1. (a) Vi vill räkna ut sannolikheten att det inte regnar på lördagen givet att det är sol på torsdagen. Vi vet alltså att Markovkedjan befinner sig i tillståndet "soligt" och vill räkna ut sannoliketen för de olika tillstånden två dagar senare. Låt $p^{(2)}$ vara en radvektor som innehåller sannolikheterna för tillstånden "soligt", "mulet" resp "regnigt" för vädret på lördagen. Enligt Chapman-Kolmogorovs sats ges $p^{(2)}$ av

$$p^{(2)} = p^{(0)}P^2 = (1\ 0\ 0)\begin{pmatrix} 0.7 & 0.3 & 0 \\ 0.2 & 0.4 & 0.4 \\ 0 & 0.5 & 0.5 \end{pmatrix}^2 = (0.55\ 0.33\ 0.12).$$

Sannolikheten att åka till stranden ges av $0.55 + 0.33 = 0.88$.

Alternativ lösning: P("Stranden på lördag")=P("ej regn på lördag")=1-P("regn på lördag"). För att det skall regna på lördag måste Markovkedjan göra övergångarna "Soligt" till "Mulet" följt av "Mulet" till "Regnigt". Detta ger att P("regn på lördag")=$P_{12}P_{23}$=$0.3 \cdot 0.4 = 0.12$ ger P("Stranden på lördag")=1-0.12=0.88.

(b) Låt radvektorn $\pi = (\pi_1\ \pi_2\ \pi_3)$ beteckna den stationära fördelningen för sommarvädret. Vi har då att π fås som lösningen till ekvationsystemet $\pi P = \pi$. Sätter vi värdena från P fås

$$\begin{aligned} -0.3\pi_1 + 0.2\pi_2 &= 0 \\ 0.3\pi_1 - 0.6\pi_2 + 0.5\pi_3 &= 0 \\ 0.4\pi_2 - 0.5\pi_3 &= 0 \end{aligned}$$

Från den första ekvationen får vi $\pi_1 = (2/3)\pi_2$ och från den sista fås $\pi_3 = (4/5)\pi_2$. Vi använder nu detta samt att $\pi_1 + \pi_2 + \pi_3 = 1$ detta ger $(2/3 + 1 + 4/5)\pi_2 = 1$ vilket ger $\pi_2 = 15/37$. Detta ger i sin tur att $\pi_1 = (2/3)(15/37) = 10/37$ och $\pi_3 = (4/5)(15/37) = 12/37$. Vi får alltså att den stationära fördelningen ges av $\pi = (10/37\ 15/37\ 12/37)$. Så i genomsnitt är det soligt 10/37 av tiden mulet 15/37 av tiden och regnigt 12/37 av tiden.

2. X = antal felaktiga i stickprovet $\in Bin(1000, p)$, $x = 25$, $p^* = x/n = 0.025$. Vi vill testa om andelen felaktiga minskat. Vi kan då använda följande hypoteser H_0: $p = 0.03$, H_1: $p < 0.03$.

Om H_0 sann så $1000 \cdot 0.03 \cdot 0.97 = 29.1 > 10$ och $p^* \underset{\sim}{\in} N\left(0.03, \sqrt{\frac{0.03 \cdot 0.97}{1000}}\right) = N(0.03,\ 0.0054)$

Eftersom $\frac{p^* - 0.03}{0.0054} = \frac{0.025 - 0.03}{0.0054} = -0.927 \not< -\lambda_\alpha = -\lambda_{0.05} = -1.64$ kan H_0 inte förkastas. Andelen felaktiga är inte signifikant mindre än tidigare.

Alternativ lösning: Gör ensidigt uppåt begränsat konfidensintervall för p. Vi kollar först att $np^*(1-p^*) > 10$. Vi har $p^* = 25/1000 = 0.025$ ger $np^*(1-p^*) = 1000*0.025*0.975 = 24.375 > 10$ så normalapproximation OK. Bilda approximativt ensidigt konfidens interval enligt

$$I_p = [0,\ p^* + \lambda_\alpha \cdot d(p^*)] = [0,\ 0.0025 + 1.6449\sqrt{0.025 * 0.975/1000}] = [0,\ 0.0331]$$

Eftersom 0.03 ligger i konfidensintervallet kan vi inte förkasta att andelen felaktiga är samma som innan.

Alternativ lösning2: Använd direktmetoden efter normalapproximation. Vi kollar först att $np^*(1-p^*) > 10$. Vi har $p^* = 25/1000 = 0.025$ ger $np^*(1-p^*) = 1000 * 0.025 * 0.975 = 24.375 > 10$ så normalapproximation OK. Vi räknar approximativt ut sannolikheten att få 25 eller färre felaktiga givet att andelen felaktiga är 3%. Utan halvkorrektion

$$P(X \leq 25, p = 0.03) \approx \Phi((25-30)/\sqrt{0.03 \cdot 0.97 \cdot 1000}) = \Phi(-0.9269) = 1-\Phi(0.9269) \approx 0.1762 > 0.05.$$

Med halvkorrektion

$$P(X \leq 25, p = 0.03) \approx \Phi((25.5-30)/\sqrt{0.03 \cdot 0.97 \cdot 1000}) = \Phi(-0.8342) = 1-\Phi(0.8342) \approx 0.2033 > 0.05.$$

Vi kan inte förkasta att andelen felaktiga är samma som innan, (vare sig med eller utan halvkorrektion).

3. Eftersom skillnaden i järnhalt är stor mellan de olika groparna och värdena på A och C-nivå ser ut att samvariera ansätter vi modellen stickprov i par:
x_k – Fe-halt på A-nivå i grop $k \in N(\mu_k, \sigma_1)$,
y_k – Fe-halt på C-nivå i grop $k \in N(\mu_k + \Delta, \sigma_2)$,
och de intressanta hypoteserna är H_0: Δ=0, H_1: $\Delta \neq 0$.
Bilda differenser $z_k = y_k - x_k$:

$$2.81, 4.35, 2.83, 2.32, 0.41, -40.35, 4.62, 3.72, -0.63, -0.57$$

Enligt modellen är $z_1, \dots, z_{10} \in N(\Delta, \sqrt{\sigma_1^2 + \sigma_2^2}) = N(\Delta, \sigma)$
$\bar{z}$=1.951 och $s_z = \sqrt{\frac{1}{10-1}\sum_{k=1}^{10}(z_k - \bar{z})^2}$=2.064.
Testkvantitet $t = \frac{\bar{z}-0}{\frac{s_z}{\sqrt{10}}} = \sqrt{10}\frac{1.951}{2.064}$=2.989.
Eftersom $|t| > t_{0.025}(9) = 2.26$ förkastas H_0 på nivå 0.05. Det finns alltså en signifikant skillnad i genomsnittlig järnhalt mellan olika nivåer.

Alternativt kan ett 95 % konfidensintervall för Δ göras:

$$\begin{aligned} I_\Delta &= (\bar{z} \pm t_{0.025}(9)\frac{s_z}{\sqrt{10}}) = \\ &= (1.951 \pm 2.26\frac{2.064}{\sqrt{10}}) = (1.951 \pm 1.475) = \\ &= (0.48, 3.43). \end{aligned}$$

Eftersom intervallet ej täcker över 0 är slutsatserna de samma som ovan vid hypotestestet.

4. Låt T vara en stokastisk variabel som anger typen av lampa. T kan anta värdena L, N och E. Låt X vara en stokastisk variabel som beskriver en slumpmässigt vald lampas lystid.

   (a) Enligt satsen om total sannolikhet ges tätheten för X av

   $$f_X(x) = f_{X|T=L}(x)P(T=L) + f_{X|T=N}(x)P(T=N) + f_{X|T=E}(x)P(T=E).$$

   Nu använder vi att de betingade fördelningarna för X är exponentialfördelningar med väntevärden 400, 700 och 1000 för L, N respektive E-lampor. Vi vet också att andelarna är 0.4, 0.35 och 0.25 för L, N respektive E-lampor. Detta ger att

   $$f_X(x) = 0.4\frac{1}{400}e^{-\frac{x}{400}} + 0.35\frac{1}{700}e^{-\frac{x}{700}} + 0.25\frac{1}{1000}e^{-\frac{x}{1000}}.$$

   (b) Vill beräkna $P(T = y|X = 600)$ för $y = L$, $y = N$ och $y = E$. Bayes sats ger

   $$P(T = y|X = x) = \frac{f_{X|T=y}(x)P(T=y)}{f_{X|T=L}(x)P(T=L) + f_{X|T=N}(x)P(T=N) + f_{X|T=E}(x)P(T=E)}$$

   Sätter vi in siffrorna från uppgiften fås

   $$P(T = L|X = 600) \approx 0.39,\ P(T = N|X = 600) \approx 0.37,\ P(T = E|X = 600) \approx 0.24.$$

   (c) Först behöver vi räkna ut $P(T = y|X \geq 600)$ för $y = L$, $y = N$ och $y = E$. Precis som i (b) kan vi använda Bayes-sats vilket ger

$$P(T = y|X \geq x) = \frac{P(X \geq x|T = y)P(T = y)}{P(X \geq x|T = L)P(T = L) + P(X \geq x|T = N)P(T = N) + P(X \geq x|T = E)P(T = E)}.$$

För beräkna detta uttryck behöver vi först beräkna de ingående sannolikheterna

$$P(X \geq x | T = L) = \int_x^\infty \frac{1}{400} e^{-\frac{y}{400}} \mathrm{d}y = e^{-\frac{x}{400}}$$

På samma sätt fås

$$P(X \geq x | T = N) = e^{-\frac{x}{700}}, \; P(X \geq x | T = E) = e^{-\frac{x}{1000}}.$$

Sätter vi nu in dessa uttryck med $x = 600$ i utrycket för $P(T = y | X \geq 600)$ fås

$$P(T = L | X \geq 600) \approx 0.24, \; P(T = N | X \geq 600) \approx 0.40, \; P(T = E | X \geq 600) \approx 0.36.$$

Vi ser att den största betingade sannolikheten ges av $P(T = N | X \geq 600)$. Vi utgår alltså då från att $T = N$. Låt X_N vara en exponentialfördelad stokastisk variabel med väntevärde 700. Vi vill beräkna sannolikheten

$$\begin{aligned} P(X_N \geq 600 + 500 | X_N \geq 600) &= \frac{P(X_N \geq 600 + 500 \cap X_n \geq 600)}{P(X_N \geq 600)} = \frac{P(X_N \geq 1100)}{P(X_N \geq 600)} \\ &= \frac{e^{-1100/700}}{e^{-600/700}} = e^{-500/700} \approx 0.49 \end{aligned}$$

Alternativt kunde vi utgått från att exponentialfördelingen är så kallat minneslös och direkt fått $P(X_N \geq 600 + 500 | X_N \geq 600) = P(X_N \geq 500) = e^{-500/700} \approx 0.49$.

5. Vi har efter vår transformation överfört problemet på en standard linjär regressionsmodell. Så vi behöver beräkna de olika summorna som behövs för att beräkna skattningarna. Först har vi att $\bar{x} = \sum_{i=1}^{20} x_i/20, \; \bar{z} = \sum_{i=1}^{20} z_i/20$. Sedan kan vi beräkna

$$\begin{aligned} S_{xz} &= \sum_{i=1}^{20}(x_i - \bar{x})(z_i - \bar{z}) = \left(\sum_{i=1}^{20} x_i z_i\right) - \left(\sum_{i=1}^{20} x_i\right)\bar{z} - \left(\sum_{i=1}^{20} z_i\right)\bar{x} + 20\bar{x}\bar{z} \\ &= \left(\sum_{i=1}^{20} x_i z_i\right) - \left(\sum_{i=1}^{20} x_i\right)\left(\sum_{i=1}^{20} z_i\right)/20 \end{aligned}$$

På samma sätt fås

$$S_{xx} = \left(\sum_{i=1}^{20} x_i^2\right) - \left(\sum_{i=1}^{20} x_i\right)^2/20$$

och

$$S_{zz} = \left(\sum_{i=1}^{20} z_i^2\right) - \left(\sum_{i=1}^{20} z_i\right)^2/20$$

Sätter vi nu in siffrorna från uppgiften fås

$$\bar{x} \approx 11.397, \; \bar{z} \approx 0.6275, \; S_{xx} \approx 134.08, \; S_{xz} \approx 62.698, \; S_{zz} \approx 29.695.$$

(a) Skattingarna av α, β och σ fås enligt formelsamlingen som

$$\alpha^* = \bar{z} - \bar{x}S_{xz}/S_{xz} \approx -4.7006, \; \beta^* = S_{xz}/S_{xx} \approx 0.4676, \; \sigma^* = s = \sqrt{(S_{zz} - S_{xz}^2/S_{xx})/(n-2)} \approx 0.1455.$$

(b) Vi vill nu göra att 99% intervall för uppmätt transformerad effekt (Z) då $x = 11m/s$, dvs vi vill göra ett prediktionsintervall för Z. Enligt formelsamling gäller

$$I_{Z(x_0)} = \left[\alpha^* + x_0\beta^* \pm t_{0.005}(n-2)s\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}\right].$$

Sätter vi in siffror fås

$$I_{Z(11)} = \left[-4.7006 + 11 \cdot 0.4676 \pm 0.1455\sqrt{1 + \frac{1}{20} + \frac{0.1560}{134.08}}\right] = [0.0134 \; 0.8722].$$

(c) För att få ett intervall för uppmätt effekt Y utnyttjar vi att Z är en monotont växande transformation av Y och därmed är Y en monotont växande transformation av Z. Vi kan därmed få att prediktionsintervall för Y genom att transformera gränserna i intervall för Z. För att göra detta måste vi först bestämma hur Y beror på Z. Vi har att

$$\begin{aligned}
\ln(Y/(1-Y)) &= Z \\
Y/(1-Y) &= e^Z \\
Y &= e^Z(1-Y) \\
Y(1+e^Z) &= e^Z \\
Y &= \frac{e^Z}{1+e^Z}
\end{aligned}$$

Applicerar vi det här sambandet på Z:s intervall fås

$$I_{Y(11)} = [0.5034\ 0.7052],$$

dvs effekten blir mellan 503 W och 705 W med 99% sannolikhet.

6. Antag att $x_1,\ x_2$ och x_3 är oberoende observationer av en stokastisk variabel med täthet

$$f_X(x) = \begin{cases} \frac{1}{9!\theta^{10}} x^9 e^{-x/\theta} & x \geq 0 \\ 0 & x < 0. \end{cases}$$

(a) Likelihood funktionen ges nu av

$$L(\theta, x_1, x_2, x_3) = f_X(x_1) f_X(x_2) f_X(x_3).$$

För att hitta ML-skattingen av θ skall vi maximera L map på θ. Detta görs lättast genom att vi först bildar log-likilihood funktionen $l(\theta, x_1, x_2, x_3) = \ln(L(\theta, x_1, x_2, x_3))$. Sätter in uttrycket för f_X fås (givet att $x_1,\ x_2$ och $x_3 > 0$)

$$l(\theta, x_1, x_2, x_3) = 3\ln(9!) - 30\ln(\theta) + 9\ln(x_1 x_2 x_3) - \frac{x_1 + x_2 + x_3}{\theta}$$

För att maximera deriverar vi map θ då fås

$$\frac{\partial}{\partial\theta} l(\theta, x_1, x_2, x_3) = -\frac{30}{\theta} + \frac{x_1 + x_2 + x_3}{\theta^2}.$$

Vi sätter uttrycket lika med noll och löser för θ då fås

$$\theta^*_{ML} = \frac{x_1 + x_2 + x_3}{30} = (11.045 + 10.474 + 13.593)/30 \approx 1.1704.$$

(b) Vi vill nu räkna ut medelfelet för θ^*_{ML}, dvs standardavvikelsen med eventuella parametrar ersatta av sina skattningar. Tittar vi närmare på X:s täthetsfunktion ser vi att X kommer från en gamma-fördelning med parametrarna $p = 10$ och $\lambda = 1/\theta$. Eftersom $\theta^*_{ML} = (x_1 + x_2 + x_3)/30$ fås

$$D(\theta^*_{ML}) = \sqrt{V(\theta^*_{ML})} = 1/30\sqrt{V(X_1) + V(X_2) + V(X_3)} = \sqrt{3V(X)}/30 = \sqrt{3}D(X)/30.$$

Vi uttnyttjar nu att X är gamma-fördelad. Formelsamlingen ger $V(X) = p/\lambda^2 = 10\theta^2$ vilket ger $D(X) = \sqrt{10}\theta$. Sätter vi in detta i uttrycket för $D(\theta^*_{ML})$ och ersätter θ med sin skattning θ^*_{ML} fås medelfelet som

$$D(\theta^*_{ML}) = 1.1704/\sqrt{30} \approx 0.21368.$$

(c) En gamma-fördelning med parameter p=10 kan ses som en summa av tio exponentialfördelade oberoende variabler. Eftersom X_1, X_2 och X_3 är oberoende och likafördelade gamma variabler med p=10, $\lambda = 1/\theta$ fås att $X_1 + X_2 + X_3$ är gamma-fördelad med p=30 och $\lambda = 1/\theta$. Vi kan således se θ^*_{ML} kan ses som en omskalning av en fördelning som kan ses som en summa av 30 oberoende likafördelade exponential variabler. Så vi kan enligt CGS anta att θ^*_{ML} approximativt normalfördelad med väntevärde $p\lambda/30 = 30\theta/30 = \theta$ och standardavvikelse $\sqrt{p}\theta/30 = \theta/\sqrt{30}$. Vi vill testa $H_0 : \theta = 1$ mot $H_1 : \theta > 1$ på nivån 5%. Vi gör nu ett nedåt begränsat approximativt 95% konfidensintervall för θ enligt

$$I_\theta = [\theta^*_{ML} - \lambda_\alpha d(\theta^*_{ML}), \infty] = [0.81893,\ \infty].$$

Eftersom vårt 95% konfidensintervall täcker över punkten $\theta = 1$ kan vi inte förkasta H_0.
