

LÖSNINGAR till tentamen: Statistik och sannolikhetslära (LMA120)

Tid och plats: 8.30-12.30 den 30 maj 2015

Hjälpmedel: Typgodkänd miniräknare, formelblad

Betygsgränser: 3: 12 poäng, 4: 18 poäng, 5: 24 poäng. Maximalt antal poäng: 30

Examinator: Petter Mostad

Telefonvakt: Anders Hildeman, telefonnummer: 0708-193630

1. Ett företag beräknar att tiden det tar i minuter att tillverka en viss produkt är normalfördelad med väntevärde 8 och varians 2. Din chef bestämmer sig för att kontrollera om väntevärdet stämmer, och mäter under en dag tillverkningstiden för fyra produkter. Din chef mäter upp tiderna 5.6, 7.2, 8.1 och 7.5 och konstaterar att företaget verkar ha blivit snabbare på att tillverka produkten.
 - (a) Om chefs mätningar är oberoende och tillverkningstiden är normalfördelad med väntevärde 8 så borde antalet tider som är mindre än 8 följa en viss fördelning. Vilket fördelning är det, och vad är dess parametrar?
 - (b) Hur sannolikt är det att få minst 3 tider som är mindre än 8 om antalet tider som är mindre än 8 följer fördelningen i (a)-uppgiften?
 - (c) Om du antar att tiden det tog att tillverka en produkt då chefen gjorde sina mätningar är normalfördelad med variansen 2, kan du beräkna ett konfidensintervall för fördelningens väntevärde? Om du svarar ja på uppgiften, beräkna ett konfidensintervall med konfidensgrad 0.98, och om du svarar nej, förklara varför det inte går.
 - (d) Om stickprovsmedelvärdet inte förändras då din chef gör fler mätningar, hur många mätningar behöver han göra för att kunna dra den slutsats han drog med hjälp av ett konfidensintervall med konfidensgrad 0.98? (2+2+3+3 poäng)

Lösning

- (a) Antalet tider X som är mindre än åtta borde vara binomialfördelat med parametrarna $n = 4$ och $p = 0.5$.
- (b) Tre av tiderna chefen mätte upp var mindre än åtta, och

$$P(X \geq 3) = P(X = 3) + P(X = 4) = \binom{4}{3} \cdot 0.5^3 \cdot 0.5^1 + \binom{4}{4} \cdot 0.5^4 \cdot 0.5^0 = 5 \cdot 0.5^4$$

- (c) Ja! Eftersom att vi antar att tiden är normalfördelad så kan vi beräkna ett konfidensintervall även om stickprovet inte är särskilt stort. Om vi använder konfidensnivå 0.98 så blir konfidensintervallet

$$\left[\bar{x} - z \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z \cdot \frac{\sigma}{\sqrt{n}} \right] = \left[7.1 - 2.33 \cdot \frac{\sqrt{2}}{\sqrt{4}}, 7.1 + 2.33 \cdot \frac{\sqrt{2}}{\sqrt{4}} \right] = [5.4524, 8.7476]$$

- (d) Vi behöver räkna ut hur stort n måste vara för att talet 8 inte ska ligga i intervallet

$$\left[7.1 - 2.33 \cdot \frac{\sqrt{2}}{\sqrt{n}}, 7.1 + 2.33 \cdot \frac{\sqrt{2}}{\sqrt{n}} \right]$$

vilket är samma sak som att titta vilket som är det minsta heltal n sådant att

$$7.1 + 2.33 \cdot \frac{\sqrt{2}}{\sqrt{n}} < 8.$$

Om vi löser ut n så får vi att

$$\frac{\sqrt{2} \cdot 2.33}{8 - 7.1} < \sqrt{n},$$

dvs.

$$\left(\frac{\sqrt{2} \cdot 2.33}{8 - 7.1} \right)^2 = 13.4047 < n$$

så chefen behöver göra minst 14 mätningar.

2. Antalet fartyg som ankommer till en viss hamn under loppet av en dag har visat sig vara Poissonfördelat med parametern $\lambda = 2$. Hamnen kan maximalt betjäna 3 tankfartyg dag. De tre först ankomna blir betjänade, och eventuella övriga blir omdirigerade till en annan hamn.
- (a) Bestäm sannolikheten för att det en slumpvist vald dag kommer att omdirigeras fartyg till en annan hamn.
 - (b) Hur stor kapacitet bör hamnen byggas ut till för att med minst 99%:s sannolikhet kunna betjäna samtliga fartyg som ankommer en slumpvist vald dag?
 - (c) Antag att hamnen utökar sin kapacitet med en plats till en kostnad av 5 miljoner kronor. Om varje betjänat fartyg ger hamnen en inkomst av 50 000 kronor, hur lång tid bör hamnen räkna med att det tar innan de börjar gå plus på den extra platsen?
 - (d) Kommunen räknar med att hamnen bara kommer att vara i bruk i tre år till. Om hamnen utökar sin kapacitet med en plats; hur stor är risken att de inte kommer ha gått plus på den innan hamnen läggs ner? (2+2+3+3 poäng)

Lösning

- (a) Låt X vara antalet fartyg som ankommer till hamnen under en slumpvist vald dag, så att $X \sim \text{Poisson}(2)$. Då är

$$\begin{aligned} P(X > 3) &= 1 - P(X \leq 3) \\ &= 1 - \left(\frac{e^{-2} \cdot 2^0}{0!} + \frac{e^{-2} \cdot 2^1}{1!} + \frac{e^{-2} \cdot 2^2}{2!} + \frac{e^{-2} \cdot 2^3}{3!} \right) \\ &= 1 - \left(\frac{e^{-2}}{1} + \frac{e^{-2} \cdot 2}{1} + \frac{e^{-2} \cdot 4}{2} + \frac{e^{-2} \cdot 8}{6} \right) \\ &= 0.1429 \end{aligned}$$

- (b) Frågan är hur stort k behöver vara för att $P(X > k)$ ska vara mindre än 0.01.

$$P(X > 4) = 0.0527$$

$$P(X > 5) = 0.0166$$

$$P(X > 6) = 0.0045$$

dvs. hamnen behöver kunna ta emot minst 6 fartyg per dag.

- (c) Sannolikheten att den extra hamnplatsen används på en slumpvist vald dag är $P(X > 3) = 0.1429$ enligt (a). Den förväntade inkomsten från den nya hamnplatsen per dag blir därför

$$50000 \cdot P(X > 3) = 50000 \cdot 0.1429 = 7145$$

Eftersom att $5000000/7145 = 699.7901$ så kommer det i snitt ta ca 700 dagar innan hamnen går med vinst på att utöka sin kapacitet med ett fartyg.

- (d) Eftersom att hamnen tjänar 50 000 per betjänat fartyg så behöver de betjäna totalt $5000000/50000 = 100$ extra fartyg under den närmsta tre åren för att tjäna på att öka sin kapacitet. Antalet dagar Y under de närmsta tre åren som de kommer att behöva den extra kapaciteten är en binomialfördelad slumpvariabel med parametrarna $n = 3 \cdot 365$ och $p = P(X > 3) = 0.1429$, dvs. $Y \sim \text{bin}(1095, 0.1429)$. Vi behöver räkna ut $P(Y < 100)$. Om vi använder att binomialfördelningen är lik normalfördelningen då n är stort så får vi att

$$\begin{aligned} P(Y < 100) &\approx \Phi\left(\frac{100 - np}{\sqrt{np(1-p)}}\right) \\ &= \Phi\left(\frac{100 - 156.555}{\sqrt{1095 \cdot 0.1429(1 - 0.1429)}}\right) \\ &= \Phi(-4.8835) \\ &< 0.00001 \end{aligned}$$

enligt tabellen (mer exakta beräkningar ger $P(Y < 100) \approx 5.2110 \cdot 10^{-7}$), dvs. hamnen kommer nästan säkert att tjäna på att utvidga hamnen med en extra plats.

3. Ett byggföretag misstänker att hur många olyckor som sker beror på när på dygnet ett visst arbete utförs. Företaget har därför samlat in data från 100 olyckor och sammanställt den i följande tabell. Testa hypotesen att risken för olyckor inte varierar över en arbetsdag.

Tid	8-9	9-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17
Antal olyckor	5	7	8	18	17	12	3	12	18

(5 poäng)

Lösning Vi ställer först upp en tabell:

Tid (x_i)	Antal olyckor (O_i)	Förväntat antal olyckor (E_i)	$O_i - E_i$	$(O_i - E_i)^2$	$(O_i - E_i)^2/E_i$
8-9	5	11.11	-6.11	37.35	3.36
9-10	7	11.11	-4.11	16.90	1.52
10-11	8	11.11	-3.11	9.68	0.87
11-12	18	11.11	6.89	47.46	4.27
12-13	17	11.11	5.89	34.68	3.12
13-14	12	11.11	0.89	0.79	0.07
14-15	3	11.11	-8.11	65.79	5.92
15-16	12	11.11	0.89	0.79	0.07
16-17	18	11.11	6.89	47.46	4.27

Med hjälp av tabellen ser vi att $\sum_i \frac{(O_i - E_i)^2}{E_i} = 23.4800$. Eftersom att vi vet att $\sum_i \frac{(O_i - E_i)^2}{E_i}$ är χ^2 -fördelad med $9 - 1 = 8$ frihetsgrader om hypotesen att antalet olyckor inte varierar över en arbetsdag stämmer så ser vi i tabell att sannolikheten att få data som avviker så här mycket från en likformig fördelning är ungefär 0.5%, dvs. vi bör förmodligen förkasta vår hypotes.

4. En stokastisk variabel X har frekvensfunktionen

$$f(x) = \begin{cases} Cx^2 + 100Ce^{-x}, & 0 \leq x \leq 10 \\ 0, & x < 0 \text{ eller } x > 10 \end{cases}$$

- (a) Vad är C ?
- (b) Vad är $\mathbb{P}(2 < X \leq 4)$?
- (c) Vad är väntevärdet av X ?
- (d) Vad är standardavvikelsen av X ? (1+1+1+2 poäng)

Lösning

- (a) En frekvensfunktion måste integreras till 1.

$$\int_0^{10} f(x)dx = 1 = C \left[\frac{x^3}{3} - 100e^{-x} \right]_0^{10} = 100C \left(\frac{10}{3} + 1 - e^{-10} \right)$$

$$\Leftrightarrow C = \frac{1}{100} \frac{3}{10 + 3 - 3e^{-10}} \approx 0.0023$$

- (b)

$$\mathbb{P}(2 < X \leq 4) = \int_2^4 f(x)dx = \frac{3}{1300 - 300e^{-10}} \left[\frac{4^3 - 2^3}{3} - 100(e^{-4} - e^{-2}) \right] \approx 0.07$$

- (c)

$$\mathbb{E}[X] = \int_0^{10} xf(x)dx = C \left(\left[\frac{x^4}{4} \right]_0^{10} + 100 \int_0^{10} xe^{-x}dx \right) = \{ \text{partiell integration} \} =$$

$$= C \frac{10^4}{4} + 100C (-10e^{-10} - e^{-10} + 1) \approx 6.00$$

- (d)

$$\mathbb{E}[X^2] = \int_0^{10} x^2 f(x)dx = C \left[\frac{x^5}{5} \right]_0^{10} + 100C \int_0^{10} x^2 e^{-x}dx =$$

$$= C \frac{10^5}{5} - 100C [x^2 e^{-x} + 2x e^{-x} + 2e^{-x}]_0^{10} =$$

$$= C \left(\frac{10^5}{5} - 100e^{-10}(10^2 + 20 + 2) + 200 \right) = 100C(202 - 122e^{-10}) \approx 46.62$$

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \approx 46.62 - 6^2 \approx 10.62$$

$$\sqrt{\text{var}(X)} = \sqrt{10.93} \approx 3.26$$