

KONSTRUKCE PRIMÁRNÍCH DAT – DOTAZNÍK, ROZHOVOR

Sampling – výběr dat

Celý set relevantních dat, ze kterých se provádí výběr, můžeme nazývat populací. Samplingem rozumím výběr podsouboru z určité definované populace, která je příliš široká na to, aby mohlo být zkoumáno každé individuum. Metodika výběru by měla být zvolena tak, že údaje o celkové populaci (souboru) mohou být získány na základě omezeného podsouboru – vzorku s určitou danou mírou přesnosti/chyby. Platí samozřejmě, že čím větší je vzorek, tím je přesnější (schopnější reprezentovat kompletní soubor). V případě tzv. vyčerpávajícího výzkumu se vzorek shoduje s populací, v případě tzv. reprezentativního výzkumu jsou výsledky získané na základě zkoumání vzorku uplatnitelné na celou populaci (struktura vzorku musí tedy co nejlépe popisovat strukturu populace).

Je tedy nutné:

1. definovat populaci – pokud zpracováváme volební průzkum, vyloučíme např. osoby pod 18 let bez práva volit. Je také nutné definovat populaci v čase a místě;
2. určit jednotky výběru (sampling units) – např. ulice, domácnost;
3. definovat rámec výběru – jde o určení seznamu entit, ze kterých bude vzorek konstruován – např. telefonní seznam, seznam firem, apod. V humánní geografii se nepoužívá prostorové vyjádření rámce tak často, jako ve FG výzkumu. Pokud využíváme jako rámec mapu volíme mezi bodovým vzorkem (např. průsečíky koordinát), liniovým vzorkem (řez) či oblastí. Často se vyskytuje problém nekompletních seznamů;
4. zvolit metodiku výběru;
5. stanovit velikost vzorku;
6. určit způsob sběru dat.

Případ špatného samplingu – průzkum časopisu *Literary Digest* před prezidentskými volbami v USA v roce 1936. Kandidáti Landon vs. Roosevelt. Předpovídán byl výsledek 57 ku 43 pro Landona, skutečný výsledek 38 ku 62 pro Roosevelta. Příčina - vzorek 2,4 mil. lidí byl brán z telefonních seznamů a seznamů klubů. Šlo o nereprezentativní výběr spíše bohatších lidí (vlastníci telefonů, členové klubů).

Nepravděpodobnostní (kvótní) sampling

Lze vyčlenit zcela náhodný nepravděpodobnostní sampling (tzv. convenience sampling) – např. interviewujeme prvních n zákazníků, kteří vstoupí do obchodu (většinou předpokládáme, že jsme zvolili typické místo a nashromážděný vzorek nám představuje typickou skladbu zákazníků, jde o výběr typických individualit). Typickým příkladem mohou být ankety televizních reportérů mezi náhodně vybranými lidmi na ulici.

Jiným typem je účelový nepravděpodobnostní výběr – vybíráme např. určité typy osob; spadá sem i výběr dle kvóty (může být i pravděpodobnostní) – oslovíme určitý počet žen, určitý počet mužů, důchodců či studentů. Většinou se používá v průzkumech politických preferencí, veřejného mínění.

V případě kvótního výběru tak postupujeme u populací, o jejichž struktuře máme dostatek informací. Specifickými dalšími typy nepravděpodobnostního samplingu je tzv. snowball sampling (respondent dává kontakt na dalšího respondenta) či experní účelový sampling (výzkumník sám určí vhodné respondenty).

Pravděpodobní (náhodný) sampling

Mezi pravděpodobnostní metody patří sampling využívající nějaké formy náhodného výběru, která zaručí, že každá jednotka v populaci má přibližně stejnou šanci na to být vybrána.

Nezákladnější metodou je využití náhodných čísel (generovaných počítačem či např. pomocí tabulky náhodných čísel).

Při stratifikovaném pravděpodobnostním samplingu rozdělíme na základě našich předběžných znalostí výchozí populaci. Populace se rozdělí do podskupin, v rámci kterých se provádí oddělený výběr.

Při systematickém výběru vybíráme každou n -tou jednotku populace, poté co je náhodně vybrána první jednotka (je nutné dbát, abychom se nedostali do konfliktu s vnitřní periodicitou či pravidelností souboru).

Geografové běžně využívají tzv. shlukový, resp. areálový sampling, kdy uplatňují pravděpodobnostní sampling pouze ve vybraných geografických oblastech.

Výběr bude vždy spojen s určitou chybou. V závislosti na použité metodě můžeme chybu odhadnout už před zahájením výzkumu. Stejně tak lze stanovit předem velikost vzorku.

Díky rafinovaným metodám výběru vzorku se podařilo v moderních výzkumech významným způsobem zmenšit nutnou velikost vzorku. Gallupův ústav pro výzkum veřejného mínění ve Velké Británii používá např. vzorku pouhých 1000 osob k zachycení předvolebních nálad, a chyba u jednotlivých stran při výzkumech před volbami od roku 1945 od roku nepřesáhla 1,4 %. Například v roce 1980 se zpracovával vzorek 1000 osob z počtu 43 miliónů oprávněných voličů.

Typy dat

1. **Nominální data (výčtová)** – každé individuum je zařazováno do jedné z kategorií (nominální data zařazují 3 mil. lidí do Birminghamu a 9 mil. do Londýna). Populaci můžeme rozdělit např. na mužskou a ženskou kategorii (2), na muslimy, židy a křesťany (3). kategorie mohou vyčerpávat danou populaci, mohou se navzájem překrývat či vylučovat.
2. **Pořadová data (ordinální)** – individua či kategorie, do kterých náležejí, jsou řazeny podle nějakého kritéria (ordinální měření nám řekne, že Londýn je větší než Birmingham). Jde tedy o data, která mají mezi sebou nějaký vztah vyjádřitelný jako „větší než“, „preferovanější“, apod.). Hodnoty se dají rozlišovat pořadím, ale ne velikostí. S těmito typy dat pracují např. výzkumy preferencí, přístupů, zařazování sebe sama do sociálních skupin (rozhodně ano, spíše ano, nevím, apod.).
Příklad: předložíme 10 výroků, se kterými může respondent souhlasit, spíše souhlasit, spíše nesouhlasit a nesouhlasit. Za každý výrok může tedy respondent dostat od 4 do 1 bodu. Dle získaných lze respondenty seřadit do pořadí. Toto pořadí je ordinálními daty, vyjadřuje opravdu jen pořadí a nic víc, neznamená např. že mezi respondenty na 2 a 4 místě a respondenty na 5 a 7 je stejný rozdíl v názorech.
Dovoleny jsou takové statistické operace, které nezasahují do pořadí.
3. **Intervalová data** – kvalitativně vyjádřená vzdálenost mezi dvěma měřeními na předurčené škále (intervalové měření dává informaci, že Londýn má o 6 mil. obyvatel více než Birmingham). Přitom naměřená vzdálenost je u intervalových dat nezávislá na zvolených jednotkách. Pomocí intervalových dat měříme např. příjmy, inteligenční kvocient, trestnou činnost, apod. Často charakterizují relativní pozici v prostoru, čase nebo velikost - zeměpisná šířka, nadmořská výška.
4. **Poměrová data (ratio)** – dovolují relativní kvantitativní ohodnocení na určité škále (Londýn je třikrát větší než Birmingham). Tyto proměnné mají přirozenou nulu (váha, délka, plocha), jsou také nezávislé na jednotce měření.

Dalším členěním může být např. dělení dat na lokalizační (vztahující se k pouhé lokalizaci jevu) a atributová (popisující samotné jevy či procesy).

Dotazníkové šetření, rozhovor

Před spuštěním dotazníkového průzkumu je nutné si zodpovědět několik otázek: jaký je náš výzkumný problém?; je opravdu nutné použít tento typ získávání dat? Např. studujeme-li proces gentrifikace vnitřního města je nutné si upřesnit zda studujeme, kde se gentrifikace odehrává, kdo je nositelem či jaké jsou motivy gentrifikace. Pouze pro některé z uvedených otázek je ospravedlnitelné využití dotazníku. Dotazník by neměl být postaven jako jediný zdroj dat o výzkumném problému a je vhodné jej doplnit i daty z sekundárních, resp. jiných statistických zdrojů. Zkoumáme-li gentrifikaci je vhodné pracovat i se sčítáním či s informacemi realitních kanceláří, což jsou nepoměrně lacinější zdroje.

Úspěšnost dotazníku/rozhovoru je založena na několika faktorech:

- a) na adekvátně vybraném vzorku (viz. předchozí přednáška);
- b) na správné skladbě a sestavení dotazníku, na provedení rozhovoru či vyplnění dotazníku;
- c) na analýzách a interpretaci výsledků.

Jakékoli šetření je získáváním přímých informací od lidí o jejich názorech, pocitech, plánech, motivacích – tyto kvalitativní údaje musíme vztáhnout k určitým skupinám, proto současně musíme od lidí získávat i data o jejich osobním, vzdělanostním či finančním pozadí. Nabízí se nám dvě možnosti – dotazník nebo interview.

Rozhodnutí zda použít dotazník či rozhovor (interview) je determinováno povahou výzkumu a praktickými možnostmi výzkumníka. Pro příklad: pokud potřebujeme získat informace od respondentů z celé země je téměř nemožné použít metodu rozhovoru.

Závažné důvody proti rozhovoru jsou: zkreslení vyvolané tazatelem a vysoká nákladnost tohoto postupu (mnohonásobně vyšší).

Závažné důvody proti dotazníku jsou: nesmírně nízká návratnost (platí především pro poštou distribuované dotazníky), což může výrazně zkreslit výsledky, které bychom dostali od vzorku skutečně reprezentujícího zkoumanou populaci (řešení: dotazníky mohou být distribuovány a sbírány osobně, mizí tím ale finanční efektivita).

Základem všech dotazníků je otázka. Kvalita dotazu pak ovlivňuje hodnotu odpovědi.

Dle předmětu zjišťování:

- fakta (objektivní údaje);
- otázky na postoje, názory, mínění, důvody, apod. (subjektivní údaje).

Podle typu proměnné:

- otázky na kvantitativní údaje (numerické, číselné proměnné) - např. počet členů domácnosti, výše nájemného, počet odpracovaných hodin apod.;
- otázky na kvalitativní údaje (kategoriální, verbální proměnné) - např. úroveň dosaženého vzdělání, druh vlastnictví bytu, spokojenost s bydlením apod.

Dle funkce, kterou otázky plní v dotazníku:

- otázky meritorní - zjišťují údaje o předmětu šetření;
- otázky analytické (identifikační, cenzové) - třídící a identifikační parametry, dle kterých se provádí analýza (u respondentů: pohlaví, věk, vzdělání, zaměstnání, rodinný stav, místo bydliště atd. dle záměru zjišťování u zpravodajských jednotek: IČO, počet pracovníků, druh vlastnictví, kraj apod.); filtrační otázky – používají se v případech, kdy na následující otázku by měla odpovídat jen část respondentů, která např. vlastní určitou proměnnou nebo znak. Tyto otázky pomohou odlišit ty respondenty, kteří jsou o určitém problému informováni lépe než ostatní, a tak jejich odpovědi mohou být pokládány za závažnější. Negativní odpověď na otázku (filtr) vede k pokynu, aby respondent přeskočil jednu nebo sérii otázek pro něj irelevantních;
- otázky kontrolní - cílem těchto otázek je prověřit pravdivost odpovědí na základní otázky dotazníku. S jejich pomocí se oceňuje kvalita získané informace. Zpravidla se kontrolují otázky, které těsně souvisí se základním cílem a předmětem výzkumu, protože v těchto případech je stupeň hodnověrnosti (validity) nesmírně důležitý. Obvykle se základní a kontrolní otázky uspořádají v dotazníku tak, aby dotazovaný nemohl při jejich vyplňování najít přímou souvislost.

Rozhovor	Dotazník
Velice pracná a nákladná technika sběru	Postihuje velký počet jedinců při relativně nízkých nákladech
Časově velice náročný	Děje se v relativně krátkém čase
Vyžaduje velký počet vyškolených tazatelů v terénu	Nízké požadavky na počet a zaškolení spolupracovníků
Výzkumna prostorově rozptýleném vzorku je nákladná	Náklady šetření na rozptýleném vzorku jsou relativně nízké
Anonymita výzkumu je pro respondenty nálo přesvědčivá	Anonymita je relativně přesvědčivá
Rozdíly mezi tazateli a rozdíly v jejich chování mohou vyvolat „interviewer bias“	„Interviewer bias“ je prakticky vyloučen
Méně nároky na iniciativu respondenta, pro respondenta je důležitější vynechat odpověď na některé otázky	Kladě vysoké nároky na ochotu dotazovaného, je snadné přeskočit otázky nebo neodpovědět vůbec
Je méně jisté, že dotazovaná obsahují ta, která byla vybrána do vzorku	U dotazníku je možné, že otázky byly zodpovězeny jiným děnem rodiny, nebo, a to nejčastěji, celým rodinným týmem
Podpora úspěšně dokončených dotazníků je podstatně vyšší než návratost dotazníků	Návratost je velice nízká

pozn. interviewer bias znamená, že tazatel podává určité otázky emocionálně zabarvené, čímž může ovlivnit odpovědi

Uzavřené otázky

V případě uzavřených otázek jsou respondenti nabídnuty alternativy odpovědí na danou otázku. Kategorie použité pro uzavřené otázky musí představovat soubor vyčerpávající všechny možné alternativy odpovědí, které se vzájemně vylučují.

studna	<input type="checkbox"/>	
veřejný vodovod	<input type="checkbox"/>	
obojí	<input type="checkbox"/>	

podnikatel (fyzická osoba)	<input type="checkbox"/>	akciová společnost	<input type="checkbox"/>
státní podnik	<input type="checkbox"/>	společnost s ručením omezeným	<input type="checkbox"/>
družstvo	<input type="checkbox"/>	jiné	<input type="checkbox"/>

Výhodou uzavřené je relativní snadnost odpovědi pro respondenta, snadnost a jednoznačnost zpracování odpovědí pro výzkumníka. Na druhou stranu mohou uzavřené otázky příliš schematizovat problém, resp. podsouvat respondentovi možné odpovědi.

Uzavřené otázky se dále dělí na:

- otázky dichotomické (two-choice) s možností výběru ze dvou odpovědí;
- otázky alternativní (polytomické, množné, multiple choice) s možností výběru jedné z několika alternativ;
- otázky výčtové (checklist) s možností označení více odpovědí;
- otázky pořadové (s uvedením pořadí vybraných alternativ, ranking format)
- škály pořadové (rating scale).

Otevřené otázky

Jde o formu otázky, kdy respondentovi nejsou předloženy žádné varianty odpovědi a tazatel nebo sám respondent zapíše jejich úplné nebo zkrácené znění. Tento typ otázek umožňuje spontánní sebevyjádření respondenta, nicméně je pracnější, jak pro respondenta, tak pro výzkumníka (složitá analýza a interpretace).

Pořadí otázek

Začátek dotazníku v mnohém ovlivňuje věrohodnost odpovědi. Proto bychom se měli v první fázi tázání především věnovat vytvoření motivace respondenta k účasti v šetření, vysvětlení jeho cílů a důležitosti, ujistění respondenta o zachování anonymnosti a ochrany jím sdělených údajů. Je dobré členit dotazník do tematických oddílů a pečlivě volit načasování citlivých otázek.

Monotónní sérii otázek je nejlépe umístit uprostřed dotazníku, otázky osobního rázu nebo otázky, kde je nebezpečí, že budou neupřímně zodpovězeny, se dávají spíše na konec; otázky jejichž zodpovězení se zdá snadno ovlivnitelné, se dávají na začátek, dříve než vlastní téma dotazníku začne působit.

Při sestavování dotazníku je nezbytné si uvědomit:

- je otázka nezbytná?
- bude otázka srozumitelná všem členům vzorku?
- neptá se otázka na dvě věci najednou? (two-barrels question – dvouhlavňová otázka)
- není naše otázka sugestivní? (za sugestivní lze považovat takový výrok, jenž navádí respondenta k jiné odpovědi, než kterou by uvedl na jinak formulovanou otázku stejného obsahu).

U dotazníkové metody platí, že výsledky lze velice snadno kvantifikovat (na rozdíl od jiných metod používaných zejména v sociologii). Kvantifikace však může být sporná, protože byť shodné odpovědi v dotaznících nemusí vždy vyjadřovat shodné názory a postoje respondentů (*jestliže dva lidé v dotazníku souhlasí s tvrzením „...Velká Británie by měla omezit svou vojenskou přítomnost v zahraničí.“ může jeden podporovat návrat k tradičnímu britskému izolacionismu, zatímco druhý být fanatickým přívržencem odzbrojení*).

Související literatura

- DISMAN, M., (2000): Jak se vyrábí sociologická znalost. Praha, Karolinum, UK.
- NACHMIAS, F.N., NACHMIAS, D., (1996): Research Methods in the Social Sciences. London, Arnold, s.223 – 278.
- ROBINSON, G.M., (1998): Methods & Techniques in Human Geography. Chichester, Wiley.