F1

# Statistikens grunder

2013 HT, dagtid

Statistiska institutionen

Stockholms universitet

## Webbsajt och kursinfo

- http://www.statistics.su.se
  ⟹ Student ⟹ Kurshemsidor ⟹ Statistikens grunder, dagtid
- Läs kursbeskrivningen!

- MONDO:
- https://mondo.su.se/portal
- ⟹ Logga in, välj kursen/fliken ”Statistikens grunder, dag”



Stockholms universitet

## Kurslitteratur

- Nyquist, H., *"Statistikens grunder, kompendium"* **(N)**
- ***finns att ladda ner på kurshemsidan & Mondo***
- Thurén, T. (2007), *"Vetenskapsteori för nybörjare"*, 2:a upplagan, Liber: Stockholm **(T)**
- Övrigt kursmaterial såsom övningstentor, instruktioner till datorövningarna m.m. läggs löpande ut på ***Mondo***



Stockholms universitet

## Lärare på kursen

| | |
|---|---|
| Michael Carlson | Kursansvarig, föreläsare, examinator |
| Karl Hellström | Övningslärare Grupp A |
| Lisa Nikiforova | Övningslärare Grupp B |
| Karin Stål | Övningslärare Grupp C |
| Mikael Havasi | Övningslärare Grupp D |



Stockholms universitet

## Kursens upplägg

Kursen består av fyra moment i två block:

- **SG1**:
  - Moment 1: tentamen (6 poäng)
    - Salsskrivning + frivillig uppgift
  - Moment 2: inlämningsuppgift (1,5 poäng)
    - Två deluppgifter som redovisas skriftligt och första även muntligt
- **SG2**:
  - Moment 3: tentamen (6 poäng)
    - Salsskrivning + frivillig uppgift
  - Moment 4: inlämningsuppgift (1,5 poäng)
    - Två deluppgifter som redovisas skriftligt



Stockholms universitet

## Kursens upplägg

- Frivillig inlämningsuppgift
  - Ger bonuspoäng på Moment 1 resp. 3
  - Övningsuppgifter som lämnas in skriftligt; sedan rättar ni någon kurskamrats lösning
- Betyg:
  - Moment 1&3: A, B, C, D, E, Fx ,F
  - Moment 2&4: Godkänd, Underkänd
  - Slutbetyg på hela kursen enligt Kursbeskrivningen



Stockholms universitet

## Betyg

| SG1 \ SG2 | *A* | *B* | *C* | *D* | *E* |
|---|---|---|---|---|---|
| *A* | A | A | B | C | C |
| *B* | A | B | B | C | C |
| *C* | B | B | C | C | D |
| *D* | C | C | C | D | D |
| *E* | C | C | D | D | E |

+ Godkänt på Moment 2 och 4



Stockholms universitet

## Kursens innehåll

**SG1:**

- Vetenskap, modeller, kunskapsbildning
- Lite matematik, repetition
- Sannolikheter
  - utfallsrum, händelser, sannolikheten för en händelse
- Stokastiska variabler, slumpvariabler
  - diskreta och kontinuerliga



Stockholms universitet

## Kursens innehåll

**SG2:**

- Deskription, beskrivande statistik
- Sampling, centrala gränsvärdessatsen (CGS)
- Estimation, punktskattning och intervallskattning
- Hypotesprövning
- Chi-två metoden ($\chi^2$-metoden)
- Tidsserier och index
- Beslutsteori



Stockholms universitet

## Vad är det som krävs i tid?

| | | |
|---|---|---|
| Lärarledd undervisning | | |
| - *föreläsningar* | 2×23 | |
| - *räkneövningar* | 2×17 | |
| - *datorlabbar* | 2×6 | 92 h |
| Läsning egen tid | | 100 h |
| Övningar/inluppar | | 100 h |
| Restid | | ca 70 h |
| **Totalt** | | **362 h** |
| **Nio veckor heltid** | | **360 h** |



Stockholms universitet

## Matematik

- Allmän räknefärdighet, använda lite sunt förnuft, t.ex. bedöma om en lösning verkar rimlig eller ej
- Förstå vad en formel säger, kunna uttrycka sig med hjälp av formler
- Elementär algebra, funktioner, summatecken, mängdlära, kombinatorik, potenser och logaritmer

- ***Lästips: "Mot bättre vetande i matematik", Dunkels et al. Studentlitteratur, 3.e uppl.***
  - Ej kurslitteratur på Statistik I, men se avsnitt 1-3 och 5
  - Kurslitteratur på Statistik II



Stockholms universitet

## Sannolikheter och slumpförsök

SG1

- Vi kommer att tala om ***sannolikheter*** i samband med ***slumpförsök***.
- Ett slumpförsök är ett försök, som kan ***upprepas*** under likartade förhållanden, och där resultatet vid varje enskild upprepning inte kan förutsägas med säkerhet.
- "Försök" i vid mening (aktivitet, process, förlopp)



Stockholms universitet

## Sannolikheter, forts.

SG1

Exempel på slumpförsök:

- Tärningskast (1, 2, 3, 4, 5 eller 6)
- Lottdragning (vinst eller förlust)
- Befruktning av äggcell (pojke eller flicka)
- Radioaktivt sönderfall (antal partiklar under ett visst tidsintervall)
- Industriell tillverkning av en enhet (fungerar eller trasig)
- Slumpmässigt urval från en population (vilka blir utvalda)



Stockholms universitet

## Stokastisk variabel

SG1

- En stokastisk variabel är en ***kvantitativ*** variabel (numerisk, tal, siffror) vars värde bestäms av ett ***slumpförsök.***
- Annat namn: ***slumpvariabel***
- ***Utfallet*** av slumpförsöket bestämmer vilket värde (tal) den stokastiska variabeln ska anta.
- Vi kan i förväg ange vad som möjliga värden och sannolikheterna för dessa.



Stockholms universitet

## Stokastisk variabel, forts.

SG1

Exempel på stokastiska variabler:

- Antal prickar vid ett kast med en tärning
- Summan av antal prickar vid två tärningskast
- Antal kast tills man för första gången får en sexa
- Antal flickor i en slumpmässigt vald trebarnsfamilj
- Längden hos ett slumpmässigt valt nyfött barn
- Livslängden hos en slumpmässigt vald glödlampa
- Årsinkomsten i ett slumpmässigt valt hushåll

Vilka är de möjliga värdena för dessa stokastiska variabler?



Stockholms universitet

## Stokastisk variabel, forts.

SG1

- En ***diskret*** stokastisk variabel kan anta ett ändligt antal möjliga värden eller oändligt men uppräkneligt antal värden (”listbara”).

- En ***kontinuerlig*** stokastisk variabel kan anta alla värden inom ett intervall på den reella talaxeln (intervallet kan ha ändlig eller oändlig utsträckning).



Stockholms universitet

## Sannolikhetsfördelningar

SG1

Utifrån ganska standardmässiga slumpförsök och experiment kan vi definiera de möjliga utfallen och dessas sannolikheter.

⟹ dvs. stokastiska variabler

En funktion där man för ett givet utfall *x* kan beräkna sannolikheten att just det ska inträffa mha sannolikhetsfördelningen:

$$f(x) = P(X = x) = \text{sannolikheten}$$



Stockholms universitet

## Statistisk inferens

SG2

***Inferens = slutledning***, att dra en slutsats

- En slutledning är en process vid vilken man från ett antal premisser och i kraft av en slutledningsregel framställer en slutsats.
- ***Statistisk inferens*** handlar om slutledning om det generella mha av en ändlig uppsättning observationer, dvs. ett ***stickprov***, samt ***osäkerheten*** kring slutsatsen.



Stockholms universitet

## Statistisk inferens, forts.

SG2

- Inferensteori baseras på sannolikhetsteorin.
- Kvantifiering av ***osäkerheten*** i slutsatserna, dvs. ge ett mått på hur säkert/osäkert ett resultat är.
- Underlag för prediktioner och beslut.

- *Vad är den genomsnittliga inkomsten i Stockholm?*
- *Är det säkert att Centern har gått upp?*
- *Ska jag köpa aktie A eller B? Eller en portfölj?*



Stockholms universitet

## Statistisk inferens, forts.

SG2

- ***Punktskattning***
  - *skatta genomsnittlig inkomst för män och kvinnor*
- ***Intervallskattning***
  - *ge ett osäkerhetsintervall för skillnaden i inkomst*
- ***Hypotesprövning***
  - *pröva om det är statistiskt säkerställt att män har högre inkomst än kvinnor*
  - *pröva om val av bilmärke är oberoende av ålder*



Stockholms universitet

## Studiebesök på SCB

- 7 oktober kl 13-15, anmälan till studievagledare@stat.su.se senast 1 oktober

  Först till kvarn, mer information kommer!



Stockholms universitet

## Inledning till Statistik, N Kap 1

Att lära sig något från ***observationer***

- Sammanfatta erfarenheter
- Dra slutsatser (inferens)
- Göra förutsägelser (prediktion)
- Fatta beslut

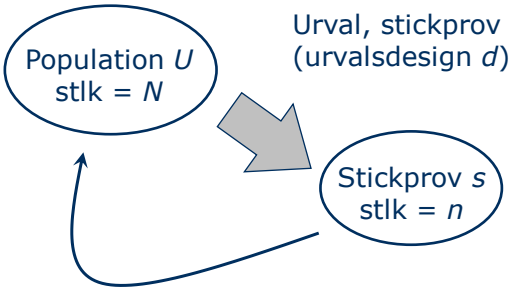
- Typiskt ***ofullständig information***
  - Vi kan inte fråga alla, vi har inte tid att pröva varje kombination

    ⟹ Statistiska metoder!



Stockholms universitet

## Ofullständig information

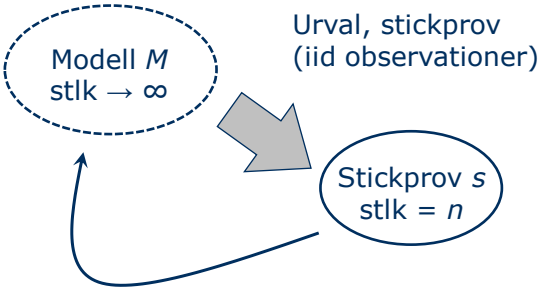



**Inferens:** säga något generellt gällande för hela populationen med ledning av informationen i stickprovet



Stockholms universitet

## Ofullständig information




**Inferens:** säga något generellt gällande om en universell egenskap, datagenererande process eller ”superpopulation” (som typiskt beskrivs med hjälp av en modell) med ledning av informationen i stickprovet



Stockholms universitet

## Varför observerar vi?

Typ av studier:

- Deskriptiva, beskrivande *("så här ser det ut")*
- Förklarande, kausalitet *("för att")*
- Normativa , preskriptiva *("gör så här")*

- Explorativa syften, sökande efter (ny) kunskap
- Konfirmativa syften, bekräftande av kunskap

***Oavsett typ så är syftet att öka vår kunskap om omvärlden***



Stockholms universitet

## En idé

Ta en titt i dagstidningen, webben, Text-TV …

- ***Hur många nyheter verkar bygga på en statistisk undersökning? Inferens?***

- ***Är nyheten baserad på beskrivande, förklarande, eller normativ studie?***

Tänk brett: opinionsundersökning, medicinska experiment, registerdata …



Stockholms universitet

## Inledning, forts.

***Statistiska undersökningar***, dvs. insamling av data, observationer som studerar och analyserar för att (förhoppningsvis) ge oss de svar som vi söker.

*Ordet **statistik** kan avse själva metoderna men kanske oftare används det som benämning på samlingen av observationer eller snarare sammanfattningar av data med beskrivande mått och grafiska presentationer.*



Stockholms universitet

## Lite vetenskapsteori

Thurén Kap 2:

- Vetenskapen söker sanningen
- Vetenskapen går ständigt framåt

- ***Dogmatism***
  - ”Detta är den (absoluta sanningen)”
- ***Relativism***
  - ”Sanningen förändras (på flera sätt) hela tiden, allstå finns ingen absolut sanning”



Stockholms universitet

## Lite vetenskapsteori, forts.

Thurén Kap 3: Definitioner

- Klargör ***alltid*** hur ni har definierat begreppen och vilka antaganden (premisser) som ligger bakom påståendena
- Annars kan inte din omgivning ta ställning till validiteten och reliabiliteten i dina påståenden



Stockholms universitet

## Lite vetenskapsteori, forts.

Thurén Kap 4: Iakttagelse och logik

- Empirisk kunskap
  - dvs. det vi ser, hör, smakar, …
  - vi drar slutsatser om det allmänna genom det vi ser
  - per definition är detta ofullständig information
  - slutsatsen är mer eller mindre trolig eller sannolik



Stockholms universitet

## Lite vetenskapsteori, forts.

Thurén Kap 4: Iakttagelse och logik

- Logik
  - utgår ifrån
    - premisser
    - lagar som styr hur vi härleder slutsatser
  - slutsatsen är alltid sann eller falsk
  - premisserna kan vara mer eller mindre verklighetsförankrade



Stockholms universitet

## Vetenskap, N Kap 2

Ett litet försök att bringa lite ordning bland begreppen.

Kunskapstyper:

- Propositionell kunskap
  - När föddes Astrid Lindgren?
- Icke propositionell kunskap
  - Färdigheter (simning)

*Proposition = Påstående*
*Ett begrepp inom logiken.*
*Ett påstående kan vara sant eller falskt.*

Den här kursen: *Kunskap = Propositionell kunskap*



Stockholms universitet

## Sanningsteori

Vad är sanning?

- ***Korrespondensteori***
  - Sant om det finns en motsvarande verklighet eller fakta
- ***Koherensteori***
  - Sant om det ”hänger ihop” med andra etablerade sanningar, ska inte leda till paradoxer
  - Ska höra till det system som är ”bäst”
- ***Pragmatism***
  - Satsens sanning ligger i dess brukbarhet



Stockholms universitet

## Sanning och kunskap

Ett påstående måste vara ***sant*** för att vara ***kunskap***.

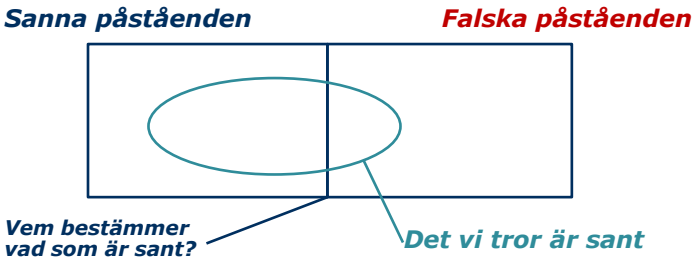
Låter som en självklarhet?

Motiveringen är att vi kan välja våra handlingar och nå de mål vi satt upp endast om vi väljer utifrån det vi vet är sant.



Stockholms universitet

## Sanning




1. Påstående som verkar vara sant
2. Nya observationer, nya fakta
3. Reviderat påstående



Stockholms universitet

## Epistemologi

Läran om vad man kan veta och hur man når kunskap

- ***Rationalism***
  - Ren tankeverksamhet
- ***Empirism***
  - Genom observationer och erfarenheter



Stockholms universitet

## Vetenskap

Ordet vetenskap kan avse:

- ***Processen***, dvs. hur får vi kunskap
- ***Resultatet*** av processen, den kunskap vi erhållit

Vetenskaper kan delas in i:

- ***Generaliserande, nomotetiska***
  - Fysik, matematik, kemi, ekonomi, beteendevetenskaper, …
  - Inte bara naturvetenskaperna!
- ***Partikulariserande, idiokratiska***
  - Enskilda händelser
  - Historia, konsterna



Stockholms universitet

## Vetenskap, forts.

Även efter vilken typ av objekt som studeras:

- ***Formella***
  - Studerar ***konstruerade*** objekt
  - Logik, matematik, (fysik)
- ***Empiriska***
  - Studerar ***verkliga*** objekt
  - Medicin, ekonomi, historia, (fysik)

*Med objekt avses till vardags ofta "**saker**" men även **händelser** och **relationer** studeras givetvis inom båda klasserna.*



Stockholms universitet

## Vetenskap, forts.

Var placerar vi ämnet ***statistik***?

Som vetenskap har ämnet sin grund i matematik och logik:

- Generell och formell
  - dvs. allmängiltig, nomotetisk
  - konstruerade objekt (tal, sannolikheter)

Men tillämpningen av statistiska metoder:

- Typiskt inom de generella och empiriska vetenskaperna
  - baseras på empiriska observationer



Stockholms universitet

## Sannolikheter

”Om gud har gjort världen till en fullkomlig mekanism, har han åtminstone givit så mycket till vårt ofullkomliga intellekt att vi, för att kunna förutsäga små delar av den, inte behöver lösa oräkneliga differentialekvationer, utan med hygglig framgång kan använda tärningar.”

MAX BORN



Stockholms universitet

## Vad är en teori?

Betyder något mer än bara ett antagande eller hypotes

*Vardagligt: ”Månen är gjord av ost” är en proposition och inte en teori*

- En teori är ett ***logiskt sammanhängande system*** av satser (påståenden) som beskriver ***relationer*** mellan väldefinierade ***objekt*** el. begrepp samt ***tolkningar*** av dessa relationer och objekt



Stockholms universitet

## Teorier

- ***Formella vetenskaper***
  - Axiom dvs. elementära grundantaganden som antas vara sanna
  - Logiska härledningar ur sanna påståenden till nya sanningar
  - ***Rationalism, koherens***
- ***Empiriska vetenskaper***
  - Vedertagna sanningar, påståenden
  - Logiska härledningar ur sanna påståenden till nya sanningar och prediktioner
  - Måste verifieras empiriskt
  - ***Empirism, korrespondens, korherens***



Stockholms universitet

## Bra teorier?

En bra (empirisk) teori ska

- Vara så ***generell*** som möjligt
- ***Förklara*** så mycket som möjligt
- Möjliggöra verifierbara ***förutsägelser***
- Ange ***riktlinjer*** (handling, beslutsunderlag)

Men även

- ***Enkelhet*** och ***tydlighet*** (Occam's razor)
- ***Objektivitet***

En teori brukar inte betraktas som sann eller falsk, snarare bedöms efter sin användbarhet (pragmatism)



Stockholms universitet

## Vetenskapens utveckling

- ***Kumulativitet***
  - Att alla nya forskningsresultat (dvs. sanningar) läggs till den etablerade teorin
  - Står ej i konflikt med det etablerade (koherens)
  - Ny pusselbit som passar in
- ***Paradigmskiften***
  - Nya fakta som står i konflikt med etablerade sanningar (bristande koherens, korrespondens)
  - Gamla påståenden ger "falska" resultat eller felaktiga prediktioner
  - Krävs en helt ny teori



Stockholms universitet