

LÖSNINGAR

1. (a) Medelvärde $\bar{x} = 0.6376$, stickprovsstandardavvikelsen $s = 0.8638$ och $t_{0.005}(6) = 3.7074$, så ett 99% konfidensintervall ges av

$$I_\mu = (\bar{x} \pm s \cdot t_{0.005}(6)/\sqrt{7}) = (-0.57, 1.85).$$

- (b) Intervallet ovan innehåller 0.2 så misstanken kan ej förkastas, med felrisk på högst 1%.
2. (a) Sökt sannolikhet: $P(X \geq 1) = 1 - P(X = 0) = 1 - 0.55 = 0.45$.
- (b) Inför händelser $A = \{X = 2\}$, $B = \{X \geq 1\}$. Sökt sannolikhet:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(X = 2)}{P(X \geq 1)} = \frac{0.15}{0.45} = \frac{1}{3} \doteq 0.33.$$

- (c) Inför slumpvariabeln $Y =$ "Kostnad under en månads drift (kr)". Man finner dess fördelning i form av sannolikhetsfunktionen

y	10	10	25	40
$p_Y(y)$	0.55	0.25	0.15	0.05

Sökt är förväntad kostnad:

$$E[Y] = \sum_k k p_Y(k) = 10 \cdot 0.55 + 10 \cdot 0.25 + 25 \cdot 0.15 + 40 \cdot 0.05 = 13.75,$$

dvs. 13 750 kr.

3. Vi har att

$$X_1 + \dots + X_5 \sim N(5, 10), \quad Y_1 + \dots + Y_4 \sim N(-8, 4)$$

(linjärkombinationer av oberoende normalfördelade variabler blir normalfördelade).

Det följer att

$$\tilde{Z} = X_1 + \dots + X_5 - (Y_1 + \dots + Y_4) \sim N(13, 14)$$

(samma motivering som ovan), och därmed att

$$Z = \frac{\tilde{Z} - 13}{\sqrt{14}} \sim N(0, 1).$$

Sökt sannolikhet:

$$P(\tilde{Z} > 0) = P(Z > -13/\sqrt{14}) = \Phi(3.47) \doteq 0.99974.$$

4. (a) Rätt svar: A-(iii); B-(iv); C-(i); D-(ii). Resonera i termer av väntevärden, likformig fördelning, skev fördelning.

- (b) Vi har att $X \sim N(10, \sigma^2)$ och därmed $Z = (X - 10)/\sigma \sim N(0, 1)$.

Olika sätt att resonera:

Alt. 1. För $X \sim N(\mu, \sigma^2)$ gäller $P(\mu - 2\sigma < X < \mu + 2\sigma) \approx 0.95$, så $4\sigma \approx 14 - 5 = 9$, $\sigma \approx 2.25$.

Alt. 2. Lådans kanter är 0.25- resp 0.75-kvantil. Vi finner från tabell över $Z \sim N(0, 1)$ att $\lambda_{0.25} = 0.6745$, alltså

$$0.25 = P(Z > \lambda_{0.25}) = P(X > 10 + \lambda_{0.25}\sigma).$$

Lådans lock 12 ger $\sigma \approx (12 - 10)/\lambda_{0.25} \doteq 2.97$. För att använda lådans botten betraktas

$$0.25 = P(Z < -\lambda_{0.25}) = P(X < 10 - \lambda_{0.25}\sigma)$$

Lådans botten 8 ger $\sigma \approx (10 - 8)/\lambda_{0.25} \doteq 2.97$.

(I simuleringen sattes $\sigma = 3$.)

5. (a) Energiförbrukningen kan antas bero på utomhustemperaturen, vilken normalt uppvisar starkt beroende mellan olika dagar. Oberoendeantagande är kanske därför mindre lämpligt.
- (b) Låt X_i = "Energiförbrukningen under dag i ". Givet är att $E[X_i] = 68$ och $\sqrt{V[X_i]} = 12.5$. December, januari och februari har sammanlagt 90 dagar ett vanligt år, så den sammanlagda energiförbrukningen är

$$Y = X_1 + X_2 + \cdots + X_{89} + X_{90}$$

Antag, något orealistiskt, att förbrukningen de olika dagarna är oberoende. Då ger centrala gränsvärdesatsen att $Y \approx N(\mu_Y, \sigma_Y^2)$ med

$$\mu_Y = 90 \cdot 68 = 6120, \quad \sigma_Y^2 = 90 \cdot 12.5^2 = 14062.5$$

så

$$P(Y \leq 6000) = [\text{Standardisera till } N(0, 1)] \doteq 0.1562.$$

6. Vi vill undersöka om *andelen* kuber med en tryckhållfasthet under 30 MPa (kalla andelen p) överstiger 5 %. Antalet kuber med för låg tryckhållfasthet är $\text{Bin}(30, p)$, så en punktskattning för andelen blir $p^* = x/n = 8/30$. Eftersom $np^*(1 - p^*) \approx 5.9 > 5$ är tumregeln för normalapproximation uppfylld och vi får konfidensintervallet

$$I_p = (p^* \pm \lambda_{\alpha/2} \sqrt{p^*(1 - p^*)/30})$$

som för valet $\alpha = 0.05$ blir $(0.11, 0.42)$. Eftersom hela konfidensintervallet ligger ovanför 0.05 dras slutsatsen att den karakteristiska tryckhållfastheten nog inte är 30 MPa.

Med konfidensgraden 99 % blir $I_p = (0.0584, 0.475)$. Alternativt kan man även konstruera ut ett konfidensintervall för p_2 =andelen med en tryckhållfasthet *över* 30 MPa. Vid 95 % konfidensgrad blir $I_{p_2} = (0.575, 0.891)$. Eftersom intervallet ligger under 0.95 tror vi inte att andelen block med tryckhållfasthet över 30 MPa är 0.95.

Observera att det faktum att man vill undersöka om andelen är 0.05 inte har något att göra med vilken konfidensgrad man bör välja! Konfidensgraden säger något om felrisken α – risken att säga att andelen inte är 0.05 om den i själva verket är just 0.05. Felrisken α behöver inte väljas till samma värde som p .

7. (a) Skattningarna ges av

$$\beta^* = \frac{S_{xy}}{S_{xx}} = -\frac{5630.652}{22073176} \doteq -2.55 \cdot 10^{-4}, \quad \alpha^* = \bar{y} - \beta^* \bar{x} \doteq 0.630.$$

(b) Förväntad kvicksilverhalt (ppm) vid elevation $x = 500$ (ft) ges av

$$\alpha^* + \beta^* \cdot 500 = 0.502.$$

(c) Man finner $s_r^2 = (S_{yy} - S_{xy}^2/S_{xx})/(120 - 2) = 0.0998$. Eftersom $\beta^* \sim N(\beta, \sigma^2/S_{xx})$ följer ett 95% konfidensintervall som

$$[\beta^* \pm t_{0.025}(118)s_r\sqrt{1/S_{xx}}] \doteq [-4.9 \cdot 10^{-4}, -2.06 \cdot 10^{-5}]$$

För beräkning av kvantilen kan man använda $t_{0.025}(100)$ eller $t_{0.025}(120)$ eller $\lambda_{0.025}$ (för stort antal frihetsgrader kan t -fördelningen approximeras med $N(0, 1)$). Konfidensintervallet I_β blir i dessa fall

$$\begin{array}{ll} t_{0.025}(100) & (-3.89 \cdot 10^{-4}, -1.22 \cdot 10^{-4}) \\ t_{0.025}(120) & (-3.88 \cdot 10^{-4}, -1.22 \cdot 10^{-4}) \\ \lambda_{0.025} & (-3.87 \cdot 10^{-4}, -1.23 \cdot 10^{-4}) \end{array}$$

Intervallet innehåller inte 0, men å andra sidan är förklaringsgraden låg ($R^2 \approx 0.1$). Vi kan förkasta eventuella hypoteser om noll lutning på signifikansnivån $\alpha = 0.05$. Det finns en svag negativ lutning vilket indikerar att kvicksilverhalten minskar med ökad elevation (ett resultat som stämmer överens med naturvetenskapen: tungmetaller färdas nedåt med vattenströmmar).

8. Låt Y beteckna det totala antalet klave.

Lösning 1: Ett mynt visar krona efter båda omgångarna med sannolikhet $\frac{1}{2} \cdot \frac{1}{2}$, och då mynten är oberoende betyder detta att $Y \sim \text{Bin}(10, \frac{3}{4})$. Därför är $E(Y) = 10 \cdot \frac{3}{4} = 7.5$.

Lösning 2: Låt X_1 vara antal klave efter första kastet, och låt X_2 vara antal nya klave efter den andra kastomgången. Då är $Y = X_1 + X_2$ och $X_1 \sim \text{Bin}(10, \frac{1}{2})$. För varje utfall av X_1 är $X_2 \sim \text{Bin}(10 - X_1, \frac{1}{2})$ med väntevärde $\frac{1}{2}(10 - X_1)$. Genom att summera över alla möjliga utfall för X_1 får vi att

$$E(Y) = E(X_1) + E(X_2) = 5 + E\left(\frac{1}{2}(10 - X_1)\right) = 5 + 2.5 = 7.5.$$

Obs! I lösning 2 är det inte korrekt att säga att $X_2 \sim \text{Bin}(5, \frac{1}{2})$, dvs man kan inte byta ut slumpvariabeln $10 - X_1$ mot dess väntevärde 5. Att man får rätt svar är ett sammanträffande. Att denna lösningsmetod är felaktig kan man se genom att betrakta fallet med 11 mynt, då man skulle fått $X_2 \sim \text{Bin}(5.5, \frac{1}{2})$, vilket är nonsens.

Kommentar: Väntevärdet av en diskret slumpvariabel behöver inte vara ett heltal.