STAT 511A Homework 11

Kathleen Wendt

12/09/2019

Load packages

```
library(readxl)
library(janitor)
library(tidyverse)
library(broom)
```

Question 1

Review problem 11.22 from Ott & Longnecker regarding treadmill "time to exhaustion" (X) and 10km race times (Y).

```
run_data <- readxl::read_xlsx("ex11-22.xlsx") %>%
  janitor::clean_names() %>%
  dplyr::rename("race_time" = "x10_k")
```

Part 1A

Regress 10.K (Y) on Treadmill (X) and include the "summary" information in your assignment.

```
run_fit <- lm(race_time ~ treadmill, data = run_data)
summary(run_fit)
```

```
##
## Call:
## lm(formula = race_time ~ treadmill, data = run_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.9440 -1.5788  0.1860  0.7863  4.5603
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  59.9211     3.1166  19.226 1.90e-13 ***
## treadmill    -1.9601     0.3164  -6.194 7.59e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.921 on 18 degrees of freedom
## Multiple R-squared:  0.6807, Adjusted R-squared:  0.6629
## F-statistic: 38.37 on 1 and 18 DF,  p-value: 7.589e-06
```

Part 1B

Create a scatterplot of 10-K vs Treadmill with fitted regression line overlaid.

Part 1C

Give the estimate, 95% confidence interval and interpretation of the slope. (4 pts)

Part 1D

Give the R^2 value and interpretation in terms of this scenario.

Part 1E

Give the predicted 10.K time for a runner with Treadmill = 11. Also provide a corresponding prediction interval.

Part 1F

Create the plots of (1) residuals vs fitted values and (2) qqplot of residuals

Part 1G

Based on the plots above, subject 13 appears to be a bit of an outlier. Run a formal outlier test for this observation. Provide the p-value and make a conclusion. Note that since we identified this observation after looking at the data, a Bonferonni adjustment is appropriate.

Question 2

Data on age in coating Thickness (X) and Strength (Y) from an experiment involving steel are available from Canvas as Steel.csv.

Part 1A

Regress Strength (Y) against Thick (X) and look at (1) the plot of Strength versus Thick (2) residuals versus predicted values and (3) qqplot of residuals. Include these plots in your assignment. Do the regression assumptions appear to be met? Discuss. (4 pts)

Part 1B

Perform an F-test for "lack of fit". Give your p-value and make a conclusion. (4 pts)

Part 1C

Now perform a quadratic regression and create a scatterplot with the fitted curve overlaid. Include the "summary" table and plot in your assignment. This can be done with code like the following. (4 pts)

```
# QFit <- lm(Strength ~ Thick + I(Thick^2), data = Steel)
# summary(QFit)
# plot(Strength ~ Thick, data = Steel)
# curve(b0 + b1*x + b2*x^2, add = TRUE)
```

Note that b0,b1,b2 need to be *replaced* with estimates from the quadratic regression.

Question 3

Review problem 11.50 from Ott & Longnecker regarding SAT Scores.

Part 1A

Create pairwise scatterplots for all 4 variables (Male.Verbal, Female.Verbal, Male.Math, Female.Math)

Part 1B

Calculate pairwise (Pearson) correlations for all 4 variables. Which pair of variables has the strongest correlation? (4 pts)

Part 1C

Provide a test of the correlation for Female.Verbal vs Female.Math. Give the p-value and conclusion in your assignment.