

Téma 7.: Úvod do testování hypotéz, ověřování normality

Základní poznatky o testování hypotéz

Předpokládáme, že testujeme nulovou hypotézu $H_0: h(\vartheta) = c$, kde $c \in \mathbb{R}$ buď proti oboustranné alternativě $H_1: h(\vartheta) \neq c$ nebo proti levostranné alternativě $H_1: h(\vartheta) < c$ nebo proti pravostranné alternativě $H_1: h(\vartheta) > c$.

Testování pomocí kritického oboru

Najdeme testovou statistiku $T_0 = T_0(X_1, \dots, X_n)$. Množina všech hodnot, jichž může testová statistika nabýt, se rozpadá na obor nezamítnutí nulové hypotézy (značí se V) a obor zamítnutí nulové hypotézy (značí se W a nazývá se též kritický obor). W a V jsou odděleny kritickými hodnotami (pro danou hladinu významnosti α je lze najít ve statistických tabulkách).

Jestliže číselná realizace t_0 testové statistiky T_0 padne do kritického oboru W , pak nulovou hypotézu zamítáme na hladině významnosti α a znamená to skutečné vyvrácení testované hypotézy. Jestliže t_0 padne do oboru nezamítnutí V , pak jde o pouhé mlčení, které platnost nulové hypotézy jenom připouští.

Stanovení kritického oboru pro danou hladinu významnosti α :

Označme t_{\min} (resp. t_{\max}) nejmenší (resp. největší) hodnotu testového kritéria.

Kritický obor v případě oboustranné alternativy má tvar

$W = (t_{\min}, K_{\alpha/2}(T)) \cup (K_{1-\alpha/2}(T), t_{\max})$, kde $K_{\alpha/2}(T)$ a $K_{1-\alpha/2}(T)$ jsou kvantily rozložení, jímž se řídí testové kritérium T_0 , je-li nulová hypotéza pravdivá.

Kritický obor v případě levostranné alternativy má tvar:

$$W = (t_{\min}, K_{\alpha}(T)).$$

Kritický obor v případě pravostranné alternativy má tvar:

$$W = (K_{1-\alpha}(T), t_{\max}).$$

Testování pomocí intervalu spolehlivosti

Sestrojíme $100(1-\alpha)\%$ empirický interval spolehlivosti pro parametrickou funkci $h(\vartheta)$.

Pokryje-li tento interval hodnotu c , pak H_0 nezamítáme na hladině významnosti α , v opačném případě H_0 zamítáme na hladině významnosti α .

Pro test H_0 proti oboustranné alternativě sestrojíme oboustranný interval spolehlivosti.

Pro test H_0 proti levostranné alternativě sestrojíme pravostranný interval spolehlivosti.

Pro test H_0 proti pravostranné alternativě sestrojíme levostranný interval spolehlivosti.

Testování pomocí p-hodnoty

p-hodnota udává nejnižší možnou hladinu významnosti pro zamítnutí nulové hypotézy:

je-li $p \leq \alpha$, pak H_0 zamítáme na hladině významnosti α , je-li $p > \alpha$, pak H_0 nezamítáme na hladině významnosti α .

Způsob výpočtu p-hodnoty:

Pro oboustrannou alternativu $p = 2 \min\{P(T_0 \leq t_0), P(T_0 \geq t_0)\}$.

Pro levostrannou alternativu $p = P(T_0 \leq t_0)$.

Pro pravostrannou alternativu $p = P(T_0 \geq t_0)$.

Příklad 1.: Víme, že výška hochů ve věku 9,5 až 10 let má normální rozložení s neznámou střední hodnotou μ a známým rozptylem $\sigma^2 = 39,112 \text{ cm}^2$. Dětský lékař náhodně vybral 15 hochů uvedeného věku, změřil je a vypočítal realizaci výběrového průměru $m = 139,13 \text{ cm}$. Podle jeho názoru by výška hochů v tomto věku neměla přesáhnout 142 cm s pravděpodobností 0,95. Lze tvrzení lékaře akceptovat?

Řešení: Testujeme $H_0: \mu = 142$ proti $H_1: \mu < 142$ na hladině významnosti 0,05.

a) Test provedeme pomocí kritického oboru.

Pro úlohy o střední hodnotě normálního rozložení při známém rozptylu používáme pivotovou statistiku $U = \frac{M - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$. Testová statistika tedy bude $T_0 = \frac{M - c}{\frac{\sigma}{\sqrt{n}}}$ a bude mít rozložení

$N(0, 1)$, pokud je nulová hypotéza pravdivá. Vypočítáme realizaci testového kritéria:

$$t_0 = \frac{139,13 - 142}{\frac{\sqrt{39,112}}{\sqrt{15}}} = -1,7773.$$

Stanovíme kritický obor: $W = (-\infty, u_\alpha) = (-\infty, u_{0,05}) = (-\infty, -u_{0,95}) = (-\infty, -1,6449)$.

Protože $-1,7773 \in W$, H_0 zamítáme na hladině významnosti 0,05. Tvrzení lékaře lze tedy akceptovat s rizikem omylu 5 %.

Výpočet pomocí systému STATISTICA:

Otevřeme nový datový soubor o dvou proměnných a jednom případě. První proměnná nazvaná t0 bude sloužit pro výpočet hodnoty testové statistiky. Do jejího Dlouhého jména napíšeme:

$$=(139,13-142)/\text{sqrt}(39,112/15)$$

Druhá proměnná nazvaná kvantil bude obsahovat 5. percentil rozložení $N(0,1)$. Do jejího Dlouhého jména napíšeme:

$$=\text{VNormal}(0,05;0;1)$$

	1	2
	t0	kvantil
1	-1,77735	-1,64485

Protože číslo -1,77735 patří do kritického oboru $(-\infty, -1,64485)$, H_0 zamítáme na hladině významnosti 0,05.

b) Test provedeme pomocí intervalu spolehlivosti.

Meze $100(1-\alpha)\%$ empirického pravostranného intervalu spolehlivosti pro střední hodnotu μ

při známém rozptylu σ^2 jsou: $(-\infty, h) = (-\infty, m + \frac{\sigma}{\sqrt{n}} u_{1-\alpha})$.

$$\text{V našem případě dostáváme: } h = 139,13 + \frac{\sqrt{39,112}}{\sqrt{15}} u_{0,95} = 139,13 + \frac{\sqrt{39,112}}{\sqrt{15}} 1,645 = 141,79.$$

Protože $142 \notin (-\infty; 141,79)$, H_0 zamítáme na hladině významnosti 0,05.

Výpočet pomocí systému STATISTICA:

Otevřeme nový datový soubor o jedné proměnné a jednom případě. Tato proměnná nazvaná h bude obsahovat horní mez 95% empirického pravostranného intervalu spolehlivosti pro μ ,

Do Dlouhého jména proměnné h napíšeme

$$=139,13+\text{sqrt}(39,112/15)*\text{VNormal}(0,95;0;1)$$

	1
	h
1	141,7861

Protože $142 \notin (-\infty; 141,7861)$, H_0 zamítáme na hladině významnosti 0,05.

c) Test provedeme pomocí p-hodnoty

$$p = P(T_0 \leq t_0) = \Phi(-1,7773) = 0,0378$$

Jelikož $0,0378 \leq 0,05$, nulovou hypotézu zamítáme na hladině významnosti 0,05.

Výpočet pomocí systému STATISTICA:

Otevřeme nový datový soubor o jedné proměnné a jednom případě. Tato proměnná nazvaná p bude obsahovat p-hodnotu. Do jejího Dlouhého jména napíšeme:

=INormal(-1,77735;0;1)

	1
	p
1	0,037755

Protože $0,037755 \leq 0,05$, H_0 zamítáme na hladině významnosti 0,05.

Příklad k samostatnému řešení: Necht' X_1, \dots, X_n je náhodný výběr z rozložení $N(\mu; 0,01)$. Realizace výběrového průměru je $m = 3$. Na hladině významnosti 0,05 testujte hypotézu $H_0: \mu = 2,93$ proti $H_1: \mu \neq 2,93$ všemi třemi metodami, je-li a) $n = 4$, b) $n = 9$, c) $n = 16$.

Výsledek:

ad a)

Test pomocí kritického oboru:

Protože realizace testové statistiky 1,4 nepatří do kritického oboru $(-\infty; -1,96) \cup (1,96; \infty)$, H_0 nezamítáme na hladině významnosti 0,05.

Test pomocí intervalu spolehlivosti:

Protože konstanta 2,93 patří do intervalu (2,902; 3,098), H_0 nezamítáme na hladině významnosti 0,05.

Test pomocí p-hodnoty:

Protože $0,161513 > 0,05$, H_0 nezamítáme na hladině významnosti 0,05.

ad b)

Test pomocí kritického oboru:

Protože realizace testové statistiky 2,1 patří do kritického oboru $(-1,96, 1,96)$, H_0 zamítáme na hladině významnosti 0,05.

Test pomocí intervalu spolehlivosti:

Protože konstanta 2,93 nepatří do intervalu (2,935; 3,065), H_0 zamítáme na hladině významnosti 0,05.

Test pomocí p-hodnoty:

Protože $0,035729 \leq 0,05$, H_0 zamítáme na hladině významnosti 0,05.

ad c)

Test pomocí kritického oboru:

Protože realizace testové statistiky 2,8 patří do kritického oboru $(-1,96, 1,96)$, H_0 zamítáme na hladině významnosti 0,05.

Test pomocí intervalu spolehlivosti:

Protože konstanta 2,93 nepatří do intervalu (2,951; 3,049), H_0 zamítáme na hladině významnosti 0,05.

Test pomocí p-hodnoty:

Protože $0,00511 \leq 0,05$, H_0 zamítáme na hladině významnosti 0,05.

Vidíme, že s rostoucím rozsahem výběru se snáz odhalí rozdíl mezi realizací výběrového průměru a předpokládanou střední hodnotou μ .

Grafické ověřování normality

Normální pravděpodobnostní graf (NP-plot)

NP-plot umožňuje graficky posoudit, zda data pocházejí z normálního rozložení.

Způsob konstrukce:

na vodorovnou osu vynášíme uspořádané hodnoty $x_{(1)} \leq \dots \leq x_{(n)}$,

na svislou osu vynášíme kvantily u_{α_j} , kde $\alpha_j = \frac{3j-1}{3n+1}$,

příčemž j je pořadí j -té uspořádané hodnoty (jsou-li některé hodnoty stejné, pak za j bereme průměrné pořadí odpovídající takové skupince).

Pocházejí-li data z normálního rozložení, pak všechny dvojice $(x_{(j)}, u_{\alpha_j})$ budou ležet na přímce.

Kvantil-kvantilový graf (Q-Q plot)

Umožňuje graficky posoudit, zda data pocházejí z nějakého známého rozložení (např. systém STATISTICA nabízí 8 typů rozložení: beta, exponenciální, Gumbelovo, gamma, log-normální, normální, Rayleighovo a Weibulovo). Pro nás je nejdůležitější právě normální rozložení.

Způsob konstrukce:

na svislou osu vynášíme uspořádané hodnoty $x_{(1)} \leq \dots \leq x_{(n)}$,

na vodorovnou osu kvantily $K_{\alpha_j}(X)$ vybraného rozložení, kde $\alpha_j = \frac{j-r_{adj}}{n+n_{adj}}$,

příčemž r_{adj} a n_{adj} jsou korigující faktory $\leq 0,5$, implicitně $r_{adj} = 0,375$ a $n_{adj} = 0,25$.

(Jsou-li některé hodnoty $x_{(1)} \leq \dots \leq x_{(n)}$ stejné, pak za j bereme průměrné pořadí odpovídající takové skupince.)

Pokud vybrané rozložení závisí na nějakých parametrech, pak se tyto parametry odhadnou z dat nebo je může zadat uživatel. Body $(K_{\alpha_j}(X), x_{(j)})$ se metodou nejmenších čtverců proloží přímkou. Čím méně se body odchylují od této přímky, tím je lepší soulad mezi empirickým a teoretickým rozložením.

Příklad 2.: Při nanášení tenkých kovových vrstev stříbra na polymerní materiál se vyžaduje, aby tloušťka vrstvy byla $0,020 \mu\text{m}$. Pomocí atomové absorpční spektroskopie se zjistily hodnoty, jež jsou uvedeny v tabulce a uloženy v souboru vrstva_stribru.sta. Posuďte N-P grafem a Q-Q grafem, zda výsledky měření se řídí normálním rozložením.

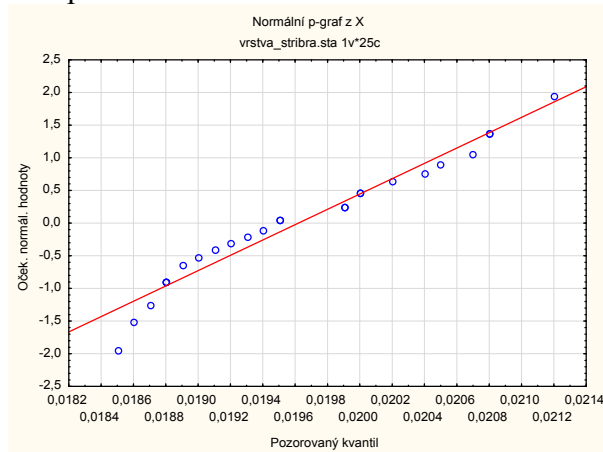
tloušťka vrstvy				
0,0212	0,0186	0,0192	0,0207	0,0200
0,0200	0,0190	0,0188	0,0208	0,0194
0,0188	0,0193	0,0204	0,0185	0,0187
0,0195	0,0191	0,0195	0,0199	0,0205
0,0189	0,0188	0,0199	0,0202	0,0208

Výpočet pomocí systému STATISTICA:

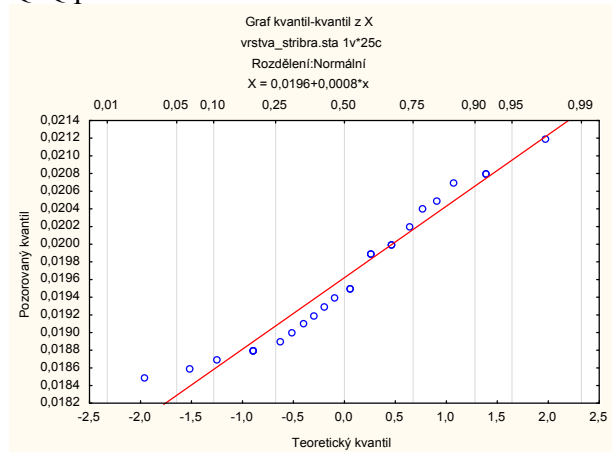
N-P plot: Grafy – 2D Grafy – Normální pravděpodobnostní grafy – Proměnná X – OK - odškrtneme Neurčovat průměrnou pozici svázaných pozorování - OK.

Q-Q plot: Grafy – 2D Grafy – Grafy typu Q-Q – Proměnná X – OK - odškrtneme Neurčovat průměrnou pozici svázaných pozorování - OK.

N-P plot



Q-Q plot



Dle vzhledu obou diagramů lze soudit, že data vykazují jen lehké odchylky od normality.

Testy normality

a) Kolmogorovův – Smirnovův test normality dat

Testujeme hypotézu, která tvrdí, že náhodný výběr X_1, \dots, X_n pochází z normálního rozložení s parametry μ a σ^2 . Distribuční funkci tohoto rozložení označme $\Phi_T(x)$. Nechť $F_n(x)$ je výběrová distribuční funkce. Testovou statistikou je statistika $D_n = \sup_{-\infty < x < \infty} |F_n(x) - \Phi_T(x)|$.

Nulovou hypotézu zamítáme na hladině významnosti α , když $D_n \geq D_n(\alpha)$, kde $D_n(\alpha)$ je tabelovaná kritická hodnota. Pro $n \geq 30$ lze $D_n(\alpha)$ aproximovat výrazem $\sqrt{\frac{1}{2n} \ln \frac{2}{\alpha}}$.

V případě, že neznáme parametry μ a σ^2 normálního rozložení, musíme je odhadnout z dat (střední hodnotu odhadneme pomocí m a rozptyl pomocí s^2). Tím se změní rozložení testové statistiky D_n . Příslušné modifikované kvantily byly určeny pomocí simulačních studií. V této situaci používáme Lilieforsovu variantu Kolmogorovova – Smirnovova testu.

b) Shapiroův – Wilkův test normality dat

Testujeme hypotézu, která tvrdí, že náhodný výběr X_1, \dots, X_n pochází z rozložení $N(\mu, \sigma^2)$. Testová statistika má tvar:

$$W = \frac{\sum_{i=1}^m a_i^{(n)} [X_{(n-i+1)} - X_{(i)}]^2}{\sum_{i=1}^m (X_i - M)^2}, \text{ kde } m = n/2 \text{ pro } n \text{ sudé a } m = (n-1)/2 \text{ pro } n \text{ liché. Koeficienty}$$

$a_i^{(n)}$ jsou tabelovány.

Na testovou statistiku W lze pohlížet jako na korelační koeficient mezi uspořádanými pozorováními a jim odpovídajícími kvantily standardizovaného normálního rozložení.

V případě, že data vykazují perfektní shodu s normálním rozložením, bude mít W hodnotu 1. Hypotézu o normalitě tedy zamítneme na hladině významnosti α , když se na této hladině neprokáže korelace mezi daty a jim odpovídajícími kvantily rozložení $N(0,1)$. Lze také říci, že $S - W$ test je založen na zjištění, zda body v Q-Q grafu jsou významně odlišné od regresní přímky proložené těmito body.

Příklad 3.: Pro data z příkladu 2 proveďte na hladině významnosti 0,05 Lilieforsovu variantu K-S testu a Shapirův – Wilkův test.

Výpočet pomocí systému STATISTICA:

Statistiky – Základní statistiky/tabulky – Tabulky četností – OK, Proměnné X – OK. Na záložce zvolíme Normalita a zaškrtneme Lilieforsův test a Shapiro – Wilksův W test – Testy normality.

Proměnná	Testy normality (vrstva_stribra.sta)				
	N	max D	Lilliefors p	W	p
X	25	0,119734	p > .20	0,939785	0,146346

Vidíme, že testová statistika K-S testu je $d = 0,119734$, odpovídající Lilieforsova p-hodnota je větší než 0,2, tedy hypotézu o normalitě nezamítáme na hladině významnosti 0,05. Testová statistika S-W testu je $W = 0,939785$, odpovídající p-hodnota je 0,146346, tedy hypotézu o normalitě nezamítáme na hladině významnosti 0,05.

Příklad k samostatnému řešení: Byla zaznamenána přesná doba (v minutách) montáže stejných dílů u 4 pracovníků, označme je A, B, C, D. Data jsou uložena v souboru doba_montaze.sta.

Na hladině významnosti 0,05 testujte pomocí Lilieforsovy varianty K-S testu a S-W testem hypotézu, že

- a) doby montáže všech pracovníků dohromady pocházejí z normálního rozložení
- b) doby montáže jednotlivých pracovníků zvlášť pocházejí z normálních rozložení.

Pro obě situace rovněž sestrojte N-P grafy a Q-Q grafy.

Výsledek:

Ad a) Lilieforsův test zamítá normalitu na hladině významnosti 0,05, S-W test nikoli.

Ad b) Lilieforsův test i S-W test zamítají normalitu na hladině významnosti 0,05 pouze u pracovníka D.