

Cvičení 4: Úlohy o více nezávislých náhodných výběrech

Příklad na ANOVU

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.

Návod:

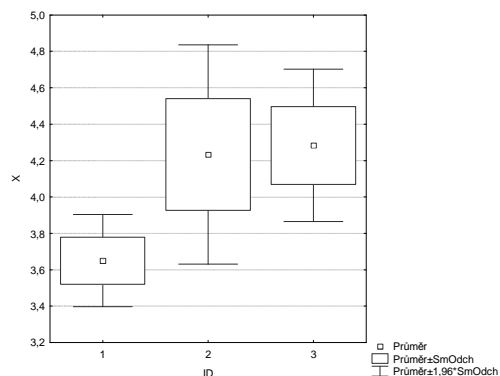
Úloha vede na analýzu rozptylu jednoduchého třídění. Načteme datový soubor cas_delniku.sta. Proměnná X obsahuje zjištěné časy, proměnná ID nabývá hodnoty 1 pro 1. dělníka, hodnoty 2 pro 2. dělníka a hodnoty 3 pro 3. dělníka.

Statistiky – 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

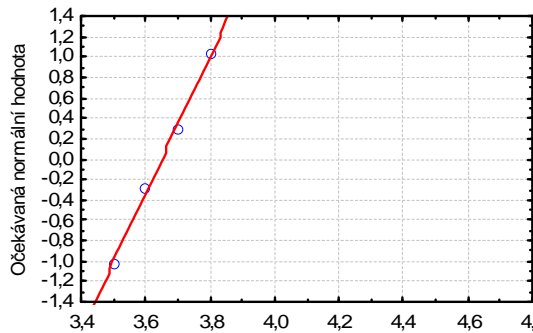
Komentář: Na uskutečnění daného pracovního úkonu potřebuje nejkratší čas 1. dělník. Podává také nejvyrovnanější výkony – směrodatná odchylka proměnné X je u něj nejmenší. Naopak nejpomalejší je 3. dělník.

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

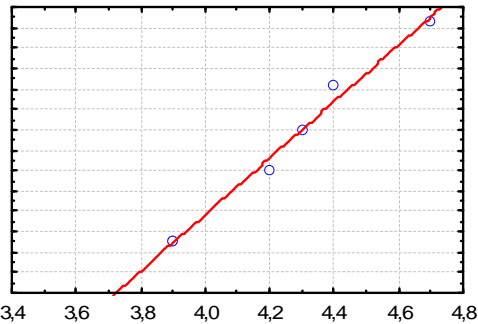


Pomocí N-P plot orientačně posoudíme normalitu všech tří výběrů:

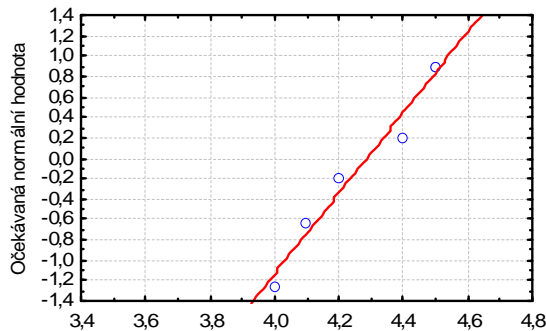
Návrat do Statistiky podle skupin – ANOVA & testy – Kategoriz. norm. pravd. grafy



ID: 1



ID: 2



ID: 3

Komentář: Ve všech třech případech se tečky jen málo odchyľují 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 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		
		{1}	{2}	{3}
ID		M=3,6500	M=4,2333	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 na Kruskalův – Wallisův test a mediánový test

Příklad: V roce 1980 byly získány tři nezávislé výběry obsahující údaje o průměrných ročních příjmech (v tisících dolarů) čtyř sociálních skupin ve třech různých oblastech USA.

jižní oblast: 6 10 15 29

pacifická oblast: 11 13 17 131

severovýchodní oblast: 7 14 28 25

Na hladině významnosti 0,05 testujte hypotézu, že příjmy v těchto oblastech se neliší.

Návod:

Načteme datový soubor `prijmy_v_USA.sta`. Je zřejmé, že kvůli vybočující hodnotě 131 ve 2. výběru bude porušena normalita, proto musíme použít neparametrický test.

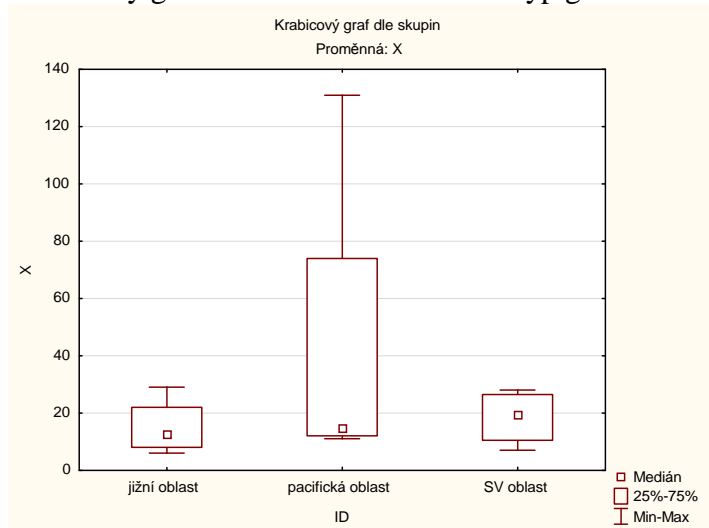
Statistiky – Neparametrická statistika – Porovnání více nezávislých vzorků - OK – Seznam závislých proměnných X, Nezáv. (grupovací) proměnná ID – OK – Shrnutí: Kruskal-Wallisova ANOVA a mediánový test. Ve dvou výstupních tabulkách se objeví výsledky K-W testu a mediánového testu.

		Kruskal-Wallisova ANOVA založ. na poř.; X (prijmy_v_USA.sta) Nezávislá (grupovací) proměnná : ID Kruskal-Wallisův test: $H(2, N=12) = 5,500000$ $p = ,7788$			
Závislá: X		Kód	Počet platných	Součet pořadí	Prům. Pořadí
jižní oblast		1	4	22,00000	5,500000
pacifická oblast		2	4	29,00000	7,250000
SV oblast		3	4	27,00000	6,750000

		Mediánový test, celk. medián = 14,5000; X (prijmy_v_USA.sta)			
Závislá:		Nezávislá (grupovací) proměnná : ID			
X		Chi-Kvadr. = 0,000000 sv = 2 p = 1,000			
		jižní oblast	pacifická oblast	SV oblast	Celkem
<= Medián: pozorov.		2,000000	2,000000	2,000000	6,00000
očekáv.		2,000000	2,000000	2,000000	
poz.-oč.		0,000000	0,000000	0,000000	
> Medián: pozorov.		2,000000	2,000000	2,000000	6,00000
očekáv.		2,000000	2,000000	2,000000	
poz.-oč.		0,000000	0,000000	0,000000	
Celkem: oček.		4,000000	4,000000	4,000000	12,00000

Testy nezamítají hypotézu o shodě mediánů v daných třech skupinách.

Grafické znázornění výsledků: návrat do Kruskal-Wallisova ANOVA a mediánový test – Krabicový graf – Proměnná X – OK – Typ grafu: Medián/kvartily/Rozpětí – OK.



Je vidět, že úroveň příjmů je nejvyšší pro pacifickou oblast, zatímco pro jižní oblast je nejnižší.

Příklady k samostatnému řešení

Příklad 1.: Studenti byli vyučováni předmětu za využití pěti pedagogických metod: tradiční způsob, programová výuka, audiotechnika, audiovizuální technika a vizuální technika. Z každé skupiny byl vybrán náhodný vzorek studentů a všichni byli podrobeni těmto písemnému testu. Výsledky testu:

metoda	počet bodů								
tradiční	76,2	48,3	85,1	63,7	91,6	87,2			
programová	85,2	74,3	76,5	80,3	67,4	67,9	72,1	60,4	
audio	67,3	60,1	55,4	72,3	40				
audiovizuální	75,8	81,6	90,3	78	67,8	57,6			
vizuální	50,5	70,2	88,8	67,1	77,7	73,9			

Na hladině významnosti 0,05 testujte hypotézu, že znalosti všech studentů jsou stejné a nezávisí na použité pedagogické metodě. V případě zamítnutí hypotézy zjistěte, které výběry se liší na hladině významnosti 0,05. Data jsou uložena v souboru pet_metod.sta.

Výsledek: ANOVA neprokázala rozdíl v účinnosti jednotlivých pedagogických metod.

Příklad 2.: Pan Novák může cestovat z místa bydliště do místa pracoviště třemi různými způsoby: tramvají (způsob A), autobusem (způsob B) a metrem s následným přestupem na tramvaj (způsob C). Máme k dispozici jeho naměřené časy cestování do práce v době ranní špičky (včetně čekání na příslušný spoj) v minutách:

způsob A: 32, 39, 42, 37, 34, 38:

způsob B: 30, 34, 28, 26, 32,

způsob C: 40, 37, 31, 39, 38, 33, 34

Pro všechny tři způsoby dopravy vypočítejte průměrné časy cestování. Na hladině významnosti 0,05 testujte hypotézu, že doba cestování do práce nezávisí na způsobu dopravy. V případě zamítnutí nulové hypotézy zjistěte, které způsoby dopravy do práce se od sebe liší na hladině významnosti 0,05. Data jsou uložena v souboru doby_cestovani.sta.

Výsledek: ANOVA zamítla hypotézu o shodě středních hodnot. S rizikem omylu nejvýše 5% se liší cestování tramvají a autobusem a dále cestování autobusem a metrem.

Příklad 3.: Voda po holení jisté značky se prodává ve čtyřech různých lahvičkách stejného obsahu. Údaje o počtu prodaných lahviček za týden v různých obchodech:

1.typ: 50 35 43 30 62 52 43 57 33 70 64 58 53 65 39

2.typ: 31 37 59 67 44 49 54 62 34 42 40

3.typ: 27 19 32 20 18 23

4.typ: 35 39 37 38 28 33.

Posuďte na 5% hladině významnosti, zda typ lahvičky ovlivňuje úroveň prodeje.

Data jsou uložena v souboru voda po holeni.sta.

Výsledek: Vzhledem k porušení homogenity rozptylu nelze použít ANOVU. Mediánový test zamítl hypotézu o shodě mediánů. Liší se typy (1, 3) a (2, 3).