Test korelace

- náhodné proměnné *x* a *y*
- naměříme hodnoty $x_1, x_2, \dots, x_N$ a $y_1, y_2, \dots, y_N$
- vypočítáme odhad korelace $\hat{\rho}(x,y) = \dfrac{\langle xy \rangle - \langle x \rangle \langle y \rangle}{\hat{\sigma_x}\hat{\sigma_y}}$ s chybou $\sigma_\rho \approx \dfrac{1-\rho^2}{\sqrt{N-1}}$
- Je korelace statisticky významná?

- použijeme testovací statistiku $f(t|H_0)$

  známá hustota pravděpodobnosti $f(t)$ a distribuční funkce $F(t)$

  transformace $t \to \rho$ (testovací proměnná)

  nulová hypotéza H_0 (předpoklad nulové korelace proměnných *x* a *y*)

  hladina signifikance P_α (typicky 5 % nebo 1 %), pro $P < P_\alpha$ odmítneme hypotézu H_0

$$P = P\left(t > \hat{t}(\hat{\rho})\right) = 1 - F\left(\hat{t}(\hat{\rho})\right)$$

Test korelace – Fisherova transformace

- **Fisherova transformace**

transformace $$\boxed{z = \frac{1}{2} \ln \frac{1+\rho}{1-\rho}}$$

Pokud platí nulová hypotéza, má proměnná z normální rozdělení $N(\mu, \sigma)$,

kde očekávaná hodnota $\mu_z = 0$ a standardní odchylka $\sigma_z = \dfrac{1}{\sqrt{N-3}}$

testovací proměnná $t = \dfrac{z}{\sigma}$

Pokud platí nulová hypotéza, má proměnná t normální rozdělení $N(0,1)$.

hladina signifikance $P < P_\alpha$

$$\boxed{P = 2\left[1 - F\left(\left|\hat{t}(\hat{\rho})\right|\right)\right] = 2\left\{1 - \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{\left|\hat{t}(\hat{\rho})\right|}{\sqrt{2}}\right)\right]\right\}}$$

Test korelace – Fisherova transformace

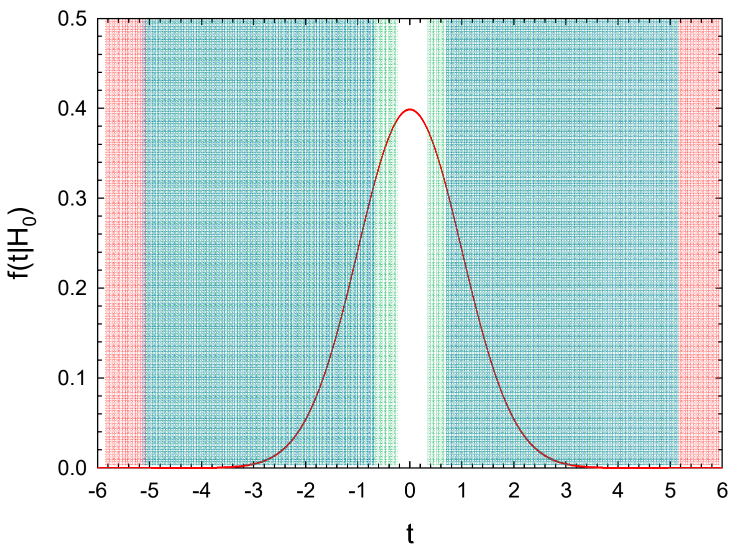
- **Fisherova transformace**

Příklad: N = 39, proměnné x, y, z

$$\hat{\rho}(x,y) = 0.694 \qquad |t_{x,y}| = 5.137 \qquad P_{x,y} = 3 \times 10^{-7}$$

$$\hat{\rho}(x,z) = -0.111 \qquad |t_{x,z}| = 0.669 \qquad P_{x,z} = 0.50$$

$$\hat{\rho}(y,z) = -0.059 \qquad |t_{y,z}| = 0.355 \qquad P_{y,z} = 0.72$$



proměnné x, y závislé

proměnné x, z a y, z nezávislé

Test korelace – studentovo rozdělení

- **studentovo rozdělení**

  transformace $$t = \rho\sqrt{\frac{N-2}{1-\rho^2}}$$

  Pokud platí nulová hypotéza, má proměnná t studentovo rozdělení s $N-2$ stupni volnosti.

  studentovo rozdělení s ν stupni volnosti $$f(x|\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

  William Sealy Gosset („student“)

  statistika na malém počtu vzorků

Test korelace – studentovo rozdělení

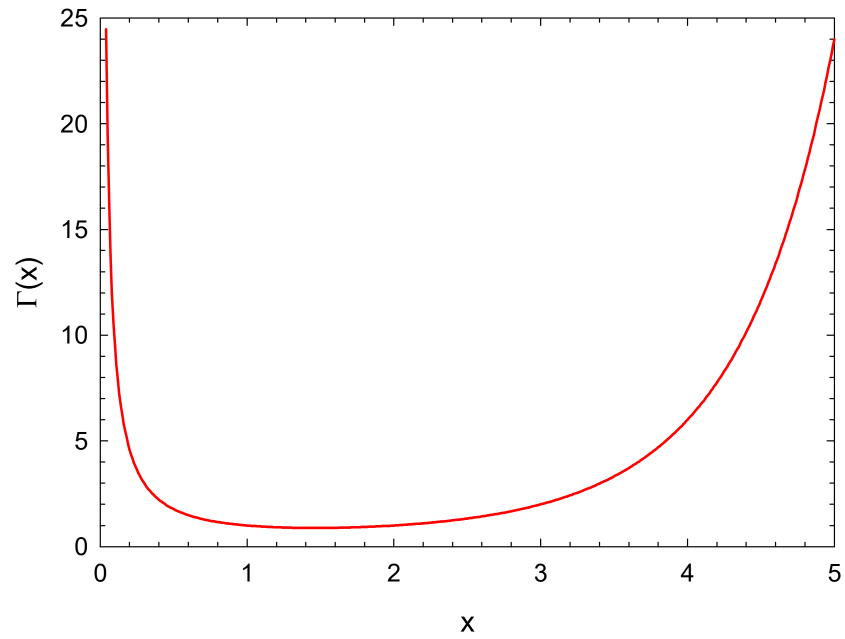
- **studentovo rozdělení** s ν stupni volnosti $$f(x|\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

gama funkce $$\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\mathrm{d}t \qquad x \geq 0$$

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

$$\Gamma(n) = (n-1)! \qquad n \in \mathbf{N}$$

$$\Gamma(x+1) = x\Gamma(x) \qquad x \in \mathbf{R}$$



ROOT: `ROOT::Math::tgamma(x)`

Excel: `EXP(GAMMALN(x`

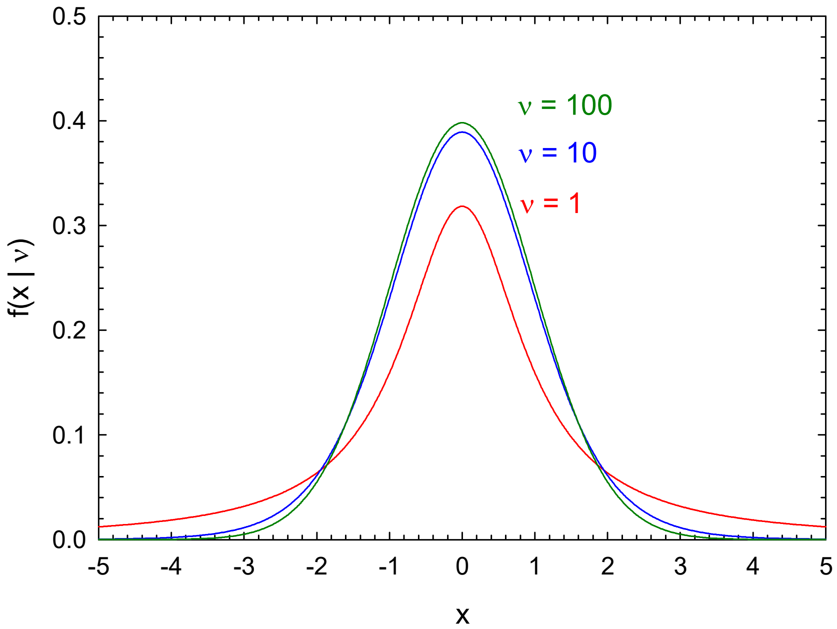
Test korelace – studentovo rozdělení

- **studentovo rozdělení** s ν stupni volnosti

$$f(x|\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

Python:
```
from scipy.stats import t
t.pdf(x,nu)
```

ROOT:
```
ROOT::Math::tdistribution_pdf(x,nu)
```



$$f(x|\nu=1) = \frac{1}{\pi}\frac{1}{1+x^2}$$

$$f(x|\nu\to\infty) \to \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)$$

$$E[x] \equiv \mu = 0$$

$$V[x] \equiv \sigma^2 = \frac{\nu}{\nu-2} \qquad \nu > 2$$

Test korelace – studentovo rozdělení

- **studentovo rozdělení**

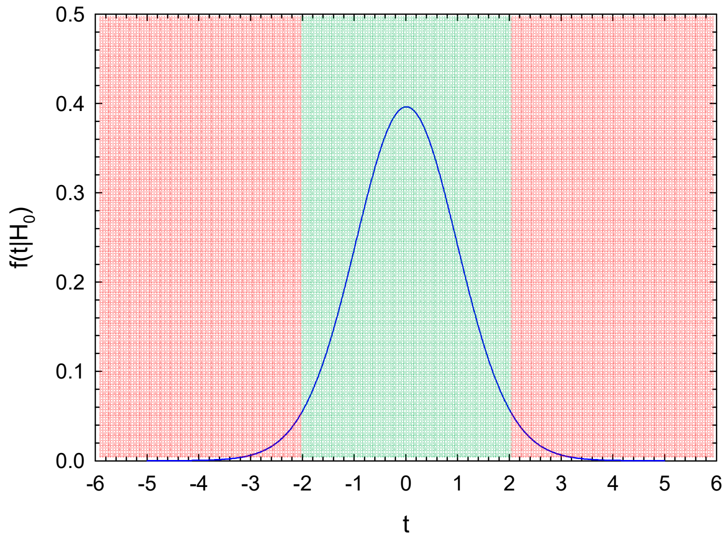
  transformace
  $$\boxed{t = \rho\sqrt{\frac{N-2}{1-\rho^2}}}$$

  Pokud platí nulová hypotéza, má proměnná t studentovo rozdělení s $N-2$ stupni volnosti.

hladina signifikance $P < P_\alpha$

$$P = 2\left[1 - T_\nu\left(\left|\hat{t}(\hat{\rho})\right|\right)\right]$$

distribuční funkce studentova rozdělení



konfidenční interval $\left(-T_\nu^{-1}(P_\alpha), T_\nu^{-1}(P_\alpha)\right)$

inverzní funkce k distribuční funkci studentova rozdělení

Test korelace – studentovo rozdělení

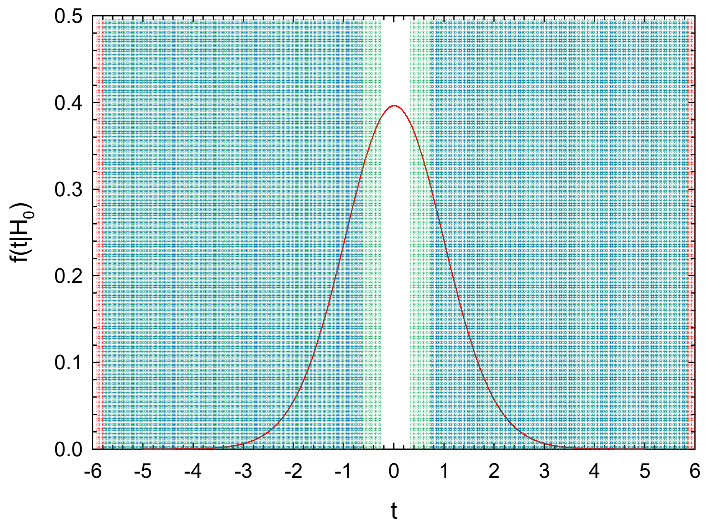
- **studentovo rozdělení**

Příklad: *N* = 39, proměnné *x*, *y*, *z*

$$\hat{\rho}(x,y) = 0.694 \qquad |t_{x,y}| = 5.867 \qquad P_{x,y} = 9 \times 10^{-7}$$

$$\hat{\rho}(x,z) = -0.111 \qquad |t_{x,z}| = 0.679 \qquad P_{x,z} = 0.50$$

$$\hat{\rho}(y,z) = -0.059 \qquad |t_{y,z}| = 0.360 \qquad P_{y,z} = 0.72$$



proměnné *x*, *y* závislé

proměnné *x*, *z* a *y*, *z* nezávislé

Přenos chyb

- náhodné proměnné $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ $E[x_i] = \mu_i$ $\boldsymbol{\mu} = \mu_1, \mu_2, ..., \mu_n$
- výsledná veličina $y(\boldsymbol{x})$ $V_{ij} = \mathrm{cov}(x_i, x_j)$
- Taylorův rozvoj $y(\boldsymbol{x}) \approx y(\boldsymbol{\mu}) + \sum_{i=1}^{n} \left.\frac{\partial y}{\partial x_i}\right|_{x=\mu} (x_i - \mu_i)$
- **očekávaná hodnota** $\boxed{E\left[y(\boldsymbol{x})\right] \approx y(\boldsymbol{\mu})}$

Přenos chyb

- náhodné proměnné $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ $E[x_i] = \mu_i$ $\boldsymbol{\mu} = \mu_1, \mu_2, ..., \mu_n$
- výsledná veličina $y(\boldsymbol{x})$ $V_{ij} = \mathrm{cov}(x_i, x_j)$
- Taylorův rozvoj $y(\boldsymbol{x}) \approx y(\boldsymbol{\mu}) + \sum_{i=1}^{n} \left.\frac{\partial y}{\partial x_i}\right|_{x=\mu} (x_i - \mu_i)$
- **rozptyl** $V\left[y(\boldsymbol{x})\right] = E\left[y^2(\boldsymbol{x})\right] - \left(E\left[y(\boldsymbol{x})\right]\right)^2$

$$y^2(\boldsymbol{x}) \approx y^2(\boldsymbol{\mu}) + 2y(\boldsymbol{\mu}) \sum_{i=1}^{n} \left.\frac{\partial y}{\partial x_i}\right|_{x=\mu} (x_i - \mu_i) + \sum_{i=1}^{n} \left.\frac{\partial y}{\partial x_i}\right|_{x=\mu} (x_i - \mu_i) \sum_{j=1}^{n} \left.\frac{\partial y}{\partial x_j}\right|_{x=\mu} (x_j - \mu_j)$$

$$E\left[y^2(\boldsymbol{x})\right] \approx y^2(\boldsymbol{\mu}) + \sum_{i=1}^{n} \sum_{j=1}^{n} \left.\frac{\partial y}{\partial x_i}\right|_{x=\mu} \left.\frac{\partial y}{\partial x_j}\right|_{x=\mu} \mathrm{cov}(x_i, x_j)$$

$$\boxed{V\left[y(\boldsymbol{x})\right] \approx \sum_{i=1}^{n} \sum_{j=1}^{n} \left.\frac{\partial y}{\partial x_i}\right|_{x=\mu} \left.\frac{\partial y}{\partial x_j}\right|_{x=\mu} \mathrm{cov}(x_i, x_j)}$$

Přenos chyb – nezávislé náhodné proměnné

- náhodné proměnné $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ $E[x_i] = \mu_i$ $\boldsymbol{\mu} = \mu_1, \mu_2, ..., \mu_n$
- výsledná veličina $y(\boldsymbol{x})$ $V_{ij} = \mathrm{cov}(x_i, x_j)$
- nezávislé proměnné $\mathrm{cov}(x_i, x_j) = \begin{cases} \sigma_i^2 & \text{pro } i = j \\ 0 & \text{pro } i \neq j \end{cases}$

- **očekávaná hodnota** $E\left[y(\boldsymbol{x})\right] \approx y(\boldsymbol{\mu})$
- **rozptyl** $V\left[y(\boldsymbol{x})\right] \approx \sum_{i=1}^{n} \left(\left. \frac{\partial y}{\partial x_i} \right|_{x=\mu} \right)^2 \sigma_i^2$

Přenos chyb – součet náhodných proměnných

- náhodné proměnné $\boldsymbol{x} = (x_1, x_2)$ $\boldsymbol{\mu} = (\mu_1, \mu_2)$ $\boldsymbol{\sigma} = (\sigma_1, \sigma_2)$
- výsledná veličina $y(\boldsymbol{x}) = x_1 + x_2$

1. nezávislé proměnné

   očekávaná hodnota $E\left[y(\boldsymbol{x})\right] = \mu_1 + \mu_2$

   rozptyl $V\left[y(\boldsymbol{x})\right] = V[x_1] + V[x_2] = \sigma_1^2 + \sigma_2^2$

2. závislé proměnné

   očekávaná hodnota $E\left[y(\boldsymbol{x})\right] = \mu_1 + \mu_2$

   rozptyl
$$V\left[y(\boldsymbol{x})\right] = V[x_1] + V[x_2] + 2\,\mathrm{cov}(x_1, x_2)$$
$$= \sigma_1^2 + \sigma_2^2 + 2\,\mathrm{cov}(x_1, x_2)$$

Přenos chyb – aritmetický průměr

- náhodné proměnné $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ $\quad \boldsymbol{\mu} = \mu_1, \mu_2, ..., \mu_n$ $\quad \boldsymbol{\sigma} = (\sigma_1, \sigma_2, ...\sigma_n)$
- výsledná veličina $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
- nezávislé proměnné

  očekávaná hodnota $$E[\bar{x}] = \frac{1}{n}\sum_{i=1}^{n} \mu_i$$

  rozptyl $$V[\bar{x}] = \frac{1}{n^2}\sum_{i=1}^{n} V[x_i] = \frac{1}{n^2}\sum_{i=1}^{n} \sigma_i^2$$

- všechny σ_i stejné: $\sigma_i = \sigma \quad \Rightarrow \quad V[\bar{x}] = \frac{\sigma^2}{n}$

⇒ chyba aritmetického průměru $\boxed{\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}}$

Přenos chyb

- nezávislé náhodné proměnné

- součet / rozdíl

$$y = a + b$$

$$y = a - b$$

$$\boxed{\sigma_y = \sqrt{\sigma_a^2 + \sigma_b^2}}$$

- součin/podíl

$$y = a \cdot b$$

$$y = \frac{a}{b}$$

$$\boxed{\frac{\sigma_y}{y} = \sqrt{\frac{\sigma_a^2}{a^2} + \frac{\sigma_b^2}{b^2}}}$$

- mocnina

$$y = a^n$$

$$\boxed{\frac{\sigma_y}{y} = n\frac{\sigma_a}{a}}$$