Linear Regression Project

In this project you will perform regression analysis on data to develop a mathematical model that relates two variables. Then you will use this model to make predictions.

Objectives

- Find and use data directly from the internet
- Produce a scatter plot of the data
- Perform a regression analysis to find the equation of the line that best fits the data
- Display the results, plotted data and the regression equation together for visual comparison
- Use the model to make predictions
- Make any conclusions about the data

Example:

Is the expression “if you shoot on goal you will score” true?

The game of hockey is based on passing the puck and shooting at the net. You often hear that if you want to score, you have to shoot on goal. It seems like a logical assumption. You will try and verify if you can truly rely on that statement.

First, you need data to support the argument. You can find data on the internet through a number of sources, such as CBS sports or the NHL’s website. You can choose statistics from any year. For this example, we will use the preseason statistics from 2001.

Team	Goals	Shots
Philadelphia Flyers	32	231
New Jersey Devils	29	215
Detroit Red Wings	26	300
New York Rangers	24	208
Montreal Canadiens	26	229
Boston Bruins	24	221
New York Islanders	22	196
Edmonton Oilers	23	216
Phoenix Coyotes	22	185
Ottawa Senators	23	192
Anaheim Mighty Ducks	20	191
Toronto Maple Leafs	21	243
Vancouver Canucks	22	205
Atlanta Thrashers	23	204
Washington Capitals	19	179
Minnesota Wild	20	128

Calgary Flames	18	175
Nashville Predators	18	209
Dallas Stars	17	162
Los Angeles Kings	18	163
St. Louis Blues	16	115
Pittsburgh Penguins	18	159
Tampa Bay Lightning	15	148
San Jose Sharks	17	195
Florida Panthers	13	166
Chicago Blackhawks	13	196
Carolina Hurricanes	13	174
Buffalo Sabres	14	174
Columbus Blue Jackets	15	217
Colorado Avalanche	12	130

For this example, do the following:

1. Input the data into your calculator or Excel
2. Create a scatter plot of the data points
3. Perform regression analysis to determine a regression equation and the correlation coefficient.
4. Plot the line of the regression equation on your scatter plot.
5. Use the model to make conclusions.

By using regression analysis on the example data, you should be able to make conclusions about several things:

- **Is the expression "if you shoot on goal you will score" true?** If you can create model with a correlation coefficient (r-squared) close to 1 or -1 it is likely that the model is a good "fit" and some correlation exists. If not, then there is little correlation between shots and goals.
- **Are shots and goals directly proportional or inversely proportional?** If the two are directly proportional (the number of shots increases when the number of goals increases), the regression equation will have a positive slope. If the two are inversely proportional (where the number of shots increases when goals decrease and vice versa), the equation will have a negative slope.
- **What is the rate of change?** If the absolute value of the slope is far from zero, then one value must increase or decrease much more to get a small change in the second value. This could tell you about how many more shots must be taken to score an extra goal.

Just as you can use the data to reach meaningful conclusions about hockey, other real world data exists that will allow you to apply models that can provide insight into how things relate to one another.

Steps for the Project

1. **Find and select your data.**

   You are responsible for finding data. Check with Mr. Stone to be sure the data you have selected would be a good fit for the project. Some sites to check for possible data include the following:

   Varied Statistics - http://college.cengage.com/mathematics/brase/understandable_statistics/7e/students/datasets/slr/frames/frame.html

   Major League Baseball Stats – mlb.mlb.com/stats

   NFL Stats – www.nfl.com/stats

   NBA Stats – www.nba.com/statistics

2. **Create a scatter plot.**
   You need to verify the presence of a relationship between the two variables. Make a scatter plot with these two variables, and show your independent and dependent variables. Label the axes and the graph accordingly (*y* vs. *x*).
3. **Regression analysis.**
   Input your data in either a calculator or Excel . Calculate the regression equation and the correlation coefficient. Add the regression line to your scatter plot.
4. **Make conclusions.**
   You will write a two- to three-page paper explaining the significance of your results and how you can interpret them (next step). Just as you interpreted the results of the goals vs. shots in the example above, you will need to examine the results of your regression and describe what sort of correlation exists. Is there a strong, weak, or no correlation in the data?
5. **Apply it to the real world.**
   How might your conclusions impact the real world? What sorts of useful applications might you be able to make from your model? Write about ways that you might take advantage of the data. If you feel your data was not particularly useful due to a low correlation coefficient, write about what other patterns you may see in the scatter plot or how it is useful to know that there is little correlation between the variables.
6. **Present.**
   You will present your results and interpretation to the class during the final examination period. You should be prepared to explain your data, conclusions, and interpretations as well as answer questions by the class and instructor. Be prepared to speak for about 10-15 minutes.

This project will be used as your final examination grade for this semester.

Checklist

- Data table
- Scatter plot
- Scatter plot with regression line
- Regression equation and correlation coefficient
- Two- to three-page essay on results and interpretations
- *Extra Credit: PowerPoint Presentation*

Rubric

	1	2	3	4
Representation of Data	Minimal data; data/scatter plots missing	Some missing data/scatter plots	Data present and accurate save for minor errors	Data accurately portrayed in tables and graphs
Application of Regression	Regression not performed	Regression results inaccurate due to faulty calculation	Regression results slightly off due to data entry	Regression results perfectly accurate
Conclusions from Results	Displays no evidence of understanding of conclusions	Misrepresents results by drawing conflicting conclusions	Results largely interpreted correctly with some minor misinterpretations	Shows clear understanding of the results and why they were reached
Real World Interpretation	Makes no connection to the real world	Has some ideas of how data connects to real world situation	Shows connections to real world but could be more thorough	Strongly connects results to real world applications
Presentation	Student seems ill-prepared and largely unable to present	Student displays some confidence, but lacks for time or cannot explain ideas thoroughly	Students displays good understanding of the project and is able to communicate well	Student displays thorough understanding of project and shows considerable time and thought invested

Based on a project by Clairie Vassiliadis, 2003