Testování hypotéz

13. cvičení

Jednovýběrový test o střední hodnotě

Provádíme průzkum, jaký skutečný objem piva točí v nejmenované hospodě. Zakoupeno bylo 10 piv a jejich objem byl (v litrech):

$$0.510,\ 0.462,\ 0.491,\ 0.466,\ 0.451,\ 0.503,\ 0.475,\ 0.487, 0.512, 0.505.$$

Předpokládejte, že natočený objem piva se řídí normálním rozdělením a jednotlivá měření jsou nezávislá. Testy provádějte na hladině $\alpha = 0.05$.

1. Odhadněte střední hodnotu objemu natočeného piva. Dále odhadněte, s jakou pravděpodobností dostaneme „podmírák".
2. Otestujte, zda se skutečný objev piva liší od požadované hodnoty 0.5 litru. Z předchozí zkušenosti víme, že směrodatná odchylka objemů piv natočených tímto hostinským je rovna 0.02.
   - Spočítejte hodnotu vhodné testové statistiky a porovnejte ji s příslušným kvantilem.
   - Výsledek testu řádně interpretujte.
3. Z pohledu zákazníka by nás ale spíše zajímalo, zda hostinský netočí pod míru. Chceme proto provést test proti jednostranné alternativě. Jaký je náš závěr?
4. Otestujte, zda se skutečný objev piva liší od požadované hodnoty 0.5 litru. Předpokládejte, že σ^2 neznáme.
   - Spočítejte hodnotu vhodné testové statistiky a porovnejte ji s příslušným kvantilem.
   - Výsledek testu řádně interpretujte a porovnejte s výsledkem v úkolu 2.
5. Otestujte, zda hostinský netočí pod míru (pro σ^2 neznámé). Výsledek porovnejte se závěrem předchozího bodu.
6. Porovnejte graficky hustoty rozdělení $\mathsf{N}(0, 1)$ a t-rozdělení s n-stupni volnosti pro $n = 2, 5, 10$ a 50.
   - Co lze říci o chování t-rozdělení pro $n \to \infty$?
   - Porovnejte kvantily výše uvedených rozdělení na hladinách 0.95, 0.975 a 0.99.
7. Proveďte testy v bodech 4. a 5. pomocí funkce `t.test`. Ve výstupu si mimo jiné všimněte intervalu spolehlivosti a porovnejte jej se vzorečkem, který znáte z přednášky.

Posouzení normality dat

8. Budeme se zabývat předpokladem, že data pochází z normálního rozdělení.
   (a) Vykreslete si histogram objemů piva. Uvažte, zda rozdělení vypadá jako normální.
   (b) Jelikož pro takto malý rozsah výběru ($n = 10$) není histogram velmi vypovídající, je lepší posoudit normalitu na základě tzv. QQ-diagramu. Jaký je náš závěr na základě tohoto obrázku?

Testování hypotéz

- ověřování platnosti nějakého výroku
- rozhodujeme na základě dat a statistického testu

Hypotéza je výrok, o jehož platnosti chceme rozhodnout na základě nasbíraných dat:
- nulová hypotéza H_0,
- alternativní hypotéza H_1.

Statistický test
= rozhodovací pravidlo,
- charakterizován testovou statistikou a kritickým oborem.

Možná rozhodnutí:
- **zamítáme** H_0 ve prospěch H_1 (*naše data svědčí proti* H_0, **prokazujeme** *platnost* H_1)
- **nezamítáme** H_0 (*na základě našich dat nelze* H_0 *zamítnout, naše data nejsou v rozporu s* H_0)

Můžeme se dopustit **chyby**:

	Naše rozhodnutí	
Skutečnost	zamítáme H_0	nezamítáme H_0
H_0 platí	chyba 1.druhu	OK
H_0 neplatí	OK	chyba 2.druhu

Chyba 1.druhu je závažnější (falešně něco prokazujeme) $\rightarrow$ její pravděpodobnost musíme kontrolovat:
- volíme **hladinu testu** α = maximální přípustná pravděpodobnost chyby 1.druhu (většinou $\alpha = 0.05$) $\rightarrow$ test na hladině α,
- $\beta = \mathsf{P}(\text{chyba 2. druhu})$
- $1-\beta$ = síla testu = $\mathsf{P}(\text{prokázání platné } H_1)$ – obecně o ní moc nevíme (můžeme ovlivnit volbou testu, počtu měření ...),

P-hodnota (angl. p-value) = dosažená hladina testu
- výsledek testů ve statistických softwarech,
- pravděpodobnost, že dostaneme výsledek, který stejně nebo ještě více svědčí proti H_0, jestliže H_0 ve skutečnosti platí,
- „stupeň důvěry“ v platnost H_0
- pravidlo: p-hodnota $\leq \alpha \Rightarrow$ zamítáme H_0 a prokazujeme tak H_1

Interval spolehlivosti

- interval s náhodnými mezemi, který pokryje neznámý parametr s předepsanou pravděpodobností $1-\alpha$
- oboustranný nebo jednostranný (levostranný, pravostranný)

Jednovýběrové testy o střední hodnotě:

Situace: $X_1, \dots, X_n$ je náhodný výběr z rozdělení se střední hodnotou μ a rozptylem $\sigma^2 \in (0, \infty)$. Chceme testovat

$$H_0 : \mu = \mu_0 \quad \text{proti} \quad H_1 : \mu \neq \mu_0,$$

kde $\mu_0 \in \mathbb{R}$ je nějaké dané číslo.

(a) Jestliže mají X_i **normální** rozdělení a rozptyl σ^2 **známe**, pak hypotézu H_0 zamítneme na hladině α, jestliže

$$\sqrt{n}\frac{|\overline{X} - \mu_0|}{\sigma} \geq q_{1-\alpha/2},$$

kde $q_{1-\alpha/2}$ je kvantil $\mathsf{N}(0,1)$ na hladině $1 - \alpha/2$.

(b) Jestliže mají X_i **normální** rozdělení a rozptyl σ^2 **neznáme**, pak hypotézu H_0 zamítneme na hladině α, jestliže

$$\sqrt{n}\frac{|\overline{X} - \mu_0|}{S_n} \geq t_{n-1}(1 - \alpha/2),$$

kde S_n^2 je výběrový rozptyl a $t_{n-1}(1 - \alpha/2)$ je kvantil t-rozdělení s $n - 1$ stupni volnosti na hladině $1 - \alpha/2$.

(c) Je-li n dostatečně velké (např. $n \geq 50$), pak pro **obecně rozdělené** X_i zamítneme hypotézu H_0 na přibližné (asymptotické) hladině α, jestliže

$$\sqrt{n}\frac{|\overline{X} - \mu_0|}{S_n} \geq q_{1-\alpha/2}.$$

Poznámka: Pro velké n máme $q_{1-\alpha/2} \doteq t_{n-1}(1 - \alpha/2)$.

Podobně lze konstruovat i testy proti **jednostranné alternativě**. Např. pro test

$$H_0 : \mu = \mu_0 \quad \text{proti} \quad H_1 : \mu > \mu_0$$

dostaneme v (a) kritický obor

$$\sqrt{n}\frac{\overline{X} - \mu_0}{\sigma} \geq q_{1-\alpha}$$

apod. pro další situace nebo opačné alternativy.

Podobně jako uvedené kritické oborů lze konstruovat **intervaly spolehlivosti** pro μ. Např. v (b) máme oboustranný interval spolehlivosti pro μ na hladině $1 - \alpha$

$$\left(\overline{X} - \frac{S}{\sqrt{n}} t_{n-1}(1 - \alpha/2), \overline{X} + \frac{S}{\sqrt{n}} t_{n-1}(1 - \alpha/2) \right).$$

Párový test:

Situace: $(X_1, Y_1)' \dots, (X_n, Y_n)'$ je náhodný výběr z dvourozměrného rozdělení, kde chceme testovat

$$H_0 : \mathsf{E}X = \mathsf{E}Y \quad \text{proti} \quad H_1 : \mathsf{E}X \neq \mathsf{E}Y$$

(resp. proti jednostranné H_1).

Tuto situaci snadno převedeme na předchozí případ zavedením $Z_i = X_i - Y_i$ a použijeme na Z_i jednovýběrový test s $\mu_0 = 0$.

Filozofii testování hypotéz

lze přirovnat k soudnímu procesu, který ctí presumpci neviny:

H_0: „Obžalovaný je nevinen.“ a H_1: „Obžalovaný je vinen.“

Potom

- zamítneme-li H_0, odsoudíme obžalovaného k trestu,
- nezamítneme-li H_0, je obžalovaný propuštěn,
- data = důkazy svědčící pro nebo proti vině,
- test = soudce,
- hladina testu α = pravděpodobnost odsouzení nevinného,
- síla testu $1 - \beta$ = pravděpodobnost odsouzení skutečného pachatele.

Hladinu α zvolíme malou (např. 5 %), abychom chránili nevinné před odsouzením. Je zřejmé, že chyba 1. druhu (odsouzení nevinného) je závažnější než chyba 2. druhu (propuštění zločince).

Výsledek procesu může být následující:

- Odsouzení obžalovaného. Pak lze tvrdit, že je obžalovaný skutečně vinen.
- Propuštění obžalovaného. To může být buď z důvodu, že je opravdu nevinen, anebo je vinen, ale soud neměl dostatečné důkazy k jeho odsouzení.