

Zpracování výsledků

Po ukončení sběru dat probíhalo jejich zpracování, a to v následujících základních krocích:

- validace sebraných údajů probíhala ve dvou krocích. V okamžiku pořízení dat probíhaly základní logické kontroly dat, které sledovaly validitu, konzistenci a úplnost sbíraných údajů pro konkrétní domácnost. V dalším kroku byla s využitím statistického software SAS prověřována všechna odlehlá pozorování,
- dopočet položkové non-response byl prováděn s využitím speciálně vytvořených statistických modelů či tzv. hot-deck metodou (doplnění chybějícího údaje od jiné náhodně vybrané domácnosti s podobnými charakteristikami). Dopočet položkové non-response se týkal částečně vyplněných dotazníků, jejichž počet byl v řádu jednotek procent,
- k dopočtu na základní soubor (přepočítání dat na úhrn populace) bylo využito váhové schéma domácností z roku 2021. Jako základ pro sestavení tohoto váhového schématu byla aplikována metoda kalibrace vah, která minimalizuje rozdíl mezi dostupnými a z výběru přepočítanými hodnotami zvolených charakteristik. Byl využit odhad počtu obvykle obydlených bytů, který primárně vychází z aktualizace údaje ze SLDB 2011 promítnutím každoročních přírůstků (resp. odhadů úbytků) počtu (dokončených) bytů ze statistiky bytové výstavby. Tento aktualizovaný odhad byl však ještě navýšen o expertní odhad počtu bytů, v nichž někdo obvykle bydlí přesto, že nejsou statisticky podchycené jako obydlené. Expertní odhad se opírá zejména o rozdíl mezi počtem osob z bilance obyvatelstva v demografické statistice a počtem osob dle SLDB 2011, přičemž je v něm navíc zohledněno, že část osob administrativně přihlášených na obecních úřadech, tj. formálně s bydlištěm mimo byty, ve skutečnosti v bytech bydlí.
Protože šetření podléhaly pouze osoby žijící v soukromých domácnostech a bytech, byly na úrovni ČR od všech údajů z demografie odečteny odhady počtu osob (např. mladistvých nebo důchodců) žijících v ústavních domácnostech (nápravných zařízeních pro mladistvé, ústavech sociální péče, domovech důchodců apod.) podle údajů statistiky sociálního zabezpečení, resp. ve věznicích podle údajů Ministerstva spravedlnosti. Dále byly odečteny osoby bydlící v zařízeních (např. ubytovny a svobodárny, obydlí mimo bytový fond) a odhad za osoby žijící mimo byty a zařízení odvozený z dat SLDB 2011.

Přesnost výsledků

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.

Při interpretaci a analýze výsledků šetření ENERGO 2021 je třeba mít na paměti, že vznikly zpracováním dat získaných z výběrového šetření. To znamená, že všechny publikované údaje představují odhady, které jsou zatíženy výběrovou chybou.

S ohledem na to, že uživatelům této publikace chceme poskytnout maximální množství informací získaných v rámci šetření ENERGO 2021, rozhodli jsme se publikovat i statistické údaje, které jsou zatíženy větší výběrovou chybou. Kvalitu publikovaných výsledků lze v tabulkách identifikovat pomocí následujících symbolů:

- (7698) Číselný údaj v závorce = ze statistického hlediska méně spolehlivý údaj
(výsledek byl spočítán z méně než 50 pozorování).
- .
- Tečka = ze statistického hlediska nespolehlivý údaj.
- 0
- Nula = statisticky nepodchyceno
(v rámci domácností, které byly zahrnuty do výběrového souboru, se daný jev nevyskytoval).
- x
- Údaj v daném třídění nepřichází do úvahy.
-
- Pomlčka = údaj nebyl z důvodu logiky tabulky vypočítáván.
- n.a
- Not available = údaj není k dispozici.

