

1974 skrev Dalenius om Total Survey Design: Vad är aktuellt i dag?

TRADITIONELLA SURVEYDATA OCH NYA DATAKÄLLOR: HUR KAN DE KOMBINERAS?

ANDERS HOLMBERG

2019-03-22



Statistisk sentralbyrå
Statistics Norway

"Ends and Means of Total Survey Design" (1974)

Tore Dalenius (1917 – 2002)

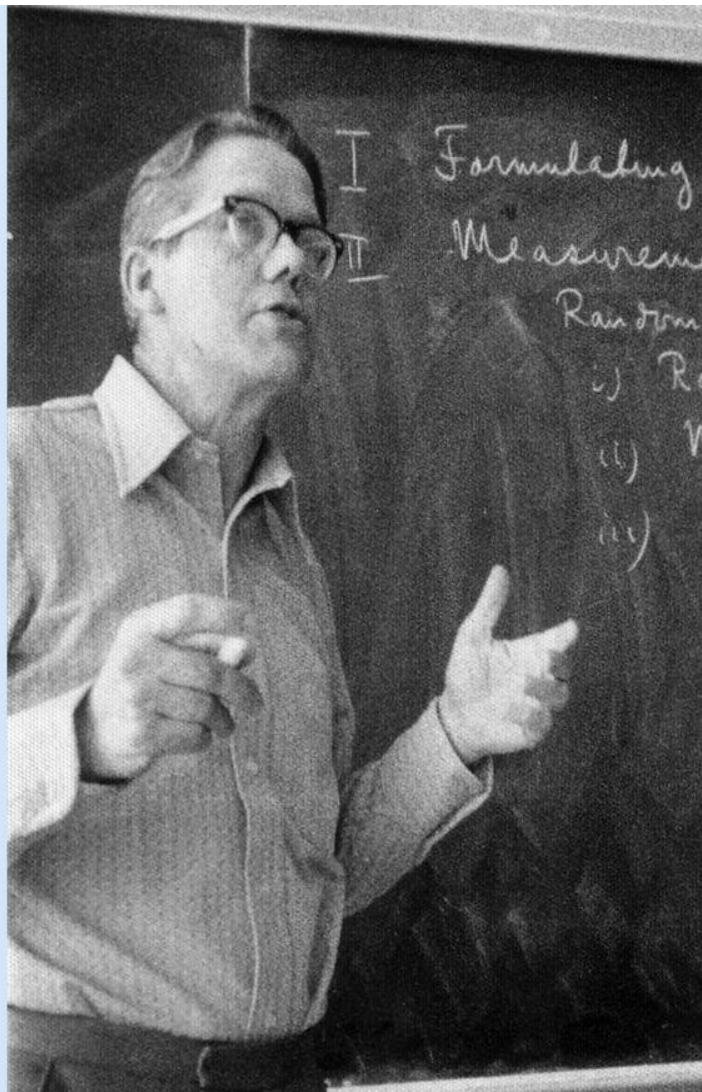
Worked at Statistics Sweden 1951 -1955

Active interest ever after

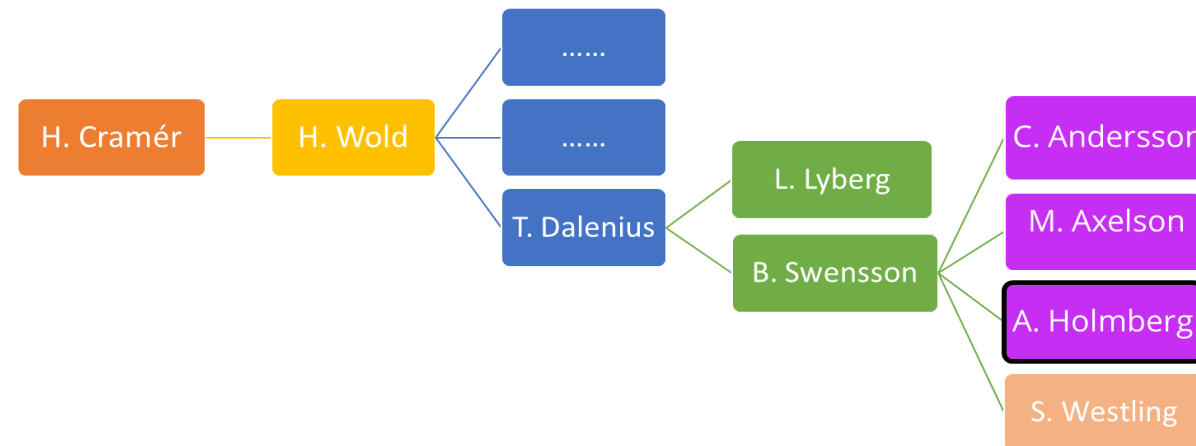
Criticism (sometimes unrealistic)

Important contributions, e.g. total survey error

Theory of Disclosure Control and confidentiality



- 1951 chef för SCB:s då nystartade Utredningsinstitut (UI)
- 1957 fil.dr. Uppsala
- 1962 en av grundarna av Svenska Statistikersamfundet
- 1965 professor i statistik vid Stockholms Universitet med inriktning mot officiell statistik, första i sitt slag
- från 1970-talets början verksam vid Brown Univ. ,USA
- 1983 Pension

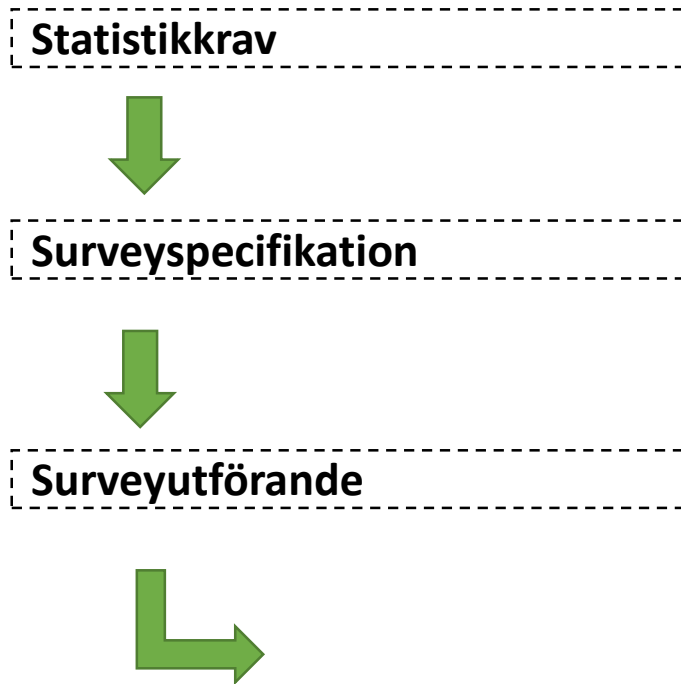


Surveydefinition

1. Väldefinierad mängd objekt (en population)
2. Populationen har flera mätbara egenskaper (attribut, variabler).
3. Vi vill beskriva populationen genom parametrar som är definierade utifrån egenskaperna i (2) och som kräver någon form av observation på objekten.
4. För att kunna observera behöver vi en operationaliserad definition av populationen. (en ram)
5. Ett sampel av enheter, punkter etc. väljs från ramen enligt en samplingdesign som specificerar sannolikhetsmekanismerna för att erhålla samplet. (Att välja alla i ramen är ett specialfall!)
6. Observationer görs på samplet i enlighet med en specificerad mätdesign: (En mätmetod och föreskrifter hur den ska användas)
7. På basis av gjorda observationer appliceras en plan för att beräkna storheter/kvantiteter som är estimerat av parametrarna i punkt (3).

Total survey design och totalfelet i en survey

”.....the conceptual and methodological prerequisite for a survey design aiming at an economic balance between sampling and non-sampling errors in a survey, a Total Survey Design.”



$\bar{Z}$: Idealt mål för en survey, s

$\bar{X}$: Definierat mål för en survey, s

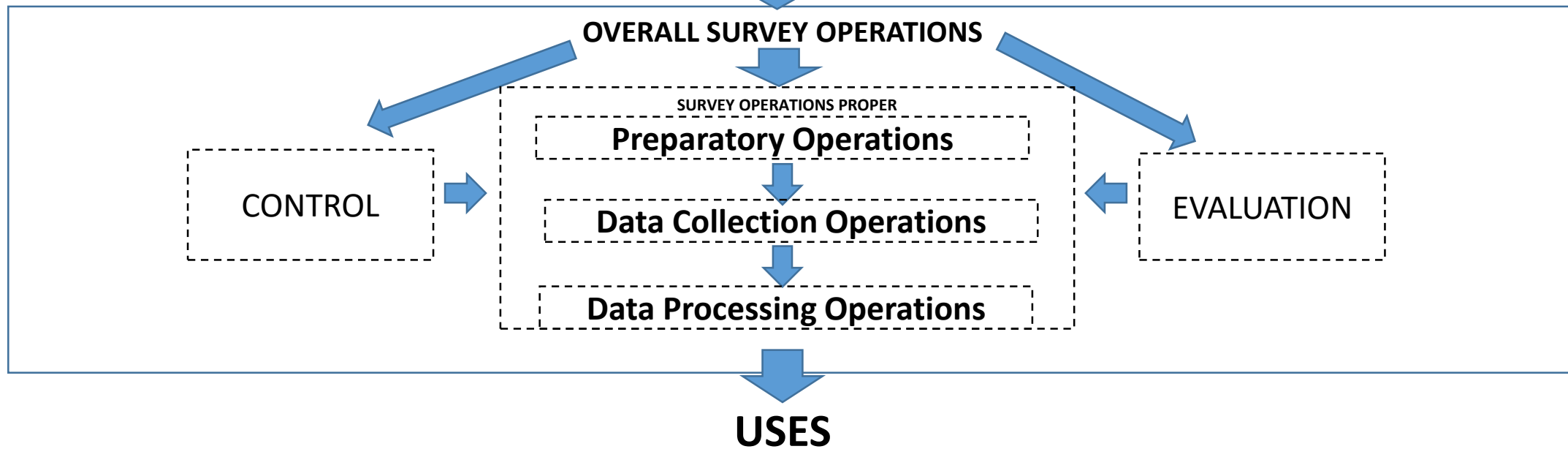
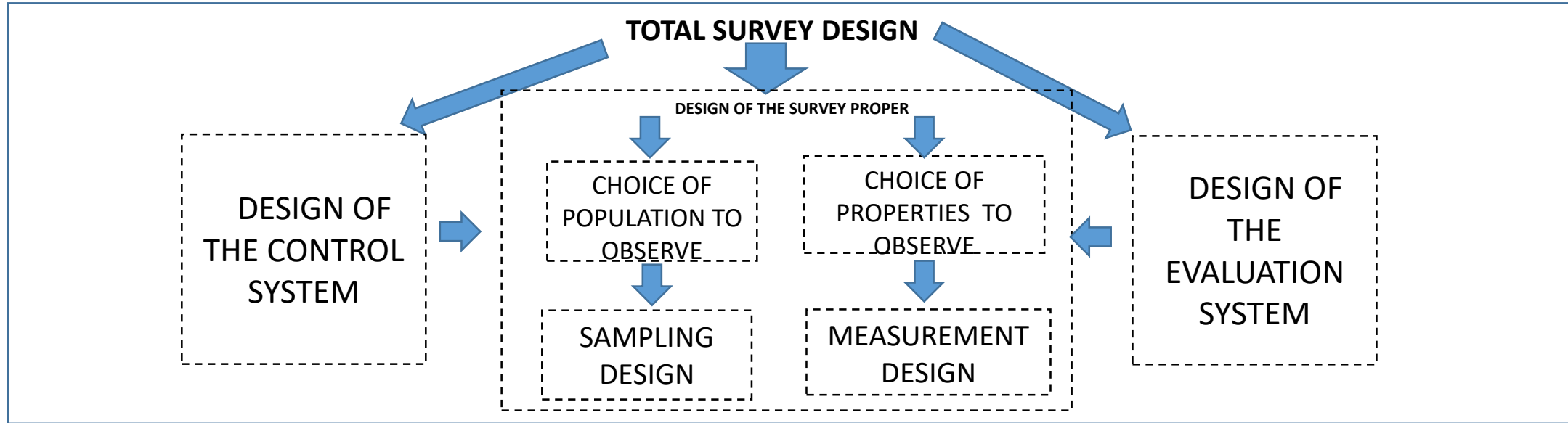
$\bar{y}$: Observerat resultat för en survey, s

$E = \bar{y} - \bar{X}$: Surveyfelet.

Nyttan $U(\bar{y}) = k/E$ ($k > 0$) vs. Kostnaden $C(\bar{y})$

DEMANDS FOR STATISTICS

ANALYSIS OF THE REQUIREMENTS



Upplägg

1. Grundläggande Surveyredskap

- I. Samplingteori och metoder
- II. Mätteori och metoder

2. Specialiserade Surveyredskap

- I. Surveymodeller (US Bureau of the Census' survey model)
- II. Modeller för randomisering av frågor/svar

3. Verktyg för felkontroll i surveyer

- I. Mål och medel med (kvalitets-)kontroll
- II. Metoder för kontroll av surveyspecifikationer
- III. Metoder för kontroll av surveyutförandet - Statistisk kvalitetskontroll
- IV. Metoder för kontroll av surveyutförandet – Specifika metoder

4. Verktyg för evalvering i surveyer

- I. Mål och medel med evalvering
- II. Analytiska evalveringsmetoder
- III. Experimentella evalveringsmetoder

Exempel Granskning / Editering (Dalenius 74)

The impact of the computer on today's editing practice has been especially large with respect to the increase in the speed by which the editing takes place, and the reliability of the editing; unlike clerks, the computer carries out routine operations exactly in accordance with the instructions given. There is, however, need for precautions: computerized editing is no better than determined by the rules applied, whether they are carried out fast and reliably or not. In view of the importance of this point, we will present here the philosophy which has been developed at the U.S. Bureau of the Census in this respect; this philosophy is summarized in Pritzker et al (1966).

Granskning / Editering (Dalenius 74)

First: Editing is no substitute for adequate original measurement. A corollary is that the best solution of the problem of controlling observational errors may be to prevent the respondents from making them in the first place.

Second: In spite of taking preventive action as just mentioned, deficiencies will occur, calling for such actions as editing changes, imputations and alterations. Such action should be governed by a conscientious endeavour to preserve the sense of the originally recorded observations insofar as possible.

Osv. Selektiv granskning, Evalvering av modeller och regler, Dokumentation och Transparens

Upplägg

5. Total Survey Design

- I. Strategi för Total Survey Design
- II. Kravidentifikation
- III. Utveckling av surveyspecifikationer

Systems Analysis and Design som arbetsmetod

- Relevanskriterie
- Mätbarhetskriterie
- Kriterie för Operationell utförbarhet
- Optimalitetskriterie

SSB och okonventionella datakällor

- Transaktionsdata
 - Betalnings- och Girotransaksjoner
 - Kassadata / (Scanner) data
 - Lojalitetskort / (Medlemsdata)
- Mobilposition- och CDR-data
(call detail records)
- Smartmeterdata, Sensorer
- Satellitdata



Statistisk sentralbyrå
Statistics Norway

Grafer og Nätverk: Schenkel & Zhang 2018

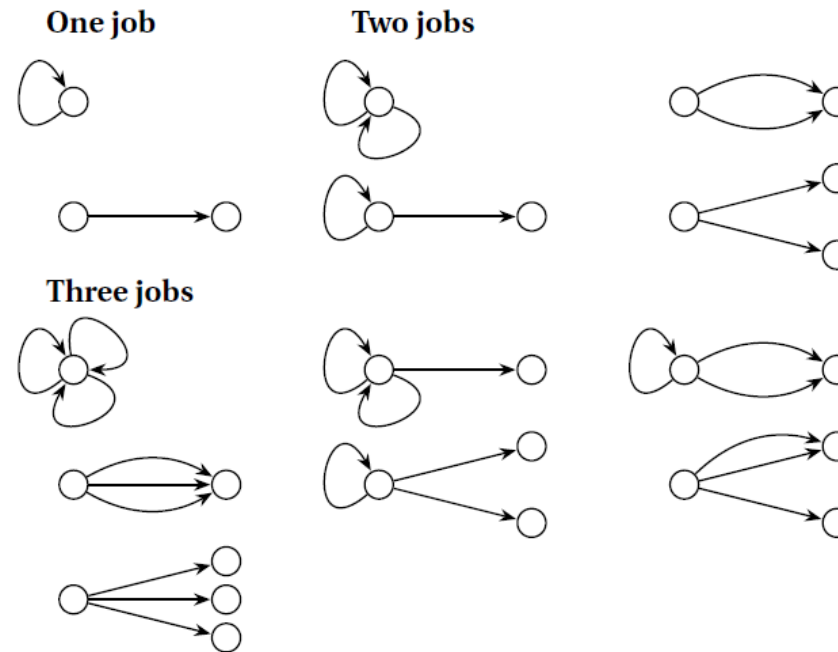


Figure 1: Commuting motifs for workers with maximum three jobs. Each node is a municipality, an edge goes from the municipality where the worker lives to the municipality where he/she works. A loop is thus that the worker lives and works in the same municipality.

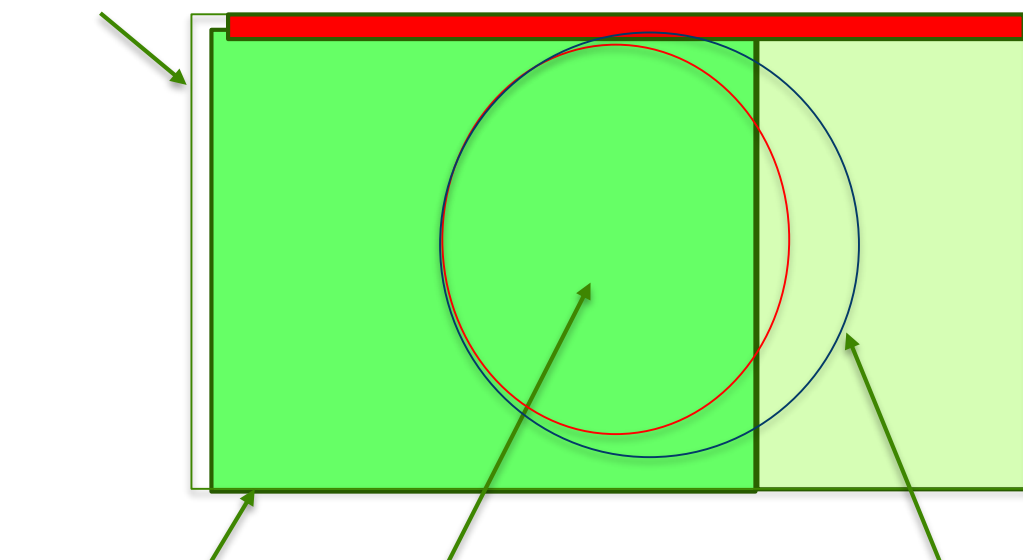
Kan källor berika varandra? (Zhang 2018)

Discrepancy between sources

	A-ordning	Telenor
Location	-/+	=
Responsiveness (timely)	-	=
Person/job attribute	$\approx$	-
Population coverage	=	-
[NB. Telenor: both under- and over-coverage]		
Job (relevance)	$\approx$	-/+
[e.g. sysselsatt i midlertidig fravær]		
Commuting	-	$\approx$
[e.g. pendlingsmønster: hvor ofte, når, fordeling flere jobber]		
[e.g. deltidsjobb: korte vs. færre dager?]		

Framtid: En annan designansats och en ny roll för surveyer?

Target
Population



Frame
Population

Response Set

(Designed Sample Data Set)

Ett ämnesområde för samhällsundersökningar

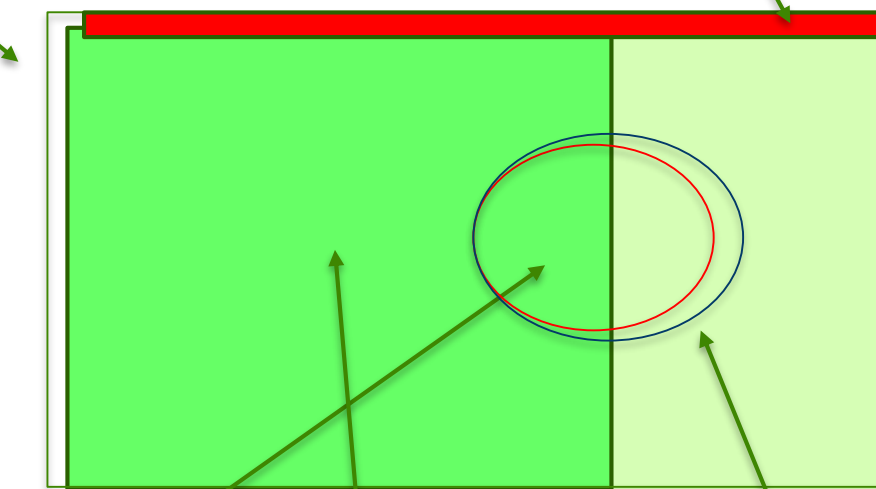
Grön: Delpopulation med små eller inga statusändringar.

Ljusgrön: Delpopulation med hög frekvens av statusändringar

Röd: Delpopulation som är speciellt svår att nå med samhällsundersökningar.

Framtid: En annan designansats och en ny roll för surveyer?

Target Population Storlekskattning med Återkommande Validerings- / audit surveys



Response Set

Designed Sample Data Set

**Administrative Data /
Andra källor**

Ett ämnesområde för samhällsundersökningar

Grön: Delpopulation med små eller inga statusändringar.

Ljusgrön: Delpopulation med hög frekvens av statusändringar

Röd: Delpopulation som är speciellt svår att nå med samhällsundersökningar.

1974 - 2019

Fortfarande högst aktuellt

- Grundteori och ramverket
- Ett helhetsperspektiv på surveyundersökningar, både enheter, variabler och operationalisering. Statistikproduktionsmodellen.
- Processkvalitet för att få statistikkvalitet (både kontroll- och evalveringsmetoder), betoning av relevans!
- Surveykvalitet / Datakvalitet: Bortfall, Editering

Saknas (i stort sett)

- Fler än en källa, fokus är på en datakälla som är 'kontrollerbar från' start till mål.
- Andra 'surveymodeller' än den designbaserade, frekventistiska modellen
- Fler sampling- och estimeringsmetoder.
- Mer teknologidrivnen mätmetodik och datainsamling
- Teknologikutveckling, metadata, (konfidentialitet)

TAKK

