Group #7

Professor Cremona

November 8th, 2017

**Research Questions and Exploratory Data Analysis**

**Introduction for the Data set:**

* Reference:

<http://ww2.amstat.org/publications/jse/jse_data_archive.htm>

* Description:

This dataset contains 21 body dimension measurements as well as age, weight, height, and gender on 507 individuals. The 247 men and 260 women were primarily individuals in their twenties and thirties, with a scattering of older men and women, all exercising several hours a week.

* Source:

Measurements were initially taken by the first two authors - Grete Heinz and Louis J. Peterson - at San Jose State University and at the U.S. Naval Postgraduate School in Monterey, California. Later, measurements were taken at dozens of California health and fitness clubs by technicians under the supervision of one of these authors.

* Variable Descriptions:

The dataset is about the body dimensions of 507 individual with 247 men and 260 women and contains the following variables:

* biacromialD: biacromial diameter (cm)
* biiliacD: biiliac diameter, or "pelvic breadth" (cm)
* bitrochantericD: bitrochanteric diameter (cm)
* chest\_depth: chest depth between spine and sternum at nipple level, mid-expiration (cm)
* chestD: chest diameter at nipple level, mid-expiration (cm)
* elbowD: elbow diameter, sum of two elbows (cm)
* wristD: wrist diameter, sum of two wrists (cm)
* kneeD: knee diameter, sum of two knees (cm)
* ankleD: ankle diameter, sum of two ankles (cm)
* shoulderG: shoulder girth over deltoid muscles (cm)
* chestG: chest girth, nipple line in males and just above breast tissue in females, mid-expiration (cm)
* waistG: waist girth, narrowest part of torso below the rib cage, average of contracted and relaxed position (cm)
* navelG: navel (or "Abdominal") girth at umbilicus and iliac crest,  iliac crest as a landmark (cm)
* hipG: hip girth at level of bitrochanteric diameter (cm)
* highG: thigh girth below gluteal fold, average of right and left girths (cm)
* bicepG: bicep girth, flexed, average of right and left girths (cm)
* forearmG: forearm girth, extended, palm up, average of right and left girths (cm)
* kneeG: knee girth over patella, slightly flexed position, average of right and left girths (cm)
* calfMaxG: calf maximum girth, average of right and left girths (cm)
* ankleMinG: ankle minimum girth, average of right and left girths (cm)
* wristMinG: wrist minimum girth, average of right and left girths (cm)
* age: age (years)
* weight: weight (kg)
* height: height (cm)
* gender: gender (1 - male, 0 - female)

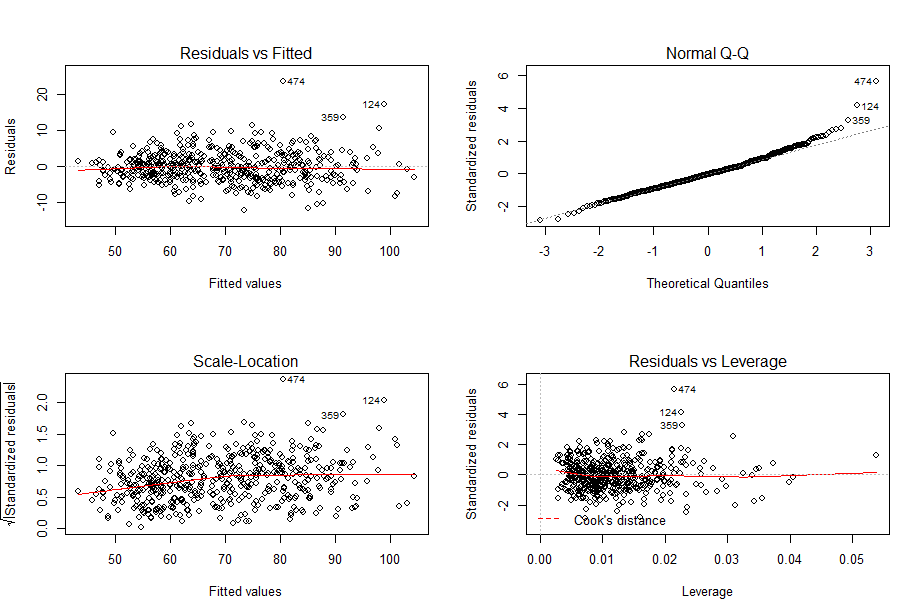
**Research Questions:**

* Objective:

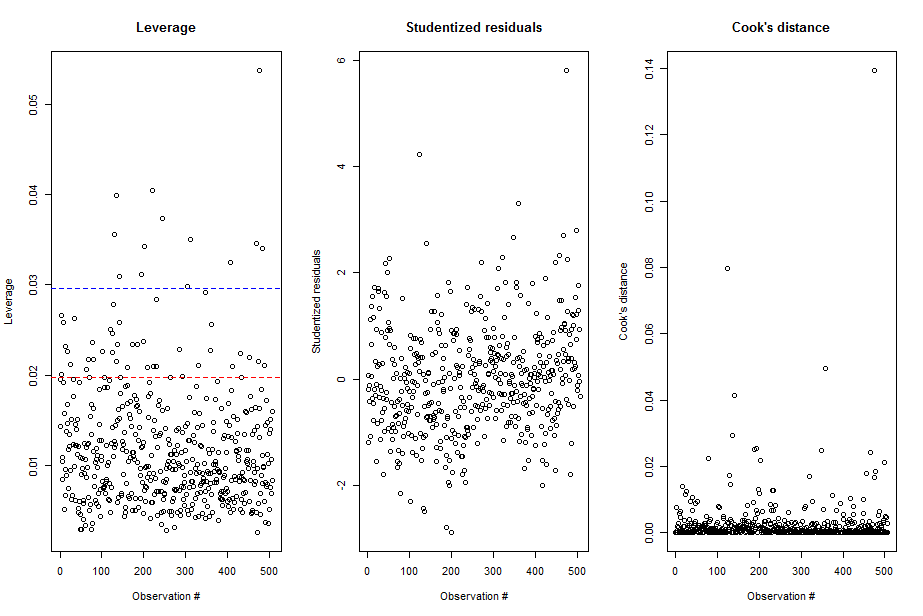
Setting up models for how one’s weight can be related to some factors of one’s body features (biacromial diameter, chest diameter, chest depth, waist girth, height).

* Questions:
* How do we set up our model?
* Which variables have linear relationship with Weight?
* Do there exist any errors or unusual data?
* Do there exist any interaction effects between different predictors,like colinearity?
* Do we have categorical variables?

**Exploratory data analysis:**



These graphs on the left are plotted by the model without predictors age, height, and gender. We can tell from the scatter plot that the model has approximately constant variance and linearity. However, the Q-Q plot shows a heavy tail on the upper corner. Thus, assumptions on error term are not all satisfied and we want to adjust our data by checking outliers.



The graph above indicates that we need to check outliers and influential point for our data.

> summary(lm\_full)

Call:

lm(formula = weight ~ biacromialD + chest\_depth + chestD + waistG +

height, data = Body\_updated)

Residuals:

Min 1Q Median 3Q Max

-12.0725 -2.7252 -0.1878 2.3284 23.5107

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -76.62971 3.49974 -21.896 < 2e-16 \*\*\*

biacromialD -0.10152 0.11376 -0.892 0.373

chest\_depth 0.62899 0.12852 4.894 1.33e-06 \*\*\*

chestD 0.93390 0.13414 6.962 1.05e-11 \*\*\*

waistG 0.64397 0.03489 18.456 < 2e-16 \*\*\*

height 0.36184 0.03086 11.724 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.222 on 501 degrees of freedom

Multiple R-squared: 0.9009, Adjusted R-squared: 0.8999

F-statistic: 911 on 5 and 501 DF, p-value: < 2.2e-16

> summary(Body\_updated[,c(23,1,4,5,12,24)])

weight biacromialD chest\_depth chestD waistG height

Min. : 42.00 Min. :32.40 Min. :14.30 Min. :22.20 Min. : 57.90 Min. :147.2

1st Qu.: 58.40 1st Qu.:36.20 1st Qu.:17.30 1st Qu.:25.65 1st Qu.: 68.00 1st Qu.:163.8

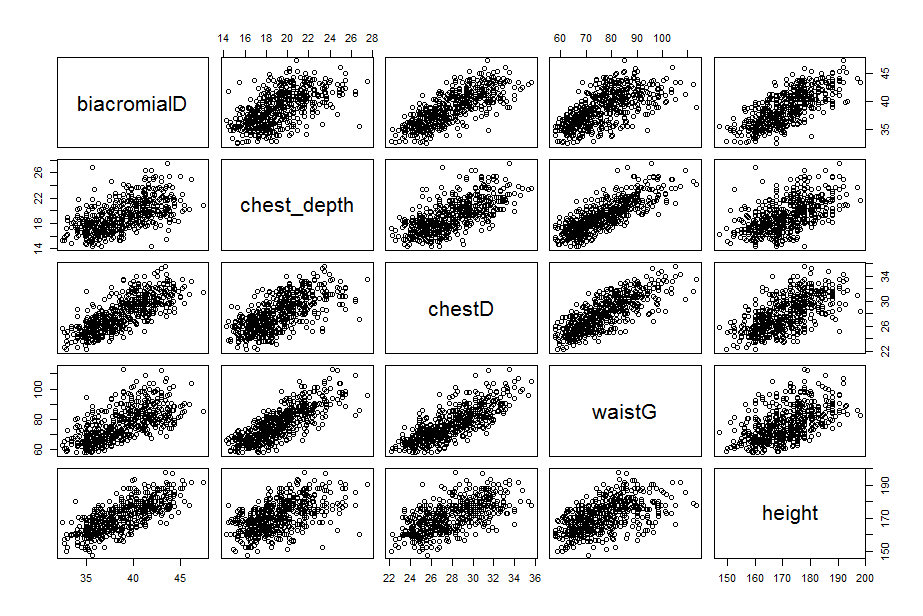
Median : 68.20 Median :38.70 Median :19.00 Median :27.80 Median : 75.80 Median :170.3

Mean : 69.15 Mean :38.81 Mean :19.23 Mean :27.97 Mean : 76.98 Mean :171.1

3rd Qu.: 78.85 3rd Qu.:41.15 3rd Qu.:20.90 3rd Qu.:29.95 3rd Qu.: 84.50 3rd Qu.:177.8

Max. :116.40 Max. :47.40 Max. :27.50 Max. :35.60 Max. :113.20 Max. :198.1

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Weight | Biacromial diameter | Chest depth | Chest diameter | Waist Diameter | Height |
| Standard Deviation | 13.34576 | 3.059132 | 2.515877 | 2.74165 | 11.01269 | 9.407205 |



> cor(Body\_updated[,c(23,c(1,4,5,12,24))])

weight biacromialD chest\_depth chestD waistG height

weight 1.0000000 0.7254145 0.8007315 0.8314645 0.9039908 0.7173011

biacromialD 0.7254145 1.0000000 0.5832585 0.7691406 0.6416072 0.7489218

chest\_depth 0.8007315 0.5832585 1.0000000 0.6650702 0.8037549 0.5529111

chestD 0.8314645 0.7691406 0.6650702 1.0000000 0.7880334 0.6268931

waistG 0.9039908 0.6416072 0.8037549 0.7880334 1.0000000 0.5529605

height 0.7173011 0.7489218 0.5529111 0.6268931 0.5529605 1.0000000