AKADEMIE VĚD ČESKÉ REPUBLIKY
Psychologický ústav
Veveří 97, 602 00 Brno
telefon, fax 05/ 74 46 67

Ψ

# Výzkumné zprávy

## Receptář jednoduchých metod statistické indukce

Lída Osecká & Pavel Osecký



1996
č. 3

ISSN: 1211-2631

# Obsah

# A Třídění indukčních receptů

Následující text obsahuje základní hlediska, podle nichž se můžeme orientovat v receptech statistické indukce. Neočekávejme však, že tak dospějeme k jejich vyčerpávající a jednoznačné systematice.

## A.1 Typy porovnávání dat

Recepty jsou především roztříděny podle situace, ve které porovnáváme data získaná za různých podmínek.

I = **jednovýběrové vyšetřování:** Vyšetřujeme jediný datový soubor, který neumožňuje porovnávat data popř. jejich závislost za různých podmínek. Datový soubor má tvar

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \text{popř.} \quad \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_n & y_n \end{bmatrix}$$

II = **dvouvýběrové porovnávání:** Porovnáváme data popř. jejich závislost ve dvou nezávisle získaných datových souborech, které mají tvar

$$\begin{bmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n_1 1} \end{bmatrix} \begin{bmatrix} x_{12} \\ x_{22} \\ \vdots \\ x_{n_2 2} \end{bmatrix} \quad \text{popř.} \quad \begin{bmatrix} x_{11} & y_{11} \\ x_{21} & y_{21} \\ \vdots & \vdots \\ x_{n_1 1} & y_{n_1 1} \end{bmatrix} \begin{bmatrix} x_{12} & y_{12} \\ x_{22} & y_{22} \\ \vdots & \vdots \\ x_{n_2 2} & y_{n_2 2} \end{bmatrix}$$

**III** = **vícevýběrové porovnávání:** Porovnáváme data popř. jejich závislost v $g$ nezávisle získaných datových souborech, které mají tvar

$$\begin{bmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n_1 1} \end{bmatrix} \begin{bmatrix} x_{12} \\ x_{22} \\ \\ \vdots \\ \\ x_{n_2 2} \end{bmatrix} \cdots \begin{bmatrix} x_{1g} \\ x_{2g} \\ \vdots \\ x_{n_g g} \end{bmatrix}$$

popř.

$$\begin{bmatrix} x_{11} & y_{11} \\ x_{21} & y_{21} \\ \vdots & \vdots \\ x_{n_1 1} & y_{n_1 1} \end{bmatrix} \begin{bmatrix} x_{12} & y_{12} \\ x_{22} & y_{22} \\ \\ \vdots & \vdots \\ \\ x_{n_2 2} & y_{n_2 2} \end{bmatrix} \cdots \begin{bmatrix} x_{1g} & y_{1g} \\ x_{2g} & y_{2g} \\ \vdots & \vdots \\ x_{n_g g} & y_{n_g g} \end{bmatrix}$$

**IV** = **párové porovnávání:** Porovnáváme data v prvním a druhém sloupci jediného párového datového souboru, který vznikl buď zkoumáním nějakých přirozených párů nebo dvojím vyšetřováním týchž objektů. Párový datový soubor má tvar

$$\begin{bmatrix} x_{11} & x_{12} \\ x_{12} & x_{22} \\ \vdots & \\ x_{n1} & x_{n2} \end{bmatrix}$$

## A.2 Stupně kvantifikace

Recepty jsou dále tříděny podle úrovně kvantifikace, které dosahuje zkoumaná proměná. Mohou být použity i na vyšších úrovních měření, ne však na nižších.

a = **alternativní proměnná:** Nabývá jen hodnot 0 a 1, znamenajích nepřítomnost nebo výskyt nějakého jevu. Podle okolností může být ztotožněna s kterýmkoliv dále uvedeným typem proměnné.

n = **nominální proměnná:** Věcný význam má jen různost a totožnost dvou číselných hodnot, které jsou pouhým kódem kvalitativních pojmenování.

o = **ordinální proměnná:** Navíc má věcný význam i větší nebo menší číselná hodnota, ne však velikost jejich rozdílu.

i = **intervalová proměnná:** I větší nebo menší délka intervalu mezi dvěma číselnými hodnotami má věcný význam, ne však poměr obou těchto hodnot.

p = **poměrová proměnná:** Konečně lze připsat věcný význam i poměru dvou číselných hodnot. Jde o kvantitativní měření v plném slova smyslu.

## A.3 Seznam charakteristik

Detailní třídění receptů odpovídá teoretické statistické charakteristice, na niž se induktivní úloha zaměřuje.

| | | |
|---|---|---|
| $\theta$ | = | pravděpodobnost výskytu |
| $\theta_1 - \theta_2$ | = | rozdíl dvou pravděpodobností výskytu |
| $\varrho''$ | = | koeficient alternativní korelace |
| $\pi(x)$ | = | pravděpodobnostní funkce jedné proměnné |
| $\pi(x, y)$ | = | simultánní pravděpodobnostní funkce dvou proměnných |
| $\pi_1(x), \pi_2(y)$ | = | marginální pravděpodobnostní funkce obou složek |
| $\pi_1(x), \ldots, \pi_g(x)$ | = | porovnávané pravděpodobnostní funkce |
| $\varrho'$ | = | Spearmanův koeficient pořadové korelace |
| $\mu$ | = | střední hodnota |
| $\mu_1 - \mu_2$ | = | rozdíl dvou středních hodnot |
| $\sigma$ | = | směrodatná odchylka |
| $\sigma_1/\sigma_2$ | = | podíl dvou směrodatných odchylek |
| $\varrho$ | = | koeficient lineární korelace |
| $\varrho_1 - \varrho_2$ | = | rozdíl dvou koeficientů lineární korelace |

## A.4 Typy procedur

Pro různé typy otázek jsou ve statistice vypracovány odpovídající procedury, které představují poslední hledisko třídění statistických receptů.

S = **stanovení intervalu spolehlivosti:** Protože informace o přesné hodnotě teoretické charakteristiky $\vartheta$ je zpravidla nedostupná, bývá aspoň z výběrových dat sestrojován takový interval o dolní popř. horní hranici $d$ popř. $h$, který by s velkou pravděpodobností $1-\alpha$ tuto neznámou charakteristiku pokrýval; číslo $\alpha$ představuje přijatelné riziko nepravdivého stanovení tohoto intervalu a volí se nejčastěji rovno 0,05 nebo 0,01. Podle potřeby vystupuje interval spolehlivosti v levostranné, oboustranné popř. pravostranné formě, charakterizované postupně jednou ze tří pravděpodobností

$$\begin{aligned} P(d \le \vartheta \quad\quad) &\ge 1-\alpha \\ P(d \le \vartheta \le h) &\ge 1-\alpha \\ P(\quad \vartheta \le h) &\ge 1-\alpha \end{aligned}$$

— Ve standardních statistických programech se zpravidla uvádějí oboustranné intervaly spolehlivosti.

N = **určení potřebného rozsahu výběru:** Je žádoucí, aby oboustranný interval spolehlivosti vycházel co možno úzký. Jestliže uživatel zvolí jeho maximální ještě přijatelnou šířku $\delta$ při stanoveném riziku $\alpha$, poskytne mu v některých případech statistická procedura vzorec pro minimální rozsah výběru, kterého je zapotřebí ke s plnění jeho požadavků. Někdy se přijatelná šířka intervalu spolehlivosti vyjadřuje podílem $\kappa$ jeho horní a dolní hranice.

P = **stanovení predikčního intervalu:** Tak jako má interval spolehlivosti pokrýt s velkou pravděpodobností neznámou teoretickou charakteristiku $\vartheta$, pokrývá predikční interval o dolní popř. horní hranici $d_*$ popř. $h_*$ nějaké dosud nepozorované statistiky $m^*$. Vyskytuje se rovněž ve třech formách charakterizovaných postupně pravděpodobnostmi

$$\begin{aligned} P(d_* \le m_* \quad\quad) &\ge 1-\alpha \\ P(d_* \le m_* \le h_*) &\ge 1-\alpha \\ P(\quad m_* \le h_*) &\ge 1-\alpha \end{aligned}$$

kde číslo $\alpha$ vyjadřuje riziko nepravdivého stanovení predikčního intervalu a volí se nejčastěji jako 0,05 nebo 0,01.

**H = testování statistické hypotézy:** Nějaký vnější, na analyzovaných datech nezávislý důvod nás nutí uvažovat o domněnce, že neznámá hodnota teoretické charakteristiky $\vartheta$ se rovná danému číslu $c$. Toto tvrzení

$$\vartheta = c$$

se nazývá testovanou hypotézou a konkuruje s hypotézou alternativní, která opět nesmí záviset na analyzovaných datech a bývá uváděna v levostranné, oboustranné popř. pravostranné formě

$$\vartheta < c \quad \vdots \quad \vartheta \neq c \quad \vdots \quad c < \vartheta$$

Do role testované hypotézy dosazujeme často skeptické tvrzení, které si ve skutečnosti přejeme vyvrátit, nebo naopak vztah, který podmiňuje zamýšlené použití nějakého matematického modelu. Alternativní hypotézy se nejčastěji užívá v její oboustranné formě, kdežto levostrannou či pravostrannou formu zvolíme v případě, že opačné jednostranné tvrzení je předem vyloučeno nebo nás nezajímá. V této situaci matematická statistika odvozuje vzorce pro kritický obor $W$, což je nějaká číselná množina, a pro testovou statistiku $t$, což je náhodná veličina. Platí pro ně implikace tvrdící, že v případě pravdivosti testované hypotézy nepřesáhne pravděpodobnost

$$P(t \in W) \leq \alpha$$

číslo $\alpha$, které se nazývá rizikem neoprávněného zamítnutí testované hypotézy a volí se nejčastěji rovno 0,05 nebo 0,01. Jestliže se tedy testová statistika $t$ navzdory této nepatrné pravděpodobnosti přece jenom realizuje v kritickém oboru $W$, zamítáme testovanou hypotézu pro její rozpor s daty. Realizuje-li se testová statistika mimo kritický obor, k žádnému rozporu nedochází a testovanou hypotézu přijímáme jako nadále podržitelnou. Kritické obory pro shora uvedené formy alternativních hypotéz mají formu intervalů nebo jejich sjednocení a jsou vymezeny po řadě kritickými podmínkami

$$t < t_\alpha \quad \vdots \quad t < t_{\alpha/2} \text{ nebo } t_{1-\alpha/2} < t \quad \vdots \quad t_{1-\alpha/2} < t$$

Testování statistické hypotézy se nemusí týkat jen číselných, ale i funkcionálních teoretických charakteristik, jako jsou např. pravděpodobnostní funkce či hustoty pravděpodobnosti: tím se mohou obměnit i uváděné formy testovaných a alternativních hypotéz. — Ve standardních statistických programech se netisknou kritické obory, ale ekvivalentně tzv. dosažené hladiny průkaznosti $p$ vyjadřující riziko, jež by bylo třeba volit, aby testová statistika padla právě na hranici kritického oboru. Testovaná hypotéza se pak zamítá v případě

$$p < \alpha$$

$n \to \infty \Rightarrow h_j \to 0 \Rightarrow f(x) \to \varphi(x)$ hustota pravděpodobnosti

$n \to \infty \Rightarrow r_{xy} \to \rho_{xy}$

$n \to \infty \Rightarrow h_r \to$ entropie

**T** = **stanovení intervalu tolerance:** Označme symbolem $G(d_{\#}, h_{\#})$ podíl všech hodnot zkoumané náhodné veličiny, který leží v intervalu o dolní popř. horní hranici $d_{\#}$ popř. $h_{\#}$. Tyto hranice však jsou sestrojeny z výběrových hodnot tak, aby zmíněný podíl dosahoval alespoň hodnoty $\gamma$, a to s velkou pravděpodobností $1-\alpha$. Riziko nepravdivého stanovení intervalu tolerance $\alpha$ volíme nejčastěji 0,05 nebo 0,01, hodnotu $\gamma$ podle potřeby kdekoliv mezi hodnotami od 0 do 1. Interval tolerance se opět vyskytuje ve třech formách charakterizovaných pravděpodobnostmi

$$P(G(-\infty, h_{\#}) \geq \gamma) \geq 1-\alpha$$

$$P(G(\, d_{\#}, h_{\#}\,) \geq \gamma) \geq 1-\alpha$$

$$P(G(\, d_{\#},\ \infty\,) \geq \gamma) \geq 1-\alpha$$

## A.5 Kód receptu

Na základě čtyř právě uvedených hledisek byly recepty statistické indukce v kapitole **B** roztříděny takto:

— Nadpisy částí receptáře obsahují římské číslo vyjadřující typ porovnávání dat a malé latinské písmeno odpovídající stupni kvantifikace zkoumané proměnné. Zdvojení tohoto písmene odkazuje na charakteristiku závislosti mezi dvěma proměnnými.

— Kódy jednotlivých receptů jsou uvedeny v rámečku a sestávají z římského čísla vyjadřujícího typ porovnávání dat, z řeckého označení pojednávané teoretické charakteristiky a konečně z velkého latinského písmene odpovídajícího hledanému typu statistické procedury.

$$r_{xy} = \frac{n_{11} n_{22} - n_{12} n_{21}}{\sqrt{n_{1.} n_{2.} n_{.1} n_{.2}}}$$ koef. asociační korelace

$r_{xy} = 1$, $r_{xy} = -1$

entropie

$$h_r = -\sum_{j=1}^{r} p_j \,\mathrm{ld}\, p_j$$ $\mathrm{ld}(p_j)$ – log. dvojkový; entropie

$$\chi^2 = \sum_{j=1}^{r} \frac{(n_j - \hat{n}_j)^2}{\hat{n}_j}$$ $\hat{n}_j = n\varepsilon_j$ ($\varepsilon_j$ = ordinální pravděpodobnost) binomické

$0 \leq \chi^2$

$\chi^2$ (2 rozměr.) očekávané

$$\hat{p}_{jk} = p_{j.} \times p_{.k}$$

$$\chi^2 = n \sum_{j=1}^{r} \sum_{k=1}^{w} \frac{(p_{jk} - \hat{p}_{jk})^2}{\hat{p}_{jk}}$$



$$K_{xy} = \sqrt{\frac{\chi^2}{n \cdot (r-1)(w-1)}}$$ koeficient kontingence; těsnost vztahu

$0 \leq K_{xy} \leq 1$

# B Recepty statistické indukce

## Tabelované hodnoty

| | | | |
|---|---|---|---|
| $u_{1-\alpha/2}$, | $u_{1-\alpha}$ | = | kvantily stand. normálního rozložení |
| $\chi^2_{1-\alpha/2}(\nu)$, | $\chi^2_{1-\alpha}(\nu)$ | = | kvantily Pearsonova rozložení |
| $t_{1-\alpha/2}(\nu)$, | $t_{1-\alpha}(\nu)$ | = | kvantily Studentova rozložení |
| $F_{1-\alpha/2}(\nu_1,\nu_2)$, | $F_{1-\alpha}(\nu_1,\nu_2)$ | = | kvantily Fisherova-Snedecorova rozl. |
| $a$ = | $2\arcsin(\sqrt{p})$ | = | arkussinová transformace |
| $z$ = | $\frac{1}{2}\ln(\frac{1+r}{1-r})$ | = | Fisherova transformace |

## Ia. Jednovýběrové vyšetřování alternativní proměnné

| | | |
|---|---|---|
| $\theta$ | = | pravděpodobnost výskytu |
| $n$ | = | rozsah výběru |
| $n_1$ | = | absolutní četnost výskytu |
| $n_0$ | = | absolutní četnost absence |
| $p$ | = | relativní četnost výskytu |
| | = | $n_1/n$ |

| $X$ | - |
|---|---|
| 1 | $n_1$ |
| 0 | $n_0$ |
| $\Sigma$ | $n$ |

$\boxed{I\ \theta\ S}$ **Přesný interval spolehlivosti pro pravděpodobnost výskytu $\theta$**

- Najdi označení Ia.
- Zvol jednu ze tří forem intervalu spolehlivosti a riziko $\alpha$.
- Vypočti pomocné údaje a vyhledej kvantily:

$$\begin{array}{llll} \nu_1' = 2(n_0+1) & \nu_2' = 2n_1 & F_{1-\alpha/2}(\nu_1',\nu_2') & F_{1-\alpha}(\nu_1',\nu_2') \\ \nu_1'' = 2(n_1+1) & \nu_2'' = 2n_0 & F_{1-\alpha/2}(\nu_1'',\nu_2'') & F_{1-\alpha}(\nu_1'',\nu_2'') \end{array}$$

- Dosaď do zvolené formy intervalu spolehlivosti:

$$\frac{n_1}{n_1+(n_0+1)F_{1-\alpha}(\nu_1',\nu_2')} \le \theta$$

$$\frac{n_1}{n_1+(n_0+1)F_{1-\alpha/2}(\nu_1',\nu_2')} \le \theta \le \frac{(n_1+1)F_{1-\alpha/2}(\nu_1'',\nu_2'')}{n_0+(n_1+1)F_{1-\alpha/2}(\nu_1'',\nu_2'')}$$

$$\theta \le \frac{(n_1+1)F_{1-\alpha}(\nu_1'',\nu_2'')}{n_0+(n_1+1)F_{1-\alpha}(\nu_1'',\nu_2'')}$$

$\boxed{I\ \theta\ S}$ **Přibližný interval spolehlivosti pro pravděpodobnost výskytu $\theta$**

- Najdi označení Ia a ověř předpoklady receptu: $n > 30$ a $0 \not\approx \theta \not\approx 1$
- Zvol jednu ze tří forem intervalu spolehlivosti a riziko $\alpha$.
- Vypočti pomocný údaj a vyhledej kvantily:

$$s_0 = \sqrt{\frac{p(1-p)}{n}} \qquad u_{1-\alpha/2} \qquad u_{1-\alpha}$$

- Dosaď do zvolené formy intervalu spolehlivosti:

$$\begin{array}{rcccl} p - u_{1-\alpha}\, s_0 & \leq & \theta & & \\ p - u_{1-\alpha/2}\, s_0 & \leq & \theta & \leq & p + u_{1-\alpha/2}\, s_0 \\ & & \theta & \leq & p + u_{1-\alpha}\, s_0 \end{array}$$

$\boxed{I\ \theta\ N}$ **Minimální rozsah výběru pro pravděpodobnost výskytu $\theta$**

- Najdi označení Ia a ověř předpoklady receptu: $0 \not\approx \theta \not\approx 1$
- Zvol přípustnou délku oboustranného intervalu spolehlivosti $\delta$ a riziko $\alpha$.
- Vyhledej kvantil $u_{1-\alpha/2}$ a pokud není o $\theta$ předem nic známo, polož

$$\lambda = 1/2$$

Víš-li předem, že nemůže platit $\theta \approx 1/2$, stanov hranici $\lambda$ tak, aby byla splněna jedna z nerovností

$$0 < \theta < \lambda < 1/2 \quad \vdots \quad 1/2 < \lambda < \theta < 1$$

- Dosaď do vzorce pro minimální počet pozorování:

$$n \approx \frac{4\, u_{1-\alpha/2}^2\, \lambda(1-\lambda)}{\delta^2}$$

## $\boxed{I\ \theta\ H}$ Přesnější test hypotézy o pravděpodobnosti výskytu $\theta$

- Najdi označení Ia a ověř předpoklady receptu: $n > 10$
- Zvol riziko $\alpha$, specifikuj číslo $c$ v testované hypotéze

$$\theta = c$$

  a zvol jednu ze tří alternativních hypotéz:

$$\theta < c \quad \vdots \quad \theta \neq c \quad \vdots \quad c < \theta$$

- Vyhledej transformované hodnoty a kvantily:

$$a = 2\arcsin(\sqrt{p}) \qquad b = 2\arcsin(\sqrt{c}) \qquad u_{1-\alpha/2} \qquad u_{1-\alpha}$$

- Vypočti testovou statistiku:

$$u = (a-b)\sqrt{n}$$

- Zamítni testovanou hypotézu, pokud je splněna ta ze tří kritických podmínek, která se pořadím shoduje se zvolenou hypotézou alternativní:

$$u < -u_{1-\alpha} \quad \vdots \quad u < -u_{1-\alpha/2} \text{ nebo } u_{1-\alpha/2} < u \quad \vdots \quad u_{1-\alpha} < u$$

## $\boxed{I\ \theta\ H}$ Přibližnější test hypotézy o pravděpodobnosti výskytu $\theta$

- Najdi označení Ia a ověř předpoklady receptu: $n > 30$ a $0 \not\approx \theta \not\approx 1$
- Zvol riziko $\alpha$, specifikuj číslo $c$ v testované hypotéze

$$\theta = c$$

  a zvol jednu ze tří alternativních hypotéz:

$$\theta < c \quad \vdots \quad \theta \neq c \quad \vdots \quad c < \theta$$

- Vypočti pomocné údaje a vyhledej kvantily:

$$s_0 = \sqrt{\frac{p(1-p)}{n}} \qquad u_{1-\alpha/2} \qquad u_{1-\alpha}$$

- Vypočti testovou statistiku:

$$u = \frac{p-c}{s_0}$$

- Zamítni testovanou hypotézu, pokud je splněna ta ze tří kritických podmínek, která se pořadím shoduje se zvolenou hypotézou alternativní:

$$u < -u_{1-\alpha} \quad \vdots \quad u < -u_{1-\alpha/2} \text{ nebo } u_{1-\alpha/2} < u \quad \vdots \quad u_{1-\alpha} < u$$

## Iaa. Jednovýběrové vyšetřování závislosti alternativních proměnných

| | | |
|---|---|---|
| $\pi(x,y)$ | = | pravděpodobnostní funkce dvojice alternativních proměnných |
| $\pi_1(x)$ | = | pravděpodobnostní funkce první proměnné |
| $\pi_2(y)$ | = | pravděpodobnostní funkce druhé proměnné |
| $\varrho$ | = | teoretický koeficient alternativní korelace |
| $n$ | = | rozsah výběrového souboru |
| $n_{jk}$ | = | absolutní četnost dvojice hodnot $j, k$ |
| $n_{j.}$ | = | absolutní četnost hodnoty $j$ u první proměnné |
| $n_{.k}$ | = | absolutní četnost hodnoty $k$ u druhé proměnné |

| $X\backslash Y$ | 1 | 0 | $\Sigma$ |
|---|---|---|---|
| 1 | $n_{11}$ | $n_{10}$ | $n_{1.}$ |
| 0 | $n_{01}$ | $n_{00}$ | $n_{0.}$ |
| $\Sigma$ | $n_{.1}$ | $n_{.0}$ | $n$ |

$\boxed{I\ \pi(x,y)\ H}$ **Test nezávislosti dvou alternativních proměnných**

- Najdi označení Iaa a ověř předpoklady receptu: $n_{jk} \geq 3$ pro všechna $j, k$
- Zvol riziko $\alpha$, testovanou hypotézu o nezávislosti obou proměnných

$$\pi(x,y) = \pi_1(x)\,\pi_2(y) \quad \text{pro všechna } x, y$$

  a alternativní hypotézu, že předešlá rovnost neplatí pro některá $x.y$.
- Vyhledej kvantil $\chi^2_{1-\alpha}(\nu)$ pro $\nu = 1$.
- Vypočti testovou statistiku:

$$\chi^2 = n\,\frac{(|n_{11}n_{00} - n_{10}n_{01}| - 1/2|)^2}{n_{1.}n_{0.}n_{.1}n_{.0}}$$

- Zamítni testovanou hypotézu, pokud je splněna tato kritická podmínka:

$$\chi^2 > \chi^2_{1-\alpha}(\nu)$$

$\boxed{I\ \varrho\ H}$ **Test hypotézy o koeficientu alternativní korelace $\varrho$**

- Najdi označení Iaa a ověř předpoklady receptu: $n_{jk} \geq 3$ pro všechna $j, k$
- Zvol riziko $\alpha$, užij testované hypotézy

$$\varrho = 0$$

a zvoljednu ze tří alternativních hypotéz:

$$\varrho < 0 \quad \vdots \quad \varrho \neq 0 \quad \vdots \quad 0 < \varrho$$

- Vyhledej kvantily: $u_{1-\alpha/2}$ $\quad u_{1-\alpha}$
- Vypočti testovou statistiku:

$$u = \sqrt{n}\,\frac{n_{11}n_{00} - n_{10}n_{01}}{\sqrt{n_{1.}n_{0.}n_{.1}n_{.0}}}$$

- Zamítni testovanou hypotézu, pokud je splněna ta ze tří kritických podmínek, která se pořadím shoduje se zvolenou hypotézou alternativní:

$$u < -u_{1-\alpha} \quad \vdots \quad u < -u_{1-\alpha/2} \text{ nebo } u_{1-\alpha/2} < u \quad \vdots \quad u_{1-\alpha} < u$$

# In. Jednovýběrové vyšetřování nominální proměnné

| | | |
|---|---|---|
| $\pi(x)$ | = | pravděpodobnostní funkce |
| $n$ | = | rozsah výběru |
| $n_j$ | = | absolutní četnost hodnoty $x_{[j]}$ |

| $X$ | $-$ |
|---|---|
| $x_{[1]}$ | $n_1$ |
| $x_{[1]}$ | $n_2$ |
| $\vdots$ | $\vdots$ |
| $x_{[v]}$ | $n_v$ |
| $\Sigma$ | $n$ |

$\boxed{I\ \pi(x)\ H}$ **Test hypotézy o rozložení jedné nominální proměnné**

- Najdi označení In a ověř předpoklady receptu:
  $n_1 > 4$ a $n_2 > 4$ a ... a $n_v > 4$
- Zvol riziko $\alpha$, specifikuj čísla $c_1, c_2, \ldots, c_v$ v testované hypotéze

$$\pi(x_{[1]}) = c_1 \quad \text{a} \quad \pi(x_{[2]}) = c_2 \quad \text{a} \quad \ldots \quad \text{a} \quad \pi(x_{[v]}) = c_v$$

  a užij alternativní hypotézy, že aspoň jedna z předešlých rovností neplatí.
- Vypočti pomocné hodnoty a vyhledej kvantil:

$$\hat{n}_j = nc_j \quad \text{pro} \quad j = 1, \ldots, v \qquad \nu = v - 1 \qquad \chi^2_{1-\alpha}(\nu).$$

- Vypočti testovou statistiku:

$$\chi^2 = \sum_{j=1}^{v} \frac{(|n_j - \hat{n}_j| - 1/2)^2}{\hat{n}_j}$$

- Zamítni testovanou hypotézu, je-i splněna tato kritická podmínka:

$$\chi^2 > \chi^2_{1-\alpha}(\nu)$$

## Inn. Jednovýběrové vyšetřování závislosti dvou nominálních proměnných

| | | |
|---|---|---|
| $\pi(x,y)$ | = | pravděpodobnostní funkce dvojice alternativních proměnných |
| $\pi_1(x)$ | = | pravděpodobnostní funkce první proměnné |
| $\pi_2(y)$ | = | pravděpodobnostní funkce druhé proměnné |
| $n$ | = | rozsah výběru |
| $n_{jk}$ | = | absolutní četnost dvojice $x_{[j]}, y_{[k]}$ |
| $n_{j.}$ | = | absolutní četnost hodnoty $x_{[j]}$ první proměnné |
| $n_{.k}$ | = | absolutní četnost hodnoty $y_{[k]}$ druhé proměnné |

| $X\backslash Y$ | $y_{[1]}$ | $y_{[2]}$ | $\ldots$ | $y_{[w]}$ | $\Sigma$ |
|---|---|---|---|---|---|
| $x_{[1]}$ | $n_{11}$ | $n_{12}$ | $\ldots$ | $n_{1w}$ | $n_{1.}$ |
| $x_{[1]}$ | $n_{21}$ | $n_{22}$ | $\ldots$ | $n_{2w}$ | $n_{2.}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | | $\vdots$ | $\vdots$ |
| $x_{[v]}$ | $n_{v1}$ | $n_{v2}$ | $\ldots$ | $n_{vw}$ | $n_{v.}$ |
| $\Sigma$ | $n_{.1}$ | $n_{.2}$ | $\ldots$ | $n_{.w}$ | $n$ |

| $I \; \pi(x,y) \; H$ | **Test hypotézy o nezávislosti dvou nominálních proměnných** |
|---|---|

- Najdi označení Inn a ověř předpoklady receptu: $n_{jk} > 2$ pro všechna $j, k$
- Zvol riziko $\alpha$, užij testované hypotézy o nezávislosti dvou proměnných

$$\pi(x,y) \;=\; \pi_1(x)\,\pi_2(y) \quad \text{pro všechna} \quad x, y$$

  a alternativní hypotézy, že předešlá rovnost je pro některá $x, y$ porušena.
- Vypočti pomocné hodnoty a vyhledej kvantil:

$$\frac{n_{jk}^2}{n_{j.}n_{.k}} \;=\; n_{j.}n_{.k} \text{ pro všechna } j,k \qquad \nu \;=\; (v-1)(w-1) \qquad \chi^2_{1-\alpha}(\nu)$$

- Vypočti testovou statistiku:

$$\chi^2 = n\left(\left(\sum_{j=1}^{v}\sum_{k=1}^{w}\frac{n_{jk}^2}{n_{j.}n_{.k}}\right) - 1\right)$$

- Zamítni testovanou hypotézu, je-li splněna tato kritická podmínka:

$$\chi^2 > \chi^2_{1-\alpha}(\nu)$$

# II. Jednovýběrové vyšetřování intervalové proměnné

| | | |
|---|---|---|
| $\mu$ | = | teoretická střední hodnota |
| $\sigma$ | = | teoretická směrodatná odchylka |
| $n$ | = | rozsah vvýběru |
| | | |
| $m$ | = | výběrový průměr |
| $s$ | = | výběrová směrodatná odchylka |

## $\boxed{I\ \mu\ S}$ **Interval spolehlivosti pro střední hodnotu $\mu$**

- Najdi označení Ii a ověř předpoklady receptu: *normální rozložení* nebo $n > 30$
- Zvol jednu ze tří forem intervalu spolehlivosti a riziko $\alpha$.
- Vypočti pomocné údaje a vyhledej kvantily:

$$s_0 = \frac{s}{\sqrt{n}} \qquad \nu = n - 1 \qquad t_{1-\alpha}(\nu) \qquad t_{1-\alpha/2}(\nu)$$

- Dosaď do zvolené formy intervalu spolehlivosti:

$$\begin{array}{rcccl} m - t_{1-\alpha}(\nu)\, s_0 & \leq & \mu & & \\ m - t_{1-\alpha/2}(\nu)\, s_0 & \leq & \mu & \leq & m + t_{1-\alpha/2}(\nu)\, s_0 \\ & & \mu & \leq & m + t_{1-\alpha}(\nu)\, s_0 \end{array}$$

## $\boxed{I\ \mu\ N}$ **Minimální rozsah výběru pro střední hodnotu $\mu$**

- Najdi označení Ii a ověř předpoklady receptu: *normalita rozložení*
- Zvol riziko $\alpha$ a přípustnou délku oboustranného intervalu spolehlivosti $\delta$.
- Z předběžného výběru o rozsahu $n_*$ v mezích asi od 5 do 30 vypočti výběrový rozptyl $s_*^2$ a v tabulkách vyhledej kvantil

$$\nu_* = n - 1 \qquad t_{1-\alpha/2}(\nu_*)$$

- Doplň předběžný výběr na rozsah daný vzorcem:

$$n \approx \frac{4\, t_{1-\alpha/2}^2(\nu_*)\, s_*^2}{\delta^2}$$

$\boxed{I\ \mu\ H}$ **Test hypotézy o střední hodnotě $\mu$**

- Najdi označení Ii a ověř předpoklady receptu:
  *normální rozložení* nebo $n > 30$
- Zvol rlniziko $\alpha$, specifikuj číslo $c$ v testované hypotéze
  $$\mu = c$$
  a zvol jednu ze tří alternativních hypotéz:
  $$\mu < c \quad \vdots \quad \mu \neq c \quad \vdots \quad c < \mu$$
- Vypočti pomocné hodnoty a vyhledej kvantily:
  $$s_0 = \frac{s}{\sqrt{n}} \qquad \nu = n - 1 \qquad t_{1-\alpha/2}(\nu) \qquad t_{1-\alpha}(\nu)$$
- Vypočti testovou statistiku:
  $$t = \frac{m - c}{s_0}$$
- Zamítni testovanou hypotézu, pokud je splněna ta ze tří kritických podmínek, která se pořadím shoduje se zvolenou hypotézou alternativní:
  $$t < -t_{1-\alpha}(\nu) \quad \vdots \quad t < -t_{1-\alpha/2}(\nu) \text{ nebo } t_{1-\alpha/2}(\nu) < t \quad \vdots \quad t_{1-\alpha}(\nu) < t$$

$\boxed{I\ \sigma\ S}$ **Interval spolehlivosti pro směrodatnou odchylku $\sigma$**

- Najdi označení Ii a ověř předpoklady receptu:
  *normální rozložení* nebo $n > 30$
- Zvol jednu ze tří forem intervalu spolehlivosti a riziko $\alpha$.
- Vypočti pomocné údaje a vyhledej kvantily:
  $$\nu = n - 1 \qquad \chi^2_{\alpha/2}(\nu) \qquad \chi^2_{\alpha}(\nu) \qquad \chi^2_{1-\alpha}(\nu) \qquad \chi^2_{1-\alpha/2}(\nu)$$
- Dosaď do zvolené formy intervalu spolehlivosti:
  $$s\sqrt{\frac{n-1}{\chi^2_{1-\alpha}(\nu)}} \leq \sigma$$
  $$s\sqrt{\frac{n-1}{\chi^2_{1-\alpha/2}(\nu)}} \leq \sigma \leq s\sqrt{\frac{n-1}{\chi^2_{\alpha/2}(\nu)}}$$
  $$\sigma \leq s\sqrt{\frac{n-1}{\chi^2_{\alpha}(\nu)}}$$

## $\boxed{I\ \sigma\ N}$ Minimální rozsah výběru pro směrodatnou odchylku $\sigma$

- Najdi označení Ii a ověř předpoklady receptu: *normalita rozložení*
- Zvol riziko $\alpha$ a přípustný poměr horní a dolní hranice oboustranného intervalu spolehlivosti $\kappa$.
- Zkusmo vypočítávej kvantily $\chi^2_{\alpha/2}(\nu)$ $\chi^2_{1-\alpha/2}(\nu)$ a hledej takové $\nu$, pro něž platí

$$\frac{\chi^2_{1-\alpha/2}(\nu)}{\chi^2_{\alpha/2}(\nu)} \approx \kappa^2$$

- Minimální rozsah výběru je dán vzorcem:

$$n = \nu + 1$$

## $\boxed{I\ \sigma\ H}$ Test hypotézy o směrodatné odchylce $\sigma$

- Najdi označení Ii a ověř předpoklady receptu: *normalita rozložení*
- Zvol riziko $\alpha$, specifikuj číslo $c$ v testované hypotéze

$$\sigma = c$$

a zvol jednu ze tří alternativních hypotéz:

$$\sigma < c \quad \vdots \quad \sigma \neq c \quad \vdots \quad c < \sigma$$

- Vypočti pomocné hodnoty a vyhledej kvantily:

$$\nu = n - 1 \qquad \chi^2_{\alpha/2}(\nu) \qquad \chi^2_{\alpha}(\nu) \qquad \chi^2_{1-\alpha}(\nu) \qquad \chi^2_{1-\alpha/2}(\nu)$$

- Vypočti testovou statistiku:

$$\chi^2 = \frac{(n-1)\,s^2}{c^2}$$

- Zamítni testovanou hypotézu, pokud je splněna ta ze tří kritických podmínek, která se pořadím shoduje se zvolenou hypotézou alternativní:

$$\chi^2 < \chi^2_{\alpha}(\nu) \quad \vdots \quad \chi^2 < \chi^2_{\alpha/2}(\nu) \text{ nebo } \chi^2_{1-\alpha/2}(\nu) < \chi^2 \quad \vdots \quad \chi^2_{1-\alpha}(\nu) < \chi^2$$

## Iii. Jednovýběrové vyšetřování závislosti intervalových proměnných

| | | |
|---|---|---|
| $\varrho$ | = | teoretický koeficient lineární korelace |
| $n$ | = | rozsah výběru |
| $r$ | = | výběrový koeficient lineární korelace |

**I ϱ H** **Test hypotézy pro koeficient korelace $\varrho$**

- Najdi označení Iii a ověř předpoklady receptu: *normalita rozložení dvojice proměnných*
- Zvol riziko $\alpha$, specifikuj číslo $c$ v testované hypotéze

$$\varrho = c$$

*a zvol jednu ze tří alternativních hypotéz:*

$$\varrho < c \quad \vdots \quad \varrho \neq c \quad \vdots \quad c < \varrho$$

- Vypočti pomocné hodnoty a vyhledej kvantily

$$z = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) \qquad d = \frac{1}{2}\ln\left(\frac{1+c}{1-c}\right) \qquad u_{1-\alpha/2} \qquad u_{1-\alpha}$$

- Vypočti testovou statistiku:

$$u = (z-d)\sqrt{n-3}$$

- Zamítni testovanou hypotézu, pokud je splněna ta ze tří kritických podmínek, která se pořadím shoduje se zvolenou hypotézou alternativní:

$$u < -u_{1-\alpha} \quad \vdots \quad u < -u_{1-\alpha/2} \text{ nebo } u_{1-\alpha/2} < u \quad \vdots \quad u_{1-\alpha} < u$$

## IIa. Dvojvýběrové porovnávání alternativní proměnné

| | | |
|---|---|---|
| $\theta_1, \theta_2$ | = | porovnávané pravděpodobnosti výskytu |
| $p_1, p_2$ | = | porovnávané relativní četnosti výskytu |

$\boxed{II\ \theta_1 - \theta_2\ S}$ **Interval spolehlivosti pro rozdíl pravděpodobností výskytu $\theta_1 - \theta_2$**

- Najdi označení IIa ověř předpoklady receptu:

$$n_1 > 30 \quad \text{a} \quad n_2 > 30 \quad \text{a} \quad 0 \not\approx \theta_1 \not\approx 1 \quad \text{a} \quad 0 \not\approx \theta_2 \not\approx 1$$

- Zvol jednu ze tří forem intervalu spolehlivosti a riziko $\alpha$.
- Vypočti pomocnou hodnotu a najdi kvantily:

$$s_0 ' = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \qquad u_{1-\alpha/2} \qquad u_{1-\alpha}$$

- Dosaď do zvolené formy intervalu spolehlivosti:

$$\begin{aligned} p_1 - p_2 - u_{1-\alpha}\, s_0 &\leq \theta_1 - \theta_2 & & \\ p_1 - p_2 - u_{1-\alpha/2}\, s_0 &\leq \theta_1 - \theta_2 &\leq\ & p_1 - p_2 + u_{1-\alpha/2}\, s_0 \\ & \quad\ \ \theta_1 - \theta_2 &\leq\ & p_1 - p_2 + u_{1-\alpha}\, s_0 \end{aligned}$$

$\boxed{II\ \theta_1 - \theta_2\ N}$ **Minimální počet pozorování pro rozdíl pravděpodobností výskytu $\theta_1 - \theta_2$**

- Najdi označení IIa a ověř předpoklady receptu:

$$0 \not\approx \theta_1 \not\approx 1 \quad \text{a} \quad 0 \not\approx \theta_2 \not\approx 1$$

- Zvol riziko $\alpha$ a přípustnou délku oboustranného intervalu spolehlivosti $\delta$.
- Vyhledej kvantil $u_{1-\alpha/2}$.
- Minimální rozsah výběru je dán vzorcem:

$$n_1 = n_2 \approx \frac{2u^2_{1-\alpha/2}}{\delta^2}$$

## $\boxed{II\ \theta_1 - \theta_2\ H}$ **Přesnější test hypotézy o rozdílu dvou pravděpodobností výskytu $\theta_1 - \theta_2$**

- Najdi označení IIa a ověř předpoklady receptu:

$$n_1 > 10 \quad \text{a} \quad n_2 > 10 \quad \text{a} \quad 0 \not\approx \theta_1 \not\approx 1 \quad \text{a} \quad 0 \not\approx \theta_2 \not\approx 1$$

- Užij testované hypotézy

$$\theta_1 - \theta_2 = 0$$

  zvol riziko $\alpha$ a jednu ze tří alternativních hypotéz:

$$\theta_1 - \theta_2 < 0 \quad \vdots \quad \theta_1 - \theta_2 \neq 0 \quad \vdots \quad 0 < \theta_1 - \theta_2$$

- Vypočti pomocné hodnoty a vyhledej kvantily:

$$s_0 = \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$a_1 = 2\arcsin(\sqrt{p_1}) \quad a_2 = \arcsin(\sqrt{p_2}) \quad u_{1-\alpha/2} \quad u_{1-\alpha}$$

- Vypočti testovou statistiku:

$$u = \frac{a_1 - a_2}{s_0}$$

- Zamítni testovanou hypotézu, pokud je splněna ta ze tří kritických podmínek, která se pořadím shoduje se zvolenou hypotézou alternativní:

$$u < -u_{1-\alpha} \quad \vdots \quad u < -u_{1-\alpha/2} \text{ nebo } u_{1-\alpha/2} < u \quad \vdots \quad u_{1-\alpha} < u$$

$\boxed{II\ \theta_1 - \theta_2\ H}$ **Přibližnější test hypotézy o rozdílu dvou pravděpodobností výskytu $\theta_1 - \theta_2$**

- Najdi označení IIa a ověř předpoklady receptu:

$$n_1 > 30 \quad \text{a} \quad n_2 > 30 \quad \text{a} \quad 0 \not\approx \theta_1 \not\approx 1 \quad \text{a} \quad 0 \not\approx \theta_2 \not\approx 1$$

- Specifikuj číslo $c$ v testované hydpotéze

$$\theta_1 - \theta_2 = 0$$

zvol riziko $\alpha$ a jednu ze tří alternativních hypotéz:

$$\theta_1 - \theta_2 < 0 \quad \vdots \quad \theta_1 - \theta_2 \neq 0 \quad \vdots \quad 0 < \theta_1 - \theta_2$$

- Vypočti pomocné hodnoty a vyhledej kvantily:

$$s_0 = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \qquad u_{1-\alpha/2} \qquad u_{1-\alpha}$$

- Vypočti testovou statistiku:

$$u = \frac{p_1 - p_2 - c}{s_0}$$

- Zamítni testovanou hypotézu, pokud je splněna ta ze tří kritických podmínek, která se pořadím shoduje se zvolenou hypotézou alternativní:

$$u < -u_{1-\alpha} \quad \vdots \quad u < -u_{1-\alpha/2} \text{ nebo } u_{1-\alpha/2} < u \quad \vdots \quad u_{1-\alpha} < u$$

# IIn. Dvouvýběrové porovnávání nominální proměnné

| | | |
|---|---|---|
| $\pi_1(x)$, $\pi_2(x)$ | = | porovnávané pravděpodobnostní funkce |
| $n_{jk}$ | = | absolutní četnost hodnoty $x_{[j]}$ v $k$–tém výběru |
| $n_{j.}$ | = | celková absolutní četnost hodnoty $x_{[j]}$ v obou výběrech |
| $n_{.k}$ | = | rozsah $k$–tého výběru |
| $n$ | = | celkový rozsah obou výběrů |

| $X\backslash\#$ | 1 | 2 | $\Sigma$ |
|---|---|---|---|
| $x_{[1]}$ | $n_{11}$ | $n_{12}$ | $n_{1.}$ |
| $x_{[1]}$ | $n_{21}$ | $n_{22}$ | $n_{2.}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $x_{[v]}$ | $n_{v1}$ | $n_{v2}$ | $n_{v.}$ |
| $\Sigma$ | $n_{.1}$ | $n_{.2}$ | $n$ |

$\boxed{II\ \pi_1(x), \pi_2(x)\ H}$ **Test hypotézy o shodnosti dvou rozložení $\pi_1(x)$, $\pi_2(x)$**

- Najdi označení IIn a ověř předpoklady receptu:

$$n_{jk} > 2 \text{ pro všechna } j, k$$

- Zvol riziko $\alpha$, užij testované hypotézy o shodnosti obou pravděpodobnostních funkcí

$$\pi_1(x) = \pi_2(x) \text{ pro všechna } x$$

  a alternativní hypotézy, že tato rovnost je pro některá $x$ porušena.

- Vyhledej kvantily

$$\chi^2_{1-\alpha/2}(\nu) \qquad \chi^2_{1-\alpha}(\nu)$$

- Vypočti testovou statistiku

− ?

$$\chi^2 = \frac{1}{n_{.1} n_{.2}} \sum_{j=1}^{v} \frac{(n_{.1} n_{j1} + n_{.2} n_{j2})^2}{n_{j1} + n_{j2}}$$

- Zamítni testovanou hypotézu, pokud je splněna kritická podmínka

$$\chi^2 > \chi^2_{1-\alpha}(\nu)$$

## III. Dvouvýběrové porovnávání intervalové proměnné

| | | |
|---|---|---|
| $\mu_1,\ \mu_2$ | = | porovnávané střední hodnoty |
| $\sigma_1,\ \sigma_2$ | = | porovnávané teoretické směrodatné odchylky |
| $n_1,\ n_2$ | = | rozsahy porovnávaných výběrů |
| $m_1,\ m_2$ | = | porovnávané výběrové průměry |
| $s_1,\ s_2$ | = | porovnávané výběrové směrodatné odchylky |

$\boxed{II\ \mu_1 - \mu_2\ S}$ **Interval spolehlivosti pro rozdíl středních hodnot $\mu_1 - \mu_2$**

- Najdi označení IIi a ověř předpoklady receptu:
  *normální rozložení* nebo $n_1 > 30$ u prvního výběru a
  *normální rozložení* nebo $n_2 > 30$ u druhého výběru
- Zvol jednu ze tří forem intervalu spolehlivosti a riziko $\alpha$.
- Je-li předem známo, že variabilita u obou výběrů je stejná, tj. že $\sigma_1 = \sigma_2$, vypočti pomocné hodnoty:

$$s_0 = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$\nu = n_1 + n_2 - 2 \qquad t_{1-\alpha/2}(\nu) \qquad t_{1-\alpha}(\nu)$$

Není-li takové předběžné informace, vypočti pomocné hodnoty:

$$s_0 = \sqrt{s_1^2/n_1 + s_2^2/n_2}$$

Pak vyhledej kvantily:

$$\nu \approx \left(\frac{1}{n_1 - 1}\left(\frac{s_1^2/n_1}{s_1^2/n_1 + s_2^2/n_2}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2/n_2}{s_1^2/n_1 + s_2^2/n_2}\right)^2\right)^{-1}$$

$$t_{1-\alpha/2}(\nu) \qquad t_{1-\alpha}(\nu)$$

- Dosaď do zvolené formy intervalu spolehlivosti:

$$\begin{aligned} m_1 - m_2 - t_{1-\alpha}(\nu)s_0 \leq\ & \mu_1 - \mu_2 \\ m_1 - m_2 - t_{1-\alpha/2}(\nu)s_0 \leq\ & \mu_1 - \mu_2 \leq m_1 - m_2 + t_{1-\alpha/2}(\nu)s_0 \\ & \mu_1 - \mu_2 \leq m_1 - m_2 + t_{1-\alpha}(\nu)s_0 \end{aligned}$$

$\boxed{II\ \mu_1 - \mu_2\ N}$ **Minimální počty pozorování pro rozdíl středních hodnot $\mu_1 - \mu_2$**

- Najdi označení IIi a ověř předpoklady receptu: *normalita rozložení*
- Zvol riziko $\alpha$ a přípustnou délku oboustranného intervalu spolehlivosti $\delta$.
- Ze dvou předběžných výběrů o rozsahu $n_{*1}, n_{*2}$ v mezích asi od 5 do 30 vypočti výběrové rozptyly $s_{*1}^2, s_{*2}^2$ a vyhledej stupně volnosti a kvantil:

$$\nu \approx \left( \frac{1}{n_{*1} - 1} \left( \frac{s_{*1}^2/n_{*1}}{s_{*1}^2/n_{*1} + s_{*2}^2/n{*2}} \right)^2 + \frac{1}{n_{*2} - 1} \left( \frac{s_{*2}^2/n_{*2}}{s_{*1}^2/n_1 + s_{*2}^2/n_{*2}} \right)^2 \right)^{-1}$$

$$t_{1-\alpha/2}(\nu)$$

- Doplň předběžné výběry na rozsahy dané vzorci:

$$n_1 \approx n_{*1}\, t_{1-\alpha/2}^2(\nu)\, \frac{s_{*1}^2/n_{*1} + s_{*2}^2/n_{*2}}{\delta^2}$$

$$n_2 \approx n_{*2}\, t_{1-\alpha/2}^2(\nu)\, \frac{s_{*1}^2/n_{*1} + s_{*2}^2/n_{*2}}{\delta^2}$$

$\boxed{II\ \mu_1 - \mu_2\ H}$ **Test hypotézy o rozdílu středních hodnot** $\mu_1 - \mu_2$

- Najdi označení IIi a ověř předpoklady receptu:
  *normální rozložení* nebo $n_1 > 30$ u prvního výběru a
  *normální rozložení* nebo $n_2 > 30$ u druhého výběru

- Specifikuj číslo $c$ v testované hypotéze

$$\mu_1 - \mu_2 = c$$

  zvol jednu ze tří alternativních hypotéz

$$\mu_1 - \mu_2 < c \quad \vdots \quad \mu_1 - \mu_2 \neq c \quad \vdots \quad c < \mu_1 - \mu_2$$

  a riziko $\alpha$.

- Je-li předem známo, že variabilita u obou výběrů je stejná, tj. že $\sigma_1 = \sigma_2$, vypočti pomocné hodnoty

$$s_0 = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}(\frac{1}{n_1} + \frac{1}{n_2})}$$

$$\nu = n_1 + n_2 - 2 \qquad t_{1-\alpha/2}(\nu) \qquad t_{1-\alpha}(\nu)$$

  Není-li takové předběžné informace, vypočti pomocné hodnoty

$$s_0 = \sqrt{s_1^2/n_1 + s_2^2/n_2}$$

$$\nu \approx \left(\frac{1}{n_1 - 1}\left(\frac{s_1^2/n_1}{s_1^2/n_1 + s_2^2/n_2}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2/n_2}{s_1^2/n_1 + s_2^2/n_2}\right)^2\right)^{-1}$$

  Pak vyhledej kvantily

$$t_{1-\alpha/2}(\nu) \qquad t_{1-\alpha}(\nu)$$

- Vypočti testovou statistiku:

$$t = \frac{m_1 - m_2 - c}{s_0}$$

- Zamítni testovanou hypotézu, pokud je splněna ta ze tří kritických podmínek, která se pořadím shoduje se zvolenou hypotézou alternativní:

$$t < -t_{1-\alpha}(\nu) \quad \vdots \quad t < -t_{1-\alpha/2}(\nu) \text{ nebo } t_{1-\alpha/2}(\nu) < t \quad \vdots \quad t_{1-\alpha}(\nu)t$$

$\boxed{II\ \sigma_1/\sigma_2\ S}$ **Interval spolehlivosti pro podíl směr. odchylek $\sigma_1/\sigma_2$**

- Najdi označení IIi a ověř předpoklady receptu: *normalita rozložení*
- Zvol jednu ze tří forem intervalu spolehlivosti a riziko $\alpha$.
- Vypočti pomocné hodnoty a vyhledej kvantily:

$$\nu_1 = n_1 - 1 \quad \nu_2 = n_2 - 1$$

$$F_{1-\alpha}(\nu_1, \nu_2) \quad F_{1-\alpha}(\nu_2, \nu_1) \quad F_{1-\alpha/2}(\nu_1, \nu_2) \quad F_{1-\alpha/2}(\nu_2, \nu_1)$$

- Dosaď do zvolené formy intervalu spolehlivosti:

$$\frac{s_1}{s_2}\sqrt{\frac{1}{F_{1-\alpha}(\nu_1, \nu_2)}} \le \frac{\sigma_1}{\sigma_2}$$

$$\frac{s_1}{s_2}\sqrt{\frac{1}{F_{1-\alpha/2}(\nu_1, \nu_2)}} \le \frac{\sigma_1}{\sigma_2} \le \frac{s_1}{s_2}\sqrt{F_{1-\alpha/2}(\nu_2, \nu_1)}$$

$$\frac{\sigma_1}{\sigma_2} \le \frac{s_1}{s_2}\sqrt{F_{1-\alpha}(\nu_2, \nu_1)}$$

$\boxed{II\ \sigma_1/\sigma_2\ N}$ **Minimální rozsah výběru pro podíl směrodatných odchylek $\sigma_1/\sigma_2$**

- Najdi označení IIi a ověř předpoklady receptu: *normalita rozložení*
- Zvol přípustn podíl horní a dolní meze oboustranného intervalu spolehlivosti $\kappa$ a riziko $\alpha$.
- V tabulkách kvantilů najdi zkusmo co nejmenší počet stupňů volnosti $\nu$ vyhovující vztahu

$$F^2_{1-\alpha/2}(\nu, \nu) = \kappa^2$$

- Dosaď do vzorce pro minimální počet pozorování:

$$n_1 = n_2 \approx \nu + 1$$

## $\boxed{II\ \sigma_1/\sigma_2\ H}$ Test hypotézy o podílu směrodatných odchylek $\sigma_1/\sigma_2$

- Najdi označení IIi a ověř předpoklady receptu: *normalita rozložení*
- Zvol riziko $\alpha$, specifikuj číslo $c$ v testované hypotéze

$$\sigma_1/\sigma_2 = c$$

a vyber jednu ze tří alternativních hypotéz:

$$\sigma_1/\sigma_2 < c \quad \vdots \quad \sigma_1/\sigma_2 \neq c \quad \vdots \quad c < \sigma_1/\sigma_2$$

- Vypočti pomocné hodnoty a vyhledej kvantily:

$$\nu_1 = n_1 - 1 \quad \nu_2 = n_2 - 1$$

$$F_{1-\alpha}(\nu_1, \nu_2) \quad F_{1-\alpha}(\nu_2, \nu_1) \quad F_{1-\alpha/2}(\nu_1, \nu_2) \quad F_{1-\alpha/2}(\nu_2, \nu_1)$$

- Vypočti testovou statistiku:

$$F = \frac{s_1^2/s_2^2}{c^2}$$

- Zamítni testovanou hypotézu, pokud je splněna ta ze tří kritických podmínek, která se pořadím shoduje se zvolenou hypotézou alternativní:

$$F < \frac{1}{F_{1-\alpha}(\nu_2,\nu_1)} \vdots F < \frac{1}{F_{1-\alpha/2}(\nu_2,\nu_1)} \text{ nebo } F_{1-\alpha/2}(\nu_1, \nu_2) < F \vdots F_{1-\alpha}(\nu_1, \nu_2) < F$$

## IIii. Dvouvýběrové porovnávání závislosti intervalových proměnných

| | | |
|---|---|---|
| $\varrho_1$, $\varrho_2$ | = | porovnávané teoretické koeficienty lineární korelace |
| $n_1$, $n_2$ | = | rozsahy porovnávaných výběrů |
| $r_1$, $r_2$ | = | porovnávané výběrové koeficienty lineární korelace |

### $\boxed{II\ \varrho_1 - \varrho_2\ H}$ Test hypotézy o rovnosti dvou koeficientů korelace $\varrho_1 = \varrho_2$

- Najdi označení IIii a ověř předpoklady receptu: *normalita rozložení*
- Zvol riziko $\alpha$, užij testované hypotézy

$$\varrho_1 - \varrho_2 = 0$$

a vyber jednu ze tří alternativních hypotéz:

$$\varrho_1 - \varrho_2 < 0 \quad \vdots \quad \varrho_1 - \varrho_2 \neq 0 \quad \vdots \quad 0 < \varrho_1 - \varrho_2$$

- Vyhledej v tabulkách a vypočti pomocné hodnoty

$$z_1 = \frac{1}{2} ln \left( \frac{1+r_1}{1-r_1} \right) \qquad z_2 = \frac{1}{2} ln \left( \frac{1+r_2}{1-r_2} \right)$$

$$s_0 = \sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}$$

- Vypočti testovou statistiku:

$$u = \frac{z_1 - z_2}{s_0}$$

- Zamítni testovanou hypotézu, pokud je splněna ta ze tří kritických podmínek, která se pořadím shoduje se zvolenou hypotézou alternativní:

$$u < -u_{1-\alpha} \quad \vdots \quad u < -u_{1-\alpha/2} \text{ nebo } u_{1-\alpha/2} < u \quad \vdots \quad u_{1-\alpha} < u$$

## IIIn. Vícevýběrové porovnávání nominální proměnné

| | | |
|---|---|---|
| $\pi_1(x), \ldots, \pi_g(x)$ | = | porovnávané pravděpodobnostní funkce |
| # | = | pořadové číslo výběru |
| $n_{jk}$ | = | absolutní četnost hodnoty $x_{[j]}$ v $k$–tém výběru |
| $n_{j.}$ | = | úhrnná absolutní četnost hodnoty $x_{[j]}$ |
| $n_{.k}$ | = | rozsah $k$–tého výběru |
| $n$ | = | úhrnný rozsah všech výběrů |

| $X\backslash\#$ | 1 | 2 | ... | g | $\Sigma$ |
|---|---|---|---|---|---|
| $x_{[1]}$ | $n_{11}$ | $n_{12}$ | ... | $n_{1w}$ | $n_{1.}$ |
| $x_{[1]}$ | $n_{21}$ | $n_{22}$ | ... | $n_{2w}$ | $n_{2.}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | | $\vdots$ | $\vdots$ |
| $x_{[v]}$ | $n_{v1}$ | $n_{v2}$ | ... | $n_{vw}$ | $n_{v.}$ |
| $\Sigma$ | $n_{.1}$ | $n_{.2}$ | ... | $n_{.w}$ | $n$ |

$\boxed{III\ \pi_1(x), ..., \pi_g(x) H}$ **Test shody několika rozložení** $\pi_1(x), ..., \pi_g(x)$

- Najdi označení IIIn a ověř předpoklady receptu:

$$n_{jk} > 3 \text{ pro všechna } j, k$$

- Zvol riziko $\alpha$ a testuj hypotézu o shodě všech rozložení

$$\pi_1(x) = ... = \pi_g(x)$$

proti alternativní hypotéze, že alespoň jedna z rovností je pro některá $x$ porušena.

- Vypočti pomocné hodnoty a vyhledej kvantily:

$$\nu = (v-1)(g-1) \qquad \chi^2_{1-\alpha}(\nu)$$

- Vypočti testovou statistiku:

$$\chi^2 = n\left(\left(\sum_{j=1}^{v}\sum_{k=1}^{g}\frac{n_{jk}^2}{n_{j.}n_{.k}}\right) - 1\right)$$

- Zamítni testovanou hypotézu, pokud je splněna kritická podmínka

$$\chi^2 > \chi^2_{1-\alpha}(\nu)$$

## IVa. Párové porovnávání alternativní proměnné

| | | |
|---|---|---|
| $\theta_1, \theta_2$ | = | porovnávané pravděpodobnosti |
| $n$ | = | rozsah výběrového souboru |
| $n_{jk}$ | = | absolutní četnost dvojice hodnot $j, k$ |
| $n_{j.}$ | = | absolutní četnost hodnoty $j$ u první proměnné |
| $n_{.k}$ | = | absolutní četnost hodnoty $k$ u druhé proměnné |

| $X\backslash Y$ | 1 | 0 | $\Sigma$ |
|---|---|---|---|
| 1 | $n_{11}$ | $n_{10}$ | $n_{1.}$ |
| 0 | $n_{01}$ | $n_{00}$ | $n_{0.}$ |
| $\Sigma$ | $n_{.1}$ | $n_{.0}$ | $n$ |

## $\boxed{IV\ \theta_1 - \theta_2\ H}$ Test hypotézy o rovnosti dvou pravděpodobností výskytu $\theta_1 - \theta_2$

- Najdi označení IVa a ověř předpoklady receptu: $n_{01} > 2$ a $n_{10} > 2$
- Zvol riziko $\alpha$, užij testované hypotézy

$$\theta_1 - \theta 2 = 0$$

  a alternativní hypotézy

$$\theta_1 - \theta 2 \neq 0$$

- Vyhledej kvantil

$$\nu = 1 \qquad \chi^2_{1-\alpha}(\nu)$$

- Vypočti testovou statistiku

$$\chi^2 = \frac{(|n_{10} - n_{01}| - 1)^2}{n_{10} + n_{01}}$$

- Zamítni testovanou hypotézu, pokud je splněna kritická podmínka

$$\chi^2_{1-\alpha}(\nu) < \chi_2$$

## IVi. Párové porovnávání intervalové proměnné

| | | |
|---|---|---|
| $\mu_1,\ \mu_2$ | $=$ | porovnávané střední hodnoty |
| $\sigma_1,\ \sigma_2$ | $=$ | porovnávané teoretické sm. odchylky |
| $n$ | $=$ | rozsah porovnávaného párového výběru |
| $m_1,\ m_2$ | $=$ | porovnávané výběrové průměry |
| $s_1,\ s_2$ | $=$ | porovnávané výběrové směrodatné odchylky |
| $r_{12}$ | $=$ | koeficient lineární korelace v párovém výběru |
| $m$ | $=$ | průměr rozdílového výběru |
| | $=$ | $m_1 - m_2$ |
| $s$ | $=$ | směrodatná odchylka rozdílového výběru |
| | $=$ | $\sqrt{s_1^2 + s_2^2 - 2\, s_1\, s_2\, r_{12}}$ |

(První a druhá složka párového výběru je odlišena indexy 1, 2. Vytvoříme-li z každého páru rozdíl, vzniká rozdílový výběr.)

## $\boxed{IV\ \mu_1 - \mu_2\ S}$ Interval spolehlivosti pro rozdíl středních hodnot $\mu_1 - \mu_2$

- Zvol jednu ze tří forem intervalu spolehlivosti a riziko $\alpha$.
- Najdi označení IVi a ověř předpoklady receptu: *normalita rozložení* nebo $n > 30$
- Vypočti pomocné hodnoty a vyhledej kvantily:

$$s_0 = \frac{s}{\sqrt{n}} \qquad \nu = n - 1 \quad t_{1-\alpha}(\nu) \quad t_{1-\alpha/2}(\nu)$$

- Dosaď do zvolené formy intervalu spolehlivosti:

$$\begin{aligned} m - t_{1-\alpha}(\nu) s_0 \leq & \quad \mu_1 - \mu_2 & \\ m - t_{1-\alpha/2}(\nu) s_0 \leq & \quad \mu_1 - \mu_2 & \leq m + t_{1-\alpha/2}(\nu) s_0 \\ & \quad \mu_1 - \mu_2 & \leq m + t_{1-\alpha}(\nu) s_0 \end{aligned}$$

## $\boxed{IV\ \mu_1 - \mu_2\ N}$ Minimální rozsah výběru pro rozdíl středních hodnot $\mu_1 - \mu_2$

- Najdi označení IVi a ověř předpoklady receptu: *normalita rozložení* nebo $n > 30$
- Zvol riziko $\alpha$ a přípustnou délku oboustranného intervalu spolehlivosti $\delta$.
- Z předběžného rozdílového výběru o rozsahu $n_*$ v mezích asi od 5 do 30 vypočti výběrový rozptyl $s_*^2$ a v tabulkách vyhledej kvantil

$$\nu_* = n_* \qquad t_{1-\alpha/2}(\nu_*)$$

- Doplň předběžný párový výběr na rozsah daný vzorcem:

$$n \approx \frac{4\,t^2_{1-\alpha/2}(\nu_*)\,s_*^2}{\delta^2}$$

## $\boxed{IV\ \mu_1 - \mu_2\ H}$ Testování hypotézy pro rozdíl středních hodnot $\mu_1 - \mu_2$

- Najdi označení IVi a ověř předpoklady receptu: *normalita rozložení* nebo $n > 30$
- Zvol riziko $\alpha$, užij testované hypotézy

$$\mu_1 - \mu_2 \ = \ 0$$

a vyber jednu ze tří alternativních hypotéz:

$$\mu_1 - \mu_2 < 0 \ \ \vdots \ \ \mu_1 - \mu_2 \neq 0 \ \ \vdots \ \ 0 < \mu_1 - \mu_2$$

- Vypočti pomocné hodnoty a vyhledej kvantily:

$$\nu \ = \ n - 1 \qquad t_{1-\alpha}(\nu) \qquad t_{1-\alpha/2}(\nu)$$

- Vypočti testovou statistiku:

$$t \ = \ \frac{m_1 - m_2 - c}{s_0}$$

- Zamítni testovanou hypotézu, pokud je splněna ta ze tří kritických podmínek, která se pořadím shoduje se zvolenou hypotézou alternativní:

$$t < -t_{1-\alpha}(\nu) \ \ \vdots \ \ t < -t_{1-\alpha/2}(\nu) \text{ nebo } t_{1-\alpha/2}(\nu) < t \ \ \vdots \ \ t_{1-\alpha}(\nu) < t$$

## $\boxed{IV\ \sigma_1/\sigma_2\ H}$ Testování hypotézy o podílu směrodatných odchylek

- Najdi označení IVi a ověř předpoklady receptu: *normalita rozložení* nebo $n > 30$

- Zvol riziko $\alpha$, užij testované hypotézy

$$\sigma_1/\sigma_2 \;=\; 1$$

a vyber jednu z alternativních hypotéz

$$\sigma_1/\sigma_2 < 1 \quad \vdots \quad \sigma_1/\sigma_2 \neq 1 \quad \vdots \quad 1 < \sigma_1/\sigma_2$$

- Vypočti pomocné hodnoty a vyhledej kvantily:

$$\nu = n - 2 \qquad t_{1-\alpha}(\nu) \qquad t_{1-\alpha/2}(\nu)$$

- Vypočti testovou statistiku:

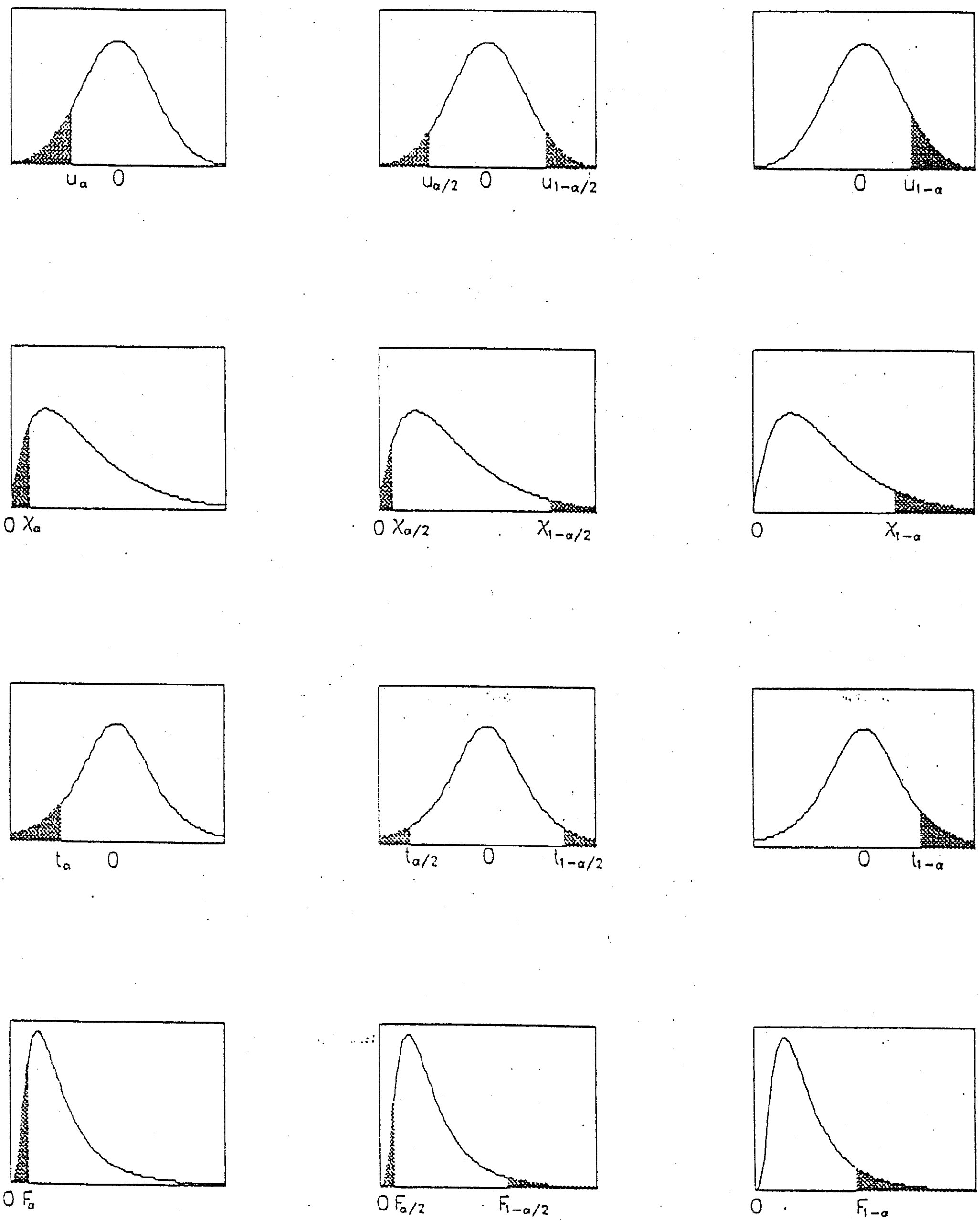
$$t \;=\; \frac{(s_1^2 - s_2^2)\sqrt{n-2}}{2\sqrt{s_1^2 s_2^2(1 - r_{12}^2)}}$$

- Zamítni testovanou hypotézu, pokud je splněna ta ze tří kritických podmínek, která se pořadím shoduje se zvolenou hypotézou alternativní:

$$t < -t_{1-\alpha}(\nu) \quad \vdots \quad t < -t_{1-\alpha/2}(\nu) \text{ nebo } t_{1-\alpha/2}(\nu) < t \quad \vdots \quad t_{1-\alpha}(\nu) < t$$

# Kritické oblasti pro testování hypotéz

$\alpha$ = riziko neoprávněného zamítnutí testované hypotézy




$u_\alpha = -u_{1-\alpha}$

$t_\alpha(\nu) = -t_{1-\alpha}(\nu)$

$F_\alpha(\nu_1, \nu_2) = \dfrac{1}{F_{1-\alpha}(\nu_2, \nu_1)}$

# C Statistické tabulky