Lösningar till Tentamen i Matematisk Statistik, 7.5p

Distanskurs
17 april, 2010 kl. 9.00–13.00

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

1. Antag att $X \in N(0,1)$ och $Y \in N(\mu, 2)$ oberoende av varandra[1]. Beräkna μ sådant att $P(2X < Y) = 0.99$. (3p)

   Lösning: $P(2X < Y) = P(2X - Y < 0) =$
   $$\begin{cases} E(2X - Y) = 2E(X) - E(Y) = 2 \cdot 0 - \mu = -\mu \\ V(2X - Y) = 4V(X) + V(Y) = 2 + 2 = 6 \end{cases}$$
   $= P(\frac{2X-Y-(-\mu)}{\sqrt{6}} < \frac{0-(-\mu)}{\sqrt{6}}) = \Phi(\frac{\mu}{\sqrt{6}}) = 0.99 \;\Rightarrow\; \frac{\mu}{\sqrt{6}} = 2.33 \;\Rightarrow\; \mu = 5.7073.$ □

2. Antalet bilar som passerar ett övergångsställe under 1 minut är Poissonfördelat med $\lambda = 6$. Vad är sannolikheten att det under 1 minut passerar exakt 7 bilar? (2p)

   Lösning: $P(X = 7) = e^{-6}\frac{6^7}{7!} = e^{-6}\frac{3888}{70} = 0.137677.$ □

3. Bosse säljer majblommor. Under 8 dagar får han följande försäljningssiffror:

ons	tors	fre	lör	sön	mån	tis	ons
10	18	13	23	22	19	15	12

   Antag att antalet sålda majblommor är oberoende från dag till dag.

   (a) Beräkna medianen för stickprovet. (2p)

   (b) Gör ett test på 5% signifikansnivå av om Bosse säljer fler än 14 majblommor per dag. (3p)

   (c) Gör ett test av om antalet sålda majblommor inte är likformigt fördelat bland de 8 dagarna på 1% signifikansnivå. (3p)

   (d) Beräkna ett 95% konfidensintervall för variansen av antalet sålda majblommor per dag. (3p)

   (e) Hur stor styrka har ett test på 1% signifikansnivå av hypotesen $H_0 : \mu = 14$ mot $H_1 : \mu = 17$ om $\sigma^2 = 25$? (3p)

   Lösning:

   (a) Det ordnade stickprovet: 10, 12, 13, 15, 18, 19, 22, 23. Eftersom stickprovsstorleken är ett jämnt tal är medianen $m = \frac{15+18}{2} = 16.5$.

   (b) Låt X beteckna antalet majblommor som Bosse säljer under en dag och beteckna $\mu = E(X)$. Det vi ska testa är då

[1] Fördelningarna är angivna på formen $N(\mu, \sigma^2)$ vilket innebär att $V(X) = 1$ och $V(Y) = 2$.

$$\begin{cases} H_0 : \mu = 14 \\ H_1 : \mu > 14 \end{cases}$$

Vi får $\bar{x} = \frac{1}{8}\sum_i x_i = 16.5$ och $\sum_i x_i^2 = 2336$ varmed $s^2 = \frac{2336-8\cdot 16.5^2}{7} = 22.57$. $\frac{\bar{x}-\mu_0}{s/\sqrt{n}} = \frac{16.5-14}{\sqrt{22.57/8}} = 1.488 \not> 1.8946 = t_{0.05}(7)$. Därmed kan H_0 ej förkastas, dvs Bosse kan ej visa att han säljer fler än 14 majblommor per dag på 5% signifikansnivå.

(c) Vi ska här testa $\begin{cases} H_0 : \text{likformigt fördelat} \\ H_1 : \text{ej likformigt fördelat} \end{cases}$

I fallet "likformigt fördelat" skulle $E(X_i)$ vara konstant m.a.p. i där X_i är antalet majblommor sålda dag i. Totalt har Bosse sålt $10+18+13+23+22+19+15+12 = 132$ majblommor så hypoteserna kan formuleras som $H_0 : E(X_i) = 16.5$ för alla i mot $H_1 : E(X_i) \neq E(X_j)$ för något par i, j. Vi får $\sum_i \frac{(O_i-E_i)^2}{E_i} = \frac{(10-16.5)^2}{16.5} + \frac{(18-16.5)^2}{16.5} + \ldots + \frac{(12-16.5)^2}{16.5} = \frac{1}{16.5}(42.25 + 2.25 + 12.25 + 42.25 + 30.25 + 6.25 + 2.25 + 20.25) = \frac{158}{16.5} = 9.5757 \not> 18.4753 = \chi^2_{0.01}(7)$. Därmed kan man ej förkasta H_0, dvs man kan ej visa att antalet majblommor ej är likformigt fördelade bland de 8 dagarna på 1% signifikansnivå.

(d) $\chi^2_{0.95}(7) = 2.1673$ så $\frac{(n-1)s^2}{\chi^2_{0.95}(n-1)} = \frac{7\cdot 22.57}{2.1673} = 72.902$ varmed det 95% konfidensintervallet blir $(0, 72.902)$.

(e) Styrkan $= P(\text{förkasta } H_0 \,|\, H_1) = P(\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} > \lambda_{0.01} \,|\, \mu = \mu_1) = P(\frac{\bar{X}-14}{5/\sqrt{8}} > 2.3263 \,|\, \mu = 17) = P(\bar{X} > \frac{5}{\sqrt{8}} \cdot 2.3263 + 14 \,|\, \mu = 17) = P(\frac{\bar{X}-17}{5/\sqrt{8}} > \frac{(5/\sqrt{8})\cdot 2.3263+14-17}{5/\sqrt{8}} \,|\, \mu = 17) = 1 - \Phi(\frac{5\cdot 2.3263-3\sqrt{8}}{5} = 1 - \Phi(0.63) = 0.2643$. □

4. Försäkringsbolaget *Skandsam* har följande statistik över trafikolyckor från en undersökning med 11% fotgängare, 71% bilister och 18% cyklister. Av dessa var

- 6% skadade bland fotgängarna
- 2% skadade bland bilisterna
- 9% skadade bland cyklisterna

Vad är sannolikheten att en skadad trafikant är cyklist? (3p)

Lösning: $P(\text{fotg}) = 0.11$, $P(\text{bilist}) = 0.71$, $P(\text{cykl}) = 0.18$
$P(\text{S}|\text{fotg}) = 0.06$, $P(\text{S}|\text{bilist}) = 0.02$, $P(\text{S}|\text{cykl}) = 0.09$.

$$\begin{aligned} P(\text{cykl}|\text{S}) &= \frac{P(\text{cykl} \cap \text{S})}{P(\text{S})} \\ &= \frac{P(\text{S}|\text{cykl})P(\text{cykl}}{P(\text{S}|\text{cykl})P(\text{cykl})P(\text{S}|\text{bilist})P(\text{bilist}) + P(\text{S}|\text{fotg})P(\text{fotg})} \\ &= \frac{0.09\cdot 0.18}{0.09\cdot 0.18 + 0.02\cdot 0.71 + 0.06\cdot 0.11} \\ &= 0.4378378 \end{aligned}$$

□

5. I ett lotteri utlovas 10% vinstchans på varje lott. Åsa köper 50 lotter men bara 1 av dessa ger vinst. Finns det anledning att betvivla "vinstgarantin"? Gör ett lämpligt test på 5% signivikansnivå. (4p)

Lösning: Om p betecknar sannolikheten att en lott ger vinst. Då är hypotesen vi vill pröva
$\begin{cases} H_0 : p = 0.1 \\ H_1 : p < 0.1 \end{cases}$
Om nu X är antalet vinster bland de 50 valda lotterna har vi att testets signifikansnivå ska vara $0.05 > P(X \leq 1 \,|\, H_0)$. Vi får att $> P(X \leq 1 \,|\, H_0) = P(X \leq 1 \,|\, X \in Bin(50, 0.1)) = \binom{50}{0} 0.1^0 (1-0.1)^{50-0} + \binom{50}{1} 0.1^1 (1-0.1)^{50-1} = 0.9^{50} + 50 \cdot 0.1 \cdot 0.9^{49} = 0.033786 < 0.05$. Alltså kan vi förkasta H_0, dvs ja, det *finns* anledning att betvivla vinstgarantin på 5% signifikansnivå. $\square$

6. Antag att variablerna $X_1, X_2, \ldots, X_n$ är fördelade enligt $P(X_i = -1) = P(X_i = 1)$ för alla $i = 1, 2, \ldots, n$ oberoende av varandra. Visa att

$$1 \ + \ \frac{2}{n} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_i X_j$$

är asymptotiskt $\chi^2(1)$ (dvs chi-två-fördelad med 1 frihetsgrad). (4p)

Tips: Om $X_1, X_2, \ldots, X_n$ är oberoende och $X_i \in N(0,1)$ för alla i, så är $\sum_{i=1}^{n} X_i^2 \in \chi^2(n)$ (dvs chi-två-fördelad med n frihetsgrader).

Lösning: Vi har att $\sum_{i=1}^n X_i^2 = n$ eftersom $X_i^2 = 1$ oavsett om $X_i = 1$ eller $X_i = -1$. Därför är $(\sum_{i=1}^n X_i) = \sum_{i=1}^n X_i^2 + \sum_{i=1}^n \sum_{j \neq i} X_i X_j = n + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n X_i X_j = n(1 + \frac{2}{n} \sum_{i=1}^{n-1} \sum_{j=i+1}^n X_i X_j)$. Enligt CGS är $P(\frac{\sqrt{n}}{\sigma}(\bar{X} - \mu) \leq x) \to \Phi(x)$. I denna uppgift är $\mu = E(X_i) = (-1) \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = 0$ och $\sigma^2 = E(X_i^2) - E(X_i)^2 = (-1)^2 \cdot \frac{1}{2} + 1^2 \cdot \frac{1}{2} - 0 = 1$. Eftersom $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ har vi enligt CGS att $P(\sqrt{n} \cdot \frac{1}{n} \sum_{i=1}^n \leq n) \to \Phi(x)$, dvs $\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i$ är asymptotiskt standard normalfördelad. Enligt tipset är därmed $(\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i)^2$ asymptotiskt $\chi^2(1)$-fördelad. Men $(\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i)^2 = \frac{1}{n} (\sum_{i=1}^n X_i)^2 = \frac{1}{n} \cdot n(1 + \frac{2}{n} \sum_{i=1}^{n-1} \sum_{j=i+1}^n X_i X_j) = 1 + \frac{2}{n} \sum_{i=1}^{n-1} \sum_{j=i+1}^n X_i X_j$. $\square$