

F2

Planering av en undersökning
Olika datainsamlingsmetoder
Olika slag av variabler

Förra gången (F1)...

- Vad är statistik?
 - Siffror i tabeller och diagram
 - Metoder
- Begrepp
 - Element, enhet, individ
 - Population
 - Variabel
- Icke-experimentella vs experimentella
- Beskrivande vs förklarande undersökningar

Förra gången (F1)...

- Olika slag av undersökningar
 - Beskrivande
 - Observation
 - Survey
 - Förklarande
 - Experiment
 - Prognoser
 - Register
 - Longitudinella
 - Kvasiexperimentella

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
- Två strategier
 - Expertgrupp
 - Statistisk undersökning
- HÄR: Vi ska göra en statistisk undersökning!

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 vs 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 svåranskräffbara – 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 och RTB-S
 - RTB=registret över totalbefolkningen
 - Telefonabbonnenter
 - 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*)
- 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. medelvärdet, tex medelålder, medellön, hos alla individer i populationen*

Medelvärde och totalvärde

Om vi har N st enheter i populationen, och X är variabeln, definieras medelvärdet som

$$\mu = \frac{x_1 + x_2 + \dots + x_N}{N} = \frac{\sum_{i=1}^N x_i}{N} = \frac{\sum x_i}{N} \quad (\text{"My"})$$

och totalvärdet som

$$\tau = \sum_{i=1}^N x_i = N \times \mu \quad (\text{"Tau"})$$

Andel

Andelen enheter som har en viss egenskap, tex andelen personer som har körkort i populationen

$$P = \frac{\sum_{i=1}^N x_i}{N}$$

där $x_i = 1$ om enheten har egenskapen
0 annars

P är ett specialfall av ett vanligt medelvärde

Varians och standardavvikelse

Varians

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N} \quad (\text{"Sigma-kvadrat"})$$

Standardavvikelse

$$\sigma = \sqrt{\sigma^2}$$

Parametrar forts.

- Exempel
 - *Medel*inkomst 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 survey, 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

Olika datainsamlingsmetoder

- Enkäter – till enskilda personer, företag
 - Postenkäter
 - Internet(webb)-enkäter
 - Gruppenkäter
 - Besöksenkäter
- Intervjuer
 - Besöksintervjuer
 - Telefonintervjuer

Olika datainsamlingsmetoder

- Datorstöd vid intervjuerna
 - CATI – computer Assisted Telephone Interviewing
 - CASI – Computer Assisted Self Interviewing
 - Svar via tonvalstelefon
 - Touchtone Data Entry (TDE)
 - Användning av elektroniska formulär
 - Electronic Data Interchange (EDI)
 - Skicka frågor/svar via disk
 - Direkt överföring av mätvärden från dator till dator
 - Överföring via e-post och/eller Internet
 - Problem
 - Tekniken
 - Integriteten

Olika datainsamlingsmetoder

- Bokföring eller dagbok över tex
 - Tidsanvändning
 - Inköp
 - Matvanor
- Direkta mätningar/observationer tex
 - Medicinska mätningar
 - "räkningar" av olika slag
 - Besökare på en hemsida, ett museum...
 - Antal djur (vargar tex)
 - Antal defekta enheter vid kvalitetskontroll

Val av datainsamlingsmetod

- Faktorer som kan påverka val av metod
 - Ämnet
 - Förväntat stort intresse, tex för barnomsorg
 - Välj billig metod (tex postenkät)
 - Känsligt
 - Enkät eller intervju med anonymitetsskydd
 - Ointressant ämne, övertalning krävs
 - Välj tex intervju, men kan bli dyrt

Val av datainsamlingsmetod forts.

- Faktorer som kan påverka...
 - Populationen
 - Äldre personer
 - Överväg besöksintervju
 - Företag
 - Gärna ny teknologi, internetenkät eller TDE
 - Låt företaget välja bland alternativa insamlingsmetoder
 - Tillgängliga resurser
 - Tid, pengar, datorstöd..
 - Postenkät billigast, kräver tid för påminnelser
 - Telefonintervjuer snabba
 - Internetenkäter, men problem finns
 - Intervjuer, intervjuare måste få utbildning, dyrt, tar tid

Val av datainsamlingsmetod forts.

- Faktorer som kan påverka...
 - Omfattningen av frågorna
 - Många frågor och svarsalternativ
 - Enkät bäst
 - Hoppfrågor
 - Om stor undersökning, överväg datorstöd så att rätt fråga besvaras av rätt person
 - Ofta bra med en kombination av olika datainsamlingsmetoder
 - Postenkät + telefonintervju
 - Internetenkät + postenkät/telefonintervju
 - Besöksenkät + telefonintervju/postenkät
 - Kan minska bortfall, få ner kostnader, minska uppgiftslämnarbördan, öka täckningsgraden (register över tex e-mailadresser ej komplett)

Olika slag av variabler

- En egenskap som kan variera mellan olika enheter i populationen
- Egenskaperna hos variablerna avgör
 - Val av läges- och spridningsmått
 - Val av mått på samband
 - Val av diagram

Olika slag av variabler forts.

- Vi vill ha *mätbara* variabler
 - Kvantitativa (numeriska)
 - Ålder, längd, vikt, inkomst,..
 - Kvalitativa (icke-numeriska)
 - Måste kodas, åsättas siffror
 - Kön, partisympati, rökare/ej rökare,...
 - Vissa måste operationaliseras innan användning
 - Ange hur de ska mätas
 - Bestäm indikatorer för begreppen
 - Ex. kvalitet, service, stress, attitydvariabler som kräver en skala

Datanivåer

- Variablerna innehåller olika mycket information: är på olika *datanivåer*
 - Olika bearbetningar/beräkningar kan göras för variabler på olika nivåer
 - Nominalskala (klassificering)
 - Ordinalskala (klass., rangordning)
 - Intervallskala (klass., rang., identiska skalsteg)
 - Kvotskala (klass., rang., id. skalst., absolut nollpunkt)
 - Variabler som har kodats i sifferkategorier innehåller inte mer information än innan

Olika slag av variabler forts.

- Vill ha variabler som mäts med hög *validitet*
- Mätinstrument
 - Våg, termometer, måttband, frågeformulär,...
 - Ett mätinstrument har hög *validitet* om det mäter det man avsett att mäta (inga systematiska fel)
 - ... hög *reliabilitet* om det har liten grad av slumpmässiga fel
 - En våg avsedd att mäta vikt ska inte variera från gång till gång om vikten ej förändras
 - Svaret på en attitydfråga som ges flera gånger till samma person ska inte variera
 - Men, finns alltid viss variation i svaren! Använd medelvärdet av upprepade mätningar för att förbättra reliabiliteten (minska osäkerheten i svaren)

Kvantitativa (numeriska) variabler

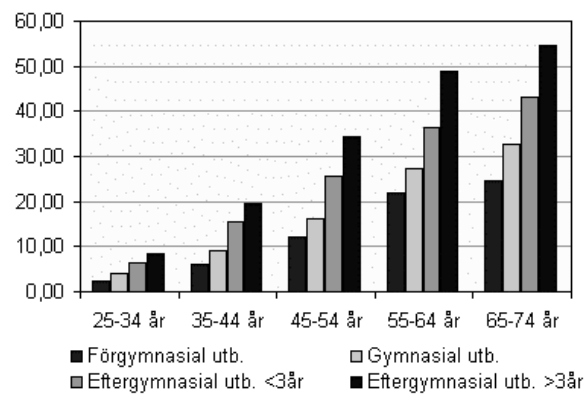
- *Diskreta* variabler
 - Kan bara anta ett ändligt antal värden
 - Kan vara heltal eller inte, positiva/negativa
 - Ex. X =antal barn, bilar,... i en familj
 X =poäng på en skrivning
 - Variabeln antar olika *variabelvärden*
 - 0 barn, 1 barn osv.
 - *Frekvenser*
 - Hur många familjer som har 0 barn, 1 barn osv.
 - Skilj på variabelvärden och frekvenser!

Diagram för kvantitativa diskreta variabler

- Stolpdiagram
- Stapeldiagram

Exempel på grupperat stapeldiagram

Andel i befolkningen med en nettoförmögenhet över 1 miljon kronor fördelad efter utbildningsnivå1 och ålder (25-74 år) 2005. Procent.



Kvantitativa *kontinuerliga* variabler

- Kan anta alla värden i ett intervall, teoretiskt *oändligt* många värden
 - Ex längd, vikt, ålder, temperatur,...
 - Värdena är i praktiken *avrundade*, men avgörande är de teoretiska egenskaperna
 - Diagram: Histogram

Bakgrundsvariabler vs undersökningsvariabler

- BV
 - Kön
 - Ålder
 - Utbildning
 - Inkomst...
- UV
 - Attityd
 - Sparande
 - Restid till universitet...
- Undersökningsplan:
 - Tänk igenom vilka BV som behövs i enkäten för att spegla frågeställningen med hjälp av UV
 - Tänk på att undersökningsvariabler i en viss undersökning kan vara bakgrundsvariabler i en annan.