

## 8. Technické poznámky

### A. Odhady intervalů spolehlivosti

S výběrovými šetřeními jsou spjaty tzv. výběrové a nevýběrové chyby. Nevýběrové chyby vznikají například administrativními odpady bytů z výběru, záměrným odmítnutím odpovědi nebo chybami při pořizování dotazníku. U těchto chyb nelze bez větších znalostí o základním souboru stanovit případné vychýlení odhadu. Naproti tomu výběrové chyby, vznikající vztažením charakteristik výběrového souboru na celý základní soubor, lze interpretovat pomocí tzv. intervalů spolehlivosti, což jsou intervaly zkonstruované kolem odhadu tak, že s určitou pravděpodobností skutečná hodnota odhadované charakteristiky leží právě v tomto intervalu. Nejčastěji se u odhadů konstruuje 95% interval spolehlivosti (vynásobením příslušného kvantilu normovaného normálního rozdělení a směrodatné odchylky odhadu), tedy interval, v němž s 95% pravděpodobností leží skutečná hodnota odhadované charakteristiky.

V teorii výběrových šetření se rozlišují dva nejčastější typy úhrnů, a to úhrny základní a dílčí. **Základní úhrny** jsou nějaké primární úhrny (zaměstnanost, nezaměstnanost ...) za určitý základní soubor (muži, ženy, osoby v produktivním věku, 20-24 letí muži, ...). **Dílčí úhrny** jsou potom nějaké podúhrny v základním úhrnu. Například členění OKEČ v souboru zaměstnaných osob jsou dílčí úhrny. Úhrny podle věkových skupin nejsou dílčí úhrny, ale základní úhrny v populaci 15-19 leté, v populaci 20-24 leté atd.

Intervaly spolehlivosti v **přílohových tabulkách I a II** jsou počítány pro velikost výběru v aktuálním čtvrtletí.

### B. Způsob použití přílohových tabulek

**Tab I Odhady 95% intervalu spolehlivosti odhadů základních úhrnů za populaci 15ti letých a starších** (v tisících)

Varianty: Ia pro základní úhrny ve 2. čtvrtletí 2004 celkem  
Ib pro základní úhrny ve 2. čtvrtletí 2004 za jedno pohlaví

Tabulka je určena ke zjištění přibližného 95% intervalu spolehlivosti základních úhrnů za základní soubor populace 15-ti leté a starší ve všech krajích a České republice. Chceme-li se například dozvědět, jaká je spolehlivost odhadu počtu zaměstnaných, jichž bylo ve 2. čtvrtletí 4 700,6 tis., najdeme si ve sloupci Česká republika řádek nejbližší číslu 4 700,6. Pro velikost odhadu 4 500 tis. je to 36,6 tisíce. Další sousední hodnota je uvedena pro odhad o velikosti 5 000 tis., což je 36,2 tis. Protože rozdíl čísel 4 700,6 a 4 500 tvoří dvě pětiny rozdílu 5 000 a 4 500, odečteme od čísla 36,6 ještě dvě pětiny rozdílu 36,6 a 36,2 a získáme nakonec hodnotu 36,4. Tedy 95% interval spolehlivosti pro odhad počtu zaměstnaných ve 2. čtvrtletí 2004 je cca 4 700,6 +/- 36,4 tis., tzn., že s 95% pravděpodobností nebyl skutečný počet zaměstnaných v ČR nižší než 4 664,2 tis. a zároveň větší než 4 737,0 tisíc osob.

**Tab II Odhady 95% intervalu spolehlivosti odhadů dílčích úhrnů za populaci 15ti letých a starších na úrovni ČR**

Varianty: IIa pro dílčí úhrny ve 2. čtvrtletí 2004 celkem  
IIb pro dílčí úhrny ve 2. čtvrtletí 2004 za jedno pohlaví

Tabulka je určena ke zjištění přibližného 95% intervalu spolehlivosti **dílčích úhrnů** za základní soubor populace 15ti leté a starší a to **pouze na úrovni České republiky**. Chceme-li například zjistit spolehlivost odhadu počtu podnikatelů (viz tabulka 101), jichž bylo 761,9 tis. z celkového počtu 4700,6 tis. zaměstnaných (tj. 16,2 %), vyhledáme v tabulce hodnotu v řádku přibližně odpovídajícímu hodnotě 4700,6 a ve sloupci přibližně odpovídajícímu hodnotě 16,2. Můžeme provést i následující korekci jednoduchou lineární interpolací:

|               | 16                                                       | 16,2                                                            | 18                                                       |
|---------------|----------------------------------------------------------|-----------------------------------------------------------------|----------------------------------------------------------|
| 4500          | 0,43                                                     |                                                                 | 0,45                                                     |
| <b>4700,6</b> | cca 0,42<br>=0,43-(4700,6-4500)/(5000-4500)* (0,43-0,41) | <b>cca 0,42</b><br><b>=0,42+ (16,2-16)/(18-16)* (0,44-0,42)</b> | cca 0,44<br>=0,45-(4700,6-4500)/(5000-4500)* (0,45-0,43) |
| 5000          | 0,41                                                     |                                                                 | 0,43                                                     |

Tzn., že s 95% pravděpodobností nebylo podnikatelů méně než 16,2 %-0,42 % (741,8 tis.) a více než 16,2 %+0,42 % (781,2 tis.).

Tabulka II se může použít i pro základní úhrny ve věkových skupinách a pohlaví na úrovni České republiky, když základní úhrn nahradíme velikostí základního souboru a dílčí úhrn příslušným odhadem.

V této kapitole jsme chtěli čtenáři podat přibližný návod, jak zhruba určit chybu, která vzniká vztažením charakteristik výběrového souboru na soubor základní. Tato chyba závisí na třech proměnných (v případě dílčích úhrnů na čtyřech) a to na velikosti výběrového souboru a velikosti odhadu a nepříliš na velikosti základního souboru. Podat objektivní přehled o chybách všech odhadů by znamenalo vytvořit velmi rozsáhlou tabulkovou přílohu, ve které by se běžný čtenář ekonomických publikací jen stěží orientoval. Z tohoto důvodu jsou všechny použité metody značně aproximující, ale pro představu přesnosti odhadů plně postačující.

**Všeobecně v celé publikaci platí, že úhrny menší než 1 tis. se považují za údaje s velmi nízkou spolehlivostí. V praxi to znamená, že jejich směrodatná odchylka je příliš vysoká nebo dokonce větší než samotný odhad.**

### **C. Použité zdroje a číselníky**

- CZ-NUTS** Územní členění je prováděno v souladu s Klasifikací územních statistických jednotek (CZ-NUTS) platnou od 1. ledna 2000.
- Obyvatelstvo** Údaje o počtu a struktuře obyvatelstva jsou převzaty z demografické statistiky podle definitivního stavu k 31. 12. 2001 a 31. 12. 2002 s promítnutím příslušného časového podílu změn v přirozeném a mechanickém pohybu obyvatelstva v průběhu roku 2002 na středy jednotlivých čtvrtletí roku 2002 resp. 2003.
- ISCED 97** Údaje o úrovni (stupni) a o skupinách oborů, resp. oborech vzdělání podle mezinárodní klasifikace ISCED 97, UNESCO, listopad 1997.
- OKEČ** Údaje o odvětví činnosti jsou členěny podle kategorií průběžně aktualizované Odvětvové klasifikace ekonomických činností (OKEČ). Klasifikace je kompatibilní s mezinárodní klasifikací NACE, rev. 1.1.
- KZAM** Zaměstnání je klasifikováno podle KZAM vydané ČSÚ v roce 1995 (1. svazek), 1. upravené vydání. Klasifikace je kompatibilní s mezinárodní klasifikací ILO ISCO-88.
- CZ-ICSE** Postavení v zaměstnání je klasifikováno podle skupin CZ-ICSE z roku 1998, které odpovídají jednotlivým kategoriím mezinárodní klasifikace ISCE-93.

### **D. Charakteristika číselníků**

- CZ-NUTS :** Klasifikace NUTS (La Nomenclature des Unités Territoriales Statistiques) byla zavedena Statistickým úřadem Evropských společenství ve spolupráci s ostatními orgány EU pro potřeby klasifikování jednotné unifikované struktury územních jednotek. Od roku 1988 je používána v legislativě EU, zejména pro úkoly spojené s čerpáním ze Strukturálních fondů EU.

Klasifikace obsahuje 6 úrovní NUTS (NUTS 0, NUTS 1, NUTS 2, NUTS 3, NUTS 4 a NUTS 5), které představují velikostní skupiny území. Vymezení jednotlivých úrovní NUTS je charakterizováno počtem obyvatel a rozlohou. Klasifikace CZ-NUTS popisuje územní uspořádání České republiky v územních jednotkách, které vyhovují kritériím Evropské unie a pro statistické účely jsou schváleny Eurostatem. Pro účely této publikace jsou užity stupně: pro Českou republiku (NUTS 1) a kraje (NUTS 3).

**ISCED 97:** Publikované údaje o úrovni (dříve stupni) a o skupinách oborů, resp. oborech vzdělání jsou v souladu s mezinárodní klasifikací ISCED 97 (International Standard Classification of Education) vydanou UNESCO v listopadu 1997. Od 1. ledna 2003 je ve VŠPS plně aplikována klasifikace oborů vzdělání na tři místa, kterou ČSÚ převzal od ÚIV MŠ ČR. Vzhledem k náročnosti zařazení respondentů do jednotlivých oborů dosaženého nebo navštěvovaného vzdělání, vyžádá si přechod na podrobnější klasifikaci delší časové období.

Dosažená (tj. dokončená) úroveň vzdělání podle ISCED 97 (Level of education) má následující členění:

- 0 preprimární vzdělávání (KKOV - stupeň A) - vzdělávací programy předškolní výchovy. Jsou zde zahrnuty i osoby bez jakéhokoliv vzdělání.
- 1 primární vzdělávání (KKOV - stupeň B) - zahrnuje první stupeň základního vzdělávání, tj. ukončený 5. ročník základní školy.
- 2 nižší sekundární vzdělávání (KKOV - stupeň C, D, E) - obsahuje především druhý stupeň základního vzdělávání ukončený obvykle 9. třídou.
- 3 sekundární vzdělávání - (KKOV - stupeň H, J, K, L, M (částečně)) - zahrnuje odborné i všeobecné středoškolské vzdělání na středních odborných a všeobecných školách a v odborných učilištích ukončené většinou maturitní nebo závěrečnou zkouškou. Do tohoto stupně jsou zařazeni i absolventi nižších praktických škol, protože zejména v případě osob, které jsou již delší dobu mimo vzdělávací systém, měla tato škola charakter středního vzdělání.
- 4 postsekundární vzdělávání - (KKOV - stupeň M (částečně) a N (částečně)) - pomaturitní kvalifikační a pomaturitní specializační a inovační studium nepatřící do systému terciárního vzdělávání.
- 5 první stupeň terciárního vzdělávání - (KKOV - stupeň N (částečně), R, T) - zahrnuje bakalářské a magisterské studijní programy, které nevedou přímo k udělení vědecké kvalifikace.
- 6 druhý stupeň terciárního vzdělávání - (KKOV - stupeň V) - pokrývá terciární vzdělávací programy, které vedou k udělení diplomu vědeckého stupně.

V rámci jednotlivých úrovní se uplatňuje kritérium typu "následného vzdělání nebo určení". V publikaci je toto hledisko použito u stupně 3 (např. absolventi skupiny 3A mají možnost následně pokračovat v klasickém bakalářském nebo magisterském studiu, naopak programy skupiny 3C přímo směřují absolventy na trh práce).

**OKEČ:** Odvětvová klasifikace ekonomických činností (OKEČ) s ohledem na informační závazky České republiky vůči Evropské unii, OSN, Mezinárodnímu měnovému fondu a dalším mezinárodním organizacím, vychází plně ze statistické klasifikace ekonomických činností EU (NACE). V současné době byla podle aktualizací změn a úprav evropského standardu NACE rev. 1.1 provedena aktualizace klasifikace OKEČ (3. vydání platné od 1. ledna 2003).

**KZAM:** Předmětem klasifikace je zaměstnání, tj. konkrétní činnost, kterou pracovník vykonává (i když není jeho povoláním) a která je zdrojem jeho hlavních pracovních příjmů. Klasifikace zaměstnání - KZAM byla vypracována na podkladě mezinárodního

standardu ISCO-88 (International Standard Classification of Occupations) přijatého 14. mezinárodní konferencí statistiků práce v listopadu 1987.

**CZ-ICSE:** Klasifikace postavení v zaměstnání CZ-ICSE byla vypracována na bázi revidované Mezinárodní klasifikace postavení v zaměstnání (International Classification of Status in Employment) - ICSE-93, schválené 15. mezinárodní konferencí statistiků práce v lednu 1993. Klasifikace ICSE-93 má závazné třídění jenom na úrovni 1. místa, podrobnější třídění je doporučeno. Klasifikace CZ-ICSE má vytvořené podrobnější třídění závazné do hloubky čtyř míst. Pojednává jenom o osobách ekonomicky aktivních.

### **E. Ostatní poznámky**

Číselné údaje v absolutním vyjádření jsou uváděny v tisících. Rozdíly mezi údajem za celek a součtem dílčích údajů v přílohových tabulkách vyplývají ze skutečnosti, že úhrn není součtem zaokrouhlených dílčích údajů, ale zaokrouhleným celkovým údajem. Absolutní i relativní údaje ve všech textových a přílohových tabulkách a v textu jsou vypočítány z nezaokrouhlených čísel.

V tabulkách jsou standardní statistickou symbolikou odlišeny marginální případy ve významu :

- daný jev se ve výběrovém souboru nevyskytl,
- x zápis hodnoty není možný z logických důvodů,
- údaj není k dispozici nebo je nespolehlivý.

Údaje "nezjištěno" v jednotlivých tabulkách zahrnují odmítnuté odpovědi, odpověď "nevím" resp. jiný případ nezjištěné odpovědi respondenta. Pokud otázka umožňuje více možností odpovědi, jsou údaje tříděny zásadně podle hlavní varianty odpovědi.

Při práci s tabulkami je třeba mít stále na zřeteli, že údaje jsou zjišťovány výběrovými metodami a proto jejich přesnost klesá se snižující se velikostí vzorku.

### **F. Dostupnost publikace**

Publikace je vydána ve dvou jazykových verzích, české a anglické následujícími formami:

- a) **Tisk :**
  - česká verze kód publikace : 1531 - 05
  - anglická verze kód publikace : 1531 - 05
- b) **Elektronická podoba :**
  - česká verze Excel (\*.xls), Adobe Acrobat (\*.pdf), JPEG (\*.jpg)  
153104q2.exe ctimne.txt
  - anglická verze 153104q2.exe readme.txt
- c) **Internet :**
  - česká i anglická verze Excel (\*.xls), Adobe Acrobat (\*.pdf), JPEG (\*.jpg)  
<http://www.czso.cz>