

# ONE-WAY TABULATION

This represents the most basic tabulation. It goes by various names including **marginal tabulation, one-way tabulation and the frequency distribution**. It consists of a simple count of the number of responses that occur in each of the data categories that comprise a variable.

We run one-way tabulations in order to:

- determine the degree of non-response
- locate blunders
- locate outliers
- determine the empirical distribution of the variable in question
- calculate summary statistics

**Statistics**

Sealtest: Fresh Taste

N	Valid	264
	Missing	43
Mean		3.06
Median		3.00
Mode		3
Std. Deviation		1.17
Variance		1.38
Skewness		-.082
Std. Error of Skewness		.150
Kurtosis		-.762
Std. Error of Kurtosis		.299
Range		4
Minimum		1
Maximum		5
Sum		807

Measures of central tendency

Measures of dispersion/variation

Number of people who responded or provided data

Number of people who did not answer or provide data

Shape of the distribution

Frequency divided by total # of respondents who provided data (does not include missing values)

Variable label (name) → Sealtest: Fresh Taste

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Does Not Describe Well	30	9.8	11.4	11.4
	2	52	16.9	19.7	31.1
	3	87	28.3	33.0	64.0
	4	63	20.5	23.9	87.9
	Describes Well	32	10.4	12.1	100.0
	Total	264	86.0	100.0	
Missing	System	43	14.0		
	Total	307	100.0		

Value label

Values

Total # people who answered question

Number of people not answering question

Total # of respondents

Sum of the valid percent and all rows before it

Count - # of times response chosen

Frequency divided by total # of respondents

# A Review of Descriptive Statistics

## Measures of Central Tendency

The measure of central tendency is a report of the “average” response. That is, the most typical response, or a person’s “best guess” of someone picked at random from the population sampled would respond. The measure you use depends upon the type of scale data (remember the charts for choosing the appropriate analytical technique). The three most commonly used measures of central tendency are:

- Mode
  - The mode is simply the category or value with the greatest frequency of cases. It is used to describe the most typical case for nominal data.
- Median
  - The median value is the preferred measure of central tendency for ordinal data. In a sample with an odd number of cases, the median is the middle observation when the data values are ordered from smallest to largest. When the sample size is even, the median is the average of the two middle values.
  - Since it’s merely the middle value, the median isn’t affected by extreme values of the existence of outliers. To put it another way, the middle value or median will be the same, no matter how much greater or lesser the values of the extremes turn out to be. Consequently, the median is likely to be more typical of the *majority* of cases and a better average to use than the mean with asymmetrical distributions where there are outliers.
- Mean
  - The mean is the arithmetic average.
  - It is the most common average used to indicate the most typical response. While the median and the mode can also be used to describe central tendency for interval or ratio data, the mean is usually the most meaningful statistic. It is computed by dividing the sum of the values by the number of values or cases. It is not appropriate for nominal or ordinal data.
  - The mean tends to be overly sensitive to influence by only one or a few extreme values in the distribution (outliers) – meaning an extremely low value (or values) can lower the mean and an extremely high value (or values) can raise the mean. With outliers, the mean isn’t very typical of either the vast majority of cases or of the outliers. When this occurs, most people use the median (or sometimes the mode) in place of the mean.

# A Review of Descriptive Statistics

## Measures of Dispersion/Variation

Measures of central tendency provide information only about “typical” values. They tell you nothing about how much the values vary within the sample. Measures of dispersion describe how close to the measure of central tendency the rest of the values in the distribution fall. The most commonly used measures of dispersion include:

- Range
  - The range identifies the distance between the lowest value (minimum) and the highest value (maximum) in an ordered set of values. It shows the extent of the spread between extremes. The range provides some information on the dispersion by indicating how far apart the extremes are found (it’s basically the width or spread of the distribution).
  
- Variance and Standard Deviation
  - Variance and standard deviation are measures of dispersion that permit a person to estimate a report the proportion of respondents or cases within certain ranges in the center part of the distribution, as well as toward the extremes, *providing that the distribution doesn’t deviate very markedly from the normal, bell-shaped curve.*
  - The variance and standard deviation are measures of the deviation or spread away from the mean. It’s a single value that indicates the amount of spread in a distribution, or the average distance of the distribution values from the mean.
  - The standard deviation (the square root of the variance) allows you to obtain a measure in the same units as the original data.

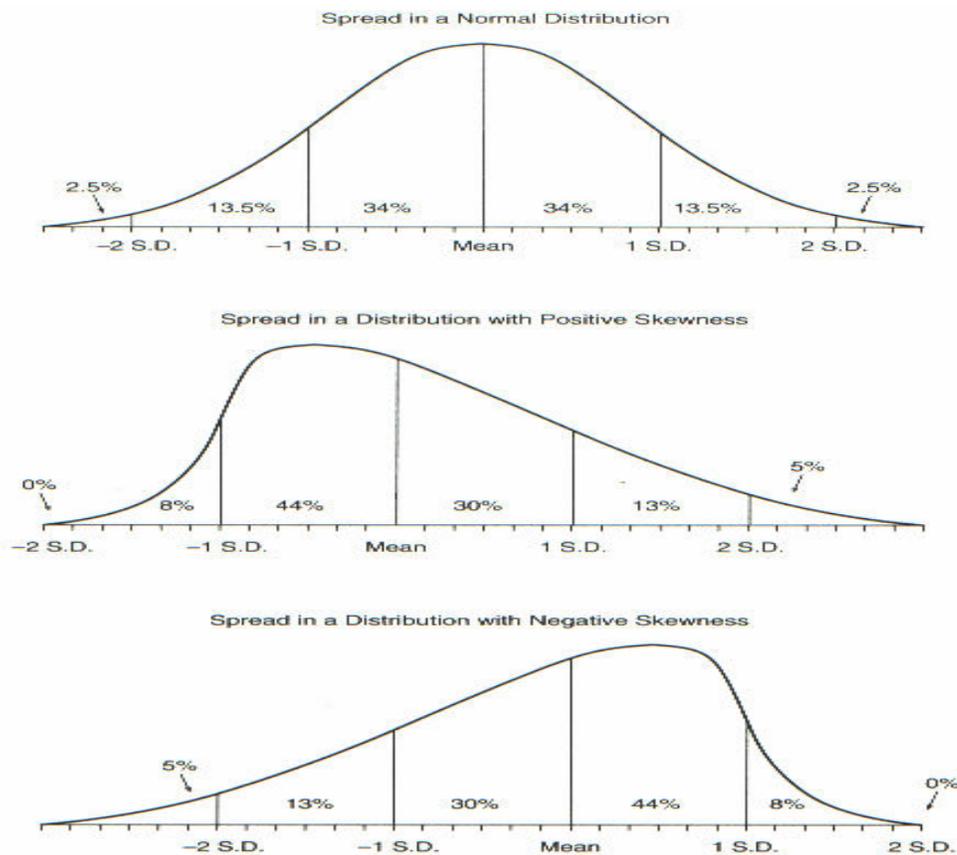
# A Review of Descriptive Statistics

## Shape of the Distribution

The shape of a distribution can be inspected by displaying it graphically. That's often useful but it can take substantial time, effort and space to produce the graphs. Additionally, graphs cannot be manipulated statistically. Because of this, two coefficients are typically used which indicate the degree and direction of the directions of interval and ratio data with large numbers of values that deviate from a normal distribution. They are:

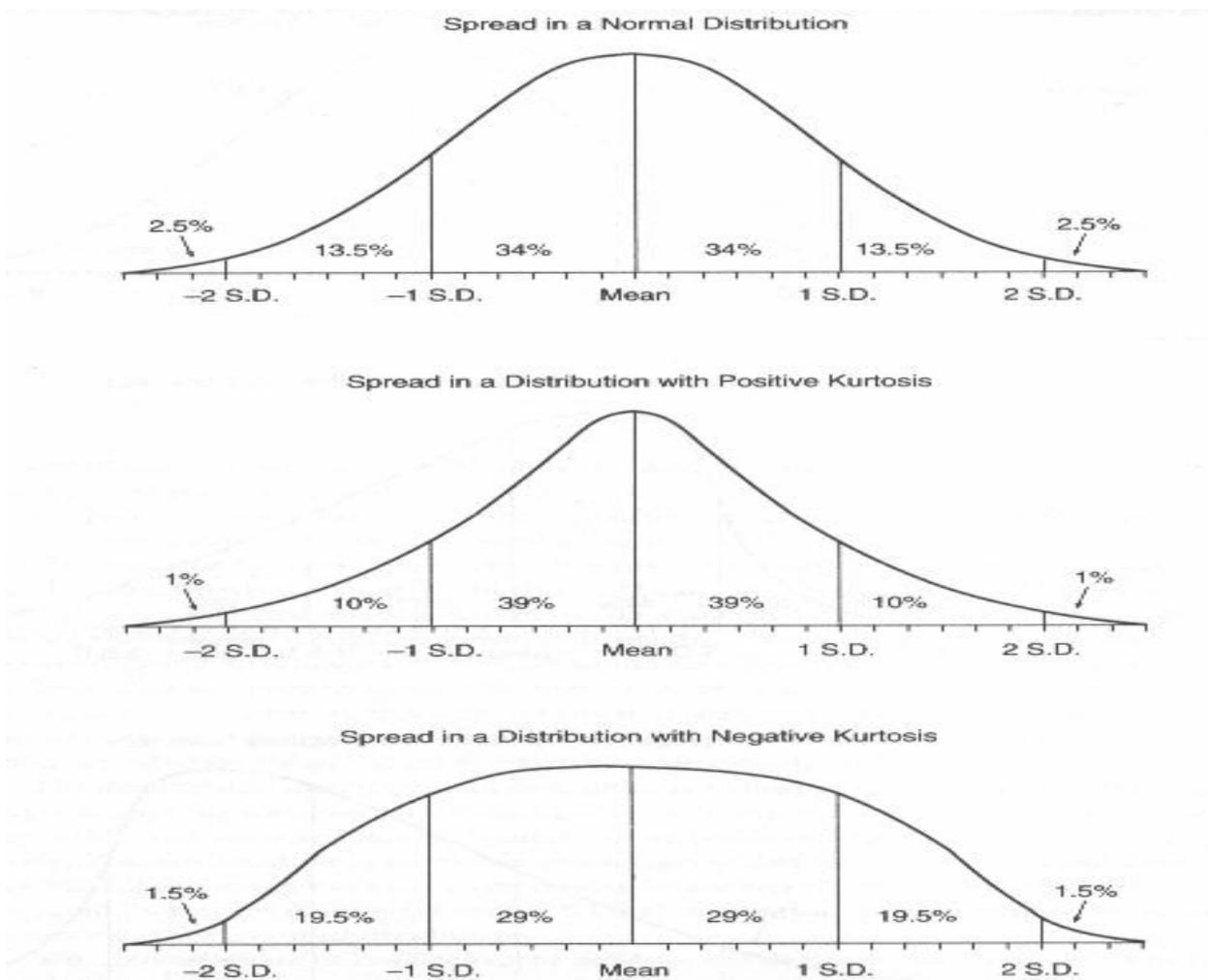
- **Skewness of the Distribution**

- Skewness of a distribution is the degree and direction of its asymmetry.
- If a distribution is symmetrical, such as the normal curve, one side of the distribution is precisely the “mirror-image” of the other and the coefficient of skewness will be zero.
- Often distributions of data for variables are somewhat asymmetrical; they “lean” toward one direction or the other.



- **Kurtosis of the Distribution**

- The kurtosis of a distribution is an indication of peaked or flat it is, compared to the normal curve.
- The value of the coefficient of kurtosis for a normal curve is zero.
- The coefficient of kurtosis will be positive if the distribution has a higher, narrower peak, with the tails extending out in each direction at a low level.
- By contrast, a negative value of kurtosis indicates a curve with a broad, low peak and “wings” that extend out at a rather high level before falling off.



## TWO-WAY TABULATION

A two-way tabulation, more commonly referred to as a **cross tabulation** is the most widely used data analysis technique in marketing research. It involves the simultaneous counting of the number of observations that occur in each of the data categories of two or more variables.

Characteristics:

- ❖ Banners – the columns – often demographic variables (e.g. gender) – that are used as groups for difference testing
- ❖ Stubs – the rows – usually the variables in which we’re using to measure a difference
- ❖ Cells – the intersection of the banner and stub

**Case Processing Summary**

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
How often do you eat ice cream? * Gender	304	99.0%	3	1.0%	307	100.0%

Stub (row)  
Variable

Banner (column)  
Variable

Total # respondents  
who answered both  
questions

# respondents who  
did not answer  
either or both  
questions

Total number of  
respondents/records  
in database

Percentage of total  
respondents

Percentage of total  
respondents

**Stub Variable**

**Banner Variable**

**How often do you eat ice cream? \* Gender Crosstabulation**

			Gender		Total
			Female	Male	
How often do you eat ice cream?	4-7 times per week	Count	62	88	150
		% within How often do you eat ice cream?	41.3%	58.7%	100.0%
		% within Gender	50.0%	48.9%	49.3%
		% of Total	20.4%	28.9%	49.3%
1-3 times per week	Count	Count	19	37	56
		% within How often do you eat ice cream?	33.9%	66.1%	100.0%
		% within Gender	15.3%	20.6%	18.4%
		% of Total	6.3%	12.2%	18.4%
1-3 times per month	Count	Count	31	46	77
		% within How often do you eat ice cream?	40.3%	59.7%	100.0%
		% within Gender	25.0%	25.6%	25.3%
		% of Total	10.2%	15.1%	25.3%
Less than once a month	Count	Count	12	9	21
		% within How often do you eat ice cream?	57.1%	42.9%	100.0%
		% within Gender	9.7%	5.0%	6.9%
		% of Total	3.9%	3.0%	6.9%
Total	Count	Count	124	180	304
		% within How often do you eat ice cream?	40.8%	59.2%	100.0%
		% within Gender	100.0%	100.0%	100.0%
		% of Total	40.8%	59.2%	100.0%

**Row %**

**Column %**

**Total %**

**Of those who eat ice cream 1 – 3 times per week, 33.9% are female**

**3.9% of those responding are female AND eat ice cream less than once a month**

**25.6% of males eat ice cream 1 - 3 times per month**

# Choosing the Appropriate Analytical Technique

## What is the goal of the analysis?

There are four possible alternatives; each which will lead you down the appropriate path for choosing the correct analytical technique.

- ❖ If the goal of the analysis is to determine the “average” response (the measure of central tendency) or understand the distribution of responses (measures of variation/dispersion), you want to **Summarize the Data** – go to page 9.
  
- ❖ If the goal of the analysis is to determine whether or not there are statistically significant differences (1) between different groups (e.g., men and women, etc.) on an individual question or (2) between two or more individual questions/attributes (also known as variables) or between one variable and a normative benchmark, your goal is to **Determine Differences** – go to page 11.
  
- ❖ If the goal of the analysis is to examine the existence and/or strength of the relationship between two or more variables (e.g., Is the amount of income a person earns related to the number of years of schooling obtained?), your goal is to **Measure Association** – go to page 12.
  
- ❖ If the goal of the analysis is to determine if one or more variables predict the outcome of another variable (e.g. Is the number of season ticket sales for the football team this year dependent on the number of games won in the previous season?), your goal is to **Assess the Impact of One Variable on Another** – go to page 13.

## Goal of the Analysis: Summarizing Data

1. Determine the level of measurement used for the answer categories of the question.
2. If you want to determine the “average” response, refer to the Measure of Central Tendency column and run the analytical measure listed.
3. If you want to understand the distribution of responses, refer to the Measures of Variation/Dispersion column and run the analytical measure listed.

Level of Measurement	Measures of Central Tendency	Measures of Variation/Dispersion
<b>Nominal</b>	<ul style="list-style-type: none"> <li>• Mode</li> </ul> <p>This statistic can be run in the Frequency procedure in SPSS</p>	
<b>Ordinal</b>	<ul style="list-style-type: none"> <li>• Median</li> </ul> <p>This statistic can be run in the Frequency procedure in SPSS</p>	<ul style="list-style-type: none"> <li>• Range</li> </ul> <p>This statistic can be run in either the Frequency or Descriptives procedures in SPSS</p>
<b>Interval or Ratio</b>	<ul style="list-style-type: none"> <li>• Arithmetic Mean</li> </ul> <p>This statistic can be run in either the Frequency or Descriptives procedures in SPSS</p>	<ul style="list-style-type: none"> <li>• Standard Deviation</li> <li>• Variance</li> </ul> <p>These statistics can be run in either the Frequency or Descriptives procedures in SPSS</p>

# Goal of the Analysis: Determining Differences

Use the **Determining Differences Analytical Test Selection** table on the following page and answer the following questions.

1. What is the nature of the difference?
  - If you are trying to determine if a difference exists between either (a) two or more individual questions/attributes/variables (e.g., customer satisfaction in 2002 and customer satisfaction in 2003) or (b) one variable and a normative benchmark (e.g. your company's customer satisfaction compared to the industry average), go to the **1 Group of Respondents** section.
    - ✓ If you are trying to determine if there is a difference between one variable and a normative benchmark, go to the row labeled "looking for differences between 1 group and a norm (or expected value)".
    - ✓ If you are trying to determine if there is a difference between two variables, go to the row labeled "looking for differences where the 1 group evaluates two attributes".
    - ✓ If you are trying to determine if there is a difference between three or more variables, go to the row labeled "looking for differences where the 1 group evaluates three or more attributes".
  - If you are trying to determine if there is a difference between 2 or more groups on a particular question/attribute/variable, go to the **2 or More Groups of Respondents** section.
    - ✓ If you have only 2 groups (e.g. males vs. females), go to the row labeled "looking for a difference in a single attribute between 2 groups".
    - ✓ If you have 3 or more groups (e.g., freshman, sophomore, junior, senior), go to the row labeled "looking for a difference in a single attribute between 3 or more groups".
2. Determine the level of measurement used for the answer categories of the variable(s) for which you want to determine whether or not there is a difference (the level of measurement of your grouping variable, if you are comparing differences between two or more groups, is not relevant). Go to the appropriate column.
  - Note: If you are looking for differences between two or more variables and the variables use different levels of measurement, use the column for the lowest level of measurement (e.g., if you are looking for a difference between an interval variable and an ordinal variable, use the ordinal column).
3. If more than one test is given, you can select the statistical test you wish to run.

# Determining Differences Analytical Test Selection

Nature of the Difference		Level of Measurement		
		Nominal	Ordinal	Interval or Ratio
<b>1 Group of Respondents</b>	Looking for differences between 1 group and a norm (or expected value)	<ul style="list-style-type: none"> <li>Binomial Test</li> <li>Chi-Square Test</li> </ul> <p>Both of these tests can be found in SPSS under Analyze, Nonparametric Tests in SPSS</p>	<ul style="list-style-type: none"> <li>Kolmogorov-Smirnov Test</li> </ul> <p>This test can be found in SPSS under Analyze, Nonparametric Tests, 1-Sample K-S</p>	<ul style="list-style-type: none"> <li>One-Sample T-Test</li> </ul> <p>This test can be found in SPSS under Analyze, Compare Means</p>
	Looking for differences where the 1 group evaluates two attributes	<ul style="list-style-type: none"> <li>McNemar Test</li> </ul> <p>This test can be found in SPSS under (1) Analyze, Nonparametric Tests, 2 Related Samples or (2) Analyze, Descriptive Statistics, Crosstabs</p>	<ul style="list-style-type: none"> <li>Sign Test</li> <li>Wilcoxon Test</li> </ul> <p>Both of these tests can be found in SPSS under Analyze, Nonparametric Tests, 2 Related Samples</p>	<ul style="list-style-type: none"> <li>Paired Samples t-Test</li> </ul> <p>This test can be found in SPSS under Analyze, Compare Means</p>
	Looking for differences where the 1 group evaluates three or more attributes	<ul style="list-style-type: none"> <li>Cochran Q Test</li> </ul> <p>This test can be found in SPSS under Analyze, Nonparametric Tests, K Related Samples</p>	<ul style="list-style-type: none"> <li>Friedman Test</li> </ul> <p>This test can be found in SPSS under Analyze, Nonparametric Tests, K Related Samples</p>	<ul style="list-style-type: none"> <li>Repeated Measures Analysis of Variance (ANOVA)</li> </ul> <p>This test can be found in SPSS under General Linear Model, Repeated Measures</p>
<b>2 or More Groups of Respondents</b>	Looking for a difference in a single attribute between 2 groups	<ul style="list-style-type: none"> <li>Chi-Square Test</li> </ul> <p>This test can be found in SPSS under Analyze, Descriptive Statistics, Crosstabs</p>	<ul style="list-style-type: none"> <li>Mann-Whitney Test</li> </ul> <p>This test can be found in SPSS under Analyze, Nonparametric Tests, 2 Independent Samples</p>	<ul style="list-style-type: none"> <li>Independent Samples t-Test</li> </ul> <p>This test can be found in SPSS under Analyze, Compare Means</p>
	Looking for a difference in a single attribute between 3 or more groups	<ul style="list-style-type: none"> <li>Chi-Square Test</li> </ul> <p>This test can be found in SPSS under Analyze, Descriptive Statistics, Crosstabs</p>	<ul style="list-style-type: none"> <li>Kruskal-Wallis Test</li> </ul> <p>This test can be found in SPSS under Analyze, Nonparametric Tests, K Independent Samples</p>	<ul style="list-style-type: none"> <li>One-Way Analysis of Variance (ANOVA)</li> </ul> <p>This test can be found in SPSS under Analyze, Compare Means</p>

# Goal of the Analysis: Measuring Association

1. How many variables do you want to determine are related/associated?
  - ✓ If you are trying to see if two variables are related, go to the row labeled “2 Variables – Bivariate Analysis”.
  - ✓ If you are trying to determine the association or relationship between 3 or more variables, go to the row labeled “More than 2 Variables – Analysis of Interdependence”.
  
2. Determine the level of measurement for the answer categories of the variables.
  - ✓ If the variables all have the same level of measurement, go to the appropriate column.
  - ✓ If the variables have different levels of measurement, use the column for the lowest level of measurement (e.g., if you are looking to determine the relationship between two variables, one interval and one ordinal, use the Ordinal column).

## Measuring Association Analytical Test Selection

Number of Variables	Level of Measurement		
	Nominal	Ordinal	Interval or Ratio
2 Variables - Bivariate Analysis	<ul style="list-style-type: none"> <li>• Phi Coefficient</li> <li>• Contingency Coefficient</li> </ul> <p>This test can be found in SPSS under Analyze, Descriptive Statistics, Crosstabs</p>	<ul style="list-style-type: none"> <li>• Spearman Rank Correlation</li> </ul> <p>This test can be found in SPSS under Analyze, Correlate, Bivariate</p>	<ul style="list-style-type: none"> <li>• Pearson's r Correlation</li> </ul> <p>This test can be found in SPSS under Analyze, Correlate, Bivariate</p>
More than 2 Variables – Analysis of Interdependence	<ul style="list-style-type: none"> <li>• Latent Structure Analysis</li> </ul>	<ul style="list-style-type: none"> <li>• Non-Metric Multidimensional Scaling (MDS)</li> </ul>	<ul style="list-style-type: none"> <li>• Factor Analysis* This test can be found in SPSS under Analyze, Data Reduction</li> <li>• Cluster Analysis* This test can be found in SPSS under Analyze, Classify, K-Means Cluster</li> </ul>

\* **Factor Analysis is used when you are trying to assess the relationship between individual variables and group variables that are similar; Cluster Analysis is used when you are trying to assess the relationship between individual variables and group (or segment) individual respondents based on their similarity of response.**

## Goal of the Analysis: Assessing the Impact of One Variable on Another

1. Determine the level of measurement of the dependent (or criterion) variable (remember the dependent variable is what you are trying to predict, e.g. your starting salary when you graduate). This will be the column you will use.
2. Determine the level of measurement of the independent (or predictor) variable(s) (these are the variables you believe have an impact on the dependent variable, e.g., your GPA, number of internships, etc. may predict your starting salary when you graduate). This will be the row you will use.

## Assessing the Impact of One Variable on Another Analytical Test Selection

		Level of Measurement of Dependent or Criterion Variable		
		Nominal	Ordinal	Interval or Ratio
Level of Measurement of Independent or Predictor Variable	Nominal	<ul style="list-style-type: none"> <li>• CHAID</li> </ul>	<ul style="list-style-type: none"> <li>• MONANOVA</li> </ul>	<ul style="list-style-type: none"> <li>• MCA</li> <li>• ANOVA</li> <li>• CHAID</li> </ul>
	Ordinal	<ul style="list-style-type: none"> <li>• Nonmetric Discriminant Analysis</li> </ul>	<ul style="list-style-type: none"> <li>• Guttman-Lingoes CM2 Regression</li> </ul>	<ul style="list-style-type: none"> <li>• MCA</li> <li>• CHAID</li> </ul>
	Interval or Ratio	<ul style="list-style-type: none"> <li>• Discriminant Analysis This test can be found in SPSS under Analyze, Classify</li> <li>• Logistic Regression* This test can be found in SPSS under Analyze, Regression, Binary Logistic</li> </ul>	<ul style="list-style-type: none"> <li>• Ordinal Regression</li> </ul>	<ul style="list-style-type: none"> <li>• Simple Regression</li> <li>• Multiple Regression</li> </ul> <p>The test for both these analyses can be found in SPSS under Analyze, Regression, Linear</p>

\* Use Logistic Regression when you are trying to predict a dichotomous outcome (e.g. will the customer purchase our new product or not).