**5.Expandování kategorizovaných dat**

Často máme k dispozici data jen v agregované podobě ve formě kontingenčních tabulek. Pokud máme k dispozici dostatek vysvětlujících proměnných, můžeme z kontingenční tabulky snadno rekonstruovat původní data. Používá se k tomu příkaz expand, který je komplementární s příkazem contract.

**5.1 Expandování dat**

Data máme k dispozici v podobě následující kontingenční tabulky. Víme přitom, že proměnná sex je kódována tak, že hodnota 1 odpovídá muži, hodnota 2 odpovídá ženě. Proměnná part je kódována tak, že 1 odpovídá popisku „má partnera“, hodnota 2 pak vyjadřuje „nemá partnera“. Levé horní pole tabulky tak lze označit indexem 11 (muž, má partnera), pravé spodní pole indexem 22 (žena, nemá partnera). Indexy jednotlivých polí jsou uvedeny na druhém řádku.

|  |  |  |
| --- | --- | --- |
|  | Má partnera  *1* | Nemá partnera  *2* |
| Muž  *1* | 2439  *11* | 537  *12* |
| Žena  *2* | 2979  *21* | 1124  *22* |

Nyní přepíšeme tuto tabulku do datové matice tak, abychom získali následující podobu:

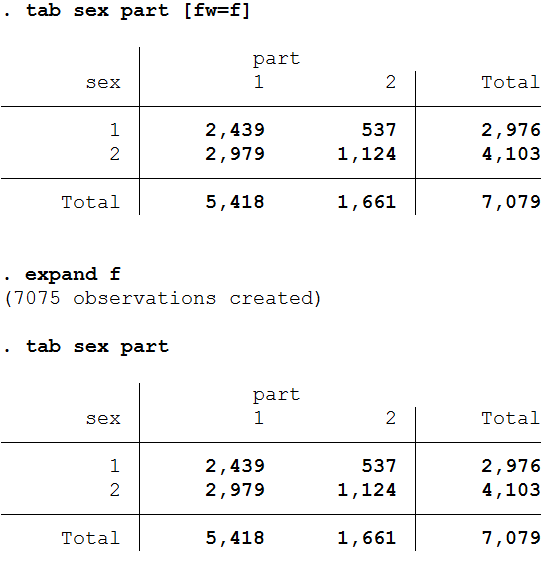
|  |  |  |
| --- | --- | --- |
| sex | part | f |
| 1 | 1 | 2439 |
| 1 | 2 | 537 |
| 2 | 1 | 2979 |
| 2 | 2 | 1124 |

Pokud použijeme proměnnou f jako váhy, můžeme nyní provádět veškeré výpočty – vypisovat kontingenční tabulky, počítat korelace, regrese apod.

tab sex part [fw=f]

Expandování dat provedeme pomocí příkazu EXPAND, kterému se jako parametr zadá proměnná f, která vyjadřuje počet pozorování odpovídajících daným vysvětlujícím proměnným (v našem případě sex a part).

expand f



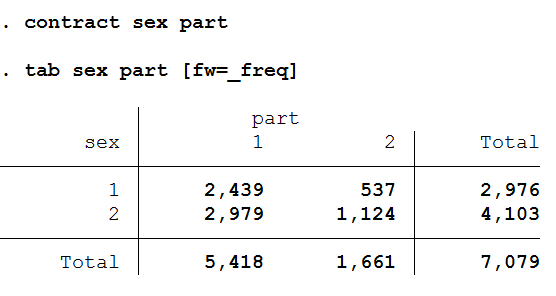
Obrázek 1 Ukázka expandování proměnných

**5.2 Stahování dat**

Pokud naopak potřebujeme data převést do zhuštěné podoby s váhami, použijeme k tomu příkazu CONTRACT. V tomto případě zadáme jako parametry všechny vysvětlující proměnné, které chceme pro stahování dat použít. V našem případě tedy

contract sex part

Frekvenční váhy jsou uloženy v proměnné \_freq. Stahování dat se může hodit například v případě velkého datového souboru. Stata dokáže pracovat se staženými daty mnohonásobně rychleji, jen je potřeba nezapomenout při všech výpočtech na zapnuté frekvenční váhy!



Obrázek 2 Ukázka stahování dat