

# Algebra III

## Regression Project:

### Putting the “Fun” in Functions

Regression analysis allows us to quantify relationships between multiple variables and make predictions about one variable (our dependent variable) based on the value of other variables (our independent variables). While regressions are generally studied in statistics courses, we can exploit them in our study of functions.

## 1 Define Your Function

$$y = f(x_1, x_2, x_3, \dots)$$

In the first part of this course, we will define various functional forms. Without context, it can be difficult to understand how this knowledge is useful. Here we will estimate a regression (ignoring statistical implications...shhh!) to quantify and describe the relationship between our dependent and independent variables (yes, you will need more than one).

You will choose your own research topic; this will give you an opportunity to dive into a subject matter than interests you. I will be here to assist you throughout the process. Research as a process can be daunting or frustrating at times, but this process will (hopefully) deepen your understanding of and appreciation for functions.

Describe the background for your topic. Why is it worth studying? Why does it interest you? What do you hope to learn from this study? Your research proposal must be approved before proceeding to step 2.

## 2 Find Your Data

This part of the project will probably be the most time consuming as data is hard to find. In fact, some data simply do not exist. Researchers get around this inevitable problem by

making up (defining) their own variables to “proxy” what they are actually looking for. I will help you with this section if you find it difficult to find data. I do, however, expect you to put in a solid effort to find the appropriate data on your own.

### 3 Describe Your Model

Based on your understanding of your subject, what do you expect your model to show us? If you are not familiar with your topic, do some quick researching. What do you think the relationship should be between your dependent and each of your independent variables? Why do you think that?

### 4 Estimate Your Function

Estimate your function with your computer. This can easily be done in Excel. We will work through an example together. Don't freak out!

### 5 Interpret Your Results

Present your regression results and explain how to interpret these results. Are the signs as you predicted? If not, explain what may have gone wrong. Again, I am here to assist you with this.

**NOTE:** Wrong signs do not mean you have arrived at the “wrong” answer. A clear and well-thought out reasoning can still earn you an A for this project.

## Determination of Grades

Each section of this project should be turned in by the due date assigned. It is important for you to keep up with these deadlines. I will discuss each component of this project as the term progresses and will assign a due date at least one week in advance, giving you ample time to complete the assignment, or to seek assistance. **Each component will be graded separately.**

**Late Work:** Unless otherwise agreed upon, your overall grade (for that section) will be decreased by 20% for every day it is late.

### **Section 1: Define Your Function (10 points possible)**

Two points are awarded for each of the following:

- You are able to show an understanding for the concept of a function.
- You have reasonably defined a function.
- You incorporated at least 4 independent variables.
- You have clearly described your topic and its background.
- You have explained the purpose of your project and what you hope to gain from this.

### **Section 2: Find Your Data (10 points possible)**

Five points are awarded for each of the following:

- You found reasonable data to fit your model in section 1.
- You have properly cited your data sources.

### **Section 3: Describe Your Model (10 points possible)**

- You are able to show an understanding for your topic. (2 pts)
- You identify the (expected) relationship between your variables. (4 pts)
- You clearly explain why you expect the relationships to be so. (4 pts)

### **Section 4: Estimate Your Function (10 points possible)**

- You are able to show you can properly run your regression in Excel by turning in the results. You should also include the command(s) used to get those results.

### **Section 5: Interpret Your Results (20 points possible)**

Five points are awarded for each of the following:

- You present your regression results in a clear and easy-to-understand way.
- You clearly explain how your results compare to your “expected” results from Section 3. If the results differ from your predictions in Section 3, you clearly and thoughtfully explore reasons for the discrepancies.
- Your work shows full grasp and use of the central mathematical ideas used throughout this project.
- You clearly explain the implications of your findings.

## **Section 6: Presentation (10 points possible)**

Two points are awarded for each of the following:

- You are prepared for your presentation.
- You clearly present and explain your topic.
- You clearly present your results.
- You explain your findings and properly interpret your results.
- You are able to field (reasonable) questions about your topic.

## **Searching for Data**

I am mainly interested in your research as a process. As long as your data comes from a reasonable source, you may incorporate it into your project.

### **Google Scholar**

This is a great way to find reliable, academic papers on a particular topic. These papers, however, are not always free. Once you have the paper title, you can search for the paper copy in other sources, such as Google. Academic papers almost always provide a data section and describe what data they used, and where they collected their data.

### **Government Websites**

Depending on your topic, you may be able to find data from government websites. The Census often has general statistics that may be of use in your study.

## Examples

Below are several examples. I encourage you to think of a topic that most interests you; you are by no means limited to the following list. Again, feel free to ask for assistance, but do not put this off. The data work in particular will take time.

### Example 1: Baseball

There has been quite a bit of research done by economists (believe it or not), studying the relationship between average salaries and contract length. One such function may be defined as follows,

$$CONTRACT = f(SALARY, HEALTH, AGE, \dots)$$

### Example 2: Movies

What is the best way to increase box office returns? Does it matter when (during a given year) a film is released?

$$returns = f(BUDGET, RELEASE, RATEDR, \dots)$$

### Example 3: Crime

Are there certain attributes about a community (or a state) that make it subject to higher crime rates than other communities? Does it matter if there are schools, police stations or young families in the area?

$$crime = d(POP, DISTANCE_{station}, SCHOOL, AGE_{district}, \dots)$$

### Example 4: Salary

Can we measure the effect variables such as education have on a person's salary? What is the added value of getting a college degree?

$$SALARY = h(age, edu, experience, gpa, \dots)$$

### **Example 5: Transportation**

We can estimate how many miles a household drives as follows,

$$MILE = f(P_{gas}, DIST_{work}, HH_{size}, \dots)$$

How much would the price of gas have to increase for a household to reduce their total miles driven significantly?

### **Example 6: Smoking Rates**

Adult smoking rates ranged from 11.7% in Utah to 28.1% in Kentucky in 2007, according to the Centers for Disease Control and Prevention. Why are adults more than twice as likely to be smokers in Kentucky and other parts of the U.S. than in Utah?

$$smokerate = s(age, income, education, \dots)$$