College of Engineering and Computer Science
Mechanical Engineering Department

California State University Northridge

Mechanical Engineering 209
Programming for Mechanical Engineers

Spring 2017 Class number: 15236 Instructor: Larry Caretto

Programming Assignment Six (and last) Due 11:59 pm, Tuesday, May 9[1]

Objective

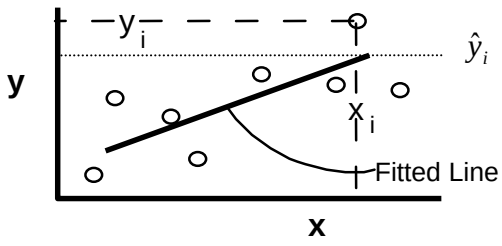
This assignment provides an example in the use of one-dimensional arrays and introduces the concept of regression analysis, which is used to estimate a relationship based on experimental data from two variables.

Background

If several measurements are made on pairs of experimental data $\{(x_i,y_i), i = 1,...,N\}$, we can use a technique, known as regression analysis, to determine an approximate equation of a straight line that gives a best fit to the data. The equation of this best-fit line is written as follows.

$$\hat{y} = a + b\,x$$

In this equation, we use the symbol $\hat{y}$ instead of y to indicate that the predicted value found from the equation, $\hat{y} = a + b\,x$ is an approximate result. For a given data point, (x_i,y_i), the value of y_i represents the actual data and we would obtain the predicted value of y, at the point $x = x_i$ from the equation $\hat{y}_i = a + b\,x_i$. The difference between the measured and predicted value is $|y_i - \hat{y}_i|$.



In the chart at the left, the data points are indicated by the small ellipses. The coordinates of one of a typical data point are shown by the dashed lines indicating the coordinates x_i and y_i. The solid line is the fitted regression line, $\hat{y} = a + b\,x$. The point where the dotted line at $x = x_i$ crosses the regression line has the coordinates $(x_i,\hat{y}_i)$. In this particular example the value of $\hat{y}_i$ is less than the value of y_i. There is a large scatter of data points about the regression line in this example.

The example plot above might represent calibration data on an instrument. The x values would denote the instrument reading and the y values would indicate the true value of the quantity being measured. Once the calibration tests were completed, it would be useful to have a simple equation to relate the instrument reading (x) to the actual quantity being measured (y).

In addition to finding the values of a and b that give the best-fit line, we would also like to have some measure of how well the line fits the data. Two different goodness-of-fit measures, the standard error and the coefficient of variation are presented in the equations section below.

Equations used

The equations used to calculate a and b can be found by an analysis which minimizes the distances between the actual data points, y_i, and the fitted points, $\hat{y}_i = a + b\,x_i$. The results of this analysis are shown below. The equations to compute the intercept, a, and the slope, b, in terms of the entire set of data, $\{x_i,y_i\}$, use the following the definitions of mean values:

[1] Assignments may be submitted by 11:59 pm on Friday, May 12, with a 30% late penalty. No credit will be given for assignments submitted after this time.

Jacaranda Hall 3308 Mail Code Phone: N/A
E-mail: lcaretto@csun.edu 8348 Fax: 818.677.7062

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i \qquad \textit{and} \qquad \bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

With these definitions, the slope, b, and the intercept, a, are found as follows.

$$b = \frac{\sum_{i=1}^{N} x_i y_i \; - N(\bar{x})(\bar{y})}{\sum_{i=1}^{N} x_i^2 - N(\bar{x})^2} \qquad \textit{and} \qquad a = \bar{y} - b\bar{x}$$

A statistical estimate of the variability can be found from the difference between the actual data points y_i and the estimated value $\hat{y}_i = a + b\, x_i$. This measure, which is called the standard error and has the symbol $s_{y|x}$, is defined as follows:

$$s_{y|x} = \sqrt{\frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{N-2}}$$

Another measure, called the R^2 value or the coefficient of variation is considered to be a measure of the amount of variation in the data which is explained by the regression equation. An R^2 value of zero means that the regression cannot explain any of the variation in y; an R^2 value of one means that all the variation in y can be explained by the regression equation. The value of R^2 is computed from the following equation:

$$R^2 = 1 - \frac{(N-2)s_{y|x}^2}{\sum_{i=1}^{N} y_i^2 - N\left(\bar{y}\right)^2} = 1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} y_i^2 - N\left(\bar{y}\right)^2}$$

Specific Tasks

1. Download the workbook with the data for this assignment, pa6Start.xlsm. This workbook contains data on two variables x and y on a worksheet in columns A and B, respectively. Write a VBA function that takes the values of x and y from the worksheet and returns values of a, b, and R^2 at the same time by using an array function. The workbook has the following shell for your code for this assignment in VBA.

```
Function <putYourOwnFunctionNameHere>( <arguments> ) as Variant
    Dim output(1 To 3, 1 To 2) As Variant

    <put code for computations here>

    output(1, 1) = "Slope = "
    output(2, 1) = "Intercept = "
    output(3, 1) = "R-squared = "
    output(1, 2) = <Result for b>
    output(2, 2) = <Result for a>
    output(3, 2) = <Result for R^2>
    <yourFunctionName> = output
End Function
```

This shell shows how to write the code for an array function. A computational array (named "output" in this example) is used and all elements of the array are defined in the code shown above. In the final step, the function name is set equal to the array name. This makes the return value of the function an array. We see that this example code has three rows and two columns.

Once you have written the code, you will have to select a region with three rows and two columns, enter the array formula show below and **press control+shift+enter** to enter the array formula, which has the form shown below.

=<putYourOwnFunctionNameHere>(<rangeForX>,<rangeForY>)

2. Review the help information for the use of the LINEST function in Excel. Use this function to determine the slope, intercept and R^2 value for the same data that you used for the code you wrote. Compare the results of the LINEST function slope, intercept and R^2 to the values you found with your code.

Submission Requirements

Submit your Excel workbook with your VBA code and the comparison of your results for a,b, and R^2, with those of the LINEST function. Your submissions are due on or before 11:59 pm on Tuesday, May 9. Assignments may be submitted by 11:59 pm on Friday, May 12, with a 30% late penalty. No credit will be given for assignments submitted after this time.