Распределение Стьюдента

Пусть $\xi, \xi_1, \ldots, \xi_n$ независимые нормально распределенные величины $\mathcal{N}(0, \sigma)$. Случайная величина вида

$$\zeta = \frac{\xi}{\sqrt{\frac{1}{n}\sum\limits_{i=1}^{n} \xi_i^2}} = \frac{\xi}{\eta}, \tag{1}$$

имеет распределение Стьюдента с n степенями свободы.

Вычисление плотности распределения Стьюдента

Не умаляя общности, будем считать $\sigma = 1$.

$\sum\limits_{i=1}^{n} \xi_i^2 \sim \chi^2(n)$ с плотностью $\gamma(x|\alpha = \frac{1}{2}, \lambda = \frac{n}{2})$

$$\kappa_n(x) = \frac{2^{-\frac{n}{2}}}{\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \,. \tag{2}$$

$$F_{a\xi}(x) = P\{a\xi < x\} = P\left\{\xi < \frac{x}{a}\right\} = F_\xi\left(\frac{x}{a}\right),$$

$$F_{\sqrt{\xi}}(x) = P\left\{\sqrt{\xi} < x\right\} = P\{\xi < x^2\} = F_\xi(x^2), \implies$$

$$f_{a\xi}(x) = \frac{1}{a} f_\xi\left(\frac{x}{a}\right), \qquad f_{\sqrt{\xi}}(x) = 2x f_\xi(x^2).$$

$$\frac{1}{n}\sum_{i=1}^{n} \xi_i^2 \sim n\kappa_n(nx), \qquad \eta = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \xi_i^2} \sim 2nx\kappa_n(nx^2)\,.$$

Плотность совместного распределения ξ и η

$$\kappa_n(x) = \frac{2^{-\frac{n}{2}}}{\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \qquad \eta = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \xi_i^2} \sim 2nx\kappa_n(nx^2)$$

$$\frac{2n\eta}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} \left(n\eta^2\right)^{\frac{n}{2}-1} e^{-\frac{n\eta^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}} = c_n \eta^{n-1} e^{-\frac{\xi^2+n\eta^2}{2}},$$

$$\text{где} \quad c_n = \frac{\sqrt{2}\left(\frac{n}{2}\right)^{\frac{n}{2}}}{\sqrt{\pi}\Gamma\left(\frac{n}{2}\right)}.$$

Плотность распределения $\zeta = \frac{\xi}{\eta}$

Например, 0.95-квантиль распределения Стьюдента со степенью свободы 20 равна 1.72, то есть $\mathrm{P}\{t < 1.72\} = 0.95$. В R можно использовать функцию $\mathrm{qt}(0.95, 20)$. Наоборот $\mathrm{pt}(1.72, 20)$ приведет к значению 0.95.

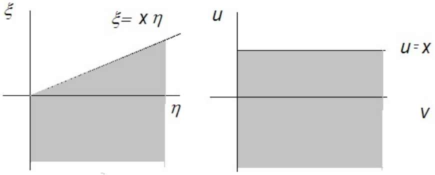
Функция распределения $\zeta = \frac{\xi}{\eta}$

$$\mathrm{P}\{\zeta < \mathrm{x}\} = \mathrm{P}\left\{\frac{\xi}{\eta} < \mathrm{x}\right\} = \mathrm{P}\{\xi < \mathrm{x}\eta\} =$$

$$= \mathrm{c_n} \iint\limits_{\substack{\eta>0 \\ \xi<\mathrm{x}\eta}} \eta^{\mathrm{n}-1} \mathrm{e}^{-\frac{\xi^2+\mathrm{n}\eta^2}{2}} \mathrm{d}\xi \mathrm{d}\eta = \mathrm{c_n} \int\limits_{-\infty}^{\mathrm{x}} \mathrm{du} \int\limits_{0}^{\infty} \mathrm{v}^{\mathrm{n}-1} \mathrm{e}^{-\frac{\mathrm{u}^2\mathrm{v}^2+\mathrm{n}\mathrm{v}^2}{2}} \mathrm{vdv},$$

используем замену $\xi = \mathrm{uv}$, $\eta = \mathrm{v}$ с якобианом преобразования

$$\mathrm{J} = \begin{vmatrix} \frac{\partial\xi}{\partial \mathrm{u}} & \frac{\partial\xi}{\partial \mathrm{v}} \\ \frac{\partial\eta}{\partial \mathrm{u}} & \frac{\partial\eta}{\partial \mathrm{v}} \end{vmatrix} = \begin{vmatrix} \mathrm{v} & \mathrm{u} \\ 0 & 1 \end{vmatrix} = \mathrm{v}\,.$$



В интегале $\int\limits_0^\infty v^{n-1} e^{-\frac{u^2v^2+nv^2}{2}} vdv$ заменим

$$\frac{(u^2+n)v^2}{2} = t, \quad v = \frac{\sqrt{2t}}{\sqrt{u^2+n}}, \quad vdv = \frac{dt}{u^2+n}.$$

$$P\{\zeta < x\} = c_n \int\limits_{-\infty}^{x} du \int\limits_0^\infty \left(\frac{\sqrt{2t}}{\sqrt{u^2+n}}\right)^{n-1} e^{-t} \frac{dt}{u^2+n} =$$

$$= c_n 2^{\frac{n-1}{2}} \int\limits_{-\infty}^{x} \frac{du}{(u^2+n)^{\frac{n+1}{2}}} \Gamma\left(\frac{n+1}{2}\right) =$$

$$= \frac{\sqrt{2}\left(\frac{n}{2}\right)^{\frac{n}{2}}}{\sqrt{\pi}\Gamma(\frac{n}{2})} 2^{\frac{n-1}{2}} \int\limits_{-\infty}^{x} \frac{du}{(u^2+n)^{\frac{n+1}{2}}} \Gamma\left(\frac{n+1}{2}\right)$$

Отсюда плотность распределения имеет вид $f_\zeta(x,n) =$

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

График плотности распределения Стьюдента

Асимптотическое свойство распределения Стьюдента

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

При любом фиксированном x имеем

$$-\frac{n+1}{2} \ln\left(1+\frac{x^2}{n}\right) \underset{n\longrightarrow\infty}{\longrightarrow} -\frac{x^2}{2} ,$$

Покажем, что при достаточно больших n

$$\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\frac{n}{2}}\Gamma(\frac{n}{2})} \approx 1$$

Воспользуемся формулой Стирлинга

$$\ln\Gamma(\lambda) = \left(\lambda - \frac{1}{2}\right)\ln\lambda - \lambda + \frac{1}{2}\ln(2\pi) + \frac{1}{12\lambda} + \mathrm{O}\left(\frac{1}{\lambda^3}\right).$$

$$\text{При}\quad \mathrm{p} = \frac{\mathrm{n}}{2} \longrightarrow \infty \qquad \ln\Gamma\left(\mathrm{p} + \frac{1}{2}\right) - \frac{1}{2}\ln\mathrm{p} - \ln\Gamma(\mathrm{p}) \approx$$

$$\approx \left(\mathrm{p} + \frac{1}{2} - \frac{1}{2}\right)\ln\left(\mathrm{p} + \frac{1}{2}\right) - \mathrm{p} - \frac{1}{2} - \frac{1}{2}\ln\mathrm{p} -$$

$$-\left(\mathrm{p} - \frac{1}{2}\right)\ln\mathrm{p} + \mathrm{p} = \mathrm{p}\ln\left(\mathrm{p} + \frac{1}{2}\right) - \frac{1}{2} - \mathrm{p}\ln\mathrm{p} =$$

$$= \mathrm{p}\ln\left(1 + \frac{1}{2\mathrm{p}}\right) - \frac{1}{2} \approx 0\,.$$

откуда

$$\mathrm{f}_\zeta(\mathrm{x},\mathrm{n}) \underset{\mathrm{n}\longrightarrow\infty}{\longrightarrow} \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{\mathrm{x}^2}{2}}\,.$$

Распределенная по Фишеру случайная величина имеет вид

$$\zeta = \frac{\frac{1}{\mathrm{n}} \sum\limits_{\mathrm{i}=1}^{\mathrm{n}} \xi_{\mathrm{i}}^2}{\frac{1}{\mathrm{m}} \sum\limits_{\mathrm{j}=1}^{\mathrm{m}} \eta_{\mathrm{j}}^2} = \frac{\mathrm{m}}{\mathrm{n}} \zeta_0 \,, \tag{3}$$

где ξ_{i}, η_{j} независимы, одинаково распределены $\mathcal{N}(0, \sigma)$. Не умаляя общности, будем считать $\sigma^2 = 1$. Тогда $\xi = \sum\limits_{\mathrm{i}=1}^{\mathrm{n}} \xi_{\mathrm{i}}^2$ и $\eta = \sum\limits_{\mathrm{j}=1}^{\mathrm{m}} \eta_{\mathrm{j}}^2$ имеют распределение (2) хи-квадрат с n и m степенями свободы соответственно.

Вычисление плотности распределения Фишера

$$\mathrm{P}\{\zeta_0 < \mathrm{x}\} = \mathrm{P}\left\{\frac{\xi}{\eta} < \mathrm{x}\right\} = \mathrm{P}\{\xi < \mathrm{x}\eta\}\,.$$

Совместное распределение величин ξ и η имеет вид

$$\kappa_{\mathrm{n}}(\xi)\kappa_{\mathrm{m}}(\eta) = \frac{2^{-\frac{\mathrm{n}}{2}}}{\Gamma(\frac{\mathrm{n}}{2})}\xi^{\frac{\mathrm{n}}{2}-1}\mathrm{e}^{-\frac{\xi}{2}}\frac{2^{-\frac{\mathrm{m}}{2}}}{\Gamma(\frac{\mathrm{m}}{2})}\eta^{\frac{\mathrm{m}}{2}-1}\mathrm{e}^{-\frac{\eta}{2}} = $$

$$= \mathrm{a}_{\mathrm{nm}}\xi^{\frac{\mathrm{n}}{2}-1}\eta^{\frac{\mathrm{m}}{2}-1}\mathrm{e}^{-\frac{\xi+\eta}{2}},$$

$$\text{где}\quad \mathrm{a}_{\mathrm{nm}} = \frac{2^{-\frac{\mathrm{m}+\mathrm{n}}{2}}}{\Gamma(\frac{\mathrm{n}}{2})\Gamma(\frac{\mathrm{m}}{2})}\,.$$

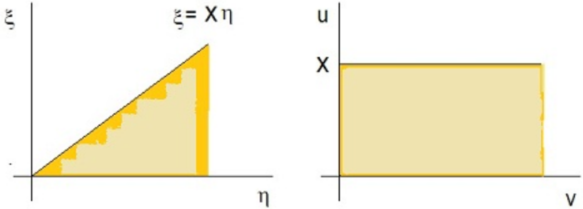
$$\mathrm{P}\{\zeta_0 < \mathrm{x}\} = \mathrm{a}_{\mathrm{nm}}\iint\limits_{\substack{\eta>0\\ 0<\xi<\mathrm{x}\eta}} \xi^{\frac{\mathrm{n}}{2}-1}\eta^{\frac{\mathrm{m}}{2}-1}\mathrm{e}^{-\frac{\xi+\eta}{2}}\mathrm{d}\xi\mathrm{d}\eta\,. \tag{4}$$

$$P\{\zeta_0 < x\} = a_{nm} \iint\limits_{\substack{\eta>0 \\ 0<\xi<x\eta}} \xi^{\frac{n}{2}-1}\eta^{\frac{m}{2}-1}e^{-\frac{\xi+\eta}{2}}d\xi d\eta\,. \qquad (5)$$

Замена $\xi = uv$, $\eta = v$ с якобианом преобразования

$$J = \begin{vmatrix} \frac{\partial\xi}{\partial u} & \frac{\partial\xi}{\partial v} \\ \frac{\partial\eta}{\partial u} & \frac{\partial\eta}{\partial v} \end{vmatrix} = \begin{vmatrix} v & u \\ 0 & 1 \end{vmatrix} = v\,.$$

Область интегрирования из $\eta > 0$, $0 < \xi < x\eta$ перейдет в область $v > 0, 0 < u < x$.



Используем $a_{nm} = \frac{2^{-\frac{m+n}{2}}}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})}$. Замена $\xi = uv$, $\eta = v$

$$P\{\zeta_0 < x\} = a_{nm} \int\int_{\substack{\eta>0 \\ 0<\xi<x\eta}} \xi^{\frac{n}{2}-1}\eta^{\frac{m}{2}-1}e^{-\frac{\xi+\eta}{2}}d\xi d\eta =$$

$$= a_{nm}\int_0^x u^{\frac{n}{2}-1}du\int_0^\infty v^{\frac{n+m}{2}-2}e^{-\frac{uv+v}{2}}vdv =$$

$$= a_{nm}\int_0^x u^{\frac{n}{2}-1}du\int_0^\infty v^{\frac{n+m}{2}-1}e^{-\frac{(u+1)v}{2}}dv\,.$$

Замени переменные $t = \frac{u+1}{2}v,\ v = \frac{2t}{u+1}, dv = \frac{2}{u+1}dt.$

$$a_{nm}\int_0^x u^{\frac{n}{2}-1}du\int_0^\infty t^{\frac{n+m}{2}-1}\left(\frac{2}{u+1}\right)^{\frac{n+m}{2}-1}\frac{2}{u+1}e^{-t}dt =$$

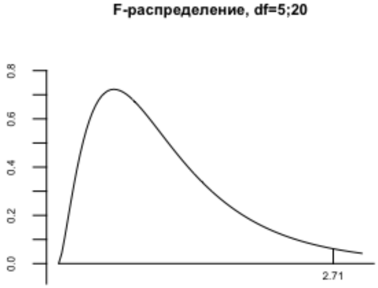
$$= \frac{2^{-\frac{m+n}{2}}}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})}\int_0^x u^{\frac{n}{2}-1}\left(\frac{2}{u+1}\right)^{\frac{n+m}{2}}du\Gamma\left(\frac{n+m}{2}\right) =$$

$$= \frac{\Gamma\left(\frac{n+m}{2}\right)}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})}\int_0^x \frac{u^{\frac{n}{2}-1}}{(u+1)^{\frac{n+m}{2}}}du\,.$$

Плотность распределения Фишера

$$f_0(x) = C_{nm} \frac{x^{\frac{n}{2}-1}}{(x+1)^{\frac{n+m}{2}}}, \quad \text{где} \quad C_{nm} = \frac{\Gamma\left(\frac{n+m}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)}$$

$$f(x) = \frac{n}{m} f_0\left(\frac{n}{m}x\right) = \frac{n}{m} C_{nm} \frac{\left(\frac{n}{m}x\right)^{\frac{n}{2}-1}}{\left(\frac{n}{m}x+1\right)^{\frac{n+m}{2}}}.$$



Распределение квадратичных форм

Лемма Фишера

Пусть $x_1, \dots, x_n$ — независимые и нормально распределенные $\mathcal{N}(0, \sigma)$ компоненты вектора X. В результате ортогонального преобразования $Y = CX$ имеем также независимые и нормально распределенные $\mathcal{N}(0, \sigma)$ компоненты $y_1, \dots, y_n$. Тогда квадратичная форма

$$Q(X) = \sum_{i=1}^{n} x_i^2 - y_1^2 - \dots - y_p^2$$

независима от $y_1, \dots, y_p$ и имеет плотность $\frac{1}{\sigma^2} \kappa_{n-p}\left(\frac{x}{\sigma^2}\right)$, где $\kappa_n(x)$ — плотность распределения хи-квадрат с n степенями свободы (2).

Распределение $\bar{x}$ и m_2

Теорема

Пусть $x_1, \dots, x_n \sim N(\mu, \sigma)$, $\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j$ и $m_2 = \frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})^2$ – выборочные среднее и второй центральный момент. Тогда

1. $\bar{x}$ и m_2 независимы;
2. $\bar{x} \sim N(\mu; \frac{\sigma}{\sqrt{n}})$;
3. статистика $\frac{nm_2}{\sigma^2} \sim \chi^2(n-1)$ имеет распределение хи-квадрат с $(n-1)$ степенью свободы.

Доказательство

Без ограничения общности будем считать $\mu = 0$.

$$\mathrm{n}\mathrm{m}_2 = \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{x}_\mathrm{i}^2 - \mathrm{n}\bar{\mathrm{x}}^2.$$

Выражение $\mathrm{n}\bar{\mathrm{x}}^2 = \left(\frac{\mathrm{x}_1}{\sqrt{\mathrm{n}}} + \ldots + \frac{\mathrm{x}_\mathrm{n}}{\sqrt{\mathrm{n}}}\right)^2$ есть квадрат линейной формы $\mathrm{c}_1\mathrm{x}_1 + \ldots + \mathrm{c}_\mathrm{n}\mathrm{x}_\mathrm{n}$, для которой $\mathrm{c}_1^2 + \ldots + \mathrm{c}_\mathrm{n}^2 = 1$. Поэтому применим лемму Фишера, положив $\mathrm{p} = 1$ и $\mathrm{y}_1 = \sqrt{\mathrm{n}}\bar{\mathrm{x}}$. Отсюда

$$\frac{\mathrm{n}\mathrm{m}_2}{\sigma^2} \sim \chi^2(\mathrm{n}-1). \tag{6}$$

Другая версия доказательства

Введем новые случайные величины $y_k = \frac{x_k - \mu}{\sigma} \sim \mathcal{N}(0, 1)$. Вектор $Y = (y_1, \ldots, y_n)^T$ имеет нулевое математическое ожидание $EY = 0$ и ковариационную матрицу $DY = EYY^T = I$. Рассмотрим $\xi = (\xi_1, \ldots, \xi_n)^T$ вида $\xi = CY$, где $CC^T = C^T C = I$, то есть C ортогональная матрица вида

$$C = \begin{bmatrix} \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ c_{21} & \cdots & c_{2n} \\ \cdots & \cdots & \cdots \\ c_{n1} & \cdots & c_{nn} \end{bmatrix}. \tag{7}$$

$$D\xi = E\xi\xi^T = E(CY)(CY)^T = ECYY^TC^T = C\underbrace{EYY^T}C^T = $$
$$= CDYC^T = CIC^T = CC^T = I.$$

Следовательно, компоненты вектора ξ независимы.

$$\xi_1 = \frac{1}{\sqrt{n}}(y_1 + \ldots + y_n) = \sqrt{n}\bar{y} = \frac{\bar{x} - \mu}{\sigma}\sqrt{n} \sim \mathcal{N}(0; 1). \tag{8}$$

Выразим nm_2/σ^2 через компоненты вектора ξ.

$$\frac{\mathrm{nm}_2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{\mathrm{k}=1}^{\mathrm{n}}(\mathrm{x_k} - \bar{\mathrm{x}})^2 = \sum_{\mathrm{k}=1}^{\mathrm{n}}\left(\frac{\mathrm{x_k} - \mu}{\sigma} - \frac{\bar{\mathrm{x}} - \mu}{\sigma}\right)^2 = \sum_{\mathrm{k}=1}^{\mathrm{n}}(\mathrm{y_k} - \bar{\mathrm{y}})^2 =$$
$$= \sum_{\mathrm{k}=1}^{\mathrm{n}}(\mathrm{y_k^2} - 2\mathrm{y_k}\bar{\mathrm{y}} - \bar{\mathrm{y}}^2) = \sum_{\mathrm{k}=1}^{\mathrm{n}}\mathrm{y_k^2} - 2\bar{\mathrm{y}}\sum_{\mathrm{k}=1}^{\mathrm{n}}\mathrm{y_k} + \mathrm{n}\bar{\mathrm{y}}^2 = \sum_{\mathrm{k}=1}^{\mathrm{n}}\mathrm{y_k^2} - \mathrm{n}\bar{\mathrm{y}}^2 =$$
$$= \mathrm{Y^TY} - \xi_1^2 = (\mathrm{C^T}\xi)^{\mathrm{T}}(\mathrm{C^T}\xi) - \xi_1^2 = \xi^{\mathrm{T}}\mathrm{CC^T}\xi - \xi_1^2 = \xi^{\mathrm{T}}\xi - \xi_1^2 =$$
$$= \sum_{\mathrm{k}=1}^{\mathrm{n}}\xi_{\mathrm{k}}^2 - \xi_1^2 = \sum_{\mathrm{k}=2}^{\mathrm{n}}\xi_{\mathrm{k}}^2.$$

Отсюда получаем, что

$$\frac{\mathrm{nm}_2}{\sigma^2} = \xi_2 + \ldots + \xi_{\mathrm{n}}^2 \ \sim \ \chi^2(\mathrm{n} - 1). \tag{9}$$

Выражение nm_2/σ^2 не зависит от ξ_1, а следовательно от $\bar{\mathrm{x}} = \mu + \xi\frac{\sigma}{\sqrt{\mathrm{n}}}$.

Доверительный интервал

Интервал, накрывающий истинное значение параметра распределения с заданной вероятностью P называется $P \cdot 100\%$-доверительным интервалом.

Пусть $x_1, \dots, x_n \sim \mathcal{N}(0, \sigma)$ при известной дисперсии σ

$$\bar{x} \sim \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \iff \frac{\bar{x} - \mu}{\sigma}\sqrt{n} \sim \mathcal{N}(0, 1)\,.$$

Обозначим $(1 - \frac{\alpha}{2})$- квантиль нормального распределения $\mathcal{N}(0, 1)$ через $z_{1-\frac{\alpha}{2}}$, $\xi = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$.

$$P\{|\xi| < z_{1-\frac{\alpha}{2}}\} = 1 - \alpha \iff P\{-z_{1-\frac{\alpha}{2}} < \xi \leq z_{1-\frac{\alpha}{2}}\} = 1 - \alpha$$

$$P\{-z_{1-\frac{\alpha}{2}} < \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \leq z_{1-\frac{\alpha}{2}}\} = 1 - \alpha,$$

отсюда

$$P\left\{\bar{x} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \mu \leq \bar{x} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right\} = 1 - \alpha\,.$$

Статистика Стьюдента

Пусть $x_1, \ldots, x_n$ — выборка из нормальной совокупности со средним μ и дисперсией σ^2. Тогда отношение

$$\tau = \frac{\bar{x} - \mu}{\sqrt{m_2}} \sqrt{n - 1} \qquad (10)$$

имеет распределение Стьюдента с $n - 1$ степенью свободы. Действительно, в (1) полагая $\xi_1 = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ и $\sum\limits_{i=1}^{n} \xi_i^2 = \frac{nm_2}{\sigma^2}$, получим

$$\tau = \frac{\xi_1}{\sqrt{\frac{1}{n-1} \sum\limits_{k=2}^{n} \xi_k^2}} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}} \sqrt{\frac{1}{n-1} \frac{nm_2}{\sigma^2}}} = \frac{\bar{x} - \mu}{\sqrt{m_2}} \sqrt{n - 1}.$$

Так как $m_2 = \frac{n-1}{n} S^2$, $\sqrt{m_2} = \frac{\sqrt{n-1}}{\sqrt{n}} S$,

$$\tau = \frac{\bar{x} - \mu}{S} \sqrt{n} \sim T(n - 1). \qquad (11)$$

Статистика Фишера для проверки гипотезы о равенстве дисперсий

$x_1, \dots, x_n \sim \mathcal{N}(\mu_1; \sigma)$ и $y_1, \dots, y_m \sim \mathcal{N}(\mu_2; \sigma)$

$$F = \frac{S_x^2}{S_y^2} \sim F(n; m), \tag{12}$$

где S_x^2 и S_y^2 – несмещенные оценки их дисперсий соответствующих выборок. Действительно, так как $(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$, то в соответствии с (3)

$$F = \frac{\frac{1}{n-1}(n-1)S_x^2/\sigma^2}{\frac{1}{m-1}(m-1)S_y^2/\sigma^2} = \frac{S_x^2}{S_y^2} \sim F(n; m).$$

Распределение разности выборочных средних

Пусть имеется две независимые выборки $x_1, \ldots, x_{n_1} \sim \mathcal{N}(\mu_1, \sigma)$ и $y_1, \ldots, y_{n_2} \sim \mathcal{N}(\mu_2, \sigma)$. Выборочные средние обозначим через $\bar{x}$ и $\bar{y}$, а несмещенные оценки дисперсии через S_1^2 и S_2^2. Тогда статистика

$$T = \frac{(\bar{x} - \bar{y} - (\mu_1 - \mu_2))\sqrt{n_1 + n_2 - 2}}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\sqrt{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}} \sim T(n_1 + n_2 - 2)$$

имеет распределение Стьюдента с числом степеней свободы, равным $n_1 + n_2 - 2$.

Доказательство

$\bar{x} \sim \mathcal{N}(\mu_1, \sigma/\sqrt{n_1})$ и $\bar{y} \sim \mathcal{N}(\mu_2, \sigma/\sqrt{n_2})$ независимы. Следовательно,

$$\bar{x} - \bar{y} \quad \sim \mathcal{N}\left(\mu_1 - \mu_2, \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right), \text{ откуда имеем}$$

$$\xi = \frac{\bar{x} - \bar{y} - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim \mathcal{N}(0; 1).$$

С другой стороны, из аддитивности распределения χ^2

$$(n_1 - 1)S_1^2/\sigma^2 \sim \chi^2(n_1 - 1), \quad (n_2 - 1)S_2^2/\sigma^2 \sim \chi^2(n_2 - 1).$$

$$\eta = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{\sigma^2} \sim \chi^2(n_1 + n_2 - 2)$$

$$T = \frac{\xi}{\sqrt{\frac{1}{n_1+n_2-2}\eta}} = \frac{(\bar{x} - \bar{y} - (\mu_1 - \mu_2))\sqrt{n_1 + n_2 - 2}}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\sqrt{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}} \sim T(n_1 + n_2 - 2).$$

Проверка равенства средних в случае неодинаковых известных дисперсий

В случае известных дисперсий σ_1^2 и σ_2^2 (возможно неодинаковых), учитывая нормальность средних

$$\bar{\mathrm{x}} \sim \mathcal{N}(\mu_1, \sigma_1/\sqrt{\mathrm{n}_1}), \quad \bar{\mathrm{y}} \sim \mathcal{N}(\mu_2, \sigma_2/\sqrt{\mathrm{n}_2}),$$

вычисляем дисперсию разности

$$\mathrm{D}(\bar{\mathrm{x}} - \bar{\mathrm{y}}) = \frac{\sigma_1^2}{\mathrm{n}_1} + \frac{\sigma_2^2}{\mathrm{n}_2},$$

откуда получаем

$$\mathrm{Z} = \frac{\bar{\mathrm{x}} - \bar{\mathrm{y}} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{\mathrm{n}_1} + \frac{\sigma_2^2}{\mathrm{n}_2}}} \sim \mathcal{N}(0; 1). \tag{13}$$

Проверка равенства средних в случае неодинаковых неизвестных дисперсий

В случае неизвестных неодинаковых дисперсий $\sigma_1^2 \neq \sigma_2^2$ статистика T имеет приближенно распределение Стьюдента с нецелым числом степеней свободы.

$$T = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim T(k), \tag{14}$$

$$\text{где } k = \frac{(S_1^2/n_1 + S_2^2/n_2)^2}{\frac{(S_1^2/n_1)^2}{n_1-1} + \frac{(S_2^2/n_2)^2}{n_2-1}}.$$

Заметим, что

$$\mathbb{E}\left(\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}\right) = \sigma^2,$$

так как

$$\mathbb{E}\left(\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{\sigma^2}\right) = n_1 + n_2 - 2\,.$$

Критерий однородности для двух зависимых выборок

$(x_i, y_i), i = 1, 2, \ldots, n$, парная выборка одних и тех же наблюдений при разных условиях
Требуется проверить значимость изменений.
Вводим $z_i = y_i - x_i$ и проверяем гипотезу $H_0 : \mu_z = 0$.

Необходимые функции в R

- равенство дисперсий var.test
- равенство средних
  t.test(x, y = NULL, alternative =
  c("two.sided "less "greater"),
  mu = 0, paired = FALSE, var.equal = FALSE, conf.level =
  0.95, ...)

Отношение правдоподобия

Если H_0, H_1 две гипотезы о принадлежности наблюдения x статистической популяции с плотностью f_0 или f_1, то из теоремы Байеса

$$P(H_i|x) = \frac{P(H_i)f_i(x)}{P(H_0)f_0(x) + P(H_1)f_1(x)}, \ i = 0, 1\,,$$

отсюда

$$\frac{P(H_1)f_1(x)}{P(H_0)f_0(x)} = \frac{P(H_1|x)}{P(H_0|x)}$$

$$\log\frac{f_1(x)}{f_0(x)} = \log\frac{P(H_1|x)}{P(H_0|x)} - \log\frac{P(H_1)}{P(H_0)}$$

разность между логарифмами отношения шансов в пользу H_1 до и после наблюдения x.

Информацию в точке $\xi = x$ для различения в пользу H_1 против H_0 выражаем в виде $\log\frac{f_1(x)}{f_0(x)}$

Средняя информация для различения в пользу H_1 против H_0

$$I(1:0) = \int \log \frac{f_1(x)}{f_0(x)} f_1(x) dx =$$

$$-I(0:1) = \int \log \frac{f_1(x)}{f_0(x)} f_0(x) dx$$

Расхождение $J(1:0)$

$$J(1:0) = I(1:0) + I(0:1) = \int (f_1(x) - f_0(x)) \log \frac{f_1(x)}{f_0(x)} dx$$