HW1

Zeqiu.Yu

2022-10-02

Question 1.5
No. The simple linear model should be stated as

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

or

$$E\{Y_i\} = \beta_0 + \beta_1 X_i$$

Question 1.19

```r
input1 <- read.table("./CH01PR19.txt")
names(input1) <- c("GPA","ACT")
```

(a.)

```r
fit1 <- lm(GPA~ACT, data = input1)
summary(fit1)
```

```
##
## Call:
## lm(formula = GPA ~ ACT, data = input1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.74004 -0.33827  0.04062  0.44064  1.22737
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.11405    0.32089   6.588  1.3e-09 ***
## ACT          0.03883    0.01277   3.040  0.00292 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6231 on 118 degrees of freedom
## Multiple R-squared:  0.07262,    Adjusted R-squared:  0.06476
## F-statistic:  9.24 on 1 and 118 DF,  p-value: 0.002917
```

$$b_0 = 2.11405, b_1 = 0.03883.$$

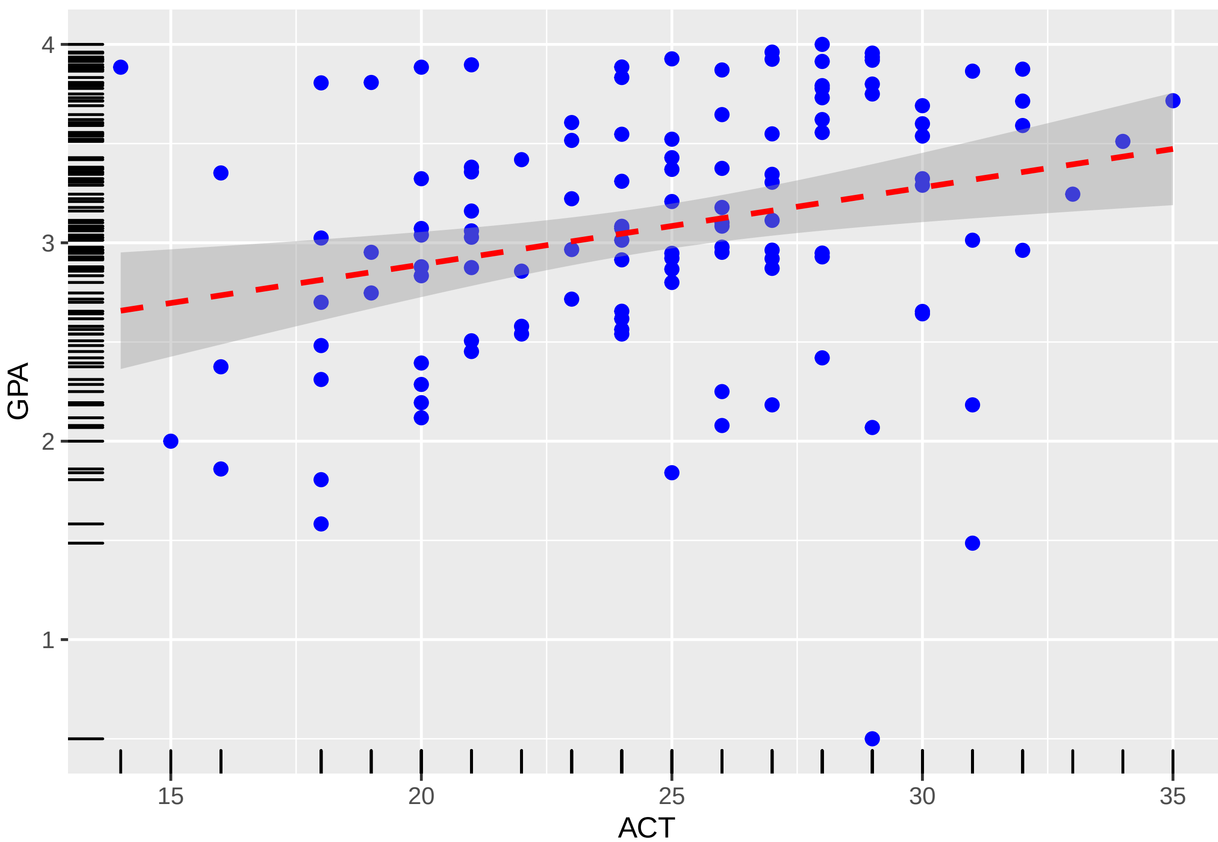
Then we get:

$$\hat{Y} = 2.11405 + 0.03883X$$

(b.) Plot the estimated regression function and the data. Does the estimated regression function appear to fit the data well.

```
## Loading required package: ggplot2

## `geom_smooth()` using formula 'y ~ x'
```



The estimated regression funnction doesn't seem to fit it well. The distribustion of the points is so spread and the linear model has a relative low adjusted R-squared value 0.06476.

(c.) Obtain a point estimate of the mean fresh man GPA for students with ACT test score X = 30.

$$\hat{Y} = 2.11405 + 0.03883X = 3.27895$$

(d) What is the point estimate of the change in the mean response when the entrance test score increase by one point?

$$1 \times b_1 = 0.03883$$

1.27

```r
input2 <- read.table("./CH01PR27.txt")
names(input2) <- c("muscle_mass","Age")
```

(a.)

```r
fit2 <- lm(muscle_mass~Age, data = input2)
summary(fit2)
```

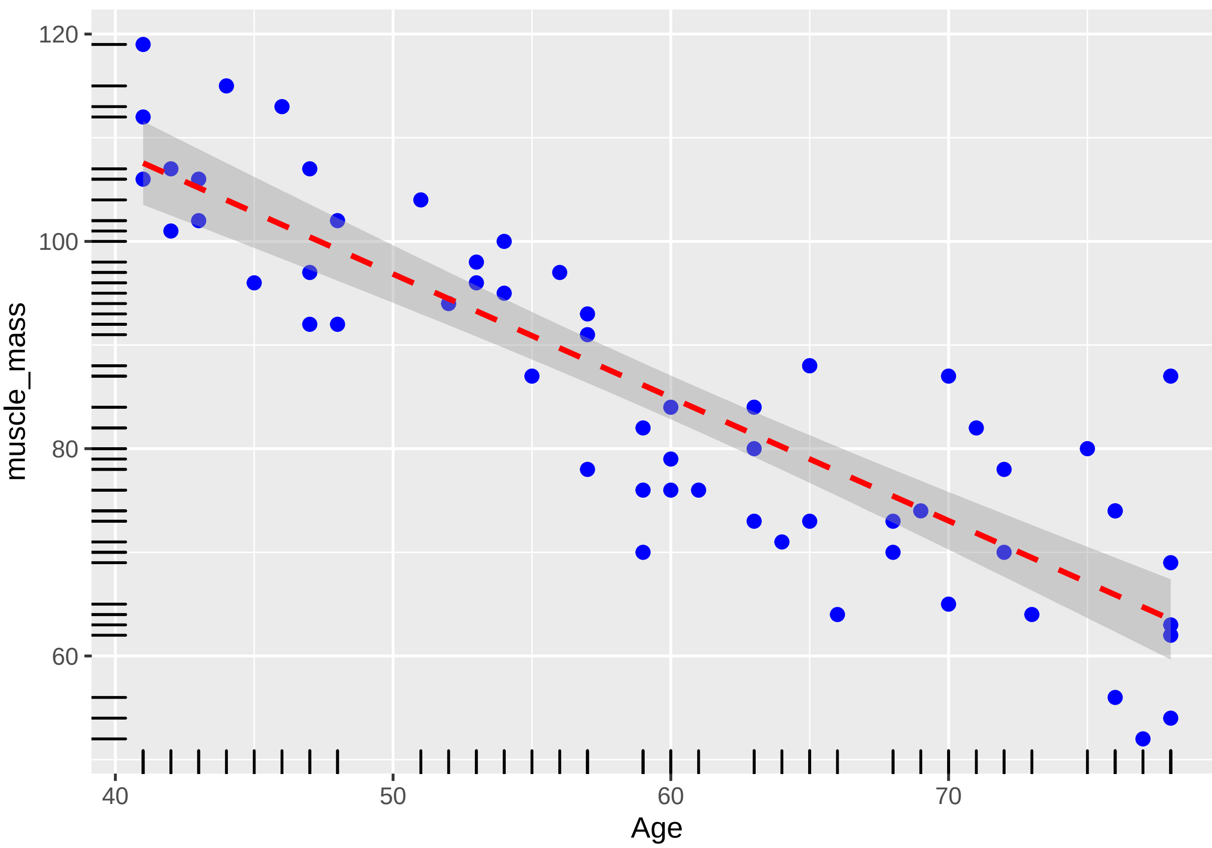
```
##
## Call:
## lm(formula = muscle_mass ~ Age, data = input2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -16.1368  -6.1968  -0.5969   6.7607  23.4731
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 156.3466     5.5123   28.36   <2e-16 ***
## Age          -1.1900     0.0902  -13.19   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.173 on 58 degrees of freedom
## Multiple R-squared:  0.7501, Adjusted R-squared:  0.7458
## F-statistic: 174.1 on 1 and 58 DF,  p-value: < 2.2e-16
```

The esitimated regression function is:

$$\hat{Y} = 156.3466 - 1.1900X$$

```
## `geom_smooth()` using formula 'y ~ x'
```



The linear function appear to give good fit and the plot proves the anticipation that the muscle mass decreases with age.

(b.) (1)

$$-1.1900 \times 1 = -1.19$$

(2)

$$\hat{Y} = 156.3466 - 1.1900 \times 60 = 84.9466$$

(3) For the eighth case, $X_8 = 41$ and $Y_8 = 112$

$$e_8 = Y_8 - \hat{Y}_8 = 112 - (156.3466 - 1.1900 \times 41) = 4.4434$$

(4) We use MSE to estimate σ^2.

$$s^2 = \frac{1}{n-2}\sum_{i=1}^{n} e_i^2$$

```
squaredError <- (input2$muscle_mass - (156.3466 -1.1900*input2$Age))^2
s2 <- sum(squaredError)/(dim(input2)[1] - 2)
print(s2)
```

```
## [1] 66.80082
```

Hence, a point estimation of σ^2 is 66.80082.

(Remark: the values of b_0 and b_1 is consistent with the values calculated by the formula proved in the textbook.)