Linköpings universitet
Institutionen för teknik och naturvetenskap
Michael Hörnquist

Statistik och sannolikhetslära, TNIU66

Tentamen, 7 juni 2022, Svar och lösningsskisser

1. (a) Man inser $\mu_A < \mu_C = 3,5 < \mu_B$, eftersom datamängd A har fler låga värden (mellan 1 och 3) medan datamängd C har färre låga värden. (b) Exempelvis populationsstandardavvikelsen blir $\sigma = \sqrt{\sum_i (x_i - \mu_C)^2/N} = \sqrt{((1-3,5)^2 + (2-3,5)^2 + (3-3,5)^2) \cdot 2 \cdot 3/18} = \sqrt{35/12} \approx 1,708$.

2. Låt $X_i, i = 1, 2, 3$ vara slumpvariabler som står för graviditetslängd, var en av dem är fördelad $N(\mu = 266, \sigma = 16)$. (a) $\Pr(X_i < 266 - 14) = \Pr(Z < -0,875) = (0,19215 + 0,18943)/2 = 0,19079 \approx 19\%$. (b) Enligt komplementregeln och multiplikationssatsen för oberoende händelser fås $\Pr(\text{minst en}) = 1 - \Pr(\text{ingen}) = 1 - \Pr(\text{inte Sigrid}) \Pr(\text{inte Eva}) \Pr(\text{inte Katarina}) = 1-(1-p)^3$, där $p = \Pr(X_i < 244-14) = 0,19079$ från (a). Vi landar i $\Pr(\text{minst en}) = 1-(1-0,19079)^3 = 0,47011 \approx 47\%$. (c) Enligt betingad sannolikhet fås $\Pr(266 - 14 < X_3 < 266 + 14 | X_3 > 260) = \Pr((252 < X_3 < 280) \cap (X_3 > 260))/\Pr(X_3 > 260) = \Pr(260 < X_3 < 280)/\Pr(X_3 > 260) = \Pr(-0,375 < Z < 0,875)/(1 - \Pr(Z < 0,375)) = (0,80921 - 0,35383)/(1 - 0,35383) = 0,70474 \approx 70\%$.

3. Täthetsfunktioner som hör till de kontinuerliga slumpvariablerna är $f_X(x) = 0,5$ om $-1 \leq x \leq 1$ och $f_X(x) = 0$ för övrigt, samt $f_Y(x) = e^{-x}$ om $x \geq 0$ och $f_Y(x) = 0$ för övrigt.
(a) $\mu_X = E(X) = ((-1) + 1)/2 = 0$, $\sigma_X = SD(X) = (1 - (-1))/\sqrt{12} = 1/\sqrt{3} \approx 0,577$. Kan även räknas ut för kontinuerliga slumpvariabler enligt $E(X) = \int_{-\infty}^{\infty} x f_X(x)\, dx$ och $SD(X) = \sqrt{\int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x)\, dx}$. (b) $\mu_Y = E(Y) = \lambda = 1$, $\sigma_X = SD(X) = \lambda = 1$ Standardresultat för exponentialfördelade slumpvariabler, kan även räknas fram på samma sätt som i (a). (c) $\Pr(\mu_X - 2,5\sigma_X \leq X \leq \mu_X + 2,5\sigma_X) = \int_{\mu_X - 2,5\sigma_X}^{\mu_X + 2,5\sigma_X} f_X(x)\, dx = \int_{-1}^{1} 0,5\, dx = 1$. (d) $\Pr(\mu_Y - 2,5\sigma_Y \leq Y \leq \mu_Y + 2,5\sigma_Y) = \int_{\mu_Y - 2,5\sigma_Y}^{\mu_Y + 2,5\sigma_Y} f_Y(x)\, dx = \int_{0}^{1+2,5\cdot 1} e^{-x}\, dx = 1 - e^{-3,5} \approx 0,9698$.

4. Vi kan använda normalfördelningsapproximationen, $np(1-p) \geq 1003 \cdot 0,35(1 - 0,35) = 228 > 5$. Konfidensintervallet ges av $p \pm z_{1-\alpha/2}\sqrt{p(1-p)/n}$. Opinion *för* medlemsskap: $p = 0,41$, ger intervallet $37,0\% < \pi < 45,0\%$. Opinion *mot* medlemsskap: $p = 0,35$, ger intervallet $31,1\% < \pi < 38,9\%$. I bägge fallen har tabellvärdet $z_{0,995} = 2,576$ använts.

5. (a) Falskt. Sannolikheten för att en hypotes är sann eller falsk är aldrig det som beräknas.
   (b) Sant.
   (c) Falskt. Bara ord staplade på varandra för att det ska låta bra.
   (d) Falskt. Se (e).
   (e) Sant.
   (f) Sant. Vi har $\beta = 1 - \gamma$ med kursens beteckningar.

6. (a) Förklaringsgraden fås som ”R-kvadrat”, dvs 54%. (b) Stickprovsmodellens regressionsparametrar avläses från ”Konstant” och ”X-variabel 1” så att ekvationen för regressionslinjen blir $\hat{y}_{x^*} = 1176,97 + 140,61x^*$. (c) Punktskattning för $x^* = 30$ grader fås enligt $\hat{y}_{30} = 5395,27 \approx 5395$ kronor. (d) Felmarginalen för prognosintervall[1] ges av $t_{n-2;1-\alpha/2} s \sqrt{1 + (1/n) + (x^* - \bar{x})^2 / \sum_{i=1}^{n} (x_i - \bar{x})^2}$. Med $n = 365$ och $1 - \alpha = 0,90$ fås tabellvärdet $t_{n-2;1-\alpha/2} = 1,650$ och ur utdatasammanfattningen $s = 1321,19$. Ur texten hittar vi $\bar{x} = 9,46$ och $\sigma_x = 10,42$, detta ger bidragen $(x^* - \bar{x})^2 = (30 - 9,46)^2 = 421,8916$ och $\sum_{i=1}^{n} (x_i - \bar{x})^2 = n\sigma_x^2 = 139630,386$. Därmed blir prognosintervallet [3189, 7601] kr.

[1] I uppgiften stod ”konfidensintervall”, även prognosintervall är konfidensintervall.