Module 7: Linear regression

Yaqi Shi

07/20/2024

Linear regression in R

```
library(tidyverse) #ggplot2, dplyr, etc.
library(reshape2) #need this for melt()
library(knitr) #need this for kable
library(MASS) #contains dataset
```

Load the birthwt data. This data contains 189 observations, 9 predictors, and an outcome, birthweight, available both as a continuous measure and a binary indicator for low birth weight.

```
data(birthwt)
head(birthwt)
```

```
##    low age lwt race smoke ptl ht ui ftv  bwt
## 85   0  19 182    2     0   0  0  1   0 2523
## 86   0  33 155    3     0   0  0  0   3 2551
## 87   0  20 105    1     1   0  0  0   1 2557
## 88   0  21 108    1     1   0  0  1   2 2594
## 89   0  18 107    1     1   0  0  1   0 2600
## 91   0  21 124    3     0   0  0  0   0 2622
```

1. Plot a scatterplot of birthweight (bwt) and mother's weight (lwt).
2. Use OLS to fit the regression of birthweight on mother's weight.
3. Extract the following: estimated coefficients, standard errors, variance-covariance matrix, and confidence intervals.
4. Plot the regression line and interpret the intercept and slope
5. Does the interpretation of the intercept make sense? How might we change this?
6. Now, we want to fit a model that includes race, mother's age, and smoking status in the model. Race takes on value 1 for white, 2 for black, and 3 for other. Mother's age is continuous. Smoking status is binary. Write out the regression function we may be interested in.
7. Use OLS to calculate the coefficient estimates in this model.
8. Interpret all the coefficient estimates.
9. Print the results in Rmarkdown using kable().