Obligatorisk innlevering 1 – STK1000 – Geologi versjon

Oppgave 1

a)

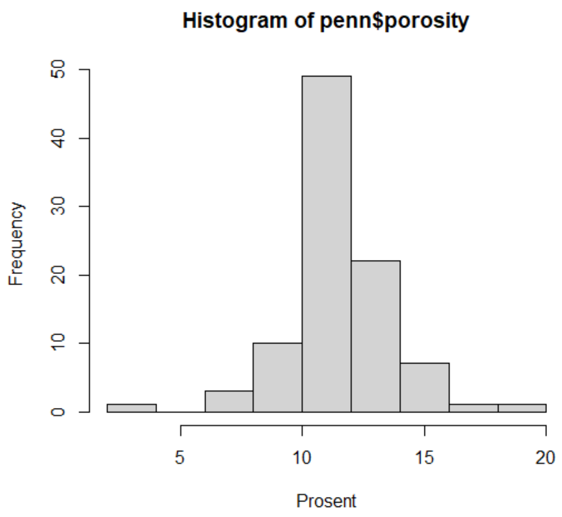
R-kode:

```
#Henter inn dataen og legger dataen i datasettet penn
data = "http://www.uio.no/studier/emner/matnat/math/STK1000/data/obligdata/oblig1/pennsylvania.txt"
penn <- read.table(data, header=TRUE)
#Plotter et histogram over porosity, som gir et plott over frekvensen
hist(penn$porosity, xlab='Prosent')
```

Output fra R/ kjøreeksempel:

```
> #Plotter et histogram over porosity, som gir et plott over frekvensen
> hist(penn$porosity, xlab='Prosent')
```

Plottet fra R er inkludert i Figur 1.



Figur 1: Plottet et histogram av porosity

Vi ser fra histogrammet at av de fleste målingene gjort, så har den høyeste frekvensen av målinger en porøsitet mellom ca. 10 og 14%. Histogrammet er også unimodalt, dvs at det har en topp. Den er også tilnærmet symmetrisk.

b)

R-kode:

```
#Regner ut gjennomsnittet av alle porosity målingene
mean(penn$porosity)
#Regner ut medianen av alle porosity målingene
median(penn$porosity)
```

Output fra R/ kjøreeksempel:

```
> #Regner ut gjennomsnittet av alle porosity målingene
> mean(penn$porosity)
[1] 11.57128
> #Regner ut medianen av alle porosity målingene
> median(penn$porosity)
[1] 11.4
```

For variabelen porosity beregner vi gjennomsnittet, som er 11.57%. Dette betyr at av de 94 målingene gjort, vil en gjennomsnittlig bergart ha en andel hulrom på 11.57%.

For variabelen porosity måler vi også medianen, som er 11.4%. Dette betyr at av de 94 målingene gjort, vil like mange individer ha porosity verdier på under 11.4% som over 11.4%.

c)

R-kode:

```
#Regner ut standardavviket av alle porosity mÃ¥lingene
sd(penn$porosity)
#Regner ut interkvartile avstand (IQR) av mÃ¥lingene
IQR(penn$porosity)
```

Output fra R/ kjøreeksempel:

```
> #Regner ut standardavviket av alle porosity mÃ¥lingene
> sd(penn$porosity)
[1] 2.059106
> #Regner ut interkvartile avstand (IQR) av mÃ¥lingene
> IQR(penn$porosity)
[1] 2.15
```

Standardavviket er 2.05%, og den interkvartile avstanden er 2.15%.

Standardavviket forteller at den gjennomsnittlige avstanden fra gjennomsnittet på 11.57% er 2.05%.

Den interkvartile avstanden forteller at avstanden mellom tredje og første kvartil er 2.15%, som betyr at av de ca. 50% av alle målingene hadde en spredning på 2.15%.

d)

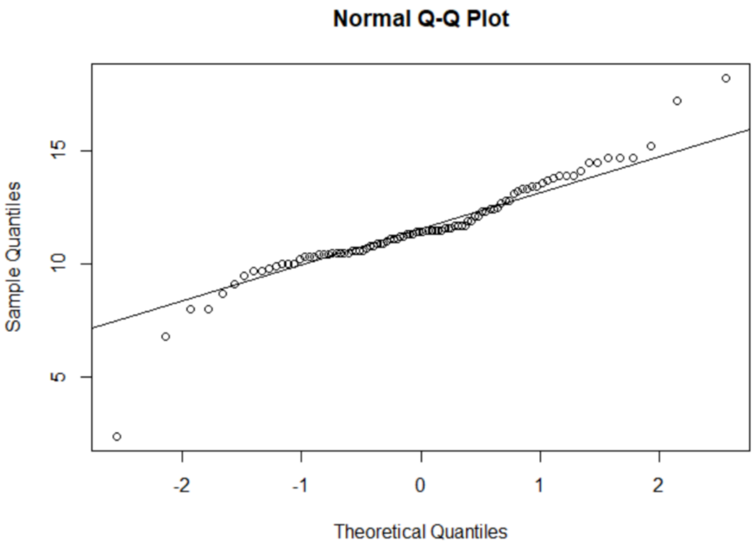
R-kode:

```
#Sjekker om porosity er normalfordelt ved qqnorm() og qqline()
qqnorm(penn$porosity)
qqline(penn$porosity)
```

Output fra R/ kjøreeksempel:

```
> #Sjekker om porosity er normalfordelt ved qqnorm()
> qqnorm(penn$porosity)
> qqline(penn$porosity)
```

Plottet fra R er inkludert i Figur 2.



Figur 2: plottet qqplot over porosity

Normalfordelingsplottet viser at porosity har en tilnærmet rett linje (lineær vekst). Dette tyder sterkt på at porosity er normalfordelt. Ser vi tilbake på Figur 1, kan vi se en tilnærmet symmetrisk histogram. Det er dermed rimelig å anta at porosity er normalfordelt.

e)

R-kode:

```r
#Gir uttrykk for den standardiserte verdien av porosity, der x er porosity verdier
(x - mean(penn$porosity))/sd(penn$porosity)
#Regner ut den standardiserte verdien av porosity = 14%
(14 - mean(penn$porosity))/sd(penn$porosity)
```

Output fra R/ kjøreeksempel:

```
> #Gir uttrykk for den standardiserte verdien av porosity, der x er porosity verdier
> (x - mean(penn$porosity))/sd(penn$porosity)
Error: object 'x' not found
> #Regner ut den standardiserte verdien av porosity = 14%
> (14 - mean(penn$porosity))/sd(penn$porosity)
[1] 1.179504
```

Den standardiserte verdien (z-skår) til porosity er 1.17. Dette forteller oss at en porosity observasjon på 14% har en z-skår på 1.17 som er positiv. En positiv z-skår betyr at porosity observasjonen på 14% er større enn forventningsverdien for målingene (gjennomsnittet av målingene). Dermed er målingen 1.17 større enn gjennomsnittlig verdi. Dermed er denne målingen høyere enn forventningen.

f)

R-kode:

```r
#Regner ut andelen av boreprøver som har porøsitet lavere enn 8% med pnorm()
pnorm(8, mean(penn$porosity), sd(penn$porosity), lower.tail = TRUE, log.p = FALSE)
```

Output fra R/ kjøreeksempel:

```
> #Regner ut andelen av boreprøver som har porøsitet lavere enn 8% med pnorm()
> pnorm(8, mean(penn$porosity), sd(penn$porosity), lower.tail = TRUE, log.p = FALSE)
[1] 0.04142518
```

Fra pnorm() får vi svaret 0.0414, som gir en andel på 4.14% av boreprøvene. Det betyr at det finnes få boreprøver med porøsitet under 8%, det kan vi også se i datasettet vårt.

g)

R-kode:

```
#Regner ut andelen av boreprøver som har porøsitet høyere enn 15%
pnorm(15, mean(penn$porosity), sd(penn$porosity), lower.tail = FALSE, log.p = FALSE)
```

Output fra R/ kjøreeksempel:

```
> #Regner ut andelen av boreprøver som har porøsitet høyere enn 15%
> pnorm(15, mean(penn$porosity), sd(penn$porosity), lower.tail = FALSE, log.p = FALSE)
[1] 0.04794129
```

Fra pnomr() får vi svaret 0.0479, som gir en andel på 4.79% av boreprøvene. Det betyr at det finnes få boreprøver med porisøsitet over 15%, som vi også kan se i datasettet vårt

Oppgave 2

a)

En kategorisk variabel plasserer observasjoner i en eller flere grupper, også kategorier. En kvantitativ variabel er en numerisk verdi som det er mulig å utføre aritmetiske operasjoner på. Ser vi på datasettet med kjerneprøvene vi har nå hentet inn, er *depth* en kategorisk variabel. Dette fordi den plasserer hver observasjon under kategorien tatt på dypt vann eller ikke tatt på dypt vann. *Weight* beskriver vekten av kjerneprøven i gram, og *sand* beskriver prosentandel sand i kjerneprøven. Disse to er derfor kvantitative variabler ettersom de er numeriske verdier.

b)

R-kode:

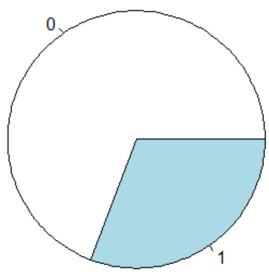
```
data="http://www.uio.no/studier/emner/matnat/math/STK1000/data/obligdata/oblig1/kjerneprover.txt"
kjerneprover <- read.table(data,header=TRUE)
#Lager en oppsummering av den kategoriske variabelen depth
#Først et table
table(kjerneprover$depth)
#Deretter en pie chart av tablet vårt
pie(table(kjerneprover$depth))
```

Output fra R/ kjøreeksempel:

```
> data="http://www.uio.no/studier/emner/matnat/math/STK1000/data/obligdata/oblig1/kjerneprover.txt"
> kjerneprover <- read.table(data,header=TRUE)
> #Lager en oppsummering av den kategoriske variabelen depth
> #Først et table
> table(kjerneprover$depth)

  0   1
137  61
> #Deretter en pie chart av tablet vårt
> pie(table(kjerneprover$depth))
```

Plottet fra R er inkludert i Figur 3.



Figur 3: kakediagram over kjerneprøvene, enten om de er tatt på dypt vann (1) eller ikke tatt på dypt vann (0)

Vi ser fra kakediagrammet at det er flere prøver som ikke er tatt på dypt vann, kontra de som er tatt på dypt vann. Det er dermed flere 0-verdier av *depth*.

c)

R-kode:

```
#Deler opp dataene i to datasett, ett som inneholder prøvene tatt på dypt vann
kjerneprover.dypt <- kjerneprover[kjerneprover[, "depth"]==1,]
#Og ett som inneholder prøvenes som ikke er tatt på dypt vann
kjerneprover.grunt <- kjerneprover[kjerneprover[, "depth"]==0,]
```

Output fra R/ kjøreeksempel:

```
> #Deretter en pie chart av tablet vårt
> pie(table(kjerneprover$depth))
> #Deler opp dataene i to datasett, ett som inneholder prøvene tatt på dypt vann
> kjerneprover.dypt <- kjerneprover[kjerneprover[, "depth"]==1,]
> #Og ett som inneholder prøvenes som ikke er tatt på dypt vann
> kjerneprover.grunt <- kjerneprover[kjerneprover[, "depth"]==0,]
```

Her er dataene delt i to datasett, ett for prøvene tatt på dypt vann og de som ikke er tatt på dypt vann.

d)

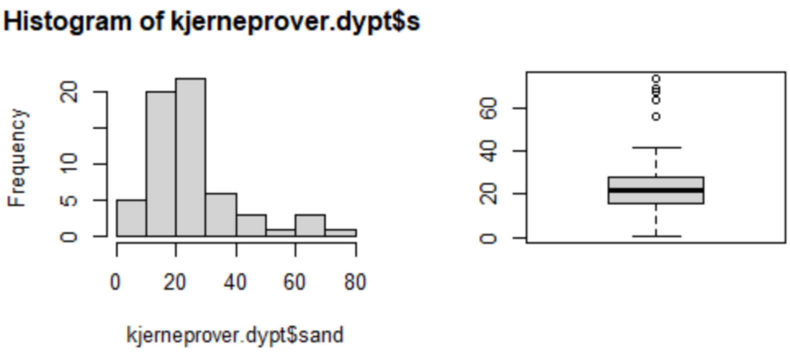
R-kode:

```
par(mfrow=c(2,2))
#Histogram for kjerneprøver tatt i dypt vann
hist(kjerneprover.dypt$sand)
#Boksplott for kjerneprøver tatt i dypt vann
boxplot(kjerneprover.dypt$sand)
#Histogram for kjerneprøver som ikke er tatt i dypt vann
hist(kjerneprover.grunt$sand)
#Boksplott for kjerneprøver som ikke er tatt i dypt vann
boxplot(kjerneprover.grunt$sand)
```

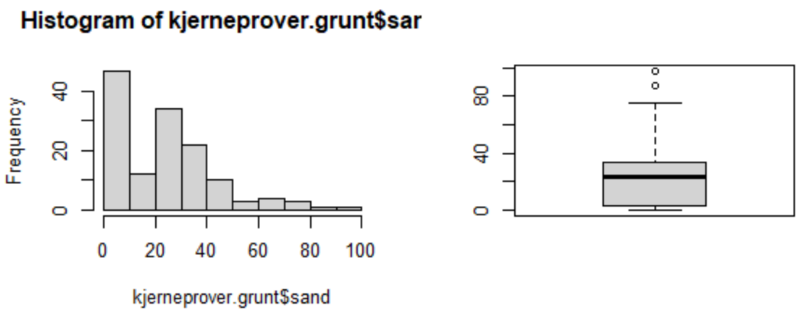
Output fra R/ kjøreeksempel:

```
> par(mfrow=c(2,2))
> #Histogram for kjerneprøver tatt i dypt vann
> hist(kjerneprover.dypt$sand)
> #Boksplott for kjerneprøver tatt i dypt vann
> boxplot(kjerneprover.dypt$sand)
> #Histogram for kjerneprøver som ikke er tatt i dypt vann
> hist(kjerneprover.grunt$sand)
> #Boksplott for kjerneprøver som ikke er tatt i dypt vann
> boxplot(kjerneprover.grunt$sand)
```

Plottet fra R er inkludert i Figur 4 og Figur 5.



Figur 4: Histogram og boksplott for kjerneprøver tatt i dypt vann



Figur 5: Histogram og boksplott for kjerneprøver som ikke er tatt i dypt vann

Fra Figur 4 kan vi se at kjerneprøvene tatt i dypt vann har en høyere frekvens av prøver med 10-30% andel sand,. Fra Figur 5 kan vi se at kjerneprøvene tatt i dypt vann har en høyere frekvens av prøver med 0-10% andel sand, men det er også flere topper ved 20-40% andel sand.

Boksplottet i Figur 4 viser at toppunktet for prøvene tatt i dypt vann er mer en uteligger, enn toppunktet for boksplottet i Figur 5. Det er også en større interkvartil avstand i boksplottet for prøver som ikke er tatt i dypt vann. Men begge boksplott i Figur 4 og Figur 5, har veldig lik median.

e)

R-kode:

```
#Finner gjennomsnitt og median av andel sand i prøvene for begge datasett
summary(kjerneprover.dypt$sand)
summary(kjerneprover.grunt$sand)
```

Output fra R/ kjøreeksempel:

Bruker femtallsoppsummeringen, så slipper jeg å lage median() og mean() for begge datasett.

```
> #Finner gjennomsnitt og median av andel sand i prøvene for begge datasett
> summary(kjerneprover.dypt$sand)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
   0.64   15.53   21.71   25.02   27.66   73.80 
> summary(kjerneprover.grunt$sand)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
   0.02    3.17   23.46   22.97   33.12   97.86 
```

Gjennomsnittet for andel sand av prøvene tatt i dypt vann er 25.02, og 22.97 for de prøvene som ikke er tatt i dypt vann. Dette viser at det er mer sand i prøvene tatt i dypt vann enn de som ikke er tatt i dypt vann.

Medianen for andel sand av prøvene tatt i dypt vann er 21.71, og 23.46 for de prøvene som ikke er tatt i dypt vann. Dette viser at halvparten av observasjonene er over 23.46 i ikke dypt vann, og er dermed større enn halvparten av de større enn 21.71 tatt i dypt vann.

f)

R-kode:

```
#Tar summary av begge datasett
summary(kjerneprover.dypt$sand)
summary(kjerneprover.grunt$sand)
```

```
#Finner gjennomsnitt og standardavvik
mean(kjerneprover.dypt$sand)
mean(kjerneprover.grunt$sand)
sd(kjerneprover.dypt$sand)
sd(kjerneprover.grunt$sand)
```

Output fra R/ kjøreeksempel:

```
> #Tar summary av begge datasett
> summary(kjerneprover.dypt$sand)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
   0.64   15.53   21.71   25.02   27.66   73.80 
> summary(kjerneprover.grunt$sand)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
   0.02    3.17   23.46   22.97   33.12   97.86 
```

```
> #Finner gjennomsnitt og standardavvik
> mean(kjerneprover.dypt$sand)
[1] 25.01574
> mean(kjerneprover.grunt$sand)
[1] 22.97161
> sd(kjerneprover.dypt$sand)
[1] 15.49115
> sd(kjerneprover.grunt$sand)
[1] 20.37242
```

Fra femtallsoppsummeringen for dypt vann kan vi se at det er stort sprik i målingene, helt fra minste verdi 0.64 til max verdi 73.80. Fra gjennomsnittet og standardavviket for dypt vann, får vi ikke fram det store spriket unntatt fra standardavviket på 15.49. Men det er en enklere måte å få oversikt på at målingene har et høyt standardavvik, og vil derfor også kanskje ha et stort sprik. Jeg vil derfor si at femtallsoppsummeringen er en mer oversiktlig måte å se spriket og de høyeste verdiene i målingene på.

Fra femtallsoppsummeringen for grunt vann kan vi se enda et stort sprik i målingene. Det kan også sees i standardavviket, men standardavviket får ikke fram uteliggere. For femtallsoppsummeringen kan vi se at vi har målinger helt fra 0.02 til 97.86, som er et stort sprik. Dermed vil jeg også benyttet femtallsoppsummeringen som en mer oversiktlig måte å utforske målingene på.

g)

R-kode:

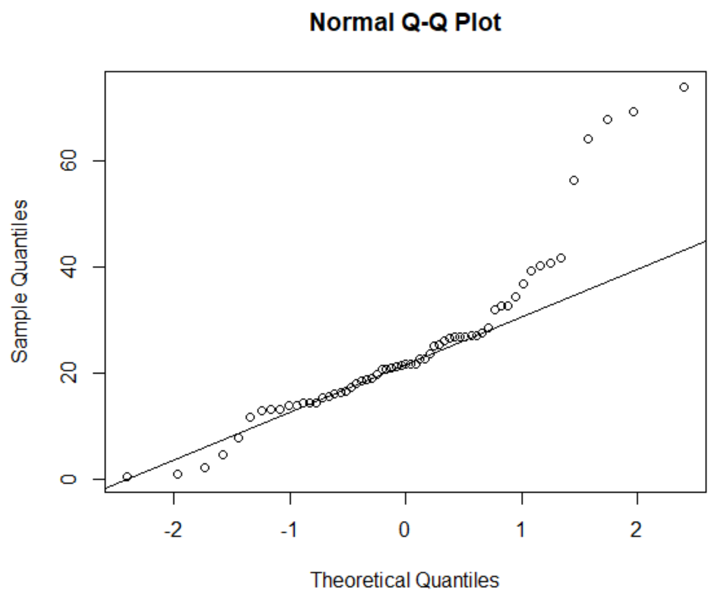
```
#Sjekker først med qqnorm på prøvene tatt på dypt vann
qqnorm(kjerneprover.dypt$sand)
qqline(kjerneprover.dypt$sand)


#Sjekker med qqnorm på prøvene som ikke er tatt på dypt vann
qqnorm(kjerneprover.grunt$sand)
qqline(kjerneprover.grunt$sand)
```

Output fra R/kjøreeksempel:

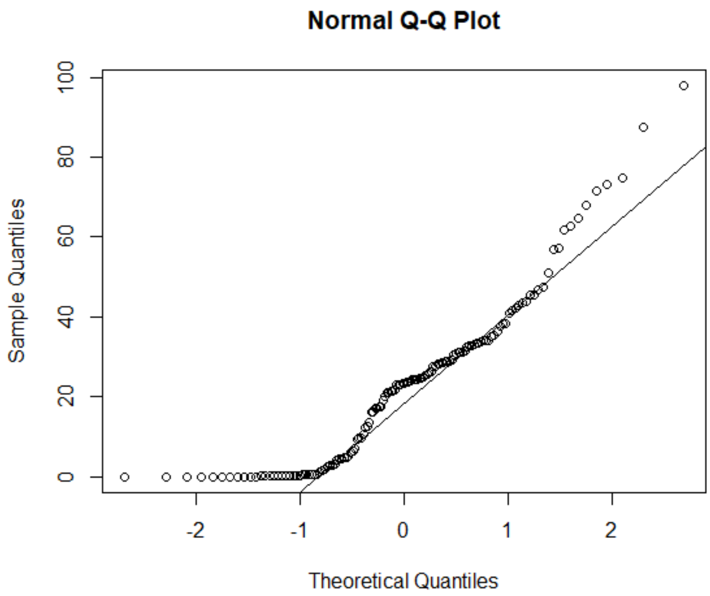
```
> #Sjekker først med qqnorm på prøvene tatt på dypt vann
> qqnorm(kjerneprover.dypt$sand)
> qqline(kjerneprover.dypt$sand)
>
>
> #Sjekker med qqnorm på prøvene som ikke er tatt på dypt vann
> qqnorm(kjerneprover.grunt$sand)
> qqline(kjerneprover.grunt$sand)
```

Plottet fra R er inkludert i Figur 6 og 7.



Figur 6: Normalfordelingsplott for prøver tatt på dypt vann

For prøvene som er tatt på dypt vann i Figur 6, kan vi se at det bare er de midterste målingene som holder seg til den lineære linjen i normalfordelingsplottet, og danner et slags toppunkt. Man ser også at det er flere verdier som er godt over linjen. Det er derfor ikke rimelig å betegne datasettet for andel sand i prøver tatt på dypt vann som normalfordelt.



Figur 7: Normalfordelingsplitt for prøver som ikke er tatt på dypt vann

For prøvene som ikke er tatt på dypt vann i Figur 7, kan vi se at flesteparten av målingene holder seg til den lineære linjen i normalfordelingsplottet. Man ser også at det få verdier som er over linjen. Det er derfor rimelig å betegne datasettet for andel sand i prøver som ikke er tatt på dypt vann som normalfordelt.

Oppgave 3

a)

R-kode:

```
data = "http://www.uio.no/studier/emner/matnat/math/STK1000/data/obligdata/oblig1/vitruvisk.txt"
vitruvisk <- read.table(data,header=TRUE)

#Bruker table funksjonen for å finne antallet av hver kategorisk variabel
table(vitruvisk$kjonn)
#Bruker summary for fem-punkts oppsummering for kroppslengde og fot.navle
summary(vitruvisk)
```

Output fra R/kjøreeksempel:

```
> #Bruker table funksjonen for å finne antallet av hver kategorisk variabel
> table(vitruvisk$kjonn)

  K   M
150  73
> #Bruker summary for fem-punkts oppsummering for kroppslengde og fot.navle
> summary(vitruvisk)
    kjonn             kroppslengde      fot.navle         navle.isse         favn
 Length:223         Min.   :152.0   Min.   : 87.0   Min.   :52.00   Min.   :146.0
 Class :character   1st Qu.:166.0   1st Qu.:101.0   1st Qu.:65.00   1st Qu.:165.0
 Mode  :character   Median :172.0   Median :104.0   Median :67.00   Median :171.0
                    Mean   :172.3   Mean   :104.8   Mean   :67.34   Mean   :172.4
                    3rd Qu.:178.0   3rd Qu.:109.0   3rd Qu.:70.00   3rd Qu.:180.0
                    Max.   :196.0   Max.   :125.0   Max.   :81.00   Max.   :202.0
```

Fra kjøringen ser vi at det ble utført målinger på 150 kvinner og 73 menn. Av en eller annen merkelig grunn ville ikke R gi meg denne totalen ved bruk av summary(). Men det fungerte ved bruk av table().

Fem-punkts-oppsummeringene for kroppslengde ser vi i outputet fra kjøreeksempelet. Dette svarer for seg selv i kjøreeksempelet.

Fem-punkts-oppsummeringene for fot.navle ser vi i outputet fra kjøreeksempelet. Dette svarer for seg selv i kjøreeksempelet.

b)

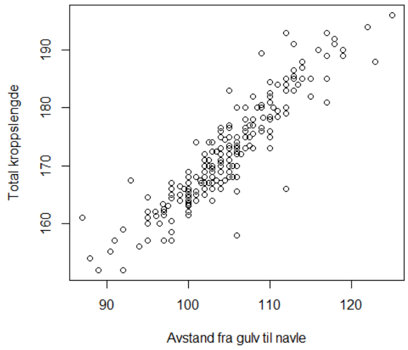
R-kode:

```
#Lager et spredningsplott for fot.navle på x-aksen og kroppslengde på y-aksen.
plot(vitruvisk$fot.navle, vitruvisk$kroppslengde, xlab="Avstand fra gulv til navle", ylab="Total kroppslengde")
```

Output fra R/kjøreeksempel:

```
> #Lager et spredningsplott for fot.navle på x-aksen og kroppslengde på y-aksen.
> plot(vitruvisk$fot.navle, vitruvisk$kroppslengde, xlab="Avstand fra gulv til navle", ylab="Total kroppslengde")
```

Plottet fra R er inkludert i Figur 8.



Figur 8: Spredningsplott over avstand fra gulv til navle på x-aksen og total kroppslengde på y-aksen

Vi ser fra spredningsplottet i Figur 8 at økningen av den totale kroppslengden vokser tilnærmet lineært sammen med avstanden fra gulv til navle. Dette er rimelig ettersom man som regel har en større total kroppslengde når avstanden fra gulv til navle øker. Det sier at man ved en høyere navlehøyde medfører en lengre kroppslengde.

c)

R-kode:

```
#Regner ut korrelasjonen mellom fot.navle og kroppslengde
cor(vitruvisk$fot.navle, vitruvisk$kroppslengde)
```

Output fra R/kjøreeksempel:

```
> #Regner ut korrelasjonen mellom fot.navle og kroppslengde
> cor(vitruvisk$fot.navle, vitruvisk$kroppslengde)
[1] 0.9140397
```

Korrelasjonen mellom avstanden fra gulv til navle og den totale kroppslengden er lik r = 0.914. Fra før vet vi at en korrelasjonsverdi på r = 1 er en perfekt korrelasjon, og siden r = 0.914, er det rimelig å si at det er en sterk korrelasjon mellom avstanden fra gulv til navle og den totale kroppslengden til en person.

d)

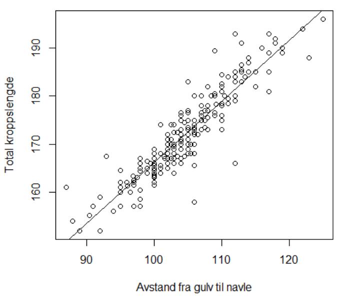
R-kode:

```
#Tilpasser en lineær modell for sammenheng mellom navlehøyde og kroppslengde
fit <- lm(vitruvisk$kroppslengde ~ vitruvisk$fot.navle)
abline(fit)
```

Output fra R/kjøreeksempel:

```
> #Tilpasser en lineær modell for sammenheng mellom navlehøyde og kroppslengde
> fit <- lm(vitruvisk$kroppslengde ~ vitruvisk$fot.navle)
> abline(fit)
```

Plottet fra R er inkludert i Figur 9.



Figur 9: Lineær modell for sammenhengen mellom navlehøyde og kroppslengde

Som plottet viser er det en fin linje med punkter som vokser tilnærmet lineært.

e)

R-kode:

```
#Finner skjæringspunkt og stigningstall til den lineære modellen fit
summary(fit)
```

Output fra R/kjøreeksempel:

```
> #Finner skjæringspunkt og stigningstall til den lineære modellen fit
> summary(fit)

Call:
lm(formula = vitruvisk$kroppslengde ~ vitruvisk$fot.navle)

Residuals:
     Min       1Q   Median       3Q      Max 
-15.7843  -1.9667  -0.0568   2.1695  11.8981 

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)    
(Intercept)         38.89675    3.98888   9.751   <2e-16 ***
vitruvisk$fot.navle  1.27252    0.03799  33.499   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.605 on 221 degrees of freedom
Multiple R-squared:  0.8355,	Adjusted R-squared:  0.8347 
F-statistic:  1122 on 1 and 221 DF,  p-value: < 2.2e-16
```

Vi skal se på koeffisientene. Fra outputet kan vi se at skjæringspunktet er ved 38.89675, og stigningstallet er 1.27252.

Ved en økning på en cm navlehøyde, vil den predikerte kroppslengden økes med ca. 1.27 cm, ettersom dette er stigningstallet vårt og forteller om grafens vekst.

f)

Her skal modellen fra oppgave d) benyttes for å finne kroppslengden til en person med en navle som er 121 cm over bakken.

$F(x) = 1.27252x + 38.89675$

$F(121) = 1.27252*121 + 38.89675$

$F(121) = 193.202$ cm

Kroppslengden til en person med en navle som er 121 cm over bakken, er 193.202 cm.

g)

R-kode:

Bruker opp igjen koden fra oppgave a).

```r
data = "http://www.uio.no/studier/emner/matnat/math/STK1000/data/obligdata/oblig1/vitruvisk.txt"
vitruvisk <- read.table(data,header=TRUE)

#Bruker table funksjonen for å finne antallet av hver kategorisk variabel
table(vitruvisk$kjonn)
#Bruker summary for fem-punkts oppsummering for kroppslengde og fot.navle
summary(vitruvisk)
```

Output fra R/ kjøreeksempel:

Bruker opp igjen eksempelet i oppgave a)

```
> #Bruker table funksjonen for å finne antallet av hver kategorisk variabel
> table(vitruvisk$kjonn)

  K   M 
150  73 
> #Bruker summary for fem-punkts oppsummering for kroppslengde og fot.navle
> summary(vitruvisk)
    kjonn            kroppslengde      fot.navle        navle.isse         favn      
 Length:223         Min.   :152.0   Min.   : 87.0   Min.   :52.00   Min.   :146.0  
 Class :character   1st Qu.:166.0   1st Qu.:101.0   1st Qu.:65.00   1st Qu.:165.0  
 Mode  :character   Median :172.0   Median :104.0   Median :67.00   Median :171.0  
                    Mean   :172.3   Mean   :104.8   Mean   :67.34   Mean   :172.4  
                    3rd Qu.:178.0   3rd Qu.:109.0   3rd Qu.:70.00   3rd Qu.:180.0  
                    Max.   :196.0   Max.   :125.0   Max.   :81.00   Max.   :202.0  
```

Her ser vi fra summary, at når kroppslengden øker så øker fot.navle og vice versa. Man ser også at de følger den lineære modellen for økningen av kroppslengde, og øker lineært slik den skal. Ser man på minste verdiene og maks verdiene, ser man at ved en lav kroppslengde medfører en lav navlehøyde og vice versa.

h)

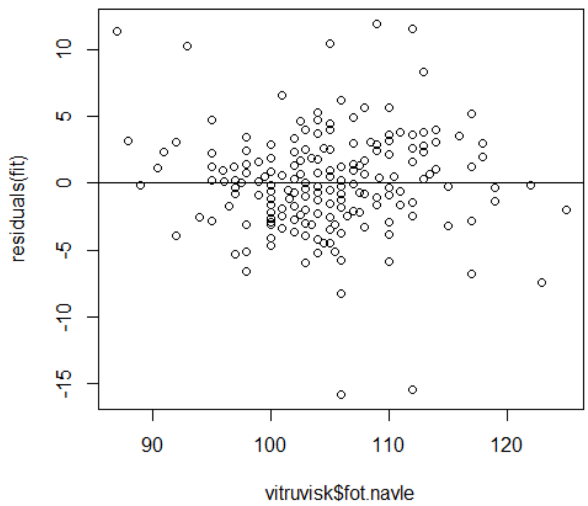
R-kode:

```
#Lager et plott av residualene i modellen
plot(vitruvisk$fot.navle,residuals(fit))
abline(h=0)
```

Output fra R/kjøreeksempel:

```
> #Lager et plott av residualene i modellen
> plot(vitruvisk$fot.navle,residuals(fit))
> abline(h=0)
```

Plottet fra R er inkludert i Figur 10.



Figur 10: Plott av residualene i modellen

I Figur 10 kan man identifisere flere uteliggere. På positiv y-akse ser vi at disse ligger på rundt 10 horisontalt, og på negativ y-akse at disse ligger på rundt -15 horisontalt. Figur 10 viser at denne modellen passer bra til å beskrive dataene. Vi ser at det er en samling av punkter rundt midten i plottet, og verdiene ligger som antatt i standardavviket, det er også få uteliggere.