lab1

Claudius Taylor, Junpu Zhao, Thomas Wilson

September 8, 2018

```r
#install.packages('data.table')
library(data.table)
```

STAT 5310

Simple Linear Regression

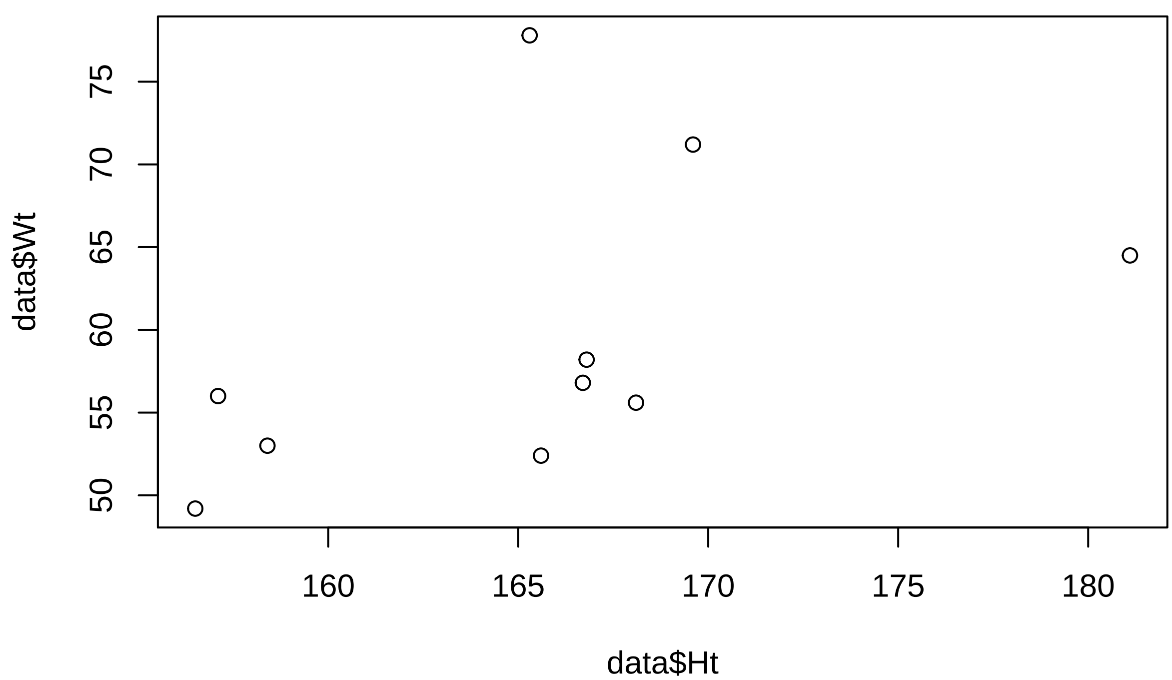
Lab 1 / Homework #1

1. Height and weight data (Data file: Htwt) The data file give ht = height in centimeters and wt = weight in kilograms for a sample of n = 10 18-year-old girls. The data are taken from a larger study. Interest is in predicting weight from height.

```r
data = fread('https://raw.githubusercontent.com/wilsonify/AppliedRegression/master/data/htwt.txt')
```

a. Draw a scatterplot of wt on the vertical axis versus ht on the horizontal axis. On the basis of this plot, does a simple linear regression model make sense for these data? Why or why not?

```r
plot(x=data$Ht,y=data$Wt)
```



Visually, there is a small positive correlation between Ht and Wt. Using simple linear regression to predict weight from height may not be very accurate.

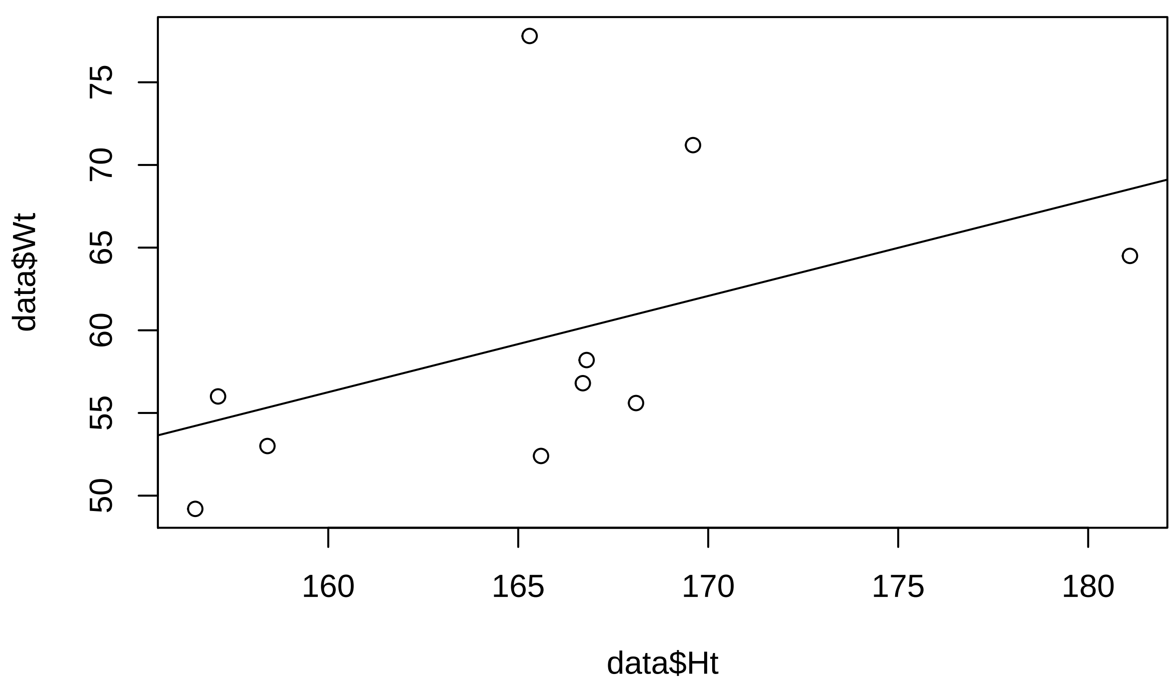
b. Compute estimates of the slope and the intercept for the regression of Y on X. Draw the fitted line on your scatterplot.

```
fit <- lm(data = data,formula = Wt~Ht)
summary_of_fit <- summary(fit)
coef_of_summary <- coef(summary_of_fit)
anova_of_fit <- anova(fit)
```

```
part_b <- fit$coefficients
part_b
```

```
## (Intercept)          Ht
##   -36.87588     0.58208
```

```
plot(x=data$Ht,y=data$Wt)
abline(fit)
```




```
<dt>(Intercept)</dt>
    <dd>-36.8758822731935</dd>
<dt>Ht</dt>
    <dd>0.582080004067143</dd>
```

c. Obtain the estimate of σ^2 and find the estimated standard errors of $\hat{\beta}_0$ and $\hat{\beta}_1$. Compute the t-tests for the hypotheses that $\beta_0 = 0$ and that $\beta_1 = 0$ and find the appropriate p-values using two-sided tests.

```r
part_c <- summary_of_fit$sigma^2
part_c
```

```
## [1] 71.5017
```

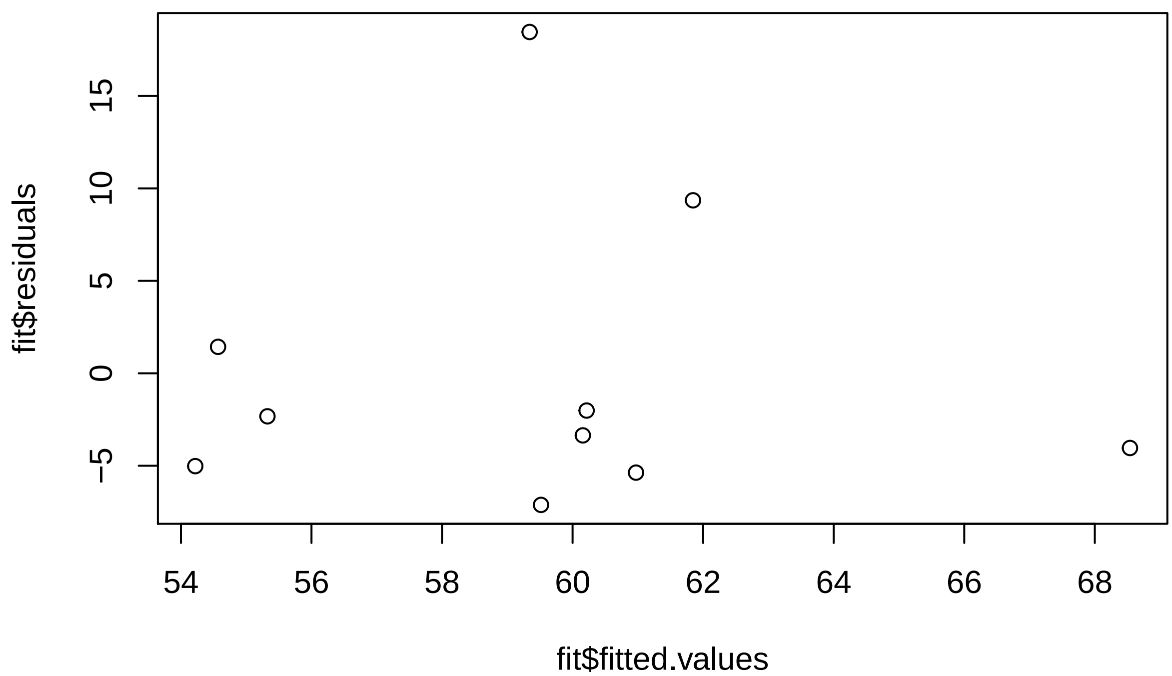
```r
coef(summary_of_fit)
```

```
##                Estimate Std. Error    t value  Pr(>|t|)
## (Intercept) -36.87588 64.4728000 -0.5719603 0.5830589
## Ht            0.58208  0.3891815  1.4956517 0.1731089
```

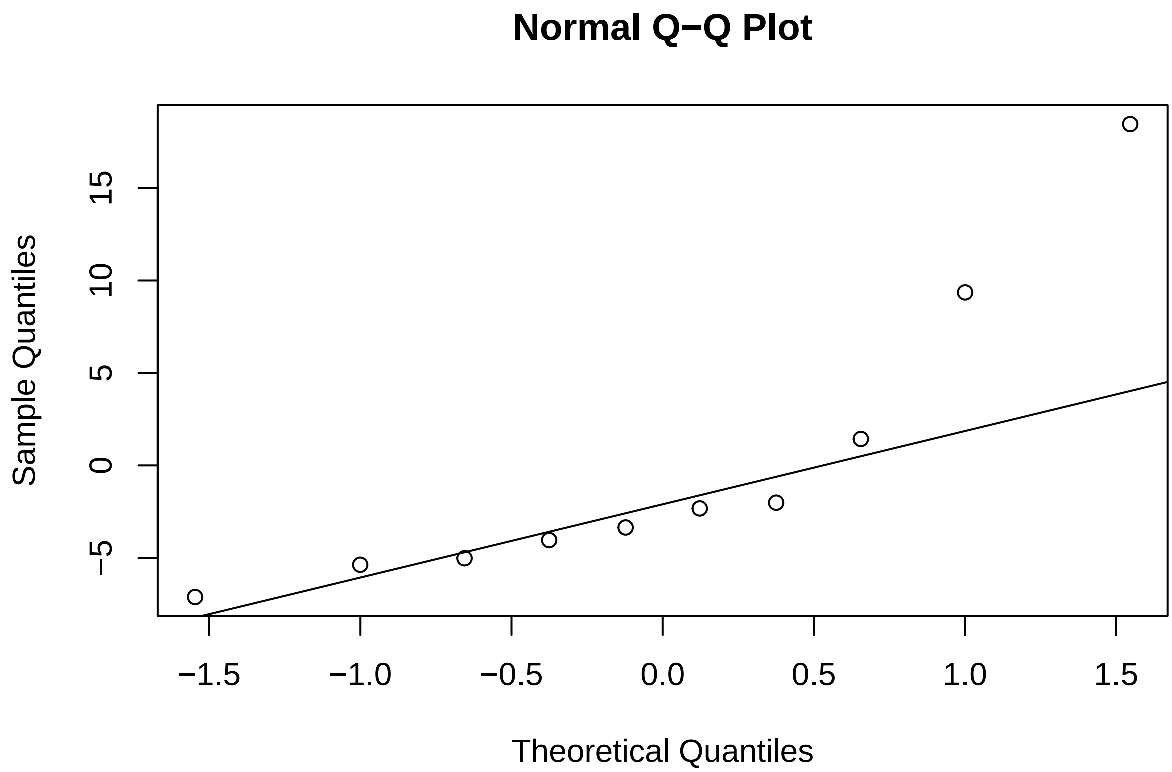
71.5016955003007

d. Examine the residuals and discuss.

```r
plot(y=fit$residuals,x=fit$fitted.values)
```




```r
qqnorm(y=fit$residuals)
qqline(y=fit$residuals)
```



The residuals are mostly less than zero with 2 or three large outliers above zero. This does not conform well to the assumptions of simple linear regression.

```
#check answers
x <- data$Ht
y <- data$Wt
n <- length(x)
df <- n-2
sxx   <- sum((x-mean(x))^2)
syy   <- sum((y-mean(y))^2)
sxy   <- sum((x-mean(x))*(y-mean(y)))
B1    <- sxy/sxx
B0    <- mean(y)-B1*mean(x)
predict_y <- B0+B1*x
rss <- sum((y-predict_y)^2)
s <- sqrt(rss/df)
s_B1 <- s/sqrt(sxx)
B1_0 <- 0 #null hypothesis
t <- (B1-B1_0)/s_B1
p_value <- pt(-abs(t),df)+(1-pt(abs(t),df)) #two-tails
```