
Vad är statistik?

Praktisk statistik kapitel 1

Kapitel 1 handlar om ...

- **Vad är statistik?**
- **Grundbegrepp**
 - population, individ, stickprov
 - variabel, datanivåer
 - olika urvalsmetoder

Vad är statistik?

- **"Siffror", deskriptiv statistik**
data, information, sammanställningar
- **Metoder**
Hur gör man ett slumpmässigt urval i en viss situation? Hur designar man ett experiment? Är skillnaden säkerställd?
- **Kunskap!**
Statistik = strukturerad, sammanställd information = kunskap

Historik

"Siffror"

- 3-4000 f Kr: folkräkningar
- år 0: "... för att skattskrivas"
- 1650: Graunt, "*Bills of Mortality*"
- 1749: lag om kyrkobokföring
- 1754: Tabellverket, Wahrgentin
- 1858: SCB

Metoder

- 1654: Pascal ☒ Fermat
- 1713: Ars conjectandi, Bernoulli
- 1830: Gauss, normalfördelningen
- 1908: "Student", Guinness, Dublin
- 1920: R.A. Fisher
- 1930-talet: Gallup och Mahalanobis

Officiell statistik

- **SOS, Sveriges officiella statistik**
 - Lagen om officiell statistik (SFS 2001:99)
 - objektiv, allmänt tillgänglig, hänsyn till skyddet för den enskilde
- **Sekretess, PuL, Datainspektionen**
 - individuppgifter i myndigheters register är sekretesskyddade
 - Personuppgiftslagen från 1998
 - Datainspektionen är en myndighet som genom sin tillsyns- verksamhet ska bidra till att behandlingen av personuppgifter inte leder till otillbörliga intrång i enskilda individers personliga integritet

Primärdata vs Sekundärdata

- **Primärdata** är data som samlats in av "mig själv" för just denna undersökning, den här utredningen.
- **Sekundärdata** är data som någon annan tagit fram. Färdiga sammanställningar eller "rådata" som jag får tillgång till.

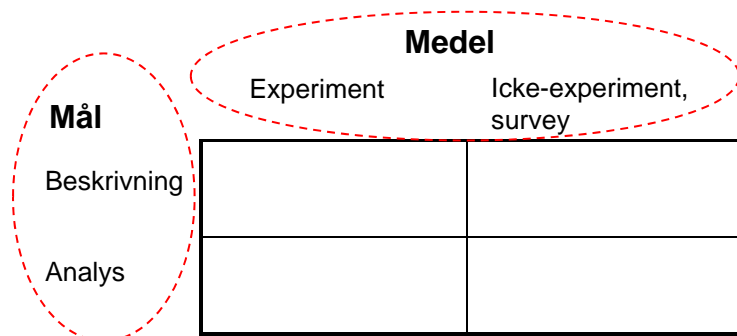
Sekundärdata

- **Sammanställningar: Statistikansvariga myndigheter t.ex.**
 - Statistiska Centralbyrån, SCB (www.scb.se)
 - Arbetsmiljöverket (www.arbetsmiljoverket.se)
 - Kemikalieinspektionen (www.kemi.se)
 - CSN (www.csn.se) och ytterligare drygt 20 myndigheter
- **Databaser: bland annat SCB och SND**
 - Statistiska Centralbyrån, SCB (www.scb.se). Databaser med t.ex. "Antal personer i olika inkomstklasser uppdelat på kommun, kön och ålder" eller "Antal anställda efter sektor, kön och ålder".
 - Svensk Nationell Datatjänst (www.snd.gu.se). Här finns en mängd datamaterial man efter ansökan kan få tillgång till för forskningsändamål, t.ex. ValU, SOM och enskilda undersökningar.

LWn

7

Mål och medel



LWn

8

Experiment vs Survey

Experiment

- kan upprepas
- försöksbetingelserna kan kontrolleras
- randomisering
- singel-, dubbel-, trippel-blind
- odiskutabla slutsatser?

Survey, observationsundersökning

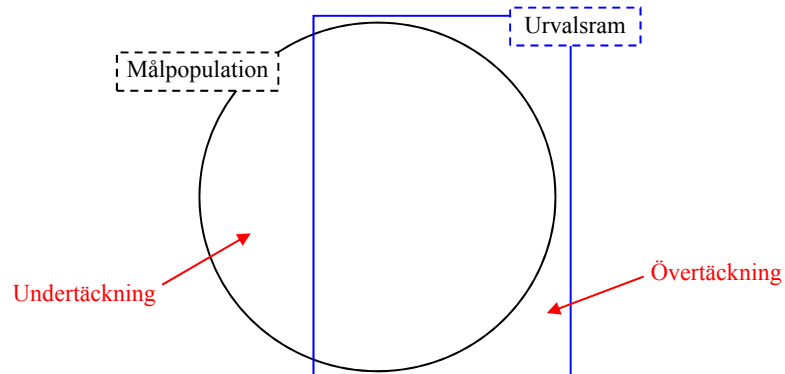
- "man tager vad man haver"
- "behandlingar" tilldelas inte slumpmässigt
- diskutabla slutsatser!

Etiska regler

- **Helsingforsdeklarationen**
 - WMA Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects (www.wma.net)
- **Vetenskapsrådet (VR)**
 - Etik under www.vr.se – Codex
- **ISI, International Statistical Institute**
 - <http://www.isi-web.org> – ISI Declaration on Professional Ethics
- **ESOMAR**
 - www.esomar.org - European Society of Opinion and Marketing Research, International Code on Market and Social Research

Målpopulation – Urvalsram

En urvalsram är den "lista" ur vilken urvalet görs. I bästa fall ska ramen vara helt identisk med målpoulationen. I praktiken är det inte alltid så.



Population – Individ – Variabel

- **Population:** *den grupp "individer" vi vill skaffa oss kunskap om*
 - alla 18-åringar i Sverige
 - alla skånska företag i en viss bransch
 - alla Volvo av 2004 års modell
- **Individ:** *undersökningsenhet*
 - människa
 - företag
 - bil
- **Variabel:** *egenskap som "mäts" hos individerna*
... och dessa egenskaper kan mätas på olika nivåer

Vilka uppgifter ska samlas in?

De här kanske:

1. Kön: Man Kvinna
2. Födelseår: _____
3. Hur många universitetspoäng har du tagit *före den här terminen*? _____
4. Hur var svårighetsgraden på senaste tentan?
 Jättelätt Lätt Svår Kanonsvår

Olika slags variabler

- **Kategorivariabel** (icke-numerisk)
 - ex: civilstånd, utbildningsprogram
- **Kvantitativ variabel** (numerisk)
 - **diskret**, ex: antal hemmavarande barn
 - **kontinuerlig**, ex: ålder, inkomst

Datanivåer, skaltyper

Dikotom, binär, 0/1-variabel

- **Nominalskala** (kategorivariabel)
anger endast kategori eller grupptillhörighet (ex. kön, utbildning)
- **Ordinalskala** ("ordnade kategorier")
möjligt att rangordna värden (ex. symptom: inga, små, måttliga, svåra)
- **Intervallskala** } **Scale** i viss programvara
- **Kvotskala** }
kvantitativ, "riktiga" numeriska mätvärden (ex. ålder, inkomst)
... och datanivån avgör vilken analysmetod som kan användas

Datanivåer, skaltyper

	Gruppering	Rang- ordning	Summor kan beräknas	Kvoter kan beräknas	
Kvotskala					Ålder (år)
Intervall-					Temp (°C)
Ordinal-					Betyg
Nominal-					Civilstånd

Några urvalsmetoder

- **Sannolikhetsurval**
 - OSU, obundet slumpmässigt urval
 - systematiskt urval
 - stratifierat urval
 - gruppurval (kallas också klusterurval)
- **Icke-sannolikhetsurval**
 - "på stan"-urval
 - självurval
 - pop-up enkäter på webbsidor

Fel i undersökningar

- **Slumpmässiga fel**
 - urvalet blir inte en exakt kopia av populationen
- **Annat som kan ge upphov till systematiska fel**
 - Bortfallsfel; gör alltid en bortfallsredovisning i rapporten
 - Individbortfall (eller externt bortfall)
 - Partiellt bortfall (eller internt bortfall)
 - Glömska
 - Prestige
 - Oklara frågor
 - Känsliga frågor
 - ...