

Kvantitativní výzkum – Studijní opora k tématu sekundární analýza dat

V tomto tématu se budeme věnovat sekundární analýze dat. Jedná se o přístup, kdy výzkumníci nebo i lidé v praxi využívají data, která sami nevytvořili (viz SL 6, 7). Sekundární analýza má řadu přínosů (SL 7). Jednak je hospodárná a umožňuje nám dostat data a tím i informace, které bychom vlastním výzkumem jen obtížně získávali. Naopak rizika spojená se sekundární analýzou dat (SL 8) jsou především spojena s nevhodným formátem dat nebo s tím, že datům špatně rozumíme, protože o nich nemáme dostatek informací.

Je velmi důležité rozumět rozdílu mezi tzv. individualizovanými a agregovanými daty (SL 10, 11, 12). Individualizovaná data jsou data, kdy máme informace o každém případě na řádku v datovém souboru (nebo jinde, jinak = např. dotazníky od každého respondenta). Agregovaná data jsou data, která vznikají slučováním individualizovaných dat do kategorií. V některých případech dochází ke vzniku agregovaných dat i složitějšími výpočty a statistickými operacemi. Proto je velmi důležité porozumět tomu, jak agregovaná data vznikla. Důležité je také reflektovat, že pro vlastní výzkumnou práci je zpravidla mnohem lepší mít data individualizovaná než data agregovaná. Je to z toho důvodu, že možnosti další práce s agregovanými daty jsou velice omezené (viz SL 21).

Pro sekundární analýzu dat můžeme získat data různého typu a z různých zdrojů (viz SL 16-20).

S analýzou individualizovaných dat jsou spojené specifické otázky a rizika (SL 23). Obecně řečeno je třeba data prozkoumat, očistit o prapodivné hodnoty, dobře se seznámit s povahou dat, data vhodně transformovat a pokud možno si všechny transformace udělat znovu sám (vycházet z originálních dat).

Velmi obezřetní musíme být při analýze a interpretaci agregovaných dat (SL 24-25). Základní problém je, že můžeme z nepozornosti přehlédnout, že jsou data nekompletní nebo nespolehlivá nebo že špatně porozumíme tomu, co data vlastně znamenají. Např. je někdy velký rozdíl, jestli uvádíme údaje z prosince nebo průměr za celý rok (prodej opalovacích krémů). Důležité je být při interpretaci opatrný i z jiných důvodů. Např. často se setkáváme s tím, že data jsou záměrně manipulována (novináři, politici, neprofesionální výzkumníci) nebo zkreslena z důvodu špatného měření nebo proto že se jedná o intimní nebo jinak citlivé otázky (potraty, domácí násilí). Specifickou chybou je usuzování na menší jednotky na základě větších jednotek (Ecological Fallacy). Tedy k prokázání vztahu zpravidla nestačí vědět agregované informace a ty si dodatečně propojit, protože nemůžeme prokázat tu vazbu na individuální úrovni, a tedy že právě rozdíl v bohatství způsobil odlišný počet motocyklů v obou městech (viz příklad SL 25).

Kontrolní otázky:

- Vysvětlíte, jaký je rozdíl mezi primární a sekundární analýzou dat (SL 6, 7)
- Jaké jsou přínosy, a naopak rizika využívání sekundární analýzy dat? (SL 8, 9)
- Vysvětlíte rozdíl mezi individualizovanými a agregovanými daty a jaké má tento rozdíl důsledky pro sekundární analýzu dat (SL 10, 11, 12, 21).
- Co a z jakých zdrojů můžeme získat pro sekundární analýzu dat? (SL 16-20)
- Jaké jsou kroky potřebné při sekundární analýze individualizovaných dat? (SL 23)
- Jaká rizika hrozí (na co si musíme dát pozor) při analýze a interpretaci agregovaných dat. Uveďte alespoň tři (SL 24, 25).