Lab 8: Scatter Plots, Bootstap and Correlation

The following worksheet is due by 8pm one day after this lab. You can find the submission dropbox in Brightspace by clicking on Content − > Lab Content.

0. Open a new R Markdown file.

Note: Your worksheet is to be submitted as the output of an R Markdown file (you can knit it to HTML and then convert it to PDF, or you can knit it to PDF if you have LaTeX on your computer, or you can knit it to Word and then convert that to a PDF).

1. Suppose you have a dataset called "sales" that contains the monthly sales (in thousands of dollars) for a company over the past year. You want to estimate the median monthly sales and its standard error using bootstrapping.

2. Creating a scatterplot, calculating the correlation coefficient, and computing a confidence interval in R using a "mtcars" dataset in R. (Using cat() to concatenate output together like [cat("95% Confidence Interval:", round(lower_ci, 2), "-", round(upper_ci, 2), "\n"])

3. Load the nba_player_data.csv dataset into R and save it to df.

(a) Copy and paste each of the following three lines into your code to eliminate rows with zeros or NA’s.

- df = na.omit(df)
- row sub = apply(df, 1, function(row) all(row !=0))
- df = df[row sub,]

(b) We are only concerned with the categories listed above so use the following code to save only the relevant columns to a new dataframe called dfc:

dfc = df[, c(13:16, 18:20)]

4. Generate six scatter plots. Each will have 3-point percentage (column 1) as the y-axis and the x-axis will be one of each of the other columns in dfc. Recall that the basic function for scatter plots is plot(x, y).

(a) (b) We will be using a for-loop to populate the plots. Each plot will have the same title:

“3pt Shooting Correlation.”, and same y-axis label, “3pt %”. Each plot should be a different color and each x-axis should be properly labeled with the appropriate category.

You may use the following code to set up vectors for different colours and labels:

- colours = c("dodgerblue", "firebrick1", "green3", "orange", "salmon", "slateblue1")
- cnames = c("2pt makes", "2pt attempts", "2pt %", "FT makes", "FT attempts", "FT %")

(c) Write a for-loop that produces the six necessary scatter plots. Above the for-loop code,

use the following to set up a grid so that your plots are presented nicely: par(mfrow = c(2, 3))

5. Use a for-loop to calculate the correlations between 3pt % and each of the other columns.

(a) Create an empty numeric vector to store the correlations and name it cor_vec.

(b) Write a for-loop to populate cor_vec with the corresponding correlations.

(c) Run the following code to properly name the elements in cor_vec: names(cor_vec) = c("3pt %", cnames)

(d) Print out cor_vec.