

Téma 10: Analýza rozptylu jednoduchého třídění

Úkol 1.: 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:

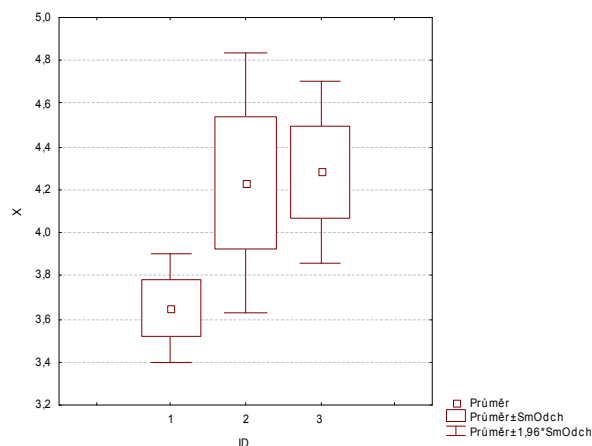
Úloha vede na analýzu rozptylu jednoduchého třídění. Načteme datový soubor cas_delniku.sta. Proměnná X obsahuje zjištěné časy, proměnná ID nabývá hodnoty 1 pro 1. dělníka, hodnoty 2 pro 2. dělníka a hodnoty 3 pro 3. dělníka.

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

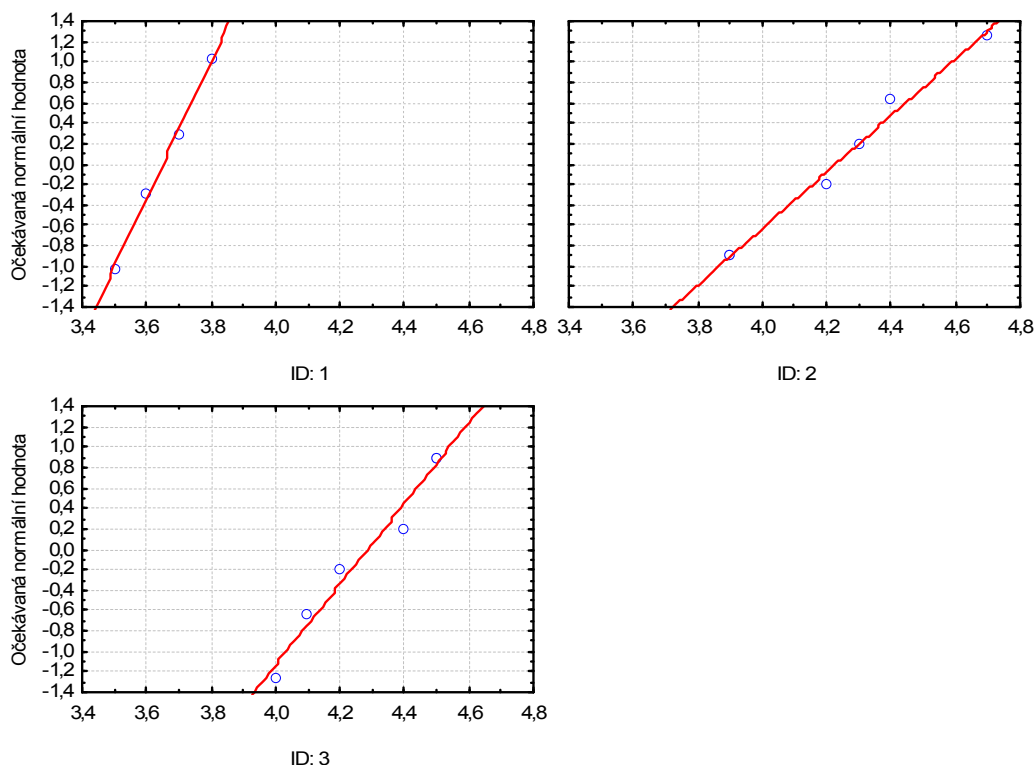
Komentář: Na uskutečnění daného pracovního úkonu potřebuje nejkratší čas 1. dělník. Podává také nejvyrovnanější výkony – směrodatná odchylka proměnné X je u něj nejmenší. Naopak nejpomalejší je 3. dělník.

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



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 metody mnohonásobného porovnávání (Scheffého test – viz skriptu Základní statistické metody, věta 8.2.2.1.):
Návrat do do Statistiky podle skupin – Post- hoc – Scheffěův test.

Scheffeho test; proměn.:X(cas_delniku.sta) Označ. rozdíly jsou významné na hlad. p < ,05000				
ID	{1}	{2}	{3}	
	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).

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 témuž 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.

Řešení:

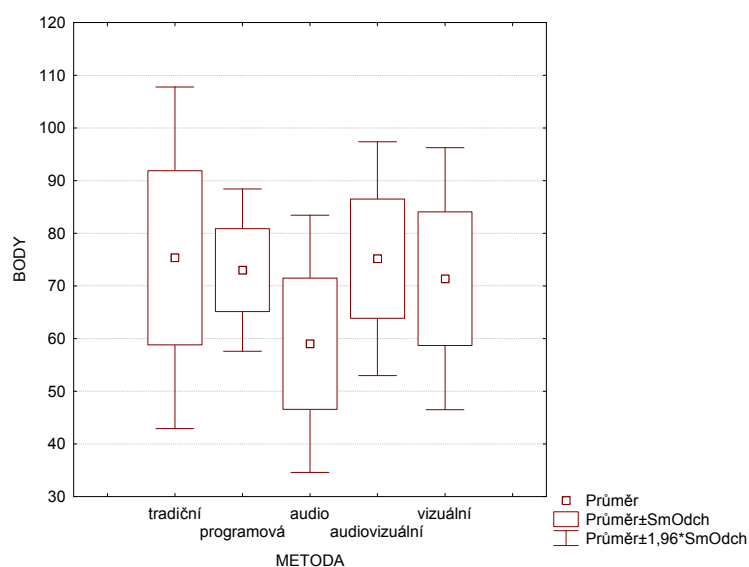
Načteme datový soubor pet_metod.sta. Proměnná BODY obsahuje dosažené počty bodů a proměnná METODA označení příslušné pedagogické metody.

Nejprve vypočteme průměry, směrodatné odchylky a rozsahy všech tří výběrů:

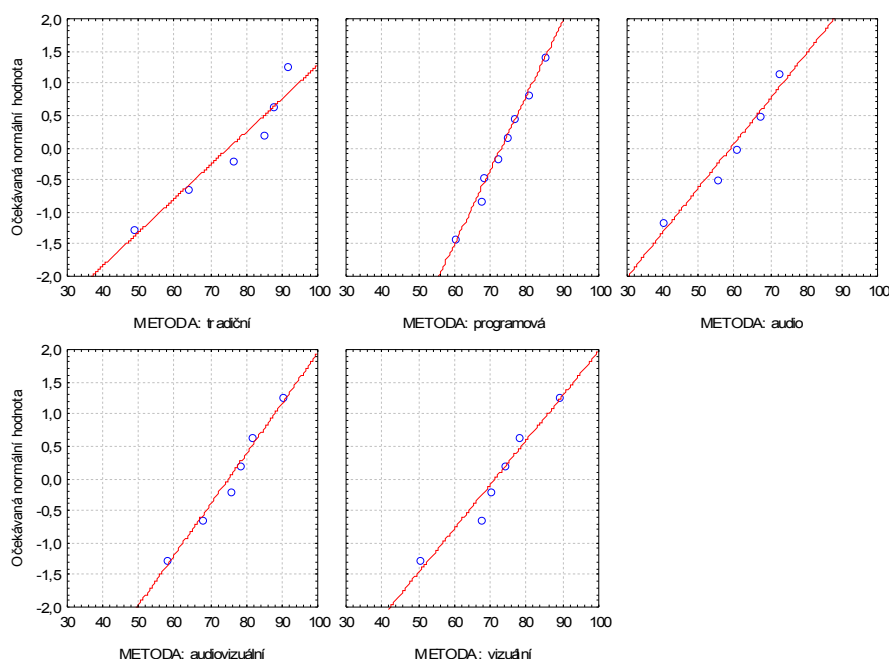
Rozkladová tabulka popisných statistik (pet_metod.sta)			
N=31 (V seznamu záv. prom. nejsou ChD)			
METODA	BODY průměr	3ODY N	BODY Sm.odch.
tradiční	75,35000	6	16,53901
programová	73,01250	8	7,86501
audio	59,02000	5	12,45941
audiovizuální	75,18333	6	11,32862
vizuální	71,36667	6	12,69199
Vš.skup.	71,30968	31	12,69534

Komentář: Nejlepších výsledků dosahují studenti vyučování tradiční metodou, podávají však nejméně vyrovnané výkony (počty bodů v této skupině mají největší směrodatnou odchylku). Naopak nejhoršího výsledku dosáhli studenti vyučování audio metodou. Nejvyrovnanější výkony pozorujeme u studentů vyučovaných programovou metodou.

Vytvoříme krabicové diagramy:



Pomocí N-P grafů vizuálně posoudíme normalitu všech pěti výběrů:



Komentář: Ze vzhledu N-P grafů je patrné, že předpoklad normality je ve všech pěti případech oprávněný.

Provedeme Levenův test (testování homogenity rozptylů všech pěti výběrů)

Levenův test homogenity rozptylů (pet_metod.sta)								
Označ. efekty jsou význ. na hlad. $p < ,05000$								
Proměnná	SČ efekt	SV efekt	PČ efekt	SČ chyba	SV chyba	PČ chyba	F	p
BODY	162,4883	4	40,62208	1289,544	26	49,59783	0,819029	0,524791

Komentář: Testová statistika F se realizuje hodnotou 0,819, počet stupňů volnosti čitatele = 4, jmenovatele = 26, odpovídající p-hodnota = 0,5248, na hladině významnosti 0,05 tedy nezamítáme hypotézu o shodě rozptylů.

Budeme testovat hypotézu o shodě středních hodnot všech pěti výběrů:

Analýza rozptylu (pet_metod.sta)								
Označ. efekty jsou význ. na hlad. $p < ,05000$								
Proměnná	SČ efekt	SV efekt	PČ efekt	SČ chyba	SV chyba	PČ chyba	F	p
BODY	966,3737	4	241,5934	3868,773	26	148,7990	1,623623	0,198252

Komentář: Testová statistika F se realizuje hodnotou 1,6236, počet stupňů volnosti čitatele = 4, jmenovatele = 26, odpovídající p-hodnota = 0,1983, na hladině významnosti 0,05 tedy nezamítáme hypotézu o shodě středních hodnot. Znamená to, že s rizikem omylu nejvýše 5% se neprokázal 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.

Řešení:

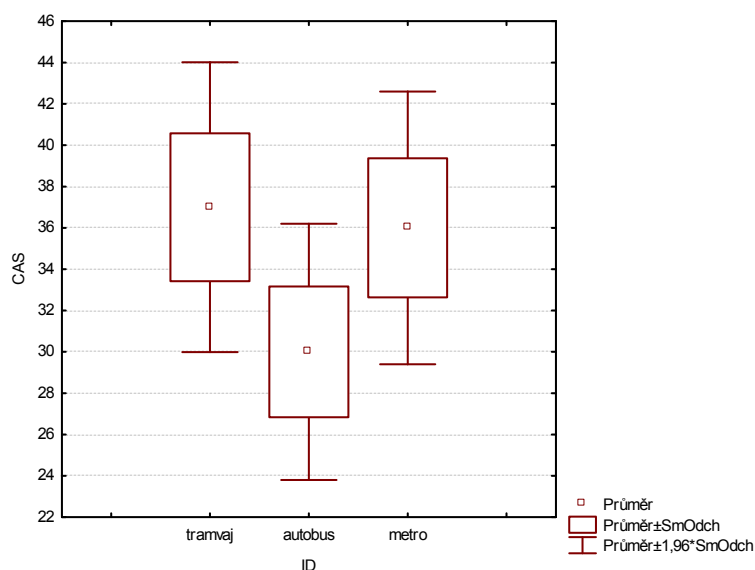
Načteme datový soubor doby_cestovani.sta. Proměnná CAS obsahuje zjištěné doby cestování a proměnná ID označení příslušného způsobu dopravy.

Nejprve vypočteme průměry, směrodatné odchylky a rozsahy všech tří výběrů:

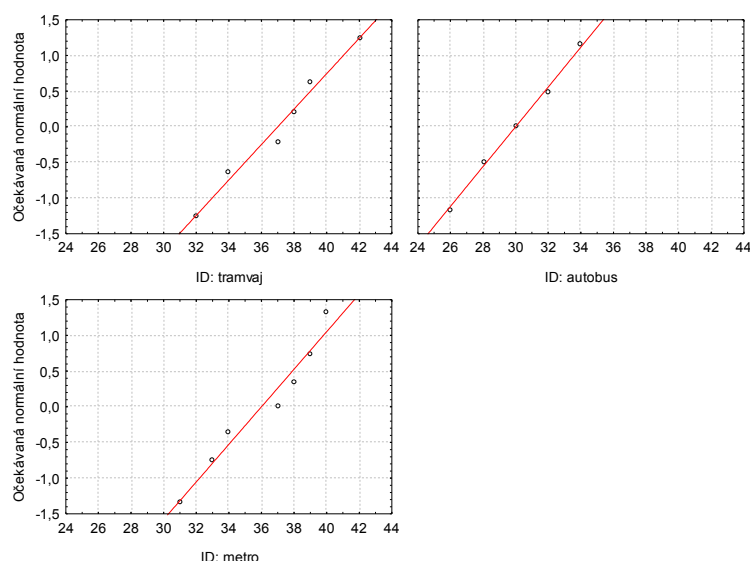
Rozkladová tabulka popisných statistik (doby_cestovani.sta)			
N=18 (V seznamu záv. prom. nejsou ChD)			
ID	CAS průměr	CAS N	CAS Sm.odch.
tramvaj	37,00000	6	3,577709
autobus	30,00000	5	3,162278
metro	36,00000	7	3,366502
Vš.skup.	34,66667	18	4,379095

Komentář: Nejkratší průměrnou dobu do zaměstnání pan Novák cestuje, když použije autobus, naopak nejdéle cestuje tramvají. Variabilita dob jednotlivých způsobů cestování je vcelku vyrovnaná.

Vytvoříme krabicové diagramy:



Pomocí N-P grafů vizuálně posoudíme normalitu všech tří výběrů:



Komentář: Ze vzhledu N-P grafů je patrné, že předpoklad normality je ve všech třech případech oprávněný.

Provedeme Levenův test (testování homogenity rozptylů všech tří výběrů)

Levenův test homogenity rozptylů (doby_cestovani.sta) Označ. efekty jsou význ. na hlad. $p < ,05000$								
Proměnná	SČ efekt	SV efekt	PČ efekt	SČ chyba	SV chyba	PČ chyba	F	p
CAS	0,609524	2	0,304762	43,39048	15	2,892698	0,105356	0,900665

Komentář: Testová statistika F se realizuje hodnotou 0,1054, počet stupňů volnosti čitatele = 2, jmenovatele = 15, odpovídající p-hodnota = 0,9007, na hladině významnosti 0,05 tedy nezamítáme hypotézu o shodě rozptylů.

Budeme testovat hypotézu o shodě středních hodnot všech tří výběrů:

Analýza rozptylu (doby_cestovani.sta) Označ. efekty jsou význ. na hlad. $p < ,05000$								
Proměnná	SČ efekt	SV efekt	PČ efekt	SČ chyba	SV chyba	PČ chyba	F	p
CAS	154,0000	2	77,00000	172,0000	15	11,46667	6,715116	0,008267

Komentář: Testová statistika F se realizuje hodnotou 6,7151, počet stupňů volnosti čitatele = 2, jmenovatele = 15, odpovídající p-hodnota = 0,0083, na hladině významnosti 0,05 tedy zamítáme hypotézu o shodě středních hodnot. Znamená to, že s rizikem omylu nejvýše 5% se prokázal rozdíl v dobách cestování pana Nováka do zaměstnání autobusem, tramvají a metrem.

Scheffého metodou mnohonásobného porovnávání zjistíme, které dvojice způsobů cestování do zaměstnání se liší na hladině významnosti 0,05:

Scheffeho test; proměn.:CAS (doby_cestovani.sta) Označ. rozdíly jsou významné na hlad. p < ,05000			
ID	{1} M=37,000	{2} M=30,000	{3} M=36,000
tramvaj {1}		0,013410	0,869732
autobus {2}	0,013410		0,028046
metro {3}	0,869732	0,028046	

Komentář: Z tabulky vyplývá, že s rizikem omylu nejvýše 5% se liší cestování tramvají a autobusem a dále cestování autobusem a metrem.