Rnash-Assing11.rmd

Raphael Nash

4/10/2017

Problem 1

**Using R's lm function, perform regression analysis and measure the significance of the independent variables for the following two data sets. In the first case, you are evaluating the statement that we hear that Maximum Heart Rate of a person is related to their age by the following equation:

MaxHR = 220 - Age**

```
hr_df= data.frame(
  Age = c(18,23,25,35,65,54,34,56,72,19,23,42,18,39,37),
  MaxHR = c(202,186,187,180,156,169,174,172,153,199,193,174,198,183,178)
)

hr_model <- lm (hr_df$MaxHR~ hr_df$Age)
summary(hr_model)
```

```
##
## Call:
## lm(formula = hr_df$MaxHR ~ hr_df$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 ***
## hr_df$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
```

What is the resulting equation?

$$y = -0.7977266 \times Age - 210.0484584$$

```
hr_model$coefficients[[2]]
```

```
## [1] -0.7977266
```

```
hr_model$coefficients[[1]]
```

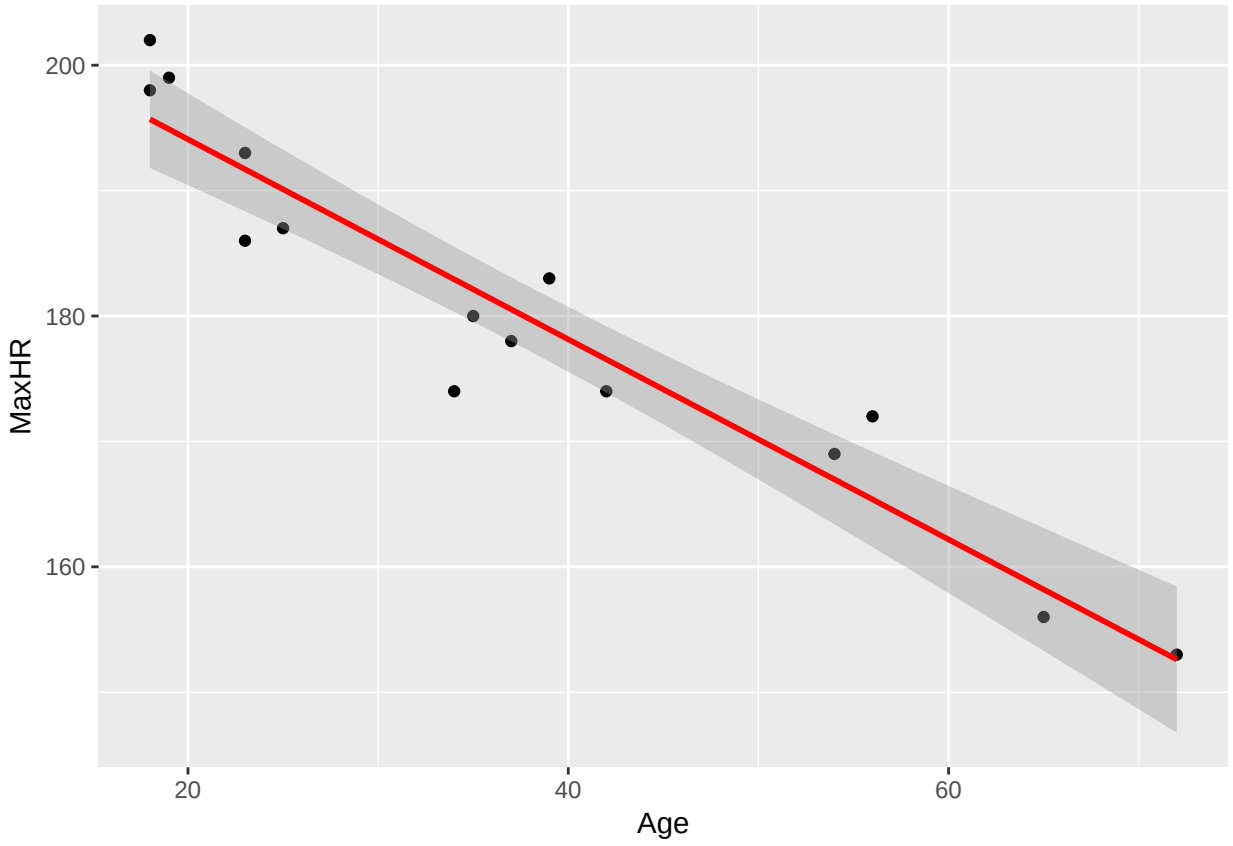
```
## [1] 210.0485
```

What is the signiffcance level?

```
r hr_model_stats$coefficients[2,4]
```

Please also plot the fitted relationship between Max HR and Age.

```
ggplot(data=hr_df, aes(y=MaxHR, x= Age)) +geom_point() + stat_smooth(method = "lm", col = "red")
```



Problem 2

Using the Auto data set from Assignment 5 (also attached here) perform a Linear Re- gression analysis using mpg as the dependent variable and the other 4 (displacement, horse- power, weight, acceleration) as independent variables. What is the final linear regression fit equation?

Perform the following by a) using the all the data, b) Using a sample of 40 data points:

```
mpg_data <- read.table("https://raw.githubusercontent.com/RaphaelNash/CUNY-DATA-605-CompMath/master/HW1
names(mpg_data) <- c('displacement', 'horsepower', 'weight', 'acceleration', 'mpg')
```

```
set.seed(1)
mpg_sample_data <- mpg_data[sample(nrow(mpg_data), 40), ]
mpg_sample_model <- lm(mpg ~ ., mpg_sample_data)
mpg_sample_model_ci <- confint(mpg_sample_model)
mpg_sample_model_summary <- summary(mpg_sample_model)
mpg_sample_model_coef <- mpg_sample_model_summary$coefficients


mpg_sample_model_sig <- mpg_sample_model_coef[, "Pr(>|t|)"]
mpg_sample_model_summary
```

```
##
## Call:
## lm(formula = mpg ~ ., data = mpg_sample_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.1192 -2.3122 -0.1993  2.2986 11.4903
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  33.108223   8.973636   3.689 0.000759 ***
## displacement  0.014435   0.017632   0.819 0.418513
## horsepower   -0.044045   0.049293  -0.894 0.377677
## weight       -0.006401   0.002457  -2.606 0.013376 *
## acceleration  0.744863   0.456644   1.631 0.111822
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.656 on 35 degrees of freedom
## Multiple R-squared:  0.7682, Adjusted R-squared:  0.7417
## F-statistic:    29 on 4 and 35 DF,  p-value: 1.118e-10
```

```
mpg_model <- lm(mpg ~ ., mpg_data)
mpg_model_ci <- confint(mpg_model)
mpg_model_summary <- summary(mpg_model)
mpg_model_coef <- mpg_model_summary$coefficients
mpg_model_sig <- mpg_model_coef[, "Pr(>|t|)"]
mpg_model_summary
```

```
##
## Call:
## lm(formula = mpg ~ ., data = mpg_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -11.378  -2.793  -0.333   2.193  16.256
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  45.2511397  2.4560447  18.424  < 2e-16 ***
## displacement -0.0060009  0.0067093  -0.894  0.37166
## horsepower   -0.0436077  0.0165735  -2.631  0.00885 **
## weight       -0.0052805  0.0008109  -6.512  2.3e-10 ***
## acceleration -0.0231480  0.1256012  -0.184  0.85388
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.247 on 387 degrees of freedom
## Multiple R-squared:  0.707,  Adjusted R-squared:  0.704
## F-statistic: 233.4 on 4 and 387 DF,  p-value: < 2.2e-16
```

What is the final linear regression fit?

Sample Model: mpg = (0.0144347) displacement + (-0.0440451) horsepower + (-0.0064011) weight + (0.7448631) acceleration + (33.1082233)

Full Model: mpg = (-0.0060009) displacement + (-0.0436077) horsepower + (-0.0064011) weight + (-0.023148) acceleration + (45.2511397)

Which of the 4 independent variables have a significant impact on mpg?

```
sig <- cbind(mpg_sample_model_sig, mpg_model_sig)
sig
```

```
##              mpg_sample_model_sig mpg_model_sig
## (Intercept)          0.0007585585  7.072099e-55
## displacement         0.4185132624  3.716584e-01
## horsepower           0.3776765861  8.848982e-03
## weight               0.0133762653  2.302545e-10
## acceleration         0.1118217353  8.538765e-01
```

Sample: weight only

Full Model: displacement, weight, acceleration

What are the standard errors on each of the coeffcients?

Sample Model:

```
standard_errors <- rbind ( mpg_sample_model_coef[,'Std. Error'], mpg_model_coef[,'Std. Error'])
rownames(standard_errors) <- c('sample_model', 'full_model')
standard_errors
```

```
##              (Intercept) displacement horsepower       weight acceleration
## sample_model    8.973636  0.017631914 0.04929332 0.0024566133    0.4566436
## full_model      2.456045  0.006709306 0.01657346 0.0008108541    0.1256012
```

measure the 95% confidence intervals.

```
colnames( mpg_sample_model_ci) <-c('low_sample_model', 'high_sample_model')
colnames( mpg_model_ci) <-c('low_full_model', 'high_full_model')
ci <- cbind(mpg_sample_model_ci, mpg_model_ci)
```

Conclusions 1) The size of the confidence intervals is smaller on the full model 2) The standard error is smaller on the full model 3) The number of signifigant factors is higher on the full model