

# Статистические методы в психологии.

Введение в дисциплину. Понятие измерения.

Виды измерительных шкал. Описательные статистики. Закон нормального распределения

в психологии. Общие принципы проверки статистических гипотез в психологии.

Непараметрические, параметрические критерии. Корреляционный анализ

**Григорьев Павел Евгеньевич**

Доктор биологических наук, профессор кафедры общей и социальной психологии Института психологии и педагогики ТюмГУ

[grigorievpe@cfuv.ru](mailto:grigorievpe@cfuv.ru)

[www.vk.com/int.research](http://www.vk.com/int.research)

## Дисциплина «Статистические методы в психологии»

- **Цель курса:** научиться грамотному использованию методов статистической обработки результатов экспериментальных, научно-практических исследований.
- **Задачи курса:**
  - ознакомить магистров с основными методами статистической обработки психологических данных;
  - сформировать навыки применения статистических методов;
  - изучить многомерные методы обработки результатов психологического исследования;
  - сформировать навыки создания математических моделей в психологии.

# Математические методы в психологическом исследовании

- Научная проверка гипотез экспериментального психологического исследования возможна лишь с привлечением методов теории вероятности и статистики. Поэтому корректное количественное описание психологических явлений, а также верификация и проверка гипотез являются важнейшими сферами применения элементарных математических методов в психологии.
- Любое исследование предполагает получение качественных результатов (да/нет?) и количественных результатов (насколько?). Наиболее содержательные и точные результаты предполагают количественное (численное) выражение. Но просто собрать данные психологического исследования и представить их в виде чисел недостаточно. Исследователю необходимо организовать данные, обработать их и проинтерпретировать, что невозможно без применения современных математических методов.

# Математические методы в психологическом исследовании

Для их корректного и результативного использования необходимо:

- 1) **организовать психологическое исследование** так, чтобы его результаты были доступны математической обработке в соответствии с проблемами исследования;
- 2) правильно **выбрать метод математической обработки;**
- 3) содержательно **интерпретировать результаты обработки.**

Ценность математического метода определяется теми **статистически значимыми, однозначно определенными, клинически или психологически значимыми** выводами (об исследованных психологических показателях или явлениях), которые вытекают из результатов математической обработки.

## Объект, предмет, свойство, признак, измерение...

- Следует различать **объекты исследования** (например, испытуемые с определенными характеристиками), их **свойства** – то, что в действительности интересует исследователя, составляет предмет изучения, (например, агрессивность) и **признаки**, отражающие в числовой шкале выраженность свойств (например, число баллов в опроснике Басса-Дарки измерения уровня агрессивности).
- **Признак** – реально измеряемое явление, в той или иной степени отражающее интересующее исследователя **свойство** изучаемого **объекта**.
- **Измерение** – это приписывание объекту числа или знака по определенному правилу. Это правило устанавливает соответствие между измеряемым свойством объекта и результатом измерения – признаком. Точность, с которой признак отражает измеряемое свойство, зависит от процедуры измерения.

## Объект, предмет, свойство, признак, измерение...

Любое исследование в зависимости от того, насколько надежны полученные в нем результаты и насколько они применимы на практике, можно охарактеризовать с двух точек зрения:

- **достоверности (внутренней обоснованности);**
- **обобщаемости (внешней обоснованности, применимости).**

Достоверность исследования определяется тем, в какой степени структура и методы исследования соответствует поставленным задачам, а полученные результаты справедливы в отношении изучавшейся выборки.

Обобщаемость результатов исследования отражает, в какой мере результаты данного исследования применимы к другим (прежде всего, аналогичным, но и другим в некоторых отношениях) группам.

## Особенности статистического описания и метода

- Статистическое описание совокупности объектов занимает промежуточное положение между **индивидуальным описанием** каждого из объектов совокупности (*например, описание конкретного случая работы с клиентом*), с одной стороны, и **описанием совокупности** по её общим свойствам (*например, только общие черты, присущие всем представителям некоторой субпопуляции, например, подросткам с возбудимой психопатией*), совсем не требующим её расчленения на отдельных субъектов, – с другой.
- **Статистические данные** всегда в большей или меньшей степени обезличены и имеют лишь ограниченную (особенно в практическом отношении) ценность в случаях, когда наиболее существенны именно индивидуальные данные.

## Особенности статистического описания и метода

Обычно применение статистического метода предусматривает:

- 1) подсчёт числа объектов, входящих в те или иные группы;
- 2) рассмотрение распределения признаков;
- 3) применение **выборочного метода** (в случаях, когда детальное исследование всех объектов обширной **генеральной совокупности** затруднительно);
- 4) использование теории вероятностей при оценке достаточности числа наблюдений для тех или иных выводов.

## Особенности статистического описания и метода

- Психологическое исследование обычно начинается с некоторой гипотезы, требующей проверки с привлечением фактов. *Гипотеза формулируется в отношении **связи явлений или свойств** в некоторой совокупности объектов.* Например, исследователь может предположить, что женщины в среднем более тревожны, чем мужчины (тревожность связана с полом). Или что аддикция компьютерными играми со сцены насилия, повышает агрессивность подростков. В первом случае исследователя интересуют связь свойств тревожности и пола, во втором – связь явления увлечения играми со свойством агрессивности. Объектами-носителями свойств в первом случае в пределе будут все мужчины и женщины, во втором все подростки.
- Для проверки подобных предположений на фактах необходимо измерить соответствующие свойства у их носителей-объектов. Но невозможно измерить тревожность у ВСЕХ мужчин и женщин в мире, как невозможно измерить агрессивность у ВСЕХ подростков, в разной степени увлеченными или не увлеченными играми с насилием! Поэтому при проведении исследования ограничиваются лишь относительно небольшой группой представителей соответствующей совокупности пюлей

# Генеральная совокупность и выборка

- **Генеральная совокупность** – это все множество объектов, в отношении которого формулируется исследовательская гипотеза.

Необязательно генеральные совокупности огромны по численности объектов. Например, при изучении профессионального самоопределения студентов-выпускников факультета психологии в конкретном вузе генеральная совокупность невелика и допускает **сплошное** (а не **выборочное**) исследование. Но исследователь обычно надеется, что выводы исследования будут справедливы (ДОСТОВЕРНЫ) не только в отношении выпускников этого, но и последующих годов (ОБОБЩАЕМЫ).

Таким образом, **генеральная совокупность** – это хотя и не бесконечное по численности, но, как правило, недоступное по тем или иным причинам для сплошного исследования множество потенциальных испытуемых.

- **Выборка** – это ограниченная по численности группа объектов (в психологии – испытуемых, респондентов), специально отбираемая из генеральной совокупности для изучения ее свойств. Соответственно, изучение на выборке свойств генеральной совокупности называется **выборочным исследованием**. Почти все психологические исследования являются выборочными, а их выводы распространяются на генеральные совокупности.

## Генеральная совокупность и выборка

- Таким образом, после того, как сформулирована гипотеза и определены соответствующие генеральные совокупности, перед исследователем возникает проблема организации выборки. Выборка должна быть такой, чтобы было обосновано обобщение выводов выборочного исследования, распространение их на генеральную совокупность.
- **Основные критерии обоснованности выводов исследования:**
  - репрезентативность выборки;
  - статистическая значимость или иначе говоря достоверность (эмпирических) результатов.
- ***Репрезентативность (представительность) выборки*** – это способность выборки представлять изучаемые явления достаточно полно.
- Конечно, полное представление об изучаемом явлении, во всем его диапазоне и нюансах изменчивости, может дать только генеральная совокупность. Поэтому репрезентативность всегда ограничена в той мере, в какой

# Репрезентативность выборки

- Приемы, позволяющие получить достаточную репрезентативность выборки:
- **Простой случайный (рандомизированный) отбор.** Он предполагает обеспечение таких условий, чтобы каждый член генеральной совокупности имел равные с другими шансы попасть в выборку. Случайный отбор обеспечивает возможность попадания в выборку самых разных представителей генеральной совокупности. При этом принимаются специальные меры, исключающие появление какой-либо закономерности при отборе. Например, изучая агрессивность подростков, исследователь может *случайным* образом остановить свой выбор на 3 классах разных школ и затем *случайным* образом отобрать по 10 учащихся из каждого класса. Если же исследователь просит испытуемого пригласить на обследование своих друзей, он грубо нарушает принцип случайности отбора.
- **Стратифицированный случайный отбор,** или отбор по свойствам генеральной совокупности. Он предполагает предварительное определение тех качеств, которые могут влиять на изменчивость изучаемого свойства (это может быть пол, уровень дохода или образования и т. д.). Затем определяется процентное соотношение численности различающихся по этим качествам групп (страт) в генеральной совокупности и обеспечивается идентичное процентное соотношение соответствующих групп в выборке. Далее в каждую подгруппу выборки испытуемые подбираются по принципу простого случайного отбора. Например, исследователь может предположить, что мальчики и девочки различаются как по агрессивности, так и по восприимчивости демонстрируемых по телевидению сцен насилия. Если исследователь планирует обобщить результат исследования влияния телевидения на агрессивность всех подростков, то, руководствуясь социально-демографическими данными, он должен

## Статистическая достоверность (значимость)

- **Статистическая достоверность (значимость)** результатов исследования определяется при помощи методов статистического вывода (рассмотрим это далее). Они предъявляют определенные требования к численности, или **объему выборки**.
- Однозначных рекомендаций по предварительному определению требуемого объема выборки не существует. Тем не менее, можно сформулировать наиболее общие рекомендации:
- при разработке диагностической методики необходим наибольший объем выборки – от 200 до 1000-2500 человек;
- при сравнении 2 выборок их общая численность должна быть не менее 50 человек, численность сравниваемых выборок должна быть приблизительно одинаковой;
- при изучении взаимосвязи между какими-либо свойствами, то объем выборки должен быть не меньше 30-35 человек.
- Кроме того, чем больше **изменчивость** изучаемого свойства, тем больше должен быть объем выборки. Поэтому изменчивость можно уменьшить, увеличивая однородность выборки например по полу, возрасту и т.д. При этом

## Зависимые и независимые выборки

- Обычна ситуация исследования, когда интересующее исследователя свойство изучается на двух или более выборках с целью их дальнейшего сравнения. Эти выборки могут находиться в различных соотношениях — в зависимости от процедуры их организации. **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. Наиболее типичный пример зависимых выборок — повторное измерение свойства (свойств) на одной и той же выборке после воздействия (ситуация «до-после»). В этом случае выборки (одна – до, другая – после воздействия) зависимы в максимально возможной степени, так как они включают одних и тех же испытуемых. Могут быть и более слабые варианты зависимости. Например, мужья – одна выборка, их жены – другая выборка (при исследовании, например, их предпочтений). Или: дети 5-7 лет – одна выборка, а их братья или сестры-близнецы – другая выборка.
- В общем случае **зависимые выборки** предполагают попарный подбор испытуемых в сравниваемые выборки, а независимые выборки — независимый отбор испытуемых.

# Обзор классификаций признаков

- *Качественные, количественные.*
- *Метрические, неметрические.*
- *Принадлежность к одной из шкал: Номинативная, порядковая, интервальная, абсолютная.*
- *По роли в статистической совокупности: факторные, результирующие.*
- *Признаки сходства (общие для статистической совокупности признаки), признаки различия (индивидуальные особенности каждой единицы наблюдения).*
- *По выбору в качестве единицы наблюдения случая (например, заболевания) или полицевого учета.*

## Различные шкалы в психологических исследованиях

- В зависимости от того, какая операция лежит в основе измерения признака, выделяют различные типы измерительных шкал. Шкалы разделяют на *метрические* (если может быть установлена единица измерения) и *неметрические* (если единицы измерения не могут быть установлены).
- **Номинативная шкала** или шкала наименований (относится к неметрическим шкалам). Пользуясь определенным правилом, объекты группируются по различным классам так, чтобы внутри класса они были идентичны по измеряемому свойству. Затем каждому классу и объекту дается наименование и обозначение. Примеры номинативных признаков: «пол» (1 – мужской, 0 – женский); «национальность» (1 – русский, 2 – украинец, 3 – белорус); «предпочтение домашних животных» (1 – собаки, 2 – кошки, 3 – крысы, 4 – попугаи, 0 – никакие). В этом случае учитываются только одно свойство чисел – то, что это разные символы. Операции с числами, такие как упорядочивание, сложение-вычитание, умножение-деление, при измерении в номинативной шкале теряют смысл. При сравнении объектов мы можем сделать вывод только о том, принадлежат они к одному или разным классам, тождественны они или нет по

## Различные шкалы в психологических исследованиях

- **Порядковая (ранговая) шкала** (относится к неметрическим шкалам). Измерение в этой шкале предполагает приписывание объектам чисел в зависимости от степени выраженности измеряемого свойства. Примеры признаков, выраженной в порядковой шкале: «место в рейтинге» (от первого до последнего); «оценка за экзамен» (от 2 до 5); и большинство иных измерений в психологических исследованиях (где не проведена стандартизация и обоснование процедуры равноинтервальности)! При сравнении испытуемых друг с другом мы можем сказать, больше или меньше выражено свойство, но не можем сказать, насколько или во сколько раз больше или меньше оно выражено. При измерении в порядковой шкале из всех свойств чисел учитывается лишь то, что они разные и то, что одно число больше другого. Например, если шкалой является распределение результатов участников гонки по местам от первого до последнего, то это вовсе не значит, что они достигали финиша через равные интервалы времени.

## Различные шкалы в психологических исследованиях

- **Интервальная шкала** (относится к метрическим шкалам). При таком измерении числа отражают не только различия между объектами в уровне выраженности свойства (как в порядковой шкале), но и то, насколько больше или меньше выражено свойство. Равным разностям между числами в этой шкале соответствуют равные разности в уровне выраженности измеренного свойства, т.е. измерение в этой шкале предполагает возможность применения *единицы измерения (метрики)*. Объекту присваивается число единиц измерения, пропорциональное степени выраженности измеряемого свойства. При этом нулевая точка не соответствует полному отсутствию измеряемого свойства. При сравнении двух объектов мы можем судить, насколько больше или меньше выражено свойство, но не можем судить о том, во сколько раз больше или меньше выражено свойство. Примеры измерения в интервальной шкале: температурная шкала Цельсия (где точка нуля – не полное отсутствие температуры, а всего лишь температура замерзания воды); многие тестовые шкалы, вводимые при **обосновании равноинтервальности**, как шкала IQ Векслера (ноль по этой шкале вовсе не соответствует полному отсутствию интеллекта как такового), некоторые виды семантического

## Различные шкалы в психологических исследованиях

- **Шкала отношений или абсолютная шкала** (относится к метрическим шкалам). Отличается от интервальной только тем, что в ней устанавливается нулевая точка, соответствующая полному отсутствию измеряемого свойства. Примеры измерений в шкале отношений: измерение времени реакции, роста, веса, температуры по абсолютной шкале (где «0» означает совершенное отсутствие температуры) и т.п. Сравнивая результаты, измеренные в этой шкале, между собой, можно сказать не только, насколько больше или меньше выражено данное свойство, но и во сколько раз.
- Вышеперечисленные шкалы отличаются по их мощности – сколько информации о различии объектов можно получить при помощи признака, выраженного в разных шкалах. По мере возрастания мощности в этом отношении данные шкалы располагаются следующим образом: номинативная, порядковая, интервальная, шкала отношений. Очевидно, наиболее тонкие различия между объектами можно выявить с помощью метрических шкал, неметрические шкалы дают более грубые интерпретации при сравнении объектов.

## Различные шкалы в психологических исследованиях

- **Определение того, в какой шкале измерено явление – первостепенный момент анализа данных:** любой последующий шаг, выбор метода зависит именно от этого. Обычно идентификация номинативной шкалы от других шкал не представляет трудностей. Сложнее определить различие между порядковой и интервальной шкалами. В психологии часто по степени выраженности непосредственно измеряемой величины (например, количество правильных ответов на вопросы) выносят суждения о некотором скрытом свойстве (например, интеллекте), недоступном прямому наблюдению. Таким образом, большинство измерений в психологии являются косвенными. Например, количество правильных ответов на вопросы – это измерение в метрической шкале, но соответствуют ли равные разности количества правильных ответов равным разностям выраженности интеллекта? Если это так, то шкала интервальная и метрическая, если нет – то шкала порядковая и неметрическая. В ряде случаев имеет смысл обосновывать метричность (равноинтервальность) шкалы для того, чтобы иметь возможность использовать более мощные инструменты анализа данных.

## **Некоторые элементарные типы задач психологического исследования (с точки зрения статистического метода)**

- 1. Выявление различий в уровне исследуемого признака.** Сопоставляются различные группы испытуемых по какому-то признаку, чтобы выявить различия между ними по этому признаку (*например, сопоставляются показатели вербального интеллекта у студентов-психологов и студентов-физиков*).
- 2. Оценка сдвига значений исследуемого признака.** Чаще всего у одной и той же группы испытуемых сопоставляются уровни признака «до» и «после» экспериментальных или иных воздействий (*или же по прошествии определенного времени, или в разных условиях, например, обычных и экстремальных*), чтобы определить эффективность этих влияний (*например, сопоставляются уровни самооценки*

## **Некоторые элементарные типы задач психологического исследования (с точки зрения статистического метода)**

### **3. Выявление различий в распределении признака.**

Сопоставляется эмпирическое распределение значений признака с каким-либо теоретическим законом распределения или два эмпирических распределения между собой (например, отличается ли распределение показателей показателя тревожности от равномерного по дням недели; отличается ли соотношение успевающих и неуспевающих школьников в зависимости от того, полная ли у них семья, и т.д.).

### **4. Выявление степени согласованности изменений нескольких признаков или профилей.**

Могут быть сопоставлены два признака, измеренные на одной и той же выборке испытуемых для того, чтобы установить степень согласованности их изменений (корреляцию) между ними (например, выясняется наличие связи уровней социального интеллекта и тревожности у студентов-психологов).

Следует отметить, что наличие различий или корреляционной связи вовсе не означает автоматически наличия причинно-следственных связей.

# Упражнения

Определите, в какой шкале представлено каждое из приведенных ниже измерений: номинативной, порядковой, интервальной, отношений? Обоснуйте ответ.

1. порядковый номер испытуемого в списке (для его идентификации);
2. количество вопросов в анкете как мера трудоемкости опроса;
3. упорядочивание испытуемых по времени решения тестовой задачи;
4. академический статус (ассистент, доцент, профессор) как указание на принадлежность к соответствующей категории;
5. академический статус (ассистент, доцент, профессор) как мера продвижения по службе;
6. телефонные номера;
7. время решения задачи;
8. количество агрессивных реакций за рабочий день;
9. количество агрессивных реакций за рабочий день как показатель агрессивности;
0. цвет глаз;
1. числа, кодирующие темпераменты;
2. метрическая система измерения расстояний.

## Упражнения

- Определите, к какому типу задач на сопоставление следует отнести нижеперечисленные задачи и почему?
1. Установить эффективность лечения депрессии, сравнивая ее показатели до и после применения определенной терапии в группе испытуемых;
  2. Определить характер связи между агрессивностью и тревожностью у группы подростков;
  3. Как отличаются студенты-физики от студентов-психологов по уровню вербального интеллекта?
  4. Как отличаются между собой по уровню тревожности дети из полных и неполных семей?
  5. Различны ли показатели настроения у студентов до и после экзаменационной сессии?
  6. Существует ли связь между ростом человека и его заработной платой?
  7. Достигнуть вершины можно по нескольким маршрутам. Существуют ли предпочтения относительно выбора какого-либо из путей?
  8. Равномерно ли распределяются частоты обращений в службу психологической помощи по разным дням недели?
  9. Зависят ли показатели воспроизведения слов испытуемыми, которые предъявлялись им на слух, от скорости их предъявления?

Научно-исследовательская работа: помощь, участие, направления квалификационных (курсовых, дипломных) работ, выполняемых под руководством профессора Григорьева Павла Евгеньевича и доцента Васильевой Инны Витальевны

- Исследования различных аспектов интуиции у представителей различных групп (возраст, пол, потребности, особенности физиологического статуса и проч.) с помощью психологических, программно-аппаратных методов и методик
- Психофизиологические исследования функционального состояния представителей различных групп в зависимости от нагрузок, ситуации, деятельности
- Разработка новых методов улучшения эффективности деятельности представителей опасных и связанных с повышенной ответственностью профессий в условиях дефицита времени и/или информации

Вступайте в группу «Intuition & Intention» [www.vk.com/int.research](http://www.vk.com/int.research)

Принимайте участие в обсуждениях, играйте в игру «Мир магии», а также других исследованиях интуиции (18+)

# Вопросы для проработки и самостоятельного изучения

1. Понятие измерения.
2. Виды измерительных шкал и свойства психологических объектов измерения.
3. Номинативная шкала как способ классификации или распределения объектов.
4. Порядковая шкала как способ расположения измеряемых признаков по рангу. Правила ранжирования.
5. Шкала интервалов и её свойства. Распределение значений по принципу: «больше на определенное количество единиц – меньше на определенное количество единиц».
6. Шкала (равных) отношений, ее особенности. Наличие фиксированного нуля.
7. Понятие генеральной совокупности. Понятие выборки как подгруппы элементов (испытуемых), выделенной из генеральной совокупности для проведения эксперимента.
8. Объем выборки. Полное (сплошное) и выборочное исследование. Зависимые и независимые выборки.
9. Требования к выборке при решении различных задач.
0. Репрезентативность выборки. Формирование и объем

## Рекомендованная для закрепления материала лекций литература

- Р. Майкл Фер, Верн Р. Бакарак. Психометрика: Введение; пер. с англ... Челябинск: Издательский центр ЮУрГУ, 2010. 445 с. <https://yadi.sk/d/rXZGxdvxuxB6n>
- Лакин Г.Ф. Биометрия. М.: Высшая школа, 1990. 352 с. <https://yadi.sk/d/gOx4ndnSuxBHB>
- Сидоренко Е.В. Методы математической обработки в психологии. СПб: Речь, 2006. 348 с. <https://yadi.sk/i/0SV19sefuxBY9>
- Остапенко Р.И. Математические основы психологии. Учебно-методическое пособие. Воронеж: ВГПУ, 2010. 76 с. <https://yadi.sk/i/LoZ57-O6uxDfg>
- Червинская К.Р. Компьютерная психодиагностика. СПб: Речь, 2003. 336 с. [https://yadi.sk/d/8ln7\\_Hk9uzxLA](https://yadi.sk/d/8ln7_Hk9uzxLA)

# Способы представления исходных данных

- Хотя существуют различные способы представления исходных данных (табличный, графический, аналитический) в математической статистике обычно используют табличный способ представления исходных данных.

Пример табличного представления данных.

	1	2	3	4
№	пол	время простой слухо-моторной реакции (мс)	Уровень общего интеллекта IQ	сила нервной системы (1-5)
1	женский	221.51	107	3
2	мужской	215.52	109	4
3	мужской	223.12	128	4
4	женский	161.46	111	5
5	мужской	130.13	111	1
6	мужской	155.15	105	2
7	мужской	166.27	109	2
8	женский	220.98	111	3
9	женский	201.42	120	3
10	мужской	148.32	108	5
<i>i</i>	$X_{i1}$	$X_{i2}$	$X_{i3}$	$X_{i4}$

# Вариационный ряд, частоты

Как правило, строками таблиц данных психологического исследования являются наборы признаков (переменных), характеризующих определенного испытуемого, а столбцами – переменные (признаки). Очевидно, пол является номинативной переменной, уровень общего интеллекта – интервальной, а сила нервной системы – порядковой.

Анализ данных начинается с изучения того, как часто встречаются те или иные значения интересующего исследователя признака в исходных данных. Для этого данные ранжируются, строятся таблицы и графики распределения абсолютных, относительных, накопленных частот.

**Вариационный ряд.** Пусть из генеральной совокупности извлечена выборка, причем значение  $x_1$  наблюдалось  $n_1$  раз, значение  $x_2$  –  $n_2$  раз,  $x_k$  –  $n_k$  раз. В данном случае  $n_1x_1$

$+ n_2x_2 + \dots + n_kx_k = \sum_{i=1}^k n_i x_k$ ,  $k$  – количество различных значений признака  $x$  выборки.

Отметим, что объем выборки  $n \geq k$ , т.к. некоторые значения  $x$  встречаются более одного

раза. Знак  $\sum_{i=1}^k n_i x_k$  обозначает, что производится суммирование всех (различных) значений  $x_i$ ,

имеющие порядковые номера, умноженных на их частоты  $x_i$ . Наблюдаемые различные значения  $x_i$  называются **вариантами**, а последовательность вариантов, проранжированных в возрастающем порядке, сопоставленных с их частотами – **вариационным рядом**.

Числа наблюдений ( $n_1, n_2, \dots, n_k$ ) называются **частотами**, а их отношение к объему

выборки  $\frac{n_i}{n}$  – **относительными частотами**.

# Вариационный ряд, ранжирование

- Предположим, что исследователя в нашем примере интересует распределение уровня интеллекта учащихся. Для этого исходный ряд упорядочивается от максимального до минимального значения или наоборот. При этом большему значению может быть приписан больший или меньший ранг (место по порядку). Так, в таблице на следующем слайде ряд значений IQ расположен в убывающем порядке, наибольшему значению приписан ранг 1. Если несколько значений одинаковы, они имеют одинаковый ранг, равный среднему арифметическому тех рангов, которые они получили бы, если бы не были равны.

Абсолютная и относительная частота связаны соотношением  $f_{отн} = \frac{f_{абс}}{N}$ , где  $f_{абс}$  -

абсолютная частота определенного значения признака,  $f_{отн}$  - относительная частота этого значения признака,  $N$  - число наблюдений (в данном случае 10). Сумма всех абсолютных частот равна числу наблюдений  $N$ , а сумма всех относительных частот равна 1. Относительная частота может служить оценкой вероятности встречаемости данного значения.

## Вариационный ряд, ранжирование

<b>№</b>	<b>IQ</b>	<b>Ранг</b>	<b>Абсолютная частота <math>f_{\text{абс}}</math></b>	<b>Относительная частота <math>f_{\text{отн}}</math></b>	<b>Накопленная частота <math>f_{\text{нак}}</math></b>
<b>1</b>	<b>128</b>	<b>1</b>	<b>1</b>	<b>1/10=0.1</b>	<b>0.1</b>
<b>2</b>	<b>120</b>	<b>2</b>	<b>1</b>	<b>0.1</b>	<b>0.2</b>
<b>3</b>	<b>111</b>	$\frac{3+4+5}{3} = 4$	<b>3</b>	<b>3/10=0.3</b>	<b>0.5</b>
<b>4</b>	<b>111</b>	<b>4</b>			
<b>5</b>	<b>111</b>	<b>4</b>			
<b>6</b>	<b>109</b>	$\frac{6+7}{2} = 6.5$	<b>2</b>	<b>2/10=0.2</b>	<b>0.7</b>
<b>7</b>	<b>109</b>	<b>6.5</b>			
<b>8</b>	<b>108</b>	<b>8</b>	<b>1</b>	<b>0.1</b>	<b>0.8</b>
<b>9</b>	<b>107</b>	<b>9</b>	<b>1</b>	<b>0.1</b>	<b>0.9</b>
<b>10</b>	<b>105</b>	<b>10</b>	<b>1</b>	<b>0.1</b>	<b>1.0</b>

# Таблицы распределения накопленных частот

Еще одной разновидностью таблиц распределения являются таблицы **распределения накопленных частот** : напротив каждого значения (интервала) указывается сумма частот встречаемости наблюдений, накопленная от первого по порядку по данный интервал.

Во многих случаях признак может принимать множество различных значений, например, если измеряется время решения задачи в секундах у значительного числа испытуемых. В этом случае о распределении признака позволяет судить **таблица сгруппированных частот** , в которых частоты группируются по разрядам. Пусть в группе испытуемых 50 человек измерено время решения тестовой задачи. Максимальное время составило  $t_{\max}=67$  секунд, минимальное –  $t_{\min}=32$  секунды. Построение таблицы сгруппированных частот в этом случае производится поэтапно.

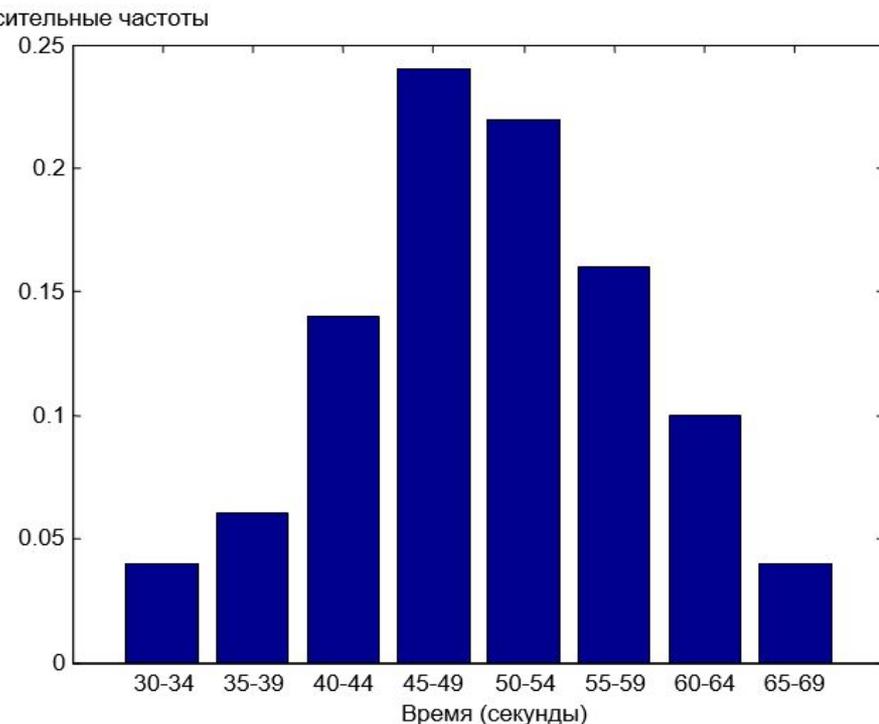
- 1) Определение **размаха**:  $R = t_{\max} - t_{\min} = 35$ .
- 2) Выбор желаемого количества разрядов и интервала разрядов. Определяется произвольно, как правило, от 6 до 20. В нашем случае удобно взять интервал разрядов взять 5, тогда количество разрядов равно  $35:5=7$ . Учитывая, что начинать лучше с 30 или 31 и заканчивать на 69 или 70, уточняем размах ( $70 - 30 = 40$ ) и число разрядов ( $40:5 = 8$ ).
- 3) Определение границ разрядов. Если начинаем с 30, то первый разряд будет с 30 до 34, второй – с 35 до 39 и т.д., до восьмого – с 65 до 69. Границы соседних разрядов не должны совпадать.
- 4) Подсчет частот встречаемости значений признака для каждого интервала.

# Гистограмма

Группировка частот по разрядам (интервалам) измеренного признака.

Интервал времени, с	$f_{\text{абс}}$	$f_{\text{отн}}$	$f_{\text{нак}}$
30-34	2	0,04	0,04
35-39	3	0,06	0,10
40-44	7	0,14	0,24
45-49	12	0,24	0,48
50-54	11	0,22	0,70
55-59	8	0,16	0,86
60-64	5	0,10	0,96
65-69	2	0,04	1
$\Sigma$ (сумма)	50	1,000	–

Для наглядного представления строится столбиковая диаграмма распределения относительных либо абсолютных частот – *гист ограмма*. Для приведенного выше примера гистограмма относительных частот представлена на рис.



. Гистограмма относительных частот времени решения задачи по данным

# Описательные статистики

К первичным описательным статистикам относят числовые характеристики распределения измеренного на выборке признака. Каждая такая характеристика отражает в одном числовом значении свойства распределения множества результатов измерения: с точки зрения их расположения на числовой оси либо с точки зрения их изменчивости. Основное назначение каждой из первичных описательных статистик – замена множества значений признака, измеренного на выборке, одним числом. Компактное описание группы при помощи описательных статистик позволяет интерпретировать результаты измерений путем сопоставления первичных статистик разных групп или разных серий измерений одной группы. Среди первичных статистик выделяют **меры центральной тенденции, меры положения, меры изменчивости**.

**Меры центральной тенденции** используются для сравнения групп по уровню выраженности признака.

**Мера центральной тенденции** – это число, характеризующее выборку по уровню выраженности измеренного признака.

Основные меры центральной тенденции – **мода, медиана, выборочное среднее**.

## Меры центральной тенденции

**Модой** называется наиболее часто встречающееся значение признака. Например, в ранжированном ряде наблюдений 1, 1, 1, 2, 3, 4, 5, 6 модой является 1. Ряд может не иметь моды, если ни одно из значений не встречается чаще другого, или иметь несколько мод, если существуют несколько значений, встречающихся чаще остальных.

**Медианой** называется значение признака, которое делит ранжированный ряд пополам. Если ряд имеет нечетное количество наблюдений, то медиана есть центральное значение. В нашем ряде 1, 1, 1, 2, 3, 4, 5, 6 четное количество наблюдений, тогда медианой будет среднее двух центральных значений  $Md = (2+3)/2 = 2,5$

**Выборочным средним** является сумма всех значений измеренного признака ( $x_i$ ), деленная на количество значений ( $n$ ): 
$$M_x = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i$$

Целесообразность использования той или иной меры центральной тенденции зависит от параметров конкретного исследования и числовых признаков.

Так, для номинативных данных единственной мерой центральной тенденции может служить мода.

# Меры центральной тенденции

У порядковых и метрических переменных, распределение частот признака у которых симметрично и имеет одну моду, – медиана, мода и среднее совпадают. Однако, при нарушении симметричности и наличии выбросов (экстремально малых или больших значениях переменной), эти равенства нарушаются. Среднее чувствительно к выбросам, а мода и медиана – нет. С другой стороны, использование средней может дать наибольшую информацию, поскольку среднее зависит от каждого индивидуального значения.

Средние можно использовать для сравнений уровней признака в различных группах, если:

- 1) группы достаточно большие, чтобы судить о форме распределения;
- 2) распределения симметричны;
- 3) отсутствуют выбросы.

В противном случае необходимо использовать моду, медиану, или сравнивать средние, вычисленные для рангов этих групп.

## Меры положения

В психологии также широко используются **меры положения**, которые называются квантилями распределения. **Квант иль** – это точка на числовой оси, которая делит всю совокупность упорядоченных измерений на две группы с известным соотношением их численности. Например, медиана является квантилем, который делит совокупность измерений на две равные части. Кроме медианы используют процентиля и квартили.

**Процент или** – это 99 точек – значений признака ( $P_1, P_2, \dots, P_{99}$ ), которые делят упорядоченное по возрастанию множество наблюдений на 100 частей, равных по численности. Для определения конкретного процентиля, например,  $P_{10}$ , все значения признака упорядочиваются по возрастанию, затем отсчитываются 10% значений со стороны меньших значений.  $P_{10}$  соответствует значению признака, который отделяет эти 10% испытуемых от остальных 90%.

**Кварт или** – это 3 точки – значения признака ( $P_{25}, P_{50}, P_{75}$ ), которые делят упорядоченное по возрастанию множество наблюдений на 4 равные по численности части. Первый квартиль соответствует 25-му процентилю, второй – 50-му процентилю или медиане, третий квартиль соответствует 75-му процентилю.

Процентиля и квартили используются для определения частоты встречаемости тех или иных значений (или интервалов) измеренного признака или для выделения подгрупп и отдельных испытуемых, наиболее типичных или нетипичных для данного множества наблюдений.

## Меры изменчивости

Наряду с мерами центральной тенденции и мерами положения используются и **меры изменчивости**, которые характеризуют разброс значений признака в данной выборке.

Наиболее простой мерой изменчивости является **размах**: разность максимального и минимального значений:  $R_x = x_{\max} - x_{\min}$

Разброс является малоинформативной мерой изменчивости, поскольку на его значение влияют лишь крайние значения, которые могут сильно отличаться от большинства других значений (быть «выбросами»). Более устойчивыми являются разновидности размаха, такие как **размах от 10-го до 90-го процента** или **( $P_{90}-P_{10}$ )**, или **междуквартильный размах ( $P_{75}-P_{25}$ )**. Эти меры изменчивости используются в основном для порядковых данных. Для метрических данных чаще используется **дисперсия**.

**Дисперсия** – мера изменчивости для метрических данных, пропорциональная сумме квадратов отклонений измеренных значений от их арифметического среднего:

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

**Стандартное отклонение** – значение квадратного корня из дисперсии:  $\sigma = \sqrt{D}$ .

## Меры изменчивости, стандартизация

На практике в качестве меры изменчивости чаще используют стандартное отклонение, а не дисперсию, поскольку стандартное отклонение, в отличие от дисперсии, выражает изменчивость в исходных единицах измерения признака, а не их квадратах.

При объединении двух групп в одну к внутригрупповой дисперсии каждой группы добавляется дисперсия, обусловленная различием между группами. Чем больше различие между средними значениями признака в двух группах, тем больше увеличивается дисперсия объединенных групп. Например, одна группа содержит значения 1,1,1,1,1, а другая группа 3,3,3,3,3. Дисперсии этих групп одинаковы и равны 0. Если же объединить эти две группы, то дисперсия будет не 0, а 1.

**Стандартная ошибка** – корень квадратный из отношения дисперсии и количества

наблюдений:  $m = \sqrt{\frac{D}{n}}$

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

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

# Стандартизированные шкалы, асимметрия

В результате преобразованные значения (z-значения) непосредственно выражаются в единицах стандартного отклонения от среднего. Если для одной выборки несколько признаков переведены в z-значения, то появляется возможность сравнения уровня выраженности разных признаков у того или иного испытуемого. Для того, чтобы избавиться от неизбежных отрицательных и дробных значений, можно перейти к любой другой известной шкале: IQ (среднее 100, стандартное отклонение 15); Т-оценок (среднее 50, стандартное отклонение 10), 10-балльной – стенов (среднее 5,5, стандартное отклонение 2) и т.д. Перевод в новую шкалу из z-значений осуществляется путем умножения каждого z-значения на заданное стандартное отклонение и прибавление среднего:  $S_i = \sigma_s z_i + M_s$

Асимметрия – степень отклонения графика распределения частот от симметричного вида относительно среднего значения (рис.).

Если исходные данные переведены в z-значения, показатель асимметрии вычисляется по

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

# Асимметрия

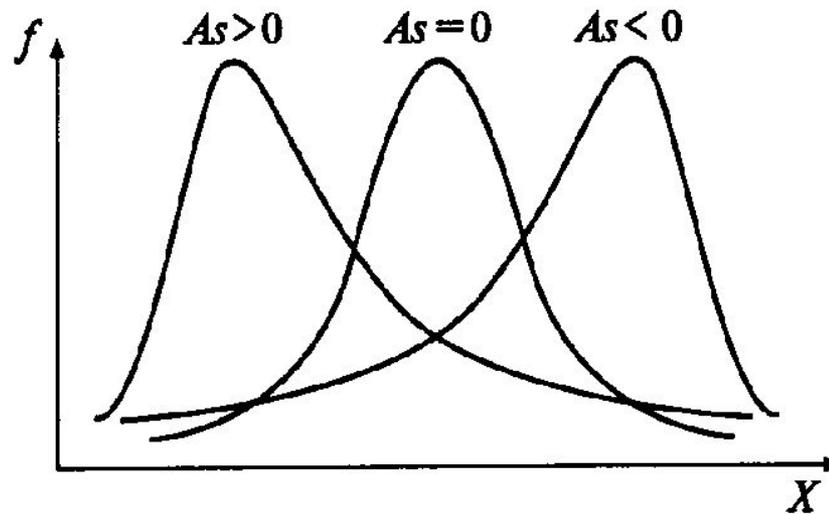


Рис. Распределения с различными значениями асимметрии.

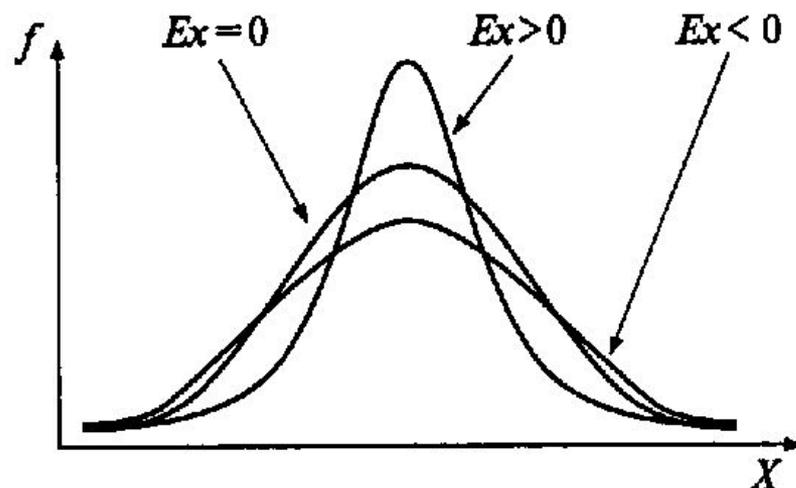
Для симметричного распределения асимметрия равна 0. Если чаще встречаются значения меньше среднего, то говорят о левосторонней или положительной асимметрии ( $As > 0$ ). Если же чаще встречаются значения больше среднего, то асимметрия — правосторонняя или отрицательная ( $As < 0$ ).

# Эксцесс

**Эксцесс** – мера плосковершинности или остроконечности графика распределения измеренного признака (рис.). Вычисляется по формуле:

Если исходные данные переведены в z-значения, показатель эксцесса определяется

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



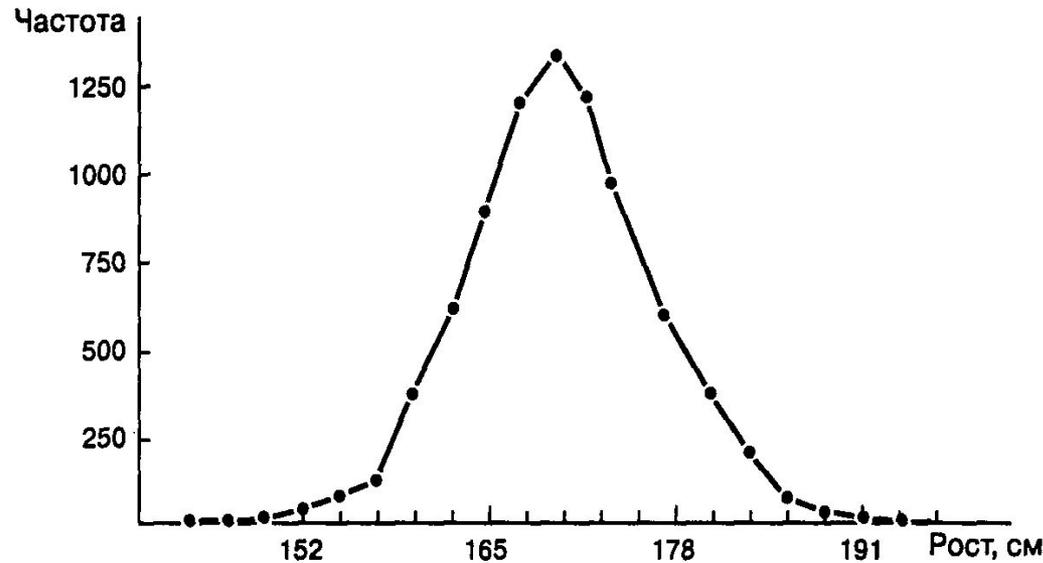
**Рис. Распределения с различными значениями эксцесса.**

Острове́ршинное распределение характеризуется положительным эксцессом, а плосковершинное – отрицательным. Средневершинное распределение имеет нулевой эксцесс.

# НОРМАЛЬНЫЙ ЗАКОН РАСПРЕДЕЛЕНИЯ И ЕГО ПРИМЕНЕНИЕ В ПСИХОЛОГИИ

- Нормальный закон распределения играет важнейшую роль в применении численных методов в психологии. Он лежит в основе измерений, разработки тестовых шкал, методов проверки гипотез. История применения закона нормального распределения в социальных и биологических науках начинается с работы бельгийского ученого А.Кетле «Опыт социальной физики» (1835 г.). В ней он доказывал, что такие явления, как продолжительность жизни, возраст вступления в брак и появления первого ребенка и т.д., подчиняются строгой закономерности. Она проявляется в том, что чаще всего встречаются средние значения соответствующих показателей, и чем больше отклонение от этой средней величины, тем реже встречаемость таких отклонений. Одинаковые отклонения от среднего в меньшую и в большую сторону встречаются одинаково редко, чем среднее значение. В его исследованиях, и позднее — в исследованиях Ф. Гальтона, было доказано, что распределение частот встречаемости любого демографического (продолжительность жизни и пр.) или антропометрического (рост, вес и пр.) показателя, измеренного на большой выборке людей, имеет одну и ту же «колоколообразную» форму.

# НОРМАЛЬНЫЙ ЗАКОН РАСПРЕДЕЛЕНИЯ И ЕГО ПРИМЕНЕНИЕ В ПСИХОЛОГИИ



Полигон частот для роста 8585 взрослых людей, родившихся в Англии в XIX в.

Форма таких распределений может быть описана математической формулой, которую предложил в XVIII веке математик де Муавр:

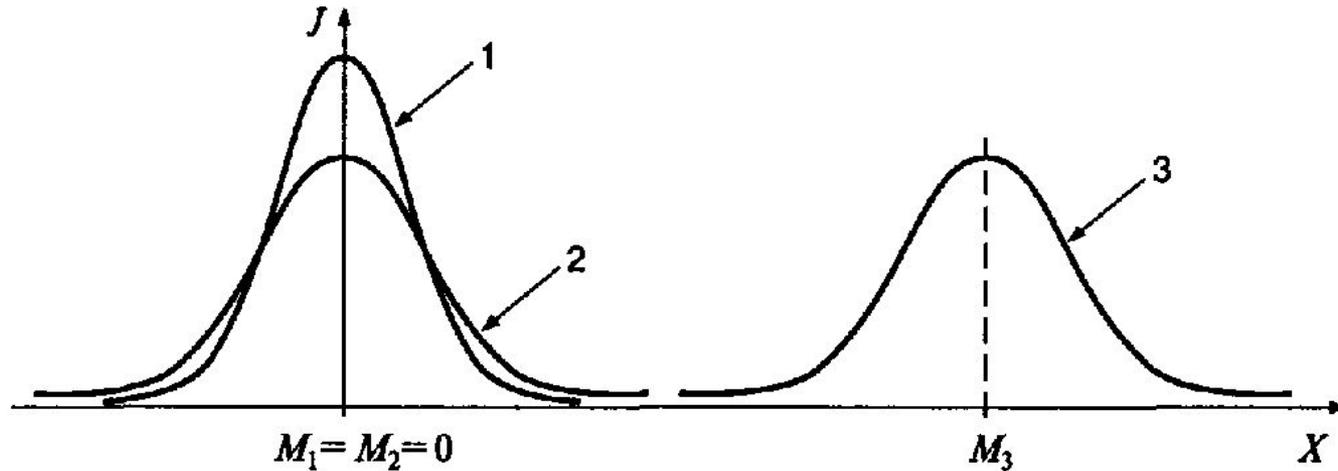
$$f(x_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x_i-M)^2/2\sigma^2}, \text{ где}$$

$f(x_i)$  – высота подъема кривой,  $e$  – основание натурального логарифма (2,718),  $M$  и  $\sigma$  – среднее арифметическое и стандартное отклонение для переменной  $x$ , которые определяют положение кривой на числовой оси и задают ее размах. Эта формула и кривая получили название закона нормального распределения случайной величины.

# НОРМАЛЬНЫЙ ЗАКОН РАСПРЕДЕЛЕНИЯ И ЕГО ПРИМЕНЕНИЕ В ПСИХОЛОГИИ

- В дальнейшем трудами Ф. Гальтона и его последователей было доказано, что и психологические особенности, например, способности, подчиняются нормальному закону. Поэтому дальнейшее развитие измерительного подхода в психологии и статистического аппарата проверки гипотез происходило на базе этого общего закона. Начиная со второй половины XIX столетия измерительные и вычислительные методы в психологии разрабатываются на основе следующего принципа. ***Если индивидуальная изменчивость некоторого свойства есть следствие действия множества причин, то распределение частот для всего многообразия проявлений этого свойства в генеральной совокупности соответствует кривой нормального распределения. Это и есть закон нормального распределения.***
- **Нормальное распределение как стандарт.** Каждому психологическому (или шире – биосоциальному) свойству соответствует свое распределение в генеральной совокупности. Чаще всего оно является нормальным и характеризуется своими параметрами: средним ( $M$ ) и стандартным отклонением ( $\sigma$ ). ***Только эти два значения полностью определяют форму кривой нормального распределения.*** Среднее задает положение кривой на числовой оси и выступает как некоторая исходная, нормативная величина измерения. Стандартное отклонение задает ширину этой кривой, зависит от единиц измерения и выступает как масштаб измерения.

# НОРМАЛЬНЫЙ ЗАКОН РАСПРЕДЕЛЕНИЯ И ЕГО ПРИМЕНЕНИЕ В ПСИХОЛОГИИ

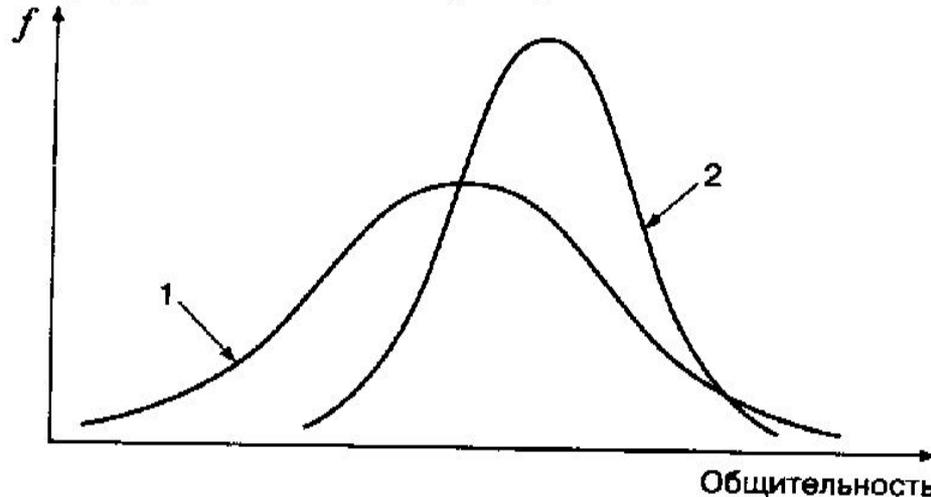


Семейство нормальных кривых, 1-е распределение отличается от 2-го стандартным отклонением ( $\sigma_1 < \sigma_2$ ), 2-е от 3-го средним арифметическим ( $M_2 < M_3$ ).

Все многообразие нормальных распределений может быть сведено к одной кривой, если применить z-преобразование ко всем возможным измерениям свойств. Тогда каждое свойство будет иметь среднее 0 и стандартное отклонение 1. Нормальное распределение с  $M=0$  и  $\sigma=1$  — единичное нормальное распределение, которое используется как стандарт — эталон.

## Упражнение

1. По результатам измерения общительности у юношей (1) и девушек (2) были построены сглаженные графики распределения частот (рис.).



Определите по графику: а) как различаются средние  $M_1$  и  $M_2$ ; б) как различаются дисперсии  $D_1$  и  $D_2$ ?

# **НОРМАЛЬНЫЙ ЗАКОН РАСПРЕДЕЛЕНИЯ И ЕГО ПРИМЕНЕНИЕ В ПСИХОЛОГИИ**

***Рассмотрим свойства нормального распределения.***

- 1) Единицей измерения единичного нормального распределения является стандартное отклонение.***
- 2) Кривая приближается к оси  $Z$  по краям асимптотически – никогда не касаясь ее.***
- 3) Кривая симметрична относительно  $M=0$ . Ее асимметрия и эксцесс равны нулю.***
- 4) Кривая имеет характерный изгиб: точка перегиба лежит точно на расстоянии в одну  $\sigma$  от  $M$ .***
- 5) Площадь между кривой и осью  $Z$  равна 1.***

Последнее свойство объясняет название «единичное нормальное распределение» и имеет исключительно важное значение. Благодаря этому свойству площадь под кривой интерпретируется как вероятность, или относительная частота. Действительно, вся площадь под кривой соответствует вероятности того, что признак примет любое значение из всего диапазона его изменчивости (от  $-\infty$  до  $+\infty$ ). Площадь под единичной нормальной кривой слева или справа от нулевой точки равна 0,5. Это соответствует тому, что половина генеральной совокупности имеет значение признака больше 0, а половина — меньше 0. Относительная частота встречаемости в генеральной совокупности значений признака в диапазоне от  $z_1$  до  $z_2$  равна площади под кривой, лежащей между соответствующими точками. Отметим еще раз, что любое нормальное распределение может быть сведено к единичному нормальному распределению путем  $z$ -преобразования.

# НОРМАЛЬНЫЙ ЗАКОН РАСПРЕДЕЛЕНИЯ И ЕГО ПРИМЕНЕНИЕ В ПСИХОЛОГИИ

Таким образом:

- 1) если  $x_i$  имеет нормальное распределение со средним  $M$  и стандартным отклонением  $\sigma$ , то  $z = (x - M_x)/\sigma$  характеризуется единичным нормальным распределением со средним 0 и стандартным отклонением 1;
- 2) площадь между  $x_1$  и  $x_2$  в нормальном распределении со средним  $M_x$  и стандартным отклонением  $\sigma$  равна площади между  $z_1=(x_1-M_x)/\sigma$  и  $z_2=(x_2-M_x)/\sigma$  в единичном нормальном распределении.

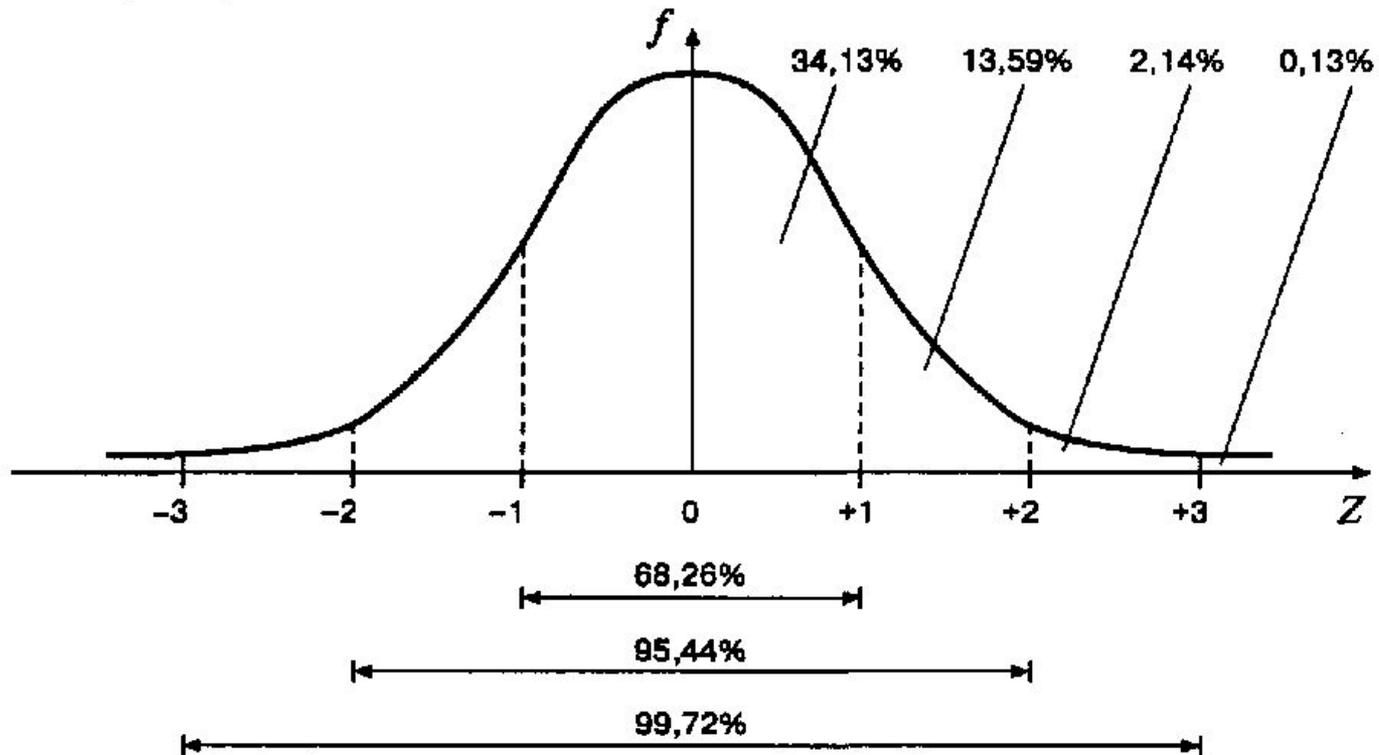


Рис. Стандартное нормальное распределение.

# НОРМАЛЬНЫЙ ЗАКОН РАСПРЕДЕЛЕНИЯ И ЕГО ПРИМЕНЕНИЕ В ПСИХОЛОГИИ

Полезно помнить, что для любого нормального распределения существуют следующие соответствия между диапазонами значений и площадью под кривой:

- $M \pm \sigma$  соответствует 68,26% площади;
- $M \pm 2\sigma$  соответствует 95,44% площади;
- $M \pm 3\sigma$  соответствует 99,72% площади.
- Если распределение является нормальным, то:
- 90% всех случаев располагается в диапазоне значений  $M \pm 1,64\sigma$ ;
- 95% всех случаев располагается в диапазоне значений  $M \pm 1,96\sigma$ ;
- 99% всех случаев располагается в диапазоне значений  $M \pm 2,58\sigma$ .
- Единичное нормальное распределение устанавливает четкую взаимосвязь стандартного отклонения и относительного количества случаев в генеральной совокупности для любого нормального распределения.
- Существуют правила перевода (специальная таблица, например), позволяющая определять площадь под кривой справа от любого положительного  $z$ . Пользуясь ею, можно определить вероятность встречаемости значений признака из

## НОРМАЛЬНЫЙ ЗАКОН РАСПРЕДЕЛЕНИЯ И ЕГО ПРИМЕНЕНИЕ В ПСИХОЛОГИИ

Несмотря на исходный постулат, в соответствии с которым свойства в генеральной совокупности имеют нормальное распределение, реальные данные, полученные на выборке, не всегда распределены нормально. Более того, разработано множество методов, позволяющих анализировать данные без всякого предположения о характере их распределения как в выборке, так и в генеральной совокупности. Тем не менее, существуют по крайней мере три важных аспекта применения нормального распределения в психологии:

1. Разработка тестовых шкал.
2. Проверка нормальности выборочного распределения для принятия решения о том, в какой шкале измерен признак – в метрической или порядковой.
3. Статистическая проверка гипотез, в частности – при определении риска принятия неверного решения

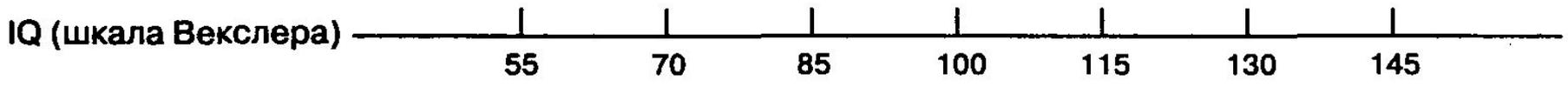
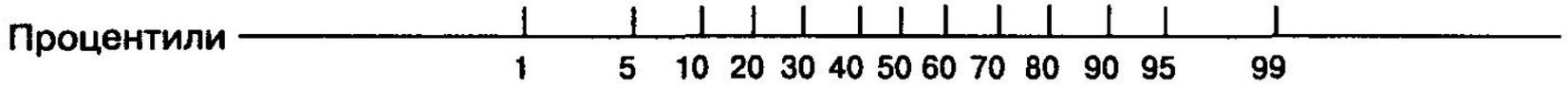
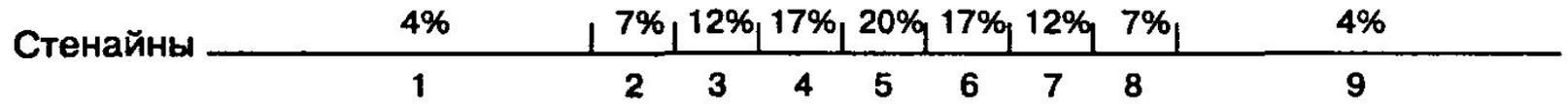
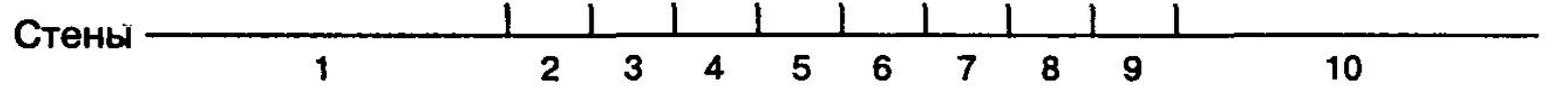
# Разработка тестовых шкал

- **Разработка тестовых шкал.** Тестовые шкалы разрабатываются для того, чтобы оценить индивидуальный результат тестирования путем сопоставления его с тестовыми нормами, полученными на выборке стандартизации. Выборка стандартизации специально формируется для разработки тестовой шкалы — она должна быть репрезентативна генеральной совокупности, для которой планируется применять данный тест. Впоследствии при тестировании предполагается, что и тестируемый, и выборка стандартизации принадлежат одной и той же генеральной совокупности.
- Исходным принципом при разработке тестовой шкалы является предположение о том, что измеряемое свойство распределено в генеральной совокупности в соответствии с нормальным законом. Соответственно, измерение в тестовой шкале данного свойства на выборке стандартизации также должно обеспечивать нормальное распределение. Если это так, то тестовая шкала является метрической — точнее, равных интервалов. Если это не так, то свойство удалось отразить в лучшем случае — в шкале порядка. Естественно, что большинство стандартных тестовых шкал являются метрическими, что позволяет более детально интерпретировать результаты тестирования — с учетом свойств нормального распределения — и корректно применять любые методы статистического анализа. Таким образом, основная проблема стандартизации теста заключается в разработке такой шкалы, в которой распределение тестовых показателей на выборке стандартизации соответствовало бы нормальному распределению.

## Разработка тестовых шкал

- Исходные тестовые оценки — это количество ответов на те или иные вопросы теста, время или количество решенных задач и т. д. Они еще называются первичными, или «сырыми» оценками. Итогом стандартизации являются тестовые нормы – таблица пересчета «сырых» оценок в стандартные тестовые шкалы. Существует множество стандартных тестовых шкал, основное назначение которых — представление индивидуальных результатов тестирования в удобном для интерпретации виде. Некоторые из этих шкал представлены на рис.
- Общим для них является соответствие нормальному распределению, а различаются они только двумя показателями: средним значением и масштабом (стандартным отклонением  $\sigma$ ), определяющим дробность шкалы.

# Некоторые из известных равновальных шкал в психологии



# Последовательность стандартизации – разработки

## тестовых норм

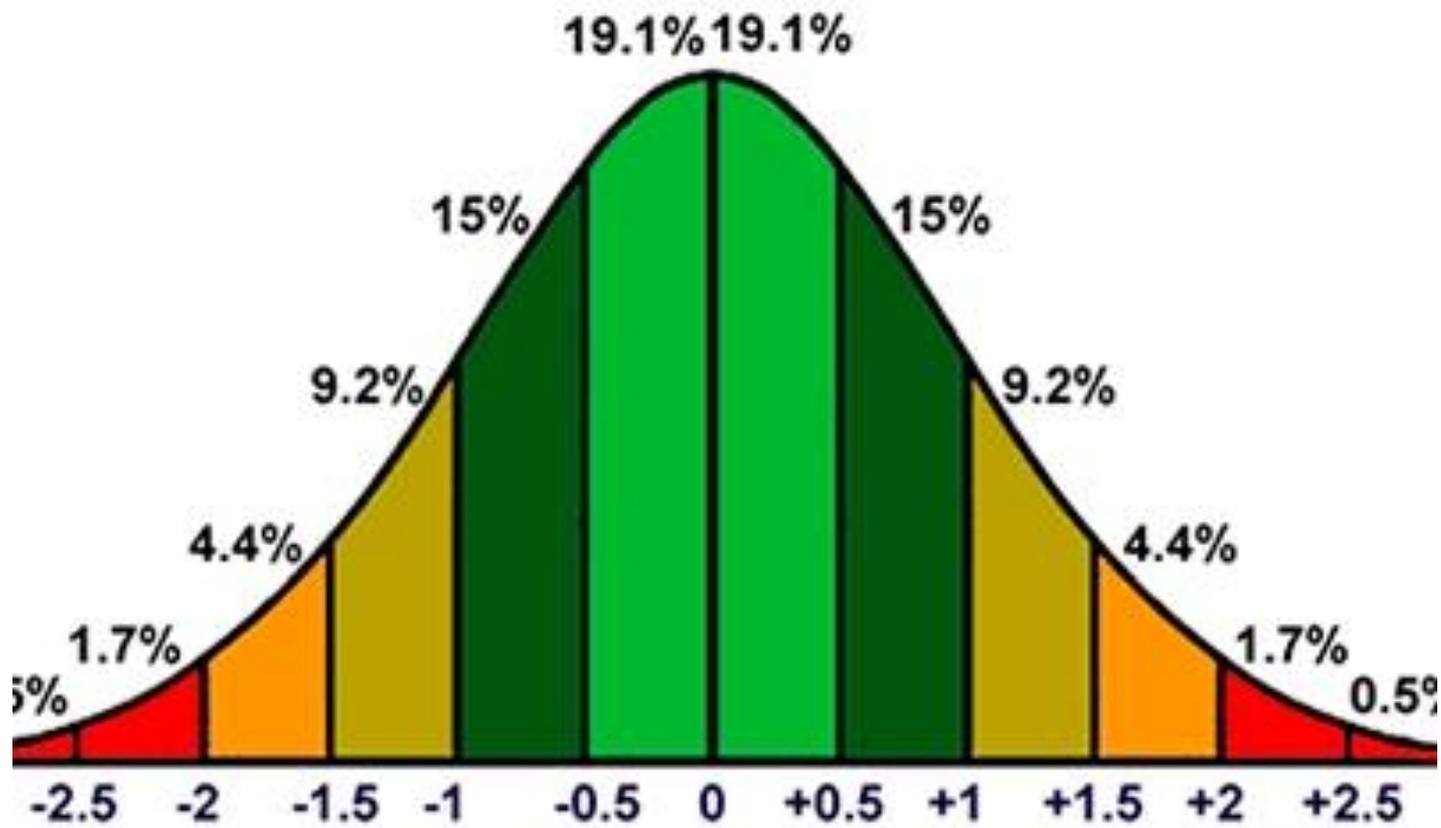
Общая последовательность стандартизации (разработки тестовых норм — таблицы пересчета «сырых» оценок в стандартные тестовые) состоит в следующем:

1. определяется генеральная совокупность, для которой разрабатывается методика и формируется репрезентативная выборка стандартизации;
2. по результатам применения первичного варианта теста строится распределение «сырых» оценок;
3. проверяют соответствие полученного распределения нормальному закону;
4. если распределение «сырых» оценок соответствует нормальному, производится линейная стандартизация;
5. если распределение «сырых» оценок не соответствует нормальному, то возможны два варианта:
  - перед линейной стандартизацией производят эмпирическую нормализацию;
  - проводят нелинейную нормализацию.

**Более подробно примеры и варианты стандартизации будут рассмотрены на лабораторных занятиях, а также упражнения на стандартизацию и тестовые шкалы.**

# Вопросы для проработки и самостоятельного изучения

1. Первичные описательные статистики. Меры центральной тенденции: среднее арифметическое. Преимущества и недостатки. Понятие моды как наиболее часто встречаемого признака в выборке. Правила нахождения моды для разных случаев. Бимодальные и мультимодальные выборки. Медиана как значение, делящее упорядоченное множество пополам.
2. Меры изменчивости. Разброс выборки. Дисперсия как характеристика отклонения от среднего. Стандартное отклонение. Стандартная ошибка для количественных признаков и долей. Квантили распределения (процентили, квартили).
3. Понятие нормального распределения и его параметры: среднее арифметическое и стандартное отклонение. Идеальная кривая нормального распределения К. Гаусса. Свойства кривой. Совпадение значений среднего арифметического, моды и медианы. Ассиметричные распределения: левосторонние, правосторонние.
4. Разработка тестовых норм. Процедура стандартизации. Различные шкалы, применяемые в тестах в результате стандартизации. Их связь с нормальным распределением.



# СТАТИСТИЧЕСКИЕ ГИПОТЕЗЫ И КРИТЕРИИ ИХ ПРОВЕРКИ

- Формулирование гипотез систематизирует предположения исследователя и представляет их в четком и лаконичном виде.  
**Статистические гипотезы подразделяются на нулевые и альтернативные, направленные и ненаправленные.**
- **Нулевая гипотеза** – это гипотеза об отсутствии различий или значимых связей (что одно и то же, ниже мы поясним это). Она обозначается как  $H_0$  и называется нулевой потому, что содержит число 0:  $X_1 - X_2 = 0$ , где  $X_1, X_2$  – 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  $H_1$ .
- Чаще в исследованиях требуется доказать наличие статистически значимых различий. Однако, бывают задачи, когда желательно доказать как раз отсутствие статистической значимости различий, то есть подтвердить нулевую гипотезу, – например, если исследователю нужно убедиться, что разные испытуемые получают хотя и различные, но уравновешенные по трудности задания, или что экспериментальная и контрольная выборки не различаются между собой по каким-то характеристикам помимо исследуемого фактора.
- **Нулевая и альтернативная гипотезы могут быть направленными и ненаправленными, а также с двусторонней или односторонней критической областью (последнее мы поясним ниже).**

# СТАТИСТИЧЕСКИЕ ГИПОТЕЗЫ И КРИТЕРИИ ИХ ПРОВЕРКИ

## Классификация типов статистических гипотез

	Нулевая гипотеза $H_0$	Альтернативная гипотеза $H_1$
Направленные гипотезы	$X_1$ не превышает $X_2$	$X_1$ превышает $X_2$
Ненаправленные гипотезы	$X_1$ не отличается $X_2$	$X_1$ отличается от $X_2$

Если исследователь отметил, что в одной из групп индивидуальные значения испытуемых по какому-либо признаку, например по социальной смелости, выше, а в другой ниже, то для проверки значимости этих различий необходимо сформулировать направленные гипотезы.

Если необходимо доказать, что в 1-й группе под влиянием каких-то экспериментальных воздействий произошли более выраженные изменения, чем во 2-й группе, – также необходимо сформулировать направленные гипотезы.

Если же необходимо доказать, что различаются формы распределения признака в группах 1 и 2, то формулируются ненаправленные гипотезы.

Проверка гипотез осуществляется с помощью критериев статистической оценки различий.

# **СТАТИСТИЧЕСКИЕ ГИПОТЕЗЫ И КРИТЕРИИ ИХ ПРОВЕРКИ**

**Статистические критерии. *Статистический критерий – это решающее правило, обеспечивающее принятие истинной и отклонение ложной гипотезы с высокой вероятностью.***

Статистические критерии обозначают также метод расчета определенного числа и само это число.

***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 того, что  $H_0$  верна.***

Когда говорят, что статистическая значимость различий определялась по критерию  $\chi^2$ , то имеется в виду, что использовали метод  $\chi^2$  для расчета определенного числа.

Когда говорят, что  $\chi^2=42,676$ , то имеем в виду определенное число, рассчитанное по методу  $\chi^2$ . Это число обозначается как эмпирическое значение критерия

# СТАТИСТИЧЕСКИЕ ГИПОТЕЗЫ И КРИТЕРИИ ИХ ПРОВЕРКИ

- В большинстве случаев для того, чтобы мы признали различия статистически значимыми, необходимо, чтобы эмпирическое значение критерия превышало критическое, хотя есть критерии (например, критерий Манна-Уитни или критерий знаков), в которых мы должны придерживаться противоположного правила.
- В некоторых случаях расчетная формула критерия включает в себя количество наблюдений в исследуемой выборке, обозначаемое как  $n$ . В этом случае эмпирическое значение критерия одновременно является тестом для проверки статистических гипотез. По специальной таблице вручную мы определяем, какому уровню статистической значимости различий соответствует данная эмпирическая величина. Примером такого критерия является  $t$ -критерий Стьюдента. Или компьютерный пакет выдает уровень статистической значимости. В большинстве случаев, однако, одно и то же эмпирическое значение критерия может оказаться значимым или незначимым в зависимости от количества наблюдений в исследуемой выборке ( $n$ ) или от так называемого количества степеней свободы, которое обозначается как  $v$  или как  $df$ .
- Для каждого случая определение количества степеней свободы имеет свою специфику, поэтому каждая формула для расчета эмпирического значения критерия обязательно сопровождается правилом (формулой) для определения числа степеней свободы. Зная  $n$  и/или число степеней свободы, мы по специальным таблицам можем определить критические значения критерия и сопоставить с ними полученное эмпирическое

# СТАТИСТИЧЕСКИЕ ГИПОТЕЗЫ И КРИТЕРИИ ИХ ПРОВЕРКИ

- **Критерии делятся на параметрические и непараметрические.** Параметрические критерии включают в формулу расчета параметры распределения, то есть средние и дисперсии (t-критерий Стьюдента, критерий  $F$  Фишера и др.) Непараметрические критерии не включают в формулу расчета параметров распределения и основаны на оперировании частотами или рангами (критерий  $U$  Манна-Уитни, критерий  $T$  Вилкоксона и др.).
- Параметрические критерии могут быть более мощными; чем непараметрические, но только в том случае, если признак измерен по интервальной шкале и нормально распределен. С интервальной шкалой есть определенные проблемы. Лишь с некоторой натяжкой мы можем считать данные, представленные не в стандартизованных оценках, как интервальные. Кроме того, проверка распределения на нормальность требует достаточно сложных расчетов, результат которых заранее неизвестен. Может оказаться, что распределение признака отличается от нормального, и нам так или иначе все равно придется обратиться к непараметрическим критериям.
- Непараметрические критерии лишены всех этих ограничений, и не требуют таких длительных и сложных расчетов. По сравнению с параметрическими критериями они ограничены лишь в одном - с их помощью невозможно оценить взаимодействие двух или более условий или факторов, влияющих на изменение признака. Эту задачу может решить только дисперсионный двухфакторный анализ.

# **СТАТИСТИЧЕСКИЕ ГИПОТЕЗЫ И КРИТЕРИИ ИХ ПРОВЕРКИ**

- **Уровни статистической значимости. Уровень значимости – это вероятность того, что исследователь счел различия существенными, а они на самом деле случайны.**
- **Когда мы указываем, что различия достоверны на 5%-ом уровне значимости, или при  $p < 0,05$ , то мы имеем виду, что вероятность того, что они все-таки недостоверны, составляет 0,05 (или  $0,05 \cdot 100\% = 5\%$ ).**
- **Когда мы указываем, что различия достоверны на 1%-ом уровне значимости, или при  $p < 0,01$ , то мы имеем в виду, что вероятность того, что они все-таки недостоверны, составляет 0,01.**
- **Уровень значимости - это вероятность отклонения нулевой гипотезы, в то время как она верна.**
- **Ошибка, состоящая в том, что мы отклонили нулевую гипотезу, в то время как она верна, называется ошибкой 1 рода.**
- **Если вероятность ошибки - это  $\alpha$ , то вероятность правильного решения:  $1 - \alpha$ . Чем меньше  $\alpha$ , тем больше вероятность правильного решения.**

# СТАТИСТИЧЕСКИЕ ГИПОТЕЗЫ И КРИТЕРИИ ИХ ПРОВЕРКИ

- Исторически сложилось так, что в психологии принято считать низшим уровнем статистической значимости 5%-ый уровень ( $p < 0,05$ ): достаточным - 1%-ый уровень ( $p < 0,01$ ) и высшим 0,1%-ый уровень ( $p < 0,001$ ), поэтому в таблицах критических значений обычно приводятся значения критериев, соответствующих уровням статистической значимости  $p < 0,05$  и  $p < 0,01$ , иногда –  $p < 0,001$ . Для некоторых критериев в таблицах указан точный уровень значимости их разных эмпирических значений.
- До тех пор, однако, пока уровень статистической значимости не достигнет  $p = 0,05$ , мы еще не имеем права отклонить нулевую гипотезу.
- **Правило отклонения  $H_0$  и принятия  $H_1$ .** Если эмпирическое значение критерия равняется критическому значению, соответствующему  $p < 0,05$  или превышает его, то  $H_0$  отклоняется, но мы еще не можем определенно принять  $H_1$ . Если эмпирическое значение критерия равняется критическому значению, соответствующему  $p < 0,01$  или превышает его, то  $H_0$  отклоняется и принимается  $H_1$ .
- **Исключения:** критерий знаков  $G$ , критерий Т Вилкоксона и критерий U Манна-Уитни. Для них устанавливаются обратные соотношения.
- Практически, однако, исследователь может считать достоверными уже те различия, которые не попадают в зону незначимости, заявив, что они статистически значимы при  $p < 0,05$ , или указав точный уровень значимости полученного эмпирического значения критерия, например:  $p = 0,02$ .

# СТАТИСТИЧЕСКИЕ ГИПОТЕЗЫ И КРИТЕРИИ ИХ ПРОВЕРКИ

**Мощность критерия** – это его способность выявлять различия, если они есть. Иными словами, это его способность отклонить нулевую гипотезу об отсутствии различий, если она неверна. Ошибка, состоящая в том, что мы приняли нулевую гипотезу, в то время как она неверна, называется ошибкой II рода.

Вероятность такой ошибки обозначается как  $\beta$ . Мощность критерия - это его способность не допустить ошибку II рода, поэтому:

$$\text{Мощность} = 1 - \beta$$

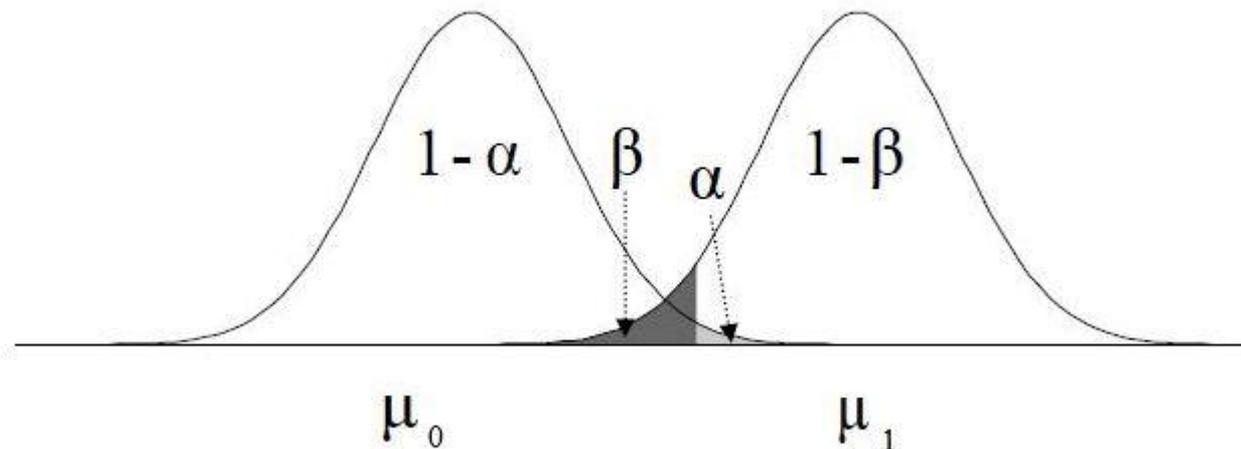
Мощность критерия определяется эмпирическим путем. Одни и те же задачи могут быть решены с помощью разных критериев, при этом обнаруживается, что некоторые критерии позволяют выявить различия там, где другие оказываются неспособными это сделать, или выявляют более высокий уровень значимости различий. Возникает вопрос: а зачем же тогда использовать менее мощные критерии? Дело в том, что основанием, для выбора критерия может быть не только мощность, но и другие его характеристики, а именно:

- простота;
- более широкий диапазон использования (например, по отношению к данным, определенным по номинативной шкале, или по отношению к большим  $n$ );
- применимость по отношению к неравным по объему выборкам;
- большая информативность результатов.

# Статистическая мощность

Величина мощности при проверке статистической гипотезы зависит от следующих факторов:

- величины уровня значимости, обозначаемого греческой буквой альфа, на основании которого принимается решение об отвержении или принятии альтернативной гипотезы;
- величины эффекта (то есть разности между сравниваемыми средними);
- размера выборки, необходимой для подтверждения статистической гипотезы.

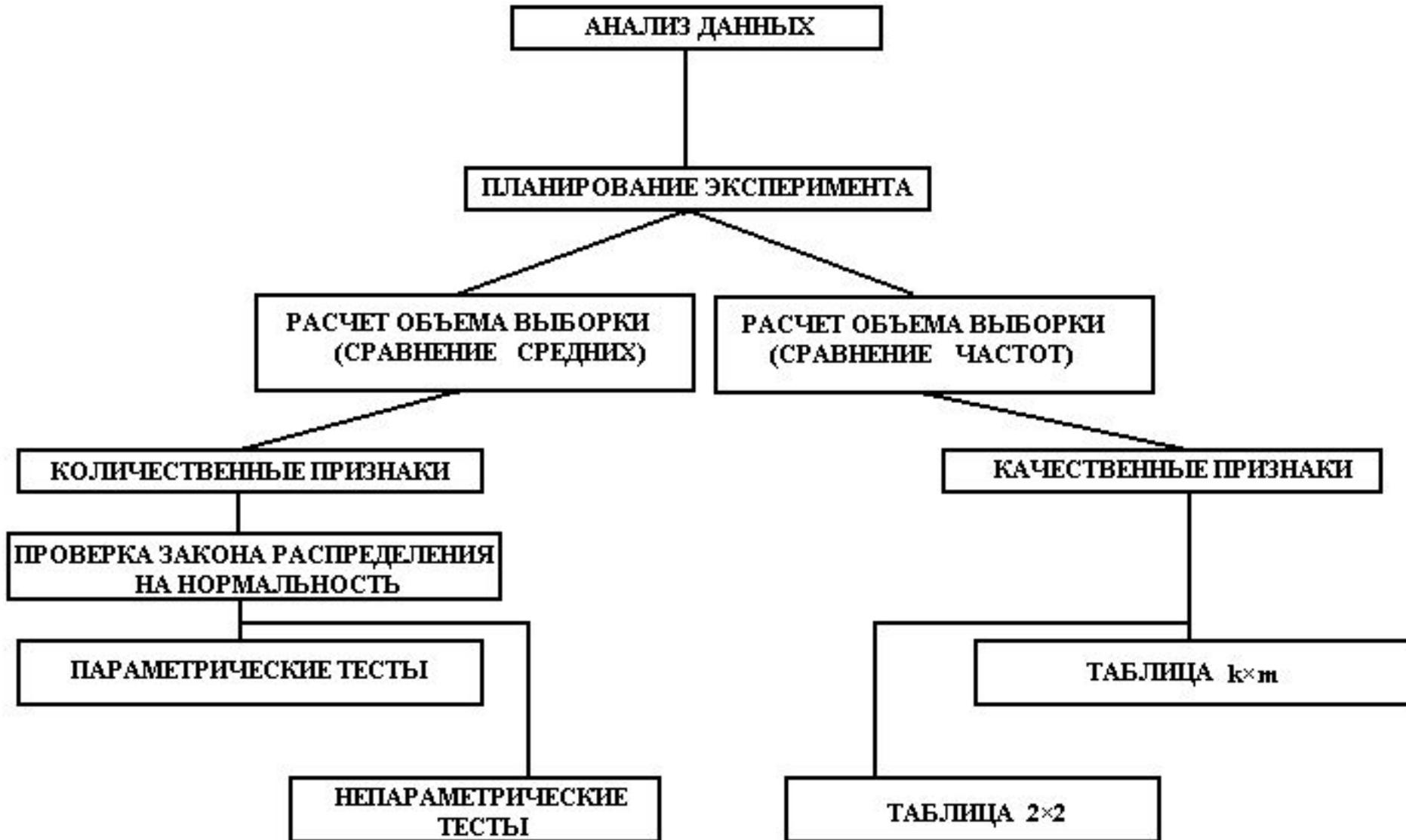


## Размер эффекта

- Величина эффекта определяет вероятность совершения ошибки второго рода. Коэффициент величины эффекта называется мерой эффекта  $d$ . Был введён в употребление Дж. Коэном и вычисляется как отношения разности между сравниваемыми средними к стандартному отклонению.
- Размер выборки, необходимой для подтверждения статистической гипотезы, влияет на статистическую мощность, так как с увеличением выборки уменьшается стандартная ошибка, а следовательно, увеличивается мощность.
- Понятия «размер эффекта», которым должен руководствоваться исследователь помимо собственно статистической значимости, будут рассмотрены на лабораторных занятиях, по отношению к различным типам переменных и характеристиках связи или различия.

# Задачи статистического сравнения двух средних или двух частот

## Возможный алгоритм действий



# Планирование эксперимента: расчет объема выборок

МОДУЛЬ ПЛАНИРОВАНИЯ ЭКСПЕРИМЕНТА статистических пакетов позволяют провести оценку размера выборки, достаточной для выявления клинически или биологически значимого эффекта с учетом заданной мощности статистического критерия и уровня значимости.

Например, можно провести расчет размера выборки для экспериментов, направленных на обнаружение статистически значимого различия между выборками. В модуле используются методы оценки объема выборки для сравнения двух частот и для сравнения двух средних. Расчеты справедливы только для случая, когда две группы имеют один и тот же размер



# Проведение расчетов в модуле «Планирование эксперимента» (пример из программы «Медстат»)

**РАСЧЕТ ОБЪЕМА ВЫБОРКИ ДЛЯ СРАВНЕНИЯ ДВУХ ЧАСТОТ**

Частоты События, %

Контрольная Группа	Исследуемая Группа
<input type="text" value="2"/>	<input type="text" value="8"/>

Уровень значимости

5%  
 1%

Мощность

80%  
 90%  
 95%

Выборки в  $n=203$  в каждой группе будут достаточными для того, чтобы выявить различия в частоте События в 6.00% с 80% мощностью на 5% уровне значимости. 6.00% различия равны разности между 2.00% частотой События в Контрольной группе и 8.00% частотой в Исследуемой группе.

**РАСЧЕТ ОБЪЕМА ВЫБОРКИ ДЛЯ СРАВНЕНИЯ ДВУХ СРЕДНИХ**

Различия средних

Стандартное отклонение	Клинически важный эффект
<input type="text" value="6"/>	<input type="text" value="5"/>

Уровень значимости

5%  
 1%

Мощность

80%  
 90%  
 95%

Выборки в  $n=43$  в каждой группе будут достаточными для выявления различий в 5 единицы с 90% мощностью на 1% уровне значимости, принимая стандартное отклонение равным 6 единицы.

# Обоснование задачи сопоставления и сравнения

Очень часто перед исследователем в психологии стоит задача выявления различий между двумя, тремя и более выборками испытуемых. Это может быть, например, задача определения психологических особенностей хронически больных детей по сравнению со здоровыми, юных правонарушителей по сравнению с законопослушными сверстниками или различий между работниками государственных предприятия и частных фирм, между людьми разной национальности или разной культуры и, наконец, между людьми равного возраста в методе «поперечных срезов». Сопоставление уровней показателей в разных выборках может быть необходимой частью комплексных диагностических, учебных, психокоррекционных и иных программ. Оно помогает нам обратить внимание на те особенности обследованных выборок, которые должны быть учтены и использованы при адаптации программ к данной группе в процессе их конкретного воплощения.

# Обоснование задачи статистической значимости сдвига в значениях исследуемого признака

- В психологических исследованиях часто бывает важно доказать, что в результате действия каких-либо факторов произошли достоверные изменения (сдвиги) в измеряемых показателях.
- Сопоставление показателей, полученных у одних и тех же испытуемых по одним и тем же методикам, но в разное время, определяет **временной сдвиг**. Сопоставление показателей, полученных по одним и тем же методикам, но в разных условиях измерения (например, «покоя» и «стресса»), дает нам ситуационный сдвиг. Условия измерения могут изменяться не только реально, но и умозрительно. Например, мы можем попросить испытуемого "представить себе", что он оказался в других условиях измерения: в будущем, в позиции других людей, которые оценивают его как бы со стороны, в состоянии разгневанного отца и т. п. Сопоставляя показатели, измеренные в обычных и воображаемых условиях, мы получаем **умозрительный сдвиг**. Мы можем создать специальные экспериментальные условия, предположительно влияющие на те или иные показатели, и сопоставить замеры, произведенные до и после экспериментального воздействия. Если сдвиги окажутся статистически достоверными, это позволит нам утверждать, что экспериментальные воздействия были существенными.

# Обоснование задачи статистической значимости сдвига в значениях исследуемого признака

- Например, мы можем сделать вывод о том, что данная программа тренинга действительно способствует развитию уверенности, или что данный способ внушающего воздействия влияет на изменение отношения испытуемых к той или иной проблеме, или что психодраматическая замена ролей подтверждает постулат Дж.Л. Морено о сближении позиций спорщиков после того, как им пришлось играть роль своего оппонента и т.п.
- Во всех этих случаях мы говорим о **сдвиге под влиянием** контролируемых или не контролируемых воздействий.
- Можно рассмотреть еще особую категорию **структурных сдвигов**. Так, мы можем сопоставлять между собой разные показатели одних и тех же испытуемых, если они измерены в одних и тех же единицах, по одной и той же шкале. Например, мы можем исследовать перепад между вербальным и невербальным интеллектом, измеренными по методике Д. Векслера, или сопоставлять экспертные оценки эмпатичности и наблюдательности, измеренные по одинаковой 10-балльной шкале, или время решения двух задач, измеренное в секундах, или экзаменационную успешность по разным дисциплинам и т.п.
- Для установления достоверности сдвигов в значениях признака в связанных выборках (чаще всего, как указано выше, это те же самые испытуемые) используются специальные *статистические критерии для связанных выборок*.

# ИССЛЕДОВАНИЕ СВЯЗИ МЕЖДУ ПЕРЕМЕННЫМИ

Еще Гиппократ в VI в. до н. э. обратил внимание на наличие связи между телосложением и темпераментом людей, между строением тела и предрасположенностью к тем или иным заболеваниям. Определенные виды подобной связи выявлены во всех предметных областях, включая психологию. Взаимосвязи на языке математики обычно описываются при помощи функций, которые графически изображаются в виде линий. На рис. изображены графики функций. 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*.

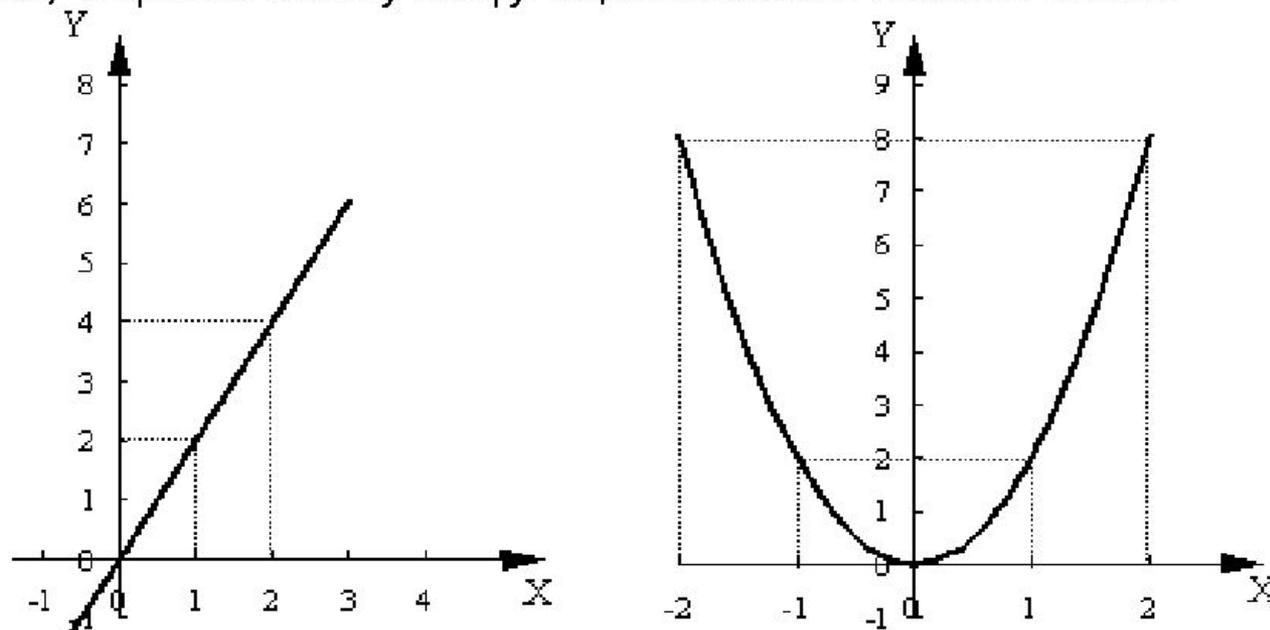


Рис. Примеры функциональных связей – линейная и квадратичная.

# ИССЛЕДОВАНИЕ СВЯЗИ МЕЖДУ ПЕРЕМЕННЫМИ.

- **Функциональные связи**, подобные изображенным на рис. выше, являются идеализациями. Их особенность заключается в том, что одному значению одной переменной соответствует строго определенное значение другой переменной. Например, такова взаимосвязь двух физических переменных — веса и длины тела (линейная положительная). Однако даже в физических экспериментах эмпирическая взаимосвязь будет отличаться от функциональной связи в силу неучтенных или неизвестных причин: колебаний состава материала, погрешностей измерения и пр.
- В психологии, как и во многих других науках, при изучении взаимосвязи признаков из поля зрения исследователя неизбежно выпадает множество возможных причин изменчивости этих признаков. Результатом является то, что даже *существующая в реальности функциональная связь между переменными выступает эмпирически как вероятностная (стохастическая): одному и тому же значению одной переменной соответствует распределение различных значений другой переменной (и наоборот)*.
- Будем говорить, что между двумя признаками  $X$  и  $Y$  существует корреляционная зависимость (взаимосвязь), при которой с изменением одного признака изменяется и другой, но каждому значению признака  $X$  могут соответствовать разные, заранее непредсказуемые значения признака  $Y$ , и наоборот.
- Для различия направленности влияния одного признака на другой введены понятия положительной и отрицательной связи.

# ИССЛЕДОВАНИЕ СВЯЗИ МЕЖДУ ПЕРЕМЕННЫМИ.

- Если с увеличением (уменьшением) одного признака в основном увеличиваются (уменьшаются) значения другого, то такая корреляционная связь называется прямой или положительной.
- Если с увеличением (уменьшением) одного признака в основном уменьшаются (увеличиваются) значения другого, то такая корреляционная связь называется обратной или отрицательной.
- В ряде случаев необходимо определить связь между двумя признаками, установить характер зависимости (прямая или обратная), количественно выразить достоверность связи. Для решения этих задач вычисляют **коэффициент корреляции  $r$** . Величина коэффициента корреляции лежит в пределах от  $-1$  до  $+1$ . Если коэффициент корреляции близок по модулю единице, то между изменением величины  $X$  и  $Y$  существует линейно пропорциональная зависимость. Если  $r > 0$ , то с ростом величины  $X$  величина  $Y$  также в среднем растет. Если  $r < 0$ , то с ростом величины  $X$  величина  $Y$  в среднем убывает. Если коэффициент корреляции по модулю близок нулю, то между величинами  $X$  и  $Y$  отсутствует линейная связь.
- Таким образом, коэффициент корреляции – важный показатель, показывающий взаимосвязь между двумя наборами данных. Отрицательное значение указывают на обратную корреляцию, положительное – на прямую. Чем ближе к 1 значение  $r$ , тем вероятнее наличие связи между показателями.

# ИССЛЕДОВАНИЕ СВЯЗИ МЕЖДУ ПЕРЕМЕННЫМИ.

**Общая классификация корреляционных связей** (по Ивантер Э.В., Коросову А.В., 1992):

- |                        |   |
|------------------------|---|
| 1) сильная, или тесная | при коэффициенте корреляции $r \geq 0,70$ ; |
| 2) средняя             | при $0,50 \leq r \leq 0,69$ ;               |
| 3) умеренная           | при $0,30 \leq r \leq 0,49$ ;               |
| 4) слабая              | при $0,20 \leq r \leq 0,29$ ;               |
| 5) очень слабая        | при $r \leq 0,19$ .                         |

**Частная классификация корреляционных связей:**

- |                                |   |
|--------------------------------|---|
| 1) высокая значимая корреляция | при $r$ , соответствующем уровню статистической значимости $p < 0,01$ ; |
| 2) значимая корреляция         | при $r$ , соответствующем уровню статистической значимости $p < 0,05$ ; |
| 3) тенденция достоверной связи | при $r$ , соответствующем уровню статистической значимости $p < 0,10$ ; |
| 4) незначимая корреляция       | при $r$ , не достигающем уровня статистической значимости.              |

# ИССЛЕДОВАНИЕ СВЯЗИ МЕЖДУ ПЕРЕМЕННЫМИ.

- Две эти классификации не совпадают. Первая ориентирована только на величину коэффициента корреляции, а вторая определяет, какого уровня значимости достигает данная величина коэффициента корреляции при данном объеме выборки. Чем больше объем выборки, тем меньшей величины коэффициента корреляции оказывается достаточно, чтобы корреляция была признана достоверной. В результате при малом объеме выборки может оказаться так, что сильная корреляция окажется недостоверной. В то же время при больших объемах выборки даже слабая корреляция может оказаться достоверной.
- Обычно принято ориентироваться на вторую классификацию, поскольку она учитывает объем выборки. Вместе с тем, необходимо помнить, что сильная, или высокая, корреляция – это корреляция с

# АНАЛИЗ КОЛИЧЕСТВЕННЫХ ДАННЫХ, РАСПРЕДЕЛЕНИЕ КОТОРЫХ НЕ ОТЛИЧАЕТСЯ ОТ НОРМАЛЬНОГО (ПАРАМЕТРИЧЕСКИЕ ТЕСТЫ).

Параметрические тесты (ПТ) предполагают известным закон распределения анализируемой величины, который описывается определенными параметрами. ПТ применяются в случае, когда закон распределения признака в генеральной совокупности подчиняется некоторому известному, в данном случае нормальному закону распределения. Нормальность распределения должна быть статистически доказана до применения этого критерия, в противном случае его применения может привести к ложным выводам. ПТ строго обоснованы и хорошо изучены, поэтому если существуют предпосылки, их использование отдается предпочтение перед соответствующими непараметрическими критериями.



# АНАЛИЗ КОЛИЧЕСТВЕННЫХ ДАННЫХ, РАСПРЕДЕЛЕНИЕ КОТОРЫХ ОТЛИЧАЕТСЯ ОТ НОРМАЛЬНОГО

Непараметрические тесты (НТ) не требуют знания конкретного закона распределения анализируемой величины. Обычно они достаточно строго обоснованы. НТ применяются в случае, когда закон распределения отличается от нормального или данные измеряют в дискретной шкале измерения. Особенно эти критерии полезны при анализе малых выборок. НТ слабо чувствительны к отклонениям от стандартных условий (робастные). Однако они в большинстве случаев являются менее мощными, чем их параметрические аналоги.

Данные методы анализа оперируют не с числовыми величинами, а с их рангами (порядковым номером элемента в упорядоченном по возрастанию вариационном ряду). Следует отметить, что в случае, когда в анализируемых данных содержится большое количество совпадающих значений (большие связки), применение этих методов сомнительно, если же количество связок невелико, то это учитывается в расчетах путем введения соответствующих поправок.



## АНАЛИЗ КАЧЕСТВЕННЫХ ДАННЫХ (ТАБЛИЦА $k \times m$ )

Если набор данных показывает, какой из нескольких нечисловых категорий принадлежит каждый из объектов, то такие данные являются качественными (поскольку они регистрируют определенное качество, которым обладает объект). Если имеется несколько классов, то можно оперировать процентами (частотами) событий в каждом классе (получив, таким образом, числовое подобие из представленных категориями данных). Если есть в точности две категории, их можно обозначить цифрами 1 и 0, приписав эти значения соответственно каждому из объектов и затем (в достаточно многих случаях) обрабатывать полученные данные как количественные. Номинальные качественные данные определяются в терминах категорий, которые нельзя содержательно упорядочить, а они просто имеют названия. Для таких категорий нет чисел, с которыми можно производить вычисления, и нет оснований для ранжирования. Например, можно сказать, что 2 индивидуума различимы в терминах переменной  $A$  (например, болен или здоров). Типичные примеры номинальных переменных - пол, национальность, цвет, диагноз и т.д. Часто номинальные переменные называют категориальными. Категориальная (качественная) переменная - это



## МОДУЛЬ ОЦЕНКИ ЭФФЕКТА ТЕРАПИИ (ИЛИ ЭФФЕКТИВНОСТИ ИНЫХ МЕТОДОВ)

Расчет рисков позволяет провести оценку относительной эффективности двух методов воздействия в случае, когда эффект выражается признаком:

«Отрицательный результат» («Событие произошло»), «Положительный результат» («Не событие произошло»). В этом случае наиболее часто для указания обобщенной оценки клинического эффекта в «Группе контроля» по сравнению с «Группой исследования» используются следующие показатели:

- «Снижение Абсолютного Риска» (САР)
- «Снижение Относительного Риска» (ОР),
- «Число Больных, которых Необходимо Лечить» (ЧБНЛ),

Абсолютный риск (АР) – отношение числа больных, у которых возник определенный клинический исход (благоприятный или неблагоприятный), в группе лечения или контроля к общему числу больных в соответствующей группе. Соответственно САР – обычно разница рисков заболевания в группе пациентов, подверженных некоторому фактору и не подверженных ему. Если САР не отличается от 0, то это свидетельствует об отсутствии различий между сравниваемыми группами.

ОР – соотношение двух рисков, обычно риск заболевания в группе пациентов, подверженных некоторому фактору, деленный на риск неподверженных пациентов. ОР не несет информации о величине абсолютного риска (заболеваемости). Даже при высоких значениях ОР абсолютный риск может быть совсем небольшим, если заболевание редкое. ОР показывает силу связи между воздействием и заболеванием. В клинических исследованиях это отношение частоты определенного исхода в экспериментальной группе к частоте таких же исходов в контрольной группе. Если ОР равен единице, то это свидетельствует об отсутствии различий между сравниваемыми группами.

**ЧБНЛ:** способ оценки относительной эффективности двух методов лечения. Показывает, какое количество больных необходимо подвергнуть лечению определенным методом в течение определенного времени, чтобы достичь одного благоприятного исхода или предотвратить один неблагоприятный исход. Этот показатель является величиной, обратной **САР**. Если, например, в испытании излечились 30% больных от лечения А и 10% - от лечения Б, то абсолютное снижение риска составляет  $30\% - 10\% = 20\%$  (0,2) и ЧБНЛ =  $1/0,2 = 5$ . Иными словами, необходимо подвергнуть лечению А ещё 5 больных, чтобы получить дополнительно излеченного больного по сравнению с лечением Б.

В программе предусмотрен расчет этих показателей с указанием их доверительного интервала (на 95% уровне доверительной вероятности). Для расчета доверительных интервалов при оценке **САР** и **ЧБНЛ**



## Некоторые критерии для сравнения выборок

Более подробно будут рассмотрены на лабораторных работах.

1. Критерии проверки на нормальность: критерий «хи-квадрат» Пирсона  $\chi^2$ , тест Шапиро-Уилка
2. Параметрические критерии для независимых и зависимых выборок – критерии Стьюдента, дисперсионный анализ, апостериорные критерии для попарных сравнений (например, Шеффе, Даннета – при сравнениях с контрольной группой).
3. Непараметрические критерии для независимых выборок (Манна-Уитни), зависимых (Вилкоксона), критерий Крускала-Уоллиса для независимых и Фридмана для зависимых выборок, апостериорные критерии (например, Данна).
4. Параметрический Критерий Фишера для сравнений дисперсий в двух выборках.

## Некоторые критерии для сравнения выборок

6. Многофункциональные критерии «Угловое преобразование Фишера», «Хи-квадрат Фишера» для независимых выборок, критерий Мак-Немара для зависимых выборок – для сравнения категориальных данных.

7. Методы корреляционного анализа: коэффициент корреляции Пирсона (параметрический метод); Кендалла, Спирмена (непараметрические методы).

8. Критерии тенденций Пейджа, Джонкира; анализ «выживаемости», и еще очень-очень много разных критериев, лишь некоторые из них мы с Вами будем разбирать, но и их достаточно для 95% задач на сравнения, с которыми Вы столкнетесь. Для остальных задач на сравнения существуют многомерные методы.

# Вопросы для проработки и самостоятельного изучения

1. Понятие статистической гипотезы.
2. Сущность проверки статистической гипотезы – установить, согласуются ли экспериментальные результаты и выдвинутая гипотеза; допустимо ли отнести расхождение между ними за счет случайных величин.
3. Нулевая гипотеза. Понятие уровня статистической значимости как вероятности ошибки при принятии решения об отклонении нулевой гипотезы. Уровни статистической значимости. Этапы принятия статистической гипотезы (решения).
4. Ошибка второго рода. Мощность критерия. Статистическая и содержательная (психологическая, клиническая) значимость.

## Рекомендованные к использованию в рамках дисциплины статистические компьютерные пакеты

Не поленитесь установить эти пакеты, каждый из них имеет свои преимущества и недостатки, уникальные возможности. Подробно будем знакомиться на лабораторных работах. Работают под Windows.

- Primer of Biostatistics. Легчайшая, portable, содержит большую часть методов (кроме многомерных), а также расчет размеров выборки, мощности, то есть планирование эксперимента <https://yadi.sk/d/9jixsf2uzxsM>
- StatMed. Легчайшая, portable, содержит базовые методы в основном для количественных данных, четкий алгоритм принятия решений – какие критерии применять – параметрические или непараметрические <https://yadi.sk/d/qsIQJ4IGuzyVa>
- MedStat. Легчайшая, portable, содержит базовые методы для количественных и качественных (ограничения демо-версии на множественные сравнения), По сути – расширенный вариант StatMed. Для большинства случаев – наиболее годный и пошаговый инструмент. <https://yadi.sk/d/vXnsrGNWrngzk>

# Рекомендованные к использованию в рамках дисциплины статистические компьютерные пакеты

- Biostat. Более сложный инструмент по сравнению с представленными выше, в то же время прекрасно русифицирован и позволяет решать большинство задач начиная от планирования эксперимента, за исключением некоторых многомерных методов, например факторного и кластерного анализа <https://yadi.sk/d/kGEWv92Wv22SX>
- StatSoft Statistica. Версия 10 – довольно новая, много улучшений. Непросто установить, не сразу поддается пониманию, особенно самостоятельному. В большинстве случаев – эталонный инструмент. Есть прекрасные обучающие материалы в Интернете [https://yadi.sk/d/oZD1Un\\_hv22xi](https://yadi.sk/d/oZD1Un_hv22xi)
- SPSS Statistics 21 версия. Плюс модуль AMOS. В программе есть всё и даже больше. С другой стороны, не всем (мне в том числе) нравится интерфейс. Преимущества – прекрасная русификация, много обучающих материалов в Интернете. <https://yadi.sk/d/FyntqfDMv23Gp>
- Кстати, Microsoft Excel и его аналоги (другие табличные процессоры) никто не отменял. Они содержат существенные возможности статистической обработки данных