

Tentamen i TNIU66, **Statistik och sannolikhetslära**, 8 juni 2020, 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.

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. 14.15. Skrivningsresultat meddelas senast femton arbetsdagar efter tentamenstillfället.

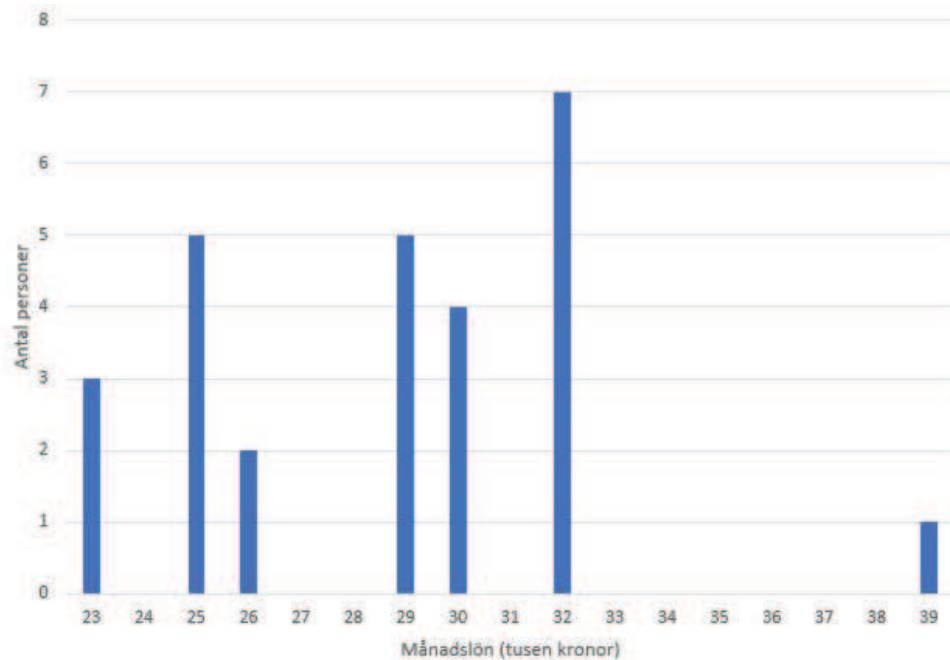
Varje uppgift ger 0 – 6 poäng. Ej behandlad uppgift ges en (1) poäng, för att markera betydelsen av att veta att man inte vet. Eventuell erhållen bonus från UPG1 påförs vid rättningen och ingår i den totala poängsumman. För betyget n krävs minst $6n - 1$ poäng.

Svaren förs in på bifogad svarsblankett, det förväntas att det kommer att bli mer än en sida eftersom vissa svar kan bli långa. Löpande text skrivs med dator (Urkund kommer att användas). Bakomliggande uträkningar skrivs för hand och skannas för att bifogas svarsblanketten så det är klart att det inte är gissningar som har gett svaren. Inlämning sker via Lisam, men om tekniken krånglar skickas de per epost till michael.hornquist@liu.se

Lycka till!

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

1. Lönerna hos ett mindre företag i västra Sverige (samma företag som vid marstentan 2020) fördelade sig år 2018 enligt:



Medelvärde var 28 667:- och standardavvikelsen var 3702:-.

- (a) Vid 2019 års lönerrevision beslutade ledningen att löneökningen skulle vara 2,5% för alla medarbetare. Vad kommer då medelvärde, median, typvärde och standardavvikelse att bli efter lönerrevisionen?
- (b) Året därpå klagade några personer att det var orättvist att de som redan hade en bra lön skulle få större ökning i kronor än de som hade låg lön. Företagsledningen lyssnade på argumentet, och beslutade att 2020 skulle allas löner öka med 800 kronor, oavsett vad de tjänade innan. Vad kommer då medelvärde, median, typvärde och standardavvikelse att bli efter lönerrevisionen (minns att lönerna redan är reviderade en gång 2019)?
2. Lisa och Liza funderar på hur många olika stavningar av deras förnamn det kan finnas. Först har vi "s" eller "z", börjar Lisa, våra varianter, det blir två, sedan "i" eller "ii", det blir två till dvs fyra. Så kan du till sist ha ett "h" på slutet, eller inte, det blir ytterligare två, dvs sex olika stavningsversioner totalt. Nej, nej, nej, kontrar Liza, så kan du inte tänka. Istället så måste du ...

Skriv med egna ord vad Liza kan tänkas svara Lisa, givet att Lisa kan sin kombinatorik. Om du skriver mindre än en mening är det antagligen för kortfattat, om du skriver mer än en sida är det antagligen mycket mer än nödvändigt. Om du vill bifoga en skiss går det bra.

3. I den här uppgiften studerar vi en 60-sidig tärning (en s.k. D60), där sidorna är markerade med heltalen mellan 1 och 60.
- Bestäm väntevärdet för kast med en D60.
 - Standardavvikelsen för tärningskastet är $\approx 17,32$, ange en formel som bygger på att det är en diskret likformig fördelning för att beräkna det värdet.²
 - Beräkna sannolikheten $\Pr(\mu - \sigma \leq X \leq \mu + \sigma)$, där X är slumpvariabeln för vilken sida som kommer upp på tärningen, μ är väntevärdet och σ är standardavvikelsen.
 - Beräkna sannolikheten $\Pr(\mu - 2\sigma \leq X \leq \mu + 2\sigma)$, samma beteckningar som i (c).
 - Vad är sannolikheten att ett enskilt tärningskast ska ge precis väntevärdet?
 - Vad är sannolikheten att medelvärdet av två kast ska ge precis väntevärdet?
4. Ett bryggeri fyller 33 cl flaskor med läsk, hinner med 238 stycken om dagen. Tappningsmaskinen har en noggrannhet (=standardavvikelse) om 1,1 cl för varje flaska den fyller. Ingen flaska får fyllas med mer än 35 cl, då rinner det över.
- Bestäm vilket väntevärde som tappningsmaskinen behöver ställas in på så att risken för överfyllning av en enskild flaska blir maximalt en på tusen (0,1%). Antag att mängden läsk som fylls på per flaska är normalfördelad.
 - Om maskinen ställs in på det väntevärde som beräknades i (a),³ hur mycket läsk behövs för att det ska räcka till att fylla en dagsproduktion med 95% säkerhet? Antag att påfyllningarna av varje flaska sker oberoende av varandra.

5. Kiosken Glädje-o-Lycka säljer bland mycket annat även Trisslotter. Man tycker sig märka att många kunder vinner, och använder det i sin marknadsföring med en slogan: "Glädje-o-lycka, den lyckosamma kiosken, här vinner många fler."

Både Konsumentverket och Lotteriinspektionen reagerar starkt på detta. Konsumentverket vill säkerställa att påståendet verkligen är *sant*, så inga konsumenter luras, och Lotteriinspektionen vill säkerställa att påståendet verkligen är *falskt*, då det annars betyder att någon har lyckats manipulera systemet.

Konsumentverket och Lotteriinspektionen kommer överens om att genomföra ett hypotestest för att undersöka fallet. Eftersom det är känt att andelen vinstlotter i Trisslotteriet är 20% sätter man upp följande hypoteser, där π är den sanna men okända andelen vinstlotter hos kiosken:

$$H_0 : \quad \pi \leq \pi_0 = 0,20, \text{ dvs inte fler vinster än andra}$$

$$H_a : \quad \pi > \pi_0 = 0,20, \text{ dvs fler vinster än andra}$$

En medarbetare skickas ut att införskaffa och skrapa 100 lotter och se hur många som ger vinst. Hon finner att 25 av dem är vinstlotter.

- Vilken slutsats kan Konsumentverket dra av testet?
- Vilken slutsats kan Lotteriinspektionen dra av testet?

Beräkningar redovisas handskrivna på inskannade sidor, slutsatserna skrivs på svarsblanketten. Välmotiverade slutsatser krävs för poäng på uppgiften.

²När denna tentamen gavs 8 juni 2020 stod av någon obegriplig orsak talet 12,62 här istället.

³Om du inte fann värdet i (a) kan du istället använda värdet 33 cl.

6. Vilket eller vilka påståenden är sanna? Vilket eller vilka är falska?
- (a) Ju brantare regressionslinje, desto högre värde på absolutbeloppet av korrelationskoefficienten.
 - (b) Korrelationskoefficienten kan sägas vara ett mått på hur väl punkterna i en tvådimensionell datamängd ligger längs en rät linje.
 - (c) Korrelationskoefficienten kallas också för förklaringsgrad, och är ett mått på hur stor andel av variationen i datamängden som fångas upp av en rät linje i regressionsanalys.
 - (d) Konfidensintervallet för genomsnittsnivån μ_{x^*} i en regressionsanalys blir hälften så brett om antalet datapunkter n blir fyra gånger fler.
 - (e) En anpassning av ett andragradspolynom till en mängd talpar kommer alltid att ha minst lika hög förklaringsgrad som anpassning av en rät linje till samma mängd talpar.
 - (f) Det går alltid att beräkna riktningskoefficienten i en regressionsanalys, även om förklaringsgraden är exakt lika med noll.

Kort motivering till varje svar krävs i denna uppgift, skrivs på svarsblanketten. OBS: Ändrade krav på redovisning mot tidigare tentamina, inga minuspoäng för felaktiga svar. Motiveringen måste vara med egna ord, får inte vara direkt citat ur kurslitteraturen.