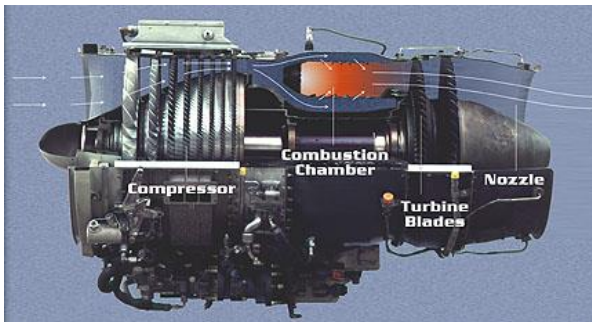# LAB ASSIGNMENT 5

SIMPLE LINEAR REGRESSION

In this assignment, you will use linear regression features available in Excel to examine the relationship between the thrust of a jet turbine engine and three predictor variables: fuel flow rate, exhaust temperature, and ambient temperature. In particular, you will develop simple regression models that adequately approximate the thrust as a function of each of the three predictor variables, and allow for the questions of interest to be investigated. Moreover, you will apply appropriate diagnostic tools in Excel to verify the regression model assumptions.

The Thrust of a Jet Turbine Engine

Jet engines are used for high-speed flight within the atmosphere. They are extremely reliable and provide lots of thrust for their weight. A jet engine goes through the same sort of process as a piston engine, except that combustion occurs continuously, rather than once, during each four-step cycle.

In the lab assignment, you will use simple linear regression to make predictions about thrust of a jet turbine engine using one of the three predictor variables: fuel flow rate, exhaust temperature, and ambient temperature. Which of the variables is the most important predictor of thrust? How reliable are predictions about thrust using a regression model based on one of the three predictors? You will answer the questions with Excel.

The data are available in the Excel file *lab5.xls* located at http://www.stat.ualberta.ca/statslabs/index.htm (click *Stat 235* link, and *Data* for *Lab 5*). The data are not to be printed in your submission.

The following is a description of the four variables in the data file:

Variable Name	Description of Variable
thrust	Thrust of a jet turbine engine,
rate	Fuel flow rate,
extemp	Exhaust temperature,
ambtemp	Ambient temperature at time of test.

1.   First examine the relationship between the response variable (thrust) and each of of the three predictor variables with scatterplots.

   (a)   Obtain a scatterplot of thrust versus each of the three predictors. The format of each of the three scatterplots should be consistent with the format used in *Lab 1 Instructions* (no lines or grids, axes rescaled to display only the observed values, title, names of the axes). Paste the three scatterplots into your report.

(b)   Comment briefly on the relationship between thrust and each of the three predictors. In particular, comment on the overall shape (line, curve), direction (positive, negative), and strength (size of the scatter and its inclination) of the relationship. Which of the three predictor variables seems to be the best predictor of thrust?

2.   Examining the array of all possible pairwise correlation coefficients is another way to understand the relationships among variables. Use the *Correlation* tool in the *Data Analysis* menu to obtain the correlation matrix for the four variables.

   (a)   Paste the correlation matrix into your report. Make sure that the matrix contains the names of all variables.

   (b)   Identify the regressor variables (predictors) having the largest and smallest absolute values of correlation with the response variable (thrust). Do the signs and magnitudes of the correlation coefficient between thrust and each predictor confirm your conclusions you have reached in Question 1? Explain briefly.

3.   Define a simple linear regression model with thrust as the response variable and fuel flow rate as the predictor variable. State the model assumptions.

4.   Perform a simple linear regression analysis using thrust as the response variable and fuel flow rate as the predictor variable using the *Regression* tool. Then use the computer output to answer the following questions:

   (a)   What is the estimated regression equation? Are there any influential observations in the sense that removing them would change the fitted line? Are there any outliers (large residuals)?

   (b)   What is the value of *s*, the estimate of the model standard deviation (σ)?

   (c)   What percent of the variation in thrust is explained by fuel flow rate? What other possible predictor variables may explain the remaining variation?

   (d)   Is linear regression on fuel flow rate of any value in explaining thrust? In particular, state the null and alternative hypotheses in terms of the population slope of the regression line, obtain the value of the test statistic, specify the distribution of the test statistic under the null hypothesis, and obtain the *p*-value of the test. What do you conclude?

   (e)   Use the output to obtain a 95% confidence interval for the mean change in thrust as fuel flow rate increases by 1 unit.

   (f)   Use the regression model to predict the mean thrust when fuel flow rate is 30,250 (case 1). What is the value of the residual in this case?

   (g)   Obtain and paste the plot of residuals against fuel flow rate. Describe the pattern of the residuals. Do the residuals appear to be randomly scattered about a horizontal line at zero? Does the assumption of normality for the residuals seem appropriate? Paste the plot into your report and provide explanations.

5.   Perform a simple linear regression analysis using thrust as the response variable and exhaust temperature as the predictor variable. Use the computer output to repeat parts (a) – (d) and (g) of Question 4 with exhaust temperature instead of fuel flow rate as the predictor variable. You may copy the appropriate statistics from the output to answer the questions. However, in your answer to part (g), you should include the residual plot.

6. Compare the results of the regressions performed in Questions 4 and 5. In particular, construct a table with R^2 values, the estimates of the standard deviation (s), the values of the t-statistic, and the p-values for each model. Which of the two models do you think is better (more reliable) at predicting thrust? Explain briefly.

LAB 5 ASSIGNMENT: MARKING SCHEMA

Proper cover page (includes your name, with the surname capitalized, course section and lab assignment number) and appearance (i.a. lab report must be **typed**): 10 marks

Question 1 (13)

(a) Scatterplots of thrust versus the 3 predictors: 3 points each (9 points total)
(b) Relationship between thrust and each of the three predictors: 3 points
The best predictor: 1 point

Question 2 (7)

(a) Correlation matrix: 3 points
(b) Explanatory variables with the highest and the lowest correlation with thrust: 2 points
Confirm analysis of Question 1: 2 points

Question 3 (4)

Definition of multiple regression model: 2 points
Assumptions: 2 points

Question 4 (23)

(a) Estimated regression equation: 2 points
Influential observations: 1 point
Outliers: 2 points
(b) Estimate of the model standard deviation: 1 point
(c) Percent of variation: 1 point
Other variables: 1 point
(d) Test of rate explaining thrust: 6 points
(hypotheses: 1, distribution: 1, test statistic: 2, p-value: 1, conclusion: 1)
(e) 95% confidence interval: 2 points
(f) Prediction and residual: 1 point each (2 points total)
(g) Residual plot: 3 points
Analysis of the assumptions: 2 points

Question 5 (18)

(a) Estimated regression equation: 2 points
Influential observations: 1 point
Outliers: 2 points
(b) Estimate of the model standard deviation: 1 point
(c) Percent of variation: 1 point
(d) Test of exhaust temperature explaining thrust: 6 points
(hypotheses: 1, distribution: 1, test statistic: 2, p-value: 1, conclusion: 1)
(g) Residual plot: 3 points
Analysis of the assumptions: 2 points

Question 6 (5)

Brief comparison (table): 4 points
Best predictor: 1 point

TOTAL = 80