Лекция 9

Проверка статистических гипотез

Проверка статистических гипотез. Основные определения

- ***Статистической гипотезой*** называется всякое высказывание о генеральной совокупности, проверяемое по выборке.
- Статистические гипотезы делятся на:
  - ***параметрические*** - это гипотезы, сформулированные относительно параметров (среднего значения, дисперсии и т. д.) распределения известного вида;
  - ***непараметрические*** - это гипотезы, сформулированные относительно вида распределения (например, определение по выборке степени нормальности генеральной совокупности).
- Процесс использования выборки для проверки гипотезы называется *статистическим доказательством*.
- Основную выдвигаемую гипотезу называют *нулевой* H_0.
- Наряду с нулевой гипотезой рассматривают *альтернативную* ей H_1.

Например: H_0: $M(X)=1$, математическое ожидание генеральной совокупности равно 1;

H_1: $M(X)>1$, или $M(X)<1$, или $M(X) \neq 1$ (математическое ожидание больше 1, или меньше 1, или не равно 1).

Проверка статистических гипотез. Ошибки первого и второго родов

- Выбор между гипотезами H_0 и H_1 может сопровождаться ошибками двух родов.
- Ошибка *первого рода α* означает вероятность принятия H_1, если верна гипотеза H_0: $\alpha = p(H_1/H_0)$.
- Ошибка *второго рода β* означает вероятность принятия H_0, если верна гипотеза H_1: $\beta = p(H_0/H_1)$.
- Существует правильное решение двух видов:

$$p(H_0/H_0) = 1 - \alpha \quad \text{и} \quad p(H_1/H_1) = 1 - \beta$$

Принятая гипотеза	H_0	H_1
H_0 -верна	$P(H_0/H_0) = 1 - \alpha$	$P(H_1/H_0) = \alpha$
H_0 - неверна	$P(H_0/H_1) = \beta$	$P(H_1/H_1) = 1 - \beta$

Проверка статистических гипотез. Статистический критерий

- Правило, по которому принимается решение о том, верна или не верна гипотеза H_0, называется *критерием,* где:

  $\alpha = P(H_1/H_0)$ - *уровень значимости критерия;*

  $M = 1 - \beta = P(H_1/H_1)$ - *мощность критерия.*

- ***Статистическим критерием*** K называют случайную величину, с помощью которой принимают решение о принятии или отклонении H_0.
- Для проверки *параметрических* гипотез используют *критерии значимости,* основанные на статистиках: u, χ^2, t, F .
- *Непараметрические* гипотезы проверяют с помощью *критериев согласия,* использующих статистики распределений: χ^2, Колмогорова-Смирнова и т. д.

Проверка статистических гипотез. Статистический критерий

- Например, H_0: $M(X)$=10. В зависимости от альтернативной гипотезы рассматривают три случая.

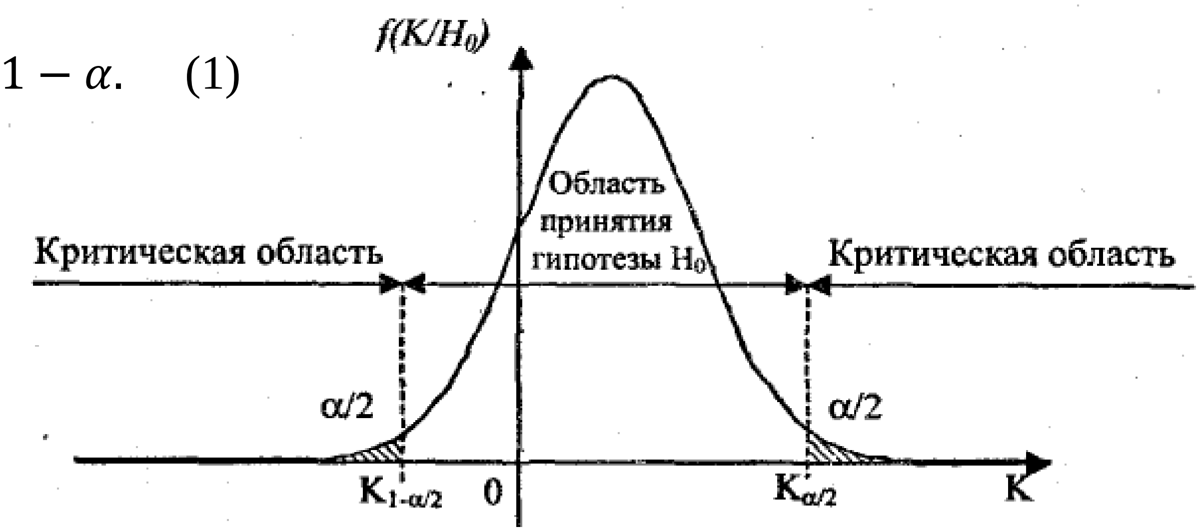
1) Если H_1 : $M(X) \neq 10$.

В этом случае рассматривают двустороннюю критическую область и используют дифференциальную функцию $f(K/H_0)$, для определения соответствующих квантилей (границ области принятия гипотезы - левой ($K_{1-\alpha/2}$) и правой ($K_{\alpha/2}$)).

Площадь под криволинейной трапецией дифференциальной функции слева от $K_{1-\alpha/2}$ и справа от $K_{\alpha/2}$ равна α/2.

Общая площадь ограниченная криволинейной трапецией дифференциальной функции, квантилями и осью абсцисс, равна (1 - α)

$$P\left(K_{1-\alpha/_2} < К < К_{\alpha/_2}\right) = 1 - \alpha. \quad (1)$$

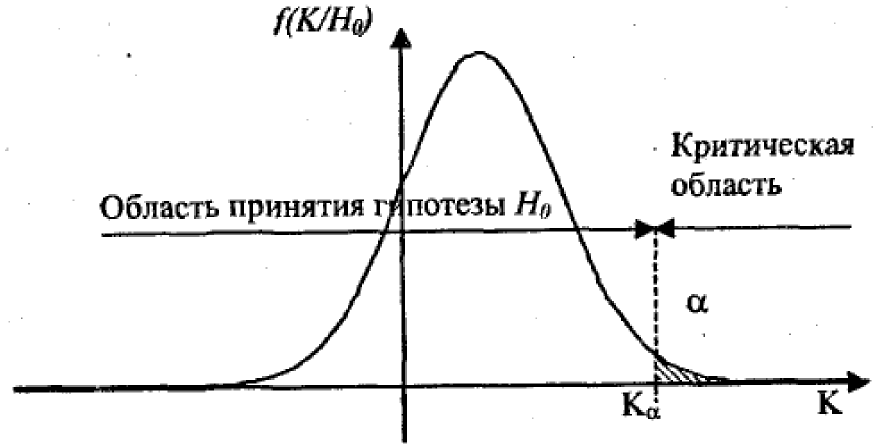


Двусторонняя критическая область

Проверка статистических гипотез. Статистический критерий

2) Если H_1:*M(X)*> 10, то рассматривается правосторонняя критическая область (площадь под криволинейной трапецией справа от K_α равна α):

$$P(K > K_{\alpha}) = \int_{K_\alpha}^{+\infty} f\left(\frac{K}{H_0}\right) dK = \alpha. \quad (2)$$



Правосторонняя критическая область

3) Если H_0: *M(X)<10,* то рассматривается левосторонняя критическая область (площадь под криволинейной трапецией слева от $K_{1-\alpha}$ равна α):

$$\text{P}(\text{K} < \text{K}_{1\_\alpha}) = \int_{-\infty}^{K_{1-\alpha}} f(K/H_0)\, dK = \alpha. \, (3)$$

Левосторонняя критическая область

Проверка статистических гипотез. Алгоритм

1. Располагая выборочными данными $(x_1, x_2, ..., x_n)$, формируют нулевую гипотезу H_0 и конкурирующую гипотезу H_1.
2. Задают уровень значимости α (обычно принимают α =0,1; 0,01; 0,05; 0,001).
3. Рассматривается выборочная статистика наблюдений (критерий) К, обычно одна из перечисленных ниже:

   u - нормальное распределение;

   χ^2- распределение Пирсона (хи - квадрат);

   t - распределение Стьюдента;

   F - распределение Фишера - Снедекора.
4. На основании выборки $x_1, x_2, ..., x_n$ - определяют значение критерия (статистики) K. В зависимости от вида альтернативной гипотезы выбирают по соответствующей таблице квантили критерия для двусторонней $\left(K_{1-\frac{\alpha}{2}}\text{ и } K_{\frac{\alpha}{2}}\right)$ или односторонней области ($K_{1-\alpha}$ или K_{α}).

Если значения критерия попадают в критическую область, то H_0 отвергается; в противном случае принимается гипотеза H_0 и считается, что H_0 не противоречит выборочным данным (при этом существует возможность ошибки с вероятностью, равной α).

Проверка статистических гипотез. Алгоритм

- **Пример.** Два сорта озимой пшеницы испытывались на одинаковом количестве участков на протяжении семи лет (табл.) При уровне значимости $a = 0{,}05$ проверить нулевую гипотезу о существенности различий в урожайности двух сортов озимой пшеницы.

Годы	Урожайность, ц/га	
	x_{2i}	x_{1i}
1995	47	53
1996	48	43
1997	46	45
1998	51	56
1999	52	58
2000	48	55
2001	52	59

Решение. Так как имеются две зависимости выборки, т.е. существует определенная корреляция между урожайностью сортов по годам, то необходимо оценить значимость не разности двух выборочных средних, а средней разности.

Выдвигаем нулевую гипотезу: средняя величина различий в урожайности пшеницы равна нулю, H_0: $\overline{X_1 - X_2} = 0$ при H_1: $\overline{X_1 - X_2} = 0$.

Проверка статистических гипотез. Алгоритм

- По данным таблицы найдем среднюю разность $\bar{d}$ и ошибку средней разности $S_{\bar{d}}$:

$$\bar{d} = \frac{\sum d_i}{n} = \frac{35}{7} = 5; \quad S_{\bar{d}} = \sqrt{\frac{\sum\left(d_i - \bar{d}\right)^2}{n(n-1)}} = \sqrt{\frac{46}{7(7-1)}} = 1{,}047;$$

где $d_i = x_{1i} - x_{2i}$, n - число пар наблюдений. Для двусторонней критической области при α=0,05; $k = n - 1 = 7 - 1 = 6$, $t_{кр.} = 2{,}45$.

- Сопоставив расчётное значение t с критическим, можно сделать вывод, что два сорта существенно различаются по уровню урожайности.

Вспомогательная таблица для расчета ошибки средней разности

Годы	Урожайность, ц/га		Разность $d_i = x_{1i} - x_{2i}$	$\left(d_i - \bar{d}\right)$	$\left(d_i - \bar{d}\right)^2$
	x_{2i}	x_{1i}			
1995	47	53	6	1	1
1996	48	43	5	0	0
1997	46	45	-1	-6	36
1998	51	56	5	0	0
1999	52	58	6	1	1
2000	48	55	7	2	4
2001	52	59	7	2	4
Сумма			35	0	46

Критерий Шапиро-Уилкса (Shapiro–Wilks): проверка на нормальность

Этот критерий применяется для проверки на нормальность, когда число вариант выборки мало $(3 \le n \le 50)$.

Пусть $(x_1, x_2, \ldots, x_n)$ – случайная выборка из распределения $F(x)$. Вычислим

$$S^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2,$$

где $\bar{x}$ - выборочное среднее, и найдем вариационный ряд $(x_{(1)}, x_{(2)}, \ldots, x_{(n)})$. Если n-четное число, принимаем $k = n/2$, в противном случае $k = (n - 1)/2$. Затем вычислим

$$b = \sum_{i=1}^{k} a_{n-i+1}(x_{(n-i+1)} - x_{(i)}),$$

где коэффициенты a_{n-i+1} для $i=1,2,\ldots,k$ берутся из соответствующей таблицы Шапиро-Уилкса.

Проверка статистических гипотез. Критерии

Сравним значение статистики критерия, вычисляемое по формуле

$$W = b^2 / S^2,$$

с критическим значением $W_{кр}$. Если $W > W_{кр}$, то распределение можно считать нормальным.

Критерий Стьюдента используется для сравнения средних значений двух нормальных выборок.

Пусть $x_1, x_2, \ldots, x_n$ и $y_1, y_2, \ldots, y_m$ – нормальные независимые выборки из законов распределения с параметрами $(a_1; \sigma_1^2)$ и $(a_2; \sigma_2^2)$ соответственно (где a – среднее, σ^2 – дисперсия).

Проверим гипотезу H_0: $a_1 = a_2$.
Заметим, что более общий случай H_0: $a_1 = a_2 + \Delta$, где Δ - заданное число, сводится к предыдущему путем преобразования выборки $y_1, \ldots, y_m$ в выборку $y_1 + \Delta, \ldots, y_m + \Delta$.

Проверка статистических гипотез. Критерии

Относительно параметров σ_1^2 и σ_2^2 выделим следующие четыре варианта предположений:

1) обе дисперсии известны и равны между собой;

2) обе дисперсии известны, но не равны между собой;

3) обе дисперсии неизвестны, но предполагается, что они равны между собой;

4) обе дисперсии неизвестны, их равенство не предполагается.

- Для построения критерия проверки гипотезы H_0 проведем следующие рассуждения. От выборок $x_1, x_2, \dots, x_n$ и $y_1, y_2, \dots, y_m$ перейдем к выборочным средним $\bar{x}$ и $\bar{y}$.
- Согласно свойствам нормального распределения и выдвинутой гипотезе, выборочные средние имеют нормальное распределение с одним и тем же средним и дисперсиями σ_1^2/n и σ_2^2/m.

Проверка статистических гипотез. Критерии

- Далее перейдем к статистике, основанной на выборочных средних и дисперсиях σ_1^2, σ_2^2 (если они известны) или их оценках s_1^2, s_2^2 (если дисперсии неизвестны). Статистику выбираем так, чтобы ее распределение не зависело от неизвестных нам значений параметра.
- Наиболее естественными статистиками для перечисленных выше случаев будут следующие:

1. $$\frac{\bar{x}-\bar{y}}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}}.$$

Эта статистика имеет стандартное нормальное распределение, так как является линейной комбинацией независимых нормальных величин.

Гипотеза H_0 принимается на уровне значимости α, если

$$\left|\frac{\bar{x}-\bar{y}}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}}\right| < z_\alpha;$$

В противном случае гипотеза отвергается в пользу альтернативы $a_1 \neq a_2$

Проверка статистических гипотез. Критерии

2. $\dfrac{\bar{x}-\bar{y}}{\sqrt{\sigma_1^2/n+\sigma_2^2/m}}$. Эта статистика имеет также стандартное нормальное распределение.

Правило принятия гипотезы аналогично правилу пункта 1.

3. Так как обе дисперсии неизвестны, но предполагаются равными между собой, целесообразно перейдем к объединенной оценке дисперсии:

$$s^2 = \frac{s_1^2(n-1)+s_2^2(m-1)}{(n-1)+(m-1)}.$$

Оценка s^2 имеет распределение χ^2 с (n+m-2) степенями свободы.

Проверка статистических гипотез. Критерии

- Критерий для проверки гипотезы H_0 опирается на статистику

$$\frac{\bar{x}-\bar{y}}{s\sqrt{\frac{1}{n}+\frac{1}{m}}},$$

которая имеет распределение Стьюдента с (n+m-2) степенями свободы.

Если значение этой статистики меньше табличного значения t_α, гипотезу H_0 принимаем.

4. В случае неизвестных дисперсий, равенство которых не предполагается, используется аналог статистики пункта 2 с заменой неизвестных дисперсий их оценками:

$$\frac{\bar{x}-\bar{y}}{\sqrt{\frac{s_1^2}{n}+\frac{s_2^2}{m}}}.$$

Проверка статистических гипотез. Критерии

- В этой ситуации указать точное распределение введенной статистики затруднительно.

- Известно, однако, что это распределение близко к распределению Стьюдента с числом степеней свободы, равным
$$\frac{(s_1^2/n+s_2^2/m)^2}{\frac{(s_1^2/n)^2}{n-1}+\frac{(s_2^2/m)^2}{m-1}}.$$
- Критерий проверки гипотезы устроен так же, как и в пункте 3.

Проверка статистических гипотез. Критерии

Критерий Фишера используется для проверки гипотезы о равенстве дисперсий двух нормальных выборок

(H_0: $\sigma_1^2 = \sigma_2^2$).

Рассмотрим отношение оценок дисперсий

$$F = {s_1^2}/{s_2^2},$$

называемое *дисперсионным отношением Фишера* или *статистикой Фишера*. Эта величина имеет F-распределение с числом степеней свободы

(n -1;m -1).

Если $F<F_{кр}$ на заданном уровне значимости α, то нулевую гипотезу о равенстве дисперсий двух нормальных выборок принимаем.

Проверка статистических гипотез. Критерии

Критерий согласия Пирсона (хи-квадрат - χ^2) используется для проверки гипотезы о законе распределения.

Требуется проверить гипотезу H_0 о том, что исследуемая случайная величина X подчиняется закону распределения $F(x)$.

- Для проверки гипотезы произведём выборку, состоящую из n независимых наблюдений над случайной величиной X. Область изменения значений выборки разобьем на k интервалов.
- По выборке можно построить эмпирическое распределение $F^*(x)$ исследуемой случайной величины. Сравнение эмпирического и теоретического распределений производится с помощью критерия согласия Пирсона.
- Для проверки критерия вычисляется статистика:

$$\chi^2 = \sum_{i=1}^{l} \frac{(m_i - np_i)^2}{np_i},$$

где m_i – эмпирическое число значений случайной величины, попавшей в i-й интервал ($i=1,\ldots,n$); np_i - теоретическая частота.

Проверка статистических гипотез. Критерии

Эта статистика обладает распределением Пирсона с (k-r-1) степенями свободы (где r – число оцениваемых параметров распределения, рассчитанных по выборке).

- Правило применения критерия сводится к следующему. Рассчитав значения χ^2 и выбрав уровень значимости критерия α, по таблице распределения Пирсона определяют $\chi_{кр}{}^2$. Если $\chi^2 > \chi_{кр}{}^2$, то гипотезу H_0 отвергают.

- Следует отметить, что необходимым условием применения критерия Пирсона является наличие в каждом из интервалов по меньшей мере 10 наблюдений. Если количество наблюдений в отдельных интервалах мало, то следует объединить некоторые интервалы.

Проверка статистических гипотез. Критерии

Пример. Взято 100 проб сточной жидкости и подсчитаны отклонения значений концентраций загрязняющих веществ в жидкости от норм ПДК.

Проверить при уровне значимости α=0,01 гипотезу H_0 о том, что отклонения от норм ПДК подчиняются нормальному закону распределения.

- **Решение**. Случайную величину (отклонения от норм ПДК) обозначим через X. Имеем x_{min}= -2,5, x_{max}= 5.

Область изменения случайной величины X равна 5-(-2,5)=7,5. В соответствии с формулой Стэрджеса оптимальная длина интервала

$$h=(x_{max}-x_{min})/(1+3{,}322\ln(n))=1.$$

Вариационный ряд приведем в виде следующей таблицы:

xi÷xi+1	-3÷-2	-2÷-1	-1÷0	0÷1	1÷2	2÷3	3÷4	4÷5
mi	3	10	15	24	25	13	7	3

Так как в крайних интервалах число наблюдений меньше 10, эти интервалы следует объединить.

Проверка статистических гипотез. Критерии

В результате получим таблицу:

$x_i \div x_{i+1}$	-3÷-1	-1÷0	0÷1	1÷2	2÷3	3÷5
m_i	13	15	24	25	13	10

Далее вычисляем вероятности p_i, предварительно вычислив параметры, определяющие закон нормального распределения: $\bar{x} = 0{,}6;\ s = 1{,}6.$

$$p_1 = P(-3 < x < -1) = P\left[\frac{-3-0{,}6}{1{,}6} < \frac{X-\bar{x}}{\sigma} < \frac{-1-0{,}6}{1{,}6}\right] =$$

$$= \tfrac{1}{2}\Phi(-1) - \tfrac{1}{2}\Phi(-2{,}25) = 0{,}4880 - 0{,}3415 = 0{,}1465.$$

Аналогично вычисляются p_2, p_3, p_4, p_5, p_6.

Проверка статистических гипотез. Критерии

- Полученные результаты приведем в виде таблицы.

$x_i \div x_{i+1}$	-3÷-1	-1÷0	0÷1	1÷2	2÷3	3÷5
m_i	13	15	24	25	13	10
np_i	14,64	19,33	24,67	21,19	12,26	6,38

- Вычислим статистику χ^2:

$$\chi^2 = \frac{(13-14{,}64)^2}{14{,}64} + \frac{(15-19{,}33)^2}{19{,}33} + \frac{(24+24{,}67)^2}{24{,}67} + \frac{(25-21{,}19)^2}{21{,}19} +$$

$$+ \frac{(13-12{,}26)^2}{12{,}26} + \frac{(10-6{,}38)^2}{6{,}38} = 5{,}53.$$

Количество интервалов равно k=6. По выборке рассчитаны два параметра, r=2. Число степеней свободы равно (k-r-1)=3. По соответствующей таблице находим $\chi_{кр}^2 = 11{,}3$.

Имеем 11,3>5,53. Следовательно, нет оснований отвергать проверяемую гипотезу, т.е. отклонения от норм ПДК подчиняются нормальному закону распределения.

Проверка статистических гипотез. Критерии

Критерий согласия Колмогорова–Смирнова оценивает близость фактического распределения к теоретическому путем нахождения величины D, т.е. максимальной разности накопленных частостей фактического и теоретического распределений.

$$D=\max|F_n(x)-F(x)| \text{ при } -\infty<x<\infty,$$

где $F_n(x)$-эмпирическая (выборочная) функция распределения;

$F(x)$ – гипотетическая функция распределения;

n – объем выборки.

Гипотеза о том, что истинная функция распределения есть $F(x)$, записывается в виде: H_0: $G(x)=F(x)$.

Если полученное значение статистики $D<D_{кр}$, то гипотезу принимаем, т.е. расхождение между фактическим распределением и теоретическим не существенно.

Проверка статистических гипотез. Критерии

Критерий Бартлета – самый мощный из критериев, привлекаемых для проверки нулевой гипотезы, называемой в данном случае гипотезой об однородности дисперсий.

- С помощью этого критерия оценивается существенность различия нескольких дисперсий.
- Использование критерия Бартлета основано на предположении о нормальности (или близости к ней) распределения изучаемого признака в совокупностях, для которых исчислены дисперсии.
- Бартлет предложил сравнивать взвешенные среднюю арифметическую и среднюю геометрическую из дисперсий. Если при этом сопоставляемые дисперсии окажутся равными друг другу ($\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2$), то арифметическая и геометрическая средние из дисперсий совпадут.

Проверка статистических гипотез. Критерии

- Покажем использование критерия Бартлета при выборке, когда n_i>50. Взвешенная средняя арифметическая из дисперсий равна:

$$\bar{\sigma}_{ap}^2 = \frac{\sum \sigma_i^2 \cdot n_i}{\sum n_i}.$$

- Взвешенная средняя геометрическая из дисперсии равна:

$$\overline{\sigma_{геом}^2} = \sqrt[\sum n_i]{\prod[(\sigma_i^2)^{n_i}]}.$$

- В качестве критерия Бартлета использовал величину

$$M = \ln \frac{\overline{s_a^2}}{s_{геом}^2} \sum n_i.$$

- Отношение M/C, где

$$C = 1 + \frac{\sum \frac{1}{n_i} - \frac{1}{\sum n_i}}{3(m-1)},$$

распределено как χ^2 с числом степеней свободы равным m-1 (где m – число сравниваемых дисперсий).

Проверка статистических гипотез. Критерии

- Кроме того, величина М табулирована. Так как по правилу мажорности $\sigma_a^2 > \sigma_{геом}^2$, то критическую точку находят по уровню значимости α и по числу степеней свободы $k = m - 1$.
- Если фактическое значение критерия меньше табличного, то отклонять нулевую гипотезу нет основания. Дисперсии отличаются незначимо.

Критерий Кохрана используется для проверки существенности отличия самой большой дисперсии, полученной при выборках из совокупностей, подчиняющихся нормальному закону распределения, от всех остальных дисперсий.

- При равенстве численностей всех выборок ($n_1=n_2=...=n_k$) вместо критерия Бартлета предпочтительно использовать критерий Кохрана, фактическое значение которого ($q_ф$) представляет отношение максимальной из сравниваемых дисперсий к сумме всех дисперсий:

$$q_ф = \frac{\sigma_{\max}^2}{\sigma_1^2 + \sigma_2^2 + ... + \sigma_k^2}.$$

- Распределение $q_ф$ зависит от числа сравниваемых дисперсий и от числа степеней свободы $k=n-1$.

Проверка статистических гипотез. Критерии

- Рассмотренные ранее критерии относятся к **параметрическим критериям**, т.к. основаны на сравнении числовых характеристик распределений (средних, дисперсий и др.).
- Эти критерии используются в случае, когда изучаемая совокупность подчиняется нормальному закону распределения или закону, приводящему к нему после соответствующих преобразований.
- Бывают ситуации, когда применять параметрические методы нельзя. Как правило, это обусловлено тем, что распределение не является нормальным или слишком велико различие дисперсий, либо исследуемый признак не является числовым. Во всех таких случаях следует использовать ***непараметрические методы***, основанные на анализе рангов.
- ***Ранг*** – это число, указывающее на порядковое место конкретной реализации наблюдаемой величины в выборке.
- Природа порядковых признаков такова, что мы можем, сравнивая два значения, сказать о них только то, что одно из них больше, а другое меньше, но дать количественную характеристику их соотношения нельзя.

Проверка статистических гипотез. Критерии

- Количественные значения мы также можем расположить в порядке возрастания или убывания (ранжировать) и каждому значению присвоить определенный ранг. Таким образом, любой количественный признак можно рассматривать как порядковый, но не наоборот.
- Главное достоинство критериев, основанных на рангах, – нет необходимости знать, к какому типу распределения принадлежат данные и каковы его параметры. Вместе с тем чувствительность и информативность таких критериев ниже, чем параметрических.
- Использование непараметрических критериев ограничено только одним условием – теоретически распределение должно быть непрерывным и исследуемые выборки должны иметь одинаковое распределение.
- Таким образом, сфера применения непараметрических критериев шире, чем параметрических.