Studijní materiál ke kurzu

IBM IBM® SPSS® STATISTICS BASE: ZÁKLADY MODULU II.

Ondřej Brom

2018

Strategický projekt UTB ve Zlíně, reg. č. CZ.02.2.69/0.0/0.0/16_015/0002204

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Informace o autorovi:

Ondřej Brom, Ing.

ACREA CR, spol. s r.o.

obrom@acrea.cz

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

OBSAH

OBSAH **3**
1 CROSSTABS – KONTINGENČNÍ TABULKY **5**
1.1 KONTINGENČNÍ TABULKY 5
1.2 VOLÁNÍ PROCEDURY V IBM SPSS STATISTICS 5
1.2.1 Nastavení dialogu 5
1.2.2 Tlačítko Exact 6
1.2.3 Tlačítko Statistics 6
1.2.4 Tlačítko Cells 7
1.2.5 Tlačítko Format 9
1.3 VÝSTUPY 9
1.3.1 Přehled o počtu platných a chybějících případů 9
1.3.2 Kontingenční tabulka 10
1.3.3 Řádková procenta 11
1.3.4 Test Chí-kvadrát 11
1.3.5 Adjustovaná standardizovaná rezidua 13
2 MEANS – PRŮMĚRY **15**
2.1 TABULKY PRŮMĚRŮ 15
2.2 VOLÁNÍ PROCEDURY V IBM SPSS STATISTICS 15
2.2.1 Nastavení dialogu 15
2.2.2 Tlačítko Options 16
2.3 VÝSTUPY 17
2.3.1 Přehled o počtu platných a chybějících případů 17
2.3.2 Přehled statistik 18
2.3.3 Analýza rozptylu 18
2.3.4 Testování linearity 19
3 OLAP CUBES – DATOVÉ KOSTKY OLAP **21**
3.1 DATOVÉ KOSTKY OLAP 21
3.2 VOLÁNÍ PROCEDURY V IBM SPSS STATISTICS 22
3.2.1 Nastavení dialogu 22
3.2.2 Tlačítko Statistics 23
3.2.3 Tlačítko Differences 24
3.2.4 Tlačítko Title 26
3.3 VÝSTUPY 26
3.3.1 OLAP kostka 27
SEZNAM POUŽITÉ LITERATURY **29**
SEZNAM OBRÁZKŮ **30**
SEZNAM TABULEK **31**

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

ÚVOD

Tento materiál slouží jako studijní materiál ke Strategickému projektu UTB ve Zlíně, reg. č. CZ.02.2.69/0.0/0.0/16_015/0002204.

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

1 CROSSTABS – KONTINGENČNÍ TABULKY

1.1 Kontingenční tabulky

Kontingenční tabulky užíváme ke zjišťování vztahu dvou kategorizovaných proměnných. Výstupní tabulka obsahuje jedno pole pro každou kombinaci hodnot těchto proměnných. Tabelovat můžeme:

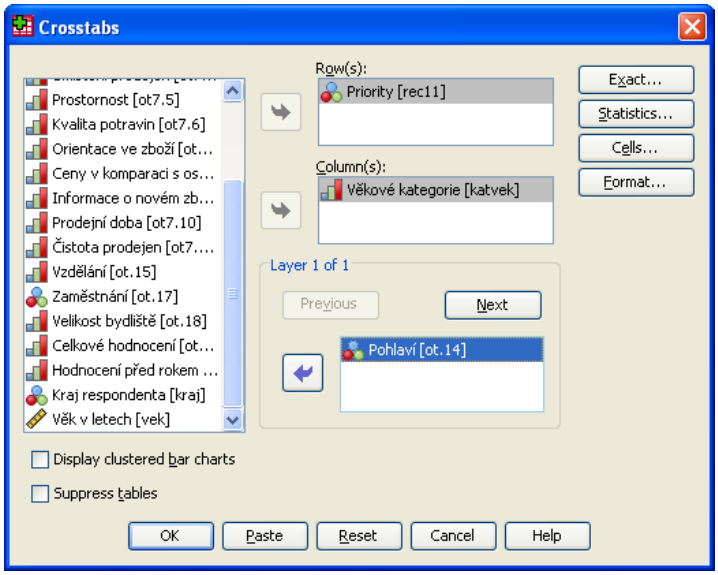
- Četnosti (pozorované i očekávané)
- Procenta (řádková, sloupcová nebo vzhledem k celé tabulce)
- Rezidua (nestandardizovaná, standardizovaná, adjustovaná standardizovaná)

Tabulku lze rovněž doplnit sloupcovým grafem četností. Pro zjišťování závislosti řádkové a sloupcové proměnné jsou k dispozici různé typy statistických testů, z nichž nejčastěji je užíván Pearsonův test chí-kvadrát.

1.2 Volání procedury v IBM SPSS Statistics

Analyze → Descriptive Statistics → Crosstabs

1.2.1 Nastavení dialogu



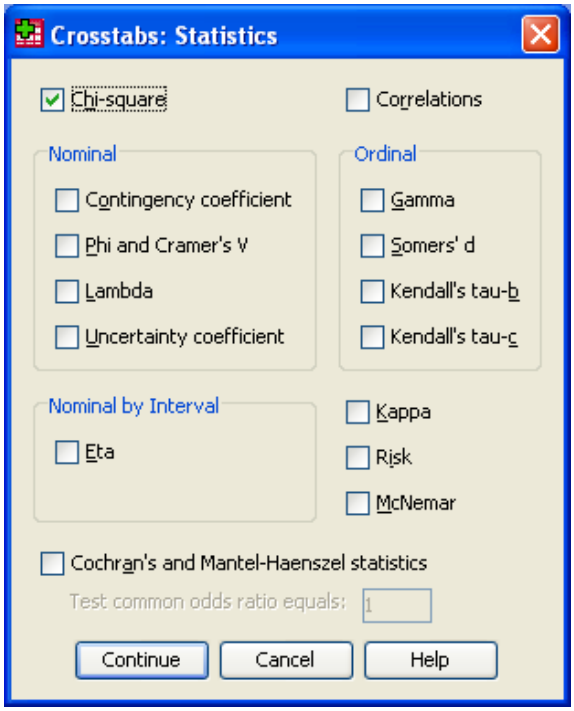
Obrázek 1 Nastavení dialogu Crosstabs

- Do políček *Row(s)* resp. *Column(s)* přeneseme řádkovou resp. sloupcovou proměnnou. Jestliže do některého z polí vložíme více proměnných, vytvoří se pro každou z nich samostatná tabulka.
- Pokud potřebujeme získat tabulku vyššího stupně třídění, využijeme políčko *Layer* a pomocí tlačítka *Next* přecházíme mezi dalšími vrstvami třídění.
- Zaškrtneme-li tlačítko *Display clustered bar charts*, zobrazí se ve výstupu také skupinkový sloupcový graf.
- Při označení *Suppress tables* se ve výstupu nevytvoří kontingenční tabulka, ale jenom zvolené statistiky nebo grafy.

1.2.2 Tlačítko Exact

Tlačítko je dostupné jen při nainstalování modulu *Exact Tests*. Umožňuje provádět spolehlivé statistické testy nezávislosti i pro tabulky, které vykazují malé očekávané četnosti v některých buňkách.

1.2.3 Tlačítko Statistics

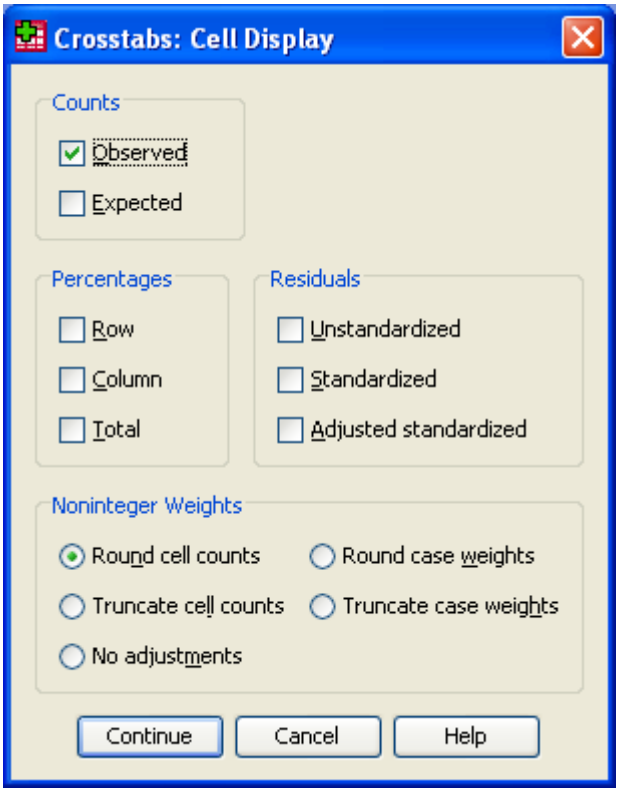


Obrázek 2 Nastavení dialogu Crosstabs, tlačítko Statistics

Pomocí tlačítka *Statistics* zaškrtneme požadované statistiky a testy. Procedura obsahuje celkem 15 různých měr závislosti dvou proměnných. Statistiky jsou rozděleny do několika skupin podle typu proměnných.

Nejčastěji užívaným testem je Pearsonův chí-kvadrát (*Chi-square*). Pro ordinální proměnné je dále k dispozici například Spearmanův koeficient pořadové korelace rho (*Correlations*).

1.2.4 Tlačítko Cells



Obrázek 3 Nastavení dialogu Crosstabs, tlačítko Cells

Tlačítkem *Cells* nastavíme, co se má zobrazit v buňkách kontingenční tabulky.

Counts (četnosti)

Observed (pozorované četnosti) – kolik záznamů obsahuje danou kombinaci hodnot.

Expected (očekávané četnosti) – kolik záznamů s danou kombinací hodnot očekáváme, za předpokladu, že zkoumané proměnné jsou statisticky nezávislé.

Percentages (procenta)

Row (řádková procenta),

Column (sloupcová procenta),

Total (procenta vzhledem k celé tabulce).

Residuals (rezidua)

Unstandardized (nestandardizovaná) – rozdíl četnosti a očekávané četnosti.

Standardized (standardizovaná) – nestandardizovaná rezidua vydělená odhadem své směrodatné odchylky.

Adjusted standardized (adjustovaná standardizovaná) – nestandardizovaná rezidua vydělená odhadem své standardní chyby, porovnávají se vzhledem k hodnotám standardizovaného normálního rozdělení.

Noninteger Weights (neceločíselné váhy)

Hodnoty, udávající četnosti v kontingenční tabulce, jsou obvykle celočíselné – jedná se o počty. Jsou-li však případy váženy neceločíselnými vahami, mohou vycházet také tyto četnosti neceločíselné. Z toho důvodu je k dispozici několik možností, jak se s touto situací vyrovnat:

Round cell counts – váhy jednotlivých případů jsou užity k výpočtu četností, avšak před výpočtem statistik jsou výsledné hodnoty zaokrouhleny.

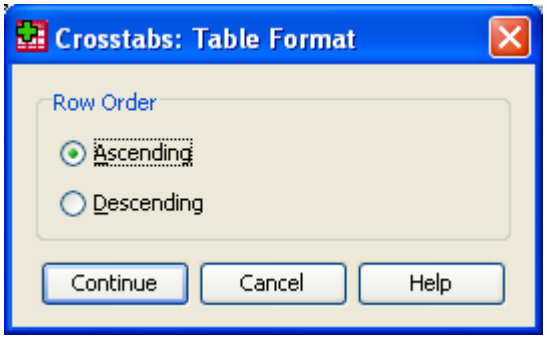
Truncate cell counts – váhy jednotlivých případů jsou užity k výpočtu četností, avšak před výpočtem statistik jsou výsledné hodnoty oříznuty.

Round case weights – váhy jednotlivých případů jsou před výpočtem zaokrouhleny.

Truncate case weights – váhy jednotlivých případů jsou před výpočtem oříznuty.

No adjustments – pracujeme s neceločíselnými vahami i četnostmi. Jestliže však využíváme tlačítko *Exact (*je k dispozici pouze máme-li nainstalovaný modul *Exact Tests*), hodnoty četností v kontingenční tabulce musí být zaokrouhleny nebo oříznuty dříve, než dojde k výpočtu exaktních statistik.

1.2.5 Tlačítko Format



Kategorie řádkové proměnné můžeme setřídit vzestupně nebo sestupně podle kódů.

1.3 Výstupy

1.3.1 Přehled o počtu platných a chybějících případů

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
PS 1998 * Vzdělání respondenta	1434.298[a]	35.7%	2583.011	64.3%	4017.309	100.0%

a. Number of valid cases is different from the total count in the crosstabulation table becaus the cell counts have been rounded.

Tabulka 1 Přehled o počtu platných a chybějících případů

Informace o počtu platných a chybějících hodnot. V tomto případě je pod tabulkou poznámka, která upozorňuje na to, že případy jsou váženy neceločíselnými vahami a v kontingenční tabulce budou hodnoty zaokrouhleny.

1.3.2 Kontingenční tabulka

PS 1998 * Vzdělání respondenta Crosstabulation

Count

		Vzdělání respondenta				
		ZŠ	SŠ bez maturity	SŠ s maturitou	VŠ	Total
PS 1998	ČSSD	67	195	154	53	469
	KDU-ČSL	27	52	33	18	130
	DŽJ	13	12	3	0	28
	KSČM	34	57	26	7	124
	ODS	47	162	168	62	439
	SPR-RSČ	7	6	9	1	23
	US	15	32	62	30	139
	jiná strana	16	26	33	7	82
Total		226	542	488	178	1434

Tabulka 2 Kontingenční tabulka

Kontingenční tabulka ukazuje, jak jsou rozloženy preference pro jednotlivé strany mezi kategorie vzdělání. Řádky tabulky znázorňují strany volené v parlamentních volbách v roce 1998, sloupce kategorie vzdělání. Buňky vyjadřují počty respondentů, z dané věkové kategorie, kteří volili uvedenou stranu. Poslední řádek resp. sloupec, je součtem předchozích řádků resp. sloupců.

1.3.3 Řádková procenta

PS 1998 * Vzdělání respondenta Crosstabulation

% within PS 1998

		Vzdělání respondenta				
		ZŠ	SŠ bez maturity	SŠ s maturitou	VŠ	Total
PS 1998	ČSSD	14.3%	41.6%	32.8%	11.3%	100.0%
	KDU-ČSL	20.8%	40.0%	25.4%	13.8%	100.0%
	DŽJ	46.4%	42.9%	10.7%	.0%	100.0%
	KSČM	27.4%	46.0%	21.0%	5.6%	100.0%
	ODS	10.7%	36.9%	38.3%	14.1%	100.0%
	SPR-RSČ	30.4%	26.1%	39.1%	4.3%	100.0%
	US	10.8%	23.0%	44.6%	21.6%	100.0%
	jiná strana	19.5%	31.7%	40.2%	8.5%	100.0%
Total		15.8%	37.8%	34.0%	12.4%	100.0%

Tabulka 3 Kontingenční tabulka – řádková procenta

Tabulka udává řádková procenta. V tomto případě můžeme porovnat, jak se liší procentuální zastoupení věkových kategorií pro jednotlivé politické strany. Z tabulky například poznáme, že mezi voliči US a ODS je více vysokoškolsky vzdělaných osob, než je tomu u jiných stran.

Poslední sloupec je součtem předchozích, je však spíše kontrolní – hodnota je vždy rovna 100 %. Poslední řádek udává celkové procentuální zastoupení kategorií vzdělání.

1.3.4 Test Chí-kvadrát

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	100.486[a]	21	.000
Likelihood Ratio	100.072	21	.000
Linear-by-Linear Association	9.649	1	.002
N of Valid Cases	1434		

a. 4 cells (12.5%) have expected count less than 5. The minimum expected count is 2.85.

Tabulka 4 Kontingenční tabulka – test Chí-kvadrát

Pearsonův test Chí-kvadrát (*Pearson Chi-Square)* na prvním řádku umožňuje testovat nezávislost řádkové a sloupcové proměnné. Rozhodující jsou hodnoty asymptotické signifikance. Porovnáváme je s předem danou hodnotou hladiny spolehlivosti *a*, zpravidla 0,05. V našem případě je signifikance menší než 0,05, a na této hladině významnosti tedy lze prohlásit, že zkoumané proměnné jsou na sobě závislé. Druhý řádek obsahuje alternativní test téhož. Třetí řádek zjišťuje existenci lineárního vztahu a má tedy smysl pouze v případě, že jsou obě proměnné ordinální.

Poznámka pod čarou nás informuje o málo početných kombinacích – prázdné či málo zaplněné buňky bývají zdrojem nespolehlivosti asymptotických statistik. Pro test Chí-kvadrát by neměla mít žádná buňka očekávanou četnost menší než 1 a více než 20 % buněk by nemělo mít očekávanou četnost menší než 5. V našem případě lze tedy považovat výsledek testu za adekvátní.

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

1.3.5 Adjustovaná standardizovaná rezidua

PS 1998 * Vzdělání respondenta Crosstabulation

Adjusted Residual

		Vzdělání respondenta			
		ZŠ	SŠ bez maturity	SŠ s maturitou	VŠ
PS 1998	ČSSD	-1.1	2.1	-.7	-.9
	KDU-ČSL	1.6	.5	-2.2	.5
	DŽJ	4.5	.6	-2.6	-2.0
	KSČM	3.7	2.0	-3.2	-2.4
	ODS	-3.5	-.5	2.2	1.3
	SPR-RSČ	1.9	-1.2	.5	-1.2
	US	-1.7	-3.8	2.8	3.5
	jiná strana	1.0	-1.2	1.2	-1.1

PS 1998 * Vzdělání respondenta Crosstabulation

Adjusted Residual

		Vzdělání respondenta			
		ZŠ	SŠ bez maturity	SŠ s maturitou	VŠ
PS 1998	ČSSD	o	+	o	o
	KDU-ČSL	o	o	-	o
	DŽJ	+++	o	--	-
	KSČM	+++	+	--	-
	ODS	---	o	+	o
	SPR-RSČ	o	o	o	o
	US	o	---	++	+++
	jiná strana	o	o	o	o

Tabulka 5 Kontingenční tabulka – Znaménkové schéma

Pro specifikaci u kterých kategorií nastal významný rozdíl můžeme použít adjustovaná standardizovaná rezidua. Tyto hodnoty porovnáváme s kvantily standardizovaného normálního rozložení pro námi zvolenou hladinu spolehlivosti.

Výsledek této analýzy lze graficky znázornit znaménkovým schématem (viz obrázek). Tabulku s adjustovanými rezidui snadno upravíme do uvedené podoby pomocí skriptu *Znaménkové schéma*, který je volně k dispozici na stránkách www.acrea.cz. Skript porovnává adjustovaná residua s 95%, 99% a 99,9% kvantily standardizovaného normálního rozdělení (tj. zaokrouhleně ±1,96 / ±2,58 / ±3,29)

a hodnoty vně těchto intervalů indikují (na příslušné hladině spolehlivosti) narušení nezávislosti v dané buňce. V každé buňce je potom znaménko rezidua uvedeno tolikrát, kolik z uvedených mezí bylo překročeno.

Z naší tabulky tedy například zjistíme, že mezi respondenty, kteří volili DŽJ a KSČM je výrazně více osob se základním vzděláním, než bychom očekávali a naopak mezi voliči ODS je takových respondentů méně. Mezi voliči US je výrazně více vysokoškoláků.

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

2 MEANS - PRŮMĚRY

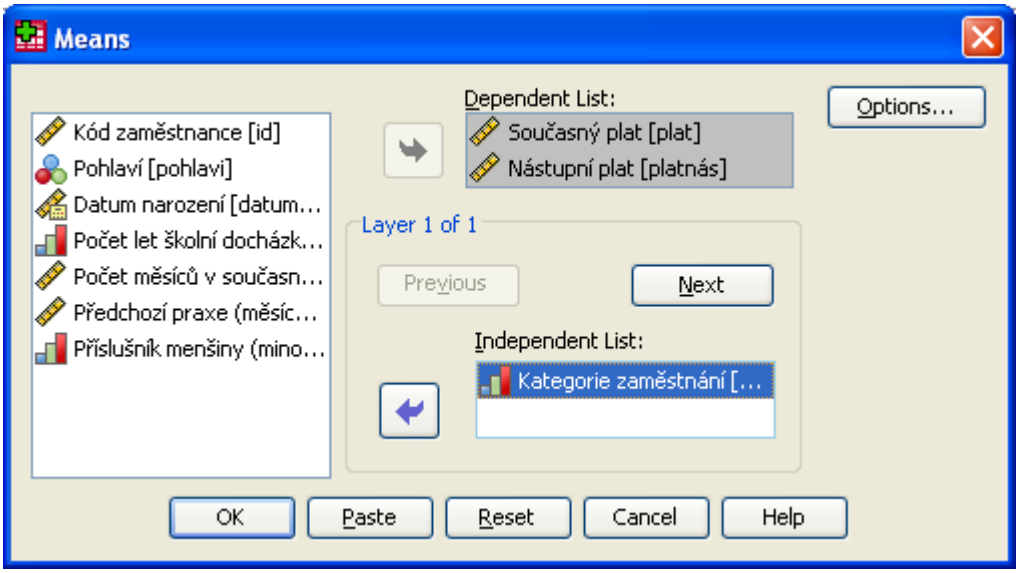
2.1 Tabulky průměrů

Procedura *Means* se užívá pro porovnání průměrů a dalších popisných statistik v podskupinách. Analyzovat můžeme současně několik číselných proměnných. Zadáme-li více kategorizovaných faktorů určujících podskupiny, lze výstup organizovat buď v samostatných tabulkách pro každou třídící proměnnou zvlášť, nebo jako třídění vyššího stupně. K dispozici je rovněž tabulka analýzy rozptylu (ANOVA) a test lineární závislosti průměrů pro první stupeň třídění.

2.2 Volání procedury v IBM SPSS Statistics

Analyze → Compare Means → Means

2.2.1 Nastavení dialogu



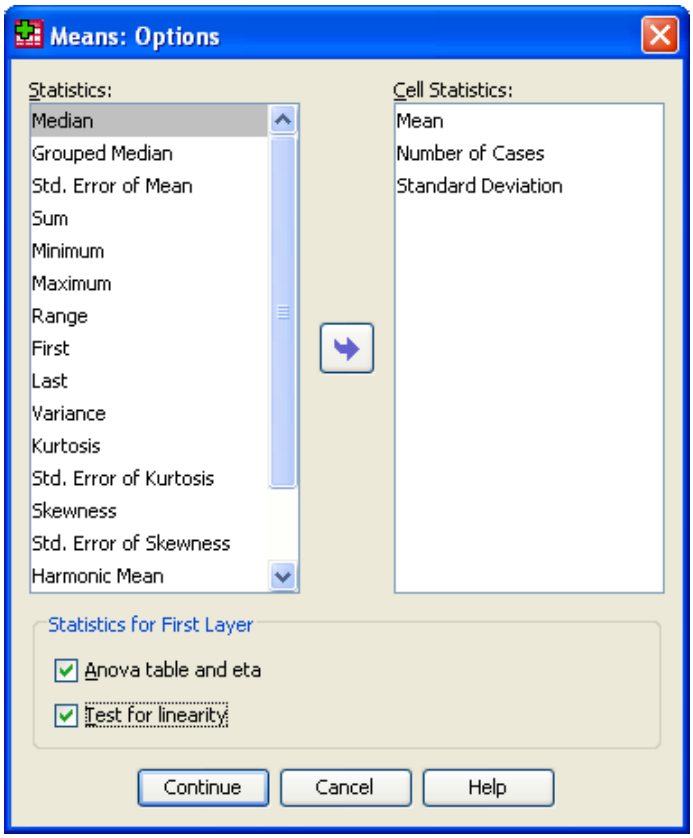
Obrázek 4 Nastavení dialogu Means

- Do pole Dependent list přeneseme číselné proměnné, jejichž popisné statistiky počítáme.
- Proměnné určující skupiny zadáme do pole Independent list. Jestliže sem vložíme více proměnných současně, vytvoří se pro každou z nich samostatná tabulka.

- Pokud potřebujeme získat tabulku vyššího stupně třídění, využijeme tlačítka Previous a Next, která přepínají mezi jednotlivými vrstvami třídění.

Chceme-li například vytvořit tabulku třídění druhého stupně podle *kategorií zaměstnání* a *pohlaví*, přeneseme do pole *Independent List* proměnnou *kategorie zaměstnání*, stiskneme tlačítko *Next* a ve druhé vrstvě zadáme opět do pole *Independent List* proměnnou *pohlaví*.

2.2.2 Tlačítko Options



Obrázek 5 Nastavení dialogu Means, tlačítko Options

Pomocí tlačítka *Options* volíme statistiky, které se zobrazí ve výstupové tabulce, a případně doplňující statistické testy.

Statistics, Cell Statistics (Zadávání statistik)

V levé části okna *(Statistics)* se nachází široký výběr popisných statistik. Požadované statistiky přeneseme šipkou do pole *Cell Statistics*.

Anova table and eta (Analýza rozptylu)

Po zaškrtnutí políčka *Anova table and eta* se na výstupu kromě tabulky popisných statistik zobrazí také tabulka analýzy rozptylu (ANOVA) pro třídící proměnné uvedené v první vrstvě. Tabulka umožňuje posoudit na základě statistického testu, zda jsou rozdíly průměrů ve skupinách významné. Koeficient *eta*, který nabývá hodnot od nuly do jedné, pak vyjadřuje míru závislosti třídících a analyzovaných proměnných (čím větší *eta*, tím silnější závislost).

Test for linearity (Test linearity)

Při označení políčka *Test for linearity* se modifikují tabulky analýzy rozptylu, kde je navíc testována lineární závislost průměrů na skupinách první vrstvy. Kromě toho se spočítá také Pearsonův korelační koeficient. Tyto analýzy mají smysl v případě, že třídící proměnná je ordinální (kategorie jsou uspořádané).

2.3 Výstupy

2.3.1 Přehled o počtu platných a chybějících případů

Case Processing Summary

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
Současný plat * Kategorie zaměstnán	474	100.0%	0	.0%	474	100.0%
Nástupní plat * Kategorie zaměstnání	474	100.0%	0	.0%	474	100.0%

Tabulka 6 Přehled o počtu platných a chybějících případů

Informace o počtu platných a chybějících hodnot pro uvedené kombinace proměnných.

2.3.2 Přehled statistik

Report

Kategorie zaměstnání		Současný plat	Nástupní plat
úředník	Mean	$27,838.54	$14,096.05
	N	363	363
	Std. Deviation	$7,567.995	$2,907.474
manuál. zaměstnání	Mean	$30,938.89	$15,077.78
	N	27	27
	Std. Deviation	$2,114.616	$1,341.235
manažer	Mean	$63,977.80	$30,257.86
	N	84	84
	Std. Deviation	$18,244.776	$9,980.979
Total	Mean	$34,419.57	$17,016.09
	N	474	474
	Std. Deviation	$17,075.661	$7,870.638

Tabulka 7 Přehled statistik procedury Means

Tabulka zobrazuje popisné statistiky zadaných proměnných uvnitř definovaných skupin. Strukturu tabulky můžeme dále upravit pivotováním. Pomocí této techniky lze modifikovat hierarchii sloupců, řádků a vrstev tak, aby tabulka přehledně zobrazovala požadovanou informaci.

2.3.3 Analýza rozptylu

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
Současný plat * Kategorie zaměstnání	Between Groups	(Combined)	89438483926	2	44719241963	434.5	.000
	Within Groups		48478011510	471	102925714		
	Total		137916495436	473			
Nástupní plat * Kategorie zaměstnání	Between Groups	(Combined)	17925544532	2	8962772266	371.1	.000
	Within Groups		11375360433	471	24151508		
	Total		29300904965	473			

Tabulka 8 Analýza rozptylu (1)

Tabulku analýzy rozptylu využijeme pro testování, zda kategorie zaměstnání je významným faktorem ovlivňujícím současný nebo nástupní plat zaměstnance. Nulová hypotéza je formulována tak, že průměrný plat na celé populaci je u všech

kategorií zaměstnání stejný. Vzhledem k tomu, že hodnota signifikancí v posledním sloupci tabulky je v obou případech velmi malá, můžeme na 95% hladině spolehlivosti zamítnout nulovou hypotézu.

Measures of Association

	Eta	Eta Squared
Současný plat * Kategorie zaměstnání	.805	.648
Nástupní plat * Kategorie zaměstnání	.782	.612

Tabulka 9 Analýza rozptylu (2)

Ze sloupce *Eta Squared* zjistíme, že *kategorie zaměstnání* vysvětlují variabilitu proměnné *současný plat* z 64,8 %, *nástupní plat* z 61,2 %.

2.3.4 Testování linearity

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
Současný plat * Kategorie zaměstnání	Between Groups	(Combined)	89438483926	2	44719241963	434.5	.000
		Linearity	83933110461	1	83933110461	815.5	.000
		Deviation from Linearity	5505373465	1	5505373465	53.5	.000
	Within Groups		48478011510	471	102925714		
	Total		137916495436	473			
Nástupní plat * Kategorie zaměstnání	Between Groups	(Combined)	17925544532	2	8962772266	371.1	.000
		Linearity	16687316491	1	16687316491	690.9	.000
		Deviation from Linearity	1238228041	1	1238228041	51.3	.000
	Within Groups		11375360433	471	24151508		
	Total		29300904965	473			

Tabulka 10 Analýza rozptylu, testování linearity

Budeme-li považovat kategorie zaměstnání za ordinální proměnnou, můžeme rovněž využít test linearity pro posouzení, zda v datech existuje lineární vztah mezi současným nebo nástupním platem a kategoriemi zaměstnání. V našem případě jsou všechny hodnoty signifikancí v posledním sloupci velice nízké, což napovídá, že lineární závislost v datech existuje, je zde však také výrazná odchylka od linearity.

Measures of Association

	R	R Squared	Eta	Eta Squared
Současný plat * Kategorie zaměstnání	.780	.609	.805	.648
Nástupní plat * Kategorie zaměstnání	.755	.570	.782	.612

Tabulka 11 Analýza rozptylu, testování linearity

Ve sloupci tabulky označeném *R* jsou uvedeny hodnoty Pearsonova lineárního korelačního koeficientu daných proměnných. Následující sloupec *R Squared* udává druhou mocninu tohoto koeficientu. Jedná se o míru, která charakterizuje kvalitu lineárního modelu. (Přesněji *R Squared*, někdy též *koeficient determinace*, udává procento variability vysvětlené pomocí lineárního regresního vztahu.)

3 OLAP CUBES – DATOVÉ KOSTKY OLAP

3.1 Datové kostky OLAP

Datové kostky OLAP (*Online Analytical Processing*) slouží k zobrazení popisných statistik (průměr, medián, rozptyl, rozpětí apod.) do vícerozměrných tabulek. Metoda umožňuje rozdělit datový soubor do několika skupin a pro každou z nich spočítat popisné statistiky samostatně. Skupiny definujeme na základě hodnot jedné nebo více kategorizovaných proměnných.

Pomocí kostek OLAP vypočteme popisné statistiky vybraných číselných proměnných v určených skupinách a zobrazíme je v tabulce, která obsahuje kromě řádků a sloupců také vrstvy. Jednotlivé vrstvy odpovídají skupinám (kombinacím kategorií třídících proměnných) a můžeme mezi nimi snadno přecházet pomocí rozbalovacího menu. Pro každou třídící proměnnou lze v rozbalovacím menu rovněž zadat volbu „celkem“. Vícerozměrnou tabulku můžeme přirovnat k několika klasickým tabulkám, které leží na sobě – vždy je vidět pouze horní tabulka, ale snadno zařídíme, aby se nahoru dostala jiná.

Procedura *OLAP Cubes* zahrnuje široké spektrum popisných statistik od jednoduchých až po složitější statistiky pro centrální tendenci a disperzi.

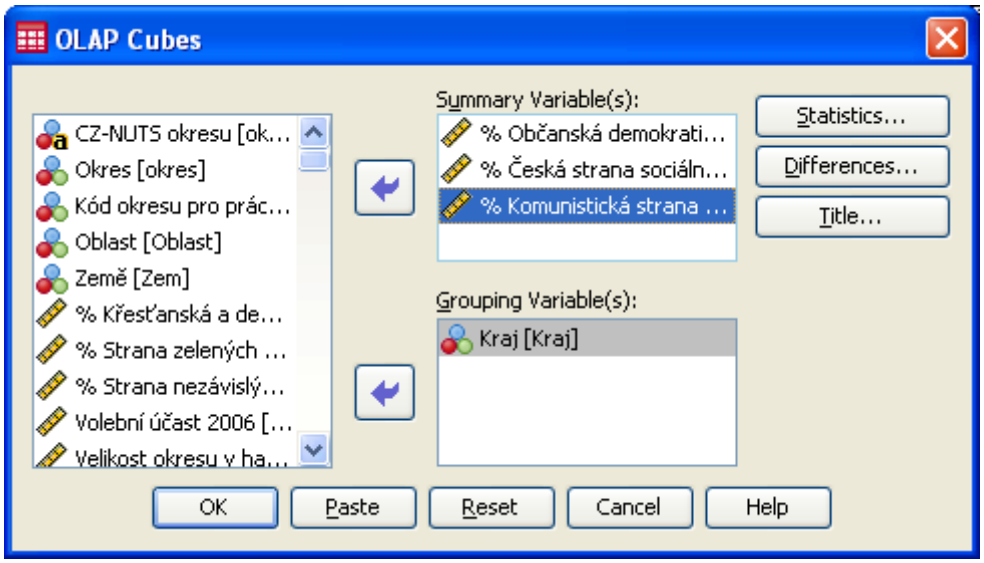
EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

3.2 Volání procedury v IBM SPSS Statistics

Analyze → Reports → OLAP Cubes

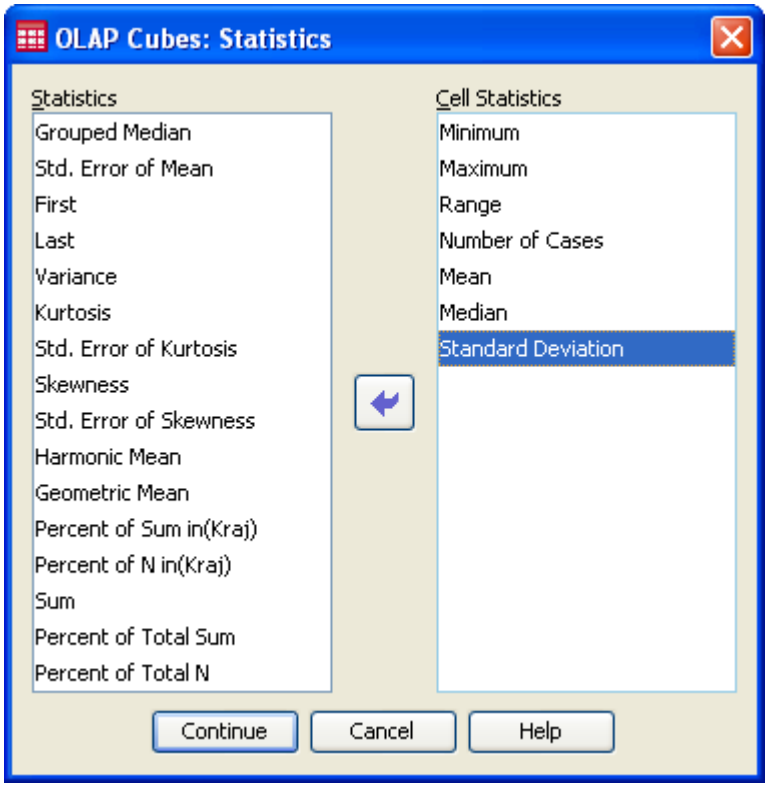
3.2.1 Nastavení dialogu



Obrázek 6 Nastavení dialogu OLAP Cubes

- Do pole *Summary Variable(s)* přeneseme číselné proměnné, jejichž statistiky počítáme.
- Do pole *Grouping Variable(s)* zadáme kategorizované proměnné, které rozdělí data do skupin.

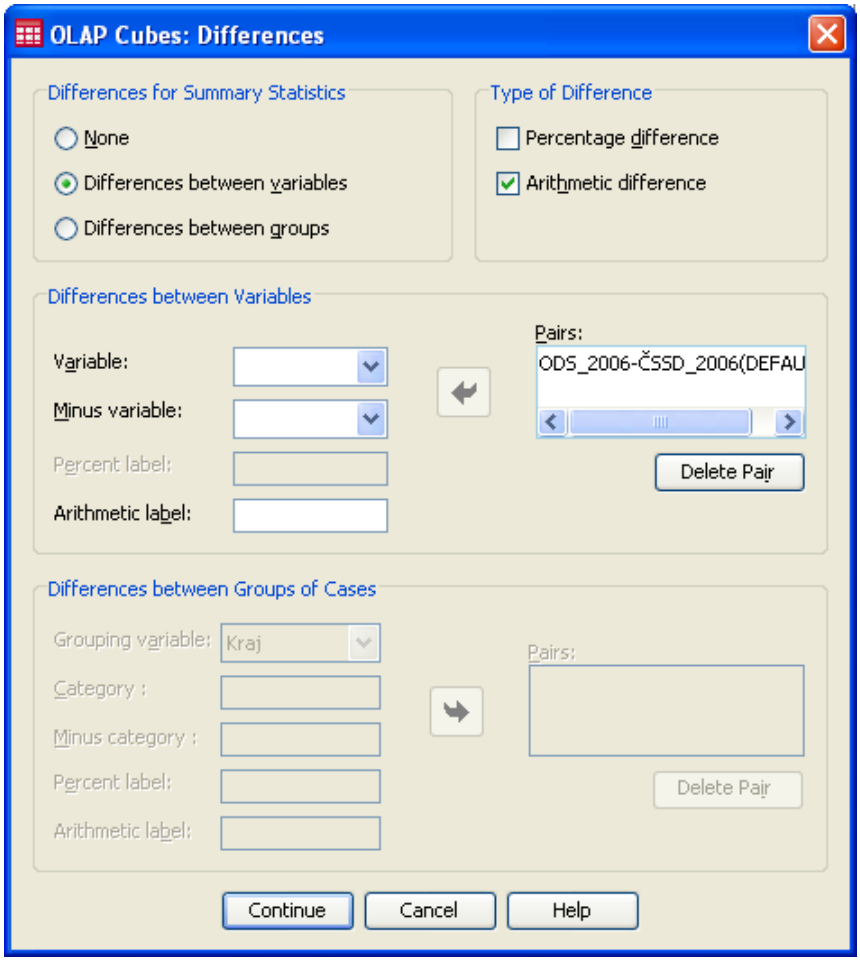
3.2.2 Tlačítko Statistics



Obrázek 7 Nastavení dialogu OLAP Cubes, tlačítko Statistics

Pomocí tlačítka *Statistics* zadáme požadované popisné statistiky. V levé části okna je k dispozici celá řada různých charakteristik. Z tohoto seznamu vybereme statistiky, které má obsahovat výsledná tabulka, a šipkou je přeneseme do pole *Cell Statistics*.

3.2.3 Tlačítko Differences



Obrázek 8 Nastavení dialogu OLAP Cubes, tlačítko Differences

Tlačítko *Differences* dovoluje přidat do tabulky informace o rozdílu vybraných proměnných (objeví se v tabulce jako další řádek) nebo vyjádřit rozdíl statistik mezi dvěma skupinami určenými třídící proměnnou (zobrazí se jako další volba rozbalovacího menu).

Differences for Summary Statistics

V této části nastavujeme typ požadovaných diferencí:

- *None* (diference se nepočítají)
- *Differences between variables* (rozdíly mezi proměnnými)
- *Differencies between groups* (rozdíly mezi skupinami).

Type of Difference

V části *Type of Difference* volíme způsob výpočtu diferencí: *Percentage difference* (procentní rozdíl), *Arithmetic difference* (aritmetický rozdíl) nebo obojí.

Differences between Variables

Toto pole je určeno k podrobnější specifikaci rozdílů mezi proměnnými a je aktivní pouze v případě, že jsme v horní části okna označili volbu *Differences between variables*.

Pomocí rozbalovacího menu zadáme proměnnou, od které se bude odečítat (*Variable*) a proměnnou, která se bude odečítat (*Minus variable*). Do pole *Percent label*, resp. *Arithmetic label* můžeme zadat název, pod kterým se tento rozdíl zobrazí ve výsledné tabulce. V případě, že pole nevyplníme, automaticky se za název dosadí vzorec, podle kterého byla diference vypočítána. Takto definovanou dvojici přeneseme šipkou do pole *Pairs,* čímž nastavení potvrdíme, a můžeme pokračovat zadáním další dvojice proměnných. V případě potřeby odstraníme vybraný pár z pole *Pairs* tlačítkem *Delete Pair*.

Differences between Groups of Cases

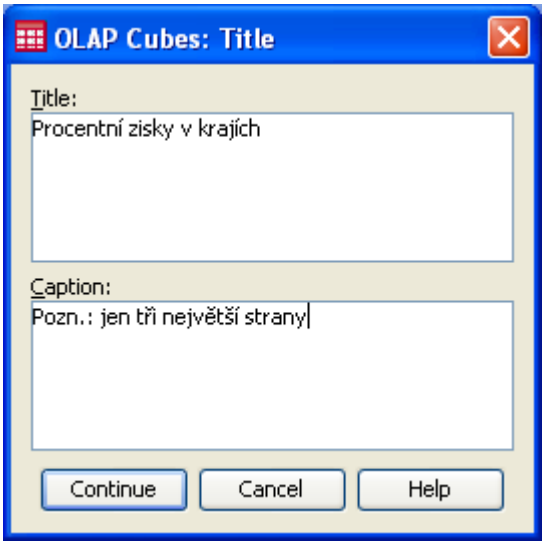
Tato část okna je určena k podrobnější specifikaci rozdílů mezi kategoriemi a je k dispozici pouze v případě volby *Differences between groups*.

Z rozbalovacího menu v poli *Grouping variable* vybereme proměnnou, na základě jejíž kategorií bude rozdíl definován. Do pole *Category* zadáme hodnotu kategorie, od které se bude odečítat a do pole *Minus category* hodnotu kategorie, která bude odečítána. Název, pod kterým se tento rozdíl zobrazí v rozbalovacím seznamu výsledné tabulky, můžeme zadat do pole *Percent label*, resp. *Arithmetic label*. Tímto způsobem zadáme všechny požadované dvojice kategorií. Tlačítkem *Delete Pair* odstraníme v případě potřeby již definovanou dvojici.

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

3.2.4 Tlačítko Title



Obrázek 9 Nastavení dialogu OLAP Cubes, tlačítko Title

Pomocí tlačítka *Title* můžeme upřesnit název tabulky nebo doplnit poznámku pod tabulku. Text nadpisu zadáváme v horní části okna, text poznámky do pole *Caption*.

3.3 Výstupy

Datový soubor obsahuje výsledky voleb do Poslanecké sněmovny Parlamentu České republiky v roce 2006 za okresy (každý případ představuje jeden okres). Proceduru *OLAP Cubes* využijeme k zobrazení vybraných popisných statistik za kraje pro tři nejsilnější strany.

Přehled o počtu zahrnutých a vyloučených případů

Case Processing Summary

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
% Občanská demokratická strana * Kraj	77	100,0%	0	,0%	77	100,0%
% Česká strana sociálně demokratická * Kraj	77	100,0%	0	,0%	77	100,0%
% Komunistická strana Čech a Moravy * Kraj	77	100,0%	0	,0%	77	100,0%

Tabulka 12 Přehled o počtu zahrnutých a vyloučených případů

Tabulka poskytuje informaci o celkovém počtu případů a dále o počtu platných a chybějících případů. Případ je považován za chybějící a vyloučen z analýzy, pokud obsahuje vynechanou hodnotu (systémovou nebo uživatelem definovanou) u některé proměnné z dvojice *strana*kraj*. Každá dvojice je přitom posuzována samostatně.

3.3.1 OLAP kostka

Procentní zisky v krajích

Kraj:Total

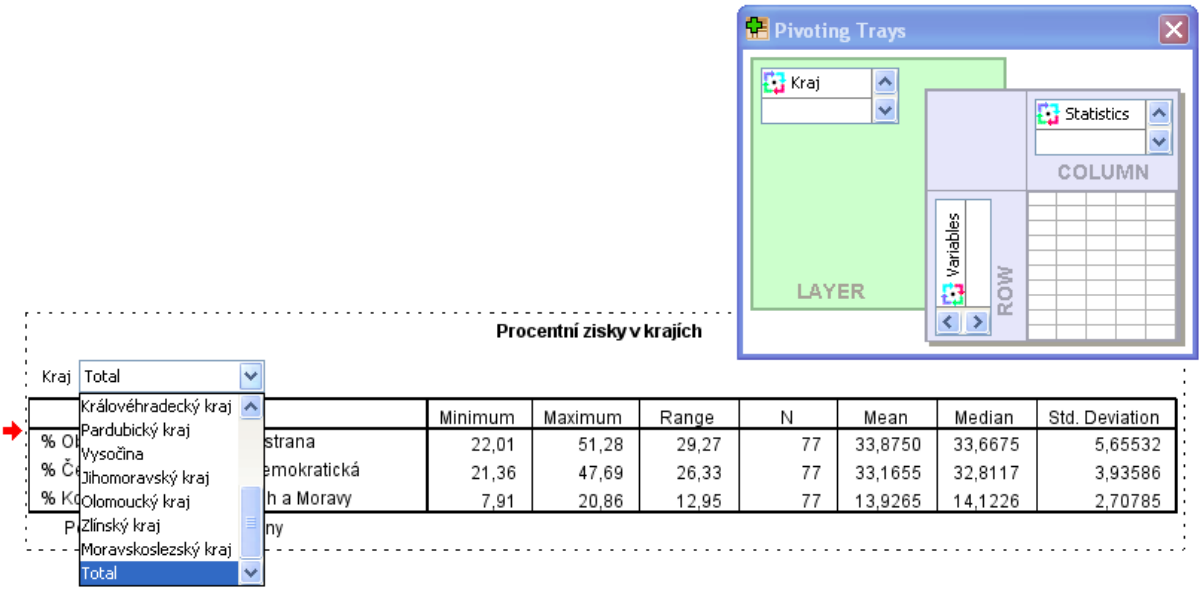
	Minimum	Maximum	Range	N	Mean	Median	Std. Deviation
% Občanská demokratická strana	22,01	51,28	29,27	77	33,8750	33,6675	5,65532
% Česká strana sociálně demokratická	21,36	47,69	26,33	77	33,1655	32,8117	3,93586
% Komunistická strana Čech a Moravy	7,91	20,86	12,95	77	13,9265	14,1226	2,70785

Pozn.: jen tři největší strany

Tabulka 13 OLAP kostka (1)

Kostka OLAP zobrazuje zvolené popisné statistiky pro vybrané proměnné, v našem případě pro procentní zisky stran ČSSD, ODS a KSČM. Tabulka je doplněna námi definovaným názvem a poznámkou.

V tabulce jsou zobrazeny informace, které se vztahují k celku, tj. ke všem krajům, jak je uvedeno vlevo nad okrajem tabulky. Ačkoliv ve výstupovém okně vidíme pouze tuto jednoduchou tabulku, jsou pod ní skryté tabulky pro další kategorie třídící proměnné (*Kraj*). Mezi jednotlivými vrstvami se můžeme pohybovat pomocí rozbalovacího menu. Po dvojitém poklikání na tabulku seznam rozbalíme a zvolíme požadovanou kategorii.



Obrázek 10 Úprava OLAP kostky pivotováním

V případě, že jsme požadovali také výpočet statistik pro rozdíl mezi proměnnými (zde ODS - ČSSD), zobrazí se jako další řádek tabulky.

Procentní zisky v krajích

Kraj:Total

	Minimum	Maximum	Range	N	Mean	Median	Std. Deviation
% Občanská demokratická strana	22,01	51,28	29,27	77	33,8750	33,6675	5,65532
% Česká strana sociálně demokratická	21,36	47,69	26,33	77	33,1655	32,8117	3,93586
% Komunistická strana Čech a Moravy	7,91	20,86	12,95	77	13,9265	14,1226	2,70785
% Občanská demokratická strana - % Česká strana sociálně demokratická	,65	3,58	2,94	0	,7095	,85577	1,71946

Pozn.: jen tři největší strany

Tabulka 14 OLAP kostka (2)

Uspořádání výsledné tabulky lze měnit pivotováním. Pomocí této techniky můžeme upravit strukturu tabulkových dimenzí, to znamená modifikovat hierarchii sloupců, řádek a vrstev. Tabulku je třeba nejprve dostat do editačního modu (dvakrát na ni poklikáme) a pomocí nabídky *Pivot*, *Pivoting Trays* zobrazíme pivotní diagram. Přenášením objektů v diagramu potom podle potřeby upravíme strukturu tabulky.

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

SEZNAM POUŽITÉ LITERATURY

Řehák, J., Brom, O. (2015). SPSS - Praktická analýza dat, Brno, Česká republika: Computer Press

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

SEZNAM OBRÁZKŮ

Obrázek 19 Nastavení dialogu Crosstabs 5
Obrázek 20 Nastavení dialogu Crosstabs, tlačítko Statistics 6
Obrázek 21 Nastavení dialogu Crosstabs, tlačítko Cells..... 7
Obrázek 22 Nastavení dialogu Means 15
Obrázek 23 Nastavení dialogu Means, tlačítko Options 16
Obrázek 24 Nastavení dialogu OLAP Cubes 22
Obrázek 25 Nastavení dialogu OLAP Cubes, tlačítko Statistics 23
Obrázek 26 Nastavení dialogu OLAP Cubes, tlačítko Differences.....24
Obrázek 27 Nastavení dialogu OLAP Cubes, tlačítko Title.....26
Obrázek 28 Úprava OLAP kostky pivotováním 28

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

SEZNAM TABULEK

Tabulka 12 Přehled o počtu platných a chybějících případů 9
Tabulka 13 Kontingenční tabulka 10
Tabulka 14 Kontingenční tabulka – řádková procenta 11
Tabulka 15 Kontingenční tabulka – test Chí-kvadrát 11
Tabulka 16 Kontingenční tabulka – Znaménkové schéma 13
Tabulka 17 Přehled o počtu platných a chybějících případů 17
Tabulka 18 Přehled statistik procedury Means 18
Tabulka 19 Analýza rozptylu (1) 18
Tabulka 20 Analýza rozptylu (2) 19
Tabulka 21 Analýza rozptylu, testování linearity 19
Tabulka 22 Analýza rozptylu, testování linearity 20
Tabulka 23 Přehled o počtu zahrnutých a vyloučených případů 26
Tabulka 24 OLAP kostka (1) 27
Tabulka 25 OLAP kostka (2) 28

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY