Lab 10: Multiple Linear Regression

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).

background Download the data sets `randle_stats.csv` and `randle_advanced.csv` and the RMarkdown shell `tutorial_10_shell.rmd`. Save all three files to whichever directory you are using for this course.

Today's datasets give statistical information for a current NBA player named Julius Randle who plays for the New York Knicks. `randle_stats.csv` contains game by game basic stats for Randle and `randle_advanced.csv` contains "advanced" stats for him.

In the `functions` and `clean` chunks I've included code that cleans the datasets as well as calculates the fantasy points scored for each game (based on FanDuel's scoring structure).

Since fantasy points (`fpts`) is calculated directly from many of the columns in `randle_stats.csv` it doesn't make any sense to perform a regression to try to predict his fantasy points. A player's advanced stats, however, can be very helpful in predicting a player's fantasy points because the advanced ratings over time can be quite stable.

Our goal today, therefore, is to use Randle's advanced stats to create a regression formula we could use to predict Randle's fantasy points scored based on his advanced statistics

Note, you will need to have the `dplyr` library installed in R. You have had plenty of time to ensure that dplyr is already installed, but in case you didn't, run the code `install.packages("dplyr")` before starting this lab.

set-up (a) Open `tutorial_10_shell.rmd`. This is the file you will complete your tutorial using. You will need to change the `setwd()` to one that works for your folder system You will also need to change "author" and "output" to your name and the style of output you normally use for these labs, respectively.

    (b) Run all of the chunks provided. When a bunch of warning messages appear in the console, simply click the "x" in the top right corner and ignore them.

    (c) In your console (not a chunk), run the code `head(dfa)` to take a look at the dataset we'll use. You may want to increase the width of your console before running this so that you can see all of the data at once.

1. (a) We are trying to predict `fpts` using the rest of the columns as the explanatory variables. Perform the regression call to do this and save the results to `fit1`. Print out the summary for `fit1`

(b) List the column names that are significant based on `Pr(>|t|)` just as you did in lecture 15 and in your homework. (Note: this does not need to be done in a chunk)

(c) What is the Adjusted R-squared value for this regression.

2. (a) Run a new regression but this time only include the variables that you found in 1a). Save the regression to `fit2`. Print out the summary for `fit2`

(b) What is the Adjusted R-squared value for this regression?

(c) Compare the Adjusted R-squared value to the one from question 1. Do you think we should include more variables in the `fit2` regression based on this comparison? Explain your answer.

Congratulations! You are done. Enjoy the rest of your day and the rest of your studies at UVic.

It has been a pleasure working with you this semester.