WORCESTER POLYTECHNIC INSTITUTE MECHANICAL ENGINEERING DEPARTMENT

Engineering Experimentation ME-3901, D'2012

Lecture 04 21 March 2012

General information

Office hours

Instructors: Cosme Furlong Christopher Scarpino Office: HL-151 Office: HL-153 **9:00 to 9:50 am sessions**

Everyday: During laboratory

Teaching Assistants: **During laboratory sessions**

Curve fitting: introduction

The most common and most efficient ways to present quantitative data are by the use of tables, plots (figures), and mathematical equations. Which method of presentation is the most appropriate form of reporting depends on a variety of factors, which include the purpose of the presentation, the quantity of data, the degree of precision of the results, and the number of variables. Because so many factors are involved, only a few general remarks can be made. Your choice of the particular form of presentation must be made largely by applying common sense guided by experience.

For example, four sets of data are shown in Tables 6.1 to 6.4, respectively. Each one of these data sets has exactly the same mean value of $X's = 9.0$ and of $Y's = 7.5$. Their standard deviations are also exactly the same and equal to $\sigma_x = 3.16$ and $\sigma_y = 1.94$. Furthermore, each of these four data sets is represented by the same equation of a straight line, obtained by the least-squares analysis, which is

$$
Y = 0.50X + 3.00 \quad , \tag{6.1}
$$

and which results in a correlation coefficient of 0.77, for each data set. However, graphical representations of the four data sets are drastically different, as shown in Figs.

The foregoing discussion indicates that mere tabulation of the data and computation of statistical parameters representing these data are not adequate for meaningful presentation of results. Graphical representation of the results is extremely important and aided with statistics and mathematical equations provides complete presentation of data. The subject of this note is graphical, statistical, and mathematical presentation of data and is illustrated with a representative problem.

6.2. Method of least squares

Table 6.1 represents variables X and Y over a range of values. What is needed is a mathematical expression that will represent Y as a function of X. The simplest type of function is a linear one, which establishes Y as a linear function of X expressed as

 $Y = aX + b$,

 $\frac{1}{2}$

⊐

$$
\quad \text{and} \quad
$$

$$
b = \frac{\left(\sum_{i=1}^{n} Y_i\right)\left(\sum_{i=1}^{n} X_i^2\right) - \left(\sum_{i=1}^{n} X_i\right)\left(\sum_{i=1}^{n} X_i Y_i\right)}{n \sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2}
$$
 (6.15)

Equations 6.14 and 6.15 give values of coefficients a and b which best fit the given data set. Following the procedure used above, equations for constants corresponding to higher order curve fits can be determined.

6.3. General form of least squares

The method shown in Section 6.2 is specific to linear, or first order, least squares curve fits. It should be noted that any order polynomial can be fit to data, provided there are more data points than the order of the polynomial (i.e. four points are needed to perform a third order curve fit).

To fit *n* data points to a general mth order polynomial given by

$$
Y = A_1 X^m + A_2 X^{m-1} + \dots + A_m X + A_{m+1} \qquad , \qquad (6.16)
$$

an expanded form of the matrix given by Eq. 6.11 is used. This matrix takes the general form of

$$
\begin{bmatrix}\n\sum_{i=1}^{n} X_i^{2m} & \sum_{i=1}^{n} X_i^{2m-1} & \sum_{i=1}^{n} X_i^{2m-2} & \cdots & \sum_{i=1}^{n} X_i^{m} \\
\sum_{i=1}^{n} X_i^{2m-1} & \sum_{i=1}^{n} X_i^{2m-2} & \cdots & \cdots & \vdots \\
\sum_{i=1}^{n} X_i^{2m-2} & \cdots & \cdots & \sum_{i=1}^{n} X_i^{2} \\
\vdots & \ddots & \ddots & \sum_{i=1}^{n} X_i^{2} & \sum_{i=1}^{n} X_i \\
\vdots & \ddots & \sum_{i=1}^{n} X_i^{2} & \sum_{i=1}^{n} X_i \\
\sum_{i=1}^{n} X_i^{m} & \cdots & \sum_{i=1}^{n} X_i^{2} & \sum_{i=1}^{n} X_i & n\n\end{bmatrix}\n\begin{bmatrix}\nA_1 \\
A_2 \\
\vdots \\
A_m \\
A_m\n\end{bmatrix} = \n\begin{bmatrix}\n\sum_{i=1}^{n} X_i^{m} Y_i \\
\sum_{i=1}^{n} X_i^{m-1} Y_i \\
\vdots \\
\sum_{i=1}^{n} X_i Y_i \\
\vdots \\
\sum_{i=1}^{n} Y_i\n\end{bmatrix},
$$
\n(6.17)

when considering the case of the polynomial given by Eq. 6.16.

To illustrate this general least squares approach, consider the 3rd order polynomial given by

 $Y = aX^3 + bX^2 + cX + d$

 (6.18)

To solve for the four coefficients a, b, c , and d , Eq. 6.17 is applied to Eq. 6.18 and results in

6.4. Example problems

6.4.1. Problem statement

A set of data, representing object's velocity measured as a function of time, is given in Table 6.5.

- 1) represent these graphically on a T-t (i.e., velocity versus time) diagram.
- 2) determine mathematical formula, where V is velocity, t is time, and a and b are constants to be determined, for the data of Table 6.5, using the method of least squares. Show all work.
- 3) compute errors, E, by determining differences between measured and curve fitted points.
- 4) plot results of step 3 on an E-t diagram.
- 5) discuss your results.

V (ft/sec) t (sec) 1.45 0 1 2.60 \overline{c} 3.60 4.50 3 4.95 4 5 5.20 6 5.10 7 4.75 8 4.10

Table 6.5. Data set representing object's velocity measured as a function of time.

ring Departmen

and

$$
a(36) + b(9) = (36.25) \tag{6.22}
$$

Solving Eqs. 6.21 and 6.22 for a and b results in values of $a = 0.3458$ and $b = 2.6442$. Using these values in Eq. 6.2 produces the equation for a linear least squares curve fit of the points in Table 6.5. This resultant equation is

$$
Y = 0.3458 X + 2.6442 \tag{6.23}
$$

Equation 6.23 can be used to compute the an fitted set of points corresponding to those listed in Table 6.5. These fitted points are given in Table 6.6.

> Table 6.6. Data set representing object's velocity measured as a function of time shown with data from first order least squares curve fit.

6.4.2. Problem statement

A set of data representing an object's velocity measured as a function of time, is given in Table 6.5.

- 1) Determine mathematical formula $V = V(t) = a t^2 + b t + c$, where V is velocity, t is time, and a, b, and c are constants to be determined, for the data Table 6.1, using the method of least squares. Show all work.
- 2) Compute errors, E, by determining differences between measured and curve fitted points.
- 3) Plot results of step 3 on an E-t diagram.

The polynomial to which the data is to be fit is in the form of

$$
Y = a X^2 + b X + c \t . \t (6.25)
$$

 (6.27)

The three simultaneous equations to be used to fit the measured data to Eq. 6.25 are

ring Departmen

$$
a \sum_{i=1}^{9} X_i^4 + b \sum_{i=1}^{9} X_i^3 + c \sum_{i=1}^{9} X_i^2 = \sum_{i=1}^{9} X_i^2 Y
$$

$$
a \sum_{i=1}^{9} X_i^3 + b \sum_{i=1}^{9} X_i^2 + c \sum_{i=1}^{9} X_i = \sum_{i=1}^{9} X_i Y
$$
, and (6.28)

$$
a\sum_{i=1}^{9} X_i^2 + b\sum_{i=1}^{9} X_i + c\sum_{i=1}^{9} 1 = \sum_{i=1}^{9} . \tag{6.29}
$$

By putting the values from Eqs. 6.20 and 6.24 into the Eqs. 6.27 to 6.29, the relations

$$
a(8772) + b(1296) + c(204) = 945.45 , \t(6.30)
$$

$$
a(1296) + b(204) + c(36) = 165.75
$$
, and (6.31)

$$
a(204) + b(36) + c(9) = 36.25
$$
 (6.32)

Solving Eqs. 6.30 to 6.32 for a, b, and c, the values $a = -0.13706$, $b = 1.442370$, and $c = 1.365151$ are obtained. Substituting these values into Eq. 6.25, the equation representing the second order curve fit of the points given in Table 6.5 is obtained. This equation is

$$
Y = -0.13706 X^2 + 1.44237 X + 1.36515
$$
 (6.33)

Equation 6.33 can be used to compute the an fitted set of points corresponding to those listed in Table 6.5. These fitted points are given in Table 6.7.

> Table 6.7. Data set representing object's velocity measured as a function of time shown with data from second order least squares curve fit.

References

Anscombe, F. J., "Graphs in statistical analysis," American Statistician, 27:17-21, 1973. Pryputniewicz, R. J., "Engineering Experimentation," Worcester Polytechnic Institute, 1993.

Curve fitting: matrix approach Discussed in class: estimation of \pm ranges for fitting coefficients

Curve fitting can be expressed as:

 $[X'X]\beta = (X'Y)$ $([X']$ is transpose of $[X]$)

Obtained from model: $[X]\beta = (Y)$ (Over determined system of equations)

Vector of unknowns (coefficients of model used, e.g., slope and dc values in a linear fit)

Solution is: $\beta = [X'X]^{-1}(X'Y)$

Components of inverse matrix are expressed as:

$$
\left[\begin{array}{c} \left[X' X \right]^{-1} = \left[c_{ij} \right] \end{array} \right]
$$

Curve fitting: matrix approach Discussed in class: estimation of \pm ranges for fitting coefficients

Recall that sum of the square of the errors (*SCE*) is:

Curve fitting: matrix approach

Discussed in class: estimation of \pm ranges for fitting coefficients

Confidence interval $(1-\alpha)$ 100% for fitting coefficient $\boldsymbol{\beta}_\textnormal{i}$:

Interval for β_i fitting coefficients where their values are $(1 - \alpha)100\%$ reliable

Curve fitting: matrix approach

Discussed in class: estimation of \pm ranges for fitting coefficients

(a) shows tabulated values

TABLE of CRITICAL VALUES for STUDENT'S t DISTRIBUTIONS

degrees of

page 114 of Holman

You can apply this to analyses done in $Lab#2$

cal Engineering Department

