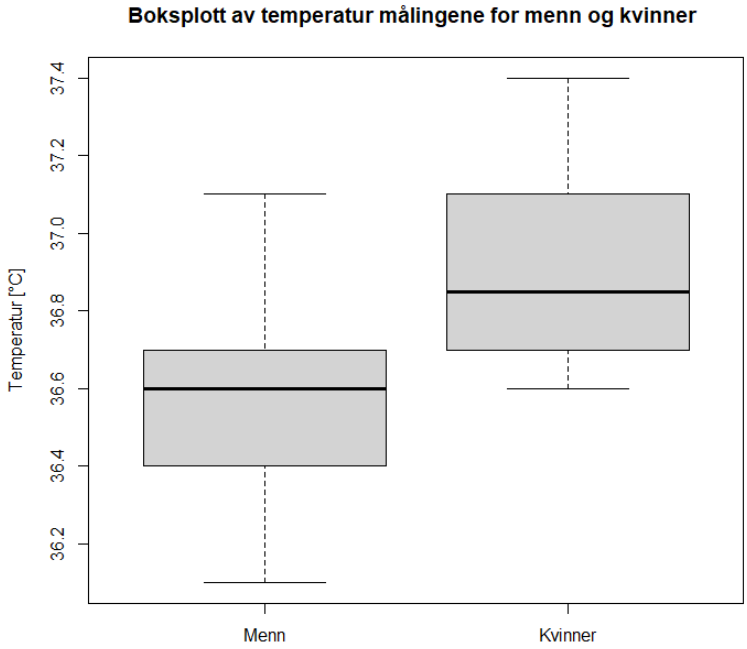
Oblig 2 - STK1110

Kevin Alexander Aslesen

November 2022

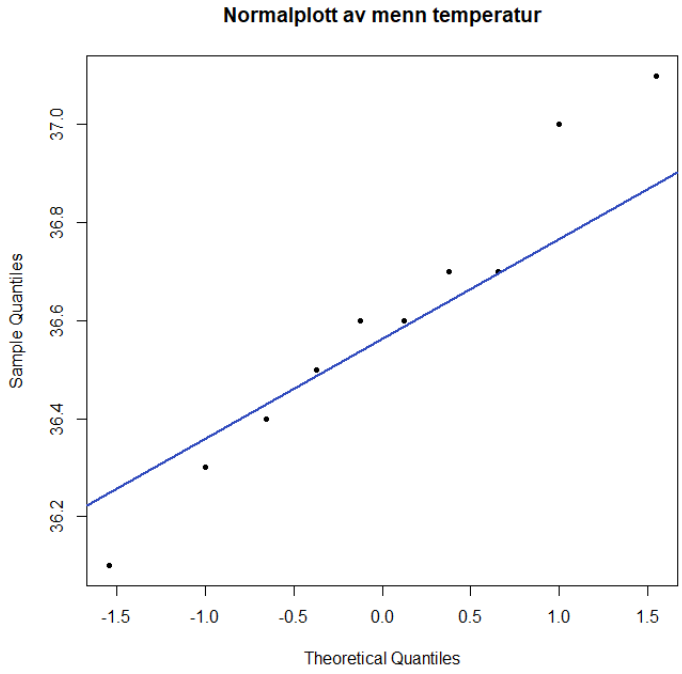
Oppgave 1

a) Boksplott av de målte temperaturene er vist i Figur 1. Vi ser fra boksplot-



Figur 1: *Boksplott for målingene av kroppstemperaturen til menn og kvinner.*

tene at kroppstemperaturene for kvinner ligger litt høyere enn for menn. Den største temperaturen for menn tilsvarer øvre kvartil for kvinnene, mens den laveste verdien for kvinne temperaturene tilsvarer gjennomsnitts temperaturen for menn. Ved å bare se på disse plottene alene så kan man fort prøve å konkludere med at kvinner har høyere kroppstemperatur enn menn. Dette skal vi se nærmere på.



Figur 2: *Normalfordelingsplott for kroppstemperaturene til menn.*

b) Normalfordeligsplott for målingene av temperaturen for menn og kvinner er vist hver for seg i henholdsvis Figur 2 og Figur 3. Vi ser fra begge normalplottene at vi får en ganske rett linje, som tyder på at målingene kommer fra en normalfordelt populasjon.

c) Vi vil teste $H_0 : \mu_1 - \mu_2 = 0$ mot $H_A : \mu_1 - \mu_2 \neq 0$ med signifikantnivå 5% ($\alpha = 0.05$). Hvis variansen ikke er det samme for de to utvalgene, så benytter vi testobservatoren:

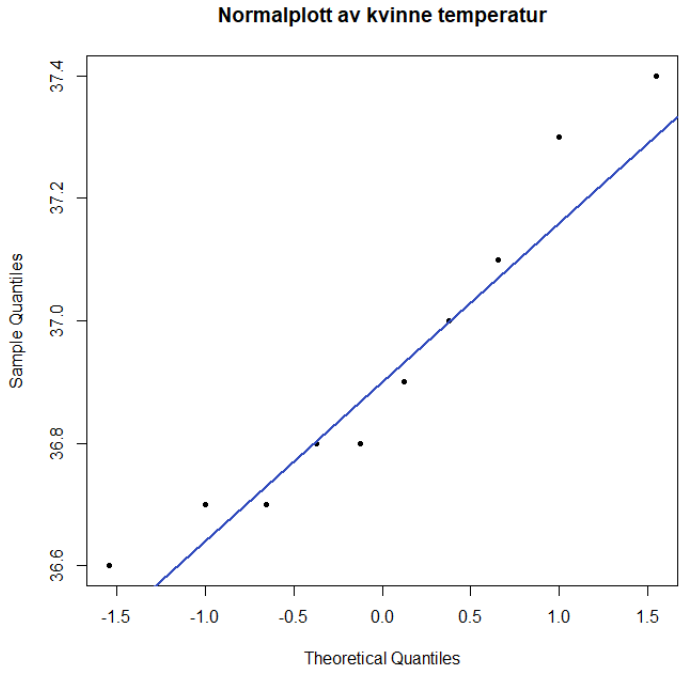
$$t_p = \frac{(\bar{x} - \bar{y}) - \Delta_0}{\sqrt{s_p^2/m + s_p^2/n}} = \frac{\bar{x} - \bar{y}}{\sqrt{s_p^2/m + s_p^2/n}}$$

hvor s_p^2 er den kombinerte estimatoren for variansen σ^2 gitt ved:

$$s_p^2 = \frac{m-1}{m+n-2} s_1^2 + \frac{n-1}{m+n-2} s_2^2$$

Siden $m = n$ i vårt tilfelle, så reduseres dette til:

$$s_p^2 = \frac{s_1^2 + s_2^2}{2}$$



Figur 3: *Normalfordelingsplott for kroppstemperaturene til kvinner.*

hvor s_1 og s_2 er det målte standardavviket til temperaturen hos menn og kvinner, respectfully. Vi antar først at variansen er det samme for de to utvalgene, så vil lar $s_p = s_1$. Benytter da følgende kode for å finne denne testobservasjoenen, t_p, gitt målingene våre:

```
> temp_menn = c(36.1, 36.3, 36.4, 36.6, 36.6, 36.7, 36.7, 37.0,
    36.5, 37.1)
> temp_kvin = c(36.6, 36.7, 36.8, 36.8, 36.7, 37.0, 37.1, 37.3,
    36.9, 37.4)
>
> m = length(temp_menn)
> n = length(temp_kvin)
>
> x.bar = mean(temp_menn)
> y.bar = mean(temp_kvin)
>
> s1 = sd(temp_menn)
> s2 = sd(temp_kvin)
> s_p = s1
>
> t_p = (x.bar - y.bar) / (s_p * sqrt(1/m + 1/n))
> print(t_p)
[1] -2.444631
```

Vi får altså $t_p = -2.44$. Vi finner så ut om denne verdien er tilstrekkelig nok

for å forkaste H_0 eller ikke ved å finne forkastningsområdet. Den kritiske t-verdien for $\alpha/2 = 0.025$ og $m + n - 2 = 10 + 10 - 2 = 18$ frihetsgrader finner vi fra Tabell A.6 i lærerboka til være $t_{\alpha/2,m+n-2} = t_{0.025,18} = 2.101$. Siden vi benytter en tosidig test, så er forkastningsområdet da lik

$$t_p \leq -2.101 \qquad \text{og} \qquad 2.101 \leq t_p$$

Ettersom $t_p = -2.44 < -2.101$, så forkaster vi da H_0 med signifikansnivå $\alpha = 0.05$. Testen vår konkluderer da med at det er forskjell mellom temperaturen til kvinner og menn. P-verdien finner vi ved hjelp av Tabell A.7 i lærerboka, hvor vi finner arealet under grafen etter 2.44 og multipliserer dette med to (siden vi har en tosidet test):

$$\begin{aligned} \text{P-verdi} &= 2[1 - \Phi_t(2.44)] \\ &= 2[0.014] \\ &= 0.028 \end{aligned}$$

Vi ser at P-verdi $= 0.028 < 0.05 = \alpha$, som også resulter i å forkaste H_0. Lar vi heller $s_p = s_2$, så får vi $t_p = -2.76$ og P-verdi $= 0.012$, som også resulterer i forkastelsen av H_0. Et 95% konfidensintervall for temperaturforskjellen er:

$$\begin{aligned} (\bar{x} - \bar{y}) &\pm t_{0.025,18} \cdot s_1 \sqrt{1/m + 1/n} \\ (36.60 - 36.93) &\pm 2.101 \cdot 0.30 \sqrt{1/10 + 1/10} \\ -0.330 &\pm 0.263 \end{aligned}$$

Dette tilsvarer til intervallet $(-0.593, -0.067)$. Vi ser at intervallet ikke inneholder 0, som betyr at vi i dette tilfellet også konkluderer med at kroppstemperaturen mellom menn og kvinner er forskjellig ved å forkaste H_0.

d) I dette tilfellet antar vi forskjellige varianser, så testobservatoren blir nå:

$$t = \frac{(\bar{x} - \bar{y}) - \Delta_0}{\sqrt{s_1^2/m + s_2^2/n}} = \frac{\bar{x} - \bar{y}}{\sqrt{s_1^2/m + s_2^2/n}}$$

Dette får vi til å bli

```
> t = (x.bar - y.bar) / sqrt(s1^2/m + s2^2/n)
> print(t)
[1] -2.590062
```

Altså $t = -2.59$. Forkastningsområdet i dette tilfellet er litt annerledes ettersom t fordelingen er bestemt av Welchs formel for frihetsgradene:

$$\nu = \frac{(s_1^2/m + s_2^2/n)^2}{((s_1^2/m)^2/(m-1) + (s_2^2/n)^2/(n-1))}$$

```
> welch = (s1^2/m + s2^2/n)^2 / ((s1^2/m)^2 / (m - 1) + (s2^2/n
    )^2 / (n - 1))
> print(welch)
[1] 17.7338
```

For våre målinger får vi altså $\nu = 17.7$ som vi runder ned til 17. Da er den kritiske t-verdien $t_{0.025,17} = 2.110$, som gir oss forkastningsområdet:

$$t \geq t_{0.025,17} = 2.110$$

$$t \leq -t_{0.025,17} = -2.110$$

Siden $-2.59 < -2.110$ så forkaster vi H_0 også i dette tilfellet. P-verdien i dette tilfellet er

$$\begin{aligned} \text{P-verdi} &= 2[1 - \Phi_t(2.59)] \\ &= 2[0.009] \\ &= 0.018 \end{aligned}$$

Vi har altså at P-verdi $= 0.018 < 0.05 = \alpha$, som betyr at vi forkaster H_0. Sammenlikner dette med **t.test()** funksjonen i R:

```
> t.test(temp_menn, temp_kvin)

        Welch Two Sample t-test

data:  temp_menn and temp_kvin
t = -2.5901, df = 17.734, p-value = 0.01863
alternative hypothesis: true difference in means is not equal
    to 0
95 percent confidence interval:
 -0.59796699 -0.06203301
sample estimates:
mean of x mean of y
    36.60     36.93
```

Vi ser at funksjonen genererer en P-verdi på 0.01863. Dette stemmer godt med det vi regnet, hvor den lille forskjellen skyldes grovere avrundingen i utregningen vår.

e) Tester $H_0 : \sigma_1 = \sigma_2$ mot $H_A : \sigma_1 \neq \sigma_2$ ved F-test med signifikansnivå $\alpha = 0.02$. Da benytter vi testobservasjonen:

$$f = s_1^2/s_2^2$$

Målingene våre får dette til å bli:

```
> s1^2/s2^2
[1] 1.279251
```

Forkastningsområdet i dette tilfellet tilsvarer de kritiske F-verdiene hvor arealet under F-grafen er 0.02/2=0.01 på hver side. Disse verdiene finner vi i Tabell A.8 i lærerboka med frihetsgradene $\nu_1 = 9$ og $\nu_2 = 9$:

$$f \geq F_{0.01,9,9} = 5.35$$

$$f \leq F_{0.99,9,9} = 0.187$$

Testobservasjonen $f = 1.279$ ligger ikke i forkastningsområdet, så vi forkaster ikke $H_0 : \sigma_1 = \sigma_2$. Det er derfor rimelig å anta at variansen mellom utvalgene er like. Sjekker dette mot **var.test()** funksjonen i R:

```
> var.test(temp_menn, temp_kvin)

        F test to compare two variances

data:  temp_menn and temp_kvin
F = 1.2793, num df = 9, denom df = 9, p-value = 0.7197
alternative hypothesis: true ratio of variances is not equal to
    1
95 percent confidence interval:
 0.3177479 5.1502577
sample estimates:
ratio of variances
          1.279251
```

Vi ser at R funksjonen regnet ut en P-verdi på 0.7197. Dette ligger over signifikansnivået $\alpha = 0.02$ som betyr at H_0 ikke blir forkastet i dette tilfellet. Dette stemmer overens med konklusjonen fra vår F-test.

f) Et rimelig anslag for en ny måling X_{n+1} gitt et tilfeldig målingssett $X_1, X_2, ..., X_n$ er snittet av målingssettet, $\bar{X}$. Dermed er et rimelig anslag for $X_{11} - Y_{11}$, med gitt målingssett $(X_1 - Y_1, X_2 - Y_2, ..., X_{10} - Y_{10})$ lik snittet (forventningen) av målingssettet, $E(X - Y) = E(X) - E(Y) = \bar{X} - \bar{Y}$.

Siden X_{11}, Y_{11}, $\bar{X}$ og $\bar{Y}$ er alle normalfordelt, så er lineær kombinasjonen

$$X_{11} - Y_{11} - (\bar{X} - \bar{Y})$$

også normalfordelt. Med dette kan vi lage et prediksjonsintervall for differansen $X_{11} - Y_{11}$, ved å se på $X - Y$ som ett utvalg. Da finner vi et 95% prediksjonsintervall med $n - 1 = 9$ frihetsgrader for differansen $X_{11} - Y_{11}$ med formelen

$$(\bar{x} - \bar{y}) \pm t_{0.025,9} \cdot s\sqrt{1 + 1/n}$$

hvor s er standardavviket til målingene x og y som vi antar er like. Den kritiske t-verdien finner vi til å være $t_{0.025,9} = 2.262$. Intervallet beregnes da med koden:

```
> x = temp_kvin
> y = temp_menn
> n = length(x)
> s = sd(x)
> t.crit = 2.262
> x.bar = mean(x)
> y.bar = mean(y)
> lower = (x.bar - y.bar) - t.crit * s * sqrt(1 + 1/n)
> upper = (x.bar - y.bar) + t.crit * s * sqrt(1 + 1/n)
> print(c(lower, upper))
[1] -0.3031356  0.9631356
```

Med dette kan vi si med 95% sikkerhet at forskjellen $X_{11} - Y_{11}$ vil befinne seg mellom -0.30 og 0.96. I prediksjonsintervall som dette så får vi sannsynlige verdier for en eventuelt ny måling, sammenlignet med konfidensintervall hvor man får sannsynlige verdier for parametere som er av interesse.

Oppgave 2

a) I dette tilfellet vil det passe best å se på en paret sammenligning ettersom det gir mening at tvillingene i hvert par ikke er helt uavhengige av hverandre gitt at tvillingene er biologisk tilknyttet. For videre analyse så må vi anta at selve parrene (A_1, B_1), (A_2, B_2), ..., (A_{31}, B_{31}) er uavhengige fra hverandre, og at forskjellen mellom parrene er tilnærmet normalfordelt.

b) Vil teste $H_0 : \mu_D = 0$ mot $H_A : \mu_D \neq 0$ med signifikansnivå $\alpha = 0.05$. Testobservatoren for denne testen er

$$t = \frac{\bar{d} - \Delta_0}{s_D/\sqrt{n}} = \frac{\bar{d}}{s_D/\sqrt{n}}$$

hvor vi har fra tabellen $\bar{d} = -3.26$ og $s_D = 8.81$. Setter inn disse verdiene med $n = 31$ og får testobservatoren:

$$t = \frac{-3.26}{8.81\sqrt{31}} = -2.06$$

Den kritiske t-verdien med $\alpha/2 = 0.025$ og antall frihetsgrader $\text{df} = n - 1 = 31 - 1 = 30$ finner vi til å være $t_{0.025,30} = 2.042$. Det betyr at forkastnigsområdet for denne testen er:

$$t \leq -2.042 \qquad \text{og} \qquad 2.042 \leq t$$

Ettersom vi fikk $t = -2.06 < -2.042$ så forkaster vi altså H_0 med signifikansnivå $\alpha = 0.05$. Siden testen vår er tosidet, så er den tilhørende P-verdien lik $2[1 - \Phi_t(2.06)]$ hvor Φ_t er arealet til venstre for t fordelingen vår med 30 frihetsgrader. Vi beregner:

```
> p = 2*(1-pt(2.06, 30))
> print(p)
[1] 0.04816536
```

P-verdien er altså 0.048, som er akkurat mindre enn signifikansnivået $\alpha = 0.05$. Vi konkluderer dermed også utifra P-verdien at H_0 forkastes. Fra testen vår så betyr dette at det er en forskjell i IQ mellom tvillingene som vokser opp med biologiske foreldrene og de som vokser opp med adoptivforeldrene.

c) Lager et 95% konfidensintervall for μ_D med formelen

$$\bar{d} \pm t_{0.025,30} \frac{s_D}{\sqrt{n}}$$

hvor disse faktorene ble bestemt i c). Beregner dette med:

```
> n = 31
> d.bar = -3.26
> s.D = 8.81
> t.crit = 2.042
> lower = d.bar - t.crit * s.D/sqrt(n)
> upper = d.bar + t.crit * s.D/sqrt(n)
> print(c(lower, upper))
[1] -6.49110298 -0.02889702
```

Med 95% sikkerhet har vi altså at μ_D ligger i intervallet (-6.49, -0.03). Vi ser at konfidensintervallet bare inneholder negative verdier, som betyr at 0 ikke befinner seg i intervallet. Dette speiler bra med konklusjonen $\mu_D \neq 0$ som vi fikk fra hypotesetesten. Bare negative verdier tyder også på at differansen må være negativ, som isåfall betyr at tvillingene med biologiske foreldre har lavere IQ enn de med adoptivforeldre.

Vi ser at i denne oppgaven så konkluderer både den tosidige hypotesetesten og konfidensintervallet at $\mu_D \neq 0$. Dette skyldes av at hypotesetesten hadde signifikantnivå 5% og konfidensintervallet var (100% - 5%) = 95%. Dette gjelder generelt hvor en hypotesetest med nivå α tilsvarer et 100(1 - α)% konfidensintervall.

Oppgave 3

a) Vil teste $H_0 : p_1 - p_2 = 0$ mot $H_A : p_1 - p_2 > 0$ med signifikansnivå $\alpha = 0.05$. Antar at målingene kommer fra to tilfeldig, uavhengige populasjoner. Vi ser at $n_1\hat{p_1} = 486$ og $n_2\hat{p_2} = 441$ er begge større enn 10 og at $\hat{p_1} = 0.162$ og $\hat{p_2} = 0.147$. Dette betyr at vi kan benytte to-proporsjon z testen med testobservator:

$$z = \frac{\hat{p_1} - \hat{p_2}}{\sqrt{\hat{p}\hat{q}(1/m + 1/n)}}$$

hvor $\hat{p} = \frac{m}{m+n}\hat{p_1} + \frac{n}{m+n}\hat{p_2}$ og $\hat{q} = 1 - \hat{p}$. Vi finner disse og beregner z med koden:

```
> p1 = 0.162
> p2 = 0.147
> m = 3000
> n = 3000
>
> p = m/(m+n) * p1 + n/(m+n) * p2
> q = 1 - p
>
> z = (p1 - p2) / sqrt(p*q*(1/m + 1/n))
> print(z)
[1] 1.60737
```

Vi får altså $z = 1.607$. Den kritiske z-verdien med $\alpha = 0.05$ finner vi til å være $z_{0.05} = 1.74$. Siden testen vår er upper-tailed ($H_A : p_1 - p_2 > 0$), så er forkastningsområdet $z \geq 1.74$. Ettersom $1.607 < 1.74$, forkaster vi ikke H_0. P-verdien i dette tilfellet tilsvarer $1 - \Phi(1.607)$, som vi finner ved:

```
> p = 2*(1-pnorm(1.607))
> print(p)
[1] 0.1080544
```

Altså P-verdi $= 0.108 > 0.05 = \alpha$. Siden P-verdien er større enn signifikansnivået så forkaster vi ikke H_0 her heller. Testen vår konkluderer altså med at $H_0 : p_1 - p_2 = 0$ stemmer, som betyr at det ikke er noen signifikant forskjell i tidsklemmeproblemer mellom fedre og mødre.

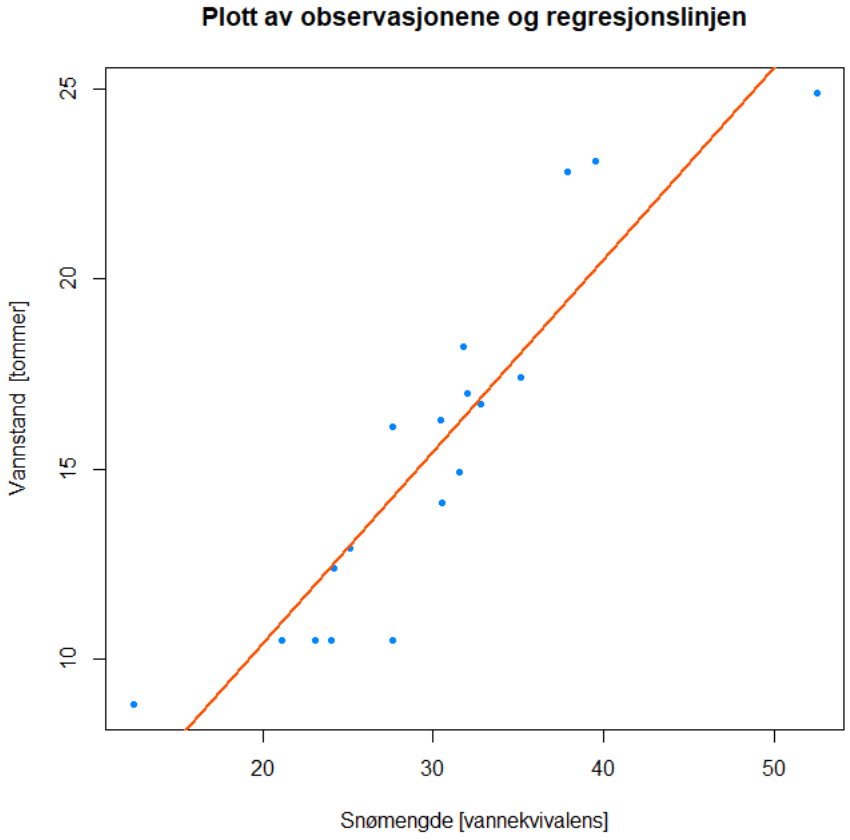
b) Kontrollerer svarene fra a) ved å benytte prop.test() funksjonen i R:

```
> prop.test(x = c(486, 441), n = c(3000, 3000))

        2-sample test for equality of proportions with
    continuity correction

data:  c(486, 441) out of c(3000, 3000)
X-squared = 2.4701, df = 1, p-value = 0.116
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.003619808  0.033619808
sample estimates:
prop 1 prop 2
 0.162  0.147
```

Vi ser at P-verdien her er lik 0.116, som betyr at H_0 ikke forkastes. Dette samsvarer med det vi fikk i a). Legger merke til at P-verdien her er noe høyere enn den vi fikk i a), altså 0.116 mot 0.108. Denne forskjellen skyldes mest sannsynlig av avrundingen av z-verdien fra 1.60737 til 1.607.



Figur 4: *Spredningsplott av målingene (blå prikker) og regresjonslinjen vi fikk ved hjelp av lm() funksjonen (oransje linje).*

Oppgave 4

a) Figur 4 viser spredningsplott av observasjonene sammen med regresjonslinjen. Vi ser at punktene former en rett linje med jevn spredning. Estimatene vi får for koeffisientene fra lm()-funksjonen i R er $\hat{\beta}_0 = 0.28$ og $\hat{\beta}_1 = 0.5056$. At stigningstallet er ca. 0.5 vil si at når snømengden øker én vannekvivalens så øker vannstanden med ca. 1/2 tomme. Gitt at vi ikke vet noe særlig om hvor stor elven er og hvor mye én vannekvivalens tilsvarer i forhold til elven, så er det litt vanskelig å si noe om hvor rimelig disse tallene er. Men de virker ikke helt på bærtur ihvertfall, så vi tar oss til gode med disse verdiene.

```
> url = "https://www.uio.no/studier/emner/matnat/math/STK1110/
    data/snoe_vann.txt"
> data = read.table(url, header=FALSE)
>
```

```
> snö = data$V1
> vann = data$V2
>
> lin.fit = lm(vann ~ snö)
> print(lin.fit)

Call:
lm(formula = vann ~ snö)

Coefficients:
(Intercept)          snö
     0.2800       0.5056


>
> plot(snö, vann,
+      xlab="Snömengde [vannekvivalens]",
+      ylab="Vannstand  [tommer]",
+      main="Plott av observasjonene og regresjonslinjen",
+      pch=20, col="#0084ff")
> abline(lin.fit, col="#ff5100", lwd=2)
```

b) Plotter residualene mot forklaringsvariabelen (snømengde) i Figur 5.

```
> # Finner standarissert residualene og plotter disse
> stan.res = rstandard(lin.fit)
> plot(snö, stan.res,
+      xlab="Snömengde [vannekvivalens]",
+      ylab="Standardissert residualer",
+      pch=20, col="#0084ff")
```

Et normalfordelingsplott av residualene er vist i Figur 6. Vi ser at plottet er veldig rett, som tyder på at normalitet var rett å anta.

```
> qqnorm(stan.res, col="#000000", pch=20)
> qqline(stan.res, col="#314bbe", lwd=2)
```

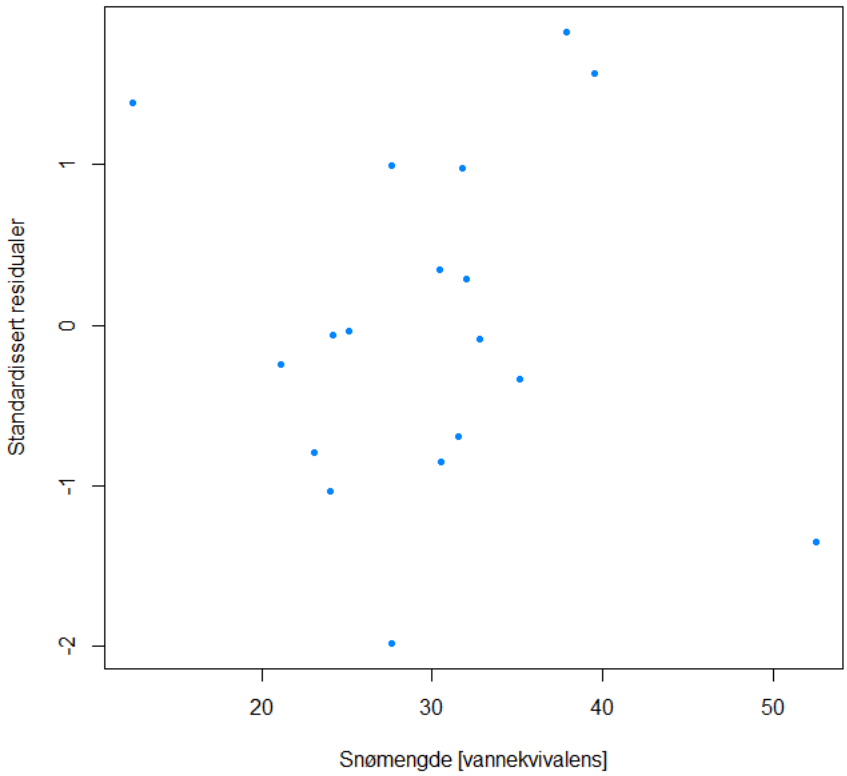
Den linære regresjonsmodellen virker altså godt egnet til målingene våre.

c) Finner et estimat for variansen til feilleddene ved å benytte

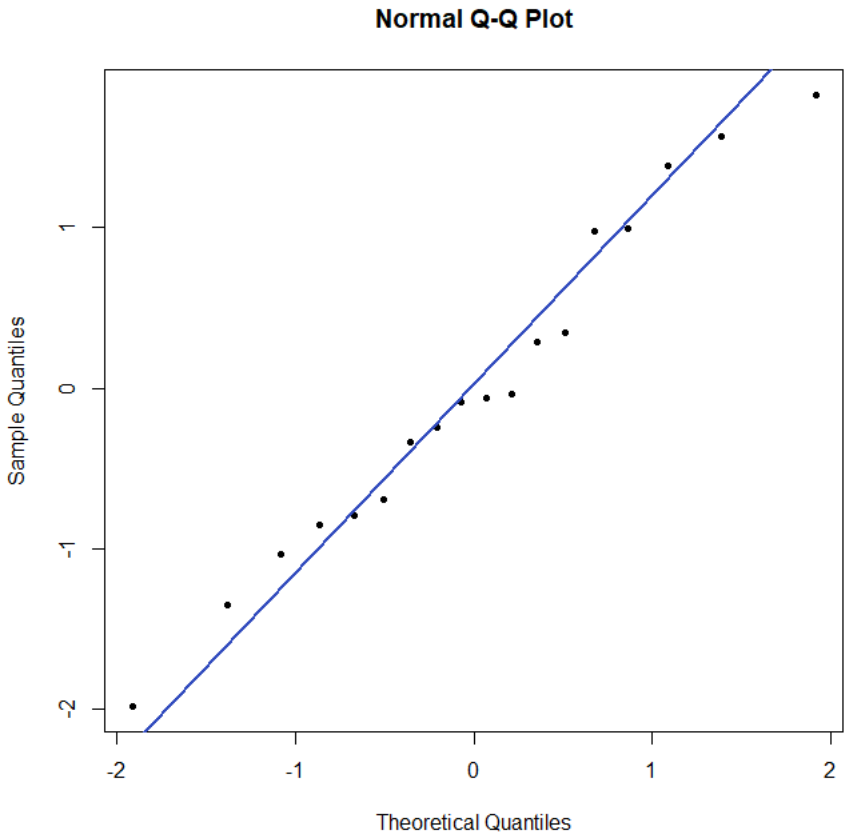
$$s_e^2 = \text{SSE}/(n-2)$$

hvor n er antall målinger og SSE er residdualkvadratsummen $\sum_n (y_i - \hat{y}_i)^2$.

```
> n = length(snö)
> y = vann
> y.hat = lin.fit$fitted.values
> SSE = sum((y - y.hat)^2)
> s.e2 = SSE/(n-2)
> print(s.e2)
[1] 3.774598
```



Figur 5: *Residualplott for målingene. Vi ser at residualene ikke har noe spesielt mønster for seg og er godt fordelt over og under y=0 linjen.*



Figur 6: *Normalfordelingsplott av residualene. Vi ser at punktene følger en rett linje ganske godt.*

Estimatet for variansen er altså $s_e^2 = 3.77$. Lager så et 95% konfidensintervall for stigningstallet β_1 ved å benytte formelen

$$\hat{\beta}_1 \pm t_{\alpha/2,n-2} \cdot \frac{s_e}{\sqrt{S_{xx}}}$$

hvor $S_{xx} = \sum(x_i - \bar{x})^2$ og $\hat{\beta}_1$ er estimatet for stigningstallet som vi fant i a) til å være $\hat{\beta}_1 = 0.5056$. Beregningene blir gjort med følgende koden

```
> x = snö
> beta = 0.5056
> t.crit = qt(0.025, n-2, lower.tail = FALSE)
> Sxx = sum((x - mean(x))^2)
> lower = beta - t.crit * sqrt(s.e2)/sqrt(Sxx)
> upper = beta + t.crit * sqrt(s.e2)/sqrt(Sxx)
> print(c(lower, upper))
[1] 0.3888443 0.6223557
```

Vi kan altså si med 95% sikkerhet at stigningstallet for den lineære relasjonen mellom vannstand og snømengde ligger innenfor intervallet $(0.389, 0.622)$. Legger merke til at det bare inneholder positive verdier, som betyr at vannstanden øker når snømengden øker, noe som virker rimelig.

d) Vil teste $H_0 : \beta_0 = 0$ mot $H_A : \beta_0 \neq 0$ med signifikansnivå 5% ($\alpha = 0.05$). Testobservasjonen for testen er

$$T = \frac{\hat{\beta}_0 - \beta_{00}}{S_{\hat{\beta}_0}} = \frac{\hat{\beta}_0}{S_{\hat{\beta}_0}}$$

hvor $S_{\hat{\beta}_0} = S_e\sqrt{1/n + \bar{x}^2/S_{xx}}$ og $\hat{\beta}_0$ er estimatet for skjæringspunktet som vi fant i a) til å være $\hat{\beta}_0 = 0.28$. Med signifikansnivå 5% så får vi en kritisk t-verdi på $t_{\alpha/2,n-2} = t_{0.025,16} = 2.120$. Forkastningsområdet med $\alpha = 0.05$ blir da:

$$t \leq -2.120 \qquad \text{og} \qquad t \geq 2.120$$

Testobservasjonen som kommer av våre målinger beregner vi ved:

```
> beta0 = 0.28
> s_b0 = sqrt(s.e2)*sqrt(1/n + mean(x)^2/Sxx)
> t = beta0 / s_b0
> print(t)
[1] 0.1635603
```

Vi får altså testobservasjonen $t = 0.164$. Denne verdien ligger ikke i forkastningsområdet, så vi konkluderer med at $H_0 : \beta_0 = 0$ ikke skal forkastes. P-verdien finner vi ved

```
> p = 2*(1-pt(0.164, 16))
> print(p)
[1] 0.871785
```

Altså P-verdi $= 0.87$, som er mye høyere enn signifikansnivået $\alpha = 0.05$. Vi konkluderer også her at H_0 ikke skal forkastes. Som sagt så fikk vi fra a) at $\hat{\beta}_0 = 0.28$ ved hjelp av lm() funksjonen i R:

```
> lin.fit = lm(vann ~ snö)
> print(lin.fit)

Call:
lm(formula = vann ~ snö)

Coefficients:
(Intercept)          snö
     0.2800       0.5056
```

Vi har altså to forskjellige antagelse for $\hat{\beta}_0$. Fra lm() funksjonen får vi $\hat{\beta}_0 = 0.28$, mens fra hypotesetesten får vi at $\hat{\beta}_0 = 0$. Dette virker kanskje litt rart, men hvis vi sammenligner dette med verdiene for vannstand (10.5, 16.7, 18.2, 17.0, ...) og snømengde (23.1, 32.8, 31.8, 32.0, ...), så ser vi at $\hat{\beta}_0 = 0.28$ ikke er så forskjellig fra $\hat{\beta}_0 = 0$. Hypotesetesten vår samsvarer altså godt med resultatene fra a).