UPPSALA UNIVERSITET
Matematiska institutionen
Sven Erick Alm, Uwe Menzel

Lösning till prov i Matematisk statistik
Sannolikhet och statistik, 1MS005
10–10–21

1. Låt $F :=$ "Finjusteras" och $K :=$ "Krånglar". Då gäller $P(F) = 0.8$, $P(K|F) = 0.05$ och $P(K|F^c) = 0.10$.

   (a) $P(K) = P(F) \cdot P(K|F) + P(F^c) \cdot P(K|F^c) = 0.8 \cdot 0.05 + 0.2 \cdot 0.10 = 0.06$.

   (b) $P(F|K) = \dfrac{P(F \cap K)}{P(K)} = \dfrac{P(F) \cdot P(K|F)}{P(K)} = \dfrac{0.8 \cdot 0.05}{0.06} = \dfrac{2}{3} = 0.667$.

2. Kalla slumpvariabeln X. Då vet vi att $P(X \geq 0.5) = 2/3$, dvs. $P(X < 0.5) = 1/3$, och att

$$f(x) = A \cdot x + B \quad \text{för } 0 \leq x \leq 1.$$

   (a) Att f är en täthetsfunktion ger att

$$1 = \int_{-\infty}^{\infty} f(x)\,dx = \int_0^1 (Ax + B)\,dx = \frac{A}{2} + B,$$

   så att $B = 1 - A/2$. Vidare gäller

$$\frac{1}{3} = P(X < 0.5) = \int_0^{0.5} (Ax + B)\,dx = \frac{A}{8} + \frac{B}{2} = \frac{1}{2} - \frac{A}{8},$$

   så att $A = 8 \cdot \left(\frac{1}{2} - \frac{1}{3}\right) = \frac{4}{3}$ och $B = 1 - \frac{2}{3} = \frac{1}{3}$.

   (b)

$$E(X) = \int_0^1 x \cdot \left(\frac{4x}{3} + \frac{1}{3}\right) dx = \frac{1}{3} \cdot \int_0^1 (4x^2 + x)\,dx = \frac{1}{3} \cdot \left(\frac{4}{3} + \frac{1}{2}\right) = \frac{1}{3} \cdot \frac{11}{6} = \frac{11}{18} = 0.611.$$

$$E(X^2) = \int_0^1 x^2 \cdot \left(\frac{4x}{3} + \frac{1}{3}\right) dx = \frac{1}{3} \cdot \int_0^1 (4x^3 + x^2)\,dx = \frac{1}{3} \cdot (1 + \frac{1}{3}) = \frac{4}{9},$$

$$\text{så att} \quad V(X) = \frac{4}{9} - \left(\frac{11}{18}\right)^2 = \frac{23}{324} = 0.0710.$$

3. Låt $X :=$ "responstiden" $\sim \mathrm{N}(8, 0.5^2)$.

   (a) $P(X > 9) = 1 - P(X \leq 9) = 1 - \Phi\left(\dfrac{9-8}{0.5}\right) = 1 - \Phi(2) = 1 - 0.9772 = 0.0228$.

   (b) $P(7.5 < X < 8.5) = \Phi\left(\dfrac{8.5-8}{0.5}\right) - \Phi\left(\dfrac{7.5-8}{0.5}\right) = \Phi(1) - \Phi(-1) = 2 \cdot \Phi(1) - 1 = 2 \cdot 0.8413 - 1 = 0.6826$.

   (c) $X \sim \mathrm{N}(8, \sigma^2)$. $0.01 \geq P(X > 9) = 1 - P(X \leq 9) = 1 - \Phi\left(\dfrac{9-8}{\sigma}\right) = 1 - \Phi\left(\dfrac{1}{\sigma}\right)$, vilket ger att $\dfrac{1}{\sigma} \geq \lambda_{0.01} = 2.3263$ och $\sigma \leq \dfrac{1}{2.3263} = 0.430$.

4. (a) $Y :=$ "# tröga reläer" $\sim \mathrm{Bin}(50, 0.01) \approx \mathrm{Po}(0.5)$ (eftersom $p = 0.01 < 0.1$).
   $P(Y > 2) = 1 - P(Y \leq 2) \approx$ (Po-tab.)$1 - 0.9856 = 0.0144 = 0.014$.
   (Exakt erhålls $P(Y > 2) = 0.0138$.)

   (b) $Z :=$ "# tröga reläer" $\sim \mathrm{Bin}(2000, 0.01) \approx \mathrm{N}(20, 19.8)$ (eftersom $19.8 > 5$).

$$P(Z \geq 25) = 1 - P(Z \leq 24) = 1 - P(Z < 24.5) \approx 1 - \Phi\left(\frac{24.5 - 20}{\sqrt{19.8}}\right) = 1 - \Phi(1.01)$$
$$= 1 - 0.8438 = 0.1562 = 0.156.$$

Var god vänd!

5. $x_1, \ldots, x_{30}$ är ett stickprov från X med $E(X) = \mu_A$ och $V(X) = \sigma_A^2$.
$y_1, \ldots, y_{40}$ är ett stickprov från Y med $E(Y) = \mu_B$ och $V(Y) = \sigma_B^2$.
Enligt Centrala gränsvärdessatsen gäller: $\bar{x} = \bar{x}_A$ är en observation av $\bar{X} \approx \mathrm{N}(\mu_A, \sigma_A^2/30)$ och $\bar{y} = \bar{x}_B$ är en observation av $\bar{Y} \approx \mathrm{N}(\mu_B, \sigma_B^2/40)$. Då är $\mu_A^* - \mu_B^* = \bar{x} - \bar{y}$ en observation av $\bar{X} - \bar{Y} \approx \mathrm{N}(\mu_A - \mu_B, \sigma_A^2/30 + \sigma_B^2/40)$.
Standardavvikelsen skattas med medelfelet $d = \sqrt{s_A^2/30 + s_B^2/40} = 0.2874$, vilket ger konfidensintervallet (med appr. konfidensgrad 95 %):

$$I_{\mu_A - \mu_B} = (\bar{x} - \bar{y} \pm \lambda_{0.025} \cdot d) = (8.2 - 7.8 \pm 1.96 \cdot 0.2874) = (0.4 \pm 0.56) = (-0.16, 0.96).$$

Eftersom $0 \in I_{\mu_A - \mu_B}$ kan man *inte*, på basis av dessa data, påstå att det finns en signifikant skillnad i förväntad responstid mellan företagens reläer (vid 5 % felrisk).

6. $x = 42$ är en observation av $X \sim \mathrm{Hyp}(N, 100, p_1) \approx \mathrm{Bin}(100, p_1)$ ($\frac{100}{N} \approx 0.001 < 0.1$), $y = 56$ är en observation av $Y \sim \mathrm{Hyp}(N, 200, p_2) \approx \mathrm{Bin}(200, p_2)$ ($\frac{200}{N} \approx 0.002 < 0.1$).
$p_1^* = \dfrac{x}{100} = 0.42$ är en obs. av $X100 \approx \mathrm{N}(p_1, \dfrac{p_1(1-p_1)}{100}$ $(100 \cdot p_1^*(1 - p_1^*) = 24.36 > 5)$,
$p_2^* = \dfrac{y}{200} = 0.28$ är en obs. av $Y200 \approx \mathrm{N}(p_2, \dfrac{p_2(1-p_2)}{200})$ $(200 \cdot p_2^*(1 - p_2^*) = 40.32 > 5)$,
$p_1^* - p_2^* = \dfrac{x}{100} - \dfrac{y}{200} = 0.14$ är en obs. av $\dfrac{X}{100} - \dfrac{Y}{200} \approx \mathrm{N}\left(p_1 - p_2, \dfrac{p_1(1-p_1)}{100} + \dfrac{p_2(1-p_2)}{200}\right)$.
Standardavvikelsen skattas med medelfelet $d = \sqrt{\dfrac{p_1^*(1-p_1^*)}{100} + \dfrac{p_2^*(1-p_2^*)}{200}} = 0.0587$, så att

$$I_{p_1 - p_2} = (p_1^* - p_2^* \pm \lambda_{0.025} \cdot d) = (0.14 \pm 1.96 \cdot 0.0587) = (0.14 \pm 0.115) = (0.035, 0.255).$$

Eftersom intervallet bara innehåller positiva värden kan man (vid 5 % felrisk) säga att kampanjen lyckats.
Anm. Man kan också göra en undre konfidensgräns för $p_1 - p_2$ *som blir* $0.14 - \lambda_{0.05} \cdot d = 0.043 > 0$, *med samma slutsats.*

7. (a) Stickprov i par!
Före: $x_1, \ldots, x_6$ oberoende observationer av $X_1, \ldots, X_6$, där $X_i \sim \mathrm{N}(\mu_i, \sigma_1^2)$.
Efter: $y_1, \ldots, y_6$ oberoende observationer av $Y_1, \ldots, Y_6$, där $Y_i \sim \mathrm{N}(\mu_i - \Delta, \sigma_2^2)$.
Bilda differenserna $z_i := x_i - y_i$. Då är $z_1, \ldots, z_6$ ett stickprov från $Z \sim \mathrm{N}(\Delta, \sigma^2)$.

(b) $\bar{z} = 2.483$, $s_z = 2.134$, $t_{0.025}(5) = 2.5706$ ger

$$I_\Delta = \left(\bar{z} \pm t_{0.025}(5) \cdot \frac{s_z}{\sqrt{6}}\right) = \left(2.483 \pm 2.5706 \cdot \frac{2.134}{\sqrt{6}}\right) = (2.483 \pm 2.239) = (0.24, 4.72).$$

Eftersom intervallet bara innehåller positiva värden kan man (vid 5 % felrisk) säga att det lönar sig att cykla (om man vill gå ner i vikt).

8. (a) $\sum_{i=1}^{10} x_i = 3.5$ ger $\bar{x} = 0.35$, $\sum_{i=1}^{10} x_i^2 = 1.234$ ger $S_{xx} = 0.009$,
$\sum_{i=1}^{10} y_i = 497.5$ ger $\bar{y} = 49.75$, $\sum_{i=1}^{10} y_i^2 = 24761.07$ ger $S_{yy} = 10.445$ och
$\sum_{i=1}^{10} x_i y_i = 174.428$ ger $S_{xy} = 0.303$.
Detta ger $\beta^* = S_{xy}/S_{xx} = 33.667$ och $\alpha^* = \bar{y} - \bar{x} \cdot \beta^* = 37.97$, så att den skattade regressionslinjen är $y = 37.97 + 33.667 \cdot x$.

(b) $R^2 = \dfrac{S_{xy}^2}{S_{xx} \cdot S_{yy}} = 0.9766$.

(c) σ^2 skattas med $s_r^2 = Q_0/(n-2) = 0.244/8 = 0.0305$, där $Q_0 = S_{yy} - \dfrac{S_{xy}^2}{S_{xx}} = 0.244$.

$$I_\beta = \left(\beta^* \pm t_{0.025}(8) \cdot \frac{s_r}{\sqrt{S_{xx}}}\right) = \left(33.667 \pm 2.3060 \cdot \sqrt{\frac{0.0305}{0.009}}\right) = (33.667 \pm 4.245)$$
$$= (29.42, 37.91).$$