Parametrické testy hypotéz o středních hodnotách spojitých náhodných veličin

© EuroMISE Centrum

I. ÚVOD

- V této přednášce budeme hovořit o jednovýběrových a dvouvýběrových testech týkajících se střední hodnoty (teoretického průměru) spojité náhodné veličiny, jakou je například porodní váha novorozenců.
- Ve všech případech budeme vycházet z předpokladu, že sledovaná náhodná veličina má normální rozdělení N(μ,σ^2).

Obecné informace

Kontakt:

EuroMISE centrum
Doc. Zdeněk Valenta, Ph.D.
Tel.: 266 053 640 (sekretariát)
Fax: 286 581 453
http://www.euromise.cz
valenta@euromise.cz

Literatura:

Zvárová, J.: *Základy statistiky pro biomedicínské obory I. Vydavatelství Karolinum, UK Praha 2001*

Zvára, K.: *Biostatistika. Vydavatelství Karolinum, UK Praha 2001*

Rosner, B.: *Fundamentals of Biostatistics, 4th Edition Duxbury Press, International Thomson Publishing Company, 1995*

II. Formalismus Testů Hypotéz

- Formalismus testů hypotéz začíná vyslovením medicínské hypotézy, která poskytuje věcný základ formulaci statistické nulové hypotézy (H_0), zde hypotézy o střední (průměrné) hodnotě sledované náhodné veličiny (tj. parametru μ). Vedle nulové hypotézy je dále třeba zformulovat tzv. alternativní hypotézu (H_1), která popisuje, jaká situace nastává v případě neplatnosti H_0.
- Dalším krokem je zvolení hladiny pravděpodobnosti chyby I. druhu, kterou jsme ochotni v našem testu tolerovat. Značíme ji obvykle α a nazýváme ji také hladinou významnosti testu („*significance level*“ *of a test*). Jedná se o pravděpodobnost, s jakou zamítneme nulovou hypotézu v H_0 případě, že platí. Obvyklou volbou hladiny významnosti je α = 0.05, případně 0.01 nebo 0.001.

II. Formalismus Testů Hypotéz (pokr.)

- Dalším krokem je volba testové statistiky. Jedná se o náhodnou veličinu, která je navržena tak, abychom uměli určit její statistické rozdělení, např. N(0,1).
- Hodnocení statistické hypotézy vychází z dosažené hodnoty testové statistiky, kterou porovnáváme s tzv. *kritickou hodnotou* odpovídajícího rozdělení na hladině α. Překročí-li hodnota testové statistiky kritickou hodnotu, zamítáme H_0 na hladině významnosti α.

II. Formalismus Testů Hypotéz (pokr.)

- Ekvivalentně můžeme posoudit výsledek testu na základě tzv. *dosažené hladiny významnosti* (*p-value, p*), která udává pravděpodobnost, s jakou bychom za předpokladu platnosti nulové hypotézy mohli obdržet výsledek přinejmenším tak extrémní, jako je výsledek pozorovaný.
- Má-li dosažená hladina významnosti *p* hodnotu nižší než je zvolená hladina významnosti testu α, pak obdržený výsledek je významný na této hladině významnosti a zamítáme nulovou hypotézu H_0.

III. Jednovýběrový *z*-test při známé hodnotě σ

- Řekněme, že máme k dispozici dlouhodobé údaje o porodní váze novorozenců v České republice, na jejichž základě jsme zjistili, že průměrná porodní váha novorozenců v ČR má hodnotu μ_0 = 3.2 kg a směrodatná odchylka má hodnotu σ_0 = 0.6 kg. V tomto případě můžeme předpokládat, že porodní váha novorozenců v ČR se řídí normálním rozdělením, tedy $N(\mu_0,\sigma_0^2) = N(3.2, 0.36)$.
- Data získaná z místní spádové nemocnice ukazují, že průměrná porodní váha 100 novorozenců zde narozených v průběhu posledních tří let má hodnotu 2.8 kg.

III. Jednovýběrový *z*-test při známé hodnotě σ

Motivace: naším cílem bude testovat hypotézu, zda je průměrná porodní váha novorozenců narozených v místní spádové nemocnici v průběhu posledních tří let nižší než hodnota dlouhodobého republikového průměru.

- H_0: $\mu = \mu_0$, přičemž $\sigma = \sigma_0$.
- H_1: $\mu < \mu_0$, přičemž $\sigma = \sigma_0$.
- Zvolme hladinu významnosti α
- Testová statistika: $z = \dfrac{\bar{x} - m_0}{s / \sqrt{n}} \sim N(0,1) \mid H_0$
- Testové kritérium: zamítneme nulovou hypotézu na hladině α v případě, že platí $z < z_\alpha$, nebo ekvivalentně v případě, že dosažená hladina významnosti *p* jednostranného testu bude menší než α.

III. Jednovýběrový *z*-test při známé hodnotě σ

Provedení testu:

- H_0: μ = 3.2, přičemž σ = 0.6.
- H_1: μ < 3.2, přičemž σ = 0.6.
- Hladina významnosti α = 5%.
- Testová statistika: $z = \frac{\bar{x}-m_0}{s/\sqrt{n}} = \frac{2.8-3.2}{0.6/\sqrt{100}} = -6.667 \sim N(0,1)$
- Testové kritérium: $-6.667 < z_{0.05} = -1.645$

  Alternativně, dosažená hladina významnosti má hodnotu p = $3.8 \cdot 10^{-8} < 0.05$.
- Závěr: Zamítáme nulovou hypotézu, že průměrná porodní váha novorozenců v místní spádové nemocnici se neliší od celostátního průměru České republiky, ve prospěch H_1.

IV. Jednovýběrový *t*-test při neznámé hodnotě σ

Motivace: testujme hypotézu, že průměrná porodní váha novorozenců narozených v místní spádové nemocnici v průběhu posledních tří let nižší než hodnota dlouhodobého republikového průměru.

- H_0: μ = μ_0, přičemž σ odhadujeme pomocí *s*.
- H_1: μ < μ_0, přičemž σ odhadujeme pomocí *s*.
- Zvolme hladinu významnosti α
- Testová statistika: $t = \frac{\bar{x}-m_0}{s/\sqrt{n}} \sim t_{n-1} \mid H_0$
- Testové kritérium: zamítneme nulovou hypotézu na hladině α v případě, že platí $t < t_{n-1}(\alpha)$, nebo ekvivalentně v případě, že dosažená hladina významnosti jednostranného testu bude menší než α.

IV. Jednovýběrový *t*-test při neznámé hodnotě σ

- Pokračujme v našem příkladě o porodní váze novorozenců v České republice, ve kterém jsme uvedli, že průměrná porodní váha novorozenců v ČR má hodnotu μ_0 = 3.2 kg. Předpokládejme ovšem, že směrodatná odchylka není známa, což je v praxi nejčastější případ. Hodnota *σ* musí tedy být odhadována z dat pomocí výběrové směrodatné odchylky *s* a modifikovaná testová statistika vede v tomto případě na studentovo *t*-rozdělení.

- Předpokládejme tedy, že data získaná z místní spádové nemocnice ukazují, že průměrná porodní váha 100 novorozenců zde narozených v průběhu posledních tří let má hodnotu 2.8 kg a odhad směrodatné odchylky σ má hodnou *s* = 0.85 kg.

III. Jednovýběrový *t*-test při neznámé hodnotě σ

Provedení testu:

- H_0: μ = 3.2, přičemž odhad σ má hodnotu s = 0.85.
- H_1: μ < 3.2, přičemž odhad σ má hodnotu s = 0.85.
- Hladina významnosti α = 5%.
- Testová statistika: $t = \frac{\bar{x}-m_0}{s/\sqrt{n}} = \frac{2.8-3.2}{0.85/\sqrt{100}} = -4.706 \sim t_{99}$
- Testové kritérium: $-4.706 < t_{99}(0.05) = -1.66$

  Alternativně, dosažená hladina významnosti má na základě *t*-rozdělení hodnotu p = $4.1 \cdot 10^{-6} < 0.05$.
- Závěr: Zamítáme nulovou hypotézu, že průměrná porodní váha novorozenců v místní spádové nemocnici se neliší od celostátního průměru České republiky, ve prospěch H_1.

V. Dvouvýběrové testy

- Předpokládejme nyní, že jsme získali data o porodní váze novorozenců ze dvou různých spádových nemocnic, přičemž odhady směrodatných odchylek naznačují odlišnou variabilitu v obou souborech.
- V případě, že hodnoty σ_1 a σ_2 nejsou známy, vede problém porovnání průměrných hodnot porodních vah novorozenců narozených v těchto dvou nemocnicích k tzv. dvouvýběrovému *t*-testu. Kdyby hodnoty σ_1 a σ_2 známy byly, mohli bychom opět použít *z*-test založený na normálním rozdělení rozdílů průměrných hodnot x_1 a x_2.

V. Dvouvýběrové testy

a) <u>Dvouvýběrový *z*-test</u> při známých hodnotách σ_1 a σ_2:

Při známých hodnotách σ_1 a σ_2 má testová statistika normální rozdělení:

$$z = \bar{x}_1 - \bar{x}_2 \sim N\left(m_1 - m_2, \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)$$

b) <u>Dvouvýběrový *t*-test</u> při neznámých hodnotách σ_1 a σ_2:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \approx t_{d'},$$

přičemž testová statistika t má na základě Satterthwaitovy aproximace počet stupňů volnosti d' roven přibližně:

$$d' \approx \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{(s_1^2/n_1)^2/(n_1-1) + (s_2/n_2)^2/(n_2-1)}$$

V. Dvouvýběrový *t*-test při neznámých hodnotách σ_1 a σ_2

<u>Motivace</u>: testujme hypotézu rovnosti průměrných porodních vah novorozenců narozených ve dvou různých spádových nemocnicích v průběhu let 2001-2003.

- H_0: $\mu_1 = \mu_2$, přičemž $\sigma_1 \neq \sigma_2$.
- H_1: $\mu_1 \neq \mu_2$, přičemž $\sigma_1 \neq \sigma_2$.
- Zvolme hladinu významnosti α
- Testová statistika: $t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \approx t_{d'}, \quad d' = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{(s_1^2/n_1)^2/(n_1-1) + (s_2/n_2)^2/(n_2-1)}$
- Testové kritérium: zamítneme nulovou hypotézu na hladině α v případě, že platí: $|t| > t_{d'}(1-\alpha/2)$,
  nebo ekvivalentně v případě, že dosažená hladina významnosti oboustranného testu bude menší než α.

V. Dvouvýběrový *t*-test při neznámých hodnotách σ_1 a σ_2

<u>Provedení</u>: předpokládejme, že údaje o porodních vahách z obou nemocnic jsou následující:

$$\bar{x}_1 = 2.8, s_1 = 0.85, n_1 = 75, \bar{x}_2 = 3.2, s_2 = 0.6, n_2 = 60$$

- H_0: $\mu_1 = \mu_2$, přičemž $\sigma_1 \neq \sigma_2$.
- H_1: $\mu_1 \neq \mu_2$, přičemž $\sigma_1 \neq \sigma_2$.
- Zvolme hladinu významnosti $\alpha = 0.05$
- <u>Testová statistika</u>: $t = \frac{|2.8 - 3.2|}{\sqrt{\frac{0.85^2}{75} + \frac{0.6^2}{60}}} = 3.20, \quad d' = \frac{(.85^2/75 + .6^2/60)^2}{(.85^2/75)^2/74 + (.6^2/60)^2/59} = 131.1$
- <u>Závěr</u>: Vzhledem k tomu, že $t = 3.20 > t_{131}(0.975) = 1.978$,
  zamítáme nulovou hypotézu na hladině významnosti 5%.
  Podobně dosažená hladina významnosti oboustranného testu má *p*-hodnotu 0.00172 < 0.05, takže i na základě tohoto kritéria zamítáme H_0 o shodnosti středních hodnot.

VI. Párový *t*-test

- Motivace: Je všeobecně známo, že používání orální antikoncepce (OA) je u žen obecně doprovázeno snížením jejich tělesné hmotnosti. V důsledku známé souvislosti mezi tělesnou hmotností a krevním tlakem bychom dále mohli například usuzovat, že ženy, které používají OA mají v průměru také nižší hodnoty systolického krevního tlaku.
- Hodnocení velikosti změn (diferencí d_i) hodnot spojité náhodné veličiny, jakou je např. systolický krevní tlak, u týchž subjektů v různých experimentálních podmínkách provádíme obecně pomocí párového *t*-testu. Tato situace vede opět k jednovýběrovému *t*-testu aplikovanému na hodnoty rozdílů sledované náhodné veličiny. I zde je třeba mít na paměti, že požadujeme normalitu, v tomto případě normalitu rozdílů. Ta je ovšem často splněna i v případech, kdy původní veličina předpoklad normálního rozdělení striktně nesplňuje.

VI. Párový *t*-test

- Příklad: Souvislost používání orální antikoncepce (OA) a průměrné hodnoty systolického krevního tlaku

ID	STK (OA-)	STK (OA+)	d
1	115	128	13
2	112	115	3
3	107	106	-1
4	119	128	9
5	115	122	7
6	138	145	7
7	126	132	6
8	105	109	4
9	104	102	-2
10	115	117	2

VI. Párový *t*-test

- H_0: $d = 0$, přičemž σ_d odhadujeme pomocí s_d.
- H_1: $d \neq 0$, přičemž σ_d odhadujeme pomocí s_d.
- Zvolme hladinu významnosti α (např. 0.05)
- Označme $\bar{d} = \sum_{i=1}^{n} d_i / n$ průměrnou hodnotu rozdílů d_i
- Testová statistika: $t = \frac{\bar{d} - 0}{s_d / \sqrt{n}} \sim t_{n-1} \mid H_0$
- Testové kritérium: zamítneme nulovou hypotézu na hladině α v případě, že platí: $|t| > t_{n-1}(1 - \alpha/2)$, resp. ekvivalentně v případě, že dosažená hladina významnosti jednostranného testu bude menší než α.

VI. Párový *t*-test

- Příklad – řešení v R:

```
x <- c(115,112,107,119,115,138,126,105,104,115)
y <- c(128,115,106,128,122,145,132,109,102,117)

> t.test(y,x,paired=TRUE)

        Paired t-test

data:  y and x
t = 3.3247, df = 9, p-value = 0.008874
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 1.533987 8.066013
sample estimates:
mean of the differences
              4.8
```