

Statistisk översiktscurs - Föreläsning 1

Anders Fredriksson

Statistiska Institutionen
Stockholms Universitet

24 mars, 2025



Stockholm
University

- **Ulf Högnäs**, kursansvarig och examinator (ulf.hognas@stat.su.se)
- **Anders Fredriksson** (anders.fredriksson@stat.su.se)
- Vi har vardera hälften av kursinnehållet

Föreläsning 1

- Inledning – några exempel
- Vad är statistik?
- Några grundläggande begrepp
- Vilseledande statistik
- Kursformalia



Dataanalys exempel 1: Penningtvätt

- Upptäcka och åtgärda penningtvätt
- Studera penningtransaktionsdata, leta efter mönster
- Exempel liknar analysen i Uppdrag Granskning (SVT, 2019)
- Inhämta data, "städa" data, bygga upp en databas
- Tusentals eller miljontals transaktioner
- Analysera belopp, antal, kontonummer, etc. i överföringarna
- Exempel på hur data skulle kunna se ut:

Transaktion	Belopp (EUR)	Datum	Tid	Sändare	Sändaradress	Mottagare	Mottagaradress	...
1	10000	250129	171310	A-bank	Bancovägen 2, Bankland	ABC Inc.	Bankroad 33, Moneyland	
2	200000	250129	171711	B-bank	Hypoteksägen 3, Bankland	XYZ Inc.	Bankroad 33, Moneyland	
3	333003	250129	171834	A-bank	Bancovägen 2, Bankland	ABC Inc.	Bankroad 33B, Moneyland	
...								



Exempel 2: Politisk påverkan vid fördelning av statsbidrag och offentliga tjänster?

- **Ex1:** Studie av Dahlberg och Johansson (2000, 2002) om hur ett (stats-)bidrag inom miljöområdet fördelades mellan kommuner
- Påverkade kommunens rösthistorik hur mycket bidrag som erhöles?
- **Ex2:** Fördelning av "närakuter" i delstaten São Paulo
- Påverkar en regions rösthistorik om regionen får en närakut?
- Testa teorier: Kärnväljare, swingväljare

Exempel 2: Politisk påverkan vid fördelning av statsbidrag och offentliga tjänster?

- Korrelation eller kausalitet?
- Exempel på datatabell (från São Paulo, 57 administrativa hälsoregioner):

	Samma parti i regionen* som i delstatsregeringen	Inte samma parti	Totalt
Fick närakut	14	17	31
Fick inte närakut	4	22	26
Totalt	18	39	57

* Valsystemets utformning påverkar vilken analys som kan göras

Exempel 3: Kaffeodling och produktivitet

- Projekt för utbildning av bönder i jordbrukstekniker
- Kan produktivitet och inkomster öka med utbildning? (en typ av statistisk utvärdering)
- Utbildningskomponenter inom exv. gödsling, bevattning, skadedjursbekämpning
- Data på bönder och jordbruk, odlingsareal, etc.
- Exempel på några av ett stort antal variabler:

	Namn	Kommun	Distrikt	Lat.	Long.	Areal	Gröda 1	Gröda 2	Utbildning (?)				Skörd (kg kaffeböror)				...
									1	2	3	...	År 0	År1	År2	...	
Bonde 1	B. Bonde	Kommun A	Veracruz	xxx	yyy	1 ha	Kaffe	-	JA	JA	NEJ	...	195	170	222	...	
Bonde 2	F. Farmer	...															
Bonde 3																	
....																	

Exempel 4: Vallokalsundersökning

- Förutspå / Skatta / Prediktera valresultat
- Exempel: SVT Valu
- Enkäter till ett urval av röstande efter röstning vid ett urval av förtidsröstningslokaler och vallokaler
- Användning av Valu från tidigare val
- Användning av tidigare valresultat
- Skattning av "valvinden"
- Data och modell för att skatta valresultatet i det ögonblick vallokalerna stänger
- Ytterligare användning av Valu: Studera hur röstning beror på geografi, kön, ålder, inkomst, etc.



Aspekter av dataanalys och statistik i ovanstående

- Stora datamängder - vi behöver metoder och verktyg
- Välstrukturerade data ("tidy data"): en variabel - en kolumn, en observation - en rad
- Leta efter "mönster" (korrelation, ..) i data
- Variabler - kategoriska och numeriska
- Deskriptiv statistik - korstabell med kategoriska variabler, fördelning
- Testa hypoteser, analysera, utreda, utvärdera
- Prediktion
- Primärdata och sekundärdata
- Enkäter / intervjuer / datainsamling
- Korrelation eller kausalitet?
- Tillämpningar inom politik och statsvetenskap



Referenser till några av exemplen ovan (ingår INTE)

- Dahlberg, M., & Johansson, E. (2000). Om regeringars röstköpsbeteende. Ekonomisk Debatt 28(4): 305-316.
- Dahlberg, M., & Johansson, E. (2002). On the vote-purchasing behavior of incumbent governments. American Political Science Review 96(1): 27-40.
- GU (2024). Vallokalsundersökningar (SVT/VALU). Valforskningsprogrammet, Göteborgs Universitet.
<https://www.gu.se/valforskningsprogrammet/undersokningar/vallokalsundersokningar-svtvalu>
- IFPRI (2024). Uganda Coffee Agronomy Training. Impact Evaluation Report. International Food Policy Research Institute.
- SVT (u.å.). Så analyserade vi materialet.
<https://www.svt.se/special/swedbank/metoden/>
- SVT (2020). Dirty Banking. <https://vimeo.com/481178797>

Ytterligare exempel på statistiska analyser

- Används inom **ekonomi, sociologi, statsvetenskap, psykologi, medicin, genetik, fysik, meteorologi, kryptografi, ...**
- **Ekonomiska prognoser** - Konjunkturinstitutet, Riksbanken
- **Kvalitetskontroll** - Tillverkningsindustrin
- **Läkemedelstestning** - Pfizer, Astra Zeneca
- **Risikanalys** - Försäkringsbolag, samhällsfunktioner
- **Infektionsspridning** - Smittskyddsintitutet
- Nobelpriset i fysik och ekonomipriset till Nobels minne 2024 erhöles delvis för statistiska metoder



Vad är statistik?

"Definition": Statistik är en vetenskaplig disciplin som handlar om

- Data
- Metoder för att arbeta med data

Ex: Data från en opinionsundersökning, respektive de slutsatser vi kan dra från opinionsundersökningen

En relaterad definition: Statistikområdet kan delas upp i

- **Deskriptiv statistik** - att beskriva data
- **Inferens** - att dra slutsatser från data -

Ex: 1054 personer har intervjuats om partipreferens.

Deskription - Presentation av undersökningen exv. i tabeller, diagram

Inferens - Slutsatser vi kan dra om befolkningens partipreferenser



Vad är statistik?

- Inom statistikområdet studeras metoder för att samla in, analysera och dra slutsatser från observationer när det finns slumpmässig variation.
- Slump, sannolikhet, urval och population är några centrala begrepp
- I exv. opinionsundersökningen vill vi, från ett slumpmässigt urval av röstberättigade, dra slutsatser om hur populationen "alla röstberättigade" skulle rösta
- Mer om dessa begrepp längre fram



■ Kursplan

- Kursinnehåll, lärandemål, juridiskt bindande dokument
- Finns i mappen Kursinformation på Athena

■ Kursbeskrivning

- Vad som gäller just den här terminen, allmän info, undervisningsplan, bedömningskriterier, viktiga datum, examination (inlämningsuppgift och tentamen)
- Finns i mappen Kursinformation på Athena

■ Kursmaterial (föreläsningar, labbar, inlämningsuppgift, etc.)

- Se kursens GitHub-sida: <https://statisticssu.github.io/SOK/>



Läroplattformen Athena (It's learning)

- Kursinformation
- Meddelanden, inklusive schemaändringar med kort varsel
- Inlämning av inlämningsuppgiften
- Chattfunktion
- Vi räknar med att ni har koll på meddelanden på Athena.
- Finns även som mobilapp.



- Föreläsningar (12) - Nya begrepp och metoder
- Dataövningar (5) - Lös uppgifter med R, förberedelse för inlämningsuppgiften, hjälp med inlämningsuppgiften
- Räkneövningar (3) - Genomgång av och hjälp med räkneuppgifter
 - viktigt att träna på att lösa uppgifter själv
 - förberedelse inför tentan
- Jourttillfällen (4) - Fråga "vad som helst"

- Inlämningsuppgift, 3 hp, grupparbete, skriftlig rapport
- Skriftlig tentamen, 4.5 hp

- Çetinkaya-Rundel, Mine., Hardin, Johanna. Introduction to Modern Statistics. Elektronisk resurs från OpenIntro.
<https://openintro-ims.netlify.app>

Övrig litteratur:

- Dahmström, K. (2011). Från datainsamling till rapport – att göra en statistisk undersökning (ca. 15 sidor).
- De Veaux, R., Velleman, P., & Bock, D. (2021), Stats: Data and Models (SDM) (ca. 15 sidor).
- Eventuellt tillkommer någon eller några (kortare) artiklar.
- Föreläsningar (pdf-filerna), labbinstruktionerna (pdf-filerna), övningsuppgifter, etc.



- Samarbete inom arbetsgruppen är självklart tillåtet.
- Alla i gruppen ska bidra ungefär lika mycket till rapporten och arbetet som leder upp till rapporten.
- Samarbete mellan grupper är också tillåtet.
- Plagiering är inte tillåtet! – automatiskt textmatchningsverktyg används.
- Om ni använder information som ni har hittat i böcker eller på nätet, ange alltid källan.
- Använd gärna AI-verktyg för inläring och för att hitta information, men lämna inte in AI-genererade svar på inlämningsuppgifterna!

Inlämningsuppgiften (fortstn.)

- Genomförs som grupparbete, 3 studenter i varje grupp
- Excelblad för gruppindelning finns på Athena
- D1 är obligatorisk för gruppindelning. Ingen annan obligatorisk närvaro på kursen
- Två tillfällen (deadlines) finns för inlämningsuppgiften:
 - Inlämning: **Fredag 25 april kl. 16.59.** Återlämnas 2 maj.
 - Komplettering: **Fredag 9 maj kl. 16.59.**
- Om inlämningsuppgiften blir underkänd efter inlämning 1 går det att komplettera och lämna in igen vid tillfälle 2 (kompletteringen).
- Om inlämningsuppgiften blir underkänd efter inlämning 2 ges nästa inlämningstillfälle vid nästa kurstillfälle.



Mer om att använda AI

- Vi betraktar en AI-genererad text som likvärdig med en text som har skrivits av en annan person.
- Att kopiera AI-genererad text och presentera som sin egen är otillåtet, detsamma gäller kopiering från en bok eller från nätdiskussionsfora.
- Om du formulerar en egen text med utgångspunkt i en AI-genererad text så bör du beskriva hur du har använt AI-verktyget.
- Du kan generera programmeringskod med ett AI-verktyg för att lära dig. All kod som lämnas in i en examination ska du ha skrivit själv.
- Kom ihåg att svaret från ett AI-verktyg, eller programkoden, inte alltid är korrekt. För att kunna skilja korrekta från felaktiga svar måste du själv ha tillräckliga kunskaper.



- Två tillfällen
 - **Onsdag 30 april kl. 14.00–19.00**
 - **Tisdag 10 juni kl. 08.00–13.00**
- Upplägg - minst 50 poäng av 100 möjliga för godkänt – kan vara räkneuppgifter och kunskapsfrågor
- Tillåtna hjälpmedel:
 - **Miniräknare utan lagrad text och data**
 - Häftet **"Formelblad och tabellsamling"** bifogas tentamen
- Om särskilda behov finns (egen lokal, extra tid, mm.), kontakta studievägledaren på statistiska institutionen i god tid innan tentan.
- **Glöm inte att anmäla dig till tentan i god tid!**

Betyg och betygskriterier

- Inlämningsuppgift: Godkänd, Underkänd
- Salstenta: A, B, C, D, E, (Fx), F
- F och Fx är underkända betyg som kräver omtentamen.
- Går ej att komplettera vid Fx
- Minimikrav för slutbetyg på hela kursen:
 - Godkänt på inlämningsuppgiften
 - Minst E på tentan
- För betygskriterier för inlämningsuppgift och tenta, se kursbeskrivningen.



- Enkät skickas ut efter kursen
- Era svar är jätteviktiga för att vi ska kunna veta vad som fungerar bra och vad vi kan förbättra i kursen.
- Vi sammanställer en rapport som läggs upp på Athena.

- Videor
- Går igenom hur Rstudio fungerar
- Introducerar programmeringsspråket R
- Titta gärna på några av filmerna före första datorövningen.
Förslagsvis video 2, och de första 5 minuterna av video 3.



- Börja plugga direkt.
- Läs igenom bokkapitlen som hör till en föreläsning innan föreläsningen.
- Utyttja undervisningstillfällena: gå på undervisningen, diskutera med andra studenter, ställ frågor.
- Kom igång med R så snart som möjligt - testa att installera innan första datorövningen.

Denna version av dokumentet: 2025-03-24

Materialet i Statistisk översikt kurs har tagits fram av Ulf Högnäs och Anders Fredriksson, med inspiration och ibland direkt användande av material från andra kurser och personer, bland annat kurserna Statistik och dataanalys 1-3, med material av Michael Carlson, Ellinor Fackle Fornius, Jessica Franzén, Oskar Gustafsson, Oscar Oelrich, Mona Sfaxi, Karl Sigfrid, Mattias Villani, med flera.

