Project 1

Daniel Carpenter

12/13/2020

Introduction

Within the `Project1.Rmd` file and this `Project1.pdf` file, the user can create a linear regression between two variables; also, the user can use a bootstrapping function. Within the function `myslr`, the user estimates the relationship between two variables, as well show the confidence in those estimates.

Within the bootstrapping function, the user can use a limited sample to infer information about a population.

Data

Within the examples in this file, the `mtcars` dataset provides meaningful information when estimating linear regressions and bootstrapping. This dataset includes information from the 1974 *Motor Trends* US magazine, and "comprises fuel consumptions," as well as "10 apspects" of each of the thirty-two cars in the dataset. These "aspects" resemble the variables within the data.

When considering the variables in the `mtcars` dataset, it is useful to understand the types of data, which help best determine how to analyze the data. Please see the variables, their descriptions, and their data types below.

Variable Name	**Variable Description**	**Data Type**
mpg	Miles/(US) gallon	Ratio
cyl	Number of cylinders	Ratio
disp	Displacement (cu.in.)	Ratio
hp	Gross horsepower	Ratio
drat	Rear axle ratio	Ratio
wt	Weight (1000 lbs)	Ratio
qsec	1/4 mile time	Interval
vs	Engine (0 = V-shaped, 1 = straight)	Nominal
am	Transmission (0 = automatic, 1 = manual)	Nominal
gear	Number of forward gears	Ratio
carb	Number of carburetors	Ratio

Theory Used

$$Y = \beta_0 + \beta_i * X_i + \epsilon_i \tag{1}$$

`Y` Represents the dependant, or explained, variable.

$$\beta_0 \tag{2}$$

represents the estimated intercept.

$$\beta_i \tag{3}$$

represents the slope of the estimated vector.

$$X_i \tag{4}$$

represents the independent, or control, variable.

$$\epsilon_i \tag{5}$$

represents some error term.

Application of SLR to the mtcars data set

Making the SLR function: `myslr`

```
myslr <- function(data,
                    y, yName,
                    x, xName,
                    sizeVar, sizeVarName,
                    colVar, colVarName,
                    titleVar)
{
  # Open Window to View Plot
    windows(title = "Linear Estimation Graph for Y on X")

  # Create Plot
    plot <- ggplot(
                  # Data
                    data,

                  # Aesthetic Mapping
                    aes(x, y,
                       color  = colVar,
                       size   = sizeVar)) +

                  # Add Scatter Layer
                    geom_point(alpha = 2/5) +

                  # Add Linear Estimation
                    geom_smooth(method  = "lm",
                                formula = y ~ x,
                                color   = "grey35") +

                  # Titles
                    labs(title     = titleVar,
                         subtitle  = " ",
                         x         = xName,
                         y         = yName,
                         col       = colVarName,
                         size      = sizeVarName) +
```

```
                    # Theme
                      theme_get()

    # show Plot
      print(plot)

    # Save plot
      ggsave(filename  = paste0(titleVar, ".png"),
             plot      = plot,
             height    = 6,
             width     = 8)


    # Linear Estimation and Summary Output

      ## Linear Regression (returned)
        y.lm <- lm(y ~ x)

      ## Linear Regression Output (void)
        print(summary(y.lm))

      ## Confidence Interval at 95% (void)
        CI <- ciReg(y.lm)
        CI
        write.csv(CI, file = "Confidence Intervals.csv")

      ## Check assumptions and save .png
        png("Normal Interval Check.png", height = 200, width = 500)
        normcheck(y.lm)
        dev.off()

      ## Check residuals and save .png
        png("Fitted vs. residuals Plot.png", height = 300, width = 500)
        plot(y.lm, which = 1)
        dev.off()

      ## Linear Estimation
        return(y.lm)
  }
```

Invokemyslr function using the mtcars dataset

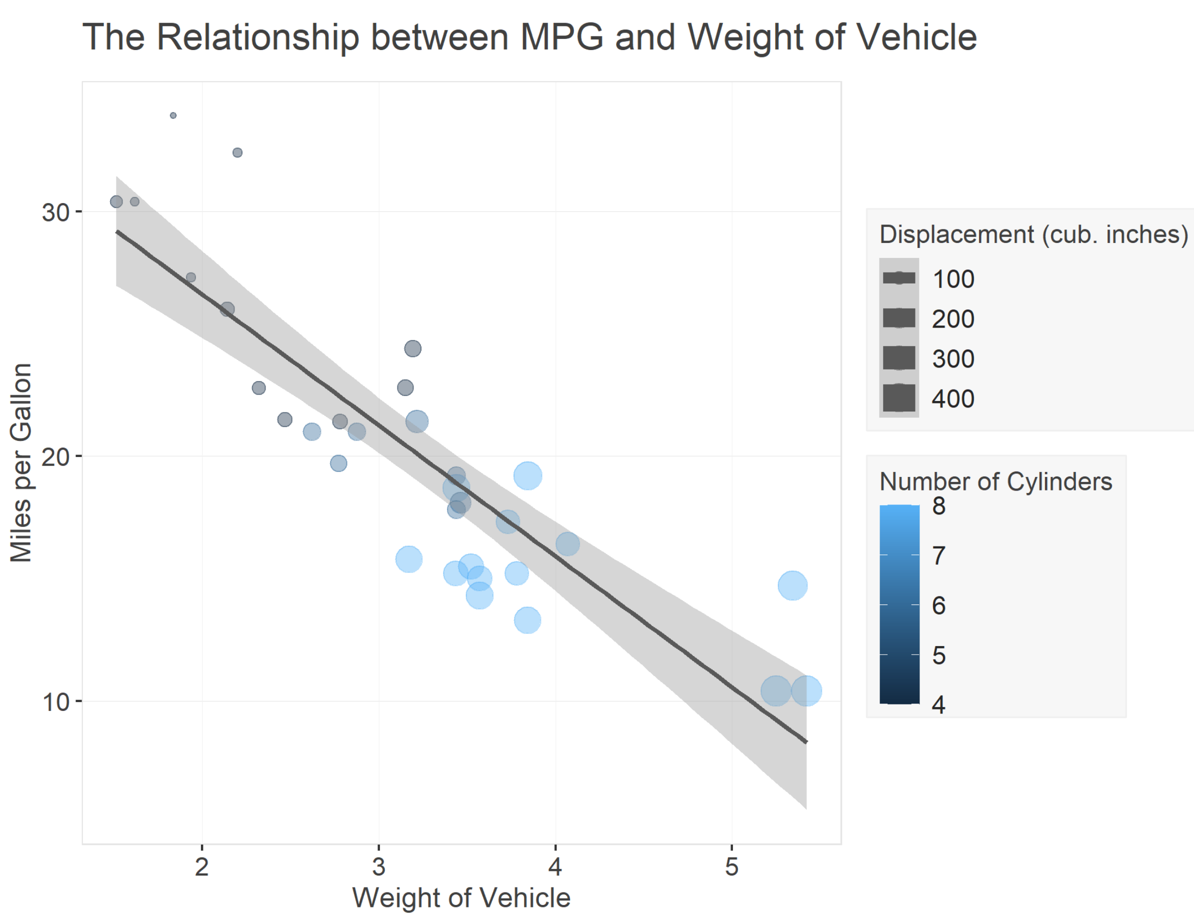
```
  # Call Get Linear Estimation for y on x
    y.lm <- myslr(data = mtcars,
                  mtcars$mpg,   "Miles per Gallon",
                  mtcars$wt,    "Weight of Vehicle",
                  mtcars$disp,  "Displacement (cub. inches)",
                  mtcars$cyl,   "Number of Cylinders",
                  "The Relationship between MPG and Weight of Vehicle")
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

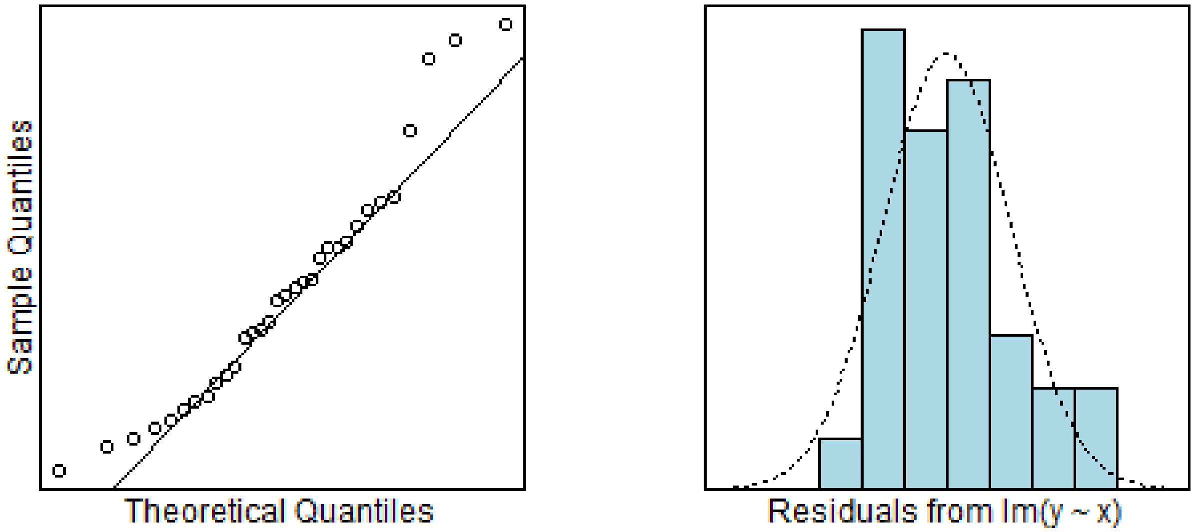
```
## -4.5432 -2.3647 -0.1252  1.4096  6.8727
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  37.2851     1.8776  19.858  < 2e-16 ***
## x            -5.3445     0.5591  -9.559 1.29e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.046 on 30 degrees of freedom
## Multiple R-squared:  0.7528, Adjusted R-squared:  0.7446
## F-statistic: 91.38 on 1 and 30 DF,  p-value: 1.294e-10
##
##             95 % C.I.lower    95 % C.I.upper
## (Intercept)       33.45050          41.11975
## x                 -6.48631          -4.20263
```

```
  # Coefficient list
    coefsList <- y.lm$coefficients
```

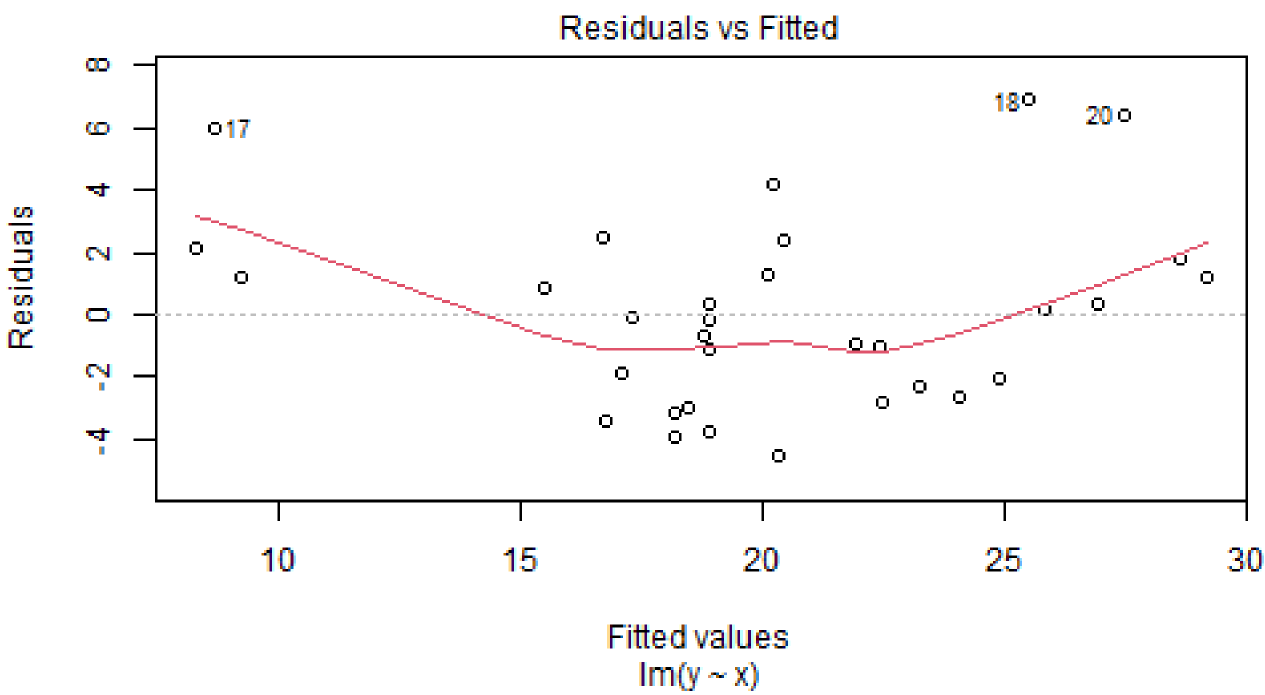
Plot Output



Normal Interval Check



Fitted vs. Residuals



Interpretation of Model

Within the model, it estimates that an increase of 1,000 pounds in weight of a given car decreases the miles per gallon of the vehicle by 5.34. Additionally, the model estimates that changes in weight account for 75.28% of the variation in the car's miles per gallon.

At a 95% level of confidence, the model estimates the lower bound to be a 6.49 decrease in miles per gallon when increasing the weight of a car by 1,000 pounds. Alternatively, the model estimates the upper bound to be a 4.20 decrease in miles per gallon when increasing the weight of a car by 1,000 pounds.

Validitiy of the model

Although the model estimates a narrow confidence interval with a relatively high R-squared value, we may not assume that soley weight accounts for the full 76.28% variation in miles per gallon. This variable could easily be correlated with another relevant variable, thus confounding the simple linear regression.

Bootstrap ==

Make Bootstrap function

Bootstrap Plots

Commandline

File

Invoke function on mtcars dataset