

Přednáška 6 – testy hypotéz obecně, testy pro jednu veličinu

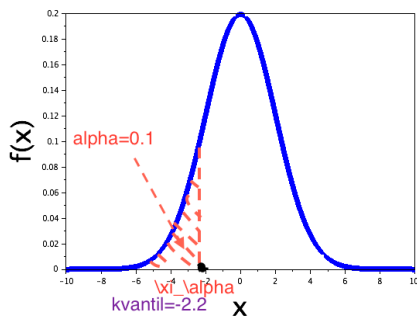
Předchozí přednáška 5 byla na téma Soubor a výběr. Je to těžká látka, proto bych chtěla tady shrnout základ, který byste si z ní měli odnést: snažíme se odhadnout parametry rozdělení náhodné veličiny pomocí výpočtů z náhodného výběru dat. Tyto výpočty jsou statistiky (např., výběrový průměr, výběrový rozptyl, atd). Abychom dostali pomocí statistik kvalitní odhad, musí statistiky mít určité vlastnosti (nestrannost, konzistence, vydatnost).

Příklad: Vláda chce nechat otestovat na koronavirus reprezentativní vzorek populace (toto je výběr). Plošné testování celé populace (toto je soubor) není podle ministra zdravotnictví na pořadu dne – nepovažují ho za efektivní a kapacity jsou omezené (zdroj [idnes.cz](https://www.idnes.cz)).

V dnešní přednášce se podíváme, jak můžeme některé statistiky použít pro odhad parametrů souboru především s normálním rozdělením $N(\mu, \sigma^2)$. Nejdřív nás bude zajímat, v jakém intervalu se nachází hodnota parametru.

Interval spolehlivosti

Interval spolehlivosti (IS) je interval, ve kterém se nachází hodnota odhadovaného parametru s pravděpodobností $1 - \alpha$. Připomeneme, že α jsme už probírali v souvislosti s kvantily. Kvantil dělí naměřená data na dvě části a α je plocha pod křivkou od $-\infty$ do hodnoty kvantilu. Např., pro normální rozdělení to může vypadat takto:



Pro účely IS se nejčastěji používá

$$\alpha = 0.05 \quad \text{nebo} \quad 0.01 \quad \text{nebo} \quad 0.1$$

což odpovídá:

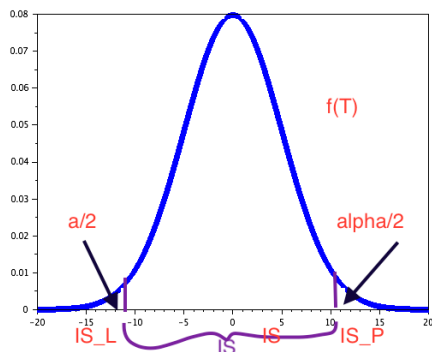
$$1 - \alpha = 0.95 \quad \text{nebo} \quad 0.99 \quad \text{nebo} \quad 0.90,$$

nebo to velmi často můžete potkat ve tvaru

$$1 - \alpha = 95\% \quad \text{nebo} \quad 99\% \quad \text{nebo} \quad 90\%.$$

Např., s pravděpodobností 0.95 (tj., $1 - \alpha$) patří parametr do IS, a jenom 0.05 (tj., α) necháváme na možnost, že tam není. Formálně α budeme říkat hladina významnosti* (tj., na kolik věříme odhadu). Například, řeknou nám, že s jistotou 95% dostaneme plat 50 tisíc a s jistotou 10% dostaneme 100 tisíc. Čemu budeme věřit víc?

Interval spolehlivosti (IS) má graficky následující tvar:



Toto je oboustranný IS. Tady IS = (IS_L, IS_P) , kde IS_L je levá hranice IS, IS_P je pravá hranice. Hranice IS jsou

kvantily a kritické hodnoty vypočtené pro určité statistiky a jejich rozdělení. Křivka $f(T)$ ukazuje hp použité statistiky T – tady je to normální rozdělení. Pozor, nezaměňujeme rozdělení souboru (náhodné veličiny), která nás zajímá) a rozdělení statistiky T ! Každá statistika má svoje rozdělení, např:

- Pokud hledáme parametr μ , ale známe σ^2 , má statistika normální rozdělení (jako na obrázku).
- Pokud hledáme μ při neznámém σ^2 , bude to Studentovo rozdělení.
- Pokud hledáme parametr rozptyl σ^2 , má statistika χ^2 -rozdělení.
- Pro podíl p – přibližně normální rozdělení.

*V literatuře můžete potkat pojem hladina významnosti i pro $1 - \alpha$, často se totiž navzájem zaměňují.

V případě oboustranného IS se hladina významnosti α dělí napůl, máme tedy plochu pod křivkou $\frac{\alpha}{2}$ vlevo od hranice intervalu a $\frac{\alpha}{2}$ vpravo, přičemž platí

$$P(\theta < IS_L) = \frac{\alpha}{2}, \quad P(\theta > IS_P) = \frac{\alpha}{2} \quad - \text{jako na obrázku}$$

To znamená, že pravděpodobnost toho, že hledaný parametr (obecně θ) je menší než levá hranice IS, se rovná $\frac{\alpha}{2}$ (to je dost malá pravděpodobnost, když např., $\alpha = 0.05$). Vpravo to je podobné. Ale s pravděpodobností $1 - \alpha$ patří parametr θ do IS, tj., od IS_L do IS_P .

Příklad Zajímá nás intervalový odhad ceny na olejový filtr pro automobil Škoda Octavia na hladině významnosti 0.05. Z dlouhodobých měření je známo, že rozptyl cen na olejové filtry $\sigma^2 = 6$.

Vezmeme výběr cen z různých eshopů:

$$X = [100 \ 154 \ 110 \ 137 \ 119 \ 109 \ 138 \ 164 \ 136 \ 85 \ 138 \ 105 \ 121 \ 149].$$

Zprůměrujeme ceny z výběru: $\bar{X} = 126.07\text{Kč}$ – to je výběrový průměr, a to je tedy naše statistika. Pomocí IS hledáme, zda \bar{X} skutečně patří do intervalu průměrné ceny olejových filtrů (to je hledaný parametr μ) s pravděpodobností 0.95.

Hranice IS by šlo určit pomocí výpočtu kvantilů pro normované statistiky. Dřív se to dělalo podle speciálních tabulek. V současné době používáme pro určení IS výhradně statistický software.

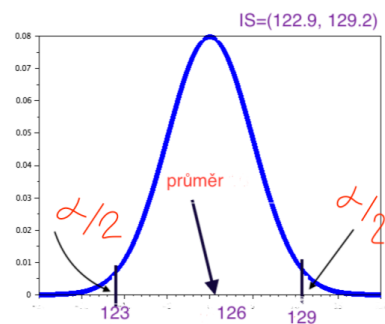
Pro konstrukci IS (v praxi volbu funkce software) potřebujeme vědět:

- hladinu významnosti α (tj., s jakou pravděpodobností). Pokud není v software potřeba zadávat α , většinou se použije 0.05.
- rozdělení souboru – předpokládáme, že cena na olejové filtry obecně je náhodná veličina s normálním rozdělením a její střední hodnota μ je neznámý parametr.
- statistiku a její rozdělení – tady to je \bar{X} , který má normální rozdělení (protože známe rozptyl σ^2).

Pro náš příklad v Matlabu nám vyšly následující hranice oboustranného IS:

$$IS_L = 122.9285, \quad IS_P = 129.2144,$$

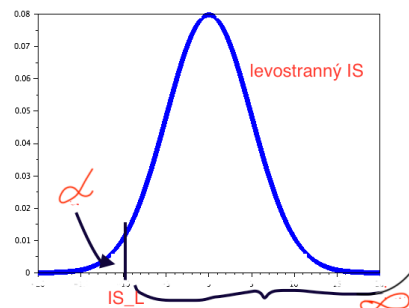
což znamená, že se průměrná cena olejových filtrů pohybuje v rozmezí od 122.9 Kč do 129.2 Kč, a náš výběrový průměr $\bar{X} = 126.07\text{Kč}$ patří do IS.



Kromě oboustranného IS, existují ještě jednostranné IS – levostranný a pravostranný. Pro náš příklad by nás mohl zajímat odhad minimální ceny filtrů – v tomto případě by šlo o levostranný IS. V tomto případě má IS pouze levou hranici:

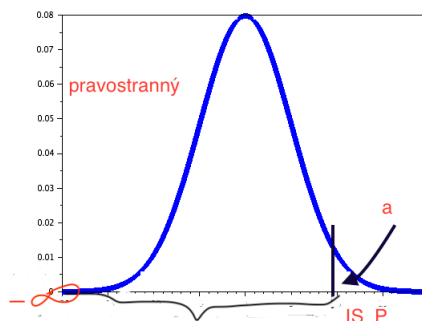
$$IS = (123.4338, \infty)$$

a plocha pod křivkou do hodnoty levé hranice (kvantilu) je celá hodnota α .



Podobnou logiku má odhad maximální ceny olejových filtrů – to je pravostranný IS:

$$IS = (-\infty, 128.7091).$$



Některé druhy software rozlišují pouze oboustranný a jednostranný IS nebo používají jinou interpretaci stran IS, tj., směr IS nemusí být jednoznačně určen. Z tohoto důvodu, je důležité při napsání zápočtových testů a zkoušení rozpoznat oboustranný a jednostranný případ.

My budeme používat IS v rámci testování hypotéz výhradně pomocí funkcí software k tomu určených.

Testy hypotéz obecně

Dá se říct, že pojem test hypotéz je prakticky totéž, co je IS, jenom “naruby”. Při testování hypotéz máme dvě hypotézy:

- nulovou hypotézu H_0 , která představuje nějaké tvrzení (podle úlohy, kterou momentálně řešíme), např. tvrdí, že hledaný parametr se rovná nějaké předpokládané hodnotě θ_0 :

$$\underbrace{\theta}_{\text{hledaný parametr}} = \underbrace{\theta_0}_{\text{hodnota}}, \text{ například, } \mu = \mu_0,$$

- dále máme alternativní hypotézu H_A , která buď popírá nulovou hypotézu H_0 , tj., tvrdí, že se nerovná:

$$\theta \neq \theta_0 \quad - \quad \text{v tomto případě se využije } \underline{\text{oboustranný}} \text{ test}$$

nebo říká, že je menší:

$$\theta < \theta_0 \quad - \quad \text{využije se } \underline{\text{levostranný}} \text{ test}$$

nebo říká, že je větší:

$$\theta > \theta_0 \quad - \quad \text{použijeme } \underline{\text{pravostranný}} \text{ test.}$$

Nás zajímá, jestli můžeme nulovou hypotézu H_0 na zvolené hladině významnosti α zamítnout. To znamená, že testujeme hypotézu H_0 .

Jak to funguje?

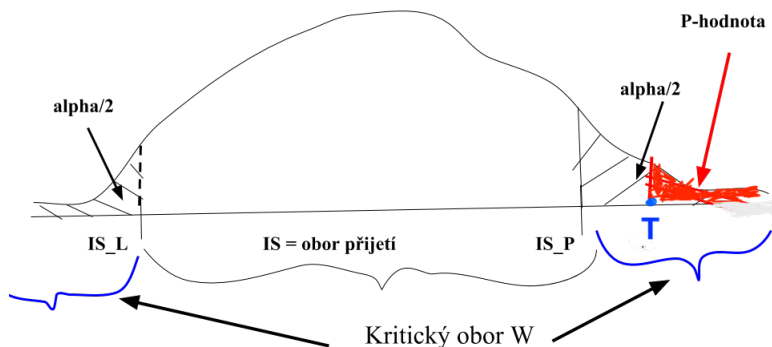
- Nejdřív vytvoříme IS.
- Spočteme statistiku T , např. výběrový průměr \bar{X} .

- Podíváme se, zda statistika T patří do IS, kterému se také z hlediska testování hypotéz říká obor přijetí.
- Pokud T padne mimo IS – v tomto případě T padne do tzv. kritického oboru

$$W = (-\infty, IS_L) \cup (IS_P, \infty),$$

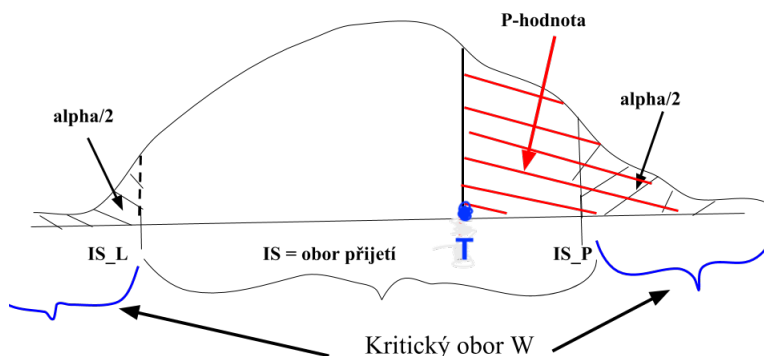
To znamená, že pravděpodobnost toho, že H_0 platí (tvrzení $\theta = \hat{\theta}$) je moc nízká. Těto pravděpodobnosti říkáme p-hodnota, tj., $p_h = P(H_0)$.

- V tomto případě H_0 zamítáme na hladině významnosti α (pozor, ale nepotvrzujeme tímto H_A !).



Na obrázku je schéma oboustranného testu hypotéz, kde se kritický obor W nachází vlevo a vpravo od IS (oboru přijetí). Vidíme statistiku T , hodnota které patří do kritického oboru vpravo (označená modrou tečkou). P-hodnota je pravděpodobnost nulové hypotézy, a to je plocha pod křivkou od hodnoty statistiky T do nekonečna (označena červenou barvou). Je vidět, že p-hodnota je menší než plocha α . V tomto případě zamítáme H_0 na hladině významnosti α .

Pokud ale statistika T patří do IS (oboru přijetí), tím pádem p-hodnota $> \alpha$. V tomto případě nulovou hypotézu H_0 nezamítáme (ale nepotvrzujeme ji, pouze nezamítáme!).



Tady vidíme zase oboustranný test hypotéz, ale hodnota statistiky T teď patří do IS. V tomto případě je p-hodnota (červená plocha pod křivkou od hodnoty T do nekonečna) velká a je větší než plocha α . To znamená, že pravděpodobnost, že nulová hypotéza platí, je dost velká a nezamítáme ji. Nemůžeme ji ale potvrdit, pouze nevyklučujeme tuto možnost.

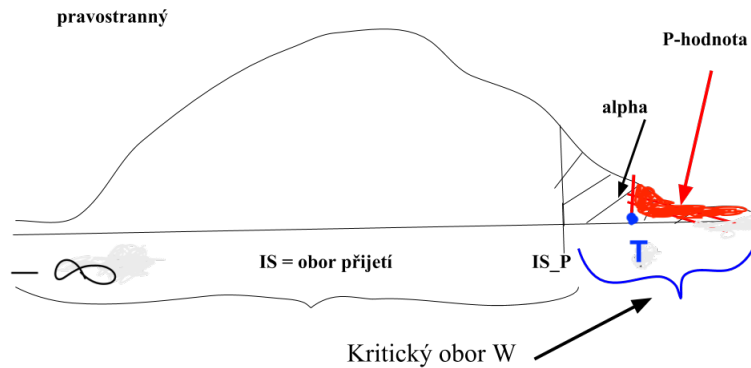
Pro jednostranné testy je princip stejný, jenom kritický obor je pouze vlevo v případě levostranného testu:

$$W = (-\infty, IS_L)$$

a je pouze vpravo pro pravostranný test:

$$W = (IS_P, \infty),$$

viz obrázek s pravostranným testem hypotéz.



Dokážete říct podle obrázku, zda v tomto případě zamítáme nulovou hypotézu?

Odověď: Na obrázku p-hodnota (červená plocha) je menší než α (plocha od nekonečna) proto nulovou hypotézu zamítáme. Protože statistika T je v kritickém oboru.

Shrnutí

Jak je vidět, **závěr**, zda **zamítáme nulovou hypotézu**, děláme **na základě p-hodnoty**. **P-hodnotu** nám spočte statistický software (podobně jako **IS** – podle rozdělení statistiky a souboru).

Proto potřebujeme pro **testování hypotéz** pouze:

- správně zvolit test (o tom později),
- H_0 – co testujeme,
- směr testu – určujeme podle H_A (neplatí pro všechny testy, některé mohou být pouze obou- nebo jednostranné),
- hladinu významnosti α .

Dále obecně soudíme podle **následujícího pravidla**:

Pokud $p_h < \alpha$, **zamítáme** H_0 .

Pokud $p_h > \alpha$, **nezamítáme** H_0 .

Je nutné ale poznamenat, že v případě **zamítnutí** nulové hypotézy **nepotvrzujeme** tímto alternativní hypotézu. Nápodobně, pokud **nezamítáme** nulovou hypotézu, **nepotvrzujeme** ji. Pouze zamítnutí H_0 má statistický význam.

Příklad

Sledujeme rychlost vozidel v tunelu Blanka na úseku s maximální povolenou rychlostí 70 km/h se směrodatnou odchylkou 1[†]. Policie tvrdí, že řidiči systematicky porušují pravidla a jezdí na úseku vyšší rychlostí. Chceme otestovat toto tvrzení na hladině významnosti 95% za předpokladu normality dat.

[†]Připomeneme, že podle pravidla 3 σ bude v tomto případě rychlost od 67 do 73 km/h.

Řešení Změřili jsme rychlost 100 náhodně vybraných vozidel – to je výběr. Spočítali jsme výběrový průměr, např., $\bar{X} = 69.79$ – to je naše statistika. Předpokládáme normální rozdělení rychlosti (je dána normalita dat), známe rozptyl (máme směrodatnou odchylku), zajímá nás průměrná rychlost řidičů (parametr střední hodnota μ).

Nulovou hypotézu formulujeme podle tvrzení. Formálně u H_0 musíme nejdřív říct, že se parametr rovná předpokládané hodnotě ($\mu = 70$), a až potom k tomu přidáme tvrzení policie, že řidiči jezdí vyšší rychlostí ($\mu > 70$). Nulová hypotéza tedy zní:

$$H_0 : \mu=70 \text{ (nebo } \mu>70) \quad - \quad \text{přidáme podle } \textit{tvrzení}.$$

Alternativní hypotéza je opačné tvrzení:

$$H_A : \mu<70 \quad - \quad \text{musí být } \textit{opak} \ H_0, \text{ a také určuje, že test je } \textit{levostranný}.$$

Ve Scilabu použijeme funkci z_test pro výpočet p-hodnoty, kterou voláme takto:

```
//z_test(mu0,av,vr,n,side), kde mu0 je 70, av je výběrový průměr, vr je známý rozptyl,
n je počet dat, side je strana testu (levá).
```

Pro náš příklad to je:

```
z_test(70,69.79,1,100,'1')
```

Výsledek je:

$p_h = 0.0178644 < 0.05$ – zamítáme H_0 . **Závěr:** není pravda, že řidiči systematicky převyšují rychlost 70 km/h.

Takto fungují testy hypotéz obecně, ale výpočet statistik a p-hodnoty se liší v závislosti na tom, co chceme testovat a kolik výběrů máme. Proto je velice důležité správně zvolit test, dále práce s ním v software je triviální. Musíme ale být schopni rozumět výsledkům testování. První, podle čeho se rozhodujeme, jaký test použijeme, je počet výběrů, se kterými pracujeme. Začneme případem, kdy máme pouze jeden výběr.

Testy hypotéz pro jeden výběr

V případě jednoho výběru se testy hypotéz liší podle toho, co testujeme. Rozlišujeme především parametrické testy, kde pracujeme s daty, která pochází z normálního rozdělení, dále neparametrické testy, kde data nesplňují předpoklad normality nebo výběr je malý, a testy, které použijeme pro určení rozdělení dat, viz tabulka.

Testy parametrů s předpokladem normality	Neparametrické testy bez předpokladu normality	Testy rozdělení
<u>Test střední hodnoty</u> při známém rozptylu – <u>z_test</u> $H_0: \mu = \mu_0(>, <)$ známý σ^2	<u>Test mediánu</u> (znaménkový test) $H_0: \tilde{x}_{0,5} = \tilde{x}_{(0,5)_0}(>, <)$	<u>w/s test normality</u> . H_0 : výběr má normální rozdělení
<u>Test střední hodnoty</u> při neznámém rozptylu – <u>t_test</u> $H_0: \mu = \mu_0(>, <)$ neznámý σ^2 – použijeme s^2	<u>Wilcoxonův test</u> – <u>wilcoxon_test</u> $H_0: \tilde{x}_{0,5} = \tilde{x}_{(0,5)_0}(>, <)$	<u>Kolmogorov-Smirnovův test</u> <u>ks_test_cont</u> , <u>ks_test_disc</u> H_0 : výběr má teoretické rozdělení Pouze pravostranný
<u>Test rozptylu</u> – <u>var_test</u> $H_0: \sigma^2 = \sigma_0^2(>, <)$		<u>Shapiro-Wilkův test normality</u> <u>shapiro_test</u> H_0 : výběr má normální rozdělení
všechny levo-, pravo-, oboustranné	všechny levo-, pravo-, oboustranné	<u>Anderson-Darlingův test</u> – <u>adtest</u> H_0 : výběr má teoretické rozdělení
		<u>χ^2 test dobré shody</u> pro normalitu – <u>normCh2_test</u> pro jiná rozd. – <u>chisquare_test</u> H_0 : výběr má teoretické rozdělení Pouze pravostranný
		<u>Q-Q graf</u> – <u>qqplot</u> grafická metoda