Теория вероятностей

Основные законы распределения случайных величин

Нормальный закон распределения

НСВ X имеет ***нормальный закон распределения (закон Гаусса)*** с параметрами a и σ^2 ($X \sim N(a;\, \sigma^2)$), если ее плотность вероятности имеет вид:

$$f_N(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-a)^2}{2\sigma^2}}$$

Кривую нормального закона распределения называют ***нормальной*** или ***гауссовой кривой.***

Свойства нормальной кривой.

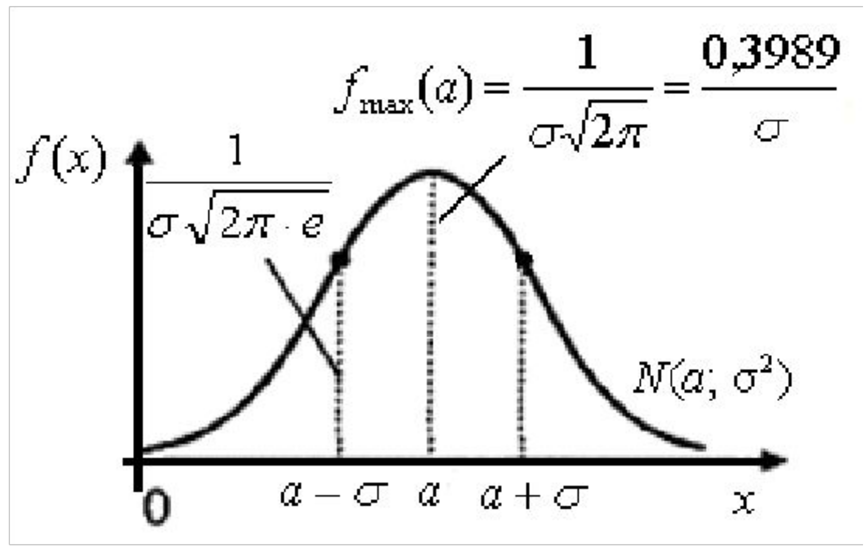
1. Определена на всей Ox.

2. $\forall x \quad f(x) \geq 0$.

3. При $|x| \to \infty$ асимптотически приближается к оси Ox.

4. Симметрична относительно прямой $x=a$, в т. $x=a$ имеет $f_{\max}(x=a)=\dfrac{1}{\sigma\sqrt{2\pi}}$ и две точки перегиба $x=a\pm\sigma$ с ординатой

$$f_{пер}(a\pm\sigma)=\frac{1}{\sigma\sqrt{2\pi e}}=\frac{0{,}2420}{\sigma}.$$

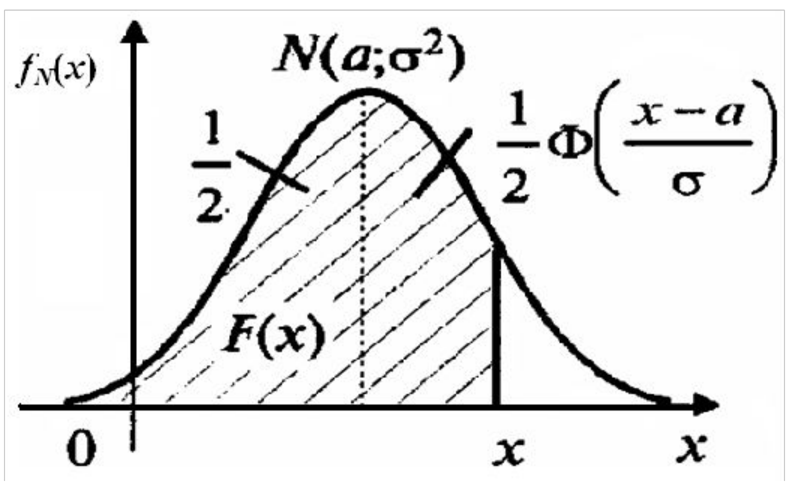


$$M(X)=a \quad D(X)=\sigma^2.$$

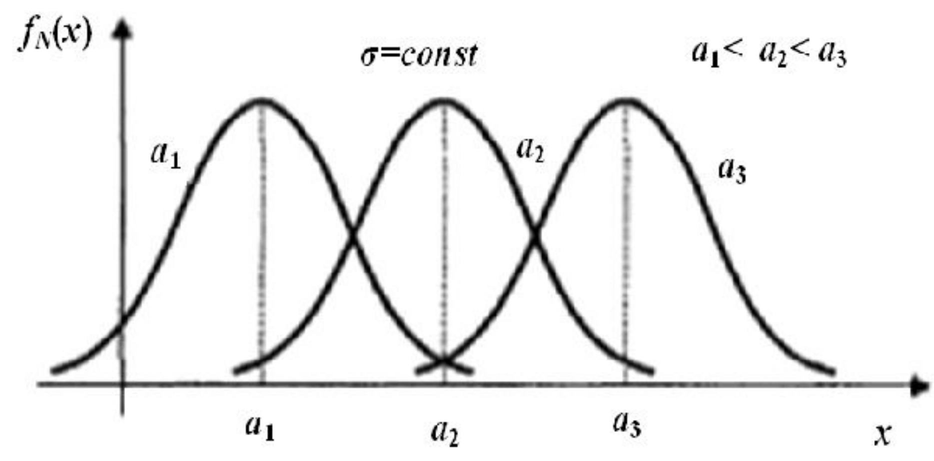
Нормальный закон распределения *СВ* $Z=\dfrac{x-a}{\sigma}$ с параметрами a=0, σ^2=1, т.е. N(0;1), называется ***стандартным*** или ***нормированным***, а соответствующая *СВ* $Z=\dfrac{x-a}{\sigma}$ называется ***стандартной*** или ***нормированной***

$$f_N(z)=\frac{1}{\sigma\sqrt{2\pi}}\cdot e^{-\frac{z^2}{2}}$$

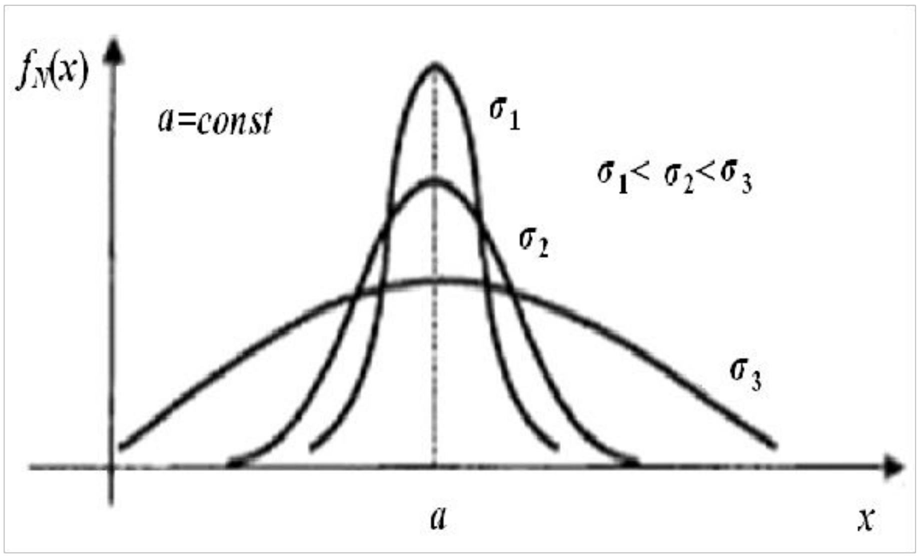
Функция распределения:



$$F_N(x)=\frac{1}{2}+\frac{1}{2}\Phi\left(\frac{x-a}{\sigma}\right)$$



Если σ=*const* и менять a ($a_1 < a_2 < a_3$) – кривая смещается вдоль оси абсцисс, не меняя формы.



Если a=*const* и менять σ – при увеличении σ_{max} значение кривой уменьшается, растягиваясь вдоль оси абсцисс.

При уменьшении σ_{max} – кривая вытягивается вверх, одновременно сжимаясь с боков.

*Параметр **a** характеризует положение, а параметр **σ** – форму нормальной кривой.*

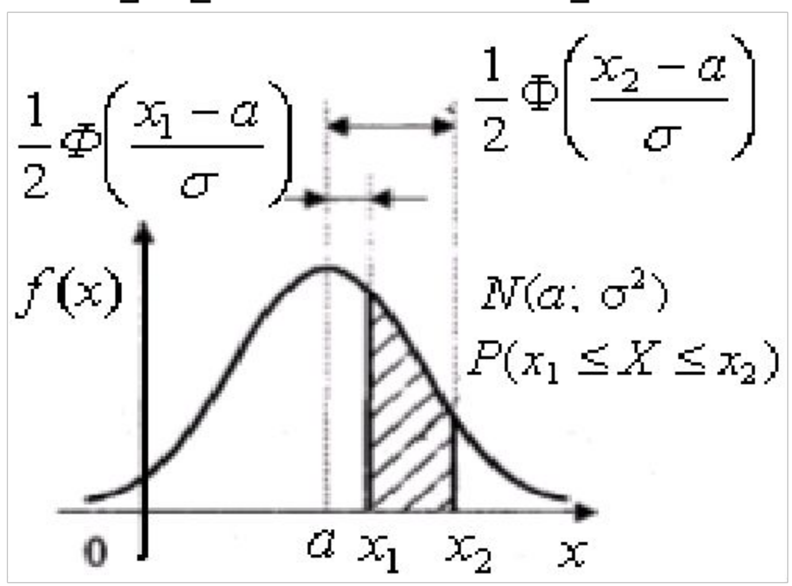
Свойства СВ, распределенной по нормальному закону.

1. $$P(x_1 \le X \le x_2) = \frac{1}{2}\left[\Phi(t_2) - \Phi(t_1)\right],$$ где

$$t_1 = \frac{x_1 - a}{\sigma} \quad t_2 = \frac{x_2 - a}{\sigma}.$$

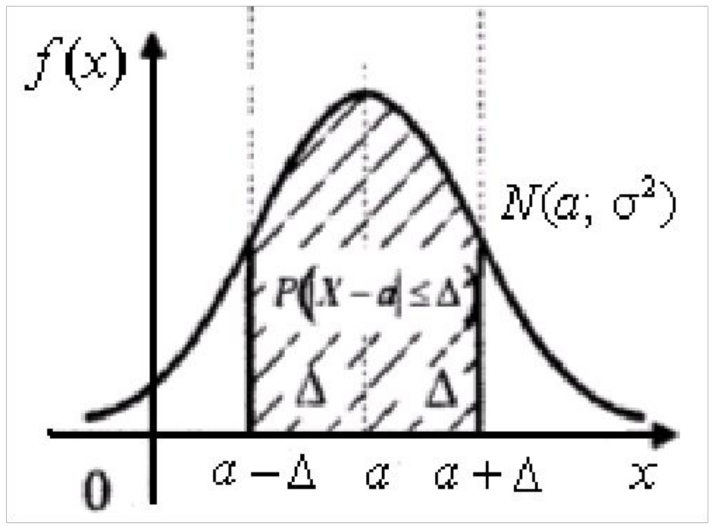
$$P(x_1 \le X \le x_2) = F(x_2) - F(x_1) =$$

$$= \left[\frac{1}{2} + \frac{1}{2}\Phi\left(\frac{x_2 - a}{\sigma}\right)\right] - \left[\frac{1}{2} + \frac{1}{2}\Phi\left(\frac{x_1 - a}{\sigma}\right)\right] = \frac{1}{2}\left[\Phi(t_2) - \Phi(t_1)\right].$$

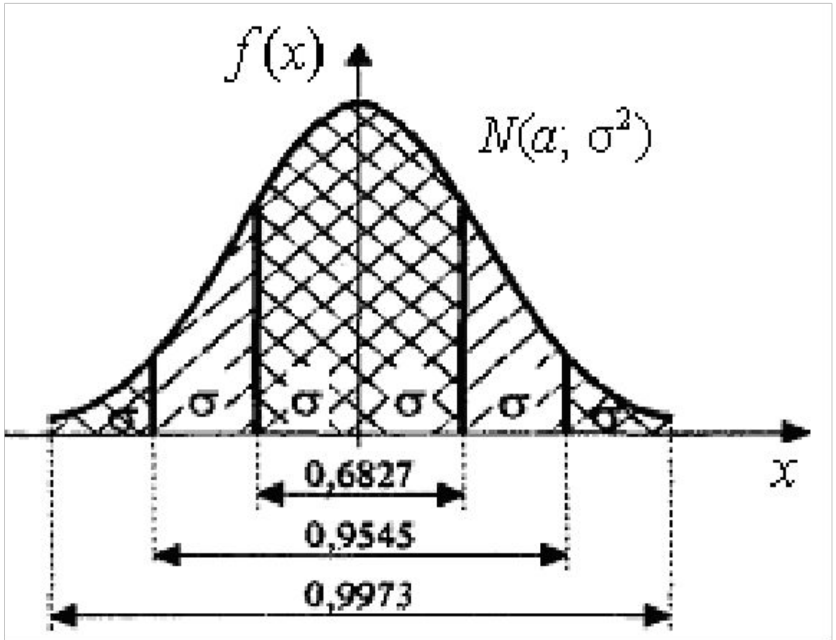


2. Вероятность того, что отклонение *СВ* *X*, распределенной по нормальному закону, от математического ожидания ***a*** не превысит величину $\Delta > 0$ (по абсолютной величине), равна

$$P(|X-a| \le \Delta) = \Phi(t)$$, где $t = \Delta/\sigma$.



Вычисляя при различных значениях Δ вероятности



$$P(|X-a|\le\Delta)=\Phi(t),$$

получим

Δ=σ

$$P(|X-a|\le\sigma)=\Phi(1)=0{,}6827,$$

Δ=2σ

$$P(|X-a|\le 2\sigma)=\Phi(2)=0{,}9545,$$

Δ=3σ

$$P(|X-a|\le 3\sigma)=\Phi(3)=0{,}9973.$$

«Правило трех сигм»:

Если СВ $X \sim N(a,\sigma^2)$, то практически достоверно, что ее значения заключены в интервале $(a-3\sigma,\ a+3\sigma)$.

Распределения, связанные с нормальным распределением

Число степеней свободы определяется как разность между числом суммируемых СВ и числом линейных связей, ограничивающих свободу изменения этих величин.

χ2-распределение (распределение Пирсона).

Распределением χ2 (хи-квадрат) с k степенями свободы называется распределение суммы квадратов k независимых CB, распределенных по стандартному нормальному закону, т.е.

$$\chi^2 = \sum_{i=1}^{k} Z_i^2 ,$$

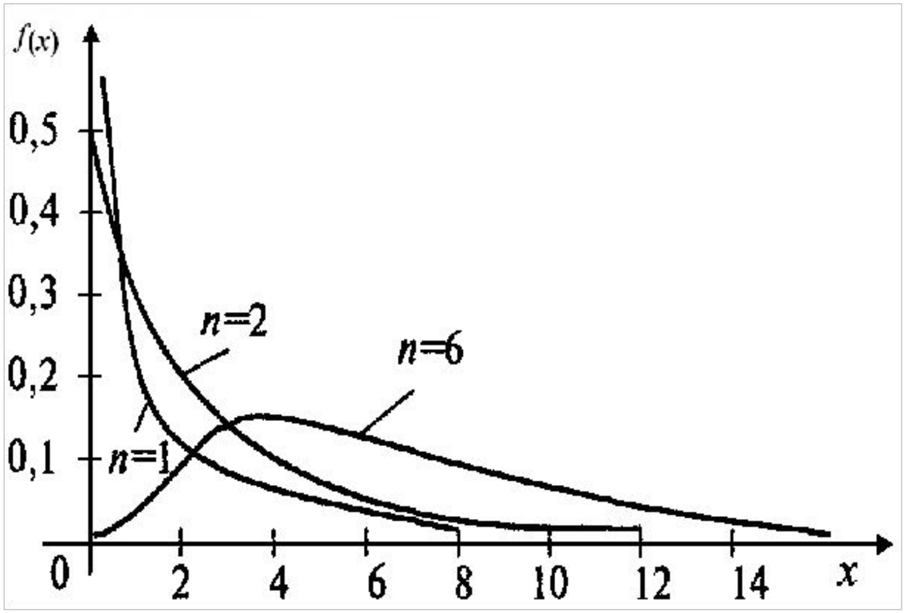
где $Z_i \ (i = \overline{1,k}) \sim N(0;1)$.

Плотность вероятности χ^2 - распределения имеет вид:

$$f(x) = \begin{cases} \dfrac{1}{2^{\frac{k}{2}}\Gamma\left(\dfrac{k}{2}\right)} \cdot x^{\frac{k}{2}-1} e^{-\frac{x}{2}}, & при \ x \geq 0 \\ 0, & при \ x < 0 \end{cases}$$

где $\Gamma(y) = \int\limits_0^{+\infty} e^{-t} t^{y-1} dt$ – гамма-функция Эйлера (для целых положительных значений $\Gamma(y)=(y-1)!$).

Кривые χ^2 – асимметричны, обладают положительной (правосторонней) асимметрией.



Распределение $\chi^2(k)$ определяется одним параметром – *числом степеней свободы k.*

С увеличением числа степеней свободы распределение медленно приближается к нормальному.

При $k>30$ распределение *СВ* $Z = 2\chi^2 - \sqrt{2k}$ близко к $N(0;1)$.

Распределение Стьюдента.

Распределением Стьюдента (или t-распределением) называется распределение *СВ*, равное отношению нормированной *СВ* к корню квадратному из независимой *СВ*, распределенной по закону χ^2.

$$t = Z \Big/ \sqrt{\frac{1}{k}\chi^2}$$

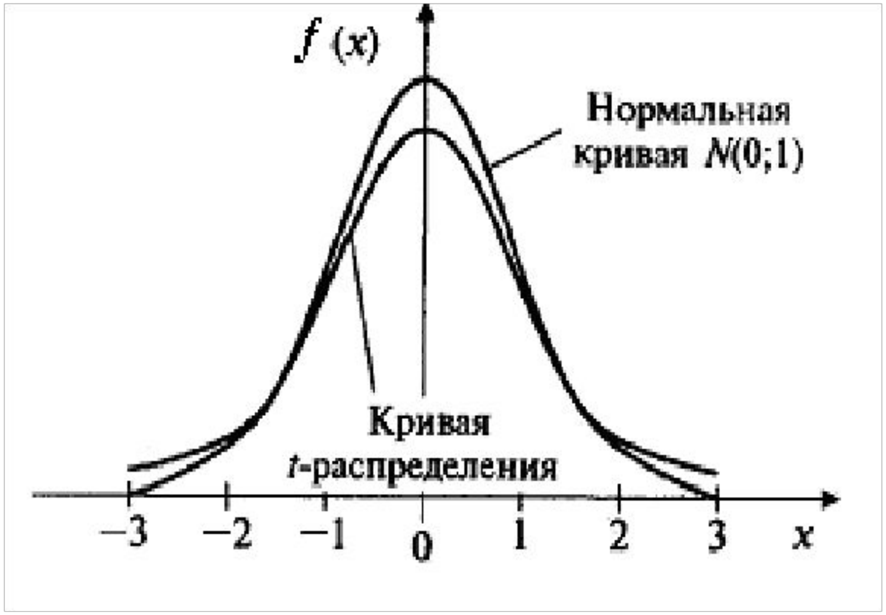
где $Z \sim N(0;1)$;

$\chi^2 \sim \chi^2(k)$,независимая от Z.

Плотность вероятности распределения Стьюдента:

$$f(x) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\sqrt{\pi k}}\left(1 + \frac{x_2}{n}\right)^{-\frac{k+1}{2}}$$

где Г(у) — гамма-функция в точке *y*.



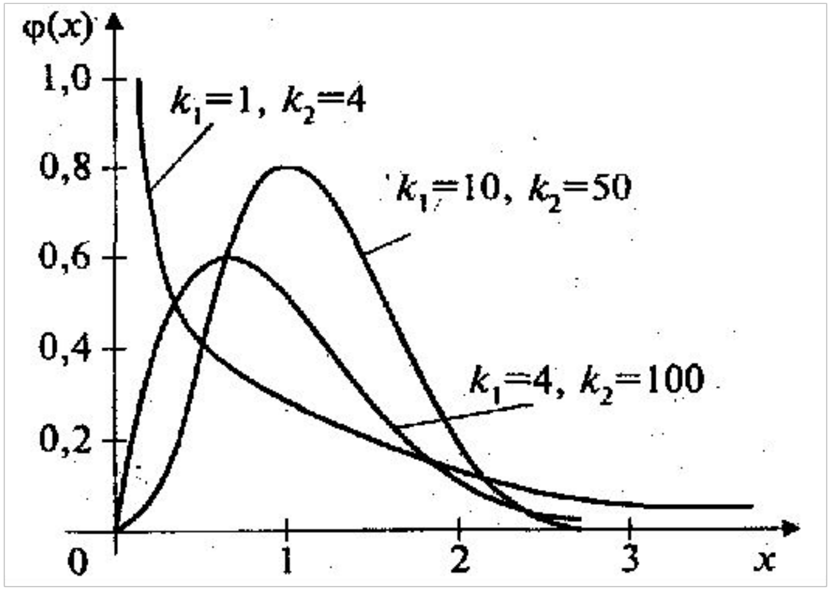
Кривая t-распределения симметрична относительно оси ординат, но по сравнению с нормальной кривой – более *пологая*.

При $k \to \infty$ t-распределение быстро приближается к нормальному.

Практически уже при $k>30$ можно считать t-распределение приближенно нормальным.

Распределение Фишера—Снедекора.

Распределением Фишера—Снедекора (F-распределением) называется распределение CB



$$F = \frac{\frac{1}{k_1}\chi^2(k_1)}{\frac{1}{k_2}\chi^2(k_2)}$$

где $\chi^2(k_1)$ и $\chi^2(k_2)$ – CB, имеющие χ^2-распределение соответственно с k_1 и k_2 степенями свободы.

Плотность F- распределения:

$$f(x) = \frac{\Gamma\left(\frac{k_1+k_2}{2}\right) k_1^{\frac{k_1}{2}} k_2^{\frac{k_2}{2}}}{\Gamma\left(\frac{k_1}{2}\right)\Gamma\left(\frac{k_2}{2}\right)} x^{\frac{k_1}{2}-1}\left(k_1 x + k_2\right)^{-\frac{k_1+k_2}{2}},$$

где $\Gamma(y)$ – гамма-функция Эйлера в точке y.

При $n \to \infty$ F-распределение приближается к нормальному закону.

Математическая статистика

Проверка статистических гипотез

ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

Статистическая гипотеза – некоторое предположение, формулируемое на основе выборки, о законе распределения СВ или о параметрах этого закона.

Проверяемую гипотезу, утверждающую, что различие между сравниваемыми характеристиками отсутствует, а наблюдаемые отклонения объясняются лишь случайными колебаниями в выборках называют ***нулевой*** (***основной***) гипотезой и обозначают H_0.

Гипотезу H_1, являющуюся логическим отрицанием H_0, называют ***альтернативной*** (***конкурирующей***), т.е. если гипотеза H_0 будет отвергнута, то будет иметь место альтернативная гипотеза H_1.

Замечание. H_0 часто специально подбирается так, чтобы отвергнуть ее и принять тем самым альтернативную.

ОШИБКИ ПЕРВОГО И ВТОРОГО РОДА

Принятие или отклонение выдвинутой гипотезы H_0 соответствует истине с некоторой вероятностью.

Ошибка первого рода возникает с вероятностью α тогда, когда отвергается верная гипотеза H_0 и принимается конкурирующая гипотеза H_1:

$$P_{H_0}(H_1) = \alpha \; .$$

Вероятность α называется ***уровнем значимости***.

Ошибка второго рода возникает с вероятностью β в том случае, когда принимается неверная гипотеза H_0, в то время как справедлива конкурирующая гипотеза H_1:

$$P_{H_1}(H_0) = \beta \; .$$

Кроме совершения ошибок может быть принято и верное решение.

Доверительная вероятность – это вероятность не совершить ошибку первого рода и принять верную гипотезу H_0:

$$P_{H_0}(H_0) = 1 - \alpha .$$

Вероятность отвергнуть ложную гипотезу H_0 называется ***мощностью критерия***:

$$P_{H_1}(H_1) = 1 - \beta .$$

Исключить ошибки первого и второго рода *невозможно* в силу ограниченности выборки.

Единственный способ уменьшения вероятности ошибок – увеличение объема выборки.

Правило, по которому гипотеза H_0 принимается или отвергается, называется ***статистическим критерием (статистика)***: $\varphi = \varphi(x_1, \ldots, x_n)$.

Статистика $\varphi = \varphi(x_1, \ldots, x_n)$ – специально подобранная выборочная характеристика, точное или приближенное распределение которой известно.

Статистику обозначают

u или ***z***, если она распределена нормально,

F – по закону Фишера-Снедекора,

t – по закону Стьюдента,

χ^2– по закону $\boldsymbol{\chi}$ – квадрат.

Вычисленное по выборке значение критерия называют ***наблюдаемым значением*** $\boldsymbol{\varphi}_{\text{набл}} = \theta^*$.

КРИТИЧЕСКАЯ ОБЛАСТЬ.
ОБЛАСТЬ ПРИНЯТИЯ ГИПОТЕЗЫ

Множество W возможных значений критерия, при которых H_0 не отклоняется, называется ***областью принятия гипотезы***, а множество $\overline{W}$ возможных значений критерия, при которых H_0 отклоняется, называется ***критической областью***.

Основной принцип проверки статистической гипотезы:

Если значение критерия $\boldsymbol{\varphi}$ *попадает в область* W, *то гипотеза принимается, а если в область* $\overline{W}$ *– гипотеза отклоняется.*

Точки, отделяющие критическую область от области принятия гипотезы, называются ***критическими*** $\boldsymbol{\varphi_{кр}}$.

Заранее задается α, считая, что в одном отдельно взятом испытании событие с вероятностью меньшей α, практически не происходит.

По α находим такое число $\varphi(\alpha)=\varphi_{кр}$, чтобы вероятность $P(\varphi_{набл} > \varphi_{кр}) = \alpha$.

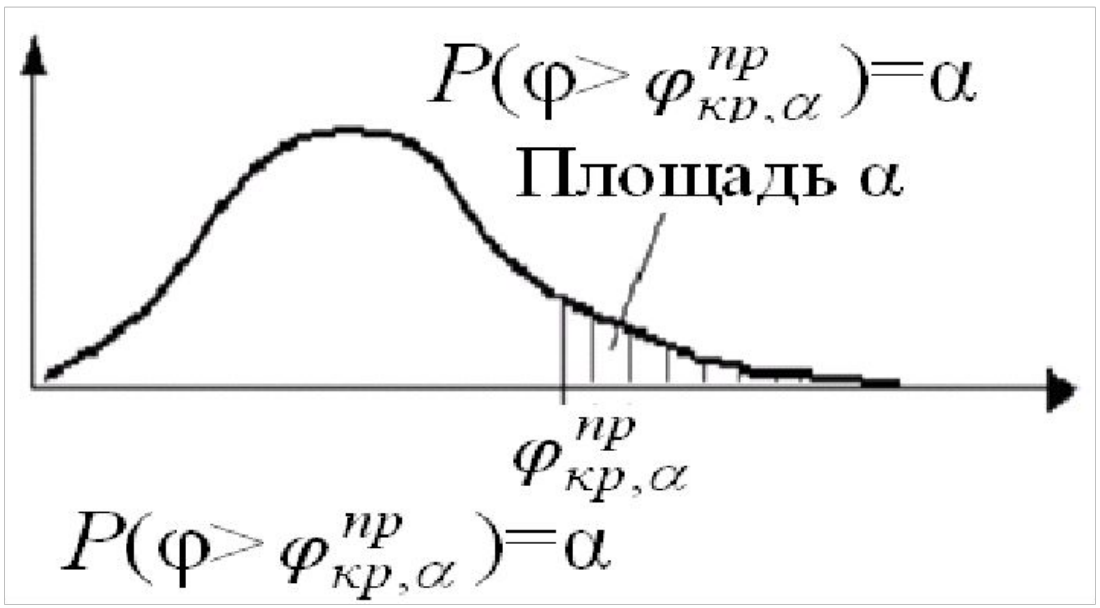
$\varphi_{кр}$ находят по таблицам критических точек соответствующих распределений.

Основная идея: если $\varphi_{набл} > \varphi_{кр}$, то это означает, что произошло «практически невозможное» событие и поэтому с вероятностью $1-\alpha$ гипотезу H_0 следует отвергнуть и принять H_1.

В противном случае можно считать, что наблюдения не противоречат гипотезе H_0.

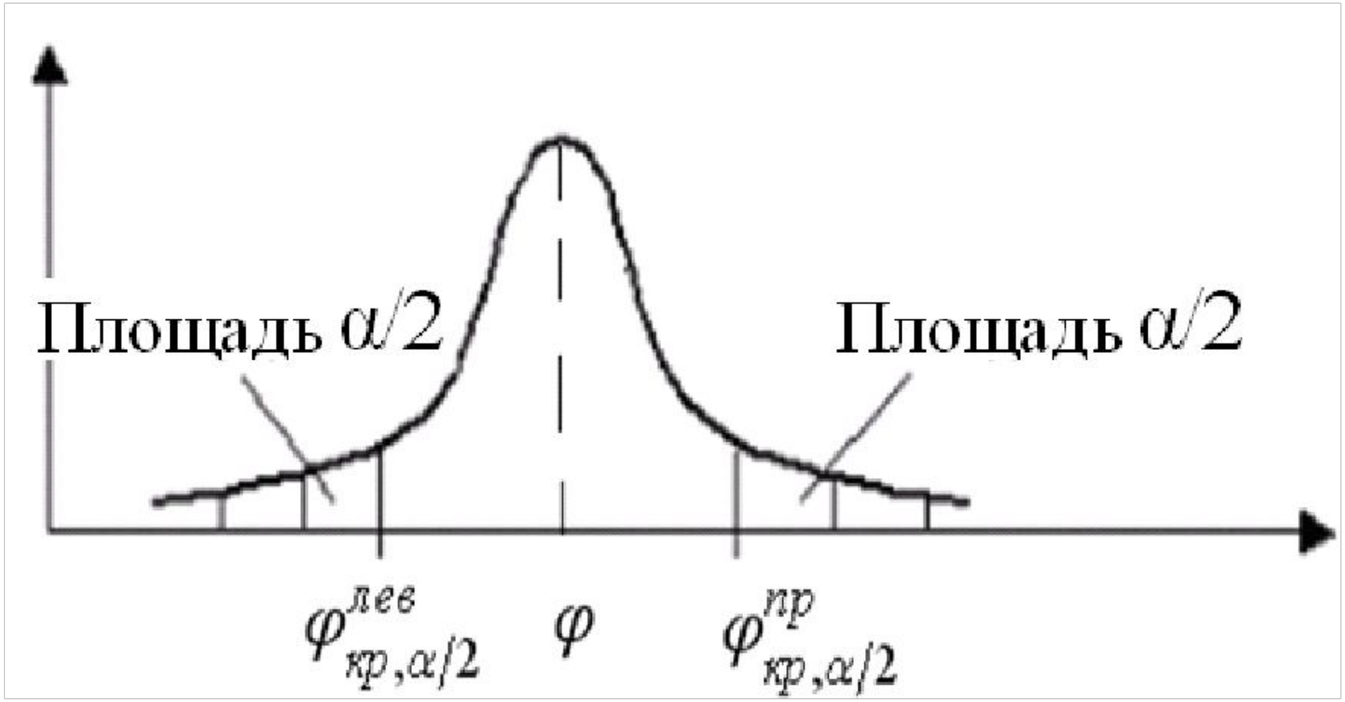
В зависимости от вида гипотезы H_1 выбирают правостороннюю, левостороннюю или двустороннюю критическую область.

Если конкурирующая гипотеза $H_1 : \varphi > \varphi_0$, то ***правосторонняя область*** определяется неравенством $\varphi_{набл} > \varphi_{кр}$.



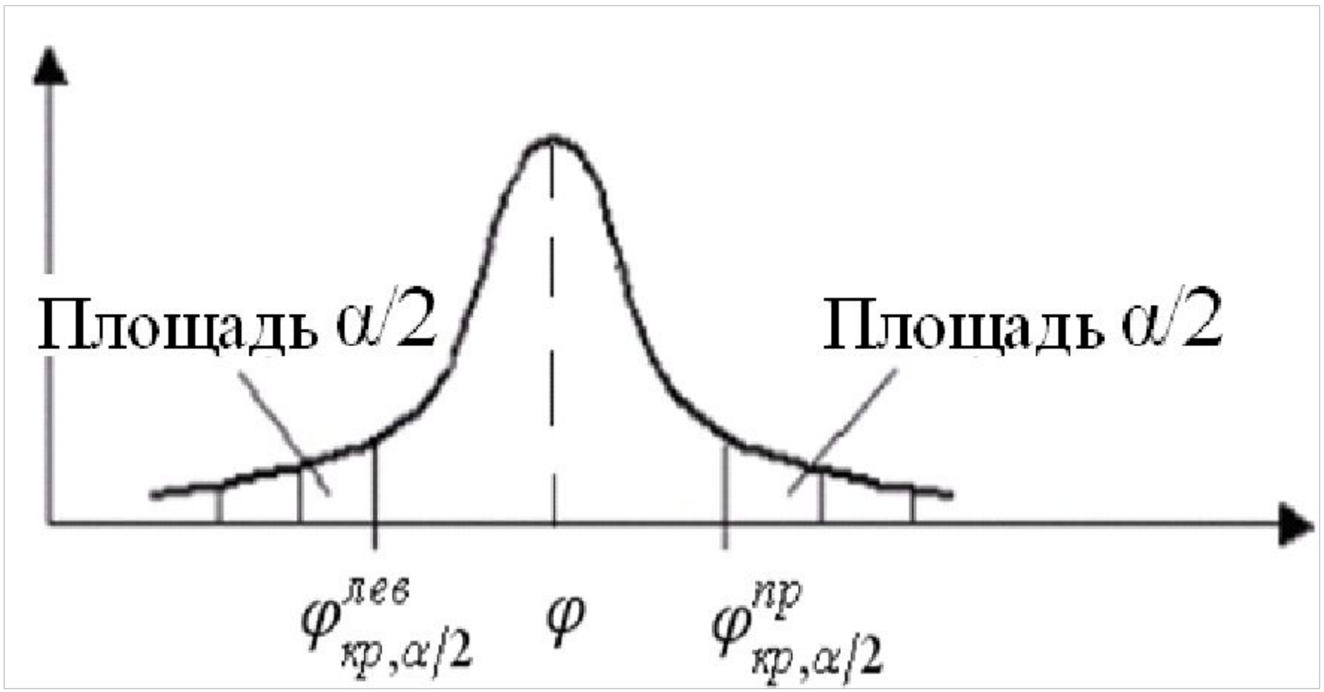
Если конкурирующая гипотеза $H_1 : \varphi < \varphi_0$, то ***левосторонняя область*** определяется неравенством $\varphi_{набл} < \varphi_{кр}$.

Если конкурирующая гипотеза $H_1 : \varphi \neq \varphi_0$, то ***двусторонняя критическая область*** состоит из интервалов $(-\infty,\ \varphi^{лев}_{кр,\alpha/2})$ и $(\varphi^{пр}_{кр,\alpha/2},\ +\infty)$.



Если распределение критерия симметрично относительно нуля, то выбирают симметричные относительно нуля критические точки $-\varphi_{кр,\alpha/2}$ и $\varphi_{кр,\alpha/2}$:

$$P(\varphi < -\varphi_{кр,\alpha/2}) = P(\varphi > \varphi_{кр,\alpha/2})$$



ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

Общая схема проверки гипотез:

1. Формулировка проверяемой (нулевой – H_0) и альтернативной (H_1) гипотез.
2. Выбор соответствующего уровня значимости α.
3. Определение объема выборки n.
4. Выбор критерия φ для проверки H_0.
5. Определение критической области и области принятия гипотезы.
6. Вычисление наблюдаемого значения критерия $\varphi_{набл}$
7. Принятие статистического решения.

Математическая статистика

Примеры проверки гипотез

Проверка гипотез о равенстве числовых характеристик генеральных совокупностей

Сравнение средних двух генеральных совокупностей при <u>известных дисперсиях</u>

Имеются 2 совокупности, характеризуемые неизвестными средними a_x и a_y и ***известными дисперсиями*** σ_x^2 и σ_y^2. Необходимо проверить $H_0: a_x = a_y$. $H_1^{(1)}: a_x > a_y$ или $H_1^{(2)}: a_x < a_y$ или $H_1^{(3)}: a_x \neq a_y$.

Для проверки H_0 взяты две независимые выборки объемом n_x и n_y, по которым найдены $\bar{x}$ и $\bar{y}$.

В случае справедливости гипотезы H_0 статистика

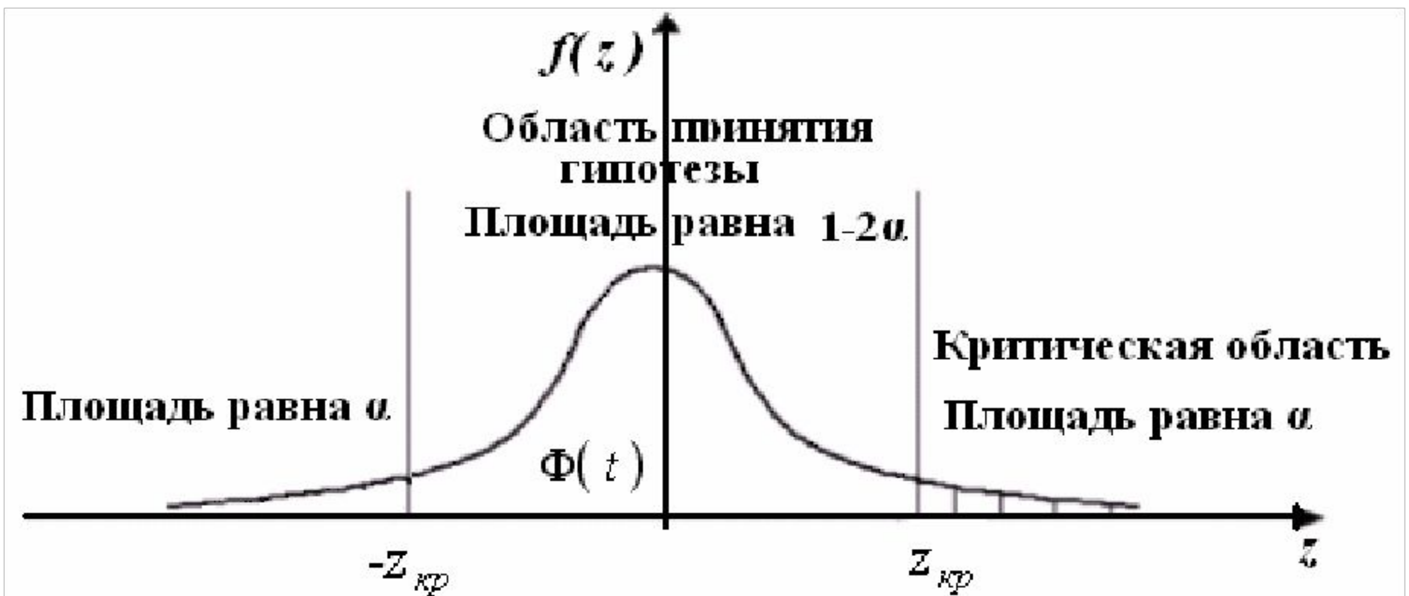
$$Z = \frac{\left(\bar{x} - \bar{y}\right)}{\sqrt{\sigma_x^2 / n_x + \sigma_y^2 / n_y}} \sim N(0,1)$$

В случае $H_1^{(1)}: a_x > a_y$ ($H_1^{(2)}: a_x < a_y$): критическую область выбирают из условия, что $P(\varphi \in \overline{W} / H_0) = \alpha$:

если $H_1^{(1)}: a_x > a_y$, то $P(z > z_{кр}) = \alpha$,

если $H_1^{(2)}: a_x < a_y$, то $P(z < -z_{кр}) = \alpha$

и $z_{кр}$ статистики находят из условия: $\Phi\left(z_{кр}\right) = \Phi\left(z_{1-2\alpha}\right) = 1 - 2\alpha$.



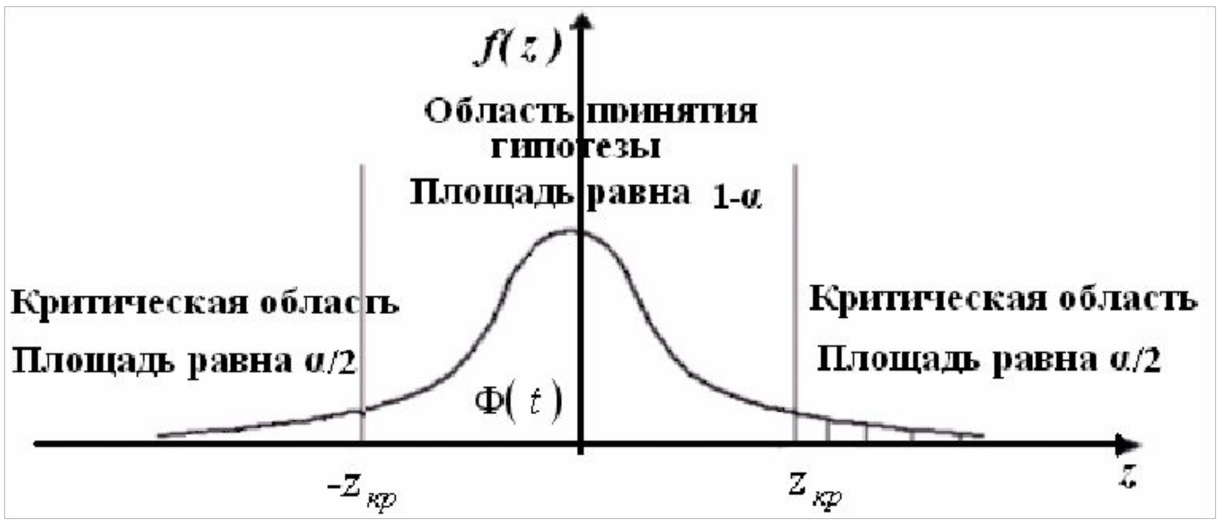
При гипотезе H_1: $a_x \neq a_y$ – двусторонняя критическая область

$(z_{кр};+\infty)\cap(-\infty;z_{кр})$ так, чтобы

$$P(z > z_{кр}) = \frac{\alpha}{2} \text{ и } P(z < -z_{кр}) = \frac{\alpha}{2}.$$

Критическое значение статистики находят из условия:

$$\Phi(z_{кр}) = \Phi(z_{1-\alpha}) = 1 - \alpha.$$



Подсчитывают $z_{набл}$. Если на уровне значимости α:

- ✓ $|z_{набл}| > z_{кр}$, то гипотеза H_0 отвергается;
- ✓ $|z_{набл}| \leq z_{кр}$, то делается вывод о том, что гипотеза H_0 не противоречит имеющимся наблюдениям.

Сравнение средних двух генеральных совокупностей при неизвестных дисперсиях

Имеются 2 совокупности, характеризуемые неизвестными средними a_x и a_y.

Дисперсии σ_x^2 и σ_y^2 ***неизвестны, но равны*** $(\sigma_x^2 = \sigma_y^2 = \sigma^2)$.

Необходимо проверить гипотезу

$$H_0 : a_x = a_y$$

$$H_1^{(1)} : a_x > a_y \text{ или } H_1^{(2)} : a_x < a_y \text{ или } H_1^{(3)} : a_x \neq a_y$$

В качестве неизвестной σ^2 берем «исправленную» выборочную дисперсию «*смешанной*» совокупности объема $n_x + n_y$

$$\tilde{s}^2 = \frac{n_x \tilde{s}_x^2 + n_y \tilde{s}_y^2}{n_x + n_y - 2}.$$

В случае справедливости гипотезы H_0 статистика

$$t=\frac{Z}{\sqrt{\chi^2/k}}=\frac{\bar{x}-\bar{y}}{\sqrt{\frac{n_x s_x^2+n_y s_y^2}{n_x+n_y-2}\cdot\left(\frac{1}{n_x}+\frac{1}{n_y}\right)}}\sim t(\gamma=1-\alpha;k=n_x+n_y-2)$$

$t_{кр}$ находится по таблице Стьюдента в зависимости от критической области: односторонняя критическая область строится из предположения, что $P(t>t_{кр})=1-2\alpha$, двусторонняя – $P(t>t_{кр})=1-\alpha$.

Если наблюдаемой значение статистики на уровне значимости α:

- ✓ $|t_{набл}|>t_{кр}|$, то гипотеза H_0 отвергается;
- ✓ $|t_{набл}|\le t_{кр}$, то делается вывод о том, что гипотеза H_0 не противоречит имеющимся наблюдениям.

ПРОВЕРКА ГИПОТЕЗ О РАВЕНСТВЕ ДИСПЕРСИЙ ДВУХ СОВОКУПНОСТЕЙ

Имеются 2 нормально распределенные совокупности, дисперсии которых σ_x^2 и σ_y^2 неизвестны. Необходимо проверить гипотезу $H_0: \sigma_x^2 = \sigma_y^2 = \sigma^2$

Взяты две независимые выборки объемом n_x и n_y. Для оценки дисперсий используются «исправленные» выборочные дисперсии $\tilde{s}_x^2$ и $\tilde{s}_y^2$.

Задача сводится к сравнению дисперсий $\tilde{s}_x^2$ и $\tilde{s}_y^2$.

В случае справедливости гипотезы H_0 статистика

$$F=\frac{\frac{1}{n_x-1}\left[(n_x-1)\cdot \tilde{s}_x^{2}/\sigma^{2}\right]}{\frac{1}{n_y-1}\left[(n_y-1)\cdot \tilde{s}_y^{2}/\sigma^{2}\right]}=\frac{\tilde{s}_x^{2}}{\tilde{s}_y^{2}}\sim F(\alpha;k_1=n_x-1;k_2=n_y-1)$$

т.к. F-распределение является несимметричным, то

- если $H_1:\sigma_x^2>\sigma_y^2$ – выбирают критическую область $(F_{кр,\alpha}^{пр};+\infty)$, где точка $F_{кр,\alpha}^{пр}$ определяется из условия

$$P(F(k_1=n_x-1;k_2=n_y-1)>F_{кр,\alpha}^{пр})=\alpha.$$

- если $H_1:\sigma_x^2<\sigma_y^2$ – выбирают критическую область $(0;F_{кр,\alpha}^{лев})$, где точка $F_{кр,\alpha}^{лев}$ определяется из условия

$$P(F(k_1=n_y-1;k_2=n_x-1)>F_{кр,1-\alpha}^{лев})=1-\alpha$$

- если $H_1: \sigma_x^2 \neq \sigma_y^2$ – выбирают двустороннюю критическую область $(0; F^{лев}_{кр,1-\alpha/2})$ и $(F^{пр}_{кр,\alpha/2}; +\infty)$, где точки $F^{лев}_{кр,1-\alpha/2}$ и $F^{пр}_{кр,\alpha/2}$ находят из условия

$$P(F(k_1 = n_y - 1; k_2 = n_x - 1) < F^{лев}_{кр,1-\alpha/2}) = 1 - \alpha/2$$

$$P(F(k_1 = n_x - 1; k_2 = n_y - 1) > F^{пр}_{кр,\alpha/2}) = \alpha/2$$

Правую критическую точку (большую единицы) находят по таблице критических точек распределения Фишера – Снедекора $F^{пр}_{кр,\alpha/2}$.

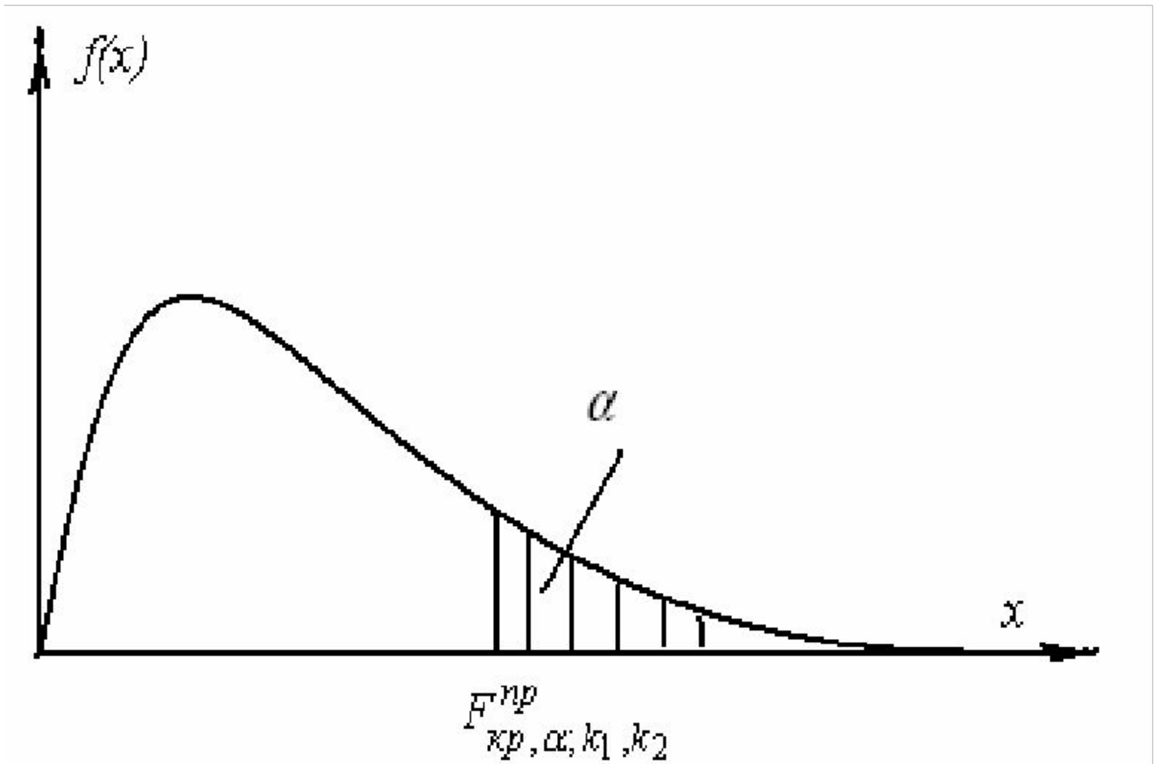
Левую же границу находят из соотношения, доказанного для критерия Фишера:

$$F^{лев}_{кр,1-\alpha/2;k_1,k_2} = \frac{1}{F^{пр}_{\alpha/2;k_2,k_1}}$$

Если на уровне значимости α:

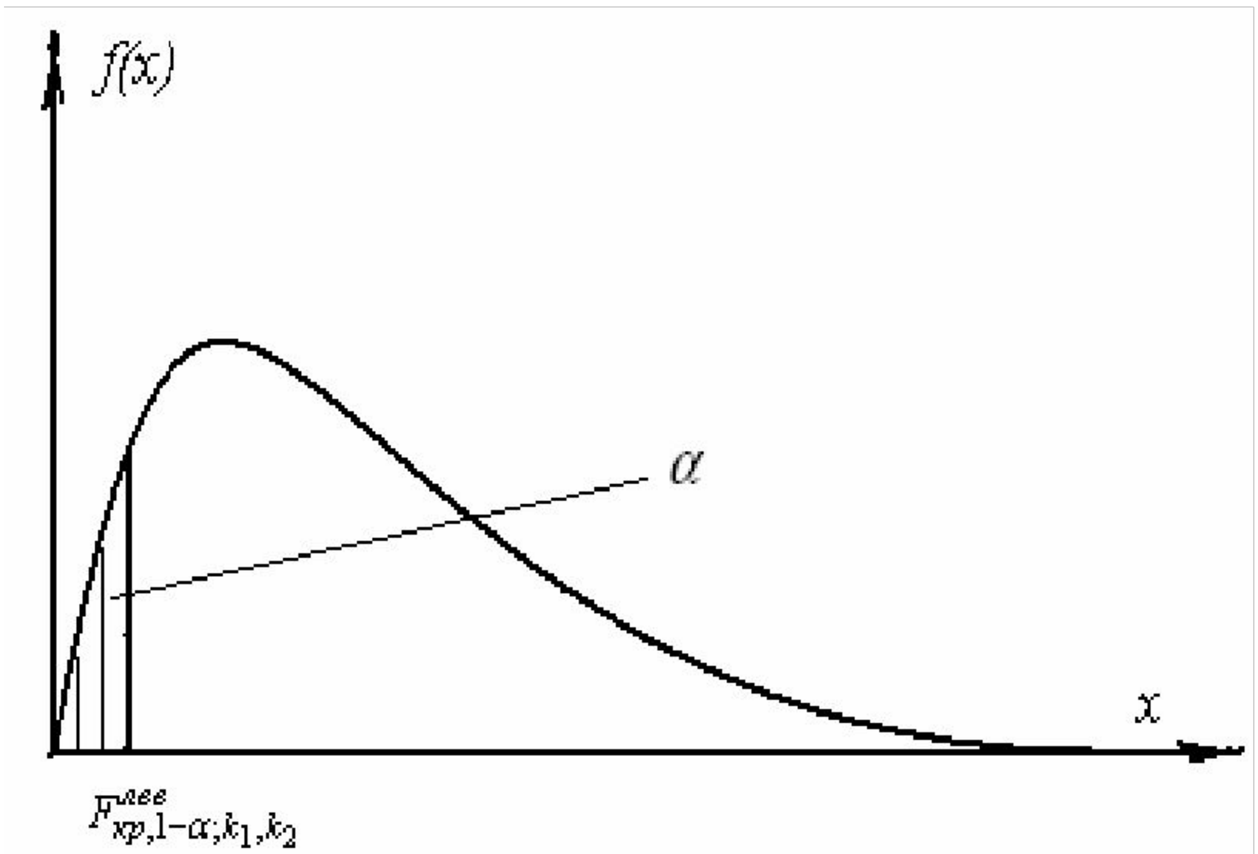
в случае правосторонней критической области

- ✓ $F_{набл} > F^{пр}_{кр,\alpha;k_1,k_2}$, то гипотеза H_0 отвергается;
- ✓ в противном случае делается вывод о том, что гипотеза H_0 не противоречит имеющимся наблюдениям.



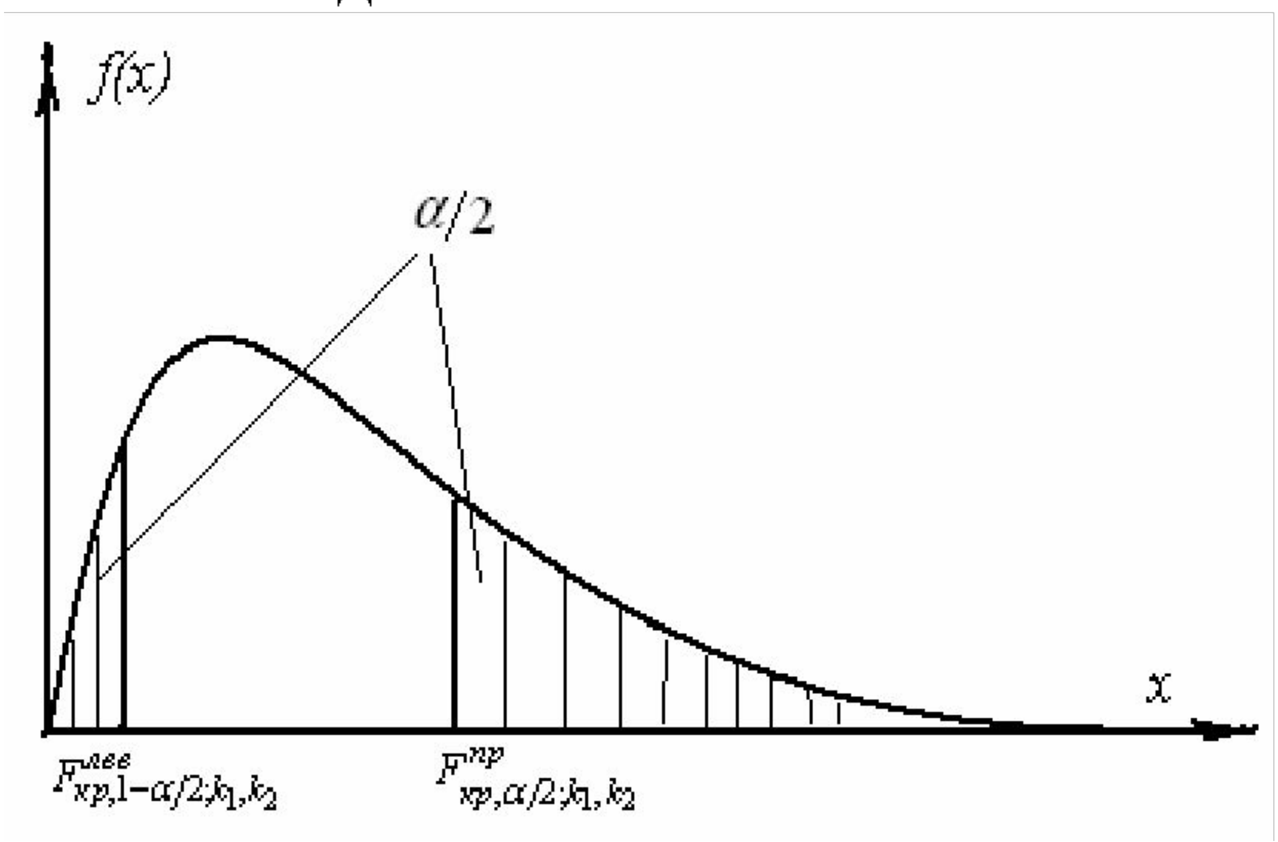
в случае левосторонней критической области

- ✓ $F_{набл} < F^{лев}_{кр,1-\alpha;k_1,k_2}$, то гипотеза H_0 отвергается;
- ✓ в противном случае делается вывод о том, что гипотеза H_0 не противоречит имеющимся наблюдениям.



в случае двусторонней критической области

- ✓ $F_{набл} < F^{лев}_{кр,1-\alpha;k_1,k_2}$ или $F_{набл} > F^{пр}_{кр,\alpha;k_1,k_2}$, то гипотеза H_0 отвергается;
- ✓ в противном случае делается вывод о том, что гипотеза H_0 не противоречит имеющимся наблюдениям.



ПРОВЕРКА ГИПОТЕЗ

О ЧИСЛОВЫХ ЗНАЧЕНИЯХ ПАРАМЕТРОВ ГЕНЕРАЛЬНЫХ СОВОКУПНОСТЕЙ

$H_0: \theta = \theta_0$, где θ – некоторый параметр исследуемого распределения, а $\boldsymbol{\theta_0}$ –его конкретное значение.

H_0	Предположение	Статистика	H_1	Критерий отклонения гипотезы
$a=a_0$	σ^2 известна	$z = \dfrac{\bar{x} - a_0}{\sigma/\sqrt{n}}$	$\left.\begin{matrix} a = a_1 > a_0 \\ a = a_1 < a_0 \end{matrix}\right\}$	$\lvert z_{набл} \rvert > z_{кр,1-2\alpha}$
			$a = a_1 \neq a_0$	$\lvert z_{набл} \rvert > z_{кр,1-\alpha}$
	σ^2 неизвестна	$t = \dfrac{\bar{x} - a_0}{s/\sqrt{n-1}}$	$\left.\begin{matrix} a = a_1 > a_0 \\ a = a_1 < a_0 \end{matrix}\right\}$	$\lvert t_{набл} \rvert > t_{кр,1-2\alpha,n-1}$
			$a = a_1 \neq a_0$	$\lvert t_{набл} \rvert > t_{кр,1-\alpha,n-1}$
$\sigma^2 = \sigma^2_0$	a неизвестно	$\chi^2 = \dfrac{ns^2}{\sigma_0^2}$	$\sigma^2 = \sigma_1^2 > \sigma_0^2$	$\chi^2_{набл} > \chi^2_{кр,\alpha;n-1}$
			$\sigma^2 = \sigma_1^2 < \sigma_0^2$	$\chi^2_{набл} < \chi^2_{кр,1-\alpha;n-1}$
			$\sigma^2 = \sigma^2_1 \neq \sigma^2_0$	$\chi^2_{набл} > \chi^2_{кр,\alpha/2;n-1}$ $\chi^2_{набл} < \chi^2_{кр,1-\alpha/2;n-1}$
$p = p_0$	достаточно большие n	$z = \dfrac{w - p_0}{\sqrt{p_0 q_0/n}}$	$\left.\begin{matrix} p = p_1 > p_0 \\ p = p_1 < p_0 \end{matrix}\right\}$	$\lvert z_{набл} \rvert > z_{кр,1-2\alpha}$
			$p = p_1 \neq p_0$	$\lvert z_{набл} \rvert > z_{кр,1-\alpha}$