

Trender på surveyområdet

Surveyföreningen 16 oktober 2024

Dan Hedlin, Statistiska institutionen



Stockholms
universitet

Datarevolutioner

”harmonisering” av klockor i Sverige

- Folkräkningar, folkbokföring
- Den tidiga, tyska statistiken innehöll inte siffror
- Keynesianism krävde statistik
- Företag började samla in data i stor omfattning
- Evidensbaserade beslut
- Nu mera *datacratic*?

Början på 2010-talet

- United Nations Fundamental Principles of Official Statistics (2014)
- AI fick väsentligt större prestanda
- Digitalisering. Ledde till att det finns många typer av data
- Öppna data

Trender nu



- Digitalisering. Stora datamängder
- Olika datatyper (bilder, ljud, text, *spatio-temporal*, "distribuerade data")
- Allt ska mätas
- Kanske mer konfidentiella data släpps, med röjandekontroll (man lägger på störningar i data)
- ❖ Diversifiering (livsstilar, ekokammare)
- ❖ Människor använder flera apparater och programvaror
- ❖ För många undersökningar
- ❖ Folk är mindre fogliga
- Internationella undersökningar (inte bara översättning)
- AI

Vad är på väg ut?

Idealet som är på väg ut

Eller den gamla bilden snarare än idealet?

- Samma frågor, samma språk, samma datainsamlingsätt för alla objekt
 - Samma urvalsdesign och estimation i alla grupper
 - Synen att i registerbaserad statistik ska det vara så lite som möjligt mellan registret och statistiken
- Jag har sagt ett tag nu att telefonundersökningar är på väg ut

Urvalsundersökningar kommer **INTE** att försvinna

1. helt icke-sannolikhetsurval problematiska
 2. surveyer i industrin, miljö och i beräkningar
- Spridningen av covid

Forskning nu

- Kombinera sannolikhetsurval och alternativa datakällor
- Finns litteratur om inferens på icke-sannolikhetsurval, och diagnostiska metoder
- Maskininlärning
- $\bar{y} = \frac{1}{N} \sum_S y_k + \sum_{U-S} \hat{y}_k$ ← prediktion
- Men neurala nätverk är inte överlägsna på tabelldata
- Inferens baserad på system av undersökningar (*systems approach*)

Varför statistiken måste baseras på vetenskap

- Någon annan gång



Trender hos statistikproducenter

Stora, statliga statistikproducenter

- Snabbare, billigare, lägre uppgiftslämnarbörda, mindre redovisningsgrupper
- Alternativa datakällor
- Mer?

Framtidens utbildning

- Universitetsutbildning måste vara bred:
Tidsserieanalys, ekonometri, maskininlärning, generaliserade linjära modeller, grafisk framställning av data etc
- Grunderna försvinner inte: urval, inferens, bortfall. Även datainsamlingsätt, och datainsamling
- SCB:s vetenskapliga råd: universitet undervisar inte om *missing values*
- Vi har ett pedagogiskt problem i undervisningen. Design-baserad inferens mot modell-baserad inferens. Båda behövs
- VVS



Vad fattas i utbildningen?

- Knyta ihop det jag nämner under "forskning nu"
- *Systems approach*
- Registerbaserad statistik
- Etiska och juridiska frågor