331a1

```
# (a)
berry=read.csv("berries.csv")
y=berry$chewiness
x=berry$sugar

plot(x,y,xlab = "sugar content(g/L)", ylab="chewiness(mJ)", main="berry data")

# (b)
M=lm(y~x)
b0.hat = M$coefficients[1]
b1.hat = M$coefficients[2]
# when b0.hat is this value, sum of square of residual is minimized, so we estimate beta_0 to be b0.hat
b0.hat
```
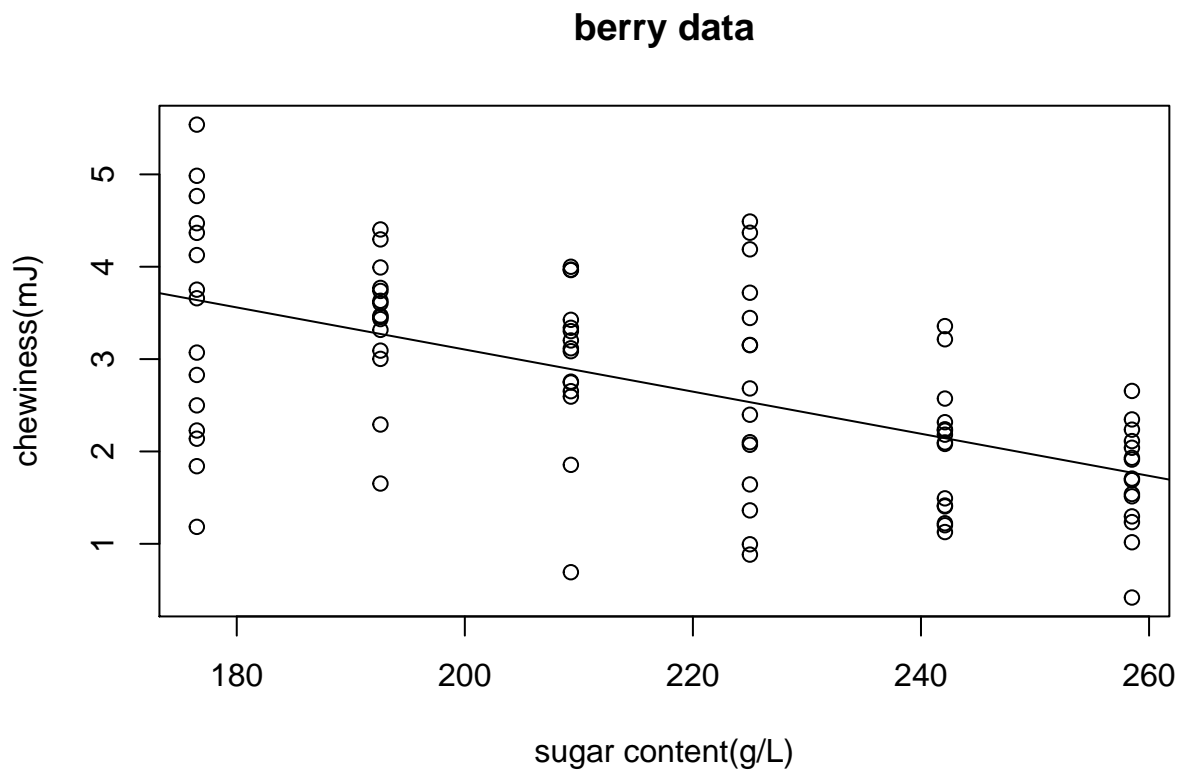
```
## (Intercept)
##    7.662878
```

```
# when b1.hat is this value, sum of square of residual is minimized, so we estimate beta_1 to be b1.hat
b1.hat
```

```
##           x
## -0.02279699
```

```
abline(b0.hat, b1.hat)
```

berry data

we have $\hat{\beta}_0 \approx 7.66$

we have $\hat{\beta}_1 \approx -0.0228$

```
# (c)
n=length(x)
e=M$residuals
sig.hat=sqrt(sum(e^2)/(n-2))
xbar=mean(x)
Sxx=sum((x-xbar)*(x-xbar))

s1 = sig.hat/sqrt(Sxx)
Tobs = b1.hat/s1
# t statistics
round(Tobs,3)
```

```
##      x
## -6.603
```

```
pval = pt(q=abs(Tobs), df=n-2, lower.tail = FALSE)
pval = 2*pval
round(pval, 9)
```

```
##     x
## 3e-09
```

we have t statistics≈ -6.603, p-value$\approx 3*10^{-9}$

if $\beta_1 = 0$, we have about $3*10^{-9}$ possibility of observing data as extreme as this or that are more extreme, we reject the null hypothesis as $10^{-9} < 0.01$
chewiness and sugar do have a significant linear relationship

```
# (d)
pred.int = function(xstar,xlevel){
  qval=qt(1-(1-xlevel)/2, df=n-2)
  mu.star = b0.hat + b1.hat * xstar
  s.star = sig.hat*(1+1/n+(xstar-xbar)^2/Sxx) # standard error
  cbind(L=mu.star - qval * s.star,
        U=mu.star + qval * s.star)
}
pred.interval= pred.int(120, 0.95)
pred.interval
```

```
##                     L       U
## (Intercept) 2.838499 7.01598
```

```
pred.estimate = b0.hat + b1.hat * 120
pred.estimate
```

```
## (Intercept)
##    4.927239
```

prediction interval of sweetness is approximately (2.838,7.016)

estimate of chewiness is approximately 4.927

```r
reptime=2000
b0.hat.vec = rep(0, reptime)
b1.hat.vec = rep(0, reptime)
b0.CI.lst = vector("list", length=reptime)
b1.CI.lst = vector("list", length=reptime)
set.seed(20704870)

for(i in 1:reptime){
  #a)
  x=c(1,1,2,3,3,3,4,5,5,6)
  y=2*x+1+rnorm(10,0,0.5)

  #b)
  xbar=mean(x)
  ybar=mean(y)
  Sxy=sum((y-ybar)*(x-xbar))
  Sxx=sum((x-xbar)*(x-xbar))
  n=length(y)
  b1.hat = Sxy/Sxx
  b0.hat=ybar-b1.hat*xbar
  b1.hat.vec[i]=b1.hat
  b0.hat.vec[i]=b0.hat

  #c)
  yfit=b0.hat+b1.hat*x
  e=y-yfit
  sig.hat=sqrt(sum(e^2)/(n-2))
  conf = 0.9
  qval = -qt(p=(1-conf)/2, df=n-2)

  s1 = sig.hat/sqrt(Sxx)

  b1.CI = b1.hat+c(-1,1)*qval*s1

  round(b1.CI, 3)

  s0=sig.hat*sqrt(1/n+xbar^2/Sxx)

  b0.CI = b0.hat + c(-1, 1) * qval * s0
  round(b0.CI, 3)

  b0.CI.lst[[i]]=b0.CI
  b1.CI.lst[[i]]=b1.CI
}

#d)
b0.hat.mean=mean(b0.hat.vec)
b0.hat.mean
```

```
## [1] 1.010285
```

```r
b0.hat.var = var(b0.hat.vec)
b0.hat.var
```

```
## [1] 0.1307359
```

```
b1.hat.mean=mean(b1.hat.vec)
b1.hat.mean
```

[1] 1.998088

```
b1.hat.var = var(b1.hat.vec)
b1.hat.var
```

[1] 0.009757099

theoretical mean of $\hat{\beta}_1$ is 2, which is true value of β_1, the approximation is very close to 1.998, the calculated mean of $\hat{\beta}_1$
theoretical mean of $\hat{\beta}_0$ is 1, which is true value of β_0, the approximation is very close to 1.010, the calculated mean of $\hat{\beta}_0$
theoretical variance of $\hat{\beta}_1$ is $\frac{\sigma^2}{S_{xx}} = \frac{0.5^2}{26.1} \approx 0.00958$, which is very close to 0.00976, the calculated variance of $\hat{\beta}_1$
theoretical variance of $\hat{\beta}_0$ is $\sigma^2(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}) \approx 0.25(0.1 + 0.4172) \approx 0.129$, which is very close to 0.131, the calculated variance of $\hat{\beta}_1$

```
correct.b1.count=0
for(i in 1:reptime){
  if(b1.CI.lst[[i]][1]<= 2 && b1.CI.lst[[i]][2]>=2){
    correct.b1.count=correct.b1.count+1
  }
}
correct.b1.count/reptime
```

[1] 0.892

```
correct.b0.count=0
for(i in 1:reptime){
  if(b0.CI.lst[[i]][1]<= 1 && b0.CI.lst[[i]][2]>=1){
    correct.b0.count=correct.b0.count+1
  }
}
correct.b0.count/reptime
```

[1] 0.893

about 89.3% interval of β_1 covers true β_1

about 89.2% interval of β_0 covers true β_0

They are very close to true CI which is 90%