

Metodologie pro Informační studia a knihovnictví 2

Modul 6: Transformace proměnných

Co se dozvíte v tomto modulu?

- Jak vytvořit novou proměnnou pomocí rekódování?
- Jak vytvořit novou proměnnou pomocí aritmetických operací?
- Jak vytvořit novou proměnnou pomocí sčítání výskytů hodnot proměnné?
- Jak vytvořit novou proměnnou pomocí seřazení položek?

Obsah

Vytváření nových proměnných	2
Rekódování	2
Aritmetické operace	4
<i>Aritmetické operace v Excelu</i>	5
<i>Aritmetické operace v SPSS</i>	5
Sčítání výskytů	6
<i>Sčítání výskytů v Excelu</i>	6
<i>Sčítání výskytů v SPSS</i>	6
Seřazení položek.....	7
<i>Seřazení položek v Excelu</i>	7
<i>Seřazení položek v SPSS</i>	8

Vytváření nových proměnných

Při analýze někdy potřebujeme zjistit a využít proměnné, které v dotazníku přímo nezkoumáme, ale můžeme si je snadno vytvořit z existujících proměnných. Například pro srozumitelnější analýzu může být výhodné kategorizovat si věkové skupiny. Nebo chceme z proměnné „rok narození“ vytvořit srozumitelnější proměnnou „věk“.

Obecně se s proměnnými dá dělat řada jednoduchých operací, my budeme využívat především:

- rekódování,
- aritmetické operace,
- kategorizace dle percentilů,
- sčítání výskytů,
- seřazení položek.

POZOR!!! Nové proměnné vždy vytváříme do nového sloupce – tak, abychom neztratili původní proměnné, kdybychom je ještě k něčemu potřebovali.

Rekódování

Rekódování nahrazuje kódy hodnot proměnných jinými kódy. Lze jej využít pro vytváření obecnějších kategorií u nominálních, ordinálních i kardinálních proměnných. Rekódování využíváme i pro otočení škály otázky (ve složitějších dotaznících bývají některé škály otočené, abychom udrželi respondentovu pozornost – při výsledné analýze, kdy vytváříme sumační indexy, je nutné reorientovat škály tak, aby byly všechny orientované jedním směrem).

Speciální příklad rekódování může být **kategorizace dle percentilů**. Ta nám umožní rozdělit respondenty na X stejně velkých skupin dle hodnot námi zvolené proměnné – například podle výše příjmů nebo dle věku (například bychom si chtěli respondenty rozdělit na „bohaté“, střední třídu“ a „chudé“). Za bohaté bychom považovali respondenty v horním kvartilu (percentil 75), za chudé respondenty v dolním kvartilu (percentil 25) a za střední třídu respondenty ve středních kvartilech.

Příklad: V dotazníku máme otázku zjišťující ekonomickou aktivitu respondentů (otázka č. 9). Třídění v této otázce je však velmi jemné a nás zajímá pouze, zda existují rozdíly v přístupu ke vzdělávání mezi lidmi, kteří jsou zaměstnaní (do této kategorie si přiřadíme i pracující na částečný úvazek a OSVČ) a ostatními. Potřebujeme proto vytvořit novou proměnnou, která bude rozdělovat respondenty jen na dvě kategorie – na ty, kteří pracují (plus OSVČ) a na ostatní.

Kategorie tedy bude nabývat nově dvou validních hodnot:

- zaměstnan/a anebo OSVČ 1
- ostatní 2

Hodnoty proměnné tedy rekódujeme takto:

- 1, 2, 3 → 1
- 4, 5, 6, 7, 8 → 2

Rekódování v Excelu

Na rekódování v Excelu existuje řada pluginů, ale je možné využít i běžný postup „Najít a nahradit“ (CTRL+H). Je potřeba však dávat pozor na to, v jakém pořadí rekódujeme, aby nedošlo k dvojitému překódování.

V našem příkladu by tedy byl postup následující:

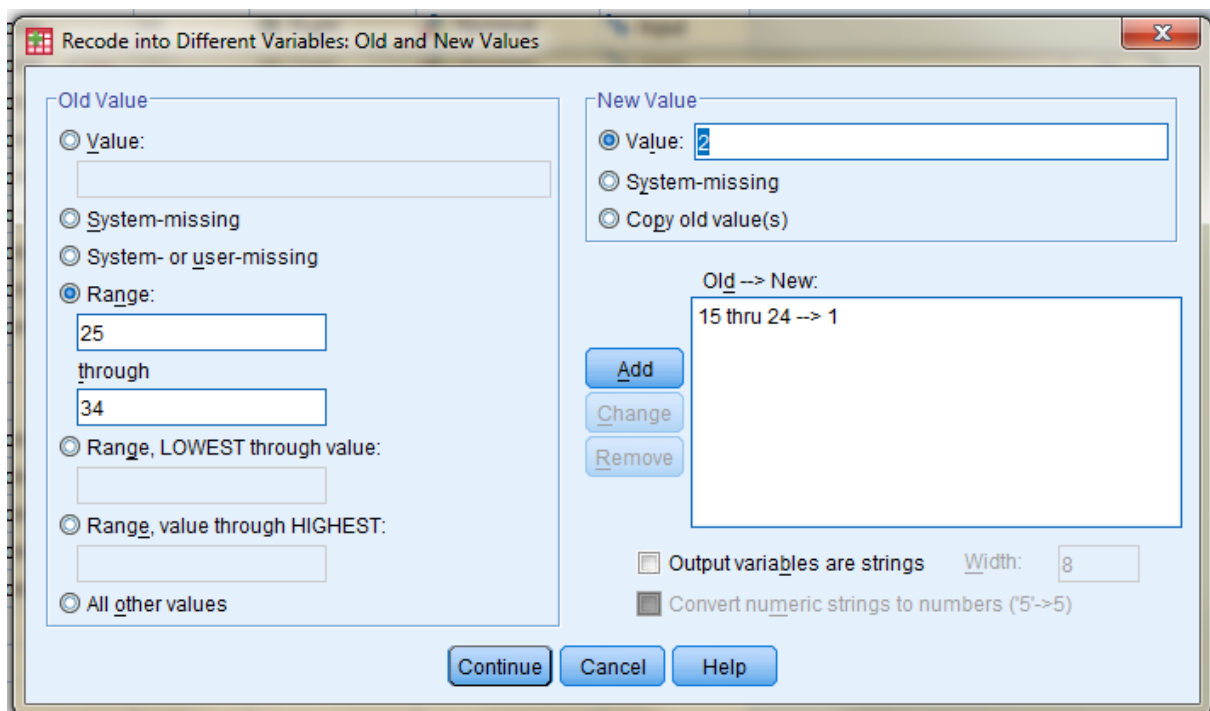
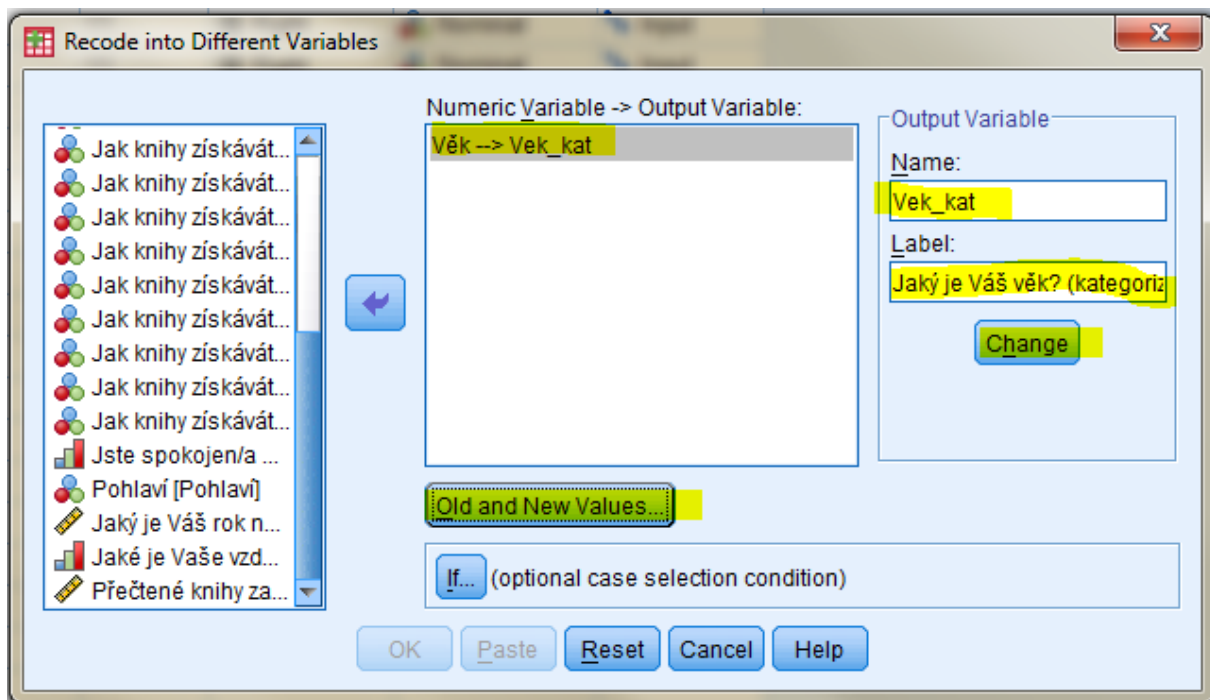
1. Zkopírujeme si sloupec s proměnnou „ekonomická činnost“ (9_ekcinnost) do nového sloupce, kde bude nová proměnná – pojmenujeme si ji třeba **9_ekcinnost2**.
2. Označíme si sloupec (dáváme pozor, abychom neměli označenou celou tabulku či na označení sloupce nezapomněli, v tom případě se nám překódují data z celého souboru).
3. Příkaz CTRL+H nám otevře dialogové okno, kde postupně budeme zadávat hodnoty k nahrazení. Vždy vybereme možnost „Nahradit vše“.
4. Nahrazujeme v pořadí:
 - a. Zaměstnaný/á na plný úvazek → Zaměstnán/a anebo OSVČ
 - b. Zaměstnaný/á na částečný úvazek → Zaměstnán/a anebo OSVČ
 - c. OSVČ → Zaměstnán/a anebo OSVČ
 - d. Nezaměstnaný/á → Ostatní
 - e. Na mateřské/rodičovské dovolené → Ostatní
 - f. Nepracující důchodce → Ostatní
 - g. V domácnosti → Ostatní
 - h. Studující → Ostatní
 - i. Jiné → Ostatní

Případně:

- a. 2 → 1
- b. 3 → 1
- c. 4 → 2
- d. 5 → 2
- e. 6 → 2
- f. atd... (hodnoty 1 zůstávají stejné)

Rekódování v SPSS

Rekódování v SPSS je snadná operace. V záložce „Transform“ zvolíme položku „Recode into different variables“. V tabulce poté naklikáme jméno a označení nové proměnné a dále staré a nové hodnoty proměnné:



Aritmetické operace

Díky aritmetickým operacím lze snadno vytvářet nové proměnné, ze starých hodnot pomocí zadaného vzorce. Typickým příkladem je vytváření sumačních indexů – sčítání prostých nebo vážených hodnot stejných variant různých znaků.

Příklad: Pokud bychom chtěli zjišťovat celkovou spokojenost s knihovnou, kterou jsme si operacionalizovali jako sumu různých měř spokojeností (spokojenosti s personálem, spokojenosti s výběrem fondu a spokojenosti s online službami), sumační index bude tvořit průměrná míra spokojenosti v těchto dílčích oblastech (získáme ji součtem hodnot jednotlivých proměnných, který vydělíme jejich počtem).

Aritmetické operace v Excelu

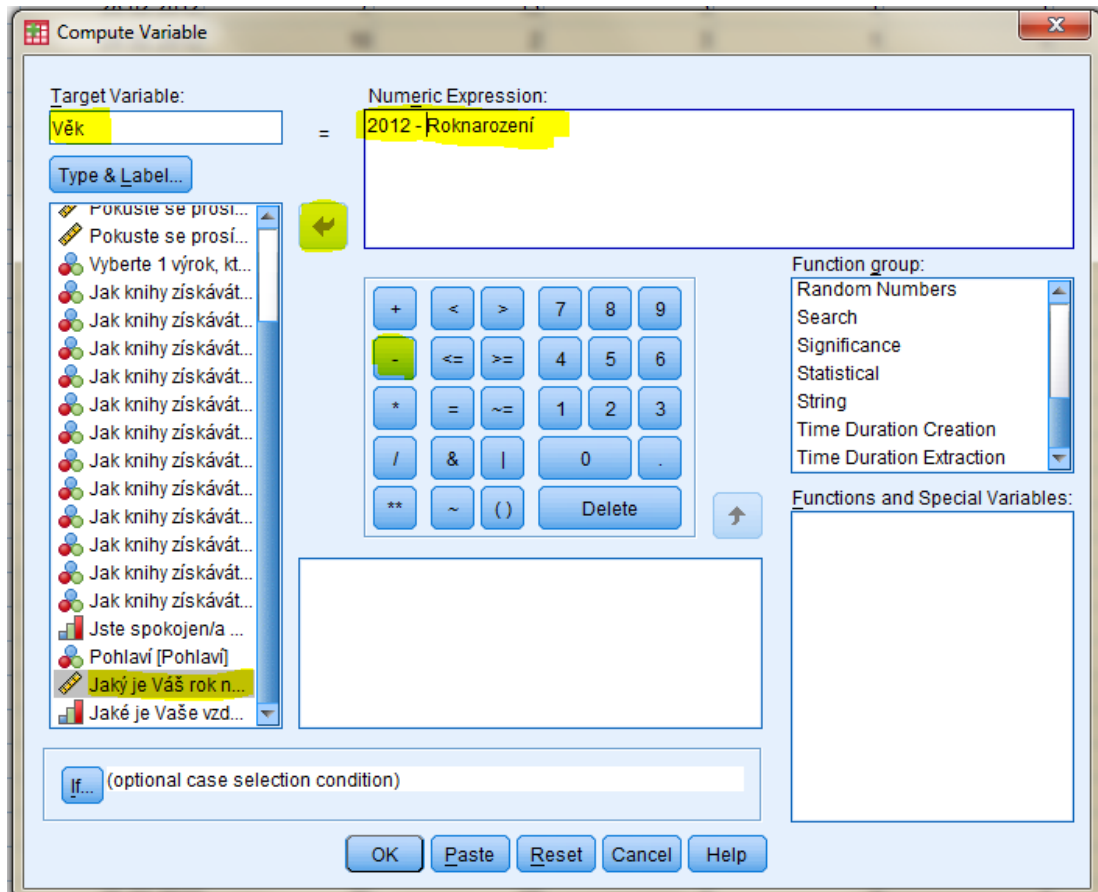
V Excelu zapisujeme vzorec přímo do tabulky.

Pokud máme např. v řádce 1 (respondent 1) proměnné spokojenost s personálem (sloupec A), spokojenost s výběrem fondu (sloupec B) a spokojenost s online službami (sloupec C), pak vzoreček pro celkovou spokojenost bude $= (A1+B1+C1)/3$.

Aritmetické operace v SPSS

V SPSS slouží k vytváření nových proměnných prostřednictvím aritmetických operací příkaz Compute (záložka Transform).

Níže je příklad vytvoření nové proměnné „věk“ z proměnné „rok narození“.



Sčítání výskytů

Sčítání výskytů je další ze způsobů jak vytvořit novou proměnnou. Někdy nás může zajímat, kolikrát respondenti například na různé otázky odpověděli „ano“.

Příklad: V našem datasetu máme otázku, o jaké oblasti se respondenti zajímají. Nás jako výzkumníky ale může také zajímat, o kolik oblastí se průměrně lidé zajímají.

Sčítání výskytů v Excelu

V Excelu slouží pro sčítání výskytů příkaz **COUNTIF**.

	A	B
1	Salesperson	Invoice
2	Buchanan	15,000
3	Buchanan	9,000
4	Suyama	8,000
5	Suyama	20,000
6	Buchanan	5,000
7	Dodsworth	22,500

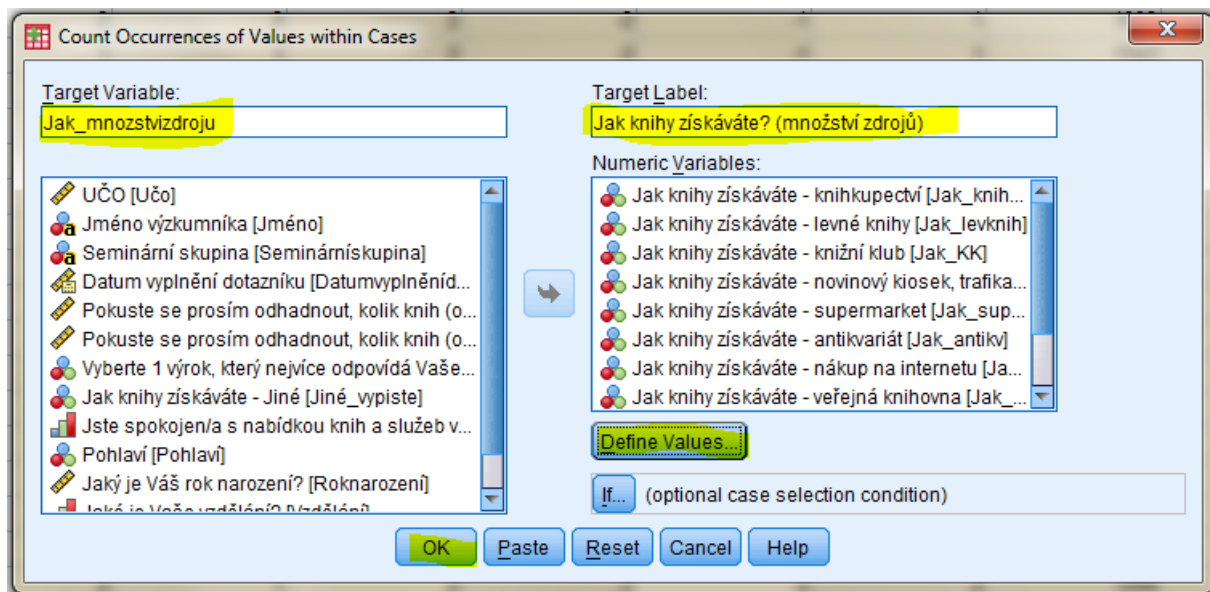
Formula	Description (Result)
=COUNTIF(A2:A7,"Buchanan")	Number of entries for Buchanan (3)
=COUNTIF(A2:A7,A4)	Number of entries for Suyama (2)
=COUNTIF(B2:B7,"< 20000")	Number of invoice values less than 20,000 (4)
=COUNTIF(B2:B7,">="&B5)	Number of invoice values greater than or equal to 20,000 (2)

Zdroj: Nápověda <http://office.microsoft.com>

Sčítání výskytů v SPSS

Příkaz COUNT v SPSS najdeme opět v záložce „Transform“.

Tabulka níže ukazuje výpočet počtu zdrojů, ze kterých respondenti získávají knihy (v dotazníku byla baterie otázek zaměřující se na různé zdroje knih, přičemž odpovědi byly vždy „ano“ nebo „ne“).



Seřazení položek

Vytváří novou proměnnou, kde řadí respondenty dle velikosti hodnoty proměnné. Například hledáte TOP 10 největších čtenářů či chcete najít 10 nejúspěšnějších studentů dle percentilu či 20 nejmladších respondentů. Seřazení položek je v tomto případě velmi elegantní řešení.

Příklad: V našem souboru chceme najít 10 nejstarších čtenářů. Budeme tedy pracovat se sloupečkem AS (rok narození). Do políčka u prvního respondenta (tedy v řádce 2) zapíšeme vzorec

$$=RANK(AS2;AS2:AS731;0)$$

Pokud máte dobře udělanou tabulku, Excel nám jej pravděpodobně přepíše na:

$$=RANK(Tabulka1[[#Tento řádek];[11_roknar]];[11_roknar];0)$$

Seřazení položek v Excelu

V Excelu na seřazení položek používáme příkaz **RANK**. RANK se zapisuje:

RANK(číslo;odkaz;pořadí)

- Číslo je číslo, jehož pořadí hledáte.
- Odkaz je matice nebo odkaz na seznam čísel. Nečíselné hodnoty jsou ignorovány.
- Pořadí je číslo určující, zda se budou hodnoty třídit vzestupně či sestupně.

Pokud je pořadí rovno 0 nebo není zadáno, určuje se v aplikaci Microsoft Excel pořadí čísla jako v sestupném seznamu. Pokud je pořadí jakákoliv nenulová hodnota, určuje se v aplikaci Microsoft Excel pořadí čísla jako ve vzestupném seznamu.

	A
1	Data
2	7
3	3,5
4	3,5
5	1
6	2

Vzorec

Popis (výsledek)

=RANK(A3;A2:A6;1) Pořadí čísla 3,5 ve výše uvedeném seznamu (3)

=RANK(A2;A2:A6;1) Pořadí čísla 7 ve výše uvedeném seznamu (5)

Zdroj: <http://office.microsoft.com>

Seřazení položek v SPSS

V Excelu na seřazení položek používá příkaz Rank Cases (opět v záložce Transform).

