

HW05 - Body Measurements

Stat 131A, Spring 2019

General Instructions

- Write your narrative and code in an Rmd (R markdown) file.
 - Name this file as `hw05-first-last.Rmd`, where `first` and `last` are your first and last names (e.g. `hw05-gaston-sanchez.Rmd`).
 - Please do not use code chunk options such as: `echo = FALSE`, `eval = FALSE`, `results = 'hide'`. All chunks must be visible and evaluated.
 - Submit your Rmd and html files to bCourses.
-

Body Measurements Data Set

You will be working with the *Body Measurements Data Set* which contains 39 variables measured on 507 individuals. The data file is available in the course's github repository. To import the data in R as a data frame, execute the following commands:

```
# assembling the URL of the CSV file
# (otherwise it won't fit within the margins of this document)
repo = 'https://raw.githubusercontent.com/ucb-introstat/introstat-fall-2018/'
datafile = 'master/data/body-measurements.csv'
url = paste0(repo, datafile)

# read in data set
body = read.csv(url)
```

In particular you will be working with three variables: `Height`, `Weight`, and `Ankle_girth`. Ankle girth and height are measured in centimeters; weight is measured in kilograms.

```
# take a peek to the variables of interest
head(body[,c('Height', 'Weight', 'Ankle_girth')])
```

```
##   Height Weight Ankle_girth
## 1  174.0   65.6         23.5
## 2  175.3   71.8         24.5
## 3  193.5   80.7         21.9
## 4  186.5   72.6         23.0
## 5  187.2   78.8         24.4
## 6  181.5   74.8         23.5
```

Research question

You will have to perform a comprehensive descriptive analysis describing the relationship between ankle girth and height, and the relationship between ankle girth and weight. More specifically, the research question to be addressed is:

Is ankle girth a better predictor of height or weight?

You will decide whether **Ankle_girth** (distance around the ankle bone) is a better predictor for **Height** or for **Weight**.

Specifications

- (1) **Thesis statement:** Develop a thesis statement that addresses the research question. This should be a clear position about which response variable is better explained by the predictor variable.
- (2) **Exploratory Data Analysis (EDA):** Explore the data by making graphs and doing calculations that make sense to you.
 - You should begin with a descriptive statistical analysis of each of the three variables of interest (individually).
 - Some of the things you can try are: plots (e.g. histograms, boxplots), and numeric summaries (e.g. centers, spreads, 5-number summary).
- (3) **Association between Ankle_girth and Height, as well as between Ankle_girth and Weight**
 - After the EDA, move on to analyze the relationship between ankle girth and height, and the relationship between ankle girth and weight.
 - Include two scatterplots with explanatory variable on the x-axis, and response variable on the y-axis. Clear and accurate description of the form, direction and strength of each relationship in the two scatterplots.
 - Correlations: clear and accurate discussion of how the sign of the correlation coefficient relates to direction and the value of the correlation relates to strength of the relationship.
- (4) **Linear Regression Equations:** Calculate the regression equations.
 - Include clear and accurate interpretation of the slope and intercept for the regression lines.
 - predicted height for an individual with an ankle girth of 26 cm
 - predicted weight for an individual with an ankle girth of 26 cm
- (5) **Regression Diagnostic:**

- Inspect the plot of residuals to make sure that the use of a regression line is justified.
- Use of r^2 to comment on the quality of the fitted lines.

(6) **Other Considerations:**

- Write a narrative to summarize your analysis. Do not just simply write R commands with their outputs. You should also write a nice narrative flow.
- Your text should illustrate or explain how your analysis of the data supports your conclusions.