

V této rubrice se budeme snažit uvádět vždy alespoň tři pravidelné části: Anotace (nové nebo málo známé součásti systému), Konzultace (odpovědi na nejčastější otázky, tj. řešení praktických problémů uživatelů), Informace (o rozšiřování systému, výuce, zajímavostech ap.). Kromě toho uvítáme krátká sdělení o těch zkušenostech uživatelů, které mohou pomoci i jiným.

Anotace: SPSS/PC+ CODEBOOK

Problémy s dokumentací datových souborů mají snad všichni výzkumníci. Způsobuje to práce s velkými počty souborů, návraty ke starším souborům, a nejvíce asi neustálá potřeba měnit a zavádět nové proměnné během analýzy. Nový produkt SPSS/PC+ CODEBOOK je výzkumným instrumentem, který má tuto situaci uživatelům SPSS/PC+ usnadnit a základní dokumentaci datových souborů poskytnout. Je to samostatný program, který poskytuje kódový klíč obsahující kompletní technickou i substantivní informaci o každé položce dotazníku (kódový klíč, jména a označení proměnných a označení hodnot, kódy vynechávané informace), a dále jakoukoli textovou dokumentaci o tom, jak byla data sebrána, kódována, resp. rekódována, instrukce tazatelů, informaci o četnostních rozloženích a základních statistikách a referenci k dalším souborům. Kromě těchto základních funkcí poskytuje CODEBOOK další, neméně důležité možnosti: vytváří instrument, který může být použit jako dotazník ve výzkumu a je navíc čitelný na počítači; může být editován a obměňován pro specifická použití, a to jak doplňováním informace, tak redukcí na podmnožinu položek apod.; poskytuje nejrozličnější rychlou informaci, např. seznamy otázek a názvů proměnných a reference, rychlá rozložení četností, statistiky ad. Výstupy lze řídit volitelnými a modifikovatelnými alternativami stylu (stylesheet). K řízení programu se používají nabídkové formy: okna, stékající menu, dialogové boxy. Výběr z nabídky je možný buď z klávesnice, nebo myší. Při realizaci programu jsou k dispozici dosti podrobné pomocné texty, které lze vyvolat na obrazovku. Manuál je přehledný, skládá se ze tří hlavních částí: A) vysvětlení pojmů a ovládání programu (23 s.), B) cvičební běhy (tutorials) které obsahují čtyři lekce: 1. "Getting Started" (ovládání menu), 2. Vytváření dokumentačního souboru na existujícím SPSS/PC+ souboru, 3. Vytváření nového dokumentačního souboru, 4. Vytváření výzkumného instrumentu (61 s.). C) "Reference Guide" obsahuje úplný popis všech možností a alternativ programu, seznam chybových hlášení a předem definované nabízené styly pro kódové klíče (68 s.).

Program SPSS/PC+ CODEBOOK je vítaným obohacením systému. Přivítají jej především uživatelé v institucích, kde se provádějí rutinní a opakované výzkumy nebo kde vytvářejí soubory systematicky a pravidelně a kde je dokumentace naprosto nezbytná (výzkumné organizace, administrativa, personalistika, finanční instituce ap.), instituce, které pravidelně získávají, resp. objednávají výzkumy pro vlastní potřebu (marketingová oddělení velkých firem) a dále ta místa, kde se vytváří archívy dat (např. při katedrách za účelem

výuky), při koordinaci komparačních projektů (kde se CODEBOOK stává také velmi efektivním prostředkem komunikace mezi koordinátorem a účastníky), ve vědeckých a výzkumných institucích, a obzvláště tam, kde se provádějí opakovaná šetření. SPSS/PC+ CODEBOOK je program nový, první zkušenosti s ním jsou dobré a zdá se, že se vzhledem ke své praktičnosti, velmi rychle rozšiřuje mezi výzkumníky i v komerční sféře. Proto je možno jej doporučit.

Konzultace

Otázka 1: *Jak je možné získat úsporné tisky pro třídění 1. stupně četností?*

Odpověď: Třídění 1. stupně četností je možné získat v SPSS/PC+ třemi základními způsoby: A) pomocí procedury FREQUENCIES, která je k tomu určena, B) pomocí REPORT, C) pomocí TABLES.

A) Procedura FREQUENCIES tabeluje každou položku zvlášť a proto je neúsporná (má však zato možnost velké řady voleb pro formáty a statistiky, pro potlačování tisků, pro sloupcové grafy). Odpověď na požadavek otázky poskytují další dvě procedury.

B) Procedura REPORT: Použijeme ji bez výpisu případů, a pokud nechceme četnostní rozložení také pro části, tak "vypneme" příkaz BREAK na dělení souboru. V seznamu proměnných zavádíme jako první DUMMY proměnnou (jinak nastane posun a ztratíme informaci o první proměnné, kterou tabelujeme). Schéma příkazu je vidět v příkladu, kde proměnné mají hodnoty 1-4. Kromě zde uvedeného procentního rozložení lze volit ovšem též absolutní četnosti a celou řadu dalších statistik.

Zadání:

```
REPORT /FORMAT AUTOMATIC /VARIABLES VALUE(DUMMY) A116 A117
A118 A119 A120
/BREAK (NOBREAK) /SUMMARY PERCENT (1,4) /SUMMARY MEAN.
```

Výstup (zkrácený):

	work	family	friends	leisure	politics
Total	100.0%	100.0%	100.0%	100.0%	100.0%
1	1.0	.4	1.1	2.9	21.3
2	2.5	1.2	17.6	20.3	42.8
3	34.5	11.1	57.4	52.6	28.3
4	62.0	87.3	24.0	24.2	7.6
Mean	3.573	3.853	3.043	2.981	2.222

C) Procedura TABLES: použijte spojení proměnných v řádcích pomocí + a místo sloupcových proměnných uveďte klíčové slovo (LABELS). Záměna (LABELS) a seznamu proměnných by vedla k transpozici tabulky (rozložení ve sloupcích).

Zadání:

```
TABLES /TABLE A116+A117+A118 BY (LABELS) /STATISTICS COUNT
CPCT.
```

Výstup:

	Very import.	Quite import.	Not very import.	Not at all
work				
Count	34	82	1116	2007
Count Percent	1.0%	2.5%	34.5%	62.0%
family				
Count	12	40	358	2825
Count Percent	.4%	1.2%	11.1%	87.3%
friends				
Count	34	570	1857	776
Count Percent	1.1%	17.6%	57.4%	24.0%
leisure time				

Výhodou REPORT je výzkumná přehlednost a větší rychlost, je také vhodnější pro objemné výstupy, TABLES poskytují přímo publikovatelné výstupy. V obou procedurách je možno využít mnoha voleb pro úpravu a doplnění výstupů.

Otázka 2: Máme publikované tabulky (nikoliv původní data o jednotkách) a chtěli bychom si na nich provést statistickou analýzu (chí-kvadrát test a koeficienty asociace). Je to možné?

Odpověď: Je to možné jen v případě, že v tabulkách jsou k dispozici absolutní četnosti, nebo je možné tyto četnosti rekonstruovat. Postup spočívá v tom, že a) každé pole tabulky budeme považovat za jednotku (případ) na vstupu, b) budeme definovat tři proměnné pro každé pole, např.: R = kód řádku tabulky pro dané pole, S = kód sloupce tabulky pro dané pole, CETN = absolutní četnost v poli, a pomocí těchto proměnných popíšeme každé pole, c) pro analýzu určíme váhu pomocí příkazu WEIGHT, kterou je proměnná CETN. Pro analýzu můžeme pojmenovat řádky a sloupce pomocí VALUE LABELS, volit libovolné statistiky. Např. pro tabulku:

```
56 73 19
41 84 68
```

můžeme použít sadu příkazů:

```
DATA LIST FREE /R S CETN.
BEGIN DATA
1 1 56
1 2 73
1 3 19
2 1 41
2 2 84
2 3 68
END DATA.
```

```

VAR LABELS R 'Řádky' S 'Sloupce'.
VALUE LABELS R 1 'řádek 1' 2 'řádek 2' / S 1 'sl1' 2 'sl2' 3 'sl3'.
WEIGHT BY CETN.
CROSSTABS /TABLES R BY S/ CELLS COUNT ROW/ STATISTICS CHI
CC.

```

Výstup:

R Řádky by S sloupce

Page 1 of 1

		S			Row Total
		sl1	sl2	sl3	
		Count			
		Row Pct			
			1.00	2.00	3.00
R	1.00	56	73	19	148
	řádek 1	37.8	49.3	12.8	43.4
	2.00	41	84	68	193
	řádek 2	21.2	43.5	35.2	56.6
Column		97	157	87	341
Total		28.4	46.0	25.5	100.0 atd.

Postup je možné použít i pro vstup vícerozměrných tabulek tak, že každé pole popisujeme kódy všech dimenzí, které jej určují. (Např. pro třídění třetího stupně zavedeme tři proměnné pro tři hlediska třídění a proměnnou "četnost".) Metodu vážení takto můžeme použít v procedurách, jako je FREQUENCIES, CROSSTABS, LOGLINEAR. U CATEGORIES takto můžeme vstupovat pouze s dvourozměrnými tabulkami do ANACOR.

Otázka 3: Mohu v SPSS/PC+ zpracovávat data z BMDP nebo ze STATGRAPHICS? Mohu použít soubory SPSS při aplikaci v programu GLIM?

Odpověď: Do SPSS/PC+ můžeme přímo překládat data z formátu LOTUS a SYMPHONY, DBASE, MULTIPLAN a EXCEL a samozřejmě můžeme vytvářet vlastní soubor ze vstupů v pevném i volném formátu ASCII. Převod do SPSS/PC+ je tedy možný tak, že soubory vytvořené v jiných souborech importujeme do některého z výše uvedených tvarů, které SPSS/PC+ umí číst. Postup převodu však je nesmírně ulehčen programem DBMS/COPY (jeho anotaci přineseme v některé z příštích rubrik), který převádí více než 70 různých formátů vzájemně na sebe a převod ze STATGRAPHICS a BMDP a do GLIM provede přímo, přičemž udrží (případně mírně upraví) veškeré převoditelné parametry (vynechané hodnoty, jména a označení).

Informace

* Na katedře statistiky a pravděpodobnosti VŠE Praha byl pořádán seminář k programovému systému SPSS každý druhý čtvrtek od března t.r. Tento seminář, který byl akcí České statistické společnosti a byl otevřen i dalším

zájemcům, vedl doc. Petr Hebák. Na pořadu bylo ovládání verze 4 a jednotlivé procedury.

- * Obdobný seminář byl pořádán na katedře matematiky přírodovědecké fakulty Masarykovy univerzity v Brně. Vedl jej doc. J. Michálek a byl otevřen zájemcům z celé univerzity i mimo ni.
- * Ministerstvo práce a sociálních věcí ČR pořádalo dva běhy školení pro pracovníky Úřadů práce v ČR, které zajišťovali učitelé fakulty společenských věd UK v Praze (první běh) a SC&C s. s. r. o. (druhý běh).
- * Výuka statistiky a analýzy dat pomocí SPSS probíhá již na fakultě společenských věd UK, Praha, na filozofické fakultě UK, Praha, na VŠE, Praha, a na MUB Brno.

Otázky, informace a příspěvky do této rubriky zasílejte na adresu redakce Sociologického časopisu, s označením Rubrika SPSS. Otázky se budeme snažit odpovědět nejen přímo adresátovi, ale také je použít v této rubrice.

Jan Řehák