

# Stats 101A Homework 1 Problem 3

Damien Ha

01/13/2023

## Contents

Loading Data	1
Part 1	1
Part 2	2
Part 3	2

## Loading Data

```
heart <- read.csv("Heart.csv")
heart = subset(heart, select = -c(X) )
```

## Part 1

Conduct simple linear regression using MaxHR as outcome variable and Chol as a predictor.

- Report the summary of your linear model, i.e. write a regression equation.
- Interpret the slope and the y-intercept in the context of data.

```
model <- lm(MaxHR ~ Chol, data = heart)
summary(model)
```

```
##
## Call:
## lm(formula = MaxHR ~ Chol, data = heart)
##
## 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
```

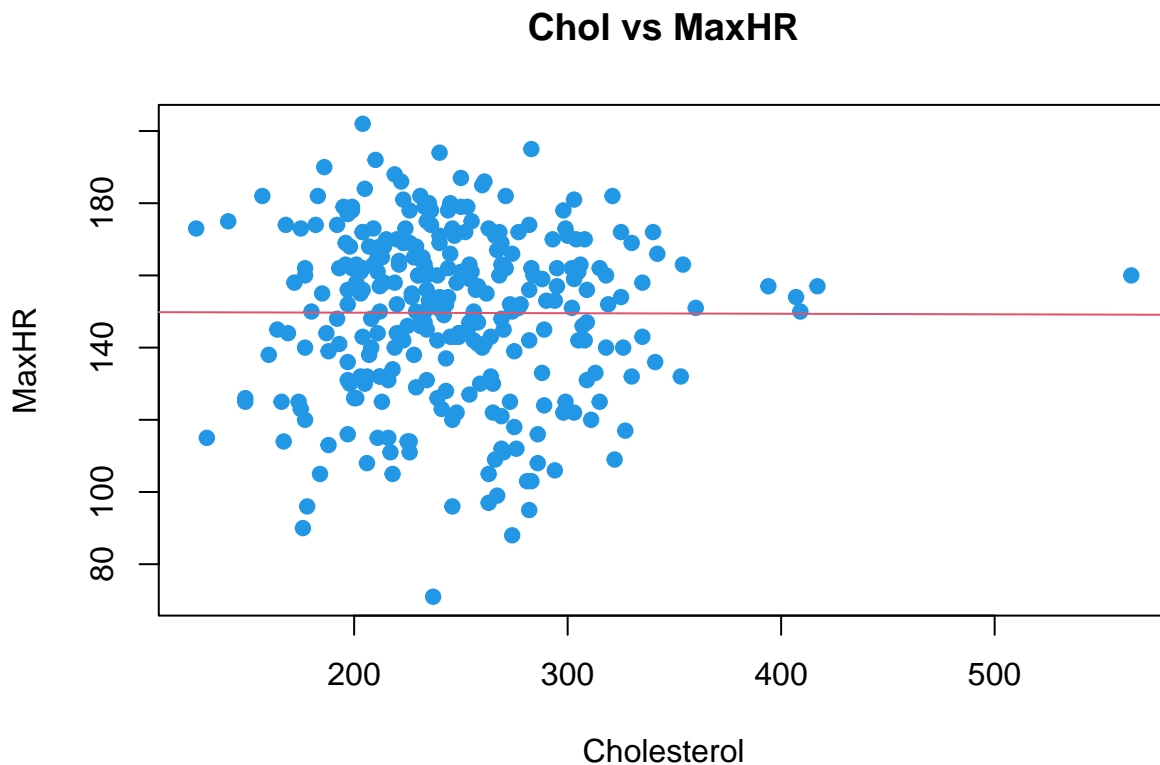
```
## Multiple R-squared:  1.178e-05, Adjusted R-squared:  -0.00331
## F-statistic: 0.003545 on 1 and 301 DF,  p-value: 0.9526
```

The slope is -0.0015 and the intercept is 149.98. For an increase of 1 unit in cholesterol, we expect a decrease of 0.0015 units in MaxHR on average. At a cholesterol level of 0, we expect a MaxHR of 149.98 on average.

## Part 2

Create a scatter plot for the Chol vs MaxHR then plot the least square regression line on the same graph

```
plot(heart$Chol, heart$MaxHR,
     xlab = "Cholesterol", ylab = "MaxHR", main = "Chol vs MaxHR",
     pch = 19, col = 4)
abline(model, col = 2)
```



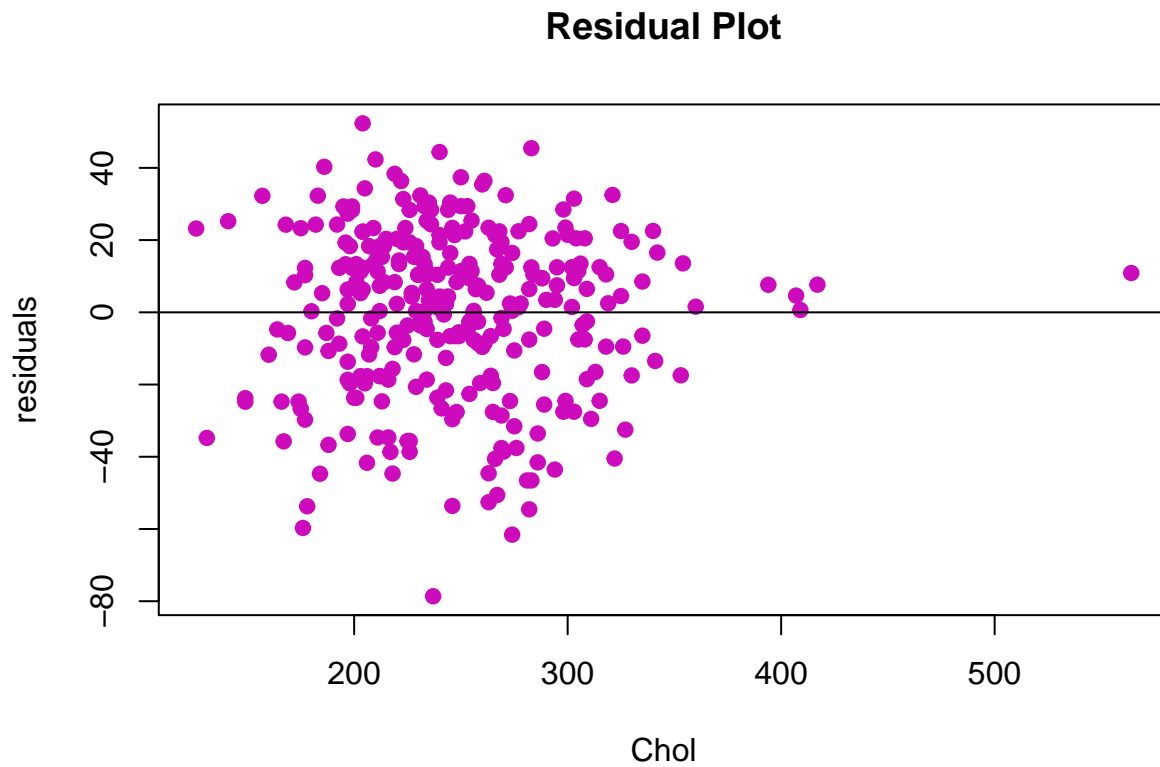
## Part 3

Report R-square and create a residual plot. How do you assess the goodness-of-fit of the model?

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

```
## [1] 1.177747e-05
```

```
plot(heart$Chol, model$residuals,
     xlab = "Chol", ylab = "residuals", main = "Residual Plot",
     pch = 19, col = 6)
abline(0,0)
```



From the low R-squared value, it seems that MaxHR is not explained very much by the linear model. The residual plot does seem to be random without any patterns, but the data points in the scatterplot are not clustered very closely to the line. The model does not have a very good fit.