# Oppg1

## A picture containing screenshot Description automatically generatedLag boksplott som viser fordelingen av observasjonene. Kommenter hva du finner.

Ser at kvinner sin snitt temperatur er høyere en menns. Og at kvinner har et høymaksverdi, mens min verdien er nærmere snittet. Dermed er kvinners fordeling skjev. Menn fordelingen er mindre skjev med maks, og min verdier ganske jevnt rundt snittet. Dog snittet er litt positivt skjevt.

## Lag normalfordelingsplott for de to observasjonssettene. Kommenter hva du ser.

A close up of a map

Description automatically generated

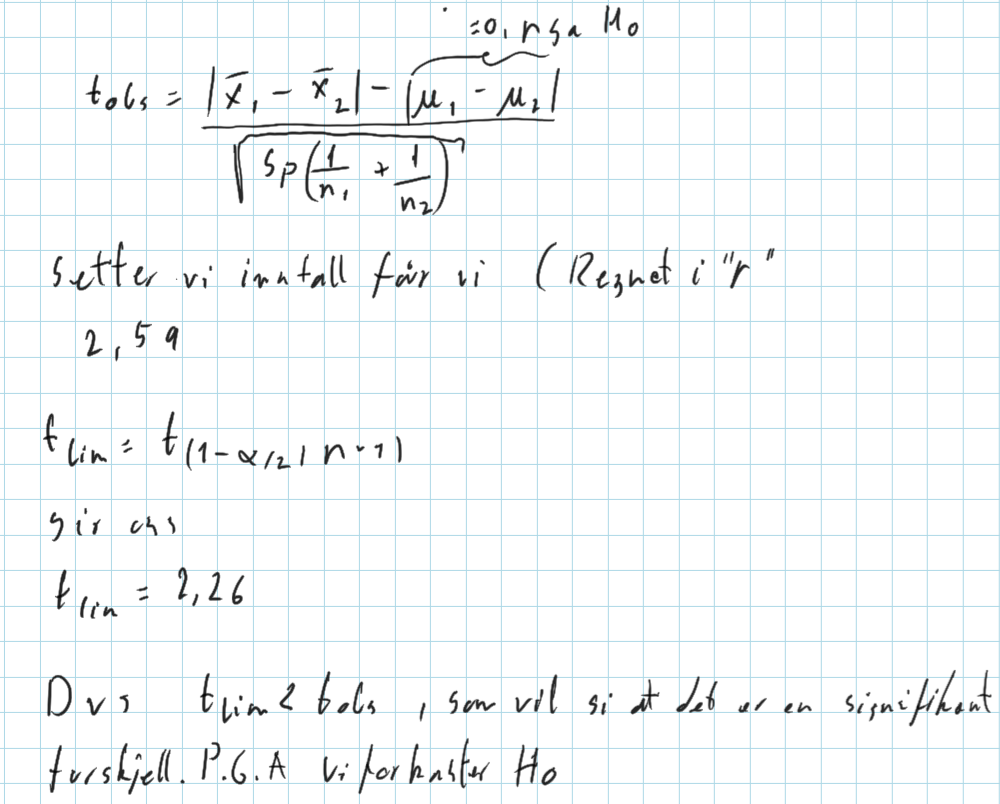
Ser at både kvinner og menn fordelingene er ganske normalt fordelte med få punkter lengere unna enn resten.

1. Anta at variansen er den samme for de to utvalgene, og test med niv˚a 5% om det er noen forskjell i forventet kroppstemperatur. Beregn ogs˚a P-verdien, og lag et 95% konfidensintervall for forventet forskjell.

Vi sette H0: mu\_menn = mu\_kvinner, Ha: mu\_menn != mu\_kvinner

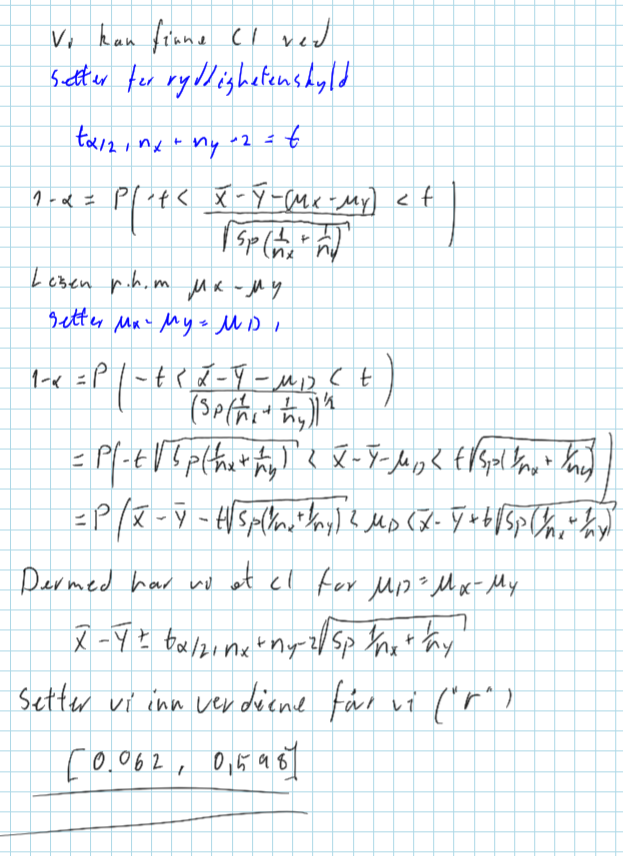
For å gjøre testen trenger jeg t\_obs, og t\_lim:

b1CI

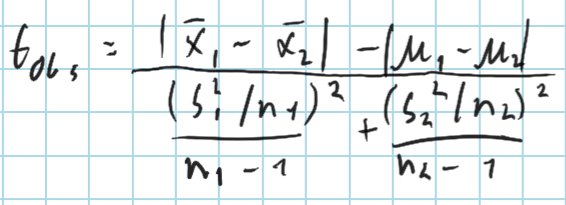


For å finne P bruker jeg r koden P = 2\* pt(t\_obs, nx+ny-2).

Dvs. P = 1.981519

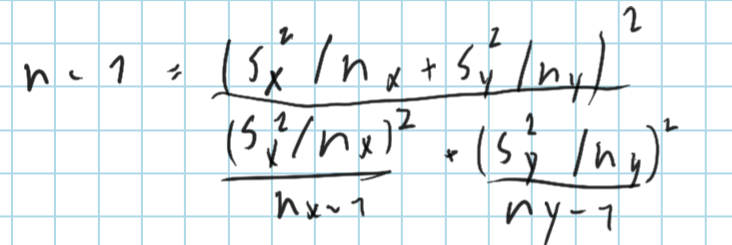


1. Gjennomfør testen og beregn P-verdien ogs˚a i det tilfellet der man ikke antar felles varians. Diskuter og forklar resultatene.

Hvis det ikke er samme varians bruker vi 

0Det gir oss en t\_obs verdi på 3.36483. som er en noe større verdi enn det vi hadde når vi anntok lik varianse. De er fortsatt ganske lik noe vi ville anntat da vaniansen for de to fordelingene er rimelig like.

Vi estimerer frihetsgraden med:

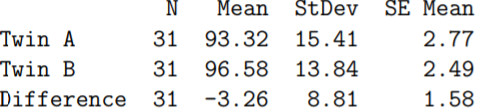
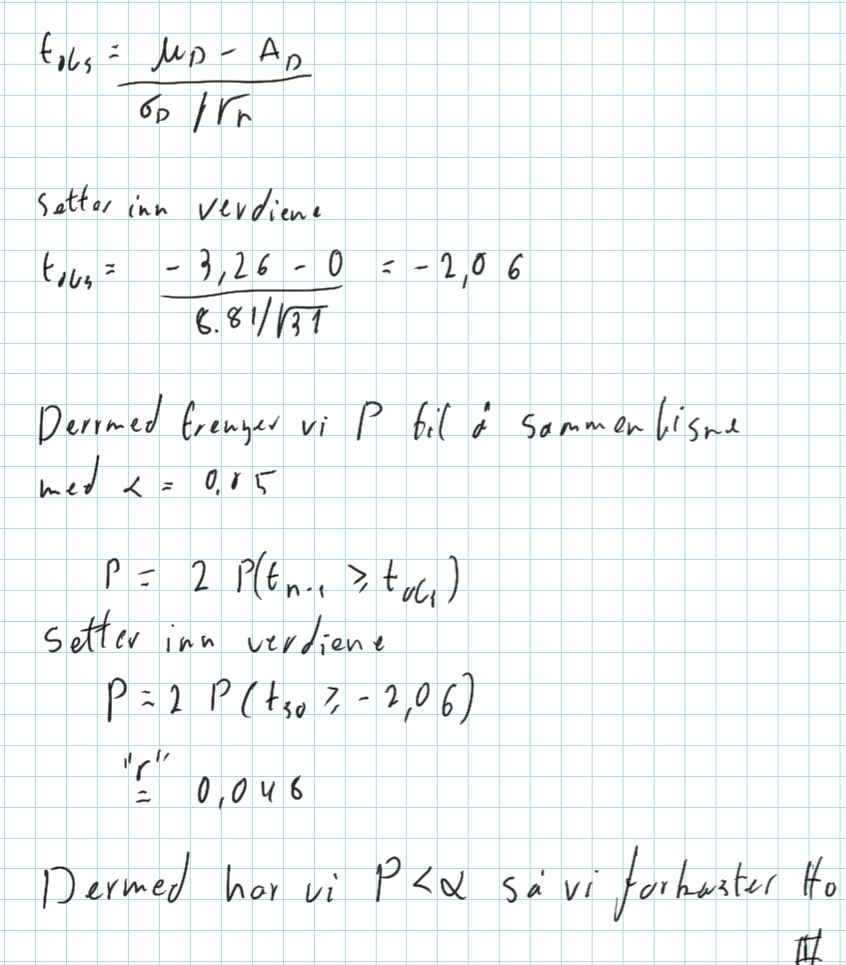


Så runder vi n-1 til nermeste hele tall. Da har vi estimerte frihetsgrader.

Estimatet er v = 18 som stemmer overens med den tidligere utregningen da vi regnet med at Variansene var det samme.

# Oppg2

### b) ) Kall forventet forskjell mellom Twin A og Twin B for µD. Sett opp nullhypotese og alternativ hypotese for ˚a besvare spørsm˚alet om forskjell i IQ. Finn en egnet testobservator, og beregn dennes numeriske verdi. Beregn s˚a tilhørende P-verdi. Spesifiser antall frihetsgrader i fordelingen du bruker. Formuler din konklusjon p˚a testen



Vi har n-1 frihetsgrader.

## c) Finn et 95% konfidensintervall for µD. Hva betyr det at dette intervallet dekker kun negative verdier? Forklar kort om sammenhengen mellom tosidig testing og konfidensintervaller.

# 

Vår forventningsverdi er negativ da får vi et konfidensintervall som er sentrer rundt forventningsverdien.

# Oppg3

### Utfør ulike typer eksplorativ data analyse (e.g. histogram, qqplot (probability plot), spredningsplott, dvs. scatterplot ) p˚a datasettet og beskriv hva du ser.

A close up of a map

Description automatically generated

Fra Normal qq Plot ser vi at Strength er tilnærmet normalfordelt. Siden det ser ut som punktene følger normalen.

Fra boxplottet ser vi at Strength er positivt skjevt fordelt. Og vi har noen veldig lave verdier sammenlignet med forventingsverdien.

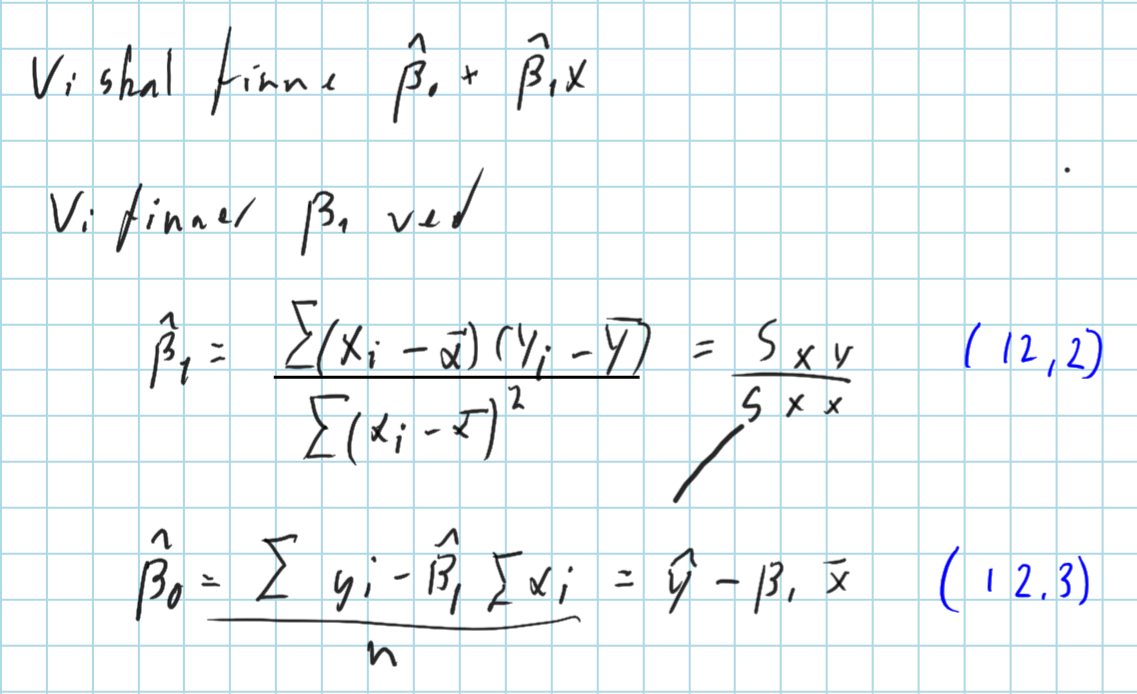
Fra de to siste plottene, er Strength(y) plottet mot Temp(x1) og Pressure(x2). Nå kan vi se hvilke som ser ut til å være mest linjer. Vi ser at y(temp) ser mer linijer ut,enn y(Pressure), som har et diskontinuerlig hop rundt x2 = 12.

## Vi vil i første omgang konsentrere oss om forklaringsvariablen Temperature. Utfør en enkel lineær regresjon i dette tilfellet. Plott data sammen med den tilpassende regresjonslinjen og komment´er resultatene.

A close up of a map

Description automatically generated

Ved å bruke minste kvadraters problem kan vi finne regresjonslinjen.

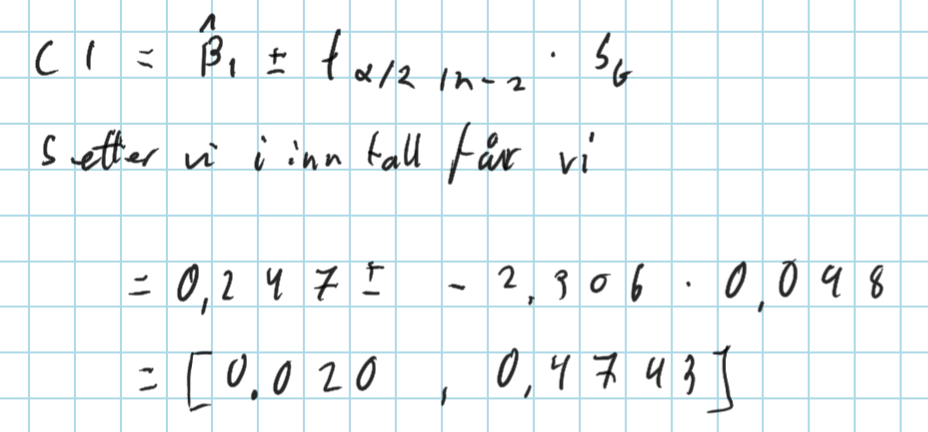


Ved r kan vi bruke lm(x1, y, xlab='Temp',ylab='Strength') for å finne beta1\_hat og beta0\_hat

(kalt betaX1hat og alfaX1hat i koden). Får da verdiene beta1\_hat=0.2474189, beta0\_hat =-29.8479549. Fra plottet ser regresjonen rimelig ut, og det virker som regresjonen er en ok estimator for y(x1).

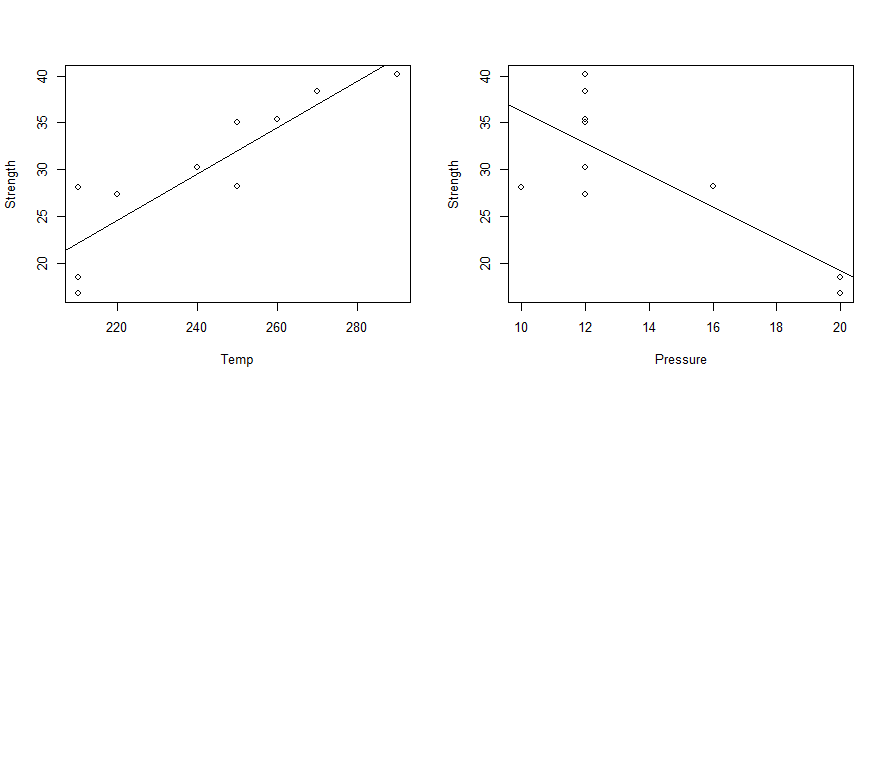
## C) Lag konfidensintervall for regresjonskoeffsienten for Temperature. Gir intervallet indikasjon p˚a om temperatur er en viktig forklaringsvariabel?

CI for regresjonsintervallet finner vi ved:



Vi ser at min tallet 0.02 ikke er veldig nerme -1 og makstallet er 0.47 som da er nermere 1, dette avkrefter ikke at vi kan bruke dette som en modell, men det er ikke veldig bra.

f)



Fra bildet ser vi at regresjonslinjen til Strength(Pressure) går nedover, og punktene til linjen ser ikke ut til å følge linjen i det hele tatt. Dermed konkluderer jeg med at Strength(Temp) er et mye bedre estimert avhengighet.

# Kode

## Oppg1

y =menn <- c(36.1, 36.3, 36.4, 36.6, 36.6, 36.7, 36.7, 37.0, 36.5, 37.1)

x = kvinner <- c(36.6, 36.7, 36.8, 36.8, 36.7, 37.0, 37.1, 37.3, 36.9, 37.4)

#a) Lag boksplott som viser fordelingen av observasjonene. Kommennter hva du finner.

boxplot(menn, kvinner, names=c('menn', 'kvinner'))

'

Vi ser fra plottet at menn har litt høyere kroppstemperatur enn menn.

Vi ser også at mennene sin fordeling er sjevt mot lavere verdier,

menns kvinners fordeling er har ingen skjevhet

'

#b) Lag normalfordelingsplott for de to observasjonssettene. Kommennter hva du ser.

limy = c(36,37.5)

par(mfrow=c(1,2))

qqnorm(menn, ylab='(C) temp menn', ylim=limy)

qqline(menn)

qqnorm(kvinner, ylab='(C) temp kinner', ylim=limy)

qqline(kvinner)

#c)

alfa = 0.05

nx = length(kvinner)

ny = length(menn)

Sx = sd(x)

Sy = sd(y)

sp\_square = ((nx-1)\*Sx\*\*2 + (ny-1)\*Sy\*\*2)/(nx+ny-2)

t\_obs = abs((mean(kvinner)-mean(menn)))/sqrt(sp\_square\*(1/nx + 1/ny))

t\_lim = qt(1-alfa/2,nx-1)

P = 2\* pt(t\_obs, nx+ny-2)

t\_CI = mean(x) - mean(y) + c(1,-1)\*qt(alfa/2, nx+ny-2)\*sqrt(sp\_square\*(1/nx + 1/ny))

#d)

t\_obs\_NonEqualVar = abs(mean(kvinner)-mean(menn))/sqrt(Sx\*\*2\*\*2/nx + Sy\*\*2/ny)

v = round((Sx\*\*2/nx + Sy\*\*2/ny)\*\*2/((Sx\*\*2/nx)\*\*2/(nx-1) + (Sy\*\*2/ny)\*\*2/(ny-1)))

P = 2\* pt(t\_obs\_NonEqualVar, v)

#e) Utled og gjennomfør en F-test for ˚a sjekke om det er noen grunn til ˚a p˚ast˚a at

#variansene er forskjellige. Sjekk mot var.test() i R.

f = Sx\*\*2/Sy\*\*2

CIf = qf(c(alfa/2, 1-alfa/2),nx-1,ny-1)

# Oppg2

AN = 31

BN = 31

AMean = 93.32

BMean = 96.58

AStDev = 15.41

BStDev = 13.84

ASE\_Mean = 2.77

BSE\_Mean = 2.49

DN = 31

DMean = -3.26

DStDev = 8.81

DSE\_Mean = 1.58

alfa = 0.05

t\_obs = DMean/(DStDev/sqrt(DN ))

P = 2\* pt(t\_obs, DN-1)

#c)

t\_CI = DMean + c(1,-1)\*qt(alfa/2, DN-1)\*DStDev/sqrt(DN)

## oppg3

path="https://www.uio.no/studier/emner/matnat/math/STK1110/data/plastic.txt"

plast=read.table(path,header=T)

y =Strength = plast[, 'Strength']

x1 = Temp = plast[, 'Temperature']

x2 = Pressure = plast[, 'Pressure']

par(mfrow=c(2,2))

qqnorm(Strength, ylab='Strength')

qqline(Strength)

#Plottet ser tilnermet normalfordelt ut.

boxplot(Strength, xlab='Strength')

#Boxplotet viser en poseti sjevhet

qqplot(x1,y, xlab='Temp')

#qqpline(y)

qqplot(x2,y, xlab='Pressure')

#Det ser ut som y er mer gjevnt avengig av Temp, mens det er en stor endring når Pressure er 12 som kan vise til at Styrken til plasten øker betydlig når trykket er 12.

#Dette kan ytteliger vises med en korrelasjonskoefisient.

#b)Regresjon, y(x1)

x\_bar = c(mean(x1), mean(x2))

y\_bar = mean(y)

#b1\_hat = sum((x1-x\_bar[1])\*(y-mean(y)))/sum((x1-x\_bar[1])\*\*2) s

#a1\_hat = y\_bar-b1\_hat\*x\_bar[1]

#b2 = sum((x2-x\_bar[2])\*(y-mean(y)))/ sum((x-x\_bar[1])\*\*2)

l.lm = lm(y~x1)

plot(x1, y, xlab='Temp',ylab='Strength')

abline(l.lm)

# ser tilsynelatende ut til å ha en korrelasjon

#c)

betaX1hat = coefficients(l.lm)[2]

alfax1hat = coefficients(l.lm)[1]

alfa = 0.05

n = length(y)

#s= sqrt((sum(y\*\*2-alfax1hat)\*sum(y-betaX1hat)\*sum(x1\*y))/(n-2))

s = sqrt(sum((y-mean(y))\*\*2) /(n-2))

Sxx = sum((x1-x\_bar[1])\*\*2)

sb = s/sqrt(Sxx)

b1CI = betaX1hat+ c(1,-1)\*qt(alfa/2, n-2) \* sb

#intervallet virker lite og viser at variasjoen er liten i betaX1hat

#d)

CI\_Strength = y\_bar + c(1,-1)\*qt(alfa/2, n-2)\*sd(y)/sqrt(n)

x\_star = c(210, 240, 270)

y\_hat = alfax1hat+betaX1hat\*x\_star

CI\_pred\_pluss = y\_hat +qt(alfa/2,n-2)\*sqrt(s\*\*2+sd(y)\*\*2)

CI\_pred\_minus = y\_hat -qt(alfa/2,n-2)\*sqrt(s\*\*2+sd(y)\*\*2)

CI\_pred\_x\_sar = cbind(CI\_pred\_pluss, CI\_pred\_minus)

#prediksjons intervallet er betydelig større en CI\_Strength siden prediksjonen basserer seg på vår estimerte regresjon

#og CI\_Strength er bassert på våre observasjoner.

#s.659

#f)

l.lm2 = lm(y~x2)

plot(x2, y, xlab='Pressure',ylab='Strength')

abline(l.lm2)

betaX2hat = coefficients(l.lm2)[2]

alfax2hat = coefficients(l.lm2)[1]

Sxx2 = sum((x2-x\_bar[2])\*\*2)

sb2 = s/sqrt(Sxx2)

b2CI = betaX2hat+ c(1,-1)\*qt(alfa/2, n-2) \* sb2