

Tentamen i TNIU66, **Statistik och sannolikhetslära**, 25 augusti 2018, kl. 8.00 – 12.00.

Kursens förväntade läranderesultat enligt kursplanen

Efter genomförd kurs skall 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 stokastiska variabler, 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

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

Frågor besvaras av Michael Hörnquist som besöker salen minst två gånger under skrivtiden, ca kl. 9 och 10.30. Svar och kortfattade lösningsförslag finns på Lisam efter skrivningens slut. Skrivningsresultat och visningstid meddelas via epost senast tio arbetsdagar efter tentamenstillfället.

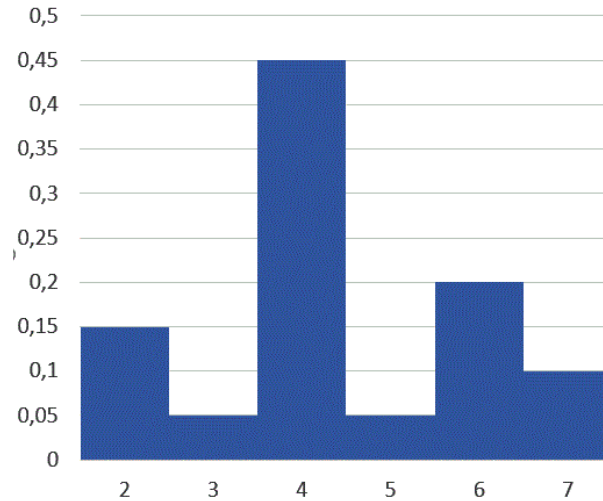
Varje uppgift ger 0 – 6 poäng. Ej behandlad uppgift ges en (1) poäng, för att markera betydelsen av att veta att man inte vet. Eventuell erhållen bonus från UPG1 påförs automatiskt och ingår i den totala poängsumman. För betyget n krävs minst $6n - 1$ poäng.

Redovisningen av uppgifterna skall vara sådan att det klart framgår vad du gör. Det innebär att förklarande text skall finnas med. Att enbart ge ett svar utan motivering, även om det är korrekt, är normalt inte tillräckligt för att uppgiften skall anses vara nöjaktigt löst. Å andra sidan behöver inte ett felaktigt svar, såvida det inte av någon orsak är orimligt, innebära att lösningen ej är godkänd. Väsentligt är att det bakomliggande resonemanget klart framgår samt är relevant.

Lycka till!

¹Kurslitteraturen ”Statistik – metoder och tillämpningar”, Lövås, Liber förlag, torde vara vanligast.

1. Nedanstående histogram visar fördelningen hos en okänt stor mängd heltal mellan 2 och 7. Bestäm mängdens medelvärde och median.



Notera att y -axeln anger andelar och att totala stapelarean är ett (1).

2. Studenten Eva läser en kurs i sannolikhetslära där det ingår sex stycken tavelredovisningsuppgifter vid lika många tillfällen. Vid varje tillfälle ska en slumpvist utvald student redovisa en uppgift; läraren väljer ut en av trettio studenter genom att kasta en trettiosidig tärning. Inför första redovisningstillfället funderar Eva över följande:

- Vad är sannolikheten att hon inte behöver redovisa någon gång alls?
- Vad är sannolikheten att hon behöver redovisa minst två av de sex gångerna?
- Vad är sannolikheten att hon får redovisa vid alla de första fem tillfällena?
- Om nu det högst osannolika i (c) skulle inträffa, att hon får redovisa vid alla de första fem tillfällena, vad är då sannolikheten att hon ska behöva redovisa även vid det sjätte tillfället?

Eftersom det här är i början av kursen saknar Eva de teoretiska verktygen för att räkna ut de sannolikheter hon funderar över. Gör uträkningarna åt henne så att hon ser nyttan av det som ska studeras.

Ovanstående uppgift har samma grundförutsättning som en uppgift på föregående tentamen. Notera dock att här frågas efter andra sannolikheter.

3. Den nye VD:n för ett konsultföretag, Elfatih, överväger att införa ett enhetspris på tjänsterna, detta då han tycker sig se att spridningen i tidsåtgång för varje enskilt uppdrag är liten och det skulle underlätta administrationen och minska faktureringskostnaderna. Från sina medarbetares tidsrapporter senaste åren kan han beräkna att medeltiden för ett uppdrag är 19,6 timmar och spridningen (standardavvikelsen) är 2,4 timmar. Tiderna förefaller också vara normalfördelade. Elfatih önskar göra några överväganden innan han är beredd att införa enhetspris:

- Hur stor andel av projekten slutförs på mindre än 15 timmar?
- Betrakta nu de en procent (1%) längsta projekten. Hur långa är de? Elfatih söker den undre gränsen.

4. Det är valår i Sverige, och opinionsundersökningarna kommer tätt. Före sommaren redovisade SIFO en undersökning baserad på ett stickprov med 2176 intervjuer av väljare. Exakt hur många röstberättigade som kommer att finnas i september är det ännu ingen som vet, men enligt en uppskattning av SCB 2017-09-11 kommer det att vara 7 268 000 personer (för riksdagsvalet).
- (a) Det största partiet, Socialdemokraterna, fick i denna undersökning 25,2%. Bestäm ett 95% konfidensintervall för den sanna andelen väljare som skulle rösta på S.
 - (b) Liberalerna har ett väljarstöd om 4,8% enligt samma undersökning. Här ligger de farligt nära gränsen 4,0% för att få lämna riksdagen. Bestäm ett ensidigt 95% konfidensintervall, nedåt begränsat, för den sanna andelen väljare som skulle rösta på L.

5. Trafikansvarig Rania på kommunen har tröttnat på dåliga ljuskällor. Minst tre år bör en LED-gatlykta hålla, anser hon. Det är därför med lika delar förtjusning och skepticism hon ser reklam från företaget "Ljus på gator och hus" som säljer ljuskällor med en garanterad livslängd (i medel) om minst tre år. Hon låter köpa in hundra sådana lampor och testar dem. Resultatet blir en medellivslängd om 26390 timmar med en standardavvikelse (*inte* standardfel) om 500 timmar. Ett år kan anses ha 8766 timmar.

Gör ett hypotestest åt Rania för att undersöka om det kan anses vara statistiskt bevisat att ljuskällorna från "Ljus på gator och hus" har en medellivslängd som överstiger tre år. Ange noga såväl signifikansnivå som vilka hypoteser du använder.

6. 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) Om den skattade riktningskoefficienten $\hat{\beta} = -0,9$ innebär det att y -värdena i medeltal *avtar* med ökande x -värden.
 - (b) Om den skattade riktningskoefficienten $\hat{\beta} = 1,1$ innebär det att y -värdena i medeltal *ökar* med ökande x -värden.
 - (c) Om den skattade riktningskoefficienten $\hat{\beta} = 1$, så vet vi att alla punkter i datamängden ligger längs en rät linje i ett spridningsdiagram.
 - (d) Om den skattade korrelationskoefficienten $r = 1$, så vet vi att alla punkter i datamängden ligger längs en rät linje i ett spridningsdiagram.
 - (e) Om alla punkter i en datamängd ligger längs en rät linje med positiv lutning i ett spridningsdiagram, så vet vi att den skattade riktningskoefficienten $\hat{\beta} = 1$.
 - (f) Om alla punkter i en datamängd ligger längs en rät linje med positiv lutning i ett spridningsdiagram, så vet vi att den skattade korrelationskoefficienten $r = 1$.

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.