

## Övningstentamen – Statistiska Metoder – ITHS AI-22

**Om tentan:** Totalt kan man få 50 p. Gränsen för G är 25p och gränsen för VG är 37p.

**Tillåtna hjälpmedel:** Bifogad formelsamling med anteckningar och miniräknare med tomt minne.

**Instruktioner:** För uppgifterna 1-3 krävs endast svar. För uppgifterna 4-7 krävs fullständiga uträkningar, motiveringar och redogörelser för full poäng. Skriv varje lösning på separata blad. Namnge alla inlämnade blad tydligt i övre högra hörnet.

**OBS:** Skriv ner alla uträkningar tydligt och noggrant! Delpoäng kan ges även vid fel svar.

1. Låt de två händelserna  $A$  och  $B$  vara två händelser med sannolikheterna  $P(A) = 0.3$ ,  $P(B) = 0.5$  samt  $P(A \cap B) = 0.14$   
Beräkna följande sannolikheter:
  - a.  $P(A^c)$  (1p)
  - b.  $P(A \cup B)$  (1p)
  - c.  $P(B|A)$  (1p)
  - d. Är händelserna  $A$  och  $B$  oberoende? Motivera ditt svar (1p)
2. Låt  $X$  vara ett stickprov med värden  $\{1, 5, -3, 0, 5, -1, 8, 3, 12\}$ . Beräkna följande deskriptiva mått:
  - a. Stickprovsmedelvärde (2p)
  - b. Stickprovsstandardavvikelse (2p)
  - c. Median (1p)
  - d. Typvärde (1p)
3. En högstadieskola för statistik över hur många timmar per månad eleverna skolkar. Datan består av en csv-fil med en kolumn där antalet skolkade timmar summeras för varje månad. Skolans rektor är intresserad av att undersöka hur skolkandet ser ut med några deskriptiva mått, samt att räkna på några sannolikheter. Rektorerna tänker använda sig av Python-paketerna Pandas, NumPy och SciPy, och försöker nu påminna sig själv om hur de grundläggande funktionerna fungerar. Rektorerna har lyckats importera sitt dataset som en Pandas DataFrame med namnet 'skolk'. Hjälp rektorerna att utföra följande uppgifter (svara med en enkel kodrad ur valfritt paket)
  - a. Utifrån beskrivningen, beräkna stickprovsmedelvärdet av antalet skolkade timmar med Pandas eller NumPy. (1p)
  - b. Utifrån beskrivningen, beräkna stickprovsstandardavvikelsen av antalet skolkade timmar med Pandas eller NumPy. (1p)
  - c. Rektorerna antar att skolkandet är en normalfördelad slumpvariabel  $X$ , med ett medelvärde på 67, och en standardavvikelse på 36. Uttryck sannolikheten att skolkandet en given månad är högre än 75 timmar, samt skriv en kodrad i SciPy som beräknar den. Antag att du redan importerat alla de fördelningar vi talat om i kursen under deras respektive fördelningsnamn. Inget uträknat svar krävs. (3p)
  - d. Enligt en annan del av undersökningen visar det sig att i medeltal 32 studenter skolkar varje månad. Låt  $X$  vara en slumpvariabel som räknar hur många studenter skolkar en given månad. Uttryck sannolikheten att högst 30 studenter skolkar en given månad, samt skriv en kodrad i SciPy som beräknar den. Antag att du redan importerat alla de fördelningar vi talat om i kursen under deras respektive fördelningsnamn. Inget uträknat svar krävs. (3p)

4. Låt  $X$  vara:
- En diskret slumpvariabel med fördelningen:  $X \sim \text{Geom}(0.2)$ . Beräkna väntevärdet  $E(X)$ , standardavvikelsen  $S(X)$ , samt sannolikheten att  $X$  antar värdet 3. (4p)
  - En kontinuerlig slumpvariabel med fördelningen:  $X \sim \text{Uniform}(0,5)$ . Beräkna väntevärdet  $E(X)$ , standardavvikelsen  $S(X)$ , samt sannolikheten att  $X$  antar värden större än 3. (4p)
5. För att bygga en enkel linjär regressionsmodell har data för responsvariabeln,  $y$ , och förklaringsvariabeln,  $x$ , samlats in. Det samlades in 30 observationer som gav stickprovskovariansen 30 mellan respons- och förklaringsvariabeln och stickprovsvariansen 10 för förklaringsvariabeln. Stickprovsmedelvärdet för responsvariabeln var 10 och stickprovsmedelvärdet för förklaringsvariabeln var 5.
- Vad blir värdena på prediktionsmodellens intercept och lutning (4p)?
  - Vad blir det predikterade  $y$ -värdet för  $x = 4$ ? (2p)
6. En apputvecklare arbetar med en smartphone-applikation som hjälper användaren att överblicka sin hushållsbudget. Nu testar utvecklaren att skicka ut en uppdatering till sina användare. 800 användare väljs slumpmässigt ut för testet och delas in i två lika stora grupper, där de två grupperna får varsin version (A respektive B) av uppdateringen. Båda grupperna tillfrågas sedan efter en vecka om de var nöjda med applikationen i fråga, vilket gav resultatet 178 nöjda användare för version A, respektive 158 nöjda användare för version B. Konstruera ett hypotestest som testar på signifikans-nivå 0.05 huruvida version A gav fler nöjda användare än version B. (8p)
7. Du har bjudit in 125 gäster till ett kalas där du skall bjuda på fördrink, och tänker köpa in alkoholfritt mousserande vin som alternativ. Du uppskattar sannolikheten att en gäst efterfrågar alkoholfritt vin som  $p = 0.2$ . Du antar att en flaska räcker till 6 personer.
- Om du köper in 4 flaskor vin, hur sannolikt är det att alla flaskorna tar slut? Förklara alla antaganden och eventuella approximationer. (4p)
  - Beräkna ett 95% konfidensintervall för förbrukningen av alkoholfritt vin. (4p)
  - Hur mycket vin behöver du köpa in för att vara 95% säker på att vinet räcker? Svara utifrån konfidensintervallet i (b) (2p)