

MASARYKOVA UNIVERZITA
 MASARYK JANA BRUNNENSTIS

Z1069 Statistické metody a zpracování dat
V. Testování statistických hypotéz

K čemu to je ? (příklad)

Má smysl se připravovat na písemný test ze statistiky?

Má to smysl!
 Průměr skupiny A je statisticky významně lepší než průměr skupiny B (na hladině významnosti 0.05)

K čemu to je?

Ověřování domněnek či předpokladů.
 Hledání odpovědi 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)

1. Formulace nulové hypotézy

- **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 difference“ či „existenci závislosti“
- Platnost hypotézy se prověřuje testem významnosti.

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

$H_0 \quad \mu = \mu_0$ Jednostranný test: $H_1 \quad \mu > \mu_0$
 $H_1 \quad \mu \neq \mu_0$ nebo: $H_1 \quad \mu < \mu_0$

2. Volba hladiny významnosti

- **Hladina významnosti** (α) – pravděpodobnost, že náhodná odchylka překročí tzv. **kritickou hodnotu**.
- Je to pravděpodobnost, že zamítneme nulovou hypotézu, ačkoliv ona platí. Je tedy vyjádřením míry ochoty, smířit se s touto chybou.
- Volíme α co nejnižší ($\alpha = 0,05$ či $0,01$ tj. 5% či 1%).
- Odchylky, které se vyskytují s menší pravděpodobností než α jsou **statisticky významné** na zvolené hladině.

3. a 4. Testovací kritérium a jeho výpočet

- Použité testovací kritérium musí odpovídat **povaze problému**.
- Testovací kritérium je **statistika vypočtená z našich dat** podle předepsaného vzorce
- Každé testovací kritérium **má své teoretické rozdělení**. Tedy každé vypočtené hodnotě testovací statistiky přísluší jistá pravděpodobnost výskytu
- **Interpretace testové statistiky**: spočetli jsme pravděpodobnost, že bychom mohli pozorovat námi zjištěnou, nebo ještě extrémnější (tj. méně pravděpodobnou) hodnotu, pokud by byla nulová hypotéza pravdivá.
- 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

5. Způsoby interpretace hodnoty testovacího kritéria

5.1 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

Příloha IV. Hodnoty, které náhodná veličina χ^2 překročí s danou pravděpodobností (p)

χ^2	0,100	0,050	0,025	0,010	0,005	0,001
1	0,445	0,675	0,988	1,637	2,706	3,841
2	0,575	0,788	1,054	1,876	2,915	3,998
3	0,713	0,900	1,213	2,179	3,219	4,348
4	0,838	1,064	1,384	2,486	3,541	4,751
5	0,949	1,237	1,549	2,793	3,858	5,051
6	1,048	1,413	1,700	3,070	4,151	5,348
7	1,136	1,591	1,848	3,341	4,431	5,617
8	1,214	1,771	1,990	3,592	4,698	5,878
9	1,282	1,953	2,124	3,838	4,957	6,131
10	1,350	2,137	2,262	4,079	5,209	6,377

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**:

Mohou nastat dvě situace:

I. Vypočtené kritérium je větší než kritická hodnota

- Jedná se o případ, který jsme očekávali s nepatrnou pravděpodobností α
- Takový případ je téměř **nemožný**.
- Testovaná odchylnka 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 $1 - \alpha$ – tedy velmi vysokou
- Takový případ můžeme považovat za téměř **jistý**.
- Mezi testovanými charakteristikami není rozdíl.
- Nulovou hypotézu **přijímáme** a rozdíl mezi testovanými charakteristikami není statisticky významný na zvolené hladině α .

Dva způsoby hodnocení vypočteného testovacího kritéria

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

(tento způsob hodnocení nabízejí počítačové programy)

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

vypočtená hodnota testovací statistiky
...a jí příslušející hodnota pravděpodobnosti (p-hodnota)

Hodnocení testovacího kritéria - výpočet p hodnoty

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

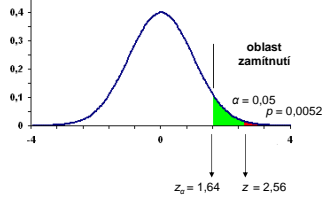
p hodnota odpovídá na otázku:

Jestliže H_0 platí, jaká je pravděpodobnost, že získáme právě vypočtenou či ještě neobvyklejší hodnotu testovací charakteristiky. Je-li p hodnota malá, máme důkaz, že H_0 neplatí.

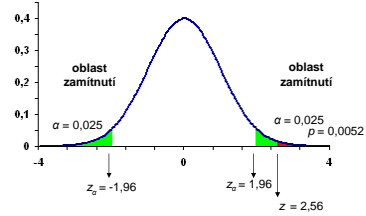
Interpretace p hodnoty

$p \leq \alpha$ důkaz pro zamítnutí H_0
 $p > \alpha$ nemáme důkaz pro zamítnutí H_0

6. Interpretace jednostranného testu



6. 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.

Rozdělení testů

Testy parametrické – testy o charakteristikách základního souboru, testy o parametrech rozdělení základního souboru (testy o průměru, rozptylu, o shodě dvou průměrů, ...). Data měřena na intervalové či poměrové škále.

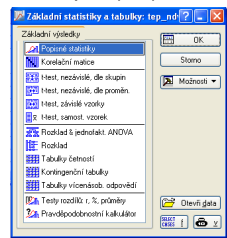
Předpokládá se, že rozdělení základního souboru z něhož pochází výběr, je určité teoretické rozdělení (normální).

Neparametrické testy - nevíme nic o rozdělení základního souboru. Data měřena na nominální či ordinální škále. Například ověřujeme předpoklad o normalitě. Patří sem:

testy dobré shody, testy nezávislosti v kombinační tabulce, ... Menší síla testů (sociologie, psychologie, ...)

Testy párové (závislé) a nepárové (nezávislé)

$$n_1 = n_2 \quad n_1 \text{ se nerovná } n_2$$

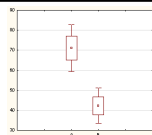


Příklad t-testu pro nezávislé vzorky

Liší se průměry skupin významně?

H₀: průměry se významně neliší

H₁: mezi průměry je významný rozdíl



Output from SPSS showing a t-test for independent samples. The data table shows scores for two groups (A and B) across 10 cases. The statistical results table shows a significant difference between the groups.

Skup. 1 vs. skup. 2	Průměr skup. 1	Průměr skup. 2	Hodnota t	df	p	Přít plat skup. 1	Přít plat skup. 2	Sm odch skup. 1	Sm odch skup. 2	F-poměr rozptyly	p rozptyly	
A (učit se) vs. B (společat na sístě)	71,00000	42,28571	3,744851	13	0,002451	0	0	7,1	16,87771	11,96821	1,988697	0,420434

Závěr: Průměr skupiny A je statisticky významně lepší než průměr skupiny B (na hladině významnosti 0,05)

F - test

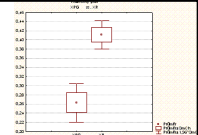
Používá se k testování významnosti rozdílů mezi dvěma rozptyly.

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: $v_1 = n_1 - 1$ $v_2 = n_2 - 1$

Kritické hodnoty veličiny F jsou tabelovány

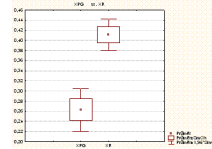
Nulová hypotéza: $\hat{\sigma}_1^2 = \hat{\sigma}_2^2$

Předpokladem použití testu je alespoň **přibližně normální rozdělení** základních souborů.

F – test: obecný postup testování

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

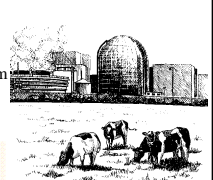
t - test



- 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 rozptýlů.
- Používá se i pro testování rozdílů **párovaných hodnot**.
- Předpokladem použití testu je alespoň **přibližně normální rozdělení** základního souboru a pro malé rozsahy souborů ($n < 30$)

Příklad t - test

Zadání: Ověřte, zda existuje statisticky významný rozdíl mezi průměrným obsahem stroncia (Sr) v mléce změněným na farmách v blízkosti jaderné elektrárny (XR) a farmách v horských oblastech (XPG).



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	p
XPG vs. XR	0,263000	0,411111	-5,41255	17	0,000047	10	9	0,067995	0,048333	1,979073	0,349267

Příklad F-test, t - test

Zadání: Zjistěte, zda existuje statisticky významný rozdíl ($\alpha = 0,05$) v mezi denními amplitudami teplot na dvou topoklimatických stanicích.

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	p
Bavory vs. Stelová	8,645161	9,435484	-1,02842	60	0,307881	31	31	2,920027	3,127464	1,147125	0,709423

Závěr: rozptýly ani průměry se významně neliší

Použití t - testu

Testování významnosti rozdílu výběrového průměru a známého průměru základního souboru:

Testovací kritérium:

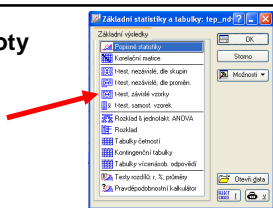
$$t = \frac{|\bar{x} - \mu| \cdot \sqrt{n-1}}{s} \quad v = n - 1$$

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.

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 rozdílu \bar{d} a směrodatnou odchylku s_d .

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

$$\bar{d} = \frac{1}{n} \sum_{i=1}^n d_i \quad s_d = \sqrt{\frac{1}{n} \sum_{i=1}^n (d_i - \bar{d})^2}$$

t - test pro párované hodnoty

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

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

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 výpusti 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

Výsledek:

Významný na hladině $\alpha = 0,05$

Pro $\alpha = 0,01$ nevýznamný

t-test pro závislé vzorky (párky, testy)							
Označ. rozdíly jsou významné na hlad. $p < .05000$							
Proměnná	Průměr	Sm. odch.	N	Rozdíl	Sm. odch. rozdílu	t	sv
NAD VYPUSTI	9,800000	3,272783					
POD VYPUSTI	7,800000	2,469818	10	1,700000	2,002776	2,684211	9
t-test pro závislé vzorky (párky, testy)							
Označ. rozdíly jsou významné na hlad. $p < .01000$							
Proměnná	Průměr	Sm. odch.	N	Rozdíl	Sm. odch. rozdílu	t	sv
NAD VYPUSTI	9,800000	3,272783					
POD VYPUSTI	7,800000	2,469818	10	1,700000	2,002776	2,684211	9

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 seřazení podle velikosti (pořadové metody)

Patří sem například:

- 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: Porovnáme 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	3
Z	2	2	2	1	6	4	4	5	4

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
C	3	5	(5+6)/2 = 5,5	5,5
Z	3	6		5,5
C	4	7	(7+8+9+10+11+12)/6 = 9,5	9,5
C	4	8		9,5
Z	4	9		9,5
C	4	10		9,5
C	4	11		9,5
C	4	12		9,5
Z	5	13	(13+14+15)/3 = 14	14
C	5	14		14
C	5	15		14
C	6	16	(16+17+18+19+20)/5 = 18	18
C	6	17		18
C	6	18		18
Z	6	19		18
C	6	20		18

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

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: $\sum R_C = 134 \quad \sum R_Z = 76$

a pro U_C tedy

$$U_C = n_C n_Z + \frac{n_C(n_C+1)}{2} - \sum R_C = 10 \cdot 10 + \frac{10(10+1)}{2} - 134 = 21$$

a analogicky pro U_Z :

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

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 .

Horní čísla v tabulce odpovídají $\alpha = 0,05$, dolní potom $\alpha = 0,01$. 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 než 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.

	7	8	9	10	11	12
7						
8						
9						
10						
11						
12						

Neparametrické testy v programu Statistika

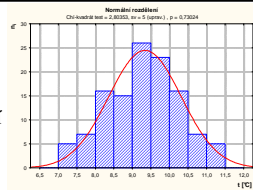
Statistika – Neparametrická statistika – Porovnání dvou nezávislých vzorků (skupiny)

Průměrná kondice	Č	Z	U	Uroven p	Z	Uroven p	N platn.	N platn. přesně p	Z1 str.
kondice	134,0000	76,00000	21,00000	2,192194	0,026368	2,246922	0,024646	10	10

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):

$$n_{e,1}, n_{e,2}, \dots, n_{e,j},$$

Četnosti získané z teoretického rozložení modelu (očekávané):

$$n_{t,1}, n_{t,2}, \dots, n_{t,j},$$

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 výrazu 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 n 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 CHI-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

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

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?)

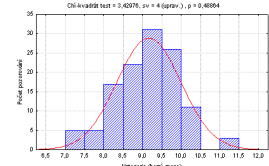
Hodnoty hranice	Pozorované četnosti		Procenta pozorované		Kumul. % pozorované		Očekávané četnosti		Kumulativní Očekávané		Procenta kumul. Očekávané		Pozorované - Očekávané	
	Četnosti	Pozorované	Pozorované	%	Pozorované	%	Očekávané	%	Očekávané	%	Očekávané	%	Pozorované - Očekávané	%
$\leq 7,00000$	0	0,00000	0,00000	0,54578	0,5458	0,45422	0,4548	0,4548	0,4548	0,4548	0,4548	0,4548	-0,54578	
7,50000	5	4,16667	4,16667	2,14286	2,6884	1,79548	2,2403	2,6884	1,79548	2,2403	1,79548	2,2403	2,85742	
8,00000	5	10	4,16667	8,33333	6,92131	5,60715	5,60715	8,3333	5,76775	8,3333	5,76775	8,3333	-1,92131	
8,50000	17	27	14,16667	23,90000	15,79393	25,3328	13,10257	21,1106	1,27691	21,1106	1,27691	21,1106	1,27691	
9,00000	22	49	18,33333	40,83333	35,12572	50,4586	20,93810	42,0487	-3,12572	42,0487	-3,12572	42,0487	-3,12572	
9,50000	31	60	25,83333	66,66667	28,24932	78,7078	23,54110	65,5896	2,75068	65,5896	2,75068	65,5896	2,75068	
10,00000	26	106	21,66667	89,33333	22,34739	101,0552	18,82282	94,2126	3,65262	94,2126	3,65262	94,2126	3,65262	
10,50000	11	117	9,16667	97,50000	12,43754	113,4927	10,36462	94,5773	-1,43754	94,5773	-1,43754	94,5773	-1,43754	
11,00000	0	117	0,00000	97,50000	4,86894	118,3617	4,05745	98,6347	4,86894	98,6347	4,86894	98,6347	4,86894	
11,50000	3	120	2,50000	100,00000	1,34023	119,7019	1,11896	99,7516	1,65977	99,7516	1,65977	99,7516	1,65977	
\leq Nekonečno	0	120	0,00000	100,00000	0,29812									

Výsledek:

Hodnota p je vysoká – není důvod zamítnout nulovou hypotézu.

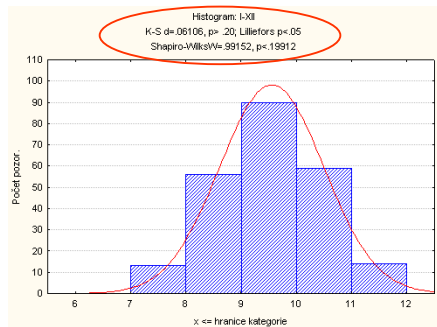
Empirické a teoretické hodnoty se na hladině $\alpha = 5\%$ významně neliší

Výběrový soubor má normální rozdělení



Posouzení normality pomocí grafů

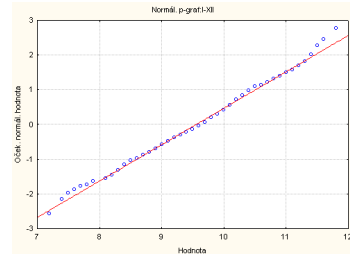
Histogram



Posouzení normality pomocí grafů

P-P graf (N-P graf)

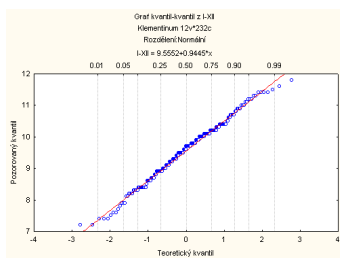
Normální pravděpodobnostní graf: umožňuje posoudit, zda data **mají normální rozdělení**. Na svislou osu vynášíme uspořádané hodnoty $x_1 < x_2 < x_3 < \dots < x_n$, a na vodorovnou osu potom kvantily normálního rozdělení. Vynesenými body se proloží přímka.



Posouzení normality pomocí grafů

Q-Q graf

Kvantil – kvantilový graf: umožňuje posoudit, zda data **pocházejí z nějakého známého rozdělení**. Na svislou osu vynášíme uspořádané hodnoty $x_1 < x_2 < x_3 < \dots < x_n$, a na vodorovnou osu potom kvantily vybraného rozdělení. Vynesenými body se proloží přímka.



Poznámky

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

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

$$P(\text{chyba I. druhu} \mid H_0 \text{ platí}) = \alpha$$

Pravděpodobnost chyby II. druhu značíme β :

$$P(\text{chyba II. druhu} \mid H_0 \text{ neplatí}) = \beta$$

Opačné jevy k chybám I. a II. druhu

Spolehlivost testu: $(1 - \alpha)$

Síla testu: $(1 - \beta)$

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

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**:

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

$$\hat{\sigma}_z = \frac{\hat{\sigma}}{\sqrt{n}} = \frac{s}{\sqrt{n-1}}$$

Příloha II. Distribuční funkce normálního rozdělení $F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{z^2}{2}} dz$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7258	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7518	.7549
.7	.7580	.7612	.7644	.7675	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7996	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9430	.9441
1.6	.9452	.9463	.9474	.9485	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9700	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9762	.9767
2.0	.9773	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9865	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9978	.9979	.9980	.9981

$\alpha = 0,05$
a tedy:
 $1 - 0,5\alpha = 0,9750$
 $Z_{1-0,5\alpha} = 1,96$

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

$Z_{1-0,5\alpha} = 1,96$

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)
 H_1 průměr je větší $\mu > \mu_0$ $\mu = \mu_0$

Testová charakteristika $Z = 2,418$

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

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

Příklad Z-testu s jednostrannou alternativou

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

Použití t - testu

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

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

$$v = n_1 + n_2 - 2$$

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

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_{α}^{\prime} značí kritickou hodnotu t-rozdělení pro $\nu_1 = n_1 - 1$

Hodnota $t_{\alpha}^{\prime\prime}$ 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 $t > t_{\alpha}^{\prime}$ nulovou hypotézu zamítáme

Na zvolené α je rozdíl průměrů významný.