

Tentamen i TNIU66, **Statistik och sannolikhetslära**, 15 mars 2018, kl. 14.00 – 18.00.

---

**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 sannolikhetslära<sup>1</sup>
- 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. 15 och 16.30. 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.

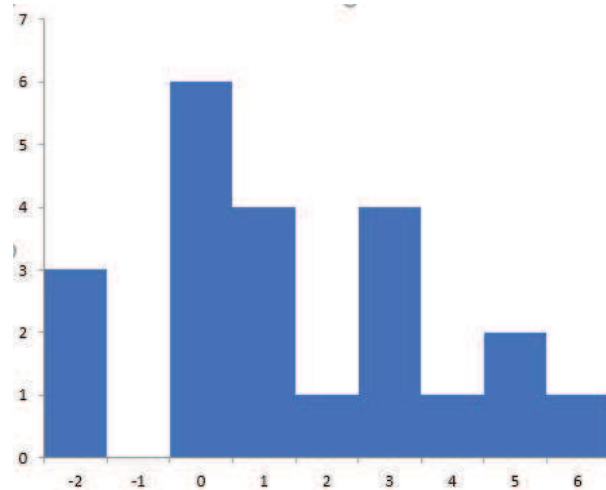
Redovisningen av uppgifterna skall vara sådan att det klart framgår vad du gör. Det innebär att förklarande text skall finnas med. Att enbart ge ett svar utan motivering, även om det är korrekt, är normalt inte tillräckligt för att uppgiften skall anses vara nöjaktigt löst. Å andra sidan behöver inte 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!**

---

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

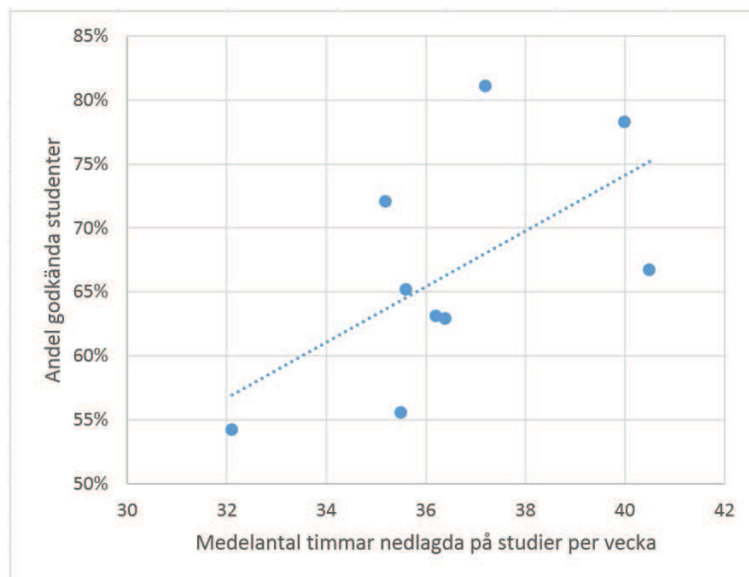
1. Nedanstående histogram visar fördelningen hos en mängd heltal. Bestäm mängdens medelvärde och median.



2. Den nyöppnade brödrestaurangen Underground gästas av de tre vännerna Margareta, Eva och Thomas. De har inget lätt val framför sig, då man här får välja mellan fem olika bröd, åtta olika fyllningar och tre olika dressingar. I en måltid ingår ett bröd, ett valfritt antal fyllningar och en dressing. Margareta är lite småmåttad och önskar enbart en fyllning, utöver bröd och dressing, medan Eva och Thomas önskar två fyllningar var, samt bröd och dressing. Bestäm:
- Antal olika kombinationer Margareta kan välja.
  - Antal olika kombinationer Eva kan välja. Hon väljer alltid två olika fyllningar, men bekymrar sig inte om vilken som läggs i brödet först.
  - Antal olika kombinationer Thomas kan välja. Han väljer alltid två olika fyllningar, och anser att "ost och skinka" är något helt annat än "skinka och ost", dvs annan ordning på fyllningarna innebär en ny kombination.
3. Nedan står  $X$  och  $Y$  för stokastiska variabler, med fördelningar som anges i förutsättningen för varje påstående. 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.
- Om  $X \in \text{Bin}(10; 0, 2)$  är  $P(X > 10) = 0$ .
  - Om  $X \in \text{Po}(10)$  är  $P(X > 10) = 0$ .
  - Om  $X \in N(22, 3, 2, 4)$  är en modell för viktfordelningen hos barn i förskoleklass, så gäller att  $P(X < 0) = 0$ .
  - Om  $X \in N(22, 3, 2, 4)$  är en modell för viktfordelningen hos barn i förskoleklass, så gäller att  $P(X > 1000) = 0$ .
  - Om  $X \in \text{Exp}(5)$  och  $Y \in \text{Exp}(2)$  så blir summa  $Z = X + Y \in \text{Exp}(7)$ .
  - Om  $X \in \text{Exp}(5)$  och  $Y \in \text{Exp}(2)$  så får deras summa  $Z = X + Y$  väntevärdet  $E(Z) = 7$ .

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.

4. Lisa samlar på stenar med så lika vikt men olika form som möjligt, samlingen har med tiden blivit ansenlig och rymmer idag tusentals stenar. Hon önskar bestämma deras medelvikt nu-förtiden, och drar därför på måfå ett stickprov med 20 stycken stenar och väger dem var och en för sig. Resultatet hon beräknar blir ett medelvärde om 57,5 gram och en stickprovsstandardavvikelse om 5,0 gram.
- Bestäm ett 95 procents konfidensintervall för var medelvikten för alla stenar i stensamlingen kan ligga.
  - Hur många fler vägningar skulle hon behöva göra om hon önskar få ett hälften så långt konfidensintervall (med bibehållen konfidensnivå, 95 procent)?
  - Fritz tycker att det är mer rationellt att istället väga de 20 kulorna tillsammans och sedan bara dela med antalet för att få medelvikten. Om han gör så, och finner att de tillsammans väger 1,15 kg, vilken slutsats kan han då dra om medelvikten, och med vilken säkerhet?
5. Det inte så välkända läkemedelsföretaget ”Friskt och sunt är inte dumt”, FOSÄID, har utvecklat en ny huvudvärksmedicin som de önskar testa. Bästa medicinen som säljs idag botar 65% av patienterna, så för att kunna sälja den nya medicinen behöver FOSÄID kunna statistiskt bevisa att deras nya medicin är bättre än så. De låter därför utföra ett dubbelblindtest enligt noggrant utprövad erfarenhet, och finner att av 128 deltagare i testet blev 90 av med sin huvudvärk.
- Genomför ett hypotestest på nivån  $\alpha = 5\%$  för att se om angivet resultat är tillräckligt för att FOSÄID ska kunna hävda att deras medicin är den bästa på marknaden. Förklara tydligt vilken slutsats du drar av hypotestestet.
6. Sambandet mellan nedlagd tid på studier och resultat är inte helt enkelt att studera. Ett försök är dock att se vilken genomsnittlig studietid studenter anger att de har lagt ned, och jämföra det med vilken andel som klarar en av de tyngre kurserna under samma läsperiod. Sådan data samlades in för åren 2007 – 2015 för ett utbildningsprogram vid ett större svenskt universitet och resultatet blev i spridningsdiagram (med regressionslinjen utritad):



Värdena matades in i Excel och dess Dataanalysverktyg, varvid följande utdata erhöles:

UTDATASAMMANFATTNING						
<i>Regressionsstatistik</i>						
Multipel-R	0,600631					
R-kvadrat	0,360758					
Justerad R-kvadrat	0,269437					
Standardfel	0,078816					
Observationer	9					
<i>ANOVA</i>						
	<i>fg</i>	<i>KvS</i>	<i>MKv</i>	<i>F</i>	<i>p-värde för F</i>	
Regression	1	0,02454	0,02454	3,95046	0,08720	
Residual	7	0,04348	0,00621			
Totalt	8	0,06802				
	<i>Koefficienter</i>	<i>Standardfel</i>	<i>t-kvot</i>	<i>p-värde</i>	<i>Nedre 95%</i>	<i>Övre 95%</i>
Konstant	-0,13031	0,40122	-0,32480	0,75482	-1,07904	0,81841
X-variabel 1	0,02179	0,01096	1,98758	0,08720	-0,00413	0,04771

Dessa data har använts på en tidigare tentamen i denna kurs, men uppgifterna som nu kommer är nya.

- Hur väl korrelerar andelen godkända studenter med medelantalet nedlagda timmar, dvs, vad är korrelationskoefficienten?
- Genomför ett hypotestest på signifikansnivån 5% för att se om undersökningen visar att det finns ett positivt samband ( $\beta > 0$ ) mellan andel godkända studenter och medelantal nedlagda timmar. Förklara tydligt vilken slutsats du drar av hypotestestet.