

Midterm BANA7038

Due at 11:30PM Dec 22nd, 2015

Instructions on R markdown:

1. Submit only the **html** file and **Rmd** file to canopy, NOT the **pdf** or **doc** file.
2. Number the sections (or questions you are answering) in **LARGE** font, so that each section/answer can be easily identified. Do not repeat the questions to avoid high percentage matching.
3. Put explanations of your results in the text section (i.e. outside of the ```{r} ...lm(y~x)... ```). And avoid using too many `print()`'s in the R code.
4. You can comment your code using `#`, but only comment on the important steps.
5. Try to use `par(mfrow=c(x,y))` when plotting multiple figures to save space. Here `x` and `y` are integers specifying how many rows and columns.
6. Make your report as concise as possible.
7. Avoid printing the entire data set or confidence intervals or too many number at once, instead, visualize them in figures.
8. You don't need to show every steps that you did, only the important steps. For the rest, use `"echo=FALSE"`.
9. Write down your last name, first name (no abbreviations), and M number.
10. **At the beginning of your report, summarize all of your findings in one page.**
11. **The readability of your report plays an important role in grading.**
12. **In canopy, I have uploaded a sample html homework using R markdown. Please refer to its style and format.**
13. Use ```{r echo=FALSE, fig.width=4, fig.height=4} ...plot(x,y)... ``` to adjust the size of the figure.

Generate an html report using R markdown + knitr to answer the following questions.

Background: Flight landing.

Motivation: To reduce the risk of landing overrun.

Goal: To study what factors and how they would impact the landing distance of a commercial flight.

Data: Landing data (landing distance and other parameters) from 800 commercial flights. See the CSV file 'Landing.csv'.

Notes about the variables:

Aircraft: The make of an aircraft (Boeing or Airbus).

Duration (in minutes): Flight duration between taking off and landing. The duration of a normal flight should always be greater than 40min.

No_pasg: The number of passengers in a flight.

Speed_ground (in miles per hour): The ground speed of an aircraft when passing over the threshold of the runway. If its value is less than 30MPH or greater than 140MPH, then the landing would be considered as abnormal.

Speed_air (in miles per hour): The air speed of an aircraft when passing over the threshold of the runway. If its value is less than 30MPH or greater than 140MPH, then the landing would be considered as abnormal.

Height (in meters): The height of an aircraft when it is passing over the threshold of the runway. The landing aircraft is required to be at least 6 meters high at the threshold of the runway.

Pitch (in degrees): Pitch angle of an aircraft when it is passing over the threshold of the runway.

Distance (in feet): The landing distance of an aircraft. More specifically, it refers to the distance between the threshold of the runway and the point where the aircraft can be fully stopped. The length of the airport runway is typically less than 6000 feet.

Suppose that you are asked by the FAA to

- 1. Do statistical analysis of the given data sets and**
- 2. Write a report to summarize your findings.**

Requirement of your report:

- Make sure the readers can understand what you write.

- I would suggest that you write the report by piecing together “chapters”. Each chapter is devoted to a specific goal or function. Each chapter should lay out the following things:
 - The specific goal (why are you doing this?)
 - The R code (how do you realize it?)
 - The R output
 - Your observations (What do you observe from the output?)
 - Your conclusion/decision made based on your observations
- Once you finish those chapters and have a complete “story”, write a summary (one paragraph) on the cover page of your report.

Hint: General procedures of data analysis

1. Data exploration and data cleaning
2. Data visualization
3. Modeling with all covariates
4. Model checking
5. Re-exploration of data
6. Re-modeling by removing insignificant variables one at a time.
7. Modeling checking
8.
9. Finalizing the model

IMPORTANT: The originality and creativity of your report plays an important role in grading (10%).