STAT 8020 R Lab 2: Simple Linear Regression II

Whitney

November 23, 2020

Contents

Maximum Heart Rate vs. Age

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)
```

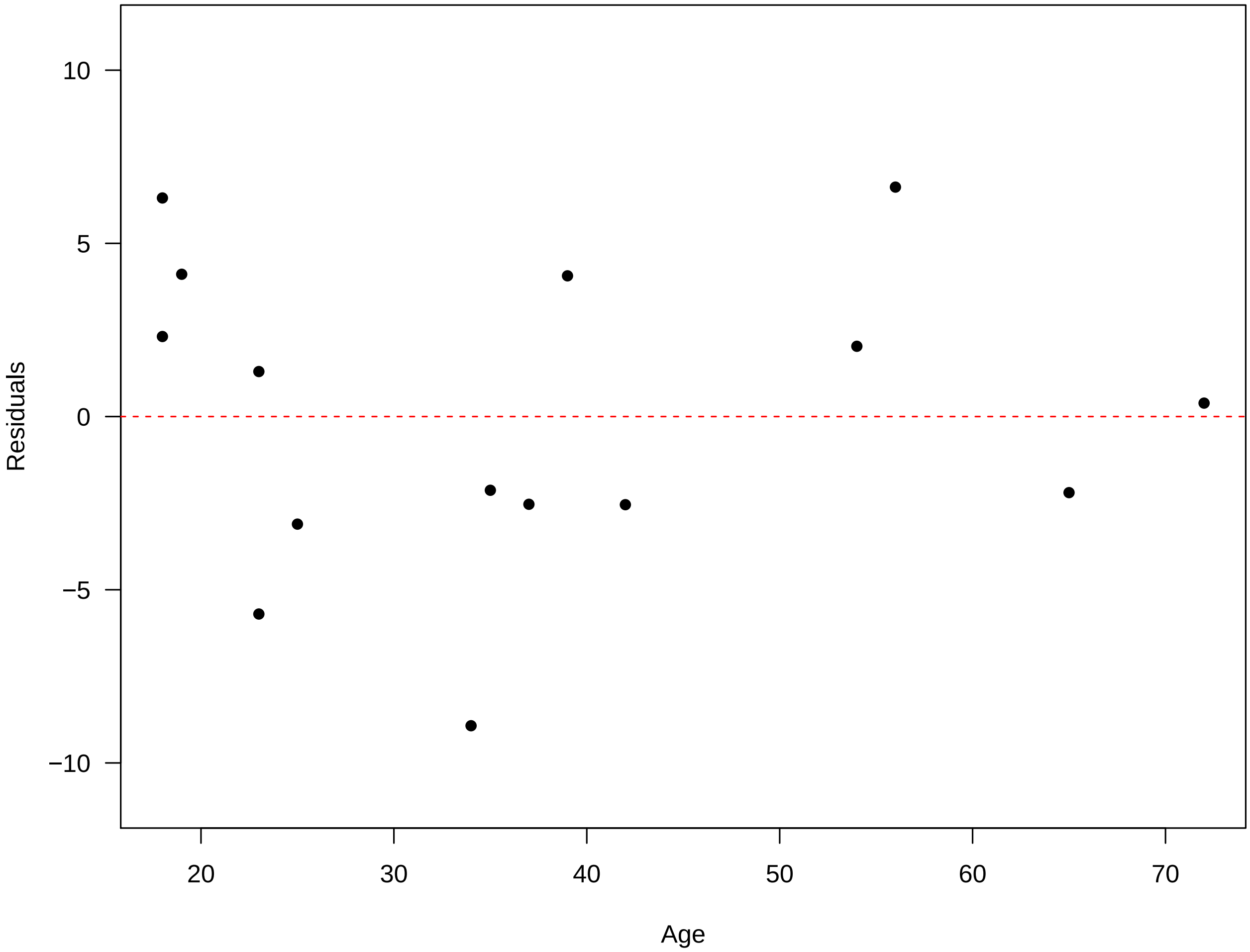
Fitting a simple linear regression

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

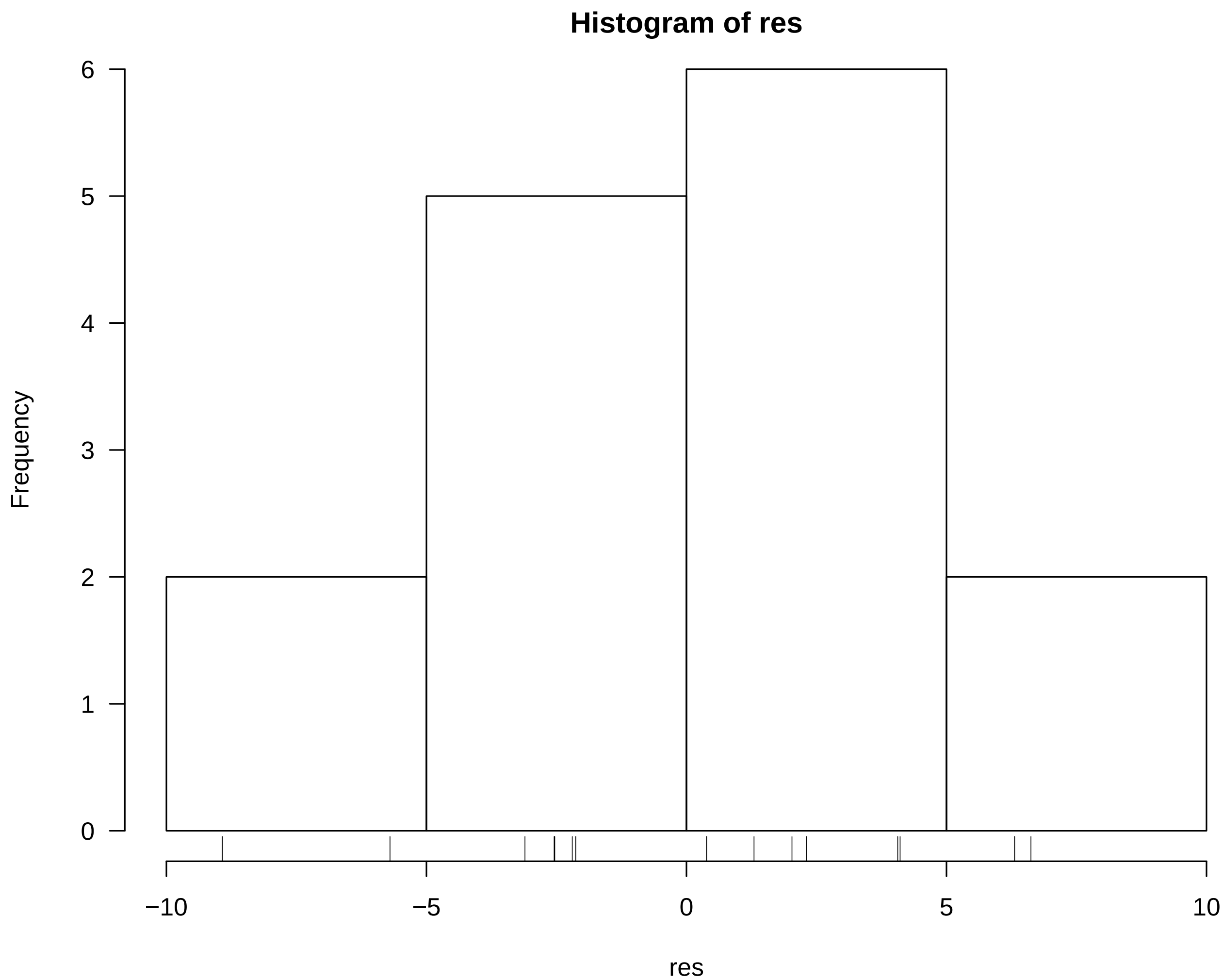
```
##
## Call:
## lm(formula = MaxHeartRate ~ Age)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.9258 -2.5383  0.3879  3.1867  6.6242
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 210.04846    2.86694   73.27  < 2e-16 ***
## Age          -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
```

Residual Analysis

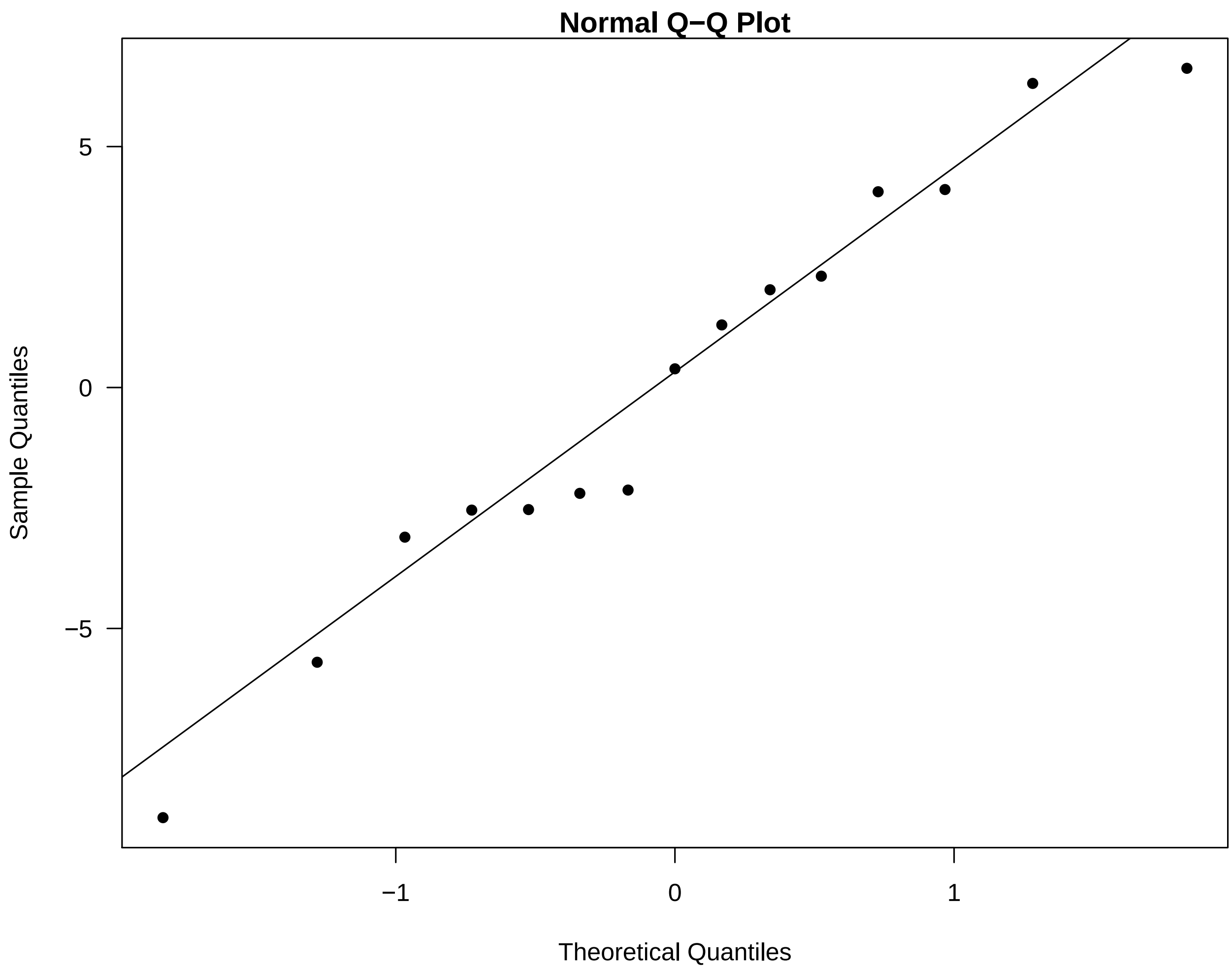
```r
par(las = 1, mar = c(4.1, 4.1, 1.1, 1.1))
plot(Age, fit$residuals, pch = 16, ylab = "Residuals", ylim = c(-11, 11))
abline(h = 0, col = "red", lty = 2)
```




```r
res <- fit$residuals
# histogram
hist(res, las = 1)
rug(res)
```




```
# QQ plot
qqnorm(res, pch = 16, las = 1)
qqline(res)
```



Confidence Interval

β_1

```r
beta1_hat <- summary(fit)[["coefficients"]][, 1][2]
se_beta1 <- summary(fit)[["coefficients"]][, 2][2]
alpha = 0.05
CI_beta1 <- c(beta1_hat - qt(1 - alpha / 2, 13) * se_beta1,
              beta1_hat + qt(1 - alpha / 2, 13) * se_beta1)
CI_beta1
```

```
##        Age        Age
## -0.9488720 -0.6465811
```

```r
confint(fit)
```

```
##                   2.5 %      97.5 %
## (Intercept) 203.854813 216.2421034
## Age          -0.948872  -0.6465811
```

$Y_h | X_h = 40$

```r
Age_new = data.frame(Age = 40)
hat_Y <- fit$coefficients[1] + fit$coefficients[2] * 40
hat_Y
```

```
## (Intercept)
##    178.1394
```

```r
predict(fit, Age_new, interval = "confidence")
```

```
##        fit      lwr      upr
## 1 178.1394 175.5543 180.7245
```

```r
predict(fit, Age_new, interval = "predict")
```

```
##        fit      lwr      upr
## 1 178.1394 167.9174 188.3614
```

Check

```r
sd <- sqrt((sum(fit$residuals^2) / 13))
ME <- qt(1 - alpha / 2, 13) * sd * sqrt(1 + 1 / 15 + (40 - mean(Age))^(2) / sum((Age - mean(Age))^2))
c(hat_Y - ME, hat_Y + ME)
```

```
## (Intercept) (Intercept)
##    167.9174    188.3614
```

Hypothesis Tests for β_1

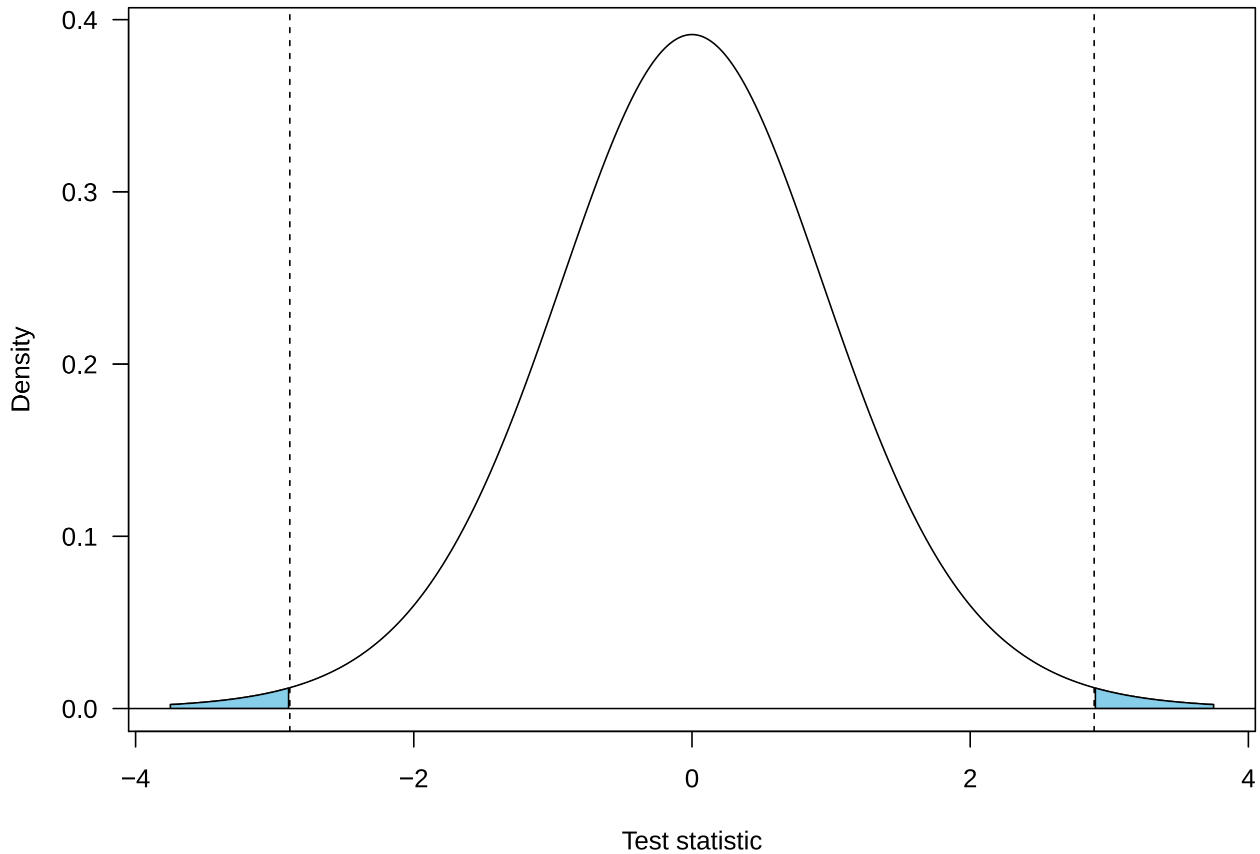
$H_0: \beta_1 = -1$ vs. $H_a: \beta_1 \neq -1$ with $\alpha = 0.05$

```r
beta1_null <- -1
t_star <- (beta1_hat - beta1_null) / se_beta1
p_value <- 2 * pt(t_star, 13, lower.tail = F)
p_value
```

```
##        Age
## 0.01262031
```

```r
par(las = 1)
x_grid <- seq(-3.75, 3.75, 0.01)
y_grid <- dt(x_grid, 13)
plot(x_grid, y_grid, type = "l", xlab = "Test statistic", ylab = "Density", xlim = c(-3.75, 3.75))
polygon(c(x_grid[x_grid < -t_star], rev(x_grid[x_grid < -t_star])),
        c(y_grid[x_grid < -t_star], rep(0, length(y_grid[x_grid < -t_star]))), col = "skyblue")

polygon(c(x_grid[x_grid > t_star], rev(x_grid[x_grid > t_star])),
        c(y_grid[x_grid > t_star], rep(0, length(y_grid[x_grid > t_star]))), col = "skyblue")
abline(v = t_star, lty = 2)
abline(v = -t_star, lty = 2)
abline(h = 0)
```


0.4
0.3
0.2
0.1
0.0
Density
−4
−2
0
2
4
Test statistic