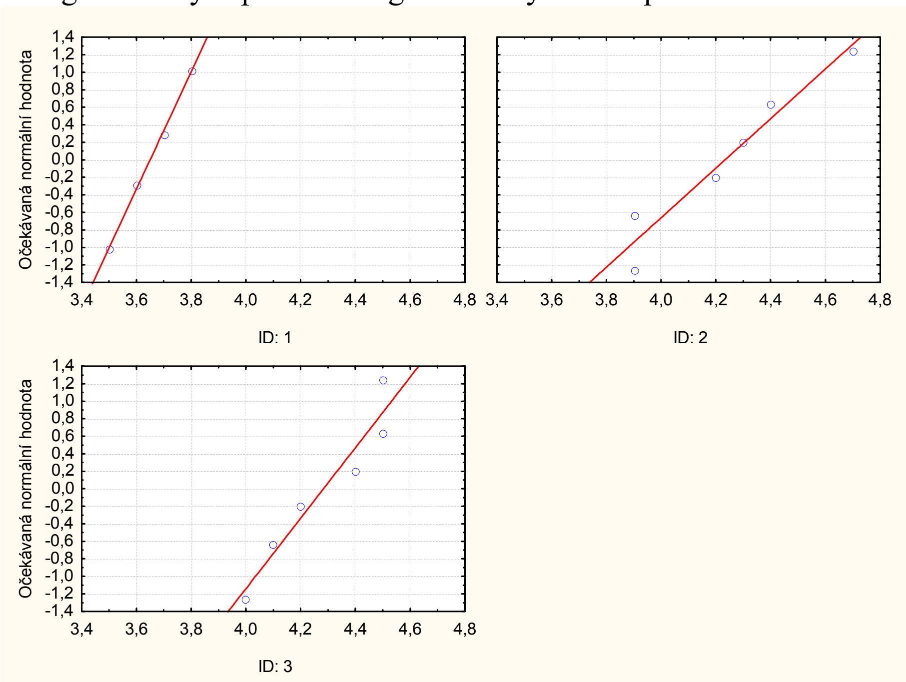
Téma 10: Analýza rozptylu jednoduchého třídění

Vzorový úkol: 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 průměrné výkony těchto tří dělníků jsou stejné. Zamítnete-li nulovou hypotézu, určete, výkony kterých dělníků se liší na dané hladině významnosti.
Návod: Vytvoříme datový soubor se dvěma proměnnými (X a ID) a 16 případy. Do 1. sloupce napíšeme změřené časy, do 2. sloupce dáme čtyřikrát jedničku, šestkrát dvojku a šestkrát trojku.
Nejprve ověříme předpoklad o normalitě všech tří výběrů:
Grafy – 2D Grafy – Normální prvaděpodobnostní grafy – Proměnné X – OK – na záložce Kategorizovaný zapneme kategorii X a vybereme proměnnou ID.

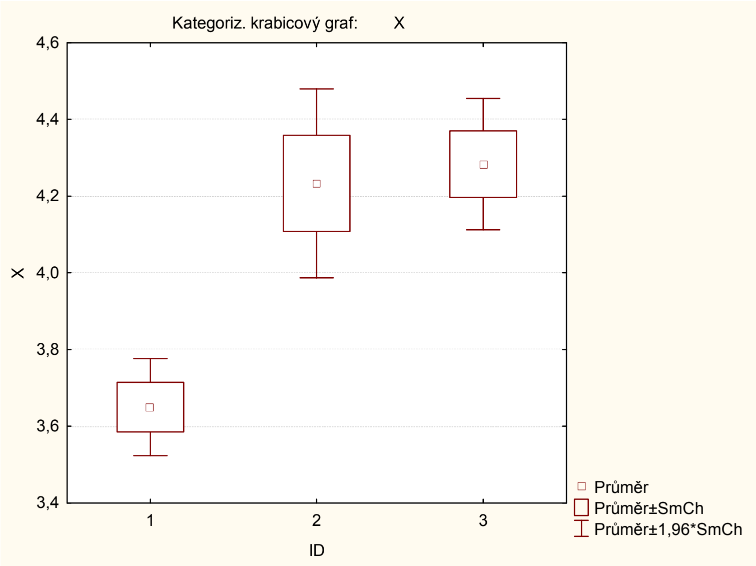


Vidíme, že rozložení dat se příliš neodlišuje od normálního rozložení.

Zjistíme rozsahy, průměry a směrodatné odchylky výkonů tří dělníků.
Statistika – Základní statsitiky a tabulky – Rozklad & jednofakt. ANOVA – OK – Proměnné Závislé X, Grupovací ID – OK – Výpočet: Tabulka statistik:

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

Nyní zobrazíme krabicové diagramy výkonů tří dělníků:
Návrat do Statistiky dle skupin – Kategoriz. krabicový graf.



Z krabicových diagramů vyplývá, že výkony 2. a 3. dělníka se liší jen málo, zatímco 1. dělník pracuje mnohem rychleji.

Před provedením analýzy rozptylu je nutné na hladině významnosti 0,05 testovat hypotézu o homogenitě rozptylů.
Pomocí Levenova testu testujeme H_0: $\sigma_1^2 = \sigma_2^2 = \sigma_3^2$ proti H_1: aspoň jedna dvojice rozptylů se liší.
Návrat do Statistiky dle skupin – na záložce ANOVA& testy vybereme Leveneovy testy

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

Testová statistika Levenova testu se realizuje hodnotou 1,5142, počet stupňů volnosti čitatele je 2, jmenovatele 13, odpovídající p-hodnota je 0,2564, tedy na hladině významnosti 0,05 nezamítáme hypotézu o shodě rozptylů.

Nyní pomocí jednofaktorové analýzy rozptylu testujeme H_0: $\mu_1 = \mu_2 = \mu_3$ proti H_1: aspoň jedna dvojice středních hodnot se liší.
Návrat do Statistiky dle skupin – na záložce ANOVA& testy vybereme Analýza arozptylu.

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

Ve výstupní tabulce je použito tohoto značení:
SČ efekt … skupinový součet čtverců S_A,
SV efekt … skupinový počet stupňů volnosti $f_A = r - 1$,
PČ efekt … $S_A/(r-1)$,
SČ chyba … reziduální součet čtverců S_E,
SV chyba … reziduální počet stupňů volnosti $f_E = n – r$,
PČ chyba … $S_E/(n-r)$,
F … testová statistika $F_A = \dfrac{S_A / f_A}{S_E / f_E}$

Protože p-hodnota = 0,00268, zamítá se na hladině významnosti 0,05 hypotézu o shodě středních hodnot.

Nyní nás bude zajímat, které dvojice dělníků se liší na zvolené hladině významnosti. Použijeme Scheffého metodu mnohonásobného porovnávání.

Návrat do Statistiky dle skupin – Post – hoc – Schefféův test:

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	

Výsledek Scheffého metody ukazuje, že na hladině významnosti 0,05 se liší výkony dělníků (1,2), (1,3) a neliší se (2,3).

Úkoly k samostatnému řešení:

Úkol 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. Na hladině významnosti 0,05 testujte hypotézu, že průměrné 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.

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		

Výsledek: Na hladině významnosti 0,05 se nulová hypotéza nezamítá.

Úkol 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 vypočtěte 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.

Výsledek: Průměrný čas cestování tramvají je 37 min, autobusem 30 min a metrem s přestupem na tramvaj je 36 min. Předpoklady normality a homogenity rozptylů všech tří výběrů jsou oprávněné. Na hladině významnosti 0,05 se zamítá hypotéza o shodě středních hodnot časů cestování do práce (testová statistika F_A se realizuje hodnotou 6,7151, odpovídající p-hodnota je 0,0083), na dané hladině významnosti se liší způsoby A a B, B a C.

Úkol 4.: Je dána neúplná tabulka ANOVA. Místo otazníků doplňte chybějící čísla a na hladině významnosti 0,05 testujte hypotézu o shodě středních hodnot.

zdroj variability	součet čtverců	stupně volnosti	podíl	F_A
skupiny	?	2	?	?
reziduální	16,033	?	?	-
celkový	17,301	30	-	-

Výsledek:

zdroj variability	součet čtverců	stupně volnosti	podíl	F_A
skupiny	**1,268**	2	**0,634**	**1,1072**
reziduální	16,033	**28**	**0,5726**	-
celkový	17,301	30	-	-

Kritický obor: $W = \langle F_{0,95}(2,28), \infty) = \langle 3{,}3404, \infty)$. Protože se F_A nerealizuje v kritickém oboru, H_0 nezamítáme na hladině významnosti 0,05.

Úkol 5.: Jsou dány čtyři nezávislé náhodné výběry postupně z rozložení $N(\mu_1, \sigma^2)$, $N(\mu_2, \sigma^2)$, $N(\mu_3, \sigma^2)$, $N(\mu_4, \sigma^2)$, přičemž každý z nich má rozsah 6. Na hladině významnosti 0,05 testujte hypotézu o shodě středních hodnot, je-li známo, že celkový součet čtverců je 114 a skupinový součet čtverců je 46,5.

Výsledek:

$S_E = S_T - S_A = 114 - 46{,}5 = 67{,}5$, testová statistika $F_A = \dfrac{S_A/(r-1)}{S_E/(n-r)} = \dfrac{46{,}5/3}{67{,}5/20} = 4{,}5926$,

příslušný kvantil $F_{0,95}(3,20) = 3{,}0984$, rozhodnutí o nulové hypotéze: zamítáme na hladině významnosti 0,05.