Lösningar till tentamen i Tillämpad Matematik och statistik för IT-forensik. Del 2: Statistik

7.5 hp

15 januari, 2014 kl. 9.00 – 13.00

Maxpoäng: 30p. **Betygsgränser**: 12p: betyg G, 21p: betyg VG.
Hjälpmedel: Typgodkänd miniräknare samt formelsamling som medföljer tentamenstexten.
Kursansvarig: Eric Järpe, telefon 0702-822 844, 035-16 76 53.

1. Antag att P(Arne har sitt paraply med sig) $= 0.25$, P(Det regnar) $= 0.4$ och att P(Arne har sitt paraply med sig och det regnar) $= 0.2$. Vad är då
   (a) sannolikheten att Arne har sitt paraply med sig eller att det regnar? (3p)
   (b) den betingade sannolikheten att Arne har sitt paraply med sig givet att det regnar? (3p)

   Lösning:
   (a) P(Arne har paraply) $= 0.25$, P(Det regnar) $= 0.4$ och P(Arne har paraply och det regnar) $= 0.2 \;\Rightarrow\; P$(Arne har paraply $\cup$ det regnar) $= P$(Arne har paraply)$+$ P(Det regnar) $-$ P(Arne har paraply $\cap$ det regnar) $= 0.25 + 0.4 - 0.2 = 0.45$.
   (b) $P(\text{Arne har paraply} \,|\, \text{Det regnar}) = \frac{P(\text{Arne har paraply} \cap \text{Det regnar})}{P(\text{Det regnar})} = \frac{0.2}{0.4} = 0.5.$ $\square$

2. En tjuv har stulit 100 mobiltelefoner. För varje mobiltelefon har tjuven 3 chanser att gissa den 4-siffriga PIN-koden. Om man gissar fel 3 gånger har man oändligt många chanser att gissa den 10-siffriga PUK-koden. Vad är tjuvens chans att lyckas
   (a) gissa PIN-koden på minst 1 av de 100 mobiltelefonerna? (3p)
   (b) knäcka koden till en mobiltelefon om tjuven använder ett datorprogram som gör 1 miljon försök att gissa PUK-koden om försöken att gissa PIN-koden misslyckas? (3p)

   Lösning:
   (a) P(gissa PIN på första försöket) $= \frac{1}{10\,000}$ så P(gissa PIN) $= P$(gissa på första)$+$ P(inte på första men på andra)$+P$(inte på första eller andra men på tredje) $=$ $\frac{1}{10\,000} + \frac{9\,999}{10\,000} \cdot \frac{1}{9\,999} + \frac{9\,999}{10\,000} \cdot \frac{9\,998}{9\,999} \cdot \frac{1}{9\,998} = \frac{1}{10\,000} + \frac{1}{10\,000} + \frac{1}{10\,000} = 0.0003$. Därmed är P(gissa PIN på minst 1 av 100) $= 1 - P$(inte gissa PIN på någon) $= 1 - (1-0.0003)^{100} = 1 - 0.9997^{100}$. (Så långt kan man lösa denna uppgift helt utan miniräknare. Vad 0.9997^{100} blir är inte huvudsaken.)
   (b) P(knäcka en telefon) $= P$(gissa PIN)$+P$(missa PIN men gissa PUK) $= 0.0003+$ $0.9997(P$(gissa PUK på första)$+P$(missa första men gissa på andra)$+P$(missa första och andra men gissa på tredje)$+\ldots+P$(missa alla tom nr 10^6-1 men gissa på nr 10^6)) $= 0.0003 + 0.9997 \cdot 10^{-10} \cdot 10^6 = 0.0003 + 0.9997 \cdot 0.0001 =$ $0.0003 + 0.00009997 = 0.00039997 \approx 0.0004$. (Även i denna uppgift är miniräknaren egentligen överfödig.) $\square$

3. Fotbollslaget IFK Göteborg slutade på tredje plats i allsvenskan efter fotbollssöongen 2013. Vid en enkät svarar 15 av 42 tränare att de rankar IFK Göteborg som 1:a eller 2:a efter 2014 års säsong.

   (a) Bilda ett 95% konfidensintervall för sannolikheten
   $\pi = P$(en slumpmässig tränare tror att IFK Göteborg skutar 1:a eller 2:a). (3p)

   (b) En annan fråga i enkäten gällde om någon allsvensk klubb skulle gå vidare till slutspel i någon av de europeiska kupperna och på denna fråga svarade 7 tränare ”ja” varav 5 svarat att de rankade IFK Göteborg som 1:a eller 2:a. Är de två åsikterna beroende? Gör ett test på 1% signifikansnivå. (3p)

Lösning:

(a) $p = \frac{15}{42} = 0.357$ (obs! avrunda inte delresultaten för mycket!) och $\lambda_{0.05/2}\sqrt{\frac{p(1-p)}{n}} = 1.96\sqrt{\frac{0.357(1-0.357)}{42}} = 0.144911$ så konfidensintervallet för π blir (0.21223, 0.50206).

(b) Låt A vara påståendet att ”tränarna tror att IFK slutar 1:a eller 2:a i tabellen” och B påståendet att ”tränarna tror att IFK går vidare till slutspel i de europeiska kupperna”. Vi vill då testa om A ej är oberoende av B, dvs

$$\begin{cases} H_0 : A \perp B \\ H_1 : A \not\perp B \end{cases}$$

Det kan vi göra med ett test av följande korstabell

		A: 1:a eller 2:a		
		Ja	*Nej*	
B: Vidare till slutspel	*Ja*	**5**	2	**7**
	Nej	10	25	35
		15	27	42

$$U = \frac{(n_{11}c_2 - n_{12}c_1)\sqrt{n}}{\sqrt{c_1 c_2 n_1 n_2}} = \frac{(5\cdot 27 - 2\cdot 10)\sqrt{42}}{\sqrt{15\cdot 27\cdot 7\cdot 35}} = 2.36$$

Detta värdde ska jämföras med normalpercentilen $\lambda_{\alpha/2} = \lambda_{0.005} = 2.575829 \not< 2.36$.

Svar: Nej, H_0 kan inte förkastas, dvs man kan inte visa att ”tränarnas rankning om de tror att IFK blir 1:a eller eller 2:a” är oberoende av ”tränarnas bedömning av IFK:s chans att gå vidare till slutspel i de europeiska kupperna”.

□

4. Låt $X \in N(3,2)$ (dvs väntevärdet är 3 och variansen är 2). Beräkna
   (a) $P(X > 2)$. (2p)
   (b) a och b sådana att $P(a \leq X \leq b) = 0.48$ och $3P(a \leq X \leq 3) = P(3 \leq X \leq b)$. (4p)

Lösning:

(a) $P(X > 2) = 1 - P(X \leq 2) = 1 - \Phi(\frac{2-3}{\sqrt{2}}) = 1 - \Phi(-0.7071) = \Phi(0.71) = 0.7611$.

(b)

$$\begin{aligned} 0.48 &= P(a \leq X \leq b) \\ &= P(a \leq X \leq 3) + P(3 \leq X \leq b) \\ &= P(a \leq X \leq 3) + 3P(a \leq X \leq 3) \\ &= 4P(a \leq X \leq 3) \end{aligned}$$

$$\Rightarrow \tfrac{0.48}{4} = P(a \leq X \leq 3) = \Phi(\tfrac{3-3}{\sqrt{2}}) - \Phi(\tfrac{a-3}{\sqrt{2}}) = 0.5 - \Phi(\tfrac{a-3}{\sqrt{2}})$$

$$\Rightarrow \Phi(\tfrac{a-3}{\sqrt{2}}) = 0.5 - \tfrac{0.48}{4} = 0.38 \Rightarrow \Phi(-\tfrac{a-3}{\sqrt{2}}) = 1 - 0.38 = 0.62$$

$$\Rightarrow -\tfrac{a-3}{\sqrt{2}} = 0.305 \Rightarrow a = -0.305\sqrt{2} + 3 = 2.5687$$

$P(3 \leq X \leq b) = 3P(a \leq X \leq 3) = 3 \cdot 0.12 = 0.36$

$$\Rightarrow 0.36 = \Phi(\tfrac{b-3}{\sqrt{2}}) - \Phi(\tfrac{3-3}{\sqrt{2}}) \Rightarrow \Phi(\tfrac{b-3}{\sqrt{2}}) = 0.36 + 0.5 = 0.86$$

$$\Rightarrow \tfrac{b-3}{\sqrt{2}} = 1.08 \Rightarrow b = 1.08\sqrt{2} + 3 = 4.5273.$$

Svar: $a = 2.569$ och $b = 4.527$. □

5. Vid en rättegång är en man (bland annat) anklagad för att ha "porrsurfat på arbetstid" vilket han förnekar. Som bevismaterial har åklagaren log-sidor där det finns dokumenterat att mannen varit inne på sidor med pornografiskt material flera gånger och även hur länge han varit där, se följande tabell:

Vecka	5	6	7	8	9
Antal timmar	7.0	6.1	4.2	6.5	6.2

där $\sum_{k=1}^{5} x_k = 30$ och $\sum_{k=1}^{5} x_k^2 = 185.54$

Mannen påstår sig dock ha hamnat på dessa sidor av misstag. Enligt tidigare undersökningar har man emellertid skattat medianen av antalet timmar folk porrsurfar till 4.22 timmar/vecka. Kan man bevisa att det förväntade antalet timmar per vecka för mannens porrsurfande är mer än 4.22 på 1% signifikansnivå? Vad blir p-värdet? (3p)

Lösning: $\begin{cases} H_0 : \mu = 4.22 \\ H_1 : \mu > 4.22 \end{cases}$

Vi får $\bar{x} = \frac{30}{5} = 6$ och $s^2 = \frac{1}{5-1}(185.54 - 5 \cdot 6^2) = 1.385$ så $U = \frac{\sqrt{n}(\bar{x}-\mu_0)}{s} = \frac{\sqrt{5}(6-4.22)}{1.177} = 3.3816$ Eftersom detta ska jämföras med t-percentilen $t_{0.01}(4) = 3.747$ och $U = 3.3816 \not> 3.747$ kan man inte förkasta H_0 på 1% signifikansnivå (även om det är ganska nära). Man får nog säga att mannens siffror indikerar ett högt värde även om man inte kan bevisa att hans förväntade antal timmar/vecka överstiger 4.22. p-värdet blir i detta fall mellan 1% och 2% (eftersom $t_{0.02}(4) = 2.9986 < 3.3816 < 3.747 = t_{0.01}(4)$).

Detta är en korrekt slutsats givet att $\sum_{k=1}^{5} x_k = 30$ och $\sum_{k=1}^{5} x_k^2 = 185.54$. Den senare uppgiften (siffran för $\sum_{k=1}^{5} x_k^2$) är dock en felräkning (från min sida) – om man istället räknar med de siffror som finns i tabellen får man $\sum_{k=1}^{5} x_k^2 = 184.54$ och med denna siffra blir $s^2 = 1.135$ och därmed $U = 3.736 \not> 3.747$ varmed H_0 inte heller i detta fall kan förkastas, dvs man kan inte visa att det förväntade antalet timmar mannen porrsurfar på jobbet överstiger 4.22 timmar/vecka på 1% signifikansnivå. I detta fall blir p-värdet bara pyttelite större än 1% (eftersom 3.736 är så nära 3.747).

Båda dessa lösningar bedöms som korrekta eftersom de beror på ett fel från min sida. $\square$

6. Antag att $X \in Poi(1+\lambda)$ där $\lambda \in Bin(2,p)$. Beräkna det värde på p som gör att $P(X = 1) = 0.3$. (4p)

Lösning:

$$\begin{aligned}
0.3 &= P(X=1) \\
&= \sum_{k=0}^{2} P(X=1, \lambda=k) \\
&= \sum_{k=0}^{2} (X=1 \mid \lambda=k) P(\lambda=k) \\
&= \sum_{k=0}^{2} \tfrac{(k+1)^1}{1!} e^{-(k+1)} \tbinom{2}{k} p^k (1-p)^{2-k} \\
&= e^{-1}(1-p)^2 + 2e^{-2}2p(1-p) + 3e^{-3}p^2 \\
&= e^{-1}((1-4e^{-1}+3e^{-2})p^2 - 2(1-2e^{-1})p + 1) \\
\Rightarrow p^2 - 2\tfrac{1-2e^{-1}}{1-4e^{-1}+3e^{-2}}p + \tfrac{e^{-1}-0.3}{(1-4e^{-1}+3e^{-2})e^{-1}} &= 0
\end{aligned}$$

$$\begin{aligned}
\Rightarrow p &= \frac{1-2e^{-1}}{1-4e^{-1}+3e^{-2}} \pm \sqrt{\left(\frac{1-2e^{-1}}{1-4e^{-1}+3e^{-2}}\right)^2 - \frac{1-0.3e}{1-4e^{-1}+3e^{-2}}} \\
&= \frac{1-2e^{-1} - \sqrt{0.3e - 1.2 + 0.9e^{-1} + e^{-2}}}{1-4e^{-1}+3e^{-2}}
\end{aligned}$$

eftersom $p > 0$. $\square$