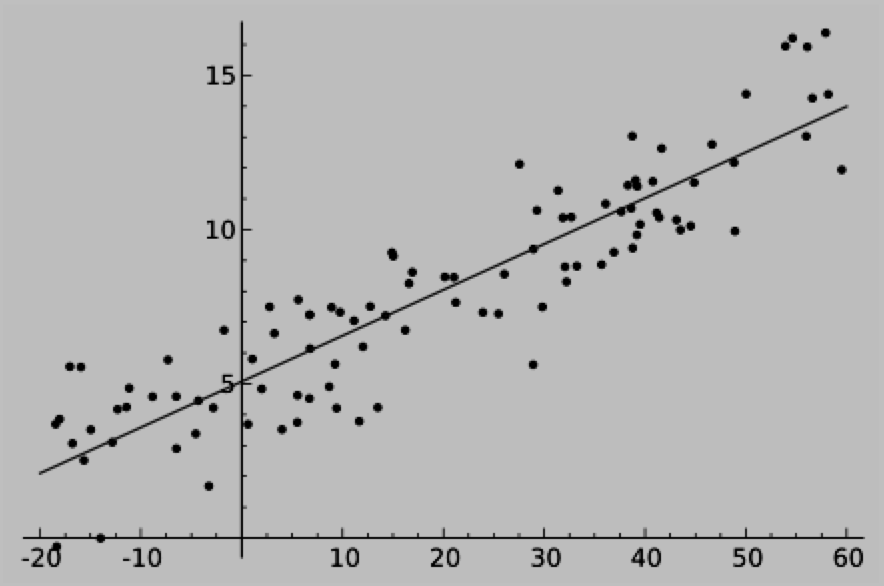
METHOD OF LEAST SQUARES



By: Varun Luthra
11CSU163

INTRODUCTION

In engineering, two types of applications are encountered:

- **Trend analysis**. Predicting values of dependent variable, may include extrapolation beyond data points or interpolation between data points.

- **Hypothesis testing**. Comparing existing mathematical model with measured data.

CURVE FITTING

There are two general approaches for curve fitting:

- **Least Squares regression:**

Data exhibit a significant degree of scatter. The strategy is to derive a single curve that represents the general trend of the data.

- **Interpolation:**

Data is very precise. The strategy is to pass a curve or a series of curves through each of the points.

OVERVIEW

•The method of **least squares** is a standard approach to the approximate solution of **overdetermined systems**, i.e., sets of equations in which there are more equations than unknowns.

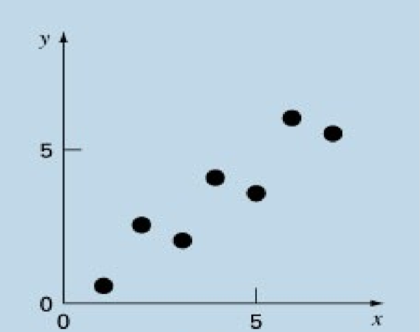
•"<u>Least squares</u>" means that the overall solution minimizes the sum of the squares of the errors made in the results of every single equation.

•The least-squares method is usually credited to **Carl Friedrich Gauss** (1795), but it was first published by **Adrien-Marie Legendre.**

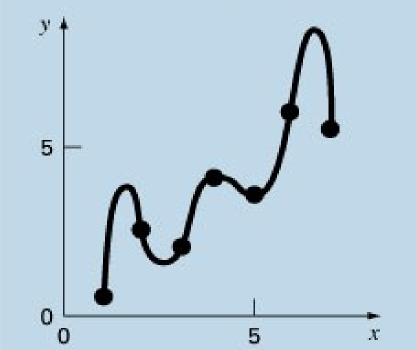
Scatter Diagram:

To find a relationship b/w the set of paired observations x and y(say), we plot their corresponding values on the graph , taking one of the variables along the x-axis and other along the y-axis i.e. $(X_1,y_1),(x_2,y_2)\ldots\ldots,(x_n,y_n)$.

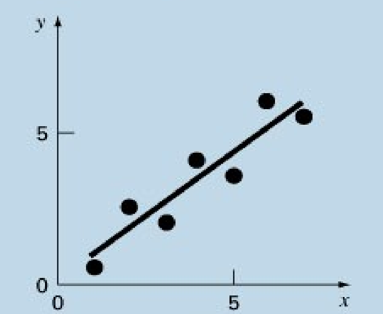
The resulting diagram showing a collection of dots is called a scatter diagram. A smooth curve that approximate the above set of points is known as the approximate curve.



(a)



(b)



(c)

WORKING PROCEDURE

- **To fit the straight line y=a+bx**

- Substitute the observed set of n values in this equation.

- Form the normal equations for each constant i.e. $\sum y=na+b\sum x$, $\sum xy=a\sum x+b\sum x^2$.

- Solve these normal equations as simultaneous equations of a and b.

- Substitute the values of a and b in y=a+bx, which is the required line of best fit.

- **To fit the parabola: $y=a+bx+cx^2$:**

- Form the normal equations $\sum y=na+b\sum x+c\sum x^2$, $\sum xy=a\sum x+b\sum x^2 +c\sum x^3$ and $\sum x^2y=a\sum x^2 + b\sum x^3 + c\sum x^4$.

- Solve these as simultaneous equations for a,b,c.

- Substitute the values of a,b,c in $y=a+bx+cx^2$, which is the required parabola of the best fit.

- **In general, the curve $y=a+bx+cx^2 +\ldots\ldots+kx^{m-1}$ can be fitted to a given data by writing m normal equations.**

- The normal equation for the unknown **a** is obtained by multiplying the equations by the coefficient of **a** and adding.

- The normal equation for **b** is obtained by multiplying the equations by the coefficients of **b** (i.e. x) and adding.

- The normal equation for **c** has been obtained by multiplying the equations by the coefficients of **c** (i.e. x^2) and adding.

Example of a Straight Line

Fit a straight line to the x and y values in the following Table:

x_i	y_i	$x_i y_i$	x_i^2
1	0.5	0.5	1
2	2.5	5	4
3	2	6	9
4	4	16	16
5	3.5	17.5	25
6	6	36	36
7	5.5	38.5	49
28	**24**	**119.5**	**140**

$$\sum x_i = 28 \quad \sum y_i = 24.0$$

$$\sum x_i^2 = 140 \quad \sum x_i y_i = 119.5$$

$$\bar{x} = \frac{28}{7} = 4$$

$$\bar{y} = \frac{24}{7} = 3.428571$$

$$a_1 = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}$$

$$= \frac{7\times 119.5 - 28\times 24}{7\times 140 - 28^2} = 0.8392857$$

$$a_0 = \bar{y} - a_1\bar{x}$$

$$= 3.428571 - 0.8392857\times 4 = 0.07142857$$

Y = 0.07142857 + 0.8392857 x

Example Of Other Curve

Fit the following Equation:

$$y = a_2 x^{b_2}$$

to the data in the following table:

x_i	y_i	$X^*=\log x_i$	$Y^*=\log y_i$
1	0.5	0	-0.301
2	1.7	0.301	0.226
3	3.4	0.477	0.534
4	5.7	0.602	0.753
5	8.4	0.699	0.922
15	19.7	2.079	2.141

$$\log y = \log(a_2 x^{b_2})$$

$$\log y = \log a_2 + b_2 \log x$$

$$\text{let } Y^* = \log y,\ X^* = \log x,$$

$$a_0 = \log a_2,\ a_1 = b_2$$

$$Y^* = a_0 + a_1 X^*$$

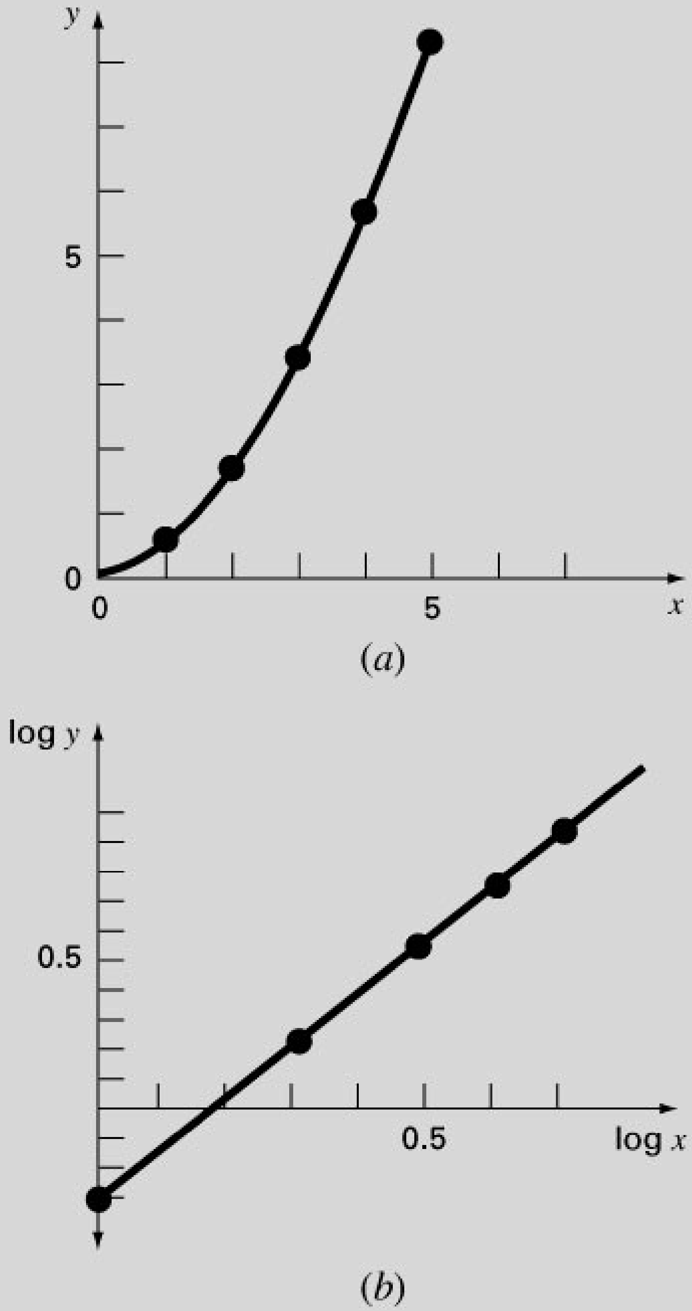
	Xi	Yi	X^*_i=Log(X)	Y^*_i=Log(Y)	X*Y*	X*^2
	1	0.5	0.0000	-0.3010	0.0000	0.0000
	2	1.7	0.3010	0.2304	0.0694	0.0906
	3	3.4	0.4771	0.5315	0.2536	0.2276
	4	5.7	0.6021	0.7559	0.4551	0.3625
	5	8.4	0.6990	0.9243	0.6460	0.4886
Sum	**15**	**19.700**	**2.079**	**2.141**	**1.424**	**1.169**

$$\begin{cases} a_1 = \dfrac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2} = \dfrac{5\times 1.424 - 2.079\times 2.141}{5\times 1.169 - 2.079^2} = 1.75 \\ a_0 = \bar{y} - a_1\bar{x} = 0.4282 - 1.75\times 0.41584 = -0.334 \end{cases}$$

Curve Fitting

$$\boxed{\log y = -0.334 + 1.75 \log x}$$

$$y = 0.46x^{1.75}$$



APPLICATION

- The most important application is in **data fitting**. The best fit in the least-squares sense minimizes the sum of squared **residuals**, a residual being the difference between an observed value and the fitted value provided by a model.

- When the problem has substantial uncertainties in the **independent variable** (the 'x' variable), then simple regression and least squares methods have problems; in such cases, the methodology required for fitting **errors-in-variables models** may be considered instead of that for least squares.

Thank You!

Have a nice day. ☺