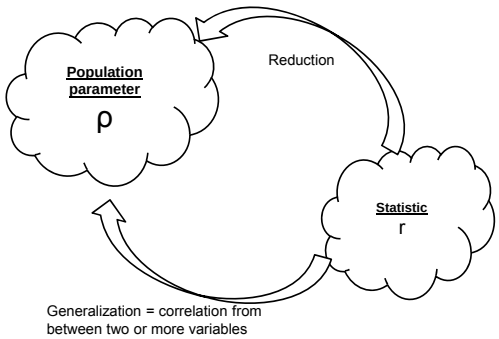
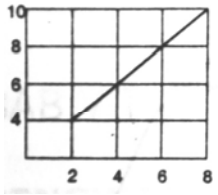
Associative Hypothesis Test: Correlation And Regression

Presented by:
Mahendra AN

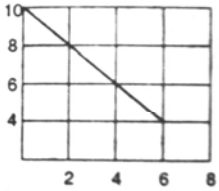
Basic Principles



Basic Concepts

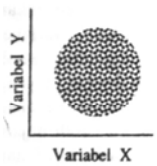


Positive correlation

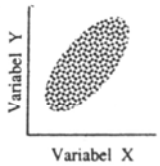


Negative correlation

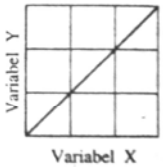
Basic Concepts



r=0



r=0.5



r=1

Statistics Techniques

Data Types	Correlation techniques
Nominal	1. Contingency coefficient
Ordinal	1. Rank Spearman 2. Tau Kendal
Interval or ratio	1. Product moment Pearson 2. Multiple correlation 3. Partial correlation

Parametric Statistics for Correlations

1. Product moment

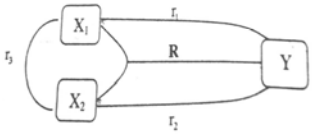
- For searching of correlation between two variables for interval or ratio data from the same source and normal distributed

$$r_{xy} = \frac{\sum xy}{\sqrt{\sum x^2 y^2}}$$

- Alternative significance test

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

2. **Multiple correlation Vs partial correlation**



Parametric Statistics for Correlations

Multiple correlation

$$R_{y.x_1x_2}=\frac{r_{yx_1}^2+r_{yx_2}^2-2r_{yx_1}r_{yx_2}r_{x_1x_2}}{1-r_{x_1x_2}^2}$$

- Alternative significance test

$$Fh=\frac{R^2/k}{(1-R^2)/(n-k-1)}$$

Partial correlation

$$R_{y.x_1x_2}=\frac{r_{yx_1}-r_{yx_2}\cdot r_{x_1x_2}}{\sqrt{1-r_{x_1x_2}^2}\cdot\sqrt{1-r_{yx_2}^2}}$$

- Alternative significance test

$$t=\frac{r_p\sqrt{n-3}}{\sqrt{1-r_p^2}}$$

Correlation Coefficient Interpretation

Coefficient Interval	Correlation
0.00 – 0.199	Very low
0.02 – 0.399	Low
0.40 – 0.599	Medium
0.60 – 0.799	Strong
0.80 – 1.000	Very strong

Non parametric statistics for correlations

1. Contingency coefficient

- For nominal data
- Strong connection with chi square

$$C=\sqrt{\frac{\chi^2}{N+\chi^2}}$$

$$\chi^2=\sum_{i=1}^{r}\sum_{j=1}^{k}\frac{(OP_{ij}+E_{ij})^2}{EP_{ij}}$$

2. Rank Spearman

- For ordinal data from different data sources and not normal distributed

$$\rho=1-\frac{6\sum b_i^2}{n(n^2-1)}$$

- Alternative significance test

$$z_h=\frac{\rho}{\frac{1}{\sqrt{n-1}}} \qquad t=r\sqrt{\frac{n-2}{1-r^2}}$$

Non parametric statistics for correlations

3. Tau Kendal

- For ordinal or rank data
- For more than 10 data
- Can be used for searching partial correlation

$$\tau=\frac{\sum A-\sum B}{\frac{N(N-1)}{2}}$$

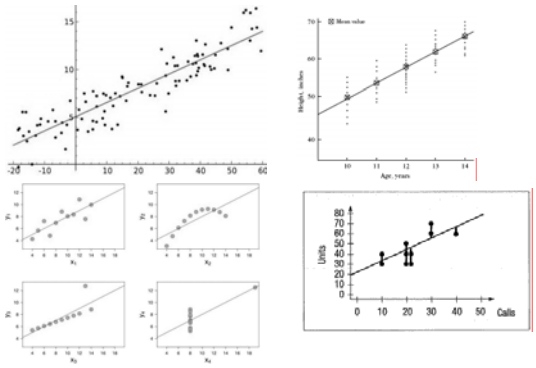
- Significance test

$$z=\frac{\tau}{\sqrt{\frac{2(2N+5)}{9N(N-1)}}}$$

Regression

- Correlation search direction and strength of symmetric , causal or reciprocal relationships between two variables while regression predict changes of dependent variables when independent variables are change

Linear Regression Lines



Linear Regression

1. Simple Linear Regression

$$Y = a + bX + \varepsilon$$

$$b = r\frac{S_y}{S_x} \quad a = Y - bX$$

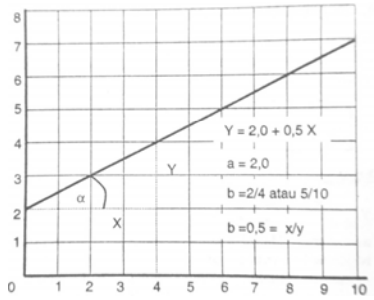
$$a = \frac{(\sum Y_i)(\sum X_i^2) - (\sum X_i)(\sum X_i Y_i)}{n\sum X_i^2 - (\sum X_i)^2}$$

$$b = \frac{n\sum X_i Y_i - (\sum X_i)(\sum Y_i)}{n\sum X_i^2 - (\sum X_i)^2}$$

2. Multiple linear Regression

$$Y = a + b_1X_1 + b_2X_2 + \ldots + b_nX_n + \varepsilon$$

Simple Linear Regression Line



Simple Linear Regression Test Steps

- Linearity test $F = \frac{S_{TC}^2}{S_G^2}$
- Count a and b
- Build regression equation and draw regression line
- Significance test $F = \frac{S_{reg}^2}{S_{res}^2}$
- Correlation hypothesis test

$$r = \frac{n\sum X_i Y_i - (\sum X_i)(\sum Y_i)}{\left(n\sum X_i^2 - (\sum X_i)^2\right)\left(n\sum Y_i^2 - (\sum Y_i)^2\right)}$$

Kita membangun langit dengan segala daya yang ada
kemudian merasa langit kita sudah cukup tinggi
bahkan mengungguli langit lain.
Tetapi akan selau ada langit di atas kita.

== MAHENDRA ADHI NUGROHO, 2008 ==