DSA 8020 R Lab 3: Multiple Linear Regression II

your name here

Contents

Percentage of Body Fat and Body Measurements

Age, weight, height, and 10 body circumference measurements are recorded for 252 men. Each man's percentage of body fat was accurately estimated by an underwater weighing technique.

Data Source: Johnson R. *Journal of Statistics Education* v.4, n.1 (1996)

Load the dataset

Code:

```
library(faraway)
data(fat)
head(fat)
```

```
##   brozek siri density age weight height adipos  free neck chest abdom   hip
## 1   12.6 12.3  1.0708  23 154.25  67.75   23.7 134.9 36.2  93.1  85.2  94.5
## 2    6.9  6.1  1.0853  22 173.25  72.25   23.4 161.3 38.5  93.6  83.0  98.7
## 3   24.6 25.3  1.0414  22 154.00  66.25   24.7 116.0 34.0  95.8  87.9  99.2
## 4   10.9 10.4  1.0751  26 184.75  72.25   24.9 164.7 37.4 101.8  86.4 101.2
## 5   27.8 28.7  1.0340  24 184.25  71.25   25.6 133.1 34.4  97.3 100.0 101.9
## 6   20.6 20.9  1.0502  24 210.25  74.75   26.5 167.0 39.0 104.5  94.4 107.8
##   thigh knee ankle biceps forearm wrist
## 1  59.0 37.3  21.9   32.0    27.4  17.1
## 2  58.7 37.3  23.4   30.5    28.9  18.2
## 3  59.6 38.9  24.0   28.8    25.2  16.6
```

```
## 4  60.1 37.3  22.8   32.4    29.4  18.2
## 5  63.2 42.2  24.0   32.2    27.7  17.7
## 6  66.0 42.0  25.6   35.7    30.6  18.8
```

For the purposes of this lab, we will use only the following variables for conducting data analysis:

1. y `brozek`: Percent body fat using Brozek's equation

$$\frac{457}{\text{Density}} - 414.2$$

2. x_1 `age`: Age (yrs);
3. x_2`weight`: Height (inches);
4. x_3 `height`: Height (inches);
5. x_4`chest`: Chest circumference (cm);
6. x_5 `abdom`: Abdomen circumference (cm) at the umbilicus and level with the iliac crest

Code:

You can use the code below to extract these variables:

```
vars <- c("brozek", "age", "weight", "height", "chest", "abdom")
data <- fat[, vars]
```

Exploratory Data Analysis

Numerical summary

1. Use `summary` commend to produce various numerical summaries of each of the 6 variables under consideration

Code:

Graphical summary

2. Make a boxplot for each variable

Code:

3. Briefly discuss the shape of the distribution of each variable

Answer:

4. Create a scatterplot matrix to explore the inter-dependence between these variables

Code:

General Linear F-Test

Suppose a researcher would like to compare the "Full" model using all the 5 predictors and a "reduce" model where only x_1 (`age`) and x_5 (`abdom`) are used by performing a general linear F-test:

5. Write down the null and the alternative hypotheses

Answer:

6. Fit the full model and write down the fitted linear regression equation.

Code:

Answer:

7. Fit the reduced model and write down the fitted linear regression equation.

Code:

Answer:

8. Perform a general linear F-test and state the conclusion at $\alpha = 0.05$

Code:

Answer:

Prediction

9. Predict a future response for an individual with `age` = 54, `weight` = 197, `height` = 72.25, `chest` = 105.375, and `abdom` = 99.325. Construct a 95% prediction interval.

Code:

Answer:

10. Construct a 95% confidence interval for the mean response of percent body fat with `age` = 54, `weight` = 197, `height` = 72.25, `chest` = 105.375, and `abdom` = 99.325.

Code:

Multicollinearity

11. Make the scatterplot matrix and compute the correlation matrix for all 6 variables (including the response).

Code:

12. Calculate VIF and briefly discuss your finding

Code:

Answer: