

Tentamen i TNIU66, **Statistik och sannolikhetslära**, 7 juni 2021, kl. 8.00 – 12.00.

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 gränsvärdessatsen, 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 ”pagemarkers” (några centimeter stora), men inga lösblad eller inklistrade sidor.

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. 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.

Svaren som lämnas in ska anges på bifogad svarsblankett, och poängsättningen kommer att utgå från att det verkligen står ett svar på den. För att säkerställa den egna prestationen ska dock den kalkyl som lett fram till givna svar bifogas svarsblanketten.

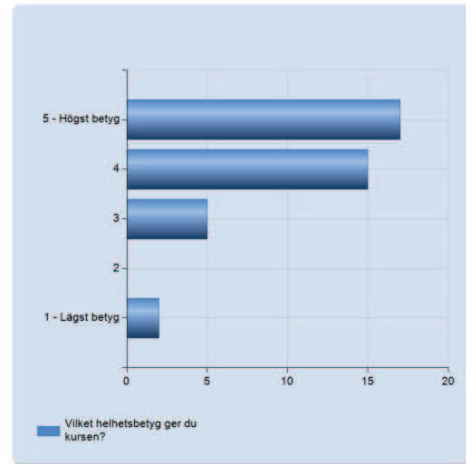
Lycka till!

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

1. På en kursvärdering i systemet Evaluate för kursen Statistik och sannolikhetslära, TNIU66, återfanns följande resultatbild efter kursomgången 2021:

9. Vilket helhetsbetyg ger du kursen?

Vilket helhetsbetyg ger du kursen?	Antal svar
5 - Högst betyg	17 (44%)
4	15 (38%)
3	5 (13%)
2	0 (0%)
1 - Lägst betyg	2 (5%)
Summa	39 (100%)



Figur 1: Resultat fråga 9 Evaluate för TNIU66 år 2021.

- Bestäm medelbetyget för kursen.
- Bestäm medianbetyget för kursen.
- Bestäm standardavvikelsen för kursen (välj själv om du bestämmer stickprovsstandardavvikelsen eller populationsstandardavvikelsen, men du måste ange vilken du väljer).
- Bonusfråga utan poäng: Det var 101 studenter som hade möjlighet att besvara kursvärderingen, svarsfrekvensen var alltså bara 38,6%. Vad kan vi göra för att få fler att svara?

2. Betrakta de fem funktionerna f_1 , f_2 , f_3 , f_4 och f_5 , definierade för alla reella tal x enligt:

$$f_1(x) = \begin{cases} \frac{1}{2} & \text{om } -\frac{1}{2} \leq x < -\frac{1}{4} \\ \frac{3}{2} & \text{om } -\frac{1}{4} \leq x < \frac{1}{4} \\ \frac{1}{2} & \text{om } \frac{1}{4} \leq x < \frac{1}{2} \\ 0 & \text{för övrigt} \end{cases} \quad f_2(x) = \begin{cases} -\frac{1}{2} & \text{om } -\frac{1}{2} \leq x < -\frac{1}{4} \\ \frac{5}{2} & \text{om } -\frac{1}{4} \leq x < \frac{1}{4} \\ -\frac{1}{2} & \text{om } \frac{1}{4} \leq x < \frac{1}{2} \\ 0 & \text{för övrigt} \end{cases}$$

$$f_3(x) = \begin{cases} \frac{1}{2} & \text{om } -\frac{1}{2} \leq x < -\frac{1}{4} \\ \frac{5}{2} & \text{om } -\frac{1}{4} \leq x < \frac{1}{4} \\ \frac{1}{2} & \text{om } \frac{1}{4} \leq x < \frac{1}{2} \\ 0 & \text{för övrigt} \end{cases} \quad f_4(x) = 2e^{2x} \quad f_5(x) = 2e^{-2x}$$

- Endast en av de fem funktionerna f_1 , f_2 , f_3 , f_4 och f_5 är täthetsfunktion till en slumpvariabel. Vilken? Motivering krävs för var och en av funktionerna.
- Bestäm väntevärde och standardavvikelse för den funktion som är en täthetsfunktion.

Ledning: Det kan vara till hjälp i vissa fall att skissa hur funktionerna ser ut.

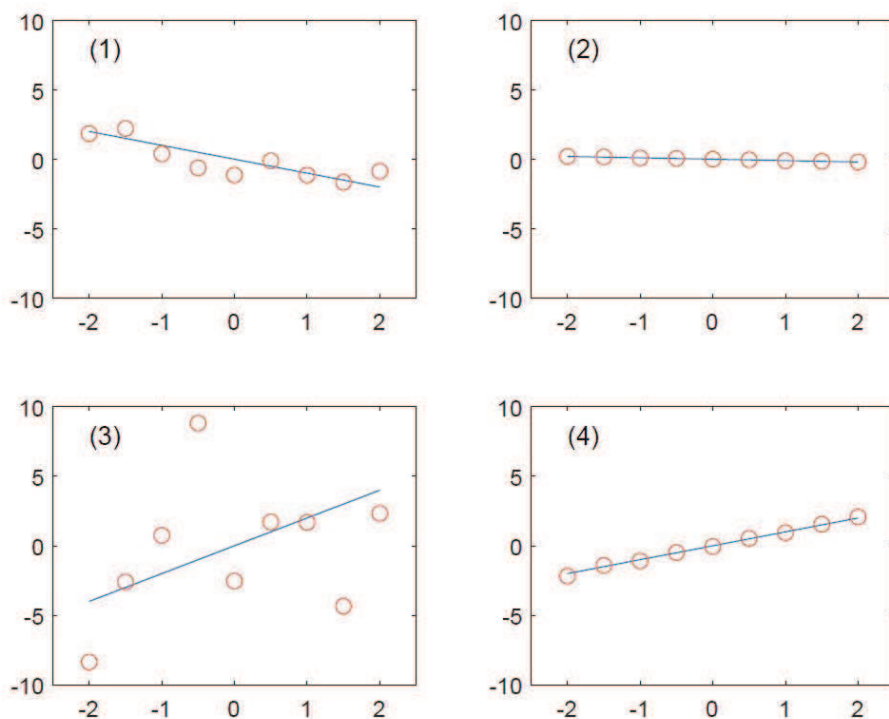
3. Begreppet "IQ", intelligenskvot, är omtvistat. Det är dock definierat så att det i populationen "alla människor" ska ha vara normalfördelat med medelvärde 100 och standardavvikelse 15.
 - (a) Vad är sannolikheten att en godtycklig person som du slumpmässigt träffar på har en IQ som överstiger 120?
 - (b) Välj ut tre personer helt på måfå, i ett obundet slumpmässigt urval. Vad är sannolikheten att medelvärdet av dessa tre personers IQ överstiger 120?

4. Två elever, Roger och Miriam, gör som del av sitt gymnasiearbete en statistisk undersökning om partisympatier inför det stundade skolvalet. De gör ett slumpmässigt urval av 40 personer bland sina skolkamrater och frågar dem vilket parti de tänker rösta på i skolvalet. Åtta personer anger Piratpartiet. Skolan har 1215 elever totalt.
 - (a) Bestäm ett dubbelsidigt konfidensintervall för andelen röster Piratpartiet förväntas erhålla i skolvalet. Använd konfidensnivån 90%.
 - (b) Miriam och Roger funderar på att utöka sina undersökningar till att gälla hela kommunen där de bor. Kommunen har 96 248 invånare som är röstberättigade i kommunalvalet. Hur stort stickprov måste de ta för att få samma felmarginal som vid undersökningen inför skolvalet?

5. Roger och Miriam från uppgiften ovan (du behöver inte ha löst den för att kunna behandla den här uppgiften) bestämmer sig för att göra slag i saken och göra en opinionsundersökning för hela kommunen (med 96 248 invånare röstberättigade i kommunalvalet). Vad de främst vill undersöka är om det finns skäl att tro att Piratpartiet har vuxit i popularitet sedan senaste valet, undersökningen vid deras skola verkar ju tyda på det. För att ge ett seriöst intryck väljer Roger och Miriam att tillfråga 400 personer i ett obundet slumpmässigt urval av röstberättigade invånare. Av dessa anger 34 personer att de skulle rösta på Piratpartiet om det vore val idag. Vid senaste valet var det 7% som röstade på Piratpartiet. Gör ett hypotestest för att se om *uppgången* är statistiskt säkerställd på signifikansnivån 10%.

På svarsblanketten skriver du slutsatsen, medan själva hypotestestet skrivs på medföljande blad med beräkningar.

6. Nedan ser du fyra spridningsdiagram, (1), (2), (3) och (4), där en regressionslinje har ritats in i var och en av dem. Notera att linjen i (2) lutar svagt nedåt.



Figur 2: Fyra spridningsdiagram med inritade regressionslinjer.

Beteckningarna är valda så att till delfigur (i) hör regressionslinjen $y = b_1^{(i)}x + b_0^{(i)}$ och korrelationskoefficienten $r^{(i)}$.

- Ordna riktningskoefficienterna $b_1^{(1)}$, $b_1^{(2)}$, $b_1^{(3)}$ och $b_1^{(4)}$ i storleksordning, från minsta till största.
- Ordna korrelationskoefficienterna $r^{(1)}$, $r^{(2)}$, $r^{(3)}$ och $r^{(4)}$ i storleksordning, från minsta till största.

Det är inte avsikten att du ska räkna ut värdena på riktningskoefficienter eller korrelationskoefficienter, utan att du ur figurerna ska kunna se deras storleksordning.

Notera att vissa av storheterna kan vara negativa, och att det *inte* är storleken på deras *absolutbelopp* som efterfrågas. Konkret står det för att exempelvis $-3 < -2$.