

Metodologie – průzkum veřejného mínění

3. 12. 2020



- Cíl: kvantifikovat postoje
- Nutnost: standardizace otázek
- Dobrá explanace
- Dobrá operacionalizace

Metoda dotazníku

- PAPI (pen and paper)
- CAPI (computer assisted personal)
- CATI (computer assisted telephone)
- IRV (interactive voice response)
- CASI (computer assisted self-interview)
- CAWI (computer aided web interviewing)

Tazatel

- Zprostředkovatel mezi dotazníkem a respondentem
- Musí znát dotazník
- Trénink!!!
- Problémy: Důvěra, interviewer bias
- Neutralita (nereagovat na odpovědi)
- Může vysvětlit
- Menší drop out rate
- Citlivé otázky – lepší self-administered
- PAPI: větší kontrola, ale nutná důvěra
- Použití karet
- **Drahý sběr**
- Kontrola?

Většina dotazníků – výběrový soubor

- Výsledek vždy nepřesný
- Vždy zkreslení díky náhodě
- Nelze jednoduše zobecnit
- Odhad – ze vzorku na populaci
- Podmínka generalizace: dobrý sampling

The Effect of Interviewer Characteristics and Expectations on Response. Singer, Frankel, Glassman. 1983.

Table 2. Screening and Response Rates, by Interviewers' Demographic Characteristics^a

	<i>N</i>	<i>Screening Rate</i>	<i>N</i>	<i>Response Rate</i>
Age				
18-21	9	81.5 (79.3)	9	63.4 (61.2)
22-34	11	90.8 (89.7)	10	69.1 (65.9)
35+	11	89.1* (92.0)	11	72.1 (76.8)
Experience				
None	13	83.2 (82.3)	12	70.0 (68.5)
Less than one year	12	91.0 (90.7)	12	73.1 (71.4)
One or more years	6	89.1* (92.3)	6	56.3* (62.7)
Number of cases				
07-30	13	91.3 (91.2)	13	77.8 (79.4)
46-79	11	85.1 (84.1)	10	62.4 (62.8)
88-224	7	84.2* (86.0)	7	60.0* (56.3)
Expectations about response rate				
Moderately easy	11	—	11	77.37 (78.9)
Neither easy nor difficult	10	—	10	65.48 (65.9)
Moderately difficult	9	—	9	60.94*(58.6)
	$R^2 = .593$		$R^2 = .800$	

* $p (F) < .05$.

^a Cell entries are mean scores, adjusted by MCA for the other variables in the equation, which has been reestimated to include only those variables identified as significant in earlier models. Unadjusted scores are shown in parentheses.

Huddy et al. 1997. The Effect of Interviewer Gender on the Survey Response

- Vliv genderu tazatele na odpovědi
- Téma: ženské hnutí, ženská témata, gender equality

TABLE 2. Mean Scores for All Items on Standard 1–10 Scale: By Gender of Interviewer

	1991 Suffolk County Survey Interviewer Gender			1993 Long Island Survey Interviewer Gender		
	Male (N = 144)	Female (N = 159)	Difference Female-Male	Male (N = 125)	Female (N = 230)	Difference Female-Male
<i>Feminists</i>						
Feminist Identity	5.19	5.95	.76***	5.48	5.91	.43
Pro-Feminist Affect	5.84	6.15	.31	6.17	6.55	.38
<i>Activism</i>						
Activity on Women's Issues	2.32	2.82	.50°	—	—	—
Pro-Collective Action	6.84	7.32	.48°	6.30	6.99	.69**
<i>Gender Equality</i>						
Egalitarian Roles	6.54	6.89	.35°	5.94	6.04	.10
Inequality	7.10	7.03	-.07	6.98	6.79	-.19
Discrimination	6.53	6.63	.10	6.54	6.49	-.05
<i>Public Policies</i>						
Pro-choice	7.45	7.35	-.10	6.94	7.76	.82***
Anti-Sexual Harassment	7.06	7.37	.31	7.10	7.73	.63**
Sexual Harassm. serious	—	—	—	6.81	7.13	.32
Pro-Affirmative Action	5.50	6.34	.84**	4.19	4.22	.03
<i>Female Political Figures</i>						
Pro-Ferraro	5.18	5.83	.65	—	—	—
Women Better Suited	5.36	5.60	.24	—	—	—
Pro-Hillary Clinton	—	—	—	7.27	7.50	.23
<i>Clarence Thomas</i>						
Pro-Thomas	6.47	6.91	.44	—	—	—
Mean Difference			.37			.31

*p < .1; **p < .05; ***p < .01.

Note: Entries are mean scores standardized on a 1-to-10 scale, with 10 scored as the most feminist response. Difference scores are calculated by subtracting the mean scores obtained for female interviewers from the mean score obtained for male interviewers; higher scores indicate a tendency to give more feminist answers to female than male interviewers. All item wording is presented in the Appendix.

Sample survey

- Element
- Populace
- Výzkumná populace
- Sampling frame
- Reprezentativita
- Výběrová chyba

Techniky vzorkování

- Nepravděpodobnostní vzorky
- Pravděpodobnostní vzorky
 - Můžete spočítat šanci, že se daný element dostane do vašeho vzorku
 - Lze spočítat pravděpodobnost, že je váš vzorek reprezentativní

Dostupné subjekty (convenience sample)

- Lidé na ulici
- Studenti

- Co nejvíce dostupný soubor (convenient)
- Riziko?

Purposive sample (záměrný vzorek)

- Na základě vlastního úsudku
- Podle **povahy a cíle výzkumu**
- Často u pilotáže
- Vlastní úsudek o tom, které subjekty jsou vhodné
- Např.: lídři protestního hnutí
- Studenti s pravicovou/levicovou orientací
- Poslanci
- Prodavači kebabu....?

Snowball

- Kvalitativní výzkum
- Explorační záměry
- Nereprezentativní

Kvótní výběr

- Jednotky vybrány na základě charakteristik
- Cíl: stejná distribuce jako u cílové populace
- Tabulka daných charakteristik a proporce
- Nepravděpodobnostní
- Lze přidávat váhy podle proporce v populaci

		18-39 let	40-59 let	60+ let
Muž	Hlavní město	3.7%	7.7%	7.5%
	Město	2.8%	5.2%	3.1%
	Venkov	7.2%	5.4%	3.9%
Žena	Hlavní město	3.9%	8.6%	10.8%
	Město	4.0%	3.9%	4.5%
	Venkov	6.2%	6.3%	5.4%

Known population/Sample populatio

Např.

V populaci ženy 51

Ve vzorku ženy 41

$51/41 = 1,24$

Kvóta

- Je reprezentativní?

Pravděpodobnostní výběr

- Každá jednotka v populaci má známou nenulovou pravděpodobnost, že bude vybrána (není stejná pro všechny)

Simple random sampling

- Potřebujeme sampling frame, vybereme náhodně
- (Excel, F = RAND nebo RANDBETWEEN)

```
77921 06907 11008 42751 27756 53498 18602 70659 90655 15053 21916 81825 44394 42880
99562 72905 56420 69994 98872 31016 71194 18738 44013 48840 63213 21069 10634 12952
96301 91977 05403 07972 18876 20922 94595 56869 69014 60045 18425 84903 42508 32307
89579 14342 63661 10281 17453 18103 57740 84378 25331 12566 58678 44947 05585 56941
85475 36857 53342 53988 53060 59533 38867 62300 08158 17983 16439 11458 18593 64952

28918 69578 88231 33276 70997 79936 56865 05859 90106 31595 01547 85590 91610 78188
63553 40961 48235 03427 49626 69445 18663 72695 52180 20847 12234 90511 33703 90322
09429 93969 52636 92737 88974 33488 38320 17617 30015 08272 84115 27156 30613 74952
10365 61129 87529 85689 48237 52267 67689 93394 01511 26358 85104 20285 29975 89868
07119 97336 71048 08178 77233 13916 47564 81056 97735 85977 29372 74461 28551 90707

51085 12765 51821 51259 77452 16308 60756 92144 49442 53900 70960 63990 75601 40719
02368 21382 52404 60268 89368 19885 55322 44819 01188 65255 64835 44919 05944 55157
01011 54092 33362 94904 31273 04146 18594 29852 71585 85030 51132 01915 92747 6-4951
52162 53916 46389 58586 23216 14513 83149 98736 23495 64350 94738 17752 35156 35749
07056 97628 33787 09998 42698 06691 76988 13602 51851 46104 88916 19509 25625 58104

48663 91245 85828 14346 09172 30168 90229 04734 59193 22178 30421 61666 99904 32812
54164 58492 22421 74103 47070 25306 76468 26384 58151 06646 21524 15227 96909 44592
32639 32363 05597 24200 13363 38005 94342 28728 35806 06912 17012 64161 18296 22851
29334 27001 87637 87308 58731 00256 45834 15398 46557 41135 10367 07684 36188 18510
02488 33062 28834 07351 19731 92420 60952 61280 50001 67658 32586 86679 50720 94953
```

Systematic sampling with a random starting point

- Sampling frame
- Každý má unikátní číslo
- Každý K-tý element bude vybrán
- Sampling interval (population n/sample n) $20\ 000/100 = 200$ (každý 200. element vybrán)
- Náhodně vybráno číslo (56)
- 56, 256, 456.....
- Je třeba mít tu populace opravdu náhodně seřazenou
- Pozor na vzorce v řazení (např ročníky, abecední řazení podlele skupin atd.)

Stratified sampling

- Tabulka s populačními poměry
- Podle toho stratifikujeme seznam
- Vybíráme náhodně ze skupin
- Je to doplňková technika v celkové strategii, ne samostatná procedura
- Lze kombinovat
- Chce to dimenze, které spolu nekorelují a jsou pro náš design přínosné

Multistage cluster sampling

- Když nemáme seznam všech elementů v populaci
- Nejprve sampling skupin elementů (clusters)
- Výběr clusterů, ze kterých vytvoříme seznamy a budeme vybírat
- Národně reprezentativní vzorky

Příklad

- Systematický stratifikovaný výběr s náhodně určeným začátečním bodem – výběr města
- 1. seřadit města podle populace
- 2. sampling interval, náhodně stanovit první číslo a pak výběr na základě výběrového intervalu
- 3. náhodný výběr lidí ze seznamu obyvatel města

Proporčnost počtu elementů

- Pokud jsou clustery města - nestejně velké – mohlo by dojít k nadreprezentaci lidí z velkých měst
- STAGE I. Výběr skupin (e.g. obcí) podle velikosti Město se 40.000 obyvateli má 4x větší šanci, že se dostane do vzorku než město se 10 000 obyvateli
- STAGE II. Výběr stejného počtu elementů z každého města. Např. 100 lidí z každého města

- Pravděpodobnost jedince z jakéhokoliv města v zemi, že bude vybrán?
- Příklad: země má 10 000 000 obyvatel a 3000 obcí:
chceme vybrat 100 obcí a 100 lidí z každé obce, dohromady 10.000 lidí

Šance, že bude vybráno velké mesto:

$$100 \cdot (40.000 / 10.000.000) = 0.4$$

Šance, bude-li vybráno velké město, že se do vzorku dostane jeho obyvatel:

$$100 / 40.000 = 0.0025$$

Celková šance, že se člověk dostane do vzorku:

$$0.4 \cdot 0.0025 = 0.001, \text{ thus } 0,1\%$$

Šance, že bude vybráno malé mesto

$$100 \cdot (10.000 / 10.000.000) = 0.1$$

Šance, bude-li vybráno malé město, že se do vzorku dostane jeho obyvatel:

$$100 / 10.000 = 0.01$$

Celková šance, že se člověk dostane do vzorku :

$$0.1 \cdot 0.01 = 0.001, \text{ thus } 0,1\%$$

CELKOVÁ CHYBA VÝZKUMU

- Survey research většinou produkuje nepřesnosti
- <https://zpravy.aktualne.cz/domaci/analyza-pruzkumy-na-pranyri-volby-ukazaly-kdo-je-umi/r~b3d2a54cb71911e79090002590604f2e/>
- <https://interaktivni.rozhlas.cz/volebni-pruzkumy/>
- Kolísavá kvalita
- Pravděpodobnost chyby na všech úrovních procesu sběru, měření, analýzy a interpretace
- Cíl co největší validita a reliabilita

Coverage error

- Pokud chci dělat generalizace
- Pracuji s populace
- Můj sampling frame vykazují systematický bias
- Např. telefonní čísla vzhledem k target population = voliči nad 18 let

Sampling error

- Pokud mám vzorek
- Když budu opakovat měření na jiném stejně velkém vzorku, budu mít jiný výsledek
- Často jako tolerance chyby, margin of error
- suma všech možných výběrových chyb kvantifikující nejistotu výsledků
- Interval spolehlivosti (např 95% - rozpětí kolem naměřené hodnoty) Jsem si jistí, že naše data budou obsahovat z 95% skutečnou hodnotu

Nonresponse error

- Vybereme jedince do vzorku
 - Nedostaneme odpověď
 - Respondenti – pouze podvzorek
 - Liší se?
-
- Unit non-response
 - Item non-response

Measurement error

- Jak dobře měříme své teoretické koncepty?
- Pretest
- Cognitive pretesting

Non response

- Neschopnost zajistit odpověď
- Někdy úplně (úplná nekooperace)
- V některých případech jen některé položky dotazníku
- Totální nonresponse = unit nonresponse
- Položky = item nonresponse
- Pokud je nonresponse systematický jev = máme v datech bias
- Korelace mezi položkou a šancí, že získáme odpovědi (spolupráci respondenta)

Non response důvody

- Neschopnost doručit dotazník respondentovi
- Neschopnost zajistit kooperaci
- Neschopnost respondenta odpovědět

Response rate

- % respondentů ve vzorku, které bylo změřeno
- V USA i Evropě významně klesá