12. cvičení z PST

11. prosince 2019

12.1 (test rozptylu normálního rozdělení)

Do laboratoře bylo odesláno $n = 5$ stejných vzorků krve ke stanovení obsahu alkoholu X (v promilích alkoholu). Výsledkem byla realizace

$$\boldsymbol{x} = (0.8,\ \ 1,\ \ 0.6,\ \ 1.4,\ \ 0.9)\ .$$

Posuďte na hladině významnosti $\alpha = 0.05$, zda směrodatná odchylka měření je nejvýše $\sigma_0 = 0.1$ promile alkoholu. Předpokládejte, že obsah alkoholu X má normální rozdělení a jednotlivá měření jsou nezávislá.

Řešení:

Naše veličina X, udávající obsah alkoholu v krvi v promilích, má normální rozdělení $N(\mu, \sigma^2)$. Místo testu směrodatné odchylky σ budeme (ekvivalentně) testovat rozptyl σ^2 a sice nulovou hypotézu tvaru

$$\mathbf{H}_0 : \sigma^2 \leq (0.1)^2\ (= \sigma_0)^2$$

proti alternativní hypotéze:

$$\mathbf{H}_1 : \sigma^2 > (0.1)^2$$

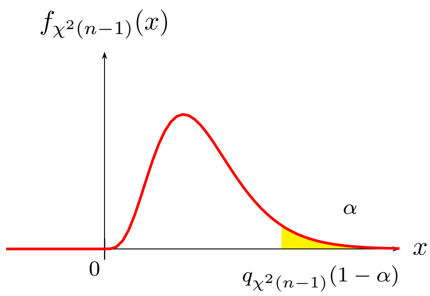
na hladině významnosti $\alpha = 0.05$.

Budeme používat statistiku

$$T = \frac{(n-1)\, S_{\boldsymbol{X}}^2}{\sigma_0^2}\ ,$$

která má pro případ $\sigma = \sigma_0$ tzv. χ^2-rozdělení s $n-1$ stupni volnosti. (V případě $\sigma^2 < \sigma_0^2$ už T nebude mít χ^2-rozdělení. To by měla teprve veličina $\frac{\sigma_0^2}{\sigma^2} \cdot T$.) Střední hodnota $\chi^2(n-1)$-rozdělení je $n-1$.

Za předpokladu nulové hypotézy, tj. pro $0 \leq \sigma \leq \sigma_0$, budou očekávané hodnoty statistiky T především kolem střední hodnoty, tj. především v intervalu $(-\infty, n-1\rangle$. (Ve skutečnosti to bude jen interval $(0, n-1\rangle$, protože T je nezáporná veličina.)



Kritérium pro **ZAMÍTNUTÍ** proto bude tvaru

$$t > q_{\chi^2(n-1)}\,(1-\alpha) \quad \Leftrightarrow \quad \text{zamítáme } \mathbf{H}_0 \text{ (na dané hladině } \alpha)\ .$$

Dosadíme opět konkrétní hodnoty:

$$\overline{\boldsymbol{x}} = \frac{0.8 + 1 + 0.6 + 1.4 + 0.9}{5} = \frac{4.7}{5} = 0.94\ ,$$

$$s_{\boldsymbol{x}}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{\boldsymbol{x}})^2 = \frac{0.14^2 + 0.06^2 + 0.34^2 + 0.46^2 + 0.04^2}{4} = \frac{0.352}{4} = 0.088 \ .$$

Realizace testovací statistiky je

$$t = \frac{(n-1)\, s_{\boldsymbol{x}}^2}{\sigma_0^2} = \frac{4 \cdot 0.088}{(0.1)^2} = 35.2$$

a hodnota kvantilu je

$$q_{\chi^2(n-1)}\,(1-\alpha) = q_{\chi^2(4)}(0.95) \doteq 9.49 \ .$$

Protože

$$t \doteq 35.2 \not> 9.49 \doteq q_{\chi^2(4)}(0.95) \ ,$$

nulovou hypotézu **ZAMÍTÁME**.

12.2 (test střední hodnoty dvou normálních rozdělení se stejným neznámým rozptylem)

Z realizací náhodných veličin X a Y (s normálním rozdělením) jsme z výběrů daného rozsahu obdrželi tyto realizace odhadů:

X	Y
$m = 11$	$n = 21$
$\overline{\boldsymbol{x}} = 10$	$\overline{\boldsymbol{y}} = 12,$
$s_{\boldsymbol{x}} = 2$	$s_{\boldsymbol{y}} = 3$

Na hladině významnosti $\alpha = 0.05$

(a) posuďte hypotézu, že rozptyly náhodných veličin X a Y jsou stejné.

(b) za předpokladu platnosti podmínky dle (a) posuďte hypotézu, že střední hodnoty náhodných veličin X a Y jsou stejné.

Řešení:

(a) **Test stejného rozptylu:** Předpokládáme, že veličiny X a Y jsou *nezávislé* s normálními rozděleními po řadě $N(\mu_1, \sigma_1^2)$ s $N(\mu_2, \sigma_2^2)$. Jednotlivá měření pro X a Y považujeme *všechna navzájem za nezávislá.*

Budeme testovat nulovou hypotézu o rovnosti rozptylů

$$\mathbf{H}_0' : \ \sigma_1^2 = \sigma_2^2$$

proti alternativní hypotéze

$$\mathbf{H}_1' : \ \sigma_1^2 \neq \sigma_2^2 \ .$$

Testovací statistika je

$$T' = \frac{S_{\boldsymbol{X}}^2}{S_{\boldsymbol{Y}}^2} \ ,$$

a má za předpokladu $\sigma_1^2 = \sigma_2^2$ tzv. Fisherovo-Snedecorovo $F(m-1, n-1)$ - rozdělení s $m-1$ a $n-1$ stupni volnosti (v tomto pořadí!).

Za předpokladu nulové hypotézy $\mathbf{H}_0'$ je očekávaná hodnota statistiky T' rovna 1 a kritérium pro **ZAMÍTNUTÍ** tak (podobně jako v některých předchozích příkladech) bude tvaru

$$\left[\ t' < q_{F(m-1,n-1)}\left(\frac{\alpha}{2}\right) \quad \text{nebo} \quad q_{F(m-1,n-1)}\left(1 - \frac{\alpha}{2}\right) < t'\ \right] \quad \Leftrightarrow \quad \text{zamítáme } \mathbf{H}_0' \text{ (na dané hladině } \alpha) \ .$$

Realizace testovací statistiky je

$$t' = \frac{s_{\boldsymbol{x}}^2}{s_{\boldsymbol{y}}^2} = \frac{4}{9} \doteq 0,444$$

a hodnoty kvantilů jsou

$$q_{F(m-1,n-1)}\left(\frac{\alpha}{2}\right) = q_{F(10,20)}(0.025) = \frac{1}{q_{F(20,10)}(0.975)} \doteq \frac{1}{3.42} \doteq 0.2924$$

a

$$q_{F(m-1,n-1)}\left(1-\frac{\alpha}{2}\right) = q_{F(10,20)}(0.975) \doteq 2.77 \ .$$

Protože

$$t' \doteq 0.444 \in \langle 0.2924,\ 2.77 \rangle \ ,$$

hypotézu $\mathbf{H}_0'$, že X *a* Y *mají stejný rozptyl*, **NEZAMÍTÁME**.

(b) **Test rovnosti středních hodnot se stejným neznámým rozptylem:**

Předpokládáme, že veličiny X a Y jsou *nezávislé* s normálními rozděleními po řadě $N(\mu_1, \sigma^2)$ s $N(\mu_2, \sigma^2)$. Tento předpoklad je podložen předchozím testem rovnosti rozptylů, který jsme nezamítli. Jednotlivá měření pro X a Y považujeme opět *všechna navzájem za nezávislá.*

Budeme testovat nulovou hypotézu o rovnosti středních hodnot

$$\mathbf{H}_0 : \ \mu_1 = \mu_2$$

proti alternativní hypotéze

$$\mathbf{H}_1 : \ \mu_1 \neq \mu_2 \ .$$

Testovací statistika je

$$T = \frac{\overline{\boldsymbol{X}} - \overline{\boldsymbol{Y}}}{S\sqrt{1/m + 1/n}} \ ,$$

kde

$$S^2 = \frac{m-1}{m+n-2} \cdot S_{\boldsymbol{X}}^2 + \frac{n-1}{m+n-2} \cdot S_{\boldsymbol{Y}}^2$$

je (vážený) odhad rozptylu. Za předpokladu nulové hypotézy $\mathbf{H}_0$, tj. $\mu_1 = \mu_2$, má statistika T Studentovo $t(m+n-2)$-rozdělení s $m+n-2$ stupni volnosti.

Kritérium pro **ZAMÍTNUTÍ** bude proto (očekávatelně) tvaru

$$|t| > q_{t(m+n-2)}\left(1-\frac{\alpha}{2}\right) \quad \Leftrightarrow \quad \text{zamítáme } \mathbf{H}_0 \text{ (na dané hladině } \alpha) \ .$$

Po dosazení máme

$$s^2 = \frac{(m-1)s_{\boldsymbol{x}}^2 + (n-1)s_{\boldsymbol{y}}^2}{m+n-2} = \frac{10 \cdot 2^2 + 20 \cdot 3^2}{10+20} = \frac{22}{3}$$

a

$$t = \frac{\overline{\boldsymbol{x}} - \overline{\boldsymbol{y}}}{s\sqrt{1/m + 1/n}} = \frac{10-12}{\sqrt{\frac{22}{3}}\sqrt{\frac{1}{11}+\frac{1}{21}}} = -\frac{3}{4}\sqrt{7} \doteq -1.984 \ .$$

Hodnota kvantilu je

$$q_{t(m+n-2)}\left(1-\frac{\alpha}{2}\right) = q_{t(30)}(0.975) \doteq 2.042 \ .$$

Protože

$$|t| \doteq 1.984 \not> 2.042 \doteq q_{t(m+n-2)}\left(1-\frac{\alpha}{2}\right) \ ,$$

hypotézu $\mathbf{H}_0$, ze X *a* Y *mají stejnou střední hodnotu*, také **NEZAMÍTÁME**.

12.3 (párový pokus)

U $n = 8$ praváků jsme změřili délku prostředníčku na pravé a levé ruce, hodnoty v milimetrech uvádí tabulka.

Levá	81	74	90	84	77	67	59	70
Pravá	84	76	89	85	80	69	58	68

Na hladině významnosti $\alpha = 5\%$ posuďte hypotézu, že praváci mají delší prostředníček na levé ruce, a uveďte předpoklady.

Řešení:
Označme si veličiny

X = *"délka prostředníčku na levé ruce"*

Y = *"délka prostředníčku na pravé ruce (u téhož člověka, zde navíc praváka)"*.

Pokud na jednom subjektu provádíme měření více veličin (zde X a Y), pak už jejich vzájemné hodnoty nemůžeme považovat za nezávislé. Za nezávislá ovšem samozřejmě považujeme měření dvojice veličin (X, Y) (tj. náhodného vektoru) u různých lidí.

U veličiny $\Delta := X - Y$, která představuje rozdíly mezi veličinami můžeme přirozeně předpokládat normální rozdělení $N(\mu, \sigma^2)$ (neboť jde o odchylky, které obvykle tuto vlastnost mají). Máme tedy nezávislá měření s hodnotami $\boldsymbol{\delta} = (x_1 - y_1, \ldots, x_n - y_n)$ a naše původní hypotéza $E(X) \geq E(Y)$ lze ekvivalentně vyjádřit pomocí $0 \leq E(X) - E(Y) = E(\Delta) = \mu$ jako nulová hypotéza

$\mathbf{H}_0 : \ \mu \geq 0$

kterou otestujeme proti alternativní hypotéze

$\mathbf{H}_1 : \ \mu < 0$.

na hladině významnosti $\alpha = 5\%$.

Půjde tedy o obvyklý test střední hodnoty (veličiny s normálním rozdělením) při neznámém rozptylu. Použijeme tudíž statistiku

$$T = \frac{\overline{\boldsymbol{\Delta}}}{S_{\boldsymbol{\Delta}}}\sqrt{n}$$

a kritérium pro **ZAMÍTNUTÍ** bude tvaru

$$t < q_{t(n-1)}(\alpha) \quad \Leftrightarrow \quad \text{zamítáme } \mathbf{H}_0 \text{ (na dané hladině } \alpha) \ .$$

Určíme si hodnoty realizace $\boldsymbol{\delta}$ veličiny $\boldsymbol{\Delta} = \boldsymbol{X} - \boldsymbol{Y}$

$\boldsymbol{x}$	81	74	90	84	77	67	59	70
$\boldsymbol{y}$	84	76	89	85	80	69	58	68
$\boldsymbol{\delta} = \boldsymbol{x} - \boldsymbol{y}$	-3	-2	1	-1	-3	-2	1	2

Spočteme její výběrový průměr a rozptyl (pro $n = 8$):

$$\overline{\boldsymbol{\delta}} = -\frac{7}{8} = -0.875 \,, \quad s_{\boldsymbol{\delta}}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\delta_i - \overline{\boldsymbol{\delta}}\right)^2 = \frac{215}{56} \doteq 3.8393 \ ,$$

určíme realizaci statistiky

$$t = \frac{\overline{\boldsymbol{\delta}}}{s_{\boldsymbol{\delta}}}\sqrt{n} = -\frac{7\sqrt{7}}{\sqrt{215}} \doteq -1.263$$

a příslušný kvantil

$$q_{t(n-1)}(\alpha) = -q_{t(n-1)}(1-\alpha) = -q_{t(7)}(0.95) \doteq -1.895 \ .$$

Protože

$$t \doteq -1.263 \not< -1.895 \doteq q_{t(7)}(0.05) \ ,$$

nulovou hypotézu, že praváci mají delší levý prostředníček než pravý, **NEZAMÍTÁME**.

12.4 (test nekorelovanosti dvou výběrů z normálních rozdělení)

Pro realizace

X	22	15	30	27	29
Y	10	6	8	4	8

náhodných výběrů z veličin X, Y testujte na hladině významnosti $\alpha = 5\,\%$ jejich korelovanost.

Řešení:

Pro test korelovanosti je potřeba předpoklad, že náhodný vektor (X, Y) má dvourozměrné normální rozdělení.

Poznámka: Každé takové rozdělení je tvaru

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \underbrace{\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}}_{\mathbb{A}} \begin{pmatrix} X' \\ Y' \end{pmatrix} + \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

kde matice $\mathbb{A}$ je regulární, $\mu_1, \mu_2 \in \mathbb{R}$ a veličiny X', Y' jsou nezávislé s normovaným normálním rozdělením $N(0,1)$.

Označme si ještě $\vec{u} = (\alpha, \beta) \in \mathbb{R}^2$ a $\vec{v} = (\gamma, \delta) \in \mathbb{R}^2$. Snadno je pak vidět, že

$$E(X) = \mu_1, \quad E(Y) = \mu_2, \quad D(X) = \|\vec{u}\|^2, \quad D(Y) = \|\vec{v}\|^2$$

a pro korelaci máme

$$\varrho(X, Y) = \frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\| \cdot \|\vec{v}\|}$$

což je právě kosinus úhlu mezi vektory $\vec{u}$ a $\vec{v}$.

Graf sdružené hustoty pro (X', Y') vznikne rotací grafu hustoty $f_{N(0,1)}$ normálního rozdělení $N(0,1)$ kolem osy vedené počátkem souřadnic (tj. je to takový "kopec"). Sdružená hustota pro (X, Y) pak vznikne prostě jen posunutím středu předchozí hustoty do bodu (μ_1, μ_2), deformací a otočením kolem nového středu, takže vrstevnice této nové hustoty mají tvar natočených soustředných elips. Konkrétně je to:

$$f_{(X',Y')}(x', y') = f_{N(0,1)}(x') \cdot f_{N(0,1)}(y') \qquad \text{pro} \ \ (x', y') \in \mathbb{R}^2$$

$$f_{(X,Y)}(x, y) = \frac{1}{\det \mathbb{A}} \cdot f_{(X',Y')}\Big((x-\mu_1, y-\mu_2) \cdot (\mathbb{A}^T)^{-1}\Big) \qquad \text{pro} \ \ (x, y) \in \mathbb{R}^2$$

My budeme testovat hypotézu o koeficientu korelace $\varrho(X, Y)$ mezi náhodnými veličinami X a Y,

$\mathbf{H}_0 :\ \varrho(X, Y) = 0$ (tj. náhodné veličiny X a Y jsou nekorelované)

proti alternativní hypotéze

$\mathbf{H}_1 :\ \varrho(X, Y) \neq 0$ (tj. náhodné veličiny X a Y jsou korelované.)

K testování použijeme výběrový koeficient korelace $R(\boldsymbol{X}, \boldsymbol{Y})$ a testovou statistiku

$$T = \frac{R(\boldsymbol{X}, \boldsymbol{Y})\sqrt{n-2}}{\sqrt{1 - R^2(\boldsymbol{X}, \boldsymbol{Y})}} \ ,$$

která má Studentovo rozdělení $\mathrm{t}(n-2)$, kde n je rozsah výběrů. Realizaci $r(\boldsymbol{x},\boldsymbol{y})$ výběrového koeficientu korelace $R(\boldsymbol{X},\boldsymbol{Y})$ vypočteme ze vzorce

$$r(\boldsymbol{x},\boldsymbol{y}) = \frac{n\sum\limits_{i=1}^{n} x_i y_i - \left(\sum\limits_{i=1}^{n} x_i\right)\cdot\left(\sum\limits_{i=1}^{n} y_i\right)}{\sqrt{\left(n\sum\limits_{i=1}^{n} x_i^2 - \left(\sum\limits_{i=1}^{n} x_i\right)^2\right)\cdot\left(n\sum\limits_{i=1}^{n} y_i^2 - \left(\sum\limits_{i=1}^{n} y_i\right)^2\right)}} = \frac{n}{n-1}\,\frac{\frac{1}{n}\sum\limits_{j=1}^{n} x_j\,y_j - \overline{\boldsymbol{x}}\,\overline{\boldsymbol{y}}}{s_{\boldsymbol{x}}\,s_{\boldsymbol{y}}}\ .$$

Za předpokladu nulové hypotézy $\mathbf{H}_0$, tj. $\varrho(X,Y) = 0$, je očekávaná hodnota statistiky T rovna 0. Kritérium pro **ZAMÍTNUTÍ** proto (podobně jako pro některé předchozí testy) bude tvaru

$$|t| > q_{t(n-2)}\left(1-\frac{\alpha}{2}\right) \quad\Leftrightarrow\quad \text{zamítáme } \mathbf{H}_0 \text{ (na dané hladině } \alpha)\ .$$

Je $n = 5$,

$$\sum_{i=1}^{n} x_i = 123,\quad \sum_{i=1}^{n} y_i = 36,\quad \sum_{i=1}^{n} x_i^2 = 3\,179,\quad \sum_{i=1}^{n} y_i^2 = 280,\quad \sum_{i=1}^{n} x_i y_i = 890\ .$$

Po dosazení hodnot dostaneme

$$r(\boldsymbol{x},\boldsymbol{y}) = \frac{4450-4428}{\sqrt{766\cdot 104}} = \frac{11}{2\sqrt{4979}} \doteq 0.07794$$

$$t = \frac{r(\boldsymbol{x},\boldsymbol{y})\,\sqrt{n-2}}{\sqrt{1-r^2(\boldsymbol{x},\boldsymbol{y})}} = \sqrt{\frac{363}{19795}} \doteq 0.1354.$$

Z tabulek nalezneme kvantil

$$q_{t(n-2)}\left(1-\frac{\alpha}{2}\right) = q_{\mathrm{t}(3)}(0.975) \doteq 3.18\ .$$

Protože

$$|t| \doteq 0.1354 \not> 3.18 \doteq q_{\mathrm{t}(3)}(0.975)\ ,$$

hypotézu $\mathbf{H}_0$ **NEZAMÍTÁME**.

Poznámky k testu dobré shody: Chceme otestovat (na hladině α), jestli daná veličina X s konečně mnoha (navzájem různými!) hodnotami $a_1,\dots,a_k$ (ne nutně číselnými) má předepsané pravděpodobnosti $(p_1,\dots,p_k)$, tedy nulovou hypotézu

$\mathbf{H}_0 : P(X = a_i) = p_i$ pro všechna $i \in \{1,\dots,k\}$

proti alternativní hypotéze:

$\mathbf{H}_A : P(X = a_{i_0}) \neq p_{i_0}$, pro alespoň jedno $i_0 \in \{1,\dots,k\}$

Při n pokusech s veličinou X si pro $i = 1,\dots,k$ označme veličiny

$$N_i = \text{„}\textit{počet výskytů případu } X = a_i \textit{ při } n \textit{ pokusech}\text{''}\ .$$

Máme tedy náhodný vektor

$$\boldsymbol{N} = (N_1,\dots,N_k)$$

a vztah $N_1+\cdots+N_k = n$. Už z toho vidíme, že veličiny N_i nejsou nezávislé, ale zase k té nezávislosti tak daleko nemají. Náhodný vektor $\boldsymbol{N}$ má tzv. multinomické rozdělení a jednotlivá marginální rozdělení veličin jsou binomická, konkrétně $N_i \sim \mathrm{Bi}(n,p_i)$. Speciálně tedy $E(N_i) = n\cdot p_i$.

Jako testovací veličinu zde používáme:

$$T = \sum_{i=1}^{k} \frac{(N_i - n \cdot p_i)^2}{n \cdot p_i}$$

která má asymptoticky (tj. pro $n \to \infty$) tzv. χ^2-rozdělení s $k-1$ stupni volnosti. Pro praktické použití této asymptotiky se obvykle požaduje, aby platilo, že

$$n \cdot p_i \geq 5 \quad \text{pro všechna} \quad i \in \{1, \ldots, k\} \ .$$

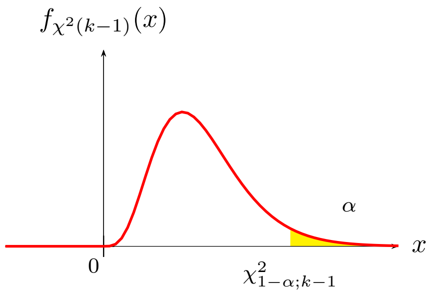
Hodnoty $n \cdot p_i$ se označují jako tzv. *teoretické četnosti.*

Pokud tedy platí nulová hypotéza $\mathbf{H}_0$, měly by být hodnoty veličiny T malé. Jestliže hodnoty T budou příliš velké, bude to důvod k zamítnutí nulové hypotézy.

Jak určit hranici, kde už nastane zamítnutí: veličina T má (přibližně) $\chi^2(k-1)$ rozdělení, tedy platí

$$P_{(H_0 \text{ platí})}\big(T > q_{\chi^2(k-1)}(1-\alpha)\big) \doteq \alpha$$

kde $q_{\chi^2(k-1)}(1-\alpha)$ je hodnota kvantilu pro $\chi^2(k-1)$ rozdělení (viz obrázek, kde α je velikost žluté plochy pod hustotou $f_{\chi^2(k-1)}(x)$ pro $\chi^2(k-1)$ rozdělení).



Kritérium pro **ZAMÍTNUTÍ** $\mathbf{H}_0$ (na hladině α) proto volíme jako

$$t > q_{\chi^2(k-1)}(1-\alpha) \quad \Leftrightarrow \quad \text{zamítáme } \mathbf{H}_0 \text{ (na hladině } \alpha) \ .$$

Z definice chyby 1. druhu, tj.

$$\text{nastává chyba 1. druhu} \quad \Leftrightarrow \quad (\text{hypotéza } \mathbf{H}_0 \text{ platí} \quad \& \quad \text{my ji zamítáme})$$

pak totiž máme, že

$$P_{(H_0 \text{ platí})}\Big(\text{nastává chyba 1. druhu}\Big) = P_{(H_0 \text{ platí})}\Big(\text{zamítáme } \mathbf{H}_0 \text{ (na hladině } \alpha)\Big) =$$

$$= P_{(H_0 \text{ platí})}\big(T > q_{\chi^2(k-1)}(1-\alpha)\big) \doteq \alpha$$

neboli pravděpodobnost chyby 1. druhu (ovšem za předpokladu platnosti $\mathbf{H}_0$!) je pak omezena hodnotou α.

- -

12.5 (test dobré shody - geometrické rozdělení)

Realizací náhodné veličiny X jsme dostali následující četnosti výsledků:

hodnota	0	1	2	3	4	5	6
pozorovaná četnost	29	15	10	5	3	0	2

Posuďte na hladině významnosti $\alpha = 0.05$ hypotézu, že náhodná veličina X má geometrické rozdělení s parametrem $q = 1/2$, tj. pravděpodobnostní funkce je

$$p_X(i) = q^i(1-q), \quad i \in \mathbb{N}_0 \ .$$

Řešení:
Veličina s geometrickým rozdělením nabývá nekonečně mnoha hodnot. Test dobré shody je ale možné dělat jen s veličinou s *konečně* mnoha hodnotami. Proto musíme některé hodnoty sloučit do jediné skupiny. Zde se přirozeně nabízí udělat to pro hodnoty 6 a výše. Pravděpodobnost pro tuto skupinu je pak součet pravděpodobností jednotlivých hodnot v této skupině. V našem případě je

$$P(X \geq 6) = 1 - P(X < 6) = 1 - \sum_{i=0}^{5} p_X(i) = 1 - \left(\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \frac{1}{32} + \frac{1}{64}\right) = \frac{1}{64} \ .$$

Při testu dobré shody porovnáváme naměřené četnosti s očekávanými četnostmi. Rozsah souboru (tj. počet měření) je $n = 29 + 15 + 10 + 5 + 3 + 0 + 2 = 64$. Naší tabulku tedy zpřesníme a doplníme o teoretické pravděpodobnosti p_i a teoretické (tj. očekávané) četnosti $n \cdot p_i$:

položka i	0	1	2	3	4	5	≥ 6
pozorovaná četnost n_i	29	15	10	5	3	0	2
teoretická pravděpodobnost p_i	1/2	1/4	1/8	1/16	1/32	1/64	1/64
teoretická četnost $n \cdot p_i$	32	16	8	4	2	1	1

Další podmínkou pro test dobré shody je to, aby jednotlivé položky měly TEORETICKÉ četnosti $n \cdot p_i \geq 5$. Pokud tomu tak není, je potřeba položky vhodně sloučit tak, abychom této hranice dosáhli. Zde se opět nabízí udělat to pro hodnoty $i \geq 3$.

Původní veličinu X tedy nakonec nahradíme veličinou X' popsanou následující tabulkou:

položka i	0	1	2	≥ 3
pozorovaná četnost n_i	29	15	10	10
teoretická pravděpodobnost p_i	1/2	1/4	1/8	1/8
teoretická četnost $n \cdot p_i$	32	16	8	8

Nyní už můžeme zformulovat naši nulovou hypotézu

$\mathbf{H}_0$: pro pravděpodobnosti hodnot veličiny X' platí $\left(p_0, \ p_1, \ p_2, \ p_{\geq 3}\right) = \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}\right)$,

kterou budeme testovat proti alternativní hypotéze:

$\mathbf{H}_1$: pro pravděpodobnosti hodnot veličiny X' platí $\left(p_0, \ p_1, \ p_2, \ p_{\geq 3}\right) \neq \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}\right)$.

Pro test dobré shody používáme určitou statistiku T (viz poznámka výše), jejíž *realizace* t se počítá vzorcem

$$t = \sum_{i \in K} \frac{(n_i - n \cdot p_i)^2}{n \cdot p_i} \ ,$$

kde K je množina položek veličiny X a $k = |K|$ je jejich počet. Rozdělení statistiky T se pro $n \to \infty$ blíží k $\chi^2(k-1)$-rozdělení s $k-1$ stupni volnosti (právě kvůli přibližnosti jsme také potřebovali teoretické četnosti ≥ 5).

Kritérium pro **ZAMÍTNUTÍ** (viz opět poznámka výše) bude podobné jako u jednostranného testu rozptylu (protože jde opět o χ^2-rozdělení). Je tedy tvaru

$$t > q_{\chi^2(k-1)}\,(1-\alpha) \quad \Leftrightarrow \quad \text{zamítáme } \mathbf{H}_0 \text{ (na dané hladině } \alpha) \ .$$

V našem případě máme $k = 4$. Hodnota statistiky je

$$t = \frac{(29-32)^2}{32} + \frac{(15-16)^2}{16} + \frac{(10-8)^2}{8} + \frac{(10-8)^2}{8} = 1.34375$$

a hodnota kvantilu je

$$q_{\chi(k-1)}(1-\alpha) = q_{\chi(3)}(0.95) \doteq 7.815 \ .$$

Protože

$$t = 1.34375 \not> 7.815 \doteq q_{\chi(3)}(0.95) \ ,$$

nulovou hypotézu $\mathbf{H}_0$ pro veličinu X' **NEZAMÍTÁME**. Tento výsledek interpretujeme tak, že hypotézu

X má geometrické rozdělení s parametrem $q = 1/2$,

rovněž **NEZAMÍTÁME**.

12.6 (test dobré shody - rozdělení dané geometrickou pravděpodobností)

Chceme zjistit, zda si jistý druh ptáka buduje hnízda rovnoměrně po krajině. K tomu jsme rozdělili testovací region na 6 souvislých částí, jejichž rozlohy v km^2 jsou uvedeny v tabulce. Tabulka udává i počet hnízd nalezených v dané části regionu. Za hladinu významnosti považujte $\alpha = 5$ %.

oblast	A	B	C	D	E	F
rozloha (v km^2)	5	10	10	5	15	15
počet hnízd	14	22	28	12	40	34

Řešení:

Využijeme test dobré shody. Počet hnízd v regionu je $n = 14 + 22 + 28 + 12 + 40 + 34 = 150$. Označíme $X_i,\ i \in \{1, ..., n\}$, veličinu nabývající hodnot $A, ..., F$, v závislosti na tom, v které části regionu leží i-té hnízdo. Neboli máme veličinu

$$X(\text{“\textit{dané hnízdo}”}) = \text{“\textit{oblast, ve které hnízdo leží}”}$$

a X_i jsou její kopie v jednotlivých pokusech (tedy jsou to nezávislé veličiny se stejným rozdělením).

Celkově rozloha testovacího regionu je $S = 5+10+10+5+15+15 = 60$ km^2. Protože předpokládáme rovnoměrné rozdělení po krajině, bude pravděpodobnost p_j nalezení hnízda v dané oblasti $j \in \{A, ..., F\}$ úměrná její velikosti, tj. daná geometrickou pravděpodobností jako

$$p_j = \frac{\text{“rozloha oblasti } j\text{”}}{S} \ .$$

Můžeme tedy doplnit tabulku následovně:

oblast	A	B	C	D	E	F
rozloha (v km^2)	5	10	10	5	15	15
počet hnízd n_j	14	22	28	12	40	34
teoretická pravděpodobnost p_j	$\frac{5}{60} = \frac{1}{12}$	$\frac{10}{60} = \frac{1}{6}$	$\frac{10}{60} = \frac{1}{6}$	$\frac{5}{60} = \frac{1}{12}$	$\frac{15}{60} = \frac{3}{12}$	$\frac{15}{60} = \frac{3}{12}$
teoretický počet hnízd np_j	12.5	25	25	12.5	37.5	37.5
příspěvek ke statistice $\frac{(n_j-np_j)^2}{np_j}$	0.18	0.36	0.36	0.02	0.167	0.327

Hypotézu tedy vyjádříme konkrétně:

$\mathbf{H}_0$: pro pravděpodobnosti hodnot veličiny X platí $(p_A, \ldots, p_F) = \left(\frac{1}{12}, \ldots, \frac{3}{12}\right)$,

a alternativní hypotéza bude:

$\mathbf{H}_1$: pro pravděpodobnosti hodnot veličiny X platí $(p_A, \ldots, p_F) \neq \left(\frac{1}{12}, \ldots, \frac{3}{12}\right)$.

Realizace testovací statistiky T je

$$t = \sum_{j=A}^{F} \frac{(n_j - np_j)^2}{np_j} = 0.18 + \cdots + 0.327 = 1.413$$

Porovnáme ji s tabulkovou hodnotou kvantilu

$$q_{\chi^2(5)}(1-\alpha) \doteq 11.1 > 1.413 = t \ .$$

Na hladině významnosti 5 % tedy nemůžeme zamítnout, že si pták buduje hnízda rovnoměrně po krajině.

Poznámky k testu nezávislosti: Máme veličiny

- X s (různými) hodnotami $\{a_1, \ldots, a_k\}$ a
- Y s (různými) hodnotami $\{b_1, \ldots, b_\ell\}$

a chceme otestovat (na hladině α), hypotézu

$\mathbf{H}_0$: rozdělení veličin X a Y jsou *nezávislá*

proti alternativní hypotéze:

$\mathbf{H}_A$: rozdělení veličin X a Y jsou *závislá*

Při n pokusech s náhodným vektorem (X, Y) si pro $i = 1, \ldots, k$ označme veličiny

$$N_{i,j} = \text{“\textit{počet výskytů případu} } (X,Y) = (a_i, b_j) \text{ \textit{při} } n \text{ \textit{pokusech}”} \ .$$

a opět máme náhodný vektor

$$\boldsymbol{N} = (N_{1,1}, \ldots, N_{k,\ell})$$

s multinomickým rozdělením. Marginální rozdělení jednotlivých veličin $N_{i,j}$ jsou opět binomická a *za předpokladu nezávislosti* X *a* Y mají střední hodnotu

$$E(N_{i,j}) = n \cdot P(X = a_i, Y = b_j) \overset{\text{(nezáv.)}}{=} n \cdot P(X = a_i) \cdot P(Y = b_j) \ .$$

Podobně jako v testu dobré shody by tyto střední hodnoty představovaly teoretické četnosti až na to, že pravděpodobnosti $P(X = a_i)$ a $P(Y = b_j)$ nemáme v hypotéze uvedeny. Proto je odhadneme jako

$$P(X = a_i) \doteq \frac{n_{i,\bullet}}{n} \quad \text{a} \quad P(Y = b_j) \doteq \frac{n_{\bullet,j}}{n}$$

kde $n_{i,\bullet}$ a $n_{\bullet,j}$ jsou naměřené hodnoty veličin

$$N_{i,\bullet} = \sum_{j=1}^{\ell} N_{i,j} = \text{“\textit{počet výskytů případu} } X = a_i \text{ \textit{při} } n \text{ \textit{pokusech}”}$$

$$N_{\bullet,j} = \sum_{i=1}^{k} N_{i,j} = \text{“\textit{počet výskytů případu} } Y = b_j \text{ \textit{při} } n \text{ \textit{pokusech}”}$$

což jsou tzv. *marginální četnosti.*

Z tohoto důvodu jako testovací veličinu volíme:

$$T = \sum_{i=1}^{k} \sum_{j=1}^{\ell} \frac{\left(N_{i,j} - \frac{N_{i,\bullet} \cdot N_{\bullet,j}}{n}\right)^2}{\frac{N_{i,\bullet} \cdot N_{\bullet,j}}{n}}$$

která má asymptoticky (tj. pro $n \to \infty$) opět χ^2-rozdělení, tentokrát ale s $(k-1) \cdot (\ell - 1)$ stupni volnosti. Pro praktické použití této asymptotiky se obvykle opět požaduje, aby platilo, že

$$\frac{n_{i,\bullet} \cdot n_{\bullet,j}}{n} \geq 5 \quad \text{pro všechna} \quad i = 1, \ldots, k \text{ a } j = 1, \ldots, \ell.$$

Kritérium pro **ZAMÍTNUTÍ** $\mathbf{H}_0$ (na hladině α) volíme podobně jako u testu dobré shody a sice

$$t > q_{\chi^2_{(k-1)\cdot(\ell-1)}}(1-\alpha) \quad \Leftrightarrow \quad \text{zamítáme } \mathbf{H}_0 \text{ (na hladině } \alpha) .$$

- -

12.7 (test nezávislosti veličin)

Na $n = 100$ lidech byla pozorována barva očí a vlasů. Data jsou shrnuta v tabulce. Na hladině $\alpha = 5\%$ testujte hypotézu o nezávislosti barvy očí a vlasů.

Oči \ Vlasy	tmavé	světlé
modré	10	20
šedé	10	10
hnědé	40	10

Řešení:

Označme si veličiny

X= *"barva očí daného člověka"*

Y= *"barva vlasů daného člověka"*

a dál budeme pracovat s náhodným vektorem (X, Y), tj. u daného člověka budeme zjišťovat barvu očí a barvu vlasů.

Budeme testovat hypotézu:

$\mathbf{H}_0$: rozdělení veličin X a Y jsou *nezávislá*

proti alternativní hypotéze:

$\mathbf{H}_1$: rozdělení veličin X a Y jsou *závislá.*

na hladině významnosti $\alpha = 5\%$.

Četnost případu $(X, Y) = (i, j)$ v tabulce označme jako $n_{i,j}$ a marginální četnosti pak budou

$n_{i,\bullet} = \sum_j n_{i,j}$ pro případ $X = i$

$n_{\bullet,j} = \sum_i n_{i,j}$ pro případ $Y = j$.

což jsou součty v řádcích a sloupcích tabulky:

$n_{i,j}$ $(Y =)\ j$ $(X =)\ i$	tmavé	světlé	$n_{i,\bullet}$
modré	10	20	30
šedé	10	10	20
hnědé	40	10	50
$n_{\bullet,j}$	60	40	

Za předpokladu $\mathbf{H}_0$ pak jako teoretické četnosti budeme chápat hodnoty $\frac{n_{i,\bullet}\cdot n_{\bullet,j}}{n}$ v této tabulce:

$\frac{n_{i,\bullet}\cdot n_{\bullet,j}}{n}$ $(Y =)\ j$ $(X =)\ i$	tmavé	světlé	$n_{i,\bullet}$
modré	$\frac{30\cdot 60}{100} = 18$	$\frac{30\cdot 40}{100} = 12$	30
šedé	$\frac{20\cdot 60}{100} = 12$	$\frac{20\cdot 40}{100} = 8$	20
hnědé	$\frac{50\cdot 60}{100} = 30$	$\frac{50\cdot 40}{100} = 20$	50
$n_{\bullet,j}$	60	40	

Podmínka na tyto teoretické (tj. očekávané) četnosti ≥ 5 je splněna, takže test nezávislosti můžeme použít. Pro hodnotu testovací statistiky dostaneme

$$t = \sum_{i,j} \frac{\left(n_{i,j} - \frac{n_{i,\bullet}\cdot n_{\bullet,j}}{n}\right)^2}{\frac{n_{i,\bullet}\cdot n_{\bullet,j}}{n}} =$$

$$= \frac{(10-18)^2}{18} + \frac{(10-12)^2}{12} + \frac{(40-30)^2}{30} + \frac{(20-12)^2}{12} + \frac{(10-8)^2}{8} + \frac{(10-20)^2}{20} = 18 + \frac{1}{18} \doteq 18.056\ .$$

Tuto hodnotu dále porovnáme s hodnotou kvantilu χ^2 pro $(k-1)(\ell-1)$ stupňů volnosti, kde k je počet položek veličiny X a ℓ je počet položek veličiny Y. Tento počet je nyní jiný, než by byl u "obvyklého" testu dobré shody s $k\cdot\ell$ položkami, protože data jsme použili k odhadu marginálních pravděpodobností.

Kritérium pro **ZAMÍTNUTÍ H**$_0$ (na hladině α) bude tedy tvaru

$$t > q_{\chi^2_{(k-1)(\ell-1)}}(1-\alpha) \quad \Leftrightarrow \quad \text{zamítáme } \mathbf{H}_0 \text{ (na hladině } \alpha)\ .$$

Hledaný kvantil je

$$q_{\chi^2_{(3-1)(2-1)}}(1-\alpha) = q_{\chi^2(2)}(0.95) \doteq 5.992\ .$$

Protože

$$t \doteq 18.056 > 5.992 \doteq q_{\chi^2(2)}(0.95)\ ,$$

hypotézu o nezávislosti **ZAMÍTÁME**.

Poznámky k testu shody rozdělení dvou veličin: Máme veličiny X a Y se *stejným* oborem hodnot tvořeným (různými) hodnotami $\{a_1, \ldots, a_k\}$ a chceme otestovat (na hladině α), hypotézu

$\mathbf{H}_0$: veličiny X a Y mají stejná rozdělení

proti alternativní hypotéze:

$\mathbf{H}_A$: veličiny X a Y nemají stejná rozdělení

Rozdělení X ani Y neznáme. Za předpokladu platnosti $\mathbf{H}_0$, tj.

$$P(X = a_i) = p_i = P(Y = a_i) \text{ pro všechna } i = 1, \dots, k$$

můžeme p_i odhadnout průměrnou hodnotou

$$p_i \doteq a_i \quad \text{kde} \quad a_i := \frac{n_i + m_i}{n + m}$$

a n_i je naměřená četnost položky i pro veličinu X a (podobně) m_i je naměřená četnost položky i pro veličinu Y. Statistika T se pak volí taková, že její realizace t je dána jako

$$t = \sum_{i=1}^{k} \frac{(n_i - n \cdot a_i)^2}{n \cdot a_i} + \sum_{i=1}^{k} \frac{(m_i - m \cdot a_i)^2}{m \cdot a_i}$$

kde hodnoty $n \cdot a_i$ a $m \cdot a_i$ chápeme jako teoretické četnosti a proto opět požadujeme, aby platilo, že

$$n \cdot a_i \geq 5 \quad \text{a} \quad m \cdot a_i \geq 5 \quad \text{pro všechna } i$$

Zamítací kritérium, kde používáme kvantil $q_{\chi^2(k-1)}$ (pozor, počet stupňů volnosti je jen $k-1$), je pak tvaru

$$t > q_{\chi^2(k-1)}(1-\alpha) \quad \Leftrightarrow \quad \text{zamítáme } \mathbf{H}_0 \text{ (na hladině } \alpha) \; .$$

12.8 (test shody dvou rozdělení)

Následující tabulka uvádí počty studentů, kteří dosáhli daného bodového ohodnocení z testu v jednotlivých skupinách (v 1. a 2. skupině).

počet bodů (i)	1	2	3	4	5	rozsah
počet studentů v 1. skupině (n_i)	4	2	1	6	8	$n = 21$
počet studentů ve 2. skupině (m_i)	2	0	6	4	13	$m = 25$

Otestujte na hladině významnosti $\alpha = 5\,\%$ hypotézu, že obě skupiny mají stejná rozdělení pravděpodobnosti dosažených výsledků (tj. že mezi tím, jak dopadl test v obou skupinách, není statisticky významný rozdíl na dané hladině α).

Řešení:
Jde o test shody dvou rozdělení veličin:

$X($ *"student z 1. skupiny"* $) :=$ *"počet bodů, kterých dosáhl"*

$Y($ *"student ze 2. skupiny "* $) :=$ *"počet bodů, kterých dosáhl"*

Obě veličiny mají pochopitelně stejný obor hodnot (jinak bychom se ani nemohli ptát na to, jestli mají stejná rozdělení). Hodnoty obou veličin (které raději budeme označovat jako "položky") nemusíme nutně chápat číselně. Mohly by být klidně označeny i písmeny (např. A,B,C,D,E namísto 1,2,3,4,5).

Chceme tedy otestovat (na hladině α), hypotézu

$\mathbf{H}_0$: veličiny X a Y mají stejná rozdělení

proti alternativní hypotéze:

$\mathbf{H}_A$: veličiny X a Y nemají stejná rozdělení

Součet rozsahu obou souborů je $n+m = 21+25 = 46$. Protože budeme potřebovat splnit podmínku pro dostatečně velké teoretické četnosti, nejdříve si je spočítáme, abychom viděli, jestli bude potřeba nějaké položky slučovat:

položka i	1	2	3	4	5
naměřené četnosti $m_i + n_i$	6	2	7	10	21
odhad pravděpodobnosti $a_i = \frac{n_i+m_i}{n+m}$	$\frac{6}{46}$	$\frac{2}{46}$	$\frac{7}{46}$	$\frac{10}{46}$	$\frac{21}{46}$
teoretické četnosti $n \cdot a_i$	$21 \cdot \frac{6}{46} \doteq 2.74$	$21 \cdot \frac{2}{46} \doteq 0.91$	3.2	4.57	9.59
teoretické četnosti $m \cdot a_i$	$25 \cdot \frac{6}{46} \doteq 3.26$	$25 \cdot \frac{2}{46} \doteq 1.09$	3.8	5.44	11.41

Vidíme, že v prvních čtyřech sloupcích nejsou dostatečně velké teoretické četnosti. Když je sloučíme, dosáhneme teoretické četnosti alespoň 5. (Také bychom ovšem mohli sloučit např. první tři sloupce do jedné položky a zbylé dva do druhé položky.) Tím samozřejmě vytvoříme nové veličiny X' a Y' a hypotézu o stejných rozděleních pak testujeme pro tyto nové veličiny. Máme tedy novou tabulku:

položka i	1až4	5
naměřené četnosti n_i' (pro X')	13	8
naměřené četnosti m_i' (pro Y')	12	13
naměřené četnosti $m_i' + n_i'$	25	21
odhad pravděpodobnosti $a_i' = \frac{n_i'+m_i'}{n+m}$	$\frac{25}{46}$	$\frac{21}{46}$
teoretické četnosti $n \cdot a_i'$	$21 \cdot \frac{25}{46} \doteq 11.41$	$21 \cdot \frac{21}{46} \doteq 9.59$
teoretické četnosti $m \cdot a_i'$	$25 \cdot \frac{25}{46} \doteq 13.59$	$25 \cdot \frac{21}{46} \doteq 11.41$

a testujeme hypotézu

$\mathbf{H}_0'$: veličiny X' a Y' mají stejná rozdělení

proti alternativní hypotéze:

$\mathbf{H}_A'$: veličiny X' a Y' nemají stejná rozdělení

Hodnota testovací statistiky je

$$t' = \sum_i \frac{(n_i' - n \cdot a_i')^2}{n \cdot a_i'} + \sum_i \frac{(m_i' - m \cdot a_i')^2}{m \cdot a_i'} =$$

$$= \frac{(13-11.41)^2}{11.41} + \frac{(8-9.59)^2}{9.59} + \frac{(12-13.59)^2}{13.59} + \frac{(13-11.41)^2}{11.41} \doteq 0.893$$

a hodnota kvantilu je

$$q_{\chi^2(2-1)}(1-\alpha) = q_{\chi^2(1)}(0.95) \doteq 3.84\ .$$

Protože

$$t = 0.893 \not> 3.84 \doteq q_{\chi^2(1)}(0.95)\ ,$$

nulovou hypotézu $\mathbf{H}_0'$ pro veličiny X' a Y' **NEZAMÍTÁME**. Tento výsledek interpretujeme tak, že hypotézu $\mathbf{H}_0$

X a Y mají stejná rozdělení,

rovněž **NEZAMÍTÁME**.

12.9 (test nezávislosti veličin)

Úspěšnost u zkoušek ve vztahu k počtu přítomných studentů udává tabulka:

termín	1.	2.	3.	4.	5.
počet přítomných	20	30	40	60	50
počet úspěšných	11	8	14	43	24

Otestujte na hladině významnosti 5 % hypotézu, že pravděpodobnost úspěchu byla u všech zkouškových termínů stejná.

Řešení:
Označme si veličiny

$X($ *"účast daného studenta na j-tém termínu"* $) :=$ *"zda uspěl/neuspěl"*

$Y($ *"účast daného studenta na j-tém termínu"* $) := \; j$

Pravděpodobnost úspěchu v j-tém termínu je dána jako $P(X = \textit{uspěl} \mid Y = j)$.

Ukážeme, že tato hodnota bude nezávislá na j, právě když veličiny X a Y budou nezávislé.

- -

Důkaz:
($\Leftarrow$): Z nezávislosti X a Y ihned máme, že $P(X = \textit{uspěl} \mid Y = j) = P(X = \textit{uspěl}\ ) = \textit{konst.}$
($\Rightarrow$): Naopak, nechť $c := P(X = \textit{uspěl} \mid Y = j) = \frac{P(X=\textit{uspěl}\ \cap\ Y=j)}{P(Y=j)}$ pro všechna j. Tedy

$$P(X = \textit{uspěl}\ \cap\ Y = j) = c \cdot P(Y = j)$$

a sečtením přes všechna j dostaneme, že

$$P(X = \textit{uspěl}\ ) = \sum_j P(X = \textit{uspěl}\ \cap\ Y = j) = \sum_j c \cdot P(Y = j) = c \cdot \underbrace{\sum_j P(Y = j)}_{=1} = c$$

neboli

$$P(X = \textit{uspěl}\ \cap\ Y = j) = P(X = \textit{uspěl}\ ) \cdot P(Y = j)\ .$$

Podobně z $P(X = \textit{neuspěl} \mid Y = j) = 1 - c$ pro všechna j odvodíme, že

$P(X = \textit{neuspěl}\ \cap\ Y = j) = P(X = \textit{neuspěl}\ ) \cdot P(Y = j)$. Tedy veličiny X a Y jsou nezávislé.

- -

Úloha tedy ve skutečnosti znamená, že budeme testovat hypotézu:

$\mathbf{H}_0$: rozdělení veličin X a Y jsou *nezávislá*

proti alternativní hypotéze:

$\mathbf{H}_1$: rozdělení veličin X a Y jsou *závislá.*

na hladině významnosti $\alpha = 5\%$.

Četnosti n_{ij} jednotlivých případů pro $X = i$ a $Y = j$ přepíšeme pomocí tabulky

$n_{i,j}$ $(Y =)\ j$ $(X =)\ i$	1	2	3	4	5	$n_{i,\bullet}$
uspěl	11	8	14	43	24	100
neuspěl	9	22	26	17	26	100
$n_{\bullet,j}$	20	30	40	60	50	200

kde

$$n_{\bullet,i} = \sum_j n_{i,j} \quad \text{a} \quad n_{j,\bullet} = \sum_i n_{i,j}$$

Za předpokladu $\mathbf{H}_0$ pak jako teoretické četnosti budeme opět chápat hodnoty $\frac{n_{i,\bullet} \cdot n_{\bullet,j}}{n}$ v této tabulce:

$\frac{n_{i,\bullet} \cdot n_{\bullet,j}}{n}$ $(Y =)\ j$ / $(X =)\ i$	1	2	3	4	5	$n_{i,\bullet}$
uspěl	$\frac{20 \cdot 100}{200} = 10$	$\frac{30 \cdot 100}{200} = 15$	20	30	25	100
neuspěl	$\frac{20 \cdot 100}{200} = 10$	15	20	30	25	100
$n_{\bullet,j}$	20	30	40	60	50	200

Podmínka na teoretické (tj. očekávané) četnosti ≥ 5 je splněna, takže položky nemusíme slučovat. Pro realizaci testovací statistiky dostaneme

$$t = \sum_{i,j} \frac{\left(n_{i,j} - \frac{n_{i,\bullet} \cdot n_{\bullet,j}}{n}\right)^2}{\frac{n_{i,\bullet} \cdot n_{\bullet,j}}{n}} =$$

$$= \frac{(11-10)^2}{10} + \frac{(9-10)^2}{10} + \frac{(8-15)^2}{15} + \frac{(22-15)^2}{15} + \frac{(14-20)^2}{20} + \frac{(26-20)^2}{20} +$$
$$+ \frac{(43-30)^2}{30} + \frac{(17-30)^2}{30} + \frac{(24-25)^2}{25} + \frac{(26-25)^2}{25} = 21.68$$

a porovnáme ji s hodnotou kvantilu χ^2 pro $(5-1) \cdot (2-1) = 4$ stupně volnosti

$$q_{\chi^2(4)}(1-\alpha) = q_{\chi^2(4)}(0.95) \doteq 9.49\ .$$

Protože

$$t \doteq 21.68 > 9.49 \doteq q_{\chi^2(4)}(0.95)\ ,$$

hypotézu o nezávislosti proto **ZAMÍTÁME**.