

Váženie, štandardné chyby a limity spoľahlivosti odhadov

Váženie

Domácnosti, ktoré sú súčasťou výberového zisťovania pracovných síl reprezentujú celú populáciu SR žijúcu v súkromných domácnostiach. Aby bolo možné na základe ich odpovedí počítať odhady za celú populáciu, je potrebné každej odpovedajúcej jednotke priradiť výberovú váhu, ktorá určuje počet jednotiek, ktoré nie sú súčasťou výberového zisťovania a sú danou jednotkou reprezentované. Výpočet váh prebieha v troch krokoch:

1. *dizajnové váhy* - každá vybraná domácnosť dostane priradenú dizajnovú váhu, ktorá je rovná obrátenej hodnote pravdepodobnosti výberu. V prípade stratifikovaného klastrového náhodného výberu sa dizajnové váhy počítajú ako

$$d_{hi} = \frac{N_{hi}}{n_{hi}}$$

kde N_{hi} je celkový počet domácností v i -tom klastri h -tej straty a n_{hi} je príslušný počet vybraných domácností.

2. *prvotné váhy* - každá odpovedajúca domácnosť dostane pridelenú dizajnovú váhu upravenú o mieru neodpovede. Tento krok sa vykonáva na úrovni stratifikácie. Ak r_h označuje mieru odpovede v h -tej strate, tak potom prvotné váhy sa počítajú ako

$$d'_{hi} = \frac{d_{hi}}{r_h}$$

3. *kalibrované váhy* - prvotné váhy sa kalibrujú, aby zreprodukovali známe populačné úhrny. Týmto krokom sa do výsledných odhadov vnesie presnosť a konzistencia. V prípade výberového zisťovania pracovných síl sa ako kalibračné kritériá použili demografická štruktúra obyvateľov a počet domácností podľa krajov. Na kalibráciu sa použil nástroj Calif, viac informácií je možné nájsť na

<https://slovak.statistics.sk/wps/portal/ext/products/software.tools>

Štandardná chyba odhadu úhrnu - absolútneho odhadu

Keďže pre výberové zisťovanie pracovných síl bol výber spravodajských jednotiek realizovaný prostredníctvom stratifikovaného klastrového náhodného výberu, musela byť táto skutočnosť zohľadnená aj pri výpočte štandardných chýb. Štandardná chyba odhadu úhrnu je druhou odmocninou rozptylu tohto odhadu a je počítaná ako

$$SE(\hat{Y}) = \sqrt{\widehat{var}(\hat{Y})} = \sqrt{\sum_{h=1}^H \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} \left(w_{hij} y_{hij} - \frac{\sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} y_{hij}}{n_h} \right)^2 \left(1 - \frac{n_h}{N_h} \right)}$$

kde H je počet strát výberu, n_h počet klastrov vybraných z h -tej straty, m_{hi} počet domácností vybraných z n_h -teho klastra a N_h počet všetkých klastrov v h -tej strate.

Weighting, standard errors and reliability limits

Weighting

Households that are part of Labour Force Survey represent the whole population of individuals living in private households in Slovakia. In order to be able to calculate estimates for the whole population, it is necessary to assign some sampling weight to each reporting unit. It determines the number of units that are not part of the survey but are represented by a given one. The process of calculation of weights consists of three steps:

1. *design weights* – each selected unit gets a design weight, which is a reverse value of probability of selection. In case of stratified cluster random sample, design weights are calculated as

$$d_{hi} = \frac{N_{hi}}{n_{hi}}$$

where N_{hi} is the total number of households in i -th cluster of h -th stratum and n_{hi} is a respective number of selected households.

2. *initial weights* – each reporting unit gets a design weight adjusted for nonresponse. This process is carried out at the stratification level. If r_h denotes the response rate in h -th stratum then initial weights are calculated as

$$d'_{hi} = \frac{d_{hi}}{r_h}$$

3. *calibration weights* – initial weights are calibrated such that they can reproduce known population totals. This step brings precision and consistence into estimates. In case of Labour Force Survey, the calibration criteria were a demographic structure and a number of households at NUTS3 level. For the process of calibration Calif tool was used, for more information visit

<https://slovak.statistics.sk/wps/portal/ext/products/software.tools>

Standard error of total - of the estimate of absolute indicator

The selection of sampling units in Labour Force Survey was carried out by stratified cluster random sampling. This fact had to be taken into account in calculation of standard errors. The standard error of the estimate of the total is a square root of its variance and is calculated as

$$SE(\hat{Y}) = \sqrt{\widehat{var}(\hat{Y})} = \sqrt{\sum_{h=1}^H \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} \left(w_{hij} y_{hij} - \frac{\sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} y_{hij}}{n_h} \right)^2 \left(1 - \frac{n_h}{N_h} \right)}$$

where H denotes the number of strata, n_h is the number of clusters selected within the h -th stratum, m_{hi} is the number of units selected from the n_h -th cluster and N_h stands for the total number of clusters in the h -th stratum.

Pri výpočte štandardných chýb odhadov mier a pomerov (miera nezamestnanosti, miera zamestnanosti, atď.) bola využitá Taylorova aproximácia/linearizácia.

Limity spoľahlivosti

Pre výberové zisťovanie pracovných síl boli stanovené limity pre určenie spoľahlivosti odhadov absolútnych hodnôt.

Limity boli vypočítané na základe vzťahu

$$L = \frac{1}{\left(\frac{n \cdot CV^2}{Def f} + 1\right)} \cdot N$$

kde N predstavuje veľkosť základného súboru, n veľkosť výberového súboru, CV hodnotu koeficientu variácie, pre ktorú sa považuje odhad za menej spoľahlivý alebo veľmi nespoľahlivý a $Def f$ dizajn efekt, čiže podiel skutočnej variácie odhadu pre daný druh výberu (v našom prípade stratifikovaný klastrový náhodný výber) k variácii odhadu pre jednoduchý náhodný výber.

Odhady absolútnych ukazovateľov, ktoré sú menšie ako 10,5 tis. sú považované za menej spoľahlivé, keďže ich koeficient variácie je v priemere väčší ako 20 % a hodnoty menšie ako 6,5 tis. sú veľmi nespoľahlivé, keďže ich koeficient variácie je v priemere väčší ako 25 %.

Taylor linearization method was used for calculation of standard errors of rates or ratios (rate of unemployment, rate of employment etc.).

Reliability limits

For Labour Force Survey reliability limits for estimates of absolute values were set. They were calculated by

$$L = \frac{1}{\left(\frac{n \cdot CV^2}{Def f} + 1\right)} \cdot N$$

where N denotes the size of the population, n is the size of the sample, CV is a value of coefficient of variation for which the estimate is considered less reliable or very unreliable and $Def f$ represents the ratio of the actual variance of the estimate for a given type of sample (in our case stratified cluster random sample) to the variance of the estimate for a simple random sample.

Estimates of absolute indicators that are less than 10.5 thousand are considered less reliable as their coefficient of variation is on average greater than 20% and values less than 6.5 thousand are very unreliable, due to the fact that their coefficient of variation is greater than 25% on average.