Unit 12 Power words

contingency table is also called a **two-way frequency table** which list bivariate data or data for two variables.

joint frequency the table cells that show the frequency or count of both variables

marginal frequency the table cells that show the total of each category.

Relative Frequency Table: A two-way frequency table in which all frequencies are divided by the total (Also called probability)

Conditional" relative frequency: a relative frequency is determined based upon a row or column (specific condition) and the cells of that raw or column is divided by the total of the row or column.

Association: The bigger the differences in the conditional relative frequencies, the stronger the association between the variables. If the conditional relative frequencies are nearly equal for all categories, there may be no association between the variables. Such variables are said to be independent.

A scatter plot is a graphical representation to display numerical data with two variables

Correlation: A relationship between two numerical variables that can be described as Positive, negative, or no correlation

A **line of best fit** (or "trend" line) is a straight line that best represents the bivariate data on a scatter plot. This line may pass through some of the points, none of the points, or all of the points.

Interpolating = predicting values within the given data

Extrapolating = predicting values outside of the given data

Correlation Coefficient is a value that describes the accuracy of the regression in representing the data. The value of the correlation coefficient is equal to or between -1 and 1. The closer the correlation coefficient to 1 and -1 the more accurate the regression. A value of 0 represent no correlation between the variables.

The residual of a data point is defined as the difference between the observed y-value and the predicted y-value from the equation of the line of best fit. Residual = Actual value – Predicted value

Residual graphs determine if a linear model is an appropriate model to use on bivariate data. We do not want residuals to make a distinct pattern. If so, then it is likely that a linear model is not appropriate to fit the data and perhaps an exponential or quadratic model is better