Pojmy a vzorce

Stránky vznikly za podpory FRVŠ (projekt číslo 1645/2007).

Kombinatorika:

• **faktoriály** $0! = 1,\ n! = n \cdot (n-1) \cdot \cdots \cdot 3 \cdot 2 \cdot 1$ • **kombinační čísla** $\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}$

• **kombinace** - neuspořádané k-tice (množiny) z n prvků - bez opakování $C_k(n) = \binom{n}{k}$

- s opakováním $C'_k(n) = \binom{n+k-1}{k}$

• **variace** - uspořádané k-tice z n prvků - bez opakování $V_k(n) = \frac{n!}{(n-k)!}$

- s opakováním $V'_k(n) = n^k$

• **permutace** - uspořádané k-tice z k prvků bez opakování $P(k) = k!$

- uspořádané k-tice z k prvků, kde je $k_1, k_2, \ldots, k_r$ stejných prvků, $k_1 + k_2 + \cdots + k_r = k$

$$P(k_1, k_2, \ldots, k_r) = \frac{(k_1 + k_2 + \cdots + k_r)!}{k_1! \cdot k_2! \cdot \cdots \cdot k_r!}$$

$$V_k(n) = C_k(n) \cdot P(k)$$

Pravděpodobnost:

• **náhodný děj** .... děj (proces, činnost...), ve kterém se projevuje náhoda (= nejsme schopni dopředu jednoznačně předpovědět, jak dopadne)

výsledky pokusu shrneme do množiny Ω (úplný a bezesporný systém)

• **náhodný jev** .... podmnožina Ω (odpovídající výroku V o výsledku náhodného pokusu), $A \subset \Omega$ $A = [V]$

• **pravděpodobnost** .... přiřazuje náhodnému jevu A číslo $P(A)$ mezi 0 a 1, vyjadřující, jak hodně či málo očekáváme, že náhodný jev A nastane

$\Omega \neq \emptyset$

$\mathcal{S}$ systém podmnožin Ω - $\sigma-$algebra

$P : \mathcal{S} \longrightarrow \mathbb{R}$ pravděpodobnostní míra

vlastnosti:

$P(\emptyset) = 0, \quad P(\Omega) = 1, \quad P(\Omega \setminus A) = 1 - P(A)$

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$

atd.

• **nezávislost náhodných jevů**:

A, B **nezávislé** .... $P(A \cap B) = P(A) \cdot P(B)$

A, B, C **nezávislé** .... $P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$

$P(A \cap B) = P(A) \cdot P(B) \qquad P(A \cap C) = P(A) \cdot P(C) \qquad P(B \cap C) = P(B) \cdot P(C)$

atd.

A, B neslučitelné ....$A \cap B = \emptyset$

• **podmíněná pravděpodobnost** (pro $P(B) \neq 0$) .... $P(A/B) = \frac{P(A \cap B)}{P(B)}$

• **věta o úplné pravděpodobnosti**:

$B_1, B_2, \ldots$ rozklad $\Omega, P(B_1) \neq 0, P(B_2) \neq 0, \ldots \quad \Longrightarrow$

$$\Longrightarrow \quad P(A) = P(A/B_1) \cdot P(B_1) + P(A/B_2) \cdot P(B_2) + \ldots$$

• **Bayesova věta**:

$$B_1, B_2, \dots \text{ rozklad } \Omega,\ P(B_1) \neq 0, P(B_2) \neq 0, \dots,\ P(A) \neq 0 \quad \Longrightarrow \quad P(B_m/A) = \frac{P(A/B_m) \cdot P(B_m)}{\sum_k P(A/B_k) \cdot P(B_k)}$$

(rozklad Ω .... jevy jsou po dvou neslučitelné a $B_1 \cup B_2 \cup \dots = \Omega$)

Náhodná veličina:

• **náhodná veličina** .... nejednoznačné vyjádření vlastnosti týkající se objektu náhodného děje
$X : (\Omega, \mathcal{S}) \longrightarrow \mathbb{R}$ Borelovsky měřitelné zobrazení
• **rozdělení pravděpodobnosti** náh. veličiny X .... každé Borelovské množině $I \in \mathbb{R}$ přiřazuje $P[X \in I]$
• **distribuční funkce** .... $F(a) = F_X(a) = P[X < a]$

vlastnosti: F neklesající, spojitá zleva, $F(-\infty) = 0$, $F(\infty) = 1$

• **diskrétní rozdělení** (náh. veličiny X) .... zadáno pomocí **hodnot** x_k a **příslušných pravděpodobností** $p_k = P[X = x_k]$, kde $\sum_k p_k = 1$

$$P[X \in I] = \sum_{k;\ x_k \in I} p_k$$

střední hodnota .... $EX = \sum_k x_k \cdot p_k$

modus $\hat{x}$ .... některé z čísel, pro něž $P[X = \hat{x}] \geq P[X = x]$ pro všechna x

rozptyl .... $DX = E((X - EX)^2) = \sum_k (x_k - EX)^2 \cdot p_k$

• **spojité rozdělení** (náh. veličiny X) .... zadáno pomocí **hustoty** f (nezáporné měřitelné funkce), pro niž $\int_{-\infty}^{\infty} f(x)\,dx = 1$

$$P[X \in I] = \int_I f(x)\,dx$$

střední hodnota .... $EX = \int_{-\infty}^{\infty} x f(x)\,dx$

modus $\hat{x}$ .... některé z čísel, pro něž $f(\hat{x}) \geq f(x)$ pro všechna x

rozptyl ....$DX = E((X - EX)^2) = \int_{-\infty}^{\infty} (x - EX)^2 f(x)\,dx$

(pro oba typy rozdělení)

p-**kvantil** .... $\tilde{x}_p = \inf\{x; F(x) \geq p\}$

medián $\tilde{x}$....... některé z čísel, pro něž $F(\tilde{x}^-) \leq \frac{1}{2} \leq F(\tilde{x})$

• **náhodný vektor** .... $(X, Y) : (\Omega, \mathcal{S}) \longrightarrow \mathbb{R} \times \mathbb{R}$ Borelovsky měřitelné zobrazení

• **rozdělení pravděpodobnosti** náhodného vektoru (X, Y) .... každé Borelovské množině $B \in \mathbb{R} \times \mathbb{R}$ přiřazuje $P[(X, Y) \in B]$

• **(sdružená) distribuční funkce** .... $F(a, b) = F_{(X,Y)}(a, b) = P[X < a, Y < b]$
• **marginální distribuční funkce** .... $F_X(a) = P[X < a] = F(a, \infty) \quad F_Y(a) = P[Y < a] = F(\infty, a)$

• **diskrétní rozdělení** (náh. vektoru (X, Y)) .... zadáno pomocí **hodnot** x_k, y_l a **příslušných (sdružených) pravděpodobností** $p_{k,l} = P[X = x_k, y = y_l]$, kde $\sum_{k,l} p_{k,l} = 1$

$$P[(X, Y) \in B] = \sum_{k,l;\ (x_k, y_l) \in B} p_{k,l}$$

marginální pravděpodobnosti .... $p_k = P[X = x_k] = \sum_i p_{ki} \quad q_k = P[X = y_k] = \sum_i p_{ik}$

kovariance .... $CX = E((X - EX)(Y - EY)) = \sum_{k,l} (x_k - EX)(y_l - EY) \cdot p_{kl}$

• **spojité rozdělení** (náh. vektoru (X, Y)) .... zadáno pomocí **(sdružené) hustoty** $f = f_{(X,Y)}$ (nezáporné měřitelné funkce), pro niž $\iint_{\mathbb{R} \times \mathbb{R}} f(x, y)\,dxdy = 1$

$$P[(X, Y) \in B] = \iint_B f(x, y)\,dxdy$$

marginální hustoty .... $f_X(a) = \int_{-\infty}^{\infty} f_{(X,Y)}(a,t)\,dt \qquad f_Y(a) = \int_{-\infty}^{\infty} f_{(X,Y)}(t,a)\,dt$

kovariance .... $C(X,Y) = E((X-EX)(Y-EY)) = \int_{-\infty}^{\infty} (x-EX)(y-EY) f(x,y)\,dxdy$

vektor středních hodnot a kovarianční matice .... $\begin{pmatrix} EX & EY \end{pmatrix} \qquad \begin{pmatrix} DX & C(X,Y) \\ C(X,Y) & DY \end{pmatrix}$

korelace .... $\rho = \dfrac{C(X,Y)}{\sqrt{DX \cdot DY}}$

- **nezávislost náhodných veličin**

$X,\ Y$ **nezávislé** ...... $[X \in B_1], [Y \in B_2]$ nezávislé pro všechny Borelovské B_1 a B_2

$F(a,b) = F_X(a) \cdot F_Y(b)$ pro všechna a a b

$f(a,b) = f_X(a) \cdot f_Y(b)$ pro všechna a a b (v případě spojitého rozdělení)

$p_{k,l} = p_k \cdot q_l$ pro všechna k a l (v případě diskrétního rozdělení)

$X,\ Y$ nezávislé $\quad\Longrightarrow\quad C(X,Y) = 0$

podobně pro náhodné vektory $(X_1, X_2, \dots, X_n)$

Typy rozdělení:

- **alternativní** $A(p) \quad 0 < p < 1$

$x_0 = 0,\ x_1 = 1 \qquad p_0 = 1-p,\ p_1 = p$

$EX = p \qquad DX = p(1-p)$

(počet černých kuliček při losování jedné z N kuliček mezi nimiž je Z černých a N-Z bílých, $p = \frac{Z}{N}$)

- **hypergeometrické** $HG(N,Z,n) \quad n < N,\ Z < N$

$x_k = k = 0, \dots, n\ , \quad k \le Z,\ n-k \le N-Z \qquad p_k = \dfrac{\binom{Z}{k} \cdot \binom{N-Z}{n-k}}{\binom{N}{n}}$

$EX = n \cdot \dfrac{Z}{N} \qquad DX = n \cdot \dfrac{Z}{N} \cdot \left(1 - \dfrac{Z}{N}\right) \cdot \dfrac{N-n}{N-1}$

(počet černých kuliček při losování n kuliček bez vracení z N kuliček mezi miniž je Z černých a N-Z bílých)

- **binomické** $Bi(n,p) \quad 0 < p < 1$

$x_k = k = 0, \dots, n \qquad p_k = \dbinom{n}{k} \cdot p^k (1-p)^{n-k}$

$EX = np \qquad DX = np(1-p)$

(počet černých kuliček při losování n kuliček s vracením z N kuliček mezi miniž je Z černých a N-Z bílých, $p = \frac{Z}{N}$)
$X_1, \dots, X_n \sim A(p)$ nezávislé $\Longrightarrow \quad X_1 + \dots + X_n \sim Bi(n,p)$
$X \sim Bi(m,p), Y \sim Bi(n,p)$ nezávislé $\Longrightarrow \quad X + Y \sim Bi(m+n,p)$
(podobně multinomické rozdělení)

- **geometrické** $G(p) \quad 0 < p < 1$

$x_k = k = 1, 2, \dots \qquad p_k = p \cdot (1-p)^{k-1}$

$EX = \dfrac{1}{p} \qquad DX = \dfrac{1}{p}$

(pořadí první černé kuličky při losování kuliček s vracením z N kuliček mezi miniž je Z černých a N-Z bílých, $p = \frac{Z}{N}$, diskrétní poruchovost, pořadí prvního impulzu)

- **Poissonovo** $Po(\lambda) \quad \lambda > 0$

$x_k = k = 0, 1, 2, \dots \qquad p_k = \mathrm{e}^{-\lambda} \cdot \dfrac{\lambda^k}{k!}$

$EX = \lambda \qquad DX = \lambda$

(počet náhodných impulsů za čas)
$X_1, X_2, \dots \sim Bi(n, p_n),\ n \cdot p_n \to \lambda \Longrightarrow \quad X_n \to X \sim Po(\lambda)$

• **exponenciální** $Exp(\lambda)$ $\quad \lambda > 0$

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases} \qquad F(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

$$EX = \frac{1}{\lambda} \qquad DX = \frac{1}{\lambda^2}$$

(doba do prvního náhodného impulsu)

pro $X \sim Exp(\lambda)$ platí $P[X > a + b/X > a] = P[X > b]$ (zapomínání)

• **gama** $\Gamma(n, \lambda)$ $\quad \lambda > 0$

$$f(x) = \begin{cases} e^{-\lambda x} \cdot \frac{\lambda^n x^{n-1}}{(n-1)!} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

$$EX = \frac{n}{\lambda} \qquad DX = \frac{n}{\lambda^2}$$

(doba do n-tého náhodného impulsu)

• **normální** $N(\mu, \sigma^2)$ $\quad \sigma > 0$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$EX = \mu \qquad DX = \sigma^2$$

(limitní rozdělení, viz centrální limitní věta)

kvantil $u(p)$; $P[X < u(p)] = p$, kde $X \sim N(0;1)$

$X \sim N(\mu, \sigma^2) \implies a \cdot X + b \sim N(a \cdot \mu + b, a^2 \cdot \sigma^2)$

$X \sim N(\mu, \sigma^2) \implies \dfrac{X - \mu}{\sigma} \sim N(0, 1)$

$X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$ nezávislé $\implies X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

$X_1, \ldots, X_n$ nezávislé $X_k \sim N(\mu_k, \sigma_k^2) \implies \sum_k a_k \cdot X_k + b \sim N(\sum_k a_k \cdot \mu_k + b, \sum_k a_k^2 \cdot \sigma_k^2)$

pro distribuční funkci Φ rozdělení $N(0, 1)$ platí $\Phi(-x) + \Phi(x) = 1$, pro kvantil $u(p) = -u(1 - p)$

(podobně vícerozměrné normální rozdělení)

• **Pearsonovo chí-kvadrát** χ_n^2

$$f_n(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{n}{2}} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

$$EX = n \qquad DX = 2n$$

kvantil $\chi_n^2(p)$; $P[X < \chi_n^2(p)] = p$, kde $X \sim \chi_n^2$

$X_1, \ldots, X_m \sim N(0, 1)$ nezávislé $\implies X_1^2 + \cdots + X_m^2 \sim \chi_m^2$

$X \sim \chi_m^2, Y \sim \chi_n^2$ nezávislé $\implies X + Y \sim \chi_{m+n}^2$

(pro $n = 2$ Raileighovo, pro $n = 3$ Maxwellovo rozdělení)

• **Studentovo** t_n

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

$$EX = 0 \qquad DX = \frac{n}{n-2}$$

kvantil $t_n(p)$; $P[X < t_n(p)] = p$, kde $X \sim t_n$

$X \sim N(0, 1), Y \sim \chi_m^2$ nezávislé $\implies \dfrac{X}{\sqrt{\frac{Y}{m}}} \sim t_m$

pro distribuční funkci F rozdělení t_n platí $F(-x) = 1 - F(x)$ pro kvantil $t_n(p) = -t_n(1 - p)$

• **Fisherovo - Snedecorovo** $F_{m,n}$

$$f_{m,n}(x) = \begin{cases} \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \left(\frac{m}{n}\right)^{\frac{m}{2}} x^{\frac{m}{2}-1} \cdot \left(\frac{m}{n}x - +1\right)^{-\frac{m+n}{2}} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

$$EX = \frac{n}{n-2} \qquad DX = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$$

kvantil $F_{m,n}(p)$; $P[X < F_{m,n}(p)] = p$, kde $X \sim F_{m,n}$

$X \sim \chi_m^2, Y \sim \chi_n^2$ nezávislé $\implies \dfrac{\frac{X}{m}}{\frac{Y}{n}} \sim F_{m,n}$

pro kvantil $F_{m,n}(p) = \frac{1}{F_{n,m}(1-p)}$

Limitní věty:

- **konvergence** $X_n \to X$ skoro jistě .... $P[|X_n - X| \to 0] = 1$
  $X_n \to X$ podle pravděpodobnosti .... pro všechna $\epsilon > 0$ $\quad P[|X_n - X| \geq \epsilon] \to 0$
  $X_n \to X$ v distribuci .... $F_{X_n}(x) \to F_X(x)$ ve všech bodech spojitosti F_X

- **Čebyševova nerovnost** $\qquad P[|X - EX| \geq \epsilon] \leq \dfrac{DX}{\epsilon^2}$ pro $\epsilon > 0$

- **slabý zákon velkých čísel**

$X_1, X_2, \ldots$ nezávislé náhodné veličiny, $\forall k$ $EX_k = a$, $DX_k = b (< \infty) \quad \Longrightarrow$

$$\Longrightarrow \quad \frac{1}{n}\sum_{k=1}^{n} X_k \longrightarrow a \text{ podle pravděpodobnosti}$$

(silný zákon velkých čísel pojednává o konvergenci skoro jistě)

- **Bernoulliova věta**

$X_1, X_2, \ldots$ nezávislé náhodné veličiny, $\forall k$ $X_k \sim A(p)$, $0 < p < 1 \quad \Longrightarrow$

$$\Longrightarrow \quad \frac{1}{n}\sum_{k=1}^{n} X_k \longrightarrow p \text{ podle pravděpodobnosti}$$

- **centrální limitní věta, Lindeberg - Lévy**

$X_1, X_2, \ldots$ nezávislé náhodné veličiny se stejným rozdělením, $\forall k$ $EX_k = a$, $DX_k = b (< \infty) \quad \Longrightarrow$

$$\Longrightarrow \quad \frac{\frac{1}{n}\sum_{k=1}^{n} X_k - a}{\sqrt{\frac{b}{n}}} \longrightarrow Y \sim N(0,1) \text{ v distribuci}$$

- **Moivrova - Laplaceova věta**

$X_1, X_2, \ldots$ nezávislé náhodné veličiny, $\forall k$ $X_k \sim A(p)$, $0 < p < 1 \quad \Longrightarrow$

$$\Longrightarrow \quad \frac{\frac{1}{n}\sum_{k=1}^{n} X_k - p}{\sqrt{\frac{p(1-p)}{n}}} \longrightarrow Y \sim N(0,1) \text{ v distribuci}$$

Základní statistické testy:

- **jednovýběrový test** $(X_k = a + \epsilon_k)$

  $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ nezávislé

  - **odhad** pro μ $\qquad \bar{X} = \dfrac{1}{n}(X_1 + \cdots + X_n) \qquad \bar{X} \sim N(\mu, \frac{\sigma^2}{n})$
  - **odhad** pro σ^2 $\qquad S_X^2 = \dfrac{1}{n-1}\left((X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2\right)$

$$U = \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1) \qquad T = \frac{\bar{X} - \mu}{\sqrt{\frac{S_X^2}{n}}} \sim t_{n-1} \qquad \approx N(0,1) \qquad V = \frac{S_X^2}{\frac{\sigma^2}{n-1}} \sim \chi^2_{n-1}$$

$\sum_k (X_k - \bar{X})^2 = \sum_k (X_k - \mu)^2 - n \cdot (\bar{X} - \mu)^2$

- oboustranný test pro μ na hladině α

H_0: $\mu = a$ zamítneme $\iff$ $|U| \geq u(1 - \frac{\alpha}{2})$ $\quad \alpha \geq 2(1 - \Phi(|U|))$ pro známé σ
$|T| \geq t_{n-1}(1 - \frac{\alpha}{2})$ $\quad \alpha \geq 2(1 - F(|T|))$ pro neznámé σ a n malé
$|T| \geq u(1 - \frac{\alpha}{2})$ $\quad \alpha \geq 2(1 - \Phi(|T|))$ pro neznámé σ a n velké

- oboustranný konfidenční interval na hladině $1 - \alpha$

$\bar{X} \pm \frac{\sigma}{\sqrt{n}} u(1 - \frac{\alpha}{2}) \qquad \bar{X} \pm \frac{S_X}{\sqrt{n}} t_{n-1}(1 - \frac{\alpha}{2}) \qquad \bar{X} \pm \frac{S_X}{\sqrt{n}} u(1 - \frac{\alpha}{2})$

- jednostranný test pro μ na hladině α

H_0: $\mu \geq a$ zamítneme $\iff$ $U \leq u(\alpha) = -u(1 - \alpha)$ $\quad \alpha \geq \Phi(U) = 1 - \Phi(-U)$ pro známé σ
$T \leq t_{n-1}(\alpha) = -t_{n-1}(1 - \alpha)$ $\quad \alpha \geq F(T) = 1 - F(-T)$
pro neznámé σ a n malé
$T \leq u(\alpha) = -u(1 - \alpha)$ $\quad \alpha \geq \Phi(T) = 1 - \Phi(-T)$
pro neznámé σ a n velké

- jednostranný konfidenční interval na hladině $1-\alpha$

$(\bar{X}-\frac{\sigma}{\sqrt{n}}u(1-\alpha);\infty)$ $\quad$ $(\bar{X}-\frac{S_X}{\sqrt{n}}t_{n-1}(1-\alpha);\infty)$ $\quad$ $(\bar{X}-\frac{S_X}{\sqrt{n}}u(1-\alpha);\infty)$

H_0: $\mu \leq a$ zamítneme $\iff$ $U \geq u(1-\alpha)$ $\quad$ $\alpha \geq= 1-\Phi(U)$ pro známé σ

$T \geq t_{n-1}(1-\alpha)$ $\quad$ $\alpha \geq 1-F(T)$ pro neznámé σ a n malé

$T \geq u(1-\alpha)$ $\quad$ $\alpha \geq 1-\Phi(T)$ pro neznámé σ a n velké

- jednostranný konfidenční interval na hladině $1-\alpha$

$(-\infty;\bar{X}+\frac{\sigma}{\sqrt{n}}u(1-\alpha))$ $\quad$ $(-\infty;\bar{X}+\frac{S_X}{\sqrt{n}}t_{n-1}(1-\alpha))$ $\quad$ $(-\infty;\bar{X}+\frac{S_X}{\sqrt{n}}u(1-\alpha))$

- oboustranný test pro σ na hladině α

H_0: $\sigma^2 = b$ zamítneme $\iff$ $V \leq \chi^2_{n-1}(\frac{\alpha}{2})$ nebo $V \geq \chi^2_{n-1}(1-\frac{\alpha}{2})$ $\quad$ $\alpha \geq 2F(V)$ nebo $\alpha \geq 2(1-F(V))$

- jednostranný test pro σ na hladině α

H_0: $\sigma^2 \geq b$ zamítneme $\iff$ $V \leq \chi^2_{n-1}(\alpha)$ $\quad$ $\alpha \geq F(V)$

H_0: $\sigma^2 \leq b$ zamítneme $\iff$ $V \geq \chi^2_{n-1}(1-\alpha)$ $\quad$ $\alpha \geq 1-F(V)$

$X_1,\dots,X_n \sim A(p)$ nezávislé

• **odhad** pro p $\quad$ $\bar{X} = \dfrac{1}{n}(X_1+\dots+X_n)$ $\quad$ $n\bar{X} \sim Bi(n,p)$

- oboustranný test pro p na hladině α

H_0: $p = a$ zamítneme $\iff \bar{X} \leq \dfrac{k_1}{n}$ nebo $\bar{X} \geq \dfrac{k_2}{n}$

kde $\sum_{k=0}^{k_1}\binom{n}{k}a^k(1-a)^{n-k} \leq \frac{\alpha}{2}$ $\quad$ $\sum_{k=k_2}^{n}\binom{n}{k}a^k(1-a)^{n-k} \leq \frac{\alpha}{2}$

$|U| \geq u(1-\frac{\alpha}{2})$ $\quad$ $U = \dfrac{\bar{X}-a}{\sqrt{\frac{a(1-a)}{n}}}$

(párový test převádíme na jednovýběrový)

• **dvouvýběrový test**

$X_1,\dots,X_m \sim N(\mu_1,\sigma_1^2)$ nezávislé, $Y_1,\dots,Y_n \sim N(\mu_2,\sigma_2^2)$ nezávislé, oba výběry nezávislé

$\bar{X}-\bar{Y} \sim N(\mu_1-\mu_2,\left(\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}\right))$

pro $\sigma_1=\sigma_2=\sigma$

$$U = \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{m}+\frac{1}{n}}} \sim N(0,1) \qquad T = \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S\sqrt{\frac{1}{m}+\frac{1}{n}}} \sim t_{m+n-2} \qquad \approx N(0,1)$$

$$S^2 = \frac{1}{m+n-2}\left((m-1)S_X^2+(n-1)S_Y^2\right) \qquad Z = \frac{S_X^2}{S_Y^2} \sim F_{m-1,n-1}$$

- oboustranné testy na hladině α

H_0: $\mu_1-\mu_2 = a$ zamítneme $\iff$ $|T| \geq t_{m+n-1}(1-\frac{\alpha}{2})$ $\quad$ $\alpha \geq 2(1-F(|T|))$ pro $\sigma_1=\sigma_2$

H_0: $\sigma_1=\sigma_2$ zamítneme $\iff$ $Z \leq F_{m-1,n-1}(\frac{\alpha}{2})$ nebo $Z \geq F_{m-1,n-1}(1-\frac{\alpha}{2})$ $\quad$ $\alpha \geq 2F(Z)$ nebo $\alpha \geq 2(1-F(Z))$

$X_1,\dots,X_m \sim A(p_1)$, $Y_1,\dots,Y_m \sim A(p_2)$ nezávislé

- oboustranný test pro p na hladině α

H_0: $p_1-p_2 = a$ zamítneme $\iff |U| \geq u(1-\frac{\alpha}{2})$ $\quad$ $U = \dfrac{\bar{X}-\bar{Y}-a}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{m}+\frac{1}{n}\right)}}$ $\quad$ $\hat{p} = \dfrac{m\bar{X}+n\bar{Y}}{m+n}$

Analýza rozptylu:

• **jednoduché třídění**

$X_{1,1},\dots,X_{1,n_1} \sim N(\mu_1,\sigma^2)$ nezávislé $\quad$ $\mu_1 = a+\alpha_1$

$X_{2,1},\dots,X_{2,n_2} \sim N(\mu_2,\sigma^2)$ nezávislé $\quad$ $\mu_2 = a+\alpha_2$

....

$X_{I,1},\dots,X_{I,n_I} \sim N(\mu_I,\sigma^2)$ nezávislé $\quad$ $\mu_I = a+\alpha_I$

výběry nezávislé, $n_1+\dots+n_I = n$, $\alpha_1+\dots+\alpha_I = 0$

H_0: $\mu_1 = \mu_2 = \cdots = \mu_I$ neboli $\alpha_1 = \alpha_2 = \cdots = \alpha_I = 0$
(průměry)

$$\bar{X}_1 = \frac{1}{n_1}\sum_k X_{1,k}, \ldots, \bar{X}_I = \frac{1}{n_I}\sum_k X_{I,k}, \qquad \bar{X} = \frac{1}{n}\sum_{k,l} X_{k,l}$$

(součty čtverců)

$$\begin{array}{lll} S_T = \sum_{k,l}(X_{k,l} - \bar{X})^2 & f_T = n-1 & \\ S_A = \sum_k(\bar{X}_k - \bar{X})^2 \cdot n_k & f_A = I-1 & F_A = \frac{\frac{S_A}{f_A}}{\frac{S_e}{f_e}} \quad s^2 = \frac{S_e}{f_e} \\ S_e = S_T - S_A = \sum_{k,l}(X_{k,l} - \bar{X}_k)^2 & f_e = n - I & \end{array}$$

H_0: $\alpha_1 = \cdots = \alpha_I = 0$ zamítneme $\iff$ $F_A \geq F_{f_A, f_e}(1-\alpha)$

(Sheffého metoda) třídy k, l se liší $\iff$ $|\bar{X}_k - \bar{X}_l| > \sqrt{\left(\frac{1}{n_k} + \frac{1}{n_l}\right)(I-1)s^2 F_{f_A,f_e}(1-\alpha)}$

- **dvojné třídění**

$$\begin{array}{llll} X_{1,1,1},\ldots,X_{1,1,p} \sim N(\mu_{1,1},\sigma^2) \text{ nez.} & X_{1,2,1},\ldots,X_{1,2,p} \sim N(\mu_{1,2},\sigma^2) \text{ nez.} & \ldots & X_{1,J,1},\ldots,X_{1,J,p} \sim N(\mu_{1,J},\sigma^2) \text{ nez.} \\ X_{2,1,1},\ldots,X_{2,1,p} \sim N(\mu_{2,1},\sigma^2) \text{ nez.} & X_{2,2,1},\ldots,X_{2,2,p} \sim N(\mu_{2,2},\sigma^2) \text{ nez.} & \ldots & X_{2,J,1},\ldots,X_{2,J,p} \sim N(\mu_{2,J},\sigma^2) \text{ nez.} \\ \cdots & \cdots & \cdots & \cdots \\ X_{I,1,1},\ldots,X_{I,1,p} \sim N(\mu_{I,1},\sigma^2) \text{ nez.} & X_{I,2,1},\ldots,X_{I,2,p} \sim N(\mu_{I,2},\sigma^2) \text{ nez.} & \ldots & X_{I,J,1},\ldots,X_{I,J,p} \sim N(\mu_{I,J},\sigma^2) \text{ nez.} \end{array}$$

výběry nezávislé, $I \cdot J \cdot p = n$

(průměry)

$$\begin{array}{lllll} \bar{X}_{1,1} = \frac{1}{p}\sum_k X_{1,1,k} & \bar{X}_{1,2} = \frac{1}{p}\sum_k X_{1,2,k} & \ldots & \bar{X}_{1,J} = \frac{1}{p}\sum_k X_{1,J,k} & \bar{X}_1^A = \frac{1}{p\cdot J}\sum_{kl} X_{1,k,l} \\ \bar{X}_{2,1} = \frac{1}{p}\sum_k X_{2,1,k} & \bar{X}_{2,2} = \frac{1}{p}\sum_k X_{2,2,k} & \ldots & \bar{X}_{2,J} = \frac{1}{p}\sum_k X_{2,J,k} & \bar{X}_2^A = \frac{1}{p\cdot J}\sum_{kl} X_{2,k,l} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \bar{X}_{I,1} = \frac{1}{p}\sum_k X_{I,1,k} & \bar{X}_{I,2} = \frac{1}{p}\sum_k X_{I,2,k} & \ldots & \bar{X}_{I,J} = \frac{1}{p}\sum_k X_{I,J,k} & \bar{X}_I^A = \frac{1}{p\cdot J}\sum_{kl} X_{I,k,l} \\ \bar{X}_1^B = \frac{1}{p\cdot I}\sum_{k,l} X_{k,1,l} & \bar{X}_2^B = \frac{1}{p\cdot I}\sum_{k,l} X_{k,2,l} & \ldots & \bar{X}_J^B = \frac{1}{p\cdot I}\sum_{k,l} X_{k,J,l} & \bar{X} = \frac{1}{n}\sum_{klm} X_{k,l,m} \end{array}$$

- s interakcemi

$\mu_{k,l} = a + \alpha_k + \beta_l + \gamma_{k,l}$ $\quad \sum_k \alpha_k = 0 \quad \sum_l \beta_l = 0 \quad \sum_{k,l} \gamma_{k,l} = 0$

H_0: $\alpha_1 = \alpha_2 = \cdots = \alpha_I = 0$

H_0: $\beta_1 = \beta_2 = \cdots = \beta_I = 0$

H_0: $\gamma_{1,1} = \gamma_{1,2} = \cdots = \gamma_{I,J} = 0$

(součty čtverců)

$$\begin{array}{lll} S_T = \sum_{klm}(X_{k,l,m} - \bar{X})^2 & f_T = n-1 & \\ S_A = \sum_k(\bar{X}_k^A - \bar{X})^2 \cdot J \cdot p & f_A = I-1 & \\ S_B = \sum_k(\bar{X}_k^B - \bar{X})^2 \cdot I \cdot p & f_A = J-1 & F_A = \frac{\frac{S_A}{f_A}}{\frac{S_e}{f_e}} \quad F_B = \frac{\frac{S_B}{f_B}}{\frac{S_e}{f_e}} \quad F_{AB} = \frac{\frac{S_{AB}}{f_{AB}}}{\frac{S_e}{f_e}} \quad s^2 = \frac{S_e}{f_e} \\ S_e = \sum_{klm}(X_{k,l,m} - \bar{X}_{k,l})^2 & f_e = n - I \cdot J & \\ S_{AB} = S_T - S_A - S_B - S_e & f_{AB} = (I-1)(J-1) & \end{array}$$

H_0: $\alpha_1 = \cdots = \alpha_I = 0$ zamítneme $\iff$ $F_A \geq F_{f_A,f_e}(1-\alpha)$

H_0: $\beta_1 = \cdots = \beta_I = 0$ zamítneme $\iff$ $F_B \geq F_{f_B,f_e}(1-\alpha)$

H_0: $\gamma_{1,1} = \cdots = \gamma_{I,J} = 0$ zamítneme $\iff$ $F_{AB} \geq F_{f_{AB},f_e}(1-\alpha)$

- bez interakcí

$\mu_{k,l} = a + \alpha_k + \beta_l$ $\quad \sum_k \alpha_k = 0 \quad \sum_l \beta_l = 0$

H_0: $\alpha_1 = \alpha_2 = \cdots = \alpha_I = 0$

H_0: $\beta_1 = \beta_2 = \cdots = \beta_I = 0$

$$\begin{array}{lll} S_T = \sum_{klm}(X_{k,l,m} - \bar{X})^2 & f_T = n-1 & \\ S_A = \sum_k(\bar{X}_k^A - \bar{X})^2 \cdot J \cdot p & f_A = I-1 & F_A = \frac{\frac{S_A}{f_A}}{\frac{S_e}{f_e}} \quad F_B = \frac{\frac{S_B}{f_B}}{\frac{S_e}{f_e}} \quad s^2 = \frac{S_e}{f_e} \\ S_B = \sum_k(\bar{X}_k^B - \bar{X})^2 \cdot I \cdot p & f_A = J-1 & \\ S_e = S_T - S_A - S_B & f_e = n - I - J + 1 & \end{array}$$

H_0: $\alpha_1 = \cdots = \alpha_I = 0$ zamítneme $\iff$ $F_A \geq F_{f_A,f_e}(1-\alpha)$

H_0: $\beta_1 = \cdots = \beta_I = 0$ zamítneme $\iff$ $F_B \geq F_{f_B,f_e}(1-\alpha)$

(Sheffého metoda) třídy k, l se liší $\iff$ $|\bar{X}_k^A - \bar{X}_l^A| > \sqrt{\frac{2(I-1)}{JP}s^2 F_{f_A,f_e}(1-\alpha)}$

$$\text{třídy } k,\ l \text{ se liší} \quad \iff \quad |\bar{X}_k^B - \bar{X}_l^B| > \sqrt{\frac{2(J-1)}{IP} s^2 F_{f_B,f_e}(1-\alpha)}$$

Korelace a regrese:

• **korelační koeficient**

$(X_1,Y_1),\dots,(X_n,Y_n) \sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})$ nezávislé, $\boldsymbol{\mu} = (\mu_1,\mu_2)$, $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$, $\rho = \dfrac{\sigma_{12}}{\sqrt{\sigma_1^2\sigma_2^2}}$

• **odhad** pro σ_{12} $\quad S_{XY} = \frac{1}{n-1}\left((X_1-\bar{X})(Y_1-\bar{Y}) + \dots + (X_n-\bar{X})(Y_n-\bar{Y})\right)$

• **odhad** pro ρ $\quad r = \dfrac{S_{XY}}{S_X \cdot S_Y}$, $T = \dfrac{r}{\sqrt{\frac{1-r^2}{n-2}}} \sim t_{n-2}$

H_0: $\rho = 0$ zamítneme $\iff$ $|T| \geq t_{n-2}(1-\frac{\alpha}{2})$ $\qquad \alpha \geq 2(1-F(|T|))$

$\boldsymbol{R_{XY}}$ korelační matice náhodných vektorů $\boldsymbol{X} = (X_1,\dots,X_I)$ a $\boldsymbol{Y} = (Y_1,\dots,Y_J)$

• **mnohonásobná korelace** X a $\boldsymbol{Y} = (Y_1,\dots,Y_k)$

(maximální korelace X a $c_1 \cdot X_1 + \dots + c_k \cdot X_k$, rovnost nastává, jsou-li $c_1,\dots,c_k$ regresní koeficienty)

$$r^2_{X,\boldsymbol{Y}} = \boldsymbol{R}_{X\boldsymbol{Y}} \cdot \boldsymbol{R}^{-1}_{\boldsymbol{YY}} \cdot \boldsymbol{R}_{\boldsymbol{Y},X}, \quad r^2_{X,(Y_1,Y_2)} = \frac{r^2_{XY_1} + r^2_{XY_2} - 2r_{XY_1}r_{XY_2}r_{Y_1Y_2}}{1 - r^2_{Y_1Y_2}}$$

$$\frac{n-k-1}{k} \cdot \frac{r^2_{X,\boldsymbol{Y}}}{1 - r^2_{X,\boldsymbol{Y}}} \sim F_{k,n-k-1}$$

• **parciální korelace** X a Y bez vlivu $\boldsymbol{Z} = (Z_1,\dots,Z_k)$

$$r_{X,Y.\boldsymbol{Z}} = \frac{r_{X,Y} - \boldsymbol{R}_{X\boldsymbol{Z}} \cdot \boldsymbol{R}^{-1}_{\boldsymbol{ZZ}} \cdot \boldsymbol{R}_{\boldsymbol{Z},Y}}{\sqrt{(1 - \boldsymbol{R}_{X\boldsymbol{Z}} \cdot \boldsymbol{R}^{-1}_{\boldsymbol{ZZ}} \cdot \boldsymbol{R}_{\boldsymbol{Z},X})(1 - \boldsymbol{R}_{Y\boldsymbol{Z}} \cdot \boldsymbol{R}^{-1}_{\boldsymbol{ZZ}} \cdot \boldsymbol{R}_{\boldsymbol{Z},Y})}}, \qquad r_{X,Y.Z} = \frac{r_{X,Y} - r_{XZ} \cdot r_{YZ}}{\sqrt{(1-r^2_{XZ})(1-r^2_{YZ})}}$$

$$\frac{r_{X,Y.\boldsymbol{Z}}}{\sqrt{1 - r^2_{X,Y.\boldsymbol{Z}}}}\sqrt{n-k-2} \sim t_{n-k-2}$$

• **lineární regrese** $(Y_k = a + b \cdot x_k + \epsilon_k)$

$Y_1 \sim N(a + b\cdot x_1, \sigma^2),\dots,Y_n \sim N(a + b \cdot x_n, \sigma^2)$ nezávislé

$$\bar{x} = \frac{1}{n}(x_1 + \dots + x_n) \qquad (n-1)s_x^2 = (x_1-\bar{x})^2 + \dots + (x_2-\bar{x})^2$$

$$(n-1)S^2_{xY} = (x_1-\bar{x})(Y_1-\bar{Y}) + \dots + (x_n-\bar{x})(Y_1-\bar{Y}) \qquad r_{xY} = \frac{S_{xY}}{\sqrt{s_x^2 S_Y^2}}$$

odhad pro b $\quad \hat{b} = \dfrac{n\sum x_k Y_k - \sum x_k \sum Y_k}{n\sum x_k^2 - (\sum x_k)^2} \qquad \hat{b} = \dfrac{S_{xY}}{s_x^2} = r_{xY} \cdot \dfrac{S_Y}{s_x}$

$$E(\hat{b}) = b \quad D(\hat{b}) = \frac{\sigma^2}{(n-1)s_x^2}$$

odhad pro a $\quad \hat{a} = \dfrac{\sum x_k^2 \sum Y_k - \sum x_k \sum x_k Y_k}{n\sum x_k^2 - (\sum x_k)^2} \qquad \hat{a} = \bar{Y} - \hat{b}\bar{x}$

$$E(\hat{a}) = a \quad D(\hat{a}) = \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{(n-1)s_x^2}\right)$$

$$\frac{\hat{b}-b}{\sqrt{\frac{s^2}{(n-1)s_x^2}}} \sim t_{n-2} \qquad \frac{\hat{a}-a}{\sqrt{s^2\left(\frac{1}{n} + \frac{\bar{x}^2}{(n-1)s_x^2}\right)}} \sim t_{n-2} \qquad s^2 = \frac{S_e}{n-2} \qquad s^2 = (1-R^2)\cdot S_Y^2 \cdot \frac{n-1}{n-2}$$

$$S_T = \sum_k (Y_k - \bar{Y})^2 = (n-1)S_Y^2 \qquad S_{REG} = \sum_k (\hat{a}x_k - \hat{b} - \bar{Y})^2 \qquad S_e = \sum_k (\hat{a}x_k - \hat{b} - Y_k)^2$$

• **koeficient determinace** ...... $R^2 = \dfrac{S_{REG}}{S_T} = 1 - \dfrac{S_e}{S_T} \qquad R^2 = r^2_{xY} \qquad R^2 = \hat{b}^2 \cdot \dfrac{S_X^2}{S_Y^2}$

pro x_0

$\hat{a} + \hat{b} \cdot x_0 \qquad E(\hat{a} + \hat{b} \cdot x_0) = a + b \cdot x_0 \qquad D(\hat{a} + \hat{b} \cdot x_0) = \sigma^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{(n-1)s_x^2} \right)$

$$\frac{\hat{a} + \hat{b} \cdot x_0 - a - b \cdot x_0}{\sqrt{\sigma^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{(n-1)s_x^2} \right)}} \sim t_{n-2}$$

interval spolehlivosti (konfidenční interval) $EY(x_0)$ pro x_0

$$EY(x_0) = \hat{a} + \hat{b} \cdot x_0 \pm s\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{(n-1)s_x^2}} \cdot t_{n-2}(1 - \frac{\alpha}{2})$$

pás spolehlivosti (Working-Hotelling) $EY(x)$ kolem regresní přímky (Sheffého metoda)

$$EY(x) = \hat{a} + \hat{b} \cdot x \pm s\sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{(n-1)s_x^2}} \cdot \sqrt{2F_{2,n-1}(1 - \alpha)} \qquad \forall x$$

interval predikce $Y(x_0)$ pro x_0

$$ax_0 + b = \hat{a}x_0 + \hat{b} \pm s\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{(n-1)s_x^2}} \cdot t_{n-2}(1 - \frac{\alpha}{2})$$

pás predikce $EY(x)$ kolem regresní přímky

$$ax + b = \hat{a}x + \hat{b} \pm s\sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{(n-1)s_x^2}} \cdot \sqrt{2F_{2,n-1}(1 - \alpha)} \qquad \forall x$$

- **případ a=0** ($Y_k = b \cdot x_k + \epsilon_k$)

odhad pro b $\qquad \hat{b} = \frac{\sum x_k Y_k}{\sum x_k^2}$

$$\frac{\hat{b} - b}{\sqrt{\frac{s^2}{\sum x_k^2}}} \sim t_{n-1} \qquad s^2 = \frac{1}{n-1}\left(\sum Y_k^2 - \hat{b} \sum x_k^2\right)$$

$$\sum x_k^2 = n\bar{x}^2 + (n-1)S_x^2 \qquad \sum Y_k^2 = n\bar{Y}^2 + (n-1)S_Y^2 \qquad \sum x_k Y_k = n\bar{x}\bar{Y} + (n-1)S_{xY}$$

Další testy:

• kontingenční tabulky

H_0: $X,\ Y$ nezávislé

$X \backslash Y$	1	$\dots$	s	
1	n_{11}	$\dots$	n_{1s}	$n_{1\cdot}$
$\vdots$		$\dots$		$\vdots$
r	n_{r1}	$\dots$	n_{rs}	$n_{r\cdot}$
	$n_{\cdot 1}$	$\dots$	$n_{\cdot s}$	n

$(n_{kl} \geq 5)$

$$\chi^2 = \sum_{kl} \frac{\left(n_{kl} - \frac{n_{k\cdot} \cdot n_{\cdot l}}{n}\right)^2}{\frac{n_{k\cdot} \cdot n_{\cdot l}}{n}} = n \sum_{kl} \frac{n_{kl}^2}{n_{k\cdot} \cdot n_{\cdot l}} - n \approx \chi^2_{(r-1)(s-1)}$$

H_0 zamítneme $\Longleftrightarrow \chi^2 \geq \chi^2_{(r-1)(s-1)}(1-\alpha)$

• Fisherův faktoriálový test

H_0: $X,\ Y$ nezávislé

$X \backslash Y$	0	1
0	n_{11}	n_{12}
1	n_{21}	n_{22}

(i pro malé četnosti)

logaritmická interakce $L = \ln \frac{n_{11} n_{22}}{n_{12} n_{21}}$

Fisherova pravděpodobnost $P = \frac{n_{1\cdot}! n_{2\cdot}! n_{\cdot 1}! n_{\cdot 2}!}{n! n_{11}! n_{22}! n_{12}! n_{21}!}$

postup:

▷ vypíšeme všechny tabulky se stejnými marginálními četnostmi

▷ vypočítáme k nim logaritmickou interakci a Fisherovu pravděpodobnost

▷ p je součet všech P, pro které je $|L| \geq L_0$ (L_0 - logaritmická interakce původní tabulky)

H_0 zamítneme $\Longleftrightarrow p \leq \alpha$

• McNemarův test

H_0: $X,\ Y$ mají stejné rozdělení

osoby	1	2	3	4	...	n	
1. otázka	0	1	1	0	...		
2. otázka	1	0	1	0	...		

$X \backslash Y$	0	1
0	n_{11}	n_{12}
1	n_{21}	n_{22}

n_{12} = počet odpovědí 01 n_{21} = počet odpovědí 10 (založeno na binomickém rozdělení) $\sum n_{kl} = n$

H_0 zamítneme $\iff \left(\frac{1}{2}\right)^m \sum_{k=0}^{s} \binom{m}{k} < \frac{\alpha}{2}$ kde $s = \min(n_{12}, n_{21})$, $m = n_{12} + n_{21}$

$\chi^2 = \frac{(n_{12} - n_{21})^2}{n_{12} + n_{21}} \approx \chi^2_1$ pro n velké

H_0: zamítneme $\iff \chi^2 \geq \chi^2_1(1-\alpha)$

• **Cochranův Q test** (zobecnění McNemarova testu)

H_0: $X_1, \ldots, X_I$ mají stejné rozdělení

osoby	1	2	...	n	
1. otázka	0	1	...		m_1
2. otázka	1	0	...		m_2
...			...		
r-tá otázka	0	1			m_r
	s_1	s_2	...	s_n	m

$$Q = \frac{r(r-1)\sum_{k=1}^{r} m_k^2 - (r-1)m^2}{rm - \sum_{k=1}^{n} s_k^2} \approx \chi^2_{r-1}$$

H_0 zamítneme $\iff Q \geq \chi^2_{r-1}(1-\alpha)$

• **Stuartův test** (zobecnění McNemarova testu)

H_0: X, Y mají stejné rozdělení

$X \backslash Y$	1	...	r	
1	n_{11}	...	n_{1r}	$n_{1\cdot}$
$\vdots$		...		$\vdots$
r	n_{r1}	...	n_{rr}	$n_{r\cdot}$
	$n_{\cdot 1}$	...	$n_{\cdot r}$	n

$a_{kk} = n_{k\cdot} + n_{\cdot k} - 2n_{kk} \quad a_{kl} = -(n_{kl} + n_{lk})$ pro $k \neq l$
$b_k = n_{k\cdot} - n_{\cdot k}$
$k, l = 1, \ldots, r-1$
$Q = \boldsymbol{b}^T \boldsymbol{A}^{-1} \boldsymbol{b} \approx \chi^2_{r-1}$

H_0 zamítneme $\iff Q \geq \chi^2_{r-1}(1-\alpha)$

• **Kolmogorův - Smirnovův test** (založen na porovnávání výběrové a hypotetické ditribuční funkce)
- pro známé hodnoty parametrů
$X_1, \ldots, X_n$ nezávislé, stejné rozdělení, uspořádané $X_{(1)}, \ldots, X_{(n)}$

H_0: X_k má rozdělení F

$$F_n(x) = \begin{cases} 0 & \text{pro } x \leq x_{(1)} \\ \frac{k}{n} & \text{pro } x_{(k)} < x \leq x_{(k+1)} \\ 1 & \text{pro } x_{(n)} < x \end{cases}$$

$D^+ = \max\left\{\frac{k}{n} - F(x_{(k)})\right\} \qquad D^- = \max\left\{F(x_{(k)}) - \frac{k-1}{n}\right\} \qquad D = \max(D^+, D^-) = \sup\{F_n(x) - F(x)\}$

H_0: X_k má rozdělení F zamítneme $\iff D \geq D_n^+(1 - \frac{\alpha}{2})$ (v tabulkách na www) $\approx \sqrt{-\frac{1}{2n} \ln \frac{\alpha}{2}}$

• **Lillieforsův test** (založen na podobném principu)
• **Shapiro - Wilkův test**
- pro normální rozdělení, neznámé parametry odhadneme $\bar{X}$, S_X^2

• **testy dobré shody**
$n_1, \ldots, n_r$ četnosti, $p_1, \ldots, p_r$ pravděpodobnosti

$$\chi^2 = \sum_{k=1}^{r} \frac{(n_k - np_k)^2}{np_k} \approx \chi^2_{r-1}$$

H_0: četnosti odpovídají pravděpodobnostem zamítneme $\iff \chi^2 \geq \chi^2_{r-1}(1-\alpha)$

- ověření rozdělení, kde neznáme m parametrů

$X_1, \dots, X_n$ nezávislé, stejné rozdělení

H_0: X_k má dané rozdělení

postup:

▷ rozdělíme na r tříd
▷ zjistíme četnosti ve třídách
▷ (případně) neznámé parametry odhadneme, např. $\bar{X}$, $S_X^2, \dots$
▷ pro každou třídu vypočítáme očekávanou četnost np_k
▷ sdružíme třídy, pro něž je $np_k < 5$ (celkový počet tříd označíme opět r)

$$\chi^2 = \sum_{k=1}^{r} \frac{(n_k - n_{p_k})^2}{np_k} \approx \chi^2_{r-1-m}$$

H_0 zamítneme $\iff \chi^2 \geq \chi^2_{r-1-m}(1-\alpha)$

• **testování stejného rozptylu**

$X_{1,1}, \dots, X_{1,n_1} \sim N(\mu_1, \sigma^2)$ nezávislé
$X_{2,1}, \dots, X_{2,n_2} \sim N(\mu_2, \sigma^2)$ nezávislé
....
$X_{I,1}, \dots, X_{I,n_I} \sim N(\mu_I, \sigma^2)$ nezávislé

výběry nezávislé, $n_1 + \dots + n_I = n$

H_0: $\sigma_1^2 = \dots = \sigma_I^2$

• **Bartlettův test**

H_0: zamítneme $\iff Q \geq \chi^2_{r-1-m}(1-\alpha)$

$$Q = \frac{(n-I)\ln s^2 - \sum\limits_{k=1}^{I}(n_k-1)\ln s_k^2}{1 + \frac{1}{s(I-1)}\left(\sum\limits_{k=1}^{I}\frac{1}{n_k-1} - \frac{1}{n-I}\right)} \approx \chi^2_{I-1} \quad \text{kde } s_k^2 = \frac{1}{n_k-1}\sum_{l=1}^{n_k} x_{kl}^2 - n_k\bar{x}_k^2 \quad s^2 = \frac{1}{n-I}\sum_{k=1}^{I}(n_k-1)s_k^2$$

(Leveneův test)

Neparametrické metody:

(pro spojité rozdělení)

• **znaménkový test**

$X_1, \dots, X_n$ nezávislé, stejné rozdělení

$H_0:\ \tilde{x} = a$

postup:

▷ utvoříme $X_1 - a, \dots, X_n - a$
▷ Y počet kladných rozdílů, $U = \frac{2Y-n}{\sqrt{n}} \approx N(0,1)$

H_0: zamítneme $\iff Y \leq k_1$ nebo $Y \geq k_2$
$\iff |U| \geq u(1-\frac{\alpha}{2})$

• **Wilcoxonův jednovýběrový test**

$X_1 - a, \dots, X_n - a$ nezávislé, stejné rozdělení

H_0 : rozdělení je symetrické kolem a ($F(a-x) + F(a+x) = 1$)

postup:

▷ hodnotám $X_1 - a, \dots, X_n - a$ přiřadíme pořadí podle absolutních hodnot
▷ $S^+ =$ součet pořadí kladných $X_k - a$, $S^- =$ součet pořadí záporných $X_k - a$, $U = \dfrac{S^+ - \frac{1}{4}n(n+1)}{\sqrt{\frac{1}{24}n(n+1)(2n+1)}} \approx N(0,1)$

H_0: zamítneme $\iff \min(S^+, S^-) \geq k$
$\iff |U| \geq u(1-\frac{\alpha}{2})$

• **Wilcoxonův dvouvýběrový test**

$X_1, \dots, X_m$ nezávislé, stejné rozdělení, $Y_1, \dots, Y_n$ nezávislé, stejné rozdělení, výběry nezávislé

H_0 : X_k a Y_k má stejné rozdělení

postup:

$\triangleright$ hodnotám $X_1, \ldots, X_m, Y_1, \ldots, Y_n$ přiřadíme pořadí
$\triangleright$ T_1 = součet pořadí X_k, T_2 = součet pořadí Y_k,
$\triangleright$ $U_1 = mn + \frac{1}{2}m(m+1) - T_1 \quad U_2 = mn + \frac{1}{2}n(n+1) - T_2 \; U = \dfrac{U_1 - \frac{1}{2}mn}{\sqrt{\frac{1}{12}mn(m+n+1)}} \approx N(0,1)$

H_0 zamítneme $\iff \min(U_1, U_2) \geq k$
$\iff |U| \geq u(1 - \frac{\alpha}{2})$

• Spearmannův korelační koeficient
$(X_1, Y_1), \ldots, (X_n, Y_n)$ nezávislé, stejné rozdělení
$H_0 : X_k \; Y_k$ nezávislé
postup:
$\triangleright$ hodnotám $X_1, \ldots, X_n$ přiřadíme pořadí $Q_1, \ldots, Q_n$
$\triangleright$ hodnotám $Y_1, \ldots, Y_n$ přiřadíme pořadí $R_1, \ldots, R_n$
$\triangleright$ $r_S = 1 - \dfrac{6}{n(n^2-1)} \sum_k (R_k - Q_k)^2$

H_0: zamítneme $\iff |r_S| \geq k$
$\iff |r_S| \geq \dfrac{u(1-\frac{\alpha}{2})}{\sqrt{n-1}}$

• Kendallův korelační koeficient
$(X_1, Y_1), \ldots, (X_n, Y_n)$ nezávislé, stejné rozdělení
$H_0 : X_k \; Y_k$ nezávislé
postup:
$\triangleright$ uspořádáme (X_k, Y_k) vzestupně podle X_k, tedy tak aby $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$
$\triangleright$ označme s_1 počet $Y_{(2)}, \ldots, Y_{(n)}$ větších než $Y_{(1)}$
s_2 počet $Y_{(3)}, \ldots, Y_{(n)}$ větších než $Y_{(2)}$
$\ldots$
s_{n-1} počet $Y_{(n)}$ větších než $Y_{(n-1)}$
$\triangleright$ $\tau = \dfrac{4}{n(n-1)} \sum_k s_k - 1 = \dfrac{1}{n(n-1)} \sum_{k \neq l} \operatorname{sgn}(Q_k - Q_l) \cdot \operatorname{sgn}(R_k - R_l)$ (značení z předchozího testu)

H_0: zamítneme $\iff |\tau| \geq k$
$\iff |\tau| \geq \dfrac{3\sqrt{n}}{2} u(1 - \dfrac{\alpha}{2})$

• Kruskalův - Walisův test (jednoduché třídění)

$X_{1,1}, \ldots, X_{1,n_1}$ nezávislé, stejné rozdělení
$X_{2,1}, \ldots, X_{2,n_2}$ nezávislé, stejné rozdělení
$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$
$X_{I,1}, \ldots, X_{I,n_I}$ nezávislé, stejné rozdělení

výběry nezávislé, $n_1 + \cdots + n_I = n$

H_0: rozdělení $X_{1k}, X_{2k}, \ldots, X_{Ik}$ jsou stejná
postup:
$\triangleright$ hodnotám $X_{11}, \ldots, X_{1n_1}, \ldots, X_{I1}, \ldots, X_{In_I}$ přiřadíme pořadí
$\triangleright$ T_1 = součet pořadí X_{1k}, T_2 = součet pořadí X_{2k},...,T_I = součet pořadí X_{Ik}
$\triangleright$ $Q = \dfrac{12}{n(n+1)} \sum_k \dfrac{T_k^2}{n_k} - 3(n+1) \approx \chi^2_{I-1}$

H_0 zamítneme $\iff |Q| \geq k$ (rozsáhlé tabulky na www)
$|Q| \geq \chi^2_{I-1}(1-\alpha)$

• Friedmannův test (dvojné třídění pro $n_k = 1$)

$X_{1,1}, \ldots, X_{1,r}$
$X_{2,1}, \ldots, X_{2,r}$
$\ldots\ldots\ldots\ldots$
$X_{n,1}, \ldots, X_{n,r}$

nezávislé

H_0: $X_{k,1}, X_{k,2}, \ldots, X_{k,r}$ mají stejné rozdělení
postup:
$\triangleright$ v každém bloku (sloupci) přiřadíme pořadí R_{kl}

$\triangleright$ $Q = \frac{12}{nr(r+1)} \sum_{l=1}^{r} \left(\sum_{k=1}^{n} R_{kl} \right)^2 - 3n(r+1) \approx \chi^2_{r-1}$

H_0 zamítneme $\iff$ $|Q| \geq k$ (tabulky na www)

$|Q| \geq \chi^2_{r-1}(1-\alpha)$