

Analyzing Survey Data with Minitab

Frequency Distributions, Cross Tabulation and Hypothesis Testing

Minitab Statistical Software makes it easy to analyze survey data you've collected and answer questions that can affect your business or organization.

Surveys are an important tool for market research. Businesses use them to systematically and objectively gather information from respondents to discover what people want and identify market needs. Surveys also have many applications beyond market research, and students in statistics (and many other types of classes) can learn a great deal from gathering and analyzing survey data.

This article highlights several basic tools in Minitab that will help you interpret your survey data accurately.

Gathering and Preparing Survey Data

You can use Minitab's Power and Sample Size tools to make sure you survey enough people to conduct a reliable analysis, while avoiding wasting resources by collecting more data than you need. Minitab's Power and Sample Size tools help you balance your need for statistical power with the expense of gathering data by answering this question: How much data do you need? For example, Minitab can quickly tell you how many people should you survey to be 95% confident that the proportion of people supporting a candidate is within 5% of its true value.

You also can use Minitab to ensure that you are selecting a truly random sample of participants. Let's say you want to survey 100 households in your community. By hand-picking them yourself, you may introduce bias into your results even though you are trying to pick households at random. To make sure you *really* have a random sample, you can place a list of all households into Minitab, then use **Calc > Random Data > Sample from Columns** to select 100 at random.

Minitab also can help you prepare your data for analysis. For example, one of your survey questions asked people to rank a product on a 7-point scale, and you want to classify responses 6 and 7 as positive, 3 to 5 as neutral, and 1 and 2 as negative. You can use **Data > Code > Numeric to Text** to assign each response to the appropriate category. The Data menu also contains a wide array of tools that you can use to sort, clean, and otherwise prepare your raw data for analysis.

Frequency Distribution

Now that you've gathered and prepared your survey data, what next? A good first step in your analysis is to conveniently summarize the data by counting the responses for each level of a given variable. These counts, or frequencies, are called the *frequency distribution* and are commonly accompanied by the percentages and cumulative percentages as well.

A frequency distribution can quickly reveal:

- the number of nonresponses or missing values
- outliers and extreme values
- the central tendency, variability and shape of the distribution.

Suppose a pet adoption and rescue agency wants to find out whether dogs or cats are more popular in a certain location. To answer this question, we survey a random sample of 100 local pet owners to find out if dogs are more popular than cats, or vice versa.

When the survey is complete, we create a frequency distribution with Minitab using **Stat > Tables > Tally Individual Variables**.

Tally for Discrete Variables: Pet

Pet	Count	Percent
Cat	47	47.00
Dog	53	53.00
N=	100	

The frequency distribution reveals that the percentages for both cats and dogs are nearly 50%, indicating that there may not be a strong local preference for one type of pet over another.

We also can summarize the data using descriptive statistics. This can be very helpful when looking at continuous variables that might have a broad range, such as the age when people got their first dog or cat. Statistics including the mean, median, mode, range, and standard deviation can all be computed in Minitab using **Stat > Basic Statistics > Display Descriptive Statistics**.

Cross Tabulation

A frequency distribution can tell you about a single variable, but it does not provide information about how two or more variables relate to one another. To understand the association between multiple variables, we can use cross tabulation.

Let's say we want to see if a gender preference exists for dogs versus cats. Are men more likely to want a dog than a cat compared to women, or vice versa?

To summarize data from both variables at the same time, we need to construct a cross-tabulation table, also known as a *contingency table*. This table lets us evaluate the counts and percents, just like a frequency distribution. But while a frequency distribution provides information for each level of one variable, cross tabulation shows results for all *level combinations* of *both* variables.

In Minitab, we can generate this table using **Stat > Tables > Cross Tabulation and Chi-Square**.

Tabulated statistics: Pet, Gender

Rows: Pet Columns: Gender

	Female	Male	All
Cat	35	12	47
	70	24	47
Dog	15	38	53
	30	76	53
All	50	50	100
	100	100	100

Cell Contents: Count
% of Column

This cross tabulation shows that women prefer cats 70 to 30%, and men prefer dogs 76 to 24%. Based on these percentages, we can conclude that males are more likely to own a dog while females are more likely to own a cat.

But what if there is third variable to consider, such as marital status? We could then create a similar cross tabulation, but break it down into two tables: one for married people, and another for those who are single.

Hypothesis Testing

Frequency distributions and cross tabulation are great starting points for survey analysis, but they may not be sufficient for a comprehensive analysis.

To get a fuller understanding of your data, we need to include hypothesis testing. For our pet survey, we want to make sure the difference we see between gender and pet preference is due to a true association, and not random chance. A hypothesis test can tell us if the difference we see in the percentages is statistically significant, and whether the pet preference and gender variables are independent or not.

To evaluate the statistical significance of cross tabulation results, we use a hypothesis test called the chi-square test. This test compares the counts observed in the data we've collected to the counts we would expect if there is no relationship between the variables.

We run this test in Minitab using **Stat > Tables > Cross Tabulation and Chi-Square**.

```
Pearson Chi-Square = 21.236, DF = 1, P-Value = 0.000  
Likelihood Ratio Chi-Square = 22.075, DF = 1, P-Value = 0.000  
Fisher's exact test: P-Value = 0.0000074
```

Minitab makes it easy to evaluate several variations of the Chi-Square test. In this analysis, we displayed the Pearson, likelihood ratio and Fisher's exact test. The p-value computed for each was less than $\alpha=0.05$, which means the difference is statistically significant. Therefore, we can reject the null hypothesis that the variables are independent and conclude that a statistically significant relationship exists between gender and pet preference.

Based on these results, we would focus on marketing a local dog rescue drive to male residents and a local cat rescue in a way that appeals to female residents.

A chi-square test was the tool we needed in this case, but there are other hypothesis tests commonly used for survey data, including t-tests and proportion tests. These types of tests can be used to compare averages or proportions to a target value, or to compare averages or proportions to each other.

These hypothesis tests can answer questions such as:

- Is one brand of cola preferred over its competitor?
- Are at least 85% of all visitors to an e-commerce site satisfied with their purchase?
- Is there a difference in the average rating given to a cell phone company by teenagers compared to parents?

Most hypothesis tests in Minitab are located in the **Stat > Basic Statistics** menu although some, like the chi-square test, are located elsewhere in the software.

Delving Deeper into Survey Data

From frequency distributions to cross tabulation to hypothesis testing to more sophisticated types of analysis, Minitab has what you need to analyze survey data and make sound conclusions about markets, customers, or whatever you're trying to assess. For more information and additional examples detailing how to use these and other useful tools, Minitab offers an extensive Help system and free Technical Support.

Michelle Paret
Product Marketing Manager, Minitab Inc.

Eston Martz
Senior Creative Services Specialist, Minitab Inc.

Visit www.minitab.com for more information about statistics.