Oblig 1 - STK1110

Kevin Alexander Aslesen

Oppgave 1: side 1 - 12   Oppgave 2: side 13 - 20

Oppgave 1

a) Tetthetsfunksjonen for en log-normal fordeling er:

$$f(x; \mu, \sigma^2) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}}$$

Gitt uavhengighet er da joint tetthetsfunksjonen lik produktet av disse:

$$f(x_1, \dots, x_n; \mu, \sigma^2) = \frac{1}{x_1\sqrt{2\pi\sigma^2}} e^{-\frac{(\ln(x_1)-\mu)^2}{2\sigma^2}} \cdot \ldots \cdot \frac{1}{x_n\sqrt{2\pi\sigma^2}} e^{-\frac{(\ln(x_n)-\mu)^2}{2\sigma^2}}$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} \frac{1}{\prod_{i=1}^{n} x_i} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\ln(x_i)-\mu)^2}$$

Dette er det vi kaller likelihood-funksjonen. Mer presist er likelihood-funksjonen for μ og σ^2 lik:

$$L(\mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \prod_{i=1}^{n} \left(\frac{1}{x_i}\right) e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\ln(x_i)-\mu)^2}$$

Tar logaritmen av denne for å få log-likelihood funksjonen:

$$l(\mu, \sigma^2) = \ln(L(\mu, \sigma^2))$$

$$= \ln\left(\frac{1}{(2\pi\sigma^2)^{n/2}} \prod_{i=1}^{n} \left(\frac{1}{x_i}\right) e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\ln(x_i)-\mu)^2}\right)$$

$$= -\ln((2\pi\sigma^2)^{n/2}) + \ln\left(\prod_{i=1}^{n}\frac{1}{x_i}\right) + \ln\left(e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\ln(x_i)-\mu)^2}\right)$$

$$= -\frac{n}{2}\ln(2\pi\sigma^2) + \sum_{i=1}^{n}\ln\left(\frac{1}{x_i}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(\ln(x_i)-\mu)^2$$

$$= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \sum_{i=1}^{n}\ln(x_i) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(\ln(x_i)-\mu)^2$$

Deriverer log-likelihood funksjonen for μ:

$$\frac{\partial l}{\partial \mu} = -0 - 0 - 0 - \frac{\partial}{\partial \mu}\left(\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\ln(x_i)-\mu)^2\right)$$

$$= -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(-1)\,2\,(\ln(x_i)-\mu)$$

produkt regelen

$$\frac{\partial \ell}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (\ln(x_i) - \mu)$$

$$\left(\sum_{i=1}^{n} \mu = n\mu \right)$$

$$= \frac{1}{\sigma^2} \left(\sum_{i=1}^{n} \ln(x_i) - n\mu \right)$$

Setter den deriverte lik null for å finne $\hat{\mu}_{mle}$:

$$\frac{\partial \ell}{\partial \mu} = 0$$

$$\frac{1}{\sigma^2} \left(\sum_{i=1}^{n} \ln(x_i) - n\mu \right) = 0$$

$$\sum_{i=1}^{n} \ln(x_i) - n\mu = 0$$

$$n\mu = \sum_{i=1}^{n} \ln(x_i)$$

$$\mu = \frac{1}{n} \sum_{i=1}^{n} \ln(x_i)$$

altså $\underline{\underline{\hat{\mu}_{mle} = \frac{1}{n} \sum_{i=1}^{n} \ln(x_i)}}$.

Deriverer så log-likelihoodfunksjonen for $\theta = \sigma^2$:

$$\frac{\partial \ell}{\partial \theta} = -0 - \frac{\partial}{\partial \theta}\left(\frac{n}{2} \ln(\theta) \right) - 0 - \frac{\partial}{\partial \theta} \left(\frac{1}{2\theta} \sum_{i=1}^{n} (\ln(x_i) - \mu)^2 \right)$$

$$= -\frac{n}{2\theta} + \frac{1}{2\theta^2} \sum_{i=1}^{n} (\ln(x_i) - \mu)^2$$

Setter den deriverte lik null for å finne $\hat{\sigma}^2_{mle}$:

$$\frac{\partial \ell}{\partial \theta} = 0$$

$$-\frac{n}{2\theta} + \frac{1}{2\theta^2}\sum_{i=1}^{n}(\ln(x_i)-\mu)^2 = 0$$

$$\frac{1}{\theta}\sum_{i=1}^{n}(\ln(x_i)-\mu)^2 = n$$

$$\theta = \frac{1}{n}\sum_{i=1}^{n}(\ln(x_i)-\mu)^2$$

Med $\theta = \sigma^2$ så har vi altså $\underline{\underline{\hat{\sigma}^2_{mle} = \frac{1}{n}\sum_{i=1}^{n}(\ln(x_i)-\mu)^2}}$.

b) Vi vet at $Y_i \sim N(\mu, \sigma^2)$ og at maksimum likelihood estimatorene for μ og σ^2 i en normal fordeling er gitt ved $\hat{\mu}_{mle} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ og $\hat{\sigma}^2_{mle} = \frac{1}{n}\sum_{i=1}^{n}(Y_i-\mu)^2$. Men siden $Y_i = \log(X_i) = \ln(X_i)$, så kan vi skrive disse om som:

$$\hat{\mu}_{mle} = \frac{1}{n}\sum_{i=1}^{n}\ln(x_i)$$

$$\hat{\sigma}^2_{mle} = \frac{1}{n}\sum_{i=1}^{n}(\ln(X_i)-\mu)^2$$

Dette var mye enklere enn metoden i a)...

c) For én enkel observasjon er log-likelihood lik:

$$\ln(f(x;\mu,\theta)) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\theta) - \ln(x_1) - \frac{1}{2\theta}(\ln(x)-\mu)^2$$

hvor $\theta = \sigma^2$. Finner de forskjellige derivert av denne

mhp. μ og θ:

$$\frac{\partial}{\partial\mu}\ln(f(x;\mu,\theta)) = -\frac{\partial}{\partial\mu}\left(\frac{1}{2\theta}(\ln(x)-\mu)^2\right)$$

produkt regel

$$= -\frac{1}{2\theta}(-1)\,2(\ln(x)-\mu)$$

$$= \frac{1}{\theta}(\ln(x)-\mu)$$

$$\frac{\partial}{\partial\theta}\ln(f(x;\mu,\theta)) = -\frac{\partial}{\partial\theta}\left(\frac{1}{2}\ln(\theta)\right) - \frac{\partial}{\partial\theta}\left(\frac{1}{2\theta}(\ln(x)-\mu)^2\right)$$

$$= -\frac{1}{2\theta} + \frac{1}{2\theta^2}(\ln(x)-\mu)^2$$

$$\frac{\partial^2}{\partial\mu^2}\ln(f(x;\mu,\theta)) = \frac{\partial}{\partial\mu}\left(\frac{1}{\theta}(\ln(x)-\mu)\right)$$

$$= -1/\theta$$

$$\frac{\partial^2}{\partial\theta^2}\ln(f(x;\mu,\theta)) = \frac{\partial}{\partial\theta}\left(-\frac{1}{2\theta} + \frac{1}{2\theta^2}(\ln(x)-\mu)^2\right)$$

$$= \frac{1}{2\theta^2} - \frac{1}{\theta^3}(\ln(x)-\mu)^2$$

$$\frac{\partial^2}{\partial\mu\partial\theta}\ln(f(x;\mu,\theta)) = \frac{\partial}{\partial\mu}\left(\frac{\partial}{\partial\theta}\ln(f(x;\mu,\theta))\right)$$

$$= \frac{\partial}{\partial\theta}\left(\frac{\partial}{\partial\mu}\ln(f(x;\mu,\theta))\right)$$

$$= \frac{\partial}{\partial\theta}\left(\frac{1}{\theta}(\ln(x)-\mu)\right)$$

$$= -\frac{1}{\theta^2}(\ln(x)-\mu)$$

Observasjons matrisen for én observasjon er gitt ved:

$$I(\mu,\theta) = \begin{pmatrix} I_{11}(\mu,\theta) & I_{12}(\mu,\theta) \\ I_{21}(\mu,\theta) & I_{22}(\mu,\theta) \end{pmatrix}$$

hvor $I_{11}(\mu,\theta) = -E\left(\frac{\partial^2}{\partial\mu^2}\ln(f(x;\mu,\theta))\right)$

$$= -E\left(-\frac{1}{\theta}\right)$$

$$= \frac{1}{\theta}$$

$$I_{12}(\mu,\theta) = I_{21}(\mu,\theta) = -E\left(\frac{\partial^2}{\partial\mu\partial\theta}\ln(f(x;\mu,\theta))\right)$$

$$= -E\left(-\frac{1}{\theta^2}(\ln(x)-\mu)\right)$$

$$= \frac{E(\ln(x)-\mu)}{\theta^2}$$

$$= \frac{E(\ln(x))-\mu}{\theta^2}$$

$$= \frac{E(Y)-\mu}{\theta^2}$$

$$= \frac{\mu-\mu}{\theta^2} = 0$$

$$
\begin{aligned}
I_{22} &= -E\left(\frac{\partial^2}{\partial\theta^2}\ln(f(x;\mu,\theta))\right) \\
&= -E\left(\frac{1}{2\theta^2}-\frac{1}{2\theta^3}(\ln(x)-\mu)^2\right) \\
&= -\frac{1}{\theta^2}+\frac{1}{2\theta^3}E\left((\ln(x)-\mu)^2\right) \\
&= -\frac{1}{\theta^2}+\frac{1}{2\theta^3}E\left((y-\mu)^2\right) \\
&= -\frac{1}{\theta^2}+\frac{1}{2\theta^3}V(y) \qquad \text{hvor } V(y)=\sigma^2=\theta \\
&= -\frac{1}{\theta^2}+\frac{1}{2\theta^2} \\
&= \frac{1}{2\theta^2}
\end{aligned}
$$

Observasjonsmatrisen er da:

$$
I(\mu,\theta) = \begin{pmatrix} \frac{1}{\theta} & 0 \\ 0 & \frac{1}{2\theta^2} \end{pmatrix}
$$

setter vi inn σ^2 for θ så har vi matrisen:

$$
I(\mu,\sigma^2) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}
$$

Siden n er stor og $X_1, \dots X_n$ ikke er avhengig av μ og σ^2, så betyr det at maksimum likelihood estimatorene $\hat{\mu}_{mle}$ og $\hat{\sigma}^2_{mle}$ har tilnærmet normalfordeling med henholdsvis forventning μ og σ^2, samt henholdsvis varians $1/n\, I_{11}(\mu,\sigma^2)$ og $1/n\, I_{22}(\mu,\sigma^2)$.

Et estimat på standardfeilen til $\hat{\mu}_{mle}$ og $\hat{\sigma}^2_{mle}$ er da:

$$\sigma_{\hat{\mu}_{mle}} = \sqrt{1/n\, I_{11}(\mu,\sigma^2)} = \sqrt{1/n\left(\frac{1}{\sigma^2}\right)} = \underline{\underline{\frac{\sigma}{\sqrt{n}}}}$$

$$\sigma_{\hat{\sigma}^2_{mle}} = \sqrt{1/n\, I_{22}(\mu,\sigma^2)} = \sqrt{1/n\left(\frac{1}{2\sigma^4}\right)} = \underline{\underline{\sqrt{\frac{2}{n}}\,\sigma^2}}$$

d) Leser inn dataen om forsikringskravene:

```
url = "https://www.uio.no/studier/emner/matnat/math/STK1110/data/forsikringskrav.txt"
data_table = read.table(url, header = FALSE)
data = data_table$V1
n = length(data)
```

Lager en bootstrap fordeling for forventningen μ for å estimere standardfeilen i $\hat{\mu}_{mle}$. Sammenlikner dette med det vi fikk i c).

```r
B = 1000
data_mean_boot = c()
for (i in 1:B){
    data_boot = c()
    for (i in 1:n){
        krav = sample(data, 1)
        data_boot = append(data_boot, krav)
    }
    data_mean_boot = append(data_mean_boot, mean(data_boot))
}
print(mean(exp(data_mean_boot)))
print(sd(exp(data_mean_boot)))
print(mean(data))
print(sd(data)/sqrt(n))
```

Får at standard feilen for $\hat{\mu}_{mle}$ ved bootstrap er $S^{\mu}_{boot} = 0.366$, mens for estimeringen fra c) er $\sigma_{\hat{\mu}_{mle}} = 0.362$. Her stemmer altså teorien med målingene.

Gjør tilsvarende for $\hat{\sigma}^2_{mle}$:

```r
B = 1000
data_sd_boot = c()
for (i in 1:B){
    data_boot = c()
    for (i in 1:length(data)){ # nolint
        krav = sample(data, 1)
        data_boot = append(data_boot, krav)
    }
    data_sd_boot = append(data_sd_boot, sd(data_boot))
}
print(sd(data_sd_boot))
print(sqrt(2/length(data)) * sd(data)^2)
```

Her får vi at $S^{\sigma^2}_{boot} = 1.56$, mens for estimeringen fra c) er $\sigma_{\hat{\sigma}^2_{mle}} = 14.81$. Dette var veldig stor forskjell, som betyr at teorien ikke stemmer med målingene.

e) Lager et 95 % konfidens intervall for $E(X_i)$. Vi har at $E(X_i) = \bar{X}$ er snittet til målingen og s er standard avviket til målingen. Siden n er stor så kan vi anta at $\bar{X}$ er normalfordelt og dermed finne konfidensintervallet ved:

$$\bar{X} \pm z_{0.025} \cdot \frac{s}{\sqrt{n}}$$

Den kritiske z-verdien er $z_{0.025} = 1.960$, snittet av målingene er $\bar{X} = 24.14$ og standardavviket av målingene er $s = 28.92$. Vi vet fra før at antall målinger er $n = 6377$. Da får vi følgende:

$$24.14 \pm 0.71$$

På intervall form er dette:

$$(23.43, \; 24.85)$$

runder av for at det skal gi mening:

$$\underline{\underline{(23, \; 25)}}$$

f) Lager så et 95 % konfidensintervall for $V(X_i)$.

Huis vi antar at X_i er normalfordelt så har vi at $V(X_i) = \sigma^2$. Konfidensintervallet finne da ved:

$$\left(\frac{n-1}{\chi^2_{\frac{\alpha}{2}, n-1}} \cdot s^2 \quad , \quad \frac{n-1}{\chi^2_{1-\frac{\alpha}{2}, n-1}} \cdot s^2 \right)$$

De kritiske χ^2-verdiene finne vi i R med funksjonen qchisq(). Vi får da $\chi^2_{0.025, 6376} = 6599$ og $\chi^2_{0.975, 6376} = 6157$.

Med $s = 28.92$ fra e), så får vi intervallet:

$$\underline{\underline{(808 \, ; \, 866)}}$$

g) Lager et 95 % konfidensintervall for $V(X_i)$ ved å benytte persentil intervallet av bootstrap-fordelingen som ble brukt i d) for σ. Sorter først fordelingen og plukker ut element #25 og #975 og kvadrerer disse:

```
data_sd_boot_sort = sort(data_sd_boot)
print(data_sd_boot_sort[25]^2)
print(data_sd_boot_sort[975]^2)
```

Det følgende intervallet ble da:

$$\underline{\underline{(685, \; 1041)}}$$

Dette intervallet var mye større enn det vi fikk i f). Dette var litt forventet ettersom vi visste at X_i ikke er tilnærmet normal fordelt, og at intervallet ville da være urealistisk. Begge intervallene inneholder (som forventet) variansen fra målingen vår $s^2 = 28.92^2 \approx 836$. Vi kan med 95 % sikkerhet si at den virkelige variansen for tilforsikringskrav vil ligge mellom (runder opp) 685 og 1041.

Oppgave 2

a) Snittet av målingene ($n = 15$) er $\bar{x} = 559.7$ og standard-avviket er $s = 28.6$. Med disse verdiene og med antagelsen om at populasjonen er normalfordelt med forventning μ og varians σ^2, så lager vi et 95% konfidens-intervall for forventning μ basert på målingene våre med:

$$\left(\bar{x} - t_{0.025, n-1} \frac{s}{\sqrt{n}} , \; \bar{x} + t_{0.025, n-1} \frac{s}{\sqrt{n}} \right)$$

hvor den kritiske t-verdien ved $\frac{\alpha}{2} = 0.025$ og frihets-grader $df = n-1 = 15-1 = 14$ er $t_{0.025, 14} = 2.145$.

Setter inn alle verdiene og får intervallet:

$$\underline{\underline{(543.9 , \; 575.5)}}$$

Vi kan da si med 95 % sikkerhet at den virkelig verdien for μ vil ligge mellom 543.9 mg/100g og 575.5 mg/100g.

b) Begynner først med å liste opp målingene og finne lengden som vi vet er $n = 15$:

```r
data = c(525,587,547,558,591,531,571,551,566,622,561,502,556,565,562)
n = length(data)
```

Lager så $m = 10000$ datasett, hvor hvert datasett inneholder de stokastiske variablene $X_1, X_2, \ldots, X_{15} \overset{uif}{\sim} N(588, 30^2)$ ved hjelp av funksjonen rnorm(15, 588, 30).

For hvert av disse datasettene lager vi et 95% konfidensintervall for μ som i a) og teller hvor mange av disse intervallene som inneholder den virkelig verdien $\mu = 558$:

```r
set.seed(1)
m <- 10000
count <- 0
for (i in 1:m) {
    sim_data <- rnorm(n, 558, 30)
    sim_mean <- mean(sim_data)
    sim_sd   <- sd(sim_data)
    t_0.025  <- 2.145
    conf_low <- sim_mean - t_0.025 * sim_sd / sqrt(n)
    conf_hig <- sim_mean + t_0.025 * sim_sd / sqrt(n)
    if (conf_low < 558 && conf_hig > 558) {
        count <- count + 1
    }
}
print(count)
```

Dette ga count = 9527.

Vi fikk altså at 95.27 % av konfidensintervallene inneholdte $\mu = 558$. Dette stemmer godt med teorien bak konfidensintervall, hvor "95% sikkerhet" tilsvarer at 95 % av intervallene vi får ved mange målinger vil inneholde den virkelige verdien for μ.

c) Det tilnærmede intervallet for store utvalg er:

$$\bar{x} \pm 1.96 \cdot \frac{s}{\sqrt{n}}$$

Dette er det samme som 95 % intervallet, utenom at 1.96 er satt inn for t-verdien. Dette skyldes av grenseverdien for t-verdien ved $\frac{\alpha}{2} = 0.025$ som går mot 1.96 når $n \to \infty$ $\left(t_{0.025, n-1} \overset{n\to\infty}{=} 1.96\right)$. Lager det tilnærmede intervallet for de 10000 data settene som i b) og finner antallet av dem som inneholder $\mu = 558$:

```r
m <- 10000
count <- 0
for (i in 1:m) {
    sim_data <- rnorm(n, 558, 30)
    sim_mean <- mean(sim_data)
    sim_sd   <- sd(sim_data)
    conf_low <- sim_mean - 1.96 * sim_sd / sqrt(n)
    conf_hig <- sim_mean + 1.96 * sim_sd / sqrt(n)
    if (conf_low < 558 && conf_hig > 558) {
        count <- count + 1
    }
}
print(count)
```

Dette gir count = 9357. Altså 93.57 % av de tilnærmede intervallene for store utvalg inneholdte $\mu = 558$. Det er 1.7% lavere enn det vi fikk i b). Dette gir mening ettersom vi effektivt sett har gjort intervallene smalere med 1.96 istedenfor 2.145, som betyr at færre intervaller vil inneholde $\mu = 558$.

d) Siden vi antar at populasjonen er normalfordelt så finner vi et 95 % konfidensintervall for variansen σ^2 ved:

$$\left(\frac{n-1}{\chi^2_{0.025,\, n-1}} \cdot s^2 ,\ \frac{n-1}{\chi^2_{0.975,\, n-1}} \cdot s^2 \right)$$

hvor de kritiske chi-verdiene er $\chi^2_{0.025,14} = 26.119$ og $\chi^2_{0.975,14} = 5.629$. Kvadratet av dette intervallet gir 95 % konfidens intervall for σ. Regner dette for $m = 10000$ datasett som i b) og teller hvor mange av disse intervallene som inneholder den virkelige verdien $\sigma = 30$:

```
m <- 10000
count <- 0
for (i in 1:m) {
    sim_data <- rnorm(n, 558, 30)
    sim_sd   <- sd(sim_data)
    chi_low  <- 26.119
    chi_high <- 5.692
    conf_low <- sqrt((n-1) / chi_low * sim_sd^2)
    conf_hig <- sqrt((n-1) / chi_high * sim_sd^2)
    if (conf_low < 30 && conf_hig > 30) {
        count <- count + 1
    }
}
print(count)
```

Dette gir count = 9482. Altså 94.82 % av konfidensintervallene inneholder $\sigma = 30$. Dette stemmer godt med det vi ønsket å få.

c) Vi har at $Z_i = \frac{X_i - \mu}{\sigma}$. Frem til nå har vi antatt at $Z_i \sim N(0,1)$, men nå antar vi at $Z_i \overset{u.i.f}{\sim} t_7$ altså Z_i er t-fordelt med 7 frihetsgrader. Med andre ord antar vi at populasjonen ikke er normalfordelt. Vi skal se hvor stor effekt dette har på antall konfidensintervaller som inneholder den virkelig verdien $\mu = 558$. Først lager vi en tilfeldig t-fordeling med 7 frihetsgrader med $n = 15$ punkter vha. funksjonen $rt(n, df=7)$ og deretter gjør den om med $X_i = \mu + \sigma \cdot Z_i = 558 + 30 Z_i$. Regner så konfidensintervaller som i b):

```
m <- 10000
count <- 0
for (i in 1:m){
    n = 15
    tdist  <- rt(n, df=7)
    xdist  <- 558 + 30 * tdist
    x_mean <- mean(xdist)
    x_sd   <- sd(xdist)
    t_0.025  <- 2.145
    conf_low <- x_mean - t_0.025 * x_sd / sqrt(n)
    conf_hig <- x_mean + t_0.025 * x_sd / sqrt(n)
    if (conf_low < 558 && conf_hig > 558) {
        count <- count + 1
    }
}
print(count)
```

Dette ga count = 9518. Altså 95.18 % av konfidensintervallene inneholder $\mu = 558$. Prosent-antallet vi fikk i b) var 95.27 %. Vi ser altså at det er liten forskjell for prosent-antallene mellom normal-antagelse og t_7-antagelse. Med dette kan vi si at metoden for å lage konfidensintervall for μ med antagelse om normal fordeling er robust. Gir mening ettersom t-fordelingen har lik midtpunkt som normalfordelingen.

f) Lager $m = 10000$ datasett som i e) og lager et 95 % konfidensintervall for standardavviket til X_i for hvert datasett som i d). Teller antall intervaller som inneholder $\tilde{\sigma} = \sqrt{1.4}\,\sigma = 30\sqrt{1.4}$:

```r
m <- 10000
count <- 0
for (i in 1:m) {
    n = 15
    tdist  <- rt(n, df=7)
    xdist  <- 558 + 30 * tdist
    x_sd   <- sd(xdist)
    chi_low  <- 26.119
    chi_high <- 5.692
    conf_low <- sqrt((n-1) / chi_low * x_sd^2)
    conf_hig <- sqrt((n-1) / chi_high * x_sd^2)
    if (conf_low < sqrt(1.4)*30 && conf_hig > sqrt(1.4)*30) {
        count <- count + 1
    }
}
print(count)
```

Dette ga count = 8887. Altså 88.87% av intervallene inneholdte $\tilde{\sigma} = 30\sqrt{1.4}$. Fra d) fikk vi prosentandel ved normalfordeling til å være 94.82 %. Dette en merkbar forskjell. Men dette gir mening ettersom t-fordelingen er en "klemt" versjon av normalfordelingen med høyere standardavvik. Dette betyr at metoden for å lage konfidensintervall for σ hvor vi antar normalfordeling ikke er robust.