LÖSNINGAR TILL TENTAMEN I MATEMATISK STATISTIK, 5P

22 mars, 2001

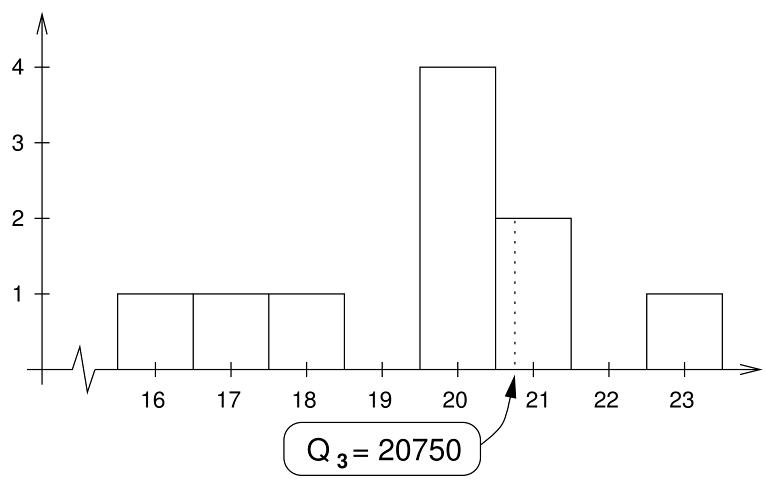
1. Månadslönerna för 10 lärare vid en viss skola är[1]

   17 700 19 800 19 900 20 200 20 800 16 100 17 000 23 500 19 700 21 100

   Beräkna medelvärdet, standardavvikelsen, medianen och tredje kvartilen? (2p)

   Lösning:
   $\bar{x} = \frac{195800}{10} = 19580$
   $\sum_i x_i^2 = 3\,875\,780\,000 \;\Rightarrow\; \frac{1}{10-1}(\sum x_i^2 - 10\bar{x}^2) = 4\,668\,444 \;\Rightarrow\; s = 2160$
   Sorterat stickprov:
   $16\,100, 17\,000, 17\,700, 19\,700, \underbrace{19\,800, 19\,900}_{\Rightarrow m=19850}, 20\,200, 20\,800, 21\,100, 23\,500$

   T.ex. med intervallen $(15500, 16500], (16500, 17500], \ldots, (22500, 23500]$ fås tredje kvartilen, Q_3, som det x-värde innan vilket 75% av den totala rektangelarean finns (se figur nedan).



   $\underline{\bar{x} = 19580,\ s = 2160,\ m = 19850,\ Q_3 = 20750}$

   Se vidare kap. 1

[1] För att denna uppgift skulle vara relevant, är det angivna stickprovet baserat på uppgifter från Lärarnas Riksförbund (hemsida: `http://www.lr.se/lrwebb~1.nsf/vHTML/Valkommen`).

2. Teofil metar i Laholmsbukten. Antalet fiskar som nappar under en timme är Poisson-fördelat, med $\lambda = 0.5$ för sill och $\lambda = 0.2$ för makrill (antalet makrillar är oberoende av antalet sillar och vice versa).

   a) Vad är Teofils chans att under 1 timme få 2 sillar och 2 makrillar? (2p)

   b) Vad är sannolikheten att Teofil under 6 veckors (2 timmar per dag, 4 dagar per vecka) fiske får *minst* 20 sillar? (Ev. approximationer ska anges!) (2p)

Lösning:

a)

$$\begin{aligned} P(\text{2 sillar och 2 makrillar}) &= p_{\text{sill}}(2)p_{\text{makrill}}(2) \\ &= e^{-0.5}\frac{0.5^2}{2!}\cdot e^{-0.2}\frac{0.2^2}{2!} \\ &= 0.0758\cdot 0.0164 \\ &= \underline{0.00124} \end{aligned}$$

Se vidare kap. 3

b) Antalet timmar Teofil fiskar är $2\cdot 4\cdot 4 = 32$. Låt X_i vara det antal fiskar han drar upp under timme i. Då är det antal fiskar han drar upp totalt $Y = \sum_{i=1}^{32} X_i$ där $X_1, \ldots, X_{32}$ är oberoende och $X_i \in Po(0.5)$. Denna summa är Poissonfördelad med $\lambda_Y = n\lambda = 32\cdot 0.5 = 16$ d.v.s. approximativt normalfördelad $N(\lambda_Y, \sqrt{\lambda_Y}) = N(16, 4)$ (enligt C.G.S.) om $\lambda_Y > 15$ och $16 > 15$ så ok! Därmed är

$$\begin{aligned} P(\text{minst 20 sillar}) &= P(Y \geq 20) \\ &= 1 - P(Y \leq 19) \\ &= 1 - \Phi\left(\frac{19-16}{4}\right) \\ &= 1 - \underbrace{\Phi(0.75)}_{0.7734} \\ &= \underline{0.2266} \end{aligned}$$

Se vidare kap. 6.3

3. Antag att U är rektangelfördelad på intervallet $(0,1)$. Vad är den betingade sannolikheten att $U > 1/5$ givet att $U \leq 2/3$? (2p)

Lösning:

$$U \in R(0,1) \;\Rightarrow\; f(x) = \begin{cases} 1 & \text{om } 0 < x < 1 \\ 0 & \text{annars} \end{cases} \quad \Rightarrow$$

$$\Rightarrow\; F(x) = \int_{-\infty}^{x} f(y)\,dy = \begin{cases} 0 & \text{om } x < 0 \\ \int_{-\infty}^{x} 1\,dy & \text{om } 0 < x < 1 \\ 1 & \text{om } x > 1 \end{cases} = \begin{cases} 0 & \text{om } x < 0 \\ x & \text{om } 0 < x < 1 \\ 1 & \text{om } x > 1 \end{cases}$$

Därmed är

$$\begin{aligned} P(X > \tfrac{1}{5} | X \leq \tfrac{2}{3}) &= \frac{P(X > \frac{1}{5}\,,\, X \leq \frac{2}{3})}{P(X \leq \frac{2}{3})} \\ &= \frac{P(\frac{1}{5} < X \leq \frac{2}{3})}{P(X \leq \frac{2}{3})} \\ &= \frac{P(X \leq \frac{2}{3}) - P(X \leq \frac{1}{5})}{P(X \leq \frac{2}{3})} \\ &= \frac{F(\frac{2}{3}) - F(\frac{1}{5})}{F(\frac{2}{3})} \\ &= \frac{2/3 - 1/5}{2/3} \\ &= \frac{2 \cdot 5 - 1 \cdot 3}{15} \cdot \frac{3}{2} \\ &= \frac{7}{10} \\ &= \underline{0.7} \end{aligned}$$

$\boxed{\text{Se vidare kap. 2.2 och 4.2}}$

4. En vaktmästare vid en skola ansvarar för att det finns tillräckligt med kontorsmatriel i förrådet. Han hävdar att det med 99% säkerhet jämt finns mellan 50 och 70 kollegieblock där. Du observerar under 6 veckor 71, 48, 53, 63, 77 och 62 block. Kan man lita på vaktmästarens ord? (3p)

Lösning:

$$\bar{x} = 62.3 \quad \text{och} \quad \sum x_i^2 = 23896 \;\Rightarrow\; \frac{1}{5}\big(23896 - 6\bar{x}^2\big) = 116.7 \;\Rightarrow\; s = 10.8$$

$$t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}} = t_{0.005}(5)\frac{s}{\sqrt{6}} = 4.03\frac{10.8}{\sqrt{6}} = 17.77$$

Ett 99% konf.int. är

$$\bar{x} \pm t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}} = (44.5, 80.1)$$

Eftersom vaktmästaren hävdade att antalet block var, med 99% säkerhet, mellan 50 och 70 så blir svaret: nej, han har fel.

Se vidare kap. 8

5. I statistisk årsbok finner man att att livslängden hos Sveriges befolkning är normalfördelad med $\mu = 80$ år och $\sigma = 10$ år. Enligt en undersökning av livslängden hos befolkningen i Blomstermåla mellan 1997 och 1999, dog 1000 människor under denna tid varav 129 var yngre än 70 år, 871 var äldre än 70 år, 540 var äldre än 80 år och 178 var äldre än 90. Finns det anledning att tro att invånarna i Blomstermåla skiljer sig från övriga landet beträffande livslängd? Gör ett test på 5% signifikansnivå. (3p)

Lösning:

Hypotes: $\begin{cases} H_0 : X \in N(80, 10) \\ H_1 : X \notin N(80, 10) \end{cases}$ Signifikansnivå: 5%

Vi vet:

Klass	Gränser	O_i
1	$x \leq 70$	129
2	$70 < x \leq 80$	$871 - 540 = 331$
3	$80 < x \leq 90$	$540 - 178 = 362$
4	$x > 90$	178

Vi vill även ha en fjärde kolumn med E_i, de förväntade antalen under nollhypotesen: $E_i = np_i$ där $p_i = P(X \text{ i klass } i|H_0)$. Alltså:

$$\begin{aligned} E_1 &= 1000p_1 \\ &= 1000P(X \leq 70|X \in N(80,10)) \\ &= 1000\Phi\left(\tfrac{70-80}{10}\right) \\ &= 1000\underbrace{(1-\Phi(1))}_{0.1587} \\ &= 159 \end{aligned}$$

$$\begin{aligned} E_2 &= 1000p_2 \\ &= 1000P(70 < X \leq 80|X \in N(80,10)) \\ &= 1000\left(\Phi\left(\tfrac{80-80}{10}\right) - \Phi\left(\tfrac{70-80}{10}\right)\right) \\ &= 1000(0.5 - 0.1587) \\ &= 341 \end{aligned}$$

$$E_3 = E_2 = 341 \quad \text{(p.g.a. symmetri)} \qquad E_4 = E_1 = 159$$

och därmed k-tabellen:

Klass	Gränser	O_i	E_i
1	$x \leq 70$	129	159
2	$70 < x \leq 80$	331	341
3	$80 < x \leq 90$	362	341
4	$x > 90$	178	159

Statistikan: $u = \sum_i \frac{(O_i-E_i)^2}{E_i} = \frac{(129-159)^2}{159} + \frac{(331-341)^2}{341} + \frac{(362-341)^2}{341} + \frac{(178-159)^2}{159} =$
$= 5.66 + 0.29 + 1.29 + 2.27 = 9.51$

Eftersom $\chi^2_\alpha(k-1) = \chi^2_{0.05}(3) = 7.81$ och $u = 9.51 > 7.81$ förkastas H_0 d.v.s. ja, livslängderna i Blomstermåla skiljer sig från resten av landet!

Se vidare kap. 10.3

6. Man gör undersökningar vid Halmstads badplatser för att få reda på om halten av gift från giftalger överstiger vissa gränsvärden. Om man (på signifikansnivå 1%) kan visa att den förväntade nivån överstiger 0.7 sätter man upp en varningsskylt som rekommenderar folk att ej bada där och om man kan visa att den förväntade nivån överstiger 1.0 utfärdar man badförbud. Antag att nivån är normalfördelad med $\sigma = 0.08$ och att man observerar nivåerna

$$1.18 \quad 0.99 \quad 0.83 \quad 0.71 \quad 1.27 \quad 0.49 \quad 1.58 \quad 1.05$$

a) Ska man sätta upp varningsskylten? (1p)

b) Ska man utfärda badförbud? (1p)

Lösning:

$\bar{x} = 1.0125$, $\sigma = 0.08$

a) Testa $\begin{cases} H_0 : \mu = 0.7 \\ H_1 : \mu > 0.7 \end{cases}$ på sign.nivå $\alpha = 0.01$

$u = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} = \frac{1.0125-0.7}{0.08/\sqrt{8}} = 11.0485 > 2.33 = \lambda_{0.01}$

så: ja, sätt upp skylten!

b) Testa $\begin{cases} H_0 : \mu = 1 \\ H_1 : \mu > 1 \end{cases}$ på sign.nivå $\alpha = 0.01$

$u = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} = \frac{1.0125-1}{0.08/\sqrt{8}} = 0.44 \not> 2.33 = \lambda_{0.01}$

så: nej, inget badförbud!

$\boxed{\text{Se vidare kap. 9.2}}$

7. Gränserna $a = 21.3$ och $b = 28.9$ är s.k. normalgränser för variabeln X. Detta innebär att

   - $X \in N(\mu, \sigma)$
   - $P(a \leq X \leq b) = 0.95$
   - a och b ligger symmetriskt över μ (d.v.s. $\mu - a = b - \mu$)

   Bestäm μ och σ. (3p)

   Lösning:
   Eftersom a, b ligger symmetriskt över μ så är $\mu - a = b - \mu$, d.v.s. $\mu - 21.3 = 28.9 - \mu$, d.v.s. $\mu = \frac{28.9+21.3}{2} = 25.1$
   Vidare är (rita gärna figur!)

$$\begin{aligned} 0.95 &= P(21.3 \leq X \leq 28.9) \\ &= \Phi\left(\frac{28.9 - \mu}{\sigma}\right) - \Phi\left(\frac{21.3 - \mu}{\sigma}\right) \\ &= \Phi\left(\frac{3.8}{\sigma}\right) - \Phi\left(\frac{-3.8}{\sigma}\right) \\ &= \Phi\left(\frac{3.8}{\sigma}\right) - \left(1 - \Phi\left(\frac{3.8}{\sigma}\right)\right) \\ &= 2\Phi\left(\frac{3.8}{\sigma}\right) - 1 \end{aligned}$$

   $\Rightarrow\ \Phi(\frac{3.8}{\sigma}) = \frac{1}{2}(0.95 + 1) = 0.975 \ \Rightarrow\ \frac{3.8}{\sigma} = 1.96 \ \Rightarrow\ \sigma = 1.94.$

   $\underline{\mu = 25.1 \text{ och } \sigma = 1.94.}$

   $\boxed{\text{Se vidare kap. 6.1}}$

8. Arne vill åka med sina kompisar på en resa till London under våren. Resan kostar 1800:- och för att tjäna ihop till det säljer han morgontidningar varje söndag och första maj-blommor under 4 veckor. Inkomsten från tidningsförsäljningen är A_i kronor vecka i där $A_i \in N(170, 50)$. Majblommorna ger $B_i \in N(30, 15)$ kronor per vecka under 4 veckor. De ska resa om 11 veckor. Vad har Arne för chans att få råd att hänga med? (3p)

Lösning:

Låt $Y = \sum_{i=1}^{11} A_i + \sum_{j=1}^{4} B_i$. (Antar intäkterna $A_1, \ldots, A_{11}, B_1, \ldots, B_4$ är oberoende.)

Eftersom

$A_i \in N(170, 50) \;\Rightarrow\; \sum A_i \in N(11 \cdot 170, \sqrt{11} \cdot 50) = N(1870, \sqrt{11} \cdot 50)$

$B_j \in N(30, 15) \;\Rightarrow\; \sum B_j \in N(4 \cdot 30, \sqrt{4} \cdot 15) = N(120, 30)$

så är $Y \in N(1870 + 120, \sqrt{11 \cdot 50^2 + 30^2}) = N(1990, \sqrt{28400})$ varmed

$$\begin{aligned}
P(\text{Arne har råd}) &= P(Y > 1800) \\
&= 1 - \Phi\left(\frac{1800 - 1990}{\sqrt{28400}}\right) \\
&= 1 - \left(1 - \Phi\left(\frac{190}{\sqrt{28400}}\right)\right) \\
&= 1 - (1 - \Phi(1.127)) \\
&= 0.8708
\end{aligned}$$

<u>Arnes chans är 87%.</u>

$\boxed{\text{Se vidare kap. 6.2}}$

9. En klass på 13 elever ska ha fysikprov. Fysikläraren tillhandahåller en speciell sorts miniräknare vid alla fysikproven. Denna gång fungerar batterierna för alla elever utom 2. Finns det anledning att tro att sannolikheten p, för att batterierna i en miniräknare skall vara slut, är större än 0.05? Välj själv signifikansnivå för ett lämpligt test. (3p)

Lösning:

Låt $X_i = \begin{cases} 1 & \text{om batterierna är slut för den } i\text{:te eleven} \\ 0 & \text{annars} \end{cases}$

och $P(X_i = 1) = p$, $P(X_i = 0) = 1 - p$ och $p_0 = 0.05$.

Vi vill testa $\begin{cases} H_0 : p = 0.05 \\ H_1 : p > 0.05 \end{cases}$ på t.ex. sign.nivå $\alpha = 0.05$.

"Alla utom 2 funkar" innebär att $u = \sum_{i=1}^{13} x_i = 2$ där U är binomialfördelad med $n = 13$ och p, dvs under H_0 är $U \in Bin(n, p_0) = Bin(13, 0.05)$.

Därmed är

$$\begin{aligned} \alpha_0 &= P(U > u|H_0) \\ &= 1 - P(U \leq u|H_0) \\ &= 1 - P(U \leq 2 \mid U \in Bin(13, 0.05)) \\ &= 1 - 0.97549 \\ &= 0.02451 \end{aligned}$$

Eftersom $\alpha_0 = 0.02451 < 0.05 = \alpha$ kan H_0 förkastas d.v.s. <u>risken att batterierna är slut är större än 0.05.</u>

(På sign.nivå 1% hade man dock ej kunnat förkasta H_0 eftersom $0.02451 \not< 0.01$.)

$\boxed{\text{Se vidare kap. 9.4}}$

10. Antag att X_1, X_2, X_3 är ett stickprov på $X \in R(0,2)$ och låt

$$m_1^* = \frac{1}{3}(X_1 + X_2 + X_3) \qquad \text{och} \qquad m_2^* = 2\min(X_1, X_2, X_3)$$

vara två punktskattningar av $\mu = E(X)$.

a) Är m_1^* och m_2^* väntevärdesriktiga? (3p)

b) Vilken skattning är effektivast? (2p)

Lösning:

a) $E(X) = \mu = \int_0^2 x\frac{1}{2}\,dx = \frac{1}{2}[\frac{x^2}{2}]_0^2 = \frac{1}{2}(\frac{4}{2} - 0) = 1$

För att kolla om m_1^* och m_2^* är väntevärdesriktiga ska vi beräkna deras väntevärde och se om de är $\mu = 1$:

$$E(m_1^*) = E(\frac{1}{3}(X_1 + X_2 + X_3)) = \frac{1}{3}(E(X_1) + E(X_2) + E(X_3)) = \frac{1}{3}3\mu = 1$$

$E(m_2^*) = 2E(\min X_i) = 2\int_0^2 x f_{\min X_i}(x)\,dx$

$F_{\min X_i}(x) = P(\min X_i \le x) = 1 - P(\min X_i > x) =$

$= 1 - P(X_1 > x)P(X_2 > x)P(X_3 > x) =$

$= 1 - (1 - P(X_1 \le x))(1 - P(X_2 \le x))(1 - P(X_3 \le x)) =$

$= 1 - (1 - \frac{x}{2})^3$

$f_{\min X_i}(x) = \frac{d}{dx}P(\min X_i \le x) = 0 - 3(-\frac{1}{2})(1 - \frac{x}{2})^2 = \frac{3}{2}(1 - \frac{x}{2})^2$, då $0 \le x \le 2$

$E(\min X_i) = \int_{-\infty}^{\infty} x f_{\min X_i}(x)\,dx = \int_0^2 x\frac{3}{2}(1 - \frac{x}{2})^2\,dx = \frac{3}{2}\int_0^2 x(1 - x + \frac{x^2}{4})\,dx =$

$= \frac{3}{2}\int_0^2 (x - x^2 + \frac{x^3}{4})\,dx = \frac{3}{2}[\frac{x^2}{2} - \frac{x^3}{3} + \frac{x^4}{16}]_0^2 = \frac{3}{\cancel{2}}(\frac{2^{\cancel{2}\,1}}{2} - \frac{2^{\cancel{3}\,2}}{3} + \frac{2^{\cancel{4}\,3}}{16} - 0) =$

$= 3(1 - \frac{4}{3} + \frac{1}{2}) = \frac{1}{2} \qquad (*)$

$E(m_2^*) = E(2\min X_i) = 2E(\min X_i) = 2 \cdot \frac{1}{2} = 1$

<u>Ja, både m_1^* och m_2^* är väntevärdesriktiga.</u>

b) Ska kolla om $V(m_1^*) < V(m_2^*)$ eller om $V(m_1^*) > V(m_2^*)$:

$$V(m_1^*) = V(\frac{1}{3}(X_1 + X_2 + X_3)) = \frac{1}{9}(V(X_1) + V(X_2) + V(X_3)) = \frac{1}{9} \cdot 3 \cdot \frac{(2-0)^2}{12} = \frac{1}{9}$$

$V(m_2^*) = V(2\min X_i) = 4\big(E((\min X_i)^2) - E(\min X_i)^2\big)$

$E((\min X_i)^2) = \int_{-\infty}^{\infty} x^2 f_{\min X_i}(x)\,dx = \int_0^2 x^2\frac{3}{2}(1 - \frac{x}{2})^2 dx =$

$= \frac{3}{2}\int_0^2 x^2(1 - x + \frac{x^2}{4})\,dx = \frac{3}{2}[\frac{x^3}{3} - \frac{x^4}{4} + \frac{x^5}{20}]_0^2 = \frac{3}{2}(\frac{2^3}{3} - \frac{2^4}{4} + \frac{2^5}{20} - 0) =$

$= 3(\frac{4}{3} - \frac{8}{4} + \frac{16}{20}) = 4 - 6 + \frac{12}{5} = \frac{2}{5}$

Från $(*)$ i a-uppg. vet vi att $E(\min X_i) = \frac{1}{3}$ varmed

$V(m_2^*) = 4\big(E((\min X_i)^2) - E(\min X_i)^2\big) = 4(\frac{2}{5} - (\frac{1}{3})^2) = 4\frac{2\cdot 9 - 5\cdot 1}{5\cdot 9} = \frac{52}{45}$

Alltså är $V(m_1^*) = \frac{1}{9} < \frac{52}{45} = V(m_2^*)$ d.v.s. <u>m_1^* är effektivare än m_2^*.</u>

Se vidare kap. 7.4