# HW2

Zeqiu.Yu

2022-10-10

## Homework2

### Question 2.27

```
input1 <- read.table("./BookDataSets/Chapter  1 Data Sets/CH01PR27.txt")
names(input1) <- c("muscle_mass","Age")
```

a. By hypothesis test: $H_0 \ : \beta_1 = 0$ and $H_a \ : \beta_1 < 0$

```
fit <- lm(muscle_mass~Age, data = input1)
summary(fit)
```

```
##
## Call:
## lm(formula = muscle_mass ~ Age, data = input1)
##
## 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
```

Then the test statistic $t^* = \frac{b_1 - 0}{s\{b_1\}} = \frac{(-1.19)-0}{0.0902} = -13.1929$
$H_0 \ : \beta_0 \geq 0$ and $H_a \ : \beta_0 < 0$

```
t_value = -qt(0.95, 58)
p_value = pt(-13.1929,58, lower.tail = TRUE)
print(sprintf("The t-value is:%.4f, and the p-value is: %.2f", t_value, p_value))
```

```
## [1] "The t-value is:-1.6716, and the p-value is: 0.00"
```

We accept $H_a$ because $t^* < t - value$

b. The answer is yes. The p-value is so small(<0.05) that can be used to reject $H_0 \ : \beta_0 = 0$. That is to say, $b_0$ provides information.

c.

```
confint(fit, "Age", level=0.95)
```

```
##          2.5 %    97.5 %
## Age -1.370545 -1.009446
```

$\hat{Y_i} = b_0 + b_1 X$ and $\hat{Y_{i+1}} = b_0 + b_1(X+1)$, $\Delta\hat{Y_i} = \hat{Y_{i+1}} - \hat{Y_i} = b_1$. Hence, it is not necessary to know the specific ages to make this estimate.

## Question 2.28

a.

```
alpha = 0.05
x_new = data.frame(Age = 60)
predict(fit, x_new, interval = "confidence", level = 1-alpha, se.fit = TRUE)
```

```
## $fit
##        fit      lwr      upr
## 1 84.94683 82.83471 87.05895
##
## $se.fit
## [1] 1.055154
##
## $df
## [1] 58
##
## $residual.scale
## [1] 8.173177
```

The confidence interval is [82.83471, 87.05895]. We are 95% confident that the mean muscle mass for women aged 60 is between 82.83471 to 87.05895.

b.

```
alpha = 0.05
x_new = data.frame(Age = 60)
predict(fit, x_new, interval = "prediction", level = 1-alpha)
```

```
##        fit      lwr     upr
## 1 84.94683 68.45067 101.443
```

The prediction interval is [68.45067, 101.443]. It's not in comparision with the result of a.

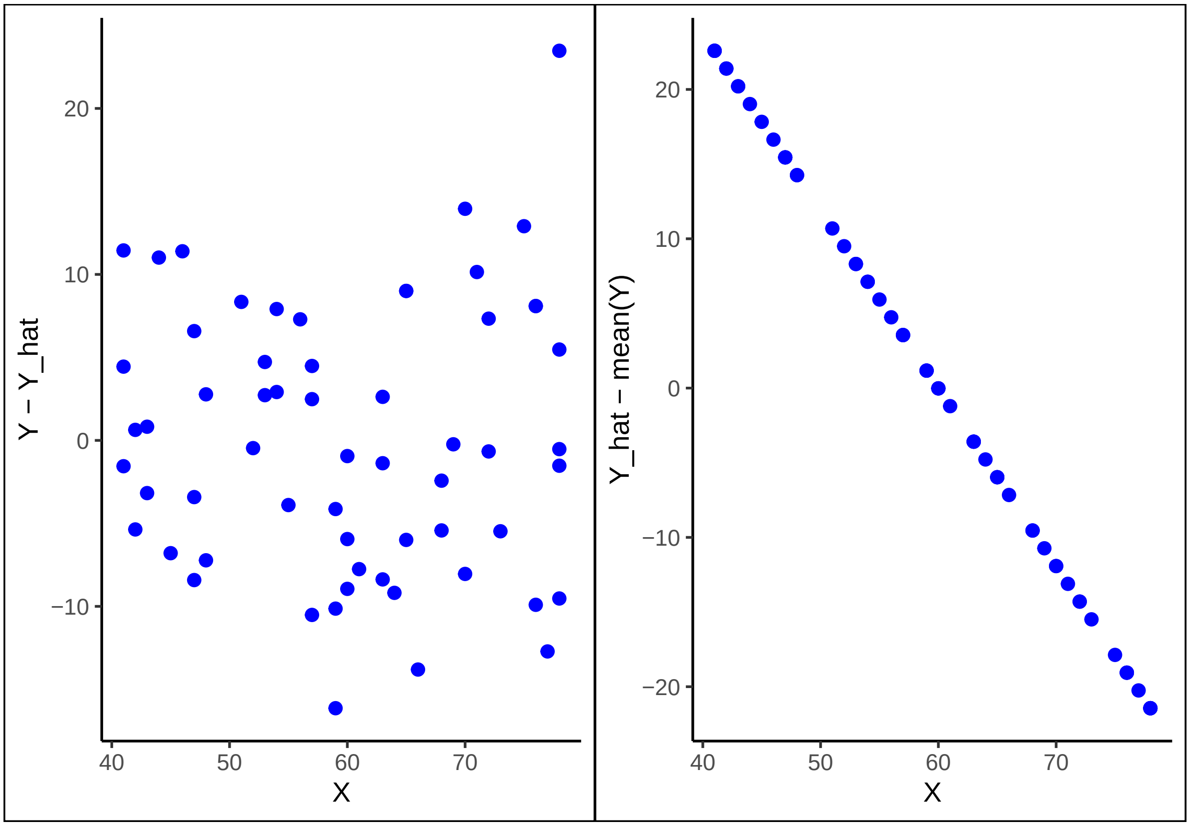
c.

```
alpha = 0.05
x_new = data.frame(Age = 60)
CI <- predict(fit, x_new, interval = "prediction", level = 1-alpha, se.fit = TRUE)
Y_h_hat <- CI$fit[1]
Y_h_hat.se <- CI$se.fit
W <- sqrt(2*qf(1-alpha, 2, 58))
Band <- c(Y_h_hat - W*Y_h_hat.se, Y_h_hat + W*Y_h_hat.se)
Band
```

```
## [1] 82.29593 87.59774
```

Yes. It is wider and it should be, because the confidence band encompass the entire regression line and one is able to draw conclusion about any values of X.

## Question 2.29

a. 


$\hat{Y}_i - \bar{Y}$ has larger range and and then SSR has larger component. $R^2 = \frac{SSR}{SSTO}$, $R^2$ tend to be larger and therefore the proportion of the variation of Y can be "explained" by X the regression on X is larger.

b.

```
anova(fit)
```

```
## Analysis of Variance Table
##
## Response: muscle_mass
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## Age        1 11627.5 11627.5  174.06 < 2.2e-16 ***
## Residuals 58  3874.4    66.8
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

c. Let $H_0 : \beta_1 = 0$ and $H_a : \beta_1 \neq 0$. $f^* \leq$ the critial value, I fail to reject $H_0$, otherwise accept $H_a$. $f^* = \frac{11627.5/1}{66.8} = 174.0623 > F(1-\alpha; 1, 58) = 4.006873$. Then we accept $H_a$.

d.

```
VA = anova(fit)
SSR <- VA$"Sum Sq"[1]
SSE <- VA$"Sum Sq"[2]
R_sq <- SSR/(SSR+SSE)
1- R_sq
```

```
## [1] 0.2499332
```

e. $R^2 = 0.7500668$, $r = -0.866064$