STAT 8020 R Lab 4: Simple Linear Regression IV

Whitney

August 30, 2020

Contents

Maximum Heart Rate vs. Age Example

First Step: Load the data

```
dat <- read.csv('http://whitneyhuang83.github.io/STAT8010/Data/maxHeartRate.csv', header = T)
head(dat)
```

```
##   Age MaxHeartRate
## 1  18          202
## 2  23          186
## 3  25          187
## 4  35          180
## 5  65          156
## 6  54          169
```

```
attach(dat)
```

Fitting a simple linear regression

```
fit <- lm(MaxHeartRate ~ Age)
summary(fit)
```

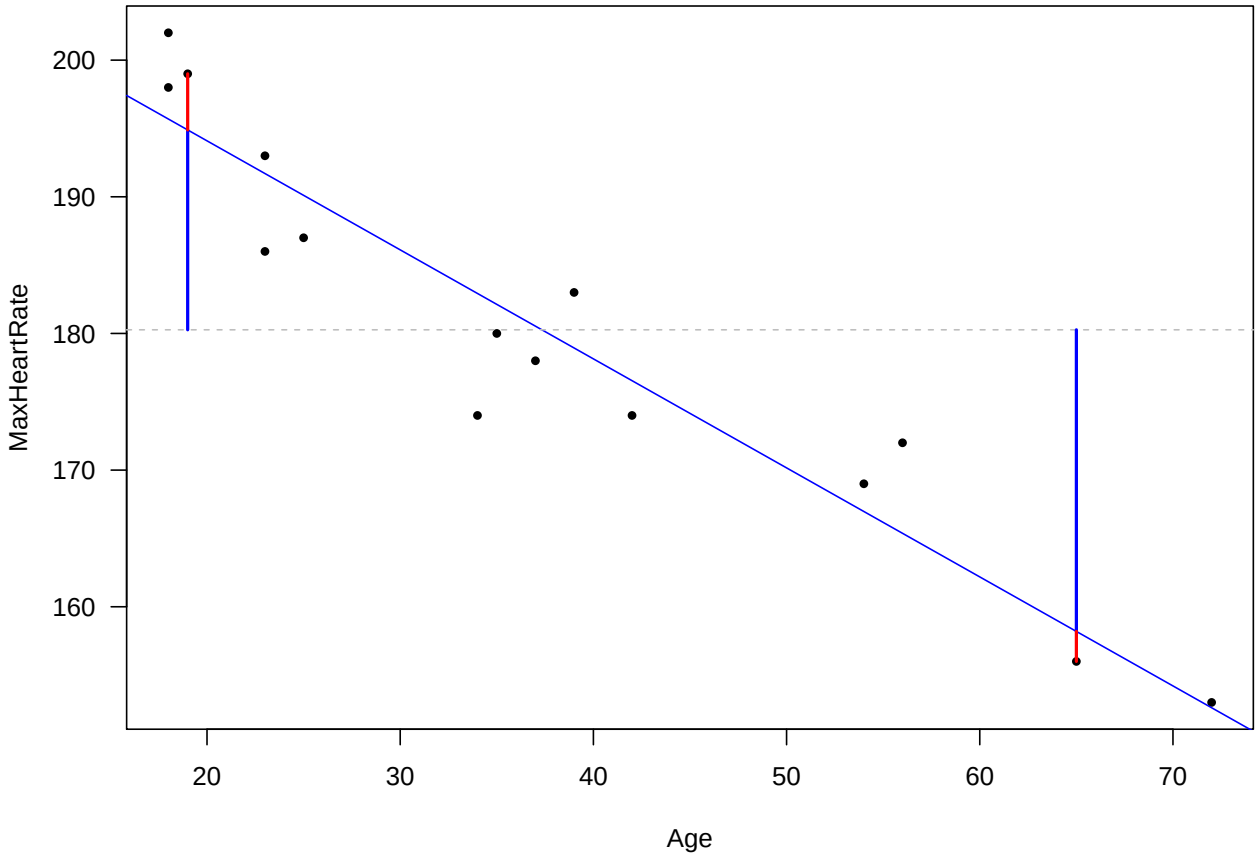
```
##
## Call:
## lm(formula = MaxHeartRate ~ Age)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.9258 -2.5383  0.3879  3.1867  6.6242
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 210.04846    2.86694   73.27  < 2e-16 ***
## Age          -0.79773    0.06996  -11.40 3.85e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.578 on 13 degrees of freedom
```

```
## Multiple R-squared:  0.9091, Adjusted R-squared:  0.9021
## F-statistic:   130 on 1 and 13 DF,  p-value: 3.848e-08
```

ANOVA

Partitioning total sums of squares

```
par(las = 1)
plot(Age, MaxHeartRate, pch = 16, cex = 0.75)
abline(fit, col = "blue")
abline(h = mean(MaxHeartRate), col = "gray", lty = 2)
pred <- fit$coefficients[1] + fit$coefficients[2] * 65
segments(65, mean(MaxHeartRate), 65, pred, col = "blue", lwd = 2)
segments(65, pred, 65, 156, col = "red", lwd = 2)
pred <- fit$coefficients[1] + fit$coefficients[2] * 19
segments(19, mean(MaxHeartRate), 19, pred, col = "blue", lwd = 2)
segments(19, pred, 19, 199, col = "red", lwd = 2)
```



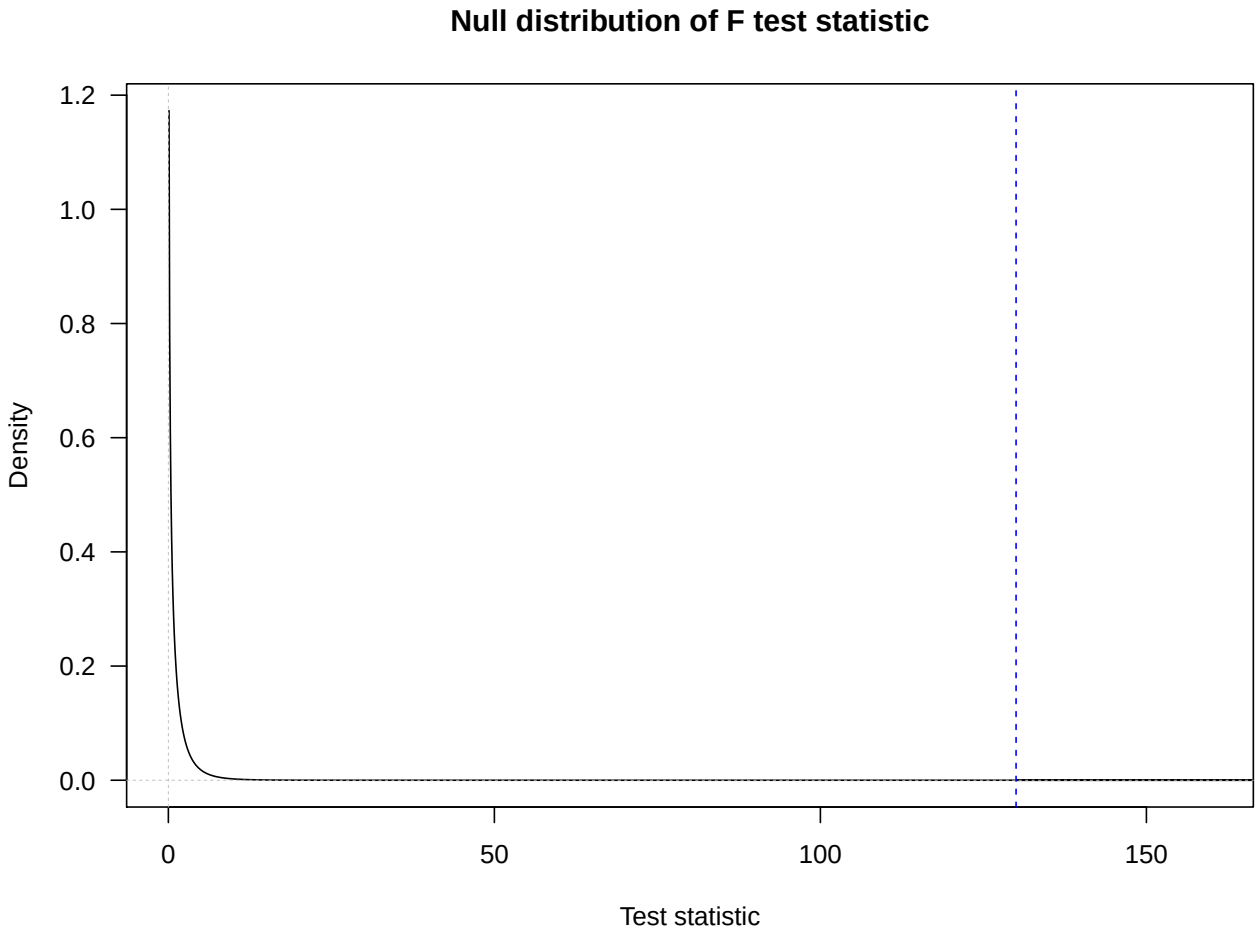

```
anova(fit)
```

```
## Analysis of Variance Table
##
## Response: MaxHeartRate
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## Age        1 2724.50 2724.50  130.01 3.848e-08 ***
## Residuals 13  272.43   20.96
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

F test

```
par(las = 1)
x_grid <- seq(0, 200, 0.1)
y_grid <- df(x_grid, 1,  13)
plot(x_grid, y_grid, type = "l", xlab = "Test statistic",
     ylab = "Density", xlim = c(0, 160),
     lwd = 1, main = "Null distribution of F test statistic")
abline(v = 0, lty = 2, col = "gray", lwd = 0.5)
F_star = 130.01
polygon(c(x_grid[x_grid > F_star], rev(x_grid[x_grid > F_star])),
        c(y_grid[x_grid > F_star], rep(0, length(y_grid[x_grid > F_star]))), col = "skyblue")
abline(v = 130.01, col = "blue", lty = 2)
abline(h = 0, col = "gray", lty = 2, lwd = 0.5)
```




```
anova(fit)$"Pr(>F)"[1] == summary(fit)[["coefficients"]][, 4][2]
```

```
##  Age
## TRUE
```

Correlation and Coefficient of Determination

```
cor(Age, MaxHeartRate)
```

```
## [1] -0.9534656
```

```
summary(fit)$ r.squared
```

```
## [1] 0.9090967
```