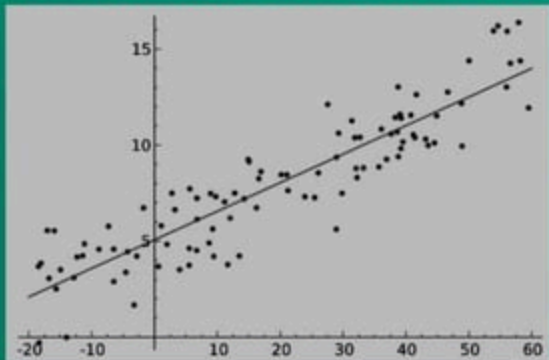


Definition

- **Curve fitting:** is the process of constructing a curve, or mathematical function, that has the best fit to a series of data points, possibly subject to constraints.
- It is a statistical technique use to drive coefficient values for equations that express the value of one(dependent) variable as a function of another (independent variable)

What is curve fitting

Curve fitting is the process of constructing a curve, or mathematical functions, which possess closest proximity to the series of data. By the curve fitting we can mathematically construct the functional relationship between the observed fact and parameter values, etc. It is highly effective in mathematical modelling some natural processes.



Interpolation & Curve fitting

- In many application areas, one is faced with the task of describing data, often measured, with an analytic function. There are two approaches to this problem:-
- 1. In Interpolation, the data is assumed to be correct and what is desired is some way to describe what happens between the data points.
- 2. The other approach is called curve fitting or regression, one looks for some smooth curve that "best fits" the data, but does not necessarily pass through any data points.

Engineering applications of curve fitting technique

- **1.Trend Analysis:-** Predicating values of dependent variable ,may include extrapolation beyond data points or interpolation between data points.

Some important relevant parameters

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.

Mathematical Background (cont'd)

- **Variance**. Representation of spread by the square of the standard deviation.

$$S_y^2 = \frac{\sum (y_i - \bar{y})^2}{n-1}$$

$$S_y^2 = \frac{\sum y_i^2 - (\sum y_i)^2 / n}{n-1}$$

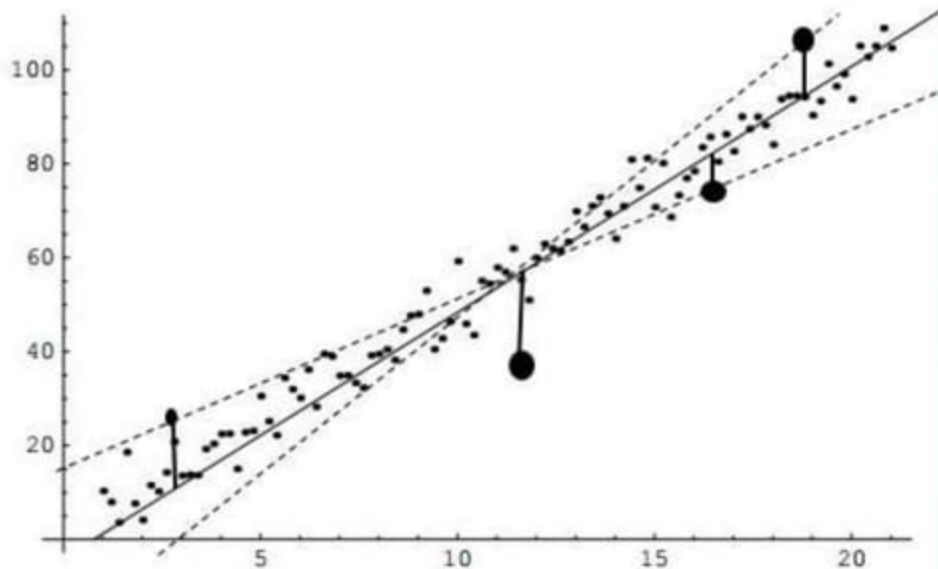
- **Coefficient of variation**. Has the utility to quantify the spread of data.

$$c.v. = \frac{S_y}{\bar{y}} 100\%$$

Least square method

- The Method of Least Squares is a procedure to determine the best fit line to data; the proof uses simple calculus and linear algebra. The basic problem is to find the best fit straight line $y = ax + b$ given that, for $n \in \{1, \dots, N\}$, the pairs (x_n, y_n) are observed. The method easily generalizes to finding the best fit of the form $y = a_1 f_1(x) + \dots + c_k f_k(x)$; it is not necessary for the functions f_k to be linearly in x – all that is needed is that y is to be a linear combination of these functions.

Least square method



Least Squares Regression

Linear Regression

Fitting a straight line to a set of paired observations:

$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$.

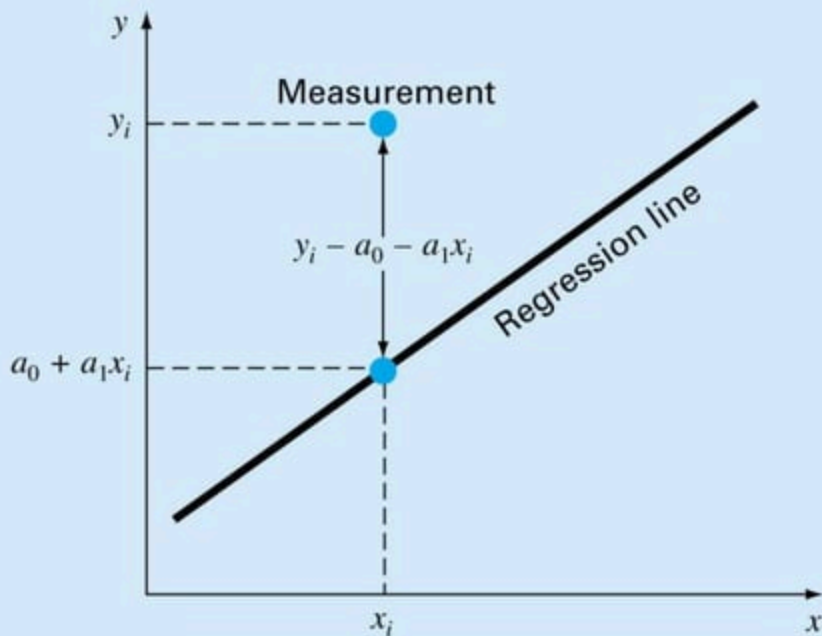
$$y = a_0 + a_1 x + e$$

a_1 - slope

a_0 - intercept

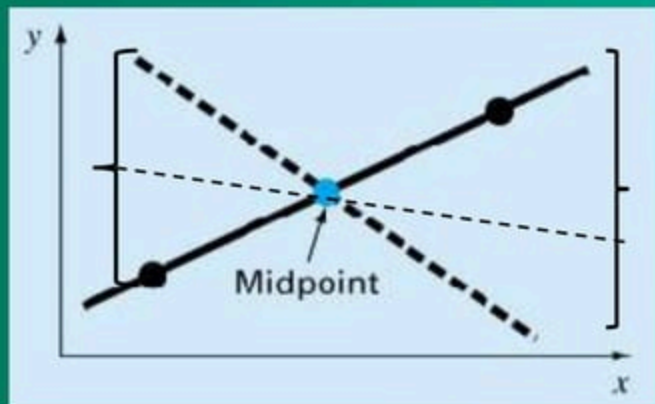
e - error, or residual, between the model and the observations

Linear Regression: Residual



Linear Regression: Criteria for a "Best" Fit

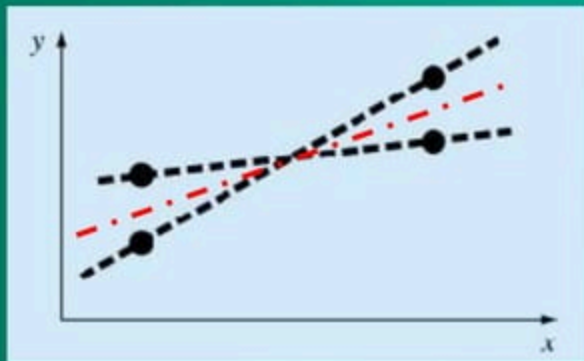
$$\min \sum_{i=1}^n e_i = \sum_{i=1}^n (y_i - a_0 - a_1 x_i)$$



$$e_1 = -e_2$$

Linear Regression: Criteria for a "Best" Fit

$$\min \sum_{i=1}^n |e_i| = \sum_{i=1}^n |y_i - a_0 - a_1 x_i|$$



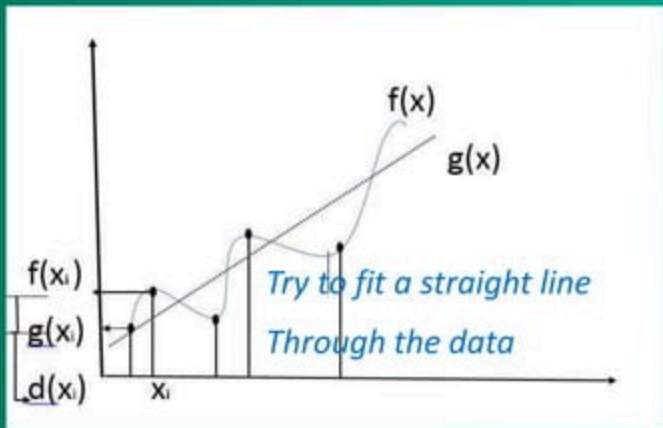
Linear Regression: Criteria for a "Best" Fit

$$\min \max_{i=1}^n |e_i| = |y_i - a_0 - a_1 x_i|$$



Linear curve fitting (Straight line)?

- Given a set of data point $(x_i, f(x_i))$ find a curve that best captures the general trend
- Where $g(x)$ is approximation function



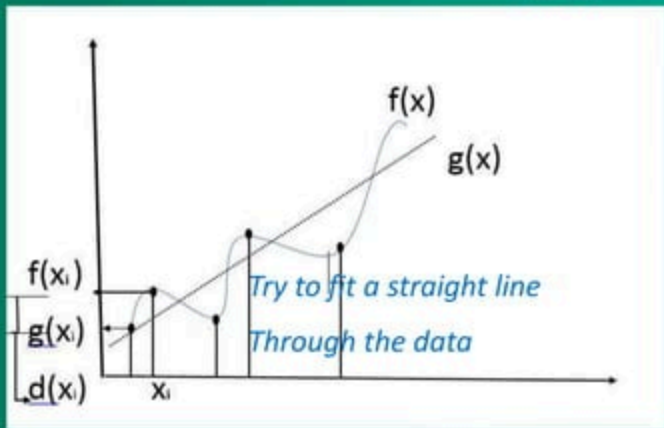
Linear curve fitting (Straight line)?

- Let $g(x) = a_0 + a_1x$

$$\sum d(x_i)^2 = \sum_{i=1}^n |f(x_i) - g(x_i)|^2$$

$$= \sum_{i=1}^n [g(x_i) - f(x_i)]^2$$

$$= \sum_{i=1}^n [a_0 + a_1x_i - f(x_i)]^2$$



Linear curve fitting(straight line)

- Error is a function of a_0 , a_1
- For error (E) to have extreme value:

- $$\frac{\delta E}{\delta a_0} = 0$$

- $$\frac{\delta E}{\delta a_1} = 0$$

- Two equation of two unknowns , solve to get a_0, a_1

Linear Regression: Least Squares Fit

$$S_r = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_{i, \text{measured}} - y_{i, \text{model}})^2 = \sum_{i=1}^n (y_i - a_0 - a_1 x_i)^2$$

$$\min S_r = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - a_0 - a_1 x_i)^2$$

Yields a unique line for a given set of data.

Linear Regression:

Determination of a_0 and a_1

$$\frac{\partial S_r}{\partial a_0} = -2 \sum (y_i - a_0 - a_1 x_i) = 0$$

$$\frac{\partial S_r}{\partial a_1} = -2 \sum [(y_i - a_0 - a_1 x_i) x_i] = 0$$

$$0 = \sum y_i - \sum a_0 - \sum a_1 x_i$$

$$0 = \sum y_i x_i - \sum a_0 x_i - \sum a_1 x_i^2$$

$$\sum a_0 = n a_0$$

$$n a_0 + \left(\sum x_i \right) a_1 = \sum y_i$$

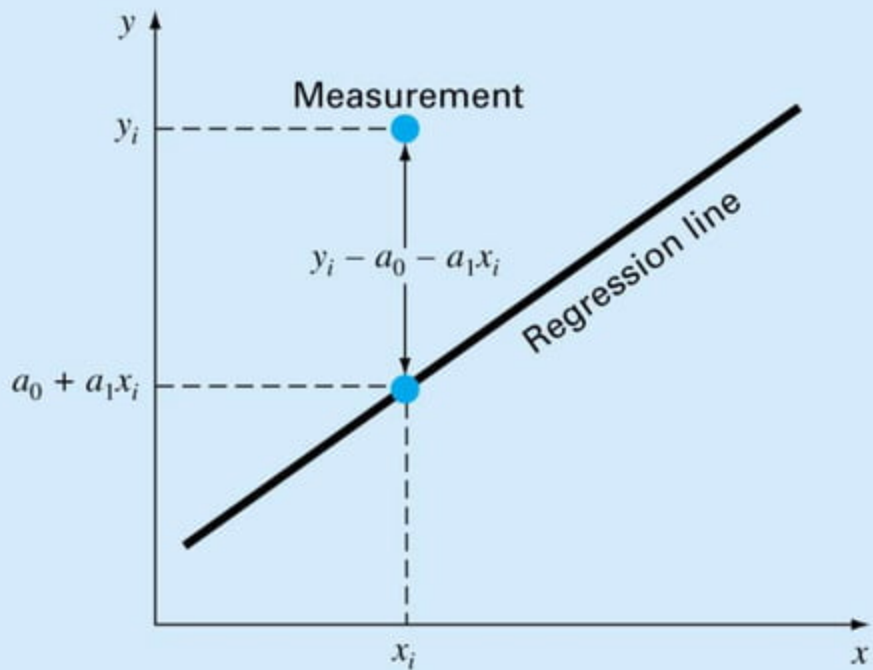
$$\sum y_i x_i = \sum a_0 x_i + \sum a_1 x_i^2$$

} 2 equations with 2 unknowns, can be solved simultaneously

Linear Regression: Determination of a_0 and a_1

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

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



Error Quantification of Linear Regression

- Total sum of the squares around the mean for the dependent variable, y , is S_t

$$S_t = \sum (y_i - \bar{y})^2$$

- Sum of the squares of residuals around the regression line is S_r

$$S_r = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - a_0 - a_1 x_i)^2$$


Example

- The table below gives the temperatures T in $^{\circ}\text{C}$ and Resistance R in Ω of a circuit if $R = a_0 + a_1 T$
- Find the values of a_0 and a_1

| | | | | | | |
|---|------|------|------|------|------|----|
| T | 10 | 20 | 30 | 40 | 50 | 60 |
| R | 20.1 | 20.2 | 20.4 | 20.6 | 20.8 | 21 |

Solution

| $T=X_i$ | $R=y_i$ | $X_i^2 = T^2$ | $X_i y_i = TR$ | $g(x_i) = Y$ |
|------------------|--------------------|---------------------|-----------------------|--------------|
| 10 | 20.1 | 100 | 201 | 20.05 |
| 20 | 20.2 | 400 | 404 | 20.24 |
| 30 | 20.4 | 900 | 612 | 20.42 |
| 40 | 20.6 | 1600 | 824 | 20.61 |
| 50 | 20.8 | 2500 | 1040 | 20.80 |
| 60 | 21 | 3600 | 1260 | 20.98 |
| $\sum x_i = 210$ | $\sum y_i = 123.1$ | $\sum x_i^2 = 9100$ | $\sum x_i y_i = 4341$ | |



Solution

- $6a_0 + 210a_1 = 123.1$ $a_0 = 19.867$
- $210a_0 + 9100a_1 = 4341$ $a_1 = 0.01857$
- $g(x) = 19.867 + 0.01857 * T$

Least Squares Fit of a Straight Line: Example

- 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 |

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

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

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

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

Least Squares Fit of a Straight Line: Example

$$a_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2}$$
$$= \frac{7 \times 119.5 - 28 \times 24}{7 \times 140 - 28^2} = 0.8392857$$

$$a_0 = \bar{y} - a_1 \bar{x}$$
$$= 3.428571 - 0.8392857 \times 4 = 0.07142857$$

$$Y = 0.07142857 + 0.8392857 x$$

Least Squares Fit of a Straight Line: Example (Error Analysis)

| x_i | y_i | $(y_i - \bar{y})^2$ | $e_i^2 = (y_i - \hat{y})^2$ |
|-----------|-------------|---------------------|-----------------------------|
| 1 | 0.5 | 8.5765 | 0.1687 |
| 2 | 2.5 | 0.8622 | 0.5625 |
| 3 | 2.0 | 2.0408 | 0.3473 |
| 4 | 4.0 | 0.3265 | 0.3265 |
| 5 | 3.5 | 0.0051 | 0.5896 |
| 6 | 6.0 | 6.6122 | 0.7972 |
| 7 | 5.5 | 4.2908 | 0.1993 |
| 28 | 24.0 | 22.7143 | 2.9911 |

$$S_t = \sum (y_i - \bar{y})^2 = 22.7143$$

$$S_r = \sum e_i^2 = 2.9911$$

$$r^2 = \frac{S_t - S_r}{S_t} = 0.868$$

$$r = \sqrt{r^2} = \sqrt{0.868} = 0.932$$

Least Squares Fit of a Straight Line: Example (Error Analysis)

- The standard deviation (quantifies the spread around the mean):

$$s_y = \sqrt{\frac{S_r}{n-1}} = \sqrt{\frac{22.7143}{7-1}} = 1.9457$$

- The standard error of estimate (quantifies the spread around the regression line)

$$s_{y/x} = \sqrt{\frac{S_r}{n-2}} = \sqrt{\frac{2.9911}{7-2}} = 0.7735$$

Because $s_{y/x} < s_y$, the linear regression model has good fitness

Algorithm for linear regression

```
SUB Regress(x, y, n, a1, a0, syx, r2)
```

```
sumx = 0: sumxy = 0: st = 0
```

```
sumy = 0: sumx2 = 0: sr = 0
```

```
DO i = 1, n
```

```
    sumx = sumx + xi
```

```
    sumy = sumy + yi
```

```
    sumxy = sumxy + xi*yi
```

```
    sumx2 = sumx2 + xi*xi
```

```
END DO
```

```
xm = sumx/n
```

```
ym = sumy/n
```

```
a1 = (n*sumxy - sumx*sumy)/(n*sumx2 - sumx*sumx)
```

```
a0 = ym - a1*xm
```

```
DO i = 1, n
```

```
    st = st + (yi - ym)2
```

```
    sr = sr + (yi - a1*xi - a0)2
```

```
END DO
```

```
syx = (sr/(n - 2))0.5
```

```
r2 = (st - sr)/st
```

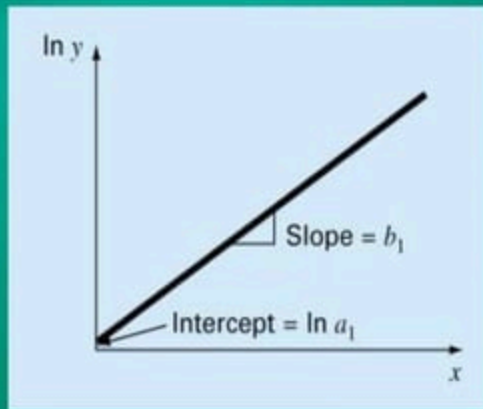
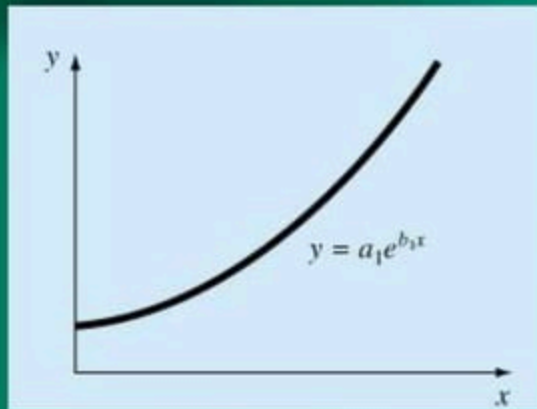
```
END Regress
```

Linearization of Nonlinear Relationships

- The relationship between the dependent and independent variables is linear.
- However, a few types of nonlinear functions can be transformed into linear regression problems.
 - The exponential equation.
 - The power equation.
 - The saturation-growth-rate equation.

Linearization of Nonlinear Relationships

1. The exponential equation.

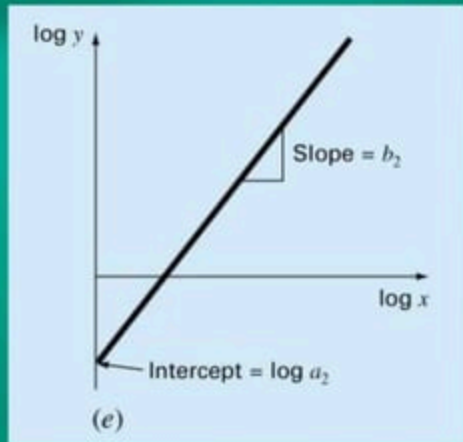
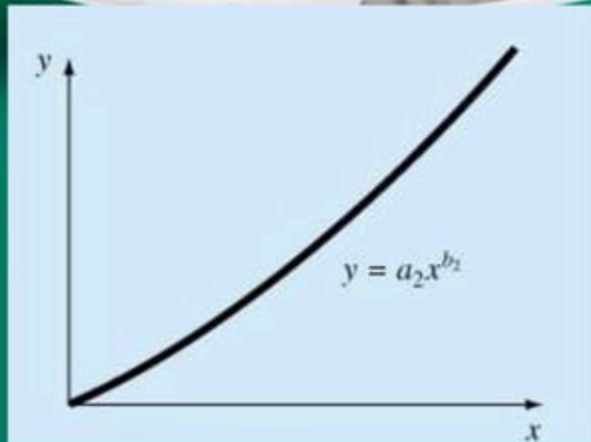


$$\ln y = \ln a_1 + b_1 x$$

$$y^* = a_0 + a_1 x$$

Linearization of Nonlinear Relationships

2. The power equation

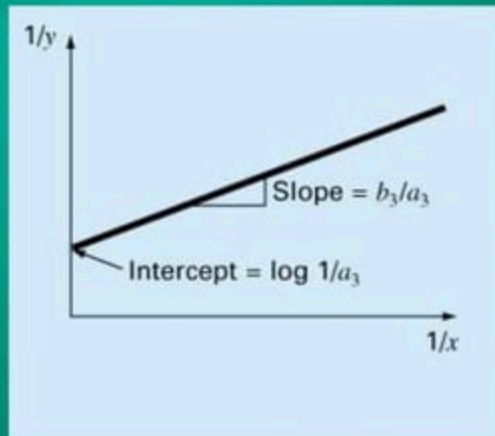
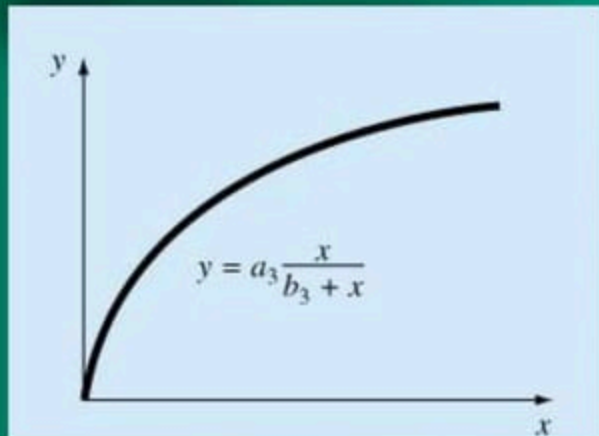


$$\log y = \log a_2 + b_2 \log x$$

$$y^* = a_0 + a_1 x^*$$

Linearization of Nonlinear Relationships

3. The saturation-growth-rate equation



$$\frac{1}{y} = \frac{1}{a_3} + \frac{b_3}{a_3} \left(\frac{1}{x} \right)$$

$$\begin{aligned} y^* &= 1/y \\ a_0 &= 1/a_3 \\ a_1 &= b_3/a_3 \\ x^* &= 1/x \end{aligned}$$

Example

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 | 0.602 |
| 2 | 1.7 | 0.301 | 0.753 |
| 3 | 3.4 | 0.301 | 0.699 |
| 4 | 5.7 | .226 | 0.922 |
| 5 | 8.7 | .447 | 2.079 |
| 15 | 19.7 | .534 | 2.141 |

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

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

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

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

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

Example

| X_i | Y_i | $X_i^* = \text{Log}(X)$ | $Y_i^* = \text{Log}(Y)$ | 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 | 2.079 | 2.141 | 1.424 | 1.169 |

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

Linearization of Nonlinear Functions: Example

$$\log y = -0.334 + 1.75 \log x$$

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

