

# TENTAMEN I TILLÄMPAD MATEMATIK OCH STATISTIK FÖR IT-FORENSIK. DEL 2: STATISTIK

7.5 HP

april, 2022

**Maxpoäng:** 30p. **Betygsgränser:** 12p: betyg 3, 18p: betyg 4, 24p: betyg 5.

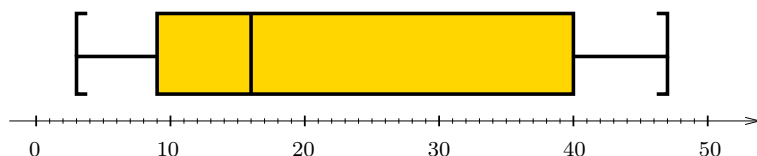
**Hjälpmedel:** Miniräknare TI-30Xa samt formelsamling som delas ut av vakterna.

**Kursansvarig:** Eric Järpe, telefon 0729-77 36 26.

Till uppgifterna skall *fullständiga lösningar* lämnas. Lösningarna ska vara *utförligt* redovisade! Bladen ska lämnas in i rätt ordning. Svara alltid med 4 decimalers noggrannhet om ej annat anges. Lösningar kommer finnas på internet: <http://dixon.hh.se/erja/teach> → Matematik och statistik för IT-forensik.

1. Vid en undersökning av datorvana i förhållande till ålder i en kommun görs observationer av 17 individers ålder,  $X$ , och hur mycket de använder dator, tablet eller mobil,  $Y$  uttryckt i timmar per vecka. För dessa data görs boxplottar och en regressionslinje beräknas.

(a) [2:1] Boxplotten av variabeln  $Y$  återfinns i diagrammet:



Beräkna kvartilavståndet för variabeln  $Y$  baserat på dessa data. (2p)

- (b) [2:1] Från observationerna  $x_1, x_2, \dots, x_{17}$  och  $y_1, y_2, \dots, y_{17}$  beräknas även  $\sum_{i=1}^{17} x_i = 629$ ,  $\sum_{i=1}^{17} x_i^2 = 28\,717$ ,  $\sum_{i=1}^{17} y_i = 391$ ,  $\sum_{i=1}^{17} y_i^2 = 13\,049$  och regressionslinjen

$$y = 51 - 0.75x.$$

Beräkna förklaringsgraden för denna modell. (3p)

2. Ett misstänkt kriminellt nätverk består av 11 personer. Av dessa har 9 personer ständig tillgång till mobiltelefon med appen Ghost installerad. Totalt har 5 av de 11 personerna en ledarfunktion i nätverket varav 4 har tillgång till mobil med Ghost. Vad är då

(a) [2:1] sannolikheten att en slumpmässigt person ur nätverket är en av ledarna? (1p)

(b) [2:1] den betingade sannolikheten att tre slumpmässigt valda personer ur nätverket är ledare givet att minst två av dem har tillgång till mobil med Ghost? (4p)

3. Låt  $X \in Poi(\lambda)$  och beräkna

(a) [2:2]  $E(X^2)$  om  $\lambda = 2$ . (2p)

(b) [2:2] approximativt  $P(1.1 \leq \bar{X} \leq 1.2)$  om  $\bar{X}$  är medelvärdet  $\frac{1}{100} \sum_{i=1}^{100} X_i$  baserat på stickprovet  $X_1, X_2, \dots, X_{100}$  av  $X$  och  $\lambda = 1$ . (3p)

(c) [2:2]  $P(X \leq 2)$  om  $\lambda = 1 + Y$  där  $Y \in Bin(2, 0.7)$ . (5p)

4. [2:3] I en studie<sup>1</sup> om personer som aldrig använt internet var det i ett stickprov 29 500 personer som var yngre än 65 år av totalt 268 400 personer. Kan man på någon rimlig signifikansnivå bevisa att denna andel är mindre än 11%? Vad blir  $p$ -värdet? (3p)

5. [2:3] Antalet personer som inte vet om de har ett säkerhetsprogram installerat på sin smartphone är vid en stor undersökning<sup>2</sup> i tusental:

481.6 423.0 374.2 447.7 392.0 319.3

per åldersgrupp. Kan man på 5% signifikansnivå bevisa att det förväntade antalet personer per åldersgrupp är fler än 300 000 personer? (3p)

6. [2:3] För att vid regressionsanalys avgöra om det är olämpligt med en linjär modell testas om residualerna kan bevisas ej normalfördelade. Följande observationer av residualerna görs:

<i>Klass</i>	1	2	3	4	5	6
<i>Antal residualer</i>	10	6	2	4	1	7

där klass 1 är intervallet  $(-\infty, -2]$ , klass 2  $(-2, -1]$ , klass 3  $(-1, 0]$ , klass 4  $(0, 1]$ , klass 5  $(1, 2]$  och klass 6  $(2, \infty)$ . Som klassmitter använd  $-2.5$ ,  $-1.5$ ,  $-0.5$ ,  $0.5$ ,  $1.6$  och  $2.5$ . Finns det någon anledning att misstro normalfördelningsantagandet vid regressionsanalysen? Gör ett hypotestest på 5% signifikansnivå. (4p)

*LYCKA TILL!*

<sup>1</sup>Data från SCB från 2021 om levnadsförhållanden och befolkningens IT-användning.

<sup>2</sup>Data från SCB från 2020 om IT-säkerhet – se <https://www.scb.se/hitta-statistik/statistik-efter-amne/levnadsforhallanden/levnadsforhallanden/befolkningens-it-anvandning>.