



Statistikproduktionssystem (SP) 4.0

Användning av "nya" datakällor och bibehållande av kvaliteten i statistiken

Thomas Laitila
Örebro universitet och SCB

Presentation, Surveyföreningens kvalitetsseminarium 2022



facebook.com/statistiskacentralbyranscb



[@SCB_nyheter](https://twitter.com/SCB_nyheter)



[statistiska_centralbyran_scb](https://www.instagram.com/statistiska_centralbyran_scb)



www.linkedin.com/company/scb





Innehåll

1. Statistik och osäkerhet
2. Inferensskolor
3. Modellbaserad inferens
4. Hur använda “nya datakällor”
5. Design för skattning av hushållens konsumtion



Table: Change (%) in number of employees by industry.

Industry	Estimate		Industry	Estimate
<i>Frmng/Hntng</i>	6,0		Transport.	-0,5
<i>Electricity</i>	5,6		Banking	1
<i>Construction</i>	8,4		Real Estate	4,7
Retail	4,7		<i>Education</i>	6,4
<i>Hotel</i>	15,4		Health Care	4,7

Source: Statistics Sweden, AM 63 SM 0703



Table: Change (%) in number of employees by industry.

Industry	95% CI		Industry	95% CI
<i>Frmng/Hntng</i>	-2,5--14,5		Transport.	-3,5--2,5
<i>Electricity</i>	-0,6--11,8		Banking	-2,5--4,5
<i>Construction</i>	4,5--12,3		Real Estate	1,2--8,2
Retail	2,2--7,2		<i>Education</i>	0,3--12,5
<i>Hotel</i>	8,1--22,7		Health Care	-1--10,4

Source: Statistics Sweden, AM 63 SM 0703



Chatterjee, 2003, p. 30

Angående kvantifiering av osäkerhet:

"A specialty of the statistical way of induction...",



Inferensskolor inom Statistik

- Randomiseringsteorin
 - Slump tillförs via randomisering
 - Full kontroll över urvalssannolikheter
- Frekventistisk teori
 - Observationer antas genererade via ett slumpmässigt försök
 - Ingen kontroll över det slumpmässiga försöket



Inferensskolor inom Statistik

- Bayesiansk teori
 - Observationer antas genererade via ett slumpmässigt försök
 - Ingen kontroll över det slumpmässiga försöket
 - Apriori-kunskap införs genom att betrakta parametrar som slumpmässiga.
 - Tillskriver olika parametervärden olika sannolikheter

Modellbaserad inferens

- För både Bayesianisk och frekventistisk teori
 - Införandet av ett antagande om slumpmässighet innebär modellbaserad inferens
 - Inferensen bygger på en statistisk modell
 - Inferens till modellen
 - Extrapolering till inferens om populationen
- Randomiseringsteorin
 - Slumpmässighet genereras via urvalsförfarandet
 - Inget antagande om slumpmässiga variabler
 - Inferens direkt till populationen

Hur använda nya datakällor?

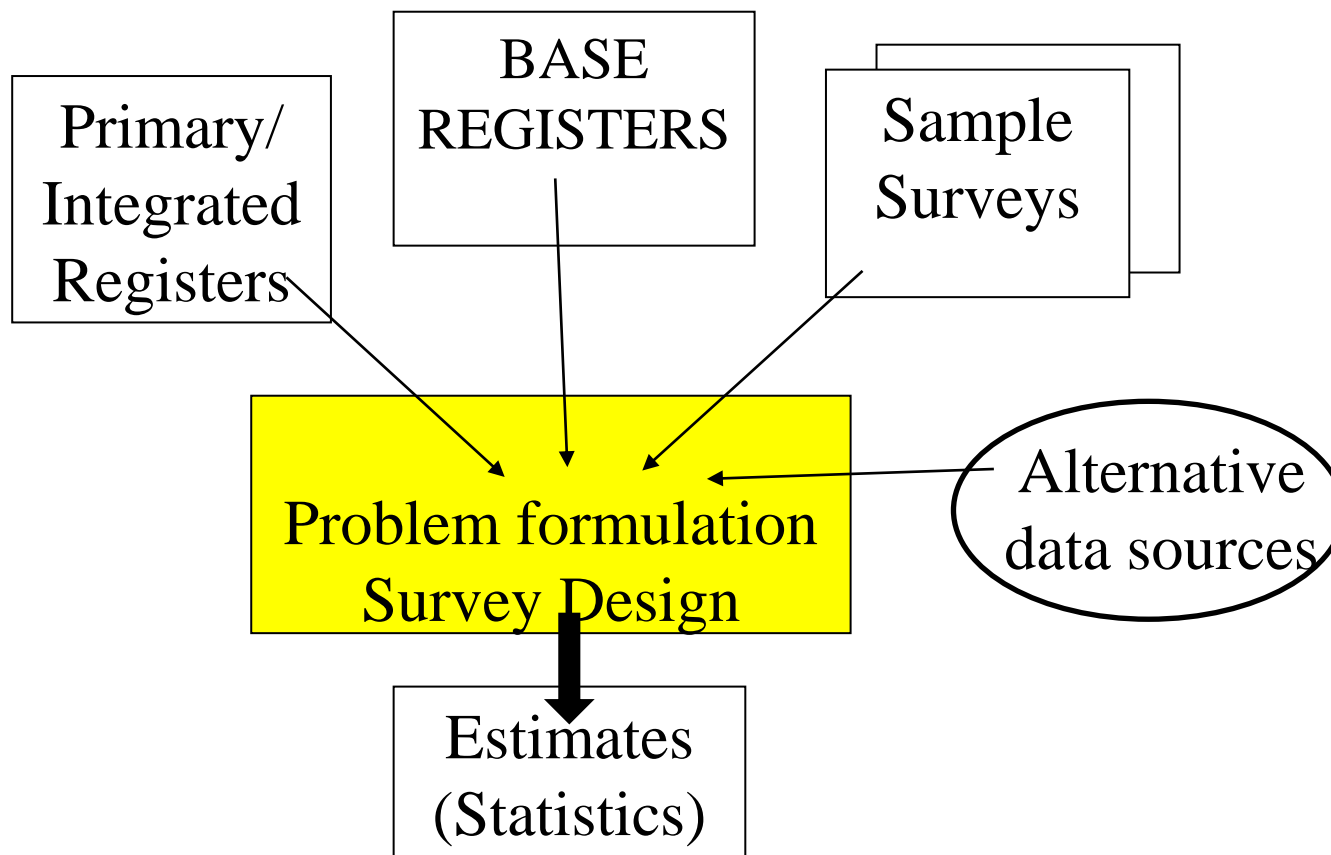
- Traditionella datakällor
 - Urvalsundersökningar
 - Register

- Nya datakällor
 - Big Data
 - Nya administrativa data
 - Observationsdata, sekundärdata, forskningsdata, etc.
 - Kräver användning av modellbaserade ansatser

- Hur bibehålla kvalitetsaspekter som följer med randomiseringsteorin?



Statistics Production System 4.0





Hur använda nya datakällor?

- Utveckla ny design av undersökningar där nya datakällor används tillsammans med de traditionella
- Exemplet Turiststatistik
- Exemplet Platsbanksdata



En ny design för skattning av hushållens konsumtionsutgifter



Uppgiften

- Utveckla en ny design för skattning av
 - Hushållens totala konsumtion på produktnivå
 - För NRs och KPIs behov

Thomas Laitila,
Surveyföreningens
kvalitetsseminarium





Huvudplanen

- A) Uppskatta total försäljning på produkt X bransch-nivå
- B) Skatta hushållens andelar av branschernas försäljning
- Kombinera A and B
- C) Lägg till nödvändiga specialundersökningar för bibehållen kvalitet – (osäkerhetsintervall)

A – försäljning

Försäljningsstatistik				
	Bransch 1	Bransch 2	Bransch 3	Summa
Livsmedel	100	0	20	120
Kläder	50	40	10	100
Möbler	45	5	90	140
Summa försäljning	195	45	120	360

Totalt 150 produktgrupper och 92 branscher

B – Hushållens inköp

Transaktionsdata – Försäljning uppdelat i privat konsumtion och övrig konsumtion - Andel (%)			
	Bransch 1	Bransch 2	Bransch 3
Privat konsumtion	80	30	50
Övrig konsumtion	20	70	50
Totalt	100	100	100

A och B kombinerat

Skattning av privat konsumtion på produktgruppsnivå				
	Bransch 1	Bransch 2	Bransch 3	Summa
Livsmedel	80	0	10	90
Kläder	40	12	5	57
Möbler	36	2	45	83
Summa	156	14	60	230

A och B kombinerat

Skattning av privat konsumtion på produktgruppsnivå				
	Bransch 1	Bransch 2	Bransch 3	Summa
Livsmedel	80	0	10	90
Kläder	40	12	5	57
Möbler	36	2	45	83
Summa	156	14	60	230

Lägg till annan tillgänglig information för bättre fördelning av privat konsumtion över produktgrupper inom respektive bransch.



Osäkerhetskällor

- Täckning – försäljning på detaljerad produktnivå (KPI, vikter)
- Täckning – transaktioner (kontantinköp, kreditkort över nätet)
- Fördelning av konsumtion över produkter inom de olika branscherna

Osäkerhetskällor

- Täckningsproblemen kan studeras med urvalsundersökningar
 - Grupp av återförsäljare som inte är representerade i använda datakällor
 - Hushållens inköp och betalningssätt



Osäkerhetskällor

- Fördelning av konsumtion på produkter är svårare
 - Sikta på bra skattningar på produktgruppsnivå
 - Enklare undersökning om hushållens utgifter – ex.vis den nya HUT-undersökningen

Osäkerhetskällor

- Potentiella möjligheter i studier av komponenterna i KPI
 - Samband mellan vikter och prisförändringar
 - Samband mellan prisförändringar för produkter inom en produktgrupp
 - Reducering av antal komponenter att skatta i KPI



Osäkerhetsmått

- Två verktyg
 - Klassiska konfidensintervall
 - Känslighetsanalys



Tack!