Tentamen i Tillämpad Matematik och statistik för IT-forensik. Del 2: Statistik

7.5 hp

24 april, 2019

Maxpoäng: 30p. **Betygsgränser**: 12p: betyg 3, 18p: betyg 4, 24p: betyg 5.
Hjälpmedel: Miniräknare och formelsamling. **Telefonvakt**: Mats Gunnarsson, tfn 031-16 72 11.

Alla svar skall ges med 4 decimalers noggrannhet där ej annat anges. Till uppgifterna skall *fullständiga lösningar* lämnas. Lösningarna ska vara *utförligt* redovisade! Varje lösning ska börja överst på nytt papper. Endast en lösning per blad. Lösningar kommer finnas på internet: `http://dixon.hh.se/erja/teach` $\to$ `Matematik och statistik för IT-forensik.`

1. *Antalet dataintrång* i västra Sverige per månad under 2017 var[1]

Månad	Jan	Feb	Mars	April	Maj	Juni
Antal intrång	242	128	312	80	111	79

   (a) **[2:1]** Beräkna medianen för dessa data. (2p)

   (b) **[2:3]** Beräkna ett 90% konfidensintervall för *andelen* månader som intrången per månad var fler än 90. (2p)

   (c) **[2:3]** Gör ett hypotestest på 1% signifikansnivå av om det förväntade *antalet* intrång per månad är fler än 90. Vad blir p-värdet (ungefär)?(3p)

   Motsvarande data för *antal bostadsinbrott* var[1]

Månad	Jan	Feb	Mars	April	Maj	Juni
Antal inbrott	802	709	756	622	557	506

   (d) **[2:3]** Skriv siffrorna från dessa båda dessa datamängder på en enda lång rad:

$$2, 4, 2, 1, 2, 8, 3, 1, \ldots, 7, 5, 0, 6 \text{ (34 siffror)}$$

   och avgör om de ej följer Zipf-fördelningen[2] med parameter $a = 1.05$ med ett lämpligt hypotestest på 5% signifikansnivå. (5p)

   (e) **[2:1]** Beräkna förklaringsgraden vid en regressionsanalys med *Antal inbrott* som kovariat och *Antal dataintrång* som respons. (4p)

2. Låt $X \in N(3, \sigma^2)$ och beräkna

   (a) **[2:2]** $P(X \leq 4.3)$ om $\sigma^2 = 1$ (2p)

   (b) **[2:2]** σ^2 om $P(1.9 \leq X \leq 4.1) = 0.97$ (3p)

[1] Detta datamaterial är hämtat från Brottsförebyggande rådet: `http://bra.se/statistik/kriminalstatistik/anmalda-brott.html`.

[2] *Zipf-fördelningen* är en sannolikhetsfördelning för förekomst av siffror bland ett stort datamaterial. Det påstås att bland siffrorna $0, 1, 2, \ldots, n$ en slumpmässigt vald siffra är den X:te mest vanliga med sannolikhet $P(X = x) = C(x + 1)^{-a}$ där C är en konstant som gör att sannolikheten för hela utfallsrummet blir 1.

3. Påskhönan Agda värper ett Poissonfördelat antal ägg med parameter $\lambda = 5$ under påskhelgen. Vad är sannolikheten

   (a) **[2:2]** att hon värper minst 3 ägg under påsken? (2p)

   (b) **[2:2]** approximativt att hela hönsgården tillsammans värper minst 43 ägg om var och en av de 10 hönorna värper Poissonfördelat med $\lambda = 5$ och oberoende av varandra? (3p)

4. **[2:1]** Jöns Påskhare har bjudit Kalle Kanin och hans tre kompisar på påskmiddag. Alla gästerna vill ha hårdkokt ägg och det vill Jösse också så Jösse går ut till kylskåpet (dit gästerna inte kan se) för att hämta detta. I kylen finns fyra löskokta och sex hårdkokta ägg men det är dock ingen ordning på vilka som är lösa och vilka som är hårda så Jösse får chansa. Han kan dock minska risken att ta fel genom att snurra alla äggen på ett bord (de två mest löskokta snurrar då långsammare och kan då elimineras från urvalsprocessen). Ibland (dvs med 30% sannolikhet) har dock Jösse bråttom och då chansar på fem ägg utan att snurra.

   Då Jösse kommer in matsalen, serverar äggen och de börjar äta märker Kalle att det blev just fem hårdkokta, precis som beställt. Hjälp Kalle att med vetskap om detta beräkna den betingade sannolikheten att Jösse snurrat äggen. (4p)

LYCKA TILL!