

Příprava dat před analýzou

- uspořádání dat
 - kontrola přesnosti dat
 - vytvoření struktury databáze
 - vkládání dat do PC
 - transformace dat
-

Uspořádání dat

- výzkumná data mohou pocházet z různých zdrojů: dotazníky, rozhovory, měření při experimentech (pretest, posttest), pozorování
 - je třeba všechny formuláře, dotazníky, záznamové archy atd. řádně označit, aby je vždy bylo možno i zpětně dohledat
-

Uspořádání dat

- původní data se u profesionálních výzkumů po určitou dobu **archivují** (min. 5-7 let)
 - je vhodné vytvořit si databázi – **přehled** všech použitých zdrojů dat (např. v MS Access nebo Excel, Statistica, SPSS)
-

Kontrola přesnosti dat

- někdy vhodné provést již při sběru dat – pak je možno příp. chybějící nebo jinak problematické odpovědi vyřešit na místě
 - **kontrola**: čitelnosti odpovědí, úplnosti odpovědí, důležitých informací (datum, jméno výzkumníka atp.)
 - je vhodné z dalších kroků vyloučit nevalidní odpovědi (na všechny otázky stejné atd.)
-

Vytvoření struktury databáze

- data je možno vkládat přímo do statistických programů nebo do databázových programů (a potom je převést)
 - rozsáhlé výzkumné projekty mají vytištěnu podrobnou kódovací příručku (*codebook*) s popisem proměnných
-

Vytvoření struktury databáze

pro každou proměnnou:

- jméno proměnné (např. *pohlaví*)
 - popis proměnné (např. *celkový skór dotazníku deprese*)
 - popis hodnot proměnné (např. *1=chlapci, 2=dívky*)
 - způsob kódování chybějících hodnot (např. 9)
 - formát proměnné (numerický, textový, datum...)
-

Vkládání dat do PC

- pro **kontrolu přesnosti vkládání** existují nejrůznější postupy
 - např. dvojitě vkládání (*double entry*)
 - speciální program, který při druhém vkládání porovnává data s původním údajem a upozorní na chyby; není příliš rozšířeno
-

Vkládání dat do PC

- jednodušší je náhodně zkontrolovat určitý počet případů (při velkém procentu chyb kódovat znovu a porovnat)
 - vždy je nutno alespoň zkontrolovat, zda rozsah zadaných hodnot odpovídá předpokladům (v programu *STATISTICA* odkaz *Data - Verify Data*)
-

Transformace dat

- **obrácení položek** –některé položky bývají formulovány v opačném směru než zbytek škály;
 - jejich hodnoty je třeba překódovat (např. u 5 bodové škály se bude 1 rovnat 5, 2 bude 4 atd. až 5 bude 1)
-

Transformace dat

- vzorec pro obrácení škály

nová hodnota = nejvyšší hodnota +1 - stará hodnota

- ve STATISTICE se vzorec zadává v definici proměnných – do rámečku pro dlouhý název proměnné
-

Transformace dat

- **výpočet celkových skóru** – většinou součtem hodnot několika proměnných (příp. průměr atp.)
 - zadá se příslušný vzorec, např. pro součet **celk = p1 + p2 + p3**
 - nebo ve STATISTICE rovnou funkce (*sum, mean*)
-

Transformace dat

- **sloučení hodnot proměnných** –
např. ze 7 bodové škály chceme udělat 3 bodovou; nebo z věku zadaného v letech vytvořit 3 věkové skupiny
 - ve STATISTICE se zadává v *Data – Recode Values of Variable*
-

Příklady

- Vytvořte strukturu databáze, obsahující tyto proměnné:
 - 1) číslo osoby
 - 2) jméno osoby
 - 3) věk (v letech)
 - 4) pohlaví
(0 = muž, 1 = žena)
 - 5) rodinný stav
(0 = svobodný, 1 = ženatý/vdaná, 2 = rozvedený, 3 =vdovec/vdova)
 - 6) příjem (Kč/měsíc)
 - 7) počet dětí
 - 8) odpověď na položku *Většinou jsem se sebou spokojen*
(1 = vůbec nesouhlasím, 2 = spíše nesouhlasím, 3 = něco mezi, 4 = spíše souhlasím, 5 = zcela souhlasím)
 - 9) odpověď na položku *Občas si připadám jako bezcenný člověk*
(1 = vůbec nesouhlasím, 2 = spíše nesouhlasím, 3 = něco mezi, 4 = spíše souhlasím, 5 = zcela souhlasím - popisy hodnot je možno kopírovat z předchozí položky)
 - 10) odpověď na položku *Mám sám sebe rád*
(1 = vůbec nesouhlasím, 2 = spíše nesouhlasím, 3 = něco mezi, 4 = spíše souhlasím, 5 = zcela souhlasím)
-

Příklady

- Data uložte do souboru pod názvem *esteem*
 - Zadejte fiktivní data 5 – 10 osob: u jedné z osob budou některé údaje chybět (např. neuvedla příjem a odpověď na položku 8)
 - Zkontrolujte, zda rozsah hodnot proměnných u zadaných dat odpovídá předpokládanému rozsahu
-

Příklady

- Vytvořte novou proměnnou – celkové sebehodnocení, která bude průměrem z položek 8 – 10
(! Ještě **předtím** je třeba obrátit položku 9, která je formulována v opačném směru než ostatní dvě)
-

Příklady

- Vytvořte novou proměnnou, která rozdělí soubor na dvě skupiny – bezdětné osoby a osoby s alespoň jedním dítětem (zkopírujte proměnnou 7 a potom rekódujte její hodnoty)
 - Vytvořte novou proměnnou, která rozdělí soubor na tři příjmové skupiny – osoby s příjmem pod 10 tis. Kč, 10 tis. – 20 tis. Kč a nad 20 tis. Kč (zkopírujte proměnnou 6 a potom rekódujte její hodnoty)
-

Příklady

- u všech proměnných (včetně nově vytvořených) určete, na jaké úrovni měření jsou měřeny
-