверки того, что распределение частот выборки является близким к нормальному распределению, используют коэффициенты асимметрии и эксцесса. Если распределение частот выборки мало отличается от нормального, то это будет необходимым условием для сопоставления статистик выборки с общепсихологическими параметрами генеральной совокупности.

И, наконец, должна быть установлена надежность оценки параметра генеральной совокупности.

Упражнения

6.1. Из студенческой группы (25 чел.) необходимо выбрать случайным образом четырех студентов. Проведите отбор методом:

а) простого случайного отбора;

б) механического случайного отбора.

6.2. Из 200 студентов, в том числе 150 студентов-психологов и 50 студентов-политологов, необходимо выбрать случайным образом 20 студентов. Проведите отбор методом типического случайного отбора.

6.3. Из 200 студентов необходимо отобрать случайным образом 20 студентов пропорционально специальности обучения и полу студентов. По направлению «Психология» обучаются 100 студентов женского и 50 студентов мужского пола. По направлению «Политология» обучаются 30 студентов женского и 20 студентов мужского пола. Проведите отбор методом ступенчатого случайного отбора.

6.4. В таблице приведены распределения частот оценок способностей студентов решать новые задачи.

Оценка способностей	1	2	3	4	5	6	7	8	9	10
Частота	2	1	11	27	31	21	20	15	14	3

Проведите проверку отклонения распределения частот выборки от нормального распределения (нормальности выборки).

6.5. Проведено исследование уровня агрессивности подростков. В группе из 28 подростков среднее выборки «Вербальная агрессия» равно 53 балла, стандартное отклонение – 5,6 балла.

Найдите доверительный интервал среднего генеральной совокупности со статистической надежностью ($p \geq 0,95$).

7. ЛИНЕЙНОЕ ШКАЛИРОВАНИЕ

Одним из методов анализа собранных исследователем выборочных эмпирических данных является линейное шкалирование. К основным типам задач, решаемых с помощью линейного шкалирования относятся:

– классификация выборки по уровням выраженности измеренного свойства;

– установление уровня выраженность этого свойства у отдельных респондентов;

– стандартизация разрабатываемого теста или опросника.

Число градаций в линейной шкале устанавливается, как правило, не больше десяти, причем число градаций может быть как четным, так нечетным.

Для определения границ градаций линейной шкалы используются количественные нормы, полученные из стандартизированных шкал, обычно из процентильных шкал или шкал, составленных из значений средних и стандартных отклонений.

Градациям линейной шкалы присваиваются либо качественные наименования (например, низкий уровень, средний уровень, высокий уровень), либо численные значения (например, 1, 2, 3, 4). В зависимости от числа градаций линейной шкалы ее называют n-балльной шкалой.

7.1. Построение шкалы с нечетным числом градаций

Из шкал с нечетным числом градаций чаще всего используются трехбалльные, пятибалльные и девятибалльные шкалы.

Градациям трёхбалльной шкалы в основном присваиваются качественные наименования, например:

– низкий уровень выраженности свойства;

– средний уровень выраженности свойства;

– высокий уровень выраженности свойства.

Градациям пятибалльной шкалы обычно присваиваются качественные наименования, например:

– очень низкий уровень выраженности свойства;

– низкий уровень выраженности свойства;

– средний уровень выраженности свойства;

– высокий уровень выраженности свойства;

– очень высокий уровень выраженности свойства.

Градациям девятибалльной шкалы чаще всего присваиваются численные наименования: 1, 2, 3, 4, 5, 6, 7, 8, 9. Девятибалльной также называется *шкала станайнов* (сокр. англ. *standard nine* – стандартная девятка).

Для построения линейной шкалы с нечетным числом градаций, как правило, используется процентильное распределение частот выборки (см. параграф 3.5).

Алгоритм построения шкалы с нечетным числом градаций

1. Составляется процентильное распределение частот выборки (*PCUM, %)*.

2. Выбирается число градаций шкалы и устанавливаются процентильные границы градаций.

Границы градаций трехбалльной шкалы: P23, P77.

Границы градаций пятибалльной шкалы: P7, P31, P69, P93.

Границы градаций девятибалльной шкалы: P4, P11, P23, P40, P60, P77, P89, P96.

Знаком *Pk* обозначен соответствующий процентиль.

Процентилем Pk называется такое значение варианты, левее (меньше) которой находится *k* процентов вариант выборки.

3. Составляется графическая модель шкалы.

Рис. 15. Модель трехбалльной шкалы

Рис. 16. Модель пятибалльной шкалы

Модель девятибалльной шкалы выглядит аналогично, только градации обозначены числами 1, 2, 3, 4, 5, 6, 7, 8, 9.

4. Проводится описание построенной шкалы.

Пример построения шкалы, содержащей три градации

В таблице представлено распределение частот оценок студентов, полученных с помощью методики «Мотивация одобрения».

Оценка, балл	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Частота	1	1	2	2	4	7	9	24	11	16	18	6	3	1

Требуется построить трехбалльную шкалу выраженности мотивации одобрения у студентов.

1. Составим процентильное распределение частот выборки.

Оценка, балл	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Частота	1	1	2	2	4	7	9	24	11	16	18	6	3	1
CUM	1	2	4	6	10	17	26	50	61	77	95	101	104	105
PCUM, %	1	2	4	6	10	16	25	48	58	73	90	9[illegible]	99	100

P23 = 6,7

23%

77%

P77 = 10,6

2. Для построения трехбалльной шкалы необходимо найти приближенные значения P23 и P77.

Отметим, что 23% находится между 16 и 25, значит, P23 больше 6 баллов, но меньше 7 баллов, приблизительно 6,7 баллов.

77% находится между 73 и 90, значит, P77 больше 10 баллов, но меньше 11 баллов, приблизительно 10,6 баллов.

3. Составим графическую модель шкалы в баллах.

Оценка, балл	1	2	3	4	5	6	**P23**	7	8	9	10	**P77**	11	12	13	14

4. Опишем построенную трехбалльную шкалу выраженности мотивации одобрения в баллах.

До 6 баллов – низкий уровень выраженности мотивации одобрения.

От 7 до 10 баллов – средний уровень выраженности мотивации одобрения.

От 11 баллов – высокий уровень выраженности мотивации одобрения.

7.1. Построение шкалы с четным числом градаций

Из шкал с четным числом градаций чаще всего используются четырехбалльные и десятибалльные шкалы.

Градациям четырехбалльной шкалы обычно присваиваются качественные наименования, например:

– низкий уровень выраженности свойства;
– умеренный уровень выраженности свойства;
– хороший уровень выраженности свойства;
– высокий уровень выраженности свойства.

Традиционная шкала оценки академической успеваемости обучающихся также четырехбалльная, с градациями:

– неудовлетворительно;
– удовлетворительно;
– хорошо;
– отлично.

Градациям десятибалльной шкалы в основном присваиваются численные наименования: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.

Десятибалльной также называется *шкала стенов* (сокр. англ. standard ten – стандартная десятка).

Для построения линейной шкалы с четным числом градаций, как правило, используются описательные статистики выборки: среднее и стандартное отклонение (см. параграфы 4.2 и 4.3).

Алгоритм построения шкалы с четным числом градаций

1. Находятся значения статистик выборки: среднего (m) и стандартного отклонения (s).

2. Выбирается число градаций шкалы и устанавливаются статистические границы градаций.

Границы для градаций четырехбалльной шкалы: $m - s$, m, $m + s$.

Границы градаций десятибалльной шкалы: $m - 2s$; $m - 1{,}5s$; $m - s$; $m - 0{,}5s$; m; $m + 0{,}5s$; $m + s$; $m + 1{,}5s$; $m + 2s$.

3. Составляется графическая модель шкалы.

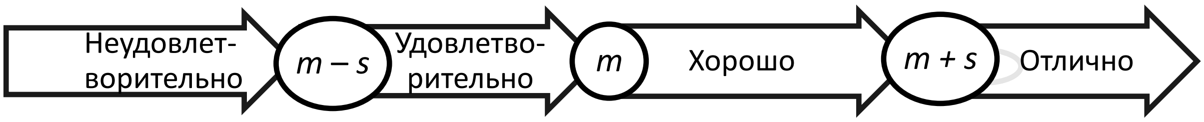


Рис. 17. Модель четырехбалльной шкалы

Модель десятибалльной шкалы выглядит аналогично, только градации обозначены числами: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.

4. Проводится описание построенной шкалы.

Пример построения шкалы, содержащей четыре градации

В таблице представлен протокол оценок студентов в баллах, полученных с помощью методики «Самооценка силы воли».

№	Оценка	№	Оценка	№	Оценка	№	Оценка	№	Оценка	№	Оценка	№	Оценка	№	Оценка
1	7	**11**	12	**21**	14	**31**	15	**41**	16	**51**	17	**61**	18	**71**	20
2	9	**12**	12	**22**	14	**32**	15	**42**	16	**52**	17	**62**	19	**72**	21
3	10	**13**	12	**23**	14	**33**	15	**43**	16	**53**	17	**63**	19	**73**	21
4	10	**14**	13	**24**	14	**34**	15	**44**	17	**54**	18	**64**	19	**74**	21
5	11	**15**	13	**25**	14	**35**	15	**45**	17	**55**	18	**65**	19	**75**	22
6	11	**16**	13	**26**	15	**36**	15	**46**	17	**56**	18	**66**	19	**76**	22
7	11	**17**	13	**27**	15	**37**	15	**47**	17	**57**	18	**67**	19	**77**	22
8	11	**18**	14	**28**	15	**38**	16	**48**	17	**58**	18	**68**	20	**78**	22
9	11	**19**	14	**29**	15	**39**	16	**49**	17	**59**	18	**69**	20	**79**	23
10	12	**20**	14	**30**	15	**40**	16	**50**	17	**60**	18	**70**	20	**80**	25

Требуется построить четырехбалльную шкалу выраженности самооценки силы воли у студентов.

1. Занесем полученные эмпирические оценки в протокол *Excel*.

Вычислим значения статистик выборки: среднего (m) и стандартного отклонения (s).

Имеем m = 16,1 балла, s = 3,5 балла.

2. Для построения четырехбалльной шкалы необходимо найти значения: $m – s$, m, $m + s$. После вычислений имеем:

$m – s$ = 12,6 баллов; m = 16,1 балла; $m – s$ = 19,6 балла.

3. Составим графическую модель шкалы в баллах.

Оценка	7	…	12	**12,6**	13	…16	**16,1**	17	…	19	**19,6**	20	…	25

4. Опишем построенную четырехбалльную шкалу выраженности самооценки силы воли в баллах.

До 12 баллов – низкий уровень выраженности самооценки силы воли.

От 13 до 16 баллов – умеренный уровень выраженности самооценки силы воли.

От 17 до 19 баллов – хороший уровень выраженности самооценки силы воли.

От 20 баллов – высокий уровень выраженности самооценки силы воли.

Резюме

Метод линейного шкалирования предъявляет различные требования к условиям задач, решаемым с его помощью. Эти требования прежде всего относятся к характеристикам выборки.

Для классификации выборки по уровням выраженности измеренного свойства выборка по объему может быть малой выборкой (менее 30 респондентов).

Для задачи установления уровня выраженность этого свойства у отдельных респондентов выборка по объему может быть малой выборкой, а также однородной по составу респондентов (по полу и возрасту).

Для решения задачи стандартизации разрабатываемого теста или опросника, в том числе построения линейной шкалы исследуемого свойства, необходимо получить выборку с оптимальными репрезентативностью и надежностью, имеющую распределение частот, близкое к нормальному (см. главу 6).

Упражнения

7.1. В таблице представлен протокол оценок студентов, полученных с помощью методики «Мотивация избегания неудач».

№	Оценка	№	Оценка	№	Оценка	№	Оценка	№	Оценка
1	5	**11**	13	**21**	16	**31**	18	**41**	19
2	5	**12**	13	**22**	16	**32**	18	**42**	19
3	9	**13**	13	**23**	17	**33**	18	**43**	20
4	9	**14**	13	**24**	17	**34**	18	**44**	20
5	10	**15**	14	**25**	17	**35**	19	**45**	20
6	10	**16**	14	**26**	17	**36**	19	**46**	23
7	11	**17**	14	**27**	18	**37**	19	**47**	23
8	12	**18**	15	**28**	18	**38**	19	**48**	24
9	12	**19**	15	**29**	18	**39**	19	**49**	24
10	12	**20**	15	**30**	18	**40**	19	**50**	24

Занесите полученные эмпирические оценки в протокол *Excel.*

Постройте трехбалльную шкалу выраженности мотивации избегания неудач у студентов.

7.2. В таблице представлен протокол оценок студентов, полученных с помощью методики «Готовность к риску».

№	Оценка	№	Оценка	№	Оценка	№	Оценка	№	Оценка
1	7	**11**	29	**21**	35	**31**	47	**41**	56
2	13	**12**	29	**22**	35	**32**	47	**42**	56
3	14	**13**	32	**23**	36	**33**	49	**43**	59
4	15	**14**	32	**24**	38	**34**	50	**44**	61
5	16	**15**	34	**25**	39	**35**	50	**45**	61
6	20	**16**	34	**26**	43	**36**	54	**46**	67
7	25	**17**	35	**27**	44	**37**	54	**47**	68
8	26	**18**	35	**28**	44	**38**	54	**48**	79
9	29	**19**	35	**29**	46	**39**	55	**49**	79
10	29	**20**	35	**30**	46	**40**	55	**50**	79

Занесите полученные эмпирические оценки в протокол *Excel*.

Постройте четырехбалльную шкалу выраженности готовности к риску у студентов.

8. ПАРАМЕТРИЧЕСКИЕ КРИТЕРИИ

С помощью параметрических критериев проверяются статистические гипотезы об отличиях средних, стандартных отклонений (дисперсий) выборок и генеральных совокупностей.

Особенностями параметрических критериев (в отличие от непараметрических) являются:

– измерения переменных, проведенные в интервальной шкале или шкале отношений;

– использование основных статистик (среднего и стандартного отклонения);

– оптимальная для генеральной совокупности репрезентативность выборки;

– соответствие распределений частот выборки нормальному распределению.

Основными параметрическими критериями, применяемыми в психологии, являются *T*-критерий Стьюдента и *F*-критерий Фишера.

8.1. *T*-КРИТЕРИЙ СТЬЮДЕНТА ДЛЯ СВЯЗАННЫХ ВЫБОРОК

Связанные выборки – множества значений двух свойств, полученные в одной группе респондентов.

Вопросы, ответы на которые можно найти с помощью T-критерия Стьюдента для связанных выборок

1. Есть ли статистически значимые различия средних выборок тестирования школьников по математике и по физике?
2. Есть ли статистически значимые различия средних генеральных совокупностей уровней общительности студентов до и после дискотеки?

Алгоритм *T*-критерия Стьюдента для связанных выборок (Excel)

1. Разместите таблицу выборок *A* и *B* в Excel и добавьте столбец «A – B». Последовательно выполните операции:

– в первой ячейке столбца выборки «А – В» наберите формулу: =B2-C2;

– перетащите ячейку «D2» в ячейки столбца «А – В»;

2. Найдите с помощью функции СРЗНАЧ в Excel среднее (m) столбца выборки «А – В».

3. Найдите с помощью функции СТАНДОТКЛОН в Excel стандартное отклонение (s) столбца выборки «А – В».

4. Вычислите значение T по формуле

$$T = \frac{|m|\sqrt{n}}{s},$$

где m – среднее столбца выборки «А – В»;

n – объем выборки «А – В»;

s – стандартное отклонение столбца выборки «А – В».

5. Найдите с помощью функции СТЬЮДРАСП в Excel уровень значимости вывода (α).

В окно «Аргументы функции» вписывают:

– в строку «X» – значение T;

– в строку «Степени свободы» – число, равное $n - 1$;

– в строку «Хвосты» – число 2.

6. Статистический вывод.

Если $\alpha > 0{,}10$, то принимается гипотеза H_0.

Если $\alpha \leq 0{,}10$, то принимается гипотеза H_1 ($p \leq \alpha$).

Пример использования Т-критерия Стьюдента для связанных выборок (Excel)

В протоколе приведены результаты измерения времени, затраченного студентами на чтение текстов, написанных шрифтом **Times** (А) и шрифтом **`Courier`** (В).

Имя	А	В
Алексей	35	45
Борис	37	45
Леонид	43	45
Марина	43	45
Мария	50	50
Михаил	55	50
Настя	47	40

Имя	А	В
Николай	37	46
Олег	35	40
Ольга	45	40
Петр	65	50
Рита	60	45
Роман	55	50
Степан	65	60

Требуется установить уровень статистической значимости различий средних значений времени, затраченного респондентами на чтение текстов, написанных шрифтом **Times** (А) и шрифтом **`Courier`** (В).

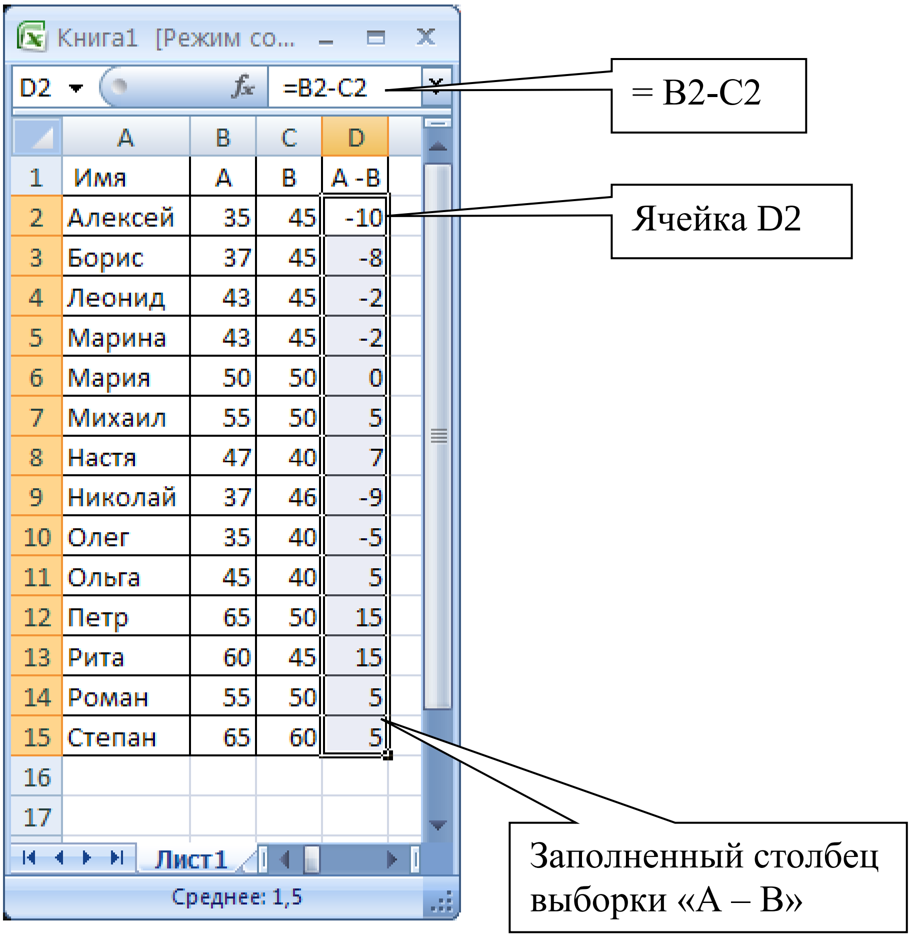
Перенесем таблицу результатов исследования в Excel.

Добавим столбец «А – В».

Выполним операции:

– в первой ячейке столбца «А – В» наберем формулу: =B2-C2;

– перетащим ячейку «D2» в пустые ячейки столбца «А – В».



	A	B	C	D
1	Имя	А	В	А -В
2	Алексей	35	45	-10
3	Борис	37	45	-8
4	Леонид	43	45	-2
5	Марина	43	45	-2
6	Мария	50	50	0
7	Михаил	55	50	5
8	Настя	47	40	7
9	Николай	37	46	-9
10	Олег	35	40	-5
11	Ольга	45	40	5
12	Петр	65	50	15
13	Рита	60	45	15
14	Роман	55	50	5
15	Степан	65	60	5
16				
17				

Найдем с помощью функции СРЗНАЧ среднее (m) столбца выборки «А – В».

Найдем с помощью функции СТАНДОТКЛОН стандартное отклонение (s) столбца выборки «А – В».

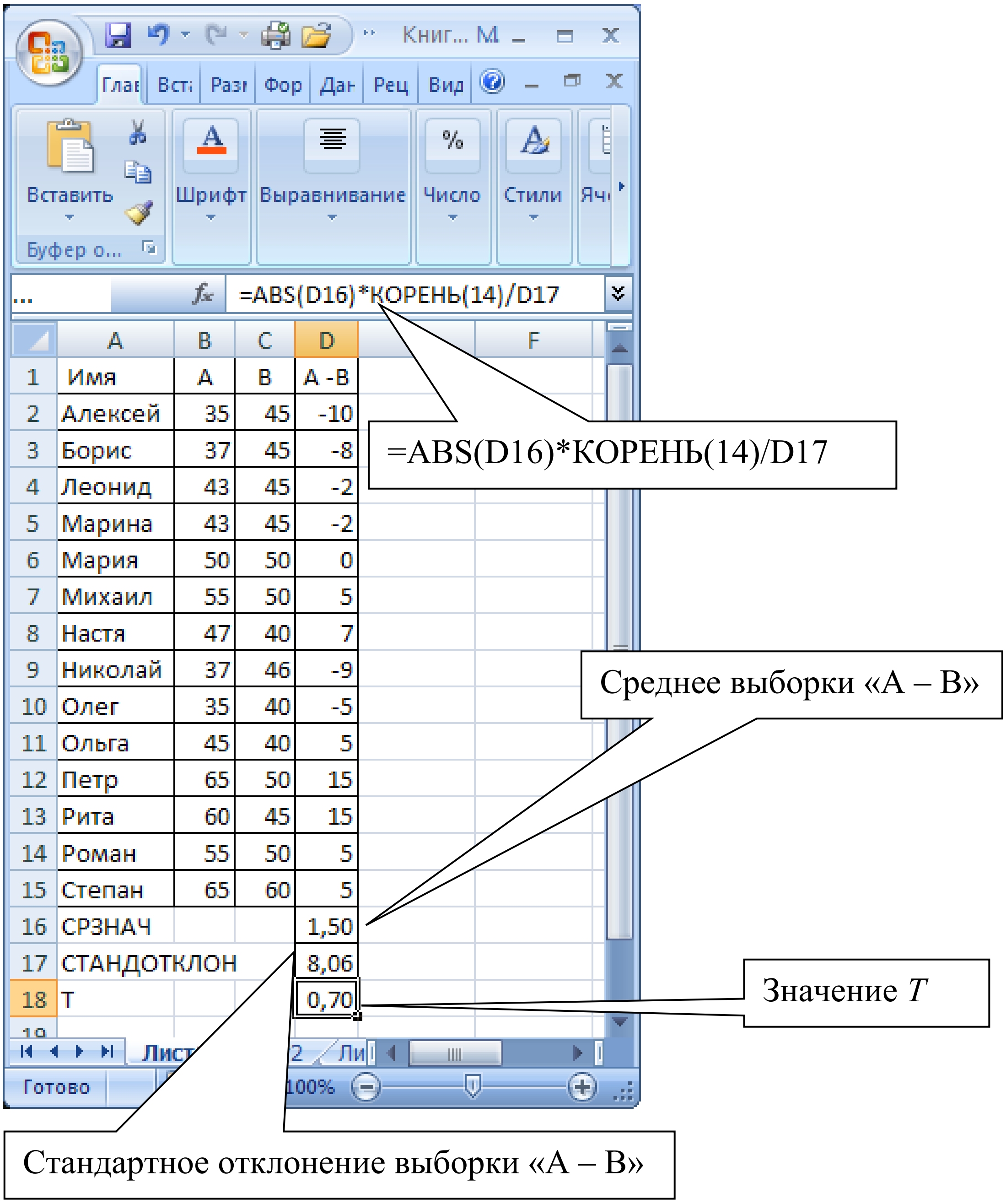
Вычислим значение T по формуле

$$T = \frac{|m|\sqrt{n}}{s},$$

где m – среднее столбца выборки «А – В»;

n – объем выборки «А – В»;

s – стандартное отклонение столбца выборки «А – В».



$T = 0{,}70$.

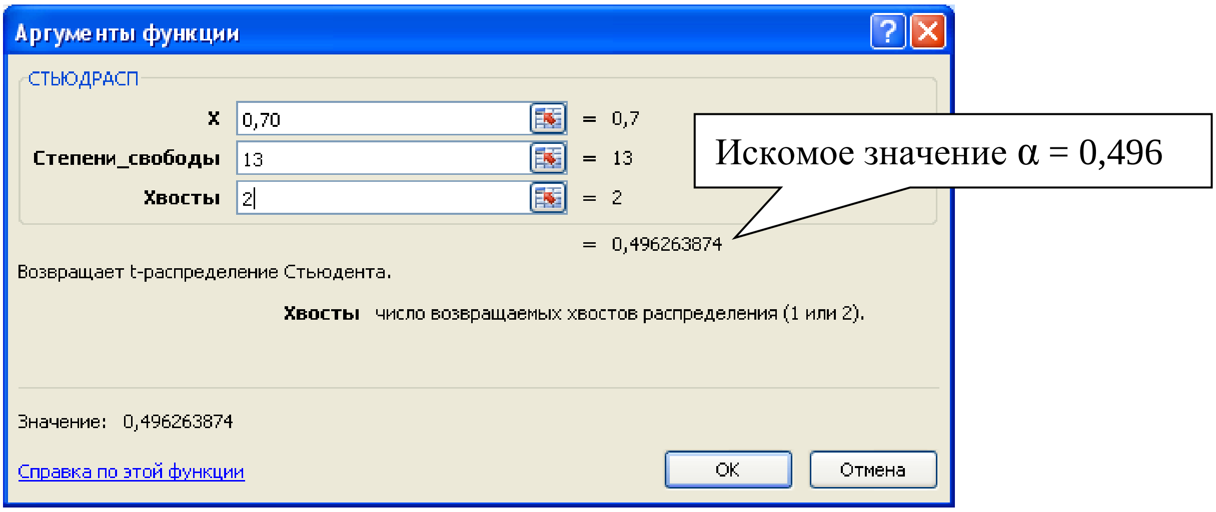
Найдем с помощью функции СТЬЮДРАСП уровень значимости вывода (α).

В окно «Аргументы функции» впишем значения:

– в строку «X» – значение T =0,70;

– в строку «Степени свободы» – число, равное $n - 1 = 13$;

– в строку «Хвосты» – число 2.



α = 0,496.

Статистический вывод.

Так как 0,496 > 0,10, то есть α > 0,10, значит, принимается гипотеза H_0.

Содержательный вывод.

Нет статистически значимых различий средних значений времени, затраченного респондентами на чтение текстов, написанных шрифтом **Times** (A) и шрифтом **`Courier`** (B).

8.2. *T*-КРИТЕРИЙ СТЬЮДЕНТА ДЛЯ НЕСВЯЗАННЫХ ВЫБОРОК

Несвязанные выборки – множества значений одного свойства, полученные в двух группах респондентов.

Вопросы, ответы на которые можно найти с помощью Т-критерия Стьюдента для несвязанных выборок

1. Есть ли статистически значимые различия средних выборок тестирования мальчиков и девочек по математике?
2. Есть ли статистически значимые различия средних генеральных совокупностей уровней общительности студентов очной и заочной форм обучения?

Алгоритм *T*-критерия Стьюдента для несвязанных выборок (Excel)

1. Разместите таблицу выборок I и II в Excel.
2. Найдите с помощью функции СРЗНАЧ в Excel:

– среднее (m_1) столбца выборки I;

– среднее (m_2) столбца выборки II.

3. Найдите с помощью функции СТАНДОТКЛОН в Excel:

– стандартное отклонение (s_1) столбца выборки I;
– стандартное отклонение (s_2) столбца выборки II.

4. Вычислите значение T по формуле

$$T = \frac{|\, m_1 - m_2 \,|}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_1^2}{n_2}}},$$

где m_1 – среднее столбца выборки I;
m_2 – среднее столбца выборки II;
s_1 – стандартное отклонение столбца выборки I;
n_1 – объем выборки I;
s_2 – стандартное отклонение столбца выборки II;
n_2 – объем выборки II.

5. Найдите с помощью функции СТЬЮДРАСП в Excel уровень значимости вывода (α).

В окно «Аргументы функции» вписывают:
– в строку «X» – значение T;
– в строку «Степени свободы» – число, равное $n_1 + n_2 – 2$;
– в строку «Хвосты» – число 2.

6. Статистический вывод.

Если α $>$ 0,10, то принимается гипотеза H_0.

Если α $\leq$ 0,10, то принимается гипотеза H_1 ($p \leq α$).

Пример использования T-критерия Стьюдента для несвязанных выборок (Excel)

В протоколе приведены результаты измерения самооценок по математике студентов очной (I) и заочной (II) форм обучения.

№	Форма	Оценка	№	Форма	Оценка	№	Форма	Оценка
1	ОФО	6	10	ОФО	6	19	ЗФО	1
2	ОФО	6	11	ОФО	6	20	ЗФО	2
3	ОФО	7	12	ОФО	7	21	ЗФО	4
4	ОФО	5	13	ОФО	7	22	ЗФО	1
5	ОФО	7	14	ОФО	8	23	ЗФО	3
6	ОФО	3	15	ЗФО	2	24	ЗФО	2
7	ОФО	3	16	ЗФО	5	25	ЗФО	7
8	ОФО	6	17	ЗФО	6	26	ЗФО	4
9	ОФО	8	18	ЗФО	9	27	ЗФО	1

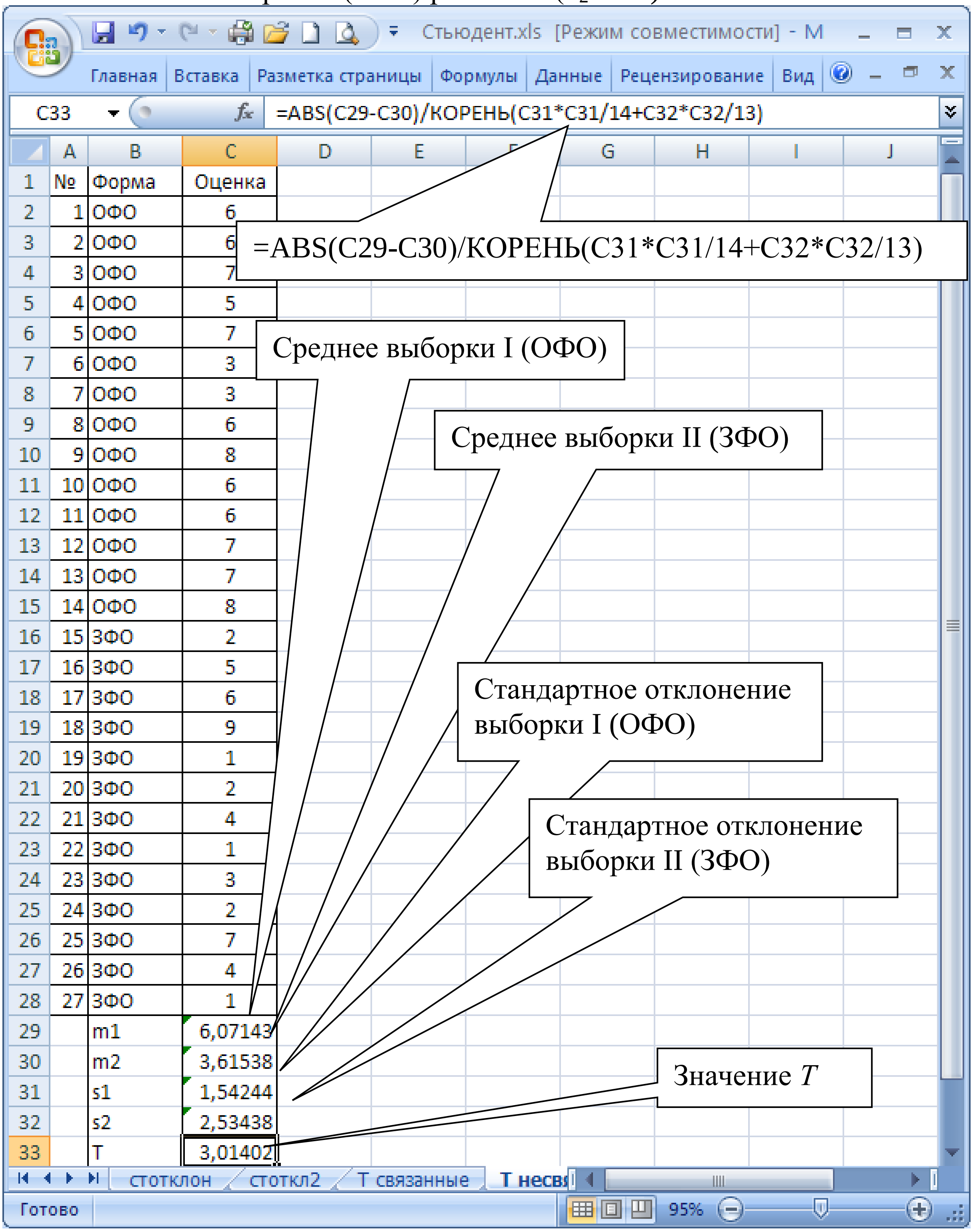
Требуется установить уровень статистической значимости различий средних значений самооценок по математике студентов очной и заочной форм обучения.

Перенесем таблицу результатов исследования в Excel.

Заметим, что:

– объем выборки I (ОФО) равен 14 ($n_1 = 14$);

– объем выборки II (ЗФО) равен 13 ($n_2 = 13$).



	A	B	C
1	№	Форма	Оценка
2	1	ОФО	6
3	2	ОФО	6
4	3	ОФО	7
5	4	ОФО	5
6	5	ОФО	7
7	6	ОФО	3
8	7	ОФО	3
9	8	ОФО	6
10	9	ОФО	8
11	10	ОФО	6
12	11	ОФО	6
13	12	ОФО	7
14	13	ОФО	7
15	14	ОФО	8
16	15	ЗФО	2
17	16	ЗФО	5
18	17	ЗФО	6
19	18	ЗФО	9
20	19	ЗФО	1
21	20	ЗФО	2
22	21	ЗФО	4
23	22	ЗФО	1
24	23	ЗФО	3
25	24	ЗФО	2
26	25	ЗФО	7
27	26	ЗФО	4
28	27	ЗФО	1
29		m1	6,07143
30		m2	3,61538
31		s1	1,54244
32		s2	2,53438
33		T	3,01402

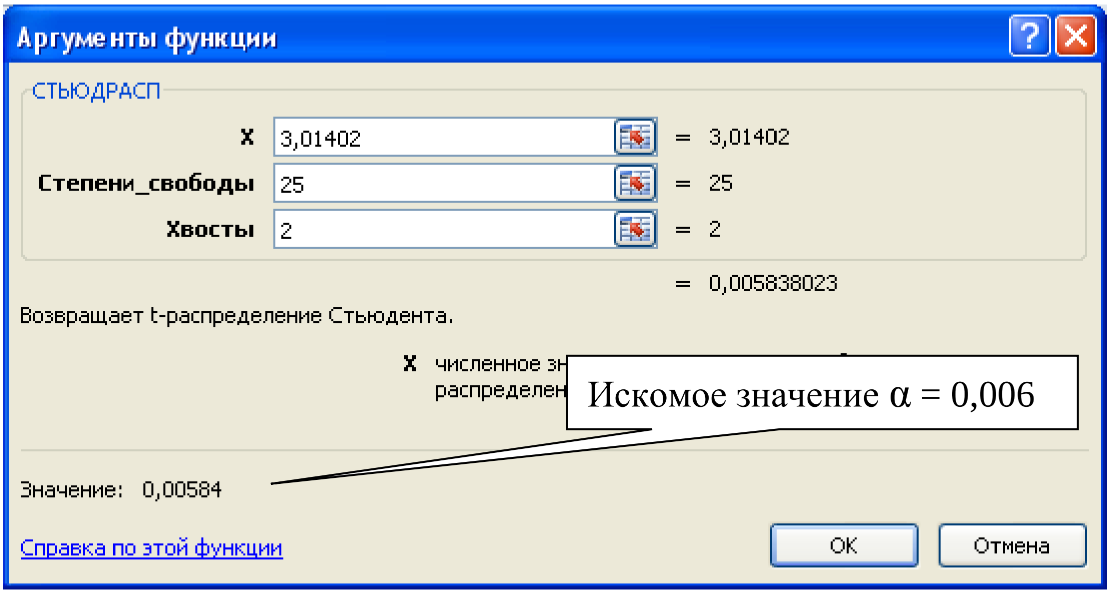
Найдем с помощью функции СРЗНАЧ:
– среднее (m_1) выборки I (ОФО);
– среднее (m_2) выборки II (ЗФО).
$m_1 = 6{,}07$; $m_2 = 3{,}62$.
Найдем с помощью функции СТАНДОТКЛОН:
– стандартное отклонение (s_1) выборки I (ОФО);
– стандартное отклонение (s_2) выборки II (ЗФО).
$s_1 = 1{,}54$; $s_2 = 2{,}53$.
Вычислим значение T по формуле

$$T = \frac{|m_1 - m_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_1^2}{n_2}}},$$

где m_1 – среднее столбца выборки I;
m_2 – среднее столбца выборки II;
s_1 – стандартное отклонение столбца выборки I;
n_1 – объем выборки I;
s_2 – стандартное отклонение столбца выборк II;
n_2 – объем выборки II.
$T = 3{,}01402$.

5. Найдем с помощью функции СТЬЮДРАСП в Excel уровень значимости вывода (α).

В окно «Аргументы функции» вписываем:
– в строку «X» – значение $T = 3{,}01402$;
– в строку «Степени свободы» – число, равное $n_1 + n_2 - 2 = 25$;
– в строку «Хвосты» – число 2.



Имеем $\alpha = 0{,}006$.

Статистический вывод.

Так как $0{,}006 < 0{,}10$, то принимается гипотеза H_1 ($p \le 0{,}006$).

Содержательный вывод.

Выявлены статистически значимые различия ($p \le 0{,}006$) средних самооценок по математике студентов очной и заочной форм обучения.

8.3. *F*-КРИТЕРИЙ ФИШЕРА

Сравнение стандартных отклонений (дисперсий) проводят с помощью F-критерия Фишера.

Вопросы, ответы на которые можно найти с помощью F-критерия Фишера

1. Есть ли статистически значимые различия стандартных отклонений (дисперсий) выборок тестирования школьников по математике и по физике?

2. Есть ли статистически значимые различия стандартных отклонений (дисперсий) генеральных совокупностей уровней общительности студентов до и после дискотеки?

Алгоритм F-критерия Фишера (Excel)

1. Разместите таблицу выборок I и II в Excel.

2. Найдите с помощью функции СТАНДОТКЛОН в Excel:

– стандартное отклонение (s_1) столбца выборки I;

– стандартное отклонение (s_2) столбца выборки II.

3. Вычислите значение F по формуле

$$F = \frac{s_1^2}{s_2^2}$$

где s_1 – большее стандартное отклонение выборок I и II;

s_2 – меньшее стандартное отклонение выборок I и II.

4. Найдите с помощью функции FРАСП в Excel уровень значимости вывода (α).

В окно «Аргументы функции» вписывают:

– в строку «X» – значение F;

– в строку «Степени свободы 1» – число, равное $n_1 - 1$;

– в строку «Степени свободы 2» – число, равное $n_2 - 1$.

5. Статистический вывод.

Если $\alpha > 0{,}10$, то принимается гипотеза H_0.

Если $\alpha \le 0{,}10$, то принимается гипотеза H_1 ($p \le \alpha$).

Пример использования F-критерия Фишера для сравнения стандартных отклонений (Excel)

В протоколе приведены результаты измерения самооценок по математике студентов очной (I) и заочной (II) форм обучения.

№	Форма	Оценка	№	Форма	Оценка	№	Форма	Оценка
1	ОФО	6	10	ОФО	6	19	ЗФО	1
2	ОФО	6	11	ОФО	6	20	ЗФО	2
3	ОФО	7	12	ОФО	7	21	ЗФО	4
4	ОФО	5	13	ОФО	7	22	ЗФО	1
5	ОФО	7	14	ОФО	8	23	ЗФО	3
6	ОФО	3	15	ЗФО	2	24	ЗФО	2
7	ОФО	3	16	ЗФО	5	25	ЗФО	7
8	ОФО	6	17	ЗФО	6	26	ЗФО	4
9	ОФО	8	18	ЗФО	9	27	ЗФО	1

Требуется установить уровень статистической значимости различий стандартных отклонений (дисперсий) значений самооценок студентов очной и заочной форм обучения.

Перенесем таблицу результатов исследования в Excel.

Заметим, что:

– объем выборки I (ОФО) равен 14 ($n_1 = 14$);

– объем выборки II (ЗФО) равен 13 ($n_2 = 13$).

Найдем с помощью функции СТАНДОТКЛОН:

– стандартное отклонение выборки I (ОФО), равное 2,53;

– стандартное отклонение выборки II (ЗФО), равное 1,89.

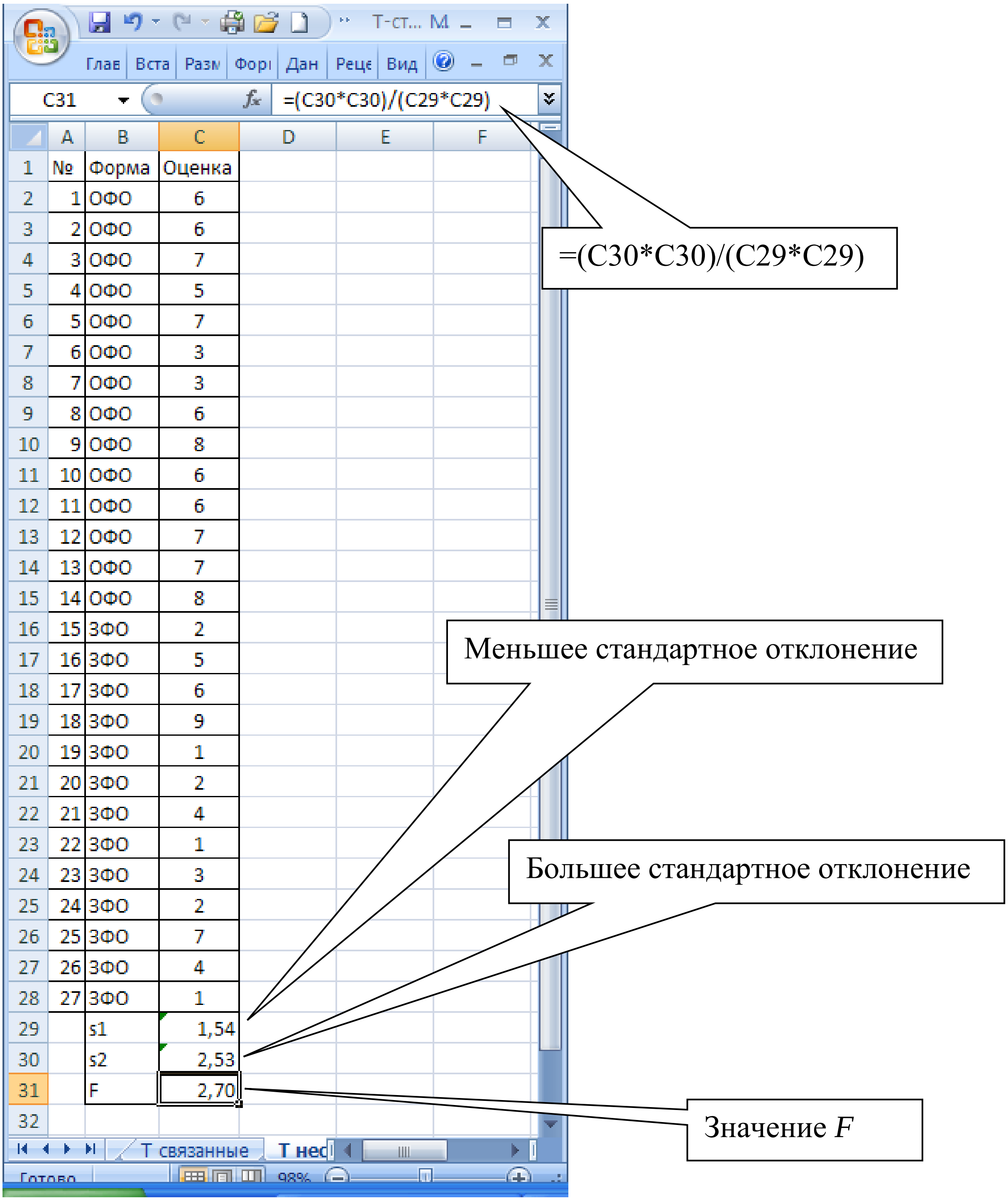
Обозначим:

– $s_1 = 2{,}53$ (большее стандартное отклонение);

– $s_2 = 1{,}89$ (меньшее стандартное отклонение).

Вычислим значение F по формуле

$$F = \frac{s_1^2}{s_2^2},$$



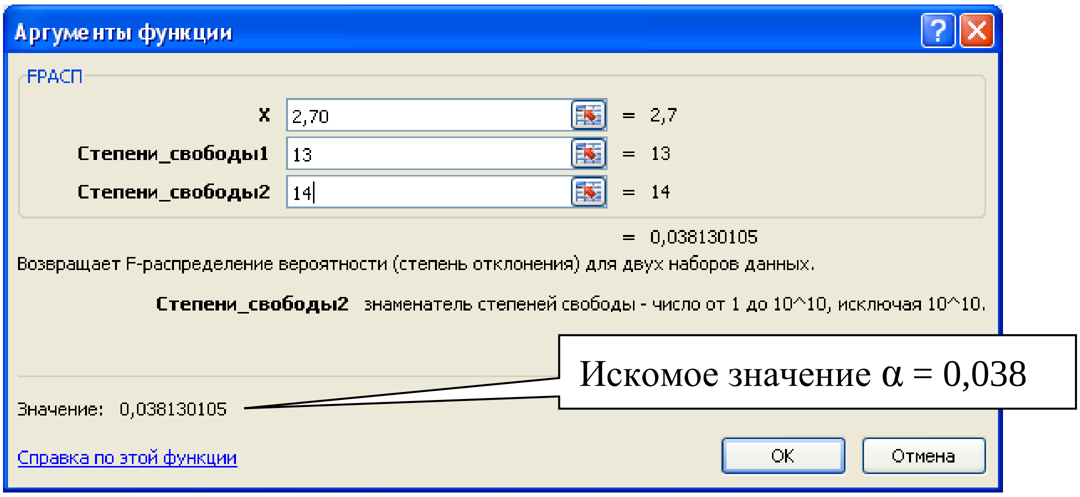
Найдем с помощью функции FРАСП в Excel уровень значимости вывода (α).

В окно «Аргументы функции» впишем:

– в строку «X» – значение F =2,70;

– в строку «Степени свободы 1» – число, равное $n_1 - 1 = 13$;

– в строку «Степени свободы 2» – число, равное $n_2 - 1 = 12$.



Имеем, $\alpha = 0{,}038$.

Статистический вывод.

Так как $0{,}038 < 0{,}10$, то принимается гипотеза H_1 ($p \leq 0{,}038$).

Содержательный вывод.

Выявлены статистически значимые различия ($p \leq 0{,}038$) стандартных отклонений самооценок по математике студентов очной и заочной форм обучения.

Резюме

Параметрическим критериям отдают предпочтение, если выборки получены из генеральной совокупности, имеющей нормальное распределение, в противном случае рекомендуется использовать непараметрические критерии.

Упражнения

7.1. В протоколе приведены результаты измерения времени, затраченного студентами на чтение текстов, написанных шрифтом **Times** (A) и шрифтом **`Courier`** (B).

Имя	A	B
Алексей	35	45
Борис	37	45
Леонид	43	45
Марина	43	45
Мария	50	50
Михаил	55	50
Настя	47	40

Имя	A	B
Николай	37	46
Олег	35	40
Ольга	45	40
Петр	65	50
Рита	60	45
Роман	55	50
Степан	65	60

Требуется установить уровень статистической значимости различий стандартных отклонений времени, затраченного респондентами на чтение текстов, написанных шрифтом **Times** (A) и шрифтом **`Courier`** (B).

7.2. В протоколе представлены результаты тестирования школьников по физике в баллах.

Ученик	Балл	Ученик	Балл	Ученик	Балл	Ученик	Балл
Алексей	60	Дмитрий	55	Леонид	46	Роман	75
Алена	55	Елена	61	Марина	86	Светлана	58
Андрей	55	Жанна	51	Мария	64	Сергей	50
Белла	31	Зина	48	Михаил	85	Стас	24
Борис	89	Игорь	92	Настя	63	Тарас	27
Вадим	69	Ирина	42	Николай	84	Татьяна	75
Вера	38	Катя	71	Олег	62	Ульяна	80
Галина	39	Клава	73	Ольга	77	Федор	49
Григорий	52	Костя	64	Петр	77	Юрий	24
Дина	54	Лариса	70	Рита	39	Яна	90

Выполните следующие задания:

а) установите уровень статистической значимости различий средних результатов тестирования школьников-мальчиков и школьников-девочек;

б) установите уровень статистической значимости различий стандартных отклонений результатов тестирования школьников-мальчиков и школьников-девочек.

7.3. В протоколе приведены самооценки студентов по психологии (А) и самооценки по математике (В).

Имя	А	В	Имя	А	В	Имя	А	В
Алексей	4	1	Рая	7	4	Николай	7	2
Борис	4	1	Роза	5	5	Олег	1	3
Вова	5	7	Тарас	8	6	Юля	8	7
Леонид	6	1	Света	5	5	Ольга	5	3
Марина	4	2	Сергей	7	5	Петр	6	3
Мария	5	2	Стас	7	6	Рита	6	3
Михаил	6	2	Толя	7	6	Роман	8	9
Настя	6	6	Таня	7	6	Таня	7	6
Юра	7	8	Федя	8	6	Вася	7	8

Выполните следующие задания:

а) установите уровень статистической значимости различий средних самооценок по психологии и самооценок по математике у студентов;

б) установите уровень статистической значимости различий стандартных отклонений самооценок по психологии и самооценок по математике у студентов.

9. ОТНОШЕНИЯ МЕЖДУ СВОЙСТВАМИ

Исследователи выделяют два основных вида отношений между свойствами: каузальное и корреляционное.

В восьмой главе представлены основные математические методы позволяющие проверить предположения о корреляционных отношениях между свойствами явлений.

9.1. Глоссарий отношений между свойствами

Корреляция

Корреляция (англ. correlation) – взаимосвязь, соответствие, взаимозависимость, связь.

Корреляционное отношение свойств

Корреляционным отношением свойств называют взаимную связь свойств.

Структура корреляционных отношений свойств A и B:

– свойство A связано со свойством B;

– свойство A взаимосвязано со свойством B;

– изменения свойств A и B происходит совместно и т.п.

$$\boxed{\text{Свойство } A} \Longleftrightarrow \boxed{\text{Свойство } B}$$

Рис. 18. Модель корреляционного отношения свойств A и B

Математическим методом выявления силы связей свойств является анализ корреляции.

Пример корреляционного отношения между свойствами

Предположение: способности школьников понимать учителя связаны с их способностями понятно выражать свои мысли.

Способности школьников понимать учителя – свойство A.

Способности школьников понятно выражать свои мысли – свойство B.

Свойство A связано со свойством B.

Отношение между свойствами A и B корреляционное.

Каузальность

Каузальность (лат. causa) – причинная обусловленность.

Каузальное отношение свойств

Каузальным отношением свойств называют причинно-следственную зависимость свойств.

Структура каузальных отношений свойств *A* и *B*:

– от свойства *A* зависит свойство *B*;

– если есть свойство *A*, то оно влияет на свойство *B* и т.п.

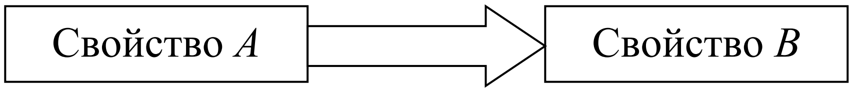


Рис. 19. Модель каузального отношения свойств *A* и *B*

Необходимым условием выявления каузального отношения свойств является теоретическое обоснование зависимости одного свойства от другого. Достаточным условием каузального отношения свойств является корреляционное отношение двух свойств.

Пример каузального отношения между свойствами

Предположение: от возраста ребенка зависит новообразование его психики.

Возраст ребенка – свойство *A*. Новообразование психики ребенка – свойство *B*. От *A* зависит *B*.

Отношение между свойствами *A* и *B* каузальное, где А – причина, *B* – следствие.

Коэффициент корреляции

Коэффициент корреляции (r; $p < \alpha$) – двумерная статистика об уровне связи (r) и уровне значимости связи ($p \leq \alpha$) между связанными переменными (свойствами).

Пример коэффициента корреляции

Способности школьников понимать учителя статистически значимо ($r = 0{,}56$; $p < 0{,}05$) связаны с их способностями понятно выражать свои мысли.

Свойства уровня связи переменных (r)

1. Уровень связи (r) вычисляется для связанных выборок.
$|\,r\,| \leq 1$.

2. Если r положительное число, то связь свойств прямая, то есть большему значению одного свойства соответствует большее значение другого.

3. Если r отрицательное число, то связь свойств обратная, то есть большему значению одного свойства соответствует меньшее значение другого.

4. Если r близко к нулю, то связь свойств отсутствует.

Для вычисленного значения r устанавливается уровень его статистической значимости.

Для определения уровня значимости связи переменных используется таблица критических значений коэффициентов корреляции (по А.Д. Наследову).

Таблица *r*-критерия Спирмена (*r*-критерия Пирсона)

n	α			n	α			n	α		
	0,10	0,05	0,01		0,10	0,05	0,01		0,10	0,05	0,01
5	0,805	0,878	0,959	20	0,378	0,444	0,561	35	0,283	0,334	0,430
6	0,729	0,811	0,917	21	0,369	0,433	0,549	36	0,279	0,329	0,424
7	0,669	0,754	0,875	22	0,360	0,423	0,537	37	0,275	0,325	0,418
8	0,621	0,707	0,834	23	0,352	0,413	0,526	38	0,271	0,320	0,413
9	0,582	0,666	0,798	24	0,344	0,404	0,515	39	0,267	0,316	0,408
10	0,549	0,632	0,765	25	0,337	0,396	0,505	40	0,264	0,312	0,403
11	0,521	0,602	0,735	26	0,330	0,338	0,496	42	0,257	0,304	0,393
12	0,497	0,576	0,708	27	0,323	0,381	0,487	44	0,251	0,297	0,384
13	0,476	0,553	0,684	28	0,317	0,374	0,479	46	0,246	0,291	0,374
14	0,458	0,532	0,661	29	0,311	0,367	0,471	50	0,235	0,279	0,361
15	0,441	0,514	0,641	30	0,306	0,361	0,463	60	0,214	0,254	0,330
16	0,426	0,497	0,623	31	0,301	0,355	0,562	70	0,198	0,235	0,306
17	0,412	0,482	0,606	32	0,296	0,349	0,449	80	0,185	0,220	0,286
18	0,400	0,468	0,590	33	0,291	0,344	0,442	90	0,174	0,207	0,270
19	0,389	0,456	0,575	34	0,287	0,339	0,436	100	0,165	0,197	0,256

Примечание: n – объем выборки; α – уровень значимости.

Шкала уровней связи переменных (r)

$0{,}70 \le r \le 1$ – сильная прямая корреляция;
$0{,}50 \le r < 0{,}70$ – средняя прямая корреляция;
$0{,}30 \le r < 0{,}50$ – умеренная прямая корреляция;
$0{,}20 \le r < 0{,}30$ – слабая прямая корреляция;
$-0{,}20 < r < 0{,}20$ – корреляция отсутствует;
$-0{,}30 < r \le -0{,}20$ – слабая обратная корреляция;
$-0{,}50 < r \le -0{,}30$ – умеренная обратная корреляция;
$-0{,}70 < r \le -0{,}50$ – средняя обратная корреляция;
$-1 \le r \le -0{,}70$ – сильная обратная корреляция.

Шкала уровней значимости связи переменных (α)

$0{,}10 < p$ – связь статистически не значимая;
$0{,}05 < p \le 0{,}10$ – связь статистически значимая (тенденция);
$0{,}01 < p \le 0{,}05$ – связь статистически значимая (достоверная);
$p \le 0{,}01$ – связь статистически значимая (высокая).

9.2. *r*-КРИТЕРИЙ СПИРМЕНА

r-критерий Спирмена является критерием ранговой корреляции, который применяется для переменных (свойств), измеренных, как правило, в шкале порядка.

Вопросы, ответы на которые можно найти с помощью r-критерия Спирмена

1. Есть ли статистически значимая связь показателей общительности (A) и академической успеваемости (B) у студентов исследуемой группы?
2. Каков уровень связи показателей общительности и академической успеваемости у студентов исследуемой группы?

Алгоритм *r*-критерия Спирмена (Excel)

1. Разместите таблицу выборок A и B в Excel. Дополните таблицу столбцами: «rA», «rB», «$rA - rB$», «$(rA - rB)^2$».

2. Проведите сортировку и ранжирование выборки A.

Проведите сортировку и ранжирование выборки B.

Замените варианты выборок A и B их рангами, заполнив столбцы «rA» и «rB».

3. Заполните столбцы:

– «$rA - rB$»;

– «$(rA - rB)^2$».

4. Вычислите сумму ячеек столбца «$(rA - rB)^2$», которую обозначьте D.

5. Вычислите r по формуле

$$r = 1 - \frac{6 * D}{n(n^2 - 1)},$$

где D – сумма квадратов разностей $(rX - rY)^2$;

n – объем выборки.

6. Найдите в строке «n» таблицы r-критерия Спирмена значения: $r0{,}10$; $r0{,}05$; $r0{,}01$.

7. Статистический вывод.

Если $|r| < r0{,}10$, то принимается гипотеза H_0.

Если $r0{,}10 \le |r| < r0{,}05$, то принимается гипотеза H_1 (r, $p < 0{,}10$).

Если $r0{,}05 \le |r| < r0{,}01$, то принимается гипотеза H_1 (r, $p < 0{,}05$).

Если $r0{,}01 \le |r|$, то принимается гипотеза H_1 (r, $p < 0{,}01$).

Пример использования r-критерия Спирмена (Excel)

В протоколе приведены самооценки студентов по психологии (А) и самооценки по математике (В).

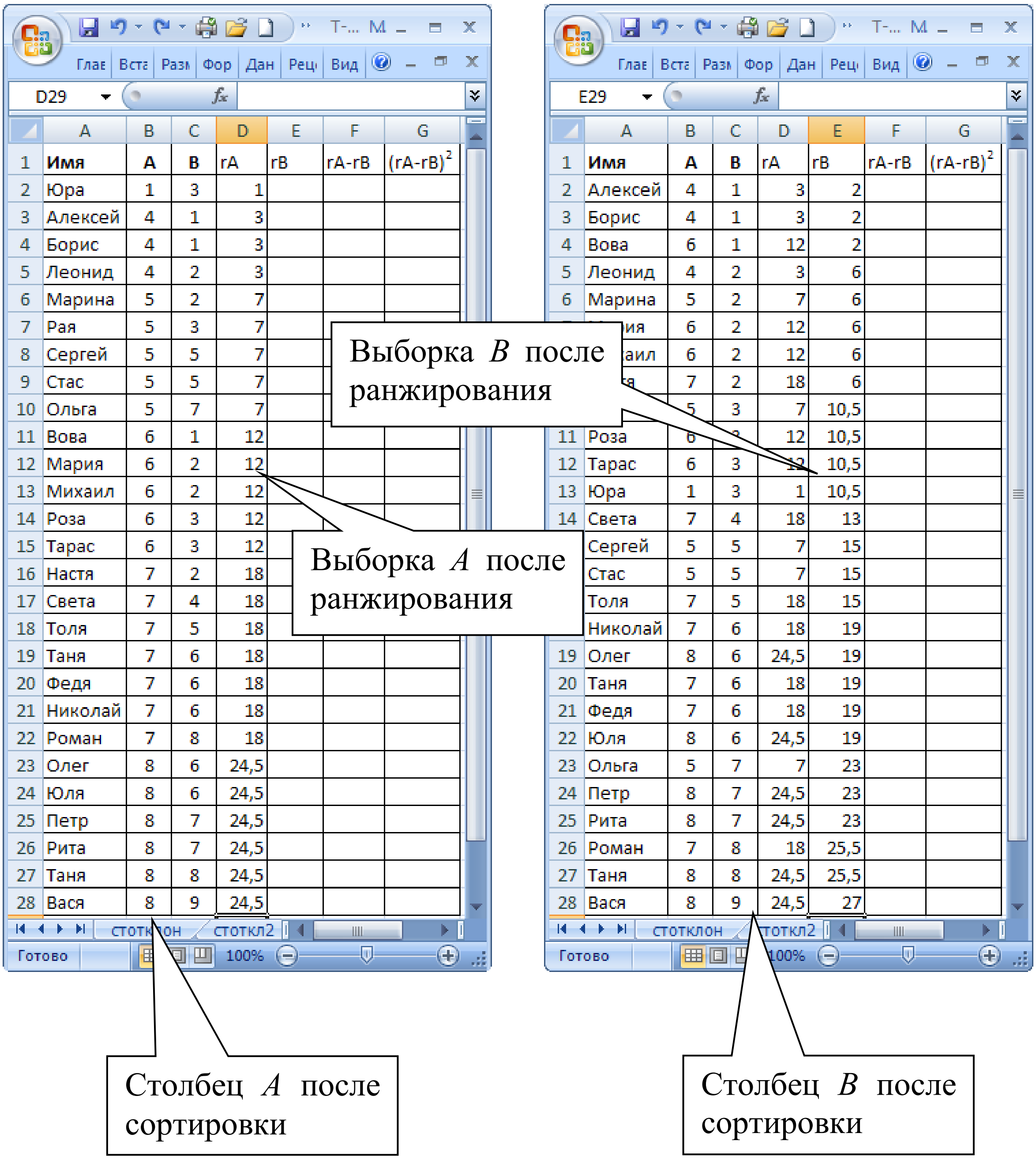
Имя	А	В	Имя	А	В	Имя	А	В
Алексей	4	1	Рая	5	3	Николай	7	6
Борис	4	1	Роза	6	3	Олег	8	6
Вова	6	1	Тарас	6	3	Юля	8	6
Леонид	4	2	Света	7	4	Ольга	5	7
Марина	5	2	Сергей	5	5	Петр	8	7
Мария	6	2	Стас	5	5	Рита	8	7
Михаил	6	2	Толя	7	5	Роман	7	8
Настя	7	2	Таня	7	6	Таня	8	8
Юра	1	3	Федя	7	6	Вася	8	9

Найдите уровни связи и ее статистической значимости самооценок по психологии и самооценок по математике у студентов.

Разместим таблицу выборок A и B в Excel и дополним ее столбцами: «rA», «rB», «$rA - rB$», «$(rA - rB)^2$».

Заметим, что объем выборки равен 27 (n = 27).

Проведем сортировку и ранжирование выборок A и B.



	A	B	C	D	E	F	G
1	Имя	A	B	rA	rB	rA-rB	$(rA-rB)^2$
2	Юра	1	3	1			
3	Алексей	4	1	3			
4	Борис	4	1	3			
5	Леонид	4	2	3			
6	Марина	5	2	7			
7	Рая	5	3	7			
8	Сергей	5	5	7			
9	Стас	5	5	7			
10	Ольга	5	7	7			
11	Вова	6	1	12			
12	Мария	6	2	12			
13	Михаил	6	2	12			
14	Роза	6	3	12			
15	Тарас	6	3	12			
16	Настя	7	2	18			
17	Света	7	4	18			
18	Толя	7	5	18			
19	Таня	7	6	18			
20	Федя	7	6	18			
21	Николай	7	6	18			
22	Роман	7	8	18			
23	Олег	8	6	24,5			
24	Юля	8	6	24,5			
25	Петр	8	7	24,5			
26	Рита	8	7	24,5			
27	Таня	8	8	24,5			
28	Вася	8	9	24,5			

	A	B	C	D	E	F	G
1	Имя	A	B	rA	rB	rA-rB	$(rA-rB)^2$
2	Алексей	4	1	3	2		
3	Борис	4	1	3	2		
4	Вова	6	1	12	2		
5	Леонид	4	2	3	6		
6	Марина	5	2	7	6		
[illegible]	[illegible]ия	6	2	12	6		
[illegible]	[illegible]аил	6	2	12	6		
[illegible]	[illegible]я	7	2	18	6		
[illegible]	[illegible]	5	3	7	10,5		
11	Роза	6	3	12	10,5		
12	Тарас	6	3	12	10,5		
13	Юра	1	3	1	10,5		
14	Света	7	4	18	13		
[illegible]	Сергей	5	5	7	15		
[illegible]	Стас	5	5	7	15		
[illegible]	Толя	7	5	18	15		
[illegible]	Николай	7	6	18	19		
19	Олег	8	6	24,5	19		
20	Таня	7	6	18	19		
21	Федя	7	6	18	19		
22	Юля	8	6	24,5	19		
23	Ольга	5	7	7	23		
24	Петр	8	7	24,5	23		
25	Рита	8	7	24,5	23		
26	Роман	7	8	18	25,5		
27	Таня	8	8	24,5	25,5		
28	Вася	8	9	24,5	27		

Заполним столбцы «$rA - rB$», «$(rA - rB)^2$».

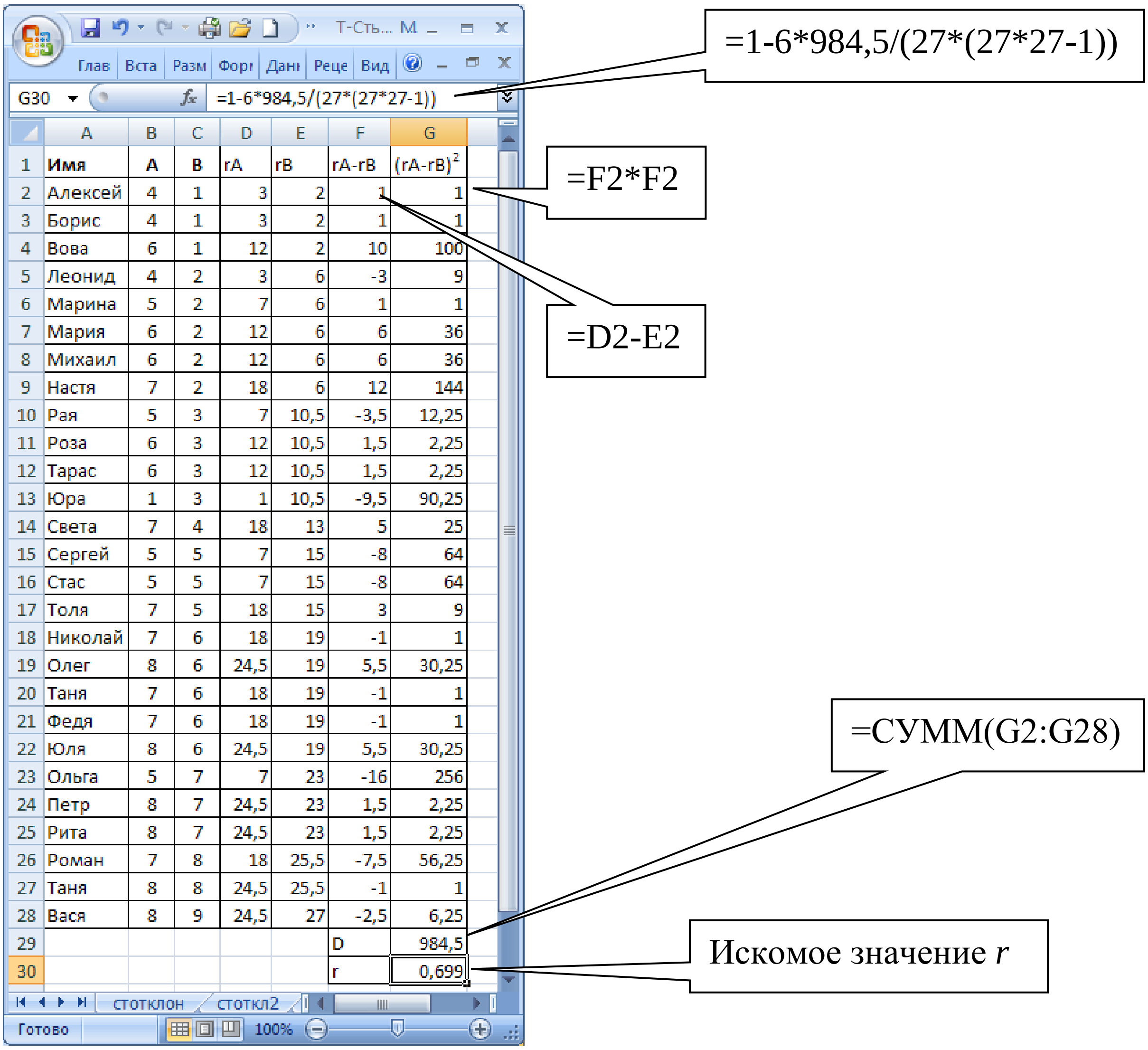
Вычислим сумму ячеек столбца «$(rA - rB)^2$», которую обозначим D. Вычислим r по формуле

$$r = 1 - \frac{6 * D}{n(n^2 - 1)},$$

где D – сумма квадратов разностей $(rX - rY)^2$;

n – объем выбори.

Имеем $r = 0{,}699$.



	A	B	C	D	E	F	G
1	**Имя**	**A**	**B**	rA	rB	rA-rB	(rA-rB)^2
2	Алексей	4	1	3	2	1	1
3	Борис	4	1	3	2	1	1
4	Вова	6	1	12	2	10	100
5	Леонид	4	2	3	6	-3	9
6	Марина	5	2	7	6	1	1
7	Мария	6	2	12	6	6	36
8	Михаил	6	2	12	6	6	36
9	Настя	7	2	18	6	12	144
10	Рая	5	3	7	10,5	-3,5	12,25
11	Роза	6	3	12	10,5	1,5	2,25
12	Тарас	6	3	12	10,5	1,5	2,25
13	Юра	1	3	1	10,5	-9,5	90,25
14	Света	7	4	18	13	5	25
15	Сергей	5	5	7	15	-8	64
16	Стас	5	5	7	15	-8	64
17	Толя	7	5	18	15	3	9
18	Николай	7	6	18	19	-1	1
19	Олег	8	6	24,5	19	5,5	30,25
20	Таня	7	6	18	19	-1	1
21	Федя	7	6	18	19	-1	1
22	Юля	8	6	24,5	19	5,5	30,25
23	Ольга	5	7	7	23	-16	256
24	Петр	8	7	24,5	23	1,5	2,25
25	Рита	8	7	24,5	23	1,5	2,25
26	Роман	7	8	18	25,5	-7,5	56,25
27	Таня	8	8	24,5	25,5	-1	1
28	Вася	8	9	24,5	27	-2,5	6,25
29						D	984,5
30						r	0,699

Найдем значения $r0{,}10$, $r0{,}05$ и $r0{,}01$ в строке $n = 27$ таблицы r-критерия Спирмена: $r0{,}10 = 0{,}323$; $r0{,}05 = 0{,}381$; $r0{,}01 = 0{,}487$.

Статистический вывод.

Так как 0,699 > 0,487, то принимается H_1 ($r = 0{,}699$; $p \leq 0{,}01$).

Содержательный вывод.

У студентов выявлена сильная прямая статистически значимая связь ($r = 0{,}699$, $p \leq 0{,}01$) самооценок по психологии и самооценок по математике. То есть, чем выше самооценка студента по психологии, тем у него выше самооценка по математике.

9.3. *r*-КРИТЕРИЙ ПИРСОНА

r-критерий Пирсона является критерием линейной корреляции, который применяется для переменных (свойств), измеренных, как правило, в шкале интервалов или шкале отношений.

Вопросы, ответы на которые можно найти с помощью r-критерия Пирсона

1. Есть ли статистически значимая связь показателей «Раздражительности» (A) и «Принятия агрессии» (B) у курсантов исследуемой группы?

2. Каков уровень связи показателей «Раздражительности» (A) и «Принятия агрессии» (B) у курсантов исследуемой группы?

Алгоритм r-критерия Пирсона (Excel)

1. Разместите таблицу выборок A и B в Excel.
2. Курсор поставьте на пустую ячейку таблицы.
3. Последовательно выполните операции:

– нажмите клавишу со знаком «f_x»;

– в появившемся окне «Мастер функций» в ячейке «Поиск функции» наберите КОРРЕЛ. Нажмите кнопку «Найти» и «ОК»;

– в появившееся окно «Аргументы функции» впишите:

  – в строку «Массив 1» – код первой ячейки выборки A: код последней ячейки выборки A;

  – в строку «Массив 2» – код первой ячейки выборки B: код последней ячейки выборки B, нажмите «ОК».

4. В ячейке появится значение r (уровня связи значений выборок A и B).

5. Находят значения $r0{,}10$, $r0{,}05$ и $r0{,}01$ в строке n таблицы r-критерия Пирсона, где n – объем выборки.

6. Статистический вывод.

Если $|r| < r0{,}10$, то принимается гипотеза H_0.
Если $r0{,}10 \le |r| < r0{,}05$, то принимается гипотеза H_1 (r, $p < 0{,}10$).
Если $r0{,}05 \le |r| < r0{,}01$, то принимается гипотеза H_1 (r, $p < 0{,}05$).
Если $r0{,}01 \le |r|$, то принимается гипотеза H_1 (r, $p < 0{,}01$).

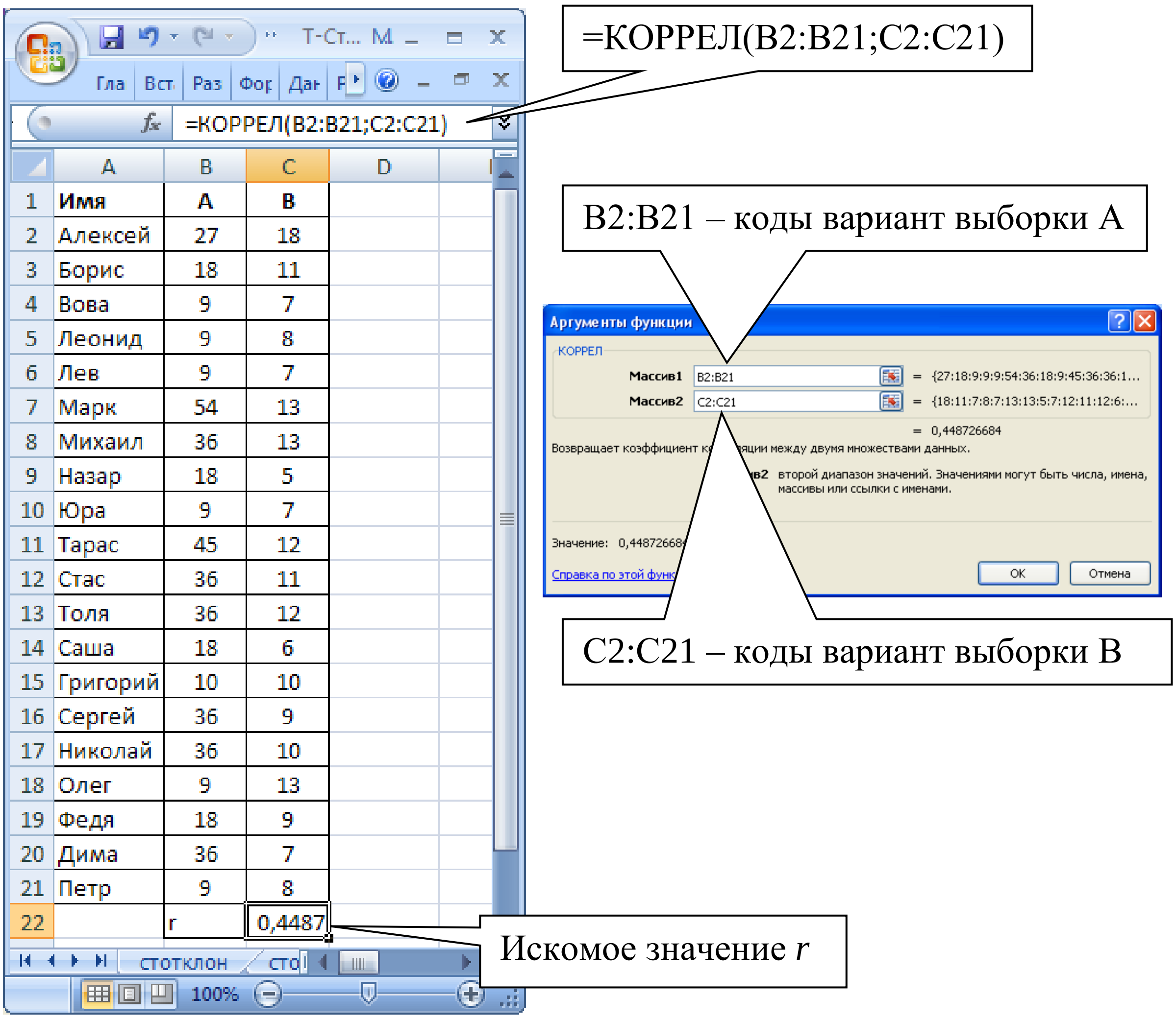
Пример использования r-критерия Пирсона (Excel)

В протоколе приведены показатели «Раздражительности» (A) и «Принятия агрессии» (B) у курсантов исследуемой группы.

Имя	A	B	Имя	A	B	Имя	A	B	Имя	A	B
Алексей	27	18	Марк	54	13	Стас	36	11	Николай	36	10
Борис	18	11	Михаил	36	13	Толя	36	12	Олег	9	13
Вова	9	7	Назар	18	5	Саша	18	6	Федя	18	9
Леонид	9	8	Юра	9	7	Григорий	10	10	Дима	36	7
Лев	9	7	Тарас	45	12	Сергей	36	9	Петр	9	8

Найдите уровни связи и ее статистической значимости показателей «Раздражительности» и «Принятия агрессии» у курсантов.

Разместим таблицу выборок A и B в Excel.



	A	B	C
1	Имя	А	В
2	Алексей	27	18
3	Борис	18	11
4	Вова	9	7
5	Леонид	9	8
6	Лев	9	7
7	Марк	54	13
8	Михаил	36	13
9	Назар	18	5
10	Юра	9	7
11	Тарас	45	12
12	Стас	36	11
13	Толя	36	12
14	Саша	18	6
15	Григорий	10	10
16	Сергей	36	9
17	Николай	36	10
18	Олег	9	13
19	Федя	18	9
20	Дима	36	7
21	Петр	9	8
22		r	0,4487

Имеем $r = 0{,}4487$.

Найдем значения $r0{,}10$, $r0{,}05$ и $r0{,}01$ в строке $n = 20$ таблицы r-критерия Пирсона: $r0{,}10 = 0{,}378$; $r0{,}05 = 0{,}444$; $r0{,}01 = 0{,}561$.

Статистический вывод. Так как $0{,}444 < 0{,}4487 \leq 0{,}561$, то принимается H_1 ($r = 0{,}449$; $p \leq 0{,}05$).

Содержательный вывод.

Выявлена статистически значимая ($r = 0{,}449$, $p \leq 0{,}05$) умеренная прямая связь показателей «Раздражительности» и «Принятия агрессии» у курсантов. Таким образом, чем выше раздражительность курсанта, тем у него значимо выше принятие агрессии.

9.4. *Z*-КРИТЕРИЙ ФИШЕРА

Z-критерий Фишера является критерием, который применяется для сравнения коэффициентов корреляции двух переменных (свойств), полученных на разных выборках.

Вопрос, ответ на который можно найти с помощью Z-критерия Фишера

Есть ли статистически значимые различия связей показателей общительности и академической успеваемости у студентов очной (r = 0,31, $p \leq$ 0,05) и заочной (r = 0,63, $p \leq$ 0,05) форм обучения?

Алгоритм *Z*-критерия Фишера (Excel)

1. В Excel занесите вычисленные значения r_1, n_1, r_2, n_2,
где r_1 – коэффициент корреляции переменных в выборке I;
n_1 – число респондентов в выборке I;
r_2 – коэффициент корреляции переменных в выборке II;
n_2 – число респондентов в выборке II.

2. Найдите с помощью функции ФИШЕР в Excel значения:
$Z_1 = Z(r_1)$;
$Z_2 = Z(r_2)$.

3. Вычислите Z по формуле

$$Z = \frac{|Z_1 - Z_2|}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}},$$

где $Z_1 = Z(r_1)$;
$Z_2 = Z(r_2)$;
n_1 – число респондентов в выборке I;
n_2 – число респондентов в выборке II.

4. Найдите с помощью функции НОРМСТРАСП в Excel значение P(Z).

5. Вычислите уровень значимости α по формуле

$$\alpha = 2*(1 - P(\mathrm{Z})).$$

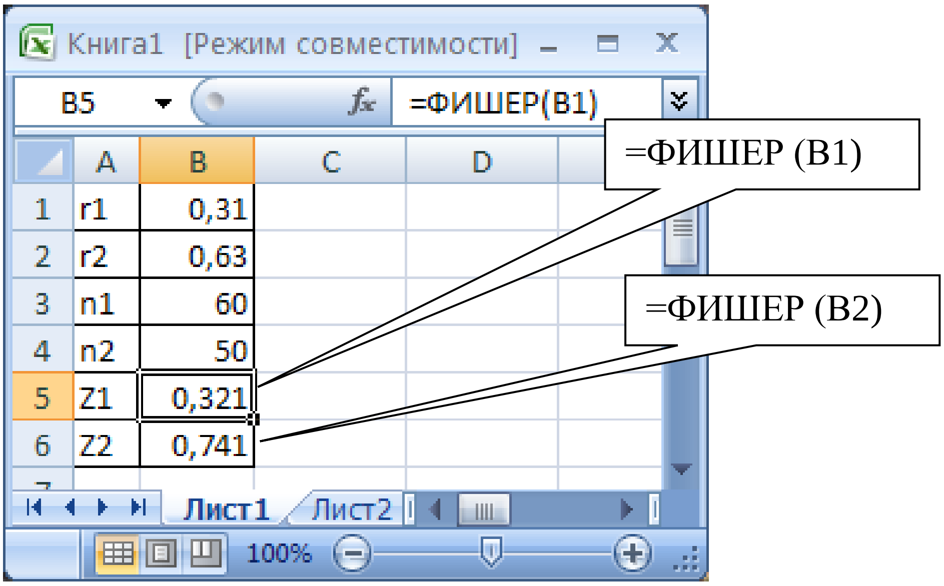
6. Статистический вывод.
Если α > 0,10, то принимается H_0.
Если α ≤ 0,10, то принимается H_1 ($p \leq \alpha$).

Пример использования Z-критерия Фишера (Excel)

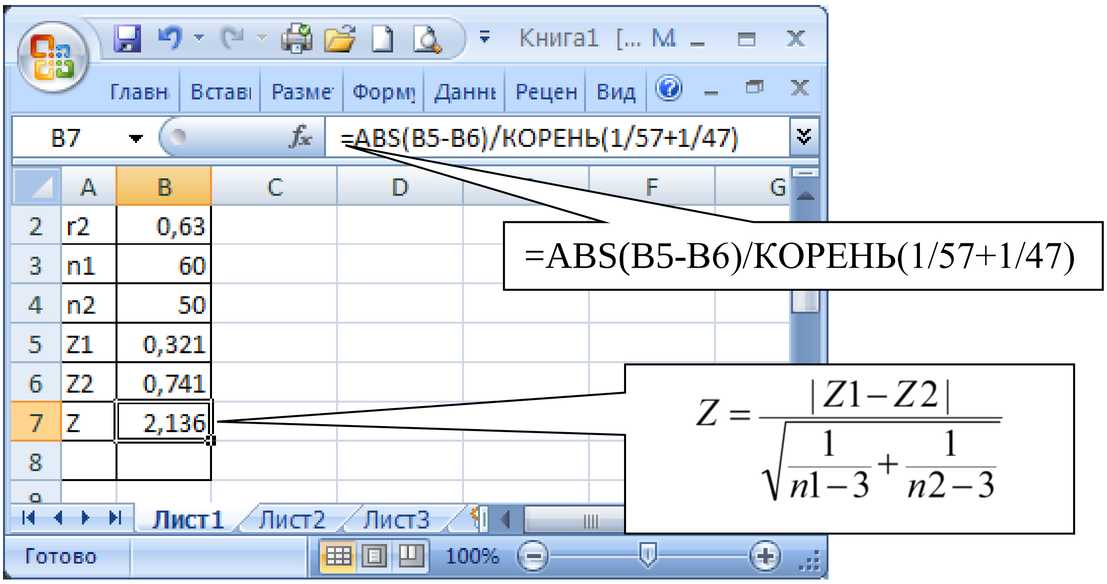
Установлены статистически значимые связи показателей общительности и академической успеваемости у студентов очной ($r = 0{,}31$, $p \leq 0{,}05$) и заочной ($r = 0{,}63$, $p \leq 0{,}05$) форм обучения. Объем выборки I (ОФО) равен 60 респондентов ($n_1 = 60$), объем выборки II (ЗФО) равен 50 респондентов ($n_2 = 50$).

Есть ли статистически значимые различия связей показателей общительности и академической успеваемости у студентов очной и заочной форм обучения?

Занесем в Excel вычисленные значения r_1, n_1, r_2, n_2. Найдем с помощью функции ФИШЕР значения: $Z_1 = Z(r_1)$; $Z_2 = Z(r_2)$.

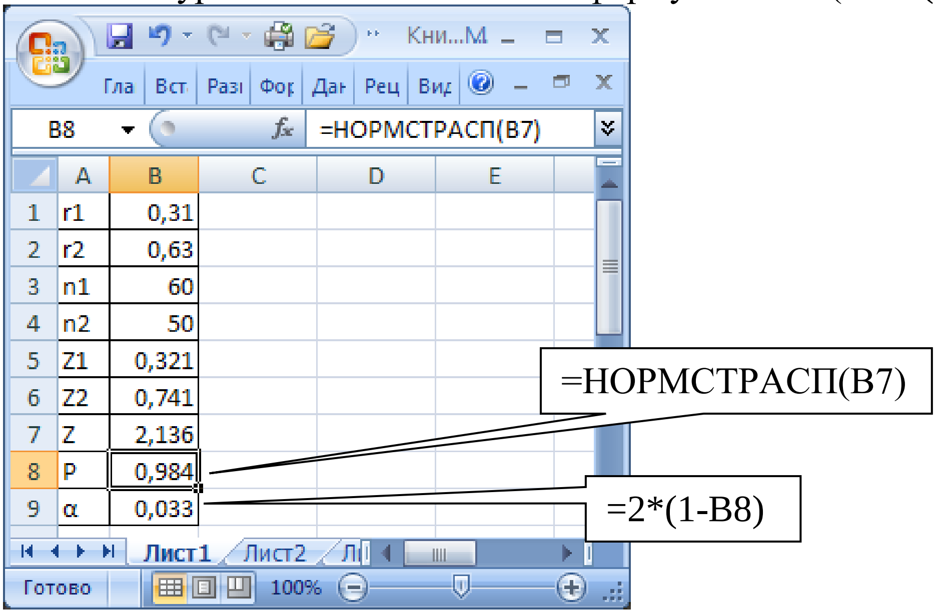


3. Вычислим Z.



Найдем с помощью функции НОРМСТРАСП значение P(Z).

Вычислим уровень значимости α по формуле $\alpha = 2*(1 - P(Z))$.



Имеем $\alpha = 0{,}033$.

Статистический вывод.

Так как 0,033 < 0,10, то принимается H_1 ($p \leq 0{,}033$).

Содержательный вывод.

Выявлены статистически значимые ($p \leq 0{,}033$) различия связей показателей общительности и академической успеваемости у студентов очной и заочной форм обучения.

Резюме

Анализ корреляции позволяет установить уровень связи (r) и уровень значимости связи ($p \leq \alpha$) между свойствами исследуемого явления.

Важно помнить, что $|r| \leq 1$, а значимыми являются связи для $p \leq 0{,}10$.

Выявленная статистическая связь свойств исследуемого явления является достаточным, но не необходимым условием каузального отношения двух свойств психологического феномена.

Необходимым условием выявления каузального отношения свойств является теоретическое обоснование зависимости одного свойства от другого, в котором сложно использовать какие-либо математические методы.

Упражнения

8.1. В протоколе приведены результаты измерения логических способностей (A) и уровней образного мышления (B) у школьников.

Имя	A	B
Алексей	42	2,2
Борис	44	2,8
Леонид	18	5,2
Марина	16	4,8
Мария	14	5,6
Михаил	38	2,2
Настя	44	2,2

Имя	A	B
Николай	28	4,6
Олег	28	4,4
Ольга	44	2,6
Петр	36	2,2
Рита	44	2,2
Роман	24	4,2
Степан	24	6,0

Имя	A	B
Стас	28	2,2
Тарас	18	2,4
Татьяна	16	2,6
Ульяна	44	2,2
Федор	36	2,8
Юрий	34	5,2
Яна	38	4,6

Найдите уровни связи и ее статистической значимости показателей логических способностей и уровней образного мышления у школьников.

8.2. В протоколе приведены самоактуализационные профили личности врачей (A) и учителей (B).

Наименование шкалы	A	B
Компетентность во времени	44	42
Поддержка	46	26
Ценностные ориентации	56	26
Гибкость поведения	45	28
Сензитивность к себе	42	28
Спонтанность	42	38
Самоуважение	53	38
Самопринятие	44	38
Представления о природе человека	55	58
Синергия	52	52
Принятие агрессии	45	42
Контактность	46	24
Познавательные потребности	46	38
Креативность	42	38

Найдите уровни связи и ее статистической значимости самоактуализационных профилей личности врачей и учителей.

8.3. В протоколе приведены результаты измерения черт личности по опроснику 16PF у мужчин (*A*) и женщин (*B*).

Наименование фактора	A	B
Открытость – скрытность (*A*)	6,47	6,56
Интеллект высокий – низкий (*B*)	6,97	7,20
Эмоциональная устойчивость – неустойчивость (*C*)	7,20	4,31
Настойчивость – покорность (*E*)	3,13	6,12
Беспечность – озабоченность (*F*)	3,93	5,74
Добросовестность – недобросовестность (*G*)	5,87	6,91
Смелость – робость (*H*)	7,93	5,84
Мягкость – твердость (*I*)	6,53	6,83
Подозрительность – доверчивость (*L*)	5,73	7,57
Мечтательность – практичность (*M*)	4,47	4,49
Проницательность – наивность (*N*)	4,73	6,51
Озабоченность – самонадеянность (*O*)	7,33	7,06
Гибкость – ригидность (Q_1)	4,47	5,76
Самостоятельность – зависимость (Q_2)	6,10	3,31
Самоконтроль высокий – низкий (Q_3)	6,42	6,05
Напряженность – расслабленность (Q_4)	6,83	5,71

Найдите коэффициент корреляции профилей черт личности мужчин и женщин.

8.4. Установлены статистически значимые связи между семейными ценностями и профессиональными ориентациями юношей ($r = 0{,}39$, $p \leq 0{,}01$) и девушек ($r = 0{,}44$, $p \leq 0{,}05$). Объем соответствующих выборок – 45 и 22 респондента.

Есть ли статистически значимые различия связей семейных ценностей и профессиональных ориентаций юношей и девушек?

10. ОДНОФАКТОРНЫЙ ДИСПЕРСИОННЫЙ АНАЛИЗ

Метод дисперсионного анализа имеет общепринятое обозначение **ANOVA** – краткое написание термина **AN**ALISIS **O**F **VA**RIANCE. В психологических исследованиях чаще всего применяются однофакторный **ANOVA** и двухфакторный **ANOVA**, реже другие виды дисперсионного анализа.

С помощью однофакторного **ANOVA** проверяются статистические гипотезы об отличиях средних значений некоторого свойства в выборке, классифицированной по одному основанию, так называемому одному фактору. То есть проверяются статистические гипотезы об отличиях средних значений одной переменной, имеющей три и более градации одного фактора.

С помощью двухфакторного **ANOVA** проверяются статистические гипотезы об отличиях средних значений некоторого свойства в выборке, классифицированной по двум основаниям, так называемым двум факторам. То есть проверяются статистические гипотезы об отличиях средних значений одной переменной (одного свойства), имеющей по две и более градации двух факторов.

Особенности дисперсионного анализа **ANOVA** заключаются в следующем:

– сравниваемая переменная (свойство) должна быть измерена в метрической шкале;

– сравниваемые выборки, соответствующие разным градациям фактора, должны мало отличаться по численности, или стандартные отклонения сравниваемых выборок отличаются незначимо;

– соответствие распределений частот для каждой градации фактора нормальному распределению, хотя исследователи (Шеффе, Гласс, Стенли и др.) утверждают, что это требование мало влияет на результаты **ANOVA.**

Однофакторный дисперсионный анализ (**ANOVA)** можно проводить с помощью ***Excel*.** Для ***двухфакторного*** дисперсионного анализа (**ANOVA)** более эффективно применений специальных статистических компьютерных программ (STATISTIKA, SPSS и др.). Поэтому рассмотрим применение однофакторного метода **ANOVA** в *Excel***.**

Однофакторный дисперсионный анализ (**ANOVA**) используется для исследования различия средних значений одной переменной (свойства) для трех и более выборок, выделенных из общей выборки путем классификации ее по одному основанию (одному фактору).

С помощью однофакторного ANOVA проверяется гипотеза H0: «Средние трех выборок статистически значимо не различаются».

Если гипотеза H0 отклоняется, то принимается гипотеза H1: «Средние по крайней мере двух выборок статистически значимо различаются ($p \le \alpha$)».

Если гипотеза H1 подтверждена, то для ответа на вопрос средние у каких групп статистически значимо различаются, применяют статистические критерии, например T-критерий Стьюдента.

Пример выборок, для которых можно применить однофакторный ANOVA

Измерено время чтения (в секундах) одного и того же текста у респондентов разных возрастных групп: до 22 лет, от 23 до 26 лет, от 27 лет. Результаты измерения представлены в протоколе.

Имя	Возраст	Время	Имя	Возраст	Время	Имя	Возраст	Время
Алекс	20	35	Гриша	23	41	Николай	27	46
Борис	20	37	Дима	23	40	Олег	27	40
Леонид	21	43	Николай	24	42	Ольга	28	40
Марина	21	35	Олег	24	39	Ольга	28	39
Мария	21	35	Олег	24	38	Петр	28	38
Михаил	21	39	Петр	25	37	Петр	29	37
Настя	21	40	Сергей	25	37	Рита	29	60
Настя	22	34	Таня	25	38	Роза	29	39
Рая	22	34	Федя	26	40	Роман	30	40
Роза	22	34	Федя	26	41	Степан	31	38
Света	22	35	Юля	26	39			
Тарас	22	36						
Юра	22	37						

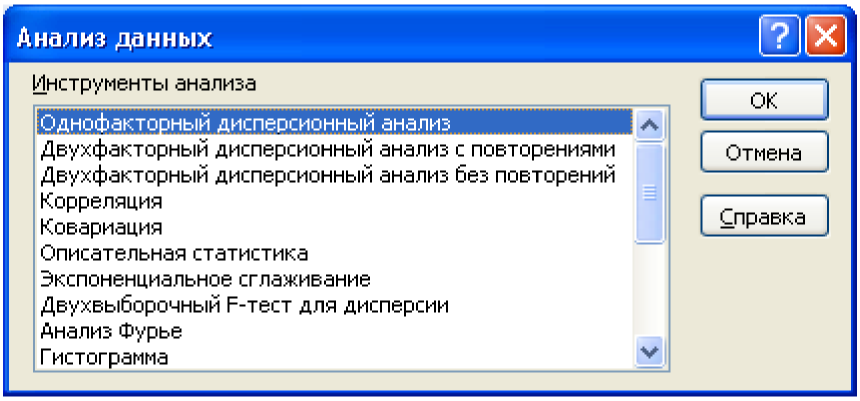
Вопрос, ответ на который можно найти с помощью однофакторного ANOVA

Есть ли статистически значимые различия времени, затраченного на чтение текста респондентами разных возрастных групп: до 22 лет, от 23 до 26 лет, от 27 лет?

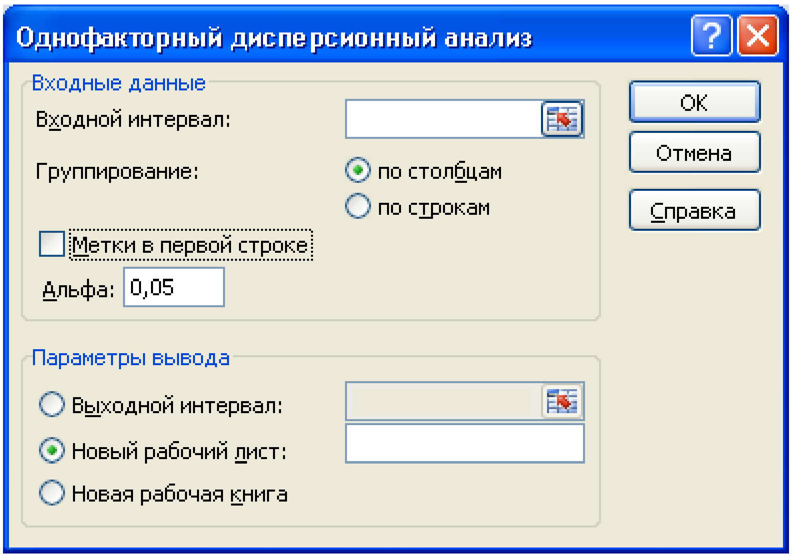
Алгоритм *однофакторного ANOVA* (Excel)

1. На отдельном листе в Excel создается таблица, столбцам которой присваиваются названия соответствующие градациям фактора. Из протокола в созданную таблицу переносятся соответствующие значения переменной (измеренного свойства).

2. Найдите в группе команд *Данные* команду *Анализ данных*.



3. В *инструментах анализа* найдите *Однофакторный дисперсионный анализ*. Нажмите ОК.



4. В окно *Входной интервал* внесите коды созданной таблицы.

5. В окне *Метки в первой строке* поставьте знак «v». Нажмите ОК.

6. В новом окне появится таблица результатов однофакторного ANOVA.

7. Найдите в таблице столбец *Р-Значение,* в котором отмечен искомый уровень значимости отличий.

Пример однофакторного ANOVA для связанных выборок (Excel)

В протоколе приведены результаты измерения времени (в секундах) чтения одного и того же текста респондентами.

Имя	Возраст	Время	Имя	Возраст	Время	Имя	Возраст	Время
Алекс	20	35	Гриша	23	41	Николай	27	46
Борис	20	37	Дима	23	40	Олег	27	40
Леонид	21	43	Николай	24	42	Ольга	28	40
Марина	21	35	Олег	24	39	Ольга	28	39
Мария	21	35	Олег	24	38	Петр	28	38
Михаил	21	39	Петр	25	37	Петр	29	37
Настя	21	40	Сергей	25	37	Рита	29	60
Настя	22	34	Таня	25	38	Роза	29	39
Рая	22	34	Федя	26	40	Роман	30	40
Роза	22	34	Федя	26	41	Степан	31	38
Света	22	35	Юля	26	39			
Тарас	22	36						
Юра	22	37						

Требуется установить уровень статистической значимости различий средних значений времени чтения текстов респондентами разных возрастных групп: до 22 лет, от 23 до 26 лет, от 27 лет.

Фактором является возраст респондентов.

Фактор имеет три градации: до 22 лет, от 23 до 26 лет, от 27 лет.

Исследуемой переменной (свойством) является время, затраченное респондентом на чтение текста.

Заменим в таблице столбец «Возраст» на столбец «Фактор».

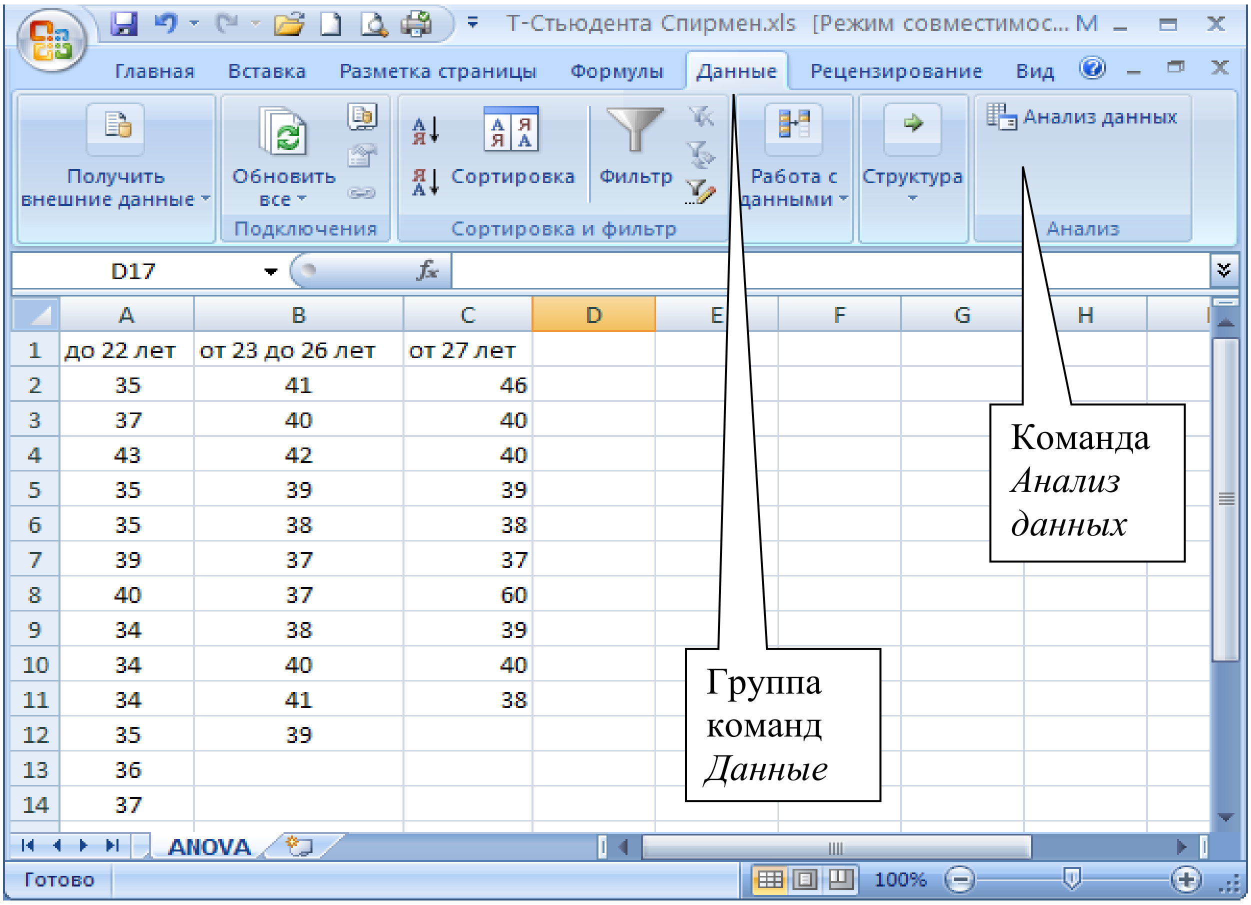
Имя	Фактор	Время
Алекс	до 22	35
Борис	до 22	37
Леонид	до 22	43
Марина	до 22	35
Мария	до 22	35
Михаил	до 22	39
Настя	до 22	40
Настя	до 22	34
Рая	до 22	34
Роза	до 22	34
Света	до 22	35
Тарас	до 22	36
Юра	до 22	37

Имя	Фактор	Время
Гриша	от 23 до 26	41
Дима	от 23 до 26	40
Коля	от 23 до 26	42
Олег	от 23 до 26	39
Олег	от 23 до 26	38
Петр	от 23 до 26	37
Сергей	от 23 до 26	37
Таня	от 23 до 26	38
Федя	от 23 до 26	40
Федя	от 23 до 26	41
Юля	от 23 до 26	39

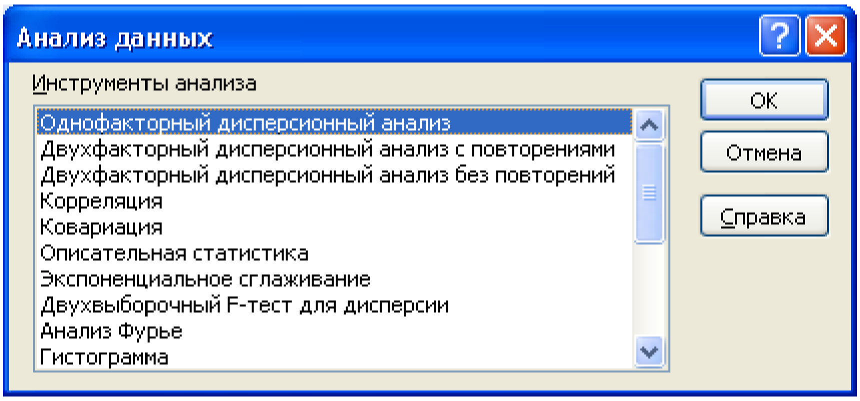
Имя	Фактор	Время
Николай	от 27	46
Олег	от 27	40
Ольга	от 27	40
Ольга	от 27	39
Петр	от 27	38
Петр	от 27	37
Рита	от 27	60
Роза	от 27	39
Роман	от 27	40
Степан	от 27	38

На отдельном лист в Excel создадим таблицу со столбцами: до 22 лет, от 23 до 26 лет, от 27 лет (названия градаций фактора).

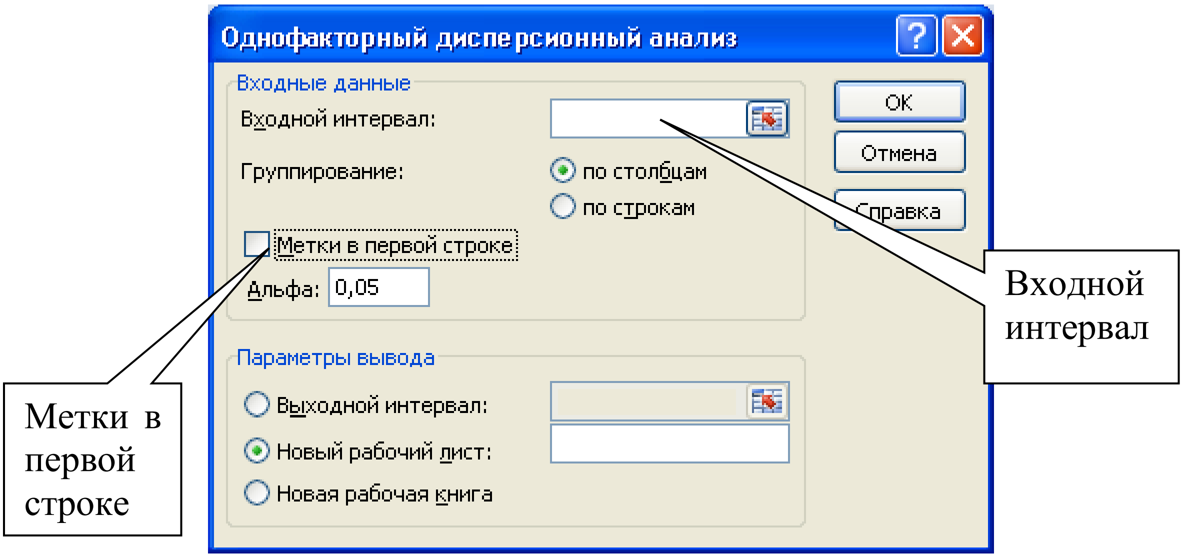
Из протокола в созданную таблицу перенесем соответствующие значения переменной (время, затраченное на чтение текста).



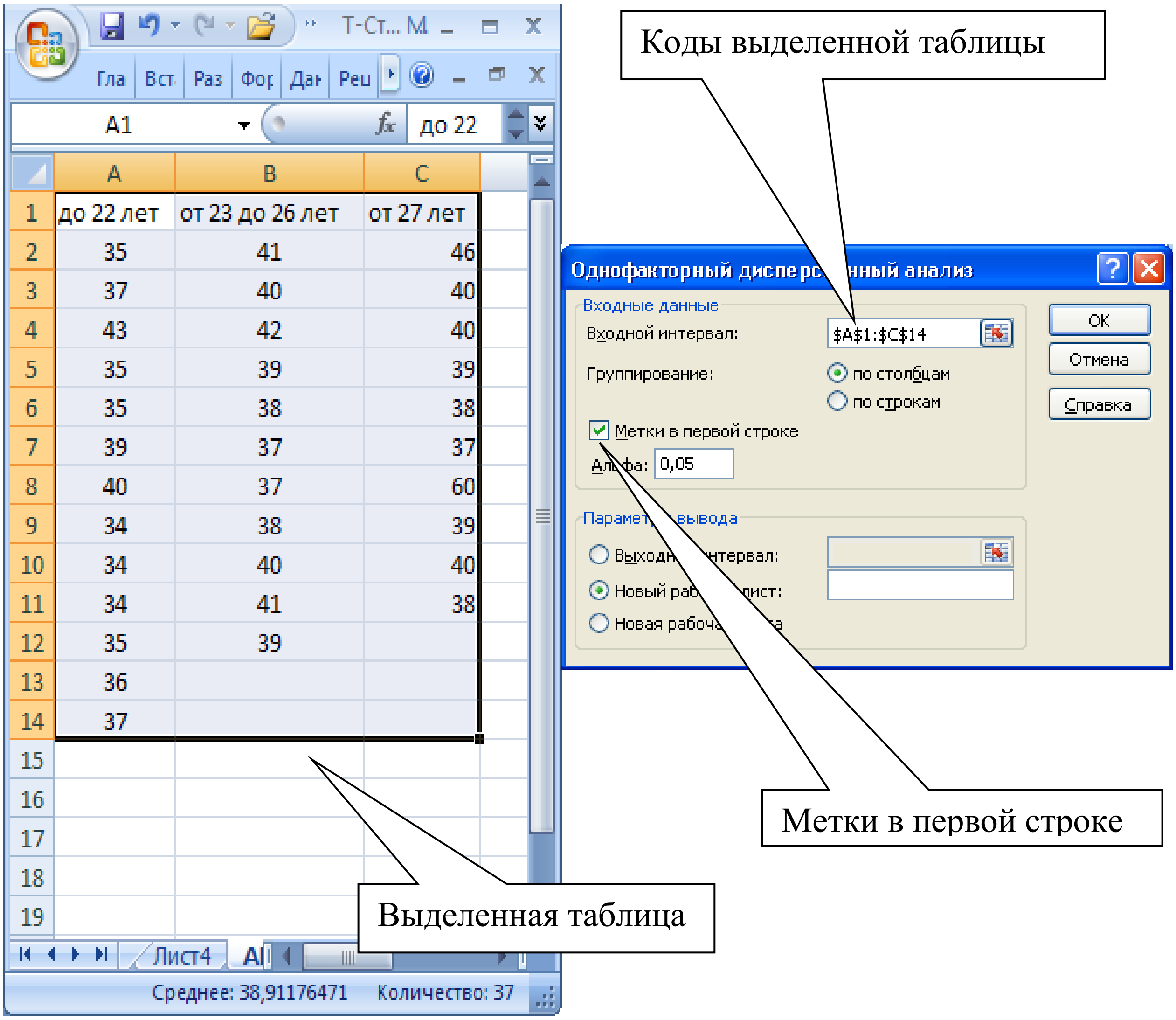
Найдем в группе команд *Данные* команду *Анализ данных*. Нажмем ОК.



В *Инструментах анализа* найдем *Однофакторный дисперсионный анализ*. Нажмем ОК.



В окно *Входной интервал* внесем коды созданной таблицы путем выделения этой таблицы. В окне *Метки в первой строке* поставим знак «v». Нажмем ОК.



В новом окне появится таблица результатов однофакторного ANOVA.

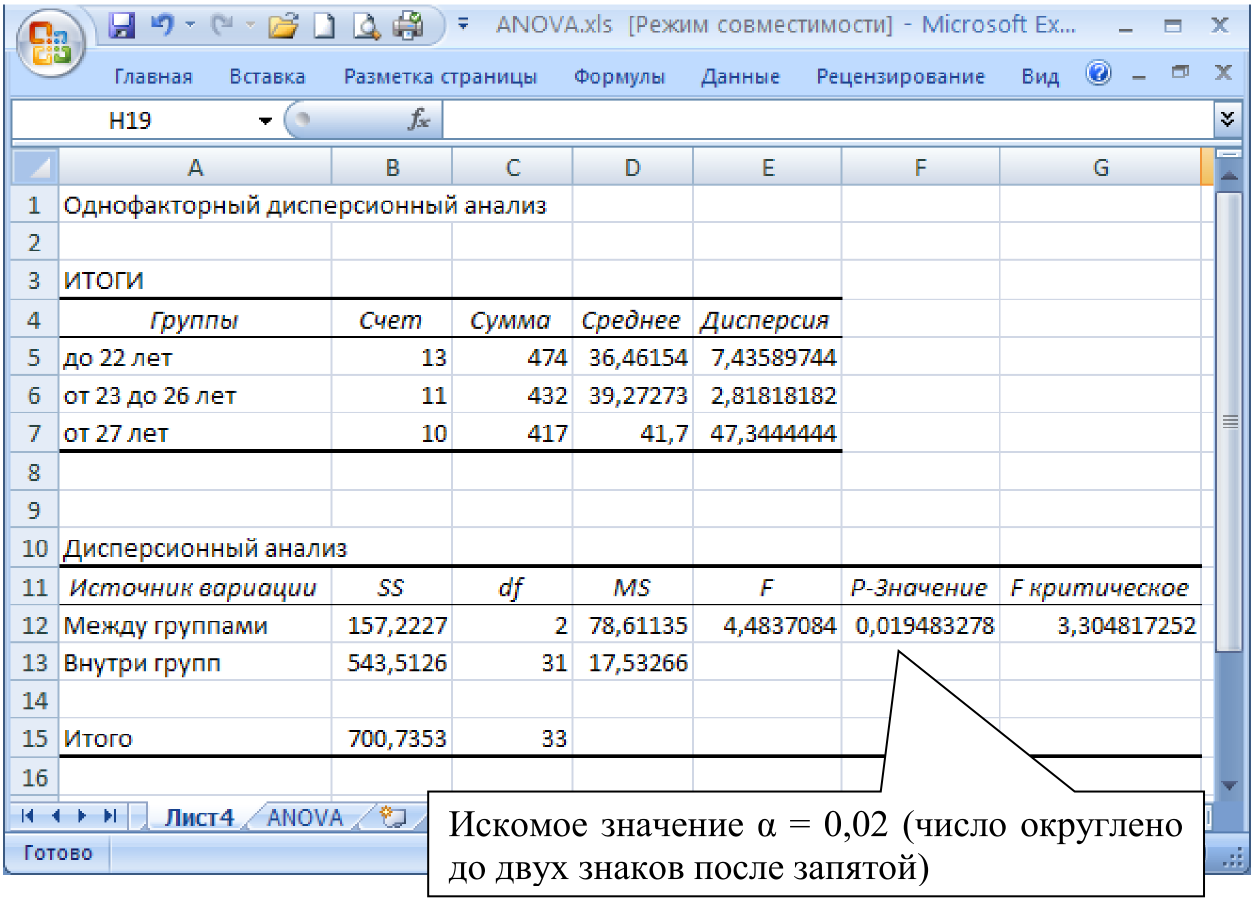
Найдем в таблице столбец *Р-Значение,* в котором отмечен искомый уровень значимости отличий: α = 0,02 (число округлено до двух знаков после запятой).

Статистический вывод.

Так как α = 0,02, т.е. р ≤ 0,02, принимается гипотеза Н1: Средние по крайней мере двух выборок статистически значимо различаются (р ≤ 0,02).

Содержательный вывод.

Выявлены статистически значимые (р ≤ 0,02) различия средних значений времени, затраченного на чтение текстов, у респондентов разных возрастных групп.



Однофакторный дисперсионный анализ						
ИТОГИ						
Группы	*Счет*	*Сумма*	*Среднее*	*Дисперсия*		
до 22 лет	13	474	36,46154	7,43589744		
от 23 до 26 лет	11	432	39,27273	2,81818182		
от 27 лет	10	417	41,7	47,3444444		
Дисперсионный анализ						
Источник вариации	*SS*	*df*	*MS*	*F*	*P-Значение*	*F критическое*
Между группами	157,2227	2	78,61135	4,4837084	0,019483278	3,304817252
Внутри групп	543,5126	31	17,53266			
Итого	700,7353	33				

Заметим, что таблице указаны значения средних для разных групп:

– 36,5 с для респондентов, возраст которых до 22 лет;
– 39,3 с для респондентов, возраст которых от 23 до 26 лет;
– 41,7 с для респондентов, возраст которых от 27 лет.

Так как гипотеза H1 подтверждена, для ответа на вопрос средние у каких групп статистически значимо различаются, применим Т-критерий Стьюдента (см. параграф 7.2).

Использование Т-критерия Стьюдента позволило выявить, что респонденты, возраст которых до 22 лет, статистически значимо меньше времени затратили на чтение текстов, чем респонденты, возраст которых от 27 лет.

Резюме

Однофакторный дисперсионный анализ позволяет установить уровень значимости отличий средних значений измеренного свойства для трех и более выборок.

Важно помнить, что результатом дисперсионного анализа в EXCEL будет подтверждение гипотезы о сходстве средних значений измеренного свойства для трех и более выборок.

Если выявлены статистически значимые различия средних для трех и более выборок, то анализ следует продолжить, чтобы установить конкретную пару выборок, у которых отличаются средние. Для анализа рекомендуется использовать критерии сравнения двух выборок, например, Т-критерия Стьюдента.

Упражнения

9.1. В протоколе приведены результаты измерения уровня самооценки (А) и значений спонтанности в общении (В) у студентов.

Имя	А	В
Алексей	н	28
Борис	н	18
Леонид	н	16
Марина	н	44
Мария	н	36
Михаил	а	34
Настя	а	38

Имя	А	В
Николай	а	42
Олег	а	44
Ольга	а	18
Петр	а	16
Рита	в	14
Роман	в	38
Степан	в	44

Имя	А	В
Стас	в	28
Тарас	в	28
Татьяна	в	44
Ульяна	в	36
Федор	в	44
Юрий	в	24
Яна	в	24

В столбце А (уровень самооценки):
«н» – самооценка заниженная;
«а» – самооценка адекватная;
«в» – самооценка завышенная.

Сравните с помощью дисперсионного анализа средние значения спонтанности в общении (В) у студентов, различающихся уровнем самооценки.

9.2. В протоколе приведены результаты измерения показателей общей интеллектуальной осведомленности (*IQ*) у студентов, обучающихся на разных направлениях высшего профессионального образования: «Психология», «Физика», «Филология».

Имя	Направление	*IQ*	Имя	Направление	*IQ*
Алексей	Психология	28	Николай	Физика	42
Борис	Психология	18	Олег	Физика	44
Леонид	Психология	16	Ольга	Физика	58
Марина	Психология	44	Петр	Физика	66
Мария	Психология	36	Рита	Физика	74
Михаил	Психология	34	Роман	Филология	38
Настя	Психология	38	Света	Филология	44
Тарас	Психология	54	Марина	Филология	24
Татьяна	Психология	68	Мария	Филология	67
Ульяна	Физика	68	Михаил	Филология	69
Федор	Физика	72	Настя	Филология	34
Дарья	Физика	84	Нина	Филология	24

Сравните с помощью дисперсионного анализа средние значения общей интеллектуальной осведомленности (*IQ*) у студентов, различающихся направлением обучения.

ЗАКЛЮЧЕНИЕ

Исследователь, наблюдая явления, формулирует предположения о свойствах явлений.

Можно выделить виды предположений о психологических свойствах явлений (психологических гипотез): описывающие отдельные свойства, сравнивающие свойства, устанавливающие отношения между свойствами. Выделенные виды предположений можно упорядочить, установить их таксономию.

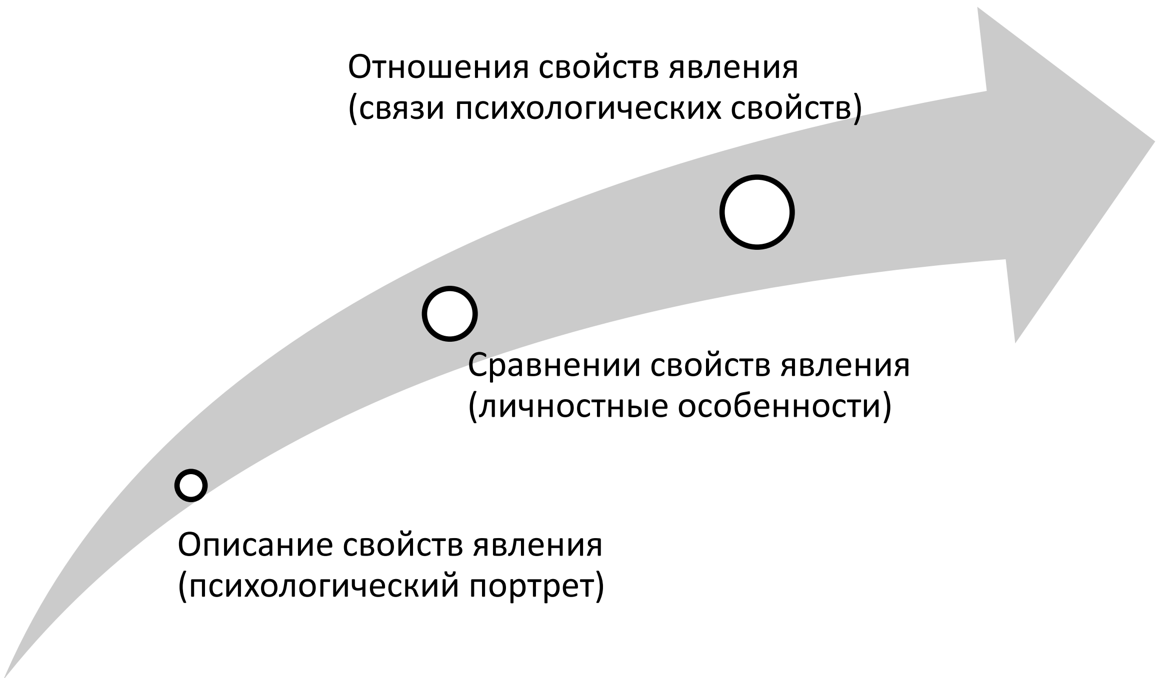


Рис. 19. Таксономия предположений

Проверка предположений начинается со сбора (протоколирования) эмпирических данных.

Для проверки гипотез первого уровня исследователь использует распределения частот (абсолютных, относительных, кумулятивных, процентильных) и описательные статистики (моду, среднее, стандартное отклонение).

Для проверки гипотез второго уровня исследователь применяет непараметрические (φ-критерий Фишера, λ-критерий Колмогорова-Смирнова, *G*-критерий знаков, *U*-критерий Манна-

Уитни) и параметрические (*T*-критерий Стьюдента, *F*-критерий Фишера) критерии сравнения выборок и генеральных совокупностей.

Для проверки гипотез третьего уровня (об отношениях между свойствами) исследователь использует коэффициенты корреляции (*r*-критерий Спирмена, *r*-критерий Пирсона) и критерии сравнения коэффициентов корреляции (*Z*-критерий Фишера).

В пособии приведен далеко не полный перечень математических методов, являющихся инструментами для проверки психологических гипотез.

Надеемся, что основные математические инструменты, представленные в пособии, дадут возможность осуществить проверку сформулированных вами предположений о психологических свойствах явлений.

РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА

Основная

Наследов А.Д. Математические методы психологического исследования: анализ и интерпретация данных. СПб., 2010.

Сидоренко Е.В. Методы математической обработки в психологии. СПб., 2012.

Дополнительная

Анастизи А., Урбина С. Психологическое тестирование. СПб., 2001.

Боровиков В.П. Программа STATISTICA для студентов и инженеров. М., 2001.

Гласс Дж., Стэнли Дж. Статистические методы в педагогике и психологии. М., 1976.

Гмурман В.Е. Теория вероятностей и математическая статистика. М., 1998.

Готтсданкер Р. Основы психологического эксперимента. М., 2005.

Гусев А.Н., Измайлов Ч.А., Михайлевская М.Б. Измерение в психологии: общий психологический практикум. М., 1997.

Ермолаев О.Ю. Математическая статистика для психологов. М., 2004.

Крамер Д. Математическая обработка данных в социальных науках. М., 2007.

Лакин Г.Ф. Биометрия. М., 1990.

Основы математической статистики / под ред. В.С. Иванова. М., 1990.

Паниотто В.И., Максименко В.С. Количественные методы в социологических исследованиях. Киев, 1982.

Суходольский Г.В. Основы математической статистики для психологов. СПб., 1997.

Тюрин Ю.Н., Макаров А.А. Анализ данных на компьютере / под ред. В.Э. Фигурнова. М., 1995.

Ядов В.А. Стратегия социологического исследования. Описание, объяснение, понимание социальной реальности. М., 1999.

ОГЛАВЛЕНИЕ

Учебное издание

Некрасов Сергей Дмитриевич

МАТЕМАТИЧЕСКИЕ МЕТОДЫ В ПСИХОЛОГИИ
(MS EXCEL)

Подписано в печать 09.01.2014 г. Формат 60х84 $^{1}/_{16}$
Усл. печ. л. 9,25. Тираж 500 экз. Заказ № 1684.

Кубанский государственный университет
350040, г. Краснодар, ул. Ставропольская, 149.
Издательско-полиграфический центр КубГУ
350040, г. Краснодар, ул. Ставропольская, 149.