

Cvičení 1 – Analýza rozptylu jednoduchého třídění

Vzorový příklad

V rámci psychologického výzkumu bylo náhodně vybráno devět dvanáctiletých dětí a to tak, že tři děti měly matku se základním vzděláním, tři se středoškolským a tři s vysokoškolským. Všechny děti byly podrobeny témuž testu. Počty bodů, které děti v testu získaly, jsou uvedeny v tabulce:

Vzdělání matky	Počet bodů		
Základní (ZŠ)	20	23	22
Středoškolské (SŠ)	24	26	25
Vysokoškolské (VŠ)	26	27	27

Na hladině významnosti 0,05 testujte hypotézu, že střední hodnota počtu bodů v testu nezávisí na vzdělání matky. Zamítnete-li nulovou hypotézu, zjistěte, které dvojice úrovní vzdělání se liší na hladině významnosti 0,05. Vypočtěte též poměr determinace. Úkol řešte nejprve pomocí modulu Základní statistiky/tabulky, poté pomocí modulu ANOVA.

Řešení:

Otevřeme datový soubor body_v_testu.sta o dvou proměnných vzdelani a Y a 9 případech.

V proměnné Y jsou zapsané body získané v testu, v proměnné vzdelani kódy pro nejvyšší dosažené vzdělání matky (1 pro ZŠ, 2 pro SŠ, 3 pro VŠ).

Vzhledem k velmi malému počtu pozorování v každé skupině nemá smysl ověřovat normalitu a vzhledem k vyváženému třídění se nemusí ověřovat ani homogenita rozptylů.

Využití modulu Základní statistiky/tabulky

Vypočteme výběrové průměry a výběrové směrodatné odchylky:

Statistiky – Základní statistiky a tabulky – Rozklad & jednofakt. ANOVA – OK – Proměnné – Závislé – Y, Grupovací – vzdelani – OK Popisné statistiky – Výpočet: Tabulka statistik.

Rozkladová tabulka popisných statistik (body_v_testu.sta) N=9 (V seznamu záv. prom. nejsou ChD)			
vzdelani	Y průměr	Y N	Y Sm.odch.
ZŠ	21,66667	3	1,527525
SŠ	25,00000	3	1,000000
VŠ	26,66667	3	0,577350
Vš.skup.	24,44444	9	2,403701

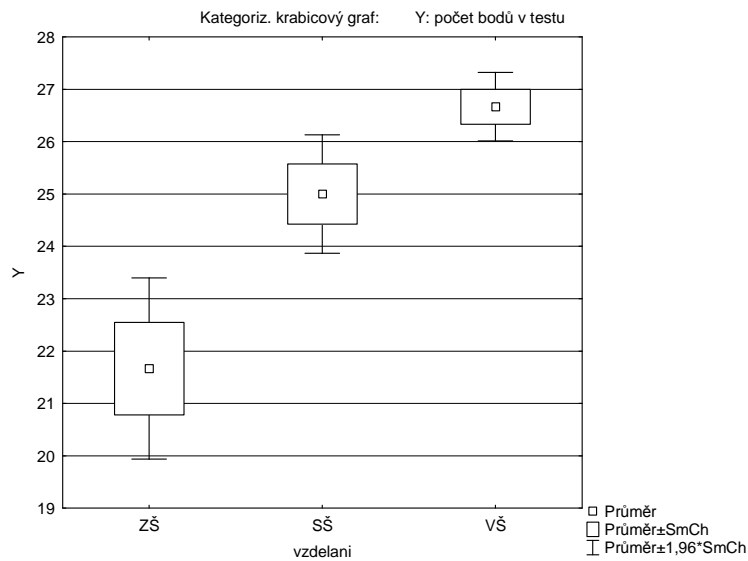
Přistoupíme k testu hypotézy o shodě středních hodnot.

Na záložce ANOVA & testy zvolíme Analýza rozptylu.

Proměnná	Analýza rozptylu (body_v_testu.sta) Označ. efekty jsou význ. na hlad. $p < ,05000$							
	SČ efekt	SV efekt	PČ efekt	SČ chyba	SV chyba	PČ chyba	F	p
Y	38,88889	2	19,44444	7,333333	6	1,222222	15,90909	0,003993

Jelikož p-hodnota = 0,003993 je menší než hladina významnosti 0,05, hypotézu o shodě středních hodnot zamítáme na hladině významnosti 0,05.

Vytvoříme krabicové diagramy: záložka Základní výsledky – Kategoriz. krabicový graf



Nyní aplikujeme Tukeyovu metodu mnohonásobného porovnávání, abychom zjistili, které dvojice úrovní vzdělání matky se liší na hladině významnosti 0,05. Na záložce Post – hoc zvolíme Tukeyův HSD.

Tukeyův HSD test; proměn.:Y (body_v_testu.sta)			
Označ. rozdíly jsou významné na hlad. p < ,05000			
vzdělání	{1}	{2}	{3}
	M=21,667	M=25,000	M=26,667
ZŠ {1}		0,023833	0,003676
SŠ {2}	0,023833		0,234043
VŠ {3}	0,003676	0,234043	

Tabulka obsahuje p-hodnoty pro vzájemné porovnání středních hodnot počtu bodů v testu všech tří úrovní vzdělání matky. Vidíme, že na hladině významnosti 0,05 se liší dvojice (ZŠ, SŠ) a (ZŠ, VŠ).

Využití modulu ANOVA

Statistiky – ANOVA – Jednofaktorová ANOVA – OK – Proměnné Seznam závislých proměnných: Y, Kategor. nezávislá proměnná (faktor): vzdělání – OK.

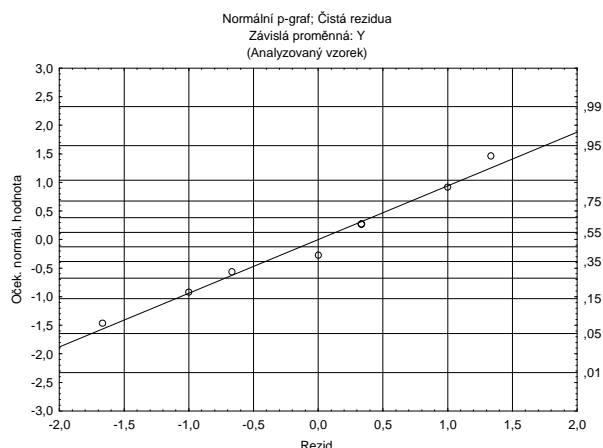
Další postup se liší podle způsobu kódování indikátorových proměnných.

a) Kódování pomocí indikátorů typu dummy

Na záložce Možnosti vypneme Sigma – omezená parametrizace – OK.

Zvolíme Více výsledků – záložka Předpoklady. Zde můžeme provést test homogenity rozptylů (v našem případě to není zapotřebí, ve všech skupinách je stejný počet pozorování) a orientačně ověřit normalitu dat v jednotlivých skupinách. Vzhledem k velmi malému rozsahu skupin to však pro nás není vhodné. V takovém případě se ověřuje normalita reziduí.

Zvolíme záložku Rezidua 1 – Pravděpodobnostní grafy reziduí – Normální.



Vypočteme výběrové průměry (s 95% intervaly spolehlivosti) a výběrové směrodatné odchylky:

Na záložce Details zvolíme Popisné st. buněk:

Efekt	Popisné statistiky (body_v_testu.sta)						
	Úroveň Faktor	N	Y Průměr	Y Sm.odch.	Y Sm.Ch.	Y -95,00%	Y +95,00%
Celkem		9	24,44444	2,403701	0,801234	22,59680	26,29209
vzdelani	ZŠ	3	21,66667	1,527525	0,881917	17,87208	25,46125
vzdelani	SŠ	3	25,00000	1,000000	0,577350	22,51586	27,48414
vzdelani	VŠ	3	26,66667	0,577350	0,333333	25,23245	28,10088

Na záložce Details zvolíme Test všech efektů a dostaneme tabulku ANOVA:

Jednorozměrné testy významnosti pro Y (body_v_testu.sta) Přeparametrizovaný model Dekompozice typu III					
Efekt	SČ	Stupně volnosti	PČ	F	p
Abs. člen	5377,778	1	5377,778	4400,000	0,000000
vzdelani	38,889	2	19,444	15,909	0,003993
Chyba	7,333	6	1,222		

Vidíme, že faktor vzdelani je významný na hladině významnosti 0,05.

Provedeme mnohonásobné porovnání. Na záložce Post-hoc zvolíme Tukeyův HSD:

Tukeyův HSD test; proměnná Y (body_v_testu.sta) Přibližné pravděpodobnosti pro post hoc testy Chyba: meziskup. PČ = 1,2222, sv = 6,0000				
Č. buňky	vzdelani	{1}	{2}	{3}
		21,667	25,000	26,667
1	ZŠ		0,023833	0,003676
2	SŠ	0,023833		0,234043
3	VŠ	0,003676	0,234043	

Na hladině významnosti 0,05 se liší dvojice skupin (ZŠ, SŠ) a (ZŠ, VŠ).

Nyní získáme odhady parametrů v modelu $Y_{ij} = \beta_0 + \beta_1 Z_{1j} + \beta_2 Z_{2j} + \varepsilon_{ij}$.

Na záložce Detaily vybereme Koeficienty:

Odhady parametrů (body_v_testu.sta) (*Vynulované prediktory neuspěly při kontrole tolerance) Přeparametrizovaný model									
Efekt	Úroveň Efekt	Sloupec	Poznámky (V/N/N)	Y Param.	Y Sm.Ch.	Y t	Y p	-95,00% LmtSpol.	+95,00% LmtSpol.
Abs. člen		1		26,66667	0,638285	41,77864	0,000000	25,10484	28,22849
vzdelani	ZŠ	2	Vychýl.	-5,00000	0,902671	-5,53912	0,001461	-7,20876	-2,79124
vzdelani	SŠ	3	Vychýl.	-1,66667	0,902671	-1,84637	0,114364	-3,87542	0,54209
vzdelani	VŠ	4	Nulov.*	0,00000					

Ve výstupní tabulce najdeme ve sloupci Y Param. odhady $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$, výsledky testů významnosti těchto parametrů a 95% intervaly spolehlivosti.

$\hat{\beta}_0 = 26,67$ představuje průměr hodnot v referenční skupině (VŠ),

$\hat{\beta}_1 = -5$ je odchylka průměru 1. skupiny (ZŠ) od průměru referenční skupiny,

$\hat{\beta}_2 = -1,67$ je odchylka průměru 2. skupiny (SŠ) od průměru referenční skupiny.

Protože p-hodnota příslušející testu významnosti parametru β_2 je 0,1144, což je větší než 0,05, znamená to, že střední hodnota 2. skupiny není významně odlišná od střední hodnoty referenční skupiny.

Poměr determinace získáme na záložce Detaily – Celkové R. Ve výstupní tabulce je pod označením Vícenás. R2:

Test SČ celého modelu vs. SČ reziduí (body_v_testu.sta)											
Závislý proměnné	Vícenás. R	Vícenás. R2	Upravené R2	SČ Model	SV Model	PČ Model	SČ Rezid.	SV Rezid.	PČ Rezid.	F	p
Y	0,917249	0,841346	0,788462	38,88889	2	19,44444	7,333333	6	1,222222	15,90909	0,003993

Hodnoty předpovědi a reziduí najdeme pomocí záložky Rezidua 1 – Předpovědi a rezidua:

Pozorované, předpovězené a reziduální hodnoty Přeparametrizovaný model (Analyzovaný vzorek)			
	Y Pozorov.	Y Předpov.	Y Rezid.
1	20,00000	21,66667	-1,66667
2	23,00000	21,66667	1,33333
3	22,00000	21,66667	0,33333
4	24,00000	25,00000	-1,00000
5	26,00000	25,00000	1,00000
6	25,00000	25,00000	0,00000
7	26,00000	26,66667	-0,66667
8	27,00000	26,66667	0,33333
9	27,00000	26,66667	0,33333

Ad b) Kódování pomocí indikátorů typu effect

Na záložce Možnosti zvolíme Sigma – omezená parametrizace.

Postup je stejný jako v případě (a), liší se však tabulka odhadů parametrů:

Efekt	Odhady parametrů (body_v_testu.sta) Sigma-omezená parametrizace							
	Úroveň Efekt	Sloupec	Y Param.	Y Sm.Ch.	Y t	Y p	-95,00% LmtSpol.	+95,00% LmtSpol.
Abs. člen		1	24,44444	0,368514	66,33250	0,000000	23,54272	25,34617
vzdelani	ZŠ	2	-2,77778	0,521157	-5,33002	0,001778	-4,05300	-1,50255
vzdelani	SŠ	3	0,55556	0,521157	1,06600	0,327427	-0,71967	1,83078

$\hat{\beta}_0 = 24,44$ představuje průměr všech hodnot veličiny Y,

$\hat{\beta}_1 = -2,78$ je odchylka průměru 1. skupiny (ZŠ) od celkového průměru,

$\hat{\beta}_2 = 0,56$ je odchylka průměru 2. skupiny (SŠ) od celkového průměru.

Odhad efektu 3. úrovně faktoru A získáme jako $\hat{\alpha}_3 = M_{3.} - M_{..} = 26,67 - 24,44 = 2,23$. Je to odchylka průměru 3. skupiny (VŠ) od celkového průměru.

Vidíme, že p-hodnota příslušející testu významnosti parametru β_2 je 0,3274, což je větší než hladina významnosti 0,05. Znamená to, že střední hodnota 2. skupiny (SŠ) není významně odlišná od celkové střední hodnoty všech skupin.

Příklad k samostatnému řešení:

Popis situace: Výzkumníci řešili problém, zda čas má vliv na pokles hladiny alkoholu v krvi. Za tímto účelem náhodně vybrali 18 dospělých mužů a náhodně je rozdělili do tří skupin označených A, B, C. Jejich rozsahy byly postupně 7, 5 a 6. U mužů skupiny A byla hladina alkoholu v krvi (udaná v promile) změřena po uplynutí 1 h, u skupiny B po 2 h a u skupiny C po 3 h. Výsledky jsou zaznamenány v tabulce:

Skupina A	1,1	1,0	0,9	0,9	1,0	1,2	1,1
Skupina B	0,8	0,9	0,7	0,7	0,6		
Skupina C	0,7	0,4	0,4	0,4	0,5	0,6	

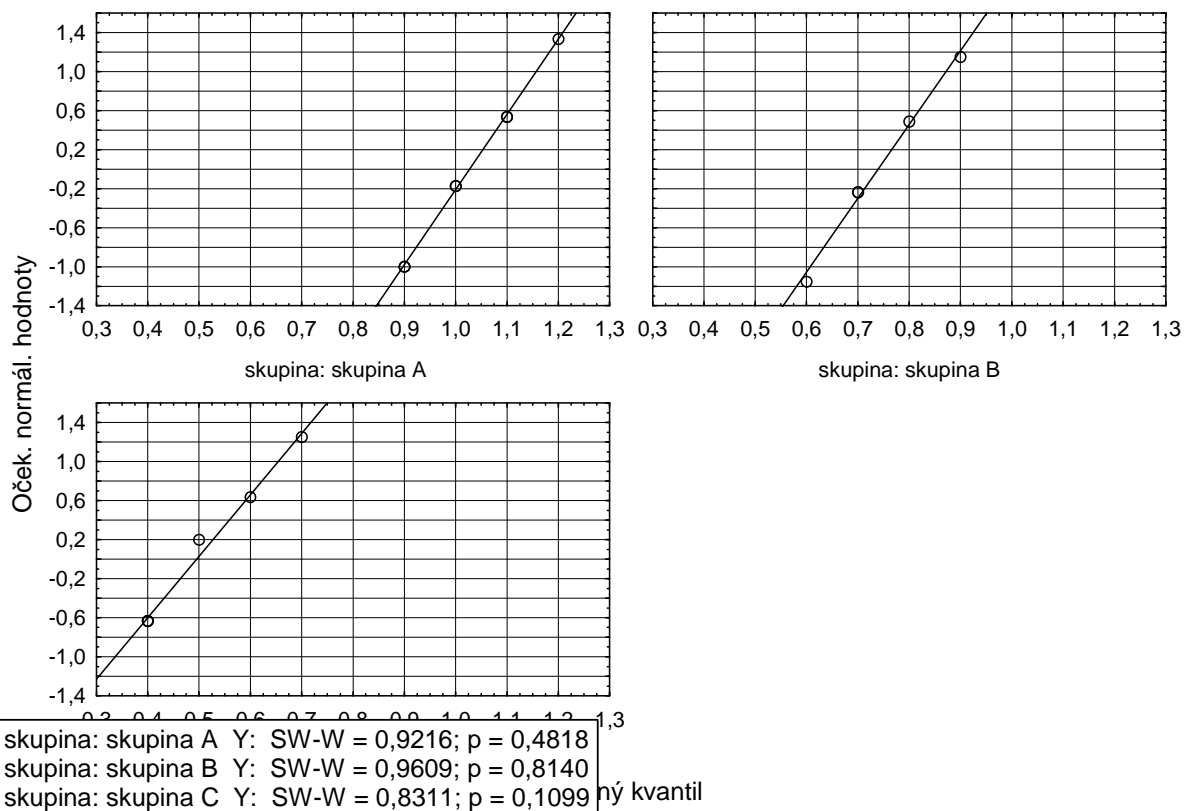
Na hladině významnosti 0,05 testujte hypotézu, že střední hodnota hladiny alkoholu nezávisí na čase. Zamítnete-li nulovou hypotézu, zjistěte, které dvojice časových kategorií se liší na hladině významnosti 0,05. Vypočtete též poměr determinace. Úkol řešte nejprve pomocí modulu Základní statistiky/tabulky, poté pomocí modulu ANOVA.

Řešení:

Otevřeme datový soubor alkohol_v_krvi.sta o dvou proměnných skupina a Y a 18 případech. V proměnné Y jsou zapsané hladiny alkoholu, v proměnné skupina kódy pro jednotlivé skupiny mužů.

Ověření normality ve všech střezech skupinách:

Normální p-graf z Y; kategorizovaný skupina
alkohol_v_krvi.sta 2v*18c

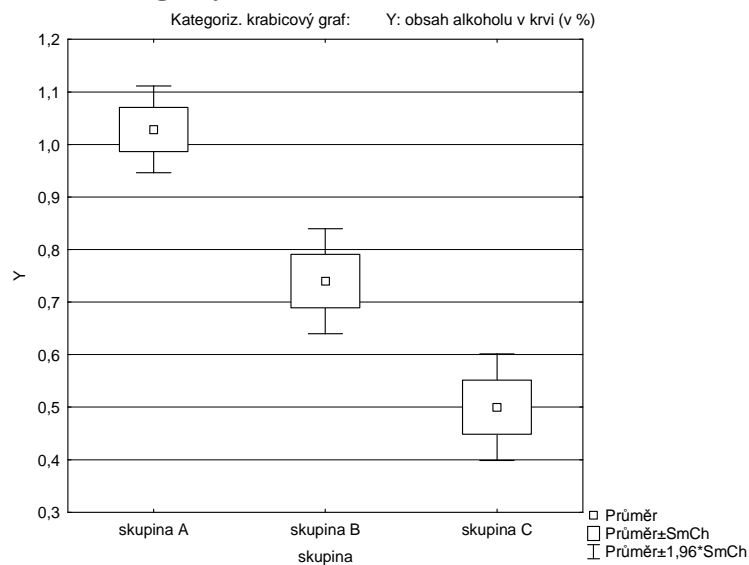


Využití modulu Základní statistiky/tabulky

Číselné charakteristiky:

Rozkladová tabulka popisných statistik (alkohol_v_krvi.sta) N=18 (V seznamu záv. prom. nejsou ChD)			
skupina	Y průměr	Y N	Y Sm.odch.
skupina A	1,028571	7	0,111270
skupina B	0,740000	5	0,114018
skupina C	0,500000	6	0,126491
Vš.skup.	0,772222	18	0,256230

Krabicové grafy:



Test hypotézy o homogenitě rozptylů:

Leveneův test homogenity rozptylů (alkohol_v_krvi.sta) Označ. efekty jsou význ. na hlad. $p < ,05000$								
Proměnná	SČ efekt	SV efekt	PČ efekt	SČ chyba	SV chyba	PČ chyba	F	p
Y	0,000489	2	0,000245	0,051123	15	0,003408	0,071772	0,931060

Test hypotézy o shodě středních hodnot:

Analýza rozptylu (alkohol_v_krvi.sta) Označ. efekty jsou význ. na hlad. $p < ,05000$								
Proměnná	SČ efekt	SV efekt	PČ efekt	SČ chyba	SV chyba	PČ chyba	F	p
Y	0,909825	2	0,454913	0,206286	15	0,013752	33,07883	0,000003

Scheffého metoda mnohonásobného porovnávání:

Scheffeho test; proměn.:Y (alkohol_v_krvi.sta) Označ. rozdíly jsou významné na hlad. $p < ,05000$			
skupina	{1}	{2}	{3}
	M=1,0286	M=,74000	M=,50000
skupina A {1}		0,002921	0,000003
skupina B {2}	0,002921		0,014317
skupina C {3}	0,000003	0,014317	

Využití modulu ANOVA

a) Kódování pomocí indikátorů typu dummy

Popisné statistiky buněk:

Efekt	Popisné statistiky (alkohol_v_krvi.sta)						
	Úroveň Faktor	N	Y Průměr	Y Sm.odch.	Y Sm.Ch.	Y -95,00%	Y +95,00%
Celkem		18	0,772222	0,256230	0,060394	0,644802	0,899642
skupina	skupina A	7	1,028571	0,111270	0,042056	0,925664	1,131479
skupina	skupina B	5	0,740000	0,114018	0,050990	0,598429	0,881571
skupina	skupina C	6	0,500000	0,126491	0,051640	0,367256	0,632744

Tabulka ANOVA:

Jednorozm. výsledky pro každou záv. proměnnou (alkohol_v_krvi.sta) Přeparametrizovaný model Dekompozice typu III					
Efekt	Stupně volnosti	Y SČ	Y PČ	Y F	Y p
Abs. člen	1	10,10044	10,10044	734,4505	0,000000
skupina	2	0,90983	0,45491	33,0788	0,000003
Chyba	15	0,20629	0,01375		
Celkem	17	1,11611			

Výsledek Scheffého metody:

Scheffeho test; proměnná Y (alkohol_v_krvi.sta) Pravděpodobnosti pro post-hoc testy Chyba: meziskup. PČ = ,01375, sv = 15,000				
Č. buňky	skupina	{1}	{2}	{3}
		1,0286	,74000	,50000
1	skupina A		0,002921	0,000003
2	skupina B	0,002921		0,014317
3	skupina C	0,000003	0,014317	

Odhady parametrů v modelu $Y_{ij} = \beta_0 + \beta_1 Z_{1j} + \beta_2 Z_{2j} + \varepsilon_{ij}$:

Odhady parametrů (alkohol_v_krvi.sta) (*Vynulované prediktory neuspěly při kontrole tolerance) Přeparametrizovaný model													
Efekt	Úroveň Efekt	Sloupec	Poznámky (V/N/N)	Y Param.	Y Sm.Ch.	Y t	Y p	-95,00% LmtSpol.	+95,00% LmtSpol.	Y Beta (β)	Y Sm.Ch. β	-95,00% LmtSpol.	+95,00% LmtSpol.
Abs. člen		1		0,500000	0,047875	10,44376	0,000000	0,397956	0,602044				
skupina	skupina A	2	Vychýl.	0,528571	0,065243	8,10154	0,000001	0,389509	0,667634	1,034806	0,127730	0,762557	1,307055
skupina	skupina B	3	Vychýl.	0,240000	0,071011	3,37977	0,004126	0,088644	0,391356	0,431696	0,127730	0,159447	0,703945
skupina	skupina C	4	Nulov.	0,000000									

$\hat{\beta}_0 = 0,5$ představuje průměr hodnot v referenční skupině C,

$\hat{\beta}_1 = 0,5286$ je odchylka průměru skupiny A od průměru referenční skupiny,

$\hat{\beta}_2 = 0,24$ je odchylka průměru skupiny B od průměru referenční skupiny.

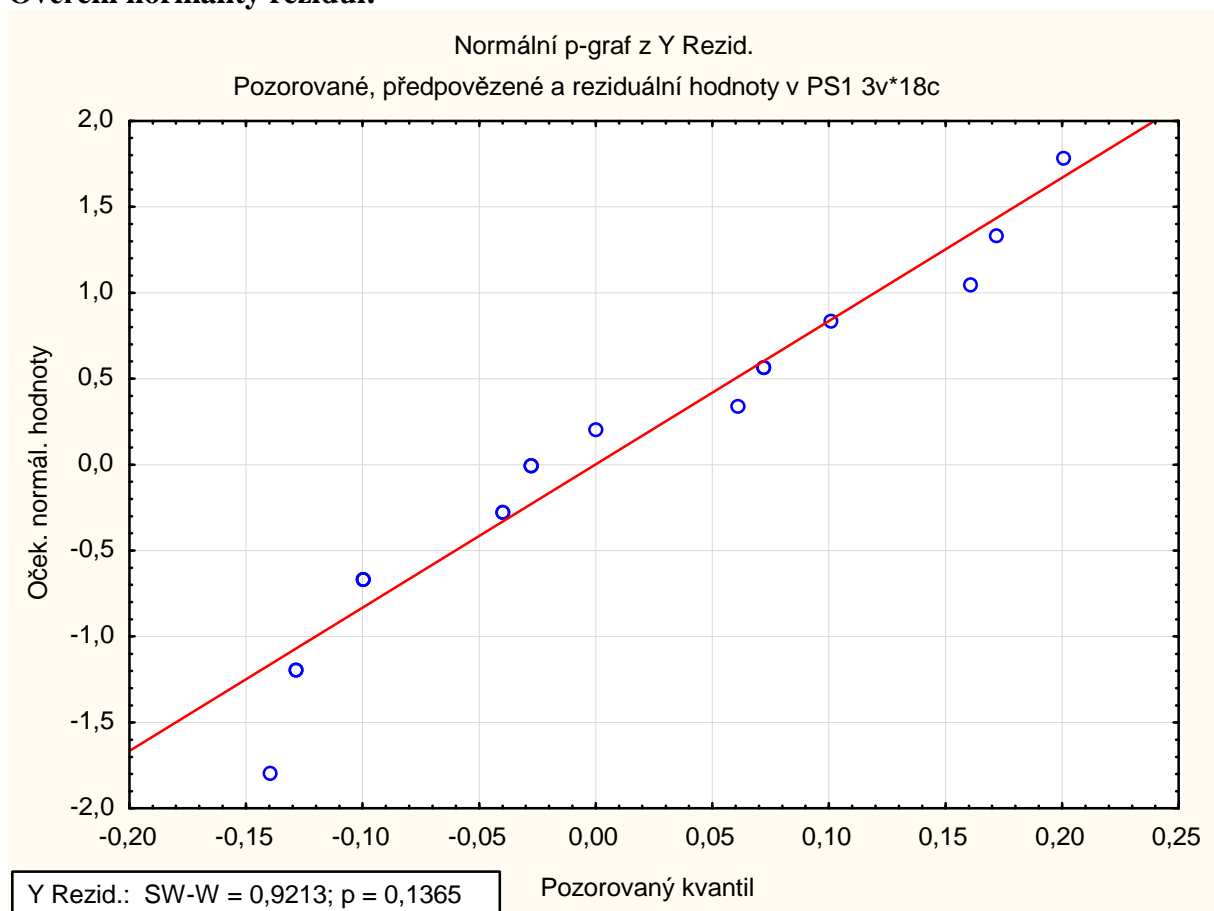
Poměr determinace:

Test SČ celého modelu vs. SČ reziduí (alkohol_v_krvi.sta)											
Závislý proměnné	Vícenás. R	Vícenás. R2	Upravené R2	SČ Model	SV Model	PČ Model	SČ Rezid.	SV Rezid.	PČ Rezid.	F	p
Y	0,902870	0,815175	0,790531	0,909825	2	0,454913	0,206286	15	0,013752	33,07883	0,000003

Hodnoty předpovědí a reziduí:

	Pozorované, předpovězené a reziduální hodnoty Přeparametrizovaný model (Analyzovaný vzorek)		
	Y	Y	Y
	Pozorov.	Předpov.	Rezid.
1	1,10000	1,02857	0,07143
2	1,00000	1,02857	-0,02857
3	0,90000	1,02857	-0,12857
4	0,90000	1,02857	-0,12857
5	1,00000	1,02857	-0,02857
6	1,20000	1,02857	0,17143
7	1,10000	1,02857	0,07143
8	0,80000	0,74000	0,06000
9	0,90000	0,74000	0,16000
10	0,70000	0,74000	-0,04000
11	0,70000	0,74000	-0,04000
12	0,60000	0,74000	-0,14000
13	0,70000	0,50000	0,20000
14	0,40000	0,50000	-0,10000
15	0,40000	0,50000	-0,10000
16	0,40000	0,50000	-0,10000
17	0,50000	0,50000	-0,00000
18	0,60000	0,50000	0,10000

Ověření normality reziduí:



b) Kódování pomocí indikátorů typu effect:

Odhady parametrů

Odhady parametrů (alkohol_v_krvi.sta) Sigma-omezená parametrizace												
Efekt	Úroveň Efekt	Sloupec	Y				-95,00%		+95,00%		Y	
			Param.	Sm.Ch.	t	p	LmtSpol.	LmtSpol.	Beta (β)	Sm.Ch. β	LmtSpol.	LmtSpol.
Abs. člen		1	0,756190	0,027903	27,10075	0,000000	0,696717	0,815664				
skupina	skupina A	2	0,272381	0,037861	7,19425	0,000003	0,191682	0,353080	0,927609	0,128937	0,652785	1,202432
skupina	skupina B	3	-0,016190	0,041175	-0,39321	0,699697	-0,103953	0,071572	-0,050699	0,128937	-0,325523	0,224124

$\hat{\beta}_0 = 0,7562$ představuje průměr všech hodnot veličiny Y,

$\hat{\beta}_1 = 0,2724$ je odchylka průměru skupiny A od celkového průměru,

$\hat{\beta}_2 = -0,0162$ je odchylka průměru skupiny B od celkového průměru.

Odhad efektu 3. úrovně faktoru A získáme jako $\hat{\alpha}_3 = M_{3.} - M_{..} = 0,5 - 0,7562 = -0,2562$. Je

to odchylka průměru skupiny C od celkového průměru.