

7. cvičení: Úlohy o více nezávislých náhodných výběrech z normálního a alternativního rozložení

Příklad 1.: Jsou známy měsíční tržby (v tisících Kč) tří prodavačů za dobu půl roku.

1. prodavač: 12 10 9 10 11 9
2. prodavač: 10 12 11 12 14 13
3. prodavač: 19 18 16 16 17 15

Na hladině významnosti 0,01 testujte hypotézu, že střední hodnoty tržeb všech tří prodavačů jsou stejné. Pokud zamítnete nulovou hypotézu, zjistěte, tržby kterých dvou prodavačů se liší na hladině významnosti 0,01.

Výsledky:

$$\text{Testová statistika: } F_A = \frac{S_A/(r-1)}{S_E/(n-r)} = \frac{142,3/2}{27,7/15} = 38,58$$

$$\text{Kritický obor: } W = \langle F_{0,99}(2,15), \infty \rangle = \langle 6,3589, \infty \rangle$$

Protože se testová statistika realizuje v kritickém oboru, nulovou hypotézu zamítáme na hladině významnosti 0,01. Tukeyova metoda prokázala, že na hladině významnosti 0,01 se liší tržby prodavačů 1, 3 a 2, 3.

Příklad 2.: Je dáno pět nezávislých náhodných výběrů o rozsazích 5, 7, 6, 8, 5, přičemž i -tý výběr pochází z rozložení $N(\mu_i, \sigma^2)$, $i = 1, \dots, 5$. Byl vypočten celkový součet čtverců $S_T = 15$ a reziduální součet čtverců $S_E = 3$. Na hladině významnosti 0,05 testujte hypotézu o shodě středních hodnot.

Výsledky:

$$\text{Testová statistika: } F_A = \frac{S_A/(r-1)}{S_E/(n-r)} = \frac{12/4}{3/26} = 26$$

$$\text{Kritický obor: } W = \langle F_{0,95}(4,26), \infty \rangle = \langle 2,7426, \infty \rangle$$

Protože se testová statistika realizuje v kritickém oboru, H_0 zamítáme na hladině významnosti 0,05.

Příklad 3.: Je dána neúplná tabulka ANOVA. Místo otazníků doplňte chybějící čísla.

zdroj variability	součet čtverců	stupně volnosti	podíl	F_A
skupiny	?	2	?	?
reziduální	16,033	?	?	-
celkový	17,301	35	-	-

Výsledek:

zdroj variability	součet čtverců	stupně volnosti	podíl	F_A
skupiny	1,268	2	0,634	1,304
reziduální	16,033	33	0,486	-
celkový	17,301	35	-	-

Příklad 4.: V jisté továrně se měřil čas, který potřeboval každý ze tří dělníků k uskutečnění téhož pracovního úkonu. Čas v minutách:

- 1. dělník: 3,6 3,8 3,7 3,5
- 2. dělník: 4,3 3,9 4,2 3,9 4,4 4,7
- 3. dělník: 4,2 4,5 4,0 4,1 4,5 4,4.

Na hladině významnosti 0,05 testujte hypotézu, že výkony těchto tří dělníků jsou stejné. Zamítnete-li nulovou hypotézu, určete, výkony kterých dělníků se liší na dané hladině významnosti 0,05.

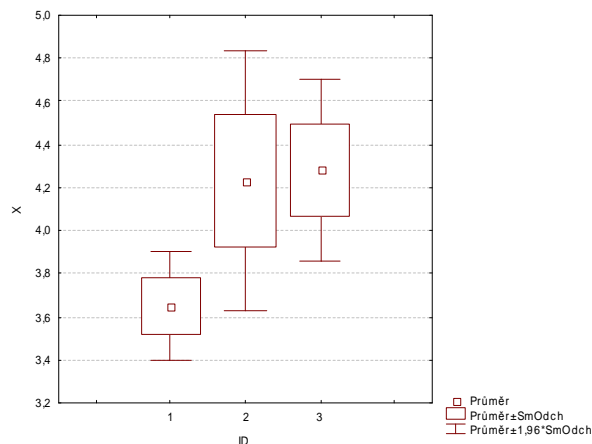
Výpočet pomocí systému STATISTICA:

Načteme datový soubor cas_delniku.sta. Proměnná X obsahuje zjištěné časy, proměnná ID nabývá hodnoty 1 pro 1. dělníka, hodnoty 2 pro 2. dělníka a hodnoty 3 pro 3. dělníka. Statistika – Základní statistiky/tabulky – Rozklad & jednofakt. ANOVA – Proměnné - Závislé X, Grupovací ID, OK, Kódy pro grupovací proměnné – Vše, OK, Výpočet: Tabulka statistik (zobrazí se průměry, směrodatné odchylky a rozsahy všech tří výběrů).

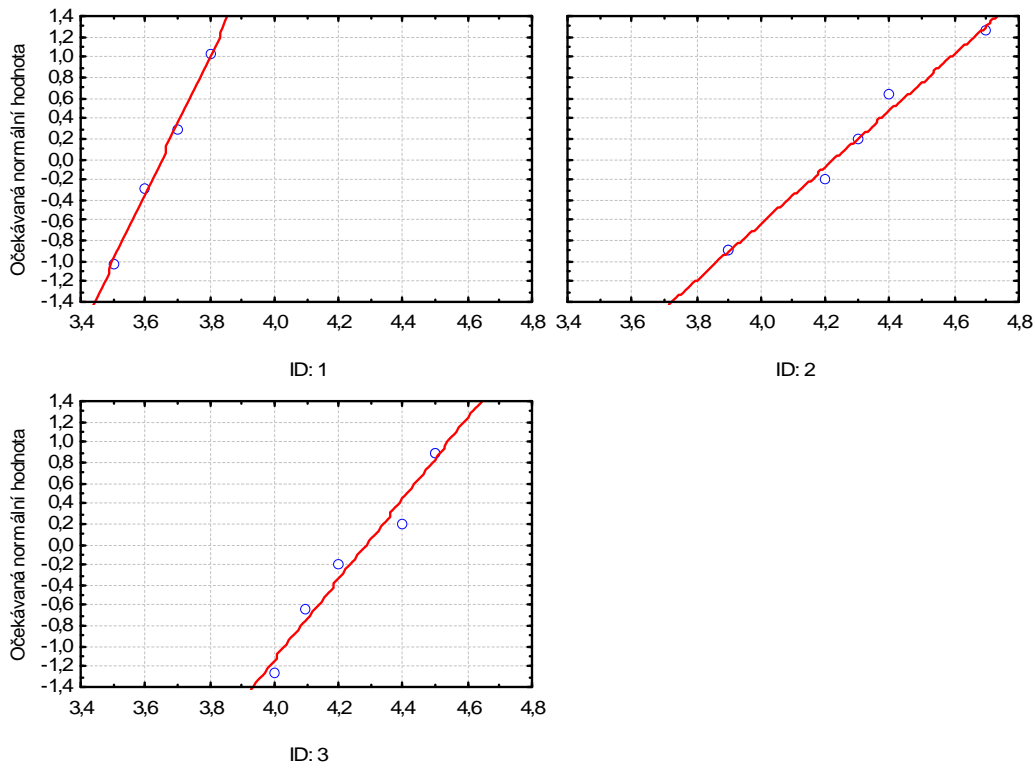
Rozkladová tabulka popisných statistik (cas_delniku.sta) N=16 (V seznamu záv. prom. nejsou ChD)			
ID	X průměr	X N	X Sm.odch.
1	3,650000	4	0,129099
2	4,233333	6	0,307679
3	4,283333	6	0,213698
Vš.skup.	4,106250	16	0,353023

Komentář: Na uskutečnění daného pracovního úkonu potřebuje nejkratší čas 1. dělník. Podává také nejvyrovnanější výkony – směrodatná odchylka proměnné X je u něj nejmenší. Naopak nejpomalejší je 3. dělník.

Nyní vytvoříme krabicové diagramy: Návrat do Statistika podle skupin – Kategoriz. krabicový graf (současné zobrazení krabicových diagramů pro všechny tři výběry)



Pomocí N-P plot orientačně posoudíme normalitu všech tří výběrů: Návrat do Statistika podle skupin – ANOVA & testy – Kategoriz. norm. pravd. grafy



Komentář: Ve všech třech případech se tečky jen málo odchylují od přímky, lze soudit, že data pocházejí z normálního rozložení.

Provedení testu o shodě rozptylů:

Návrat do Statistiky podle skupin – Leveneovy testy

Leveneův test homogenity rozptylů (cas_delniku.sta)								
Označ. efekty jsou význ. na hlad. p < ,05000								
Proměnná	SČ efekt	SV efekt	PČ efekt	SČ chyba	SV chyba	PČ chyba	F	p
X	0,042708	2	0,021354	0,183333	13	0,014103	1,514205	0,256356

Komentář: Testová statistika Levenova testu nabývá hodnoty 1,5142, stupně volnosti čitatele = 2, jmenovatele = 13, odpovídající p-hodnota = 0,256, tedy na hladině významnosti 0,05 se nezamítá hypotézu o shodě rozptylů.

Provedení testu o shodě středních hodnot:

Návrat do Statistiky podle skupin – Analýza rozptylu.

Analýza rozptylu (cas_delniku.sta)								
Označ. efekty jsou význ. na hlad. p < ,05000								
Proměnná	SČ efekt	SV efekt	PČ efekt	SČ chyba	SV chyba	PČ chyba	F	p
X	1,117708	2	0,558854	0,751667	13	0,057821	9,665327	0,002680

Komentář: Skupinový součet čtverců $S_A = 1,1177$, počet stupňů volnosti $f_A = 2$, reziduální součet čtverců $S_E = 0,7517$, počet stupňů volnosti $f_E = 13$, testová statistika $F_A = \frac{S_A/f_A}{S_E/f_E}$ nabývá hodnoty 9,6653, počet stupňů volnosti čitatele = 2, jmenovatele = 13, odpovídající p-

hodnota = 0,00268, tedy na hladině významnosti 0,05 se zamítá hypotéza o shodě středních hodnot .

Provedení Scheffého metody mnohonásobného porovnávání:
Návrat do Statistiky podle skupin – Post- hoc – Schefféův test.

		Scheffeho test; proměn.:X (cas_delniku.sta) Označ. rozdíly jsou významné na hlad. p < ,05000		
ID		{1} M=3,6500	{2} M=4,2333	{3} M=4,2833
1	{1}		0,008391	0,004705
2	{2}	0,008391		0,937504
3	{3}	0,004705	0,937504	

Komentář: Tabulka obsahuje p-hodnoty pro testování hypotéz o shodě středních hodnot všech dvojic výběrů. Výsledek Scheffého metody ukazuje, že na hladině významnosti 0,05 se liší výkony dělníků (1,2), (1,3) a neliší se (2,3).

Příklad 5.: V cestovní kanceláři zkoumali u 609 náhodně vybraných klientů, o jaké ubytování měli zájem (varianty apartmán, bungalov, hotel, stan) a zjišťovali též pohlaví klienta.

Typ ubytování	apartmán	bungalov	hotel	stan
Počet žen	12	27	208	33
Počet mužů	100	68	36	152

Na asymptotické hladině významnosti 0,05 testujte hypotézu, že rozdíly v typech ubytování mezi muži a ženami jsou způsobeny pouze náhodnými vlivy.

Výpočet pomocí systému STATISTICA:

Načteme datový soubor klienti_CK.sta. Proměnná POHLAVI obsahuje hodnotu 0 pro ženu, 1 pro muže. Proměnná TYP UBYTOVANI má hodnotu 1 pro apartmán, hodnotu 2 pro bungalov, hodnotu 3 pro hotel a hodnotu 4 pro stan.

Nejprve zjistíme podíly mužů v jednotlivých typech ubytování.

Statistiky – Základní statistiky/tabulky - Rozklad & jednofakt. ANOVA - OK - Proměnné - Závislé POHLAVI, Grupovací TYP UBYTOVANI, OK, Kódy pro grupovací proměnné – Vše, OK – Popisné statistiky - Výpočet: Tabulka statistik – ponecháme zaškrtnuto N - OK.

typ ubytovani	pohlavi průměr	pohlavi N
apartmán	0,892857	112
bungalov	0,602941	68
hotel	0,147541	244
stan	0,821622	185
Vš.skup.	0,540230	609

Komentář: Vidíme, že z těch klientů, kteří se ubytovali v apartmánu, bylo 89,3% mužů, mezi obyvateli bungalovů bylo 60,3% mužů, z ubytovaných v hotelu bylo mužů pouze 14,7% a z těch, kteří bydleli pod stanem, bylo 82,1% mužů.

Ověříme splnění podmínek dobré aproximace: $n_j m^* > 5$ pro všechna $j = 1, \dots, r$. Vážený průměr m^* se nachází v posledním řádku výstupní Rozkladové tabulky popisných statistik.

Jeho hodnotu okopírujeme do políček pro průměry relativní četnosti ubytovaných v jednotlivých typech ubytování, poslední řádek odstraníme a k tabulce přidáme jednu novou proměnnou, do jejíhož Dlouhého jména napíšeme =v2*v3.

typ ubytovani	pohlavi průměr	pohlavi N	NProm =v2*v3
apartmán	0,540230	112	60,505747
bungalov	0,540230	68	36,735632
hotel	0,540230	244	131,816092
stan	0,540230	185	99,942529

Komentář: Vidíme, že podmínky dobré aproximace jsou splněny.

Dále provedeme testování hypotézy o shodě parametrů čtyř alternativních rozložení. Statistika – Základní statistiky/tabulky – Kontingenční tabulky – OK - Specif. tabulky – List 1 POHLAVI, List 2 TYP UBYTOVANI, OK – Možnosti - Statistika dvou rozm tabulek - zaškrtneme Pearson & M-L Chi –square – Detailní výsledky – Detailní 2-rozm. tabulky

Statist.	Chí-kvadr.	sv	p
Pearsonův chí-kv.	267,6070	df=3	p=0,0000
M-V chí-kvadr.	294,9782	df=3	p=0,0000

Komentář: Testová statistika Q se realizuje hodnotou 267,6070, počet stupňů volnosti je 3, odpovídající p-hodnota = 0,0000, tedy na asymptotické hladině významnosti 0,05 hypotézu H_0 zamítáme. S rizikem omylu nejvýše 0,05 jsme tedy prokázali, že rozdíly v podílech klientů a klientek ubytovaných v různých typech ubytovacích zařízení nelze vysvětlit pouze náhodnými vlivy.

Nakonec provedeme metodu mnohonásobného porovnávání, abychom zjistili, které dvojice typů ubytování se liší na asymptotické hladině významnosti 0,05.

Návrat do Statistika podle skupin – Post- hoc – Schefféův test.

typ ubytovani	{1}	{2}	{3}	{4}
	M=,89286	M=,60294	M=,14754	M=,82162
apartmán {1}		0,000016	0,000000	0,471207
bungalov {2}	0,000016		0,000000	0,000797
hotel {3}	0,000000	0,000000		0,000000
stan {4}	0,471207	0,000797	0,000000	

Komentář: Tabulka obsahuje p-hodnoty pro testování hypotéz o shodě středních hodnot všech dvojic výběrů. Výsledek Scheffého metody ukazuje, že z hlediska podílu mužů se na hladině významnosti 0,05 neliší pouze ubytování v apartmánu a ve stanu.