Testování hypotéz, obecnější pohled, testy založené na rakových statistikách a testy poměrem věrohodností.

Testování hypotéz – opakování základních pojmů

Mějme náhodný výběr $\{x_1, x_2, \ldots, x_n\}$ testujeme hypotézu H proti alternativě A, k testu hypotézy použijeme statistiku $T(x_1, x_2, \ldots, x_n)$, která bude nazývána testovým kritériem. Obor hodnot, kterých může statistika $T(x_1, x_2, \ldots, x_n)$ nabývat, rozdělíme na dvě disjunktní části W a její doplněk. Pokud hodnota testového kritéria padne do W zamítáme hypotézu H ve prospěch alternativy A. Obor W bude nazýván kritickým oborem. Pokud hodnota T, testového kritéria padne mimo kritický obor W, testovanou hypotézu nezamítáme.

Chyby takového rozhodování

- **Chyba prvního druhu:** Hypotéza H **platí,** ale **je** na základě testového kritéria **zamítnuta** $T(x_1, x_2, \ldots, x_n) \in W$.
- **Chyba druhého druhu:** Hypotéza H **neplatí,** ale **není** na základě testového kritéria **zamítnuta** $T(x_1, x_2, \ldots, x_n) \notin W$.

Pravděpodobnosti chyb

Pravděpodobnost chyby prvního druhu: $\alpha = P(T \in W / H)$. α je pravděpodobností chyby prvního druhu = hladinou významnosti testu = velikostí kritického oboru. Běžně bývá dána tato pravděpodobnost a k ní se určuje odpovídající kritický obor. Taková skutečnost bývá značena $\alpha = P(T \in W_\alpha / H)$. Kritický obor je pak řešením této rovnice k danému α vůči W_α. Pro složené hypotézy lze zavést kritický obor následující nerovností $\alpha \geq P_{P \in H}(T \in W_\alpha / H)$. Danému α je pak přiřazen „maximální" kritický obor vyhovující $\alpha \geq P_{P \in H}(T \in W_\alpha / H); \forall P \in H$. Pravděpodobnost chyby prvního druhu je též nazývána ***hladinou významnosti testu***.

p-hodnota testu: je nejmenší hodnotou hladiny významnosti, při konkrétním testu, při které by bylo možno ještě H zamítnout. Detaily viz: http://www.jstor.org/pss/2684655

Pravděpodobnost chyby druhého druhu: $1 - \beta = P(T \notin W_\alpha / A) = 1 - P(T \in W_\alpha / A)$. Pravděpodobnost β je pak nazývána silou testu založeného na kritickém oboru W_α nebo přímo silou kritického oboru W_α. Síla kritického oboru (testu) je pravděpodobností zamítnutí hypotézy za platnosti alternativy.

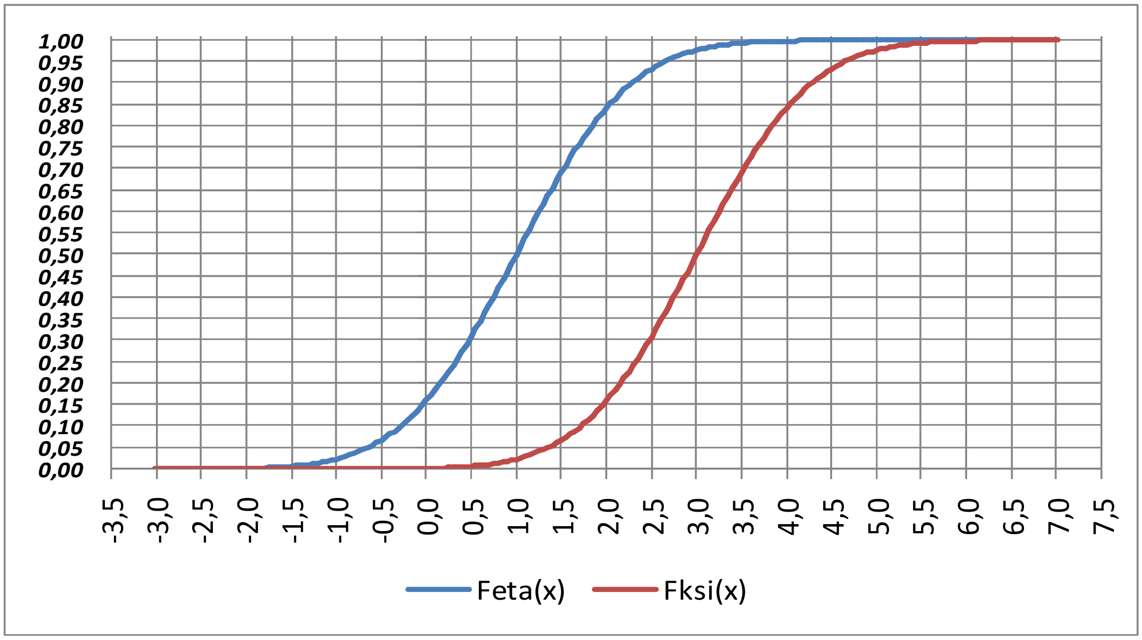
Pravděpodobnostní a statistické srovnávání.

Stochastické srovnávání

Definice: (Hájek J., Vorlíčková D.:Neparametrické metody. Skripta MFFUK, Praha 1974, str. 17) Náhodná veličina ξ je stochasticky větší než η, pokud platí: $P(\xi > x) \geq P(\eta > x)$ pro všechna

reálná x a alespoň pro jedno x_0 platí $P(\xi > x_0) > P(\eta > x_0)$. Dále budeme uvažovat náhodné proměnné se spojitými a rostoucími distribučními funkcemi.
Potom nerovnost $P(\xi > x) \geq P(\eta > x)$, znamená $1 - F_\xi(x) \leq 1 - F_\eta(x) \Leftrightarrow F_\xi(x) \leq F_\eta(x)$.

Tedy: ξ je stochasticky větší než η (značeno $\xi \overset{st}{\triangleright} \eta$) $\Leftrightarrow$ $F_\xi(x) \leq F_\eta(x)$ pro všechna reálná x, přičemž pro alespoň jedno x je taková nerovnost ostrá.



Zde distribuční funkce $F_\xi(x)$ je distribuční funkcí náhodné proměnné ξ, $\xi \overset{st}{\triangleright} \eta$. Tedy distribuční funkce $F_\xi(x)$ je „pod“ $F_\eta(x)$.

Poznámka: Distribuční funkce náhodných proměnných $\xi \overset{st}{\triangleright} \eta$ se nikde na ose R_1 neprotnou, mohou se pouze „jednosměrně dotknout“. Proto libovolné dvě normální náhodné proměnné s různými směrodatnými jsou stochasticky neporovnatelné – dokažte.

Poznámka (zobecněná inverze k distribuční funkci): $F_\xi^{-1}(u) = \inf\{x : F_\xi(x) \geq u\}$. Pokud je $F_\xi(x)$ ostře rostoucí a spojitá, pak taková „zobecněná inverze“ souhlasí s klasickou inverzní funkcí.

Poznámka: $F_\xi(x) \leq F_\eta(x); -\infty < x + \infty \Rightarrow F_\xi^{-1}(u) \geq F_\eta^{-1}(u); 0 \leq u \leq 1$. **Námět:** Dokažte (Inspirace, výše uvedený obrázek).

Wilcoxonův test

Mějme pozorování $x_1, x_2, \ldots, x_m, x_{m+1}, x_{m+2}, \ldots, x_N$, kde $x_1, x_2, \ldots, x_m$ jsou pozorování z prvního výběru a $x_{m+1}, x_{m+2}, \ldots, x_N$ jsou pozorování z druhého výběru. Těmto pozorováním přiřadíme jejich ranky $r_1, r_2, \ldots, r_m, r_{m+1}, r_{m+2}, \ldots, r_N$ (pořadí se stanoví z celé posloupnosti pozorování z obou výběrů). Wilcoxonova statistika je pak zavedena: $S_W = \sum_{i=1}^{m} r_i$, tj. jedná se o součet pořadí připadající na pozorování z prvního výběru. Pokud testujeme hypotézu $H \equiv$ jedná se o výběry z jedné a téže náhodné proměnné a za její platnosti dostaneme:

$E\{S_W\} = \sum_{i=1}^{m} E\{r_i\} = \sum_{i=1}^{m} \frac{N+1}{2} = m\frac{N+1}{2}$, obdobně pro její rozptyl:

$$\sigma^2\{S_W\} = E\left\{\left(\sum_{i=1}^{m} r_i - m\frac{N+1}{2}\right)^2\right\} = E\left\{\left(\sum_{i=1}^{m}\left(r_i - \frac{N+1}{2}\right)\right)^2\right\} = \sum_{i=1}^{m}\sigma^2\{r_i\} + 2\sum_{i=1}^{m}\sum_{j=i+1}^{m} Cov(r_i, r_j) =$$

$$= \sum_{i=1}^{m}\frac{N^2-1}{12} + 2\sum_{i=1}^{m}\sum_{j=i+1}^{m}\left(-\frac{N+1}{12}\right) = m\frac{N^2-1}{12} - 2\frac{m(m-1)}{2}\frac{N+1}{12} = m\frac{N+1}{12}(N-1-m+1) =$$

$= m\frac{(N+1)(N-m)}{12}$. Poznámka: Zatímco u $x_1, x_2, \ldots, x_m$ je dán předpoklad jejich nezávislosti, u ranků $r_1, r_2, \ldots, r_m$ již neplatí.

Hodnota statistiky $S_W = \sum_{i=1}^{m} r_i$ je zdola a shora omezená čísly $\sum_{i=1}^{m} i \le \sum_{i=1}^{m} r_i \le \sum_{i=1}^{m}\left(i + (N-m)\right)$. Dolní omezení je dosaženo, jestliže všechna pozorování $x_1, x_2, \ldots, x_m$ jsou menší než jakékoliv pozorování $x_{m+1}, x_{m+2}, \ldots, x_N$. Horní mez je dosažena, jestliže všechna pozorování $x_1, x_2, \ldots, x_m$ jsou větší než kterékoliv pozorování $x_{m+1}, x_{m+2}, \ldots, x_N$.

Proto: $\frac{m(m+1)}{2} \le \sum_{i=1}^{m} r_i \le \frac{m(m+1)}{2} + m(N-m)$.

Za platnosti hypotézy H má statistika asymptoticky $\left(\min\{m, (N-m)\} \to +\infty\right)$ normální rozdělení.

Kritický obor testu H proti alternativě výběr $x_1, x_2, \ldots, x_m$ je z rozdělení stochasticky většího než rozdělení $x_{m+1}, x_{m+2}, \ldots, x_N$ pak vypadá $\sum_{i=1}^{m} r_i > \underset{\triangleright}{k_{st}}$.

Obdobně kritický obor testu H proti alternativě výběr $x_{m+1}, x_{m+2}, \ldots, x_N$ je z rozdělení stochasticky většího než rozdělení $x_1, x_2, \ldots, x_m$ pak vypadá $\sum_{i=1}^{m} r_i < \underset{\triangleleft}{k_{st}}$. Konstanty $\underset{\triangleright}{k_{st}}$ a $\underset{\triangleleft}{k_{st}}$ se určí na základě předepsané pravděpodobnosti chyby prvního druhu.

Námět: Stanovte $\underset{\triangleright}{k_{st}}$ a $\underset{\triangleleft}{k_{st}}$ pro případ kdy je $\min\{m, (N-m)\}$ dostatečně velké.

Podobným způsobem je konstruován tzv. **mediánový test,** ve kterém se používá statistika $S_{med} = \sum_{i=1}^{m} a_i$, kde $a_i = 0, \text{pokud } r_i \le \frac{N+1}{2}$ a $a_i = 1, \text{pokud } r_i > \frac{N+1}{2}$, (*někdy se používá „přesnější“ modifikace* $a_i = 0, \text{pokud } r_i < \frac{N+1}{2}$, $a_i = \frac{1}{2}, \text{pokud } r_i = \frac{N+1}{2}$ *a* $a_i = 1, \text{pokud } r_i > \frac{N+1}{2}$ *).*

Pro mediánovou (základní verze, tj. pro N sudé) statistiku, pokud testujeme hypotézu $H \equiv$ jedná se o výběry z jedné a téže náhodné proměnné a za její platnosti dostaneme:

$$E\{S_{med}\} = \sum_{i=1}^{m} E\{a_i\} = \sum_{i=1}^{m} \frac{1}{2} = \frac{m}{2}$$

$\sigma^2\{S_W\} = \dfrac{m(N-m)}{4(N-1)}$ nebo $\sigma^2\{S_W\} = \dfrac{m(N-m)}{4N}$. Námět: Odvoďte. Odvoďte též dolní a horní mez této statistiky.

Mediánová statistika má v případě platnosti $H \equiv$ jedná se o výběry z jedné a téže náhodné proměnné hypergeometrické rozdělení (při N sudém):

$P(S_{med} = s) = \dfrac{\binom{\frac{N}{2}}{s}\binom{\frac{N}{2}}{m-s}}{\binom{N}{m}}$ a rozdělení je za uvedeného předpokladu asymptoticky normální. Důkaz: viz Hájek J., Vorlíčková D.:Neparametrické metody. Skripta MFFUK, Praha (1967) 1973, str. 34-39.

Námět: navrhněte testy mediánovou statistikou:

1. H proti alternativě výběr $x_1, x_2, \ldots, x_m$ je z rozdělení stochasticky většího než rozdělení $x_{m+1}, x_{m+2}, \ldots, x_N$
2. H proti alternativě výběr $x_{m+1}, x_{m+2}, \ldots, x_N$ je z rozdělení stochasticky většího než rozdělení $x_1, x_2, \ldots, x_m$.

A to za situace kdy se uplatní asymptotika.

Testy poměrem věrohodností

V případech testů, u kterých je hypotéza „jednoduchou“ a alternativa jednoduchou nebo složenou lze (s určitou modifikací u složené alternativy) užít Neyman-Pearsonovo lemma.Pokud je hypotéza složenou již nejde, bez dalšího použít Neyman-Pearsonovo lemma, proto byl navržen test (pro parametrické hypotézy) poměrem věrohodností. Mějme nějaký prostor možných parametrů G, a rozdělení $f(x;g)$. Budeme testovat hypotézu $g \in G_H \subset G$ obecně proti alternativě $g \in G_A \subset G; G_H \cap G_A = \varnothing$. Jako testové kritérium bude užit poměr věrohodností: $V^*(x_1, x_2, \ldots, x_n) = \dfrac{\sup_{g \in G_H} \prod_{i=1}^{n} f(x_i; g)}{\sup_{g \in G} \prod_{i=1}^{n} f(x_i; g)}$. Pokud existuje maximálně věrohodný odhad $\hat{g}$ parametru g, pak evidentně platí: $V^*(x_1, x_2, \ldots, x_n) = \dfrac{\sup_{g \in G_H} \prod_{i=1}^{n} f(x_i; g)}{\prod_{i=1}^{n} f(x_i; \hat{g})}$. Kritický obor je pak zaveden standardním způsobem $P(V^*(x_1, x_2, \ldots, x_n) \le V_\alpha^*) = \alpha$, z podmínky na pravděpodobnost chyby prvního druhu. Pro $V^*(x_1, x_2, \ldots, x_n)$ platí $0 < V^*(x_1, x_2, \ldots, x_n) \le 1$. Takový test sám o sobě je svou konstrukcí intuitivní (pokud se poměr blíží jedné, je předpoklad, že hypotéza platí, čím je menší lze očekávat, že neplatí;

Neyman, J., Pearson, E.S.: On the use and interpretation of certain test kriteria for purposes of statistical inference. Biometrika 20A, str.175-240, 263-294, rok 1928).

Příklad: Testy hypotéz o střední hodnotě u normálního rozdělení (v případě velkého výběru), „při známém σ“.

V takovém případě při náhodném výběru $\{x_1, x_2, \ldots, x_n\}$ je nestranným odhadem střední hodnoty μ průměr z pozorování $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, ten je současně maximálně věrohodným odhadem. Věrohodnostní funkce pro parametr μ bude:

$V(x_1, x_2, \ldots, x_n; \mu) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = \frac{1}{\left(\sqrt{2\pi}\sigma\right)^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$ a bude nabývat maxima pro průměr $V(x_1, x_2, \ldots, x_n; \bar{x}) = \frac{1}{\left(\sqrt{2\pi}\sigma\right)^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\bar{x})^2}$.

Odtud věrohodnostní poměr: $\frac{V(x_1, x_2, \ldots, x_n; \mu)}{V(x_1, x_2, \ldots, x_n; \bar{x})} = e^{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(x_i-\mu)^2 - \sum_{i=1}^{n}(x_i-\bar{x})^2\right)} = e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left((x_i-\mu)^2-(x_i-\bar{x})^2\right)}$.

Ale $\sum_{i=1}^{n}\left((x_i-\mu)^2 - (x_i-\bar{x})^2\right) = \sum_{i=1}^{n}(\bar{x}-\mu)\left((x_i-\bar{x})+(x_i-\mu)\right) = (\bar{x}-\mu)\sum_{i=1}^{n}\left((x_i-\bar{x})+(x_i-\mu)\right) =$

$= (\bar{x}-\mu)n(\bar{x}-\mu) = n(\bar{x}-\mu)^2$

Proto $\frac{V(x_1, x_2, \ldots, x_n; \mu)}{V(x_1, x_2, \ldots, x_n; \bar{x})} = e^{-\frac{n(\bar{x}-\mu)^2}{2\sigma^2}}$.

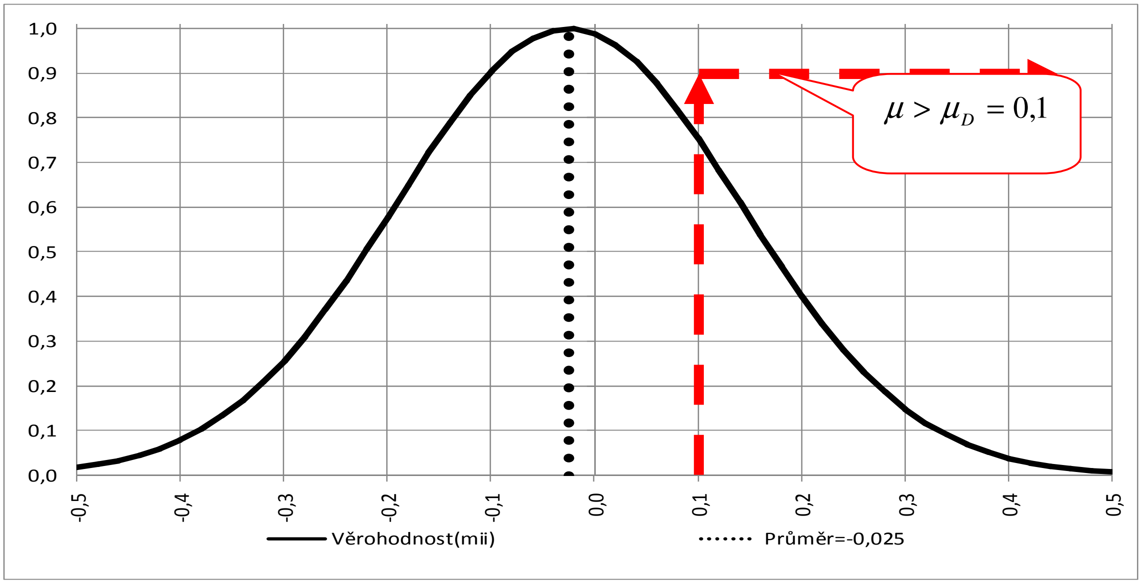
Kritickým oborem je: $\frac{\sup_{\mu\in G_H} V(x_1, x_2, \ldots, x_n; \mu)}{V(x_1, x_2, \ldots, x_n; \bar{x})} < V_\alpha < 1$, tedy $\sup_{\mu\in G_H} e^{-\frac{n(\bar{x}-\mu)^2}{2\sigma^2}} < V_\alpha$.

Nechť $\sup_{\mu\in G_H} e^{-\frac{n(\bar{x}-\mu)^2}{2\sigma^2}} = e^{-\frac{n(\bar{x}-\mu_0)^2}{2\sigma^2}}$, pak lze kritický obor přepsat do tvaru

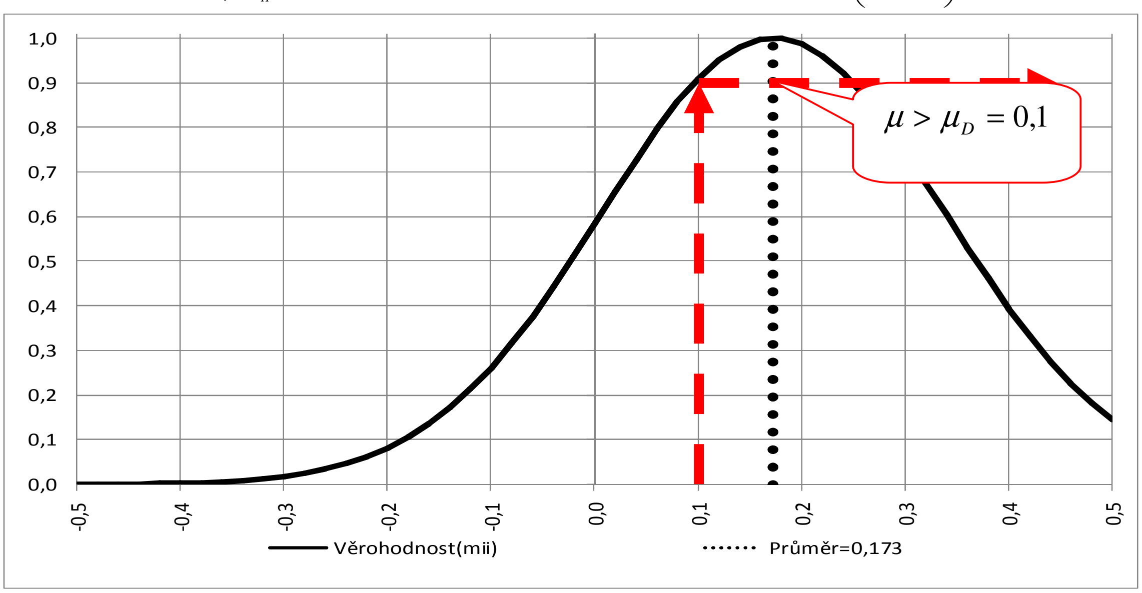
$e^{-\frac{n(\bar{x}-\mu_0)^2}{2\sigma^2}} \le V_\alpha \;\Leftrightarrow\; -\frac{n(\bar{x}-\mu_0)^2}{2\sigma^2} \le \lg V_\alpha \;\Leftrightarrow\; \left(\frac{\bar{x}-\mu_0}{\sigma}\right)^2 \ge \frac{-2\lg V_\alpha}{n}$. Za platnosti hypotézy a znalosti σ bude mít $\left(\frac{\bar{x}-\mu_0}{\sigma}\right)^2$ $\chi^2(1)$ rozdělení, proto pravděpodobnost chyby prvního druhu bude $P\left(\left(\frac{\bar{x}-\mu_0}{\sigma}\right)^2 \ge \frac{-2\lg V_\alpha}{n}\right) = 1 - F_{\chi^2(1)}\left(\frac{-2\lg V_\alpha}{n}\right) = \alpha$ a odtud $\frac{-2\lg V_\alpha}{n} = \chi^2_{1-\alpha}(1)$ a $V_\alpha = e^{-\frac{n\chi^2_{1-\alpha}(1)}{2}}$. Pokud by bylo σ neznámé, lze kritický obor „aproximovat“ $\frac{(\bar{x}-\mu_0)^2}{s^2} \ge \frac{-2\lg V_\alpha}{n}$, tj. levá strana je podíl dvou χ^2 náhodných proměnných, v čitateli s jedním stupněm volnosti a ve jmenovateli s *n-1* stupni volnosti, proto se v takovém případě užije místo $\chi^2_{1-\alpha}(1)$ kvantilu $F_{1-\alpha}(1, n-1)$ kvantil Fisher-Snedecorova rozdělení.

Stanovení μ_0

Budeme předpokládat $G_H \equiv \{\mu : \mu > \mu_D\}$, vztah věrohodnostního poměru a „oblasti“ G_H může nabývat následujících dvou „relací“:



V tomto případě je $\sup\limits_{\mu \in G_H} e^{-\frac{n(\bar{x}-\mu)^2}{2\sigma^2}} = e^{-\frac{n(\bar{x}-\mu_D)^2}{2\sigma^2}}$ a hypotézu zamítáme pokud $\left(\frac{\bar{x}-\mu_0}{\sigma}\right)^2 \geq \frac{-2\lg V_\alpha}{n}$ a



pokud data „vygenerují“ tuto situaci je $\sup\limits_{\mu \in G_H} e^{-\frac{n(\bar{x}-\mu)^2}{2\sigma^2}} = e^{-\frac{n(\bar{x}-\bar{x})^2}{2\sigma^2}} = 1$ a hypotézu nezamítáme.

Tento příklad však naznačuje, že kritický obor není takový, jaký byl výše naznačen, ale je průnikem dvou oblastí $\left(\frac{\bar{x}-\mu_D}{\sigma}\right)^2 \geq \frac{-2\lg V_\alpha}{n}$ a $\bar{x} < \mu_D$

Námět: Zjednodušte popis takového kritického oboru a stanovte V_α pro tento kritický obor.
Námět: Obdobně rozeberte případ, kdy $G_H \equiv \{\mu : \mu < \mu_H\}$.
Námět: Pro čísla $\mu_1 < \mu_2 < \mu_3 < \mu_4$ rozeberte případ kdy $G_H \equiv \{\mu : (\mu_1 \leq \mu \leq \mu_2) \vee (\mu_3 \leq \mu \leq \mu_4)\}$.

Shrnutí: Pokud padne $\bar{x}$ dovnitř nebo na kraj oblasti G_H hypotézu nezamítáme. Pokud padne $\bar{x}$ mimo oblast G_H hypotézu zamítáme, pokud je $\bar{x}$ „dostatečně daleko"

Námět: Rozeberte situaci kdy: H formulována $\mu = \mu_0$ a alternativa A je formulována $\mu = \mu_1$, pro případy, kdy $\mu_0 < \mu_1$ a $\mu_0 > \mu_1$. (Klasická úloha řešitelná pomocí Neyman Pearsonova lemmatu)
Námět: Jak by dopadla situace $\mu_0 = \mu_1$ a proč nemá smysl? Mělo by smysl testovat hypotézu: pozorovaná data se řídí normálním rozdělením se střední hodnotou μ_0 a směrodatnou odchylkou σ_0 proti alternativě pozorovaná data se řídí normálním rozdělením se střední hodnotou μ_0 a směrodatnou odchylkou σ_1, kde σ_1 je mnohem větší než σ_0? Jak by takový test mohl vypadat?

Výše uvedený příklad je demonstrací toho, jak se za poměrně silných předpokladů může v klasickém pojetí komplikovat test složené hypotézy proti jednoduché i složené alternativě.

Námět: Lze za uvedených předpokladů použít v nějaké modifikaci test jednoduché hypotézy proti složené alternativě jako test složené hypotézy proti jednoduché alternativě.

Doporučená a zdrojová literatura:

Jaroslav Hátle, Jiří Likeš	Základy počtu pravděpodobnosti a matematické statistiky. SNTL Praha 1974.
Alfréd Rényi	Teorie pravděpodobnosti, ACADEMIA, Praha 1972
C. Radhakrishna Rao	Lineární metody statistické indukce a jejich aplikace, ACADEMIA, Praha 1978
Machek J.	Teorie odhadu, SPN Praha 1974, skripta MFFUK
Jurečková J.	Pořadové testy. Skripta MFF UK, Praha 1981.
Hájek J., Vorlíčková D.	Neparametrické metody. Skripta MFFUK, Praha (1967) 1973
Blatná D.	Neparametrické metody. Testy založené na pořádkových a pořadových statistikách. Skripta VŠE, Praha 1996.
Neyman, J., Pearson, E.S.	On the use and interpretation of certain test kriteria for purposes of statistical inference. Biometrika 20A, str.175-240, 263-294, rok 1928
Mark J. Schervish	P Values: What They Are and What They Are Not. The American Statistician, Vol. 50, No. 3 (Aug., 1996), pp. 203-206. http://www.jstor.org/pss/2684655