Innlevering 4

Oppgave 1

$$s = \frac{1}{2} g t^2 \quad , \quad g = \frac{2s}{t^2} \quad , \quad G = \frac{2S}{T^2}$$

$$E[S] = \mu_s \quad \text{og} \quad Var[S] = \sigma_s^2$$
$$E[T] = \mu_t \quad \text{og} \quad Var[T] = \sigma_t^2$$

$S = 241.3\,m$
$t = 7.02\,s$

a) Vi bruker feil forplantnings formelen

$$\mu_G = E[G] = \frac{2\mu_s}{\mu_t^2}$$

$$\sigma_g^2 = \left(\frac{\partial G}{\partial S}\right)^2 \sigma_s^2 + \left(\frac{\partial G}{\partial T}\right)^2 \sigma_t^2$$

$$= \left(\frac{2}{T^2}\right)^2 \sigma_s^2 + \left(\frac{-4S}{T^3}\right)^2 \sigma_t^2$$

$$= \left(\frac{2}{\mu_t^2}\right) \sigma_s^2 + \left(\frac{4\mu_s}{\mu_T^2}\right) \sigma_t^2$$

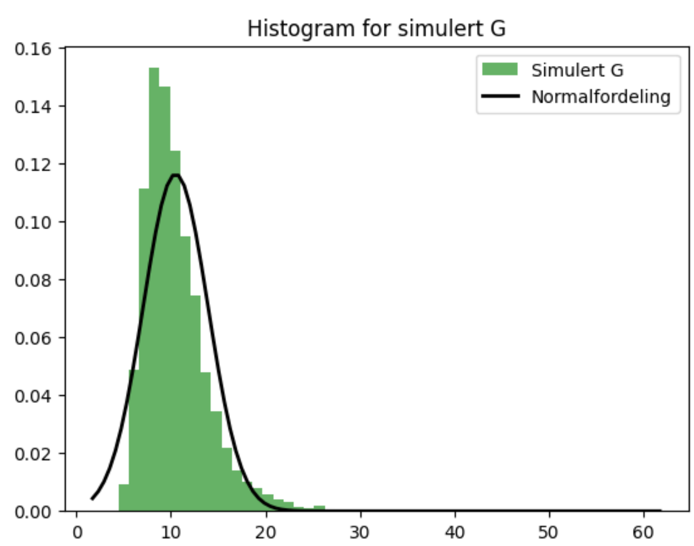
$$\sigma_g^2 = \left(\frac{2}{7.02}\right)^2 \cdot 2^2 + \left(\frac{-4 \cdot 241.3}{7.02^3}\right) \cdot 1^2 = 7.79$$

$$\sigma_g = \sqrt{7.79} = 2.79$$

b)

```python
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm

# Definisjon av parametere
mu_s = 241.3
sigma_s = 2
mu_t = 7.02
sigma_t = 1

# Antall simuleringer
n = 10000

# Simulering av S og T
S_sim = np.random.normal(mu_s, sigma_s, n)
T_sim = np.random.normal(mu_t, sigma_t, n)

# Beregning av G for hver simulering
G_sim = 2*S_sim / T_sim**2

# Beregning av approksimerte verdier for mu_g og sigma_g fra simuleringen
mu_g_sim = np.mean(G_sim)
sigma_g_sim = np.std(G_sim)

# Sannsynlighetshistogram for G
plt.hist(G_sim, bins=50, density=True, alpha=0.6, color='g', label='Simulert G')

# Plott av normalfordeling med funnede mu_g og sigma_g
xmin, xmax = plt.xlim()
x = np.linspace(xmin, xmax, 100)
p = norm.pdf(x, mu_g_sim, sigma_g_sim)
plt.plot(x, p, 'k', linewidth=2, label='Normalfordeling')

plt.title("Histogram for simulert G")
plt.legend()
plt.show()

(mu_g_sim, sigma_g_sim)
```

✓ 0.1s



(2)

$\mu = 10g$

$\sigma = 0.2g$

a) Både X og Y e normalfordelt med μ og σ^2.

gjennomsnittet av X og Y e $\bar{x} = \frac{X+Y}{2}$

dette gir oss en ny normalfordeling med $\mu = 10g$ og $\sigma = \frac{0.2}{\sqrt{2}}$

$$P(\bar{x} \geq 10.2) \rightarrow P\left(Z \geq \frac{10.2-10}{\frac{0.2}{\sqrt{2}}}\right) = P(Z \geq 1.41)$$
$$= 1 - P(Z \leq 1.41)$$
$$= 1 - 0.9207 = 0.0793$$

for at $|X-Y| > 0.4g$:

$$P(|X-Y| > 0.4) = 2 \cdot P(X-Y > 0.4)$$ siden fordelingen e symmetrisk om 0

$$= 2 \cdot P\left(Z > \frac{0.4}{\sqrt{2} \times 0.2}\right)$$
$$= 2 \cdot (1 - P(Z < 1.41))$$
$$= 2 \cdot (1 - 0.9207) = 2 \cdot 0.0793 = 0.1586$$

b) For alternativ 1:

$$\mu_A = \bar{X} = \frac{1}{5}\sum_{i=1}^{5} X_i \quad \text{og} \quad \mu_B = \bar{Y} = \frac{1}{5}\sum_{i=1}^{5} Y_i$$

$$E[\hat{\mu}_A] = \mu_A \qquad Var[\hat{\mu}_A] = \frac{1}{5^2}\sum_{i=1}^{5} Var[X_i] = \frac{5\sigma^2}{25} = \frac{\sigma^2}{5}$$

$$E[\hat{\mu}_B] = \mu_B \qquad Var(\hat{\mu}_B) = \frac{\sigma^2}{5}$$ samme som over

For alternativ 2:

$$\hat{\mu}_A = \frac{1}{2}(\bar{U} + \bar{V}) \quad \text{og} \quad \hat{\mu}_B = \frac{1}{2}(\bar{U} - \bar{V})$$

$$E[\hat{\mu}_A] = \frac{1}{2}\left(E[\bar{U}] + E[\bar{V}]\right) = \frac{1}{2}\left((\mu_A + \mu_B) + (\mu_A - \mu_B)\right) = \mu_A$$

$$E(\hat{\mu}_B) = \frac{1}{2}\left(E[\bar{U}] - E[\bar{V}]\right) = \frac{1}{2}\left((\mu_A + \mu_B) - (\mu_A - \mu_B)\right) = \mu_B$$

$$Var[N_A] = \frac{1}{2^2}\left(Var[\bar{U}] + Var(\bar{V}]\right) = \frac{1}{4}\left(\frac{\sigma^2}{5} + \frac{\sigma^2}{5}\right) = \frac{\sigma^2}{10}$$

$$Var[N_B] = \frac{1}{2^2}\left(Var[\bar{U}] + Var(\bar{V}]\right) = \frac{1}{4}\left(\frac{\sigma^2}{5} + \frac{\sigma^2}{5}\right) = \frac{\sigma^2}{10}$$

Begge alternativene gir godt estimat av N men alternativ 2 gir lavere varians og er derfor å foretrekke. //

③

a) Gitt at X_i e $\sim poi(\lambda)$, a SME for λ

$$L(\lambda) = \prod_{i=1}^{n} f(x_i|\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} = \frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} e^{-n\lambda}$$

Som gir

$$\ln(L(\lambda)) = \ln\left(\frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} e^{-n\lambda}\right) = \left(\sum_{i=1}^{n} x_i\right)\ln(\lambda) - n\lambda - \ln\left(\prod_{i=1}^{n} x_i!\right)$$

deriver med hensyn på λ

$$\frac{d}{d\lambda}\ln(L(\lambda)) = \frac{\sum_{i=1}^{n} x_i}{\lambda} - n$$

setter lik 0 og løser for λ

$$0 = \frac{\sum x_i}{\lambda} - n \rightarrow \hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

er $\hat{\lambda}$ forventningsrett?

$$E[\hat{\lambda}] = E\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[x_i] = \frac{1}{n}\sum_{i=1}^{n}\lambda = \lambda$$

$\hat{\lambda}$ er forventningsrett //

$$Var[\hat{\lambda}] = Var\left[\frac{1}{n}\sum x_i\right] = \frac{1}{n^2}\sum_{i=1}^{n} Var(x_i] = \frac{1}{n^2} n\lambda = \frac{\lambda}{n} //$$

b) For at $\tilde{\lambda}$ skal være forventningsrett må $\tilde{\lambda} = \lambda$

$$E[\tilde{\lambda}] = E[\alpha \bar{X} + \beta \bar{Y}] = \alpha E[\bar{X}] + \beta E[\bar{Y}] = \alpha\lambda + \beta\left(\frac{\lambda}{2}\right)$$

dette skal være lik λ

$$\alpha\lambda + \beta\left(\frac{\lambda}{2}\right) = \lambda$$

$$\alpha + \beta\frac{1}{2} = 1$$

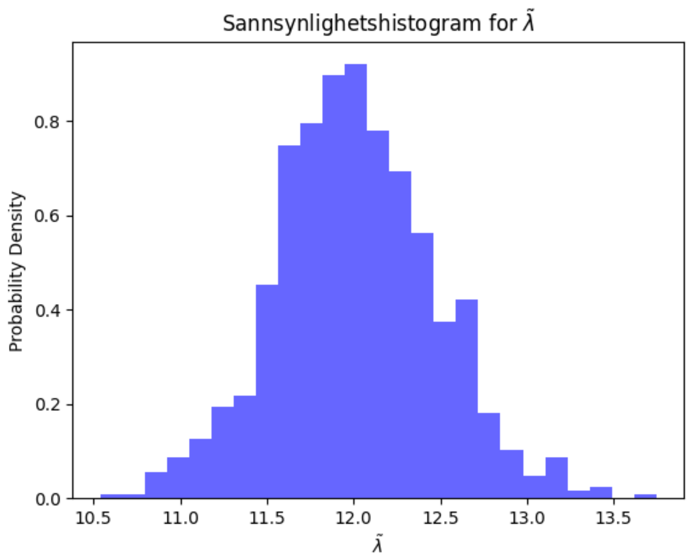
For variansen:

$$Var[\tilde{\lambda}] = Var[\alpha\bar{X} + \beta\bar{Y}] = \alpha^2 Var[\bar{X}] + \beta^2 Var[\bar{Y}]$$
$$= \alpha^2\frac{\lambda}{n} + \beta^2\frac{\lambda}{2m}$$

c)

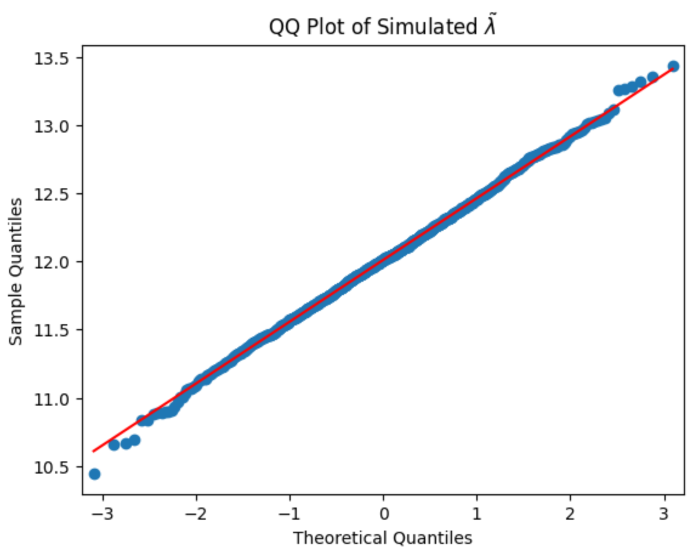
```python
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import poisson

# Gitt
lambda_ = 12
n = 42
m = 35
r = 1000

# Optimaliserte konstanter basert på forrige diskusjon
alpha = 2/3
beta = 2/3

# Simulering
lambdatilde_values = []

for _ in range(r):
    X = poisson.rvs(mu=lambda_, size=n)
    Y = poisson.rvs(mu=lambda_/2, size=m)
    lambdatilde = alpha * np.mean(X) + beta * np.mean(Y)
    lambdatilde_values.append(lambdatilde)

# Sannsynlighetshistogram
plt.hist(lambdatilde_values, bins=25, density=True, alpha=0.6, color='blue')
plt.title('Sannsynlighetshistogram for $\~{\lambda}$')
plt.xlabel('$\~{\lambda}$')
plt.ylabel('Sannsynlighetstetthet')
plt.show()
```

✓ 0.1s




```python
import statsmodels.api as sm

# QQ-plot
sm.qqplot(np.array(lambdatilde_values), line='s')
plt.title('QQ Plot of Simulated $\~{\lambda}$')
plt.show()
```

✓ 0.9s



her se vi at målepunktene våre ligger
bra på linjen som betyr at vi har
en tilnærmet normalfordelt fordeling. //

(4)

a) μ representerer gjennomsnittlig mengde koffein i 1dl coca-cola

σ indikerer hvor mye individuelle målinger kan forventes å variere fra
forventningen μ.

b) for å utlede et $(1-\alpha)\cdot 100\%$ konfidensintervall for μ der σ er kjent
bruker vi formelen for konfidensintervall for en normalfordelt populasjon med kjent varians

$$\bar{X} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

gitt $\bar{x} = 8.2$, $\sigma = 0.19$ og $n = 12$, og vi ønsker $\alpha = 0.05$
får vi $z_{\frac{0.05}{2}} = z_{0.025} = 1.960$
og et konfidensintervall på

$$\min \rightarrow 8.2 - 1.960 \cdot \frac{0.19}{\sqrt{12}} = 8.09 \,//$$
$$\max \rightarrow 8.2 + 1.960 \cdot \frac{0.19}{\sqrt{12}} = 8.31 \,//$$

b) for at bredden på 95% konfidensintervallet våres
skal være 0.1 må vi finne et uttrykk og
løse for n

feilen er gitt som $z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

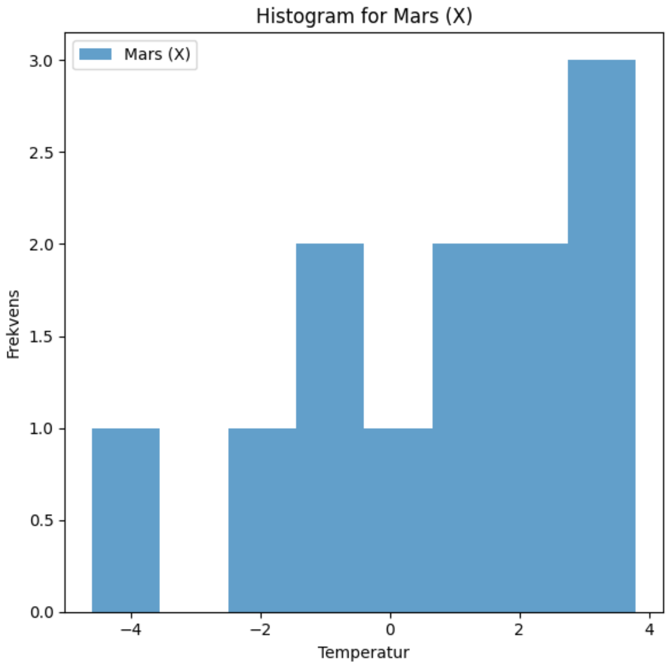
så vi vil at $\frac{0.1}{2} \geq z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

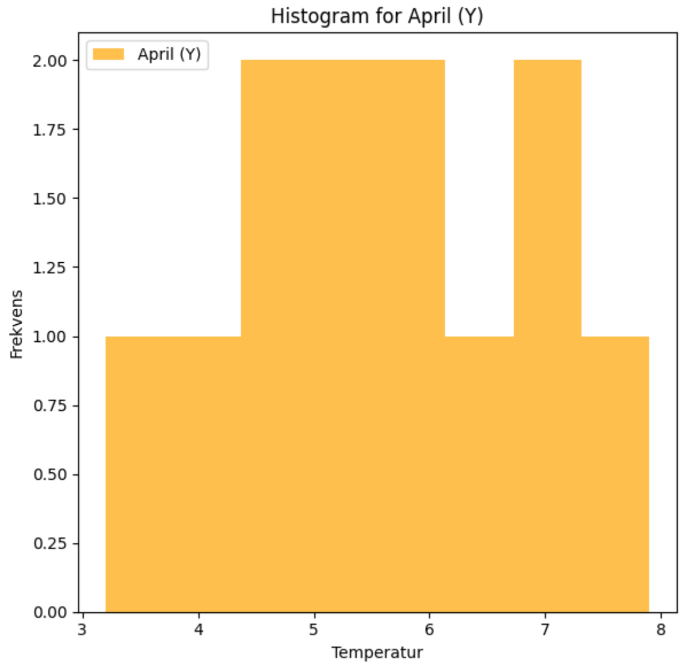
$$0.05 \geq 1.96 \cdot \frac{0.19}{\sqrt{n}}$$
$$n \geq \left(\frac{1.96 \cdot 0.19}{0.05}\right)^2$$

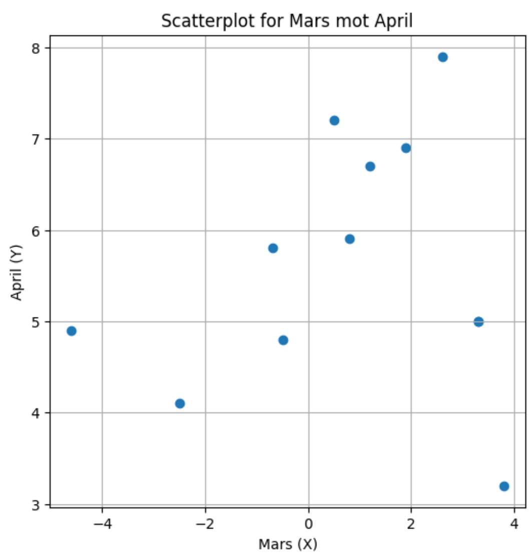
$$n \geq 55.47$$
altså $n \geq 56$ //

5

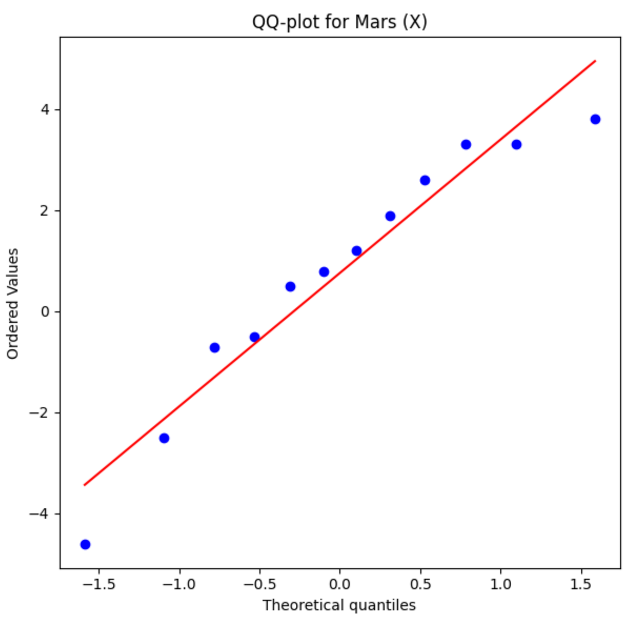
Histogrammene er ikke tilstrekkelig for å si om det er normalfordelt, men utelukker det ikke.

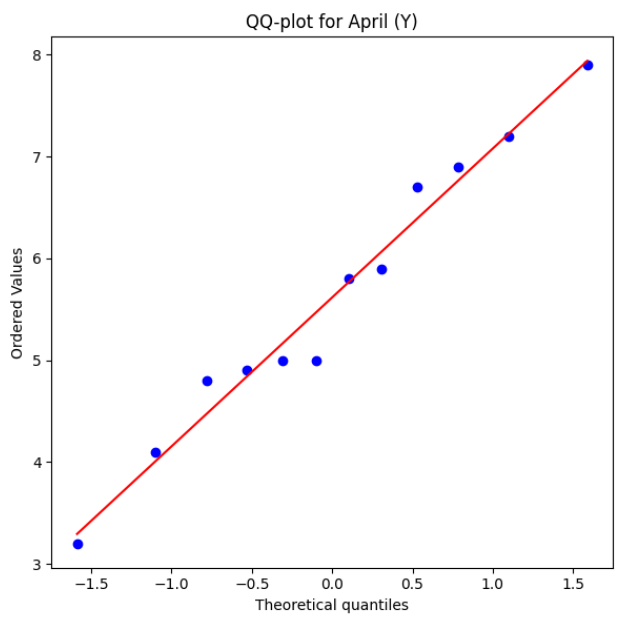






Her ser vi ingen lineær sammenheng mellom X og Y. Dermed er antakelsen om uavhengighet rimelig





Vi se at desse QQ plottene viser at det er rimelig å anta normalfordeling, men kunne vært bedre å sjekke med mer data. //

b) $\bar{x} \pm t_{\frac{\alpha}{2}, n-1} \cdot \frac{s}{\sqrt{n}}$

$t_{\frac{0.01}{2}, 12-1} = 3.106$

$\bar{x} = 0.76°C$

$s = 2.53°C$

$min = 0.76 - 3.106 \frac{2.53}{\sqrt{12}} = -1.508$ //

$max = 0.76 + 3.106 \frac{2.53}{\sqrt{12}} = 3.028$ //