LINKÖPINGS UNIVERSITET

Institutionen för teknik och naturvetenskap
Michael Hörnquist, 011-363381

Tentamen i TNIU66, **Statistik och sannolikhetslära**, 4 juni 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, träddiagram.
4. formulera en sannolikhetsmodell med hjälp av stokastiska variabler, 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 sannolikhetslära[1]
- 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!

[1] Kurslitteraturen ”Statistik – metoder och tillämpningar”, Løvås, Liber förlag, torde vara vanligast.

1. Det får anses välkänt att stickprovsstandardavvikelsen, s, för en mängd tal $x_1, x_2, \cdots, x_n$ definieras som

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2},$$

där $\bar{x}$ är det aritmetiska (dvs vanliga) medelvärdet av talen i mängden. Kvadratroten brukar motiveras med att man vill att standardavvikelsen ska ha samma enhet som elementen i mängden, till exempel om varje x_i anger en längd mätt i meter, så vill man att spridningsmåttet också ska ha enheten meter (utan kvadratroten skulle det bli kvadratmeter, eftersom varje term kvadreras).

Det här tycker Peter verkar dumt. Först kvadrera och sedan roten ur. Han genomskådar att standardavvikelsen effektivt endast är summan av alla avvikelser från medelvärdet, i kvadrat, och delat med antalet element (minus ett, av någon orsak han inte helt förstår, hade något med skillnad mellan stickprov och totalundersökning att göra). Rimligare, och enklare, och bättre på alla vis, borde vara att strunta i såväl kvadreringen som kvadratroten och "antal minus ett", tycker Peter, och föreslår därför sin egen standardavvikelse, Peterstandardavvikelsen P_s. Den definieras som

$$P_s = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}).$$

Linda väljer att förhålla sig lätt skeptisk till denna innovation. Hon vill se vad den innebär i praktiken, och väljer därför att beräkna såväl den vanliga stickprovsstandardavvikelsen som Peterstandardavvikelsen för två mängder. När hon gjort det konfronterar hon Peter med resultatet och förklarar att Peterstandardavvikelsen har en klar brist som mycket tydligt framgår av hennes beräkningar.

Gör om Lindas beräkningar med hennes val av mängder, $\{1, 2, 3\}$ och $\{0, 1, 1, 2, 3, 3, 4\}$. Skriv en mening eller två till Peter som anger vilken brist hans nyuppfunna Peterstandardavvikelse tycks[2] lida av.

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:

   (a) Vad är sannolikheten att hon får redovisa vid första tillfället?

   (b) Vad är sannolikheten att hon behöver redovisa minst en av de sex gångerna?

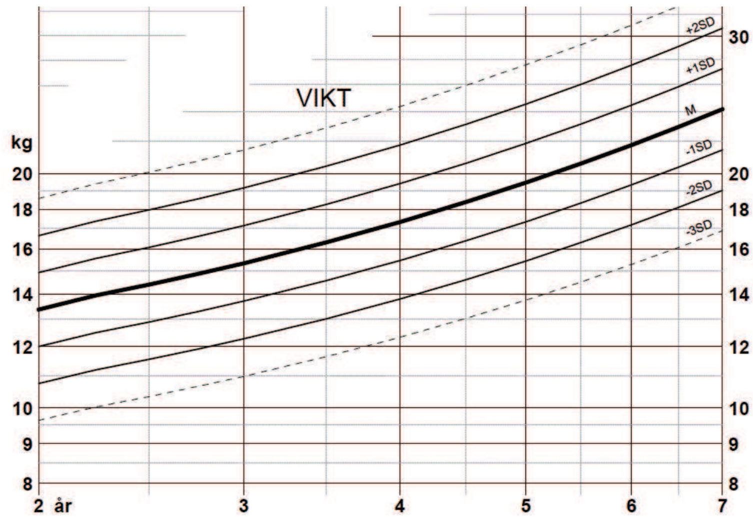
   (c) Vad är sannolikheten att hon behöver redovisa exakt två gånger?

   (d) Om hon nu "drabbas" av slumpen och får redovisa vid första tillfället, vad är då sannolikheten att hon ska behöva redovisa ytterligare minst en gång vid de fem återstående tillfällena?

   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.

[2] Notera, enbart "tycks", för att bevisa något i matematiken krävs mer än att bara studera två specialfall. Guldstjärna, men ingen ytterligare poäng, till den som bevisar att Lindas observation är en allmän sanning.

3. Vikten för pojkar i Sverige i åldrarna två till sju år följer nedanstående diagram.[3]



Här fokuserar vi på femåringarna.

(a) Ange medelvikt och standardavvikelse för femåriga pojkar genom att läsa ut detta ur diagrammet. Eftersom det inte går att se exakta värden i ett diagram räcker det med ungefärliga värden.

(b) Hur stor andel av pojkarna väger mindre än 15 kg?

(c) Om vi vill studera den grupp av pojkarna som väger mest, den översta procenten, vilken vikt svarar det mot att de måste väga mer än?

4. Olle är mycket fäst vid en kristallvas som han har ärvt av sin mormor. För att ta reda på dess vikt med stor säkerhet väger han den fem gånger på en precisionsvåg han införskaffat nyligen, och får resultaten 2383 g, 2385 g, 2382 g, 2384 g, 2385 g. I databladet för vågen står angivet att mätfelet är 5 g för vikter mellan 1 och 10 kg. Olle inser att vad som avses är att felet kan betraktas som en stokastisk variabel med väntevärdet noll (dvs inget systematiskt fel) och standardavvikelse 5 g.

(a) Bestäm ett 99% konfidensintervall för kristallvasens vikt åt Olle.

(b) Intervallet blir lite för stort, tycker Olle, och sneglar på databladets uppfattning om mätnoggrannheten. Bättre än så verkar ju hans mätningar vara. Bestäm ett nytt 99% konfidensintervall som inte tar standardavvikelsen hos mätningarna för given.

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.

(a) Signifikansnivån, α, vid en hypotesprövning är sannolikheten att godta en felaktig nollhypotes.

(b) Signifikansnivån, α, vid en hypotesprövning är sannolikheten att förkasta en sann nollhypotes.

(c) Sannolikheten för godtagandefel kan beräknas ur styrkefunktionen.

[3] Återgiven med tillstånd från PC PAL, www.tillvaxtkurvor.se.

(d) Det så kallade p-värdet anger hur stor andel av nollhypoteserna som inte kan förkastas, givet att hälften av dem är sanna.

(e) Det så kallade p-värdet anger sannolikheten att få ett resultat som är åtminstone lika extremt som det man fick, givet att nollhypotsen är sann.

(f) Det så kallade p-värdet anger sannolikheten för att nollhypotesen är falsk.

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.

6. En form av cement baseras på mineralet trikalciumsilikat, ofta förkortat C_3S, 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 utför Liza en försöksserie med tolv oberoende experiment där man lät andelen trikalciumsilikat variera mellan 27% och 70%. Resultatet blev följande mätserie:

0,27	0,30	0,32	0,32	0,40	0,47	0,51	0,55	0,56	0,56	0,67	0,70
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 värmen (angett i Joule per gram cement).

En linjär regressionsmodell ansätts, och ur Excel erhåller Liza:

UTDATASAMMANFATTNING

Regressionsstatistik	
Multipel-R	0,802
R-kvadrat	0,644
Justerad R-kvadrat	0,608
Standardfel	2,259
Observationer	12

	Koefficienter	*Standardfel*	*t-kvot*	*p-värde*	*Nedre 95%*	*Övre 95%*
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) Hur väl korrelerar andelen trikalciumsilikat med uppmätt värmeutveckling, dvs vad är korrelationskoefficienten?

(b) Nästa vecka planerar Liza att göra en ny tillblandning av 10 kg cement, och ha 50% trikalciumsilikat. Ange ett 95% konfidensintervall för hur stor värmeutveckling hon då kan förvänta sig.