

Tentamen i TNIU66, **Statistik och sannolikhetslära**, 25 mars 2020, kl. 14.00 – 18.00.

Kursens förväntade läranderesultat enligt kursplanen

Efter genomförd kurs ska du kunna:

1. analysera fördelningen hos en datamängd avseende centralvärde och spridning, såsom medelvärde och median samt standardavvikelse, samt visualisera detta.
2. redogöra för olika synsätt på begreppet sannolikhet.
3. beräkna sannolikheter för händelser, med användning av begrepp och verktyg som oberoende, betingning, oförenlighet, komplementhändelse, union, snitt, kombinatorik, trädidiagram.
4. formulera en sannolikhetsmodell med hjälp av slumpvariabler, även med centrala gränsvärdesatsen, och använda den för att bestämma egenskaper hos dess fördelning samt beräkna sannolikheter.
5. beräkna punktskattningar av väntevärde, varians, standardavvikelse, sannolikhet och intensitet samt bedöma deras lämplighet.
6. beräkna konfidensintervall för väntevärde (med och utan känd standardavvikelse), sannolikhet och intensitet samt tolka resultatet.
7. formulera och genomföra en hypotesprövning, och däri kunna tolka begreppen styrkefunktion och p -värde.
8. genomföra en korrelationsanalys och tolka resultatet.
9. ställa upp och tolka en linjär regressionsmodell med två variabler, avgöra om en linjär modell är tillämpbar, samt bedöma tillförlitligheten hos skattningar av såväl väntevärden som enskilda observationer.
10. använda datorstöd för alla beräkningar där det är relevant.

Tillåtna hjälpmedel:

- Valfri bok inom statistik och sannolikhetslära¹
- Miniräknare av valfritt slag (utan wifi-uppkoppling)

Det får finnas anteckningar och markeringar i boken, inklusive ”pagemarkeringar” (några centimeter stora), men inga lösblad.

Frågor besvaras av Michael Hörnquist som finns tillgänglig på telefon 011 - 36 33 81 under hela skrivtiden. Svar och kortfattade lösningsförslag finns på Studieinfo senast kl. 20.15. Skrivningsresultat meddelas förhoppningsvis senast femton arbetsdagar efter tentamenstillfället.

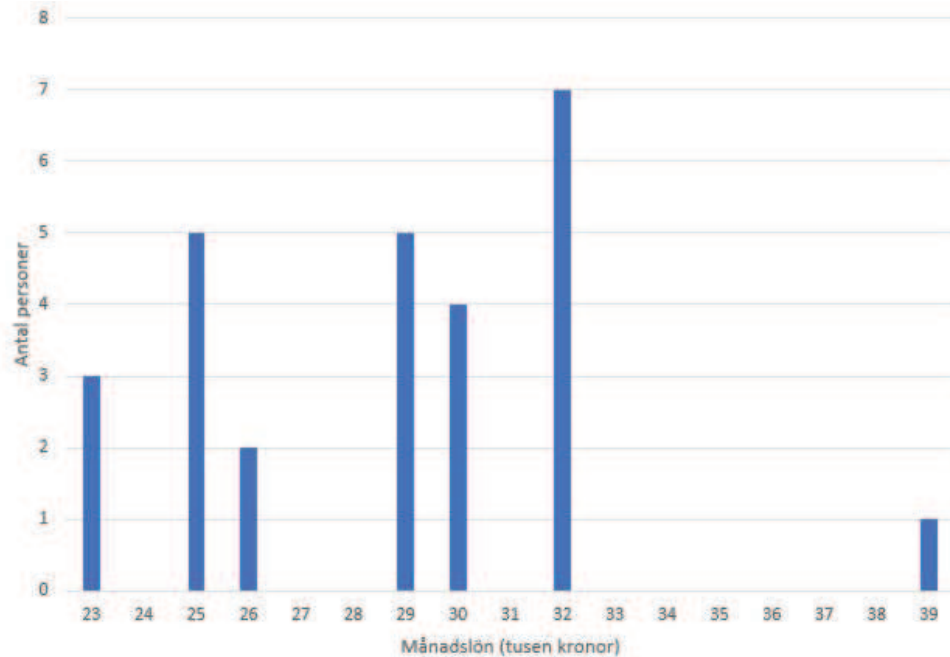
Varje uppgift ger 0 – 6 poäng. Ej behandlad uppgift ges en (1) poäng, för att markera betydelsen av att veta att man inte vet. Eventuell erhållen bonus från UPG1 påförs automatiskt och ingår i den totala poängsumman. För betyget n krävs minst $6n - 1$ poäng.

Svaren förs in på bifogad svarsblankett. Bakomliggande uträkningar ska bifogas det materiel som lämnas in så det är klart att det inte är gissningar och tur som har gett svaren. Allt laddas upp på Lisam senast 15 minuter efter skrivtidens slut.

Lycka till!

¹Kurslitteraturen ”Tillämpad statistik – en grundkurs”, Wahlin, Sanoma förlag, torde vara vanligast.

1. Lönerna hos ett mindre företag i västra Sverige fördelas sig enligt:



Ledningen anser den lägsta lönen vara alltför låg, och låter höja den med 2000 kronor för var och en av de tre medarbetare det gäller. Kommer därmed följande storheter att öka, minska, eller inte ändras alls?

- (a) Medelvärdet
- (b) Medianen
- (c) Typvärdet
- (d) Standardavvikelsen

Kort motivering krävs på varje del (men inga beräkningar).

2. En teknik vid insamlande av känsliga data är att ställa två olika frågor, och låta slumpen avgöra vilken respondenten ska besvara. I en undersökning av studenters narkotikaanvändning uppmanades de som svarade att börja med att kasta en tärning. Om sidan med ett öga eller sidan med två ögon kom upp skulle fråga 1 besvaras, annars fråga 2.

Fråga 1: Fick du upp sidan med ett öga?

Fråga 2: Har du under senaste året använt narkotika?

Vid en undersökning på en högskola svarade 240 av totalt 600 studenter "ja".

- (a) Hur stor andel har, enligt den här undersökningen, under senaste året använt narkotika?
- (b) Vad är sannolikheten att ett godtyckligt "ja"-svar avser narkotikabruk?

Bortse från att stickprov egentligen bör beskrivas med slumpvariabler, och räkna med punkt-skattningar.

3. Ett sågverk levererar brädor med medellängden 3,5 meter. En undersökning visar att 80% av dem har faktisk längd i intervallet 3,45 till 3,55 meter, och att längderna är oberoende och normalfördelade.
- (a) Bestäm standardavvikelsen för brädornas längder.
 - (b) Vad är sannolikheten att om du lägger två brädor i rad blir deras sammanlagda längd kortare än 6,9 meter? [Om du inte har löst (a) kan du använda värdet 4,0 cm som standardavvikelse för en bräda.]

4. I den halvstora staden X-köping förs en ständig diskussion huruvida den lokala flygplatsen ska få vara kvar eller om den ska stängas. Förslag om folkomröstning har framförts, men de ledande politikerna tvekar. Kan man verkligen överlåta en så viktig fråga till den "allmänna opinionen"?

En opinionsundersökning, dvs ett stickprov, vågar man sig dock på för att se om det eventuellt kan finnas en majoritet för stängning av flygplatsen. I ett obundet slumpmässigt urval svarar 110 personer "ja" på frågan om flygplatsen bör stängas, 55 personer svarar "nej" och 35 personer svarar "vet inte". Totalt i X-köping finns 103 289 personer som skulle ha rösträtt i en framtida eventuell folkomröstning.

- (a) Bestäm ett 90% nedåt begränsat konfidensintervall för andelen väljare som anser att flygplatsen ska stängas (dvs som svarar "ja").
 - (b) Bestäm ett 90% dubbelsidigt konfidensintervall för samma andel.
 - (c) Finns det på konfidensnivån 90% en majoritet för att stänga flygplatsen (motivering krävs)?
5. Vilket eller vilka påståenden är sanna? Vilket eller vilka är falska? Markera för varje påstående "S" om det är sant och "F" om det är falskt. Lämna blankt om du är osäker.

Signifikansnivån, α , vid en hypotesprövning är sannolikheten att:

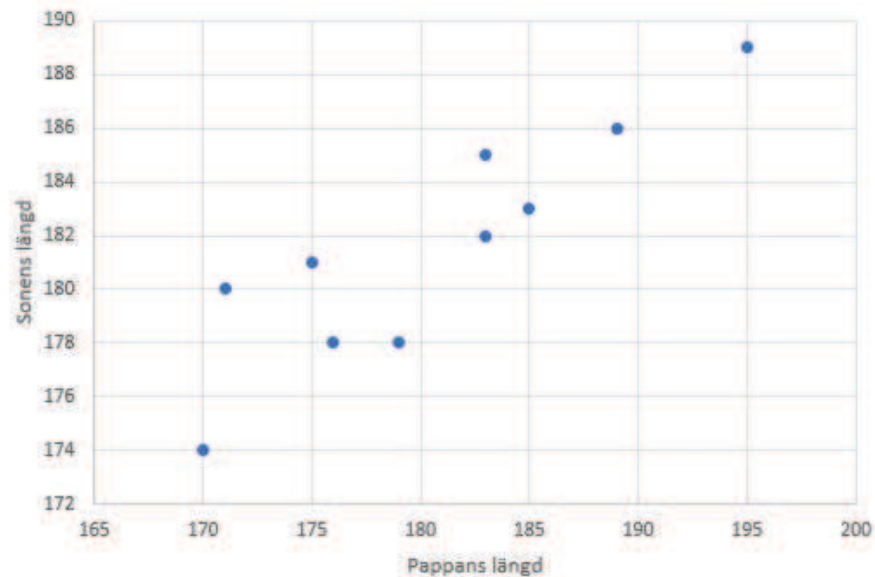
- (a) nollhypotesen är falsk.
- (b) nollhypotesen är sann.
- (c) förkasta en falsk nollhypotes.
- (d) *inte* förkasta en falsk nollhypotes.
- (e) förkasta en sann nollhypotes.
- (f) *inte* förkasta en sann nollhypotes.

Varje rätt svar ger en poäng och varje fel svar minus en poäng, dock kan totala poängsumman inte bli mindre än noll. Om du lämnar blankt blir det varken plus eller minus.

6. Ibland sägs det att "långa föräldrar får långa barn". Vi studerar det för fäder och söner genom att helt slumpmässigt välja ut tio män som alla även har minst en vuxen son. Fädernas och sönerns längder är (i centimeter):

Pappans längd:	176	183	170	195	183	189	179	175	171	185
Sonens längd:	178	182	174	189	185	186	178	181	180	183

Spridningsdiagrammet som alltid bör ritas blir:



Regressionsanalys med dataanalys-verktyget enligt Excel ger:

UTDATASAMMANFATTNING						
<i>Regressionsstatistik</i>						
Multipel-R	0,89059					
R-kvadrat	0,79315					
Justerad R-kvadrat	0,76729					
Standardfel	2,12352					
Observationer	10					
ANOVA						
	<i>fg</i>	<i>KvS</i>	<i>Mkv</i>	<i>F</i>	<i>p-värde för F</i>	
Regression	1	138,33	138,33	30,68	0,00	
Residual	8	36,07	4,51			
Totalt	9	174,40				
	<i>Koefficienter</i>	<i>Standardfel</i>	<i>t-kvot</i>	<i>p-värde</i>	<i>Nedre 95%</i>	<i>Övre 95%</i>
Konstant	92,507	16,100	5,746	0,000	55,381	129,634
X-variabel 1	0,493	0,089	5,539	0,001	0,288	0,699

Dags att reda ut vad undersökningen kan ge.

- (a) Bestäm ett uppåt begränsat konfidensintervall för riktningskoefficienten på konfidensnivån 95%, för att se om det är statistiskt säkert att den är mindre än ett.
- (b) Jonas är 181 cm lång. Ange ett 95% prognosintervall för längden hos hans son i vuxen ålder. Det kan förenkla att känna till att medellängden och stickprovsstandardavvikelsen bland fäder i det här stickprovet är 180,6 cm respektive 7,95 cm, för söner är motsvarande storheter 181,6 cm respektive 4,40 cm.

Not: En liknande undersökning som den här gav upphov till ordet "regression" för sådana här studier. Engelsmannen Francis Galton observerade 1885 att barn till särskilt långa eller korta föräldrar tenderade att få en längd i vuxen ålder närmare sin generations medellängd än deras föräldrar hade varit. Han beskrev detta som "regression to the mean", dvs som en återgång till medelvärdet. Matematiskt motsvarar det att riktningskoefficienten är mindre än ett.