

Tentamen i TNIU66, **Statistik och sannolikhetslära**, 5 juni 2023, kl. 8.00 – 12.00.

Kursens förväntade läranderesultat enligt kursplanen

Efter genomförd kurs ska du kunna:

1. analysera och visualisera fördelningen hos en datamängd.
 2. beräkna sannolikheter för vissa händelser med hjälp av teoretiska begrepp som ingår i kursinnehållet.
 3. beräkna punktskattningar och konfidensintervall.
 4. genomföra hypotesprövning
 5. genomföra enkel, linjär regressionsanalys.
 6. använda datorstöd för 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 ”pagemarkers” (några centimeter stora), men inga lösblad eller inklistrade sidor.

Frågor besvaras av Michael Hörnquist som besöker skrivsalen cirka kl. 9.00 och kl. 10.30. Svar och kortfattade lösningsförslag finns på Studieinfo senast kl. 15 på tentamensdagen. Skrivningsresultat meddelas 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 vid rättningen och ingår i den totala poängsumman. För betyget n krävs minst $6n - 1$ poäng, för betyg 3 ska dock minst 12 poäng av dessa komma från tentamen. Inget övrigt krav på fördelningen av poängen föreligger. Svaren som lämnas in ska anges på bifogad svarsblankett, och poängsättningen kommer att utgå från att det verkligen står ett svar på den. De resonemang och den kalkyl som lett fram till givna svar bifogas svarsblanketten.

Lycka till!

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

1. (a) Vad är skillnaden mellan en populationsundersökning och en stickprovsundersökning?
(b) Ge exempel på när det är viktigt att hålla ordning på om ens data kommer från en populationsundersökning eller från en stickprovsundersökning.
Ledning om längd: Om du skriver mindre än en mening är det antagligen för lite, om du skriver mer än tre meningar är det antagligen för mycket. Fortsätt på eget blad om svarsblanketten inte räcker till.

2. En fabrik producerar glödlampor och har en defekthalt på 2%. Om 50 glödlampor slumpmässigt väljs ut från produktionen, vad är sannolikheten att
 - (a) ingen av dem är defekt?
 - (b) exakt 2 av dem är defekta?
 - (c) minst 2 av dem är defekta?

3. En matematisk modell för livslängden för en viss typ av produkt säger att den har en exponentialfördelning med medelvärdet 3 år.
 - (a) Vad är sannolikheten att en produkt har en livslängd på mer än 5 år?
 - (b) Givet att produkten har klarat sina 5 första år, vad är sannolikheten att livslängden är ytterligare minst 5 år?

4. 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.
 - (a) Student t -fördelningen övergår i normalfördelningen $N(\mu = 0, \sigma = 1)$ när antalet frihetsgrader växer mot oändligheten.
 - (b) Om man vill höja konfidensnivån utan att ändra på något i sin stickprovsundersökning kan det åstadkommas genom att göra konfidensintervallet längre.
 - (c) Om man vill höja konfidensnivån utan att ändra på något i sin stickprovsundersökning kan det åstadkommas genom att göra konfidensintervallet kortare.
 - (d) En fördubblad konfidensnivå ger en fördubblad längd på konfidensintervallet.
 - (e) En fördubblad signifikansnivå ger en fördubblad längd på konfidensintervallet.
 - (f) Vid stickprovsundersökningar av andelar ger ett dubbelt så stort stickprov ett konfidensintervall som är fyra gånger smalare (antag dubbelsidigt intervall).

Endast svar krävs i denna uppgift. 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.

5. Från Novus undersökning om viktigaste politiska frågan i mars 2023 hämtar vi: "Svenskarnas viktigaste politiska fråga fortsätter att vara sjukvårdsfrågan (57%), följt av lag och ordning (51%). Den tredje viktigaste frågan fortsätter även att vara skola och utbildning (44%). Energipolitiken åker upp igen nu till en fjärdeplats (43%). Invandring/integration (42%) åker därmed ned till en femteplats. När det kommer till frågan om miljö och klimat har vi delat upp dessa politiska frågor i två så respondenterna har fått ta ställning till klimatet för sig och miljö för sig. Klimatet (31%) hamnar på en åttondeplats – en signifikant minskning jämfört med föregående mätning på 11 procentenheter, och miljö (27%) åker ned till en tiondeplats." Uppenbart summerar andelarna ovan ihop till mer än 100%, tydligen har svaranden fått ange mer än ett alternativ på "viktigaste frågan".

Genomför ett hypotestest på signifikansnivån $\alpha = 5\%$ för att se om angivet resultat är tillräckligt för att hävda att mer än hälften av svenska folket anser att "sjukvårdsfrågan" är den viktigaste politiska frågan. Skriv slutsatsen på svarsblanketten, genomför själva testet på eget blad.

Undersökningen omfattar 1034 intervjuer genomförda under perioden 16 – 22 februari 2023 med personer i åldersgruppen 18 – 79 år i ett slumpmässigt individurval av svenska folket. I aktuellt åldersintervall finns det cirka 7,5 miljoner människor i landet.

6. Följande uppgift bygger på samma grunddata som tentamen i mars 2023. Notera dock att det är andra storheter som efterfrågas denna gång.

En form av cement baseras på mineralet trikalciumsilikat, som bland annat ger snabb hållfasthetstillväxt. När cement stelnar utvecklas värme, hur mycket anses bero bland annat på mängden av det mineralet. För att närmare utreda detta utförs en försöksserie med tolv oberoende experiment där man lät andelen trikalciumsilikat variera mellan 27% och 70%. Resultatet blev följande mätserie:

x_i :	0,27	0,30	0,32	0,32	0,40	0,47	0,51	0,55	0,56	0,56	0,67	0,70
y_i :	18,8	17,8	21,0	17,3	20,0	27,7	22,9	22,3	26,1	25,0	27,1	24,6

Den övre raden anger andelen trikalciumsilikat och den undre mängden utvecklad värme (angett i Joule per gram cement). En linjär regressionsmodell ansätts, och ur Excel erhålls:

UTDATASAMMANFATTNING						
<i>Regressionsstatistik</i>						
Multipel-R	0,802					
R-kvadrat	0,644					
Justerad R-kvadrat	0,608					
Standardfel	2,259					
Observationer	12					
	<i>Koefficienter</i>	<i>Standardfel</i>	<i>t-kvot</i>	<i>p-värde</i>	<i>Nedre 95%</i>	<i>Övre 95%</i>
Konstant	13,3	2,3	5,8	1,66E-04	8,2	18,3
X-variabel 1	19,8	4,7	4,3	1,69E-03	9,4	30,1

- (a) Vad är förklaringsgraden för en linjär modell av hur värmeutvecklingen beror på andelen trikalciumsilikat?
- (b) Ange ett enkelsidigt, nedåt begränsat 95% konfidensintervall för riktningskoefficienten.
- (c) Vid vilken konfidensnivå går det inte längre att utesluta att riktningskoefficienten kan vara noll? Om du använder t -tabell får du avrunda till närmaste tabellvärde.