Математические методы в психологии

Рекомендуемая литература

- Наследов, А.Д. Математические методы психологического исследования. Анализ и интерпретация данных. – СПб. : Речь, 2004. – 392 с.
- Сидоренко, Е.В. Методы математической обработки в психологии. – СПб. : Речь, 2001. – 350 с.
- Кутейников, А.Н. Математические методы в психологии. – СПб. : Речь, 2008. – 172с.
- Тюменева, Ю.А. Психологическое измерение. – М. : Аспект Пресс, 2007. – 192 с.
- Халафян, А.А. STATISTICA 6. Статистический анализ данных. – М. : Бином-Пресс, 2007. – 512 с.
- Боровиков, И.П. Боровиков В.П. Статистический анализ и обработка данных в среде Windows. – М. : Информационно-издательский дом Филинъ, 1998. – 608 с.

Тема 1. Измерение в психологии

- Предмет и назначение дисциплины
- Измерение в психологии. Взаимоотношение параметров, признаков, показателей и переменных.
- Шкалы измерений по С. Стивенсону

Определение статистики

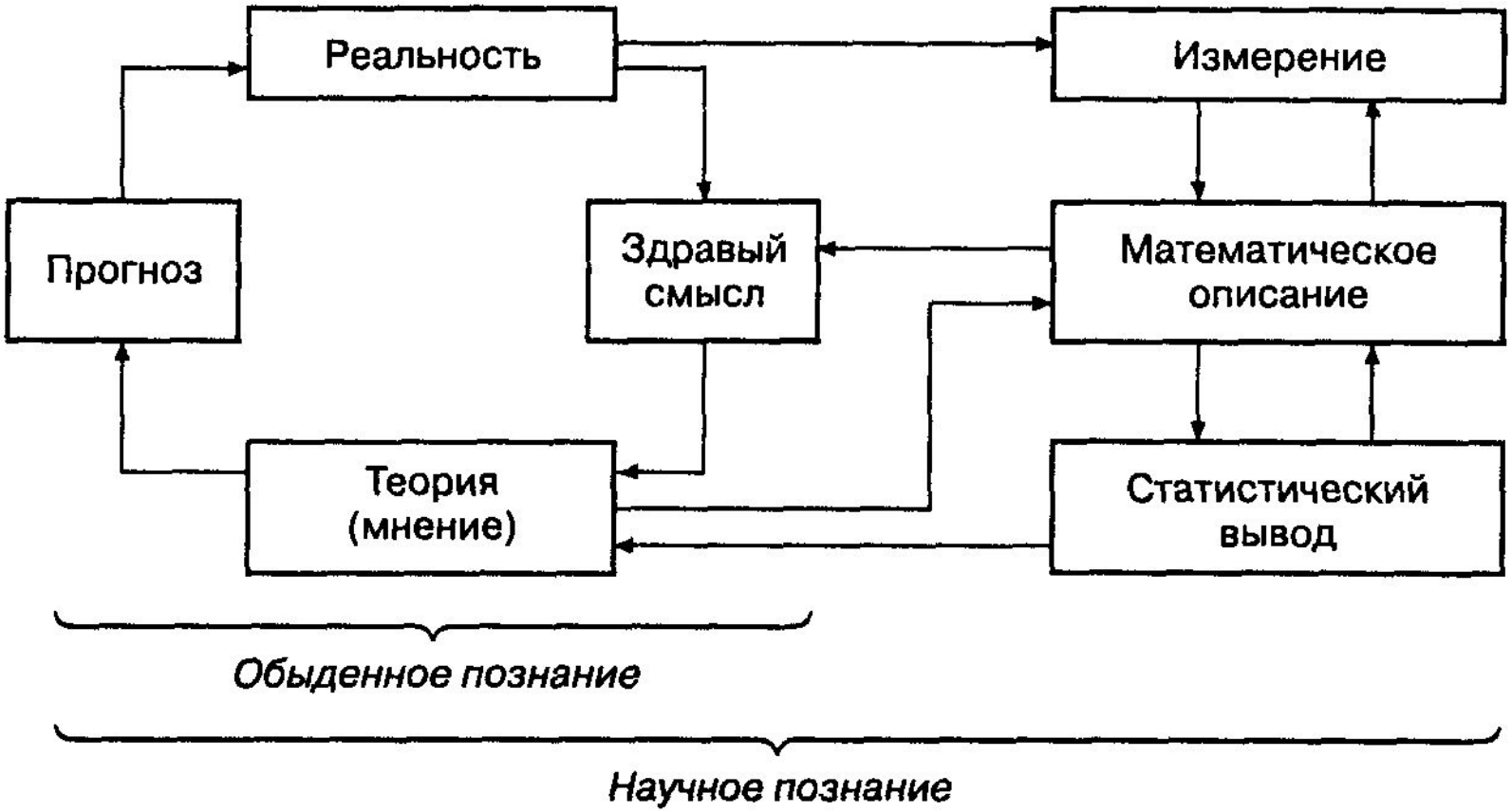
Термин «статистика» имеет несколько значений:

•• это совокупность данных и сведений, посвященных какому-либо вопросу, в этом значении он используется во многих международных и национальных изданиях, примером чего может служить «Ежегодник мировой санитарной статистики», «статистика, заболеваемости и смертности»; старое значение слова «статистика», как один из разделов науки об управлении государством, сбор, классификация и обсуждение сведений об обществе и государстве.

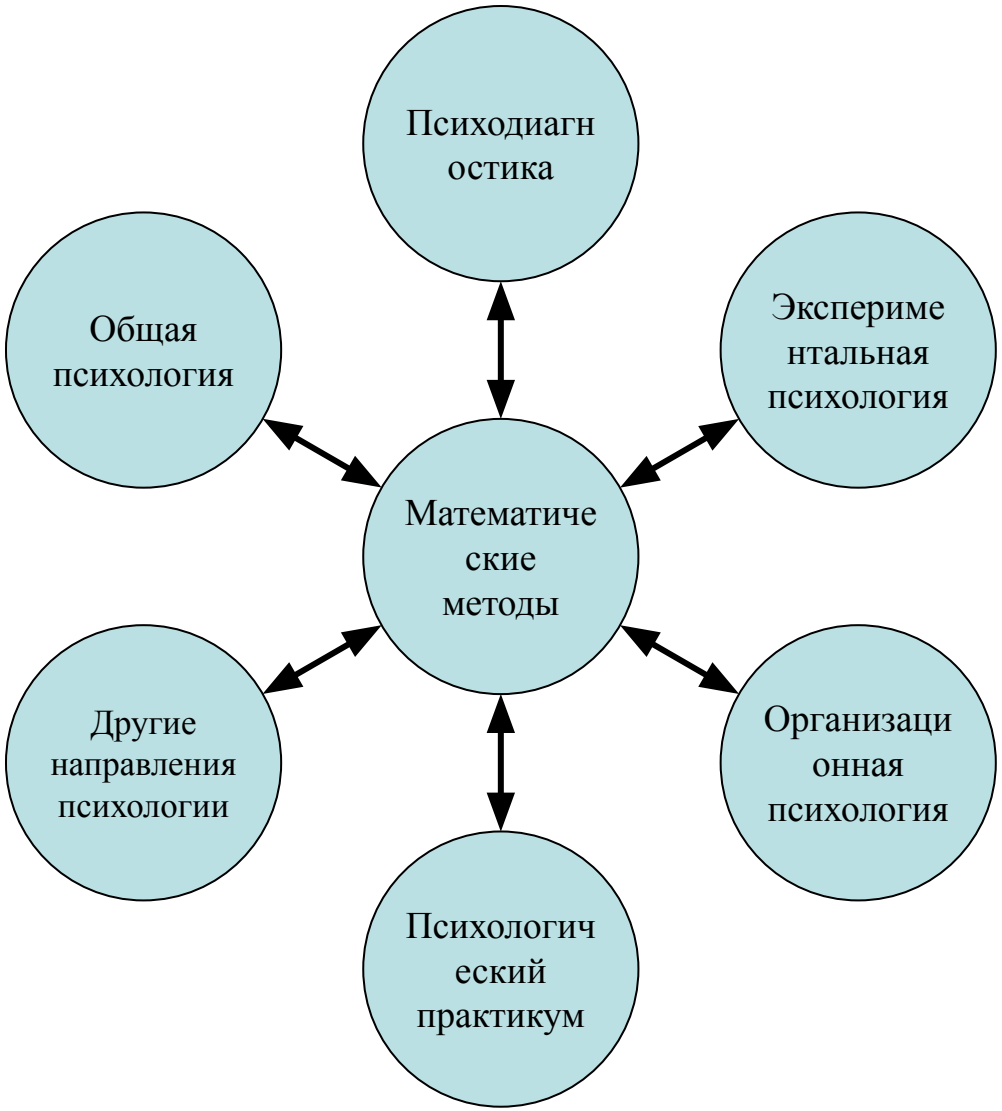
•• это описательные или дистрибутивные характеристики описывающие какую то совокупность данных, по каким то параметрам (средняя, дисперсия и так далее);

•и наконец статистика (или математическая статистика) это научная дисциплина, изучающая методы сбора и обработки фактов и данных, относящихся к человеческой деятельности и природным явлениям (из Оксфордского словаря английского языка).

Соотношение обыденного и научного познания

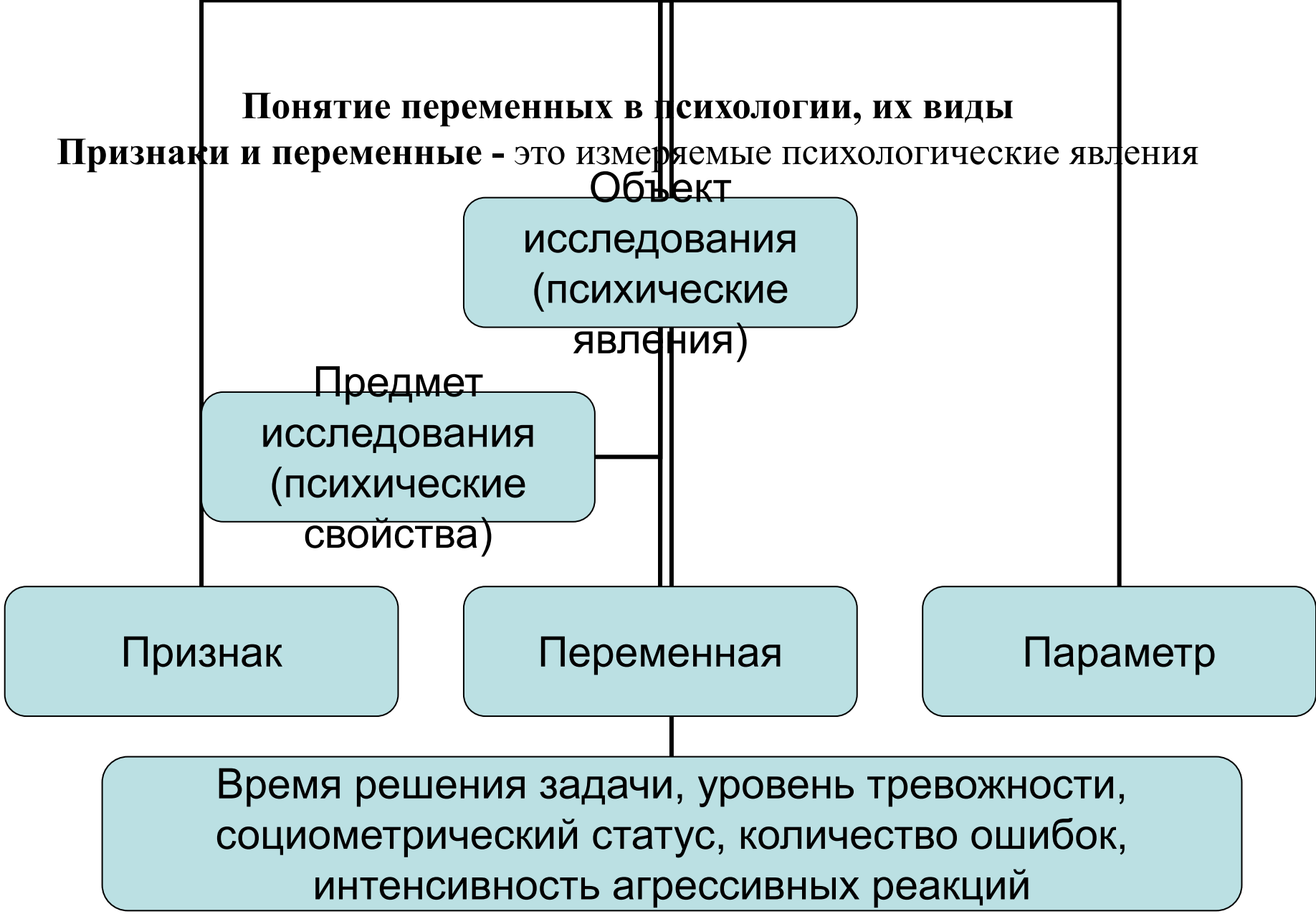


Связь «Математических методов в психологии» с другими дисциплинами



Понятие переменных в психологии, их виды

Признаки и переменные - это измеряемые психологические явления

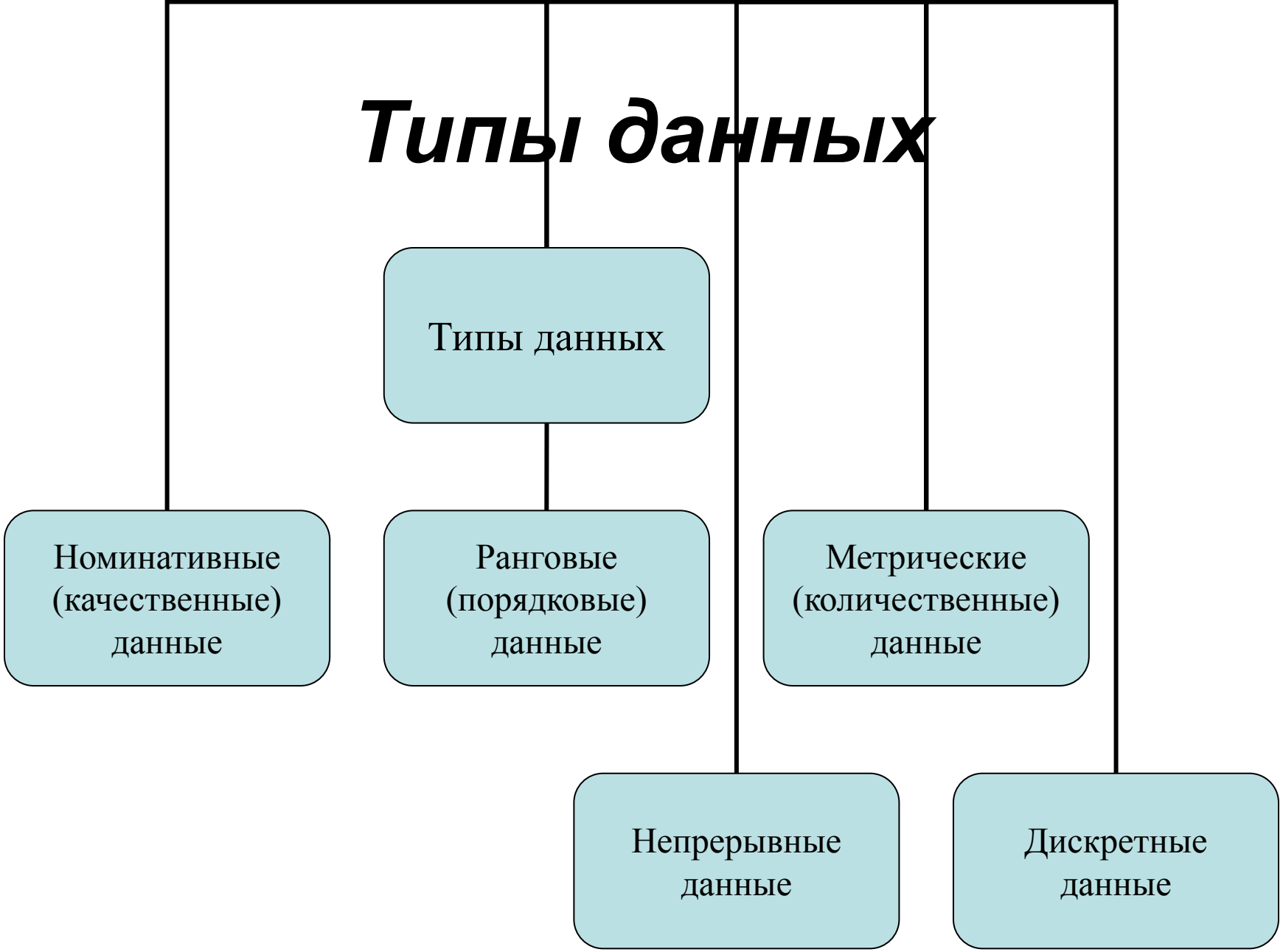


Измерение — это приписывание объекту числа по определенному правилу. Это правило устанавливает соответствие между измеряемым свойством объекта и результатом измерения — признаком.

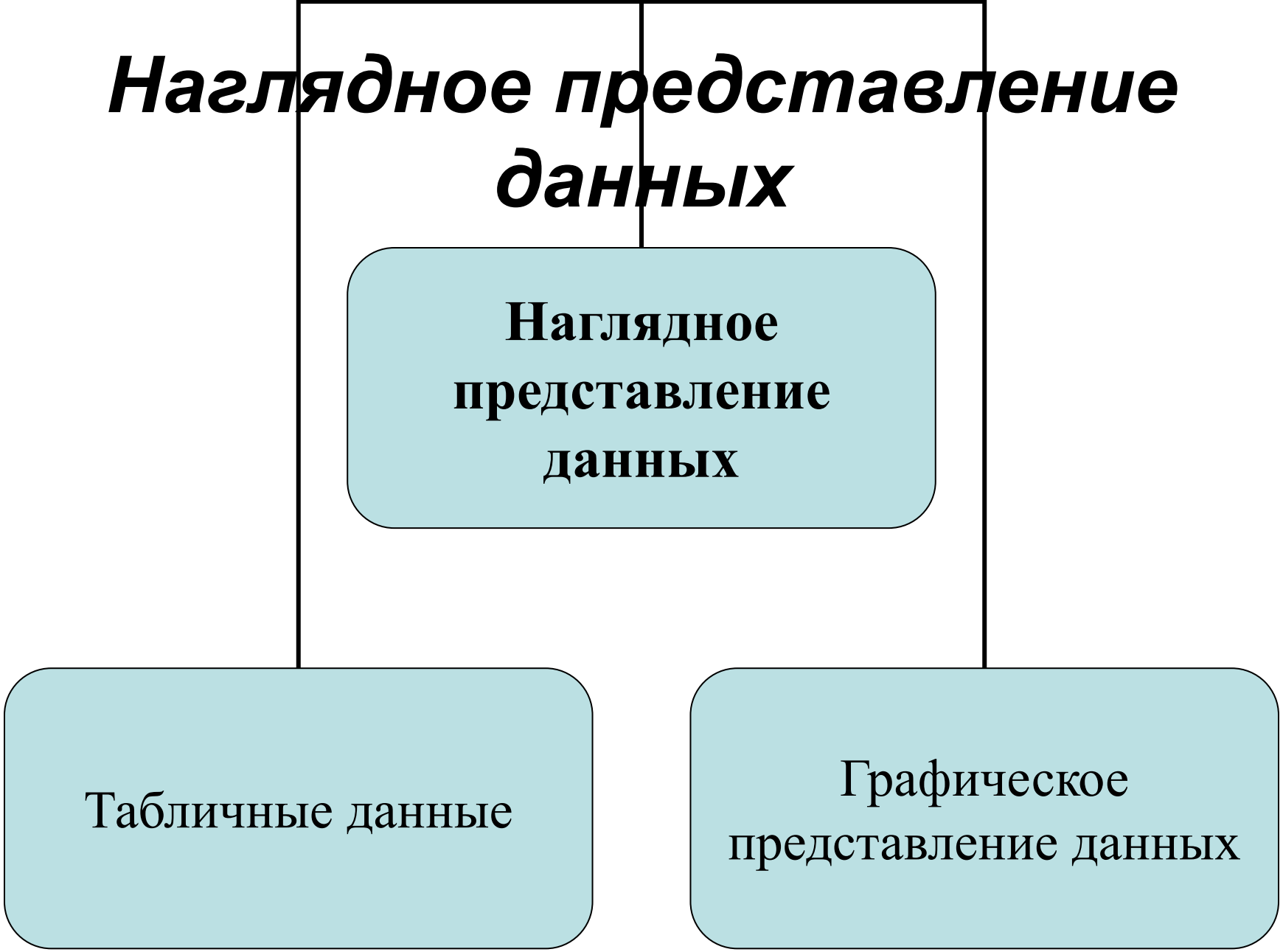
Сводка характеристик и примеры измерительных шкал

Шкала	Характеристики	Примеры
Наименований	Объекты классифицированы, а классы обозначены номерами. То, что номер одного класса больше или меньше другого, еще ничего не говорит о свойствах объектов, за исключением того, что они различаются.	Раса, цвет глаз, номера на футболках, пол, клинические диагнозы, автомобильные номера, номера страховок.
Порядковая	Соответствующие значения чисел, присваиваемых предметам, отражают количество свойства, принадлежащего предметам. Равные разности чисел не означают равных разностей в количествах свойств.	Твердость минералов, награды за заслуги, ранжирование по индивидуальным чертам личности, военные ранги.
Интервальная	Существует единица измерения, при помощи которой предметы можно не только упорядочить, но и приписать им числа так, чтобы равные разности чисел, присвоенных предметам, отражали равные различия в количествах измеряемого свойства. Нулевая точка интервальной шкалы произвольна и не указывает на отсутствие свойства.	Календарное время, шкалы температур по Фаренгейту и Цельсию.
Отношений	Числа, присвоенные предметам, обладают всеми свойствами объектов интервальной шкалы, но, помимо этого, на шкале существует абсолютный нуль. Значение нуль свидетельствует об отсутствии оцениваемого свойства. Отношения чисел, присвоенных в измерении, отражают количественные отношения измеряемого свойства.	Рост, вес, время, температура по Кельвину (абсолютный нуль).

Типы данных


Типы данных
Номинативные (качественные) данные
Ранговые (порядковые) данные
Метрические (количественные) данные
Непрерывные данные
Дискретные данные

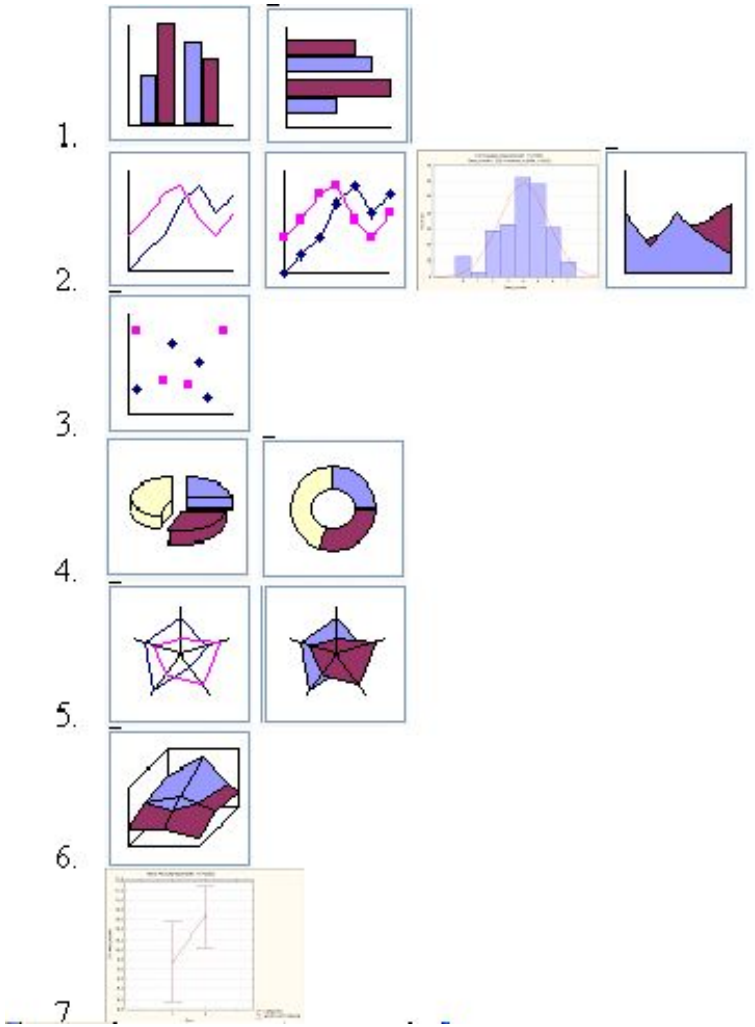
Наглядное представление данных



Графическое представление данных

В самом общем виде диаграммы делятся на:

1. Столбиковые:
   - Вертикальные;
   - Горизонтальные;
2. Линейные
   - Собственно линейные,
   - Ступенчатые,
   - Линейные с областями (профили);
3. Точечные (диаграммы рассеянья);
4. Круговые:
   - Собственно круговая,
   - Кольцевая,
5. Радиальные:
   - Звезды;
   - Лучевые;
6. Диаграммы поверхностей.
7. Комбинированные и др.



Правила графического оформления

- Вся структура графика предполагает его чтение слева направо, вертикальные шкалы — снизу вверх.
- Чтобы диаграмма не получилась сплющенной или вытянутой, выбирают такой масштаб шкалы, чтобы соотношение высоты к ширине составляли 3 к 5.
- На вертикальной шкале необходимо разместить нулевую отметку.
- Пороговые точки на шкалах желательно выделить размером или цветом, но если речь идет о временном интервале, предпочтительно не указывать начальной и конечной точек.
- Подобрать такой масштаб, чтобы кривые линии резко отличались от прямых, желательно включить в график цифровые данные и изображение формулы, а при необходимости — использовать ясные, полные заголовки и подзаголовки как для самой диаграммы, так и для ее осей.

Правила табличного представления первичных данных

- Вся структура таблицы предполагает ее чтение слева направо.
- В первом столбце предполагается размещение испытуемых.
- В последующие столбцах располагаются значения по признакам, полученные после проведения психодиагностической процедуры.

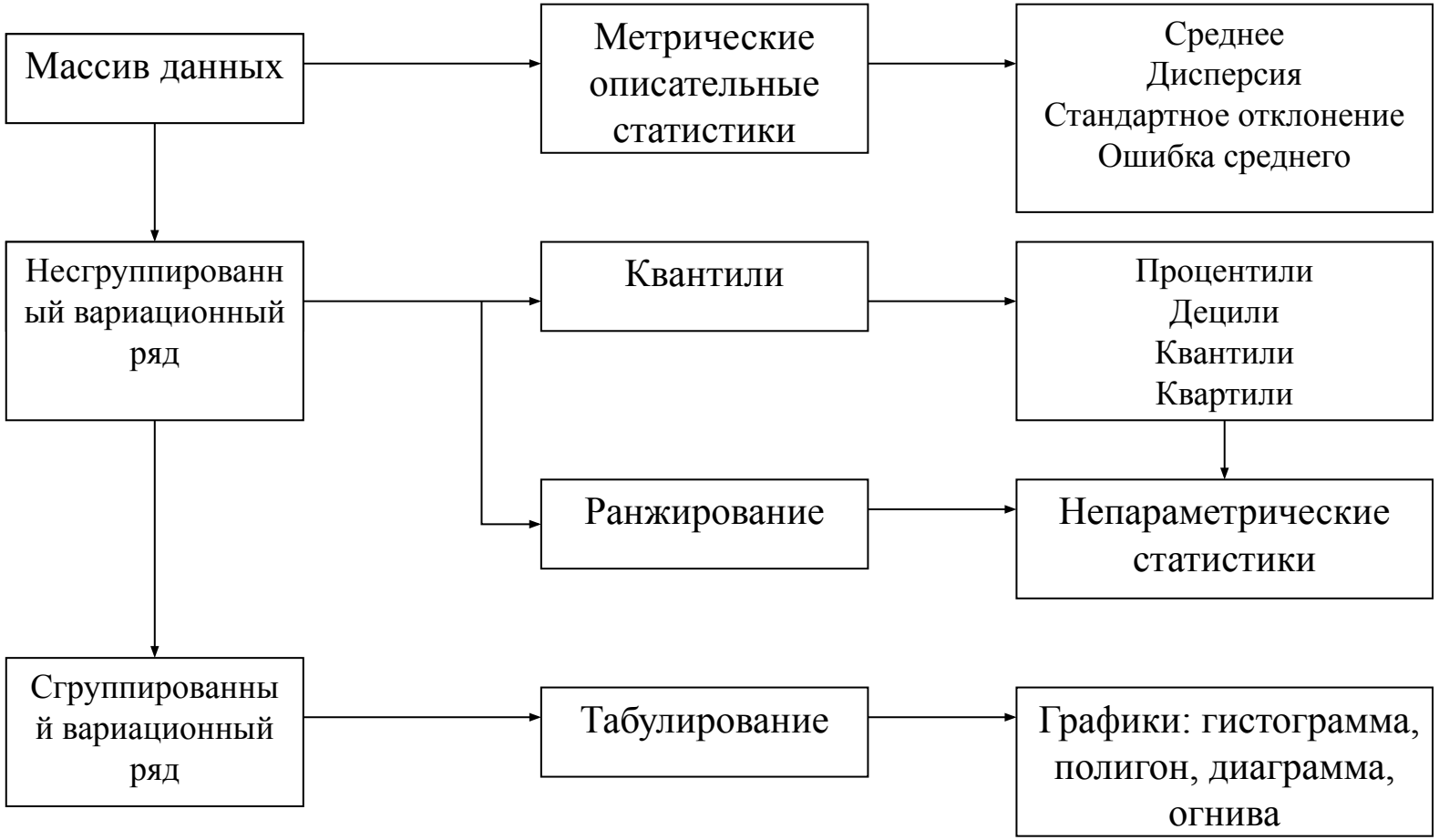
Тема 3.
Способы представления данных в психологии

- Представление данных.
- Понятие о квантилях.
- Понятие о рангах. Процедура ранжирования.
- Табулирование данных.
- Графическое представление данных.

Представление данных в психологии бывает в виде:

- **Массив данных** – первичные результаты измерения искомых параметров сводятся в одну таблицу.
- **Несгруппированный вариационный ряд** – упорядочение всех значений переменной от минимального до максимального.
- **Сгруппированный вариационный ряд** – вариационный ряд сворачивают, указывая все полученные значения однократно, а в соседнем столбце указывают частоту, с которой встречается данная оценка

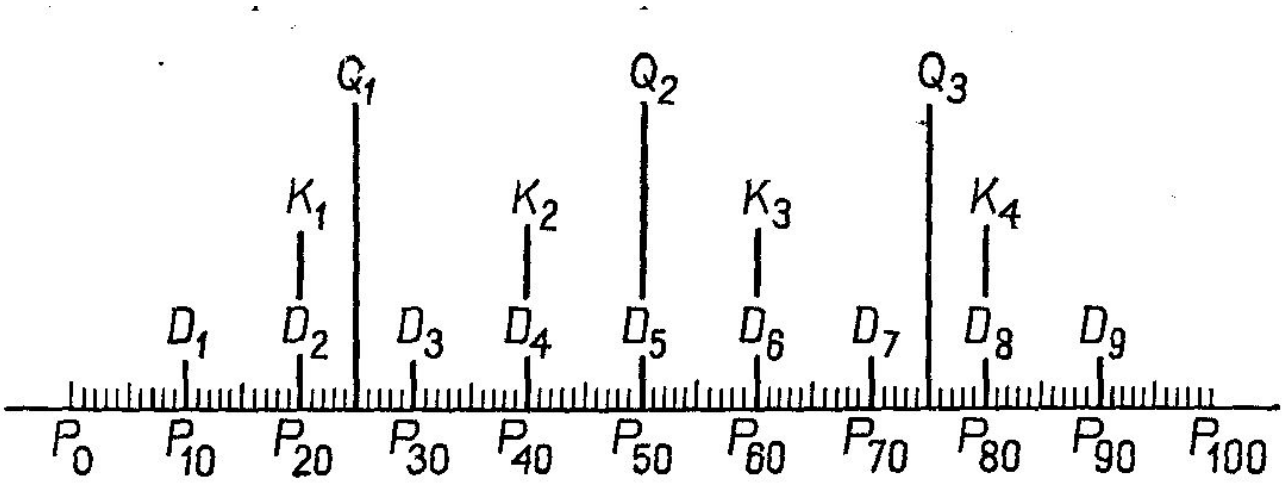
Варианты представления данных



Меры положения – квантили

Квантиль — это точка на числовой оси измеренного признака, которая делит всю совокупность упорядоченных измерений на две группы с известным соотношением их численности

- **Процентиль** *(Percentiles)* — это 99 точек — значений признака *(P1 ..., P99),* которые делят упорядоченное (по возрастанию) множество наблюдений на 100 частей, равных по численности.
- **Дециль** - это 9 точек — значений признака *(D1 ..., D9),* которые делят упорядоченное (по возрастанию) множество наблюдений на 10 частей, равных по численности.
- **Квинтель** - это 4 точки — значений признака *(K1 ..., K4),* которые делят упорядоченное (по возрастанию) множество наблюдений на 5 частей, равных по численности.
- **Квартиль** - это 3 точки — значений признака *(Q1 ..., Q3),* которые делят упорядоченное (по возрастанию) множество наблюдений на 4 части, равных по численности.



Нахождение процентиля

- ***P-й* процентиль** представляет собой точку, ниже которой лежит *P %* процентов всех наблюдений.

Формула

$$Pp = L + \frac{pn - (cum\ f)}{f},$$

где **L –** фактически нижняя граница единичного интервала оценок, содержащего частоту **pn;**

cum f - накопленная к **L** частота (до данного интервала);

f – частота оценок в интервале, содержащем частоту **pn**

Задача: Преподаватель предложил 125 учащимся контрольное задание, состоящее из 40 вопросов. В качестве оценки теста выбиралось количество вопросов, на которые были получены правильные ответы. Найти 25-й процентиль

Нахождение интервала:

- Найти между какими значениями в разряде оценок лежит накопленная pn частота (31.25 лежит между 28 и 29 значениями).
- Определить сколько единиц составляет интервал, и разделить пополам (между 28 и 29 лежит 1 / 2 = 0,5).
- Прибавить к каждому значению интервала результат второго шага (28 + 0,5 = 28,5 и 29 + 0,5 = 29,5)
- Таким образом, искомый интервал лежит между 28,5 и 29,5, а его фактически нижняя граница составляет L = 28,5.

Оценка в тесте	Частота	Накопленная частота	Вычисления
38	1	125	**Шаг 1.** $0{,}25\ n = \frac{n}{4} = \frac{125}{4} = 31{,}25$
37	1	124	
36	3	123	
35	5	120	**Шаг 2.** Найти фактическую нижнюю границу разряда оценок, содержащего оценку 31,25: $L = 28{,}5$
34	9	115	
33	8	106	
32	17	98	
31	23	81	**Шаг 3.** Вычесть накопленную к L частоту из 31,25 $31{,}25 - 16 = 15{,}25$
30	24	58	
29	18	34	
28	10	16	**Шаг 4.** Разделить результат 3-го шага на частоту f в интервале, содержащем оценку 31,25 $\frac{15{,}25}{18} = 0{,}85$
27	3	6	
26	1	3	
25	0	2	
24	2	2	
	$n = 125$		**Шаг 5.** Прибавить результат 4-го шага к L $P_{25} = 28{,}5 + 0{,}85 = 29{,}35$

Ранговый порядок

Ранжирование – это приписывание объектам чисел в зависимости от степени выраженности измеряемого свойства

- Установите для себя и запомните порядок ранжирования. Вы можете ранжировать испытуемых по их «месту в группе»: ранг 1 присваивается тому, у которого наименьшая выраженность признака, и далее — увеличение ранга по мере увеличения уровня признака. Или можно ранг 1 присваивать тому, у которого 1-е место по выраженности данного признака (например, «самый быстрый»). Строгих правил выбора здесь нет, но важно помнить, в каком направлении производилось ранжирование.
- Соблюдайте правило ранжирования для связанных рангов, когда двое или более испытуемых имеют одинаковую выраженность измеряемого свойства. В этом случае таким испытуемым присваивается один и тот же, средний ранг. Например, если вы ранжируете испытуемых по «месту в группе» и двое имеют одинаковые самые высокие исходные оценки, то обоим присваивается средний ранг 1,5: (1+2)/2 = 1,5. Следующему за этой парой испытуемому присваивается ранг 3, и т. д.

Ранжирование данных

Варианты ранжирования

Метрические данные (X_i)	Ранги (X_r)		Метрические данные (X_i)	Ранги (X_r)
15	1	← Принцип: большему значению	15	7
11	2	меньший ранг	11	6
9	3	Принцип: меньшему значению	9	5
8	4	больший ранг →	8	4
7	5		7	3
6	6		6	2
2	7		2	1

Ранжирование связанных рангов

Метрические данные (X_i)	Предварительное ранжирование (X_{R1})	Окончательное ранжирование (X_r)
12	1	1
9	2	(2+3)/2 = 2,5
9	3	(2+3)/2 = 2,5
7	4	4
6	5	5
5	6	(6+7+8)/3 = 7
5	7	(6+7+8)/3 = 7
5	8	(6+7+8)/3 = 7
4	9	9
2	10	10

Распределение частот

- **Абсолютная частота распределения (fa)** - называется частота. указывающая, сколько раз встречается каждое значение
- **Относительная частотах распределения (fo)** – называется частота, указывающая долю наблюдений, приходящихся на то или иное значение признака ($f_0 = fa / N$)
- **Накопленная частота (fcum)** – это частота показывающая, как накапливаются частоты по мере возрастания значений признака.
- **Сгруппированная частота** – это частота сгруппированная по разрядам или интервалам значений признака.

Таблица распределения частот

Значение	f_a (абсолютная частота)	f_o (относительная частота)	f_{cum} (накопленная частота)
5	3	0,05	0,05
4	12	0,20	0,25
3	21	0,35	0,60
2	15	0,25	0,85
1	9	0,15	1
Σ сумма):	60	1	—

Абсолютная и относительная частоты связаны соотношением:

$$f_0 = \frac{f_a}{N},$$

где f_a — абсолютная частота некоторого значения признака,
N — число наблюдений,
f_0 — относительная частота этого значения признака.

Табулирование данных - это методы и способы построения таблиц

Таблица 1 – Результаты исследования младших школьников

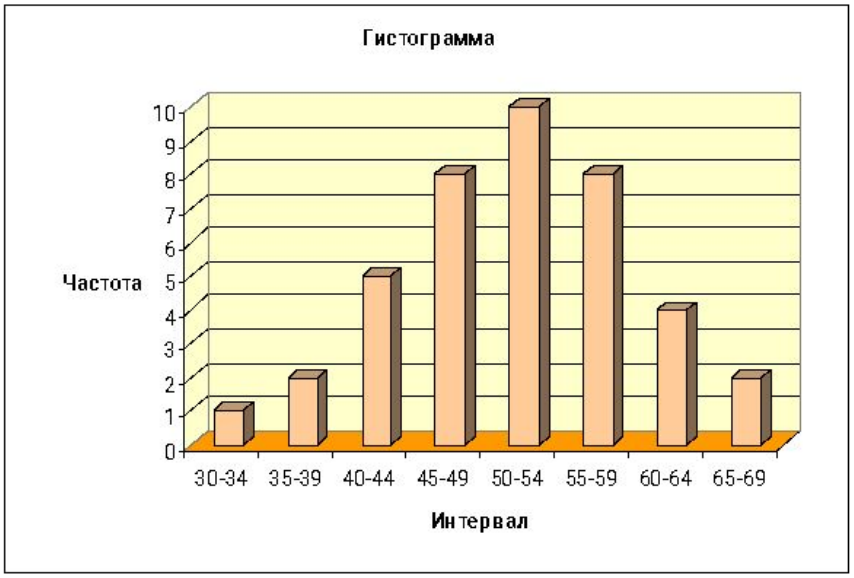
ФИО	Пол	Тревожность	Идентичность	Мотивация	Успеваемость
МИО	М	3	0	10	3
ВПР	Ж	3	1	20	5
СМТ	Ж	0	0	15	4
ВЛР	М	3	0	12	3
ЖДО	М	5	1	25	5
СТВ	М	0	1	13	3
МИН	М	4	0	18	4
КГН	М	3	1	14	3

Этапы построения распределения сгруппированных частот

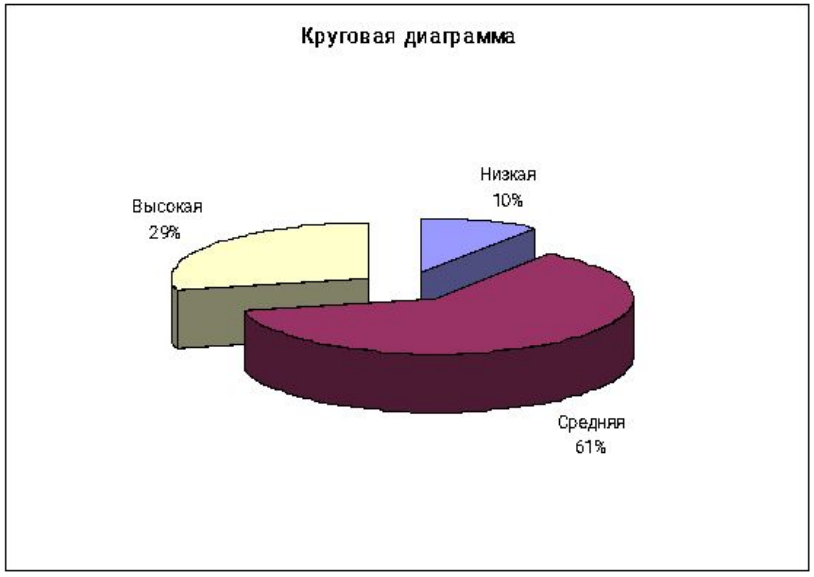
- Уточнение лимитов (крайних значений интервала) – производится округление лимитов - min и max значений: реальные лимиты max = 67и min = 32, уточненные лимиты max = 70 и min = 30.
- Определение размаха: мах – мин = 70-30 = 40
- Выбор желаемой ширины интервала разрядов l - наиболее удобной шириной интервала разрядов в является l = 5.
- Определение числа разрядов. Размах делится на интервал разряда: 40/5 = 8, получаем число разрядов — 8.
- Расчет границ интервалов, посредством прибавления к нижней границе ширину интервала.
- Подсчет абсолютной, относительной и накопленной частот

Графическое представление

- **Гистограмма** – это последовательность столбцов, каждый из которых опирается на один раздельный интервал, а высота столбца отражает количество случаев.
- **Вариационная кривая** – линия соединяющая точки, соответствующие середине каждого разрядного интервала и частоте.
- **Полигон распределения** – вариационная кривая с перпендикуляром линий до горизонтальной оси в середине каждого интервала.
- **Полигон накопленных частот (кумулята**) – на оси ординат откладывают значения суммы всех случаев лежащих в данном интервале, так и всех предыдущих интервалов. Сглаженная линия описывает все эти значения.
- **Огива (процентильная кривая)** – сглаженная линия, у которой по оси абсцисс (x) откладывают значения процентов (процентилей), а на оси ординат (y) – значения показателей.
- **Диаграмма** – отражение в долевом отношении частот на круге.


Гистограмма
Частота
10
9
8
7
6
5
4
3
2
1
0
30-34
35-39
40-44
45-49
50-54
55-59
60-64
65-69
Интервал


Полигон распределения частот
Частота
12
10
8
6
4
2
0
30-34
35-39
40-44
45-49
50-54
55-59
60-64
65-69
Интервал

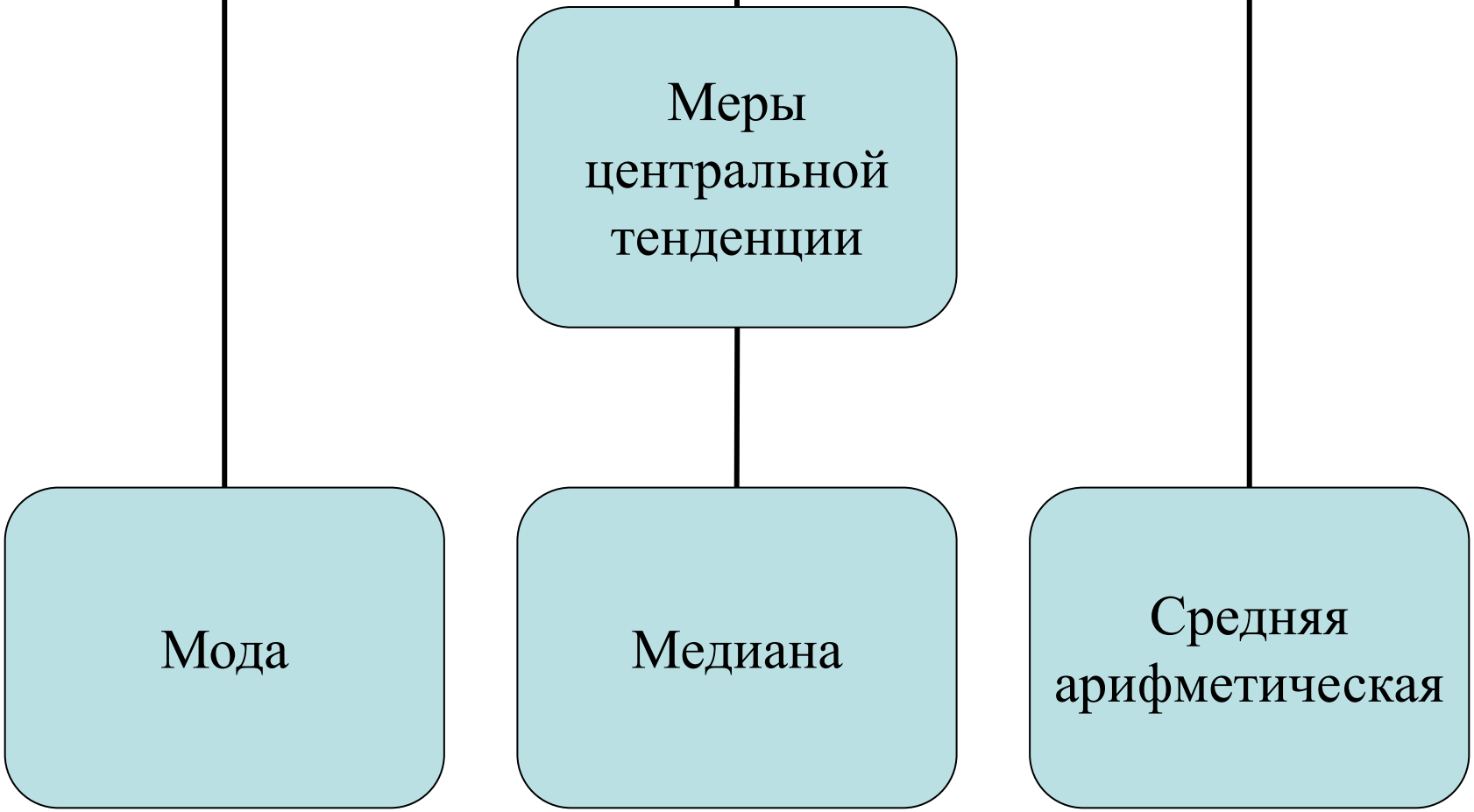

Круговая диаграмма
Высокая
29%
Низкая
10%
Средняя
61%


Кумулята
Накопленная частота
1,2
1
0,8
0,6
0,4
0,2
0
1
2
3
4
5
6
7
8
Число разрядов

Тема 4. Меры центральной тенденции

- Определение меры центральной тенденции;
- Мода;
- Медиана;
- Среднее;
- Выбор и особенности мер центральной тенденции.
- Графическое соотношение среднего, моды, медианы

Меры центральной тенденции – предназначены для замены множества значений признака, измеренного на выборке, одним числом и показывающие концентрацию группы значений на числовой шкале



Мода *(Mode)* — это такое значение из множества измерений, которое встречается наиболее часто.

- Если все значения в группе встречаются одинаково часто, то считают, что у данной выборки **моды нет** (3, 7, 4, 5, 2, 8, 1, 6 - Mo = 0).
- Если график распределения частот имеет одну вершину, то такое распределение называется **унимодальным** (3, 7, 4, 5, 7, 8, 7, 6 - Mo = 7).
- Когда два соседних значения встречаются одинаково часто и чаще, чем любое другое значение, мода есть среднее этих двух значений (3, 7, 4, 6, 7, 6, 8, 7, 6 - Mo = 6,5).
- Если два несмежных значения имеют равную и наибольшую в данной группе частоту, то у такой группы есть две моды, и распределение называют **бимодальным** (3, 7, 3, 5, 7, 3, 7, 6, 7 - Mo = 7; Mo = 3).
- Если в группе несколько значений, встречаются наиболее часто, при этом их частота может различаться, тогда выделяют наибольшую моду и локальные моды и такое распределение называют **полимодальным** (3, 7, 3, 5, 7, 3, 7, 6, 7, 10, 10. Наибольшая: Mo = 7; локальные: Mo = 3, Mo = 10).

Медиана *(Median)* — это такое значение признака, которое делит упорядоченное множество данных пополам так, что одна половина всех значений оказывается меньше медианы, а другая — больше.

- Первым шагом при определении медианы является упорядочивание (ранжирование) всех значений по возрастанию или убыванию.
- Если данные содержат нечетное число значений (8, 9, 10, 13, 15), то медиана есть центральное значение, т. е. Md= 10.
- Если данные содержат четное число значений (5, 8, 9, 11), то медиана есть точка, лежащая посередине между двумя центральными значениями, т. е. M/=(8+9)/2 = 8,5.

Среднее *(Mean)* *(M* — выборочное среднее, среднее арифметическое) — определяется как сумма всех значений измеренного признака, деленная на количество суммированных значений.

$$M_x = \frac{1}{N}\sum_{i=1}^{N} x_i \, .$$

- Если к каждому значению переменной прибавить одно и то же число c, то среднее увеличится на это число (уменьшится на это число, если оно отрицательное).

$$M_{(x_i+c)} = \frac{1}{N}\sum_{i=1}^{N}(x_i + c) = M_x + c \, .$$

- Если каждое значение переменной умножить на одно и то же число *c,* то среднее увеличится в *c* раз (уменьшится в *c* раз, если делить на c).

$$M_{(x_i \cdot c)} = \frac{1}{N}\sum_{i=1}^{N}(x_i \cdot c) = M_x \cdot c \, .$$

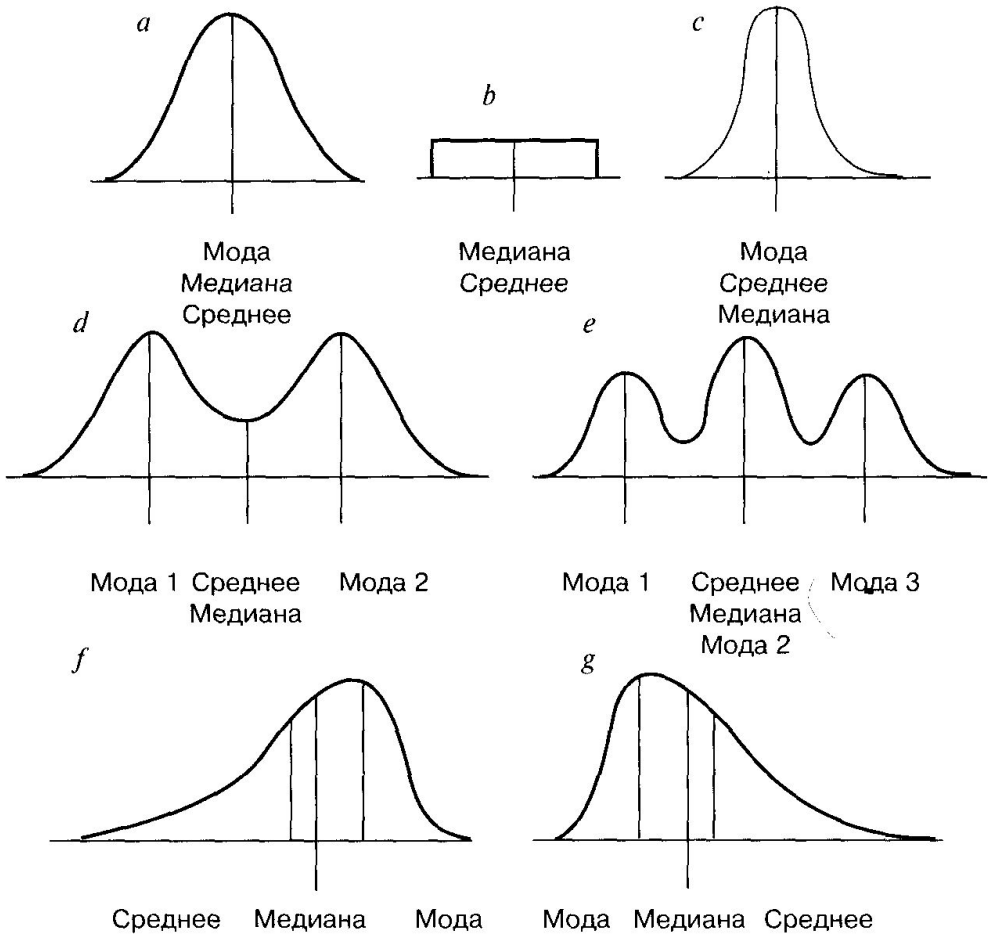
- Сумма всех отклонении от среднего равна нулю.

$$\sum_{i=1}^{N}(x_i - M_x) = 0 \, .$$

Выбор и особенности мер центральной тенденции

- Для **номинативных** данных единственной подходящей мерой центральной тенденции является мода.
- В малых группах мода нестабильна.
- Для метрических и порядковых данных наиболее подходящей мерой являются медиана и средняя арифметическая.
- На медиану не влияет величины очень больших и очень малых значений
- На величину среднего влияет каждое значение, оно чувствительно к «выбросам» — экстремально малым или большим значениям переменной.
- Наиболее устойчива к выбросам средняя гармоническая , при расчете которой используются обратные величины.
- Если распределение симметричное и унимодальное, то мода, средняя и медиана совпадают.

Графическое соотношение среднего, моды, медианы



Сравнение преимуществ и ограничений мер центральной тенденции

Мера	Преимущества	Ограничения
Среднее арифметическое. «Центр тяжести» данных. Равно сумме значений всего ряда дан ных, деленной на количество этих значений	Выборочная стабильность — менее всего изменяется от выборки к выборке. Поддается математической обработке: может быть использована при подсчете дальнейших статистик. Отражает действительную ценность каждого показателя и поэтому содержит больше информации, относящейся к полному набору данных.	Не используется: — если распределе ние скошено; — когда значение экстремальных случаев неизвестно. Не используется в номинальной и по рядковой шкалах
Медиана. Разделяет предварительно упо рядоченные данные на две равные по размеру части	Лучше всего репрезентирует центр сильно скошенного распределения (не подвержена влиянию экстремальных значений). Может быть подсчитана, когда экстремальные значения неизвестны	Зависит от величины принятого интервала (для сгруппированных данных). Редко используется в даль нейших статистиках. Не используется в номинальной шкале
Мода. Наиболее часто встречаемое явление.	Полезна для неупорядоченных качественных переменных. Быстро дает представление о типичном по группе. Ее очень легко посчитать. Малочувствительна к экстремаль ным значениям	Зависит от принятого интервала (для сгруппированных данных). Редко ис пользуется в дальнейших статистиках. Может отсутствовать для некоторых сгруппированных данных

Тема 5. Меры изменчивости

- Понятие меры изменчивости
- Лимиты. Размах вариации и его разновидности.
- Дисперсия и ее свойства.
- Стандартное отклонение.
- Асимметрия и эксцесс.

Меры изменчивости

Меры рассеяния независящие от распределения

- **Лимиты –** это характеристики, определяющие верхнюю (max) и нижнюю (min) границы значений показателя.
- **Размах *(Range)*** — это разность максимального и минимального значений: R = max – min.
- **Исключающий размах** - это разность максимального и минимального значений в группе.
- **Включающий размах** - это разность между естественной верхней границей интервала, содержащего максимальное значение, и естественной нижней границей интервала, включающей минимальное значение.
- Размах это очень неустойчивая мера изменчивости, на которую влияют любые возможные «выбросы». Более устойчивыми являются разновидности размаха: **размах от 10 до 90-го процентиля** $R = P_{90} – P_{10}$ или **полумежквартильный размах:**

$$Q = \frac{Q_3 - Q_1}{2}.$$

Меры рассеяния характеризующие нормальное распределение

Дисперсия *(Variance)* — мера изменчивости для метрических данных, пропорциональная сумме квадратов отклонений измеренных значений от их арифметического среднего:

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

Свойства дисперсии:

1. Если значения измеренного признака не отличаются друг от друга (равны между собой) — дисперсия равна нулю. Это соответствует отсутствию изменчивости в данных.
2. Прибавление одного и того же числа к каждому значению переменной не меняет дисперсию.
3. Умножение каждого значения переменной на константу *c* изменяет дисперсию в *c* раз.
4. При объединении двух выборок с одинаковой дисперсией, но с разными средними значениями дисперсия увеличивается.

Расчет дисперсии

N	x_i	$(x_i - M_x)$	$(x_i - M_x)^2$	Вычисления
1	4	1	1	$D_x = \frac{\sum_{i=1}^{N}(x_i - M_x)^2}{N-1}$.
2	2	-1	1	
3	4	1	1	
4	1	-2	4	
5	5	2	4	
6	2	-1	1	
Σ	18	0	12	$M_x = 18/6 = 3$ $D_x = 12/ (6-1) = 2,4$ $\sigma_x = \sqrt{2,4} = 1,549$

Меры рассеяния характеризующие нормальное распределение

- **Стандартное отклонение *(Std. deviation)*** (сигма, среднеквадратическое отклонение) — положительное значение квадратного корня из дисперсии, говорит о том, на сколько могут значимо отклоняться, изменяющиеся данные :

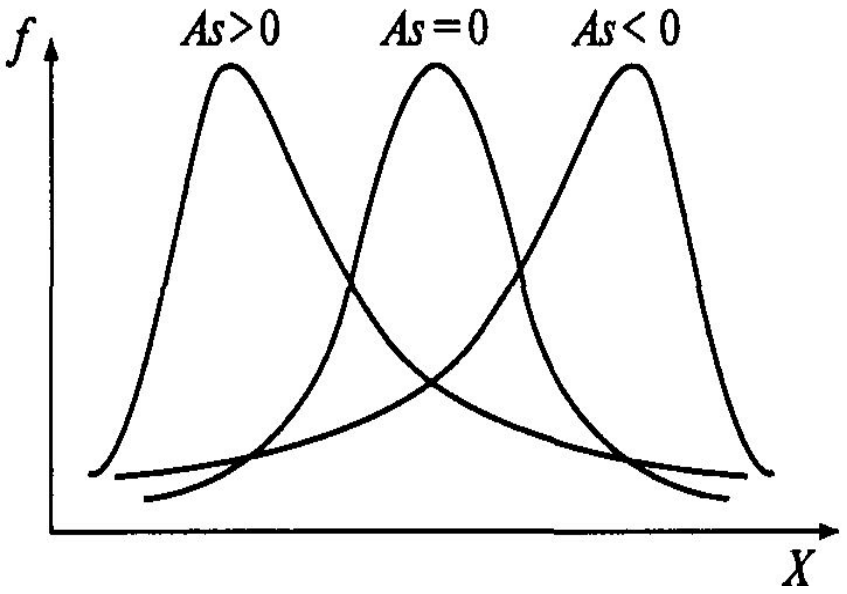
$$\sigma_x = \sqrt{D_x} = \sqrt{\frac{\sum_i (x_i - M_x)^2}{N-1}}.$$

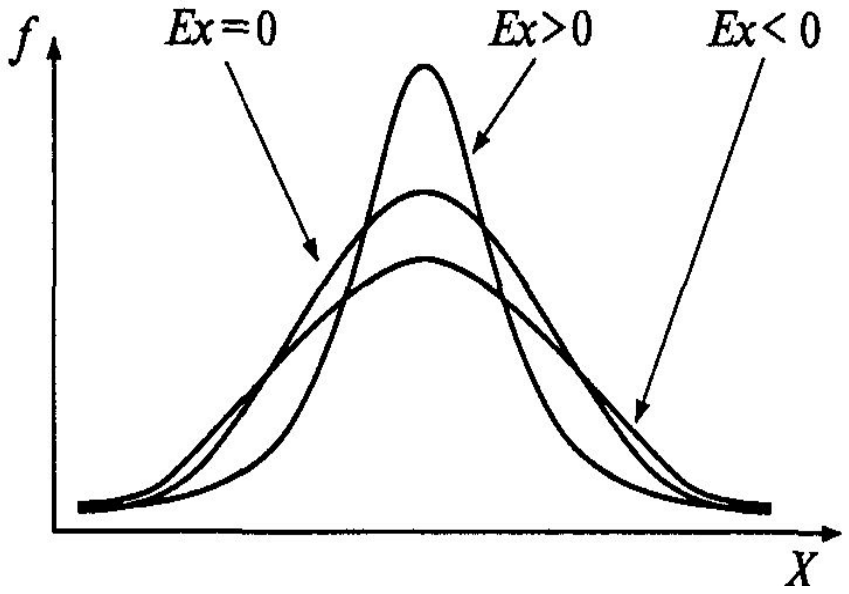
- **Ошибка среднего значения (error of mean)** - среднеарифметическое значение среднеквадратичного отклонения, она говорит о том, на сколько могут отклониться данные при повторном исследовании:

$$m = s_0 = \frac{s_x}{\sqrt{n-1}}.$$

Меры формы

- **Асимметрия** ***(Skewness)*** — степень отклонения графика распределения частот от симметричного вида относительно среднего значения: $A = \frac{\Sigma(x_i - M_x)^3}{n \times \sigma^3}$.
- **Эксцесс** ***(Kurtosis)*** — мера плосковершинности или остроконечности графика распределения измеренного признака. $E = \frac{\Sigma(x_i - M_x)^4}{n \times \sigma^4} - 3$.





Тема 6. Стандартизация данных

- Понятие стандартизации данных.
- Основные формы стандартизации.
- z-преобразование данных.

Стандартизация (англ. standard нормальный) — унификация, приведение к единым нормативам процедуры и оценок теста.

Различают две формы стандартизации

1. В первом случае под С. понимаются обработка и регламентация процедуры проведения, унификация инструкции, бланков обследования, способов регистрации результатов, условий проведения обследования, характеристика контингентов испытуемых.
2. Во втором случае под С. понимается преобразование нормальной (или искусственно нормализованной) шкалы оценок в новую шкалу, основанную уже не на количественных эмпирических значениях изучаемого показателя, а на его относительном месте в распределении результатов в выборке испытуемых.

Преобразование первичных оценок в новую шкалу

- **Центрирование** – это линейная трансформация величин признака, при котором средняя величина распределения становится равной нулю (M ±σ – нормативный диапазон).
- **Нормирование** - это переход к другому масштабу (единицам) измерения, называемый **z-преобразованием** данных. **z-преобразование** данных — это перевод измерений в стандартную *Z-шкалу* со средним Mz = 0 и Dz (или σ z) = 1.

Этапы перехода к другому масштабу

- Для переменной, измеренной на выборке, вычисляют среднее по выборке, индивидуальный показатель (или среднее каждого испытуемого) **Mx,** стандартное отклонение **σx**.
- Все значения переменной xi пересчитываются по формуле:

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

- Перевод в новую шкалу осуществляется путем умножения каждого z-значения на заданную сигму и прибавления среднего:

$$S_i = \sigma_S z_i + M_S.$$

- Известные шкалы: IQ (среднее 100, сигма 15); Т-оценки (среднее 50, сигма 10); 10-балльная — стены (среднее 5,5, сигма 2) и др.

Пример преобразования в z-значения, Т-баллы

№ п/п	Косвенная агрессия	Преобразование в z-значения		Преобразование в Т-баллы	
		$x_i - X$	$\frac{(x_i - X)}{\sigma}$	$\frac{/x_i - X/}{\sigma} *10$	$\frac{((x_i - X) *10)}{\sigma} + 50$
1	8	2,75	1,61	16	66
2	4	-1,25	-0,73	7	57
3	3	-2,25	-1,32	13	63
4	5	-0,25	-0,15	1	51
5	5	-0,25	-0,15	1	51
6	7	1,75	1,02	10	50
7	5	-0,25	-0,15	1	51
8	6	0,75	0,44	4	54
9	5	-0,25	-0,15	1	51
10	8	2,75	1,61	16	66
11	3	-2,25	-1,32	13	63
12	4	-1,25	-0,73	7	57
X	**5,25**		**0**		**56,6**
σ	**1,71**		**1**		**6,3**

Тема 7. Теоретические распределения, используемые при статистических выводах

- Нормальное распределение
- Единичное нормальное распределение и его свойства
- Соответствия между диапазонами значений и площадью под кривой
- Проверка нормальности распределения

Виды распределения данных

Нормальное распределение Распределение Стьюдента	Показывает распределения выраженности количественного признака (метрических данных)
Биномиальное распределение Пуассоновское распределение	Показывает распределения качественного признака (номинативных данных). Сортирует группы по числу объектов, обладающих признаком

- **Нормальное распределение.** Нормальный закон распределения состоит в том, что чаще всего встречаются средние значения соответствующих показателей, и чем больше отклонение от этой средней величины в меньшую или большую сторону встречаются одинаково реже чем среднее значение.

Единичное нормальное распределение и его свойства

Если применить z-преобразование ко всем возможным измерениям свойств, все многообразие нормальных распределений может быть сведено к одной кривой. Тогда каждое свойство будет иметь среднее 0 и стандартное отклонение 1. Это и есть **единичное нормальное распределение,** которое используется как стандарт — эталон.

Свойства единичного нормального распределения

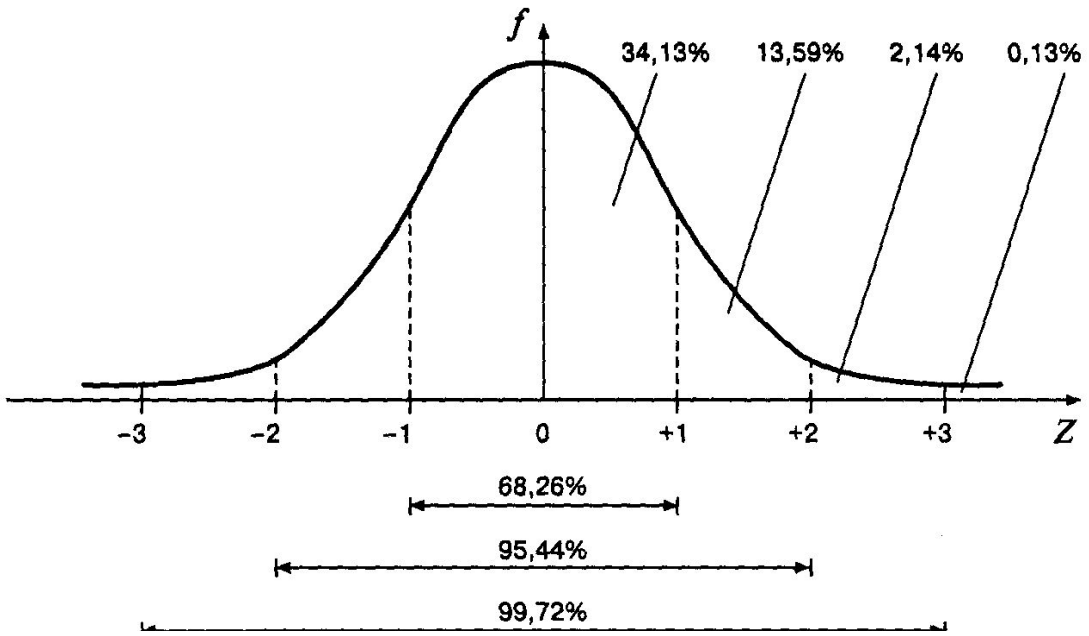
□ Единицей измерения единичного нормального распределения является стандартное отклонение.

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

□ Кривая симметрична относительно $M = 0$. Ее асимметрия и эксцесс равны нулю.

□ Кривая имеет характерный изгиб: точка перегиба лежит точно на расстоянии в одну σ от *M*.

□ Площадь между кривой и осью Z равна1.

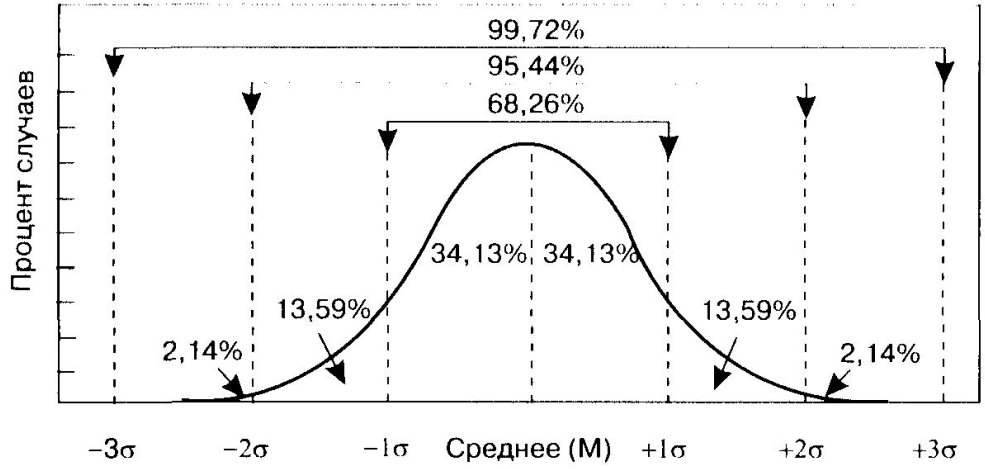


Соответствия между диапазонами значений и площадью под кривой

- *M*± σ соответствует ≈ 68% (точно — 68,26%) площади;
- M±2σ соответствует ≈ 95% (точно — 95,44%) площади;
- *M*± *3*σ соответствует ≈ 100% (точно — 99,72%) площади.

Если распределение является нормальным, то:

- 90% всех случаев располагается в диапазоне значений *M*± 1,64σ;
- 95% всех случаев располагается в диапазоне значений *M*± 1,96σ;
- 99% всех случаев располагается и диапазоне значений *M*± 2,58σ.



Проверка нормальности распределения

1. Нормальность распределения результативного признака можно проверить путем расчета показателей асимметрии и эксцесса по Н.А. Плохинскому, которые определяется по формулам:

$$t_A = \frac{|A|}{m_A} \geq 3$$

где |A| - абсолютная величина асимметрии;

$m_{A\,-}$ стандартная ошибка асимметрии.

$$t_E = \frac{|E|}{m_E} \geq 3$$

где |E| - абсолютная величина эксцесса;

$m_{E\,-}$ стандартная ошибка

Показатели асимметрии и эксцесса свидетельствуют о достоверном отличии эмпирических распределений от нормального в том случае, если они превышают по абсолютной величине свою ошибку репрезентативности в 3 и более раз. Все значения t_A и t_E не превышают свою ошибку репрезентативности в три раза, из чего можно заключить, что распределение признака не отличается от нормального.

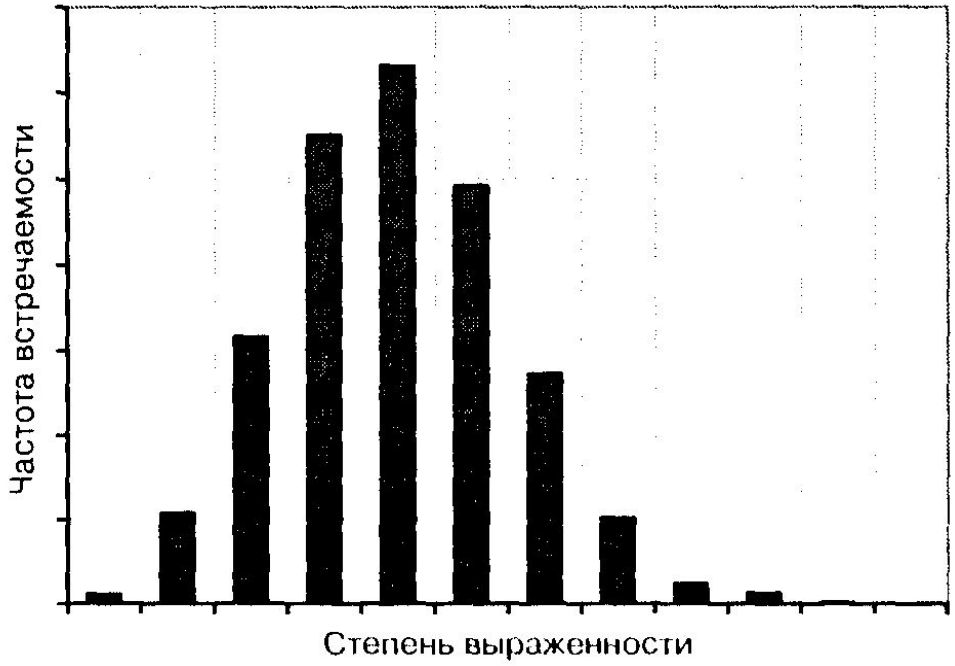
2. Еще одним из критериев проверки на нормальность - является критерий Колмагорова-Смирнова.

- Он позволяет оценить вероятность того, что данная выборка принадлежит генеральной совокупности с нормальным распределением.
- Вероятность **$p \leq 0,05$,** распределение **отличается от нормального.**
- Вероятность **$p > 0,05$**, распределение **соответствует нормальному**.

Биноминальное распределение

- **Биноминальное распределение** связано со случайными событиями, имеющими определенную постоянную степень вероятности. Оно отражает распределение вероятностей числа появления какого-либо бинарного параметра (именно бинарного, а не метрического) при повторных независимых измерениях в сходных условиях.

Кривая биномиального распределения



Распределение Пуассона

- Распределение Пуассона описывает случайные (редкие) события, вероятность появления которых в отдельных случаях мала, но число этих случаев достаточно велико.

Кривая распределения Стьюдента

- Для выборок с числом наблюдений 30 или более, распределение Стьюдента равно нормальному распределению. При меньшем количестве наблюдений оно отличается от нормального, становится более плоским.

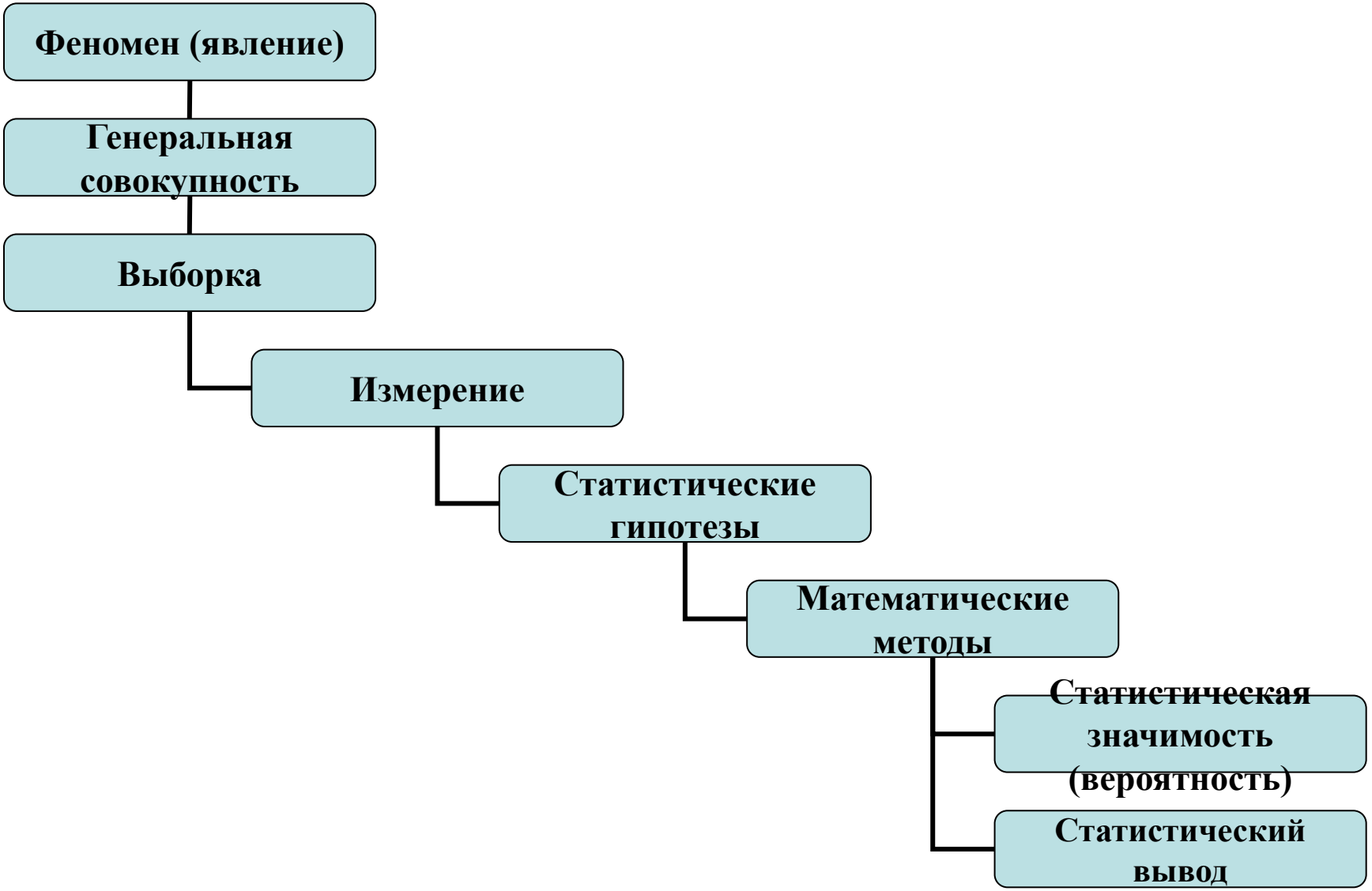
Кривая распределения Фишера

- Распределение Фишера описывает значения F при случайном выборе из одной генеральной совокупности *m* групп по *n* объектов.
- Связь с распределением Стьюдента обусловлена простым соотношением: t2 = F.

Тема 8. Статистическое оценивание и проверка гипотез

- Понятие генеральной совокупности и выборки
- Виды вероятностной выборки
- Зависимые и независимые выборки
- Определение объема выборки при нормальном распределении
- Статистические гипотезы.
- Статистический критерий.
- Степень свободы.
- Уровень значимости.
- Статистический вывод.
- Ошибки 1 и 2 рода.

Этапы статистического вывода



Понятие генеральной совокупности и выборки

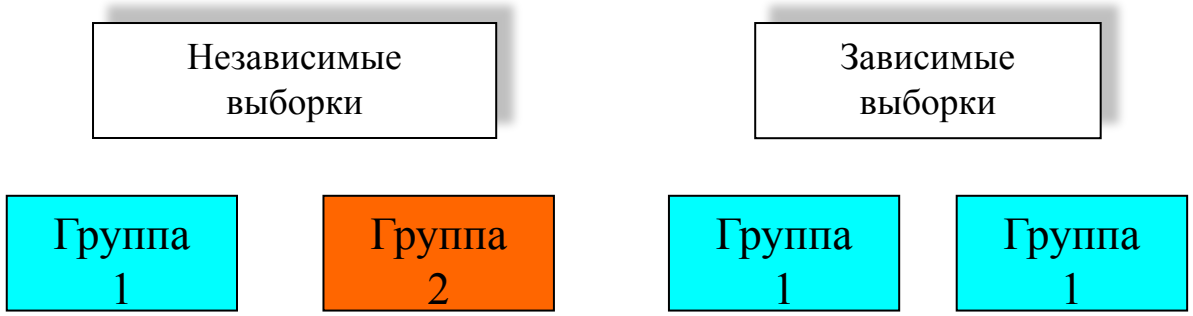
- **Генеральной совокупностью** – называется всякая большая (конечная или бесконечная) коллекция или совокупность предметов, которые мы хотим исследовать.
- **Выборка** — это часть или подмножество совокупности. Выборка называется **репрезентативной** если она адекватно отражает свойства генеральной совокупности.
- Репрезентативность достигается методом **рандомизации,** т. е. случайным отбором объектов из генеральной совокупности.

Виды вероятностной выборки

Генеральная
совокупность

- **Случайная выборка –** сформированная на основе случайного отбора.
- **Минус случайной выборки:** отобранная часть популяции может существенно отличаться от популяции в целом.
- **Стратифицированная выборка –** отражающая особенности популяции.
- **Групповая выборка (кластерная) –** это группа людей, имеющих определенную особенность, не важную с точки зрения исследуемых переменных.
- **Простая выборка –** это выборки с наиболее часто встречаемыми признаками в популяции.

Зависимые и независимые выборки



- **Независимые выборки** характеризуются тем, что вероятность отбора любого испытуемого одной выборки не зависит от отбора любого из испытуемых другой выборки.
- **Зависимые выборки** характеризуются тем, что каждому испытуемому одной выборки поставлен в соответствие по определенному критерию испытуемый из другой выборки или это тот же самый испытуемый при повторном измерении. В общем случае зависимые выборки предполагают попарный подбор испытуемых в сравниваемые выборки, а независимые выборки — независимый отбор испытуемых.

Объем выборки – определяется численностью входящих в нее элементов. Объем выборки зависит от целей и методов исследования, от гомогенности генеральной совокупности, от принимаемой исследователем погрешности.

- Объем выборки для нормального распределения определяется по формуле:

$$n = \frac{t^2 \cdot \sigma^2}{\Delta^2}$$

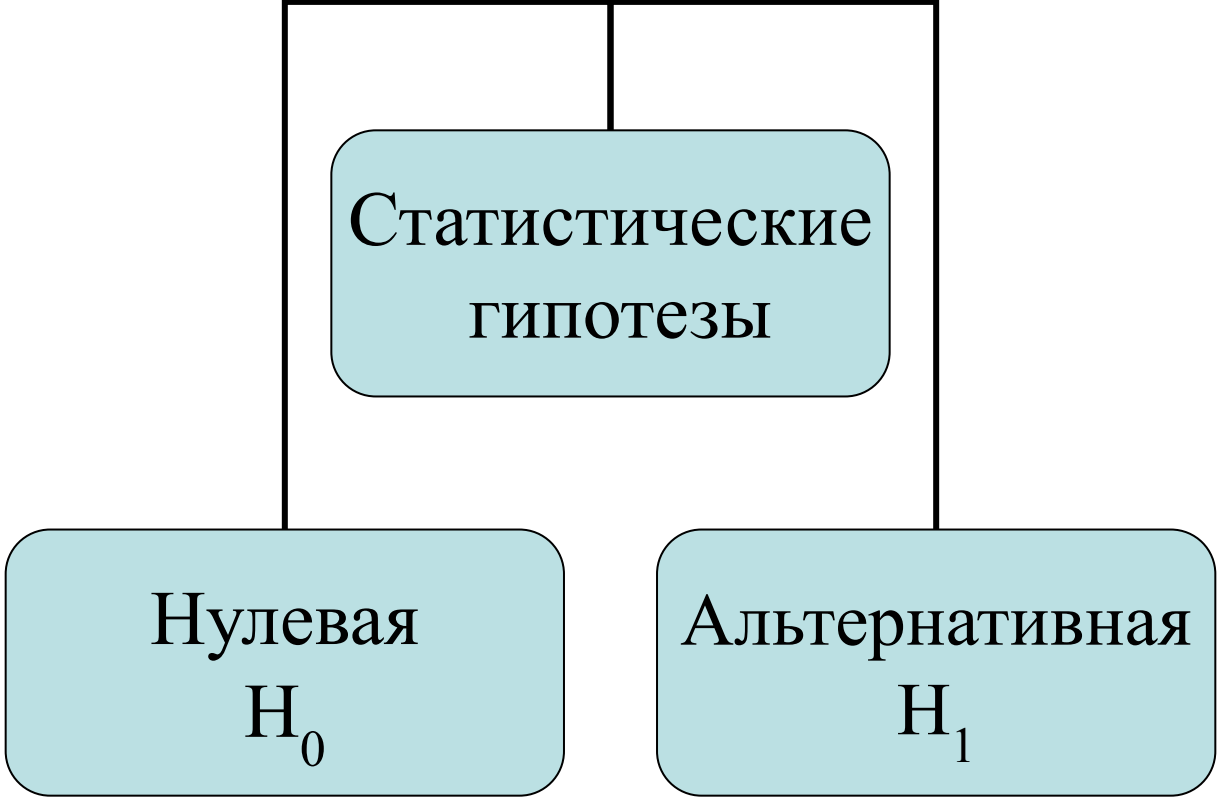
где ***n*** — объем выборки;

- ***t*** — табулированное значение абсциссы для кривой нормального распределения, определяемое желаемой точностью оценки (для наиболее распространенных p = 0,95 ***t*** = 1,96; для p = 0,99 ***t*** = 2,58);
- Δ — предельная репрезентативность выборки (обычно задается исследователем в пределах от 10% до 1% погрешности соответственно);
- σ — дисперсия признака в генеральной совокупности.

- **Гипотеза –** это утверждение, истинность или ложность которого неизвестны, но могут быть проверены опытным путем
- **Статистическая гипотеза** — это утверждение относительно неизвестного параметра генеральной совокупности, которое формулируется для проверки надежности связи и которое можно проверить по известным выборочным статистикам (критерий).

Варианты гипотез

- 1.О (различии) значении генеральных параметров;
- 2.О (взаимосвязи) отличии параметров от нуля;
- 3.О (нормальности распределения) законе распределения.



Нулевая гипотеза - это гипотеза об отсутствии различий. Она обозначается как H0 и называется нулевой потому, что содержит число 0: X1—X2=0, где X1, X2 - сопоставляемые значения признаков. Нулевая гипотеза - это то, что мы хотим опровергнуть, если перед нами стоит задача доказать значимость различий.
Альтернативная гипотеза - это гипотеза о значимости различий. Она обозначается как H1. Альтернативная гипотеза - это то, что мы хотим доказать, поэтому иногда ее называют *экспериментальной* гипотезой.

Статистический критерий

Статистический критерий – это решающее правило, обеспечивающее надежное поведение, т.е. принятие истинной и отклонение ложной гипотезы с высокой вероятностью.

Статистический критерий обозначает также метод расчета определенного числа и само это число

Мощность критерия – это его способность выявлять различия, если они есть (т.е. это его способность не допустить ошибку).

Критерий включает в себя:

- формулу расчета эмпирического значения критерия по выборочным статистикам;
- правило (формулу) определения числа степеней свободы;
- теоретическое распределение для данного числа степеней свободы;
- правило соотнесения эмпирического значения критерия с теоретическим распределением для определения вероятности того, что Но верна.


Критерии делятся
Параметрические, включающие в формулу расчета параметры распределения: средние и дисперсии (t-Стьюдента, ANOVA)
Непараметрические, основанные на оперировании частотами или рангами (Т-Уилкоксон, U=-Манна-Уитни)

Основание выбора критерия

а) в какой шкале представлены признаки;

б) мощность критерия

в) применимость по отношению к неравным по объему выборкам;

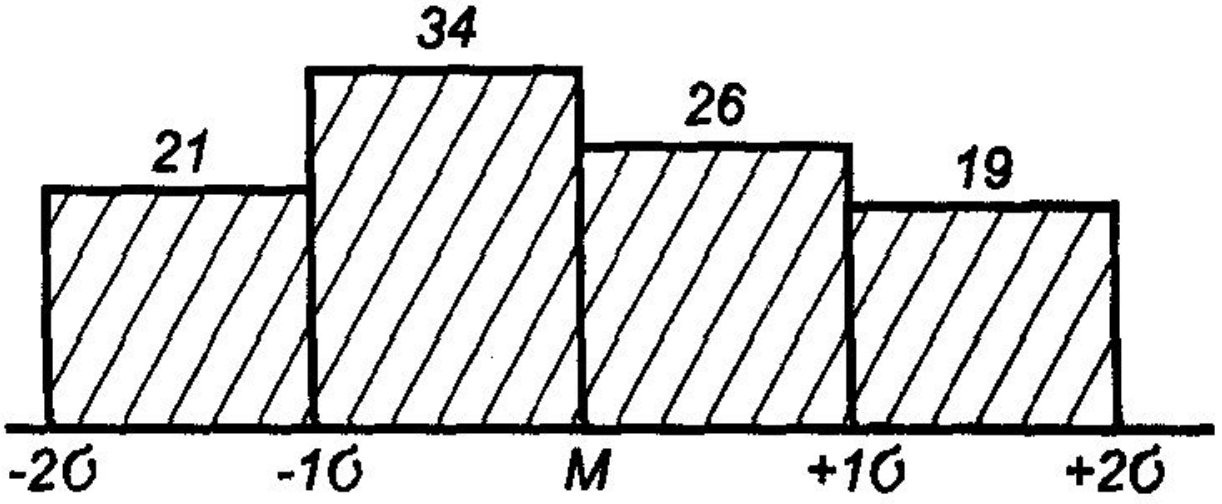
г) выполнение ограничения.

Степень свободы

Число степеней свободы – это количество возможных направлений изменчивости признака.

Это характеристика распределения, используемая при проверке статистических гипотез, отражающая степень произвольности вариантов заполнения определенных групп, на которые квантифицируется распределение **(обозначается как *df* или n-1**).

Вариант заполнения интервалов оценок в выборке из 100 обследованных степень свободы равна трем *(df* = *k*-1= 4-1=3*).*



Показатели степеней свободы для зависимых и независимых выборок

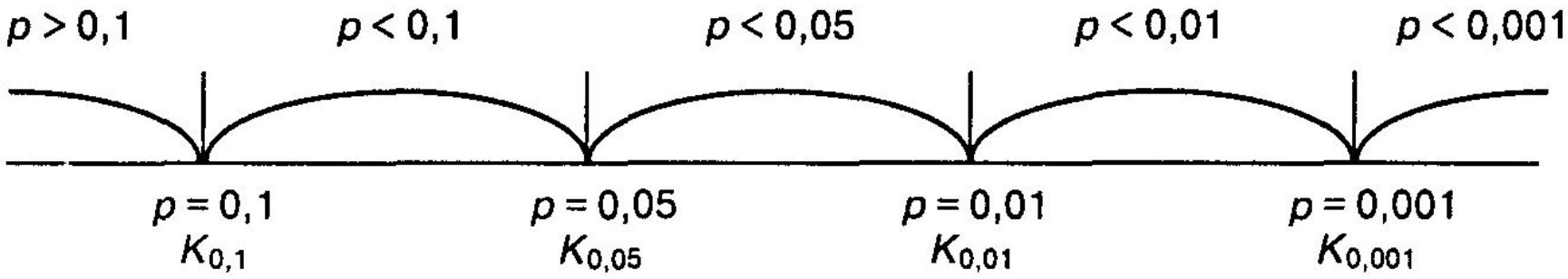
Если имеются две независимые выборки, то число степеней свободы для первой из них составляет $n_1 - 1$, а для второй $n - 1$. таким образом, число степеней свободы для этих независимых выборок будет составлять $(n_1 + n_2) - 2$.

В случае зависимых выборок число степеней свободы равно $n - 1$.

Показатель степени свободы наиболее широко используется при расчете статистических гипотез с использованием критериев Стьюдента, Фишера, z-критерия, *критерия* χ^2. При применении каждого критерия и в каждом конкретном случае его использования существуют свои правила определения количества степеней свободы.

Статистическая значимость (Significant level, сокращенно **Sig***.),* или ***p-уро*вень значимости (p-level),** — основной результат проверки статистической гипотезы, это вероятность получения различий в выборке исследования при условии, что на самом деле для генеральной совокупности верна нулевая статистическая гипотеза — то есть различий нет.

Схема определения p – уровня



Свойства статистической значимости

Чем меньше значение p-уровня, тем выше статистическая значимость результата исследования, подтверждающего научную гипотезу.

Уровень значимости при прочих равных условиях выше (значение p-уровня меньше), если:

- величина связи (различия) больше;
- изменчивость признака (признаков) меньше;
- объем выборки (выборок) больше.

Статистический вывод — это формулирование вывода на основе статистической значимости.

Статистический вывод — это рассуждение от частного к общему, от явного к неявному. Рассуждения статистического вывода помогают ответить на такой вопрос, как: «Что мне известно, если даны определенные показатели и произведен математико-статистический расчет и известен уровень значимости».

Ошибки 1 и 2 рода

- **Ошибка I рода** - ошибка, состоящая в том, что мы отклонили H_0, в то время как она верна.

Вероятность такой ошибки - α (или p), вероятность правильного решения: 1- α. Чем меньше α, тем больше вероятность правильного решения.

- **Ошибка II рода** - ошибка, состоящая в том, что мы приняли H_0, в то время как она не верна.

Вероятность такой ошибки β. Вероятность (1 — β) называется **мощностью** (чувствительностью) критерия. Эта величина характеризует статистический критерий с точки зрения его способности отклонять H_0, когда она не верна.

Традиционная интерпретация уровней значимости при $\alpha = 0{,}05$

Уровень значимости	Решение	Возможный статистический вывод
$p > 0{,}1$	Принимается H_0	«Статистически достоверные различия не обнаружены»
$p \leqslant 0{,}1$	сомнения в истинности H_0, неопределенность	«Различия обнаружены на уровне статистической тенденции»
$p \leqslant 0{,}05$	значимость, отклонение H_0	«Обнаружены статистически достоверные (значимые) различия»
$p \leqslant 0{,}01$	высокая значимость, отклонение H_0	«Различия обнаружены на высоком уровне статистической значимости»

Алгоритм проверки статистических гипотез

1. Обоснование применения критерия.
2. Выполнение ограничений (если есть).
3. Формулирование статистических гипотез (H_0 и H_1).
4. Расчет критерия (результаты в таблице).
5. Определение уровня значимости (p).
6. Принятие одной из статистических гипотез.
7. Формулирование статистического вывода.
8. Интерпретация значимых результатов ($p \leq 0,05$) + рисунок.

H_0: = 0 принимается при $p > 0,05$

H_1: ≠ 0 принимается при $p \leq 0,05$

Тема 9. Меры связи

- Понятие корреляции.
- Диаграмма рассеяния.
- Классификация коэффициентов корреляции.
- Корреляционные матрицы.
- Интерпретация коэффициентов корреляции.
- Графическое представление полученных взаимосвязей. Корреляционные плеяды.

Понятие корреляции и ее основные параметры

- **Корреляционная связь –** это согласованное изменение двух или более признаков**.**
- **Коэффициент корреляции** — это количественная мера силы и направления вероятностной взаимосвязи двух переменных; принимает значения в диапазоне от -1 до +1.

Сила связи достигает максимума при условии взаимно однозначного соответствия: когда каждому значению одной переменной соответствует только одно значение другой переменной (и наоборот). Показателем силы связи является **абсолютная** (без учета знака) **величина** коэффициента корреляции.

Направление связи определяется прямым или обратным соотношением зна чений двух переменных: если возрастанию значений одной переменной соответствует возрастание значений другой переменной, то взаимосвязь на зывается прямой (положительной); если возрастанию значений одной переменной соответствует убывание значений другой переменной, то взаимосвязь является обратной (отрицательной). Показателем направления связи является **знак** коэффициента корреляции.

Формулировка статистических гипотез

H_0: Корреляция между переменными не отличается от нуля.

H_1: Корреляция между переменными отличается от нуля.

Виды связей

Взаимосвязи на языке математики обычно описываются при помощи функций, которые графически изображаются в виде линий.

- Если изменение одной переменной на одну единицу всегда приводит к изменению другой переменной на одну и ту же величину, функция является ***линейной*** (график ее представляет прямую линию); любая другая связь — ***нелинейная.***
- Если увеличение одной переменной связано с увеличением другой, то связь — ***положительная (прямая);*** если увеличение одной переменной связано с уменьшением другой, то связь — ***отрицательная (обратная)***.
- Если направление изменения одной переменной не меняется с возрастанием (убыванием) другой переменной, то такая функция — ***монотонная****;* в противном случае функцию называют ***немонотонной.***

Примеры графиков часто встречающихся функций

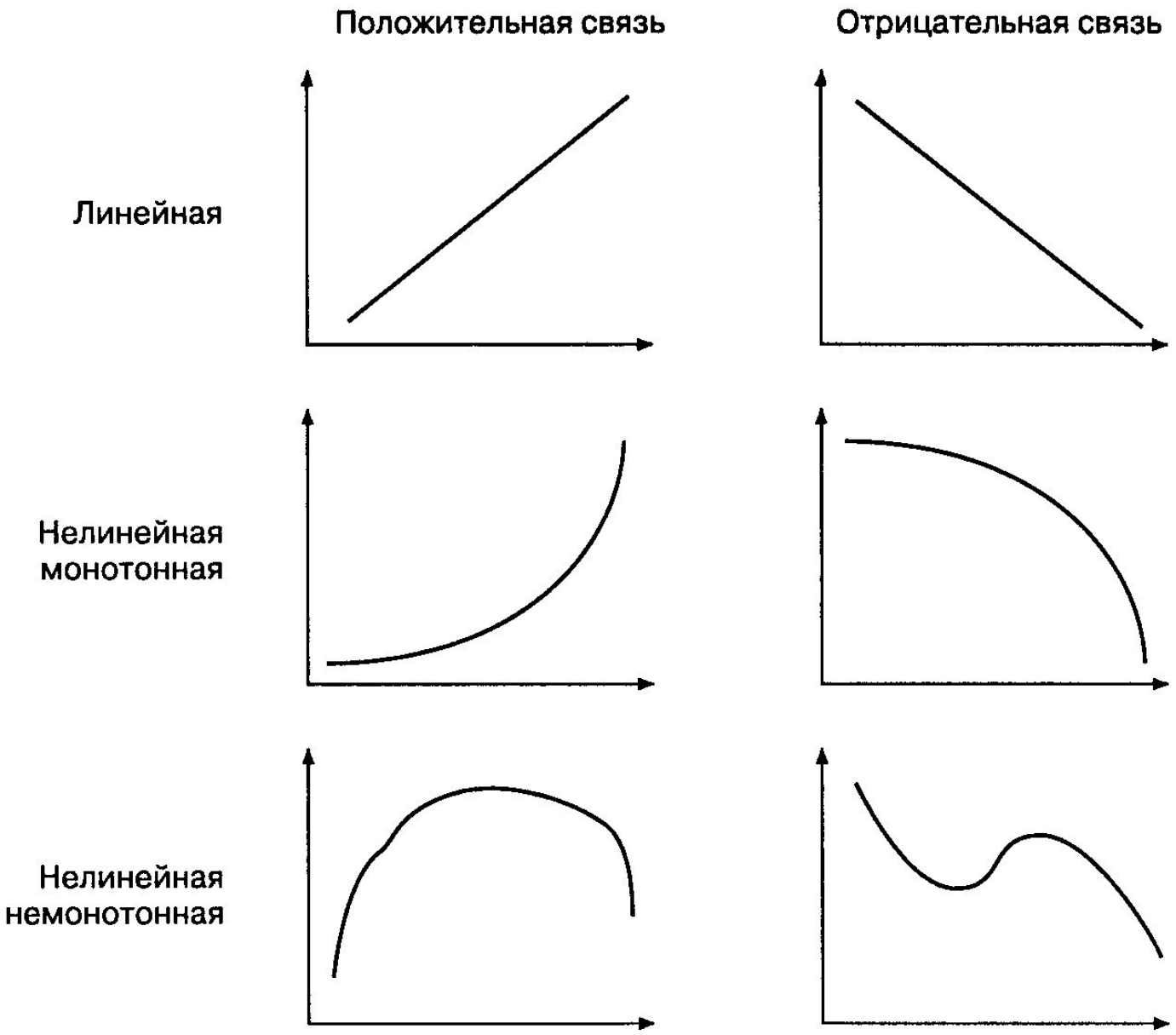
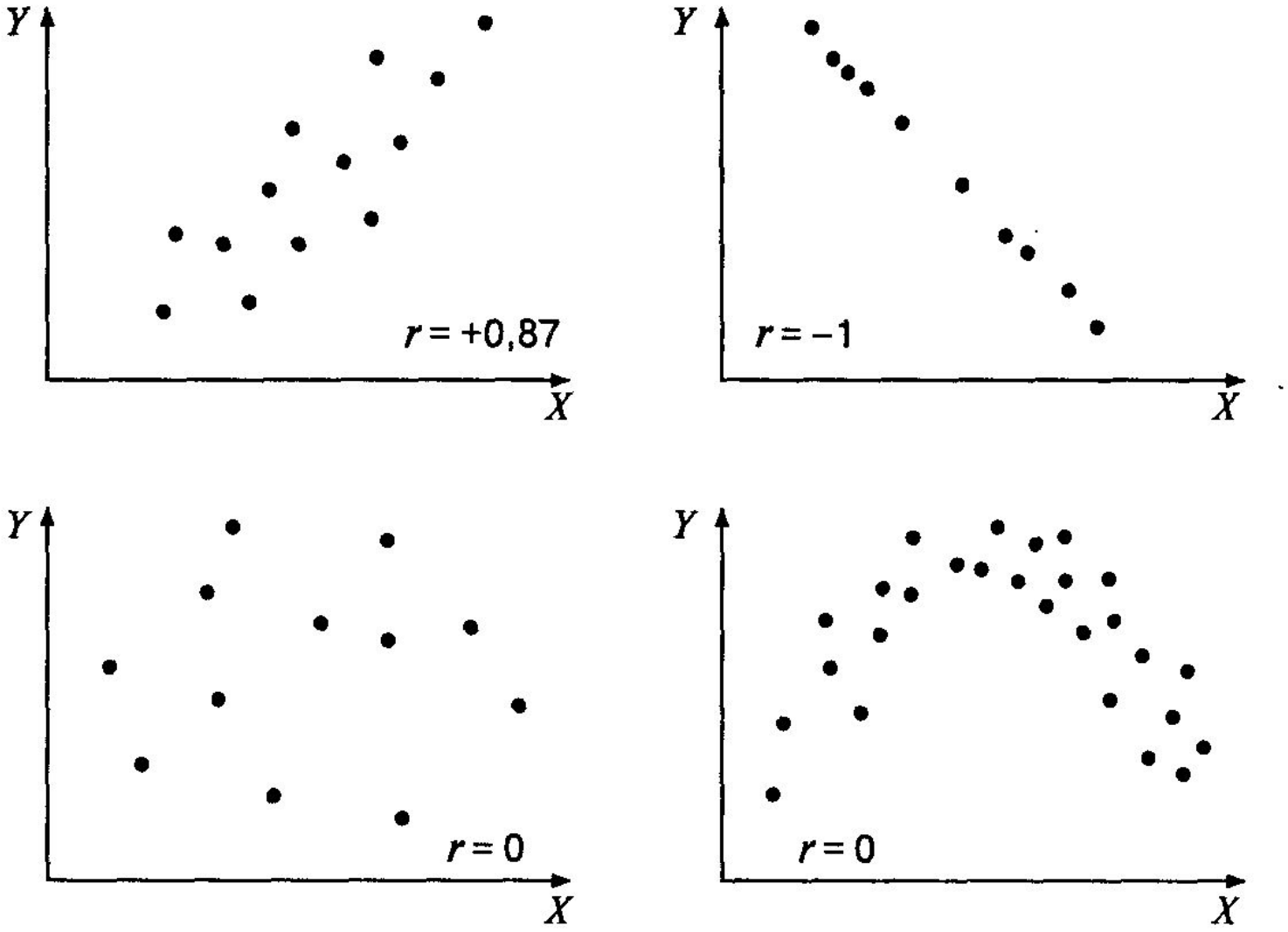


Диаграмма рассеивания — график, оси которого соответствуют значениям двух переменных, а каждый испытуемый представляет собой точку

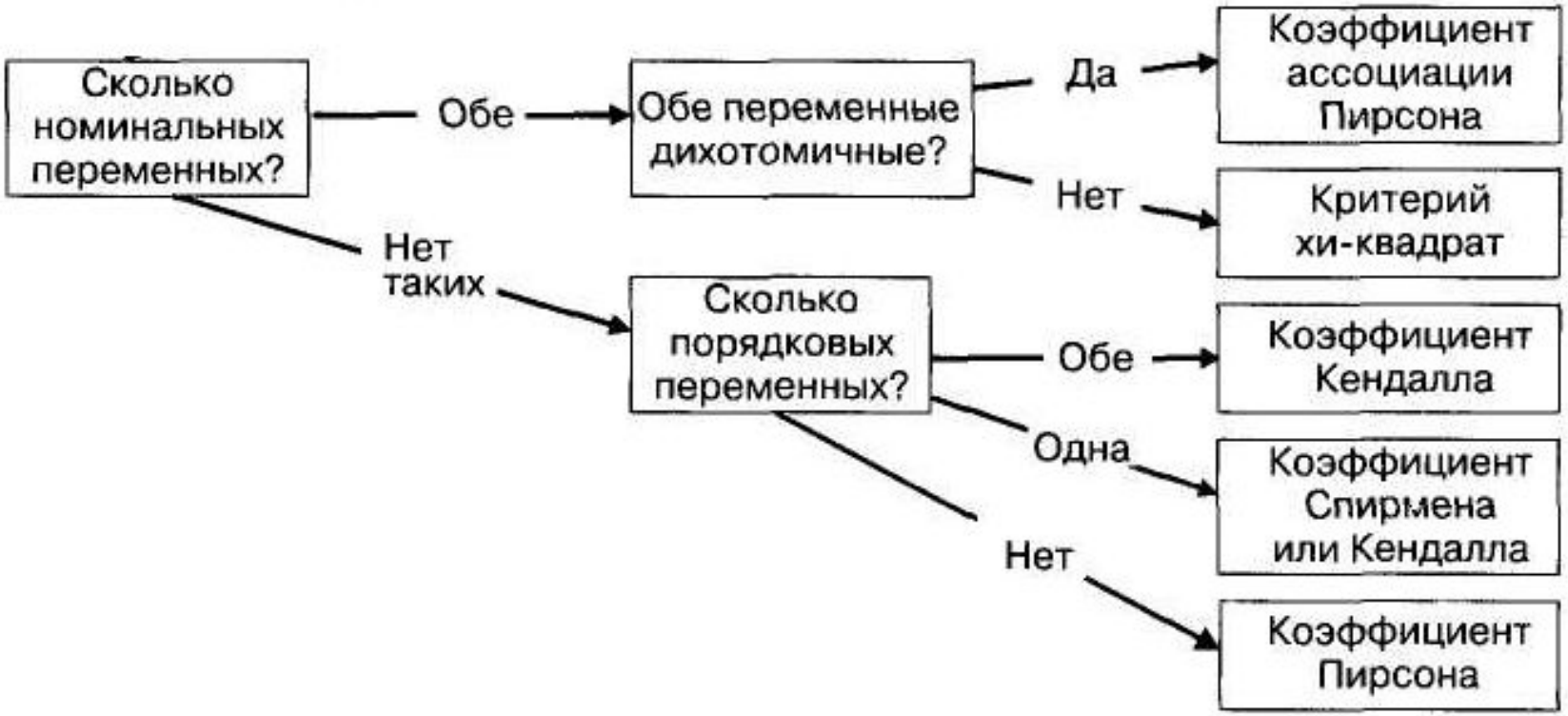


Классификация мер связи

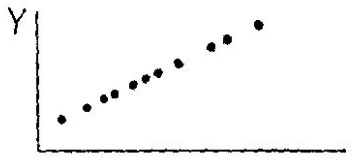
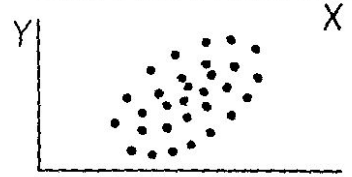
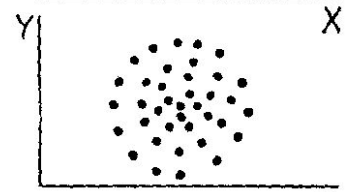
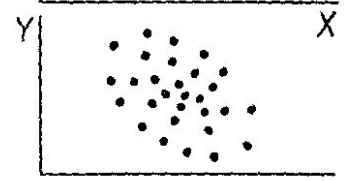
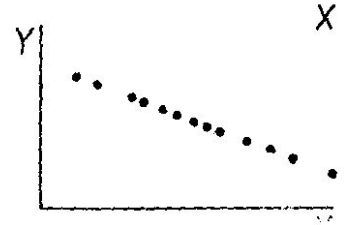
Шкала измерения					
Метрическая	**Порядковая (порядковая и метрическая)**		**Номинативная**		
			Более двух градаций	**Две градации**	
				Независимые выборки	**Зависимые выборки**
r_{xy}-Пирсона	r_s-Спирмена τ-Кенделла	τ_b-Кенделла γ-Gamma (гамма-статистика)	χ^2-Пирсона	ϕ-коэффициент сопряженности Пирсона (0,1)	McNemar (критерий Макнемара)
Нет выраженной асимметрии. Нет «выбросов»	Менее 10 % связанных рангов	Более 10 % связанных рангов	Теоретическая частота превышает 5	Только две градации	-

При $r \leq 0.3$ (слабая связь), $0,3 < r \leq 0,7$ (умеренная связь), $r > 0,7$ (сильная связь)

Алгоритм выбора коэффициента корреляции



Интерпретация значений r_{xy}

Величина r_{xy}	Описание линейной связи	Диаграмма рассеивания
+1,00	Строгая прямая связь	
Около +0,50	Умеренная прямая связь	
0,00	Нет связи (то есть ковариация X и $Y = 0$)	
Около −0,50	Умеренная обратная связь	
−1,00	Строгая обратная связь	

Представление данных корреляционного анализа

Построение корреляционных матриц и их анализ

1 вид - Квадратная матрица

Признаки	Менеджмент	Автономия	Стабильность	Служение	Вызов	Интеграция
Менеджмент	1,00	0,33	0,04	-0,35	0,69	0,14
Автономия	0,33	1,00	0,32	0,27	0,31	0,02
Стабильность	0,04	0,32	1,00	-0,21	0,15	0,53
Служение	-0,35	0,27	-0,21	1,00	0,42	0,06
Вызов	0,69	0,31	0,15	0,42	1,00	0,32
Интеграция	0,14	0,02	0,53	0,06	0,32	1,00

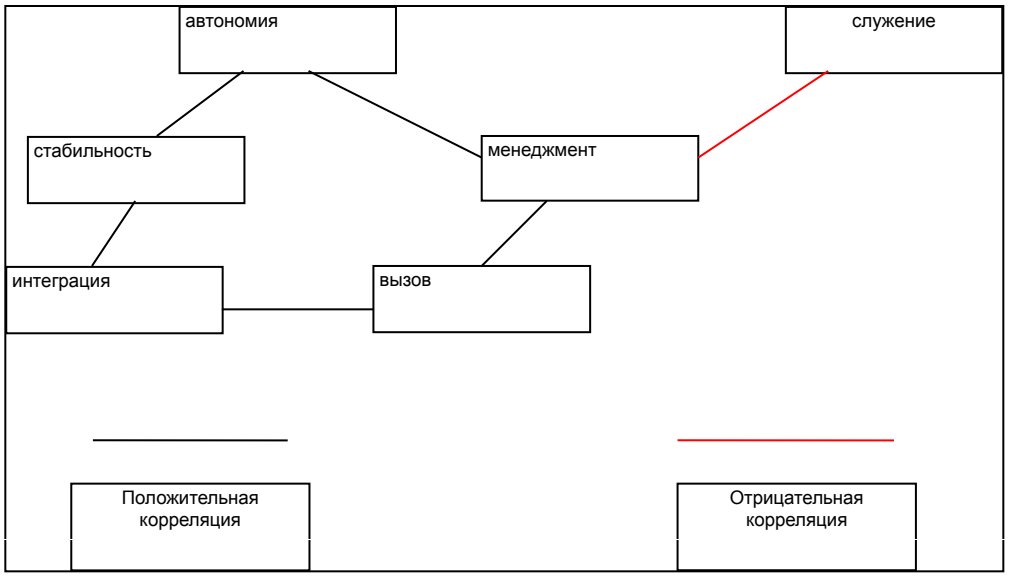
2 вид - Прямоугольная матрица

Признаки	Служение	Вызов	Интеграция
Менеджмент	-0,35	0,69	0,14
Автономия	0,27	0,31	0,02
Стабильность	-0,21	0,15	0,53

3 вид – Детализированная таблица

Признаки	N	r Spearman	p-level
Менеджмент Служение	40	-0,35	0,02
Менеджмент Вызов	40	0,69	0,00
Менеджмент Интеграция	40	0,14	0,38
Автономия Служение	40	0,27	0,10
Автономия Вызов	40	0,31	0,05
Автономия Интеграция	40	0,02	0,88
Стабильность Служение	40	-0,21	0,19
Стабильность Вызов	40	0,15	0,37
Стабильность Интеграция	40	0,53	0,00

Графическое представление данных корреляционного анализа
Поле рассеяния и Корреляционные плеяды



Классификация мер связи

Шкала измерения					
Метрическая	**Порядковая (порядковая и метрическая)**		**Номинативная**		
			Более двух градаций	**Две градации**	
				Независимые выборки	**Зависимые выборки**
r_{xy}-Пирсона	r_s-Спирмена τ-Кенделла	τ_b-Кенделла γ-Gamma (гамма-статистика)	χ^2-Пирсона M-L Chi-square (максимум правдоподобия χ^2)	Fisher exact (точный критерий Фишера), ϕ- коэффициент сопряженности Пирсона (0,1)	McNemar (критерий МакНимара)
Нет выраженной асимметрии. Связь между переменными прямолинейная.	Менее 10 % связанных рангов	Более 10 % связанных рангов	Не менее 5 наблюдений в каждом случае	-	-

Коэффициент корреляции rxy- Пирсона

- Коэффициент был создан Карлом (Чарлзом) Пирсоном (англ. Karl (Charles) Pearson), выдающимся английским математиком, статистиком, биологом и философом.
- Родился 27 марта 1857, Лондон
- Умер 27 апреля 1936, там же) —К. Пирсон считается основателем математической статистики; основные его труды по математической статистике: разработал теорию корреляции; тесты математической статистики и критерии согласия; распределение Пирсона и др.

Основные положения

- r-Пирсона (Pearson r) применяется для изучения взаимосвязи двух метрических переменных, измеренных на одной и той же выборке.

Ограничения

- Обе переменные не имеют выраженной асимметрии;
- Отсутствуют выбросы;
- Связь между переменными прямолинейная.

Пояснения к формуле

$$r_{xy} = \frac{\sum_{i=1}^{N}(x_i - M_x)(y_i - M_y)}{(N-1)\sigma_x \sigma_y},$$

- (xi – Mx), (yi – My) – отклонения соответствующих переменных от своих средних величин;
- N – количество испытуемых;
- σx, σy – соответствующие стандартные отклонения.

Интерпретация коэффициента корреляции Пирсона

- +1 – строгая прямая связь; -1 – строгая обратная связь
- +0,5 – умеренная прямая связь; -0,5 – умеренная обратная связь
- 0,0 – нет связи

Нахождение коэффициента корреляции r_{xy}-Пирсона

$$r_{xy} = \frac{25,6}{1,735 * 1,501 * 19} = 0,57 \quad p \leq 0,01$$

$$\sigma_x = \sqrt{\frac{57,2}{19}} = 1,735; \quad \sigma_y = \sqrt{\frac{42,8}{19}} = 1,501.$$

№	*X*	*Y*	$(x_i - X)$	$(y_i - Y)$	$(x_i - X)^2$	$(y_i - Y)^2$	$(x_i - X)(y_i - Y)$
1	13	12	3,2	1,6	10,24	2,56	5,12
2	9	11	-0,8	0,6	0,64	0,36	-0,48
3	8	8	-1,8	-2,4	3,24	5,76	4,32
4	9	12	-0,8	1,6	0,64	2,56	-1,28
5	7	9	-2,8	-1,4	7,84	1,96	3,92
6	9	11	-0,8	0,6	0,64	0,36	-0,48
7	8	9	-1,8	-1,4	3,24	1,96	2,52
8	13	13	3,2	2,6	10,24	6,76	8,32
9	11	9	1,2	-1,4	1,44	1,96	-1,68
10	12	10	2,2	-0,4	4,84	0,16	-0,88
11	8	9	-1,8	-1,4	3,24	1,96	2,52
12	9	8	-0,8	-2,4	0,64	5,76	1,92
13	10	10	0,2	-0,4	0,04	0,16	-0,08
14	10	12	0,2	1,6	0,04	2,56	0,32
15	12	10	2,2	-0,4	4,84	0,16	-0,88
16	10	10	0,2	-0,4	0,04	0,16	-0,08
17	8	11	-1,8	0,6	3,24	0,36	-1,08
18	9	10	-0,8	-0,4	0,64	0,16	0,32
19	10	11	0,2	0,6	0,04	0,36	0,12
20	11	13	1,2	2,6	1,44	6,76	3,12
Σ	**196**	**208**	**0,00**	**0,00**	**57,2**	**42,8**	**25,6**
X	**9,8**	**10,4**					

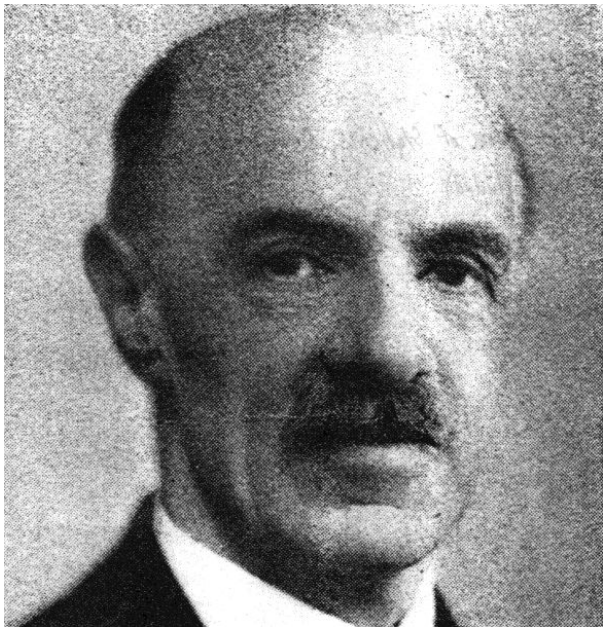
КРИТИЧЕСКИЕ ЗНАЧЕНИЯ КОЭФФИЦИЕНТОВ КОРРЕЛЯЦИИ *r*-ПИРСОНА (*r*-СПИРМЕНА)

(для проверки ненаправленных альтернатив, *n* — объем выборки)

n	*p*				*n*	*p*			
	0,10	0,05	0,01	0,001		0,10	0,05	0,01	0,001
5	0,805	0,878	0,959	0,991	46	0,246	0,291	0,376	0,469
6	0,729	0,811	0,917	0,974	47	0,243	0,288	0,372	0,465
7	0,669	0,754	0,875	0,951	48	0,240	0,285	0,368	0,460
8	0,621	0,707	0,834	0,925	49	0,238	0,282	0,365	0,456
9	0,582	0,666	0,798	0,898	50	0,235	0,279	0,361	0,451
10	0,549	0,632	0,765	0,872	51	0,233	0,276	0,358	0,447
11	0,521	0,602	0,735	0,847	52	0,231	0,273	0,354	0,443
12	0,497	0,576	0,708	0,823	53	0,228	0,271	0,351	0,439
13	0,476	0,553	0,684	0,801	54	0,226	0,268	0,348	0,435
14	0,458	0,532	0,661	0,780	55	0,224	0,266	0,345	0,432
15	0,441	0,514	0,641	0,760	56	0,222	0,263	0,341	0,428
16	0,426	0,497	0,623	0,742	57	0,220	0,261	0,339	0,424
17	0,412	0,482	0,606	0,725	58	0,218	0,259	0,336	0,421
18	0,400	0,468	0,590	0,708	59	0,216	0,256	0,333	0,418
19	0,389	0,456	0,575	0,693	60	0,214	0,254	0,330	0,414
20	0,378	0,444	0,561	0,679	61	0,213	0,252	0,327	0,411
21	0,369	0,433	0,549	0,665	62	0,211	0,250	0,325	0,408
22	0,360	0,423	0,537	0,652	63	0,209	0,248	0,322	0,405
23	0,352	0,413	0,526	0,640	64	0,207	0,246	0,320	0,402
24	0,344	0,404	0,515	0,629	65	0,206	0,244	0,317	0,399
25	0,337	0,396	0,505	0,618	66	0,204	0,242	0,315	0,396
26	0,330	0,388	0,496	0,607	67	0,203	0,240	0,313	0,393
27	0,323	0,381	0,487	0,597	68	0,201	0,239	0,310	0,390
28	0,317	0,374	0,479	0,588	69	0,200	0,237	0,308	0,388
29	0,311	0,367	0,471	0,579	70	0,198	0,235	0,306	0,385
30	0,306	0,361	0,463	0,570	80	0,185	0,220	0,286	0,361
31	0,301	0,355	0,456	0,562	90	0,174	0,207	0,270	0,341

Поле рассеяния

Коэффициенты ранговой корреляции rs-Спирмена и ι-Кендалла



- Чарльз Э́двард Спи́рмен (англ. Charles Edward Spearman) - английский психолог, профессор Лондонского и Честерфилдского университетов.
- Родился 10 сентября 1863
- Умер 17 сентября 1945 — Разработчик многочисленных методик математической статистики. Создатель двухфакторной теории интеллекта и техники факторного анализа. Кроме прочего, Спирмен открыл, что результаты даже несравнимых когнитивных тестов отражают единый фактор, который он назвал g-фактором (g factor).

Основные положения

Коэффициентов корреляции rs-Спирмена и ι-Кендалла

- Коэффициенты ранговой корреляции: r-Спирмена или ι-Кенделла применяются если обе переменные представлены в порядковой шкале, или одна из них — в порядковой, а другая — в метрической.

Ограничения

- Обе переменные представлены в количественной шкале (метрической или ранговой);
- Связь между переменными является монотонной (не меняет свой знак с изменением величины одной из переменных.
- Отсутствие повторяющихся рангов (менее 10 % связанных рангов).

Формула rs-Спирмена и пояснения к формуле

$$r_s = 1 - \frac{6\sum_i d_i^2}{N(N^2-1)},$$

- d – разность между рангами по двум переменным для каждого испытуемого;
- N – количество ранжируемых значений, в данном случае количество испытуемых

Интерпретация коэффициентов корреляции

- +0,7 и выше – тесная положительная связь; -0,7 и выше – тесная отрицательная связь;
- +0,4 и выше – умеренная положительная связь; -0,4 и выше – умеренная отрицательная связь;
- +0,2 и – выше слабая положительная связь;-0,2 и – выше слабая отрицательная связь;
- 0,0 и выше – нет связи

Нахождение коэффициента корреляции rs-Спирмена

$$r_s = 1 - \frac{6*474}{12(144-1)} = -0{,}65 \quad p \le 0{,}05$$

№	X	Y	Ранги X	Ранги Y	d_i	d_i^2
1	122	4,7	7	2	5	25
2	105	4,5	10	4	6	36
3	100	4,4	11	5	6	36
4	145	3,8	5	9	-4	16
5	130	3,7	6	10	-4	16
6	90	4,6	12	3	9	81
7	162	4,0	3	8	-5	25
8	172	4,2	1	6	-5	25
9	120	4,1	8	7	1	1
10	150	3,6	4	11	-7	49
11	170	3,5	2	12	-10	100
12	112	4,8	9	1	8	64
Σ	-	-	78	78	0	474

$$r_s = 1 - \frac{6\sum_i d_i^2}{N(N^2-1)},$$

Формула τ-Кенделла : $\tau = \frac{P - Q}{N(N-1)/2}$,

Пояснения к формуле

- *P* — общее число совпадений.
- *Q* — *общее* число инверсий
- N – количество испытуемых

Алгоритм

- Данные упорядочиваются по переменной *X*.
- Затем для каждого испытуемого подсчитывается, сколько раз его ранг по Y оказывается меньше, чем ранг испытуемых, находящихся ниже. Результат записывается в столбец «Совпадения». Сумма всех значений столбца «Совпадения» и есть **P** — общее число совпадений, подставляется в формулу.
- После чего, для каждого испытуемого подсчитывается сколько раз его ранг поYоказывается больше, чем ранг испытуемых, находящихся ниже. Сумма всех значений столбца «инверсии» и есть **Q** — общее число инверсий, которые подставляются в формулу

Нахождение коэффициента корреляции ι-Кенделла

$\iota = \frac{21-7}{8(8-1)/2} = 0{,}5 \quad p = 0{,}08$

Статистический вывод: взаимосвязь между мотивацией и эмоциональными выборами не обнаружена.

N	x	y	P	Q
1	1	**3**	***5***(5.7.8.4. 6)	***2(***1.2)
2	2	1	6	0
3	3	2	5	0
4	4	5	3	1
5	5	7	1	2
6	6	8	0	2
7	7	4	0	0
8	8	6	0	0
∑			21	7

$$\tau = \frac{P-Q}{N(N-1)/2},$$

Тема 10. Анализ качественных признаков (номинативных данных)

- Корреляция номинативных данных
  критерий $\chi 2$-Пирсона
- Корреляция бинарных данных
  фи-коэффициент сопряженности Пирсона

Анализ качественных признаков (номинативных данных)

Корреляция номинативных данных критерий χ^2-Пирсона

- Критерий χ^2-Пирсона применяется если обе переменные представлены в номинативной шкале, одна из которых или обе имеют более двух градаций.

Ограничения

- Ожидаемые частоты должны быть больше 5.
- Суммы по строкам и по столбцам должны быть больше нуля.

Формула χ^2-Пирсона и пояснения к формуле

- **fe =** $\frac{fj \times fk}{n}$ df = (k – 1)x(j – 1)
- $$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$
- fo – наблюдаемая частота (эмпирическая);
- fe – ожидаемая частота (теоретическая);
- n – общее количество наблюдений;
- k – k – й столбец;
- j – j-я строка.

Нахождение критерия χ2-Пирсона

Эмпирические частоты

Пол	**Предпочитаемый цвет**			
	синий	**зеленый**	**красный**	**Всего**
женский	4	4	0	8
мужской	0	1	6	7
Всего	4	5	6	15

Теоретические частоты fe женский и синий = $\frac{4 \times 8}{15} = 2{,}1$

Пол	**Предпочитаемый цвет**						
	синий		**зеленый**		**красный**		**Σ**
	f_e	№ ячейки	f_e	№ ячейки	f_e	№ ячейки	
женский	2,1	1	2,7	3	3,2	5	8
мужской	1,9	2	2,3	4	2,8	6	7
Σ	4		5		6		15

Нахождение критерия χ2-Пирсона

Расчет

№ ячейки	f_0	f_e	$f_0 - f_e$	$(f_0 - f_e)^2$	$\frac{(f_0 - f_e)^2}{f_e}$
1	4	2,1	1,9	3,61	1,7
2	0	1,9	-1,9	3,61	1,9
3	4	2,7	1,3	1,69	0,6
4	1	2,3	-1,3	1,69	0,7
5	0	3,2	-3,2	10,24	3,2
6	6	2,8	3,2	10,24	3,7
Σ	15	15	0	31,08	**11,8**

$\chi^2 = 11,8$

$k = 3; j = 2; df = (k - 1)x(j - 1) = (3 - 1)x(2 - 1) = 2;$

$p \leq 0,01$

Статистический вывод: существует взаимосвязь между полом и предпочтением цвета – мужчины значимо предпочитают красный цвет, а женщины синий и зеленый цвета с вероятностью ошибки менее 1 %.

Корреляция бинарных данных фи-коэффициент сопряженности Пирсона

- Коэффициент сопряженности φ-Пирсона применяется если обе переменные представлены в номинативной шкале, имеющей две градации.

Формула φ-Пирсона и пояснения к формуле

- px – доля имеющих 1 по x;
- py – доля имеющих 1 по y;
- pxy – доля тех, кто имеет 1 и по x и по y;
- qx – доля имеющих 0 по x = 1 – px
- qy – доля имеющих 0 по y = 1 – py

$$\varphi = \frac{p_{xy} - p_x p_y}{\sqrt{p_x q_x p_y q_y}}.$$

Нахождение коэффициента сопряженности φ-Пирсона $\varphi = \frac{p_{xy} - p_x p_y}{\sqrt{p_x q_x p_y q_y}}$.

N	x	y	Вычисления
1	0	0	P_x=5/12=0,42 P_y=6/12=0,5 P_{xy} = 4/12=0,33 q_x = 1 – 0,42=0,58 q_y = 1 – 0,5=0,5 $\phi = \frac{0,33 - 0,42 \times 0,5}{\sqrt{0,42 \times 0,58 \times 0,5 \times 0,5}} \approx 0,5$ p = 0,07 Статистический вывод: не подтверждается взаимосвязь между выраженной тревожностью и выполнением задачи
2	1	1	
3	0	1	
4	0	0	
5	1	1	
6	1	0	
7	0	0	
8	1	1	
9	0	0	
10	0	1	
11	0	0	
12	1	1	

Тема 11. Анализ различий между 2 группами независимых выборок

- **Классификация методов сравнения**
- **Представление данных сравнительного анализа**
- **Параметрический критерий t-Стьюдента для двух независимых выборок**
- **Непараметрический критерий U-Манна-Уитни для двух независимых выборок**

Методы сравнения

В зависимости от решаемых задач методы внутри этой группы классифицируются по трем основаниям:

- Количество градаций *X:*

а) сравниваются 2 выборки;

б) сравниваются больше 2 выборок.

- Зависимость выборок:

а) сравниваемые выборки независимы;

б)сравниваемые выборки зависимы.

- Шкала *У:*

а) *Y*— ранговая переменная;

б) *У*— метрическая переменная.

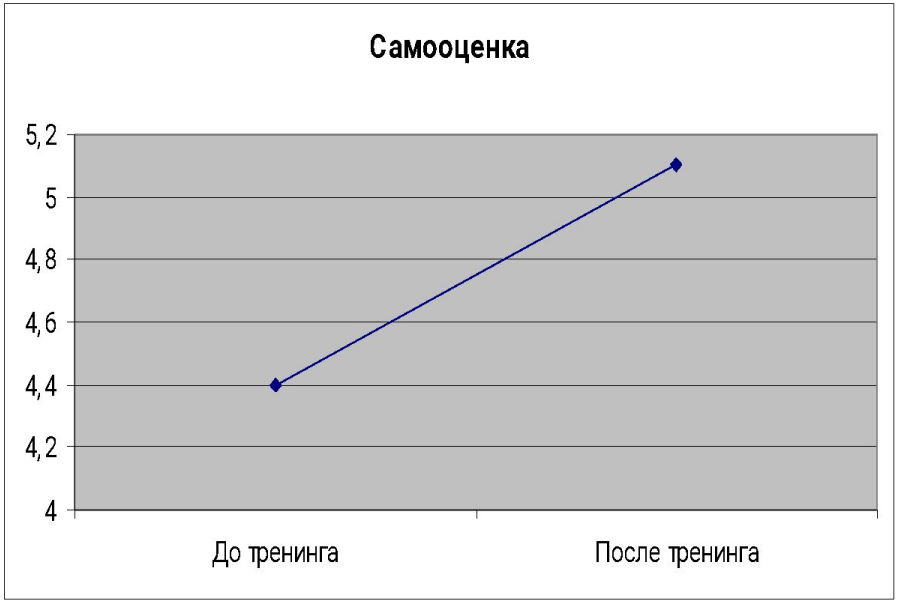
- По последнему основанию методы делятся на две большие группы: параметрические методы (критерии) — для метрических переменных и непараметрические методы (критерии) — для порядковых (ранговых) переменных. **Параметрические методы** проверяют гипотезы относительно **параметров распределения** (средних значений и дисперсий) и основаны на предположении о нормальном распределении в генеральной совокупности. **Непараметрические методы** не зависят от предположений о характере распределения и не касаются параметров этого распределения.
- **Независимые выборки** характеризуются тем, что вероятность отбора любого испытуемого одной выборки не зависит от отбора любого из испытуемых другой выборки. Напротив, **зависимые выборки** характеризуются тем, что каждому испытуемому одной выборки поставлен в соответствие по определенному критерию испытуемый из другой выборки.В общем случае зависимые выборки предполагают попарный подбор испытуемых в сравниваемые выборки, а независимые выборки — независимый отбор испытуемых.
- **Формулировка статистических гипотез**

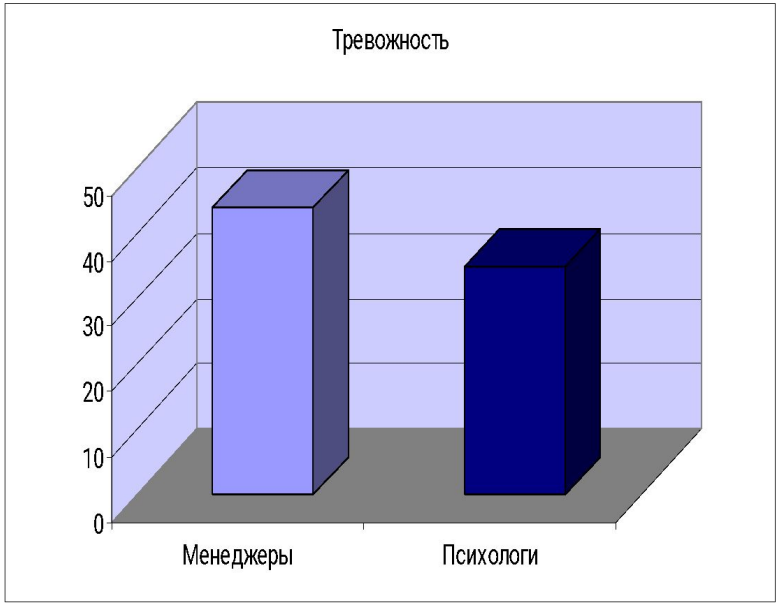
H0: Различий между выборками в уровне изучаемого признака не имеется.

H1: Различия между выборками в уровне изучаемого признака имеются.

Представление данных сравнительного анализа

Графическое представление данных





Построение таблиц

Признаки	Среднее выборка 1	Среднее выборка 2	Значение критерия	Уровень значимости
1				
2				

Классификация методов сравнения

<table>
<tr><td colspan="2">Количество выборок (градаций X)</td><td colspan="2">Две выборки</td><td colspan="2">Больше двух выборок</td></tr>
<tr><td colspan="2">Зависимость выборок</td><td>Независимые</td><td>Зависимые</td><td>Независимые</td><td>Зависимые</td></tr>
<tr><td rowspan="5">Признак Y</td><td>метрический</td><td colspan="4">Параметрические методы сравнения</td></tr>
<tr><td>t-Стьюдента, для независимых выборок</td><td>t-Стьюдента, для зависимых выборок</td><td>ANOVA (дисперсионный анализ Фишера)</td><td>ANOVA, с повторными измерениями</td></tr>
<tr><td colspan="4">Проверяют средние значения и дисперсий и зависят от нормальности распределения и генеральной совокупности</td></tr>
<tr><td rowspan="2">ранговый

Примечание</td><td colspan="4">Непараметрические методы сравнения</td></tr>
<tr><td>U-Манна-Уитни, критерий серий</td><td>T-Вилкоксона, G-критерий знаков</td><td>H-Краскала-Уоллеса</td><td>χ^2-Фридмана</td></tr>
<tr><td colspan="4">Проверяют средние значения и по уровню выраженности ранговой переменной, не зависят от нормальности распределения и генеральной совокупности</td></tr>
<tr><td colspan="4">Для t, F, χ^2 и др. - чем больше значение критерия, тем выше статистическая значимость (меньше p - уровень);
Для U, T – чем меньше значение критерия, тем выше статистическая значимость (уменьшение p – уровня).</td></tr>
</table>

Критерий t-Стьюдента

- Уи́льям Си́ли Го́ссет - известный учёный-статистик.
- Родился 13 июня 1876 г. в Кентербери (Англия)
- Умер 16 октября 1937 г. в Беконсфилд (Англия)
- Госсет совершил «логическую революцию». По иронии судьбы, t-статистика, благодаря которой знаменит Госсет, была фактически изобретением Фишера. Госсет считал статистику для $z = t/\sqrt{(n-1)}$. Фишер предложил вычислять статистику для t, потому что такое представление укладывалось в его теорию степеней свободы.
- На пивоваренном заводе, где работал Госсет работодатель запретил своим работникам публикацию материалов. Это означало, что Госсет не мог опубликовать свои работы под своим именем. Поэтому он избрал себе псевдоним Стьюдент, чтобы скрыть себя от работодателя. Поэтому его самое важное открытие получило называние Распределение Стьюдента

Параметрический критерий t-Стьюдента для двух независимых выборок

- Метод позволяет проверить гипотезу о том, что средние значения двух генеральных совокупностей, из которых извлечены две сравниваемые независимые выборки, отличаются друг от друга.

Ограничения:

- Распределения признака и в той, и в другой выборке существенно не отличаются от нормального.
- Дисперсии выборок равны.
- Признак измерен в метрической шкале.

$$t_э = \frac{M_1 - M_2}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}}$$

Формула t-Стьюдента и пояснения к формуле

- df = N1 + N2 – 2
- M1 и M2 – средние значения в соответствующих выборках;

σ1 и σ2 – ст. отклонение в соответствующих выборках;

- N1 и N2 – количество испытуемых в соответствующих выборках;
- df - число степеней свободы.

- **Гипотезы:**

- **H0:** признак в выборке 1 равен исследуемому признаку в выборке 2.
- **H1:** признак в выборке 1 не равен исследуемому признаку в выборке 2.

Нахождение критерия t-Стьюдента для двух независимых выборок

$$t_э = \frac{M_1 - M_2}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}}$$

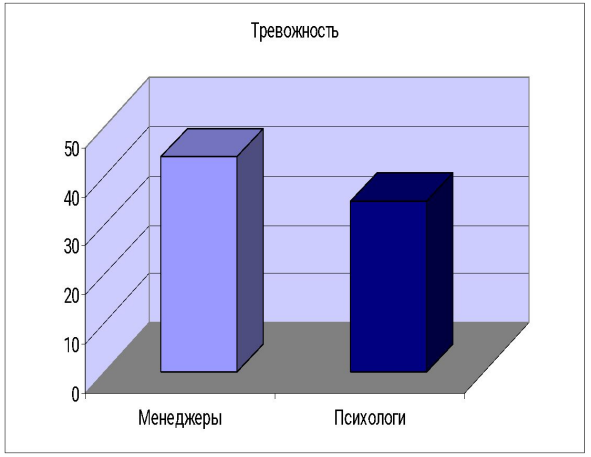
№ п/п	Тревожность	
	Менеджеры	Психологи
1	45	23
2	37	25
3	24	34
4	56	33
5	55	45
6	42	36
7	44	38
8	46	32
9	49	39
10	43	44
M	**44,1**	**34,9**
σ	**9,12**	**7,19**

$$tэ = \frac{44,1-34,9}{\sqrt{9,12/10+7,19/10}} = 2,5$$

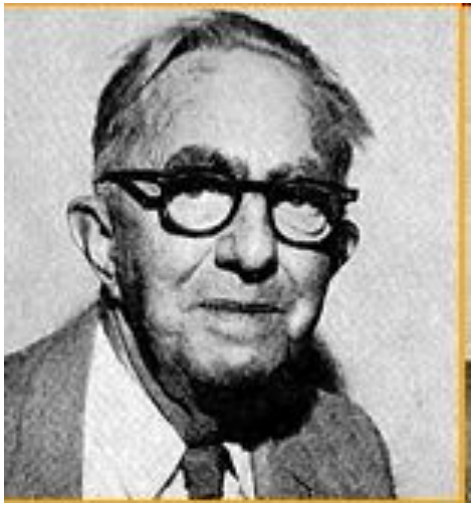
$$df = 10 + 10 - 2 = 18;\ p \le 0,05$$

Статистический вывод: Между психологами и менеджерами существуют значимые различия в уровне тревожности с вероятностью ошибки менее 5 %.

df	*p*			
	0,10	0,05	0,01	0,001
1	6,314	12,70	63,65	636,61
2	2,920	4,303	9,925	31,602
...	2,353	3,182	5,841	12,923
17	1,740	2,110	2,898	3,965
18	1,734	2,101	2,878	3,922



Критерий U-Манна-Уитни



- Настоящий статистический метод был предложен Фрэнком Вилкоксоном в 1945 году. Однако в 1947 году метод был улучшен и расширен Х. Б. Манном и Д. Р. Уитни, посему U-критерий чаще называют их именами.

Непараметрический критерий U-Манна-Уитни для двух независимых выборок

- Критерий предназначен для оценки различий между двумя выборками по уровню какого-либо признака, количественно измеренного. Он отражает степень совпадения (перекрещивания) двух рядов значений, то значение p-уровня тем меньше, чем меньше значение U.
- **Ограничения:**
- В каждой из выборок должно быть не менее 3 значений признака. Допускается, чтобы в одной выборке было два значения, но во второй тогда не менее пяти.
- В выборочных данных не должно быть совпадающих значений (все числа — разные) или таких совпадений должно быть очень мало.т.

Формула U-Манна-Уитни и пояснения к формуле

$$U_x = mn - R_x + \frac{n(n+1)}{2},$$

$$U_y = mn - R_y + \frac{m(m+1)}{2},$$

$$U_x + U_y = mn,$$

- n — объем выборки X;
- m — объем выборки У,
- Rx и Ry — суммы рангов для X и У в объединенном ряду.
- В качестве эмпирического значения критерия берется наименьшее из U_x и U_y. Чем больше различия, тем меньше эмпирическое значение U.

Гипотезы

- **H0:** Уровень признака в группе 2 не ниже уровня признака в группе 1.
- **H1:** Уровень признака в группе 2 ниже уровня признака в группе 1.

Нахождение критерия U-Манна-Уитни

Значения	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	19
Выборка	X	X	Y	X	X	X	Y	X	X	Y	X	Y	Y	Y	Y	Y
Ранги	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Ранги X	1	2		4	5	6		8	9		11					
Ранги Y			3				7			10		12	13	14	15	16

Ш а г 1. Значения двух выборок объединяются в один ряд и упорядочиваются.

Ш а г 2. Обозначается принадлежность к выборке.

Ш а г 3. Значения ранжируются.

Ш а г 4 и 5. Выписываются ранги отдельно по X отдельно по У.

Ш а г 6. Сумма рангов по X и по У подставляется в формулу:

X(RX) и по *Y(Ry): Rx* = 46; *Ry* = 90.

Ux = 8 x 8 – 46 + 8(8+1)/2 = 18 + 72/2 = 18 + 36 = 54

Uy = 8 x 8 – 90 + 8(8+1)/2 = -26 + 72/2 = -26 + 36 = 10

Наименьшая сумма сравнивается с табличной и определяется р.

На уровне α = 0,05 принимается статистическая гипотеза о различии X и *Y* по уровню выраженности признака. Уровень *Y* статистически достоверно выше уровня *X* (*p* < 0,05).

$$U_x = mn - R_x + \frac{n(n+1)}{2},$$

$$U_y = mn - R_y + \frac{m(m+1)}{2},$$

$$U_x + U_y = mn,$$

n_1	2	3	4	5	6	7	8
n_2	ρ=0,05						
3	-	0					
4	-	0	1				
5	0	1	2	4			
6	0	2	3	5	7		
7	0	2	4	6	8	11	
8	1	3	5	8	10	13	15

Тема 12. Анализ различий между 2 группами зависимых выборок

- Параметрический критерий t-Стьюдента для двух зависимых выборок
- Непараметрический критерий Т-Уилкоксона для сравнения двух зависимых групп

Параметрический критерий t-Стьюдента для двух зависимых выборок

- Метод позволяет проверить гипотезу о том, что средние значения двух генеральных совокупностей, из которых извлечены две сравниваемые зависимые выборки, отличаются друг от друга. Допущение зависимости чаще всего значит, что признак измерен на одной и той же выборке дважды, например, до воздействия и после него.

Ограничения:

- Распределения признака и в той, и в другой выборке существенно не отличаются от нормального.
- Дисперсии выборок равны.
- Признак измерен в метрической шкале.

Формула t-Стьюдента и пояснения к формуле

- Md – средняя разность значений;
- σd – стандартное отклонение разностей;
- N – количество испытуемых в выборке
- df - число степеней свободы.

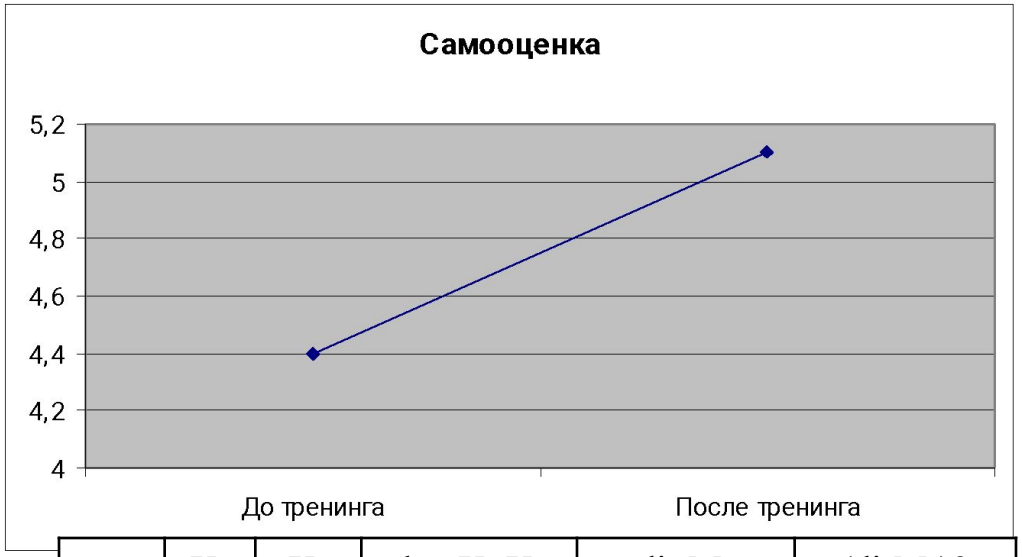
$$t_э = \frac{M_d}{\sigma_d / \sqrt{N}}, \; df = N - 1,$$

Гипотезы

H_0: Между показателями, полученными (измеренными) в разных условиях, существуют лишь случайные различия.

H_1: Между показателями, полученными в разных условиях, существуют неслучайные различия.

Нахождение критерия t-Стьюдента для двух зависимых выборок



n	X_1	X_2	$d_i = X_1 - X_2$	$di - M_d$	$(di-M_d)2$
1	3	4	-1	-0,25	0,0625
2	6	6	0	0,75	0,5625
3	5	6	-1	-0,25	0,0625
4	2	4	-2	-1,25	1,5625
5	7	6	1	1,75	3,0625
6	3	4	-1	-0,25	0,0625
7	4	5	-1	-0,25	0,0625
8	5	6	-1	-0,25	0,0625
∑	**35**	**41**	**-6**	**0**	**5,5**

Ша г 1. Эмпирическое значение критерия по формуле:

средняя разность $M_d = \sum d_i / n = -6/8 = -0,75$;

стандартное отклонение $\sigma_d = \sqrt{\frac{\sum (x_i - M_x)^2}{N-1}}$ = 0,886;

тэмп, = -2,39; df = 8-1 = 7.

Ш а г 2. Определяем по таблице критических значений критерия t-Стьюдента Для *df = 7* эмпирическое значение находится между критическими для $p = 0,05$ и $p = 0,01$. Следовательно, $p < 0,05$.

Ш а г 3. Принимаем статистическое решение и формулируем вывод. Статистическая гипотеза о равенстве средних значений отклоняется. Вывод: показатель самооценки конформизма участников после тренинга увеличился статистически достоверно ($p < 0,05$).

$$t_э = \frac{M_d}{\sigma_d / \sqrt{N}},\ df = N - 1,$$

Непараметрический критерий Т-Уилкоксона для сравнения двух зависимых групп

- Критерий предназначен для оценки различий между двумя зависимыми выборками по уровню какого-либо признака, количественно измеренного. Он отражает степень совпадения (перекрещивания) двух рядов значений.

Ограничения - нет.

Формула Т-Уилкоксона и пояснения к формуле

- Подсчитываются суммы рангов для положительных и отрицательных разностей. Затем меньшая из сумм принимается в качестве эмпирического значения критерия, значение которого сравнивается с табличным значением для данного объема выборки. Чем больше различия, тем меньше эмпирическое значение *T,* тем меньше значение p-уровня.

Гипотезы

H_0: Интенсивность сдвигов в типичном направлении не превосходит интенсивности сдвигов в нетипичном направлении.

H_1: Интенсивность сдвигов в типичном направлении превышает интенсивность сдвигов в нетипичном направлении.

Нахождение непараметрического критерия Т-Уилкоксона

№ объекта:	1	2	3	4	5	6	7	8	9	10	11	12
Условие 1	6	11	12	8	5	10	7	6	3	9	4	5
Условие 2	14	5	8	10	14	7	12	13	11	10	15	16
Разность $d_{i:}$	-8	6	4	-2	-9	3	-5	-7	-8	-1	-11	-11
Ранги$\|d_i\|$	8,5	6	4	2	10	3	5	7	8,5	1	11,5	11,5
Ранги d_i (+)		6	4			3						
Ранги d_i (-)	8,5			2	10		5	7	8,5	1	11,5	11,5

Ш а г 1. Подсчитать разности значений для каждого объекта выборки (строка 4).

Ш а г 2. Ранжировать абсолютные значения разностей (строка 5).

Ш а г 3. Выписать ранги положительных и отрицательных значений разностей (строки 6 и 7).

Ш а г 4. Подсчитать суммы рангов отдельно для положительных и отрицательных разностей: Т1 = 13; Т2 = 65. За эмпирическое значение критерия Тэмп принимается меньшая сумма: Тэмп = 13.

Наименьшая сумма сравнивается с табличной и определяется р.

Уровень выраженности признака для условия 2 статистически значимо выше, чем для условия 1 (p = 0,05).

n	0,10	0,05	0,02
11	13	10	7
12	17	13	9
13	21	17	12

Тема 13. Анализ различий между 3 и более группами независимых выборок

- Непараметрический критерий Н-Краскала-Уоллеса для сравнения 3 и более групп
- Критерий $\chi 2$-Фридмана для сравнения 3-х и более зависимых выборок

Непараметрический критерий Н-Краскала-Уоллеса для сравнения 3 и более групп

- Критерий Краскала — Уоллиса предназначен для проверки равенства медиан нескольких выборок. Данный критерий является многомерным обобщением критерия Уилкоксона — Манна — Уитни. Критерий Краскала — Уоллиса является ранговым, поэтому он инвариантен по отношению к любому монотонному преобразованию шкалы измерения.

Основные положения

- Критерий Н-Краскала-Уоллеса позволяет проверять гипотезы о различии более двух выборок по уровню выраженности изучаемого признака. Он оценивает степень пересечения (совпадения) нескольких рядов значений измеренного признака. Чем меньше совпадений, тем больше различаются ряды, соответствующие сравниваемым выборкам.

Ограничения - нет.

Формула Н-Краскала-Уоллеса и пояснения к формуле

- N — суммарная численность всех выборок;
- k — количество сравниваемых выборок;
- Ri — сумма рангов для выборки i;
- ni, — численность выборки i.

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1),$$

- Чем сильнее различаются выборки, тем больше вычисленное значение Н и тем меньше p-уровень значимости.
- При отклонении H_0 для утверждений о том, что уровень выраженности признака в какой-то из сравниваемых выборок выше или ниже, необходимо парное соотнесение выборок по критерию U-Манна-Уитни.
- **Гипотезы**
- H_0: Между выборками 1, 2, 3 и т. д. существуют лишь случайные различия по уровню исследуемого признака.
- H_1: Между выборками 1, 2, 3 и т. д. существуют неслучайные различия по уровню исследуемого признака.

Нахождение Н-Краскала-Уоллеса

Значения	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	19	$\sum_i$
Выборка	1	1	2	1	1	1	2	1	1	2	1	3	2	3	3	2	
Ранги	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
Ранги 1	1	2		4	5	6		8	9		11						46
Ранги 2			3				7			10			13			16	49
Ранги 3												12		14	15		41

Шаг 1. Значения объединяются в один упорядоченный ряд. Обозначается принадлежность каждого значения к выборке (строки 1 и 2).

Ш а г 2. Значения выборок ранжируются и выписываются отдельно ранги для каждой выборки (строки 3-6).

Ш а г 3. Вычисляются суммы рангов для каждой выборки *Rx* = 46; *R2* = 49; *R3 = 41.*

Ш а г 4. Н = 12/ 16(16 + 1) х (462/8 + 492/5 + 412/3) – 3(16 + 1) = 7,725

Шаг 5. Определяется р-уровень значимости. Хотя сравниваются 3 выборки, но объем одной из них больше 5, поэтому вычисленное Н сравнивается с табличным значением $\chi 2$ (приложение 4) для числа степеней свободы *df* = 3—1=2. Эмпирическое значение Н находится между критическими для $p = 0{,}05$ и $p = 0{,}01$. Следовательно, $p < 0{,}05$.

Ш а г 6. На уровне р = 0,05 гипотеза Но отклоняется. Содержательный вывод: сравниваемые выборки различаются статистически достоверно по уровню выраженности признака *(*$p < 0{,}05$).

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1),$$

Критерий χ2-Фридмана для сравнение 3-х и более зависимых выборок

- Критерий χ2-Фридмана позволяет проверять гипотезы о различии более двух зависимых выборок (повторных измерений) по уровню выраженности изучаемого признака. Чем больше различаются зависимые выборки по изучаемому признаку, тем больше эмпирическое значение χ2-Фридмана.

Ограничения - нет.

Формула χ2-Фридмана и пояснения к формуле

- N — число объектов (испытуемых),
- k — количество условий (повторных измерений),
- Ri — сумма рангов для условия i.

$$\chi^2_э = \left[\frac{12}{Nk(k+1)} \cdot \sum_{i=1}^{k} R_i^2\right] - 3N(k+1),\ df = k-1,$$

- При расчетах для определения p-уровня пользуются таблицами критических значений. Если k=3, N > 9 или k > 3, N > 4, то пользуются обычной таблицей для χ2, df = к — 1. Если к = 3, N < 10 или k = 4, N < 5, то пользуются дополнительными таблицами критических значений χ2- Фридмана.
- Для утверждений о том, что уровень выраженности признака в какой-то из сравниваемых выборок выше или ниже, необходимо парное соотнесение выборок по критерию Т-Вилкоксона.

Гипотезы

H_0: Между показателями, полученными (измеренными) в разных условиях, существуют лишь случайные различия.

H_1: Между показателями, полученными в разных условиях, существуют неслучайные различия.

Нахождение критерия χ2-Фридмана

№	Условие 1		Условие 2		Условие 3		Условие 4	
	X	Ранг	*X*	Ранг	*X*	Ранг	*X*	Ранг
1	6	2	14	3.5	5	1	14	3.5
2	11	3	5	2	4	1	12	4
3	12	4	8	2	7	1	10	3
4	8	1	10	2	11	3	12	4
5	5	1	14	3.5	10	2	14	3.5
6	10	3	7	2	6	1	12	4
Сумма рангов:		14		15		9		22

$$\chi^2_э = \left[\frac{12}{Nk(k+1)} \cdot \sum_{i=1}^{k} R_i^2\right] - 3N(k+1),\ df = k-1,$$

- Шаг 1. Для каждого объекта условия ранжируются (по строке).
- Ш а г 2. Вычисляется сумма рангов для каждого условия: *R1* = 14, *R2* = 15, R3 = 9, R4=22.
- Ш а г 3. Вычисляется значение *χ2*-Фридмана по формуле :
- *χ2* = [12/ 6 x 4(4 + 1) x (142+152+92+222)] –3 x 6(4 + 1) = 8,6;
- *df* =3
- Ш а г 4. Определяется р-уровень значимости. Так как *к* > 3, *N* > 4, то пользуются обычной таблицей для *χ2* (приложение 4). Эмпирическое значение *χ2* находится между критическими для р = 0,05 и *p* = 0,01. Следовательно, *p*< 0,05.
- Ш а г 5. Принимается статистическое решение и формулируется содержательный вывод. На уровне а = 0,05 гипотеза Но отклоняется. Содержательный вывод: сравниваемые условия статистически достоверно различаются по уровню выраженности признака *(p* < 0,05).

Тема 14. Дисперсионный анализ (ANOVA)

- Однофакторный дисперсионный анализ ANOVA
- Методы множественного сравнения

Дисперсионный анализ ANOVA

(от англоязычного **AN**alysis **O**f **VA**riance)

- Анализ предназначен для изучения различий у трех и более выборок в уровне выраженности признака. Типичная схема эксперимента сводится к изучению влияния независимой переменной (одной или нескольких) на зависимую переменную.
- Выделяются два вида переменных – независимая и зависимая. **Независимая переменная** *(Independent Variable)* представляет собой **качественно определенный (номинативный) признак**, имеющий две или более градации. Каждой градации независимой переменной соответствует выборка объектов (испытуемых), для которых определены значения зависимой переменной. **Зависимая переменная** *(Dependent Variable)* (должна быть представлена в **метрической шкале)** в экспериментальном исследовании рассматривается как изменяющаяся под влиянием независимых переменных.

Ограничения

- дисперсии выборок, соответствующих разным градациям фактора, равны между собой

Статистические гипотезы

- **H_0:** средние значения признака в выборках 1, 2, 3, … соответствующих разным уровням фактора не отличаются.
- **H_1:** средние значения признака в выборках 1, 2, 3, … соответствующих разным уровням фактора отличаются.

Последовательность вычислений для ANOVA

- В общей изменчивости зависимой переменной выделяются основные ее составляющие. (В однофакторном ANOVA их две: внутригрупповая (случайная) и межгрупповая (факторная) изменчивость.) После этого вычисляются соответствующие показатели в следующей последовательности:

□ суммы квадратов (SS) – общая, внутригрупповая и межгрупповая;

□ числа степеней свободы (df): dftotal=N-1; dfbg = k-1(k – группа); dfwg = df total –dfbg;

□ средние квадраты (MS);

□ F-отношения;

□ p-уровни значимости.

После отклонения H0 применяется парное сравнение групп по критерию Шеффе.

$$SS_{total} = \sum_{i=1}^{N} (x_i - M)^2$$

$$SS_{bg} = \sum_{j=1}^{k} n(M_j - M)^2 ,$$

$$MS_{bg} = \frac{SS_{bg}}{df_{bg}} \qquad MS_{wg} = \frac{SS_{wg}}{df_{wg}} \qquad F_э = \frac{MS_{bg}}{MS_{wg}}, \text{ с} \qquad R^2 = \frac{SS_{bg}}{SS_{total}} .$$

$$SS_{total} = SS_{wg} + SS_{bg}.$$

$$SS_{wg} = SS_{total} - SS_{bg} = \sum_{j=1}^{k} \sum_{i=1}^{n} (x_i - M_j)^2 .$$

Виды дисперсионного анализа (ДА)

Количество зависимых переменных	**Количество независимых переменных (факторов)**		
	Один фактор	**Два и более фактора**	
	Выборки независимые		**Выборки зависимые**
Одна	Однофакторный ДА	Многофакторный ДА	ДА с повторными измерениями
Две и более		Многомерный ДА	
Ограничения	дисперсии выборок, соответствующих разным градациям фактора, равны между собой (критерий Левина)	ковариационно-дисперсионные матрицы, соответствующих разным уровням межгрупповых факторов идентичности, должны быть идентичны (критерий М-Бокса)	
		дисперсии выборок, соответствующих разным градациям фактора, равны между собой (критерий Левина)	

Нахождение однофакторного ANOVA

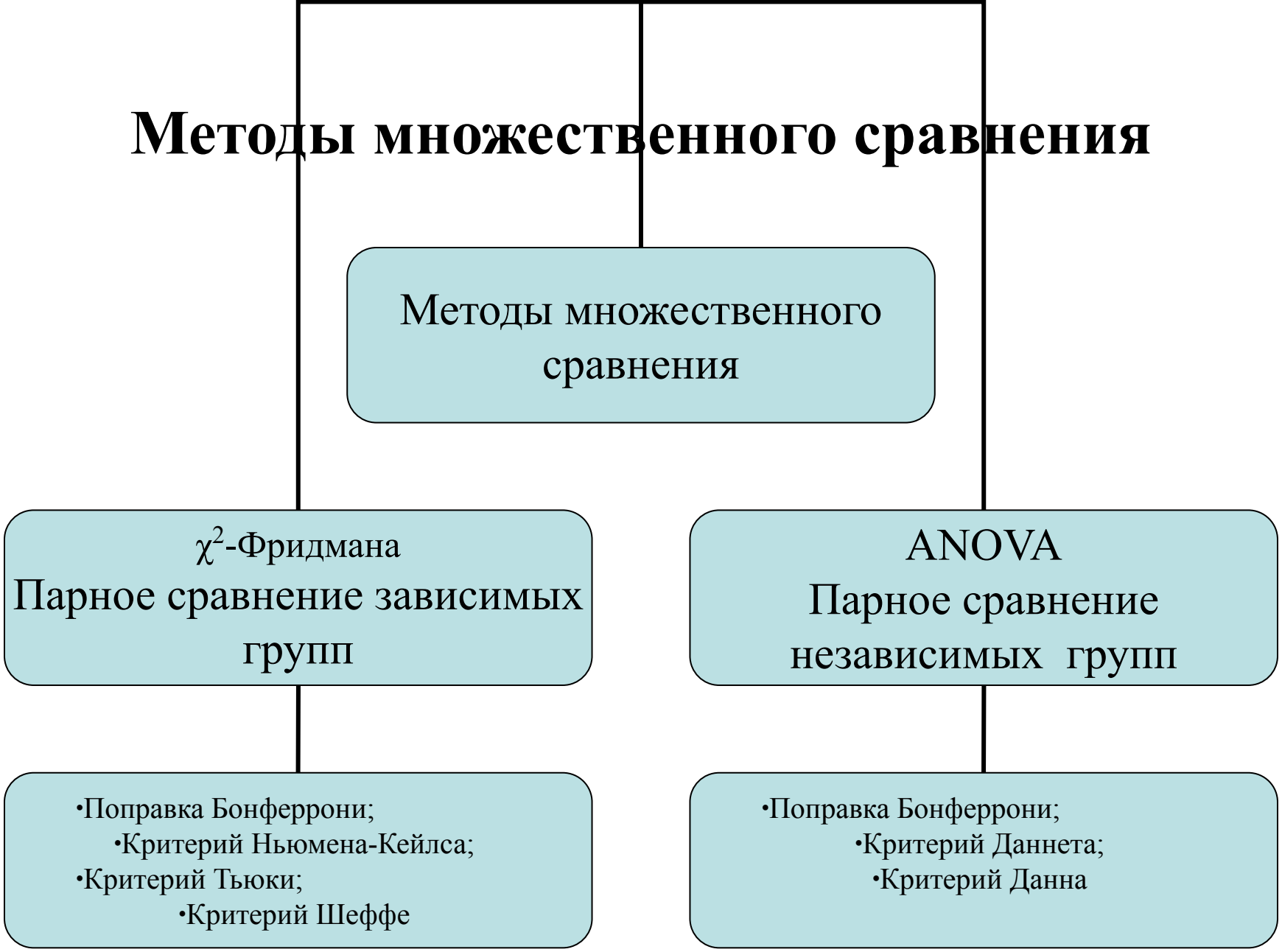
- Общее среднее: *M= 7.*
- Среднее для разных условий: M1 = 5; *M2* = 7; *M3* = 9.
- Ш а г 1. Вычислим внутригрупповые суммы квадратов:
- SStotal= (5-7)2 +(4-7)2 +... + (8-7)2 =70
- SSbg = 5[(5-7)2 +(7-7)2 +(9-7)2] = 40
- SSwg = 70 – 40 = 30
- Ш а г 2. Определим числа степеней свободы:
- *dfbg* =k- 1 = 3 - 1 = 2; *df*wg = N – k = 15 – 3 = 12
- Ш а г 3. Вычислим средние квадраты:
- MSbg = 40/2 = 20; MSwg = 30/12= 2.5
- Ш а г 4. Вычислим F-отношение:
- Шаг 5. Определим p-уровень значимости. По таблице критических значений F-распределения (для направленных альтернатив) для p = 0,01; dfчисл = 2; dfзнам = 12 критическое значение равно F— 6,927. Следовательно, $p < 0,01$, т.к.
- Дополнительно вычислим коэффициент детерминации: R2 = 0,571.
- Отклоняем Но и принимаем альтернативную гипотезу о том, что межгрупповая изменчивость выше внутригрупповой.

Условие 1		**Условие 2**		**Условие 3**	
№	***Y***	**№**	***Y***	**№**	***Y***
1	5	6	8	11	11
2	4	7	7	12	9
3	3	8	6	13	7
4	6	9	9	14	10
5	7	10	5	15	8

Источник изменчивости	*SS*	*df*	*MS*	F	p-
Межгрупповой	40	2	20	8	p<0,01
Внутригрупповой	30	12	2,5	—	—

$$R^2 = \frac{SS_{bg}}{SS_{total}}.$$

Методы множественного сравнения



Тема 15. Многомерные методы

- Определение и классификация многомерных методов
- Регрессионный анализ (частный случай множественного регрессионного анализа)
- Множественный регрессионный анализ
- Дискриминантный анализ
- Факторный анализ
- Кластерный анализ
- Многомерное шкалирование

- **Многомерные методы** - это математические модели в отношении многостороннего (многомерного) описания изучаемых явлений. ММ воспроизводят мыслительные операции человека, но в отношении таких данных, непосредственное осмысление которых невозможно в силу нашей природной ограниченности. Многомерные методы выполняют такие интеллектуальные функции, как структурирование эмпирической информации (факторный анализ), классификация (кластерный анализ), экстраполяция (множественный регрессионный анализ), распознавание образов (дискриминантный анализ) и т. д.

Классификация многомерных методов

Решаемая задача	Зависимая переменная	Независимые переменные	Используемый метод
Классификация	Номинативная, порядковая	Номинативная, порядковая	Метод условных вероятностей Байеса
	Количественная	Номинативная, порядковая	Многофакторный дисперсионный анализ
Классификация, прогноз	Номинативная, порядковая	Количественная	Дискриминантный анализ Кластерный анализ
	Количественная	Количественная	Множественная регрессия Кластерный анализ
Анализ структуры взаимосвязей		Номинативная, порядковая	Многомерное шкалирование Кластерный анализ
		Количественная	Многомерное шкалирование, Кластерный анализ Факторный анализ

Регрессионный анализ (частный случай множественного регрессионного анализа)

- **Регрессионный анализ** — основан на коэффициенте детерминации. Регрессионный анализ применяется, для предсказания значения одной переменной, если известны значения другой, т.е. для исследования взаимосвязи зависимой одной *y* и одной независимой *x* переменных.
- **Линия регрессии**, обобщает все точки рассеяния наилучшим способом из возможных. Иными словами, абсолютные значения расстояний по вертикали между каждой точкой графика и линией регрессии минимальны.
- Переменная, по которой предсказывают, называется **предикторной**. Обычно ее значения откладываются по оси *X*.
- Переменная, которую предсказывают, называется **критериальной.** Ее значения откладываются по оси *Y*.

Уравнение линейной регрессии

- Если переменные пропорциональны друг другу, то графически связь между ними можно представить в виде прямой линии с положительным (прямая пропорция) или отрицательным (обратная пропорция) наклоном. Кроме того, если известна пропорция между переменными, заданная уравнением графика прямой линии, то по известным значениям переменной X можно точно предсказать значения переменной *Y.*
- На практике связь между двумя переменными, если она есть, является вероятностной и графически выглядит как облако рассеивания эллипсоидной формы. Этот эллипсоид, однако, можно представить (аппроксимировать) в виде прямой линии, или *линии регрессии.*
- **Линия регрессии** *(Regression Line)* — это прямая, построенная методом наименьших квадратов: сумма квадратов расстояний (вычисленных по оси *Y)* от каждой точки графика рассеивания до прямой является минимальной:
- где *yi,* — истинное i-значение У, $\sum_i (y_i - \hat{y}_i)^2 = \sum_i e_i^2 = \min,$
- yi, — оценка i-значения Упри помощи линии (уравнения) регрессии,
- ei,-= yi-yi,— ошибка оценки.
- Уравнение регрессии имеет вид: $\hat{y}_i = bx_i + a,$
- где *b — коэффициент регрессии (Regression Coefficient),* задающий угол наклона прямой;
- *a — свободный член,* определяющий точку пересечения прямой оси *Y.*
- Угловой коэффициент регрессии (b) показывает, насколько в среднем величина признака y изменяется при соответствующем изменении на единицу признака x. Таким образом, если на некоторой выборке измерены две переменные, которые коррелируют друг с другом, то, вычислив коэффициенты регрессии, мы получаем принципиальную возможность предсказания неизвестных значений одной переменной (Y- зависимая переменная) по известным значениям другой переменной (X – независимая переменная).

Расчеты уравнения регрессии

Пример: Школьникам была дана тестовая задача, которую им необходимо было решить, при этом регистрировалось скорость выполнения задания и количество ошибок. Необходимо установить возможность предсказания количества ошибок в зависимости от скорости выполнения заданий теста и определить параметры уравнения линейной регрессии в зависимости от ошибок и скорости выполнения заданий теста.

N	Время выполнения (x)	Количество ошибок (y)	x^2	xy	Расчет
1	6	4	36	24	$a = \frac{\Sigma y \Sigma x^2 - \Sigma yx \Sigma x}{n\Sigma x^2 - (\Sigma x)^2} = \frac{24 \cdot 187 - 149 \cdot 29}{5 \cdot 187 - 841} = 1{,}78$ $b = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma x^2 - \Sigma x \Sigma y} = \frac{5 \cdot 149 - 29 \cdot 24}{5 \cdot 187 - 841} = 0{,}52$ $y = 1{,}78 + 0{,}52x$
2	9	7	81	63	
3	3	4	9	12	
4	5	4	25	20	
5	6	5	36	30	
∑	29	24	187	149	

Множественный регрессионный анализ

Множественный регрессионный анализ (МРА) предназначен для изучения взаимосвязи одной переменной *(зависимой, результирующей - y)* и нескольких других переменных *(независимых, исходных - x)*. Частный случай регрессионный анализ для исследования взаимосвязи зависимой одной y и одной независимой x переменных.

Ограничения

1. *Главное требование к исходным данным* — отсутствие линейных взаимосвязей между переменными, когда одна переменная является линейной производной другой переменной. Следует избегать включения в анализ переменных, корреляция между которыми близка к 1, так как сильно коррелирующая переменная не несет для анализа новой информации, добавляя излишний «шум».
2. *Следующее требование* — переменные должны быть измерены в метрической шкале (интервалов или отношений) и иметь нормальное распределение.

Основными целями МРА являются

1. Определение того, в какой мере «зависимая» переменная связана с совокупностью «независимых» переменных, какова статистическая значимость этой взаимосвязи. Показатель — коэффициент множественной корреляции (КМК - *R*) и его статистическая значимость по критерию F-Фишера,
2. Определение существенности вклада каждой «независимой» переменной в оценку «зависимой» переменной, отсев несущественных для предсказания «независимых» переменных. Показатели — регрессионные коэффициенты β, их статистическая значимость по критерию t-Стьюдента.
3. Анализ точности предсказания и вероятных ошибок оценки «зависимой» переменной. Показатель — квадрат КМК (КМД - *R2)*, интерпретируемый как доля дисперсии «зависимой» переменной, объясняемая совокупностью «независимых» переменных. Вероятные ошибки предсказания анализируются по расхождению (разности) действительных значений «зависимой» переменной и оцененных при помощи модели МРА.
4. Оценка (предсказание) неизвестных значений «зависимой» переменной по известным значениям «независимых» переменных. Осуществляется по вычисленным параметрам множественной регрессии.

Дискриминантный анализ

Предназначен для изучения взаимосвязи одной переменной *(зависимой, результирующей - y)* и нескольких других переменных *(независимых, исходных - x)*.

Ограничения

Зависимая переменная должна быть представлена в номинативной шкале, а независимые измерены в метрической шкале (интервалов или отношений) и иметь нормальное распределение.

Дискриминантный анализ позволяет решить две группы проблем:

1. **Интерпретировать различия между классам**и, то есть ответить на вопросы: насколько хорошо можно отличить один класс от другого, используя данный набор переменных; какие из этих переменных наиболее существенны для различения классов.
2. **Классифицировать объекты**, то есть отнести каждый объект к одному из классов, исходя только из значений дискриминантных переменных.

Основные результаты дискриминантного анализа

1. Определение статистической значимости различения классов при помощи данного набора дискриминантных переменных. Показатели — λ-Вилкса, χ2-тест, p-уровень значимости.
2. Выяснение вклада каждой переменной в дискриминантный анализ. Определяется по значениям критерия F-Фишера, толерантности и статистики F-удаления.
3. Вычисление расстояний между центроидами классов и определение их статистической значимости по F-критерию.
4. Анализ канонических функций, их интерпретация через дискриминантные переменные (по стандартизированным и структурным коэффициентам канонических функций).
5. Классификация «известных» и «неизвестных» объектов при помощи расстояний или значений априорных вероятностей. Качество классификации определяется совпадением действительной классификации и предсказанной для «известных» объектов. Мерой качества может служить ве роятность ошибочной классификации как соотношение количества ошибочного отнесения к общему количеству «известных» объектов.
6. Графическое представление всех объектов и центроидов классов в осях канонических функций.

Факторный анализ

- *Главная цель факторного анализа* — уменьшение размерности исходных данных.
- *Результатом* факторного анализа является переход от множества исходных переменных к существенно меньшему числу новых переменных — факторов. Фактор при этом интерпретируется как причина совместной изменчивости нескольких исходных переменных.
- *Основное назначение* факторного анализа — *анализ корреляций* множества признаков.

Область применения факторного анализа (задачи)

1. Исследование структуры взаимосвязей переменных. В этом случае каждая группировка переменных будет определяться фактором, по которому эти переменные имеют максимальные нагрузки. Нагрузки исследуемых факторов представляют корреляцию с общими факторами.
2. Идентификация факторов как скрытых (латентных) переменных — причин взаимосвязи исходных переменных.
3. Вычисление значений факторов для испытуемых как новых, интегральных переменных. При этом число факторов существенно меньше числа исходных переменных. В этом смысле факторный анализ решает задачу сокращения количества признаков с минимальными потерями исходной информации.

Основные этапы факторного анализа

1. Выбор исходных данных.
2. Предварительное решение проблемы числа факторов: используются критерий отсеивания Р. Кетелла (требует построения графика) и критерий Г. Кайзера (определяется по числу компонент, собственные значения которых больше 1).
3. Факторизация матрицы интеркорреляций, вращение факторов (Задается число факторов, производится вращение методом «Варимакс-нормализованное». Результатом данного этапа является матрица факторных нагрузок (факторная структура) .
4. Интерпретация факторов: *По каждому фактору* выписывают наименования (обозначения) переменных, имеющих наибольшие нагрузки по этому фактору — выделенных на предыдущем шаге. При этом обязательно учитывается знак факторной нагрузки переменной. Если знак отрицательный, это отмечается как противоположный полюс переменной. После такого просмотра всех факторов каждому из них *присваивается наименование,* обобщающее по смыслу включенные в него переменные.

Кластерный анализ

- **Кластерный анализ** — это процедура упорядочивания объектов в сравнительно однородные классы на основе попарного сравнения этих объектов по предварительно определенным и измеренным критериям.
- Кластерный анализ решает задачу построения *классификации,* то есть разделения исходного множества объектов на группы (классы, кластеры).
- *Классификация* объектов — это группирование их в классы так, чтобы объекты в каждом классе были более похожи друг на друга, чем на объекты из других классов.

Задачи кластерного анализа:

- разбиение совокупности испытуемых на группы по измеренным признакам с целью дальнейшей проверки причин межгрупповых различий по внешним критериям, например, проверка гипотез о том, проявляются ли типологические различия между испытуемыми по измеренным признакам;
- применение кластерного анализа как значительно более простого и наглядного аналога факторного анализа, когда ставится только задача группировки признаков на основе их корреляции.

Этапы кластерного анализа

1. *Отбор объектов для кластеризации.* Объектами могут быть, в зависимости от цели исследования: а) испытуемые; б) объекты, которые оцениваются испытуемыми; в) признаки, измеренные на выборке испытуемых.
2. *Определение множества переменных,* по которым будут различаться объекты кластеризации. Для испытуемых — это набор измеренных признаков, для оцениваемых объектов — субъекты оценки, для признаков — испытуемые.
4. *Выбор и применение метода классификации* для создания групп сходных объектов. Это вторая и центральная проблема кластерного анализа. Ее весомость связана с тем, что разные методы кластеризации порождают разные группировки для одних и тех же данных. Наиболее популярные методы: одиночной связи, полной связи и средней связи.
5. *Проверка достоверности разбиения* на классы (используются критерии сравнения).

Многомерное шкалирование

- Основная *цель многомерного шкалирования* (МШ) — выявление структуры исследуемого множества объектов
- *Главная задача МШ* — реконструкция психологического пространства, заданного небольшим числом измерений-шкал, которые интерпретируются как критерии, лежащие в основе различий стимулов.

Основные этапы многомерного шкалирования

- Определение величины стресса (φ-Stress), который является показателем точности - наиболее приемлемый для него диапазон от 0,05 до 0,2. Вычисление коэффициентов отчуждения (D-star) и напряжения (D-hat). Чем меньше эти величины тем лучше воспроизведена матрица расстояния в наблюдаемой модели.
- Построение итоговой конфигурации нагрузки объектов по выделенным шкалам.
- Построение графика.
- Интерпретация шкал по итоговой конфигурации и графику (интерпретация шкал осуществляется через входящие в них объекты).

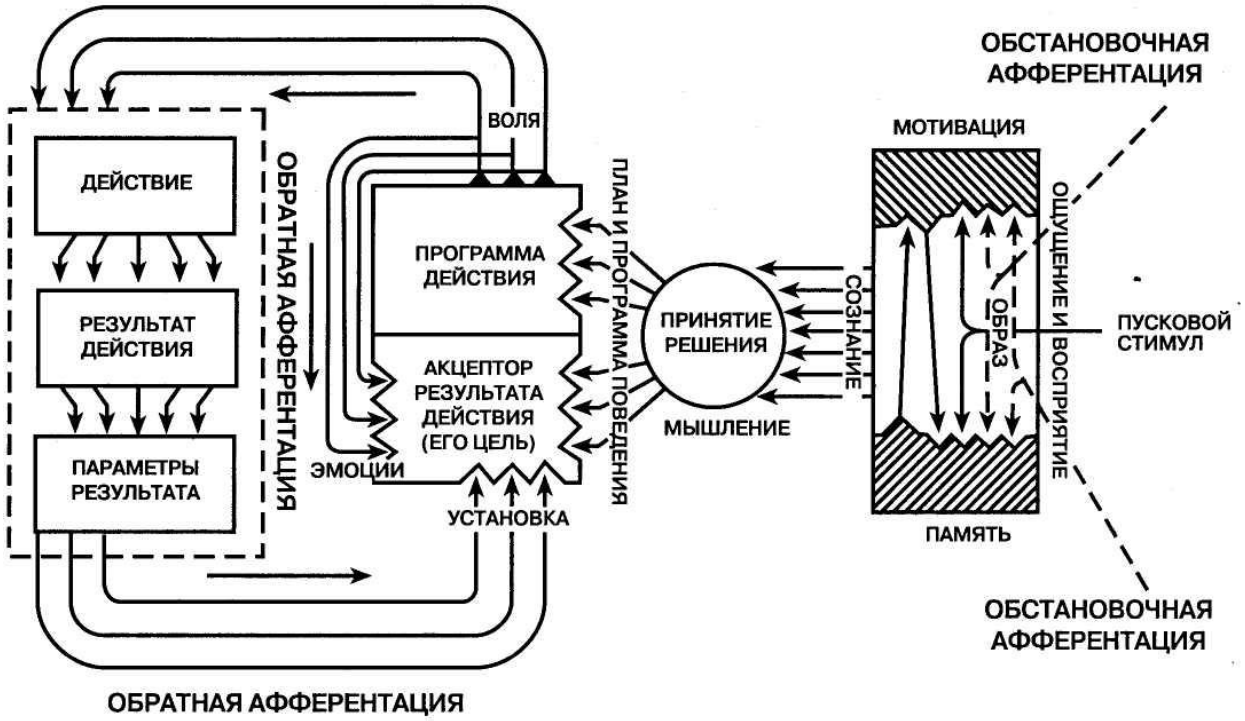
Тема 16. Математическое моделирование в психологии

- Системные подходы.
- Теория функциональных систем.
- Становление кибернетики.
- Системный анализ.
- Теория катастроф.
- Методы математического моделирования в психодиагностике: априорные и апостериорные модели.
- Проблема искусственного интеллекта.

- **Система** - множество элементов, находящихся в отношениях и связях друг с другом, которое образует определенную целостность, единство.
- **Признаки системы:**
- система обладает целостностью, все ее части служат достижению единой цели;
- система является большой как с точки зрения разнообразия составляющих ее элементов, так и с точки зрения количества одинаковых частей;
- система является сложной, что означает наличие большего количества связей между элементами как по вертикали, так и по горизонтали. Следовательно, изменение в каком - либо одном компоненте влечет за собой изменение в других;
- независимо от сложности и размера система обладает чертами «черного ящика», их поведение в любой момент недетерминировано как в силу стохастической природы входных действий, так и внутреннего ее поведения;
- большинство систем, и в первую очередь наиболее сложные системы, содержат элементы конкурентной ситуации, т.е. обязательно существуют элементы, которые стремятся уменьшить эффективность системы.

Теория функциональных систем (модель П. К. Анохина)

- Центральная нервная система представлена в виде нервной модели



Кибернетика Н. Винера

- Человек, один из самых сложных объектов реального мира, известных науке в настоящее время. Он не только самоактуализирующийся и саморегулируемый, но и саморазвивающийся объект. Его свойство как саморазвивающегося объекта состоит в том, что он в состоянии самостоятельно создавать и изменять программу своих действий.
- Другое дело технические системы. В отличие от живого организма все можно оценить и исследовать с момента их создания. Можно установить закономерности их функционирования.

Синергетика (Г. Хакена)

- По Хакену, синергетика занимается изучением систем, состоящих из большого (очень большого, «огромного») числа частей, компонент или подсистем, одним словом, деталей, сложным образом взаимодействующих между собой. Слово «синергетика» и означает «совместное действие», подчеркивая согласованность функционирования частей, отражающуюся в поведении системы как целого.
- Синергетический процесс самоорганизации материи это бесконечное чередование этапов «спокойной» адаптации и «революционных» перерождений, выводящих системы на новые ступени совершенства.

Общая теория систем Л. Фон Берталанфи

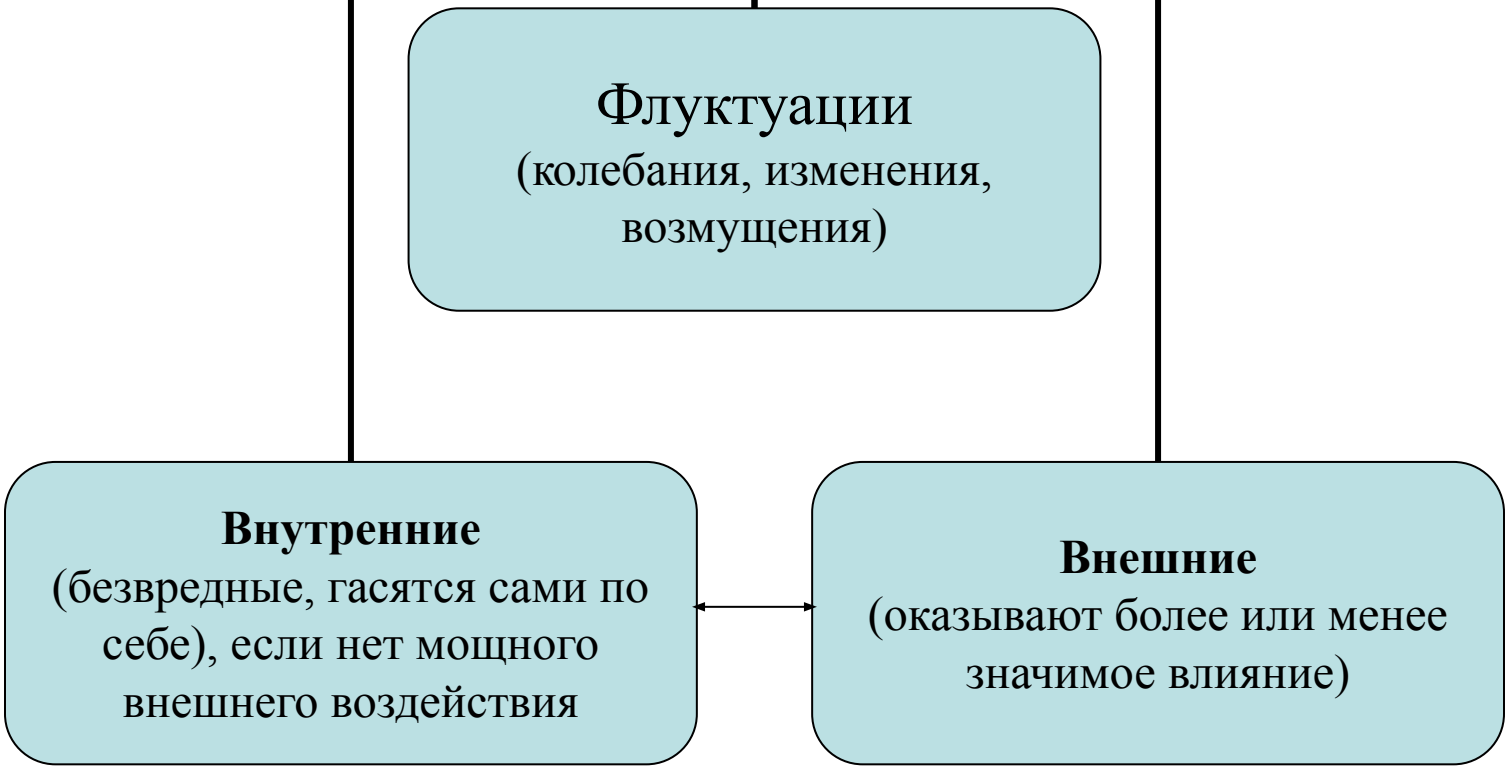
- **Общая теория систем Л. Фон Берталанфи** состоит в том, что если замкнутую систему вывести из состояния равновесия, то в ней начнутся процессы, возвращающие ее к состоянию термодинамического равновесия, в котором ее энтропия достигает максимального значения.

Теория развития И.Р. Пригожина

- Теория развития И.Р. Пригожина гласит, что если отток энтропии (меры необратимого рассеяния энергии) превышает ее внут реннее производство, то возникают и разрастаются до макроскопического уровня крупномасштабные флуктуации.

Теория катастроф

- **Катастрофами** называются скачкообразные изменения, возникающие в виде внезапного ответа объекта па плавные изменения внешних условий.



Системный анализ

- **Системный анализ -** научная дисциплина, разрабатывающая общие принципы исследования сложных объектов с учетом их системного характера.

Этапы системного анализа любого объекта:

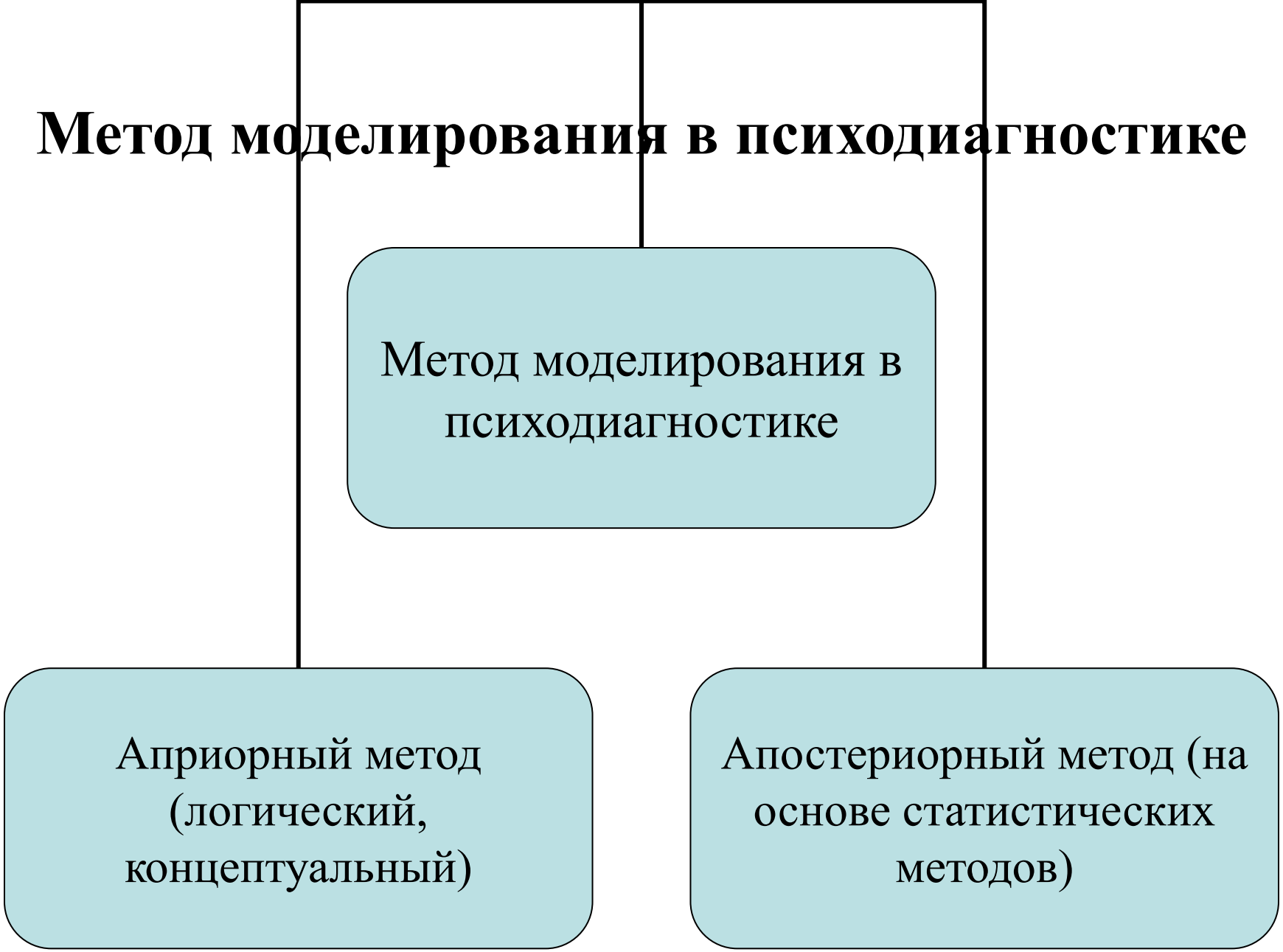
- Постановка задачи - определение объекта исследования, постановка целей, задание критериев для изучения объекта и управления им.
- Выделение системы, подлежащей изучению, и ее структуризация.
- Составление математической модели изучаемой системы: параметризация, установление зависимостей между введенными параметрами, упрощение описания системы путем выделения подсистем и определения их иерархии, окончательная функция целей и критериев.

Моделирование сложных систем

Этапы моделирования сложных процессов и явлений:

- Формулировка цели моделирования.
- Анализ объекта исследования, включающий статистическую обработку параметров для определения математического ожидания, типа распределения и других описательных статистик.
- Выявление причинно-следственных связей. Определение независимых и зависимых переменных. Для этого используется математический аппарат кластерного анализа, называемый также аппаратом поиска естественной классификации.
- Определение степени сложности и организации моделируемой системы.
- Выбор класса и вида модели. В зависимости от уровня организации объекта выбирается класс математической модели: линейная, нелинейная, детерминированная, вероятностная. Класс модели во многом определяет математический аппарат, наиболее подходящий для описания работы модели. В выбранном классе определяется вид модели. Существует множество видов внутри одного класса. Так, например, к классу нелинейных моделей относятся полиномиальные, дифференциальные уравнения и т. д.
- Синтез параметров модели или собственно моделирование.
- Верификация созданной модели с использованием независимого массива.

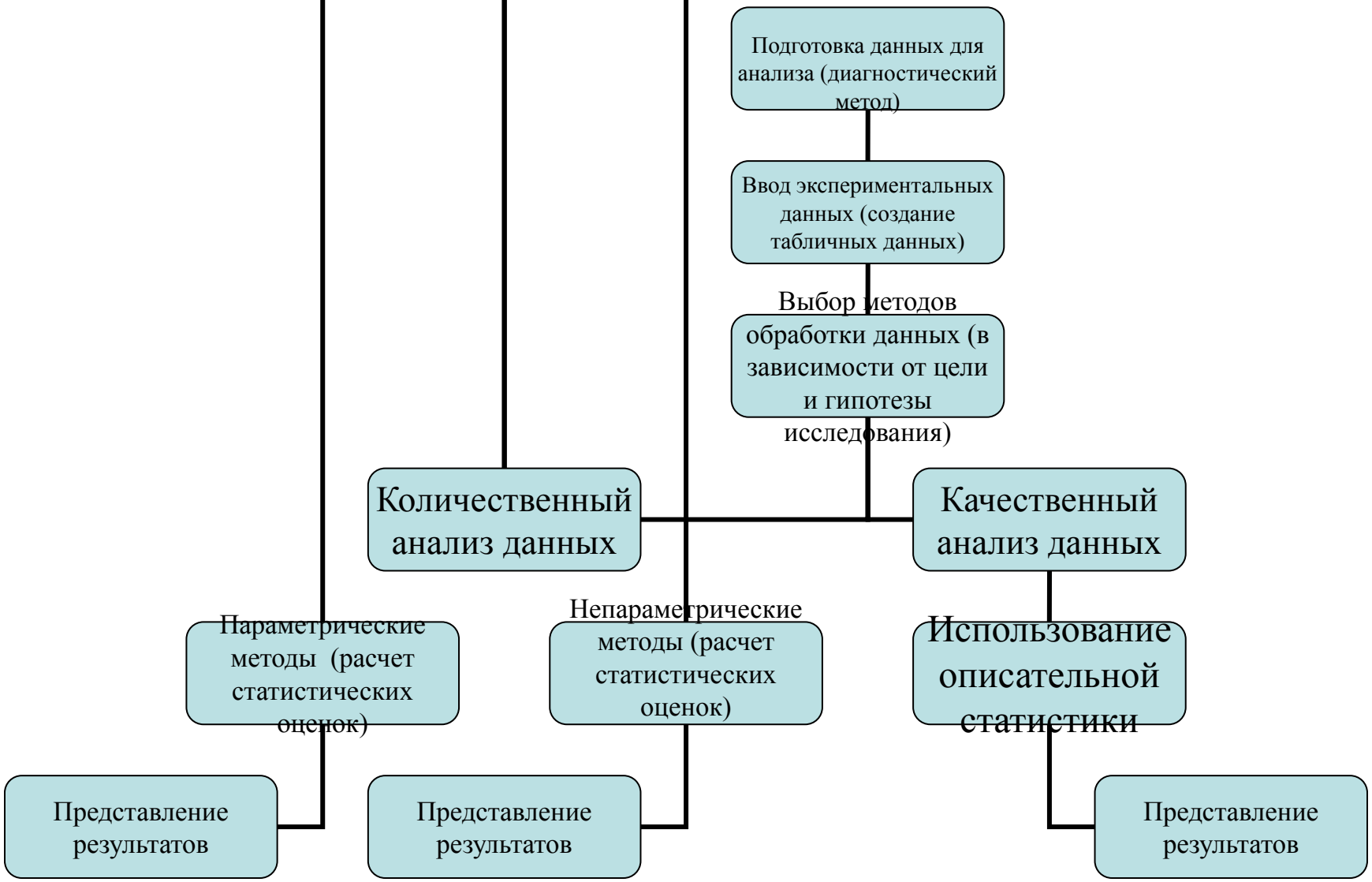
Метод моделирования в психодиагностике



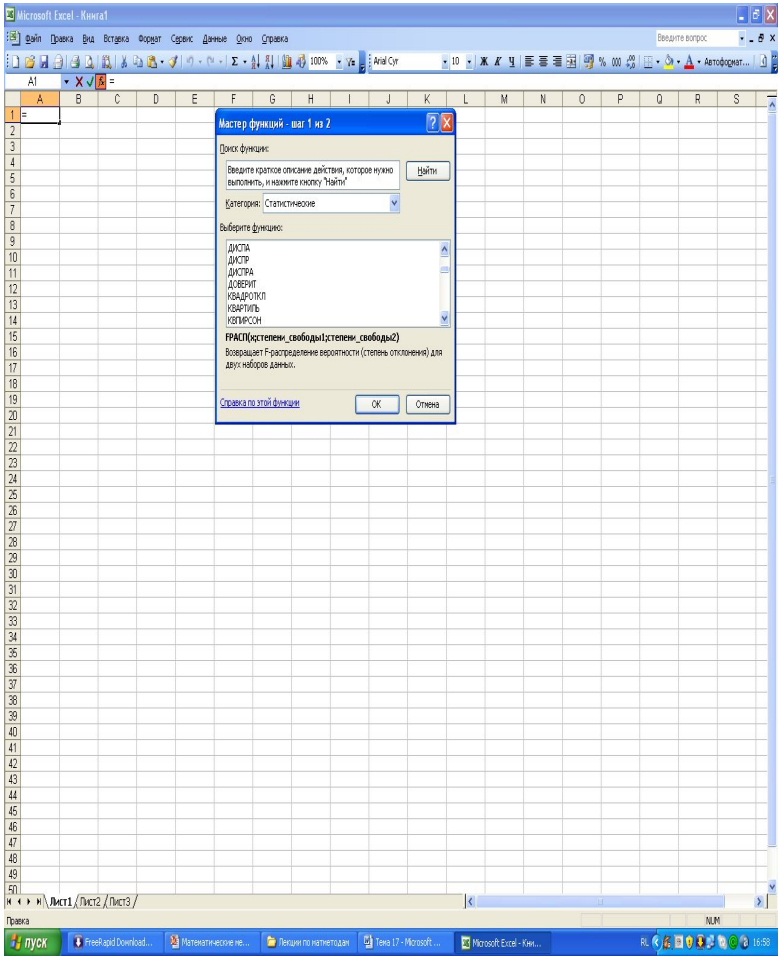
Тема 17. Анализ данных на компьютере.

- Использование MS Excel
- Статистические пакеты: SPSS, STATISTICA.
- Особенности подготовки данных для анализа на компьютере.

Алгоритм применения анализа данных на компьютере



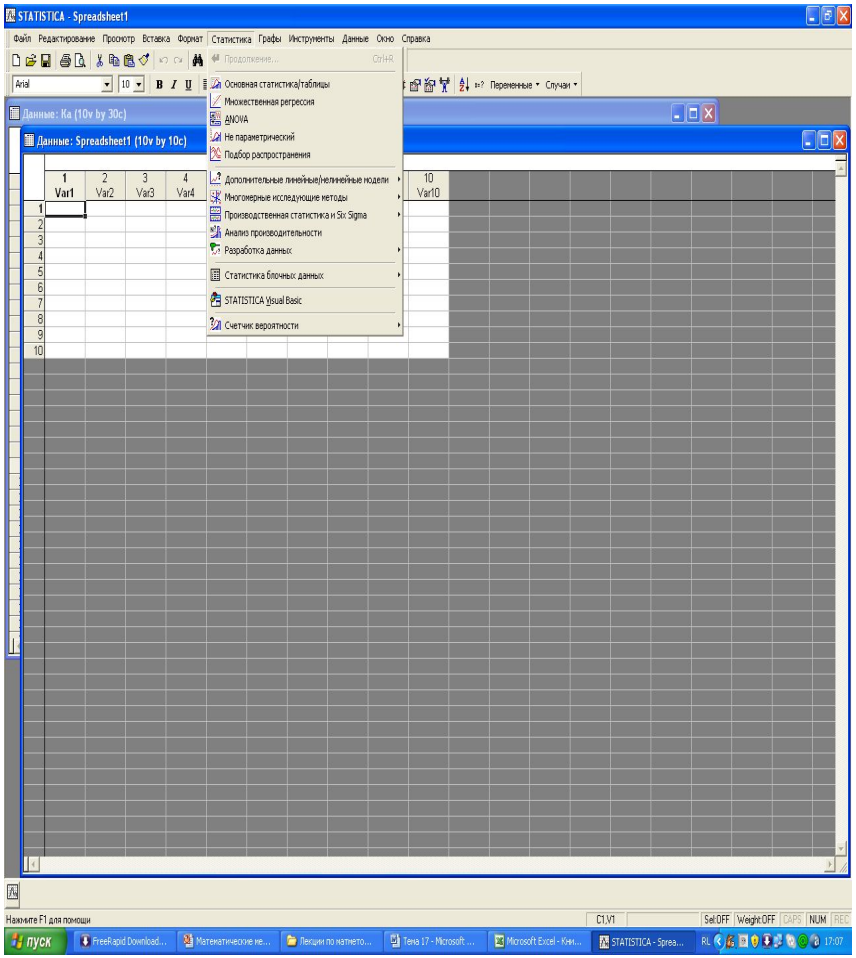
Использование MS Excel



Плюсы и минусы MC Excel

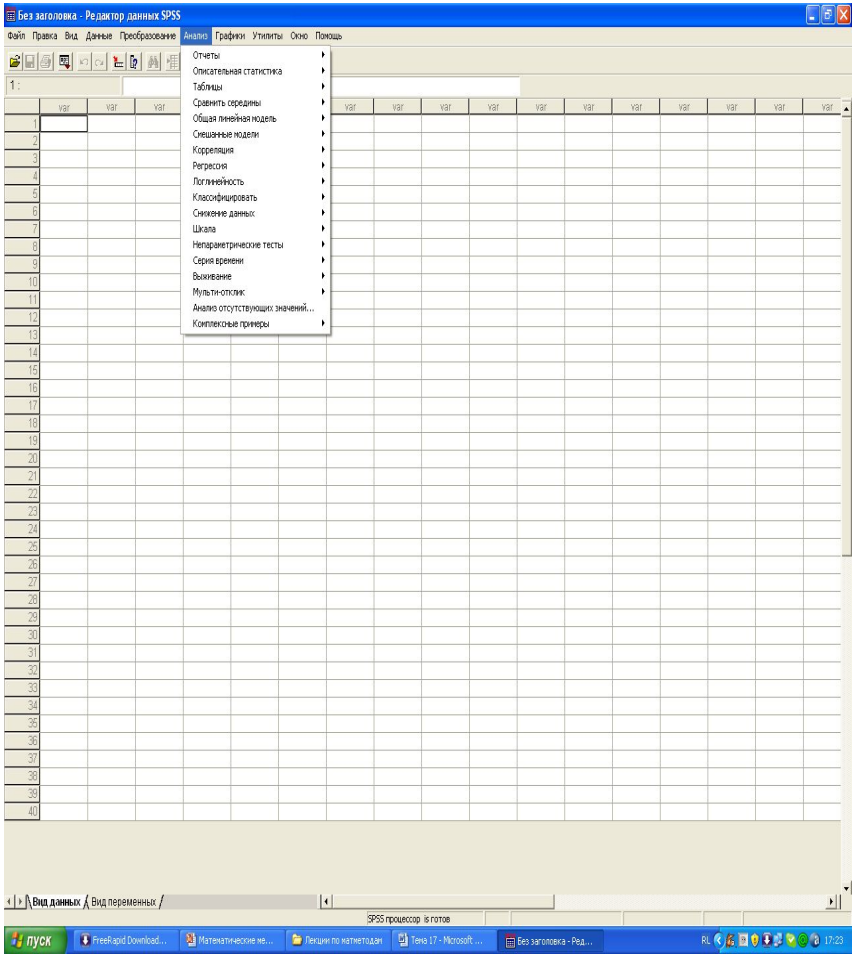
- В Microsoft Excel входит набор средств анализа данных (так называемый пакет анализа), предназначенный для решения довольно сложных статистических задач. Для проведения анализа данных с помощью этих инструментов следует указать входные данные и выбрать параметры; анализ будет проведен с помощью подходящей статистической макрофункции, а результат будет помещен в выходной диапазон. Другие инструменты позволяют представить результаты анализа в графическом виде. Статистические методы, имеющихся в пакете анализа, достаточно для обработки первичных данных.
- Однако при больших массивах данных, анализ в этой программной среде приводит к существенному увеличению ошибок. Кроме того, отсутствие в Microsoft Excel возможности кодирования номинальных и порядковых показателей приводит к необходимости многократной сортировки данных по номинальным показателям, если в исследовании их несколько. И, наконец, пакет анализа достаточно капризен. Например, если в массиве данных имеется, хотя бы один пропуск (незаполненная ячейка), Microsoft Excel отказывается считать корреляцию и т. д.

Статистические пакеты: SPSS, STATISTICA



- STATISTICA for Windows представляет собой интегрированную систему статистического анализа и обработки данных. Она состоит из следующих основных компонент, которые объединены в рамках одной системы:
- электронных таблиц для ввода и задания исходных данных, а также специальных таблиц для вывода численных результатов анализа;
- мощной графической системы для визуализации данных и результатов статис тического анализа;
- набора специализированных статистических модулей, в которых собраны группы логически связанных между собой статистических процедур;
- специального инструментария для подготовки отчетов;
- встроенных языков программирования *SCL (STATISTICA Command Language)* и *STATISTICA BASIC,* которые позволяют пользователю расширить стандартные возможности системы.

SPSS



- Альтернативное программное обеспечение SPSS включает также все процедуры ввода, отбора и корректировки данных, а также большинство предлагаемых в SPSS статистических методов, что и в STATISTICA. Наряду с простыми методиками статистического анализа, такими как частотный анализ, расчет статистических характеристик, таблиц сопряженности, корреляций, построения графиков, этот модуль включает t-тесты и большое количество других непараметрических тестов, а также усложненные методы, такие как многомерный линейный регрессионный анализ, дискриминантный анализ, факторный анализ, кластерный анализ, дисперсионный анализ, анализ пригодности (анализ надежности) и многомерное шкалирование.