BIOSTATISTICS

Dr. A. Vinoth

Assistant Professor

Department of Botany

St. Xavier's College

RESIDUALS

- The difference between the observed value of the dependent variable (y) and the predicted value ($\hat{y}$) is called the **residual** (e). Each data point has one residual.

Residual = Observed value - Predicted value

$$\mathbf{e = y - \hat{y}}$$

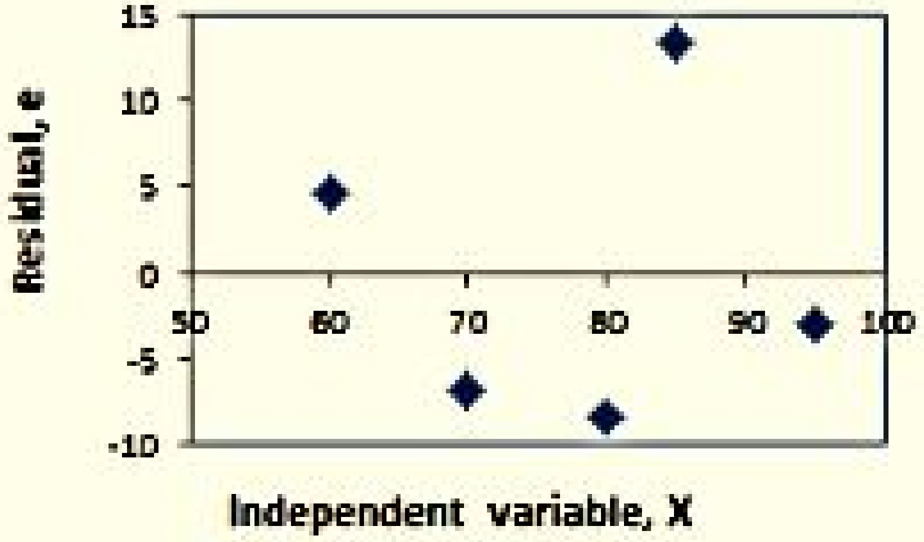
- Both the sum and the mean of the residuals are equal to zero. That is, $\Sigma\, e = 0$ and $e = 0$.

RESIDUAL PLOTS

- A residual plot is a graph that shows the residuals on the vertical axis and the independent variable on the horizontal axis.
- If the points in a residual plot are randomly dispersed around the horizontal axis, a linear regression model is appropriate for the data; otherwise, a non-linear model is more appropriate.

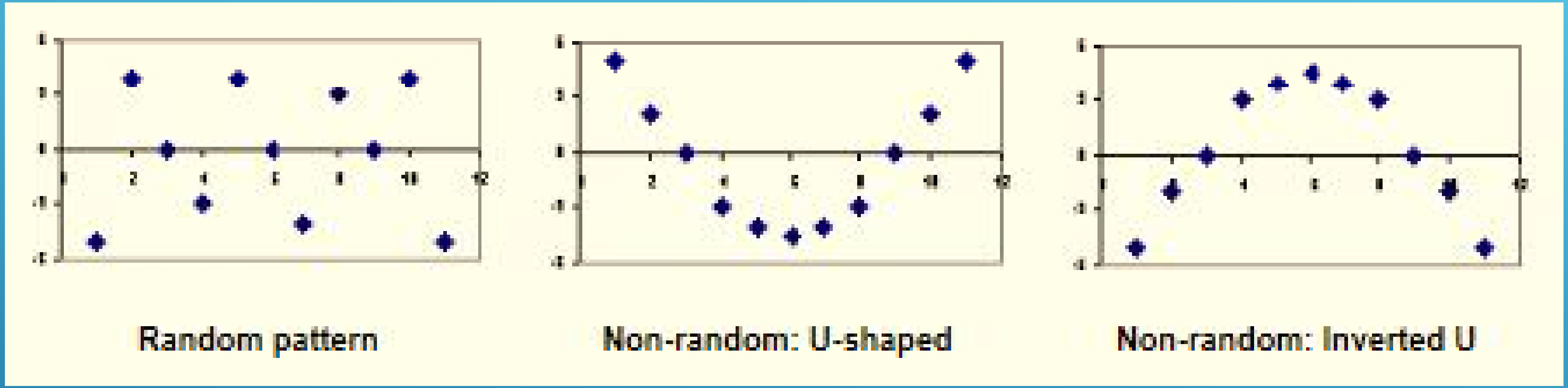
RESIDUAL PLOTS

x	60	70	80	85	95
y	70	65	70	95	85
ŷ	65.411	71.849	78.288	81.507	87.945
e	4.589	-6.849	-8.288	13.493	-2.945



- The residual plot shows a fairly random pattern - the first residual is positive, the next two are negative, the fourth is positive, and the last residual is negative. This random pattern indicates that a linear model provides a decent fit to the data.

RESIDUAL PLOTS



- The residual plots show three typical patterns. The first plot shows a random pattern, indicating a good fit for a linear model. The other plot patterns are non-random (U-shaped and inverted U), suggesting a better fit for a non-linear model.

RESIDUAL PLOTS

In the context of regression analysis, which of the following statements are true?

- I. When the sum of the residuals is greater than zero, the data set is nonlinear.
- II. A random pattern of residuals supports a linear model.
- III. A random pattern of residuals supports a non-linear model.

(A) I only

(B) II only

(C) III only

(D) I and II

(E) I and III

TRANSFORMATIONS TO ACHIEVE LINEARITY

- When a residual plot reveals a data set to be nonlinear, it is often possible to "transform" the raw data to make it more linear.
- This allows us to use linear regression techniques more effectively with nonlinear data.
- Transforming a variable involves using a mathematical operation to change its measurement scale.

TRANSFORMATIONS TO ACHIEVE LINEARITY

Linear transformation.

- A linear transformation preserves linear relationships between variables. Therefore, the correlation between x and y would be unchanged after a linear transformation.
- Examples of a linear transformation to variable x would be multiplying x by a constant, dividing x by a constant, or adding a constant to x.

TRANSFORMATIONS TO ACHIEVE LINEARITY

Nonlinear transformation

- A nonlinear transformation changes (increases or decreases) linear relationships between variables and, thus, changes the correlation between variables.
- Examples of a nonlinear transformation of variable x would be taking the square root of x or the reciprocal of x.

TRANSFORMATIONS TO ACHIEVE LINEARITY

- In regression, a transformation to achieve linearity is a special kind of nonlinear transformation. It is a nonlinear transformation that increases the linear relationship between two variables.

Method	Transformation(s)	Regression equation	Predicted value ($\hat{y}$)
Standard linear regression	None	$y = b_0 + b_1x$	$\hat{y} = b_0 + b_1x$
Exponential model	Dependent variable = log(y)	$\log(y) = b_0 + b_1x$	$\hat{y} = 10^{b_0 + b_1x}$
Quadratic model	Dependent variable = sqrt(y)	$\text{sqrt}(y) = b_0 + b_1x$	$\hat{y} = (b_0 + b_1x)^2$
Reciprocal model	Dependent variable = 1/y	$1/y = b_0 + b_1x$	$\hat{y} = 1 / (b_0 + b_1x)$
Logarithmic model	Independent variable = log(x)	$y = b_0 + b_1\log(x)$	$\hat{y} = b_0 + b_1\log(x)$
Power model	Dependent variable = log(y) Independent variable = log(x)	$\log(y) = b_0 + b_1\log(x)$	$\hat{y} = 10^{b_0 + b_1\log(x)}$

HOW TO PERFORM A TRANSFORMATIONS TO ACHIEVE LINEARITY?

- Conduct a standard regression analysis on the raw data.
- Construct a residual plot.

    If the plot pattern is random, do not transform data.

    If the plot pattern is not random, continue.

- Compute the coefficient of determination (R^2).
- Choose a transformation method.

HOW TO PERFORM A TRANSFORMATIONS TO ACHIEVE LINEARITY?

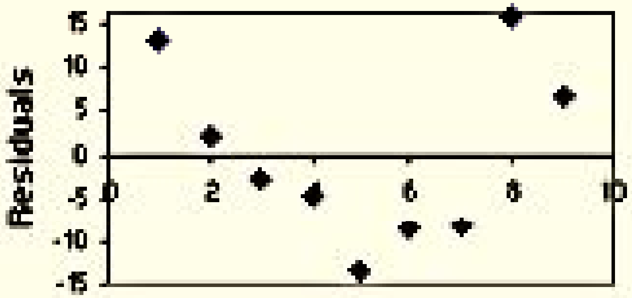
- Transform the independent variable, dependent variable, or both.
- Conduct a regression analysis, using the transformed variables.
- Compute the coefficient of determination (R^2), based on the transformed variables.

If the transformed R^2 is greater than the raw-score R^2, the transformation was successful. Congratulations!

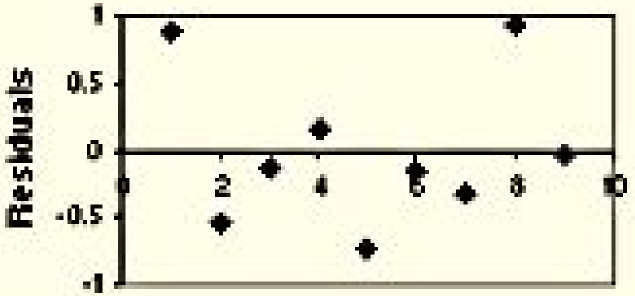
If not, try a different transformation method.

HOW TO PERFORM A TRANSFORMATIONS TO ACHIEVE LINEARITY?

x	1	2	3	4	5	6	7	8	9
y	2	1	6	14	15	30	40	74	75



x	1	2	3	4	5	6	7	8	9
yt	1.41	1.00	2.45	3.74	3.87	5.48	6.32	8.60	8.66



- The plot suggests that the transformation to achieve linearity was successful. And the coefficient of determination was 0.96 with the transformed data versus only 0.88 with the raw data. The transformed data resulted in a better model.

TRANSFORMATIONS TO ACHIEVE LINEARITY

In the context of regression analysis, which of the following statements is true?

- I. A linear transformation increases the linear relationship between variables.
- II. A logarithmic model is the most effective transformation method.
- III. A residual plot reveals departures from linearity.

(A) I only

(B) II only

(C) III only

(D) I and II only

(E) I, II, and III