

# Jak udělat z kontingenční tabulky zdrojová data


Newsletter Statistica ACADEMY



Téma: Kontingenční tabulky, makro, příprava dat  
Typ článku: Tipy a triky

*Pokud pracujete s kategorickými (kvalitativními) daty, jistě jste již prováděli analýzu kontingenčních tabulek. Dnes Vám ukážeme makro, které Vám může zjednodušit práci.*

	1 Časopis A	2 Časopis B	3 Časopis C
ZS	75	75	50
SS	40	70	40
VS	35	5	10



	1 Vzdělání	2 Časopis
1	ZS	C
2	ZS	C
3	ZS	C
4	SS	B
5	ZS	C
6	SS	B
7	SS	C
8	ZS	C
9	SS	C
10	SS	B
11	ZS	B

## Analýza kontingenčních tabulek

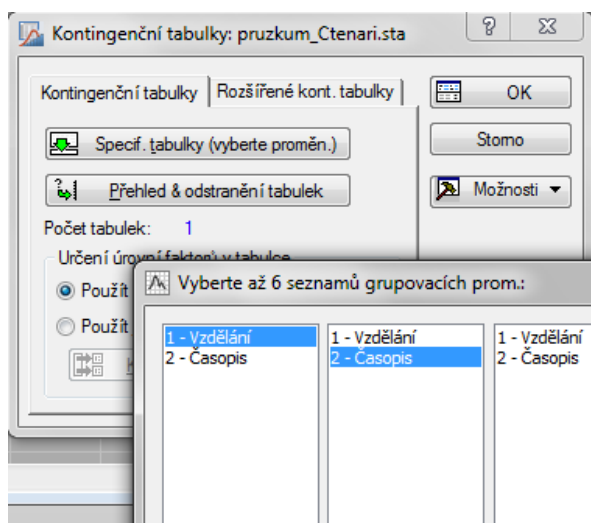
Typicky do analýz vstupují data ve formě, kdy každý řádek je měření (pozorování) dané jednotky (pacienta, výrobku,...), každý sloupec je pak jedna naměřená proměnná. Nejinak je tomu i pro analýzu kontingenčních tabulek. Tato analýza je vlastně porovnání vztahů dvou kategorických proměnných. Předpokládá se tedy, že do analýzy vstupují data, která mají dvě kategorické proměnné, přičemž každý řádek se váže k jednomu pozorování.

Uvedme příklad. Máme **data (pruzkum\_Ctenari.sta)**, kde se ptáme lidí, jaké mají dosažené vzdělání a který časopis čtou. O každém jedinci máme tedy informaci v jednom řádku:

Kontingenční tabulku bychom z těchto dat vytvořili následovně: **Statistiky ->**

	1 Vzdělání	2 Časopis
1	ZS	C
2	ZS	C
3	ZS	C
4	SS	B
5	ZS	C
6	SS	B
7	SS	C
8	ZS	C
9	SS	C
10	SS	B
11	ZS	B

**Základní statistiky -> Kontingenční tabulky.** Zde zvolíme proměnné následovně:



V této analýze program provede výpočet kontingenční tabulky a případně dalších charakteristik – typicky si například můžeme nechat spočítat hodnotu Chí-kvadrát testu, který testuje nulovou hypotézu, že pro jednotlivé hladiny jednoho faktoru je rozdělení vzhledem k druhému faktoru stejné. Jinak řečeno, že to, jaký časopis čtu, nezávisí na tom, jaké mám vzdělání,

Vzdělání	Časopis A	Časopis B	Časopis C	Řádk. součty
ZS	75	75	50	200
SS	40	70	40	150
VS	35	5	10	50
Celk.	150	150	100	400

## Motivace

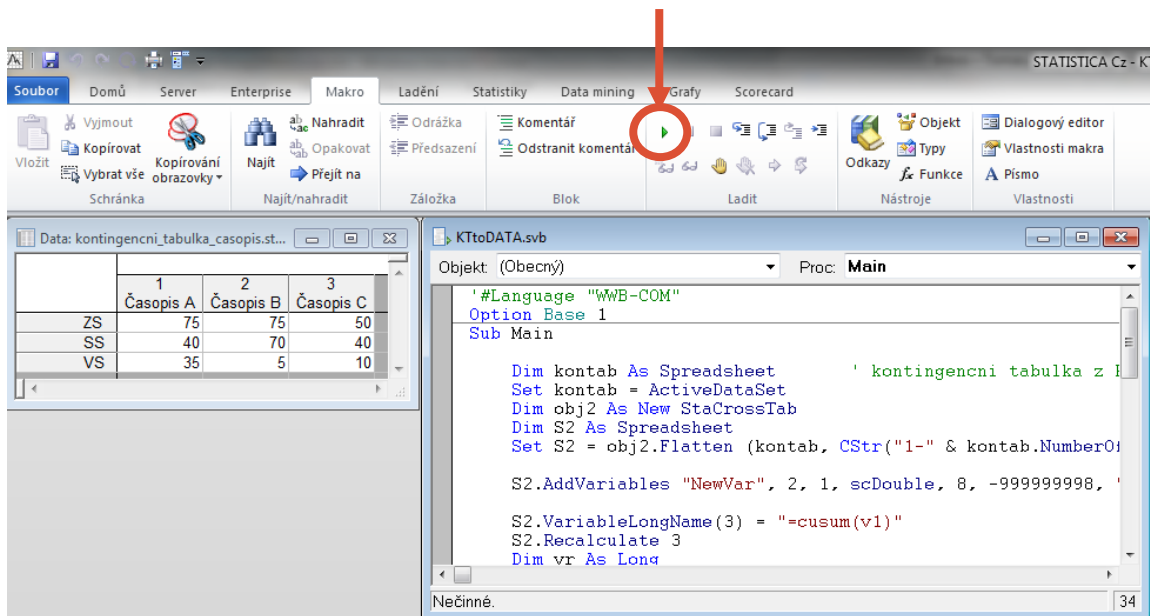
Nyní se dostáváme ke smyslu našeho článku a potřeby makra. Co když dostaneme data ke zpracování (například pro výpočet zmiňovaného testu) ve tvaru kontingenční tabulky a ne zdrojových dat? To se jistě může stát. Pro výpočet testů je kontingenční tabulka i zdrojová data naprosto rovnocenná, informace v obou je stejná a dají se na sebe vzájemně převést. Drobnou nepříjemností ale je, že Statistica vyžaduje vstup ve formě zdrojových dat. Samozřejmě si můžete převést data ručně, to je ale pracné a vzniká prostor pro chyby. Máme zde pro Vás alternativu a tou je naše makro.

## Makro

Makro, které jsme pro Vás připravili, dokáže převést kontingenční tabulku na zdrojová data a naleznete jej spolu s ukázkovými datovými soubory [zde](#). Je spustitelné pro jakýkoli typ licence, mělo by Vám tedy bez problémů fungovat. Kontingenční tabulku je potřeba mít v tvaru vpravo (také je mezi staženými daty – soubor *Ctenari\_konttab.sta*):

	1 Časopis A	2 Časopis B	3 Časopis C
ZS	75	75	50
SS	40	70	40
VS	35	5	10

Poté stačí otevřené makro spustit pomocí **F5** nebo tlačítka **Spustit makro v Makro -> Ladit**



Výsledkem je tato tabulka:

	1 Kategorie_sloupecK	2 Kategorie_r adkyKT
1	Časopis A	ZS
2	Časopis A	ZS
3	Časopis A	ZS
4	Časopis A	ZS
5	Časopis A	ZS
6	Časopis A	ZS
7	Časopis A	ZS
8	Časopis A	ZS
9	Časopis A	ZS
10	Časopis A	ZS

Názvy kategorií jsou z původní kontingenční tabulky. Pokud se Vám nelíbí názvy proměnných, tak si je jednoduše přejmenujte, podobně můžete přejmenovat i jednotlivé kategorie v **Textových popisících** proměnné.

*Poznámka:* Při spuštění makra musí být jako aktivní vstup (poslední aktivní tabulka) označena právě kontingenční tabulka.

*Poznámka:* Makro předpokládá, že v datech nejsou chybějící data – v tom případě se objeví chybová hláška.

## Závěr

Nyní máte k dispozici makro, které Vám rychle a jednoduše převede kontingenční tabulku do podoby zdrojových dat. Věříme, že se Vám takovéto makro bude hodit, stejně jako se čas od času hodí i nám.