Lösningar till tentamen i Matematisk Statistik, 5p

LGR00

13 mars, 2003 kl. 13.30 – 17.30

Kursansvarig: Eric Järpe
Maxpoäng: 30
Betygsgränser: 12p: G, 21p: VG

1. En koltrast fångar mask i gryningen. Maskarnas längder i centimeter kan betraktas som oberoende slumpvariabler var och en fördelad $N(21, 16)$. Eftersom koltrasten är mycket hungrig äter den 50 maskar. Vad är sannolikheten att maskarnas sammanlagda längd överstiger 10 meter? (2p)

   Lösning Låt X_i vara längden av mask i, $i = 1, 2, \dots, 50$.
   Då är $X \in N(21, 16)$ och den sammanlagda längden i centimeter är $S = \sum_{i=1}^{50} X_i$ där $E(S) = \sum_{i=1}^{50} E(X_i) = 50 \cdot 21 = 1050$ och $V(X_i) = 16^2$ så $V(S) = \sum_{i=1}^{50} V(X_i) = 12800$ (ty $X_1, X_2, \dots, X_{50}$ stickprov så oberoende variabler). Därmed är $S \in N(1050, \sqrt{12800})$ och $P(\text{sammanlagd längd} > 10 \text{ meter}) = P(S > 1000) = 1 - P(S \leq 1000) = 1 - \Phi(\frac{1000-1050}{\sqrt{12800}}) =$
   $= \Phi(\frac{50}{\sqrt{12800}}) = \Phi(0.44) = \boxed{0.67}$ □

2. Koltrastens lunch består av äpplen. Mätt i tuggor är "tuggkvantiteten" från 7 äpplen:

   7.1 8.6 12.5 8.7 11.2 9.3 9.7

   (a) Vilken konfidensgrad har intervallet (8.6 , 11.2) för medianen? (3p)

   Antag att tuggkvantiteten normalfördelad med väntevärde μ och varians σ^2.

   (b) Gör ett 95% konfidensintervall för μ. (2p)

   (c) En elak skata hävdar att väntevärdet av koltrastens tuggkapacitet bara är 7 tuggor. Försök bevisa att den är större på signifikansnivå 1%. (3p)

   Lösning

   (a) $(8.6\,,\,11.2) = (x_{(2)}\,,\,x_{(n-1)})$ där $n = 7$.
   Konfidensgraden $= P(X_{(2)} \leq m \leq X_{(6)}) = P(m \leq X_{(6)}) - P(m \leq X_{(2)}) =$
   $= 1 - P(X_{(6)} \leq m) - (1 - P(X_{(2)} \leq m)) = \underbrace{P(X_{(2)} \leq m)}_{II} - \underbrace{P(X_{(6)} \leq m)}_{I}$
   $I = P(X_{(6)} \leq m, X_{(7)} \leq m) + P(X_{(6)} \leq m, X_{(7)} > m) =$
   $= P(X_1 \leq m) \cdots P(X_7 \leq m) + \binom{7}{1} P(X_1 \leq m) \cdots P(X_6 \leq m) P(X_7 > m) =$
   $= 0.5^7 + 7 \cdot 0.5^6(1 - 0.5) = 8 \cdot 0.5^7 = 0.5^4 = 0.0625$
   På liknande sätt är $II = 1 - P(X_{(2)} > m) =$
   $= 1 - (P(X_1 > m) \cdots P(X_7 > m) + \binom{7}{1} P(X_1 > m) \cdots P(X_6 > m) P(X_7 \leq m) =$
   $= 1 - 0.0625 = 0.9375$ varmed $II - I = 0.9375 - 0.0625 = 0.875$
   dvs konfidensgraden är $\boxed{87.5\%}$

(b) $\bar{x} = 67.1/7 = 9.586$ $\quad s^2 = \frac{1}{6}(662.33 - 7\cdot 9.586^2) = 3.188$ $\quad t_{\alpha/2}(n-1) = t_{0.025}(6) = 2.4469$
Konfidensintervallet är $(\bar{x} - c\,,\ \bar{x} + c)$ där $c = t_{\alpha/2}(n-1)\sqrt{s^2/n} = 2.4469\sqrt{3.188/7} =$
$= 1.651$ varmed intervallet är $(7.934\ldots\ ,\ 11.237\ldots)$ vilket avrundas till $\boxed{(7.93\,,\ 11.24)}$

(c) Hypotes: $\begin{cases} \mu = 7 \\ \mu > 7 \end{cases}$ Signifikansnivå: $\alpha = 0.01$
$u = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} = \dfrac{9.586 - 7}{\sqrt{3.188}/\sqrt{7}} = 3.83$
$t_\alpha(n-1) = t_{0.01}(6) = 3.1427$
Eftersom $u = 3.83 > 3.1427 = t_\alpha(n-1)$ så förkastas H_0 dvs
<u>koltrastens tuggor har större väntevärde än 7 på 1% signifikansnivå.</u> □

3. I tid för middag får koltrasten syn på ett fågelbord där antalet gästande fåglar är 5 varav 2 är på väg därifrån och 3 just påbörjat sin måltid. Om antalet fåglar förutom koltrasten vid fågelbordet är Poissonfördelat med parameter 6, vad är den betingade sannolikheten att koltrasten slipper trängas med fler än 5 fåglar givet att tre kommer sitta där hela måltiden? (3p)

Lösning Låt X vara antalet fåglar förutom koltrasten vid fågelbordet. $X \in Poi(6)$ så

$$P(X \leq 5 \mid X \geq 3) = \frac{P(3 \leq X \leq 5)}{P(X \geq 3)} = \frac{P(X \leq 5) - P(X \leq 2)}{1 - P(X \leq 2)} \overset{\text{tabell}}{=} \frac{0.446 - 0.062}{1 - 0.062} = \boxed{0.4094}$$

□

4. För att få reda på hur populärt basket är bland eleverna på rasterna observeras

Rast	10-rasten	matrasten	2-rasten
Antal spelare	11	12	8

(a) Beräkna det genomsnittliga antalet spelare per rast, mediantalet spelare per rast och variationsbredden för antal spelare per rast. (3p)

(b) En lärare påstår att deltagarantalet halveras från rast till rast. Gör ett test på 5% signifikansnivå av om detta påstående är falskt. (3p)

Lösning

(a) Medelvärdet: $\bar{x} = \frac{1}{3}(11 + 12 + 8) = \boxed{10.33}$
Medianen: $x_{(2)} = \boxed{11}$
Variationsbredden: $x_{(3)} - x_{(1)} = 12 - 8 = \boxed{4}$

(b) Att antalet spelare halveras för varje rast innebär att om det är x spelare rast 1 så är det $0.5x$ spelare rast 2 och $0.25x$ spelare rast 3. Det totala antalet spelare är $x + 0.5x + 0.25x = 11 + 12 + 8 = 31 \ \Rightarrow\ x = \frac{31}{1.75} = 17.7$ och därmed är $E_1 = 17.71$, $E_2 = 8.86$, $E_3 = 4.43$. Därmed har vi r-tabellen:

k	O_i	E_i
1	11	17.71
2	12	8.86
3	8	4.43

varmed $U = \sum_{i=1}^{3} \frac{(O_i - E_i)^2}{E_i} = \frac{45.02}{17.71} + \frac{9.86}{8.86} + \frac{12.74}{4.43} = 6.53$

$\chi^2_{0.05}(3-1) = 5.9915$. Eftersom $u = 6.53 > 5.9915 = \chi^2_{0.05}(2)$
så förkastas H_0 dvs läraren har fel på signifikansnivå 5%. $\square$

5. Antalet sena ankomster vid en skola är oberoende från vecka till vecka. Under ett läsår (dvs 36 veckor) har man 361 sena ankomster. Antalet sena ankomster per vecka har standardavvikelse $\sigma = 3$.
   (a) Testa om det förväntade antalet sena ankomster per vecka är större än 9 på valfri signifikansnivå under normalfördelningsantagande. (2p)
   (b) Antag att det förväntade antalet sena ankomster per vecka är 11 Vilken styrka har testet ovan med signifikansnivå 1%? (3p)

Lösning

(a) Låt X_i vara antaletsena ankomster vecka i. Vi vet att $X_i \in N(\mu, 3)$.
Vi vill testa hypotesen $\begin{cases} H_0 : \mu = 9 \\ H_1 : \mu > 9 \end{cases}$ Läsåret innehåller 36 veckor så då är $u = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{361/36 - 6}{3/\sqrt{36}} = 2.0555$ och därmed $u = 2.0555 > 1.96 = \lambda_{0.025}$ men $u = 2.0555 \not> 2.33 = \lambda_{0.01}$ varmed H_0 förkastas om man använt signifikansnivå $\alpha \geq 0.025$ men H_0 förkastas ej om signifikansnivån $\alpha \leq 0.01$.

(b) $\begin{cases} H_0 : \mu = 9 \\ H_1 : \mu > 9 \end{cases}$ Under H_1 är $\mu = 11$, $\alpha = 0.01, \quad \sigma = 3$

$\alpha = 0.01$ betyder att vi ska använda oss av $\lambda_{0.01} = 2.3263$.
Vidare är $\bar{X} \in N(11, 3)$ under H_1.

$$\begin{aligned}
\text{Styrkan} &= P\,(\text{förkasta } H_0 \,|\, H_1 : \mu = 11) \\
&= P(U > \lambda_{0.01} \,|\, \mu = 11) \\
&= 1 - P\left(\frac{\bar{X} - 9}{3/\sqrt{36}} \leq 2.3263 \,|\, \mu = 11\right) \\
&= 1 - P(\bar{X} \leq 2.3263 \cdot \tfrac{3}{\sqrt{36}} + 9 \,|\, \mu = 11) \\
&= 1 - \Phi\left(\frac{2.3263 \cdot \frac{1}{2} + 9 - 11}{1/2}\right) \\
&= 1 - \Phi(2.3263 - 2 \cdot 2) \\
&= 1 - \Phi(-1.42) \\
&= \Phi(1.42) \\
&= \boxed{0.9222}
\end{aligned}$$

$\square$

6. Ålder är typiskt en <u>icke</u>-normalfördelad variabel. Vid en stor skola arbetar 247 lärare, var och en med en ålder som har väntevärde 40 år och standardavvikelse 15 år. Vad är approximativt sannolikheten att genomsnittsåldern inte överstiger 38 år? Redogör för alla antaganden du gör. (3p)

Lösning Låt X_i vara åldern av lärare i där $i = 1, 2, \ldots, 247$. Vi vet att $E(X_i) = 40$ och att $V(X_i) = 15^2$. Antag att $X_1, X_2, \ldots, X_{247}$ är oberoende av varandra.
Då är enligt <u>CGS</u> $\bar{X} \overset{\text{appr.}}{\in} N(40, 15/\sqrt{247})$ så
$P(\text{genomsnittsåldern} \leq 38 \text{ år}) = P(\bar{X} \leq 38) = \Phi(\frac{38-40}{15/\sqrt{247}}) = 1 - \Phi(2 \cdot \frac{\sqrt{247}}{15}) =$
$= 1 - \Phi(2.095) = \boxed{0.0179}$.
Om man räknar med halv-korrektion fås $P(\text{genomsnittsåldern} \leq 38 \text{ år}) =$
$= P(\bar{X} \leq 38.5) = 1 - \Phi(1.5 \cdot \frac{\sqrt{247}}{15}) = \boxed{0.0582}$.
Båda dessa svar bedöms korrekta då ålder både kan uppfattas som diskret och kontinuerlig variabel. $\square$

7. Antag att $X_1, X_2, \ldots, X_n$ är ett stickprov på $X \in R(0, a)$ och låt $X_{(n)} = \max_{1 \leq i \leq n} X_i$. Beräkna konstanten C sådant att $CX_{(n)}$ är en väntevärdesriktig skattning av medianen för X. (3p)

Lösning

$E(CX_{(n)}) = m$ där $m : P(X \leq m) = \frac{1}{2} \;\Rightarrow\; \frac{1}{2} = \int_0^m f(x)\,dx = \int_0^m \frac{1}{a-0}\,dx = \frac{m}{a} \;\Rightarrow$
$\Rightarrow\; m = \frac{a}{2}$ dvs $CE(X_{(n)}) = \frac{a}{2}$ dvs $C = \frac{a}{2E(X_{(n)})}$. $E(X_{(n)}) = \int_{-\infty}^{\infty} x f_{X_{(n)}}(x)\,dx$ där
$f_{X_{(n)}}(x) = \frac{d}{dx} P(X_{(n)} \leq x)$ och $P(X_{(n)} \leq x) = P(X_1 \leq x\,,\,\ldots\,,\,X_n \leq x) =$
$= P(X_1 \leq x) \cdots P(X_n \leq x) = (\int_0^x \frac{1}{a}\,dy)^n = (\frac{x}{a})^n$ där $0 < x < a$.
Därmed är $f_{X_{(n)}}(x) = \frac{d}{dx}(\frac{x}{a})^n = \frac{1}{a^n} n x^{n-1}$ där $0 < x < a$ och slutligen
$E(X_{(n)}) = \int_0^a x \frac{n}{a^n} x^{n-1} dx = \frac{n}{a^n} \int_0^a x^n dx = \frac{n}{a^n} [\frac{x^{n+1}}{n+1}]_0^a = \frac{n a^{n+1}}{a^n (n+1)} = \frac{an}{n+1}$.
Därmed är konstanten $C = \frac{a}{2E(X_{(n)})} = \frac{a(n+1)}{2an} = \boxed{\frac{n+1}{2n}}$ $\square$