

**Assignment 5: Part 1. Predictions and outcomes**

### Main Learning Goal

### You will learn how computers make predictions and how statistical math is used to help make predictions using computers

### **How can we use statistics to help us make better predictions?**

Linear Regression

What is linear regression?  Linear regression is one of the most commonly used mathematical algorithms to examine the relationship between two variables and is often used to make predictions about things.

You will learn how to do regression first in EXCEL and then in your programming class, you will learn more about how to do this in MATLAB.

[Linear Regression in Excel: How to Interpret a Model & Make Predictions - YouTubeLinks to an external site.](https://www.youtube.com/watch?v=wkkllZqICaQ)[](https://www.youtube.com/watch?v=wkkllZqICaQ)

We will be using this data set with this model program

[Student performance scores.xlsx](https://ufl.instructure.com/courses/495296/files/88072027?wrap=1)[Download Student performance scores.xlsx](https://ufl.instructure.com/courses/495296/files/88072027/download?download_frd=1)

TO DO:

1. Open the attached EXCEL file.

2. Rewatch the video and find the line of best fit for the regression. This is done by adding the trendline.

3.  Get the equation for the line. What is the slope? What does the slope mean in terms of the student gain for the employability score before and after the training program?

4. Follow the directions to make the regression model and make the prediction based on the before score

5. Why might a student want to know how much the training program would improve their score before they signed up for the program?

6. Will the program always accurately predict a student's performance after taking the class?

## **Deploy**

Now we will look at some Sports data.

The file you will use is from pro football  -  [2023 NFL Advanced Stats | Pro-Football-Reference.comLinks to an external site.](https://www.pro-football-reference.com/years/2023/advanced.htm#site_menu_link)

we are using the 2023 statistics. They are already loaded into a spreadsheet for you to use.

[2023 NFL data stats.xlsx](https://ufl.instructure.com/courses/495296/files/88072280?wrap=1)[Download 2023 NFL data stats.xlsx](https://ufl.instructure.com/courses/495296/files/88072280/download?download_frd=1)

The data here is for passing and receiving data.

1. What two variables would you think are related and would be good to try and do regression on?

2. Using what you learned in the last spreadhseet, do regression for at least two different comparisons. Decide what two different comparisons you want to make.

3. How well did your regressions turn out? What was your number for the trendline? How well correlated was the data?

4. Why do you think it is important to understand the data in order to be able to use it to make predictions?

## **Refine**

Think about examples for music artists and what things may be correlated from the music data you looked up in the last lesson.

5. For either music or sports data - come up with some variables you think could be related and could be analyzed with linear regression.

For example for music- Kodak Black is often in trouble with the law - I could look up his dates of incarceration and see if the number of downloads of his songs goes up (positive correlation) or down (negative correlation) each time he has to go to jail....

Resources

 