



Education Transfer Plan ABSTRACT

Directions:

- 1) Please fill out this form electronically by downloading it from <http://iisme.org> under Summer Fellowships, Fellowship Forms.
- 2) Save As and change the file name to **Abstract_yourlastname**
- 3) Print and attach to your ETP.
- 4) Acquire mentor signature, retain **ORIGINAL** and submit **COPY** to Peer Coach by 8/12/05.
- 5) Acquire administrator signature on ORIGINAL and submit to IISME by 10/3/05.

ETP TITLE: Multiple Regression - An AP Statistics Extension Project

FELLOW NAME: Scott N. Friedland

FELLOW PRIMARY E-MAIL: scott.friedland@comcast.net

SPONSOR COMPANY: Lockheed Martin

MENTOR NAME: Chris Lockett

ABSTRACT (50 words or less:) Students will learn the engineering industry standard application MATLAB in order to perform advanced statistical analysis. In particular, they will use MATLAB to perform multiple regression. Students will then present their findings to the class.

ETP TYPE: Conventional

GRADE LEVEL: High School

Subject: Math

Document Format(s): Word doc PowerPoint Excel Other:
If "Other," please describe: _____

Describe how your ETP aligns with NBC or State standard as stated in your proposal:

<http://www.cde.ca.gov/be/st/ss/mthaprobstat.asp>

11.0 Students compute the variance and the standard deviation of a distribution of data.

12.0 Students find the line of best fit to a given distribution of data by using least squares regression.

13.0 Students know what the correlation coefficient of two variables means and are familiar with the coefficient's properties.

14.0 Students organize and describe distributions of data by using a number of different methods, including frequency tables, histograms, standard line graphs and bar graphs, stem-and-leaf displays, scatterplots, and box-and-whisker plots.

Describe the connection between your ETP and the Summer Fellowship.

At Lockheed-Martin, I was responsible for developing a multiple regression model describing the interplay between the various factors experienced by the airplane and the amount of accelerative force the airborne laser was experiencing. I worked on this project in MATLAB

Checklist for sections contained in ETP:

- | | | |
|--|--|--|
| <input type="checkbox"/> Clearly Stated Outcomes & Standards | <input type="checkbox"/> List of Materials/Resources | <input type="checkbox"/> Hard Copy Turned In to Peer Coach |
| <input type="checkbox"/> Specific Procedure/Plan | <input type="checkbox"/> Rubric or plan for evaluating outcomes. | <input type="checkbox"/> Soft Copy Turned into Peer Coach |

I. FELLOW SIGNATURE-- Required Before August 12, 2005

I, the IISME **Fellow** named above, affirm that the ETP I am submitting is my own work, that I acknowledged sources where appropriate, and that I avoided including any proprietary information of the Sponsor Company. By my submission I am assigning to IISME my entire copyright in the ETP. I understand IISME is simultaneously granting me a license to use the ETP for pedagogical purposes.

Signature _____ Date _____

II. MENTOR SIGNATURE-- Required Before August 12, 2005

I, the **Mentor** named below [please select one of the following],

- have read the attached ETP, and my comments, if any, appear below.
 have read the attached ETP, and, as outlined in the IISME-Company Fellowship Agreement, have reviewed it on behalf of the Sponsor Company, and have determined that the ETP does not contain any Sponsor-proprietary information. My additional comments, if any, appear below.

Mentor Comments:

Signature _____ Date _____

Printed Name:

III. ADMINISTRATOR SIGNATURE-- Required Before October 3, 2005, submit to IISME on or before October 3 to be eligible for \$300 grant.

I, the **Administrator** named below have read the attached ETP and my comments, if any, appear below.

Administrator comments:

Signature _____ Date _____

Printed Name:



Education Transfer Plan PROPOSAL

Directions:

- 1) Please fill out this form electronically by downloading it from <http://iisme.org> under Summer Fellowships, Fellowship Forms.
- 2) Save As and change the file name to **Proposal_yourlastname**
- 3) Due to your Peer Coach Thursday, July 7, 2005

PROPOSED ETP TITLE: Multiple Regression - an Extension Project for AP Statistics

FELLOW NAME: Scott N. Friedland

FELLOW PRIMARY E-MAIL: scott.friedland@comcast.net

SPONSOR COMPANY: Lockheed-Martin

MENTOR NAME: Chris Lockett

ABSTRACT (50 words or less summarizing of the focus of your project:) Students will learn the engineering industry standard application MATLAB in order to perform advanced statistical analysis. In particular, they will use MATLAB to perform multiple regression. Students will then present their findings to the class.

PROPOSED ETP TYPE: Conventional **GRADE LEVEL:** High School

Subject: Math **Document Format(s):** Word doc PowerPoint Excel Other

If "Other," please describe:

Describe the need your ETP idea will fulfill in your classroom, teaching, or school. What are the main objectives of your plan?

Upon learning the various descriptive statistics and inferential procedures used in statistics, students have a natural curiosity regarding how these ideas are extended into more complicated realms. Multiple regression is a useful tool they otherwise would not get exposure to unless they took advanced statistics courses in college. Students also wonder what type of software applications they will be expected to know. MATLAB is an industry standard for modeling and signal processing. Exposure to this package now will make them more comfortable in their university engineering courses.

Describe how your ETP idea will align with NBC or State standards:

<http://www.cde.ca.gov/be/st/ss/mthaprobstat.asp>

11.0 Students compute the variance and the standard deviation of a distribution of data.

12.0 Students find the line of best fit to a given distribution of data by using least squares regression.

13.0 Students know what the correlation coefficient of two variables means and are familiar with the coefficient's properties.

14.0 Students organize and describe distributions of data by using a number of different methods, including frequency tables, histograms, standard line graphs and bar graphs, stem-and-leaf displays, scatterplots, and box-and-whisker plots.

Describe connections between your proposed ETP and your Summer Fellowship.

At Lockheed-Martin, I was responsible for developing a multiple regression model describing the interplay between the various factors experienced by the airplane and the amount of accelerative force the airborne laser was experiencing. I worked on this project in MATLAB.

What kinds of materials or resources will you need to develop or gather in order to make your ETP a success? (supplies, personnel, equipment, etc., needed for the project--Make sure this is realistic for your school site.)

I would need access to MATLAB, an LCD projector and a personal computer. My school site has all of this equipment.

Give a basic overview of your plan:

I will begin by teaching students the basics of multiple regression. There is an excellent website that will serve as my primer: <http://www.statsoft.com/textbook/stmulreg.html> as well as Ellen Ludlow's wonderful powerpoint presentation:

<http://www.bbn->

[school.org/us/math/ap_stats/project_abstracts_folder/proj_student_learning_folder/multiple_reg_ludlow.pps](http://www.bbn-school.org/us/math/ap_stats/project_abstracts_folder/proj_student_learning_folder/multiple_reg_ludlow.pps). Students will form groups of size 3-4 and then research topics where multiple regression might be appropriate. If they cannot come up with a topic, they can choose from the list provided here:

<http://www-unix.oit.umass.edu/~statdata/statdata/stat-mult.html>. Students will then work towards

developing an appropriate multiple regression model for their data (preferably using MATLAB) and prepare a presentation (preferably using powerpoint) of their findings including the validity of their model, the usefulness of their model, insights into the data or analysis process, general conclusions, and possible improvements. This project can be completed within two weeks although more time may be given based on the teacher's desire to implement other end-of-year activities in the class.

What kinds of assessments or evaluations do you have in mind? (rubrics, surveys, evaluation forms, etc.)

The evaluation process will consist of two parts. The first part will be a 45 point rubric addressing:

- | | |
|---|---|
| 1) Quality of Presentation: 10 points | 2) Addressed assumptions required: 5 points |
| 3) Thoroughness of analysis: 5 points | 4) Mathematical rigor/correctness: 5 points |
| 5) Quality of Conclusions: 10 points | 6) Insights into Data: 5 points |
| 7) Possible Improvements Considered: 5 points | |

The second part will be a student group evaluation worth 5 points. Each student will give the other members of the group a score ranging from 0-5. The average of the other students' evaluations will get added to the person's project score from part one for a total of 50 points. Scores lower than 4 require commentary explaining why the student thinks that the other group member deserves that score.

Mentor Signature _____

Date

Printed Name:

ETP Revision – Scott N. Friedland – 7/28/05 – P.C. Cory Gaub

ETP: Multiple Regression and MATLAB for Students

Background: I supported the Airborne Laser team at Lockheed Martin. While there, one of my main responsibilities was to analyze flight data in MATLAB in order to develop a regression model for the amount of acceleration the turret mirrors were experiencing as a function of the aerodynamics data. MATLAB is an industry standard computer algebra

system for engineers and is also used extensively at the university level. One of its built-in “toolkits” involves statistics which includes the ability to do multiple regression. In this ETP, students will be developing and analyzing a multiple regression model of their own data.

Note: This project is an extension project for a high school AP Statistics course or an introductory level college statistics course. Multiple regression is not currently an AP test topic, although the concept itself is a natural extension of the regression work the students have already completed. . Since most student calculators do NOT have this feature, the teacher will need to provide some means of doing multiple regression, whether it involves the MATLAB procedures I outline here, the use of websites that include statistical java applets (I provide one such site), or the downloading of one of the many statistical freeware programs available on the internet (recommend *versiontracker.com* or <http://members.aol.com/johnp71/javasta2.html>). This project is most naturally done as a post-AP exam project but can also be given as an extra credit assignment at any point after the students have learned basic regression techniques and analysis.

Recommended Materials:

- LCD projector
- Windows or Mac computer lab with MATLAB pre-installed
- Microsoft Office suite

Intro: Using LCD projector and computer, demonstrate and teach students:

- Concepts behind multiple regression (PowerPoint presentation) – 1 day
- MATLAB use – 1 day
 - Setting a directory
 - Creating a vector of data points
 - Saving data
 - Multiple regression feature
 - Basic command structure
 - Using multiple regression command on sample data sets
 - Find the maximum error from the data to the predicted regression value
- Students research topics of interest where multiple regression might apply – 1 day
- Students form groups based on areas of interest and complete regression project – 7-12 days

Part I: Teaching multiple regression

- Ellen Ludlow has created a nice PowerPoint presentation regarding multiple regression – it makes for an excellent introduction to the topic:
- http://www.bbn-school.org/us/math/ap_stats/project_abstracts_folder/proj_student_learning_folder/multiple_reg_ludlow.pps
 - She explains her results in the context of minitab output, which the students should already be familiar with due to the AP exam’s deciphering computer output requirements.
- There’s also an excellent website that can be used: <http://www.statsoft.com/textbook/stmulreg.html>

- This could be printed for students or presented in web form if in a computer lab
- Give special attention to how a model can go bad: multicollinearity, distribution of residuals, number of recommended data points, etc.

Part II: Teaching MATLAB

- Windows computer lab recommended with MATLAB pre-installed
- If you don't have MATLAB, this site can be used:
<http://www.shodor.org/chemviz/tools/multireg/>
- Demonstrate for students:
 - Setting the directory
 - Recommend each student have their own folder on the lab computer and then cut and paste the directory into the "Current Directory" window.
 - Defining a vector
 - General format is: `varname = [val1 val2 ... lastVal];`
 - The apostrophe creates a column vector
 - The semi-colon tells MATLAB that we don't need it to repeat the output back at us. Omitting the semicolon means MATLAB will reprint the vector on the screen
 - Saving files: **File** → **Save Workspace As**
 - Students will be able to doubleclick this file and their vectors will be loaded into memory
 - Multiple Regression Feature:
 - Suppose we have three independent variables: x_1, x_2, x_3 and one dependent variable y
 - So we'd have defined four vectors in MATLAB: x_1, x_2, x_3, y
 - Our model is essentially: $y = a_0 + a_1x_1 + a_2x_2 + a_3x_3$ where $a_i =$ constants
 - Create a matrix of the independent variables for MATLAB:
 - $X = [\text{ones}(\text{size}(x_1)) \ x_1 \ x_2 \ x_3];$
 - Tell MATLAB to find the coefficients using least squares method: $a = X \backslash y$
 - Notice the use of the backslash (\backslash) and NOT the forward slash ($/$). Be sure to use the correct operator.
 - MATLAB will print out the coefficients of the model. Notice the constant is given first.
 - But how good is the model? We need to validate it somehow. A good first-step is to find the maximum deviation of the data to the model.
 - $Y = X * a; \text{MaxErr} = \max(\text{abs}(Y-y))$
 - This should lead to a good discussion with students of how much error is acceptable.

Part III: Letting the Students Explore

- While either in the lab or as a homework assignment, have students brainstorm possible multiple regression relationships. They can use the internet for ideas but

any data they find must NOT have been already analyzed from a multiple regression standpoint

- The next day, put this list of ideas up on the board and discuss with students which ones seems feasible for research
 - If there's not enough topics, you can use some of the ones listed here: <http://www-unix.oit.umass.edu/~statdata/statdata/stat-rmult.html>
- After getting 6-10 strong ideas, students should form groups of 3-5 based on their area of interest.
- After they've formed their groups, present the rubric for the grading of the project. I've outlined a possible rubric here:
 - **50 points**
 - Quality of Presentation: 10 points
 - Did all students speak? Were they prepared and knowledgeable about what they were speaking about? Did they address the audience or just read off of notes/slides? Any special "hook" used to capture audience's attention?
 - Quality of Conclusions: 10 points
 - Did they understand what their results mean? Did they address the predictive value of their results? Did they give examples of how the model could or could not be used?
 - Assumptions appropriately addressed and discussed: 5 points
 - Did they address the necessary regression assumptions in their analysis?
 - Mathematical Rigor/Correctness: 5 points
 - Did they enter in data correctly and get the correct regression model for their data?
 - Thoroughness of analysis: 5 points
 - Did they present a superficial amount of information or did they consider all the factors that may have influenced their data? Did they consider the individual bi-variate analyses before creating the multivariate analysis?
 - Insights into Data: 5 points
 - Were there problems with data collection? Did anything interesting pop out while looking at the data? Influentials, outliers, skewness, etc.
 - Possible improvements: 5 points
 - Did they consider what might have made for better analysis or better data?
 - Evaluation Score from other group members: 5 points
 - The other members of a person's group while rate them from 0-5. The average of those scores becomes the last five points earned for the project. If a student evaluates another student at a 3 or less, a written justification must be made.
 - Students absent on presentation day automatically lose 5 points

Part IV: Students complete project

- I recommend at least a week for the students to complete the project
- I encourage powerpoint presentations or other attention grabbing presentations.
- At least one weekend should be provided for the students to collaborate.
- You can choose to have the students work in the computer lab, work in class, or complete the assignment mostly outside of class
- You can choose to establish checkpoints to evaluate the level of completion of the project or to field questions from the students