



TENTAMEN I SF1903 SANNOLIKHETSLÄRA OCH STATISTIK FÖR 3-ÅRIG MEDIA
MÅNDAGEN DEN 15 AUGUSTI 2011 KL 08.00–13.00.

Examinator: Gunnar Englund, tel. 790 7416.

Tillåtna hjälpmedel: Läroboken, Formel- och tabellsamling i Matematisk statistik. Räknare.

Införda beteckningar skall förklaras och definieras. Resonemang och uträkningar skall vara så utförliga och väl motiverade att de är lätta att följa. Numeriska svar skall anges med minst två siffrors noggrannhet. Tentamen består av 5 uppgifter. Varje korrekt lösning ger 10 poäng. Gränsen för godkänt är preliminärt 20 poäng. Möjlighet att komplettera ges för de tentander med 18–19 poäng. Tid och plats för komplettering kommer att anges på kursens hemsida. Det ankommer på dig själv att ta reda på om du har rätt att komplettera.

Tentamen kommer att finnas tillgänglig på elevexpeditionen sju veckor efter skrivnings-tillfället.

Uppgift 1

För händelserna A och B gäller att $P(A) = 0.11$, $P(B) = 0.17$ och att sannolikheten att någon av händelserna inträffar är 0.26.

Beräkna sannolikheterna

- a) att ingen av händelserna A och B inträffar, (5 p)
- b) att exakt en av händelserna A och B inträffar. (5 p)

Uppgift 2

En metallurg önskar skatta skillnaden Δ mellan halterna av ett visst ämne i två prov A och B . Hon gör sju bestämningar på prov A och elva bestämningar på prov B och får resultaten $x_1, x_2, \dots, x_7$ resp. $y_1, y_2, \dots, y_{11}$. Hon beräknar medelvärdena av sina observationer och får $\bar{x} = 9.58$, $\bar{y} = 8.35$ och standardavvikelserna till $s_x = 2.25$ respektive $s_y = 3.36$.

De sju resp. elva observationerna antas vara oberoende och komma från $N(m + \Delta, \sigma)$ - resp. $N(m, \sigma)$ -fördelningar.

- (a) Bestäm ett 95% konfidensintervall för Δ . (5 p)
- (b) Ge ett 95% konfidensintervall för den okända standardavvikelsen σ . (5 p)

Uppgift 3

Ett bussbolag som har 81 bussar önskar dimensionera en reparationsverkstad. Man önskar därför få en uppfattning om det totala antalet större reparationer som behöver göras på en månad. Av erfarenhet vet man att antalet gånger som en buss behöver genomgå en större

reparation under en månad kan beskrivas av en stokastisk variabel X med sannolikhetsfunktionen

$$p_X(k) = \begin{cases} 0.2 & \text{om } k = 0 \\ 0.5 & \text{om } k = 1 \\ 0.2 & \text{om } k = 2 \\ 0.1 & \text{om } k = 3 \\ 0 & \text{annars.} \end{cases}$$

a) Beräkna $E(X)$ och $D(X)$. (5 p)

b) Beräkna med lämplig och välmotiverad approximation sannolikheten att minst 90 och högst 110 större reparationer totalt behöver göras under en månad. (5 p)

Uppgift 4

Nationalekonomiprofessorerna Peter Bohm och Martin Dufwenberg genomförde ett experiment med 396 studenter på grundkursen i nationalekonomi vid Stockholms universitet. Dessa studenter får anses vara relativt väl förtrodda med EMU-frågan.

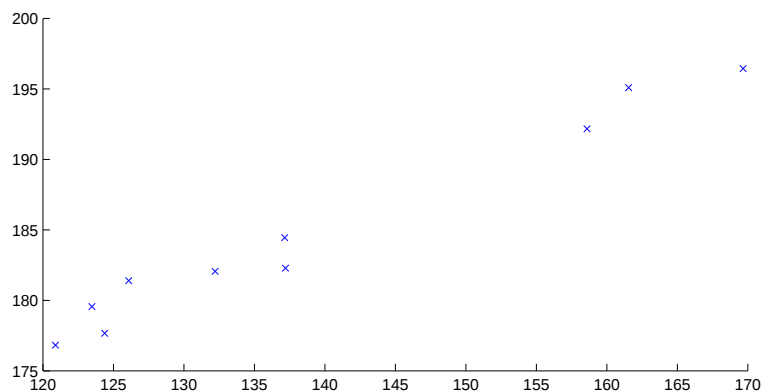
Ett slumpvis urval av hälften av studenterna fick frågan ”Skall Sverige gå med i den europeiska monetära unionen (EMU) och införa euro som betalningsmedel?” där 107 av 198 tillfrågade svarade ”Ja”, den andra hälften fick frågan ”Skall Sverige gå med i den europeiska monetära unionen (EMU) och avskaffa kronan som betalningsmedel?” där 95 av 198 tillfrågade svarade ”Ja”.

Beräkna ett (approximativt) 95%-igt konfidensintervall för skillnaden i ”Ja”-andelar och använd detta för att på nivån 5% testa om undersökningen ger belägg för att fråge-formuleringen påverkar resultatet. (10 p)

Uppgift 5

Tio mätningar av mängden spjälkat protein, $y_1, \dots, y_{10}$, är gjorda vid tio olika koncentrationer, $x_1, \dots, x_{10}$, av det aktiva enzymet. Ansätt regressionsmodellen att vid enzymkoncentration x_i beskrivs mängden protein av Y_i där Y_i är $N(\alpha + \beta x_i, \sigma)$.

De $n = 10$ mätningarna



sammanfattas av följande storheter

$$\bar{x} = 139.12 \quad \bar{y} = 184.79$$

$$\sum_{i=1}^n (x_i - \bar{x})^2 = 2834.9 \quad \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = 1131.4 \quad \sum_{i=1}^n (y_i - \bar{y})^2 = 463.36$$

- a) Skatta regressionslinjens parametrar α och β . (4 p)
- b) Skatta σ . (2 p)
- c) Testa hypotesen $H_0 : \beta = 0.30$ mot $H_1 : \beta \neq 0.30$ på signifikansnivå 0.05. (4 p)



LÖSNINGAR TILL

TENTAMEN I SF1903 SANNOLIKHETSLÄRA OCH STATISTIK FÖR 3-ÅRIG MEDIA
MÅNDAGEN DEN 15 AUGUSTI 2011 KL 08.00–13.00.

Uppgift 1

a) $P(\text{ingen av } A \text{ och } B \text{ inträffar}) = 1 - P(\text{någon av } A \text{ och } B \text{ inträffar}) = 0.74.$

b) Additionssatsen $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ ger

$$P(A \cap B) = 0.11 + 0.17 - 0.26 = 0.02.$$

Genom att teckna en figur ser man att den sökta sannolikheten blir

$$P(A \cap B^*) + P(A^* \cap B) = P(A \cup B) - P(A \cap B) = 0.26 - 0.02 = 0.24.$$

Uppgift 2

(a) Den sammanvägda variansskattningen s^2 är

$$s^2 = \frac{6s_x^2 + 10s_y^2}{16} = 8.954$$

Konfidensintervallet ges av

$$\begin{aligned} I_\Delta &= \bar{x} - \bar{y} \pm t_{\alpha/2}(n_x + n_y - 2)s \sqrt{\frac{1}{n_x} + \frac{1}{n_y}} = \\ &= 9.58 - 8.35 \pm t_{0.025}(16)s \sqrt{\frac{1}{7} + \frac{1}{11}} = 1.23 \pm 2.12 \cdot \sqrt{8.954 \left(\frac{1}{7} + \frac{1}{11} \right)} = \underline{1.23 \pm 3.07}. \end{aligned}$$

(b) Konfidensintervallet för σ^2 ges av

$$I_{\sigma^2} = \left(\frac{16s^2}{\chi_{0.025}^2(16)}, \frac{16s^2}{\chi_{0.975}^2(16)} \right) = (4.97, 20.73)$$

eftersom kvantilerna $\chi_{0.025}^2(16)$ och $\chi_{0.975}^2(16)$ är 28.8 respektive 6.91. Drar vi roten ur gränserna så erhåller vi ett konfidensintervall för σ , dvs $I_\sigma = \underline{(2.23, 4.55)}$.

Uppgift 3

a) Vi har att

$$\begin{aligned} E(X) &= \sum_{k=0}^3 k p_X(k) = 0 \cdot 0.2 + 1 \cdot 0.5 + 2 \cdot 0.2 + 3 \cdot 0.1 = \underline{1.2}. \\ E(X^2) &= \sum_{k=0}^3 k^2 p_X(k) = 0^2 \cdot 0.2 + 1^2 \cdot 0.5 + 2^2 \cdot 0.2 + 3^2 \cdot 0.1 = 2.2, \end{aligned}$$

vilket ger

$$V(X) = E(X^2) - (E(X))^2 = 2.2 - 1.2^2 = 2.2 - 1.44 = 0.76$$

och

$$D(X) = \sqrt{0.76} = \underline{0.872}.$$

b) Låt X_j vara antalet reparationer som buss nr j behöver genomgå under en månad, $j = 1, \dots, 81$. Modell: $X_1, \dots, X_{81}$ är oberoende och lika fördelade stokastiska variabler med samma fördelning som X . Sätt $Y = \sum_{j=1}^{81} X_j$. Vi har

$$E(Y) = \sum_{j=1}^{81} E(X_j) = 81 \cdot 1.2 = 97.2 = m$$

och

$$V(Y) = (\text{pga oberoendet}) = \sum_{j=1}^{81} V(X_j) = 81 \cdot 0.76 = 61.56 = \sigma^2.$$

Centrala gränsvärdessatsen ger att Y är approximativt normalfördelad $N(m, \sigma)$. Således fås

$$\begin{aligned} P(90 \leq Y \leq 110) &= P\left(\frac{90-m}{\sigma} \leq \frac{Y-m}{\sigma} \leq \frac{110-m}{\sigma}\right) \approx \Phi\left(\frac{110-m}{\sigma}\right) - \Phi\left(\frac{90-m}{\sigma}\right) \\ &= \Phi(1.63) - \Phi(-0.92) = \Phi(1.63) - (1 - \Phi(0.92)) \\ &= 0.9484 - (1 - 0.8212) = \underline{0.77}. \end{aligned}$$

Uppgift 4

Vår modell är att antalet "Ja"-anhängare är $\text{Bin}(198, p_1)$ respektive $\text{Bin}(198, p_2)$. Vi vill beräkna ett 95%-igt konfidensintervall för $p_1 - p_2$. Vi skattar p_1 med $p_1^* = 107/198 \approx 0.540$ och p_2 med $p_2^* = 95/198 \approx 0.480$. Eftersom $np_i^*(1-p_i^*)$ är approximativt 50 i båda fallen kan vi alltså approximerar binomial-fördelningarna med normalfördelningar. Vi får att p_i^* är approximativt $N(p_i, \sqrt{p_i(1-p_i)/198})$ för $i = 1, 2$ och ser att $p_1^* - p_2^*$ är approximativt

$$N\left(p_1 - p_2, \sqrt{\frac{p_1(1-p_1)}{198} + \frac{p_2(1-p_2)}{198}}\right)$$

som med approximativa metoden (FS 11.3) ger ett (approximativt) 95%-igt konfidensintervall för $p_1 - p_2$ till

$$\begin{aligned} p_1^* - p_2^* \pm \lambda_{0.025} \sqrt{\frac{p_1^*(1-p_1^*)}{198} + \frac{p_2^*(1-p_2^*)}{198}} &\approx \\ \approx 0.540 - 0.480 \pm 1.9600 \sqrt{\frac{0.54 * 0.46 + 0.49 * 0.51}{198}} &\approx 0.060 \pm 0.098. \end{aligned}$$

Eftersom 0 ingår i konfidensintervallet kan vi inte förkasta $H_0 : p_1 = p_2$ och undersökningen visar alltså inte att fråge-formuleringen spelar någon roll.

Kommentar: Testet (men inte konfidensintervallet) kan utföras som homogenitetstest.

Uppgift 5

a) Med beteckningar som i uppgiftslydelsen skattas α och β med

$$\beta_{\text{obs}}^* = \frac{S_{xy}}{S_{xx}} = \frac{1131.4}{2834.9} = 0.39909, \quad \alpha_{\text{obs}}^* = \bar{y} - \beta_{\text{obs}}^* \bar{x} = 184.79 - 0.39909 \cdot 139.12 = 129.27$$

b) Vi skattar σ^2 med

$$s^2 = \frac{1}{n-2} \left(\sum_{i=1}^n (y_i - \bar{y})^2 - (\beta_{\text{obs}}^*)^2 \sum_{i=1}^n (x_i - \bar{x})^2 \right) = 1.4777$$

dvs, σ skattas med $s = \underline{1.2156}$.

c) Vi förkastar hypotesen $H_0 : \beta = 0.30$ till förmån för $H_1 : \beta \neq 0.30$ på nivån 5% om 0.30 ej ingår i det 95%-iga konfidensintervallet för β . Detta intervall blir

$$\beta_{\text{obs}}^* \pm t_{0.025}(10-2) \frac{s}{\sqrt{S_{xx}}} = 0.39909 \pm 2.31 \frac{1.2156}{\sqrt{2834.9}} = 0.39909 \pm 0.0527.$$

Undre gränsen blir 0.3464 så $H_0 : \beta = 0.30$ förkastas.