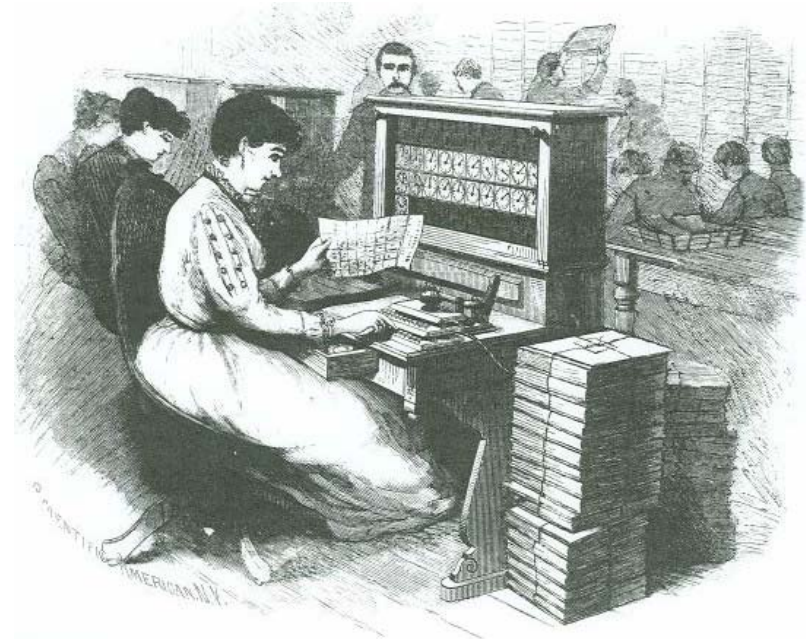


Surveys voor officiële statistieken

Betrouwbaarheid tot elke prijs?

Jelke Bethlehem



Centraal Bureau voor de Statistiek
Sector Methodologie



Officiële statistiek

- Kenmerken
 - Verzamelen, bewerken en publiceren van statistieken ten behoeve van praktijk, beleid en wetenschap.
 - De kwaliteit van de statistieken moet onomstreden zijn.
 - Feitelijke en onpartijdige informatie die door alle belangengroepen wordt geaccepteerd.
 - Cijfers moeten snel beschikbaar zijn.
 - De statistieken moeten de samenhang tussen allerlei maatschappelijke ontwikkelingen zichtbaar maken.
 - Het statistische onderzoekprogramma (van het CBS) wordt jaarlijks vastgesteld door de Centrale Commissie voor de Statistiek.
 - Veel statistieken komen voort uit Europese richtlijnen.



European Statistics Code of Practice



- Principle 6: Impartiality and objectivity
 - Statistical authorities must produce and disseminate European statistics respecting scientific independence and in an objective, professional and transparent manner in which all users are treated equitably.
- Principle 7: Sound methodology
 - Sound methodology must underpin quality statistics. This requires adequate tools, procedures and expertise.
- Principle 8: Appropriate statistical procedures
 - Appropriate statistical procedures, implemented from data collection to data validation, must underpin quality statistics.
- Principle 12: Accuracy and reliability
 - European statistics must accurately and reliably portray reality.

Het statistisch fundament

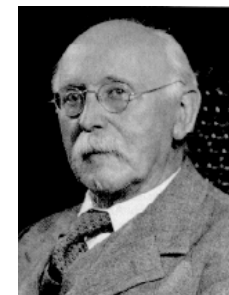
- Tot 1985 integraal onderzoek
 - 1000 voor Christus: bestuurlijke overzichten in China en Egypte.
 - 1086: Domesday Book (Willem de Veroveraar): gegevens over 13.418 nederzettingen in Engeland.
 - 1200: Quipucamayoc, de Inca-statisticus.



Rex ten in dño heret. Almar tenute in alod
de rege. E. Te p. v. hid. modo p. iii. hid. Tra. e. vi. car.
In dño. e. una car. 7 vi. uilli 7 un bord. cu. m. car.
lbi. u. serui. 7 una haga in Rudingel. 7 u. piscarie
de. vii. sol. 7 vi. den. 7 xx. ac. Tra. Silua de. lxx. poro.
T. h. e. uatb. c. sol. 7 post 7 m. 1. sol. In. h. e. h. d.

Het statistisch fundament

- Vanaf 1985 steekproeven
 - 1654: Pascal en Fermat, Kansrekening
 - 1895: Anders Kiaer (Noorwegen)
Representatieve Methode
 - 1926: Arthur Bowley (Engeland)
Steekproeven loten
 - 1934: Jerzy Neyman (Polen)
Betrouwbaarheidsinterval
 - 1952: Horvitz-Thompson-schatter



Het statistisch fundament

- Belangrijke eigenschappen
 - Man-made randomization: De onderzoeker heeft controle over het selectieproces.
 - Er hoeven geen oncontroleerbare modelveronderstellingen te worden gemaakt.
 - Als de selectiekansen bekend zijn, kan altijd een zuivere schatter worden geconstrueerd (Horvitz-Thompson).
 - De nauwkeurigheid van de uitkomsten kan worden gekwantificeerd (variantie, standaardfout, betrouwbaarheidsinterval)



Het statistisch fundament

- The proof of the pudding ...

Variabele	Uitslag	Kennisnet	RTL4	SBS6	Ned 1
Steekproef		17,000	10,000	3,000	1,200
CDA	44	29	24	42	42
LPF	8	18	12	6	7
VVD	28	24	38	28	28
PVDA	42	13	41	45	43
SP	9	22	10	11	9
GL	8	26	9	6	8
D66	6	4	7	5	6
Andere	5	14	9	7	7
MAD		12,5	5,3	1,8	0,8

Zes kwaliteitsdimensies

- **Relevantie.**
 - Is dit wat de gebruikers willen hebben?
- **Nauwkeurigheid.**
 - Hoe dicht liggen de cijfers bij de werkelijkheid?
- **Tijdigheid & punctualiteit.**
 - Hoe actueel zijn de cijfers?
 - Wordt de geplande publicatiedatum gehaald?
- **Toegankelijkheid en duidelijkheid**
 - Kunnen gebruikers de cijfers eenvoudig verkrijgen?
 - Is er voldoende en begrijpelijke uitleg (metadata)
- **Vergelijkbaarheid**
 - Hoe vergelijkbaar zijn de gegevens door de tijd en over verschillende regio's?
- **Samenhang**
 - Combinatie met ander onderzoek mogelijk?

Het speelveld

- Minder geld
 - Beschikbare budget voor het maken van goede statistieken neemt steeds verder af.
- Meer vraag
 - De roep om meer en steeds gedetailleerder informatie wordt steeds groter.
- Minder lastig vallen
 - De roep om vermindering van de administratieve lastendruk wordt steeds groter.
 - Geen gegevens verzamelen als ze al elders beschikbaar zijn.
 - Verschuiving van primaire waarneming (surveys) naar secundaire waarneming (registers).

Publiek onderzoek-moe



... ik doe 'n onderzoek naar onderzoek-moeheid ..."

Secundaire waarneming (registers)

- Voordelen
 - Geen kosten voor dataverzameling.
 - Integraal, geen steekproeven.
 - Vrijwel geen unit non-respons.
- Nadelen
 - Geen controle over de dataverzameling
 - Gegevens voor andere doeleinden door andere instanties op andere tijdstippen verzameld met andere definities.
 - Registers niet perfect. Dekkingsproblemen, meetfouten en item non-respons.
 - Kwaliteitscontrole is lastig.
- Uitdaging
 - Het combineren van gegevens uit verschillende bronnen.



Primare waarneming (surveys)

- Web-surveys
 - Eenvoudige, snelle en goedkope toegang to grote groep respondenten.
 - Kun je CAPI- en CATI- surveys vervangen door web-surveys?
- Problemen:
 - Onderdekking (niet iedereen heeft internet).
 - Zelf-selectie
 - Afwezigheid interviewer (voordeel en nadeel)
 - Controles in de vragenlijst niet mogelijk?
- Mixed-mode surveys
 - Web-survey als een van de modes
 - Sequentieel: beginnen met goedkoopste mode.
 - Simultaan: model voor beste mode.



Doe meer met minder data

- Gebruik van modellen
 - Maandcijfers voor de werkloosheid door toepassing van tijdreeksmodellen.
 - Schatters voor kleine gebieden, door combinatie van weinig of geen gegevens met modellen.
- Risico's
 - Hoe controleer je of het model correct is?
 - Ben je niet variantie aan het inruilen voor vertekening?
 - Hoe bepaal en communiceer je de nauwkeurigheid van je uitkomsten?
 - Er blijft behoefte bestaan aan 'echte' cijfers voor andere analyse.



Uitdagingen

- Data
 - Kunnen surveys worden vervangen door registers?
 - Goedkopere dataverzameling, met behoud van de kwaliteit van de uitkomsten.
 - Kunnen CAPI en CATI worden vervangen door CAWI?
 - Is er een rol voor mixed-mode surveys?
- Modellen
 - Kun je met modellen en weinig data toch nauwkeurige uitspraken doen?
- Publicaties
 - Hoe houdt je de productie van statistieken transparant?
 - Hoe leg je het uit aan de gebruikers?

