

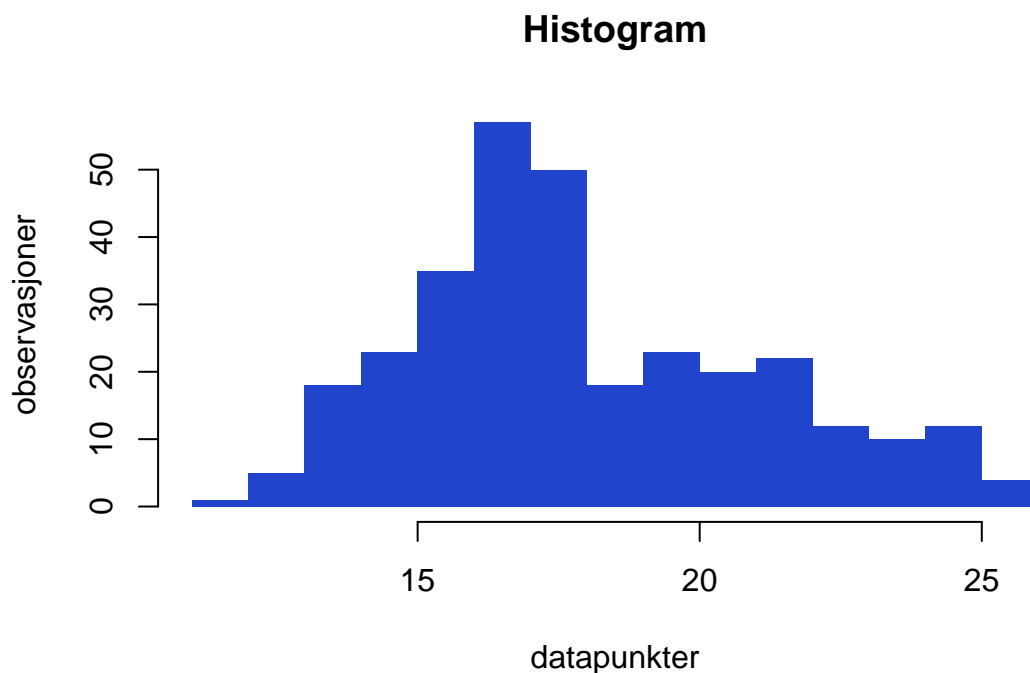
STK1000 - Innlevering 1 (Oppgave 1)

XXXXXXX-XXXXX

A) Histogram

Beskriv kort: hvor ligger mesteparten, og om det er skjevhet/hale.

```
hist(x,  
     col = "#2044cc",  
     xlab = "datapunkter",  
     ylab = "observasjoner",  
     border = NA,  
     breaks = 20,  
     main = "Histogram")
```



Tolkning A (kort): Mest rundt 16–18 C. De fleste dager ca. 14–22 C. Litt høyrehale.

B) Gjennomsnitt og median

Hva betyr tallene? Gjennomsnitt = snitt; median = midtverdi (50% under/over).

```
avgx = mean(x, na.rm = TRUE); avgx  
[1] 18.0371  
medx = median(x, na.rm = TRUE); medx  
[1] 17.5
```

```
min(x, na.rm = TRUE); max(x, na.rm = TRUE)
[1] 11.2
[1] 25.9
```

Tolkning B: Median er mer robust mot uteliggere enn gjennomsnitt.

C) Standardavvik og IQR

Forklar: sd = spredning rundt snittet; IQR = Q3 - Q1 (midterste 50%), robust.

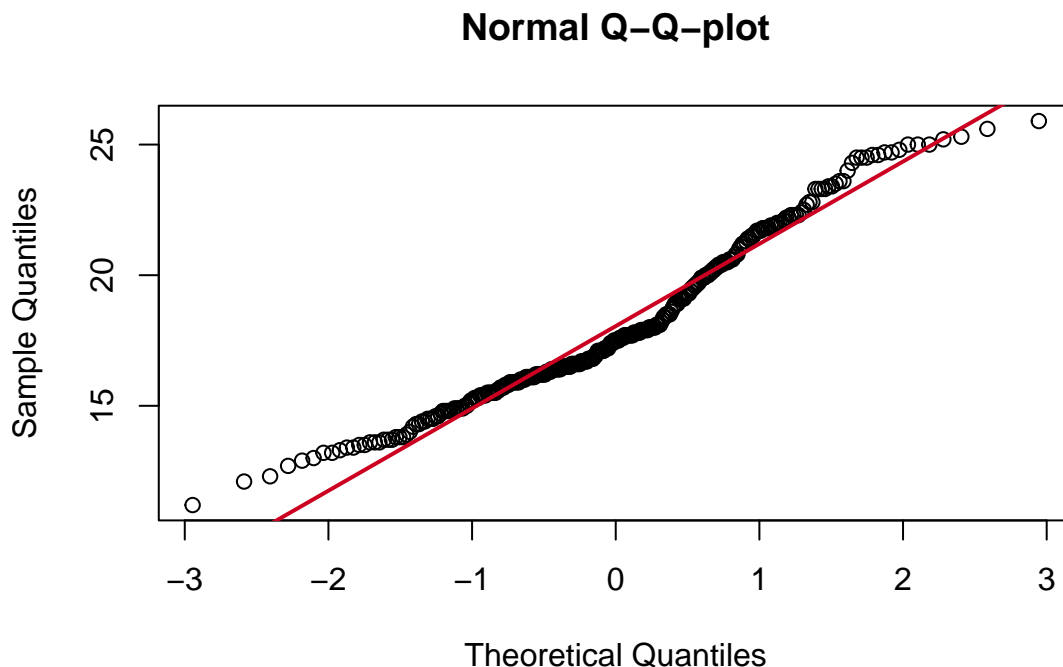
```
s = sd(x, na.rm = TRUE); s
[1] 3.07484
iqr = IQR(x, na.rm = TRUE); iqr
[1] 4.25
summary(x)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 11.20  15.93   17.50   18.04  20.18   25.90
```

Tolkning C: Stor sd = stor variasjon. IQR forteller bredden til midten av fordelingen.

D) Normalfordeling (QQ-plot)

Vurder om punktene følger linja.

```
qqnorm(x, main = "Normal Q-Q-plot")
qqline(x, col = "#cc0020", lwd = 2)
```



Tolkning D: Ligger punktene omtrent langs linja, er normalantakelsen ok (små avvik i halene er greit).

E) Standardisert verdi for 15 C

Formel: $z = (x - \text{gjennomsnitt}) / \text{sd}$. Regn ut for 15 C og tolk.

```
z = (15 - avgx)/s; z  
[1] -0.9877253
```

Tolkning E: $z < 0$ betyr at 15 C ligger under snittet; $|z|$ sier hvor mange sd unna.

F) Andel dager med temp > 20 C

Normalantakelse med `pnorm`, og den faktiske andelen i data.

```
p20 = 1 - pnorm(20, mean = avgx, sd = s); p20  
[1] 0.2616145  
p20_emp = mean(x > 20, na.rm = TRUE); p20_emp  
[1] 0.2580645
```

Tolkning F: $1 - \text{pnorm}(20, \text{mean}, \text{sd})$ gir andel over 20 C hvis vi antar normalfordeling. `p20_emp` er faktisk andel i datasettet.

G) 2018 og 2025 vs. andre år

Del opp data, tell dager > 20 C, og sammenlign.

```
ind = which(data$year %in% c(2018, 2025))  
data_18_25 = data[ind,]  
data_not18_25 = data[-ind,]  
  
n18 = sum(data_18_25$year == 2018); n18  
[1] 31  
over18 = sum(data_18_25$temp > 20 & data_18_25$year == 2018); over18  
[1] 29  
p18 = if (n18 > 0) over18 / n18 else NA; p18  
[1] 0.9354839  
  
n25 = sum(data_18_25$year == 2025); n25  
[1] 31  
over25 = sum(data_18_25$temp > 20 & data_18_25$year == 2025); over25  
[1] 17  
p25 = if (n25 > 0) over25 / n25 else NA; p25  
[1] 0.5483871  
  
n0 = nrow(data_not18_25); n0  
[1] 248  
over0 = sum(data_not18_25$temp > 20); over0  
[1] 34  
p0 = over0 / n0; p0  
[1] 0.1370968
```

Tolkning G: 2018 har mye høyere andel > 20 C enn “andre år” (uvanlig varm). 2025 er også høyere enn “andre år”, men lavere enn 2018.

Oppgave 2 – Ulv

2a) Data og variabler

```
wolfpath = "http://www.uio.no/studier/emner/matnat/math/STK1000/data/obligdata/oblig1/wolves.txt"
wolf = read.table(wolfpath, header=TRUE)
head(wolf, 3)      # sex (F/M), population (1=lett jaktet, 2=tungt), cpmg (kortisol), tpmg (testosteron)
```

	sex	population	cpmg	tpmg
1	M	2	15.86	5.32
2	F	1	20.02	3.71
3	F	2	9.95	5.30

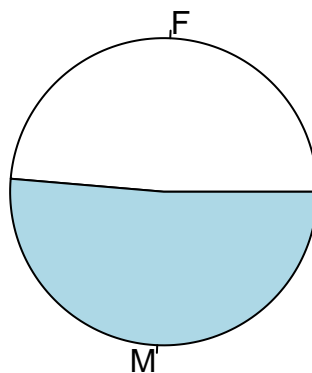
2b) Oppsummer kategoriske variabler (table + pie)

```
table(wolf$sex)
```

	F	M
	72	76

```
pie(table(wolf$sex), main = "Kjønn")
```

Kjønn

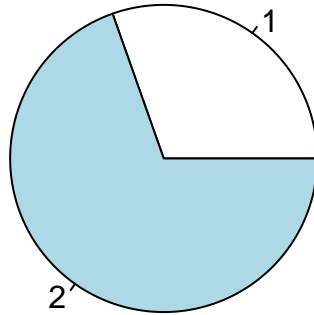


```
table(wolf$population)  # 1 = lett jaktet, 2 = tungt jaktet
```

	1	2
	45	103

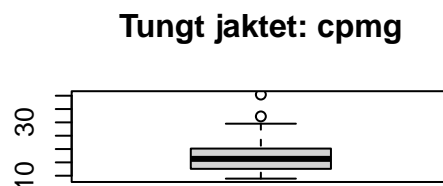
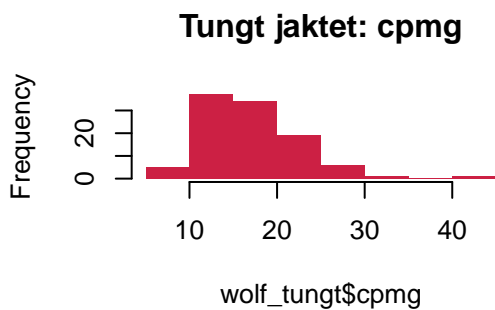
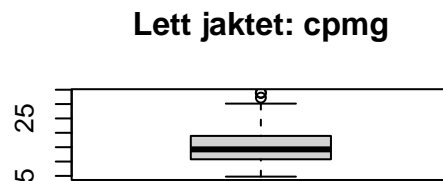
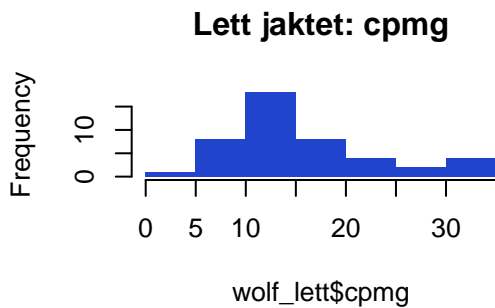
```
pie(table(wolf$population), main = "Jaktnivå (1=lett, 2=tungt)")
```

Jaktnivå (1=lett, 2=tungt)



2c) Del opp data og plott kortisol (cpmg)

```
wolf_lett = wolf[wolf$population==1,]  
wolf_tungt = wolf[wolf$population==2,]  
  
par(mfrow=c(2,2))  
hist(wolf_lett$cpmg, main="Lett jaktet: cpmg", col="#2044cc", border=NA)  
boxplot(wolf_lett$cpmg, main="Lett jaktet: cpmg")  
hist(wolf_tungt$cpmg, main="Tungt jaktet: cpmg", col="#cc2044", border=NA)  
boxplot(wolf_tungt$cpmg, main="Tungt jaktet: cpmg")
```



```
par(mfrow=c(1,1))
```

2d) Gjennomsnitt og median av cpmg (per jaktnivå) + kort sammenligning

```
m_lett = mean(wolf_lett$cpmg, na.rm=TRUE); md_lett = median(wolf_lett$cpmg, na.rm=TRUE)
m_tungt = mean(wolf_tungt$cpmg, na.rm=TRUE); md_tungt = median(wolf_tungt$cpmg, na.rm=TRUE)

m_lett; md_lett
[1] 15.56222
[1] 14.24
m_tungt; md_tungt
[1] 17.07495
[1] 16.32
```

Kort tolkning 2d:

- Lett jaktet: 15.56 vs 14.24 → gjennomsnitt > median (høyreskjev) .
- Tungt jaktet: 17.07 vs 16.32 → gjennomsnitt > median (høyreskjev) .
- Sammenlignet: tungt jaktet har høyere nivå enn lett jaktet.

2e) Hvilken oppsummering passer best?

```
sd_lett = sd(wolf_lett$cpmg, na.rm=TRUE); iqr_lett = IQR(wolf_lett$cpmg, na.rm=TRUE)
sd_tungt = sd(wolf_tungt$cpmg, na.rm=TRUE); iqr_tungt = IQR(wolf_tungt$cpmg, na.rm=TRUE)

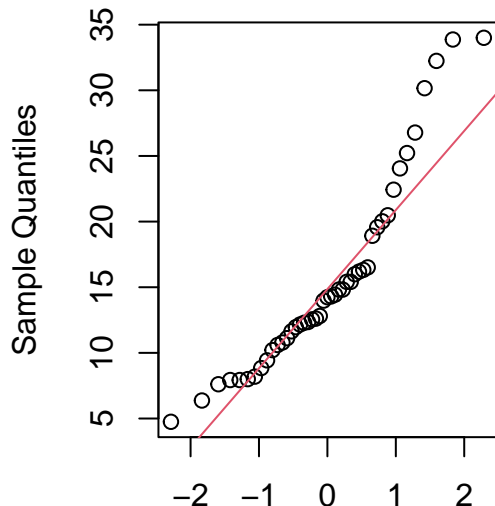
sd_lett; iqr_lett
[1] 7.298785
[1] 8.13
sd_tungt; iqr_tungt
[1] 5.543389
[1] 7.605
```

Kort svar 2e: Begge gruppene viser skjevhet/uteliggere i plott og forskjell mellom gjennomsnitt/median → bruk femtallsoppsummering (median + IQR) for å beskrive kortisolnivået i både lett og tungt jaktet.

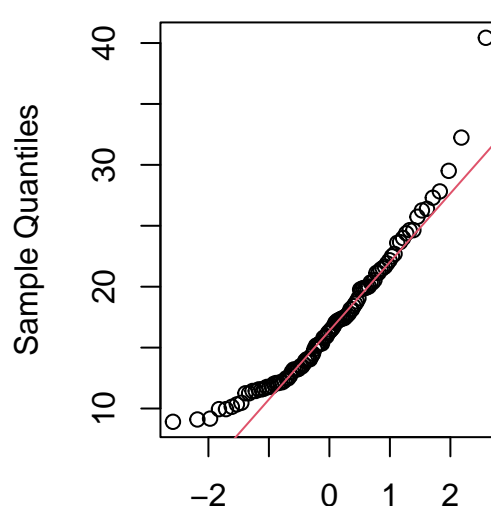
2f) Normalitet av cpmg (per jaktnivå)

```
par(mfrow=c(1,2))
qqnorm(wolf_lett$cpmg, main="QQ-plot: cpmg (lett)"); qqline(wolf_lett$cpmg, col=2)
qqnorm(wolf_tungt$cpmg, main="QQ-plot: cpmg (tungt)"); qqline(wolf_tungt$cpmg, col=2)
```

QQ-plot: cpmg (lett)



QQ-plot: cpmg (tungt)



```
par(mfrow=c(1,1))

p_lett = shapiro.test(wolf_lett$cpmg)$p.value
p_tungt = shapiro.test(wolf_tungt$cpmg)$p.value

p_lett; p_tungt
[1] 0.0007833313
[1] 3.005861e-05
```

Konklusjon 2f:

- Man kan se tydelige avvik i begge halene som vil antyde at det ikke er en normalfordeling når det kommer til starten og slutten.

Oppgave 3 – Månedstemperatur (august vs. september)

3a) Les data, opprette to datasett deretter forklare

```
mpath = "https://www.uio.no/studier/emner/matnat/math/STK1000/data/obligdata/oblig1/blindernmonthly.txt"
m = read.table(mpath, header=TRUE, sep=";")

data8 = m[m$month==8, ] # august
data9 = m[m$month==9, ] # september

summary(data8$temp)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 13.10  15.32  16.10  16.14  16.70  20.50
summary(data9$temp)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  8.80  10.85  12.00  11.78  12.57  15.00

meanAug = mean(data8$temp, na.rm=TRUE)
medianAug = median(data8$temp, na.rm=TRUE)
iqrAug = IQR(data8$temp, na.rm=TRUE)
```

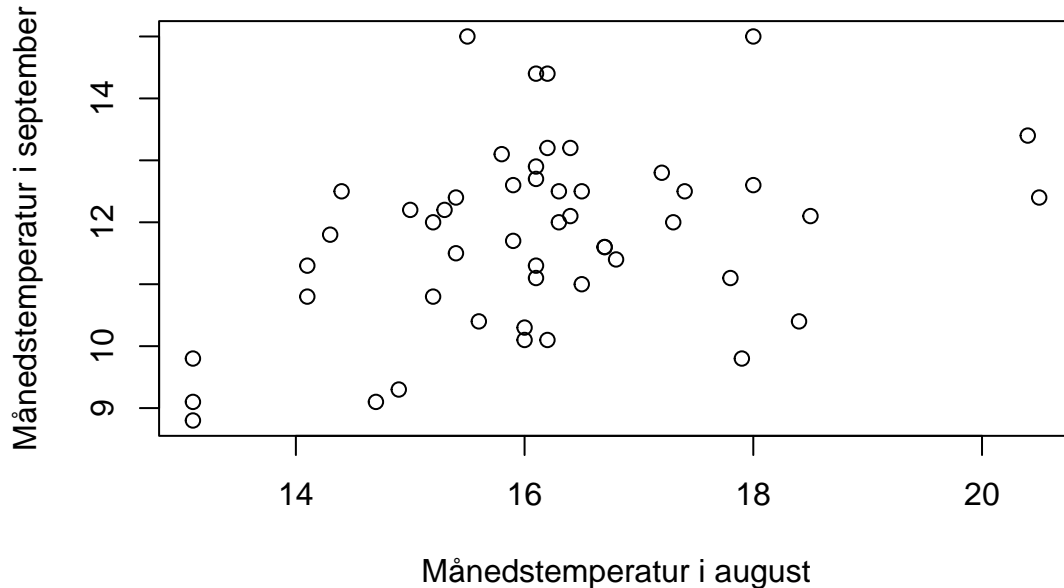
```
meanSep = mean(data9$temp, na.rm=TRUE)
medianSep = median(data9$temp, na.rm=TRUE)
iqrSep = IQR(data9$temp, na.rm=TRUE)
```

Konklusjon 3a: August er generelt varmere enn september. Medianen i august = 16.1 grader og medianen i september = 12 grader.

3b) Spredningsplott (august på x, september på y)

```
augTemp = data8$temp
septTemp = data9$temp

plot(augTemp, septTemp,
     xlab="Månedstemperatur i august",
     ylab="Månedstemperatur i september",
     main="")
```



Konklusjon 3b: Kan konkludere på generelt plan av månedstemperaturen i august er høyere enn september og at de samles rundt 14-18 grader området og generelt er positivt korrelerte.

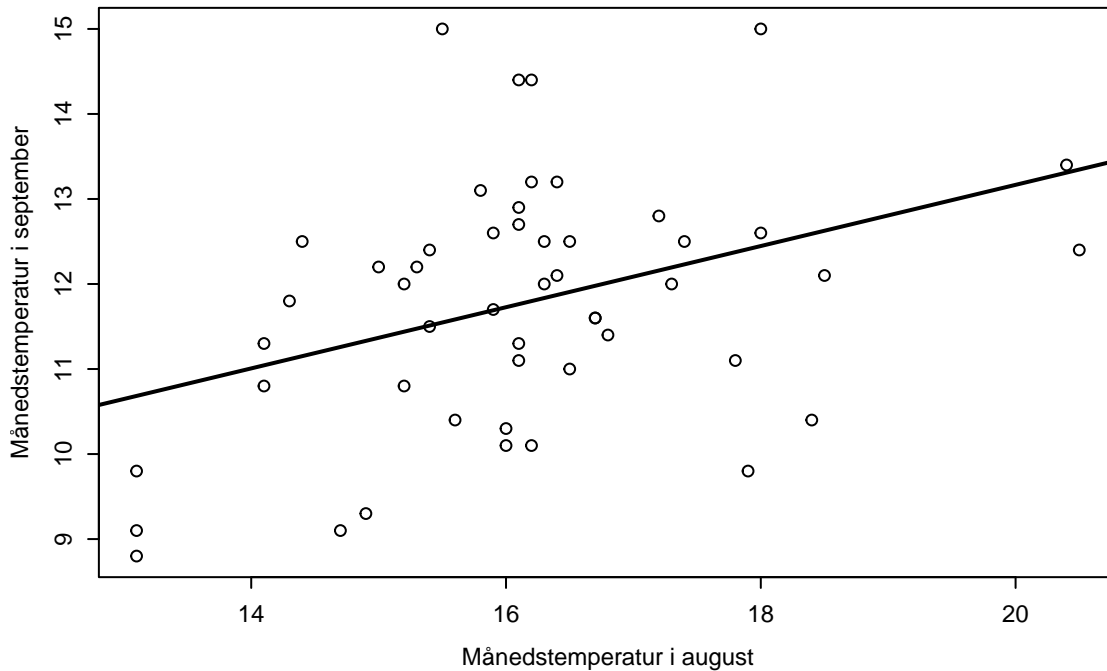
3c) Korrelasjon

```
r = cor(augTemp, septTemp, use="complete.obs")
r
[1] 0.383473
```

Konklusjon 3c: $r = 0.38$. Tolkning: svak positiv sammenheng, hvor korrelasjon er hvor mye de svinger/beveger seg sammen som et mål fra -1 til 1.

3d) Lineær modell og regresjonslinje (nytt plott + linje)

```
par(mex=0.75, cex=0.75)
plot(augTemp, septTemp,
     xlab="Månedstemperatur i august",
     ylab="Månedstemperatur i september",
     main="")
fit = lm(septTemp ~ augTemp)
abline(fit, lwd=2)
```



```
summary(fit)

Call:
lm(formula = septTemp ~ augTemp)

Residuals:
    Min       1Q   Median       3Q      Max
-2.6106 -0.9010  0.0495  0.8059  3.4530

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.9697     2.0282   2.943  0.00499 **
augTemp        0.3598     0.1251   2.877  0.00598 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.359 on 48 degrees of freedom
Multiple R-squared:  0.1471,    Adjusted R-squared:  0.1293
F-statistic: 8.275 on 1 and 48 DF,  p-value: 0.005979
```

Konklusjon 3d: Linja heller oppover, modellen fanger den positive sammenhengen.

3e) Skjæringspunkt og stigningstall

```
b0 = coef(fit)[1]    # skjæringspunkt
b1 = coef(fit)[2]    # stigningstall
b0; b1
(Intercept)
  5.969675
augTemp
  0.3598269
```

Konklusjon 3e: Skjæringspunkt $b_0 = 5.97$ (modellens september når august = 0). Stigningstall $b_1 = 0.36$; en økning på 1 grad i august gir i snitt 0.36 grader høyere i september. Intercepten er hvor den skjærer i Y-aksen altså hvor det starter, så når modellen settes i 0 så starter den på ca 6 grader.

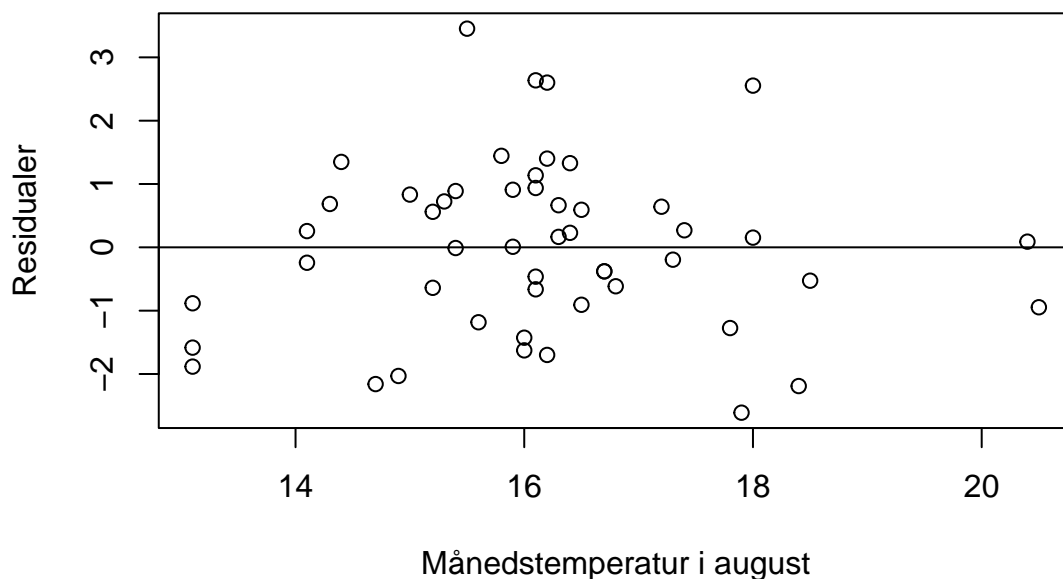
3f) Forklart variasjon (R^2) og sammenligning med r^2

```
R2 = summary(fit)$r.squared
R2
[1] 0.1470516
r^2
[1] 0.1470516
```

Konklusjon 3f: $R^2 = 14.7$ prosent av variasjonen i september som forklares av august. De stemmer greit kan man si i enkel lineær regresjon og fra hva man ser i c på sånn ca 0.38. Ved R^2 på 95 og opp er signifikant og alt under 0.4 er fortsatt ikke noe særlig men det er noe.

3g) Residualplott

```
plot(augTemp, residuals(fit),
     xlab="Månedstemperatur i august", ylab="Residualer")
abline(h=0)
```



Konklusjon 3g: Residualene ligger rundt 0 uten tydelig kurve eller trend. Spredningen virker omtrent jevn.

Modellen (rett linje) er ok, men ikke perfekt. Man kan med dette at det er mye variasjon og dermed ikke mye som forklares så modellen er ikke direkte veldig handlingsdyktig eller forkalrende.

3h) Prediksjon når august = 16.9

```
pred = predict(fit, newdata=data.frame(augTemp=16.9))
pred
      1
12.05075
```

Konklusjon 3h: Predikert september ved august 16.9 er 12.1 grader. Vanlig september (gjennomsnitt) er ca. 11.8 grader, så dette er varmere enn vanlig.