**[Assignment 4: Using statistical methods to examine factors that give rise to community engagement](https://mymasonportal.gmu.edu/webapps/assignment/uploadAssignment?content_id=_14270127_1&course_id=_432730_1&group_id=&mode=view)**

**(DUE FRIDAY, 12/3)**

The purpose of assignment 4 is designed to help students understand some basic statistical analysis methods in examining relationships between different community characteristics. If you participated in the [class activities for Week 12](https://gmuedu-my.sharepoint.com/personal/mlee89_gmu_edu/_layouts/15/onedrive.aspx?id=%2Fpersonal%2Fmlee89%5Fgmu%5Fedu%2FDocuments%2FLectures%2FAIT722%5FFall2021%5Fweek12%2Emp4&parent=%2Fpersonal%2Fmlee89%5Fgmu%5Fedu%2FDocuments%2FLectures), *(go to 1:58:00 -- ~2-hr mark)* this assignment should be very easy.

Follow the steps:

1. The data file to be used for Assignment 4 is the data used for the class activity in Week 12: <https://github.com/gmu-cil/AIT722/blob/master/week12/data_28cities.csv>
2. Using R or Python, run statistical analyses to examine the following questions:
   1. How does ethnic heterogeneity affect the poverty level for the given 28 cities, when controlling for the percentage of citizens?
   2. How does poverty affect people's participation in events (i.e., RSVPs) for the given 28 cities, when controlling for the population?
   3. How does poverty affect people's participation in events (i.e., RSVPs) for the given 28 cities, when controlling for the population and Gini index?
   4. How is socio-economic inequality (i.e., Gini index) related to poverty for the given 28 cities?
   5. How is socio-economic inequality (i.e., Gini index) related to the number of events per capita for the given 28 cities?
   6. When 28 cities are categorized into 4 groups based on poverty level (7 cities in each group, based on the order of poverty level), are there systematic differences in their RSVPs and the number of events per capita between different groups of poverty? Examine this question using ANOVA. Also, provide a Box plot for showing the differences between the groups. ANOVA was not covered in the class, but you can use the "anova()" function instead of "summary()" to see the significances in R.
   7. When 28 cities are categorized into 4 groups based on the level of socio-economic inequality, i.e., Gini index (7 cities in each group), are there any systematic differences in their RSVPs and the number of events per capita? Examine this question using ANOVA. Also, provide a Box plot for showing the differences between the groups.
3. For each question, provide (1) the model used, (2) results (e.g., beta cofficients, F-statistics, p-value, graphs, etc.), and (3) the interpretation of the results. Make a report using the results and explanations for all the questions. The number of pages should not exceed 4 pages.
4. Compress the statistical analysis scripts and the report and submit it to BlackBoard.

**Some resources for syntax and useful functions in R and Python**:

* <https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html>
* <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.f_oneway.html>
* <https://www.statmethods.net/stats/regression.html>
* <https://www.datanovia.com/en/lessons/anova-in-r/>

**Submission**: submit (1) your code to run stat analyses (one file, any language in R or Python), and (2) a report (maximum of 4 pages, 1 PDF file). Compress all these files as a ZIP file and submit the ZIP file through BlackBoard.

**Grading Criteria:**

* Whether the 7 questions were examined correctly with appropriate models (5 points \* 7 questions = 35)
* Whether the stat scripts were written in an executable way and correctly? (20)
* Whether the Box plots were generated correctly? (5 points \* 2 questions = 10)
* Whether the report provide reasonable interpretation of the results correctly? (5 points \* 7 questions = 35)
* If the number of report pages is more than 4 pages, the points will be deducted (-5).
* Late panelty: -10% of the full score for every 24 hours after the deadline (so it becomes 0 after 10 days delay).

\* This is an individual assignment, so you have to do the assignment by yourself. You can discuss general ideas with colleagues but cannot share scripts or assignment details with others. Plagiarism will be strictly enforced (the points will be 0 if there is any plagiarism).

If you have any questions, send an email to Myeong at mlee89@gmu.edu.