

1. (a) Låt  $R, B, S$  beteckna händelserna att respektive färg fungerar. Eftersom pennorna fungerar oberoende av varandra ges sannolikheten av

$$P(R^c \cap B^c \cap S^c) = P(R^c)P(B^c)P(S^c) = (1 - 0.9)(1 - 0.8)(1 - 0.4) = 0.1 \cdot 0.2 \cdot 0.6 = 0.012.$$

- (b) Låt  $P_R, P_B, P_S$  beteckna händelserna att respektive märke väljs och  $F$  att den valda pennan fungerar, vi har då

$$P(F) = P(F|P_R)P(P_R) + P(F|P_B)P(P_B) + P(F|P_S)P(P_S) = 0.9\frac{1}{3} + 0.8\frac{1}{3} + 0.4\frac{1}{3} = 0.7.$$

- (c) Vi använder Bayes formel och beräknar

$$P(P_R|F) = \frac{P(F|P_R)P(P_R)}{P(F)} = \frac{0.9\frac{1}{3}}{0.7} = \frac{9}{21}$$

$$P(P_B|F) = \frac{P(F|P_B)P(P_B)}{P(F)} = \frac{0.8\frac{1}{3}}{0.7} = \frac{8}{21}$$

$$P(P_S|F) = \frac{P(F|P_S)P(P_S)}{P(F)} = \frac{0.4\frac{1}{3}}{0.7} = \frac{4}{21}$$

2. (a) Baserat på datan beräknar vi  $\hat{\mu} = \bar{x} = 493.3$  och  $\hat{\sigma}^2 = s^2 = \frac{1}{9} \sum_{i=1}^{10} (x_i - \bar{x})^2 = 72.9$ . Från formelsamlingen har vi att  $t_{0.95}(9) = 1.8331$ . Baserat på detta kan vi beräkna att övre begränsat konfidensintervall för  $\mu$ :  $I_\mu = (-\infty, a)$  där

$$a = \bar{x} + t_{0.95}(9) \frac{s}{\sqrt{10}} = 493.3 + 1.8331 \frac{8.5381}{\sqrt{10}} = 493.3 + 1.8331 \cdot 2.7 = 498.25.$$

Eftersom intervallet inte täcker 500 kan vi förkasta  $H_0$ .

- (b) Låt  $X$  beteckna vikten, vi vill välja  $d$  så att  $P(\hat{\mu} - d < X < \hat{\mu} + d) = 0.95$ . Låt  $Z$  beteckna en  $N(0, 1)$  variabel, vi har då

$$\begin{aligned} 0.95 &= P(\hat{\mu} - d < X < \hat{\mu} + d) = P(\hat{\mu} - d < \sigma Z + \hat{\mu} < \hat{\mu} + d) \\ &= P(-d < \hat{\sigma} Z < d) = P(-d/\hat{\sigma} < Z < d/\hat{\sigma}) \end{aligned}$$

Alltså har vi att  $d/\hat{\sigma}$  är lika med  $\lambda_{0.025} = 1.96$  och alltså  $d = 1.96\hat{\sigma} = 16.7348$ .

3. (a) Om  $X \sim \text{Po}(6000)$  har vi enligt CGS att  $X$  approximativt är  $N(6000, 6000)$ . Den sökta sannolikheten ges av

$$P(X < 5800) \approx \Phi\left(\frac{5800 - 6000}{\sqrt{6000}}\right) = \Phi(-2.582) = 1 - \Phi(2.582) = 0.0049.$$

Om man beräknar sannolikheten utan normalapproximation fås 0.0048.

- (b) Väntevärdet är  $6000/1000 = 6$  blodkroppar per  $\text{mm}^3$ . Låt  $Y$  beteckna  $\text{Po}(6)$  fördelningen, den sökta sannolikheten är

$$\begin{aligned} P(Y < 5) &= p_Y(0) + p_Y(1) + p_Y(2) + p_Y(3) + p_Y(4) \\ &= \frac{6^0}{0!}e^{-6} + \frac{6^1}{1!}e^{-6} + \frac{6^2}{2!}e^{-6} + \frac{6^3}{3!}e^{-6} + \frac{6^4}{4!}e^{-6} \\ &= e^{-6}(1 + 6 + 18 + 36 + 54) = 155e^{-6} = 0.2851. \end{aligned}$$

4. (a) Vi börjar med att beräkna de kvadratsummor och medelvärden som behövs:  $\bar{x} = 3.667$ ,  $\bar{y} = 75.6167$ ,

$$S_{xx} = \sum_{i=1}^6 (x_i - \bar{x})^2 = 16.0733, \quad S_{yy} = \sum_{i=1}^6 (y_i - \bar{y})^2 = 7802, \quad S_{xy} = \sum_{i=1}^6 (x_i - \bar{x})(y_i - \bar{y}) = 351.83.$$

Vi har nu  $\beta_1^* = S_{xy}/S_{xx} = 21.8905$  och  $\beta_0 = \bar{y} - \beta_1^* \bar{x} = -4.6485$ . Vi beräknar  $Q_0 = S_{yy} - S_{xy}^2/S_{xx} = 99.743$  och  $s^2 = Q_0/4 = 24.93$ . Kvantilen vi behöver är  $t_{0.025}(4) = 2.7764$  och vi har

$$I_{\beta_0} = [\beta_0^* - t_{0.025}(4)s\sqrt{1/6 + \bar{x}^2/S_{xx}}, \beta_0^* + t_{0.025}(4)s\sqrt{1/6 + \bar{x}^2/S_{xx}}] = [-18.5344, 9.2374]$$

$$I_{\beta_1} = [\beta_1^* - t_{0.025}(4)s/\sqrt{S_{xx}}, \beta_1^* + t_{0.025}(4)s/\sqrt{S_{xx}}] = [18.4323, 25.3487]$$

- (b) Det sökta konfidensintervallet är

$$[\beta_0^* + \beta_1^* 3 \pm t_{0.025}(4)s\sqrt{1/6 + (3 - \bar{x})^2/S_{xx}}] = [54.9114, 67.1346].$$

- (c) Det sökta prediktionsintervallet är

$$\left[ \beta_0^* + \beta_1^* 3 \pm t_{0.025}(4)s\sqrt{1 + 1/6 + (3 - \bar{x})^2/S_{xx}} \right] = [45.8714, 76.1746].$$

5. (a) Om studenterna klarar tentan oberoende av varandra då är antalet studenter som klarar tentan  $\text{Bin}(n_{\text{år}}, p_{\text{år}})$ , där  $n_{\text{år}}$  är antalet studenter som skriver tentan det året.  
(b) Vi har log-likelihood funktionen

$$\begin{aligned} L(x; \text{år } p) &= \log(p(x)) \log \left( \binom{n_{\text{år}}}{x} p_{\text{år}}^x (1 - p_{\text{år}})^{n_{\text{år}} - x} \right) \\ &= \log \left( \binom{n_{\text{år}}}{x} \right) + x \log(p_{\text{år}}) + (n_{\text{år}} - x) \log(1 - p_{\text{år}}) \end{aligned}$$

Vi deriverar med avseende på  $p_{\text{år}}$  och sätter derivatan lika med noll:

$$\begin{aligned} \frac{\partial L(x; p_{\text{år}})}{\partial p_{\text{år}}} &= \frac{x}{p_{\text{år}}} - \frac{n_{\text{år}} - x}{1 - p_{\text{år}}} = 0 \Rightarrow \\ (1 - p_{\text{år}})x - p_{\text{år}}(n_{\text{år}} - x) &= 0 \Rightarrow \\ x - p_{\text{år}}n_{\text{år}} &= 0 \Rightarrow p_{\text{år}} = \frac{x}{n_{\text{år}}} \end{aligned}$$

Alltså är ML-skattaren  $p_{\text{år}}^* = x/n_{\text{år}}$ .

- (c) Vi har

$$E(p_{\text{år}}^*) = \frac{E(X)}{n_{\text{år}}} = \frac{n_{\text{år}} p_{\text{år}}}{n_{\text{år}}} = p_{\text{år}}$$

alltså är skattaren väntevärdesriktig. Variansen är

$$V(p_{\text{år}}^*) = \frac{V(X)}{n_{\text{år}}^2} = \frac{n_{\text{år}} p_{\text{år}} (1 - p_{\text{år}})}{n_{\text{år}}^2} = \frac{p_{\text{år}} (1 - p_{\text{år}})}{n_{\text{år}}}$$

- (d) För 2017 har vi  $n_{2017} = 51$  och  $x_{2017} = 28$  och får alltså  $p_{2017}^* = 28/51 = 0.549$ . För 2018 har vi  $n_{2018} = 60$  och  $x_{2018} = 39$  och får alltså  $p_{2018}^* = 39/60 = 0.65$ .

- (e) Vi vill testa  $H_0 : p_{2017} = p_{2018}$  mot  $H_1 : p_{2017} \neq p_{2018}$ . Vi har  $n_{2017}p_{2017}^*(1 - p_{2017}^*) = 12.63 > 10$  och  $n_{2018}p_{2018}^*(1 - p_{2018}^*) = 13.65 > 10$ . Alltså är det OK att använda CGS för att approximera båda fördelningarna med normalfördelningar. Vi kan nu beräkna ett 95% konfidensintervall för  $p_{2017} - p_{2018}$  enligt

$$I_{p_{2017}-p_{2018}} = \left[ p_{2017}^* - p_{2018}^* \pm z_{0.025} \sqrt{\frac{p_{2017}^*(1 - p_{2017}^*)}{n_{2017}} + \frac{p_{2018}^*(1 - p_{2018}^*)}{n_{2018}}} \right]$$

$$= [-0.101 \pm 0.1823] = [-0.2832, 0.0813].$$

Eftersom konfidensintervallet täcker noll kan vi inte förkasta  $H_0$ . Alternativ kan vi använda en poolad variansskattning för testet. Skatta

$$p^* = \frac{28 + 39}{51 + 60} = 0.6036$$

och använd

$$V(p_{2017} - p_{2018}) = p^*(1 - p^*) \left( \frac{1}{51} + \frac{1}{60} \right) = 0.0087.$$

Vi beräknar teststorheten

$$T = \frac{p_{2017}^* - p_{2018}^*}{\sqrt{0.0087}} = -1.0839$$

Vi jämför denna med det kritiska värdet  $z_{0.025} = 1.96$  och eftersom  $|T| < 1.96$  kan vi inte förkasta  $H_0$ .

6. (a) Låt  $x_{Ai}$  beteckna mätningarna av Märke A  $x_{Bi}$  mätningarna hos Märke B. Vi beräknar först stickprovsmedelvärde och stickprovsvarians hos datan för respektive märke:

|         | Väntevärde | Varians |
|---------|------------|---------|
| Märke A | 10.9286    | 1.0557  |
| Märke B | 9.5833     | 2.5777  |

Vi antar att  $x_A \sim N(\mu_A, \sigma_A^2)$  och  $x_B \sim N(\mu_B, \sigma_B^2)$  och vill testa

$$H_0 : \sigma_B = \sigma_A$$

$$H_1 : \sigma_B \neq \sigma_A$$

på nivå  $\alpha = 0.05$ . Vi beräknar teststorheten  $T_{obs} = s_B^2/s_A^2 = 2.5777/1.0557 = 2.4416$ , som under  $H_0$  är  $F(5, 6)$ -fördelad. Enligt formelsamlingen är  $F_{\alpha/2}(5, 6) = 4.38$  och eftersom  $T_{obs} < F_{\alpha/2}(5, 6)$  kan vi inte förkasta  $H_0$ .

- (b) Vi har samma modell som i (a) men antar att  $\sigma_A = \sigma_B$  och vill nu testa

$$H_0 : \mu_A = \mu_B$$

$$H_1 : \mu_A \neq \mu_B$$

på nivå  $\alpha = 0.05$ . Den poolade variansskattningen ges av

$$s_p^2 = \frac{6s_A^2 + 5s_B^2}{11} = 1.7475$$

Vi beräknar teststorheten

$$T_{obs} = \frac{\bar{x}_A - \bar{x}_B}{s_p \sqrt{\frac{1}{7} + \frac{1}{6}}} = 1.8291$$

Under  $H_0$  är teststorheten  $t(11)$ -fördelad, och enligt formelsamlingen är  $t_{\alpha/2}(11) = 2.2010$ . Eftersom  $|T_{obs}| < t_{\alpha/2}(11)$  kan vi inte förkasta  $H_0$ .

- (c) För Wilcoxons ranksummetest testas om medianen hos de två fördelningarna är lika. Låt  $m = 6$  vara antalet mätningar för det mindre stickprovet och låt  $n = 7$  vara antalet mätningar för det större. Vi börjar med att ordna de  $m + n$  värdena i ökande storleksordning och ger varje observation en rank  $R_i$  från 1 till  $m + n$ . I fallet då flera mätningar har samma värde tilldelas dess medelvärde av motsvarande ranker. Vi bildar nu teststorheten  $W_m$  som summan av rankerna för det mindre stickprovet. Vi förkastar  $W_m$  om dess observerade värde ligger i det kritiska området för testet, som anges i tabell.
- (d) För datan här har vi

|         |     |     |     |     |     |      |      |      |      |      |      |      |      |
|---------|-----|-----|-----|-----|-----|------|------|------|------|------|------|------|------|
| Mätning | 8.0 | 8.4 | 9.1 | 9.2 | 9.6 | 10.0 | 10.3 | 10.4 | 11.0 | 11.2 | 12.2 | 12.2 | 12.4 |
| Märke   | B   | B   | B   | B   | A   | A    | A    | B    | A    | A    | A    | A    | B    |
| Rank    | 1   | 2   | 3   | 4   | 5   | 6    | 7    | 8    | 9    | 10   | 11.5 | 11.5 | 13   |

Summan av rankerna för Märke B är alltså  $1 + 2 + 3 + 4 + 8 + 13 = 31$ . Enligt tabellen är undre och övre kritiska värden för  $W_m$  28 och 56. Eftersom 31 ligger i detta intervall kan vi inte förkasta  $H_0$ .