STA302 - Assignment #1

Due Date: May 25, 2017, 23:00

If you work with other students on this assignment then:

- your solutions must be written up independently (i.e., your solutions should not be the same as another students solutions).
- Solutions with high similarity will be called in or turned in and resolved at the department level.

Assignment # 1 (Write your solution using provided template A1-soln-temp.Rmd)

- Q1 (4 pts) - Learn how to type mathematical notations. In Q1-a, a detailed proof is given to show you how to type this proof with left alignment in R markdown. Learn from (a), then type your solution of (b) and (c) in the same way. Referece for the latex code to produce mathematical symbols

  http://web.ift.uib.no/Teori/KURS/WRK/TeX/symALL.html

  - Q1-a: Show that $\sum_i^n (X_i - \bar{X}) = 0$

$$\begin{aligned}\sum_{i=1}^{n}(X_i - \bar{X}) &= \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \bar{X} \\ &= \sum_{i=1}^{n} X_i - n\bar{X} \\ &= \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} X_i \\ &= 0\end{aligned}$$

  - Q1-b (2 pts): Show that $\sum_i^n (X_i - \bar{X})^2 = \sum_{i=1}^n X_i^2 - n\bar{X}^2$
  - Q1-c (2 pts): Show that $\sum_i^n (X_i - \bar{X})(Y_i - \bar{Y}) = \sum_{i=1}^n X_i Y_i - n\bar{X}\bar{Y}$

- Q2 (8 pts) - Answer the following questions

  - Q2-a (2 pts): When asked to state the simple linear regression model, a student wrote it as follows

$$E(Y_i) = \beta_0 + \beta_1 X_i + \epsilon_i.$$

    Do you agree? And give your reasoning.

  - Q2-b: The **oldfaithful.txt** data set contains data on 21 consecutive eruptions of Old Faithful geyser in Yellowstone National Park. It is believed that one can predict the time until the next eruption (next), given the length of time of the last eruption (duration). That is, Y is the "eruption" and X is the "waiting" in the data set.

    * (2 pts) Fit a simple linear regression (show R code)
    * (2 pts) Show the summary output of the simple linear regression.
    * (2 pts) What is the estimated linear regression model? (replace the following b_0 and b_1 with their estimates)

$$\widehat{eruption} = b_0 + b_1 waiting$$