Univerzita Palackého v Olomouci

Testování statistických hypotéz

Chemometrie I (ACH/CHEX1)

(c) David MILDE, 2024

1

Univerzita Palackého v Olomouci

Úvod

- Jedná se o jednu z nejpoužívanějších metod pro vyslovení závěrů o základním souboru, který nezkoumáme celý, ale pomocí náhodného výběru. Tj. základní soubor posuzujeme na základě jeho parametrů. Hodnotíme tvrzení, zda parametry nabývají určité hodnoty a testujeme, zda tvrzení je nebo není pravdivé.
- 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 roven rozptylu popsanému v normě?
- Budeme se zabývat testováním hypotéz o parametrech rozdělení základního souboru (střední hodnotě μ a rozptylu σ^2).
  - Testy týkajícími se ověřování hypotézy, zda data pocházejí z normálního rozdělení, jsme se zabývali dříve.
- Testy, ve kterých ověřujeme platnost hypotézy o parametru souboru (tj. $\bar{x}, \sigma^2$), se označují jako **parametrické testy**.

2

Univerzita Palackého v Olomouci

Úvod

- Statistická hypotéza – jakýkoliv předpoklad o rozdělení pravděpodobnosti náhodné veličiny. Týká se parametrů rozdělení náhodné veličiny v základním souboru nebo se může vztahovat k druhu rozdělení náhodné veličiny.
- Test statistické hypotézy – 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 %; což ověřujeme pomocí analýzy vybraných vzorků kyseliny.
  - Pro oboustrannou hypotézu: H_0: μ = 90 %; H_1: $\mu \neq$ 90 %.
  - Pro jednostrannou hypotézu: H_0: μ = 90 %; H_1: μ < 90 % či H_0: μ = 90 %; H_1: μ > 90 %.

3

Univerzita Palackého v Olomouci

Postup testování statistické hypotézy

1. Formulace nulové hypotézy H_0 a alternativní hypotézy H_1.
2. Volba hladiny významnosti.
3. Výpočet testační statistiky např. ***t*** nebo ***F*** (na základě náhodného výběru a splnění předpokladů pro statistický test).
4. Rozhodnutí zda se jedná jednostrannou nebo oboustrannou hypotézu.
5. Nalezení kritické hodnoty (v tabulkách či softwaru), tj. určení tzv. kritického oboru.
6. Rozhodnutí, zda
   a. **zamítnout H_0** a přijmout H_1, jestli výsledek testační statistiky padne do kritického oboru. U parametrických testů nastává, pokud je výsledek testační statistiky větší než kritická hodnota.
   b. nezamítnout H_0, jestli výsledek testační statistiky nepadne do kritického oboru. Tuto variantu budeme zjednodušeně označovat jako **přijetí H_0**, přestože korektně bychom měli uvádět, že H_0 nemůžeme zamítnout (nemáme dost „důkazů“ k zamítnutí). U parametrických testů nastává, pokud je výsledek testační statistiky menší než kritická hodnota.

4

Univerzita Palackého v Olomouci

Chyby při testování hypotéz

- Při rozhodování o přijetí či zamítnutí H_0 se můžeme dopustit jedné ze dvou chyb (aj. error):
  - Zamítneme H_0, když ve skutečnosti platí – chyba 1. druhu.
  - Přijmeme H_0, když ve skutečnosti neplatí – chyba 2. druhu.
- Chyba 1. druhu má pravděpodobnost α a ta je rovna hladině významnosti testu (obvykle 5 %).
- Chyba 2. druhu má pravděpodobnost β a její velikost neznáme.
  - Hodnota 1-β se nazývá síla testu.

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

5

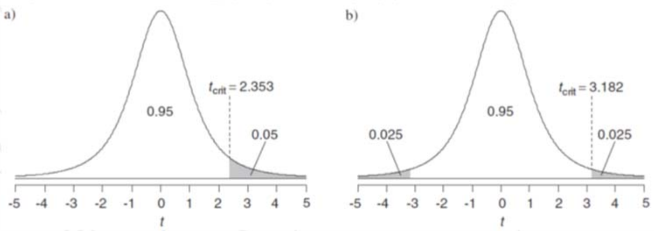
Univerzita Palackého v Olomouci

Kritické hodnoty, „*p*-value“

- Kritické hodnoty lze nalézt v tabulkách kritických hodnot nebo ve statistickém SW.
- K nalezení kritické hodnoty potřebujeme: zvolit rozdělení podle testu (***N, t, F,*** χ^2), znát hladinu významnosti, počet stupňů volnosti a určit, zda jde o jednostrannou nebo dvoustrannou hypotézu.
- Př.: *t*-rozdělení, α = 5 %, ν = 3

  a) jednostranná hypotéza

  b) oboustranná hypotéza



- ***p-value*** – představuje pravděpodobnost, že pozorovaná hodnota statistického testu je větší nebo rovna kritické hodnotě, pokud H_0 platí. V SW je často uváděna místo slovního závěru testu.
  - Př. pro α = 5 %: p-value < 0,05 ⇒ zamítáme H_0, p-value > 0,05 ⇒ přijímáme H_0.

6

Univerzita Palackého v Olomouci

Parametrické testy

Jednovýběrový *t*-test

Dvouvýběrový *t*-test

Párový *t*-test

7

Univerzita Palackého v Olomouci

Jednovýběrový *t*-test (střední hodnoty)

- Slouží k rozhodnutí, zda střední hodnota náhodného výběru ($\bar{x}$) je nebo není rovna nějaké konkrétní hodnotě (μ), či zda je $\bar{x}$ menší nebo větší než μ.
  - Test je také nazýván test správnosti.
- Testováním $\bar{x} - \mu$ zjišťujeme, je-li rozdíl menší než kritická hodnota, kdy je vysvětlitelný pouze náhodnými chybami nebo větší než kritická hodnota, což potvrzuje přítomnost systematické chyby.
- Předpoklad: základní soubor i náhodný výběr mají normální rozdělení.
- **2 základní přístupy** (ve statistickém SW):
  - použití jednovýběrového *t*-testu střední hodnoty, což je možné v případě splnění předpokladů (ZP: normalita, homogenita).
  - aplikace 100(1 – α)% intervalu spolehlivosti: pomocí EDA a ZP identifikujeme vhodný IS střední hodnoty (medián, opravený průměr po transformaci) a zjistíme zda správná hodnota (μ) leží uvnitř tohoto IS.

8

Univerzita Palackého v Olomouci

Jednovýběrový *t*-test

- **Je znám σ^2:**

  H_0: $\bar{x} = \mu$ H_1: (1.) $\bar{x} \neq \mu$

  H_1: (2.) $\bar{x} < \mu$

  H_1: (3.) $\bar{x} > \mu$

$$z = \frac{\bar{x} - \mu}{\sigma} \cdot \sqrt{n}$$

- Vypočtené *z* srovnáváme s kritickou hodnotou normovaného normálního rozdělení, oblast zamítnutí H_0 je

  (1.) $|z| > z_{(1-\frac{\alpha}{2})}$, **(2.)** $z < -z_{1-\alpha}$, **(3.)** $z > z_{1-\alpha}$

- **Není znám σ^2 nebo n < 30:**

  H_0: $\bar{x} = \mu$ H_1: (1.) $\bar{x} \neq \mu$

  H_1: (2.) $\bar{x} < \mu$

  H_1: (3.) $\bar{x} > \mu$

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

- Vypočtené *t* srovnáváme s kritickou hodnotou Studentova rozdělení, oblast zamítnutí H_0 je

  (1.) $|t| > t_{(1-\frac{\alpha}{2}, n-1)}$, **(2.)** $t < -t_{(1-\alpha, n-1)}$, **(3.)** $t > t_{(1-\alpha, n-1)}$

9

Univerzita Palackého v Olomouci

Jednovýběrový *t*-test a interval spolehlivosti

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

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

Odstraněním absolutní hodnoty a úvahou získáme:

$$\boxed{\bar{x} - \mu = \pm\, s\frac{t}{\sqrt{n}}}$$

Převedením $\bar{x}$ na pravou stranu a úvahou, že μ je v IS získáme:

$$\boxed{L_{1,2} = \bar{x} \pm s\frac{t}{\sqrt{n}}}$$

10

Univerzita Palackého v Olomouci

Test o rozptylu

- Slouží k rozhodnutí, zda neznámý odhad rozptylu s^2 je nebo není roven nějaké konkrétní číselné hodnotě (σ^2), či zda je odhad rozptylu menší nebo větší než nějaká konkrétní hodnota.
- Předpoklad: náhodný výběr, ze kterého počítáme s^2 pochází z normálního rozdělení.

H_0: $s^2 = \sigma^2$, H_1: (1.) $s^2 \neq \sigma^2$, H_1: (2.) $s^2 < \sigma^2$, H_1: (3.) $s^2 > \sigma^2$

- Testační statistika:

$$\chi^2_\sigma = \frac{(n-1)\cdot s^2}{\sigma^2}$$

- Je-li vypočtená hodnota χ^2_σ menší než kritická hodnota, přijímáme H_0. Je-li vypočtená hodnota χ^2_σ větší než kritická, zamítáme H_0 a přijímáme H_1. Oblast zamítnutí:

(1.) $\chi^2_\sigma > \chi^2_{krit(1-\frac{\alpha}{2}, n-1)}$, (2.) $\chi^2_\sigma < \chi^2_{krit(\alpha,\, n-1)}$, (3.) $\chi^2_\sigma > \chi^2_{krit(1-\alpha,\, n-1)}$

11

Univerzita Palackého v Olomouci

Dvouvýběrový *t*-test (shody středních hodnot)

- Slouží k testování shody středních hodnot dvou základních souborů (μ_1 a μ_2) pomocí průměrů vypočtených z n_1 a n_2 dat náhodných výběrů.
  - Test je také nazýván test shodnosti.
- Využívá se např.:
  - Porovnání výsledků analýzy 2 vzorků pomocí jedné metody.
  - Porovnání výsledků 2 laboratoří (či 2 metod) při opakované analýze jednoho vzorku.
- Předpoklad: oba náhodné výběry jsou na sobě nezávislé a pocházejí z normálního rozdělení.
- Dvouvýběrový *t*-test existuje ve dvou základních variantách v závislosti na shodě rozptylů s_1^2 a s_2^2. Předpoklad shody či neshody rozptylů je třeba ověřit pomocí ***F*-testu shody rozptylů**.
- Zabývat se budeme pouze situací, kdy neznáme σ^2, tj. budeme používat výběrové odhady rozptylu s^2.

12

Univerzita Palackého v Olomouci

Dvouvýběrový *t*-test

H_0: $\bar{x}_1 = \bar{x}_2$ H_1: (1.) $\bar{x}_1 \neq \bar{x}_2$ (2.) $\bar{x}_1 < \bar{x}_2$ (3.) $\bar{x}_1 > \bar{x}_2$

– $s_1^2 = s_2^2$:

T1

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

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

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

pro $n_1 = n_2$

– Vypočtené *t* srovnáváme s kritickou hodnotou Studentova rozdělení pro $\nu = (n_1+n_2-2)$ stupňů volnosti.

– $s_1^2 \neq s_2^2$:

T2

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

– Vypočtené *t* srovnáváme s kritickou hodnotou Studentova rozdělení pro *ν*, stupňů volnosti.

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

(1.) $|t| > t_{(1-\frac{\alpha}{2},\nu)}$, (2.) $t < -t_{(1-\alpha,\nu)}$, (3.) $t > t_{(1-\alpha,\nu)}$

13

Univerzita Palackého v Olomouci

F-test shody rozptylů

– Dvouvýběrový Fisher-Snedecorův test (zkráceně *F*-test) slouží k ověření shody dvou rozptylů dvou základních souborů.

– Ze základních souborů provedeme dva náhodné výběry a spočteme výběrové odhady rozptylů s_1^2 a s_2^2.

– Předpoklad: dva nezávislé výběry ze souborů s normálním rozdělením.

H_0: $s_1^2 = s_2^2$, H_1: $s_1^2 \neq s_2^2$ (oboustranná hypotéza)

– Jednostrannými hypotézami H_0: $s_1^2 = s_2^2$, H_1: $s_1^2 < s_2^2$ či H_0: $s_1^2 = s_2^2$, H_1: $s_1^2 > s_2^2$ se nebudeme zabývat.

$$F = \frac{s_{max}^2}{s_{min}^2}$$

s_{max}^2 je větší a s_{min}^2 je menší z dvojice s_1^2 a s_2^2

– Je-li vypočtená hodnota $F < F_{krit(\alpha;\, n1-1,\, n2-1)}$, přijímáme H_0. Je-li vypočtená hodnota $F > F_{krit(\alpha;\, n1-1,\, n2-1)}$, zamítáme H_0 a přijímáme H_1.

14

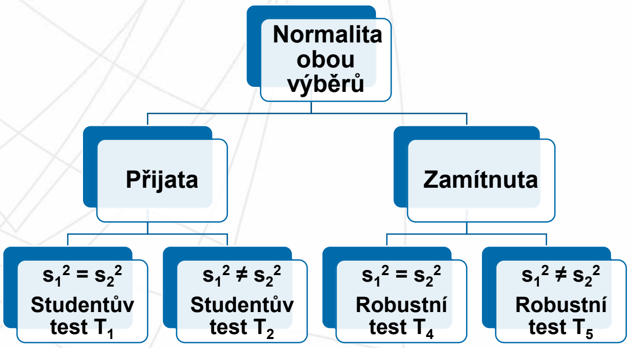
Univerzita Palackého v Olomouci

Dvouvýběrový *t*-test v softwaru

- **POSTUP:**
  1. Ověření normality obou výběrů (EDA, ZP).
  2. Testování shody rozptylů (F test).
  3. Testování shody středních hodnot (*t*-testy).
- Testy shody rozptylů:
  - „Klasický“ *F*-test shody rozptylů – oba výběry pocházejí z normálního rozdělení.

$$F = \frac{s_{max}^2}{s_{min}^2}$$

  - Robustní *F*-test shody rozptylů – pro případ, kdy jeden nebo oba výběry nepocházejí z normálního rozdělení. F se vypočte stejně jako u klasického testu, ale počty stupňů volnosti se korigují na odchylku od normality.
- Testy shody středních hodnot:



15

Univerzita Palackého v Olomouci

Dvouvýběrový *t*-test v softwaru

- Testy shody středních hodnot ve statistickém softwaru:
  - o Studentův *t*-test pro shodné rozptyly – “T_1” (normální rozdělení u obou výběrů).
  - o Studentův *t*-test pro různé rozptyly – “T_2” (normální rozdělení u obou výběrů).
  - o (Modifikovaný Studentův *t*-test – “T_3” pro odchylky od normality v šikmosti).
  - o Robustní *t*-test pro shodné rozptyly – “T_4” (zamítnutá normalita 1 či 2 výběrů).
  - o Robustní *t*-test pro různé rozptyly – “T_5” (zamítnutá normalita 1 či 2 výběrů).
- Robustní *t*-testy: v čitateli je rozdíl uřezaných průměrů ($\overline{x(v)}_i$) a ve jmenovateli winsorizované součty čtverců odchylek ($S_{w,i}{}^2$) nebo winsorizované rozptyly ($s_{w,i}{}^2$).

T_4 $$t = \frac{\overline{x(v)}_1 - \overline{x(v)}_2}{\sqrt{S_{w,1}^2 + S_{w,2}^2}}, \quad t = \frac{\overline{x(v)}_1 - \overline{x(v)}_2}{\sqrt{\frac{s_{w,1}^2}{h_1} + \frac{s_{w,2}^2}{h_2}}}$$ T_5

***h_1* a *h_2* souvisí s počtem hodnot obou výběrů**

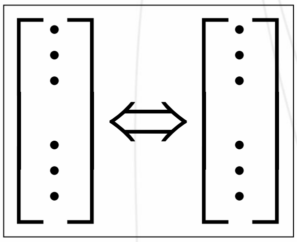
- Grafy v QC-Expertu:
  - krabicové grafy obou souborů,
  - gaussovy křivky hustoty pravděpodobnosti obou souborů – IS $|\bar{x}_1 - \bar{x}_2|$.

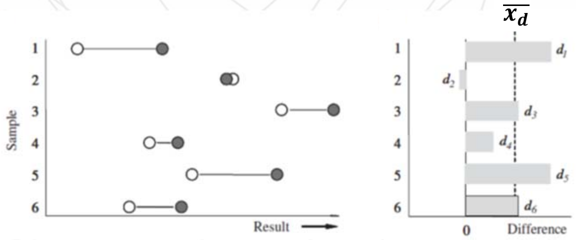
16

Univerzita Palackého v Olomouci

Párový *t*-test

- Slouží k testování shody dvou středních hodnot (tj. průměrů) pro závislé výběry – tzv. „párová data“, např.:
  - porovnání výsledků 2 metod pomocí analýzy více než 2 vzorků.
  - porovnání výsledků 2 laboratoří pomocí analýzy více než 2 vzorků.
  - srovnání životních nákladů u těch samých osob ve dvou letech.
  - vliv léku na hladinu cholesterolu před a po aplikaci u stejných (více než 2) pacientů.





Př.: analýza 6 vzorků pomocí metody A (prázdná kolečka) a metody B (plná kolečka)

- Předpoklady: párové diference (d_i) jsou nezávislé a s přibližně normálním rozdělením. Normalitu obvykle stačí ověřit graficky, např. v Q-Q grafu párových diferencí.

17

Univerzita Palackého v Olomouci

Párový *t*-test

- Párový test je obvykle formulován jako oboustranná hypotéza:

H_0: $\bar{x}_d = 0$, H_1: $\bar{x}_d \neq 0$ $\qquad$ $$t = \frac{\bar{x}_d}{s_d} \cdot \sqrt{n}$$

  - kde $\bar{x}_d$ je průměr a s_d je směrodatná odchylka párových diferencí d_i.
- Vypočtené ***t*** srovnáváme s kritickou hodnotou Studentova rozdělení $t_{krit(1-\alpha/2;\ n-1)}$. Je-li $|t| < t_{krit(1-\alpha/2;\ n-1)}$, přijímáme H_0, je-li $|t| > t_{krit(1-\alpha/2;\ n-1)}$, přijímáme H_1.
- Jednostrannými hypotézami, kdy H_0: $\bar{x}_d = 0$, H_1: $\bar{x}_d < 0$ nebo H_0: $\bar{x}_d = 0$, H_1: $\bar{x}_d > 0$ se nebudeme zabývat.
- Grafy v QC-Expertu:
  - Q-Q graf – grafické posouzení normality diferencí,
  - Bland-Altmannův graf – závislost variability na velikosti hodnot,
  - Gaussova křivka hustoty pravděpodobnosti s IS $\bar{x}_d$.

18

Univerzita Palackého v Olomouci

Neparametrické testy

Wilcoxonův jednovýběrový test

Wilcoxonův párový test

Mann-Whitney U test

Chí kvadrát test dobré shody

19

Univerzita Palackého v Olomouci

Úvod

- Výhodou neparametrických testů je jejich použitelnost bez ohledu na typ rozdělení, z něhož data pochází. K testování se totiž nepoužívají parametry výběru ($\bar{x}$, s).
- Nevýhodou neparametrických testů je menší citlivost, tj. menší schopnost odkrýt na dané hladině významnosti nesprávnost testované hypotézy.
- Použití neparametrických testů je běžnější v humanitních či ekonomických vědních oborech. V chemii a dalších přírodních oborech bychom se jejich použití měli spíše vyvarovat nebo je používat po důkladném zvážení všech dalších možností.
- Druhy neparametrických testů:
  - Pořadové testy – provádíme setřídění hodnot podle velikosti do jedné řady a místo původních dat používáme jejich pořadová čísla. Budeme se jimi zabývat.
  - Znaménkové testy – sleduje se střídání znamének výsledků. Nebudeme se jimi zabývat.
- U Wilcoxonových testů a u Mann-Whitneyova testu se používá opačný postup zamítání H_0 než u parametrických testů. Tedy, je-li vypočtená statistika menší nebo rovno kritické tabelované hodnotě, H_0 se zamítá. Pokud je vypočtená statistika větší než kritická hodnota, H_0 se přijímá (nezamítá).

20

Univerzita Palackého v Olomouci

Jednovýběrový Wilcoxonův test

- Jde o neparametrickou obdobu jednovýběrového t-testu. Budeme se zabývat pouze oboustrannou hypotézou.

$$H_0: \mu = \tilde{x}_{0,5},\ H_1: \mu \neq \tilde{x}_{0,5}$$

- POSTUP (n by mělo být > 6):
  1. Od prvků výběru se odečte správná hodnota (μ).
  2. Absolutní hodnoty rozdílů seřadíme do neklesající posloupnosti.
  3. Každé hodnotě přiřadíme pořadové číslo (pořadí).
  4. Vytvoříme sumu pořadí nezáporných prvků S^+ a sumu pořadí záporných prvků S^-. Při shodě pořadí se použije průměrné pořadí.
  5. Menší číslo z dvojice S^+ a S^- srovnáváme s kritickou tabelovanou hodnotou $w_{(n,\ 0{,}05)}$.

- **Př.:** Objemy spotřeby titračního činidla při titraci 10 ml přibližně 0,01 mol l^{-1} HCl na titrátoru RTS 622 jsou 1,10; 1,08; 1,09; 1,08; 1;10; 1,08; 1,10; 1,09; 1,11; 1,08 ml. Správná hodnota byla určena na 1,09 ml. Zjistěte, zda titrátor pracuje správně.

21

Univerzita Palackého v Olomouci

Párový Wilcoxonův test

- Jde o neparametrický párový test. Data se výpočtem párových diferencí převedou na soubor dat shodný s typem pro jednovýběrový Wilcoxonův test. Opět se budeme zabývat pouze oboustrannou hypotézou.
- POSTUP (n by mělo být > 6):
  1. Vypočtou se diference párových hodnot d_i.
  2. Diference se v absolutní hodnotě seřadí do neklesající posloupnosti.
  3. Každé hodnotě přiřadíme pořadové číslo (pořadí), přičemž případy kdy $d_i = 0$ se vynechávají.
  4. Vytvoříme sumu pořadí nezáporných prvků S^+ a sumu pořadí záporných prvků S^-. Při shodě pořadí se použije průměrné pořadí.
  5. Menší číslo z dvojice S^+ a S^- srovnáváme s kritickou tabelovanou hodnotou $w_{(n,\ 0{,}05)}$.
- **Př.:** Paralelními analýzami vzorku Cu v 8 slitinách byla získána data nové metody a standardní metody. Testujte, zda obě metody určují vždy stejný obsah. Použijte parametrický i neparametrický test. Data [%]: 11,68 11,23; 23,91 23,77; 32,27 33,04; 38,29 38,43; 47,04 46,79; 51,34 50,96; 68,23 67,85; 79,24 78,55.

22

Univerzita Palackého v Olomouci

Mann-Whitney U test

- Jedná se o neparametrickou obdobu dvouvýběrového t-testu středních hodnot. Oba soubory musí pocházet ze stejného rozdělení. Řešit budeme pouze oboustrannou hypotézu.

$$H_0: \tilde{x}_{0,5;1} = \tilde{x}_{0,5;2}, \quad H_1: \tilde{x}_{0,5;1} \neq \tilde{x}_{0,5;2}$$

- POSTUP:
  1. Označme jeden výběr X_1 a druhý X_2, jejich rozsahy n_1 a n_2.
  2. Všechny prvky z obou výběrů (tzv. sdružený výběr) uspořádáme do neklesající posloupnosti a zjistíme součet pořadí výběru X_1, který označíme T_1, obdobně určíme součet T_2.
  3. Vypočteme statistiky U_1 a U_2.

$$U_1 = T_1 - \frac{n_1(n_1+1)}{2} \qquad U_2 = T_2 - \frac{n_2(n_2+1)}{2}$$

  4. Menší číslo z dvojice U_1 a U_2 se srovnává s kritickou tabelovanou hodnotou $w_{(n1,\ n2;\ 0,05)}$.

- **Př.:** Bylo vybráno 10 polí stejné kvality. Na 4 polích byl aplikován růstový stimulátor, ostatní byla ponechána bez aplikace. Poté byla oseta pšenicí a sledoval se hektarový výnos. Na polích s aplikací stimulátoru byly získány hektarové výnosy 51, 67, 56, 63 a na polích bez aplikace 45, 54, 48, 44, 53, 50 q/ha. Zjistěte, zda jsou výnosy stejné.

23

Univerzita Palackého v Olomouci

Chí kvadrát test dobré shody

- Dosud popsány testy, které hodnotí spojité proměnné. Na rozdíl od toho se χ^2 test dobré shody zabývá četností, tj. počtem případů, kdy se určitá hodnota vyskytla. Chí-kvadrát (χ^2) test slouží k testování, zda se pozorované četnosti v konkrétním případě významně liší od četnosti, kterou očekáváme na základě H_0. Testujeme hodnoty četnosti a ne spojité proměnné, nemůžeme tedy použít *t*-testy.
- Jedná se o testy tvaru rozdělení, tj. že základní soubor má rovnoměrné rozdělení, např.: vznik zmetků ve výrobě je rovnoměrně rozložen po celé směně.

H_0: veličina má rovnoměrné rozdělení **H_1: veličina nemá rovnoměrné rozdělení**

- POSTUP:
  1. Vypočteme teoretickou (po)četnost podle H_0.
  2. Zjistíme skutečné (po)četnosti v jednotlivých skupinách.
  3. Porovnáme skutečné a teoretické početnosti a rozdíly mezi nimi použijeme pro výpočet hodnoty testační statistiky χ^2_{exp}.
  4. Pro α = 5 % porovnáme χ^2_{exp} s kritickou hodnotou $\chi^2_{krit(n-1;\ 0,95)}$, kde *n* je počet skupin,.
  5. Je-li $\chi^2_{exp} < \chi^2_{krit}$, přijímáme H_0, pokud je $\chi^2_{exp} > \chi^2_{krit}$, zamítáme H_0 a přijímáme H_1.

24

Univerzita Palackého v Olomouci

Chí kvadrát test dobré shody

$$\boxed{\chi^2_{exp} = \sum_{i=1}^{n} \frac{(O_i - E)^2}{E}}$$

kde O_i – skutečná početnost v příslušné skupině
$E = N/n$ – teoretická početnost, kde N je celková početnost.

- **Př.:** Čtyři laboranti rozbili následující počet laboratorního skla za určité období 24, 17, 11 a 9 kusů. Lze rozdíl v počtu rozbitého skla mezi pracovníky považovat za významný?

$n = 4$, $N = 24 + 17 + 11 + 9 = 61$ $E = 61/4 = 15{,}25$

$O_1 – E = 8{,}75$ $O_2 – E = 1{,}75$ $O_3 – E = -4{,}25$ $O_4 – E = -6{,}25$

$$\chi^2_{exp} = \frac{8{,}75^2 + 1{,}75^2 + (-4{,}25)^2 + (-6{,}25)^2}{15{,}25} = \frac{136{,}25}{15{,}25} = 8{,}934$$

$\chi^2_{krit(3;\,0{,}95)} = 7{,}815$ $\chi^2_{exp} > \chi^2_{krit}$ zamítáme H_0 a přijímáme H_1

25

Univerzita Palackého v Olomouci

Testy vylučování odlehlých hodnot

Dean-Dixonův test

Grubbsův test

26

Univerzita Palackého v Olomouci

Úvod, Dean-Dixonův test

- Za odlehlé považujeme výsledky, které jsou v souboru dat zatíženy hrubou chybou. Zkreslily by statistické zpracování dat a proto je před statistickou analýzou vylučujeme.
- **Dean-Dixonův test** je neparametrický test vhodný pro soubory do n ≤ 30.
  - Výsledky se seřadí podle velikosti a spočítá se rozpětí *R*.

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

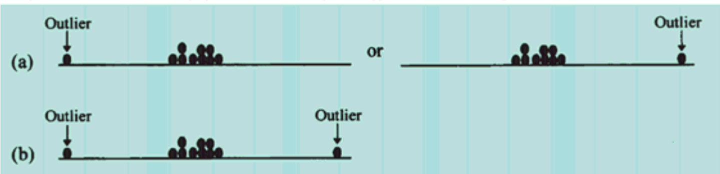
  - x_1 – nejmenší hodnota, x_2 – druhá nejmenší hodnota, x_n – největší hodnota, x_{n-1} – druhá největší hodnota.
  - Q_1 a Q_n následně srovnáme s kritickou hodnotou $Q_{krit(\alpha,\, n)}$. Je-li Q_1 anebo $Q_n < Q_{krit}$, dané hodnoty nejsou odlehlé. Je-li Q_1 anebo $Q_n > Q_{krit}$, dané hodnoty jsou odlehlé.

27

Univerzita Palackého v Olomouci

Grubbsův test

- Grubbsův test je parametrický test vhodný pro soubory do n ≤ 100.
- 2 základní varianty:



$$(a)\ T_1 = \frac{\bar{x} - x_1}{s}, T_n = \frac{x_n - \bar{x}}{s}$$

$$(b)\ T_b = \frac{x_n - x_1}{s}$$

  - **x_1 – nejmenší hodnota, x_n – největší hodnota**
- 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, pokud je $T_i < T_{krit}$, daná hodnota/ty není/nejsou odlehlé.
- Mezi testy vylučování odlehlých hodnot patří i test modifikovaných vnitřních hradeb (viz základní předpoklady o datech).

28