

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

7.5 HP

maj, 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 samlar man in data för variablerna  $X = ECTS$  betyg med utfallsrum  $\Omega_X = \{A, B, C, D, E, Fx, F\}$ ,  $Y = \text{ålder}$  med utfallsrum  $\Omega_Y = \{0, 1, 2, \dots\} = \mathbb{Z}^+$  och  $Z = \text{examen}$  med utfallsrum  $\Omega_Z = \{\text{Tekn. kand., Tekn. mag., Fil. kand., Fil. mag., Med. kand., Med. mag., Civ. ing., Civ. ek., Annan examen}\}$ . Ange om
  - (a) [2:1] variabeln  $X$  mäts i nominalskala, ordinalskala, intervallskala eller kvot-skala. (1p)
  - (b) [2:1] variabeln  $Y$  mäts i nominalskala, ordinalskala, intervallskala eller kvot-skala. (1p)
  - (c) [2:1] variabeln  $Z$  mäts i nominalskala, ordinalskala, intervallskala eller kvot-skala. (1p)
2. [2:1] Fridolina tar reda på vad hennes kollegor har i månadslön och får följande data i kSEK:  
$$71.5 \quad 37.3 \quad 28.0 \quad 29.3 \quad 32.4 \quad 29.2 \quad 115.8 \quad 31.9 \quad 25.5 \quad 30.6$$

Vilka blir morrhåren vid en boxplot av dessa data? (3p)
3. [2:1] Låt  $P(A|B) = 0.65$ ,  $P(A|B^C) = 0.95$ ,  $P(B) = 0.15$  och beräkna  $P(B|A^C)$ . (4p)
4. Antag att Anders varje dag i sin spargris sparar ett antal kronor som är binomialfördelat med  $n = 100$  och  $p = 0.1$  oberoende från dag till dag. Vad är då
  - (a) [2:2] sannolikheten att Anders under en dag sparar minst 3 kronor? (2p)
  - (b) [2:2] approximativt sannolikheten att Anders efter ett års sparande får råd att köpa sin drömcykel för 3500:-? (3p)

5. **[2:2]** För att bestämma hur många dagars internettrafik som ska observeras vid utformningen av ett pentest beräknas hypotetiskt ett konfidensintervall för det förväntade antalet cyberattacker per dag under antagandet att det är normalfördelat med standardavvikelse  $\sigma = 5$ . Hur många dagars trafik ska observeras för att ett 99% konfidensintervall ska bli högst 7 dagar brett? (4p)

6. **[2:3]** Vid en studie av cyberattacker vill man argumentera för att andelen signaler som indikerar attack i förhållande till andelen ofarliga signaler är alltför litet och att adaptiva metoder därför är nödvändiga för att inte få för många falsklarm. Man observerar 1 283 522 signaler varav 1 178 kan anses vara signaler från attacker. Kan man med dessa data bevisa att sannolikheten att observera attacksignaler är mindre än 1‰ på signifikansnivå 1‰? Beräkna även  $p$ -värdet. (3p)

7. För en grupp om 100 personer varav 43 män och 57 kvinnor har man observerat deras skostorlekar.

(a) **[2:3]** Kan man på 5% signifikansnivå bevisa att män har större förväntad skostorlek än kvinnor om

$$\sum_{i=1}^{43} m_i = 1\,847, \quad \sum_{i=1}^{43} m_i^2 = 80\,591, \quad \sum_{i=1}^{57} k_i = 2\,062 \quad \text{och} \quad \sum_{i=1}^{57} k_i^2 = 76\,116$$

(där  $m_i$  är männens och  $k_i$  är kvinnornas skostorlekar)? Redovisa alla antaganden du gör. (4p)

(b) **[2:3]** Man mäter även kvinnornas IQ och finner att 38 av kvinnorna har  $\text{IQ} > 100$  och 23 av kvinnorna hade skostorlek  $> 37$  varav 16 kvinnor hade  $\text{IQ} > 100$ . Kan man på 5% signifikansnivå bevisa att variablerna  $A = \{\text{IQ} > 100\}$  och  $B = \{\text{skostorlek} > 37\}$  är beroende bland kvinnorna? (3p)

*LYCKA TILL!*