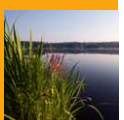


# Urvalsundersökningar – vetenskapliga aspekter och internationella utblickar

Jörgen Brewitz  
Processansvarig  
Utforma och utvärdera

Surveyföreningen, februari 2016





# Inferensperspektiv

---

- Inferens
  - att dra slutsatser om populationsstorheter utifrån observerade data, som fås från t.ex. ett urval eller ett register
- Inferensperspektiv
  - statistik är inte aritmetik, dvs. är inte exakt utan handlar om skattningar
  - statistiska storheter skattas utifrån teori och ramverk (vetenskaplig grund) med beaktande av osäkerhetskällor





# Sannolikhetsurval

---

- Urval draget med sådan slumpmekanism att det finns chans för alla objekt i den population som statistiken ska spegla att komma med i urvalet
- Storleken på den chansen (sannolikheten) ska vara känd





# Varför sannolikhetsurval?

---

- Billigare och mindre uppgiftslämnar-  
börda än totalundersökningar
- “Representativt”
- Transparent
- Ger tilltro till statistiken
- Osäkerhetsmått kan beräknas

Undantag därifrån då starka skäl





# Varför ifrågasätts sannolikhetsurval?

---

- Mätfelsproblem
- Ramtäckningsproblem
- Bortfallsproblem (svarsmängden är inte ett sannolikhetsurval)
  
- Önskemål om snabbare och billigare statistik
- Tillgång till data på internet





## Svarsandelar i några SCB-undersökningar

---

- Individ- och hushållsundersökningar

- Arbetskraftsundersökningarna: 60 %
- Undersökningar av levnadsförhållanden: 56 %
- Nationell folkhälsa: 49 %

- Företagsundersökningar

- Kortperiodisk sysselsättningsstatistik: 89 %
- Skörd av spannmål, trindsäd och oljeväxter: 92 %





# Svarsandelar i USA

---

- Officiell statistik
  - American Community Survey: 98 % (vägd)
  - Current Population Survey: 88 %
  - Survey of Income and Program Participation (wave 1): 81 %
- Marknads- och opinionsundersökningar
  - Ofta 10–20 %





## Svarsandelar i Sverige

---


- Marknads- och opinionsundersökningar via telefonintervjuer: **Ofta 20–40 %**
- Marknads- och opinionsundersökningar via sannolikhetsbaserade webbpanelundersökningar: **3–8 %**
- Marknads- och opinionsundersökningar via undersökningar från självrekryterade webbpaneler: **Inte definierbar svarsandel!**





# Icke-sannolikhetsurval

---

- Alla urval som *inte* är sannolikhetsurval
- 
- Det finns inget etablerat ramverk för icke-sannolikhetsurval
    - inte reproducerbara
    - oklara kvalitetsmått
    - inferens beroende av modellantaganden
  - Exempel: Självselekterade webbpaneler, river sampling, kvot(a)urval, snöbollsurval, på-plats-urval, experturval (bedömningsurval), deterministiska urval, crowd sourcing, cut-off-urval
  - Behövs modeller vid både urval och skattning





# Webbundersökningar

---

- ✓ Traditionella undersökningar (sannolikhetsurval) med webbenkät
  - ✓ Webbpanelundersökningar
  - ✓ Engångsundersökningar (river sampling o.d.)
- 
- Webbpanelundersökningar utgör *en* grupp av webbundersökningar
  - Knuten sitter i selektionseffekten, inte i valet av insamlingsmetod!



# Webbpanelundersökningar – olika typer

---

- I. Webbpanelerna är rekryterade utifrån sannolikhetsurval (men har stora bortfall)



Kan ses som "traditionella" undersökningar

- II. Webbpanelerna är självrekryterade (och har odefinierbara bortfall)



Saknar vanlig sannolikheteoretisk grund





# Vad säger forskningen – teori?

---

- I. Ambitiösa webbpanelundersökningar baserade på sannolikhetsurval har en vetenskaplig grund
  
- II. Webbpanelundersökningar baserade på självrekryterade paneler saknar en solid vetenskaplig grund (hittills)



# Vad säger forskningen – empiri?

---

- Empiriska resultat tyder på att undersökningar som inte baseras på sannolikhetsurval ger osäkra statistikvärden (med undantag för väljarbarometrar, möjligen)
  - ✓ Yeager et al., 2011 (Stanford-studien)
  - ✓ Scherpenzeel & Bethlehem, 2011
  - ✓ Tourangeau, Conrad & Couper, 2013





# Fallet *väljarbarometrar*

---

- Väljarbarometer: *surveying* eller *modelling*?
- En huvudfråga
- Partivägning
- Kalibrering mot valresultat eller andra undersökningar
- Delvis enklare än många andra typer av undersökningar!



# AAPOR:s rekommendationer

---

- Webbpanelundersökningar (baserade på icke-sannolikhetsurval) bör undvikas om syftet är att få tillförlitliga skattningar av populationsstorheter
- Webbpanelundersökningar kan användas i vissa andra sammanhang, t.ex. för explorativa undersökningar
- Mer forskning behövs



# AAPOR förespråkar öppenhet

---

- För marknads- och opinionsundersökningar
- Transparens om undersökningsmetoder
- Öppenhet om vald metodik är väsentlig för webbpanelundersökningar







# En ansats från Danmarks statistik

---

- Började testa webbpanelundersökningar 2014
- Samlade e-postadresser (ca 16 000) till webbpanel från traditionella, sannolikhets-baserade undersökningar
- Lyckoundersökning: kumulativ deltagarandel 8 %
- Kalibrerade mot flera sociala och ekonomiska registervariabler
- Blev svårt att justera bort skevheter





## Kontrollerad inferens genom bättre modellering?

---

- Kunna räkna upp från urval till målpopulation
- Modellbaserad kalibrering?
- Bayesianska ansatser?





## Kontrollerad inferens genom extra information?

---

- **Kombinationsansatser:**  
webbpanelundersökning ↔  
traditionell sannolikhetsbaserad undersökning
- **Dataintegration**





# Urvalsmatchning

---

- Kvoturval
- Avancerad matchning (fall-kontroll)
  - Mer än demografiska variabler





# Propensity score adjustment

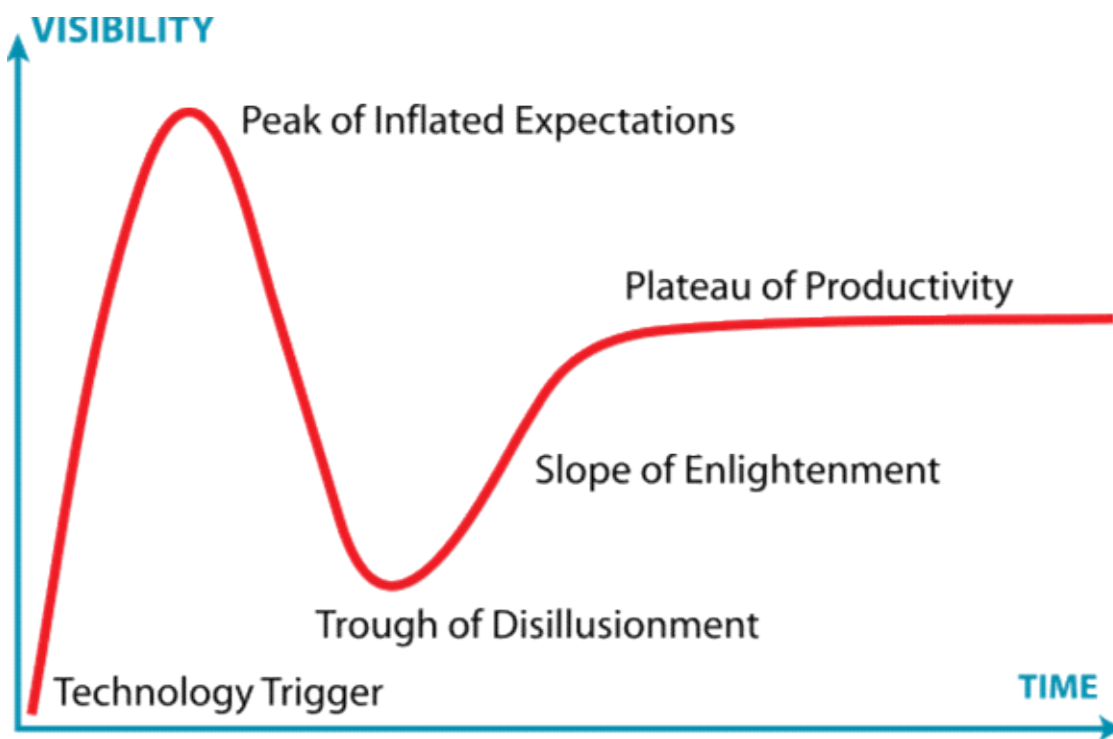
---

- Referensurval
- Psykografiska data
- Ett slags kalibreringsskattning
- Hög varians



# Annat än urvalsundersökningar?

- Registerbaserad statistik
- Big Data



Källa: Wikipedia, Gartner's hype cycle

# Framtida undersökningsmetodik

---

- En mångfald av metoder
- Adaptiv design
- Dataintegration
- ”Datafiering” (digitalisering)
  
- Inferens
  - Designbaserad
  - Modellbaserad
  - Algoritm-baserad (?)



# Några referenser för den intresserade 1

---

- Baker et al. (2010). *Research synthesis AAPOR report on online panels*. *Public Opinion Quarterly*, 74, 711–781.
- Baker et al. (2013). *Summary report of the AAPOR task force on non-probability sampling*. *Journal of Survey Statistics and Methodology*, 1, 90–143.
- Bethlehem (2010). *Selection bias in web surveys*. *International Statistical Review*, 78, 161–188.
- Bethlehem, J.G., och Cobben, F. (2013). *Web Panels for Official Statistics?* Invited paper presented at the 59th ISI World Statistical Congress, 2013, Hong Kong, China.
- Beulens et al. (2012). *Shifting paradigms in official statistics. From design-based to model-based to algorithmic inference*. Discussion paper, Statistics Netherlands.
- Callegaro et al. (2014). *Online panel research: A data quality perspective*. John Wiley & Sons, Ltd.





## Några referenser för den intresserade 2

---

- Lee (2006). *Propensity Score Adjustment as a Weighting Scheme for Volunteer Panel Web Surveys*. *Journal of Official Statistics*, 22, 329–349.
- Rivers (2007). *Sampling for web surveys*. White paper prepared from presentation given at the 2007 Joint Statistical Meetings, Salt Lake City, UT.
- Scherpenzeel and Bethlehem (2011). *How representative are online panels? Problems of coverage and selection and possible solutions*. In Das, Ester, Kaczmirek (Eds), *Social and behavioral research and the internet: Advances in applied methods and research strategies* (105–132). New York: Routledge.
- Tourangeau, Conrad & Couper (2013). *The science of web surveys*. Oxford University Press.
- Yeager et al. (2011). *Comparing the accuracy of RDD telephone surveys and internet surveys conducted with probability and non-probability samples*. *Public Opinion Quarterly*, 75, 709–747.