

**Unit 9 & Unit 10, Creating Cross-tabulations and Performing Chi-Square Analysis & Correlation →**

Cross-tabulation, chi-square analysis, and correlation are statistical tools used to examine relationships between variables.

**Cross-Tabulation**

Cross-tabulation, also known as contingency tables or cross-tabs, is a method used to analyse the relationship between two or more categorical variables. It presents the data in a matrix format, allowing for a visual comparison of the frequency distribution across categories. Cross-tabs are particularly useful in identifying patterns, trends, and potential anomalies within the data. They are often the first step in data analysis, providing a foundational understanding that guides further statistical investigation.

<b>Gender</b>	<b>Chocolate</b>	<b>Strawberry</b>	<b>Vanilla</b>	<b>Row Total</b>
<b>Female</b>	Raw: 37 Row%: 56% Col%: 63.8%	Raw: 17 Row%: 25.8% Col%: 48.6%	Raw: 12 Row%: 18.2% Col%: 28.8%	Raw: 66 Row%: 100%
<b>Male</b>	Raw: 21 Row%: 29.6% Col%: 36.2%	Raw: 18 Row%: 25.4% Col%: 51.4%	Raw: 32 Row%: 45.0% Col%: 71.2%	Raw: 71 Row%: 100%
<b>Total</b>	Raw: 58 Col%: 100%	Raw: 35 Col%: 100%	Raw: 44 Col%: 100%	137

**Chi-Square Analysis**

Chi-square analysis is a statistical test applied to sets of categorical data to evaluate how likely it is that any observed difference between the sets arose by chance. It's especially useful following a cross-tabulation to test the independence of two variables. The test calculates a chi-square statistic that reflects the differences between the observed counts in the data and the counts one would expect if there were no relationship between the variables. A significant chi-square result suggests that a relationship exists between the variables. There are two main types of chi-square tests: the chi-square test of independence and the chi-square goodness-of-fit test. The former assesses whether two categorical variables are related, while the latter tests how well an observed distribution matches an expected distribution.

**Correlation**

Correlation is a statistical technique used to measure and describe the direction and strength of the relationship between two continuous variables. The most common correlation coefficient is Pearson's  $r$ , which ranges from -1 to 1. A correlation of -1 indicates a perfect negative linear relationship, 0 indicates no linear relationship, and 1 indicates a perfect positive linear relationship. Correlation does not imply causation; it merely suggests a degree to which two variables change

together. Spearman's rank correlation is another method used when the data do not meet the assumptions required for Pearson's correlation, providing a measure of a monotonic relationship.

