

**AGB327: Data Analysis Final Project**  
**Due: Monday June 11 at Noon.**

**You may work in groups of up to 4 to complete this project but you are not required to.** You must inform me of your groups and potential determinants by May 5th. A preliminary rough draft is due May 16th in class.

You must upload one well-labelled Excel workbook with all your work to Polylearn. You will also submit a hard copy final report to my office.

**What impacts the number of organic farms in a county?**

The purpose of this project is to encourage you to explore a dataset with the objective of identifying a relationship of interest between variables. You will be graded according to the rubric below. To successfully complete the project, you must perform and report at least the following activities:

1. Table of summary statistics, including coefficient of variation, for continuous variables
2. Calculate a correlation matrix among continuous variables
3. Search for outliers for the key variables of interest
4. Conduct at least one two-population test
5. Conduct a 1-way ANOVA, including Tukey-Kramer analysis
6. Provide at least two plots or graphs of the data that illustrate a meaningful relationship or difference among variables

You must then draw on the work you have done for parts 1-6 to develop and justify two (2) multiple linear regression models to investigate your key relationship of interest. The models must differ by at least one variable, and you must explain why this is not a simple decision to make and why either model may be of interest. You must:

- (i) Create and utilize at least one new variable, based on existing variables, and use this in at least one of your models.
- (ii) Carefully compare and contrast the two sets of results and identify your preferred specification.
- (iii) Discuss the nature and magnitude of your key relationship of interest.

As a group you will submit one final report that explains your analyses, demonstrates each set of results either graphically or with tables, and uses both statistical and logical explanations. You will be graded on the statistics chosen, the presentation, an explanation of how each step builds on the previous one, the validity of your conclusions, and the writing. You need to use your knowledge and outside literature to decide what you think could explain the variation in your dependent variable.

## RUBRIC

### **Good Introduction**

Introduces Data  
Introduces Problem

### **Summary Statistics**

Has correct numbers  
Has good table  
Variables well chosen  
Clearly explained

### **Correlation Matrix**

Performed correctly  
Clear table  
Correlations well explained  
Relates to other analyses

### **Outliers**

Correctly calculated  
Presented well  
Discussion on implications  
Discussion on how to handle them

### **Two-Population Test**

Performed correctly  
Clear Table  
Explains choice of populations  
Understands hypothesis  
Presents conclusion clearly  
Discussion of implications

### **One-Way Anova**

Performed correctly  
Clear table  
Explains choice of factor  
Explains choice of variable  
Understands hypothesis  
Presents conclusions clearly  
Discussion of implications

### **Tukey Kramer**

Performed correctly  
Clear table  
Understands hypothesis  
Presents conclusions clearly  
Discussion of implications

### **Regressions**

Performed correctly  
Clear table  
Includes new variables  
Variables well justified  
Significance well discussed  
Coefficients well discussed  
Clearly compared  
Preferred model well chosen  
Preferred model justified  
Discussion of implications

### **Conclusion**

Bring together all your work  
Answers question  
Solid discussion of results

### **Overall**

Well-written  
Proofread  
Logical connection between sections  
Tables all well designed  
Two good graphs  
Tells an interesting story