(Det tas forbehold om feil i løsningsforslaget.)

Oppgave 1

a) $\mathrm{E}(X) = \sum_{x=0}^{4} x \cdot P(X = x) = 0 \cdot 0.7 + 1 \cdot 0.15 + \cdots + 4 \cdot 0.02 = 0.54$

$\mathrm{E}(X^2) = \sum_{x=0}^{4} x^2 \cdot P(X = x) = 0^2 \cdot 0.7 + 1^2 \cdot 0.15 + \cdots + 4^2 \cdot 0.02 = 1.24$

$\mathrm{Var}(X) = \mathrm{E}(X^2) - \{\mathrm{E}(X)\}^2 = 1.24 - 0.54^2 = 0.9484.$

K har fordelingen:

x	0	1	2	3	4
k	0	80	90	100	110
$P(K = k)$	0.7	0.15	0.08	0.05	0.02

Da får vi:

$\mathrm{E}(K) = \sum_{k} k \cdot P(K = k) = 0 \cdot 0.7 + 80 \cdot 0.15 + \cdots + 110 \cdot 0.02 = 26.4$

$\mathrm{E}(K^2) = \sum_{k} k^2 \cdot P(K = k) = 0^2 \cdot 0.7 + 80^2 \cdot 0.15 + \cdots + 110^2 \cdot 0.02 = 2350$

$\mathrm{Var}(K) = \mathrm{E}(K^2) - \{\mathrm{E}(K)\}^2 = 2350 - 26.4^2 \approx 1653.$

b) La $Y = X_1 + X_2 + X_3 + X_4 + X_5$ være antall produkter med feil i løpet av uke med fem dager.

$$\begin{aligned}
&P(\text{minst ett feilprodukt en uke}) \\
&= 1 - P(\text{ingen feilprodukt en uke}) \\
&= 1 - P(\text{ingen feilprodukt dag 1} \cap \cdots \cap \text{ingen feilprodukt dag 5}) \\
&= 1 - P(X_1 = 0 \cap X_2 = 0 \cap X_3 = 0 \cap X_4 = 0 \cap X_5 = 0) \\
&= 1 - P(X_1 = 0)P(X_2 = 0)P(X_3 = 0)P(X_4 = 0)P(X_5 = 0) = 1 - 0.7^5 = 0.83
\end{aligned}$$

(Her er komplementsetningen benyttet og at antall feilprodukter ulike dager er antatt å være uavhengige begivenheter.)

$P(\text{ett feilprodukt dag 1 og ingen de andre fire dagene en uke}) = 0.15 \cdot (0.7^4) = 0.0360;$
$P(\text{akkurat ett feilprodukt en uke}) = 5 \cdot 0.15 \cdot (0.7^4) = 0.18.$

c) La K_i være kostnaden som følge av produkter med feil på dag $i = 1, 2, \ldots, n = 252$. Vi har da at alle K_i'ene er uavhengige og har samme fordeling som K i punkt a). Videre har vi at kostnaden i løpet av et helt år, som følge av produkter med feil, kan skrives $Y = K_1 + K_2 + \cdots + K_n$. Da blir forventet kostnad i løpet av et helt år:

$\mathrm{E}(Y) = \mathrm{E}(K_1 + K_2 + \cdots + K_n) = n\mathrm{E}(K) = 252 \cdot 26.4 = 6652.8.$

$\mathrm{Var}(Y) = \mathrm{Var}(K_1 + K_2 + \cdots + K_n) = n\mathrm{Var}(K) = 252 \cdot 1653 = 416556$, og siden Y er en sum av u.i.f. tilfeldige variable, har vi at Y er tilnærmet normalfordelt; $N(6652.8, \sqrt{416556})$.

Da får vi: $P(Y < 8000) = P\Big(\frac{Y - \mathrm{E}(Y)}{\mathrm{SD}(Y)} < \frac{8000 - 6652.8}{\sqrt{416556}}\Big) \approx P(Z < 2.09) = 0.9817$, der $Z \sim N(0,1)$.

Oppgave 2

a) $\mathrm{E}(X) = np = 61 \cdot 0.6 = 36.6$; $\mathrm{SD}(X) = \sqrt{np(1-p)} = \sqrt{61 \cdot 0.6 \cdot 0.4} = 3.83$.

$P(X = 0) = \binom{61}{0} 0.4^{61} = 0.4^{61} = 5.3 \cdot 10^{-25} \approx 0$

$P(X \geq 30) = 1 - P(X \leq 29) \approx 1 - P(Z \leq \frac{29 + 0.5 - 36.6}{3.83})$
$= 1 - P(Z \leq -1.85) = 1 - 0.0322 = 0.9678$. (Her er $Z \sim N(0,1)$.)

b) Vi har $\mathrm{E}(I_j) = 0(1-p) + 1p = p$, og
$\mathrm{E}(X) = \mathrm{E}(I_1 + I_2 + \cdots + I_{61}) = \mathrm{E}(I_1) + \mathrm{E}(I_2) + \cdots + \mathrm{E}(I_{61}) = 61p = np$.

Vi har videre at $\mathrm{E}(I_j^2) = 0^2(1-p) + 1^2 p = p$, og derfor at $\mathrm{Var}(I_j) = p - p^2 = p(1-p)$.
I_j'ene er uavhengige som følge av at X er antatt binomisk fordelt. Og da får vi:
$\mathrm{Var}(X) = \mathrm{Var}(I_1 + I_2 + \cdots + I_{61}) = \mathrm{Var}(I_1) + \mathrm{Var}(I_2) + \cdots + \mathrm{Var}(I_{61}) = 61p(1-p) = np(1-p)$.

c) Forventning under H_0: $np = 61 \cdot 0.6 = 36.6$.
Test: forkast H_0 dersom $X > 41$. Signifikansnivå $= P(X > 41 | p = 0.6)$, og

$$P(X > 41 | p = 0.6) = 1 - P(X \leq 41 | p = 0.6) \approx 1 - p(Z \leq \frac{41 + 0.5 - 36.6}{\sqrt{61 \cdot 0.6(1 - 0.6)}})$$
$$= 1 - P(Z \leq 1.28) = 1 - 0.8997 = 0.1003 = \text{signifikansnivået til testen.}$$

Signifikansnivået er sannsynligheten for å konkludere med H_1 når i virkeligheten H_0 er riktig. Dvs.: sannsynligheten for å hevde at det har vært klimaendring i form av økt antall nedbørsdager, når det i virkeligheten ikke er noen endring.

d) Teststørrelse: $\frac{\hat{p} - 0.6}{\sqrt{0.6(1-0.6)/61}}$, Nullfordeling: tilnærmet $N(0,1)$.

Test med signifikansnivå 5%: forkast H_0 dersom $\frac{\hat{p} - 0.6}{\sqrt{0.6(1-0.6)/61}} \geq z_{0.05} = 1.645$.

(Alternativt: forkast H_0 dersom
$X = n\hat{p} \geq n \cdot 0.6 + z_{0.05}\sqrt{n \cdot 0.6(1-0.6)} = 61 \cdot 0.6 + 1.645\sqrt{61 \cdot 0.6(1-0.6)} = 42.9 \approx 43$.
)

e) Datafordelingen har betydelig mer spredning enn normalfordelingen. Det ser vi også av at dataene har estimert standardavvik på 7.94, mens det i normaltilnærmingsfordelingen er et standardavvik på kun 3.83.

Normalfordelingen er basert på at X er binomisk fordelt. Dette forutsetter igjen at hvorvidt det er *nedbør på ulike dager er uavhengige begivenheter* (I_j'ene uavhengige). Dette er neppe tilfellet. Antakelig er det heller slik at dersom det er nedbør i dag, så er det høyere sannsynlighet for nedbør imorgen, enn om det ikke var nedbør i dag. Dette vil medføre at indikatorvariablene (I_j'ene) for hvorvidt det regner ulike dager, blir *positivt korrelert.* Da får vi at $X = I_1 + I_2 + \cdots + I_{61}$, vil ha større varians enn om I_j'ene er uavhengige, som er forutsetning for at X skal være binomisk fordelt.

Siden X-data fra Bergen har en fordeling med større spredning enn det vi finner i binomisk, $B(n = 61, p = 0.6)$, fordeling, vil en test med forkastningsområde $(41, \infty)$ ha signifikansnivå større enn 0.1003 dersom den brukes på Data fra Bergen.

Oppgave 3

a)

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \epsilon_i, \qquad i = 1, \ldots, 40$$

Designmatrisen:

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{21} & x_{31} \\ 1 & x_{12} & x_{22} & x_{32} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} & x_{3n} \end{bmatrix} = \begin{bmatrix} 1 & 240 & 98 & 140 \\ 1 & 237 & 92 & 140 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 274 & 102 & 95 \end{bmatrix}$$

Regresjonslinje: $\widehat{y} = 73.8 + 0.186x_1 - 0.221x_2 - 0.058x_3$

Estimatet av β_1, koeffisienten foran x_1, er 0.186. Det betyr at for hvert sekund inneværende utbrudd varer, øker tiden til neste utbrudd med 0.186 minutt.

Prediksjon: $\widehat{y}_0 = 73.8 + 0.186 \cdot 240 - 0.221 \cdot 90 - 0.058 \cdot 120 = 91.5$ (minutter).

b) Test for koeffisient til hver av x-variablene:
$j = 1$: $H_0 : \beta_1 = 0$ mot $H_1 : \beta_1 \neq 0$; $p-$verdi $=$ 0.000; dvs. forkast H_0.
$j = 2$: $H_0 : \beta_2 = 0$ mot $H_1 : \beta_2 \neq 0$; $p-$verdi $=$ 0.065; dvs. behold H_0.
$j = 3$: $H_0 : \beta_3 = 0$ mot $H_1 : \beta_3 \neq 0$; $p-$verdi $=$ 0.491; dvs. behold H_0.

Forutsetter at Y_i'ene er uavhengige og normalfordelte med konstant varians. (Alt.: ϵ_i'ene er uif. $\mathcal{N}(0, \sigma^2)$.)

c)

$$\begin{aligned} \mathrm{Var}(\widehat{\beta}_0 + \widehat{\beta}_1 x_{10} - Y_0) &= \mathrm{Var}(\overline{Y} + (x_{10} - \overline{x})\widehat{\beta}_1 - Y_0) \\ &= \mathrm{Var}(\overline{Y}) + \mathrm{Var}\{(x_{10} - \overline{x})\widehat{\beta}_1\} + \mathrm{Var}(Y_0) \end{aligned}$$

$$
\begin{aligned}
&= \mathrm{Var}(\overline{Y}) + (x_{10} - \overline{x})^2 \mathrm{Var}(\widehat{\beta}_1) + \mathrm{Var}(Y_0) \\
&= \frac{\sigma^2}{n} + (x_{10} - \overline{x})^2 \mathrm{Var}(\widehat{\beta}_1) + \sigma^2
\end{aligned}
$$

Videre har vi:

$$
\widehat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_{1i} - \overline{x})Y_i}{S_{xx}}
$$

og da:

$$
\begin{aligned}
\mathrm{Var}(\widehat{\beta}_1) &= \mathrm{Var}\left\{\frac{\sum_{i=1}^{n}(x_{1i} - \overline{x})Y_i}{S_{xx}}\right\} = (\frac{1}{S_{xx}})^2 \sum_{i=1}^{n}(x_{1i} - \overline{x})^2 \underbrace{\mathrm{Var}(Y_i)}_{\sigma^2} \\
&= (\frac{1}{S_{xx}})^2 S_{xx}\sigma^2 = \frac{\sigma^2}{S_{xx}}
\end{aligned}
$$

Da har vi:

$$
\begin{aligned}
\mathrm{Var}(\widehat{\beta}_0 + \widehat{\beta}_1 x_{10} - Y_0) &= \frac{\sigma^2}{n} + (x_{10} - \overline{x})^2 \mathrm{Var}(\widehat{\beta}_1) + \sigma^2 \\
&= \frac{\sigma^2}{n} + (x_{10} - \overline{x})^2 \frac{\sigma^2}{S_{xx}} + \sigma^2 \\
&= \sigma^2 \left\{1 + \frac{1}{n} + \frac{(x_{10} - \overline{x})^2}{S_{xx}}\right\}
\end{aligned}
$$

d) Vi har:

$$
\frac{\widehat{\beta}_0 + \widehat{\beta}_1 x_{10} - Y_0}{\sqrt{S^2\left\{1 + \frac{1}{n} + \frac{(x_{10} - \overline{x})^2}{S_{xx}}\right\}}} \sim t(n-2)
$$

$$
S^2 = \widehat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}\{Y_i - (\widehat{\beta}_0 + \widehat{\beta}_1 x_{1i})\}^2.
$$

Fra dette kan vi utlede:

$$
P\left(\widehat{\beta}_0 + \widehat{\beta}_1 x_{10} - t_{0.025,38}\sqrt{S^2\left\{1 + \frac{1}{n} + \frac{(x_{10} - \overline{x})^2}{S_{xx}}\right\}} \le Y_0 \le \widehat{\beta}_0 + \widehat{\beta}_1 x_{10} + t_{0.025,38}\sqrt{S^2\left\{1 + \frac{1}{n} + \frac{(x_{10} - \overline{x})^2}{S_{xx}}\right\}}\right) = 0.95
$$

som gir prediksjonsintervallet:

$$
\left(\widehat{\beta}_0 + \widehat{\beta}_1 x_{10} - t_{0.025,38}\sqrt{S^2\left\{1 + \frac{1}{n} + \frac{(x_{10} - \overline{x})^2}{S_{xx}}\right\}}, \widehat{\beta}_0 + \widehat{\beta}_1 x_{10} + t_{0.025,38}\sqrt{S^2\left\{1 + \frac{1}{n} + \frac{(x_{10} - \overline{x})^2}{S_{xx}}\right\}}\right)
$$

Insatt data: $(76.17, 104.88)$