

Cvičenie 10 – Testovanie hypotéz

Riešené príklady

Pri základnom testovaní hypotéz o strednej hodnote a rozptyle (z technického hľadiska) ide o prácu s intervalovými odhadmi, ktoré sa preberali pred týždňom. Ten istý princíp, tie isté vzorce, len otázka bude položená inak. Zo suchých výpočtov urobíme tak trochu rituál „súdneho procesu“.

Príklad 1a:

Na preteky sa prihlásilo 30 plavcov. Predpokladalo sa, že trať by mali zvládnuť v priemere za 55 minút. V skutočnosti však stredná hodnota dosiahnutých časov vyšla až 72 minút. Rozptyl výsledkov je $s^2 = 400$. Na hladine významnosti 5% testujte hypotézu $H_0 (\mu=55)$.

Riešenie:

Najprv si ľudskou rečou povedzme, čo ideme robiť. Na základe viacročných skúseností s pretekmi tohto typu bolo odhadnuté, že plavci by to mohli zvládnuť v priemere za 55 minút. Muselo byť prekvapením, ak po pretekoch spočítali priemer časov a vyšlo im 72. Je to dosť veľa nad 55. Otázka znie, či sa tento rozdiel dá vysvetliť iba náhodou, teda či môžeme pripustiť, že stredná hodnota naozaj môže byť 55 a len tento rok to nejak ušlo (lebo náhoda). A pokiaľ povieme, že to nemôže byť náhoda, chceme mať len 5% rizika, že sa mýlime.

– Ak nevieme (na základe výpočtov) protirečiť tomu, že $\mu=55$, hypotézu H_0 nezamietneme. Pozor!!! Nezamietnuť neznamená potvrdiť či dokázať hypotézu H_0 . Nedokázali sme ju (ani zďaleka), len pri povolenom riziku sme si netrúfli ju zamietnuť. Niečo ako keď súd oslobodí pre nedostatok dôkazov – čo nie je nutne to isté ako skutočná nevina.

Ak odmietnutie (opäť na základe výpočtov) možnosti, že $\mu=55$, nesie v sebe riziko omylu menšie než 5%, hypotézu H_0 zamietneme. To znamená, že oficiálne vyhlásime predpoklad $\mu=55$ za „utrhnutý od reality“ nameraných čísel, pričom sme si vedomí, že sa na 5 percent môžeme mýliť. Prakticky to znamená tvrdiť, že číslo 72 miesto 55 nemôže byť výsledkom náhody a musí mať svoju príčinu napr. v inej trati, nečakaných prúdoch, inom výkonnostnom zložení plavcov a pod.

Podľa teraz dať do vzorcov namerané hodnoty rovnakým postupom ako minule. Máme bodový odhad strednej hodnoty $m=72$ a bodový odhad rozptylu $s^2=400$, tj. $s=20$. Tu musíme zdôrazniť, že 30 plavcov a ich výsledky sú len reprezentanti (výber) zo širšej reality, takže na naše úvahy použijeme Studentovo rozdelenie.

Pre spoľahlivosť $1-\alpha$, kde $\alpha = 5\%$, potrebujeme kvantil $u = t_{0.975, 29} = 2.0452$. Dosadíme do známeho vzorca

$$d = s \cdot u / \sqrt{n} = 20 \cdot 2.0452 / \sqrt{30} = 7.47.$$

Podľa nameraných dát by stredná hodnota mala na 95% byť v intervale $72 \pm 7.47 = [64.53, 79.47]$. To znamená, že je len 5% šanca, že stredná hodnota je mimo tohto intervalu.

Teraz sa vráťme k hypotéze $\mu=55$. Vidíme, že 55 nepatrí (!) do vyššie uvedeného intervalu. Je priam brutálne mimo. Pri 5% povolenom riziku (toto riziko určovalo veľkosť intervalu) omylu si preto nedovolíme pripustiť, že 55 by mohla byť stredná hodnota. Hypotézu H_0 ZAMIETAME.

Poznámka 1: Proti nulovej hypotéze H_0 , ktorú posudzujeme, sa stavia tzv. alternatívna hypotéza H_1 , ktorá je buď doslovnou negáciou H_0 alebo (vzhľadom na kontext) je jej inak formulovaný „opak“. Zamietnutím H_0 vyslovujeme, že platí H_1 . V našom prípade by H_1 mohlo prvoplánovo znieť „ $\mu \neq 55$ “, ale vzhľadom na situáciu a kontext si môžeme dovoliť tvrdenie upresniť na „ $\mu > 55$ “.

Poznámka 2: Asi tušíme, že v tomto prípade zamietnutie hypotézy nesie v sebe menšie riziko než 5%, keďže 55 bolo naozaj ďaleko od hraníc intervalu. Podrobnejšie sa na to pozrieme o kúsok ďalej.

Príklad 1b:

Iný tím odborníkov z konkurenčného klubu odhadol na základe svojich skúseností a vedomostí očakávaný priemerný čas plavcov na 67 minút. Na hladine významnosti 5% testujte hypotézu H_0 ($\mu=67$).

Riešenie:

Príklad je pokračovaním 1a. Tam sme zistili, že podľa výsledkov pretekov je stredná hodnota na 95% v intervale $72 \pm 7.47 = [64.53, 79.47]$.

Predpokladaná $\mu=67$ patrí do uvedeného intervalu. Hypotézu H_0 teda nezamietame. Nedokázali sme ju, len pripúšťame, že by mohla platiť (nemáme dosť silný argument proti nej). Tvrdiť, že neplatí, by predstavovalo riziko väčšie, než kryje naša poisťka (tj. dovolených 5%).

Poznámka 3:

Známy vzorec môžeme použiť trochu inak.

$$d = s \cdot u / \sqrt{n}$$

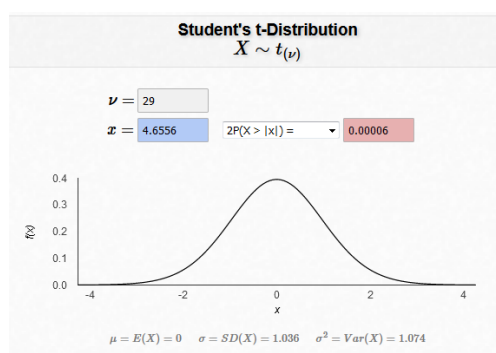
upravíme takto:

$$u = \sqrt{n} \cdot d / s$$

Za číslo d dosadíme rozdiel $|m-\mu|$.

$$1a) \quad u = |72-55| / 20 \cdot \sqrt{30} = 4.6556 .$$

Hypotézu budeme teraz hodnotiť priamo podľa hodnôt v tabuľke príslušného rozdelenia. Konkrétne v tomto prípade, v 29. riadku t-rozdelenia vidíme, že hodnotu 4.6556 ani nič blízko nej nevidíme :) je strašne ďaleko. Online kalkulačka to istí a prezradí, že



hypotézu môžeme zamietnuť s rizikom (na hladine) 0.00006.

$$1b) \quad u = \sqrt{30} \cdot |72-67| / 20 = 1.917 .$$

V tabuľkách vidíme, že v 29. riadku je ku 1.917 najbližší väčší výsledok 2.0452, čo zodpovedá kvantilu 0.975. To znamená, že na hladine 0.05 (dvakrát 0.025, kde $0.025 = 1 - 0.975$) síce tesne ale predsa len nezamietame hypotézu H_0 .

V 29. riadku je ku 1.917 vľavo najbližšie hodnota 1.6991, čo je kvantil 0.95. Na hladine 0.1 preto H_0 môžeme zamietnuť.

Inými slovami, odhad 67 tiež nie je žiaden zázrak. Na hladine 0.05 ho nezamietneme, ale akonáhle máme dovolené riziko 0.1, už sa zamietnuť dá. Presnejšia hranica medzi zamietnutím a nezamietnutím je 0.06514 (z kalkulačky).

Príklad 2a:

V obecnom kravíne v dôsledku tráviacich procesov dobytku, nedostatočného vetrania a chyby v elektroinštalácii prišlo k výbuchu, ktorý spôsobil isté škody na majetku.¹ Po analýze príčin sa okrem lepšieho vetrania namontovali senzory zvýšenej koncentrácie metánu.

Kritická hranica koncentrácie, pri ktorej môže prísť k výbuchu, je 50 až 150 ‰. Senzor je spojený s alarmom, ktorý je výrobcom nastavený na 10 ‰. To je číslo, ktoré by s malými odchýlkami malo byť strednou hodnotou možných výsledkov.

Riaditeľ kravína sa rozhodol otestovať senzory v 18 pokusoch. Hypotéza H_0 bola $\mu=10$, stanovená bola aj hladina významnosti 5%. Experiment priniesol 18 výsledkov, ktorých stredná hodnota je $m = 10.079$, s odchýlkou $s^2 = 0.0194761$ (teda $s = 0.13956$).

Zhodnoťte, ako dopadli senzory/alarmy v teste.

Riešenie:

$H_0: \mu=10$ // $H_1: \mu>10$

Podme na vec hneď “druhým” spôsobom, teda cez vzorec

$$u = \sqrt{n} * d / s$$

Dosadíme $d = |m - \mu| = |10 - 10.079| = 0.079$:

$$u = \sqrt{18} * 0.079 / 0.13956 = 2.4016$$

Pozrieme sa do 17. riadku t-rozdelenia:

17	t(v)				
	0,9	0,95	0,975	0,99	0,995
	1,3334	1,7396	2,1098	2,5669	2,8982

Hodnota je niekde medzi stĺpcami 0.975 a 0.99. To znamená, že na hladine 0.02 = $2*(1-0.99)$ hypotézu H_0 nevieme zamietnuť. Ale to iba tak pomimo – na zadaním požadovanej hladine 0.05 = $2*(1-0.975)$ ju samozrejme zamietame.

Riaditeľ kravína poslal reklamáciu výrobcovi senzora. Ten ju akceptoval, ale keďže nešlo o principiálnu nefunkčnosť senzora, vyriešil to iba zľavou a prelepením deklarovanej hranice na 10.1. Nové testovanie, opäť s 18 pokusmi, prinieslo rovnaké výsledky.

Počítajme: $d = |m - \mu| = |10.1 - 10.079| = 0.021$

$$u = \sqrt{18} * 0.021 / 0.13956 = 0.63840$$

Hodnota u je menšia než 2.1098 v stĺpci 0.975, takže na hladine 5% H_0 nezamietame. Pozornejší pohľad ukazuje, že nezamietat' by sa dalo aj na dosť vyšších hladinách (tj. ponúkame pochybovačom vyššie riziko, a oni napriek tomu hypotézu nezamietnu).

¹ <https://www.novinky.cz/koktejl/clanek/nadmira-metanu-z-kravskych-utrob-zpusobila-vybuch-v-kravine-216965>

Pokračujeme v téme stráženia metánu kravína.

V 18 meraniach spomenutých vyššie sa zistilo $s^2 = 0.0194761$, výrobca však deklaroval rozptyl $\sigma^2 = 0.0144$. Dá sa na hladine 5% deklarácia výrobcu zamietnuť (a reklamovať výrobok)?

$$H_0: \sigma^2 = 0.0144 \quad // \quad H_1: \sigma^2 > 0.0144$$

Na predošlom cviku sme stanovili hranice L a P intervalu, v ktorom sa podľa meraní môže vyskytovať skutočný rozptyl.

Ak výrobcom deklarovaná hodnota bude v tomto intervale, je všetko v poriadku.

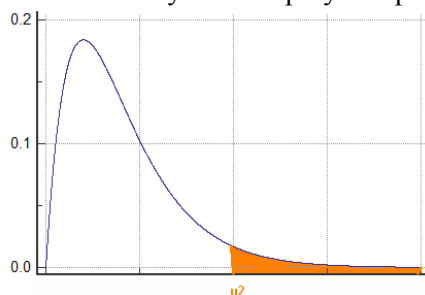
Ak je výrobcom udávaný rozptyl vpravo od intervalu LP, tak sme tiež spokojní, naše merania totiž ukazujú, že rozptyl je v skutočnosti menší a výrobok je presnejší, ako je garantované. Kto by si tu sťažoval?

Ak je výrobcom udávaný rozptyl vľavo od intervalu LP, tak spokojní nie sme, namerané hodnoty aj pri povolenej tolerancii sú vyššie než stanovené číslo.

Z uvedených možností vyplýva, že hranica P nás v podstate nezaujíma, tam sa nemá čo pokaziť. To, čo musíme ustriehnuť, je hranica L a jej vzťah k σ^2 . Aby sme neplýtvali percentami tam, kde netreba, na hranicu P celkom rezignujeme (dáme ho ako nekonečno).

Kritická oblasť – oblasť, kde sa H_0 zamietá – bude medzi 0 a L:

Tomu zodpovedá v rozdelení chí-kvadrát oblasť od u_2 do nekonečna (totož hodnoty u idú do menovateľa). Na túto kritickú oblasť vyhradíme plných 5 percent, ktoré sú stanovené v zadaní.



Po upratovaní v pojmach môžeme konečne počítať.

Pri 5% riziku bude (ako vidno aj z obrázku) $u_2 = \chi^2_{0.95,17} = 27.59$

Z toho dostávame $L = (n-1) \cdot s^2 / u_2 = 17 \cdot 0.0194761 / 27.59 = 0.012$

Výrobcom deklarované $\sigma^2 = 0.0144$ je väčšie než L , zmestí sa teda do intervalu rozptylov prijateľných na základe nameraných dát. Je to v poriadku, reklamácia výrobku sa nekoná, nulovú hypotézu nezamietame. To, že experiment ukázal vyššie hodnoty, sa dá zvaliť na náhodu.

Skúsme to teraz z inej strany. Upravíme si vzorec

$$L = (n-1) \cdot s^2 / u_2$$

takto:

$$u_2 = (n-1) \cdot s^2 / L$$

Za L rovno dosadíme hodnotu, na ktorej sa to láme, teda výrobcom deklarovaný rozptyl:

$$u_2 = 17 \cdot 0.0194761 / 0.0144 = 22.993$$

Posúdenie situácie teraz urobíme priamo v príslušnom riadku tabuľky chí-kvadrátu:

$\chi^2(v)$										
v	0,005	0,01	0,025	0,05	0,1	0,9	0,95	0,975	0,99	0,995
17	5,70	6,41	7,56	8,67	10,09	24,77	27,59	30,19	33,41	35,72

Vidíme, že 22.993 je menšie než 0.9 (menšie = mimo kritickej oblasti). Takže na hladine 0.1 (a o to skôr na požadovanej 0.05) stále nulovú hypotézu nemáme prečo zamietnuť. (To samozrejme neznamená, že vnútorne sme celkom spokojní s vlastnosťami senzorov).

Ak si pomôžeme kalkulačkou, zistíme, že nulová hypotéza sa dá zamietnuť až na hladine 0.1495 a viac.

Príklad 3:

Veľký mestský autobus s „inteligentným“ motorom by mal mať spotrebu takmer nezávislú od zvykov a zlozvykov vodiča. Výrobca deklaruje pri bežnej prevádzke rozptyl (najviac) $\sigma^2 = 0.83$. V 10-dňovom teste, kde sa na konci dňa vždy vyhodnotila spotreba, bol zistený rozptyl $s^2 = 1.51$. Na hladine významnosti 5% „preklepnite“ tvrdenia výrobcu.

Riešenie:

$$H_0: \sigma^2 = 0.83 \quad // \quad H_1: \sigma^2 > 0.83$$

Počítajme rovno „druhou cestou“:

$$u_2 = (n-1) \cdot s^2 / L$$

$$u_2 = (10-1) \cdot 1.51 / 0.83 = 16.373$$

Vidíme, že $16.373 < 16.92$, kde 16.92 zodpovedá hodnote 0.95. Preto na úrovni 5% hypotézu $\sigma^2 = 0.83$ nezamietame. (Na úrovni 0.1 by sme si však trúfli)

Neriešené príklady

Predpokladáme, že všetky náhodné veličiny, ktoré účinkujú v nasledujúcich príkladoch, majú normálne rozdelenie.

1. Basketbalisti klubu XYZ na tréningu hádžu z úctyhodnej vzdialenosti na kôš. Tréner deklaruje, že v priemere hráči trafia 7 z 10 hodov, s rozptylom najviac 1.5. Tréner konkurenčného klubu bol pozvaný na jeden z tréningov, aby si na vlastné oči overil situáciu.

Každý z 15 športovcov klubu hádzal 10-krát, počty košov boli nasledujúce:

$$x = [6 \ 7 \ 9 \ 10 \ 5 \ 9 \ 9 \ 8 \ 7 \ 10 \ 7 \ 9 \ 8 \ 10 \ 7]$$

Na hladine významnosti 2% a 1% otestujte nulovú hypotézu $\mu=7$.

Na hladine významnosti 10% a 5% otestujte nulovú hypotézu $\sigma^2=1.5$.

2. Množstvo účinnej látky ZYX v 1 tablete je výrobcom deklarovaná na 0.1 g, rozptyl deklaruje najviac 0.0007. Praktický test na 50 náhodne vybraných tabletách ukázal priemernú koncentráciu 0.093 g, s rozptylom $s^2 = 0.0009$.

a) Na akej hladine významnosti vieme zamietnuť alebo nezamietnuť nulovú hypotézu $\mu=0.1$?

b) Na akej hladine významnosti vieme zamietnuť/nezamietnuť nulovú hypotézu $\sigma^2=0.0007$?

3. Presnosť nastavenia váhy v samoobslužnej pokladni je charakterizovaná maximálnym rozptylom pri opakovanom vážení kontrolného tovaru s hmotnosťou 800 g. Ak je táto hodnota väčšia ako 20, váhu treba znova nastaviť. Kontrolné meranie dalo tieto výsledky v gramoch:

[792, 803, 790, 804, 801, 803, 798, 799, 806, 797, 802, 796, 802, 801, 798, 799, 806, 809, 797, 803]

Posúďte, či je váha v poriadku alebo treba urobiť novú kalibráciu – na hladine významnosti $\alpha=0,05$.

4. Overovalo sa nastavenie relatívneho výškomeru. 15-krát po sebe (s určitým časovým odstupom) sa merala výška bodu, ktorý je nad zemou presne 20 m. Priemer meraných výšok vyšiel 19.2 m s rozptylom 1.21. Rozhodnite, či výškomer meria správne. Hladinu významnosti si určte (rozumne) podľa vlastného uváženia.