Poznámky k předmětu Aplikovaná statistika, 9.téma

Princip testování hypotéz, jednovýběrové testy

V minulé hodině jsme si ukázali, jak sestavit intervalové odhady pro některé číselné charakteristiky normálního rozdělení. V praxi nás ale mohou zajímat i další věci. Dnes si ukážeme, na jakém principu je postaveno testování hypotéz. Uvažujme náhodný výběr $X_1, \dots, X_n$. Předpokládejme, že v rámci experimentu jsme si udělali jakousi představu o některých aspektech. Označme tuto naši hypotézu H_0. Tuto **nulovou hypotézu** budeme chtít nyní otestovat. Nulová hypotéza může být velice různorodá. Například můžeme zkoumat, zda střední hodnota náhodného výběru odpovídá nějaké hypotetické hodnotě μ_0, nebo že náhodný výběr pochází z nějakého konkrétního rozdělení. Z podstaty testování hypotéz, kterou si vysvětlíme níže na konkrétním případě, vyplývá nutnost testovat hypotézu H_0 oproti nějaké alternativě. Nelze provádět test, není-li tato **alternativní hypotéza** H_1 stanovena. Rozhodovací kritérium, tj. jestli bude nulová hypotéza zamítnuta či nikoliv, závisí právě na alternativní hypotéze. Zamítneme-li nulovou hypotézu, je to vždy ve prospěch právě alternativy. Na druhou stranu, princip testování nám nedovoluje nulovou hypotézu přijmout. Můžeme ji pouze **nezamítnout**. To v podstatě znamená, že data nejsou dostatečně průkazná proti hypotéze H_0 (lidově řečeno, neříkáme tak ani tak). Při testování nám můžou nastat tyto čtyři možnosti

1. Hypotéza H_0 ve skutečnosti neplatí a test ji zamítá.
2. Hypotéza H_0 ve skutečnosti platí a test ji nezamítá.
3. Hypotéza H_0 ve skutečnosti platí a test ji zamítá.
4. Hypotéza H_0 ve skutečnosti neplatí a test ji nezamítá.

První dvě situace jsou naprosto v pořádku. Druhé dvě možnosti již ovšem ne. Situaci 3. říkáme **chyba prvního druhu** a situaci 4. **chyba druhého druhu**. Pravděpodobnosti těchto chyb se značí

$$\begin{aligned} \alpha &= \mathbb{P}(H_0 \text{ zamítáme} | H_0 \text{ platí}), \\ \beta &= \mathbb{P}(H_0 \text{ nezamítáme} | H_0 \text{ neplatí}). \end{aligned}$$

Samozřejmě bychom si přáli, aby pravděpodobnosti chyb prvního a druhého druhu byly co nejmenší, ale toho v praxi nelze obecně dosáhnout. Proto se pravděpodobnost chyby prvního druhu α, tzv. **hladina testu**, pevně volí (zpravidla $\alpha = 1$ % nebo $\alpha = 5$ %). Hodnota $1 - \beta$ se nazývá **síla testu**. Snaha snížit hladinu testu α s sebou přináší zvětšení chyby β. To se dá eliminovat zvýšením rozsahu výběru n. Rozhodovací kritérium, resp. princip testování si ukážeme na testu hypotézy o střední hodnotě.

Odvození rozhodovacího pravidla pro testování střední hodnoty

Uvažujme náhodný výběr $X_1, \dots, X_n$ z rozdělení $\mathcal{N}(\mu, \sigma^2)$, kde $\sigma^2 > 0$ je známé. Chceme otestovat, zda střední hodnota je rovna dané hypotetické hodnotě μ_0 oproti alternativě, že tomu tak není. Tedy

$$\begin{aligned} H_0: &\quad \mu = \mu_0 \\ H_1: &\quad \mu \neq \mu_0 \end{aligned}$$

V minulé hodině jsme se dozvěděli, že výběrový průměr je nejlepším nestranným bodovým odhadem střední hodnoty μ a také víme, že $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$, respektive že

$$\frac{\bar{X} - \mu}{\sigma}\sqrt{n} \sim \mathcal{N}(0, 1).$$

Intuitivně H_0 zamítneme, jestliže hodnota $\bar{X}$ bude hodně daleko od hypotetické hodnoty μ_0, tj. když $|\bar{X} - \mu_0| > k$. Hodnotu k pak při pevně zvolené hladině testu určíme následovně:

$$\alpha = \mathbb{P}\left(|\bar{X} - \mu_0| > k | H_0 \text{ platí}\right) = \mathbb{P}\left(\frac{|\bar{X} - \mu_0|}{\sigma}\sqrt{n} > \frac{k\sqrt{n}}{\sigma} \middle| H_0 \text{ platí}\right).$$

Potom při platnosti H_0 dostáváme rovnost

$$\mathbb{P}\left(\frac{|\bar{X} - \mu_0|}{\sigma}\sqrt{n} > u_{1-\frac{\alpha}{2}}\right) = \alpha,$$

a odtud

$$\frac{k\sqrt{n}}{\sigma} = u_{1-\frac{\alpha}{2}}.$$

Tedy hypotézu H_0 zamítneme, bude-li

$$\left|\frac{\bar{X} - \mu_0}{\sigma}\sqrt{n}\right| > u_{1-\frac{\alpha}{2}}.$$

Funkce $R(X_1, \ldots, X_n) = R = \frac{\bar{X}-\mu_0}{\sigma}\sqrt{n}$ se nazývá **testová statistika**. Obor hodnot testové statistiky, při kterých zamítáme H_0, se nazývá **kritický obor** $W_\alpha := \{R : |R(X_1, \ldots, X_n)| > u_{1-\frac{\alpha}{2}}\}$.

Poznámka 1. *Povšimněme si, jak velice podobná je předchozí úvaha odvození oboustranného $(1-\alpha)\%$ intervalového odhadu pro střední hodotu. Ve skutečnosti jsou tyto dvě metody naprosto ekvivalentní. To znamená, že naši hypotézu zamítneme právě tehdy, když hodnota μ_0 spadne* ***mimo*** *$(1-\alpha)\%$ oboustranný interval spolehlivosti pro střední hodnotu.*

Analogicky lze odvodit i tzv. jednostranné hypotézy, tj. hypotézy, kde alternativa H_1 je tvaru

$$H_1 : \mu > \mu_0$$

nebo

$$H_1 : \mu < \mu_0.$$

Tvar testové statistiky a kritické obory naleznete v tabulkách. Pro nás je nyní podstatná následující úvaha.

V předchozích odstavcích jsme se dozvěděli, že nulovou hypotézu nemůžeme přijmout, ale můžeme ji zamítnout a to **vždy** ve prospěch alternativní hypotézy. Uvažujme nyní, že chceme například zkoumat průměrnou výšku mužů v České republice. Na základě náhodného výběru máme podezření, že tato hodnota je alespoň 175 cm. Bohužel nulovou hypotézu $H_0 : \mu \geq 175$ by se nám nikdy potvrdit nepodařilo. Na druhou stranu, pokud hypotézy stanovíme následovně:

$$H_0 : \mu = 175$$
$$H_1 : \mu > 175,$$

potom podaří-li se nám zamítnout H_0, je to už ve prospěch naší původní hypotézy. Tedy se nám podaří **statisticky významně na hladině** α prokázat hypotézu o tom, že očekávaná výška mužů v naší republice je větší než 175 cm. Obdobný postup se tedy v takových případech používá.

Shrňme si nyní obecný postup při testování hypotéz:

- stanovení cíle testování $\Rightarrow$ stanovení H_0 a H_1
- spočtení testové statistiky (její tvar závisí na tom, co testujeme)
- stanovení kritického oboru a porovnání statistiky R s kritickým oborem
- závěr: rozhodnutí H_0 zamítáme ve prospěch H_1 nebo H_0 nezamítáme

Test, na kterém jsme si odvodili princip testování hypotéz byl **jednovýběrový test o střední hodnotě při známém** σ^2. V praxi se nám ale moc často nestane, že bychom znali rozptyl náhodného výběru. V takovém případě se používá tzv. **jednovýběrový t–test o střední hodnotě**. V něm je hodnota σ^2 v testové statistice nahrazena S^2 a statistika R má pak za platnosti H_0 Studentovo $t-$rozdělení o $n-1$ stupních volnosti. Přesný tvar testové statistiky R a kritický obor naleznete opět v tabulkách.

Poznámka 2. *V praxi se při testování využívá tzv.* ***p–hodnota testu*** *(p-value). Tato hodnota p je dosaženou hladinou testu. Je to nejmenší hladina významnosti α, na které ještě lze hypotézu H_0 zamítnout. V dnešní době nám tuto hodnotu poskytne každý statisticky zaměřený software.*

CVIČENÍ K TOMUTO TÉMATU

(i) Rozptyl koncentrace kyseliny při výrobě byl dlouhodobě $\sigma^2 = 10$. Otestujte na hladině významnosti 5 %, zda po generální opravě výrobního zařízení je opět roven deseti, máme-li vzorky měření

x_i	86	88	90	87	85	86	84	88	89

.

Předpokládejme, že koncentrace je normálně rozdělená náhodná veličina.

(ii) Spotřeba paliva u daného typu auta je 10 l/100 km. Byla navržena úprava na snížení spotřeby a bylo vyrobeno 25 prototypů aut s touto úpravou. Automobilka chce úpravu zavést sériově pouze tehdy, když se spolehlivě (na hladině významnosti 5 %) prokáže snížení spotřeby paliva u upravených aut. Předpokládá se normální rozdělení spotřeby paliva. Rozhodněte, zda má automobilka zavést úpravu sériově, pokud víte, že z naměřených hodnot spotřeby se spočetlo

$$\bar{x} = 9,3 \text{ l/100 km}, s^2 = 4.$$

(iii) Napište 95% dolní interval spolehlivosti pro střední hodnotu náhodné veličiny $X \sim \mathcal{N}(\mu, \sigma^2)$, byly-li naměřeny následující hodnoty X

x_i	11	15	13	19	7

.

Lze pouze na základě tohoto intervalu spolehlivosti tvrdit, že

(a) s 95% spolehlivostí platí

- $\mu < 9$,
- $\mu < 5$,
- $\mu > 8$,
- $\mu > 10$,

(b) s 90% spolehlivostí platí $\mu > 8$,

(c) s 99% spolehlivostí platí $\mu > 10$,

(d) lze zamítnout na hladině významnosti 5 % nulové hypotézy (ve prospěch alternativ)

- $H_0: \mu = 7,\ H_1: \mu < 7$,
- $H_0: \mu = 10,\ H_1: \mu < 10$,
- $H_0: \mu = 7,\ H_1: \mu > 7$,
- $H_0: \mu = 9,\ H_1: \mu > 9$?

(iv) Norma požaduje, aby určitý reagující roztok měl hodnotu pH = 8, 30. Modelujme hodnotu pH náhodnou veličinou s rozdělením $\mathcal{N}(\mu, \sigma^2)$, kde $\sigma = 0,02$. Na základě získaných měření

pH	8,34	8,31	8,30	8,33	8,32

rozhodněte na hladině významnosti 5 %, zda je hodnota pH větší, než požaduje norma.

(dcv) Bylo měřeno množství odpadu (v %) při výrobě jisté součástky

x_i v %	4,1	4,0	3,8	3,9	3,8	3,8	3,5	3,7	4,0	4,0

.

Na hladině významnosti 5 % testujte následující hypotézy o průměrném odpadu při výrobě součástky

(a) $H_0: \mu = 3,9,\ H_1: \mu > 3,9$,

(b) $H_0: \mu = 3,8,\ H_1: \mu \neq 3,8$,

(c) $H_0: \mu = 4,\ H_1: \mu < 4$.

Dvouvýběrové testy, párový test

Teď už tedy víme, na jakém principu funguje testování hypotéz a uvedli jsme si základní testy pro číselné charakteristiky normálního rozdělení. Vraťme se k příkladu z minulého odstavce. Uvažujme nyní muže a ženy v České republice a podívejme se blíže na rozdělení výšky pro obě pohlaví. Asi nikdo z nás nepředpokládá, že by očekávaná výška byla pro obě pohlaví stejná. Kdybychom si ale chtěli podobnou hypotézu statisticky ověřit, potřebovali bychom porovnat obě střední hodnoty μ_1 (muži) a μ_2 (ženy). Nejjednodušším testem shody středních hodnot je tzv. **dvouvýběrový t–test**.

Mějme $X_1, \ldots, X_n$ náhodný výběr z $\mathcal{N}(\mu_1, \sigma_1^2)$ a $Y_1, \ldots, Y_m$ náhodný výběr z $\mathcal{N}(\mu_2, \sigma_2^2)$ a předpokládejme, že tyto výběry jsou na sobě nezávislé. Hypotézu $H_0: \mu_1 = \mu_2$ můžeme ekvivalentně formulovat ve tvaru $H_0: \mu_1 - \mu_2 = 0$.

Dvouvýběrový t–test shody středních hodnot při známých σ_1^2 a σ_2^2

Z vlastností normálního rozdělení plyne, že $\bar{X}$ a $\bar{Y}$ mají také normální rozdělení a tudíž i veličina $\bar{X}-\bar{Y}$ bude mít normální rozdělení. Potom (po znormování a za platnosti H_0) má testová statistika

$$R = \frac{\bar{X}-\bar{Y}}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}}$$

normované normální rozdělení. Tímto rozdělením se řídí kritický obor (viz tabulky).

Dvouvýběrový t–test shody středních hodnot při neznámých rozptylech

Neznáme-li rozptyly jednotlivých rozdělení (což praxi neznáme v podstatě nikdy), je třeba nahradit tyto rozptyly jejich bodovými odhady S_1^2 a S_2^2. Potom má (opět po znormování a za platnosti H_0) testová statistika

$$R = \frac{\bar{X}-\bar{Y}}{\sqrt{\frac{S_1^2}{n}+\frac{S_2^2}{m}}}$$

Studentovo rozdělení $t(l)$, kde počet stupňů volnosti l se získá ze vztahu

$$l = \frac{\left(\frac{S_1^2}{n}+\frac{S_2^2}{m}\right)^2}{\frac{1}{n-1}\frac{S_1^4}{n^2}+\frac{1}{m-1}\frac{S_2^4}{m^2}}.$$

Toto číslo není (skoro nikdy) celé, proto se buď zaokrouhlí na nejbližší celé číslo, nebo můžeme za kvantil volit průměr mezi kvantilem se stupni volnosti rovnými nejbližšímu menšímu celému číslu a kvantilem se stupni volnosti rovnými nejbližšímu většímu celému číslu. Bližší detaily o kritickém oboru opět naleznete v tabulce.

Test shody rozptylů

Chceme-li u dvou výběrů testovat shodu rozptylů (například k určení, zda jsou dva přístroje stejně citlivé), lze nulovou hypotézu $H_0: \sigma_1^2 = \sigma_2^2$ formulovat jako $H_0: \frac{\sigma_1^2}{\sigma_2^2} = 1$. Bodové odhady S_1^2 a S_2^2 mají χ^2–rozdělení, a proto (za platnosti H_0) bude statistika

$$R = \frac{S_1^2}{S_2^2}$$

mít Fisherovo–Snedecorovo rozdělení o $n-1, m-1$ stupních volnosti.

Poznámka 3. *V literatuře (včetně skrip Pavlík a kol.) lze nalézt i test shody středních hodnot pro případ shodných, ale neznámých rozptylů $\sigma_1^2 = \sigma_2^2$. Testová statistika je pak tvaru*

$$R = \frac{\bar{X}-\bar{Y}}{S_{12}}\sqrt{\frac{n\,m}{n+m}},$$

kde

$$S_{12} = \sqrt{\frac{(n-1)S_1^2+(m-1)S_2^2}{n+m-2}},$$

a má za platnosti H_0 Studentovo rozdělení s $n+m-2$ stupni volnosti. V praxi je ovšem jen velmi zřídka předpoklad o shodě rozptylů dostatečně opodstatněný. Dříve se často dělal předběžný test *shody rozptylů, a v případě nezamítnutí hypotézy o shodě se používal právě test z této poznámky. Tento postup ovšem nedoporučujeme, neboť nezamítnutí hypotézy není (jak již víme ze začátku této kapitoly) ekvivalentní jejímu přijetí a podobným postupem už bychom nemohli zajistit dosažení předepsané hladiny α a navíc bychom mohli výrazně ovlivnit sílu testu $1-\beta$. Je tedy lépe v takových případech užít testu pro obecně neznámé rozptyly i za cenu toho, že častěji nezamítneme.*

Předtím, než se podíváme na poslední test této kapitoly, musíme ještě zavést pojem dvourozměrného normálního rozdělení.

Definice 1. Náhodný vektor $\mathbb{X} = (X_1, X_2)$ se střední hodnotou $\mu = (\mu_1, \mu_2)$ a kovarianční maticí

$$\operatorname{var}(X_1, X_2) = \begin{pmatrix} \sigma_1^2 & , & \operatorname{cov}(X_1, X_2) \\ \operatorname{cov}(X_1, X_2) & , & \sigma_2^2 \end{pmatrix}$$

má **dvourozměrné normální rozdělení**, jestliže pro všechna $a, b \in \mathbb{R}$ má náhodná veličina $Y = aX_1 + bX_2$ normální rozdělení $\mathcal{N}(a\mu_1 + b\mu_2, \sigma^2)$, kde $\sigma^2 = a^2\sigma_1^2 + b^2\sigma_2^2 + 2ab\operatorname{cov}(X_1, X_2)$.

Párový t–test

Příklad 1. *U pěti náhodně vybraných pacientů byl naměřen krevní tlak před (x_i) a po (y_i) podání nového léku. Údaje jsou uvedeny v následující tabulce Rádi bychom ověřili, že podání tohoto léku*

i	1	2	3	4	5
x_i	105	99	109	97	115
y_i	115	103	101	108	121

má vliv na krevní tlak pacientů, jinými slovy řečeno, zda se střední hodnota tlaku před i po podání léku mění.

Jak je vidno, máme tu opět dva výběry dat, ale je zřejmé, že v tomto případě nemůžeme uvažovat jednotlivé výběry za nezávislé. Ba naopak, závislost dvojic hodnot v $i-$tém sloupci je zcela zřejmá. Dvouvýběrový t–test je tedy absolutně nevhodný. Předpokládejme, že dvourozměrný náhodný výběr $(X_1, Y_1), \ldots, (X_n, Y_n)$ (tj. množina nezávislých stejně rozdělených vektrů) pochází z dvourozměrného normálního rozdělení a $X_1, \ldots, X_n$ mají střední hodnotu μ_1 a rozptyl σ_1^2 a veličiny $Y_1, \ldots, Y_n$ mají střední hodnotu μ_2 a rozptyl σ_2^2. Definujme nyní veličiny

$$Z_i = X_i - Y_i,$$

potom Z_i mají normální rozdělení se střední hodnotou $\mu_Z = \mu_1 - \mu_2$ a výběrový rozptylem $S_Z^2 = \frac{1}{n-1}\sum_{i=1}^n (Z_i - \bar{Z})^2$. Testování hypotézy $H_0 : \mu_1 = \mu_2$ je ekvivalentní testování hypotézy $H_0 : \mu_Z = 0$ a pro tuto hypotézu lze již použít jednovýběrového t–testu o střední hodnotě. Testová statistika má pak tvar (dosaď $\mu_0 = 0$)

$$R = \frac{\bar{Z}}{S_Z}\sqrt{n}.$$

Kritický obor naleznete v tabulkách.

Poznámka 4. *Pro všechny uvedené testy byl využit předpoklad, že náhodný výběr pochází z normálního rozdělení. Co ale v případě, že se nejedná o výběr z normálního rozdělení (což v praxi ne vždycky můžeme předpokládat)? Vzhledem k tomu, že všechny uvedené testové statistiky byly založeny na náhodných veličinách* $\bar{X}$ *nebo* S^2 *a tyto veličiny lze interpretovat jako součet nezávislých veličin, můžeme díky* ***centrální limitní větě*** *předpokládat* ***asymptotickou normalitu*** *veličin* $\bar{X}$ *a* S^2 *a pro dostatečně široký rozsah výběru (CLV funguje dobře už od relatvně malého* n*) použít zmiňované testy jako* ***asymptotické****.*

<u>CVIČENÍ K TOMUTO TÉMATU</u>

(i) Náhodně se vybrala prasata a rozdělila do dvou skupin. Každá skupina byla krmena půl roku jinou dietou. Na závěr se zjistily váhové přírůstky prasat (v kilogramech)

1. dieta	62	54	55	60	53	58
2. dieta	52	56	49	50	51	

Předpokládejme normalitu váhy prasat. První dieta je nákladnější než druhá, a proto se výkrmna ptá, zdali je první dieta statisticky významně účinnější než druhá (testujte na hladině 5 %).

(ii) Firma provozující billboardy se rozhodla otestovat, zdali má nakoupit dražší či levnější barvu podle toho, jak zůstane zachována. Náhodně se vybralo 10 billboardů, 5 z nich se natřelo dražší barvou a zbytek levnější. Po čtvrt roce se zjistilo procento zachovalé barvy na každém billboardu

dražší barva	89	89	90	84	88
levnější barva	85	87	92	80	84

Předpokládáme normální rozdělení zachování barvy se stejným neznámým rozptylem. Dražší barva se nakoupí, pokud bude spolehlivě (na hladině 10 %) prokázáno, že je stálejší. V opačném případě se pořídí levnější. Rozhodněte, kterou barvu má firma koupit.

(iib) Uvažte stejné zadání jako v příkladu (ii), jen s tím rozdílem, že se vybralo pouze 5 bilboardů a každý z nich se natřel z poloviny dražší barvou a z poloviny levnější.

(iii) Laboratoř chce koupit jistý měřící přístroj. Má na výběr ze dvou možností. První přístroj lze obstarat hned, druhý až za rok. Rozhodovací strategie je následující. Druhý přístroj se koupí pouze tehdy, ukáže-li se spolehlivě ($\alpha = 10$ %), že je přesnější než první. Laboratoř má k dispozici 4 měření jistého vzorku prvním přístrojem a 5 měření téhož vzorku druhým přístrojem

1. přístroj	610	580	620	630	
2. přístroj	635	625	640	620	630

Poraďte laboratoři (na základě výsledků měření), má-li s koupí přístroje počkat, či nikoli.

(iv) Zkoumá se, zdali se u daného typu auta sjíždí pravá a levá zadní pneumatika stejně, na hladině významnosti 10 %. Náhodně se vybralo 6 aut, se kterými se jezdilo půl roku na stejných pneumatikách. Změřilo se ojetí zadních pneumatik (v mm)

auto	1	2	3	4	5	6
levá	1,8	1,0	2,2	0,9	1,5	1,6
pravá	1,5	0,9	2,0	1,1	1,0	1,4

Dále zjistěte (se spolehlivostí 90 %), zdali podhuštění levé zadní pneumatiky způsobuje větší ojetí, tj. ptáme se, zdali se levá zadní pneumatika sjíždí více, než pravá?

(v) Ve dvou cementárnách se provedla kontrola dávkovačů. Zvážily se náhodně vybrané pytle cementu z obou cementáren (jejich hmotnost uvedena v kg)

1. cementárna	50	51	48	50	51
2. cementárna	49	46	52		

.

Předpokládejme normalitu váhy pytlů cementu. Otestujte (volte hladinu 20 %), zdali dávkovače v obou cementárnách dávkují stejně.

(vi) Laboratoř má dvě elektronické váhy. Chce otestovat (na hladině 10 %), zda není systematická odchylka mezi měřeními hmotnosti první a druhou váhou. Jedno závaží se zvážilo šestkrát jednou váhou, a pak šestkrát druhou váhou

1. váha	$0,99$	$0,98$	$1,03$	$0,95$	$1,00$	$0,99$
2. váha	$1,05$	$0,97$	$1,00$	$1,02$	$1,00$	$0,96$

.

(dcv) Na hladině významnosti 5 % rozhodněte, zdali jsou dvě analytické metody srovnatelné z hlediska

(a) přesnosti,

(b) správnosti

výsledků měření, máte-li k dispozici data

metoda A	$3,28$	$3,28$	$3,29$	$3,29$
metoda B	$3,25$	$3,27$	$3,26$	$3,25$

.

V případě (b) nadto určete přibližně dosaženou hladinu testu.