

F2

Undersökningsplanering

Datakällor: officiell statistik, olika databaser, registerstatistik

Planeringen av en statistisk undersökning

- Tre huvudfrågor:
 - *Vem* ska undersökas?
 - *Vad* ska undersökas?
 - *Hur* ska undersökningen göras?

Grundläggande begrepp

- Innan vi kan göra uttalanden måste vi samla in data
- Vi har ofta inte råd/tid att undersöka en hel population
- Population
 - En mängd av enheter som vi vill studera egenskaper och attityder för.
 - Måste definieras och avgränsas i tid och rum
 - Populationsstorlek: N
- Element, enhet, individ, "unit", "object", "individual", "subject"
 - Människor, bilar, företag, olika händelser,...
- Stickprov, urval (sample)
 - En observerad delmängd av populationen
 - Stickprovsstorlek: n
- Variabel X , Y , Z ,...
 - En egenskap som kan variera mellan olika enheter i populationen

Planering av en undersökning

- Vi har ett *problem*
- Vi vill ha *mera kunskap* innan vi fattar beslut
- Vi vill kunna *bedöma kvaliteten* på det beslutsunderlag vi får

Beskrivande frågeställning

- Problemet avgör typ av undersökning
- Beskrivande frågeställning
 - Hur många (är positiva)?
 - Hur har andelen (tveksamma) ändrats sedan förra mätningen?
 - Hur skiljer sig åsikterna åt mellan könen? Mellan olika inkomstgrupper?
- Lämpligt med en survey (totalundersökning eller urvalsundersökning)

Urvalsundersökning eller totalundersökning?

- Man väljer ofta urvalsundersökning
 - Kan ge högre kvalitet
 - Totala felet kan bli mindre
 - Tid/råd att anstränga sig för att resultatet ska bli så riktigt som möjligt
 - Ex. fler påminnelser till bortfall

Förklarande frågeställning

- Förklarande frågeställning
 - Råder det orsakssamband mellan X och Y?
 - Dvs påverkar X Y eller finns det bakomliggande variabler Z som kan förklara det (numeriska) sambandet?
 - Hur stor är effekten av X på Y givet att alla andra påverkande variabler hålls konstanta?

Primärdata vs sekundärdataundersökning

- Antag beskrivande frågeställning
 - Finns uppgifterna redan insamlade, *sekundärdata*?
 - Leta i officiella statistiken, olika databaser, utredningar osv.
 - Primärdataundersökning
- Vem ska undersökas?
- Vad ska undersökas?
- Hur ska undersökningen göras?

Vem ska undersökas?

- Målpopulation
 - Den population som vi *idealt* vill undersöka, som vi vill dra slutsatser till
- Rampopulation
 - (Undersökningspopulation) den population som vi har en ram/ett register över
- Ram
 - Register eller annan förteckning över populationens enheter

Vem ska undersökas?

- Definiera målpopulationen genom att avgränsa den i tid och rum
 - Ex. Medlemmar i en förening, tidpunkt?
 - De som betalat årets medlemsavgift? Ingår de i målpopulationen? De finns antagligen kvar i registret (ingår i rampopulationen)
 - Nya medlemmar som ej hunnit bli registrerade? Bör ingå i målpopulationen, men finns ej med i rampopulationen.

Vem ska undersökas?

- Täckningsfel
 - Avvikelser mellan mål- och rampopulation
 - Över-och undertäckningsfel
 - Viktigt: man får inte definiera bort enheter i målpopulationen som man vet är tex svåränträffbara – risk för systematiska fel.
 - Ex personer med hemligt telefonnummer

Vem ska undersökas?

- Ramen kan vara
 - Register
 - Tidpunkter/intervall
 - Områden ("area sampling", områdesurval)
 - Tex alla stadens kvarter på en karta
- Register finns över
 - Befolkning, tex medlemmar, anställda
 - Folkbokförda i landet finns i RTB
 - RTB=registret över totalbefolkningen
 - Telefonabonnenter
 - Företag

Vem ska undersökas?

- Ex. Populationen "statistiker"
- Definition?
 - Medlem i någon statistisk förening?
 - "Arbetar" med "statistik"?
 - Krav på formell kompetens?
 - Statistiker enligt SSYK 96 (se KD sid 158)

Egenskaper hos en bra ram

- Identifiering av enheterna
 - För att kunna nå dessa
 - Namn, adress, tel.nr, e-mail
 - För att kunna samköra med andra register
 - Personnummer, organisationsnummer
- Aktuell
 - Inga brev i retur, hög täckningsgrad
- Kunskap om hur ramen är ordnad/sorterad
 - Ex efter bokstavsordning, ålder, bostadsområde

Egenskaper hos en bra ram

- Även viktigt att andra variabler finns i ramen som kan användas tex vid urvalet (tex som *stratifieringsvariabler*)
- Kunskap om hur många gånger varje enhet finns i ramen.
 - Vi kan bestämma chansen (sannolikheten) att bli utvald för varje enhet

Vad ska undersökas?

- Parameter
 - *Konstant* storhet som karakteriserar en variabel i *populationen*
 - *Ex. medelvärde, andel, total*
- Exempel
 - *Medelinkomst* hos en grupp anställda
 - *Andel* kvinnor i företaget
 - *Totala* utbetalda lönesumman varje månad
 - *N*=antalet anställda
- De storheter/parametrar som efterfrågas måste vara vägledande för vilka frågor som ska ställas och hur i frågeformuläret
- Utgå från VAD man vill undersöka

Hur ska undersökningen göras?

- Antag att vi bestämt oss för en urvalsundersökning
 - Slumpmässigt urval?
 - Icke-slumpmässigt urval?
- Olika urvalsmetoder: senare
- Antag slumpmässigt urval
 - Varje *urvalsenhet* har en *känd* sannolikhet (chans) *större än 0* att bli utvald

Urvalsenhet: elementet eller grupper av element

Sekundärdata

- Finns det statistik tillgänglig?
 - Finns svaren direkt?
 - Behöver nya bearbetningar göras?
- Datamaterial finns ofta sparade i register eller databaser.
- Kan även vara komplement till sekundärdata

Officiell statistik

- Officiell statistik
 - ”Skall finnas för allmän information, utredningsverksamhet och forskning. Den skall vara objektiv och allmänt tillgänglig.” (SFS2001:99)
 - Produceras av SCB samt 24 myndigheter
 - Ex Arbetsmiljöverket, Brottsförebyggande rådet, CSN, Finansinspektionen, Högskoleverket, Konjunkturinstitutet, Riksförsäkringsverket, Socialstyrelsen (se KD sid 105)
 - Allmänt tillgänglig, i elektronisk form
 - Objektiv
 - Skydd för personlig integritet enligt PuL
 - Kvalitetsdeklareras
 - Publiceras i Sveriges Officiella Statistik (SOS)
 - Finansieras via anslag, ej uppdragsfinansierad
 - Uppdelad efter kön

Var finns statistiken?

- Publikationer
 - Tidigare: tryckta publikationer
 - Numera: elektroniskt
 - www.scb.se
- Exempel på publikationer
 - Statistisk Årsbok (SÅ)
 - Årsbok för Sveriges Kommuner (ÅSK)
 - Statistiska Meddelanden (SM)
 - SCBs tidskrift Välfärd
 - Webbmagasinet artikulerat
- Andra exempel
 - GIS
 - Befolkningspyramid

Användning av registerdata

- Register: en fullständig förteckning över objekten i en viss objektmängd eller population.
 - Uppgifter kan pga kvalitetsbrister saknas för vissa objekt.
 - Uppgifter som objektens identitet ska finnas så att registret kan uppdateras och utökas med nya variabelvärden för respektive objekt.
 - För att bedöma användbarheten bör man veta hur det är upplagt och hur det hålls aktuellt
- Användning
 - Sammanställning av registerdata
 - Bas för undersökningar – som urvalsram
 - Komplement till primärdata

Registerbaserade undersökningar forts.

- Fyra basregister är grunden för SCB:s registersystem
 - Befolkningsregistret (RTB)
 - Aktivitetsregistret
 - Företagsregistret
 - Fastighetsregistret

Databaser

- LOUISE
 - Longitudinell databas för utbildning, inkomst och sysselsättning
 - Omfattar alla mellan 16-64 år varje år från 1990, från 1995 även alla över 64
- LINDA
 - OSU 300 000 personer 1994
 - Kompletteras med nya personer varje år
 - Inkomstförhållanden och förändringar för enskilda personer

För- och nackdelar med att använda sekundärdata (se KD s. 119)

- **Fördelar**
 - Billigare per element
 - Kan användas i kombination med primärdata
 - Minskad uppgiftslämnarbörda
- **Nackdelar**
 - Risk för inaktuella uppgifter
 - Undersökningens ursprungliga syfte ej detsamma
 - Ändrade definitioner på variablerna
 - Ändrade avgränsningar av population och element
 - Data ej sparade i ursprungligt skick
 - Underrapportering i administrativa register