Lösningar till Dataanalys och statistik
TMS136
omtenta 2016-08-23

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

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

Beteckna $P(A^C \cap B^C) = p$

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

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

	A	**A^C**	
B	0.20	$0.8 - \frac{p}{0.6}$	$\mathbf{1 - \frac{p}{0.6}}$
B^C	$0.8 - \frac{p}{0.4}$	p	$\mathbf{\frac{p}{0.6}}$
	$\mathbf{1 - \frac{p}{0.4}}$	$\mathbf{\frac{p}{0.4}}$	**1**

$$P(A^C \cap B) + P(A^C \cap B^C) = P(A^C) \quad \Rightarrow$$

$$0.8 - \frac{p}{0.6} + p = \frac{p}{0.4} \quad \Rightarrow \quad 0.8 = p\left(\frac{10}{4} + \frac{10}{6} - 1\right) \quad \Rightarrow \quad p = \frac{48}{190} \qquad \text{dvs}$$

$$P(A) = 1 - \frac{p}{0.4} = 1 - \frac{10}{4}p = 1 - \frac{10}{4} \cdot \frac{48}{190} = 1 - \frac{12}{19} = \frac{7}{19}$$

$$P(B) = 1 - \frac{p}{0.6} = 1 - \frac{10}{6}p = 1 - \frac{10}{6} \cdot \frac{48}{190} = 1 - \frac{8}{19} = \frac{11}{19}$$

Uppgift 2: ξ = antal gånger Anna får p-böter $\quad \xi = \text{Bin}(n, p) = \text{Bin}(300, 0.04)$
η = vinsten $\quad \eta = 300 \cdot 40 - 800 \cdot \xi$

a) $E(\eta) = 300 \cdot 40 - 800 \cdot E(\xi) = 12000 - 800 \cdot 300 \cdot 0.04 = 2400$

b) $np(1 - p) = 300 \cdot 0.04 \cdot 0.96 = 11.52 > 10 \Rightarrow$ använd normalapproximation

$\text{Var}(\eta) = \text{Var}(300 \cdot 40 - 800 \cdot \xi = 800^2\, \text{Var}(\xi) = 800^2 \cdot 11.52$

$P(\eta>0) = 1 – P(\eta<0) = 1 – P(Z< \frac{0-2400}{\sqrt{800^2 \cdot 11.52}}) = 1 – P(Z<–0.88) = P(Z<0.88) \approx 0.8106$

Uppgift 3: ξ = antal kunder $\xi = Po(\lambda = 30$ kunder/1 tim)

a) Räkna om λ till antal kunder/ 5minuter. $\lambda = 2.5$ kunder/ 5 min

$$P(\xi \geq 2) = 1 – P(\xi \leq 1) = 1 – e^{-2.5} \cdot \left(\frac{2.5^0}{0!} + \frac{2.5^1}{1!}\right) \approx 1 – 0.2872 = 0.7127$$

b) Räkna om λ till minuter. $\lambda = 0.5$ kunder/min

η = tiden i minuter mellan två kunder $\eta = Exp(\lambda = 0.5$ kunder/min)

$$P(\eta >1) = 1 – P(\eta <1) = 1 – e^{-0.5} \approx 0.3935$$

Uppgift 4: ξ = vikten av en mjölsäck ξ = Rektangelfördelad [95 – 105]

$$F(x) = \frac{x-95}{105-95} \quad \Rightarrow \quad P(\xi < 95.1) = F(95.1) = \frac{95.1-95}{105-95} = 0.01$$

η = antal säckar vars vikt är högst 95.1 kg $\eta = Bin(n, p) = Bin(10, 0.01)$

$$P(\eta \leq 2) = \binom{10}{0} 0.01^0 \cdot 0.99^{10} + \binom{10}{1} 0.01^1 \cdot 0.99^9 + \binom{10}{2} 0.01^2 \cdot 0.99^8 \approx 0.9999$$

Uppgift 5:

a) $E(\xi) = \sum_{\Omega} x \cdot P(\xi = x) = 1\cdot 0.20 + 2\cdot 0.25 + 3\cdot 0.30 + 4\cdot 0.10 + 5\cdot 0.15 = 2.75$

b) ξ_i = vikten på last ”i” $\xi_i = N(\mu = 5, \sigma = 2)$

ξ_1	ξ_2	ξ_3	ξ_4	ξ_5
1	2	3	4	5

η = genomsnittsvärdet av vikterna $\eta = \frac{1}{\sum_1^5 \xi_i}(1\cdot\xi_1 + 2\cdot\xi_2 + 3\cdot\xi_3 + 4\cdot\xi_4 + 5\cdot\xi_5)$

$P(2.4 < \eta < 3.4) = P(\eta < 3.4) – P(\eta < 2.4)$

$$P(\eta < 3.4) = P\left(\frac{1}{\sum_1^5 \xi_i}(1\cdot\xi_1 + 2\cdot\xi_2 + 3\cdot\xi_3 + 4\cdot\xi_4 + 5\cdot\xi_5) < 3.4\right) =$$

$$P(1\cdot\xi_1 + 2\cdot\xi_2 + 3\cdot\xi_3 + 4\cdot\xi_4 + 5\cdot\xi_5 < 3.4\sum_{i=1}^{5}\xi_i) =$$

Fortsättning uppgift 5 på nästa sida

<u>Fortsättning uppgift 5</u>

$$P(-2.4\cdot\xi_1 - 1.4\cdot\xi_2 - 0.4\cdot\xi_3 + 0.6\cdot\xi_4 + 1.6\cdot\xi_5 < 0)$$

$$\zeta_1 = -2.4\cdot\xi_1 - 1.4\cdot\xi_2 - 0.4\cdot\xi_3 + 0.6\cdot\xi_4 + 1.6\cdot\xi_5$$

$$E(\zeta_1) = -2.4\cdot 5 - 1.4\cdot 5 - 0.4\cdot 5 + 0.6\cdot 5 + 1.6\cdot 5 = -10$$

$$Var(\zeta_1) = (-2.4)^2\cdot 2^2 + (-1.4)^2\cdot 2^2 + (-0.4)^2\cdot 2^2 + 0.6^2\cdot 2^2 + 1.6^2\cdot 2^2 = 43.2$$

Alltså, $\zeta_1 = N(-10, \sqrt{43.2})$

På motsvarande sätt beräknar vi

$$P(\eta < 2.4) = P(\frac{1}{\sum_1^5 \xi_i}(1\cdot\xi_1 + 2\cdot\xi_2 + 3\cdot\xi_3 + 4\cdot\xi_4 + 5\cdot\xi_5) < 2.4) =$$

$$P(1\cdot\xi_1 + 2\cdot\xi_2 + 3\cdot\xi_3 + 4\cdot\xi_4 + 5\cdot\xi_5 < 2.4\sum_{i=1}^{5}\xi_i) =$$

$$P(-1.4\cdot\xi_1 - 0.4\cdot\xi_2 + 0.6\cdot\xi_3 + 1.6\cdot\xi_4 + 2.6\cdot\xi_5 < 0)$$

$$\zeta_2 = -1.4\cdot\xi_1 - 0.4\cdot\xi_2 + 0.6\cdot\xi_3 + 1.6\cdot\xi_4 + 2.6\cdot\xi_5$$

$$\zeta_2 = N(15, \sqrt{47.2})$$

$$P(2.4 < \eta < 3.4) = P(\eta < 3.4) - P(\eta < 2.4) =$$
$$P(-2.4\cdot\xi_1 - 1.4\cdot\xi_2 - 0.4\cdot\xi_3 + 0.6\cdot\xi_4 + 1.6\cdot\xi_5 < 0) -$$
$$P(-1.4\cdot\xi_1 - 0.4\cdot\xi_2 + 0.6\cdot\xi_3 + 1.6\cdot\xi_4 + 2.6\cdot\xi_5 < 0) =$$

$$P(Z < \frac{0-(-10)}{\sqrt{43.2}}) - P(Z < \frac{0-15}{\sqrt{47.2}}) = P(Z < 1.52) - P(Z < -2.18) =$$
$$P(Z < 1.52) - (1 - P(Z < 2.18)) \approx 0.9357 - (1 - 0.9854) = 0.9211$$

Uppgift 6: ξ är $Bin(n, p) = Bin(n, \frac{1+k}{2})$ $\quad E(\xi) = np = \frac{n(1+k)}{2}$

$$E\left(\frac{2\xi - n}{n}\right) = \frac{2E(\xi) - n}{n} = \frac{2\frac{n(1+k)}{2} - n}{n} = \frac{n(1+k) - n}{n} = \frac{n(1+k-1)}{n} = k$$

dvs, $\frac{2x - n}{n}$ är en väntevärdesriktig skattning av k.

Uppgift 7:

a) Frågan är mångtydig. Om vi antar vi har observationer $x_1, \ldots, x_{10}$ som är normalfördelade med väntevärde μ_1 och varians σ_1^2 så är två möjliga tolkningar att man behöver hitta ett 95%-igt konfidensintervall antingen för μ_1 eller för σ_1^2.
I första fallet:

$n = 10 \qquad df = 9 \qquad \sum_{i=1}^{10} x_i = 92 \qquad \sum_{i=1}^{10} (x_i - \bar{x})^2 = 16 \quad \Rightarrow \quad s^2 = \frac{16}{9}$

$$\bar{x} \pm t \cdot \frac{s}{\sqrt{n}} \quad \Rightarrow \quad \frac{92}{10} \pm 2.26 \cdot \sqrt{\frac{16}{9 \cdot 10}} \quad \Rightarrow \quad 9.2 \pm 0.95 \qquad [\,8.25 - 10.15\,]$$

I andra fallet är ett möjligt 95%-igt konfidensintervall

$[\,0, 9s^2/\chi_{9,0.05}] = [0, 16/3.325] = [0, 4.81]$.

Ett annat är

$[\,9s^2/\chi_{9,0.975}, 9s^2/\chi_{9,0.025}] = [16/19.02, 16/2.70] = [0.84, 5.93]$.

b) Frågan är mångtydig. Om vi antar vi har nya observationer $y_1, \ldots y_{12}$ som är normalfördelade med väntevärde μ_2 och varians σ_2^2, så är ett tillvägagångssätt att beräkna ett nytt konfidensintervall μ_2 och jämföra detta med det för μ_1:

$n = 12 \qquad df = 11 \qquad \sum_{i=1}^{10} x_i = 130 \qquad \sum_{i=1}^{10} (x_i - \bar{x})^2 = 24 \quad \Rightarrow \quad s^2 = \frac{24}{11}$

$$\bar{x} \pm t \cdot \frac{s}{\sqrt{n}} \quad \Rightarrow \quad \frac{130}{12} \pm 2.20 \cdot \sqrt{\frac{24}{11 \cdot 12}} \quad \Rightarrow \quad 10.83 \pm 0.98 \quad [\,9.85 - 11.81\,]$$

Eftersom intervallen för μ_1 och μ_2 överlappar vill vissa konkludera att vi inte kan dra slutsatsen att ”medelvärdet har ändrats”. Ett annat tillvägagångssätt är att beräkna ett 95%-igt konfidensintervall för $\mu_2 - \mu_1$: Vi får

$df = (16/(9\cdot 10) + 24/(11\cdot 12))^2/((16/(9\cdot 10))^2/9+(24/(11\cdot 12))^2/11) = 19.84$

$$\frac{130}{12} - \frac{92}{10} \pm 2.086 \cdot \sqrt{\frac{16}{9 \cdot 10} + \frac{24}{11 \cdot 12}} = 1.63 \pm 1.25$$

Eftersom intervallen för $\mu_2 - \mu_1$ inte innehåller 0 vill vissa konkludera att man kan dra slutsatsen att ”medelvärdet har ändrats”. Slutligen: Eftersom medelvärden av de två grupperna av observationer är olika, är det tydligt att medelvärdet har ändrats!

Uppgift 8: y = omsättning i miljontals kronor x = årtal t = årtal – 2013

a) Anpassa en exponentialfunktion $\hat{y} = a \cdot b^t \Rightarrow \mathbf{ln}\,\hat{y} = \mathbf{ln}a + t\,\mathbf{ln}b$

$$\sum_{i=1}^{5} t_i = 0 \quad \sum_{i=1}^{5} t_i^2 = 10 \quad \sum_{i=1}^{5} \mathbf{ln} y_i = 4.92829 \quad \sum_{i=1}^{5} t_i \cdot \mathbf{ln} y_i = 3.067584$$

$$\mathbf{ln}b = \frac{\sum_{i=1}^{5} t_i \cdot \mathbf{ln} y_i - \frac{\left(\sum_{i=1}^{5} t_i\right)\left(\sum_{i=1}^{5} \mathbf{ln}\, y_i\right)}{5}}{\sum_{i=1}^{5} t_i^2 - \frac{\left(\sum_{i=1}^{5} t_i\right)^2}{5}} = \frac{3.067584 - 0}{10 - 0} \Rightarrow b \approx 1.359$$

$$\mathbf{ln}a = \frac{\sum_{i=1}^{5} \mathbf{ln}\, y_i}{5} - \frac{\sum_{i=1}^{5} t_i}{5} \cdot \mathbf{ln}b = \frac{4.92829}{5} - 0 \approx 0.9856578 \Rightarrow a \approx 2.680$$

Funktionen blir $\hat{y} = 2.68 \cdot 1.359^t = 2.68 \cdot 1.359^{(\text{årtal - 2013})}$

b) Den årliga ökningstakten = 0.359

c) Prognos för 2016: $\hat{y} = 2.68 \cdot 1.359^{(2016 - 2013)} = 6.72$ miljoner kr