

Dotazníky

František Vávra

2007

Obecně

Dotazníkem zjišťujeme určitá fakta nebo jevy, staticky nebo dynamicky. Až na výjimky se jedná o výběrová šetření. Výsledky na dotazovacím kmeni zobecňujeme na celou populaci (ne obecnou, ale nějak vymezenou). U úplných šetření se často zjišťuje stav některé vlastnosti (znalosti, ...) – jedná se spíše o testy. Někdy nelze jednoznačně rozlišit mezi testem a dotazníkem. Dotazníkem můžeme také vyvolávat některé akce.

⇒ Před konstrukcí dotazníku musí být jasné:

1. Vymezení, zdroje a metody
 - a) Vymezení populace a její vnitřní struktury, na kterou bude zobecňováno
 - b) Vymezení toho, co má být zjišťováno nebo čeho má být docíleno
 - i) Co má být zjišťováno a jak to lze měřit (ohodnotit, označit, ...)
 - c) Vymezení metodiky, jak má být zjišťováno
 - d) Vymezení metodiky, jak bude vyhodnocováno (nelze konstruovat dotazník bez neustálé konfrontace se způsobem vyhodnocování)
 - e) Disponibilní prostředky na realizaci dotazníkové akce
 - i) Rozsah kmene (výběru)
 - ii) Složení kmene
2. Způsob předkládání dotazníku probantům (dotazník vhozený do schránky, dotazník předkládá tazatel, dotazník jako součást denního tisku, dotazník jako součást www prezentace, dotazník jako součást úředního a pseudouředního jednání (pokud přijdu k svému dodavateli plynu, předloží mi dotazník na téma, jak jsem spokojen s vůní jím dodávaného topného a nehořlavého plynu), spotřebitelský dotazník spojený se soutěží nebo losováním o ceny, dotazník na vývěsce se zásobníkem a odkládací urnou, ...).
3. Zda a jak bude určen kmen probantů. V případě určeného kmene jsou problémy s výběrem kmene dotazovaných = probantů. Jedná se o problém proporcionálního zastoupení všech (nedisjunktních) skupin. Obecně to vede na kombinatorickou úlohu (někdy zvanou úlohu o latinských čtvercích – ještě obecněji o kombinatorických konfiguracích). Obvykle se používá pro předvolení průzkumy (50% žen, 50% mužů, 20% s ukončeným VŠ vzděláním, ..., 20% mladších 20 let, ..., ...).
4. Anonymní, poloanonymní, neanonymní, skrytě neanonymní, dotazování.
5. Případná standardizace.
6. Zda bude dotazníkové šetření nebo cílení doplněno další metodou zjišťování nebo ovlivňování.
7. Odhad návratnosti, výpadky.
8. Verifikace a valorizace, testovaný a testovací kmen.

Co chceme zjistit, čeho chceme docílit, na co se ptáme

Dotazníkem lze zjišťovat daná fakta nebo stupeň (měření) dané vlastnosti nebo jím lze vyvolávat předpokládané akce.

Zjišťování daných fakt – otázky a odpovědi

Nejčastěji otázkou na výběr předem definovaných variant (výčet, ve kterém není vztah obecně lepší – obecně horší, varianty není možné mezi sebou přirozeně srovnávat).

Množinově – variant je možné vybrat více.

Příklad: V domácnosti máte:

- a) sporák
- b) ledničku
- c) myčku nádobí
- d) umývač nohou
- e) jiný kuchyňský spotřebič
- f) žádný kuchyňský spotřebič

Poznámky: Varianty by měly obsahovat alespoň jednu „utíkácí“ (únikovou) variantu (v uvedeném příkladu e), popř. f). Varianty by měly být vybrány tak, aby „utíkácí“ varianty byly málo časté, spíše jako východisko z nouze. Varianty by měly být souměřitelné (neměl by se mezi nimi vyskytovat typ logicky nepatřící do souboru - v příkladu varianty a), d). Výběr může být nadán testovací strukturou (zde varianta f) může být označena jen samostatně anebo v kombinaci s d), také v kombinaci s a) nebo bez ní). Pozor však na případy, kdy daný výběr (otázka) je testem jedince na jeho pozornost a schopnost rozlišovat nehomogenity a nesoulady.

Jednoznačně – správná odpověď obsahuje pouze jednu variantu.

Příklad: V následujících volbách budu volit:

- a) ČSSD
- b) KDU-ČSL
- c) Stranu mírného pokroku v mezích zákona
- d) ODS
- e) KSČM
- f) Stranu zelených
- g) Jinou existující stranu
- h) Nejsem dosud rozhodnut
- i) Volit nepůjdu

Poznámky: Výběr by měl umožňovat i „úplný“ nevýběr“ zde varianty h) a i). Varianty by měly mít po dvou prázdný průnik (v jedné volbě lze volit právě jednu stranu – neplatí pro české volby do místních zastupitelstev). V případě, že to nelze, mělo by být upozornění, jak se v takové situaci zachovat (např.: v případě, že byste musel do místního zastupitelstva volit kandidáty jen z jedné strany, kterou byste si vybral – to samozřejmě může vyvolat nezbytnost další varianty, zde např. gg) *Určitě nebudu volit kandidáty jen z jedné strany*). Opět, varianty

by měly splňovat podmínku homogenity (zde to nesplňuje varianta c). Kandidáti za tuto stranu kandidovali pouze ve volbách roku 1911¹).

Důležité: *Varianty u obou typů by neměly být obecně poměřovatelné nějakou stupnicí (lepší, horší, dražší, levnější, příjemnější, nepříjemnější, ...), individuálně však je to přijatelné. Jde o to, aby výběr nebyl ovlivněn takovou stupnicí. Tj. do hodnocení by neměla být vnášena předem systematická (populační) chyba.*

Chybná (v uvedeném smyslu) je např. následující otázka:

Kdybyste si měli koupit los za 500 Kč, pak si ho koupíte v loterii, kde je hlavní výhra ve výši:

- a) 5 milionů korun
- b) 5 tisíc korun
- c) 500 korun
- d) nesázím.

Jedině excentrik by si vybral variantu d), vzhledem k tomu, že se jedná o hypotetické sázení, většina si vybere variantu a). Tato poznámka ale neplatí pro případ, kdy je otázkou zjišťováno zastoupení excentriků (a poťouchlů) a zásadových odmítačů sázení v dané populaci. Pozor na otázky po hypotetickém chování – jsou obecně nevěrohodné.

U všech uvedených typů má smysl hodnocení jen pomocí četností výskytů (ne průměry a srovnatelné statistiky, ne pořadové statistiky, ne korelace) a neparametrických testů (χ^2 test dobré shody, výběrová entropie, výběrová vazební informace, rankové korelace s náhodným pořadím u shod...).

Ve stupnici

Zde připadají v úvahu otázky s odpověďmi, které lze rovnat do nějaké preferenční stupnice, kde lze rozlišit vztah preferuji více nebo méně, nikoliv však měřit změnu.

Příklad: Jste pro umístění kolotoče na své zahrádce

- a) Ano – bez výhrad.
- b) Ano – pokud dostanu zapláceno.
- c) Nevím – přizpůsobím se okolnostem.
- d) Ne – pokud budou vyhrávat z tlampače.
- e) Zásadně ne.

Podstata takových dotazů je v tom, že odpovědi mají ordinální charakter (lze je porovnávat), chybí jim však kardinální míra (nelze přirozeně a korektně odpovídat na dotazy typu „o kolik“). Poznámka: tohoto charakteru jsou i klasické známky ve škole (změna známky na vysvědčení z dvojky na trojku je něco zcela jiného než změna ze čtyřky na pětku, i když aritmetický rozdíl je shodný). Pozor, stupnice by měla vždy obsahovat odpověď typu neutrálního postoje (zde c). Krajní a neutrální varianty bývají poměrně informativní.

¹ Bude to bezmála sto let, kdy se v hostinci U zlatého litru v Balbínově ulici číslo 1 scházela bohémská společnost, jejíž duší byl Jaroslav Hašek. Právě v debatách o politice vznikla idea založit stranu, která by cele vyhovovala jejich pohledu na svět. A tak vznikla **Strana mírného pokroku v mezích zákona**. Patřili do ní anarchističtí básníci Rosenzweig-Moir a Luis Křikava, malíř Kubín, později Josef Lada, intelektuálové František Drobílek, dr. Novák, zvaný Kytíčka, dr. Grünberger, dr. Foerster, ing. Kún a řada dalších i náhodných členů. Strana se intenzivně věnovala osvětové činnosti a všichni členové šířili její ideje po celém Rakousku-Uhersku. Vrcholem politické činnosti strany byla kandidatura ve volbách do Říšské rady za Vinohradský okres v roce 1911. Celá věc se rozhodla v hostinci U Zvěřinů na Vinohradech za Národním domem, kam se společnost časem přestěhovala. V době nejintenzivnější politické činnosti strany provozovali Zvěřinovi svůj hostinec v Kravíně na Korunní třídě. <http://www.vinohrady.cz/clanek.asp?20021125181243>.

Vlivem nepřítomnosti kardinálního charakteru pak jsou nekorektní vyhodnocení průměry, směrodatnými odchylkami, , ale jsou korektní hodnocení pomocí pořadových statistik (medián, kvartily, percentily, kvartilové rozpětí, ...).

Měřitelné veličiny

Zde připadají v úvahu otázky a odpovědi, které lze ohodnotit jak do pořadí, tak i korektně měřit změnu, např.:

1. Váš věk v letech?

2. Průměrný měsíční příjem vaší domácnosti?

Odpovědi pak lze vyhodnocovat všemi dostupnými statistickými technikami, které se týkají náhodných proměnných (diskrétních i spojitých, pozor na kvantování). Pozor však na očekávání obecného publika. Např. u průměrů jest obecným statisticky nedokonalé vzdělaným lidem očekávána „vlastnost centra = průměr je ve středu“. To u výběrů z nesymetrických rozdělení neplatí. Striktně má vlastnost centra (polovina je na tom hůře a polovina lépe) pouze medián. Je půvabné, jak se projevují četní komentátoři každé čtvrtletí (když ČSÚ vyhlašuje výsledky pravidelných mzdových šetření), když hledají nápravná opatření ke skutečnosti, že 2/3 zaměstnanců mají „podprůměrný“ plat (dokonce lze dokumentovat i návrhy, že by s tím vláda měla něco dělat, a ono je to přitom zcela správně a zákonité). Je to dáno tím, že rozdělení příjmů jsou vždy výrazně nesymetrická. To, co intuitivně očekává veřejnost, je medián (polovina je pod a polovina je nad).

Měřitelné veličiny – kvantované „intervalové“

Zde se jedná o měřitelné veličiny, je však zjišťována jen příslušnost do nějakého intervalu.

Příklad: Váš měsíční plat je

- a) nižší než 10 tis. Kč
- b) od 10 tis. Kč do 15 tis. Kč
-
- k) vyšší než 350 tis. Kč

Zásady:

Intervaly by měly být voleny tak, aby uvnitř nich platil předpoklad rovnoměrného rozdělení (lépe, nedal se prokázat opak). Pro vyhodnocování se pak používá buď speciálních intervalových statistik nebo pořadové statistiky (takový intervalový výčet je obvykle ordinální nebo se využívá faktu, že se jedná o směs rovnoměrných rozdělení pravděpodobnosti).

Třídící a zjišťovací otázky

Někdy se dotazníkem zjišťuje i složení „dotazované“ populace. Pak se do dotazníků přidávají identifikační otázky. Ty mohou být charakteru třídícího, např.

- a) Váš věk
- b) Jste žena nebo muž.
- c) Máte vzdělání: základní, střední, vysokoškolské, vyšší než vysokoškolské, VÚML

nebo i skrytě identifikačního charakteru, např.:

- a) Váš rok narození
- b) Váš měsíc narození
- c) Jste žena nebo muž.
- d) Ulice, ve které bydlíte.

Z těchto údajů lze až na poměrně malou skupinu výjimek identifikovat probanta (např.) z evidence obyvatel. Také je možné použít veřejných zdrojů, jako je obchodní a živnostenský rejstřík.

Kontrola a zajišťování

Testování tazatele

U dotazníkových akcí s tazateli nabývá na významu testování jevu, zda data pocházejí od probantů nebo zda je tazatel generoval (vymyslel) sám (obvykle bývají tazatelé placeni za počet vyplněných dotazníků). Testování tazatelů má dva základní směry.

První – statistický: Využívají se testy shody dvou nebo (častěji) více výběrů. Je účelnější testovat shody ne v parametrech polohy (parametry polohy průměr, medián, ..., jsou snáze a věrohodněji manipulovatelné při generování), ale v parametrech (ukazatelích) variability (kvartilové rozpětí, směrodatná odchylka, skupinové četnosti, k jejich manipulaci je třeba poměrně náročný aparát).

Druhý – zajištěním: V testu musí být otázka nebo zjišťovaná skutečnost, kterou lze zjistit jedině od tázaného (nebo na dotazovacím místě) a kterou lze verifikovat z jeho vlastních přímých odpovědí nebo z veřejně dostupných údajů (např. u domů se zjišťuje nejen číslo orientační (bývá na domě), ale i číslo popisné – bývá v občanském průkazu nebo v obecné místní stavební evidenci. Samozřejmě je velmi těžkou úlohou formulovat takovou otázku, aby nebyla rozlišitelná od ostatních.

Je třeba si uvědomit, že žádný takový test není dokonalý a 100% efektivní. Intuitivně bývají přijímána taková zajištění (jako výborná), která mají efektivitu alespoň 60%.

Testování tázaného a vnitřní zajišťování

Toto není možné testovat statisticky. Takový test musí být individuální. Testuje se tím věrohodnost výpovědi v dotazníku, obvykle nikoliv totožnost probanta s tázaným. Většinou nelze rozlišit neschopnost tázaného od záměrného zkreslování skutečnosti. K takovému testování se využívají (vnitřní) závislosti mezi výsledky dotazů. Např.

1. Váš věk v letech
2. Rok narození

(Rok tázání – rok narození) může být nejvýše o jednotku větší než uvedený věk. Samozřejmě obě takové otázky nebudou v dotazníku u sebe.

Jako teoretického základu pro takové zajišťování je využívána teorie samoopravných kódů.

Příklad 7-místného Hammingova kódu:

z1	z2	i1	z3	i2	i3	i4
1	2	3	4	5	6	7

se zajišťovacími rovnicemi: $z1 = i1 + i2 + i4$ $z2 = i1 + i3 + i4$ $z3 = i2 + i3 + i4$
 (jedná se o binární kód, tedy výpočty modulo 2, logicky neekvivalence).

První rovnice = vazba zajišťuje následující pozice:

z1	z2	i1	z3	i2	i3	i4
-----------	-----------	-----------	-----------	-----------	-----------	-----------

Druhá rovnice = vazba zajišťuje následující pozice:

z1	z2	i1	z3	i2	i3	i4
-----------	-----------	-----------	-----------	-----------	-----------	-----------

Třetí rovnice = vazba zajišťuje následující pozice:

z1	z2	i1	z3	i2	i3	i4
-----------	-----------	-----------	-----------	-----------	-----------	-----------

Je vidět, že zajištění má ortogonální strukturu. Při testování můžeme z výsledků testů (tří) zjistit jednoznačně, na které pozici došlo k chybě (špatnému vyplnění, špatné odpovědi).

Zajišťovací rovnice lze také psát:

$$0 = z1 + i1 + i2 + i4 \quad 0 = z2 + i1 + i3 + i4 \quad 0 = z3 + i2 + i3 + i4$$

Jednotlivé pozice kódu i1, i2, i3, i4 je možno (v tomto konkrétním případě) chápat jako binární odpovědi (ano, ne) na informační otázky a pozice z1, z2, z3 jako odpovědi na zajišťovací otázky, které jsou voleny tak, aby splňovaly uvedené zajišťovací rovnice.

Samozřejmě důsledné využití této teorie povede k téměř nepřekonatelným problémům. Podstatný je ten, jak formulovat příslušné otázky na pozicích z1, z2, z3 tak, aby měly pro danou populaci srozumitelný obsah a aby z jejich formulace nebylo vidět, že jde o zajištění. To je téměř vždy neřešitelná úloha. Proto se z daného aparátu využívá jen jeho část, a to ortogonální struktura zajištění. Pozice (otázka) z1 se volí tak, aby mezi odpověďmi otázek (z1, i1, i2, i4) existoval nějaký testovatelný vztah. z2 se volí tak, aby mezi odpověďmi na otázky (z2, i1, i3, i4) byl nějaký (jiný) testovatelný vztah. Konečně z3 se volí tak, aby mezi odpověďmi na otázky (z3, i2, i3, i4) byl také nějaký testovatelný vztah. Samozřejmě výběr obecných vztahů může vést (ve skutečnosti téměř vždy povede) na ztrátu samoopravnosti (zvláště v případech, kdy nebude možné najít zobrazení odpovědí do nějakého konečného tělesa). Avšak zůstane detekční schopnost a lokalizační schopnost umístění chyby (= špatné odpovědi, včetně lokalizace nesprávností na testovacích pozicích). To je pro daný účel víc než dostatečné. Příslušné „zajišťovací“ vztahy pak mohou mít poměrně rozmanitý charakter. Důležitou otázkou je také, jak reagovat na detekci chyby a její lokalizaci. Obvyklou reakcí je vyřazení celého dotazníku z dalšího zpracování (je podezřelý). Pokud dotazníkové šetření

není tak vydatné, aby bylo možno vyřadit celý dotazník, vyřadí se pouze odpověď na lokalizované pozici.

Je však třeba poznamenat, že generování zajišťovacích vztahů je náročnou intelektuální prací s potřebou delší praxe a zkušeností (je třeba mít nápad, výběr nejde algoritmizovat). Dost často se k takovému zajišťování používá skupinový brainstorming. Další používanou technikou je vyhledávání souvislostí mezi informačními otázkami, opět je dobrým prostředkem brainstorming. Např. současná odpověď v otázce: Přednášky byly zajímavé hodnotou SOUHLASÍM a v otázce: Přednášky byly srozumitelné hodnotou PŘEVÁŽNĚ NESOUHLASÍM poněkud znehodnocuje celý dotazník. To (odhalená vnitřní souvislost) pak minimalizuje přidávané otázky. Omezíme-li se na prostou detekční schopnost bez možnosti bližší lokalizace chyby (tj. je-li detekována chyba, dotazník je celý vyřazen), může být zajišťovací struktura poměrně triviální (je to ekvivalent testu na paritu).

Triviální poznámka: Vnitřní zajištění je také dobrým prostředkem na odhalování „generovaných“ dotazníků.

Formulace a náplň otázek

Zde budou pojednány jen informační (a jiné) otázky. Zajišťovací a třídící viz výše. Otázky by měly být odvozeny přímo z projektové dokumentace o tom, co má být zjišťováno. Ideální je případ, kdy je možné jednotlivé otázky formulovat nezávisle na ostatních. To je však téměř vždy nerealizovatelné. Až na velice sporé výjimky budou otázky vytvářet nějaký kontext. To pak komplikuje vyhodnocení (závislost odpovědí by měla být testována). Forma odpovědí by měla být poplatná dostupným vyhodnocovacím prostředkům. Pokud hodnotící tým má k dispozici schopnosti jen ze základů statistického aparátu, pak lze doporučit používat jen odpovědi ve formě čísla (hodnocení průměry, variačním rozpětím, po kvantování i četnosti, případně i kontingenční tabulky) a ve formě vyjmenovaných možností s výběrem právě jedné (hodnocení pomocí četností, kontingenčních tabulek, ...). Pokud otázky generují odpovědi ve formě stupnic, je zapotřebí předpokládat hodnocení se složitějším statistickým aparátem (pořadové statistiky, vlastní nebo speciální modely).

Nezávislost otázek je přirozeným požadavkem, ale jen z pohledu hodnocení. Jejich nepřítomnost může mít však pozitivní vliv někde jinde, např. při zajišťování – viz výše.

Otázky by měly být formulovány co nejjednodušeji a co nejjednoznačněji. Při jejich formulaci je nezbytné „vmyšlení“ se do tázaného a respektování jeho intelektuální úrovně. Jednoduchost a srozumitelnost mají vyšší prioritu než přesnost (někdy i za cenu mírné nejednoznačnosti). Případná **poučení a vysvětlení** dávat **přímo k otázce**, nestačí vysvětlující kapitola. Proband u otázky přemýšlí o jejím obsahu a srovnává (někdy ji s tím, co již absolvoval nebo zažil (toto je příznačné u vysoce vzdělaných lidí). Pak ale obvykle ztrácí návaznost na již přečtené (pokud vůbec) pokyny. Časté jsou záměny mezi bodováním (1-nejhorší, 5-nejlepší) a známkováním (1-nejlepší, 5-nejhorší) a obdobné. Tento požadavek (srozumitelnosti a přiměřené jednoznačnosti) nejlépe plní výčty s právě jedním výběrem hodnoty odpovědi. Při hromadných dotazovacích akcích by měly být vynechány odpovědi typu volně psaného textu (prakticky nejdou hodnotit, mohou však být někdy neobyčejně cenné – proto: pokud, tak jako nehodnotitelný doplněk; a spíše s účelem, co chce probant tazatelům vzkázat nebo volně sdělit – pozor na grafomany).

Otázky by neměly být sugestivní (rozhodnout ale, zda je něco sugestivní, je těžké). Např. než položit otázku:

Pokud půjdete volit, budete volit stranu

- a)
- b)

.....

je lepší položit otázku s odpověďmi:

Ve volbách budu volit stranu:

- a) ...
- b)
- c)
- d) Nepůjdu volit
- e) Nejsem dosud rozhodnut(a)

Dotazníky (u nichž je jejich vlastním účelem zjišťování) by neměly obsahovat motivující (morálně vydírající) text. Např.: Výsledky této dotazníkové akce budou využity pro sponzorskou činnost ve prospěch útulku pro zatoulané veverky. To už probanty usměňuje k volbě určitých typů odpovědí (podle toho, zda jsou ochránci zatoulaných veverek nebo naopak).

Speciální povahu mají otázky, u nichž se tážeme na něco jiného, než má být vlastně zjišťováno: Např.: Je realizována dotazníková akce, která má testovat, zdali se v místě, (místech), kde je testování realizováno, vyskytují (pohybují) potenciální zákazníci „second-handu“ (ti by se na přímý dotaz mohli „stydět“). Takový dotazník bude mít otázky (převážně třídicího charakteru), jejichž výsledek pouze určí, jestli je tázaný potenciálním zákazníkem tohoto typu obchodů, např.

1. Kupujete značkové oděvy ANO – NE.
2. Jste ochotna dát za svetr dát
 - a) méně než 300 Kč
 - b) mezi 300- 1700 Kč
 - c) více než 1700 Kč
3.

Zjišťovaná informace je pouze ta: odkud, z jakého tazacího místa dotazník pochází (a samozřejmě zařídění potenciální zákazník – nákup v second handu nepřipadá v úvahu). Může (=místo odkud) být zajištěna (i zjištěna) např. tazateli nebo i barvou (motivem) dotazníku (při volném vyložení dotazníků). Kupříkladu dotazníky s kytičkou budou vyloženy na centrálním náměstí, dotazníky s nápojovou sklenicí budou vyloženy před hostincem „U Děravé pípy“. Samotný motiv nebo barva mají i efekt přitahující (tj. lákající probanty). Zde je třeba rozlišovat i skutečnost, kde je dotazník získán, od skutečnosti, kde je dotazník podán (razítko podací pošty). To může mít různý (použitelný i nepoužitelný) význam. Takový typ dotazníku (na www stránkách zpravodajského serveru) může testovat, jak aktivní populace je a v jaký čas stránky navštěvuje. Jednoduchá (dvou až tří otázková) anketa se umístí na různé stránky serveru. Zjišťuje se pak prostá návštěvnost – (na tu ale není třeba dotazník), časy a časy a četnosti reakcí na anketu z jednotlivých stránek. Otázky v anketě mohou být i skryté třídicí.

Zvláštní povahu mají dotazníky, u nichž hlavním účelem je spíše vyvolat u „probantů“ nějaké jednání než získat odpovědi a informace. Např.: Dotazník s dotazy týkající se zahrádkářské problematiky může být vhodným připomenutím zákazníkům, že mají zajít do blízké (ne tolik „na očích“ umístěné) prodejny semen. Umístění orientační tabule nebo viditelné reklamy bývá obvykle dražší než jeden nebo dva „tazatelé“ v blízkosti prodejny (než se tam zákazníci naučí chodit). Takový dotazník pak má v sobě i informaci o umístění dané prodejny.

Taktéž může být vhodné dotazníkové akce (s příslibem losovaných cen za vyplnění) použito pro vygenerování adresáře pro zaslání reklam či něčeho jiného. Dotazy jsou pak jen třídicího a lokalizačního charakteru. Navíc, kdo vyplní takový dotazník, bude lépe snášet

zasílání všelikých smyslů a nesmyslů než ten, kdo se na daný dotazník nenechá nalákat. Lze tím také „obcházet“ ustanovení všelikých norem o důvěrnosti a užití osobních údajů.

Výběr kmene

Toto je problém profesionálních tazatelských institucí. Z daného firemního kmene (smluvních i nesmluvních dotazovaných) vybrat reprezentaci dané populace. Problém a jeho řešení bude naznačeno na příkladu.

Je známo, že sledovaná (celková) populace je složena z:

1. 50,5% žen a 49,5% mužů.
2. 35% jsou idealisté a 65% jsou pragmatičtí.
3. 40% populace je vyšší než 180 cm a 60% nižší než 180 cm.
4. 25 % populace vlastní chalupu, 70 % ji nevlastní.

Z ekonomických důvodů je možné aby dotázaný kmen měl jen 200 jedinců. Aby byl „reprezentativní“, musí tedy obsahovat:

- 101 žen a 99 mužů
- 70 idealistů a 130 pragmatiků
- 80 vyšších než 180 cm a 120 menších než 180 cm.
- 50 vlastníků chalup a 150 nevlastníků.

Obvykle nejsou známy „průnikové podíly“ na úhrnu (= na celku). Tj. je obtížné, ne-li nemožné, zjistit % pragmatických žen, vyšších než 180 cm a vlastníků chalupu, ... (často je takový problém řešen při předvolebních průzkumech – % pravicových voličů, % středových voličů, % levicových, % mužů, % žen, % obyvatelů měst, % obyvatelů venkova, % z českých korunních území, % z velkovévodství slezského, ...).

Jedná se o úlohu kombinatorickou (latinské čtverce) a její řešení je obecně NP složité, existují však přijatelně přibližná heuristická řešení (zejména pokud se netrvá na přesném dodržení počtu jedinců v aktuálním = dotazovaném kmeni). Řešení takových úloh lze nalézt v textech vědního oboru „Plánování experimentů“.

Opačný problém je, když kmen není projektován a dotazníky se předkládají náhodnému výběru probantů. Pak se pomocí třídících otázek zjišťuje příslušnost tázaného k té či oné skupině. Z takových údajů se pak zjišťuje relativní (=poměrné) zastoupení jednotlivých skupin podle konkrétních třídících hledisek. Z těchto poměrů se pak realizuje některý z testů „dobré shody s teoretickým modelem“ založený na multinomickém nebo vícerozměrném hypergeometrickém rozdělení pravděpodobnosti (častěji na některé z jejich aproximací – vícerozměrné normální, Wishartovo, ...). Podle výsledku testu (není zamítnuta hypotéza o příslušnosti náhodného výběru k dané celkové populaci, je zamítnuta) se pak s příslušným šetřením nakládá. Nejčastěji takové testování končí χ^2 testem dobré shody.

Typ sběru dotazníků

Existují dva základní typy sběru dotazníků:

1) Vyložené

- a) Volně ložené na některém více či méně přístupném místě se sběrnou urnou vedle vyložených dotazníků. Např. Dotazníky o spokojenosti zákazníků u východu z obchodu.
- b) Nabízené hosteskou u vstupu na výstavu a po vyplnění odevzdávané při odchodu z výstavy.

- c) Doručované tázanému do poštovní schránky. Např. zákonná šetření ČSÚ (dnes již možno také elektronicky). ČSÚ je ale netypický příklad – probant je k vyplnění obvykle přinucen zákonem nebo jinou formou (89/1995 Sb. a prováděcími vyhláškami ČSÚ, naposled 476/2006 Sb.).
- d) Dotazníky umístěné na www stránkách s elektronickým odesláním po vyplnění.
- e)

2) Zajišťované tazatelem

- a) Tazatel přichází za probantem (vybraným do tazacího kmene) a s probantem dotazník vyplňuje (obvykle za úplatu probantovi).
- b) Tazatel náhodně oslovuje probanty a inspiruje je ke společnému vyplnění dotazníku (dotazník neopustí tazatele).
- c) Tazatel náhodně oslovuje probanty a inspiruje je k vyplnění dotazníku spolu s informací, kde a jak mají vyplněný dotazník odložit (odevzdat). Dotazník tedy opustí tazatele.
- d) Dotazník je zaslán e-mailem probantovi (vybranému do tazacího kmene) s pokyny, jak a kam vyplněný dotazník předat či odeslat (jedná se o skrytě neanonymní šetření, dá hodně práce odeslat takový dotazník anonymně).
- e) Dotazník je zaslán e-mailem probantovi (náhodně nebo dle nějakého existujícího adresáře) s pokyny, jak a kam vyplněný dotazník předat či odeslat.
- f) Dotazník je vyplňován při telefonickém kontaktu tazatele a tázaného – levné, ale náchylné k záznamovým chybám (nejistá je pak vždy identifikace a zatřídění tázaného; pan Vopička může při dobrém hudebním nadání v telefonu úspěšně zahrát paní Vopičkovou, ...)
- g)

Podstatný rozdíl u obou skupin je v motivacích. Zatímco u tazatelských dotazníků není zapotřebí vnější motivace, ale používá se vnitřní motivace (např. úplata 50 Kč za vyplnění dotazníku nebo lízátko malému dítěti – aby nalákalo matku k vyplnění dotazníku, ...), u vykládaných dotazníků je zapotřebí vnější motivace, a to ve dvou etapách. První: aby si probant dotazníku nějakým způsobem všiml a druhá: aby probant dotazník vyplnil a předal. První etapa využívá převážně některých reklamních technik. Ty by měly být přitahující, neprovokující cílovou skupinu (uvažovaný kmen, mohou však provokovat necílové jedince), avšak výrazné a viditelné. Např. obálka se zasílaným dotazníkem by měla být výrazná a zajímavá. Při umístění dotazníku na www stránky by odkaz na anketu (dotazník) měl být výrazný, neprovokující a hlavně vlevo nahoře na stránce (zcela určitě bez potřeby rolování). Na přístup k dotazníku by mělo stačit „jedno kliknutí“. Také se používá přitahujících návykových technik (na místo budoucího dotazníku se umístí fotografie (nepříliš, ale) oblečené slečny, která na kliknutí zareaguje nějakou zprávou, např. „Pokračování zítra“. Obdobný efekt může mít fotografie exotického zeměpisného místa. Na místo budoucího vyložení dotazníků se umístí výrazný plakát s tajemným neurčitým obsahem, aby se zákazníci „naučili“ na dané místo dívat, .. .

Druhá (etapa) motivace je inspirace k vyplnění dotazníku. Účast ve slosování o ceny, zde je užitečné studovat techniky zcela korektního losování s malou pravděpodobností, že bude vylosován probant, který dotazník vyplnil (např. do slosování se zařadí všechna čísla i

nepředaných dotazníků {příliš triviální, existují dokonalejší}, ...). Poukázka na slevu při koupi (neprodejného, ...) zboží. Výměna vyplněného dotazníku na odevzdávacím místě za „nákupní bony“,

Obecně: Je třeba uvažovat, že každý dotazník je svým způsobem prototypem (neplatí pro opakování). Jeho návrh je tvůrčím dílem s velkou mírou osobní nealgoritmizované invence. Pro korektní šetření jsou nezbytné znalosti a schopnosti „public relations“. A samozřejmě znalosti pravděpodobnosti, plánování experimentů, teorie sociálních šetření, sociálních výzkumů a hlavně dokonalé zvládnutí statistiky. Dále je nezbytné v každém kroku dotazníkového projektu neopomenout brát v úvahu celou trajektorii dotazníkové (anketní) akce. Nezapomínat na to, že průběhem realizace může být podstatně ovlivněno výsledné hodnocení a i výsledek. Teorie říká, že celá anketní trajektorie by měla být výsledkově neutrální (neplatí pro dotazníky s ovlivňujícím charakterem a pro šetření, ve kterých je objednán výsledek, což je totéž).

Vyhodnocení

Vyhodnocení by mělo probíhat v několika krocích:

1. Protokolace výsledků sběru s popisem metod sběru.
2. Testy a reakce na vnější zajištění (tj. testování tazatelů).
3. Testy a reakce na vnitřní zajištění (tj. testování tázaných nebo jimi zadaných výsledků).
4. Skupinové testy shody pro různé zdroje (tazatele) a reakce na jejich výsledky.
5. Testy předpokládatelných závislostí (testy závislostí jednotlivých otázek a případně i jejich skupin) a reakce na jejich výsledky.
6. Protokolace výsledků předchozích kroků.
7. Klasická popisná statistika (počty, četnosti, průměry, mediány,, vývoj v čase, vývoj v prostoru).
8. Grafická prezentace výsledků bodu 7.
9. Srovnání získaných výsledků s projektem (zadáním), formulace pracovních hypotéz o výsledcích dotazníkové akce (ankety).
10. Statistické nebo jiné testování pracovních hypotéz.
11. Validace získaných výsledků z pohledu poznatků dané vědní nebo odborné oblasti. Diskuse nesouladů. Někdy bývá zajišťováno až ve fázi interpretační.
12. Formulace doporučení pro zadavatele, formulace doporučení do další praxe tazatelského subjektu (instituce pro sociologický ... výzkum).
13. Zpracování publikace výsledků ve všech jejích částech: Výsledkové, přehledové, detailní a interpretační (vše viz dále).
14. Zpracování prezentace.
15. Archivace dat, protokolů a poznatků.

Poznámka: Prvních 10 bodů je vlastní hodnocení, zbytek jej již nadstavba.

Metody hodnocení jsou poplatné typům jednotlivých otázek. Popis možností je uveden výše u jednotlivých typů otázek. Je důležité zopakovat, že **vždy**, v jakémkoliv okamžiku dotazníkového (anketního) projektu nesmí být opomenuta **fáze vyhodnocení**. Dotazník by neměl odejít mimo pracoviště bez konzultací s těmi, kdo jej budou vyhodnocovat (také v solidních firmách poslední atest na prototyp dotazníků dává hodnotící a analytické pracoviště). Je důležité respektovat vlivy jednotlivých otázek samostatně a i v kontextu na ostatní otázky. Dále je důležité, aby byly jasně vymezeny třídící, zajišťovací a informační

otázky (v dotazníku mohou být promíchány, postavení té které otázky k uvedeným skupinám musí být ale jasné pro zpracovatelský tým). Stejně tak, jako metody hodnocení (statistické, sociologické, ..), nesmějí být opomenuty **technologie vlastního převodu informací** z vyplněných dotazníků do digitalizované formy (např. barevný motiv na dotazníku může značně zkomplikovat „scanování“ a nakonec i případný převod metodou oko ruka, také se doporučuje respektovat zvyklostní konvence, např. modrá = ztracené pozadí; při přijímání dotazníků „vyložených“ na www stránkách by měl být automaticky generován a aktualizován vyhodnocovací soubor již ve fázi pořizování, ...). Z pohledu korektního hodnocení jsou nejtěžší otázky s **množinovými výčty**. Ač obvykle velmi informativní (individuálně), mají poměrně komplikovanou pravděpodobnostní strukturu a řádově komplikovanější strukturu statistickou (vlivem velkého možného počtu kombinací se často budou vyskytovat kombinace s nulovým výskytem v dotazníkovém šetření – i když reálně budou důvody pro nenulové pravděpodobnosti výskytů takových kombinací; zde se pak používají statistiky řešící tzv. zero-frequency problem, .., velkým problémem je také možnost testování souvislosti odpovědí takových otázek s jinými, ...). Obecně jdou vždy takové typy otázek převést na více otázek s jediným přípustným výběrem. Reálně to však nemusí být jednoduché a dobře interpretovatelné (podmíněné výskyty). Proto je nutné zvážit nezbytnost otázek typu množinového výběru. Také platí empirický poznatek, čím více otázek s množinovým výběrem v dotazníku, tím méně kvalifikovaný realizátor dotazníkové akce (neví, o čem je řeč).

Prezentace

Prezentace by v přehledné formě měla obsahovat následující části (a i v následujícím pořadí).

1. Získané výsledky.
2. Použité metody.
3. Problémy při výzkumu.
4. Závěry a případná doporučení.

O náročnosti či obsažnosti prezentace podstatně rozhoduje předpokládané publikum. Jiná bude prezentace pro novináře, jiná pro odborníky v dané vědní nebo odborné oblasti. Jiná bude prezentace pro manažery. Také je zapotřebí respektovat charakter prezentace. Zda se jedná o prostou prezentaci a pak už nic, prezentaci s tiskovou konferencí (zde se doporučuje v prezentaci navodit témata pro dotazy novinářů a mít na ně připravenou dílčí prezentaci jako reakci = tak tohle jste chtěl(a) vědět, vidět, ...), prezentaci s následující odbornou diskusí (zde je dobré mít k dispozici u prezentace „stromové“ odkazy až do poměrně velkého detailu). U prezentací pro manažery je zapotřebí použít návrhů pro řešení (není to analytické, ale užitečné, analytik by neměl mluvit do výkonu). V jaké formě a jaké údaje prezentovat závisí na publiku. U publika s nepříliš velkými poznatky v sociologickém výzkumu a statistice je zapotřebí používat údajů s očekávanou interpretací. Korektní statistické výsledky se korektně prezentují na úrovni trojic:

- Reprezentativní hodnota
- Pásmo, ve kterém se reprezentativní hodnota nachází, nebo jiné vymezení její neurčitosti, volatility nebo nepřesnosti.
- Hladinu významnosti (spolehlivosti), tj. odhad pravděpodobnosti, že tomu tak není.

U reprezentativní hodnoty je běžným publikem předpokládána vlastnost centra = symetrie, tj. stejný počet jedinců je na tom hůře a stejný počet lépe než u reprezentativní hodnoty (tedy nejlépe medián a u „symetricky“ rozdělených průměr – pokud má smysl a

interpretaci). Pásmo spolehlivosti se obvykle uvádí ve formě \pm , tedy také v symetrické formě. Pokud to nelze nebo by vedlo k neplatným závěrům, je zapotřebí použít nesymetrickou formu s patřičným vysvětlením. V takovém případě u případu průměr-medián a v obdobných je zapotřebí při prezentacích dělat „statistickou“ osvětu. Také místo hladiny spolehlivosti je zapotřebí pro neodborníky používat pojmu jim přijatelného, např. „celková předpokládatelná chyba“ a pro omylem přítomné odborníky uvést i statisticky korektní pojem. Toto vše ale platí pro statisticky korektní prezentace.

Interpretace

Ryze a idealisticky teoretické pravidlo říká, že hodnotitel výstupů by je neměl interpretovat. To je oblastí odborníka z dané interpretační oblasti (sociologa toho či onoho typu, ..). To však je prakticky téměř neproveditelné. Interpretace by proto měla být zpracovávána v „konvergujících cyklech“: hodnotitelé dodají výsledky interpretům → interpretační skupina zpracuje interpretaci → hodnotitelé se k ní vyjádří a popřípadě doplní hodnocení → interpretační skupina doplní a upraví interpretaci → ta je předána k dalšímu vyjádření hodnotitelům. Pokud ani poté nedojde k jejich shodě, obvykle se zveřejní interpretace a za ní následuje „votum separatum“ analytiků nebo hodnotitelů. V takovém odděleném vyjádření se pak uvedou důvody a vyjádření analytiků k interpretacím. Toto je praxe špičkových dodavatelů – menší subjekty by se k ní měly blížit. Taková praxe ale vyžaduje hraniční vzdělání (aby se dokázali domluvit a aby si rozuměli). Interpreti by měli být sociologové (při výzkumech veřejného mínění) a obdobní odborníci (u jiných typů anket) s další aprobací statistika a informatika. Naopak hodnotitelé by měli být statistici s dalšími aprobacemi sociologie (nebo srovnatelní odborníci v daném oboru šetření) a informatika.

Publikace

Publikace každé ankety by měla mít několik základních oddílů: Výsledkový, přehledový, detailní a interpretační.

Výsledkový

Zde se přehledně a stručně uvedou dosažené výsledky, a to jak z anketního šetření, tak i z jejich interpretace. Obvykle se neuvádí důkazní materiál a zdůvodnění. Výsledkový oddíl by neměl přesáhnout 1-2 A4 při obvyklém 1,5 řádkování. Někdy se pro něj používají názvy „šéfovská stránka“, „technical summary“ nebo „non-technical summary“. Je určen pro obecnou a manažerskou veřejnost (bez velkého vynaložení času se lze dovědět, čeho bylo či nebylo dosaženo).

Přehledový

Ten je rozvinutím výsledkového oddílu s tím, že se do něho doplní zdůvodnění, a to ještě ve formě spíše popisné než exaktní. Také se v něm objevují návrhy a náměty na další konání. Existuje empirické pravidlo, že až sem by to měli dočíst novináři, další čtení už by neměli vydržet. Pokud čtou i detailní oddíl, přisvojují si obvykle právo na interpretace (většinou nekorektní a nehomogenní).

Detailní

Opět se jedná o rozvinutí předchozího oddílu. Ten je pak u každého nově získaného faktu podložen detailní diskusí a doplněním zdrojových informací, z nichž publikovaný fakt vyplývá. Detailní oddíl by měl obsahovat data z dotazníkového šetření v nějaké více či méně

agregované formě (tabulky, grafy, ...) a odkaz na syrová data. Také by měl obsahovat podrobný popis použitých zpracovatelských metodik a jejich výsledků. Empiricky: ten, kdo přečte a rozumí detailnímu oddílu, by měl umět celý anketní projekt zopakovat (pokud na to má prostředky).

Interpretační

Interpretační oddíl obsahuje interpretace = transformace výsledků do jazyka té oblasti, pro kterou je dotazníková akce realizována. Obvykle bývá doplněn i citacemi předchozích prací a děl s danou tematikou a srovnáním v nich dosažených výsledků s aktuálně získanými výsledky. Dále by zde měl být alespoň pokus o vysvětlení nesouladů s citovanými autory. Samozřejmě i přehled shod. Interpretační oddíl by měl mít dvě podseky. Vlastní výsledkovou a diskusní (diskusí se zde míní podrobný popis podmínek, předpokladů a okolností, za kterých byly výsledky získány nebo formulovány).

Literatura pro další studium:

Adámek: Kódování. SNTL, Praha 1989. Edice Matematika pro vysoké školy technické.

Alan, J., Machonin, P.: Jak se rodí sociologický výzkum. K problémům výzkumu transformace sociální struktury československé společnosti. Praha, UK 1992.

Bican: Algebra II. Skripta MFF - UK, SPN Praha 1982.

Bican, Kepka, Němec: Úvod do teorie konečných těles a lineárních kódů. Skripta MFF-UK, SPN Praha 1982.

Buriánek, J.: Systémová sociologie. Problém operacionalizace. Praha, UK 1993.

Disman, M.: Jak se vyrábí sociologická znalost. Příručka pro uživatele. Praha, UK 1993.

Disman, M.: Terénní sběr informací. In: Vybrané techniky sociologického výzkumu. Praha, Svoboda, 1969.

Gallup, G.: Průvodce po výzkumu veřejného mínění. Praha 1948.

Horáková, H. (2003): Strategický marketing. Grada Publishing

Herzmann, J.: Výzkum veřejného mínění. Praha: VŠE 1996.

Kane, E.: Doing Your Own Research. London-New York, Marion Boyars, 1985.

Kapr, J., Šafář, Z.: Sociologie nebo zdravý rozum? Praxe sociologického průzkumu. Praha 1969.

Katriak, M., Milly, S.: Sociologický výzkum. Bratislava 1971.

Katriak, M.: Metodologické zásady sociologického výzkumu. Bratislava 1968.

Kerlinger, F.N.: Základy výzkumu chování. Pedagogický a psychologický výzkum. Praha 1972.

kol.: Metody a techniky sociologického výzkumu I. (Principy, struktura a strategie empirického výzkumu) Praha, FF UK 1982.

Kolektiv autorů (ed. J. Šubrt): Kapitoly ze sociologie veřejného mínění, Karolinum, Praha, 1998

Komárková R. a kol.: Psychologie trhu, Grada Publishing 1998

Kotler, P. aj.: Principles of marketing, Prentice Hall 1996,

Lamsar, V.: Základy sociologického výzkumu. Praha 1966.

Moser, C.A. - Kalton, G.: Survey in Social Investigation. London-,Heineman 2nd ed., 1971.

Noelleová, E.: Výzkum veřejného mínění. Praha, Svoboda, 1969.

Noelleová, E.: Výzkum veřejného mínění. Úvod do metod demoskopie. Praha 1968.

Nowak, S.: Metodologie sociologických výzkumů. Obecné problémy. Praha 1975.

Pecáková, I., Herzmann, J., Novák, I.: Pořizování a vyhodnocování dat ve výzkumech veřejného

mínění. Praha VŠE, 1998.

Petrusek, M.: Sociometrie. Praha, Svoboda, 1969.

Petrusek, M.: Teorie a metoda v moderní sociologii. Praha 1993.

Schneider, M., Koudelka, F.: Úvod do základů sociologických výzkumů. Olomouc, FF UP 1993.

Szostkiewicz, S.: Sociologický výzkum. Bratislava 1965.

Velký sociologický slovník. Praha 1996.

Výrost, J.: Sociálno-psychologický výskum postojov. Bratislava 1989.