Matematika III – 14. přednáška
Bodové a intervalové odhady, testování hypotéz

Michal Bulant

Masarykova univerzita
Fakulta informatiky

18. 12. 2013

Obsah přednášky

Doporučené zdroje

- Jan Slovák, Martin Panák, Michal Bulant, *Matematika drsně a svižně*, MU Brno, 2013, 774 s. (též jako e-text).
- Karel Zvára, Josef Štěpán, **Pravděpodobnost a matematická statistika**, Matfyzpress, 4. vydání, 2006, 230 stran, ISBN 80-867-3271-1.
- Marie Budíková, Štěpán Mikoláš, Pavel Osecký, **Teorie pravděpodobnosti a matematická statistika (sbírka příkladů)**, Masarykova univerzita, 3. vydání, 2004, 117 stran, ISBN 80-210-3313-4.
- Marie Budíková, **Statistika**, Masarykova univerzita, 2004, distanční studijní opora ESF, `http://www.math.muni.cz/~budikova/esf/Statistika.zip`.
- Marie Budíková, Tomáš Lerch, Štěpán Mikoláš, **Základní statistické metody**, Masarykova univerzita, 2005, 170 stran, ISBN 80-210-3886-1.

Náhodný výběr

Definice

Náhodným výběrem rozsahu n rozumíme n-tici **nezávislých** a **stejně rozdělených** náhodných veličin $X_1, \dots, X_n \sim F_X(x)$ (někdy také hovoříme o n nezávislých kopiích náhodné veličiny X).
Náhodným výběrem rozsahu n **z** p**-rozměrného rozdělení** rozumíme n-tici **nezávislých** a **stejně rozdělených** p-rozměrných náhodných vektorů.

V matematické statistice často pracujeme s transformacemi náhodného výběru, takovým náhodným veličinám (příp. vektorům) říkáme **statistiky**. V následujícím zavedeme několik důležitých statistik a ukážeme jejich souvislost s číselnými charakteristikami náhodných veličin.

Základní statistiky

Definice

Nechť $X_1, \ldots, X_n$ je náhodný výběr. Statistiku

$$M = \frac{1}{n} \sum_{i=1}^{n} X_i$$

nazýváme **výběrový průměr**, statistiku

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - M)^2$$

výběrový rozptyl a statistiku $S = \sqrt{S^2}$ **výběrová směrodatná odchylka**. Analogicky se definují i výběrová kovariance, příp. výběrový korelační koeficient pro dvourozměrný náhodný výběr.

Vlastnosti statistik

Protože jsou uvedené statistiky náhodnými veličinami, lze se přirozeně ptát po jejich číselných charakteristikách.

Věta

Nechť $X_1, \ldots, X_n$ je náhodný výběr rozsahu n z rozdělení se střední hodnotou μ a rozptylem σ^2 . Pak platí:

- $E(M) = \mu$,
- $D(M) = \mathrm{var}(M) = \sigma^2/n$,
- $E(S^2) = \sigma^2$.

Důkaz.

Ukážeme jen (nejsložitější) 3. tvrzení.
Snadno se odvodí, že platí

$$\sum(X_i-\mu)^2=\sum(X_i-M)^2+n(M-\mu)^2.$$

Proto je

$$\begin{aligned}E(S^2)&=\frac{1}{n-1}E\Big(\sum(X_i-\mu)^2\Big)-\frac{n}{n-1}E(M-\mu)^2=\\&=\frac{1}{n-1}\sum D(X_i)-\frac{n}{n-1}D(M)=\\&=\frac{n}{n-1}\sigma^2-\frac{1}{n-1}\sigma^2=\sigma^2.\end{aligned}$$

$\square$

V předchozí větě jsme ukázali, že výběrový průměr M splňuje $E(M) = \mu$, jeho střední hodnota je tedy rovna odhadovanému parametru μ. V takovém případě říkáme, že statistika M je **nestranným odhadem** parametru μ.
Podobně jsme viděli, že S^2 je nestranným odhadem parametru σ^2.
Všimněme si rovněž, že „přirozeněji“ definovaná statistika $\frac{1}{n}\sum(X_i - M)^2$ není nestranným odhadem σ^2, její střední hodnota je totiž $\frac{n-1}{n}\sigma^2$.

Příklad

Rozmyslete si, je-li S nestranným odhadem směrodatné odchylky σ.

Náhodný výběr z normálního rozdělení

Uvažme nyní speciální případ, kdy je $X_1, \ldots, X_n$ náhodný výběr z normálního rozdělení $N(\mu, \sigma^2)$. Bez důkazu uvedeme velmi důležité tvrzení o rozdělení následujících statistik:

Věta

- *M a S^2 jsou nezávislé náhodné veličiny.*
- *$M \sim N(\mu, \sigma^2/n)$, a tedy $U = (M - \mu)/(\sigma/\sqrt{n}) \sim N(0, 1)$.*
- *$T = (M - \mu)/(S/\sqrt{n}) \sim t(n - 1)$.*
- *$K = (n - 1)S^2/\sigma^2 \sim \chi^2(n - 1)$.*
- *$\sum(X_i - \mu)^2/\sigma^2 \sim \chi^2(n)$.*

Poznámka

K odhadu μ, neznáme-li σ^2, slouží T, v opačném případě U.
K odhadu σ^2, neznáme-li μ, slouží K, v opačném případě následující (bezejmenná?) statistika, která je vlastně statistikou K, v níž místo odhadu M použijeme přímo μ.

Příklad

V roce 1951 bylo rozsáhlým statistickým průzkumem zjištěno, že střední hodnota výšky desetiletých chlapců je 136,1 cm se směrodatnou odchylkou $\sigma = 6{,}4$ cm.
V roce 1961 byla zjištěna výška pouze u 15 náhodně vybraných chlapců:

130	140	136	141	139	133	149	151
139	136	138	142	127	139	147	

Otázkou je, zda se v porovnání s rokem 1951 změnila střední výška chlapců, pokud předpokládáme, že variabilita výšek se v různých generacích příliš nemění.

Řešení

Vzhledem k tomu, že základní soubor všech desetiletých chlapců je rozsáhlý, lze zmíněná data považovat za náhodný výběr[a]. Zjistíme, že výběrový průměr $M = 139{,}133$, $n = 15$ a s využitím statistiky U dostáváme, že s 95% pravděpodobností leží hodnota μ v intervalu

$$(M - 1{,}96\sigma/\sqrt{n}; M + 1{,}96\sigma/\sqrt{n}) = (135{,}9; 142{,}4).$$

Protože i střední hodnota výšek z roku 1951 leží v tomto intervalu, nemáme vážný důvod tvrdit, že se střední výška změnila. Pokud bychom ovšem připustili vyšší možnost omylu a stanovili interval se spolehlivostí pouze 90%, pak bychom na této hladině hypotézu, že se střední výška změnila, „přijali“ – interval je nyní (136,41;141,85). Podobně, pokud nás zajímá pouze **dolní odhad** střední hodnoty výšek chlapců (a vůbec tedy nepřipouštíme možnost, že by se střední výška snížila), pak s 95% pravděpodobností je střední výška větší než 136,41, a tedy nyní opět „přijímáme“ hypotézu, že se střední výška zvýšila.

Příklady k procvičení

Příklad

Předpokládejme, že velká skupina studentů má ze zápočtové písemky ze statistiky bodové hodnoty normálně rozloženy kolem střední hodnoty 72 se směrodatnou odchylkou 9 bodů. Určete pravděpodobnost, že

a) náhodně vybraný student bude mít výsledek lepší než 80 bodů,

b) průměr výsledků náhodného výběru 10 studentů bude lepší než 80 bodů.

Příklad

Rychlost letadla byla určována v 5 zkouškách, jejichž aritmetický průměr byl $M = 870{,}3\,ms^{-1}$. Najděte 95% interval spolehlivosti pro μ, víte-li, že měření rychlosti se řídí normálním rozdělením se směrodatnou odchylkou $2{,}1\,ms^{-1}$.

Dva nezávislé výběry z normálního rozdělení

Věta

Nechť je $X_{11}, \dots, X_{m1}$ náhodný výběr rozsahu m z rozdělení $N(\mu, \sigma_1^2)$ a $X_{12}, \dots, X_{n2}$ je na něm nezávislý náhodný výběr rozsahu n z rozdělení $N(\mu, \sigma_2^2)$, přičemž $m, n \geq 2$. Označme M_1, M_2 jejich výběrové průměry a S_1^2, S_2^2 výběrové rozptyly. Dále nechť je

$$S_*^2 = \frac{(m-1)S_1^2 + (n-1)S_2^2}{m+n-2}$$

vážený průměr výběrových rozptylů. *Pak platí:*

- *$M_1 - M_2$ a S_*^2 jsou stochasticky nezávislé,*
- $M_1 - M_2 \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n})$,
- *je-li $\sigma_1^2 = \sigma_2^2 = \sigma^2$, pak* $K = (m+n-2)S_*^2/\sigma^2 \sim \chi^2(m+n-2)$,
- $F = \frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} \sim F(m-1, n-1)$.

Užití statistik dvou nezávislých výběrů

- Statistika U, vzniklá normováním $M_1 - M_2$, se používá pro odhad rozdílu $\mu_1 - \mu_2$, známe-li rozptyly σ_1^2, σ_2^2.
- Je-li $\sigma_1^2 = \sigma_2^2 = \sigma^2$, pak statistika T (vzniklá z U nahrazením teoretického společného rozptylu σ^2 váženým průměrem výběrových rozptylů S_*^2) slouží pro odhad rozdílu $\mu_1 - \mu_2$, neznáme-li rozptyl σ^2.
- Statistika $K = (m + n - 2)S_*^2/\sigma^2$ slouží k odhadu společného rozptylu σ^2.
- Statistika $F = \frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2}$ slouží k odhadu podílu rozptylů σ_1^2/σ_2^2.

Příklad

Mějme dva nezávislé náhodné výběry; první rozsahu 10 z rozdělení $N(2; 1{,}5)$ a druhý rozsahu 5 z rozdělení $N(3, 4)$. Určete pravděpodobnost, že výběrový průměr prvního výběru bude menší než výběrový průměr druhého výběru.

Řešení

$$P(M_1 < M_2) = P(M_1 - M_2 < 0) =$$

$$= P\left(\frac{(M_1 - M_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} < \frac{0 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}}\right) =$$

$$P\left(U < \frac{-2 + 3}{\sqrt{\frac{1{,}5}{10} + \frac{4}{5}}}\right) = P(U < 1{,}05) =$$

$$= \Phi(1{,}05) = 0{,}853.$$

Náhodný výběr je n-tice nezávislých náhodných veličin se stejným rozdělením, které záleží na jednom nebo více parametrech. Obvykle přitom hodnotu těchto parametrů neznáme, ale můžeme tuto hodnotu nebo hodnotu nějaké jeho funkce (tzv. parametrické funkce) z náhodného výběru odhadnout.

Definice

Mějme náhodný výběr $X_1, \dots, X_n$, které závisí na (obecně vektorovém) parametru θ. **Bodovým odhadem parametru θ** rozumíme statistiku $T(X_1, \dots, X_n)$, která je v nějakém smyslu blízko parametru θ. Rozdíl (příp. vektorový) $E(T) - \theta$ nazveme **vychýlení**, je-li $E(T) = \theta$, pak odhad T nazveme **nestranným**. **Intervalovým odhadem parametru θ** rozumíme (obecně vícerozměrný) interval (T_L, T_U), kde $T_L(X_1, \dots, X_n)$ a $T_U(X_1, \dots, X_n)$ jsou statistiky výběru $(X_1, \dots, X_n)$. Platí-li

$$P(T_L \leq \theta \leq T_U) = 1 - \alpha,$$

říkáme, že (T_L, T_U) je interval spolehlivosti $1 - \alpha$ pro parametr θ .

Definice

Jsou-li T_1, T_2 nestranné odhady parametru θ, říkáme, že odhad T_1 je **lepší** než odhad T_2, pokud $D(T_1) < D(T_2)$, příp. $\operatorname{var} T_1 < \operatorname{var} T_2$ (tj. matice $\operatorname{var} T_2 - \operatorname{var} T_1$ je pozitivně definitivní).
O posloupnosti T_n odhadů θ říkáme, že je **asymptoticky nestranná**, pokud $\lim_{n\to\infty} E(T_n) = \theta$.
O posloupnosti T_n odhadů θ říkáme, že je **konzistentní**, pokud $\lim_{n\to\infty} P(|T_n - \theta| < \epsilon) = 1$.

Příklad

Uvažujme opakovaná nezávislá měření určité konstanty μ, popsaná náhodným výběrem $X_1, \dots, X_n$ z rozdělení se střední hodnotou $E(X_i) = \mu$ a rozptylem $D(X_i) = \sigma^2$. Dokažte, že statistiky $M = \frac{1}{n}\sum_{i=1}^{n} X_i$ a $L = \frac{1}{2}(X_1 + X_n)$ jsou nestrannými odhady μ a rozhodněte, který z odhadů je lepší.

Řešení

$$
\begin{aligned}
E(M) &= \frac{1}{n}\sum E(X_i) = \frac{1}{n} n\mu = \mu \\
E(L) &= \frac{X_1 + X_n}{2} = \frac{1}{2}E(X_1 + X_n) = \frac{1}{2}(\mu + \mu) = \mu \\
D(M) &= \frac{1}{n^2}\sum D(X_i) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n} \\
D(L) &= D(1/2(X_1 + X_n)) = 1/4 D(X_1 + X_n) = \\
&= \tfrac{1}{4}(D(X_1) + D(X_n)) = \tfrac{\sigma^2}{2}.
\end{aligned}
$$

Poznámka

Odpověď na otázku, zda je výběrová směrodatná odchylka S nestranným odhadem směrodatné odchylky σ, je záporná. Kdyby totiž $E(S) = \sigma$, pak by $D(S) = E(S^2) - E(S)^2 = \sigma^2 - \sigma^2 = 0$, což by znamenalo, že S je konstanta, a to je spor, protože rozptyl S je nenulový.

Poznámka

Jak jsme viděli dříve, není statistika $s_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - M)^2$ náhodného výběru z normálního rozdělení nestranným odhadem rozptylu σ^2 – je totiž $E(s_n^2) = E(\frac{n-1}{n}S^2) = \frac{n-1}{n}\sigma^2$. Zřejmě je ale $\lim_{n\to\infty} E(s_n^2) = \sigma^2$ a protože

$$D(S^2) = \frac{2\sigma^4}{n-1},$$

je i $\lim_{n\to\infty} D(s_n^2) = \lim_{n\to\infty} D((n-1)S^2/n) = 0$, a je tedy posloupnost s_n^2 konzistentním odhadem rozptylu σ^2.

Intervaly spolehlivosti (*confidence intervals*)

Připomeňme, že pro náhodný výběr $X_1, \ldots, X_n$ závislý na parametru θ jsme definovali intervalový odhad parametru θ pomocí statistik T_L, T_U výběru tak, že $P(T_L \leq \theta \leq T_U) = 1 - \alpha$. Jde o tzv. **oboustranný interval spolehlivosti pro** θ . Podobně definujeme **levostranný interval spolehlivosti** (T_L, ∞) pomocí $P(T_L < \theta) = 1 - \alpha$, analogicky **pravostranný interval spolehlivosti**. Číslo α se nazývá riziko (obvykle se používá $\alpha = 0{,}05$), číslo $1 - \alpha$ spolehlivost.

Algoritmus konstrukce intervalu spolehlivosti

1. Zvolíme statistiku V, která je nestranným bodovým odhadem parametru θ.
2. Najdeme tzv. *pivotovou statistiku* W, která je transformací V se známým rozdělením, nezávisící na neznámé hodnotě θ (např. M, K, T, F).
3. Najdeme příslušné kvantily rozdělení statistiky W tak, že
$$P(w_{\alpha/2} \leq W \leq w_{1-\alpha/2}) = 1 - \alpha.$$
4. Nerovnost $w_{\alpha/2} \leq W \leq w_{1-\alpha/2}$ převedeme ekvivalentními úpravami na nerovnost $T_L \leq \theta \leq T_U$.
5. Z daného výběru zjistíme konkrétní číselné realizace statistik T_L, T_U a dostaneme tak intervalový odhad požadované spolehlivosti $1 - \alpha$.

Intervaly spolehlivosti pro parametry normálního rozdělení

μ (známe σ^2)	$\left(M - \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2}, M + \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2}\right)$
μ (neznáme σ^2)	$\left(M - \frac{S}{\sqrt{n}} t_{1-\alpha/2}(n-1), M + \frac{S}{\sqrt{n}} t_{1-\alpha/2}(n-1)\right)$
σ^2 (neznáme μ)	$\left(\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}, \frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}\right)$
σ^2 (známe μ)	$\left(\frac{\sum(X_i-\mu)^2}{\chi^2_{1-\alpha/2}(n)}, \frac{\sum(X_i-\mu)^2}{\chi^2_{\alpha/2}(n)}\right)$

Příklad

Nechť $X_1, \ldots, X_n$ je náhodný výběr z rozdělení $N(\mu; 0{,}1)$. Jaký musí být minimální rozsah výběru, aby velikost 95% intervalu spolehlivosti pro μ nepřesáhla číslo 0,03?

Řešení

Podle předchozí tabulky dostáváme (pro $\alpha = 0{,}05$)

$$0{,}03 \geq M + \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2} - \left(M - \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2}\right) = \\ = 2\frac{\sigma}{\sqrt{n}} u_{1-\alpha/2}.$$

Proto

$$n \geq \frac{4\sigma^2 u_{1-\alpha/2}^2}{0{,}03^2} \approx 1707{,}38$$

a rozsah výběru tedy musí splňovat $n \geq 1708$.

Intervaly spolehlivosti pro parametry 2 normálních rozdělení

$\mu_1 - \mu_2$ (známe σ_1^2, σ_2^2)	$M_1 - M_2 \pm \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}\, u_{1-\alpha/2}$
$\mu_1 - \mu_2$ (nezn. $\sigma_1^2 = \sigma_2^2$)	$M_1 - M_2 \pm S_* \sqrt{\frac{1}{m} + \frac{1}{n}}\, t_{1-\alpha/2}(m+n-2)$
společný rozptyl σ^2	$\left(\frac{(m+n-2)S_*^2}{\chi^2_{1-\alpha/2}(m+n-2)}, \frac{(m+n-2)S_*^2}{\chi^2_{\alpha/2}(m+n-2)} \right)$
podíl rozptylů σ_1^2/σ_2^2	$\left(\frac{S_1^2/S_2^2}{F_{1-\alpha/2}(m-1,n-1)}, \frac{S_1^2/S_2^2}{F_{\alpha/2}(m-1,n-1)} \right)$

Poznámka

Pokud a priori nevíme, jestli jsou rozptyly shodné, můžeme to ověřit tak, že nejprve sestrojíme interval spolehlivosti pro σ_1^2/σ_2^2. Obsahuje-li 1, lze (s pravděpodobností $1-\alpha$) považovat rozptyly za shodné a tento rozptyl odhadovat pomocí statistiky K, jak je uvedeno v tabulce.

Inteval spolehlivosti pro výběr z dvourozměrného rozdělení

Nechť $(X_1, Y_1), \ldots, (X_n, Y_n)$ je výběr z rozdělení

$$N_2\left(\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}, \begin{pmatrix}\sigma_1^2 & \sigma_{12}\\ \sigma_{12} & \sigma_2^2\end{pmatrix}\right).$$

Označíme $\mu = \mu_1 - \mu_2$ a zavedeme *rozdílový výběr* $Z_i = X_i - Y_i$. Pak statistika $T = \frac{M-\mu}{S/\sqrt{n}}$ výběru Z má t-rozdělení s $n-1$ stupni volnosti, proto jsou hranice intervalu spolehlivosti $1-\alpha$ pro μ rovny

$$M \pm \frac{S}{\sqrt{n}} t_{1-\alpha/2}(n-1).$$

Příklad

U šesti nových automobilů bylo testováno, nakolik se sjíždějí pneumatiky na předních kolech. Byly naměřeny tyto hodnoty (v mm):

číslo auta	1	2	3	4	5	6
sjetí pravé pneu	1,8	1,0	2,2	0,9	1,5	1,6
sjetí levé pneu	1,5	1,1	2,0	1,1	1,4	1,4

Předpokládejte, že jde o realizaci náhodného výběru z dvourozměrného normálního rozdělení a rozhodněte, jestli nedochází k výraznějšímu nesymetrickému sjíždění pneumatik (tj. sestrojte 95% interval spolehlivosti pro $\mu = \mu_1 - \mu_2$).

Řešení

Postupně vypočteme: $Z = (0{,}3; -0{,}1; 0{,}2; -0{,}2; 0{,}1; 0{,}2)$, $M = 0{,}0833$, $S = 0{,}1941$. Pak jsou krajními body hledaného 95% intervalu spolehlivosti $M \pm \frac{S}{\sqrt{n}} t_{1-\alpha/2}(n-1) = 0{,}0833 \pm 0{,}1941 \cdot 2{,}5706/\sqrt{6}$, tj. $(-0{,}12; 0{,}29)$.

Poznamenejme, že snadno odvodíme i míru rizika, se kterou bychom mohli tvrdit, že je $\mu_1 > \mu_2$, tj. že pravé pneumatiky se sjíždějí více než levé. Je to takové číslo α, aby příslušný interval spolehlivosti neobsahoval číslo 0 – v našem případě je $\alpha = 0{,}34$, což je riziko příliš vysoké.

Motivační úvod

Testování hypotéz umožňuje na základě náhodného výběru s danou pravděpodobností ověřovat domněnky o rozdělení, z něhož pochází daný náhodný výběr. **Hypotézou** budeme rozumět nějaké tvrzení o parametrech tohoto rozdělení.

Definice

H_0 ... nulová hypotéza (např. $\theta = c$, kde c vyjadřuje naši domněnku o hodnotě parametru θ)
H_1 ... (oboustranná) alternativní hypotéza (obvykle negace nulové)
Testováním H_0 oproti alternativní hypotéze rozumíme postup založený na náhodném výběru, s jehož pomocí platnost H_0 *zamítneme* nebo *nezamítneme* (= připouštíme).
Chyba 1. druhu ... H_0 platí a my ji zamítneme (závažnější)
Chyba 2. druhu ... H_0 neplatí a my ji nezamítneme
Pravděpodobnost chyby 1. druhu se nazývá *hladina významnosti* (α, obvykle $\alpha = 0,05$), pravděpodobnost chyby 2. druhu se značí β a číslo $1 - \beta$ se nazývá *síla testu*.

Způsoby testování nulové hypotézy

1. pomocí intervalu spolehlivosti
2. pomocí kritického oboru
3. pomocí tzv. $p-$hodnoty (p-value)

Interval spolehlivosti Na základě realizace náhodného výběru sestrojíme $100(1-\alpha)\%$ interval spolehlivosti pro neznámý parametr θ a zjistíme, zda c patří do tohoto intervalu. Pokud ano, hypotézu H_0 nezamítáme (v opačném případě zamítáme) na hladině významnosti α.

Kritický obor Stanovení kritického oboru je postup do jisté míry obrácený. Nejprve (i bez náhodného výběru) zvolíme vhodnou statistiku T a množinu hodnot, jichž může T nabývat, rozdělíme na dvě disjunktní podmnožiny: obor nezamítnutí H_0 (značíme V) a **kritický obor** W (obor zamítnutí H_0). Pokud realizace T padne do W, pak H_0 zamítneme, jinak nezamítáme.

Stanovení kritického oboru na hladině α

Pro statistiku T (*testové kritérium*) stanovíme obor nezamítnutí V jako interval, jehož hraniční body tvoří kvantil $\alpha/2$ a $1-\alpha/2$, odtud je

$$W = (-\infty, F^{-1}(\alpha/2)) \cup (F^{-1}(1-\alpha/2), \infty).$$

Způsoby testování nulové hypotézy

p-hodnota Testování pomocí p-hodnoty je jednoduchý test, umožněný rozšířením statistických balíků. p-hodnota udává nejnižší možnou hladinu významnosti, při níž H_0 zamítáme. Je-li p-hodnota $> \alpha$, hypotézu H_0 nezamítáme, pro p-hodnotu menší než α, hypotézu zamítneme.

p-hodnota se stanoví rovněž se znalostí konkrétní realizace t_0 statistiky T náhodného výběru jako

$$p = 2 \min\{P(T \leq t_0), P(T \geq t_0)\}.$$

Testování hypotézy proti jednostranné alternativě

Je-li H_0 hypotéza $\theta = c$, pak *levostranná alternativní hypotéza* je tvrzení $\theta < c$, *pravostranná alternativní hypotéza* je tvrzení $\theta > c$. Volba typu alternativní hypotézy vyplývá z konkretní situace.

Příklad

- V předmětu Matematika 3 psali studenti písemku rozdělení na 2 skupiny. Hypotéza H_0 : *obě zadání mají stejnou průměrnou obtížnost* je testována oproti oboustranné alternativní hypotéze *zadání nejsou stejně obtížná*.
- V předmětu Matematika 3 se dříve po studentech nevyžadovalo řešení domácích úloh. Toto bylo nyní nově zavedeno s cílem dosažení lepších výsledků studentů u závěrečné zkoušky.
  V tomto případě zřejmě použijeme nulovou hypotézu H_0 : *výsledné bodové hodnocení se nezlepšilo* oproti pravostranné alternativní hypotéze H_1 : *bodový výsledek studentů se zlepšil*

Jednoduchý příklad

Příklad

Náš protivník hodil 60x kostkou a padla mu 16x šestka. Testujme na hladině významnosti $\alpha = 0{,}05$ nulovou hypotézu H_0 : *kostka není upravená* oproti jednostranné alternativní hypotéze H_1 : *kostka je upravená tak, aby padalo více šestek*.

Řešení

Statistika T (*počet šestek*) ma rozdělení $T \sim Bi(60, 1/6)$. Kritický obor je dán 95. percentilem tohoto rozdělení. Snadno vypočteme, že $P(T > 14) = 0{,}065$ a $P(T > 15) = 0{,}034$, proto je $p-$hodnota rovna 0,034 (nebo jinými slovy: kritickým oborem na hladině 0,05 je interval $\langle 16, \infty)$. Hypotézu H_0 tedy zamítáme – na hladině 0,05 můžeme tvrdit, že kostka je upravená.

Jednoduchý příklad – pokr.

Řešení (pomocí aproximace)

Porovnejme předchozí řešení příkladu s řešením, při kterém využijeme aproximaci pomocí de Moive-Laplaceovy věty. Náhodnou veličinu

$$X = \frac{T - 10}{\sqrt{50/6}}$$

lze považovat za veličinu mající normální rozdělení $N(\mu, \sigma^2)$ s jednotkovým rozptylem $\sigma^2 = 1$, testovat budeme hypotézu $\mu = 0$. **Kritickým oborem** $N(0, 1)$ je interval $(1{,}65, \infty)$ (stále uvažujeme *pravostranou alternativu*). Přitom pro realizaci statistiky X platí $x = (16 - 10)/\sqrt{50/6} \approx 2{,}08$ a hypotézu tedy opět zamítáme. **Jednostranným intervalem spolehlivosti** pro X je $((2{,}08 - 1{,}65)/\sqrt{60}, \infty)$ a protože do něj nepatří hodnota 0 zamítáme nulovou hypotézu (všimněte si, že v obou případech *rozhodlo* porovnání $1{,}65 < 2{,}08$).

Jednoduchý příklad – pokr.

Řešení (pomocí aproximace a p-hodnoty)

Určeme nejmenší pravděpodobnost p, při níž stále ještě zamítáme nulovou hypotézu $\mu = 0$ oproti pravostranné hypotéze $\mu > 0$ (tj. p-hodnotu). Má-li X rozdělení $N(0, 1)$, pak
$p = P(X \geq 2{,}08) = 1 - 0{,}981 = 0{,}019$.
Protože je $\alpha = 0{,}05 > 0{,}019$, opět hypotézu zamítáme.

Základní testy hypotéz o parametrech normálního rozdělení

Podobně jako statistiky při konstrukci intervalů spolehlivosti jsou i základní testy standardizované (není divu – jak jsme viděli, jde o úzce propojené pojmy).

z-test Nechť je $X_1, \dots, X_n$ náhodný výběr z rozdělení $N(\mu, \sigma^2)$ se známým σ^2 a $n \geq 2$. Test $H_0 : \mu = c$ proti alternativní hypotéze $\mu \neq c$ se nazývá **z-test**.

jednovýběrový t-test Nechť je $X_1, \dots, X_n$ náhodný výběr z rozdělení $N(\mu, \sigma^2)$ s neznámým σ^2 a $n \geq 2$. Test $H_0 : \mu = c$ proti alternativní hypotéze $\mu \neq c$ se nazývá **jednovýběrový t-test**.

dvouvýběrový t-test Nechť je $X_{11}, \dots, X_{m1}$ náhodný výběr z rozdělení $N(\mu_1, \sigma^2)$ a $X_{12}, \dots, X_{n2}$ na něm nezávislý náhodný výběr z rozdělení $N(\mu_2, \sigma^2)$ s $m, n \geq 2$ a neznámým σ^2. Test $H_0 : \mu_1 - \mu_2 = c$ proti $H_1 : \mu_1 - \mu_2 \neq c$ se nazývá **dvouvýběrový t-test**.

Základní testy hypotéz o parametrech normálního rozdělení

párový t-test Nechť je $(X_1, Y_1)^T, \dots, (X_n, Y_n)$ výběr z rozdělení

$$N\left(\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}, \begin{pmatrix}\sigma_1^2 & \sigma_{12}\\ \sigma_{12}^2 & \sigma_2^2\end{pmatrix}\right)$$

s $n \geq 2$ a neznámými parametry. Test $H_0 : \mu_1 - \mu_2 = c$ oproti $H_1 : \mu_1 - \mu_2 \neq c$ se nazývá **párový t-test**.

F-test Nechť je $X_{11}, \dots, X_{m1}$ náhodný výběr z rozdělení $N(\mu_1, \sigma_1^2)$ a $X_{12}, \dots, X_{n2}$ na něm nezávislý náhodný výběr z rozdělení $N(\mu_2, \sigma_2^2)$ s $m, n \geq 2$. Test $H_0 : \sigma_1^2/\sigma_2^2 = 1$ proti $H_1 : \sigma_1^2/\sigma_2^2 \neq 1$ se nazývá **F-test**.

test rozptylu Nechť je $X_1, \dots, X_n$ náhodný výběr z $N(\mu, \sigma^2)$ s neznámým μ a $n \geq 2$. Test $H_0 : \sigma^2 = c$ proti $H_1 : \sigma^2 \neq c$ se nazývá **test o rozptylu**.

Kritický obor testů normálního rozdělení

z-test $|(M-c)/(\sigma/\sqrt{n})| \geq u_{1-\alpha/2}$

jednovýběrový t-test $|(M-c)/(S/\sqrt{n})| \geq t_{1-\alpha/2}(n-1)$

dvouvýběrový t-test

$$\left| \frac{M_1 - M_2 - c}{S_* \sqrt{\frac{1}{m} + \frac{1}{n}}} \right| \geq t_{1-\alpha/2}(m+n-2)$$

párový t-test sestrojením rozdílu $Z_i = X_i - Y_i$ a $\mu = \mu_1 - \mu_2$ úlohu předvedeme na jednovýběrový $t-$ test

F-test $S_1^2/S_2^2 \leq F_{\alpha/2}(m-1, n-1)$ nebo $S_1^2/S_2^2 \geq F_{1-\alpha/2}(m-1, n-1)$

test rozptylu $(n-1)S^2/c \leq \chi^2_{\alpha/2}(n-1)$ nebo $(n-1)S^2/c \geq \chi^2_{1-\alpha/2}(n-1)$

Příklad

Aktivní studenti chtěli dopravnímu podniku dokázat, že autobusy trpí většími výkyvy příjezdových dob na danou zastávku než tramvaje a provedli měření odchylek od jízdního řádu:

autobus	0	2	4	-3	2	-4	-3	0	0	5
tramvaj	4	6	3	0	-2	2	0	1	1	0

Z tabulky lze snadno vypočítat, že $S_1^2 = 9{,}12$ a $S_2^2 = 5{,}39$.

1. Na hladině 0,05 testujte nulovou hypotézu, že autobus i tramvaj jsou stejně spolehlivé oproti alternativní hypotéze, že tramvaj je spolehlivější.
2. Určete maximální pravděpodobnost, s níž můžete tvrdit, že je tramvaj spolehlivější než autobus.

Příklad

Ve dvou nádržích se zkoumal obsah chlóru. Z první bylo odebráno 22 vzorků, z druhé 10 vzorků. Byly vypočteny následující hodnoty výběrových průměrů a rozptylů: $M_1 = 34{,}23$, $M_2 = 35{,}73$, $S_1^2 = 1{,}76$, $S_2^2 = 1{,}81$. Hodnoty zjištěné z odebraných vzorků považujeme za realizace dvou nezávislých náhodných výběrů z rozdělení $N(\mu_1, \sigma^2)$, resp. $N(\mu_2, \sigma^2)$. Sestrojte 95% interval spolehlivosti pro rozdíl středních hodnot $\mu_1 - \mu_2$ a vyslovte závěr na dané hladině spolehlivosti o podstatnosti rozdílu naměřených hodnot.

Řešení

Dosadíme do vztahu $M_1 - M_2 \pm S_* \sqrt{\frac{1}{m} + \frac{1}{n}} \cdot t_{1-\alpha/2}(m + n - 2)$ hodnoty $M_1 - M_2 = -1{,}5$, $S_* = 1{,}3323$ a dostaneme interval $(-2{,}5377; -0{,}4623)$. Do tohoto intervalu 0 nepatří, proto je rozdíl $\mu_1 - \mu_2$ statisticky významně různý od nuly.

Komplexní příklad na dvouvýběrový t-test

Příklad

Uvažme bodové výsledky studentů z 2. termínu zkoušky předmětu `MB103` v roce 2008, přičemž výsledky testů skupiny A a skupiny B považujeme za dva nezávislé výběry z normálního rozdělení. Úkolem je zjistit, jestli výsledky některé ze skupin byly statisticky významně horší. Testujme nulovou hypotézu $H_0 : \mu_1 - \mu_2 = 0$ oproti alternativní hypotéze $H_1 : \mu_1 \neq \mu_2$.

Řešení

Nejprve pomocí F-testu otestujeme hypotézu o stejných rozptylech, v případě úspěchu poté použijeme dvouvýběrový t-test. Vypočteme základní statistiky:

	rozsah	výb. průměr	výb. rozptyl
A	65	10,48	22,49
B	64	7,21	29,75

Řešení (Komplexní příklad na dvouvýběrový t-test (pokr.))

Dostáváme $S_1^2/S_2^2 = 0{,}76$ a protože $F(0{,}025; 64; 63) = 0{,}61$, **nezamítáme** hypotézu o rovnosti rozptylů. O tomtéž se přesvědčíme i vypočtením intervalu spolehlivosti

$$\left(\frac{S_1^2/S_2^2}{F_{1-\alpha/2}(m-1, n-1)}, \frac{S_1^2/S_2^2}{F_{\alpha/2}(m-1, n-1)}\right) \approx (0{,}46; 1{,}24),$$

v němž leží testovaný podíl rozptylů 1.
Budeme tedy dále s výběry pracovat s předpokladem, že mají stejný rozptyl a použijeme dvouvýběrový t-test.

Řešení (Komplexní příklad na dvouvýběrový t-test (pokr.))

Vypočteme vážený průměr výběrových rozptylů

$$S_*^2 = \frac{(m-1)S_1^2 + (n-1)S_2^2}{m+n-2} \approx 5{,}11^2,$$

dále $M_1 - M_2 = 3{,}27$. V tabulkách najdeme hodnotu $t_{0{,}975}(65+64-2) = 1{,}98$, a protože

$$T = \frac{M_1 - M_2}{S_* \sqrt{\frac{1}{65} + \frac{1}{64}}} \approx 3{,}64,$$

docházíme k závěru, že můžeme hypotézu o stejné střední hodnotě obou rozdělení (tj. hypotézu $\mu_1 = \mu_2$) **zamítnout** (neboť $3{,}64 > 1{,}98$). Toto opět ověříme výpočtem intervalu spolehlivosti, který má střed v $M_1 - M_2$ a velikost rovnou dvojnásobku $S_* \sqrt{\frac{1}{m} + \frac{1}{n}} \cdot t_{1-\alpha/2}(m+n-2) \approx 1{,}78$, proto je interval spolehlivosti roven $(1{,}49; 5{,}05)$.