Recitation 3

Big Data Science

Friday April 22nd, 2022

ASSIGNMENT 3

MPI Estimation Using Nightlight Data

Objectives of the Assignment

- Setup and use a Geographic Information System (GIS) software to analyze geospatial data
- Export the results from the GIS software into a programming environment for further analysis
- Explore different models for regression like backward-stepwise, ridge regression and elastic nets
- Use results from data analysis to reconstruct maps and make comparisons

Questions 1 - 5

Execute the steps in the assignment to obtain the required excel file.

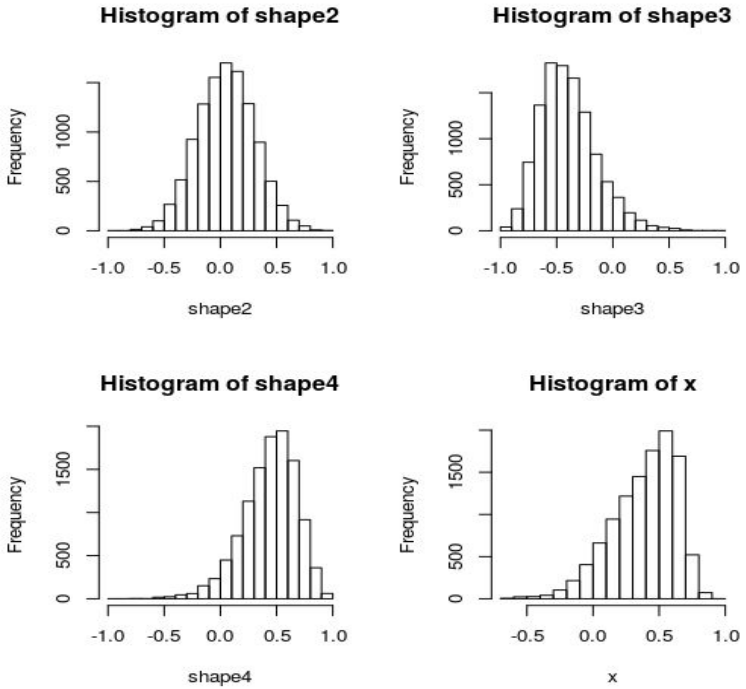
You are advised to watch this ArcGIS Tutorial Video prepared by the TAs.

Note: provide a screenshot of each step in your PDF report.

Question 6

Expected outputs

- Five (5) histograms
- Four (4) scatter plots
- A table showing the requested correlations
- Answers to the qualitative questions



Question 7

Expected outputs

- Two new features
- Three (3) histograms
- A table showing the requested correlations
- Answers to the qualitative questions

Note: The “area” is contained in the “AREA” column of file generated in Question 5.

Question 8

Use the transformations of the variables that produced the highest correlations

Expected outputs

- Three models (backward-stepwise, ridge regression, elastic nets)
- P-value for each feature, for each model
- Comment/justification of significance or insignificance of the features
- Overall p-value for each model and a comment on their significance

Question 9

Use a Lasso or LassoCV model with the same features as in the previous question and perform predictions

Other outputs

- Correlation of log yhat to log y
- R-Squared value
- Answers to qualitative questions

Question 10

Add the estimated MPI in the MPIAssignment.xlsx then in ArcGIS:

- Load the updated MPIAssignment.xlsx dataset. This will be identified as Sheet1$
- Create 2 sector layers using the Sector_Boundary.shp file. One to display the original MPI and the other to display the estimated MPI.
- Use the Add Join (Data Management Tool) to join the Sect_ID field of the Sector_Boundary_2012 layer (i.e the first layer) with the Sect_ID of the MPI excel table (alias Sheet1$)

Question 10 cont'd

To color code the map:

- Right click on the first layer and select **symbology**
- Select **graduated colors** from the dropdown
- Fill the details of the form as follows:
    - Field -> actual mpi
    - Normalization -> None
    - Method -> Quantile
    - Classes -> 10

Repeat the same process for the second layer using the estimated mpi

Notes

- **Avoid Unauthorized Assistance:**
  - Your reports will be compiled to check for plagiarism.
  - Your code files also will be checked for code similarities.
- **Students with top 5 best reports will receive bonus points:**
  - Students with top 5 best reports will be will be selected and asked to make 2 minutes video
  - Students who attend the next recitation will vote on the best videos.
- **Exam logistics**:
  - Date: May 3rd at 8:30am ET / 14:30pm CAT
  - Respondus LockDown Browser Setup (Windows and Mac only) + Test Quiz
  - Exam will be conducted in person (info about rooms will be communicated later)

Submission process:

- Put source code **file and data files** in a single folder
- Name of the folder should be the same as your andrew ID
- **Zip this folder and attach the zipped file on assignment submission page (CANVAS)**
- After attaching zipped file, click on "Add Another File" from assignment submission page and **attach your report**
- Submit your assignment

N.B. This new process will allow us to compile your reports in **Turnitin** to check for plagiarism.

Specific reasons for a submission being classified as incomplete include:

- Failure to correctly name your folder with your Andrew ID, report, and code file with andrewID-BDS-AssignmentNo. For example, mcsharry-BDS-Assignment1, mcsharry-BDS-Assignment2 and mcsharry-BDS-Assignment3.
- A missing report describing the steps, results, and insights
- A missing dataset required for running the code
- A missing code file such as .ipynb or .m file
- An error in the file path needed to run the code

Questions?