

Key Concepts and Terminology

Correlation: A relationship between two variables based on co-occurrence. Relationship between variables that is mutually inclusive. Correlation is NOT causation.

Causation: The occurrence of one event is directly responsible for the occurrence of another event.

Pearson Correlation: a version of linear correlation that is based on the Pearson Product Moment Coefficient (PPMC). Valid for 2 continuous (interval or ratio) variables.

Point Biserial Correlation: Test correlation between a dichotomous categorical variable (T or F) and a continuous variable.

Bivariate normality: Checking to see if both variables are normal.

Line of Best fit: Mathematical predictor equation. Aka the regression line. The line is only as good as the data.

Least-Squares Method: Used to determine the equation of the regression line by error minimization (using the lowest possible residual values).

$$\text{LSM} = \sum \text{Negative residuals} + \text{Positive residuals} = 0$$

Auto-correlation: Applicable in linear regression is the absence of independence of observations, as a result, absence of independence of residuals.

F-ratio: $F - ratio = \frac{MS_M}{MS_R}$, $MS_M = SOS_{regression}$, $MS_R = SOS_{residuals}$

T-test: Tells whether the coefficients of the regression equation differ from 0 thus helps understand the significance of the magnitude of coefficients.

Anova: Anova in linear regressions produces an F-ratio that tells if the regression model is a statistically significant predictor of the outcome variable from the predictor variable.

Pearson's correlation coefficient (r): Interpreted using cohen's standards and is highly sensitive to sample size (directly proportional). It is a measure of effect size.

Pearson's r-square: It's a square of r and this provides the actual percentage of the predictor variable that affects changes in the outcome variable.

Spearman's Rank order: Non-parametric equivalent to the Pearson product moment correlation. Used when assumptions of PPM are violated. Ranks of the data and not the raw data is used.

Rho (r_s): Test statistic for spearman's rank order and ranges from -1 to +1 with 0 meaning no correlation.

Key Concepts and Terminology

Point biserial correlation: A form of correlation with 1 dichotomous categorical variable and the other continuous.

Biserial PPM: A modified version of PPM. Image included in output folder. Easier to do in SPSS and also other parametric or non-parametric test can be used in lieu of this.