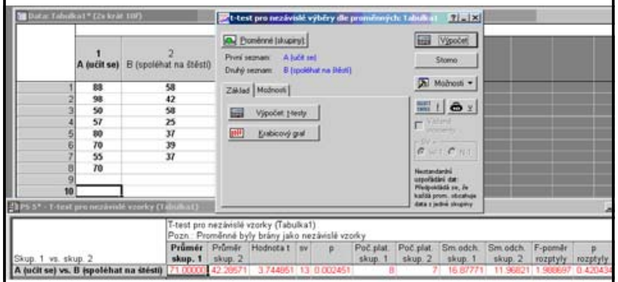
Statistické metody a zpracování dat

V. Testování statistických hypotéz

Petr Dobrovolný

K čemu to je (příklad)

Má smysl se připravovat na písemný test ze statistiky?



Má to smysl.

K čemu to je?

Ověřování domněnek či předpokladů.

Hledání odpovědí na určitým způsobem zformulované otázky.

Příklady:

- Jak mnoho se liší průměrná míra nezaměstnanosti v našem okrese od celorepublikového průměru?
- Liší se významně údaje zjištěné dvěma různými metodami?
- Pochází výběr ze základního souboru, který má určité teoretické rozdělení?
- Je jedna metoda lepší než druhá?

Obecný postup testování

1. Formulace nulové hypotézy
2. Volba hladiny významnosti
3. Volba vhodného testovacího kritéria
4. Výpočet hodnoty testovacího kritéria z empirických dat
5. Porovnání vypočtené hodnoty s hodnotou kritickou nebo její převedení do pravděpodobnostní škály
6. Vyslovení závěru o výsledku testu (přijetí či zamítnutí nulové hypotézy)

Základní pojmy

- **Statistická hypotéza** – předpoklad o neznámé vlastnosti základního souboru.
- Prověřujeme tzv. **nulovou hypotézu** (H_0). Např. průměry výběrových souborů se neliší (pocházejí z jednoho základního souboru).
- Nulová hypotéza je obvykle opakem hypotézy pracovní (je obvykle opakem toho, co chceme výzkumem prokázat, když zahajujeme studii a začínáme sbírat data). Obvykle deklaruje „žádný rozdíl“
- **Alternativní hypotéza** (H_1) – situace, kdy H_0 neplatí. Tedy obvykle vyjadřuje „existenci diference“ či „existenci závislosti“
- Platnost hypotézy se prověřuje testem významnosti.

Základní pojmy

- Hypotéza může být dvoustranná a test dvoustranným
- Existují i jednostranné (pravostranné a levostranné) hypotézy

$H_0 \quad \mu = \mu_0$

$H_1 \quad \mu \neq \mu_0$

Jednostranný test

$H_1 \quad \mu > \mu_0$

$H_1 \quad \mu < \mu_0$

Základní pojmy

- **Hladina významnosti** (α) – pravděpodobnost, že náhodná odchylka překročí tzv. **kritickou hodnotu**.
- Volíme α co nejnižší (α = 0,05 či 0,01 tj. 5 % či 1 %).
- Odchylky, které se vyskytují s menší pravděpodobností než α jsou **statisticky významné** na zvolené hladině.

Obecný tvar testovacího kritéria:

$$\text{testová statistika} = \frac{\text{pozorovaná hodnota} - \text{očekávaná hodnota}}{\text{směrodatná chyba pozorované hodnoty}},$$

Testovou statistiku vyhodnotíme tak, že spočteme pravděpodobnost, že bychom mohli pozorovat námi zjištěnou, nebo ještě extrémnější (tj. méně pravděpodobnou) hodnotu, pokud by byla nulová hypotéza pravdivá.

Testovací kritérium

- Použité testovací kritérium musí odpovídat povaze problému.
- Každé testovací kritérium má své teoretické rozdělení.
- Ve statistických tabulkách jsou uvedeny **kritické hodnoty** testovacích kritérií pro běžně používané hladiny významnosti a běžné rozsahy výběrových souborů.
- Tyto rozsahy jsou většinou tabelovány v tzv. stupních volnosti.
- Pokud nejsou kritické hodnoty tabelovány (pro velká n) lze vypočítat pomocí SW

Dva způsoby hodnocení vypočteného testovacího kritéria

1. porovnání vypočtené hodnoty s hodnotou kritickou, kterou nalezneme v **tabulkách**

   - vypočteme hodnotu testovací statistiky
   - v tabulkách nalezneme tzv. kritickou hodnotu testovací charakteristiky pro zvolené α
   - obě hodnoty porovnáme

Hodnocení testovacího kritéria s využitím statistických tabulek

Výrok o platnosti či neplatnosti nulové hypotézy vyslovujeme na základě porovnání vypočtené hodnoty testovacího kritéria s hodnotou kritickou:

I. Vypočtené kritérium je větší než kritická hodnota

- Jedná se o případ, který jsme očekávali s nepatrnou pravděpodobností
- Takový případ je téměř **nemožný**.
- Testovaná odchylka tedy nemá náhodný charakter.
- Nulovou hypotézu **zamítáme** a rozdíl mezi testovanými charakteristikami je statisticky významný na zvolené hladině α .

Hodnocení testovacího kritéria s využitím statistických tabulek

- **II. Vypočtené kritérium je menší než kritická hodnota**
- Jedná se o případ, který jsme očekávali s pravděpodobností 1- α – tedy velmi vysokou
- Takový případ můžeme považovat za téměř **jistý**.
- Mezi testovanými charakteristikami není rozdílu.
- Nulovou hypotézu **přijímáme** a rozdíl mezi testovanými charakteristikami není statisticky významný na zvolené hladině α.

Dva způsoby hodnocení vypočteného testovacího kritéria

2. převedení hodnoty testovací statistiky do pravděpodobnostní škály na tzv. ***p hodnotu*** (hodnotu významnosti)

   (tento způsob hodnocení nabízejí počítačové programy)

PS 1* - t-test pro závislé vzorky (Prehrada_STA)

t-test pro závislé vzorky (Prehrada_STA)
Označ. rozdíly jsou významné na hlad. p < ,05000

Proměnná	Průměr	Sm.odch.	N	Rozdíl	Sm.odch. rozdílu	t	sv	p
Přítok	3,580000	3,104960						
Odtok	3,758152	3,136222	184	-0,178152	1,545406	-1,56371	183	0,119612

Hodnocení testovacího kritéria - výpočet *p hodnoty*

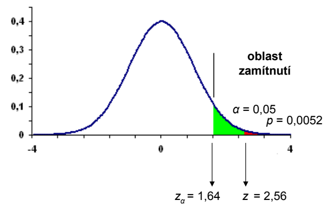
Protože má testovací kritérium určité teoretické rozdělení, každé jeho hodnotě přísluší určitá pravděpodobnost (***p hodnota***).

p hodnota odpovídá na otázku:
Jestliže H_0 platí, jaká je pravděpodobnost, že získáme právě vypočítanou či ještě neobvyklejší hodnotu testovací charakteristiky.
Je-li *p* hodnota malá, máme doklad, že H_0 neplatí.

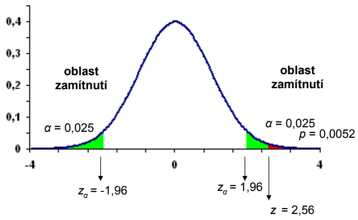
Interpretace *p hodnoty*

$p <= \alpha$ důkaz pro zamítnutí H_0
$p > \alpha$ nemáme důkaz pro zamítnutí H_0

Interpretace jednostranného testu



Interpretace dvoustranného testu



Při testování se můžeme dopustit dvou druhů chyb:

Chyba I. druhu – nulová hypotéza platí, ale zamítne se
Chyba II. druhu – nulová hypotéza neplatí, ale přijme se

		Závěr testu	
		H_0 platí	**H_0 neplatí**
Skutečnost	**H_0 platí**	správný	chyba I. druhu
	H_0 neplatí	chyba II. druhu	správný

Chyba I. druhu se omezuje volbou α. Čím menší hladinu významnosti zvolíme, tím menší je pravděpodobnost chyby I. druhu.

Naopak však ale roste pravděpodobnost chyby II. druhu.

Vztahy mezi chybami I. a II. druhu, síla testu:

Pravděpodobnost chyby I. druhu značíme α a lze ji vyjádřit jako podmíněnou pravděpodobnost:

P(chyba I. druhu | H0 platí) = α

Pravděpodobnost chyby II. druhu značíme β :

P(chyba II. druhu | H0 neplatí) = β

Opačné jevy k chybám I. a II. druhu

Spolehlivost testu: *(1- α)*
Síla testu: *(1- β)*

- Síla testu vyjadřuje, s jakou pravděpodobností zamítneme nulovou hypotézu, platí-li hypotéza alternativní
- Udává pravděpodobnost, že se nedopustíme chyby II. druhu

Vztahy mezi chybami I. a II. druhu:

Tabulka 3.5 Chyby I. a II. druhu a jejich pravděpodobnosti

Skutečnost / Úsudek o H_0	H_0 je pravdivá	Pravděpo-dobnost	H_0 je nepravdivá	Pravděpo-dobnost
Nezamítá se	správné rozhodnutí	$1-\alpha$	chyba II. druhu	$P(II.) = \beta$
Zamítá se	chyba I. druhu	$P(I.) = \alpha$	správné rozhodnutí	$1-\beta$
Celkem	x	1	x	1

Rozdělení testů

Testy parametrické – testy o charakteristikách základního souboru, testy o parametrech rozdělení základního souboru (testy o průměru, rozptylu, o shodě dvou průměrů, …).

Předpokládá se, že rozdělení základního souboru z něhož pochází výběr, je určité teoretické rozdělení (normální).

Neparametrické testy - nevíme nic o rozdělení základního souboru. Například ověřujeme předpoklad o normalitě. Patří sem:

Testy dobré shody, testy nezávislosti v kombinační tabulce, testy o shodě úrovně

Menší síla testů, sociologie, psychologie, …

Testy párové a nepárové

Příklad Z-testu, oboustranná alternativa

Ve výběru 216 vzorků byl zjišťován obsah rozpuštěných látek:

Průměr: 34,46 g/l

Směrodatná chyba: 0,397 g/l

H_0 průměr se neliší od průměru základního souboru (33,5 g/l) $\mu = \mu_0$

H_1 $\mu \neq \mu_0$

Protože měříme spojitou veličinu a rozsah výběru je velký – můžeme předpokládat normální rozdělení a použít tzv. **Z-testu**:

Testová charakteristika

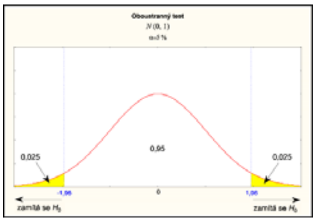
$$z = \frac{\text{výběrový průměr} - \text{očekávaný průměr při } H_0}{\text{směrodatná chyba výběrového průměru}} = \frac{34{,}46 - 33{,}5}{0{,}397} = 2{,}418.$$

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

Nalezneme kritickou hodnotu Z standardizovaného normálního rozdělení odpovídající 95% koeficientu spolehlivosti – nebo–li 5% hladině významnosti α: $Z_{1-0,5\alpha}$

$Z_{1-0,5\alpha} = 1{,}960$

Protože $Z > Z_{1-0,5\alpha}$ dostáváme na zvolené hladině významnosti významný výsledek – zamítáme H_0 – Průměr získaný ze vzorků se liší od průměru populace



Příklad Z-testu, jednostranná alternativa

Ve výběru 216 vzorků byl zjišťován obsah rozpuštěných látek:

Průměr: 34,46 g/l

Směrodatná chyba: 0,397 g/l

H_0 průměr je stejný jako průměr základního souboru (33,5 g/l) $\mu = \mu_0$

H_1 průměr je větší $\mu > \mu_0$

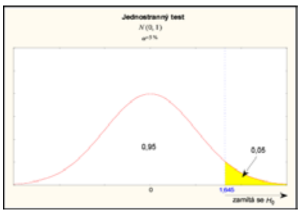
Testová charakteristika Z = 2,418

Kritická hodnota Z pro α = 0,05, tedy $Z_{1-\alpha} = 1{,}645$

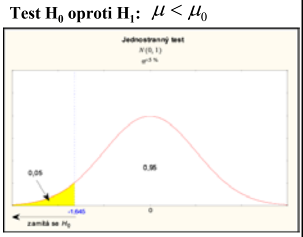
Protože $Z > Z_{1-\alpha}$ zamítáme H_0 – Průměr získaný ze vzorků je významně větší od průměru populace na 5 % hladině významnosti

Příklad Z-testu s jednostrannou alternativou

Test H_0 oproti H_1: $\mu > \mu_0$



Test H_0 oproti H_1: $\mu < \mu_0$



F - test

Používá se k testování významnosti rozdílu mezi dvěma rozptyly.

Testovací kritérium je definováno jako poměr odhadů dvou rozptylů základních souborů

$$F = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}$$

Odhady zjistíme z výběrových rozptylů ze vztahů:

$$\hat{\sigma}_1^2 = \frac{n_1}{n_1 - 1} \cdot s_1^2 \quad \text{a} \quad \hat{\sigma}_2^2 = \frac{n_2}{n_2 - 1} \cdot s_2^2$$

F - test

Do vzorce s testovacím kritériem F se dosazuje do čitatele vždy větší hodnota.

Počty stupňů volnosti: $\nu_1 = n_1 - 1 \qquad \nu_2 = n_2 - 1$

Kritické hodnoty veličiny F jsou tabelovány

Nulová hypotéza: $\hat{\sigma}_1^2 = \hat{\sigma}_2^2$

Předpokladem použití testu je alespoň přibližně normální rozdělení základních souborů.

F – test: obecný postup testování

1. zvolíme hladinu významnosti *α = 0,05* či *α = 0,01*
2. vypočteme odhady rozptylů základních souborů pomocí rozptylů výběrových souborů
3. vypočítáme hodnotu testovacího kritéria *F* (*F* musí být větší než 1)
4. určíme počty stupňů volnosti a pro daná a vyhledáme kritickou hodnotu $F_{\alpha/2}$
5. Porovnáme hodnotu *F* s kritickou hodnotou $F_{\alpha/2}$ a zhodnotíme výsledek

t - test

• Je vhodný pro testování rozdílů dvou veličin (např. průměru základního a výběrového souboru).

• Lze ho použít i pro testování rozdílu dvou výběrových průměrů jestliže *F* - testem ověříme významnost či nevýznamnost rozdílu odpovídajících rozptylů.

• Používá se i pro testování rozdílů párovaných hodnot.

• Předpokladem použití testu je alespoň přibližně normální rozdělení základního souboru a pro malé rozsahy souborů (*n < 30*)

Použití t - testu

1. Testování významnosti rozdílu výběrového průměru a známého průměru základního souboru:

Testovací kritérium:

$$t = \frac{|\bar{x} - \mu| \cdot \sqrt{n-1}}{s} \qquad \nu = n - 1$$

Protože za oblasti zamítnutí považujeme obě strany křivky t-rozdělení, je zapotřebí rozdělit zvolenou hladinu významnosti na poloviny a v tabulkách vyhledat kritické hodnoty t_α pro poloviční hodnoty.

Jestliže $t > t_\alpha$ zamítáme nulovou hypotézu – výběrový průměr se na zvolené hladině *α* statisticky významně liší od průměru základního souboru.

Použití t - testu

2. Testování významnosti rozdílu dvou průměrů pokud F-testem nezamítneme hypotézu $\hat{\sigma}_1^2 = \hat{\sigma}_2^2$.

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

$$\nu = n_1 + n_2 - 2$$

Použití t - testu

3. Testování významnosti rozdílu dvou průměrů pokud F-testem zjistíme, že mezi rozptyly je statisticky významný rozdíl $\hat{\sigma}_1^2 \neq \hat{\sigma}_2^2$

Testovací kritérium: $$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{n_1 - 1} + \frac{s_2^2}{n_n - 1}}}$$

Kritická hodnota t_α^+ $$t_\alpha^+ = \frac{t_\alpha^{'} \frac{s_1^2}{n_1 - 1} + t_\alpha^{''} \frac{s_2^2}{n_2 - 1}}{\frac{s_1^2}{n_1 - 1} + \frac{s_2^2}{n_2 - 1}}$$

Použití t - testu

Hodnota $t_{\alpha}^{'}$ značí kritickou hodnotu t-rozdělení pro $\nu_1 = n_1 - 1$

Hodnota $t_{\alpha}^{''}$ kritickou hodnotu pro $\nu_2 = n_2 - 1$

Kritické hodnoty lze najít v tabulkách (Brázdil a kol. 1995, příl. VII).

Postup testování je obdobný jako v případě výše uvedených testů.

Je-li $t > t_{\alpha}^{+}$ nulovou hypotézu zamítáme

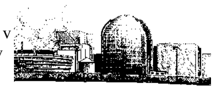
Na zvolené α je rozdíl průměrů významný.

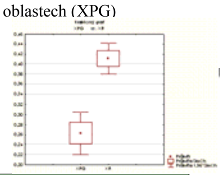
Příklad t - test

Statistika – Základní statistiky

T- test, nezávislé, dle proměnných

Zadání: Existuje statisticky významný rozdíl mezi průměrným obsahem Stroncia v mléce změřeným na farmách v blízkosti jaderné elektrárny (XR) a farmách v horských oblastech (XPG)





Data: prikl_testy* (2s krát ...

	1 XPG	2 XR
1	0,25	0,43
2	0,34	0,35
3	0,26	0,42
4	0,2	0,39
5	0,21	0,45
6	0,15	0,49
7	0,27	0,34
8	0,26	0,39
9	0,31	0,44
10	0,38	

Výsledek: Průměry se významně liší na hladině významnosti p=0,05

T-test pro nezávislé vzorky (prikl_testy)
Pozn.: Proměnné byly brány jako nezávislé vzorky

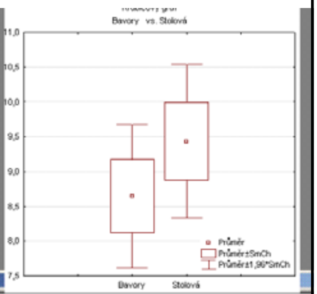
Skup. 1 vs. skup. 2	Průměr skup. 1	Průměr skup. 2	Hodnota t	sv	p	Poč.plat. skup. 1	Poč.plat. skup. 2	Sm.odch. skup. 1	Sm.odch. skup. 2	F-poměr rozptyly	p rozptyly
XPG vs. XR	0,263000	0,411111	-5,41256	17	0,000047	10	9	0,067995	0,048333	1,979073	0,349267

Příklad F-test, t - test

(Brázdil a kol. 1995, str. 114, cvičení č. 7.4)

Data: T_test (4s krát 31?)

	1 Bavory	2 Stolová	3 Prom3	4 Prom4
1	13,3	15		
2	12,2	9,9		
3	8,1	6,7		
4	8,9	9,1		
5	11,2	12,1		
6	10,3	9,4		
7	1,3	1,7		
8	6	6,2		
9	7,2	7,7		
10	8,9	9,2		
11	9,2	9		
12	7,1	9,3		
13	10,5	11,6		
14	9,3	10,7		
15	7	8,3		
16	14	13,6		



PS 3* - T-test pro nezávislé vzorky (T_test)

T-test pro nezávislé vzorky (T_test)
Pozn.: Proměnné byly brány jako nezávislé vzorky

Skup. 1 vs. skup. 2	Průměr skup. 1	Průměr skup. 2	Hodnota t	sv	p	Poč.plat. skup. 1	Poč.plat. skup. 2	Sm.odch. skup. 1	Sm.odch. skup. 2	F-poměr rozptyly	p rozptyly
Bavory vs. Stolová	8,645161	9,435484	-1,02842	60	0,307881	31	31	2,920027	3,127464	1,147125	0,709420

t - test pro párované hodnoty

Používá se v případě, že každý prvek jednoho výběru tvoří pár s určitým prvkem druhého výběru (např. provádíme dvě měření na stejném objektu za změněných podmínek).

Máme n párů na sobě závislých měření.

Postup testování: Vypočteme rozdíly d_i mezi oběma měřeními, průměr těchto rozdílů $\bar{d}$ a směrodatnou odchylku *sd*.

Předpokladem použití je opět normální rozdělení.

t - test pro párované hodnoty

Nulová hypotéza: $\mu_1 = \mu_2$

Počet stupňů volnosti: $\nu = n - 1$

Testovací kritérium: $$t = \frac{|\bar{d}|\sqrt{n-1}}{s_d}$$

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i \qquad s_d = \sqrt{\frac{1}{n}\sum_{i=1}^{n}|d_i - \bar{d}|^2}$$

t - test pro párované hodnoty

V případě zamítnutí nulové hypotézy ($t > t_{\alpha}$) lze stanovit 100.(1- α)% interval spolehlivosti rozdílu $\mu_1 - \mu_2$:

$$\bar{d} - t_{\alpha}\frac{s_d}{\sqrt{n-1}} \le \mu_1 - \mu_2 \le \bar{d} + t_{\alpha}\frac{s_d}{\sqrt{n-1}}$$

Pokud n >30, potom lze t-test nahradit tzv. z testem

Příklad t - test pro párované hodnoty

Statistika - Základní statistiky - T- test, závislé vzorky

Zadání: Existuje statisticky významný rozdíl v počtu bezobratlých živočichů zjištěných nad a pod výpustí z kanalizace (data zjištěná pro dvojice na 10 tocích)?

	1 REKA	2 NAD VYPUSTI	3 POD VYPUSTI
1	A	8	6
2	B	9	9
3	C	12	11
4	D	8	4
5	E	15	10
6	F	7	8
7	G	14	10
8	H	5	5
9	I	7	6
10	J	11	10

závislé vzorky (prikl_testy)

t-test pro závislé vzorky (prikl_testy)
Označ. rozdíly jsou významné na hlad. p < ,05000

Proměnná	Průměr	Sm.odch.	N	Rozdíl	Sm.odch. rozdílu	t	sv	p
NAD VYPUSTI	9,600000	3,272783						
POD VYPUSTI	7,900000	2,469818	10	1,700000	2,002776	2,684211	9	0,025033

t-test pro závislé vzorky (prikl_testy)
Označ. rozdíly jsou významné na hlad. p < ,01000

Proměnná	Průměr	Sm.odch.	N	Rozdíl	Sm.odch. rozdílu	t	sv	p
NAD VYPUSTI	9,600000	3,272783						
POD VYPUSTI	7,900000	2,469818	10	1,700000	2,002776	2,684211	9	0,025033

Výsledek:

Významný na hladině α =0,05

Pro α = 0,01 nevýznamný

z - test

Pokud n >30, potom lze t-test nahradit tzv. z-testem

testovací kritérium:

$$z = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Výhody z-testu:

• využití násobků směrodatné odchylky normovaného normálního rozdělení jako kritických hodnot

• kritické z hodnoty nemají stupně volnosti (normované rozdělení)

Tedy kritická hodnota 1,96 a menší indikuje pravděpodobnost větší nebo rovnu 0,05 – tedy nevýznamný výsledek

kritická hodnota větší než 2,576 indikuje pravděpodobnost menší než 0,01 – tj. vysoce významný rozdíl mezi testovanými hodnotami

Neparametrické testy

• Neznáme rozdělení základního souboru a chceme porovnávat úroveň hodnot v souboru či prokázat nezávislost znaků.
• Jsou vhodné pro hodnocení ordinálních dat či pro data intervalová nebo poměrová, která nemají normální rozdělení

Jsou založeny na těchto principech:
• počítáme četnost odchylek kladného a záporného znaménka od určité meze (**znaménkový test**)
• počítá se s **pořadovými čísly**, která jsou vstupním číselným hodnotám přiřazena po jejich setřídění podle velikosti (pořadové metody)

Patří sem například testy:
• testy dobré shody (CHI-kvadrát, K-S test)
• testy o shodě úrovně (Mann-Whitneyův test, Wilcoxonův test)
• testy nezávislosti v kombinační tabulce (CHI-kvadrát)

Mann- Whitney U - test

• Neparametrický ekvivalent t-testu. Lze ho využít i pro nenormální, silně asymetrická rozložení.

• Jako míru centrální tendence využívá ne průměr ale medián a k výpočtu testovacího kritéria využívá ne původních hodnot, ale pořadových čísel.

• Může být použit i pro data získaná na ordinální škále

Příklad: Porovnáváme zdravotní kondici stromů rostoucích v městě (Z – znečištěné prostředí) a ve volné krajině (Č – relativně čisté prostředí). Tuto zdravotní kondici posuzujeme podle stavu (barvy) olistění v šesti-stupňové škále

Mann- Whitney U test - příklad

Ordinální škála hodnocení zdravotní kondice stromů

6 – naprostá většina listů tmavě zelených
5 – ….
4 – …
3 – některé listy mají světlé skvrny
2 – ….
1 – podstatná část listoví má nažloutlou barvu

Máme k dispozici deset různých vzorků obou lokalit

Č	**4 5 4 4 5 6 6 6 6 3**
Z	**2 2 2 1 6 4 4 5 4 3**

Prvním krokem je přiřazení **pořadových čísel** jednotlivým měřením. Pro aplikaci uvedeného testu založeného na pořadí je vhodné, aby byla data uspořádána do jednoho sloupce s indikací, ke které skupině patří.

Mann- Whitney U test - příklad

Lokalita	Kondice	Pořadí	Výpočet hodnoty pořadí	Hodnota pořadového čísla
Z	1	1		1
Z	2	2	(2 + 3 +4)/3 = 3	3
Z	2	3		3
Z	2	4		3
Č	3	5	(5 + 6)/2 = 5,5	5,5
Z	3	6		5,5
Z	4	7	(7+8+9+10+11+12)/6 = 9,5	9,5
Č	4	8		9,5
Z	4	9		9,5
Č	4	10		9,5
Z	4	11		9,5
Č	4	12		9,5
Z	5	13	(13+14+15)/3 = 14	14
Č	5	14		14
Č	5	15		14
Č	6	16	(16+17+18+19+20)/5 = 18	18
Č	6	17		18
Z	6	18		18
Č	6	19		18
Č	6	20		18

$$\sum R_Z = 76 \quad \sum R_Č = 134$$

Mann- Whitney U test – testovací kritérium

Test je založen na výpočtu testovací statistiky U:

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - \sum R_1$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - \sum R_2$$

kde n_1 a n_2 jsou počty vzorků v jednotlivých výběrech

Výrazy $\sum R_1$ a $\sum R_2$ značí sumy pořadových čísel pro jednotlivé výběry.

Menší z hodnot U_1 a U_2 se bere jako testovací kritérium a porovnává se s tabulkovou hodnotou.

Mann- Whitney U test – příklad (pokrač.)

V našem příkladě: $\sum R_Č = 134$ $\sum R_Z = 76$

a pro $U_Č$ tedy

$$U_Č = n_Č n_Z + \frac{n_Č(n_Č+1)}{2} - \sum R_Č = 10 \cdot 10 + \frac{10(10+1)}{2} - 134 = 21$$

a analogicky pro U_Z:

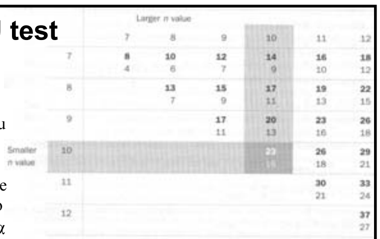
$$U_Z = n_Č n_Z + \frac{n_Z(n_Z+1)}{2} - \sum R_Z = 10 \cdot 10 + \frac{10(10+1)}{2} - 76 = 79$$

Menší z hodnot je tedy testovací kritérium U = 21

Mann- Whitney U test

Interpretace a vyslovení závěru o testování:

Statistický program určí hodnotu *p*, která přísluší vypočtené hodnotě testovacího kritéria a nebo se pro tuto hodnotu nalezne kritická hodnota v tabulkách pro zvolenou hladinu významnosti α a pro parametry n_1 a n_2.



Smaller n value \ Larger n value	7	8	9	10	11	12
7	**8** 4	**10** 6	**12** 7	**14** 9	**16** 10	**18** 12
8		**13** 7	**15** 9	**17** 11	**19** 13	**22** 15
9			**17** 11	**20** 13	**23** 16	**26** 18
10				**23** 16	**26** 18	**29** 21
11					**30** 21	**33** 24
12						**37** 27

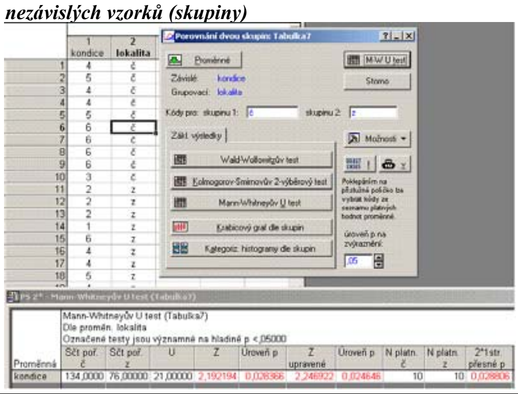
Horní čísla v tabulce odpovídají α *=0,05*, dolní potom α = *0,01*. V našem případě pro $n_1=10$ a $n_2=10$

Pro U test platí, že čím menší hodnota U, tím menší pravděpodobnost – interpretace je tedy opačná jako např. u t-testu

Na hladině významnosti 5% jsme prokázali statisticky významný rozdíl mezi zdravotní kondicí stromů rostoucích ve znečištěném a relativně čistém prostředí.

Neparametrické testy v programu Statistika

Statistika – Neparametrická statistika – Porovnání dvou nezávislých vzorků (skupiny)



Proměnná	Sčt poř. č	Sčt poř. z	U	Z	Úroveň p	Z upravené	Úroveň p	N platn. č	N platn. z	2*1str. přesné p
kondice	134,0000	76,00000	21,00000	2,192194	0,028366	2,246922	0,024646	10	10	0,028806

Test χ^2

Jedná se o test shody.

Testujeme, do jaké míry se liší rozložení četností empirického souboru od rozložení základního souboru.

Četnosti zjištěné při statistickém šetření (empirické):

$$n_{e,1}, n_{e,1}, \ldots, n_{e,j},$$

Četnosti získané z teoretického rozložení modelu (očekávané):

$$n_{t,1}, n_{t,1}, \ldots, n_{t,j},$$

Smyslem testu je hodnocení rozdílů v četnostech, tedy:

$$n_{e,j} - n_{t,j}$$

Test χ^2

Nulová hypotéza H_0: Četnosti $n_{e,j}$ a $n_{t,j}$ se liší pouze náhodně

Testovací kritérium:

$$\chi^2 = \sum_{j=1}^{k} \frac{(n_{e,j} - n_{t,j})^2}{n_{t,j}}$$

Ve výraze značí *k* počet skupin, do kterých je soubor tříděn.

Testovací kritérium má rozdělení χ^2 s $\nu = k - 1$ stupni volnosti.

Kritické hodnoty uvádí tabulky. Velké rozdíly v četnostech dávají velké hodnoty testovacího kritéria.

Test χ^2 - podmínky použití

Testu by se nemělo použít v případě, je-li a některá teoretická četnost $n_{t,j}$ je menší než 5.

Při $k > 2$ nemá být více než 20 % teoretických četností menších než 5 a žádná menší než 1.

Je možné sloučení některých četností – bez narušení smyslu úlohy.

Kolmogorovův – Smirnovův test

Tento test lze použít pro testování významnosti shody teoretického a empirického rozložení i v případech, kdy nelze použít CHÍ-kvadrát testu.

K-S test: postup testování I.

1. zvolíme hladinu významnosti α
2. roztřídíme zpracovávaná data do skupin
3. stanovíme příslušné teoretické četnosti
4. vypočítáme kumulativní četnosti empirického rozdělení $N_{e,j}$
5. vypočítáme kumulativní četnosti teoretického rozdělení $N_{t,j}$
6. stanovíme absolutní hodnoty rozdílů kumulovaných četností v odpovídajících skupinách
7. vypočteme hodnotu testovacího kritéria D

$$D = \frac{\max\left|N_{e,j} - N_{t,j}\right|}{n}$$

K-S test: postup testování II

8. Pro zvolenou hladinu významnosti p a dané n vyhledáme v tabulkách kritickou hodnotu D_α
9. V případě, že $D > D_\alpha$, potom zamítáme nulovou hypotézu a tvrdíme, že empirické a teoretické rozdělení se statisticky významně liší.

K-S test lze použít i pro srovnání dvou výběrových souborů.

Potom jako n bereme:

$$n = \frac{n_1 \cdot n_2}{n_1 + n_2}$$

Příklad použití χ^2 testu a K-S testu

Statistika – Prokládání rozdělení

Zadání: Testujeme, zda lze výběrový soubor proložit normálním rozložením (Existuje shoda empirických a teoretických četností?)

Proměnná: Praha, Rozdělení:Normální (cv 7 - nejnovější testy)
Kolmogorov-Smirnov. d = 0,06423, p = n.s., Lilliefors. p = n.s
Chí-kvadrát = 3,42976, sv = 4 (uprav.), p = 0,48864

Horní hranice	Pozorované Četnosti	Kumulativ. Pozorované	Procent Pozorované	Kumul. % Pozorované	Očekáv. Četnosti	Kumulativ. Očekáv.	Procent Očekáv.	Kumul. % Očekáv.	Pozorované - Očekáv.
<= 7,00000	0	0	0,00000	0,0000	0,54578	0,5458	0,45482	0,4548	-0,54578
7,50000	5	5	4,16667	4,1667	2,14258	2,6884	1,78548	2,2403	2,85742
8,00000	5	10	4,16667	8,3333	6,92131	9,6097	5,76775	8,0081	-1,92131
8,50000	17	27	14,16667	22,5000	15,72309	25,3328	13,10257	21,1106	1,27691
9,00000	22	49	18,33333	40,8333	25,12572	50,4585	20,93810	42,0487	-3,12572
9,50000	31	80	25,83333	66,6667	28,24932	78,7078	23,54110	65,5898	2,75068
10,00000	26	106	21,66667	88,3333	22,34738	101,0552	18,62282	84,2126	3,65262
10,50000	11	117	9,16667	97,5000	12,43754	113,4927	10,36462	94,5773	-1,43754
11,00000	0	117	0,00000	97,5000	4,86894	118,3617	4,05745	98,6347	-4,86894
11,50000	3	120	2,50000	100,0000	1,34023	119,7019	1,11686	99,7516	1,65977
< Nekonečno	0	120	0,00000	100,0000	0,29812	[illegible]	[illegible]	[illegible]	[illegible]

Výsledek:

Hodnota p je vysoká – není důvod zamítnout nulovou hypotézu.

Empirické a teoretické hodnoty se na hladině α= 5 % významně neliší

Výběrový soubor má normální rozdělení