Tentamen i Dataanalys och statistik för I den 5 jan 2016

Tentamen består av åtta uppgifter om totalt 50 poäng. Det krävs minst 20 poäng för betyg 3, minst 30 poäng för 4 och minst 40 för 5.
Examinator: Ulla Blomqvist
Hjälpmedel: Chalmersgodkänd miniräknare, Matematisk statistik (inte den ljusblå) av Ulla Dahlbom och Håkan Blomqvists formelsamling. **Boken eller formelsamlingen får inte innehålla egna anteckningar.**
Jour: Besöker tentamen c:a kl 10.00

Lycka till!

Uppgift 1: Anta att man har två händelser A och B. För dessa gäller att $P(A^C|B^C) = 0.6$, $P(B^C|A^C) = 0.4$ och $P(A \cup B) = 0.8$. Beräkna P(A) och P(B).
(6 poäng)

Uppgift 2: I en liten dagisgrupp finns 5 barn. Beräkna sannolikheten att minst 2 av barnen har samma födelsedag. Anta att året har 365 dagar och att alla födelsedagar är lika sannolika. (6 poäng)

Uppgift 3: Anta att antal bilar som kör in på en bensinstation är Poissonfördelat med en genomsnittlig ankomstfrekvens på 2 bilar på 10 minuter. En nyanställd person räknar antal bilar som anländer en viss timma.

a) Vad är sannolikheten att det kommer minst 3 bilar till bensinstationen under denna timma?
b) Anta att en bil just har kört in på bensinstationen. Hur lång tid kan man förvänta sig att det tar tills nästa bil kommer?
c) Anta att den nyanställde personen har väntat på nästa bil i 2 minuter. Vad är sannolikheten att han/hon får vänta i ytterligare 10 minuter?

(8 poäng)

Uppgift 4: Livslängden, ξ, hos en radioaktiv atom har frekvensfunktionen

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{för} \quad x \geq 0 \\ 0 & \text{för} \quad x < 0 \end{cases}$$

a) Beräkna $P(\xi < E(\xi))$ om λ är okänd.
b) Vad är väntevärdet och variansen för ξ om $\lambda = 2$?

(6 poäng)

Uppgift 5: Man har 300 reella tal som man har beräknat med 5 korrekta decimaler, vilket innebär att felet i varje tal ligger i intervallet $(-0.5 \cdot 10^{-5},\ 0.5 \cdot 10^{-5})$. Beräkna sannolikheten att felet i summan av dessa tal till sitt absolutbelopp är mindre än $0.5 \cdot 10^{-4}$. Felen i de olika talen kan antas vara oberoende och rektangelfördelade i det angivna intervallet. (6 poäng)

Uppgift 6: Längden av en bräda mäts en gång med en tumstock A, en gång med en tumstock B och en gång med en tumstock C. Kalla de uppmätta längderna för x_A, x_B och x_C. Motsvarande stokastiska variabler ξ_A, ξ_B och ξ_C har väntevärdet lika med plankans verkliga längd, men ξ_C har bara en tredjedel så stor standardavvikelse som ξ_A och ξ_B. Vilken av nedanstående förslag är bäst om man vill skatta den sanna längden? Svaret måste motiveras för att du skall få poäng.

a) $\dfrac{x_A + x_B + x_C}{3}$ b) x_C

c) $\dfrac{x_A + x_B + 3x_C}{5}$ d) $\dfrac{x_A + 7x_C}{8}$

(6 poäng)

Uppgift 7: En glassförsäljare har noterat följande försäljning under en 4-dagars period i juli:

Dag	försäljning i tusentals kr	temp °C	väderlek	kodad väderlek
1	22.4	18	solsken	1
2	20.6	21	regn	0
3	25.4	28	solsken	1
4	21.8	20	solsken	1

I tabellen är väderleken kodad i variabeln med solsken = 1 och regn = 0.

Glassförsäljaren vill kunna uppskatta morgondagens försäljning genom att titta på väderleksrapporten kvällen före där såväl temperatur som väderleken anges.

a) Hjälp honom genom att skatta koefficienterna a, b_1 och b_2 i en multipel regressionsmodell $\hat{y} = a + b_1x_1 + b_2x_2$ där x_1 är temperaturen och x_2 är den kodade väderleken.
b) Anta att väderprognosen en kväll sa att vädret nästa dag skulle vara soligt med en temperatur på 25 °C. Använd den multipla regressionsmodellen för att uppskatta försäljningen den dag som prognosen gäller.

(6 poäng)

Uppgift 8: En viss växt kan ha vita, skära eller röda blommor. Betrakta avkomman till plantorna med skära blommor. Enligt en teori bör 25% av dotterplantorna ha vita blommor, 50% skära blommor och 25% röda blommor. Man observerade vid ett tillfälle att av 140 dotterplantor hade 33 vita blommor, 81 skära blommor och 26 röda blommor. Testa om ovanstående teori kan vara falsk. Använd 5%:s signifikansnivå.

(6 poäng)

Lösningar till Dataanalys och statistik 20160105

Uppgift 1: $P(A^C|B^C) = 0.6$, $P(B^C|A^C) = 0.4$ och $P(A \cup B) = 0.8$.

Använd de Morgans sats $P(A^C \cap BC) = 1 - P(A \cup B) = 1 - 0.8 = 0.2$

$$P(B^C \mid A^C). = \frac{P(A^C \cap B^C)}{P(A^C)} = 0.4 \quad \Rightarrow \quad P(A^C) = \frac{0.2}{0.4} = \frac{1}{2}$$

$$P(A^C \mid B^C). = \frac{P(A^C \cap B^C)}{P(B^C)} = 0.6 \quad \Rightarrow \quad P(B^C) = \frac{0.2}{0.6} = \frac{1}{3}$$

dvs

$$P(A^C) = \frac{1}{2} \quad \Rightarrow \quad P(A) = \frac{1}{2}$$

$$P(B^C) = \frac{1}{3} \quad \Rightarrow \quad P(B) = \frac{2}{3}$$

Uppgift 2: A = minst 2 barn har samma födelsedag. $\Rightarrow$ A^C = inget av barnen har samma födelsedag

$$P(A) = 1 - P(A^C) = 1 - \frac{365}{365} \cdot \frac{364}{365} \cdot \frac{363}{365} \cdot \frac{362}{365} \cdot \frac{363}{365} \approx 1 - 0.9729 = 0.0271$$

Uppgift 3: ξ = antal bilar $\xi = Po(\lambda = 2 \text{ bilar}/10 \text{ min})$

a) Räkna om λ till antal bilar/ 60 minuter. $\lambda = 12$ bil/ 60 min

$$P(\xi \geq 3) = 1 - P(\xi \leq 2) = 1 - e^{-12} \cdot \left(\frac{12^0}{0!} + \frac{12^1}{1!} + \frac{12^2}{2!}\right) \approx 1 - 0.0005 = 0.9995$$

b) η = tiden mellan två bilar $\eta = Exp(\lambda = 2 \text{ bilar}/10 \text{ min})$

$$E(\eta) = \frac{1}{\lambda} = \frac{10 \text{ minuter}}{2 \text{ bilar}} = \frac{5 \text{ min}}{\text{bil}} = 5 \text{ min}$$

c) Räkna om λ till minuter. $\lambda = 0.2$ bilar/min

$$P(\xi>12 \mid \xi>2) = \frac{P(\xi>12 \cap \xi>2)}{P(\xi>2)} = \frac{P(\xi>12)}{P(\xi>2)} = \frac{1-P(\xi<12)}{1-P(\xi<2)} = \frac{e^{-12 \cdot 0.2}}{e^{-2 \cdot 0.2}} = e^{-10 \cdot 0.2} \approx 0.1353$$

Uppgift 4: ξ = livslängden ξ = exponentialfördelad $E(\xi) = 1/\lambda$

$$F(x) = \begin{cases} 0 & \text{för } x<0 \\ 1-e^{-\lambda x} & \text{för } x \geq 0 \end{cases}$$

a) $P(\xi < E(\xi)) = P(\xi < 1/\lambda) = F(\frac{1}{\lambda}) = 1 - e^{-\lambda \cdot \frac{1}{\lambda}} = 1 - e^{-1} \approx 0.633$

b) $E(\xi) = 1/\lambda = 1/2 = 0.5$ $Var(\xi) = 1/\lambda^2 = 1/4 = 0.25$

Uppgift 5: ξ_i = felets storlek ξ_i är $R[-0.5 \cdot 10^{-5},\ 0.5 \cdot 10^{-5}]$

$f(x) = \frac{1}{b-a} = \frac{1}{10^{-5}}$ i ovanstående interval

$E(\xi) = 0$ på grund av symmetrin

$$Var(\xi) = \int_a^b x^2 f(x)\,dx - [E(\xi)]^2 = \int_{-0.5\cdot 10^{-5}}^{0.5\cdot 10^{-5}} x^2 \frac{1}{10^{-5}}\,dx = \frac{1}{10^{-5}}\left[\frac{x^3}{3}\right]_{-0.5\cdot 10^{-5}}^{0.5\cdot 10^{-5}} = \frac{0.25\cdot 10^{-10}}{3}$$

Eftersom antal fel är stort (n= 300) så kommer centrala gränsvärdessatsen att användas

η = summan av 300 fel
$\eta = \xi_1 + \xi_2 + \ldots + \xi_{300}$

$E(\eta) = E(\xi_1) + E(\xi_2) + \ldots + E(\xi_{300}) = 300 \cdot 0 = 0$

$Var(\eta) = Var(\xi_1) + \ldots + Var(\xi_{300}) = 300 \cdot \frac{0.25\cdot 10^{-10}}{3} = 0.25\cdot 10^{-8}$

$P(|\eta| \leq 0.5\cdot 10^{-4}) = P\left(-\frac{0.5\cdot 10^{-4} - 0}{\sqrt{0.25\cdot 10^{-8}}} < Z < \frac{0.5\cdot 10^{-4} - 0}{\sqrt{0.25\cdot 10^{-8}}}\right) = P(-1 < Z < 1) =$

$P(Z< 1) - P(Z< -1) = P(Z< 1) - (1 - P(Z< 1) = 0.8413 - (1 - 0.8413) = 0.6826$

Uppgift 6: $S(\xi_A) = \sigma$ $S(\xi_B) = \sigma$ och $S(\xi_C) = \frac{\sigma}{3}$ $\Rightarrow$

$\Rightarrow$ $Var(\xi_A) = \sigma^2$ och $Var(\xi_B) = \sigma^2$ $Var(\xi_C) = \frac{\sigma^2}{9}$

Vi söker den sammanvägningen där summan av vikterna = 1 och variansen är minst

Fortsättning uppgift 6 på nästa sida

Fortsättning uppgift 6

a) $\frac{x_A + x_B + x_C}{3}$ Summan av vikterna: $\frac{1+1+1}{3} = 1$ väntevärdesriktig

Variansen: $Var(\frac{\xi_A + \xi_B + \xi_C}{3}) = \frac{1}{9}[Var(\xi_A) + Var(\xi_B) + Var(\xi_C)] =$

$$\frac{1}{9}(\sigma^2 + \sigma^2 + \frac{\sigma^2}{9}) = \frac{19}{81}\sigma^2 \approx 0.235\,\sigma^2$$

b) x_C Summan av vikterna: 1 väntevärdesriktig

Variansen: $Var(\xi_C) = \frac{\sigma^2}{9} \approx 0.111\sigma^2$

c) $\frac{x_A + x_B + 3x_C}{5}$ Summan av vikterna: $\frac{1+1+3}{5} = 1$ väntevärdesriktig

Variansen: $Var(\frac{\xi_A + \xi_B + 3\xi_C}{5}) = \frac{1}{25}[Var(\xi_A) + Var(\xi_B) + 9\,Var(\xi_C)] =$

$$= \frac{1}{25}(\sigma^2 + \sigma^2 + \frac{9}{9}\sigma^2) = \frac{3}{25}\sigma^2 \approx 0.12\,\sigma^2$$

d) $\frac{x_A + 7x_C}{8}$ Summan av vikterna: $\frac{1+7}{8} = 1$ väntevärdesriktig

Variansen: $Var(\frac{\xi_A + 7\xi_C}{8}) = \frac{1}{64}[Var(\xi_A) + 49\,Var(\xi_C)] =$

$$= \frac{1}{64}(\sigma^2 + \frac{49}{9}\sigma^2) = \frac{58}{576}\sigma^2 \approx 0.100\,\sigma^2$$

Välj alternativ d) dvs $\frac{x_A + 7x_C}{8}$

Uppgift 7: y = försäljning i hundratals kronor
x_1 = temperatur, x_2 = kodad vädervariabel

n = 4 $\sum_{i=1}^{4} x_{1i} = 87$ $\sum_{i=1}^{4} x_{2i} = 3$ $\sum_{i=1}^{4} y_i = 90.2$ $\sum_{i=1}^{4} x_{1i} \cdot x_{2i} = 66$

$\sum_{i=1}^{4} x_{1i}^2 = 1949$ $\sum_{i=1}^{4} x_{2i}^2 = 3$ $\sum_{i=1}^{4} x_{1i} \cdot y_i = 1983$ $\sum_{i=1}^{4} x_{2i} \cdot y_i = 69.6$

a) normalekvationerna ger följande ekvationssystem:

Fortsättning uppgift 7 på nästa sida

Fortsättning uppgift 7

$$\begin{cases} 4a + 87b_1 + 3b_2 = 90.2 \\ 87a + 1949b_1 + 66b_2 = 1983 \\ 3a + 66b_1 + 3b_2 = 69.6 \end{cases} \quad \text{E[1] – E[3]} \longrightarrow \quad \begin{cases} a + 21b_1 = 20.6 \\ 87a + 1949b_1 + 66b_2 = 1983 \\ 3a + 66b_1 + 3b_2 = 69.6 \end{cases}$$

$$\text{E[2] – 22E[3]} \longrightarrow \begin{cases} a + 21b_1 = 20.6 \\ 21a + 497b_1 = 451.8 \\ 3a + 66b_1 + 3b_2 = 69.6 \end{cases} \quad \text{E[2] – 21E[1]} \longrightarrow \begin{cases} a + 21b_1 = 20.6 \\ 56b_1 = 19.2 \\ 3a + 66b_1 + 3b_2 = 69.6 \end{cases}$$

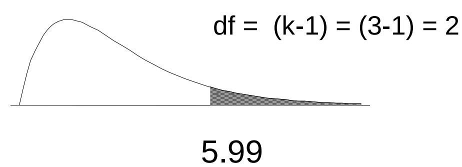
$$\Rightarrow \quad \begin{cases} b_1 = \dfrac{19.2}{56} \approx 0.343 \\ a = 20.6 - 21 \cdot \dfrac{19.2}{56} = 13.4 \\ b_2 = \dfrac{69.6 - 3 \cdot 13.4 - 66 \cdot \dfrac{19.2}{56}}{3} \approx 2.257 \end{cases}$$

Modellen blir $\hat{y} = 13.4 + 0.34x_1 + 2.26x_2$

b) $x_1 = 25^o$ $x_2 = 1$ (solsken) $\Rightarrow$ $\hat{y} = 13.4 + 0.34 \cdot 25 + 2.26 \cdot 1 = 24.16$

Uppgift 8:

Steg 1: H_0: fördelningen 0.25 0.50 0.25
H_1: <u>inte</u> fördelningen 0.25 0.50 0.25



Steg 2: $\alpha = 0.05$

Steg 3: Välj testvariabeln $\chi^2 = \sum_{i=1}^{6} \frac{(O_i - E_i)^2}{E_i}$

Steg 4:

Vi ställer nu upp de observerade och de förväntade värdena i en tabell.

Fortsättning uppgift 8 på nästa sida

Fortsättning uppgift 8

	vita	skära	röda	Totalt
Obs antal, O_i	33	81	26	140
p_i	0.25	0.50	0.25	1
Förväntat antal, E_i	35	70	35	140

$$\chi^2 = \frac{(33-35)^2}{35} + \frac{(81-70)^2}{70} + \frac{(26-35)^2}{35} = 4.16 < 5.99$$

Värdet hamnar i acceptansområdet.

Steg 5: H_0 kan inte förkastas. Undersökningen motsäger inte hypotesen att färgen på dotterplantorna har den fördelning som teorin säger.