

Wat is eigenlijk representativiteit?

Barry Schouten (CBS)



DANS Symposium, 9 april 2008

Wat is goed survey onderzoek?

- *Hoe bereiken we de gewenste doelpopulatie?*
- *Hoe zorgen we voor een representatieve respons?*
- *Hoe krijgen we goede antwoorden op de vragen?*



Voorbeeld 1

| | Respons gemiddelde | | Schatting | |
|------------------------|--------------------|--------------|------------|--------------|
| | Na 1 maand | Na 2 maanden | Na 1 maand | Na 2 maanden |
| Werkzaam | 48,6% | 50,4% | 49,6% | 50,6% |
| Huisbezitter | 63,0% | 63,3% | 59,1% | 59,4% |
| PC bezitter | 59,6% | 59,8% | 57,3% | 57,2% |
| Ontvangt een uitkering | 10,5% | 10,4% | 11,6% | 11,4% |
| Allochtoon | 12,9% | 12,5% | 14,6% | 14,4% |

12,1% ontvangt een uitkering en 15,0% is allochtoon



Voorbeeld 2

Enquête Beroepsbevolking juli – oktober 2005

Herbenaderingsmethode (Hurwitz 1949)

- Selectie uit interviewers met hoogste responscijfers
- Aanvullende training van interviewers
- Beloningen
- Papieren overzichten met gegevens huishouden

Centrale-vraag-methode (Kersten & Bethlehem 1984)

Korte telefonische, schriftelijke en web vragenlijsten



Voorbeeld 2

Herbenaderingsmethode; n=785 uit 18076 non-respondenten

| | Respons |
|------------------|---------|
| EBB | 62% |
| Na herbenadering | 77% |

Centrale-vraag-methode; n=942 uit 18076 non-respondenten

| | Respons |
|--------------------------------------|---------|
| EBB | 62% |
| EBB met vaste telefoon | 68% |
| Na centrale vraag | 76% |
| Na centrale vraag met vaste telefoon | 83% |



Voorbeeld 3

| | 0-20 | 20-65 | 65+ | Totaal |
|------------------------|--------------------|---------------------|--------------------|------------|
| Nederland | 4 miljoen 25,0% | 10 miljoen 62,5% | 2 miljoen 12,5% | 16 miljoen |
| Panel 1 - zelfselectie | 40.000 40,0% | 55.000 55,0% | 5.000 5,0% | 100.000 |
| Panel 1 - steekproef | 2.500 25,0% | 6.250 62,5% | 1.250 12,5% | 10.000 |
| Panel 2 - zelfselectie | 25.000 20,8% | 77.000 64,1% | 18.000 15,0% | 120.000 |



Steekproeftheorie

Selecteer een deelpopulatie met een steekproefmechanisme waarvan je de trekkingskansen zelf kiest.

Afweging tussen kosten en precisie. Goedkopere waarneming gaat gepaard met onzekerheid in statistiek. Echter, geen systematische vertekening omdat de selectiekansen bekend zijn.



Non-respons

Selectiekansen zijn niet meer in eigen hand. Naast een aanvullend verlies in precisie nu ook een eventuele systematische vertekening.

Wanneer is de respons “representatief”?

Wanneer kunnen we respons beschouwen als een aselechte steekproef uit de steekproef?



Representativiteit in de literatuur

Kruskal & Mosteller (1979)

1. Als aanmoediging van de data (klopje op de rug)
2. Afwezigheid van selectieve krachten
3. Als miniatuur van de populatie (quota sampling)
4. Purposive sampling
5. Afdekking van heterogeniteit van populatie
6. Representative sampling

Door het ontbreken van een heldere definitie, kan het begrip op verschillende manieren gebruikt en geïnterpreteerd worden.



Definities van representativiteit

Definitie (sterk): De respons op een onderzoek is representatief t.o.v. de steekproef als individuele responskansen gelijk zijn en als de respons van een eenheid onafhankelijk is van de respons van andere eenheid.

Definitie (zwak): De respons op een onderzoek is representatief voor een variabele X als de gemiddelde responskans over categorieën van X constant is.



Hoe een afwijking van representatieve respons te meten?

Variatie in responskansen

$$R(\tilde{\rho}) = 1 - 2 \sqrt{\frac{1}{N-1} \sum_{i=1}^N (\rho_i - \bar{\rho})^2}$$

Geschatte variatie in geschatte responskansen

$$\hat{R}(\hat{\rho}) = 1 - 2 \sqrt{\frac{1}{N-1} \sum_{i=1}^N \frac{s_i}{\pi_i} (\hat{\rho}_i - \hat{\rho})^2}$$

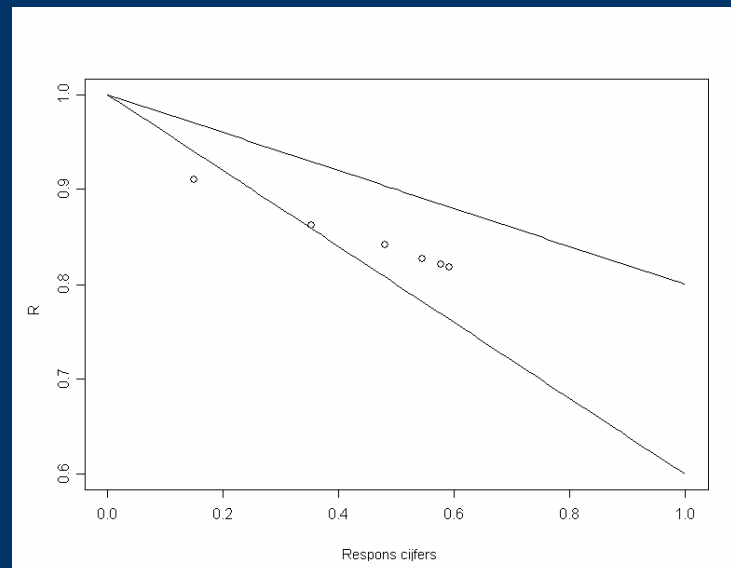


Hoe de afwijking te interpreteren?

Voorzichtigheid geboden: Afhankelijkheid van hulpvariabelen en van steekproefomvang

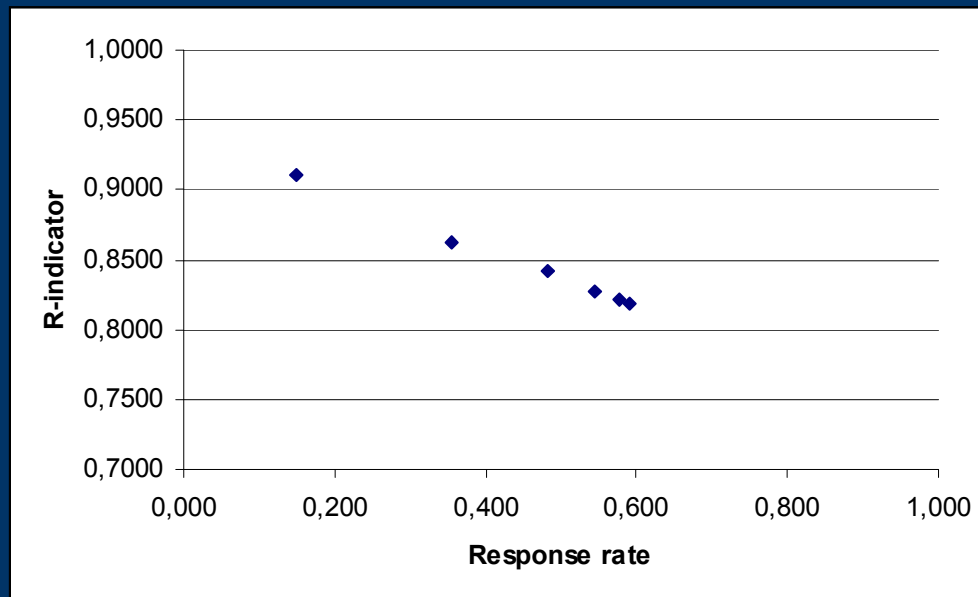
Normering van R:

Maximale systematische fout bij slechtst-denkbare scenario



Voorbeeld 1

Schatting van responskansen na 1 – 6 contactpogingen
Leeftijd, herkomstgroep, provincie



Voorbeeld 2

Herbenaderingsmethode

Telefoon, provincie, herkomstgroep, huishoudtype, stedelijkheid

| | Respons | R | Vertekening |
|---------------|---------|-----|-------------|
| EBB | 62% | 79% | 9% |
| Herbenadering | 77% | 85% | 5% |

Centrale-vraag-methode

Huishoudtype, stedelijkheid, leeftijd, geslacht, baan in loondienst, uitkering

| | Respons | R | Vertekening |
|-----------------------|---------|-----|-------------|
| EBB | 62% | 79% | 9% |
| EBB (tel) | 68% | 85% | 5% |
| Centrale vragen | 76% | 77% | 8% |
| Centrale vragen (tel) | 83% | 87% | 4% |



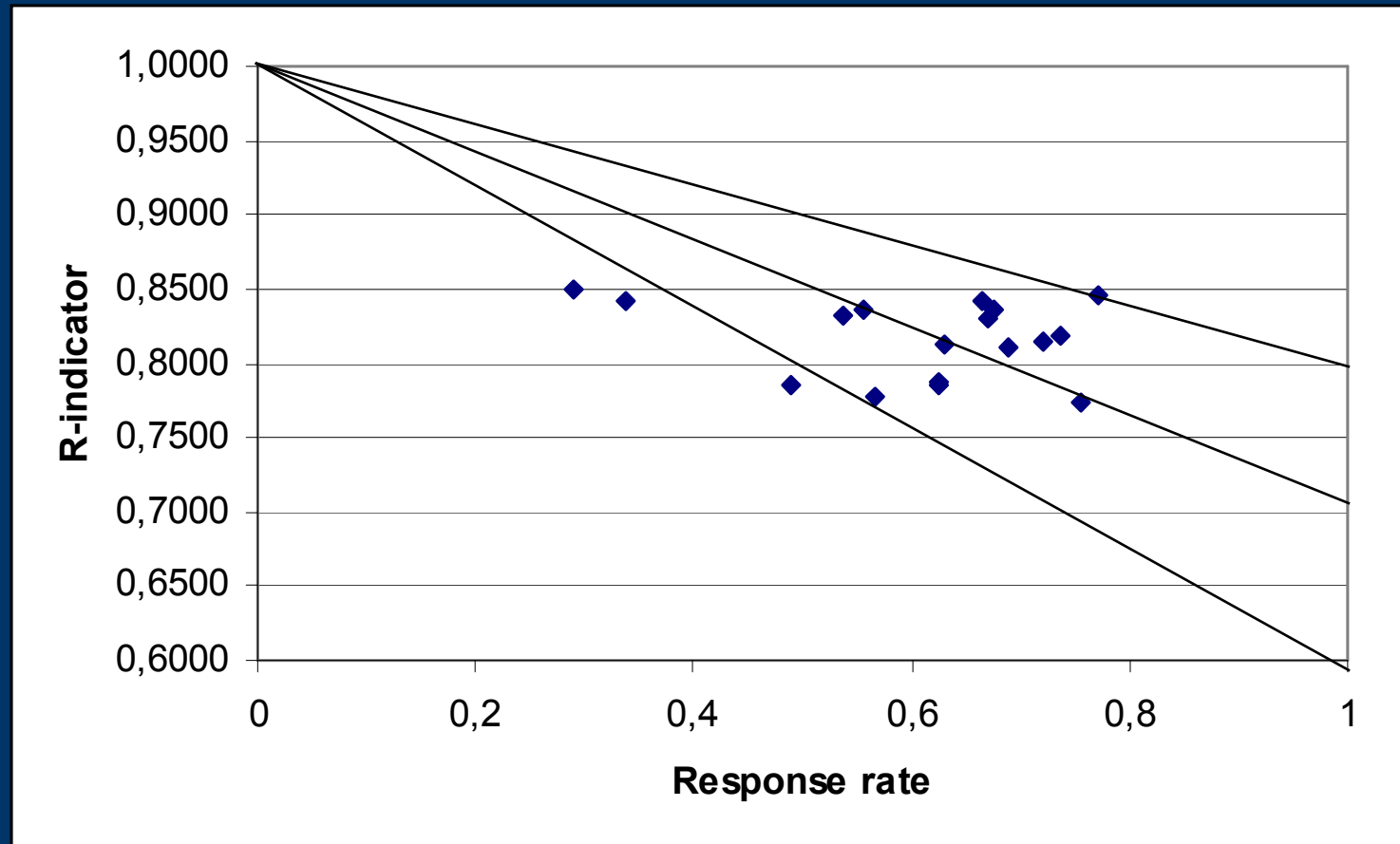
Voorbeeld 3

Responskansen zijn geschat met alleen leeftijd

| | Respons | Steekproef | R | Vertekening |
|------------------------|---------|--|-------|-------------|
| Panel 1 - zelfselectie | 0,63% | Nee | 99,5% | 19,0% |
| Panel 1 - steekproef | 0,63% | 0-20: 6,3% 20-65: 11,4% 65+: 25,0% | 99,5% | 19,0% |
| Panel 2 - zelfselectie | 0,75% | Nee | 99,8% | 5,6% |



Maximale systematische fout



Discussie

- Behoeftte aan een eenduidige definitie van representativiteit
- Behoeftte aan een heldere indicator voor afwijking van een representatieve respons
- Meet de samenstelling van respons naar een vaste verzameling demografische en sociaaleconomische kenmerken (gouden standaard)

