Случайная величина -

Генеральная совокупность X (сл. вел. X) – множество возможных значений случайной величины X

Закон распределения ген. сов. X – закон распределения сл. вел. X

Случайная выборка из ген. сов. X – совокупность независимых сл. вел. $X_1, \ldots, X_n$, каждая из которых имеет то же распределение, что и сл. вел. X. Записывается $\overline{X_n} = (X_1, \ldots, X_n)$

Выборка из ген. сов. X (реализация сл. выборки $\overline{X_n}$) – это любое возможное значение $\overline{x_n} = (x_1, \ldots, x_n)$ сл. выборки $\overline{X_n}$. Интерпретируется как результаты n независимых наблюдений над сл. величиной X.

Выборочный метод – свойства сл. вел. X устанавливаются путем изучения тех же свойств на случайной выборке.

Выборочное пространство - \chi _n χ_n – множество значений сл. выборки $\overline{X_n}$.

Как выражается функция распределения сл. выборки $F_{\overline{X_n}}(t_1, \ldots, t_n)$ через функцию распределения генеральной совокупности (т.е. через ф-цию распр. X).

$$F_{\vec{X}}(t_1, \ldots, t_n) = \mathbf{P}\{X_1 < t_1, \ldots, X_n < t_n\} = \\ = \prod_{i=1}^{n} \mathbf{P}\{X_i < t_i\} = \prod_{i=1}^{n} F(t_i), \quad (1.1)$$

где $F(t)$ — функция распределения случайной величины X (генеральной совокупности X).

стр 21. *Статистическая модель* – выборочное пространство, на котором задан класс распределений сл. выборки (если мы знаем тип функции распределения, но не знаем ее параметры).

Параметрическая модель

Статистика (выборочная характеристика) – любая функция случайной выборки $g(X_1, \ldots, X_n) = g(\overline{X_n})$ – она является случайной величиной с распределением, называемым **выборочным распределением**

Выборочное распределение выборочной характеристики

Выборочное значение выборочной характеристики – значение $g(\overline{x_n})$ выборочной характеристики $g(\overline{X_n})$, определенное по реализации $\overline{x_n}$ случайной выборки $\overline{X_n}$.

стр. 24 ***Сходимость по вероятности***

Сходимость по распределению (слабая)

Задачи мат. статистики: оценка неизв. параметров, проверка стат. гипотез, установление формы и степени связи между сл. вел.

Два подхода к оценке неизвестный параметров функции распределения генеральной совокупности – точечная оценка и интервальная оценка.

Точечная оценка (или просто **оценка**) неизвестного параметра θ ф-ции распр. генеральной совокупности – это статистка (функция) $\hat{\theta}(\overline{X_n})$, выборочное значение $\hat{\theta} = \hat{\theta}(\overline{x_n})$ которой для любой реализации $\overline{x_n}$ принимают за приближенное значение неизвестного параметра θ.

Значение точечной оценки - $\hat{\theta}$.

Интервальная оценка с коэффициентом доверия γ неизвестного параметра θ ф-ции распр. генеральной совокупности – это пара статистик (функций) $\underline{\theta}(\overline{X_n})$ и $\overline{\theta}(\overline{X_n})$ таких, что с вероятностью γ выполняется неравенство $\underline{\theta}(\overline{X_n}) \leq \theta \leq \overline{\theta}(\overline{X_n})$ (то есть таких, что $P\{\underline{\theta}(\overline{X_n}) \leq \theta \leq \overline{\theta}(\overline{X_n})\} = \gamma$).

Доверительный интервал для $\boldsymbol{\theta}$ с коэффициентом доверия $\boldsymbol{\gamma}$ - $\left(\underline{\theta}(\overline{X_n}), \overline{\theta}(\overline{X_n})\right)$.

Статистическая гипотеза – любое предположение о распределении вероятностей (о вероятностных свойствах) наблюдаемой сл.вел. (гипотеза о величине м.о., об однородности (т.е. равенстве) дисперсий, о виде распределения и т.д.).

Корреляционный и *дисперсионный* анализ – наличие связи между величинами и ее существенность. *Регрессионный* анализ – построение регрессионной модели (т.е. зависимости ср. знач. сл. величины от знач. других сл. величин).

Вариационный ряд выборки $(\boldsymbol{x_1}, \dots, \boldsymbol{x_n})$ – упорядоченная последовательность элементов выборки $(x_{(1)}, \dots, x_{(n)})$.

Вариационный ряд случайной выборки $(\boldsymbol{X_1}, \dots, \boldsymbol{X_n})$ – последовательность случайных величин $(X_{(1)}, \dots, X_{(n)})$, где $X_{(i)}$ – сл. величина, которая при каждой реализации $\overline{x_n}$ случайной выборки $\overline{X_n}$ принимает значение, равное i-му члену вариационного ряда выборки $\overline{x_n}$.

Функции распр. крайних членов вариационного ряда ($\boldsymbol{X_{(1)}}$ и $\boldsymbol{X_{(n)}}$). Вывод

показать (см. пример 2.20), что для ***крайних членов вариационного ряда*** случайной выборки $X_{(1)}$ и $X_{(n)}$ их функции распределения имеют вид

$$\mathbf{P}\{X_{(1)} < x\} = 1 - \left(1 - F(x)\right)^n$$

и

$$\mathbf{P}\{X_{(n)} < x\} = F^n(x).$$

Статистический ряд – таблица, которая в первой строчке содержит уникальные отсортированные значения элементов выборки, а во второй – количество их повторений.

Частота – количество раз, которое встречается элемент в выборке.

Относительная частота (частость) – отношение частоты значения элемента выборке к общему количеству элементов в выборке.

Интервальный статистический ряд – отрезок, содержащий все значения выборки, делят на равные части и составляют статистический ряд, в котором количество элементов подсчитывается на интервале.

Оптимальное число интервалов для гистограммы – по правилу Стёрджеса – $m = 1 + \lfloor \log_2 n \rfloor$.

Стр 32 **Выборочная функция распределения**

Эмпирическая функция распределения

Теоретическая функция распределения

Эмпирич. плотность распр.

Гистограмма, Полигон частот

Выборочные числовые моменты

Теоретические (генеральные) числовые характеристики

Выборочный начальный момент k-го порядка,

Выборочный центральный момент k-го порядка

Выборочное среднее, Выборочная дисперсия,

Выборочное ср/квадр. отклонение

Выборочный корреляционный момент,

Выборочный коэффициент корреляции

Кор. момент выборки,

Коэф. кор. выборки

из лабы 1 **Коэффициент вариации**

Стандартное отклонение

Стандартизованная асимметрия

Стандартизованный эксцесс

! Актуальные критерии нормальности распределения (асимметрия и эксцесс и что-то еще?)

из лабы 2 **Состоятельность оценки**

Состоятельная оценка – это точечная оценка, которая при увеличении объема выборки сходится по вероятности к оцениваемому параметру.

Несмещенная оценка – это точечная оценка, математическое ожидание которой равно оцениваемому параметру.

Эффективная оценка – это точечная оценка, дисперсия которой меньше дисперсии любой другой оценки данного параметра, т.е. наиболее эффективной считают ту из нескольких возможных несмещенных оценок, которая имеет наименьшую дисперсию.

Требование несмещенности на практике не всегда целесообразно, так как оценка с небольшим смещением и малой дисперсией может оказаться предпочтительнее несмещенной оценки с большой дисперсией. На практике не всегда удается удовлетворить одновременно все три этих требования, однако выбору оценки должен предшествовать ее критический анализ со всех перечисленных точек зрения.

❗Правила для определения достаточного объема выборки

Законы больших чисел

Теорема Бернулли – закон больших чисел https://studopedia.ru/12_163342_reshenie.html

При неограниченном увеличении числа однородных независимых опытов частота события будет сколь угодно мало отличаться от вероятности события в отдельном опыте.

Иначе, вероятность того, что отклонение относительной частоты m/n наступления события A от постоянной вероятности p события A очень мало при n -> +oo, стремится к 1 при любом eps > 0

$$P\left\{\left|\frac{m}{n}-p\right|<\varepsilon\right\}\underset{n\to\infty}{\to}1$$

Геометрическое распределение - распределение вероятностей случайной величины X равной количеству «неудач» до первого «успеха» в серии испытаний Бернулли и принимающей значения $n = 0,1,2,\ldots$ либо распределение вероятностей случайной величины $Y = X + 1$ равной номеру первого «успеха» и принимающей значения $n = 1,2,3,\ldots$.

Функция вероятности	$q^n p$
Функция распределения	$1 - q^{n+1}$

Экспоненциальное распределение - абсолютно непрерывное распределение, моделирующее время между двумя последовательными свершениями одного и того же события.

Обозначение	$\mathrm{Exp}(\lambda)$
Параметры	$\lambda > 0$ - интенсивность или обратный коэффициент масштаба
Носитель	$x \in [0; \infty)$
Плотность вероятности	$\lambda e^{-\lambda x}$
Функция распределения	$1 - e^{-\lambda x}$
Математическое ожидание	λ^{-1}

Распределение Бернулли – дискретное распределение вероятностей, моделирующее случайный эксперимент произвольной природы, при заранее известной вероятности успеха (p) или неудачи ($p-1$).

Функция вероятности	q	$k=0$
	p	$k=1$
Функция распределения	0	$k<0$
	q	$0 \le k < 1$
	1	$k \ge 1$

Биномиальное распределение - распределение количества «успехов» в последовательности из n независимых случайных экспериментов, таких, что вероятность «успеха» в каждом из них постоянна и равна p.

Пусть $X_1, \dots, X_n$- конечная последовательность независимых случайных величин, имеющих одинаковое распределение Бернулли с параметром p. Тогда сл. вел. $Y = X_1 + \cdots + X_n$ имеет биномиальное распределение с параметрами n и p. $Y \sim Bin(n,p)$.

Носитель	$k \in \{0, \dots, n\}$
Функция вероятности	$\binom{n}{k} p^k q^{n-k}$

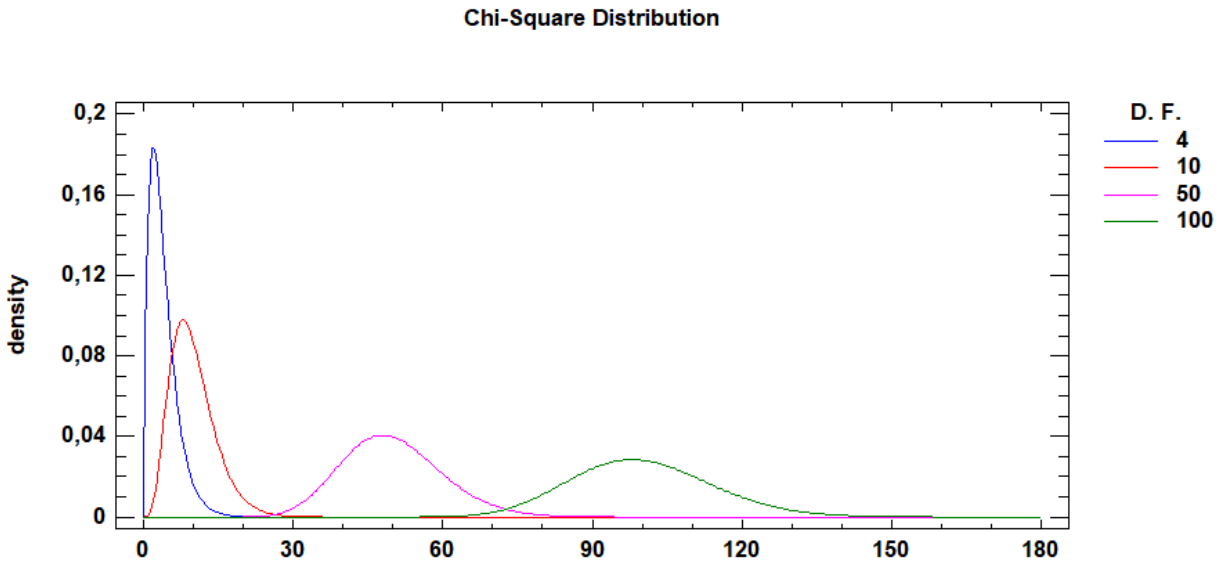
$$F_Y(y) \equiv \mathbb{P}(Y \leqslant y) = \sum_{k=0}^{\lfloor y \rfloor} \binom{n}{k} p^k q^{n-k}, \; y \in \mathbb{R},$$

Распределение Пуассона – вероятностное распределение дискретного типа, моделирует случайную величину, представляющую собой число событий, произошедших за фиксированное время, при условии, что данные события происходят с некоторой фиксированной средней интенсивностью и независимо друг от друга. $Y \sim P(\lambda)$, где $\lambda > 0$ – м.о.

Функция вероятности $p(k) \equiv \mathbb{P}(Y = k) = \frac{\lambda^k}{k!} e^{-\lambda}$, Функция распределения $\frac{\Gamma(k+1, \lambda)}{k!}$.

Распределение χ^2 (Chi-Squared) с k степенями свободы — это распределение суммы квадратов k независимых стандартных нормальных случайных величин.

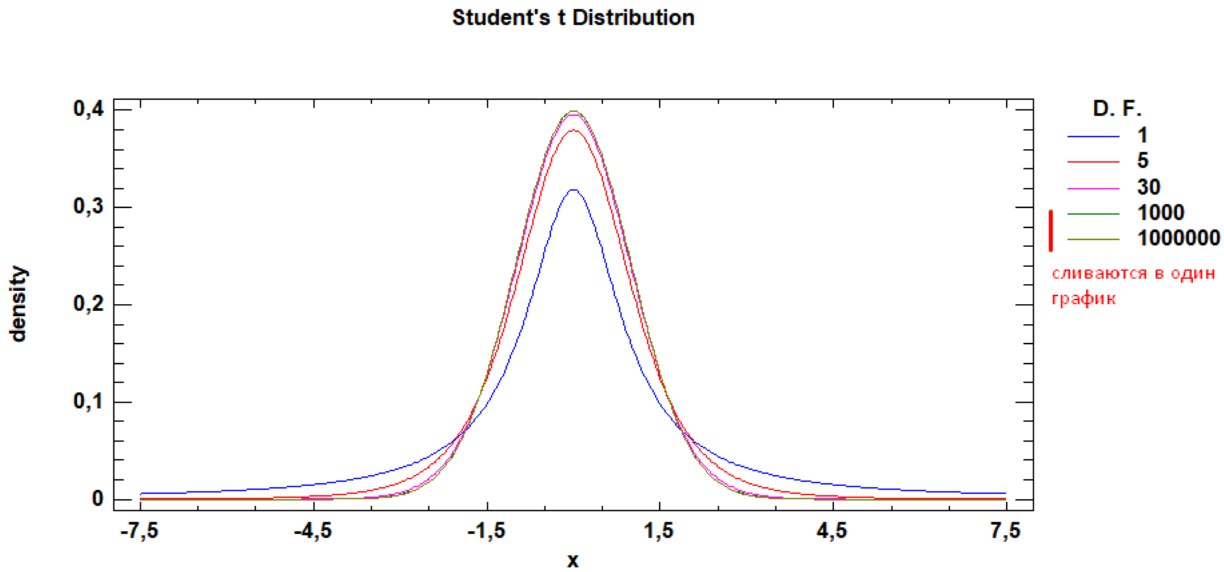
Пусть $Z_1, \dots, Z_k$ – совместно независимые стандартные нормальные случайные величины, то есть $Z_i \sim N(0{,}1)$. Тогда случайная величина $X = Z_1^2 + \cdots + Z_k^2$ имеет распределение хи-квадрат с k степенями свободы, т.е. $X \sim f_{\chi^2(k)}(X)$.



Распределение Стьюдента (Student's t) – это однопараметрическое семейство абсолютно непрерывных распределений. Пусть $Y_0, \ldots, Y_n$ – конечная последовательность независимых стандартных нормальных случайных величин, т.е. $Y_i \sim N(0,1)$. Тогда распределение сл. вел.

$$t = \frac{Y_0}{\sqrt{\frac{1}{2}\sum_{i=1}^{n} Y_i^2}}$$

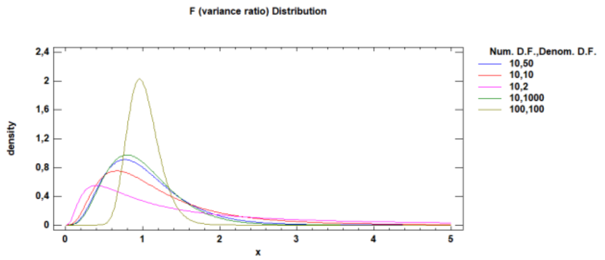
имеет распределение Стьюдента с n степенями свободы. $t \sim t(n)$.

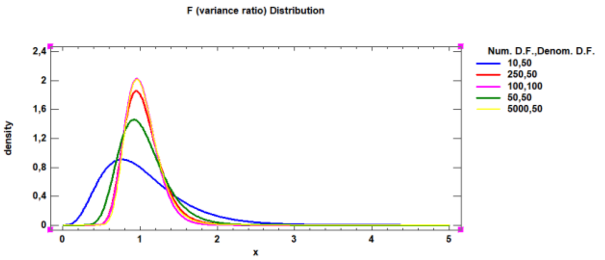


Распределение Фишера F (Variance Ratio) – это двухпараметрическое семейство абсолютно непрерывных распределений. Пусть Y_1, Y_2 — две независимые случайные величины, имеющие распределение хи-квадрат: $Y_i \sim \chi^2(d_i)$, где $d_i \in \mathbb{N}, i = 1, 2$. Тогда распределение случайной величины

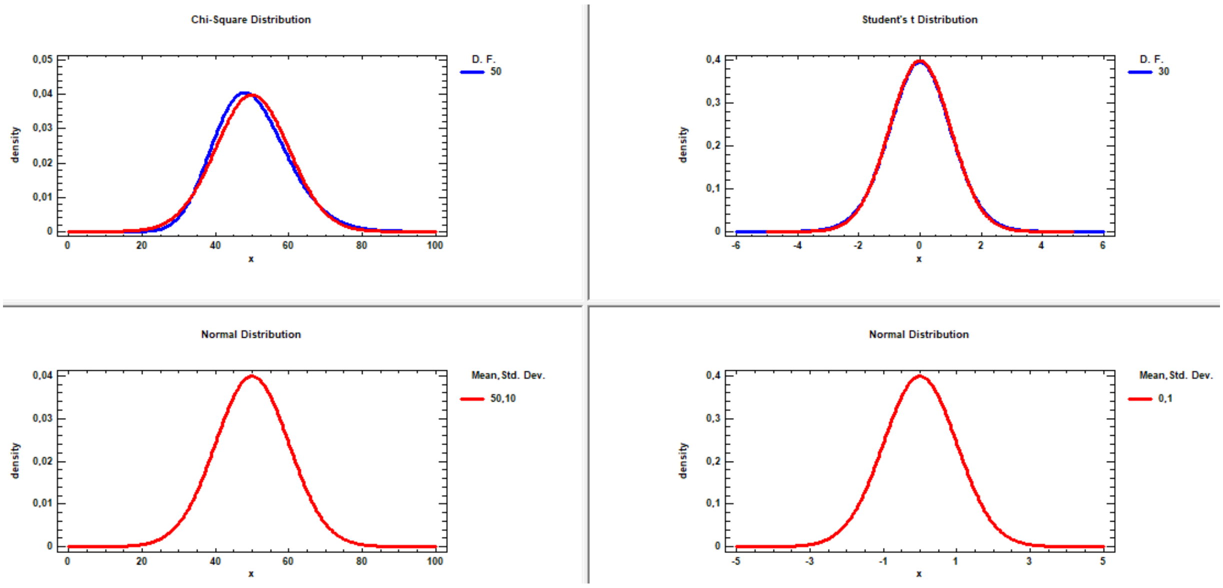
$$F = \frac{Y_1/d_1}{Y_2/d_2},$$

называется распределением Фишера со степенями свободы d_1 и d_2. Пишут $F \sim F(d_1, d_2)$.





Асимптотическая нормальность распределений Стьюдента и χ^2 – $\mathrm{t}(30) \cong \mathrm{N}(0,1)$, а $\chi^2(\mathrm{v}) \cong \mathrm{N}(\mathrm{v}, \sqrt{2 * \mathrm{v}})$ при $\mathrm{v} \geq 50$.



Bin(0.5, 100), Bin(0.01, 100), Bin(0.99, 100)

А) Для какой из выборок гистограмма «похожа» на нормальную кривую? Почему это можно было ожидать (вспомните предельные теоремы из теории вероятностей (какую???).

Б) На какое распределение должна быть «похожа» гистограмма для второго распределения? Наложите это распределение на гистограмму.

В) Почему нормальная аппроксимация дает плохой результат для третьей выборки?

Локальная теорема Муавра — Лапласа

Если в схеме Бернулли n стремится к бесконечности, величина $p \in (0,1)$ постоянна, а величина $x_m = \dfrac{m - np}{\sqrt{npq}}$ ограничена равномерно по m и n (то есть $\exists a, b : -\infty < a \leqslant x_m \leqslant b < +\infty$), то

$$P_n(m) = \frac{1}{\sqrt{2\pi npq}} \exp\left(-\frac{x_m^2}{2}\right)(1 + \alpha_n(m))$$

где $|\alpha_n(m)| < \dfrac{c}{\sqrt{n}}, c = \mathrm{const} > 0$.

Приближённую формулу

$$P_n(m) \approx \frac{1}{\sqrt{2\pi npq}} \exp\left(-\frac{x_m^2}{2}\right)$$

рекомендуется применять при $n > 100$ и при $m > 20$.

γ-доверительная интервальная оценка – $\left(\underline{\theta}(\overline{X_n}), \overline{\theta}(\overline{X_n})\right)$.

Нижняя и верхняя границы интервальной оценки - пара статистик (функций) $\underline{\theta}(\overline{X_n})$ и $\overline{\theta}(\overline{X_n})$.

Коэффициент доверия (доверительная вероятность, уровень доверия).

Односторонняя нижняя (и соответственно верхняя) – доверительная граница – $\underline{\theta}(\overline{X_n})$, когда $P\{\underline{\theta}(\overline{X_n}) \le \theta\} = \gamma$.

Пример 3.1. Пусть θ — среднее значение предела прочности X некоторого материала, которое оценивают независимо друг от друга в каждой из N различных лабораторий по результатам n независимых натурных испытаний. Иначе говоря, среднее значение предела прочности в каждой лаборатории оценивают по „своим“ *экспериментальным данным*, представленным *выборкой объема* n, и в каждой лаборатории получают „свои“ значения верхней и нижней границ γ-доверительного интервала (рис. 3.1).

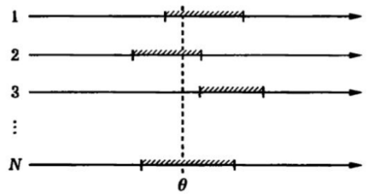


Рис. 3.1

Возможны случаи, когда γ-доверительный интервал для параметра θ не накрывает его истинного значения. Если M — число таких случаев, то при больших значениях N должно выполняться приближенное равенство $\gamma \approx (N - M)/N$. Таким образом, если опыт — получение выборки объема n в лаборатории, то уровень доверия γ — доля тех опытов (при их многократном независимом повторении), в каждом из которых γ-доверительный интервал накрывает истинное значение оцениваемого параметра.

§ 2. ИНТЕРВАЛЫ В НОРМАЛЬНОЙ МОДЕЛИ

Пример 2. Допустим, что элементы выборки X_i распределены по закону $\mathcal{N}(\theta, \sigma^2)$, причем параметр масштаба σ известен, а параметр сдвига θ — нет. Эту модель часто применяют к данным, полученным при независимых измерениях некоторой величины θ с помощью прибора (или метода), имеющего известную среднюю погрешность (стандартную ошибку) σ (рис. 3).

Пусть $\Phi(x) = (2\pi)^{-1/2} \int\limits_{-\infty}^{x} e^{-u^2/2} du$ — функция распределения закона $\mathcal{N}(0, 1)$. Для $0 < \alpha < 1$ обозначим через x_α так называемую α-квантиль этого закона, т. е. решение уравнения $\Phi(x_\alpha) = \alpha$ (см. § 3 гл. 7). Приведем некоторые значения $x_{1-\alpha/2}$ (см. также таблицу Т2):

α	0,05	10^{-2}	10^{-3}	10^{-5}
$x_{1-\alpha/2}$	1,96	2,58	3,29	4,26

Согласно примеру 4 гл. 9, эффективной оценкой для θ служит $\overline{X}$. Известно, что $\overline{X} \sim \mathcal{N}(\theta, \sigma^2/n)$. Тогда $\sqrt{n}\,(\overline{X} - \theta)/\sigma \sim \mathcal{N}(0, 1)$. Поэтому в качестве границ интервала с коэффициентом доверия $1 - \alpha$ можно взять $\widehat{\theta}_1 = \overline{X} - \sigma x_{1-\alpha/2}/\sqrt{n}$ и $\widehat{\theta}_2 = \overline{X} - \sigma x_{\alpha/2}/\sqrt{n}$:

$$\mathsf{P}(\widehat{\theta}_1 < \theta < \widehat{\theta}_2) = \mathsf{P}(x_{\alpha/2} < \sqrt{n}\,(\overline{X} - \theta)/\sigma < x_{1-\alpha/2}) = 1 - \alpha.$$

В силу четности плотности закона $\mathcal{N}(0, 1)$ верно равенство $x_{\alpha/2} = -x_{1-\alpha/2}$. Таким образом, из приведенной выше таблицы видим, что с вероятностью 0,95 истинное значение параметра сдвига θ находится в интервале $\overline{X} \pm 1{,}96\,\sigma/\sqrt{n} \approx \overline{X} \pm 2\sigma/\sqrt{n}$ (**правило двух сигм**).

Определение. Пусть $\alpha \in (0, 1)$. Две статистики $\widehat{\theta}_1$ и $\widehat{\theta}_2$ определяют границы *доверительного интервала для параметра θ с коэффициентом доверия* $1 - \alpha$, если при всех $\theta \in \Theta$ для выборки $\boldsymbol{X} = (X_1, \ldots, X_n)$ из закона распределения $F_\theta(x)$ справедливо неравенство

$$\mathsf{P}\left(\widehat{\theta}_1(\boldsymbol{X}) < \theta < \widehat{\theta}_2(\boldsymbol{X})\right) \geqslant 1 - \alpha. \quad (1)$$

Часто на практике полагают $\alpha = 0{,}05$. Если вероятность в левой части неравенства (1) стремится к $1 - \alpha$ при $n \to \infty$, то интервал называется *асимптотическим.* Как правило, длина доверительного интервала возрастает при увеличении коэффициента доверия $1 - \alpha$ и стремится к нулю с ростом размера выборки n.

Статистический критерий – правило, позволяющее принять или отвергнуть гипотезу H на основе реализации выборки $x_1, \dots, x_n$.

или *Статистическим критерием (статистикой критерия) называют случайную величину, которая служит для проверки гипотезы.*

Статистика критерия – $T(x_1, \dots, x_n)$ – статистика (функция), для которой типично принимать умеренные значения в случае, когда гипотеза H верна, и большие (малые), когда H не выполняется.

Уровень значимости – α – вероятность, с которой мы можем позволить себе отвергнуть верную гипотезу (вероятность ошибочного отклонения правильной гипотезы,).

Схема

Берем $T(x_1, \dots, x_n)$ – статистика (какая-то функция),
$(x_1, \dots, x_n)$ – реализация выборки (данные эксперимента).
Делаем гипотезу H, выбираем приемлемый уровень значимости α.

Если значение T попало в область, имеющую при выполнении гипотезы H высокую вероятность, то можно заключить, что данные *согласуются* с гипотезой H. Отсюда происходит термин *«критерии согласия»*.

Если H – верна, то у нас есть определенные ожидания от значения T.
Мы находим $x_{1-\alpha}$, такое, что $\boldsymbol{P}(T(X_1, \dots, X_n) \geq x_{1-\alpha}) \leq \alpha$, то есть $x_{1-\alpha}$ – это максимальное значение для функции T такое, что её вероятность быть больше этого значения «равна» α (точнее не больше α, то есть маленькая). Потом вычисляем реальное $T(x_1, \dots, x_n) = t_0$ и смотрим, $t_0 < x_{1-\alpha}$? (что более ожидаемо при выполнении H) или нет?

Критическое значение – $x_{1-\alpha}$ – значение статистики критерия $T(x_1, \dots, x_n)$, при превышении которого мы должны отвергнуть гипотезу (так как по факту произошло маловероятное событие и наше предположение, наша гипотеза, скорее всего, не верна).

Фактический уровень значимости – $\alpha_0 = \boldsymbol{P}\big(T(X_1, \dots, X_n) \geq t_0 = T(x_1, \dots, x_n)\big) \leq \alpha$ – вероятность, с которой значение статистики может превысить ее фактическое значение на данной реализации случайной выборки $(x_1, \dots, x_n)$.

Статистической называют гипотезу о виде неизвестного распределения или о параметрах известного распределения.

Статистическая гипотеза называется **простой**, если она содержит только одно предположение, т.е. если ей соответствует одно распределение или одна точка пространства параметров. Гипотеза называется **сложной**, если она сводится к выбору какого-либо распределения из целого множества или выбору точки из интервала конечного или бесконечного.

Параметрическая гипотеза – гипотеза, в которой сформулированы предположения относительно значений параметров функции распределения известного вида **(гипотеза о параметре распределения)**

Непараметрическая гипотеза

Нулевой (основной) гипотезой называют выдвинутую гипотезу Н0, обычно утверждающую, что различие между сравниваемыми величинами отсутствует, а наблюдаемые отклонения объясняются лишь случайными колебаниями выборки.

Альтернативной (конкурирующей) гипотезой называют гипотезу Н1 , которая противоречит нулевой.

Простая параметрическая гипотеза – гипотеза, которая содержит только одно предположение относительно параметра (например, если a – математическое ожидание нормально распределенной случайной величины, то гипотеза H0: a = 0 – простая).

Сложная параметрическая гипотеза – гипотеза, которая состоит из конечного или бесконечного числа простых гипотез (например, если a – математическое ожидание нормально распределенной случайной величины, то гипотеза H0: a > 0,5 – сложная).

Статистическая гипотеза, Простая гипотеза, Сложная гипотеза

введем формально понятие статистической гипотезы.

Напомним, что под статистической моделью в § 1 гл. 6 понималось семейство функций распределения $\{F(x,\theta), \theta \in \Theta\}$, где Θ — множество возможных значений параметра. При этом данные $x_1,\ldots,x_n$ рассматривались как реализация выборки $X_1,\ldots,X_n$, элементы которой имеют функцию распределения $F(x,\theta_0)$ с неизвестным значением $\theta_0 \in \Theta$.

Пусть выделено некоторое подмножество $\Theta_0 \subset \Theta$. Под *статистической гипотезой* H понимается предположение о том, что $\theta_0 \in \Theta_0$. Если множество Θ_0 состоит всего из одной точки, то гипотеза H называется *простой*, иначе — *сложной*. В последнем случае задача заключается в проверке принадлежности закона распределения величин X_i целому классу функций распределения $\{F(x,\theta), \theta \in \Theta_0\}$.

Двусторонняя гипотеза

Односторонняя

Ошибка первого рода – «ложная тревога» (англ. type I errors, α-errors, false positive) – отказ от верной нулевой гипотезы.

Ошибка второго рода – «пропуск цели» (англ. type II errors, β-errors, false negative) – приятие неверной нулевой гипотезы. (Вероятность ошибки второго рода обозначается β, отсюда β-errors).

Критическая область - совокупность значений статистики критерия, при которых нулевую гипотезу отвергают.

Критические точки - точки, отделяющие критические области от области принятия гипотезы.

Область принятия гипотезы (область допустимых значений) - совокупность значений критерия, при которых нулевую гипотезу принимают.

Уровень значимости (еще раз) - вероятность совершить ошибку первого рода (α-error), т.е. отвергнуть верную нулевую гипотезу.

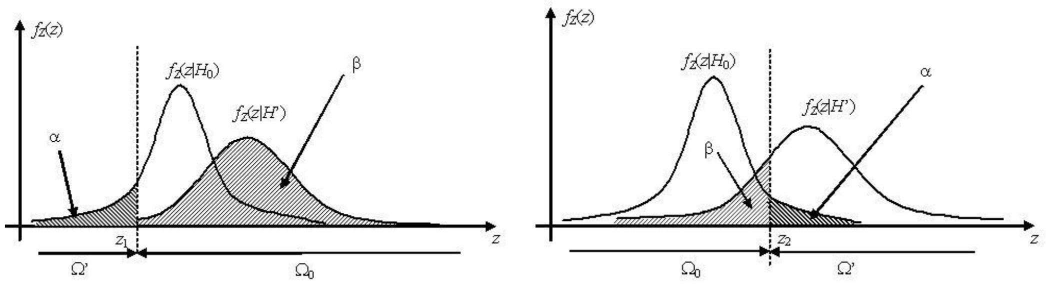
http://datalearning.ru/index.php/textbook?cid=1&mid=3&topic=0

Мощность критерия – вероятность **не** совершить ошибку второго рода $= 1 - \beta$. Вероятность $P(S|H_1) = 1 - P(\overline{S}|H_1) = 1 - \beta$ попадания точки (выборки) в критическое множество (критическую область S) при верной альтернативной гипотезе H1 = -H0. Таким образом, чем выше мощность, тем меньше вероятность принять неверную нулевую гипотезу (т.е. совершить ошибку второго рода).

Основным (хорошим) свойством статистического критерия должна быть минимальность ошибки второго рода при фиксированной ошибке первого рода. Так как $P(S|H_1) = 1 - P(\overline{S}|H_1) = 1 - \beta$, то при β_{min}, вероятность отвергнуть основную гипотезу, если верна конкурирующая, будет максимальной. Т.е. если $\max\limits_{S, P(S|H_0) \le \alpha} (1 - \beta) = \max\limits_{S, P(S|H_0) \le \alpha} P(S_0, H_1)$, то S_0 – **наилучшая критическая область** (НКО).

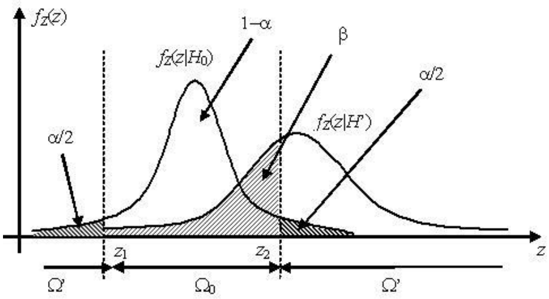
Наилучшая критическая область (область принятия решений).

Различные варианты выбора критической области



Наилучшей критической областью (*НКО*) называют критическую область, которая при заданном уровне значимости α обеспечивает минимальную вероятность β ошибки второго рода. Критерий, использующий наилучшую критическую область, имеет максимальную мощность.

Распределения вероятностей статистики критерия - Z при условии истинности основной и альтернативной гипотез



Статистическая значимость какого-либо значения
https://ru.wikipedia.org/wiki/%D0%A1%D1%82%D0%B0%D1%82%D0%B8%D1%81%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B0%D1%8F_%D0%B7%D0%BD%D0%B0%D1%87%D0%B8%D0%BC%D0%BE%D1%81%D1%82%D1%8C

Величину (значение) переменной называют **статистически значимой**, если мала вероятность случайного возникновения этой или ещё более крайних величин. Здесь под крайностью понимается степень отклонения тестовой статистики от нуль-гипотезы.

Разница называется **статистически значимой**, если появление имеющихся данных (или ещё более крайних данных) было бы маловероятно, если предположить, что эта разница отсутствует; это выражение не означает, что данная разница должна быть велика, важна, или значима в общем смысле этого слова.

Непараметрические критерии *не опираются* на дополнительные предположения о распределении.

Непараметрические критерии – статистические критерии, которые не включают в расчёт параметры вероятностного распределения и основаны на оперировании частотами или рангами.

- Q-критерий Розенбаума
- U-критерий Манна — Уитни
- Критерий Уилкоксона
- Критерий Пирсона
- Критерий Колмогорова — Смирнова

Параметрические критерии *предполагают*, что выборка порождена распределением из заданного параметрического семейства

Параметрические критерии – статистические критерии, которые включают в расчет параметры вероятностного распределения признака (средние и дисперсии).

- t-критерий Стьюдента
- Критерий Фишера
- Критерий отношения правдоподобия
- Критерий Романовского

Критерии значимости: проверка на значимость предполагает проверку гипотезы о численных значениях известного закона распределения: $H_0: a = a_0$ — нулевая гипотеза. $H_1: a > a_0$ $(a < a_0)$ или $a \neq a_0$ — конкурирующая гипотеза.

Критерии согласия – критерии проверки гипотез о типе распределения генеральной совокупности (о соответствии эмпирического распределения теоретическому закону распределения).

Критерии проверки на однородность: при проверке на однородность случайные величины исследуются на факт значимости различия их законов распределения (т.е. проверки того, подчиняются ли эти величины одному и тому же закону). Используются в факторном анализе для определения наличия зависимостей.

Критерии значимости

Обратите внимание на величину разности между $\bar{X}$ и m0 и на то, как это повлияло на результаты проверки гипотез, а также на то, что в некоторых случаях результат проверки основной гипотезы зависит от вида альтернативной.

Критерий (при уровне значимости α):

- против альтернативы H_1: $\bar{x} \neq \mu$
  если $|t| > t_{\alpha/2}$, то нулевая гипотеза отвергается;
- против альтернативы H_1': $\bar{x} < \mu$
  если $t < t_{\alpha}$, то нулевая гипотеза отвергается;
- против альтернативы H_1'': $\bar{x} > \mu$
  если $t > t_{1-\alpha}$, то нулевая гипотеза отвергается;

где t_{α} есть α-квантиль распределения Стьюдента с $m-1$ степенями свободы.

Как зависит частость ошибки второго рода от вероятности ошибки первого рода (α)?

Критерий Стьюдента

Критерий Фишера

Критерии согласия

Критерии согласия – критерии проверки гипотез о типе распределения генеральной совокупности (о соответствии эмпирического распределения теоретическому закону распределения).

1. Критерий Пирсона
2. Критерий Колмогорова
3. Критерий Андерсона — Дарлинга
4. Критерий Крамера — Мизеса — Смирнова
5. Критерий согласия Купера
6. Z-тест
7. Тест Харке — Бера
8. Критерий Шапиро — Уилка (англ.)
9. График нормальности (англ.)

Общие критерии согласия – применимы к самой общей формулировке гипотезы, а именно к гипотезе о согласии наблюдаемых результатов с любым априорно предполагаемым распределением вероятностей.

Специальные критерии согласия – предполагают специальные нулевые гипотезы, формулирующие согласие с определенной формой распределения вероятностей.

Критериями согласия называют критерии, предназначенные для проверки простой гипотезы $H\_1 = \{F = F\_1\}$ при сложной альтернативе $H\_2 = \{H\_1 - \text{неверна}\}$. Мы рассмотрим более широкий класс основных гипотез, включающий и сложные гипотезы, а критериями согласия будем называть любые критерии, устроенные по одному и тому же принципу. А именно, пусть задана некоторая ***функция отклонения*** эмпирического распределения от теоретического, распределение которой существенно разнится в зависимости от того, верна или нет основная гипотеза. Критерии согласия принимают или отвергают основную гипотезу исходя из величины этой функции отклонения.

Глазомерный метод проверки нормальности

в основе этого метода лежит тот факт, что случайная величина $F_X^{-1}(X)$ имеет равномерное распределение на отрезке [0;1]; подумайте, как это использовать

Как проверить сложную двухпараметрическую *гипотезу нормальности* о том, что выборка была взята из совокупности с функцией распределения $F(x,\mu,\sigma) = \Phi\left(\frac{x-\mu}{\sigma}\right)$ с какими-то неизвестными параметрами μ и $\sigma > 0$? (Здесь, как обычно, $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-u^2/2} du$ — функция распределения стандартного нормального закона.)

Прежде чем применять критерии, полезно посмотреть на данные на вероятностной бумаге (см. § 1 гл. 9): если точки $\left(x_{(i)}, \Phi^{-1}\left(\frac{i-0{,}5}{n}\right)\right)$ не расположены вблизи некоторой прямой, то гипотеза нормальности скорее всего ошибочна.

Критерий Шапиро-Уилка проверки нормальности: сильный критерий

http://www.machinelearning.ru/wiki/index.php?title=Критерий_Шапиро-Уилка

Критерий Шапиро-Уилка основан на оптимальной линейной несмещённой оценке дисперсии к её обычной оценке методом максимального правдоподобия.

Статистика критерия имеет вид:

$$W=\frac{1}{s^2}\left[\sum_{i=1}^{n}a_{n-i+1}(x_{n-i+1}-x_i)\right]^2, \quad \text{где} \quad s^2=\sum_{i=1}^{n}(x_i-\bar{x})^2, \bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i.$$

Числитель является квадратом оценки среднеквадратического отклонения Ллойда. Коэффициенты a_{n-i+1} *берутся из таблиц. Критические значения статистики* $W(\alpha)$ *также находятся таблично. Если* $W<W(\alpha)$*, то нулевая гипотеза о нормальности распределения отклоняется при уровне значимости* α*.. Критерий Шапиро-Уилка является очень мощным критерием для проверки нормальности, но, к сожалению, при больших значениях* n $(n>100)$ *таблицы коэффициентов* a_{n-i+1} *становятся неудобными.*

Критерий Колмогорова: (пример задачи на странице 175)

Функция отклонения $\rho(\mathbf{X})=\sqrt{n}\sup_y|F_n^*(y)-F_1(y)|.$ $D_n=\sup_{-\infty<x<\infty}|\widehat{F}_n(x)-F(x)|,$

n>50 Таблица *Критерий согласия Колмогорова (λ) определяется путем деления модуля **max** разности между эмпирическими и теоретическими кумулятивными частотами на корень квадратный из числа наблюдений:* $\lambda=\frac{D}{\sqrt{n}}=\frac{|55-61,63|}{\sqrt{125}}=\frac{6.63}{\sqrt{125}}=0.59$

По специальной таблице вероятности для критерия согласия λ определяем, что значению λ=0,59 соответствует вероятность 0,88 (λ).

Критерий согласия Пирсона χ^2

Критерий согласия Пирсона используется, если объем совокупности достаточно велик (N>50), при этом, частота каждой группы должна быть не менее 5.

Критерий согласия Пирсона χ^2 – один из основных, который можно представить как сумму отношений квадратов расхождений между теоретическими (f_T) и эмпирическими (f) частотами к теоретическим частотам:

$$\chi^2=\sum_{i=1}^{k}\frac{(f_i-f_T)^2}{f_T}$$

- k – число групп, на которые разбито эмпирическое распределение,
- f_i – наблюдаемая частота признака в i-й группе,
- f_T – теоретическая частота.

При вычислении теоретических частот их параметры вычисляются по выборке (например, μ и σ в норм. распр. используются выборочное среднее и выборочное среднее квадратическое отклонение).

Заметим, что малочисленные частоты (n_i < 5) следует объединить. В этом случае и соответствующие им теоретические частоты тоже надо сложить. При условии объединения частот, для определения числа степеней свободы следует взять число групп выборки, оставшихся после объединения частот.

Для распределения χ^2 составлены таблицы, где указано критическое значение критерия согласия χ^2 для выбранного уровня значимости α и степеней свободы df (или v).

Число степеней свободы df определяется как число групп в ряду распределения минус число связей: $df = k - z$. Под числом связей понимается число показателей эмпирического ряда, использованных при вычислении теоретических частот, т.е. показателей, связывающих эмпирические и теоретические частоты. Например, при выравнивании по кривой нормального распределения имеется три связи. Поэтому при выравнивании по кривой нормального распределения число степеней свободы определяется как $df = k - 3$. Для оценки существенности, расчетное значение сравнивается с табличным $\chi^2_{табл.}$

http://www.machinelearning.ru/wiki/index.php?title=Критерий_хи-квадрат

Критерий χ^2 - статистический критерий для проверки гипотезы H_0, что наблюдаемая случайная величина подчиняется некому теоретическому закону распределения.

Определение

Пусть дана случайная величина X .

Гипотеза H_0: с. в. X подчиняется закону распределения $F(x)$.

Для проверки гипотезы рассмотрим выборку, состоящую из n независимых наблюдений над с.в. X: $X^n = (x_1, \cdots x_n)$, $x_i \in [a,b]$, $\forall i = 1 \cdots n$. По выборке построим эмпирическое распределение $F^*(x)$ с.в X. Сравнение эмпирического $F^*(x)$ и теоретического распределения $F(x)$ (предполагаемого в гипотезе) производится с помощью специально подобранной функции — критерия согласия. Рассмотрим критерий согласия Пирсона (критерий χ^2):

Гипотеза H_0^*: X^n порождается функцией $F^*(x)$.

Разделим [a,b] на k непересекающихся интервалов $(a_i, b_i]$, $i = 1 \cdots k$;

Пусть n_j - количество наблюдений в j-м интервале: $n_j = \sum_{i=1}^{n} [a_j < x_i \le b_j]$;

$p_j = F(b_j) - F(a_j)$ - вероятность попадания наблюдения в j-ый интервал при выполнении гипотезы H_0^*;

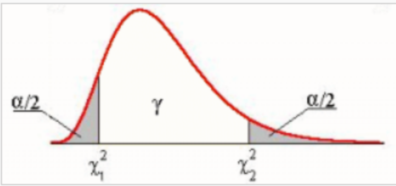
$E_j = n p_j$ - ожидаемое число попаданий в j-ый интервал;

Статистика: $\chi^2 = \sum_{j=1}^{k} \frac{(n_j - E_j)^2}{E_j} \sim \chi^2_{k-1}$ - Распределение хи-квадрат с k-1 степенью свободы.

Проверка гипотезы H_0

В зависимости от значения критерия χ^2, гипотеза H_0 может приниматься, либо отвергаться:

- $\chi^2_1 < \chi^2 < \chi^2_2$, гипотеза H_0 выполняется.
- $\chi^2 \le \chi^2_1$ (попадает в левый "хвост" распределения). Следовательно, теоретические и практические значения очень близки. Если, к примеру, происходит проверка генератора случайных чисел, который сгенерировал n чисел из отрезка [0,1] и гипотеза H_0: выборка X^n распределена равномерно на [0,1], тогда генератор нельзя называть случайным (гипотеза случайности не выполняется), т.к. выборка распределена слишком равномерно, но гипотеза H_0 выполняется.
- $\chi^2 \ge \chi^2_2$ (попадает в правый "хвост" распределения) гипотеза H_0 отвергается.



Распределение хи-квадрат

Задача 2 показывает, что, как правило, *нельзя надежно проверить сложную гипотезу по небольшой выборке* (состоящей из нескольких десятков наблюдений): критерии улавливают только очень крупные отклонения, так как за счет варьирования параметра (параметров) обычно удается достаточно хорошо подогнать $F(x,\theta)$ к эмпирической функции распределения $\widehat{F}_n(x)$.

С другой стороны, для выборок большого размера (порядка нескольких сотен наблюдений) трудно гарантировать одинаковость условий при сборе данных (однородность наблюдений).

По-видимому, хороший способ разрешения этой проблемы — использование статистических методов, не предполагающих строгую нормальность наблюдений, для которых требуется лишь непрерывность функции распределения элементов выборки. Именно такие критерии рассматриваются в гл. 14–17.

Проверить нормальность по сгруппированным данным можно

стр170 также при помощи критерия хи-квадрат (см. гл. 18).

Критерии однородности

Критерии проверки на однородность: при проверке на однородность случайные величины исследуются на факт значимости различия их законов распределения (т.е. проверки того, подчиняются ли эти величины одному и тому же закону). Используются в факторном анализе для определения наличия зависимостей.

Парные (зависимые) наблюдения – две выборки представляют собой пары наблюдений (измерений) одного и того же объекта до и после какого-либо воздействия на него. Данные должны быть организованы в виде двух столбцов. В одном столбце содержатся «первые» измерение каждого объекта, во втором столбце - «второе» измерение пары.

Предполагается, что воздействие может повлиять на признаки, сдвинув их средние значения в большую или в меньшую сторону, и это необходимо проверить. Вначале признаки объектов принимают значения x_i ,после воздействия – значения y_i .Такие наблюдения называют парными. Вычисляем их разности $d_i = y_i - x_i$. По наблюдениям d_1, d_2, ..., d_n проверяем гипотезу о равенстве нулю генерального среднего (H0:a_d=0) при неизвестной дисперсии σ_d^2 . При этом предполагают, что случайные изменения признаков распределены нормально.

Независимые (непарные) наблюдения

Двухвыборочный критерий Стьюдента (t–тест) для проверки равенства средних значений двух выборок из нормального распределения

http://www.machinelearning.ru/wiki/index.php?title=Критерий_Стьюдента

Двухвыборочный критерий Фишера (F–тест) для проверки равенства дисперсий двух независимых выборок из нормального распределения

http://www.machinelearning.ru/wiki/index.php?title=Критерий_Фишера

Слишком маленький объем выборки (например, 10) для того чтобы говорить о проверки нормальности

Если у нас есть сомнения в нормальности рассматриваемых величин, особенно если мы работаем с малыми выборками можно воспользоваться одним из непараметрических критериев — критерием Уилкоксона (Манна-Уитни).

Критерий суммы рангов.

W-test критерий Уилкоксона — (критерий знаковых рангов Уилкоксона, критерий суммы рангов Уилкоксона) непараметрический статистический тест (критерий), используемый для проверки различий между двумя выборками парных или независимых измерений по уровню какого-либо количественного признака, измеренного в непрерывной или в порядковой шкале. Уверенно критерий Уилкоксона можно использовать при объёме выборки до 25 элементов.

Тест Уилкоксона для независимых выборок также называется критерием Манна-Уитни.

Критерий Уилкоксона (Манна-Уитни) U-test **для проверки равенства медиан**

Этот метод определяет, достаточно ли мала зона перекрещивающихся значений между двумя рядами (ранжированным рядом значений параметра в первой выборке и таким же во второй выборке). Чем меньше значение критерия, тем вероятнее, что различия между значениями параметра в выборках достоверны. Ограничения применимости критерия: 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 (до 10).

Критерий Колмогорова-Смирнова для проверки однородности

Парные наблюдения

Критерий Стьюдента

Критерий знаков: используется при проверке нулевой гипотезы о равенстве медианы некоторому заданному значению (для одной выборки) или о равенстве нулю медианы разности (для двух **связанных** выборок). Это непараметрический критерий, то есть он не использует никаких данных о характере распределения, и может применяться в широком спектре ситуаций, однако при этом он может иметь меньшую мощность, чем более специализированные критерии.

Критерий знаковых ранговых сумм

Сравнение нескольких выборок

Таблица сопряженности – таблица частот (?)

Методика	Оценка			
	отлично	хорошо	удовлетвори-тельно	неудовлетвори-тельно
1	7	12	15	7
2	10	18	26	9
3	40	50	87	10

Гипотеза о независимости данных от методики. Гипотеза об однородности выборок.

НЕПАРАМЕТРИЧЕСКИЕ КРИТЕРИИ ПРОВЕРКИ ОДНОРОДНОСТИ.

Критерий Краскела–Уоллиса

Дисперсионный анализ **ANalysis** Of **VAriance** (**ANOVA**)

ОДНОФАКТОРНЫЙ ДИСПЕРСИОННЫЙ АНАЛИЗ Однофакторный дисперсионный анализ применяется для обнаружения влияния выделенного (контролируемого) набора факторов на результативный признак в случае, если величины имеют нормальное распределение с общей для всех дисперсией, которая нам неизвестна.

Используя однофакторный дисперсионный анализ, ответить на поставленный вопрос. Проверьте законность применения однофакторного дисперсионного анализа с помощью критерия Кочрена. Уровень значимости принять равным 0,05 (5 %).

Если гипотезу о равенстве дисперсий следует отклонить, применение однофакторного дисперсионного анализа нельзя считать законным.