Homework 1

Juwon Lee, Economics and Statistics, UCLA

2023-01-13

tinytex::install_tinytex()

1. Review on simple linear regression.

(1) B. They have moderate negative linear relationship.

(2)

$$\rho = \text{Corr}(X,\ Y) = \frac{\text{Cov}(X,\ Y)}{\text{sd}(x)\text{sd}(y)} = \frac{E(XY) - E(X)E(Y)}{10.84 * 5} = \frac{E(XY) - 20.97 * 20.16}{10.84 * 5}$$

(3) This is

$$\hat{\beta}_1 = R^2 * \frac{s_Y}{s_X} = 0.4117 * \frac{5}{10.84} \approx 0.1898985$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 * \bar{X} \approx 16.17783$$

It means that we can make a 'model' to estimate Y such that $Y = \hat{\beta}_0 + \hat{\beta}_1 X$.

(4) C. About 41% of the total variability in strength of deltoid muscle is explained by the model.

(5) C. It does not have any apparent pattern and it implies the good fit of the model.

(6) It means that $X = 35$, so that

$$Y = \hat{\beta}_0 + \hat{\beta}_1 X \approx 16.17783 + 0.1898985 * 35 = 22.82428$$

2. Review on Hypothesis testing.

(1) $H_0 : \mu = 5.7$ vs $H_1 :\ \mu < 5.7$.

(2) Because $n = 36 > 30$, we can use z-statistic,

$$z = \frac{5.1 - 5.7}{1.4/\sqrt{36}} \approx -2.57143$$

(3) $P(z \leq -2.57143) \approx 0.00506$.

```
pnorm((5.1 - 5.7) / (1.4 / (36)^(1/2)))
```

```
## [1] 0.005063995
```

(4) p-value $\approx 0.00506 < 0.05$, so we can ignore the null.
Thus, we can't say that the number of introverts are 5.7.

(5) $(5.7 - z_{0.05} * \frac{1.4}{\sqrt{36}},\ 5.7 + z_{0.05} * \frac{1.4}{\sqrt{36}}) \approx (5.31620,\ 6.08380)$.

(6) Thus, if the sample mean is included in this interval, then we can say that the null hypothesis can be rejected. Thus, because $\bar{X} = 5.1$, we can reject the null.

3. Review on R programming

First of all, we can make a dataset by the csv file.

```
heart <- read.csv("/Users/user/Desktop/Yonsei/Junior/3-2/Introduction to Data Analysis and Regression/H

heart1 <- heart[c(6,9)]
head(heart1)
```

```
##   Chol MaxHR
## 1  233   150
## 2  286   108
## 3  229   129
## 4  250   187
## 5  204   172
## 6  236   178
```

(1)

```
MaxHR <- as.vector(heart1[2])
Chol <- as.vector(heart1[1])

data_lm <- lm(MaxHR~Chol, data=heart1)
data_lm
```

```
##
## Call:
## lm(formula = MaxHR ~ Chol, data = heart1)
##
## Coefficients:
## (Intercept)         Chol
##  149.981292    -0.001516
```

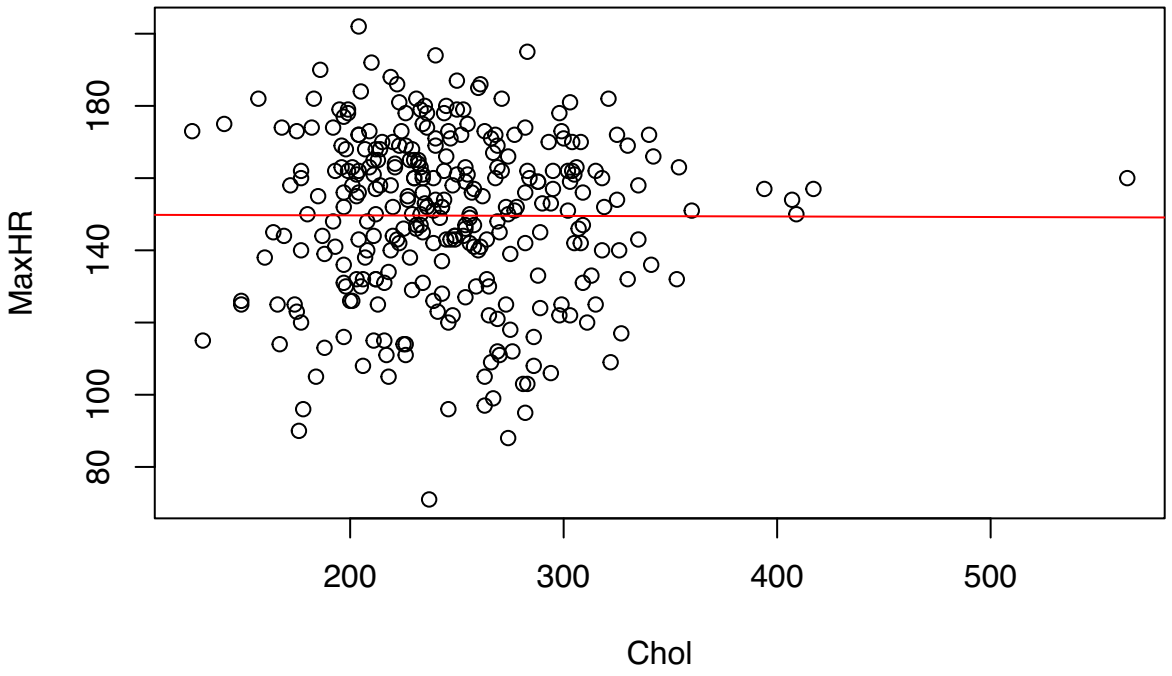
```
summary(data_lm)
```

```
##
## Call:
## lm(formula = MaxHR ~ Chol, data = heart1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -78.622 -16.079   3.375  16.412  52.328
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 149.981292   6.418400   23.37   <2e-16 ***
## Chol         -0.001516   0.025465   -0.06    0.953
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 22.91 on 301 degrees of freedom
## Multiple R-squared:  1.178e-05,  Adjusted R-squared:  -0.00331
## F-statistic: 0.003545 on 1 and 301 DF,  p-value: 0.9526
```

Thus, $y = 149.981292 - 0.001516x$.

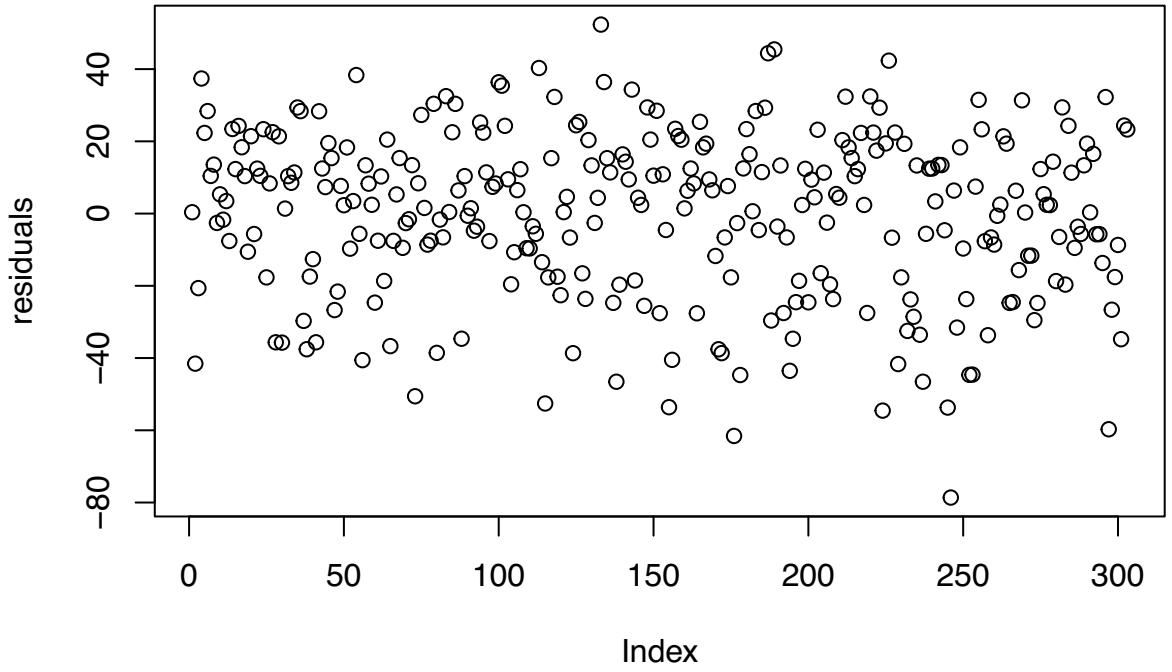
(2)

```
plot(heart1)
abline(149.981292, -0.001516, col='red')
```



(3) The R-squared in summary is 1.178e-05, which is too small to assess.

```
plot(data_lm$residuals, ylab='residuals')
```



4. Pre-Course Survey

I have completed the Pre-Course Survey

I have completed the Pre-Course Survey. Thank you.