Testování statistických hypotéz

CHEMOMETRIE I, © David MILDE

- Jedná se o jednu z nejpoužívanějších metod pro vyslovení závěrů o základním souboru, který nezkoumáme celý, ale pomocí náhodného výběru.
- **Př.:**
  - Je obsah účinné látky ve 2 tabletách léku shodný?
  - Je obsah NO_3^- v pitné vodě menší než 15 mg/l?
  - Je koncentrace kyseliny vyráběná jedním postupem jiná než druhým postupem?
  - Je rozptyl výsledků stanovení Fe naším přístrojem menší než hodnota 2,55 uvedená v normě?
- Budeme se zabývat testováním hypotéz o parametrech rozdělení základního souboru (μ, σ^2). Nebudeme se zabývat testy týkajícími se tvaru rozdělení.

- Statistická hypotéza - jakýkoli předpoklad o rozdělení pravděpodobnosti náhodné veličiny; týká se parametrů rozdělení náhodné veličiny (shoda středních hodnot či rozptylů) v základním souboru nebo se může vztahovat k rozdělení náhodné veličiny.
- Test statistické hypotézy je pravidlo, které na základě výsledků zjištěných z naměřených hodnot předepisuje rozhodnutí, má-li být testovaná hypotéza zamítnuta či nikoli.
- Hypotézu, kterou chceme testovat (rozhodnout o ní) nazýváme **nulová hypotéza H_0**. Dále se definuje **alternativní hypotéza H_1**, která se přijímá v případě zamítnutí H_0.

- Př.: vyráběná kyselina má mít koncentraci 90 %; to ověřujeme pomocí analýzy vybraných vzorků kyseliny.
  - Pro „oboustrannou alternativu“: H_0: $\mu = 90$ H_1: $\mu \neq 90$.
  - Pro jednostrannou alternativu“: H_0 : $\mu = 90$ H_1: $\mu < 90$

    H_0: $\mu \geq 90$ H_1: $\mu < 90$.
- Při testování se vymezuje kritická hodnota (nejčastěji v tabulkách kritických hodnot) pro testování nulové hypotézy. Je-li výsledek zjištěný statistickým testem menší než kritická hodnota, přijímáme H_0. Je-li výsledek větší, pak se H_0 zamítá přijímá se H_1.

Postup při testování hypotéz

1. Formulace H_0 a H_1.
2. Volba hladiny významnosti.
3. Výpočet testovací charakteristiky na základě náhodného výběru (pocházejícího z normálního rozdělení).
4. Nalezení kritické hodnoty (v tabulkách).
5. Rozhodnutí o přijmutí či zamítnutí hypotéz.

Chyby při testování hypotéz

- Při rozhodování o přijetí či nepřijetí H_0 se můžeme dopustit jedné ze dvou chyb:
  1. **Zamítneme H_0, když ve skutečnosti platí – <u>chyba 1. druhu</u>.**
  2. **Přijmeme H_0, když ve skutečnosti neplatí (platí tedy H_1) – <u>chyba 2. druhu</u>.**
- Chyba 1. druhu má pravděpodobnost α a ta je rovna hladině významnosti testu (v praxi nejčastěji 5 %).
- Chyba 2. druhu má pravděpodobnost β a její velikost obvykle neznáme.
  - Číslo 1-β se nazývá síla testu.

Chyby při testování hypotéz

		Rozhodování	
		Přijímáme H_0	Zamítáme H_0 (přijímáme H_1)
Skutečnost	Platí H_0	**O.K.**	**Chyba 1. druhu**
	Neplatí H_0 (platí H_1)	**Chyba 2. druhu**	**O.K.**

Testování správnosti

(Jednovýběrový t-test o střední hodnotě)

Test správnosti

- Slouží k rozhodnutí, zda střední hodnota náhodného výběru (= aritmetický průměr) je nebo není rovna nějaké konkrétní číselné hodnotě (nazývané správná hodnota), či zda je průměr menší nebo větší než nějaké konkrétní hodnota.
- Testováním rozdílu průměru a správné hodnoty zjišťujeme, jak velký je mezi nimi rozdíl. Je-li menší než kritická hodnota, je vysvětlitelný pouze náhodnými chybami a výsledek považujeme za správný. Předpokladem je, že základní soubor i náhodný výběr mají normální rozdělení.

Test správnosti

$$H_0 : \mu = \overline{x} \; ; \; H_1 : \mu \neq \overline{x}$$

- Je znám rozptyl σ^2:

$$\boxed{z = \frac{|\overline{x} - \mu|}{\sigma}\sqrt{n}}$$

  - Srovnáváme s kritickou hodnotou normovaného normálního rozdělení $z_{(1-\alpha/2)}$.
- Není znám rozptyl σ^2:

$$\boxed{t = \frac{|\overline{x} - \mu|}{s}\sqrt{n}}$$

  - Srovnáváme s kritickou hodnotou t-rozdělení s n-1 stupni volnosti $t_{(1-\alpha/2;\, n-1)}$.

Test správnosti

- Je-li vypočtená hodnota t nebo z menší než příslušná kritická hodnota, přijímáme H_0. Je-li vypočtená hodnota t nebo z větší než kritická, zamítáme H_0 a přijímáme H_1.

H_0	H_1
$\bar{x} = \mu$	$\bar{x} \neq \mu$
$\bar{x} \leq \mu$	$\bar{x} > \mu$
$\bar{x} \geq \mu$	$\bar{x} < \mu$

H_0 se zamítá, když:

$|t| > t_{krit}(1-\alpha/2;\ n-1)$

$t > t_{krit}(1-\alpha;\ n-1)$

$t < t_{krit}(\alpha;\ n-1)$

Test správnosti

$$L_{1,2} = \bar{x} \pm s \frac{t_{(1-\alpha/2,n-1)}}{\sqrt{n}}$$

IS obsahuje μ, ± odstraníme použitím absolutní hodnoty.

$$|\mu - \bar{x}| = s \frac{t}{\sqrt{n}}$$

$$t = \frac{|\bar{x} - \mu|}{s} \sqrt{n}$$

Test správnosti v software

- **2 základní přístupy:**
- Použitím klasického testu s právnosti, což je možné v případě splnění statistických předpokladů (ZP: normalita, homogenita).
- Aplikací intervalu spolehlivosti:
  - Pomocí EDA a ZP identifikujeme vhodný intervalový odhad střední hodnoty (průměr, medián, opravený průměr po transformaci) a zjistíme zda správná hodnota (μ) leží uvnitř intervalu spolehlivosti.

Testování shody středních hodnot

(Dvouvýběrový t-test rovnosti středních hodnot dvou souborů)

Testování shody středních hodnot

- Slouží k testování dvou průměrů vypočtených z $\underline{n_1}$ a $\underline{n_2}$ stanovení. Využívá se např.:
  - Porovnání výsledků analýzy 2 vzorků pomocí jedné metody.
  - Porovnání výsledků 2 laboratoří (či 2 metod) při opakované analýze jednoho vzorku.
- Předpokládá se, že oba náhodné výběry jsou na sobě nezávislé a pocházejí z normálního rozdělení!
- Předpoklad shody či neshody rozptylů je třeba ověřit pomocí *F testu shody 2 rozptylů*.

Test shody středních hodnot pro $s_1^2 = s_2^2$

$$\boxed{t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{(n_1-1)s_1^2 + (n_2-1)s_2^2}} \cdot \frac{\sqrt{n_1 \cdot n_2 \cdot (n_1+n_2-2)}}{n_1+n_2}}$$

„T_1“

pro $n_1=n_2$

$$\boxed{t = \frac{(\bar{x}_1 - \bar{x}_2)}{s \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}}$$

$$\boxed{s = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}}$$

- Vypočtené t srovnáváme s t_{krit} pro $(n_1 + n_2 - 2)$ stupňů volnosti.

Test shody středních hodnot pro $s_1^2 \neq s_2^2$

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

„T_2“

- Vypočtené t se srovnává s t_{krit} pro $\underline{\nu}$ stupňů volnosti.

$$\nu = \left\{ \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1+1} + \frac{(s_2^2/n_2)^2}{n_2+1}} \right\} - 2$$

Test shodnosti

- Je-li vypočtená hodnota <u>t</u> menší než příslušná kritická hodnota *<u>pro příslušný počet stupňů volnosti</u>*, přijímáme H_0. Je-li vypočtená hodnota <u>t</u> větší než kritická, zamítáme H_0 a přijímáme H_1.

H_0	H_1
$\bar{x}_1 = \bar{x}_2$	$\bar{x}_1 \neq \bar{x}_2$
$\bar{x}_1 \leq \bar{x}_2$	$\bar{x}_1 > \bar{x}_2$
$\bar{x}_1 \geq \bar{x}_2$	$\bar{x}_1 < \bar{x}_2$

H_0 se zamítá, když:

$|t| > t_{krit}(1-\alpha/2)$

$t > t_{krit}(1-\alpha)$

$t < t_{krit}(\alpha)$

F-test shody 2 rozptylů

- Kromě testů o hodnotách parametrů 1 rozdělení je v praxi často potřeba porovnávat neznámé hodnoty parametrů mezi dvěma základními soubory.
- Dvouvýběrový Fisher-Snedecorův test (zkráceně F-test) slouží k ověření shody rozptylů dvou základních souborů.
- Ze základních souborů $N(\mu_1, \sigma_1^2)$ a $N(\mu_2, \sigma_2^2)$ provedeme 2 náhodné výběry, o kterých předpokládáme, že jsou nezávislé a spočteme výběrové odhady rozptylů s_1^2 a s_2^2.

$$\boxed{F = \frac{s_1^2}{s_2^2}}$$

F-test shody 2 rozptylů

- **POZOR! F musí být vždy větší než 1!**
- Je-li vypočtená hodnota F menší než příslušná kritická hodnota F-rozdělení, přijímáme H_0. Je-li vypočtená hodnota F větší než kritická, zamítáme H_0 a přijímáme H_1.

H_0	H_1	H_0 se zamítá, když
$s^2_1 = s^2_2$	$s^2_1 \neq s^2_2$	$F > F_{krit(n1-1,\ n2-1)}$
$s^2_1 \leq s^2_2$	$s^2_1 > s^2_2$	$F > F_{krit(n1-1,\ n2-1)}$
$s^2_1 \geq s^2_2$	$s^2_1 < s^2_2$	$F < 1/F_{krit(n2-1,\ n1-1)}$

Testování shodnosti v software

- Studentovy testy vycházejí z předpokladu normálního rozdělení analyzovaných souborů. Pokud tato podmínka není splněna, nelze je použít.

- Obecný postup:

1. Ověření normality obou výběrů (EDA, ZP).
2. Testování shody rozptylů.
3. Testování shody středních hodnot.

Testování shodnosti v software

TESTY SHODY ROZPTYLŮ

H_0: $s_1^2 = s_2^2$ H_1: $s_1^2 \neq s_2^2$

- **Klasický F-test shody rozptylů** – oba výběry pocházejí z normálního rozdělení.
- **Robustní F-test shody rozptylů** – pro případ, kdy jeden nebo oba výběry nejsou z normálního rozdělení.
- *Testy shody rozptylů se používají k rozhodování, zda lze při testování shody středních hodnot vycházet z předpokladu rovnosti rozptylů.*

Testování shodnosti v software

TESTY SHODY STŘEDNÍCH HODNOT

H_0: $\mu_1 = \mu_2$ H_1: $\mu_1 \neq \mu_2$

- "T_1" **Klasický Studentův t-test pro shodné rozptyly** – normální rozdělení u obou výběrů.
- "T_2" **Klasický Studentův t-test pro různé rozptyly** – normální rozdělení u obou výběrů.
- " T_3" **Modifikovaný Studentův t-test** – pro odchylky od normality v šikmosti.

Testování shodnosti v software

TESTY SHODY STŘEDNÍCH HODNOT

- "T_4" **Robustní test pro homoskedasticitu** – v čitateli jsou uřezané průměry a ve jmenovateli winsorizované součty čtverců odchylek (*obdoba rozptylu, ale bez podělení počtem stupňů volnosti*)

$$t = \frac{\bar{x}_1(\vartheta) - \bar{x}_2(\vartheta)}{\sqrt{S_{w,1}(\vartheta) + S_{w,2}(\vartheta)}}$$

Testování shodnosti v software

TESTY SHODY STŘEDNÍCH HODNOT

- **"T_5" Robustní test pro heteroskedasticitu**

$$t = \frac{\bar{x}_1(\vartheta) - \bar{x}_2(\vartheta)}{\sqrt{\frac{s_{w,1}^2}{h_1} + \frac{s_{w,2}^2}{h_2}}}$$

$$s_{w,i}^2 = \frac{S_{w,i}(\vartheta)}{h_i - 1} \quad a \quad h_i = n_i - 2\,\mathrm{int}\left(\frac{n_i \cdot \vartheta}{100}\right), \quad pro\ i = 1, 2.$$

- Testy T_1 a T_2 jsou použitelné pro výběry z normálního rozdělení a jsou i dostatečně robustní vůči odchylkám od normality ve špičatosti.
- Robustní testy T_4 a T_5 jsou výhodné pro asymetrická rozdělení a rozdělení s výrazně vyšší špičatostí než 3. V případě normálního rozdělení však mají menší sílu než testy T_1 a T_2.

Párový test

(Párový t-test rovnosti středních hodnot dvou souborů)

Párový test

- Slouží k testování shody dvou středních hodnot pro závislé výběry – tzv. „párová data“, např.:
  - Porovnání 2 metod pomocí analýzy více než 2 vzorků.
  - Srovnání životních nákladů u těch samých osob v roce 2005 a 2006.
  - Vliv léku na hladinu cholesterolu před a po aplikaci u stejných (více než 2) pacientů.

$$\begin{bmatrix} \vdots \\ \vdots \end{bmatrix} \Leftrightarrow \begin{bmatrix} \vdots \\ \vdots \end{bmatrix}$$

- Statistické předpoklady: párové diference (d_i) jsou nezávislé s normálním rozdělením.

Párový test

- Párový test je v praxi obvykle formulován jako oboustranná alternativa s:

$$H_0: \overline{x}_d = 0; \quad H_1: \overline{x}_d \neq 0$$

$$\boxed{t = \frac{\overline{x}_d}{s_d} \cdot \sqrt{n}}$$

  - kde x_d je průměr a s_d je směrodatná odchylka párových diferencí.
  - Vypočtené t srovnáváme s kritickou hodnotou Studentova rozdělení pro n-1 stupňů volnosti $t_{krit}(1-\alpha/2; n-1)$.
- Je-li $t<t_{krit}$, platí H_0. Je-li $t>t_{krit}$, platí H_1.

Test vylučování odlehlých hodnot

Dean-Dixonův test

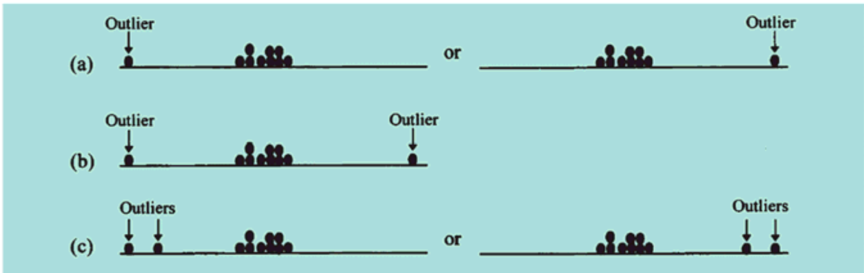
- Za odlehlé považujeme výsledky, které jsou v sérii paralelních měření zatíženy hrubou chybou. Zkreslily by nám statistické zpracování dat a proto je musíme před analýzou vyloučit.
- Dean-Dixonův test je vhodný pro soubory do n = 30.
  - Výsledky se seřadí podle velikosti a spočítá se rozpětí R.

$$Q_1 = \frac{x_2 - x_1}{R} \quad a \quad Q_n = \frac{x_n - x_{n-1}}{R}$$

  - Q_1 a Q_n následně srovnáme s kritickou hodnotou $Q_{krit}(\alpha, n)$. Je-li Q_1 nebo $Q_n < Q_{krit}$, dané hodnoty nejsou odlehlé. Je-li Q_1 nebo $Q_n > Q_{krit}$, dané hodnoty jsou odlehlé.

Grubbsův test

- Grubbsův test je vhodný pro soubory do n = 100.
  - Parametrický test.
  - Používáme parametry souboru: průměr a směrodatnou odchylku.



Grubbsův test

Varianta A $T_n = \frac{x_n - \bar{x}}{s}$ $T_1 = \frac{\bar{x} - x_1}{s}$

Varianta B $T_B = \frac{x_n - x_1}{s}$

Varianta C $T_c = 1 - \left(\frac{(n-3)s_{n-2}^2}{(n-1)s^2}\right)$

  - Vypočtené T_i srovnáme s kritickou hodnotou pro n stupňů volnosti $T_{krit}(\alpha, n)$. Je-li $T_i > T_{krit}$, daná hodnota/ty je/jsou odlehlé.
- Mezi testy vylučování OB patří i test modifikovaných vnitřních hradeb!