LINKÖPINGS UNIVERSITET

Institutionen för teknik och naturvetenskap
Michael Hörnquist, 011-363381

Tentamen (TEN1) i **Statistik och sannolikhetslära** (TNIU66) 3 juni 2024, 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 sannolikhetslära[1]
- 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, varav minst 12 poäng 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!

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

1. Betrakta de fem talen $-5, -5, 0, 5, 5$.

   (a) Bestäm deras medelvärde och stickprovsstandardavvikelse.

   (b) Rita ett stolpdiagram för talen.

   (c) Bestäm fem nya tal så att deras medelvärde blir 5 men stickprovsstandardavvikelsen inte ändras från de ursprungliga talen.

   (d) Rita ett stolpdiagram för de fem nya talen.

   Stolpdiagrammen kan ritas på det medföljande kalkylbladet där du motiverar dina svar.

2. Sigrid är en skicklig dörrförsäljare, som kan få folk att köpa det mesta. Just nu säljer hon mobilabonnemang, och lyckas hon bara få någon att öppna dörren när hon ringer på är sannolikheten att hon får sålt ett abonnemang 25%. För att få en anständig lön krävs att hon säljer minst tre abonnemang i timmen.

   Bestäm sannolikheten att hon under en timme när hon lyckas få tio personer att öppna för henne säljer minst tre abonnemang.

3. Låt slumpvaribeln X_1 vara antal ögon som kommer upp vid kast med en vanlig sexsidig tärning, X_2 slumpvariabeln för en annan likadan tärning, och $X_i, i = 3, \cdots, 50$ slumpvariabler för ytterligare 48 likadana tärningar.

   (a) Bestäm $\Pr(X_1 \leq 2)$.

   (b) Bestäm $\Pr(\frac{1}{2}(X_1 + X_2) \leq 2)$.

   (c) Bestäm $\Pr(\frac{1}{50}\sum_{i=1}^{50} X_i \leq 2)$.

4. Inför försäljning av en livsmedelsaffär försöker sig mäklaren Kirsten på att uppskatta hur stor den genomsnittliga försäljningen är. Hon kontrollerar dagskassan under sex dagar, och finner att försäljningen i medeltal har varit 32 768 kronor, med en spridning (standardavvikelse) om 4096 kronor.

   Antag dagsförsäljningarna är normalfördelade och oberoende, bestäm ett 95% dubbelsidigt konfidensintervall för hur mycket affären säljer för i medel per dag.

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) Det så kallade p-värdet anger sannolikheten att nollhypotesen är sann.

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

   (c) Det så kallade p-värdet tas fram med värden som om nollhypotesen vore sann.

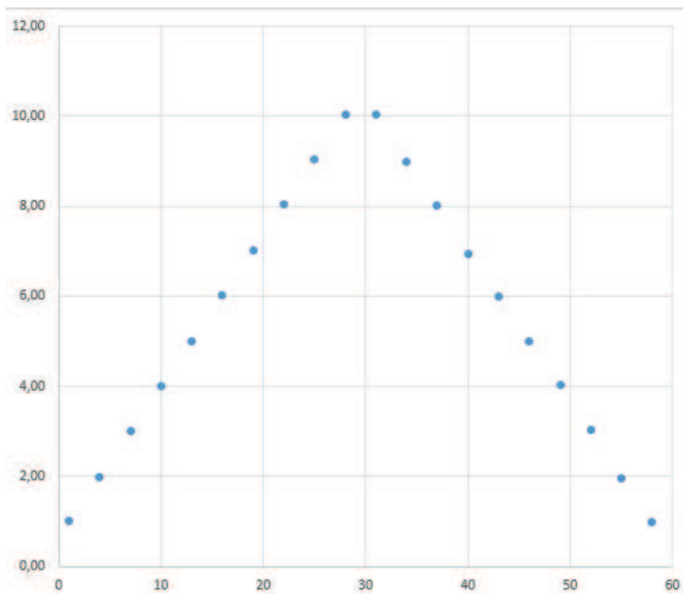
   (d) Ett sätt att göra ett dubbelsidigt konfidensintervall hälften så brett är att dubblera storleken på stickprovet.

   (e) Ett sätt att göra ett dubbelsidigt konfidensintervall hälften så brett är att dubblera konfidensnivån.

   (f) Ett sätt att göra ett dubbelsidigt konfidensintervall hälften så brett är att byta till enkelsidigt.

   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. Du får i uppgift att analysera en datamängd av talpar (x_i, y_i) för att se om det kan finnas åtminstone något approximativt samband mellan x-värdena och y-värdena. Det första du gör är (givetvis?) att rita spridningsdiagram, vilket ger:



Det ser ju ut att vara två räta linjer, och därför går du vidare med linjär regression för att låta datorn göra jobbet åt dig. Värdena matas in i Excel och dess Dataanalysverktyg, varvid följande resultat erhålls:

UTDATASAMMANFATTNING						
Regressionsstatistik						
Multipel-R	0,002036847					
R-kvadrat	4,14874E-06					
Justerad R-kvadrat	-0,055551176					
Standardfel	3,026743802					
Observationer	20					
ANOVA						
	fg	*KvS*	*MKv*	*F*	*p-värde för F*	
Regression	1	0,000684136	0,000684136	7,46777E-05	0,993200123	
Residual	18	164,9012048	9,161178045			
Totalt	19	164,901889				
	Koefficienter	*Standardfel*	*t-kvot*	*p-värde*	*Nedre 95%*	*Övre 95%*
Konstant	5,49792381	1,337961749	4,109178618	0,000658634	2,686970482	8,308877137
X-variabel 1	-0,000338095	0,03912403	-0,00864163	0,993200123	-0,082534632	0,081858442

Svara nu på följande:

(a) Vilken ekvation beskriver den räta linjen som är bästa anpassningen till datamängden?

(b) Ange konfidensintervallet som med 95% sannolikhet täcker riktningskoefficienten.

(c) Hur bra är anpassningen, dvs vilket värde har förklaringsgraden?

(d) Linjär regression *borde* fungera bra för räta linjer, kan man tycka. Vad kan göras här för att det skall bli meningsfullt att använda linjär regression?