What is a Contingency Table?

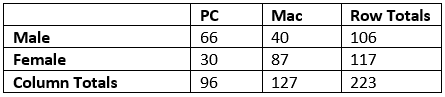
A contingency table displays frequencies for combinations of two [categorical variables](https://statisticsbyjim.com/glossary/categorical-variables/). Analysts also refer to contingency tables as crosstabulation and [two-way tables](https://statisticsbyjim.com/basics/two-way-table/).

Contingency tables classify outcomes for one variable in rows and the other in columns. The values at the row and column intersections are frequencies for each unique combination of the two variables.

Use contingency tables to understand the relationship between categorical variables. For example, is there a relationship between gender (male/female) and type of computer (Mac/PC)?

Example Contingency Table

The contingency table example below displays computer sales at our fictional store. Specifically, it describes sales frequencies by the customer’s gender and the type of computer purchased. It is a two-way table (2 X 2). I cover the naming conventions at the end.



In this contingency table, columns represent computer types and rows represent genders. Cell values are frequencies for each combination of gender and computer type. Totals are in the margins. Notice the grand total in the bottom-right margin.

At a glance, it’s easy to see how two-way tables both organize your data and paint a picture of the results. You can easily see the frequencies for all possible subset combinations along with totals for males, females, PCs, and Macs.

For example, 66 males bought PCs while females bought 87 Macs. Furthermore, there are 117 females, 106 males, 96 PC sales, 127 Mac sales, and a grand total of 223 observations in the study.

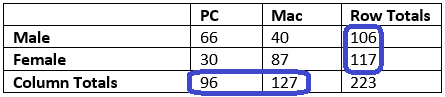
Marginal and Conditional Distributions in Contingency Tables

Contingency tables are a fantastic way of finding marginal and conditional distributions. These two distributions are types of frequency distributions. Learn more about [Frequency Tables: How to Make and Interpret](https://statisticsbyjim.com/basics/frequency-table/).

Marginal Distribution

These distributions represent the frequency distribution of one categorical variable without regard for other variables. Unsurprisingly, you can find these distributions in the *margins* of a contingency table.

The following marginal distribution examples correspond to the blue highlights.



Advertisement

For example, the marginal distribution of gender without considering computer type is the following:

* **Males**: 106
* **Females**: 117

Alternatively, the marginal distribution of computer types is the following:

* **PC**: 96
* **Mac**: 127

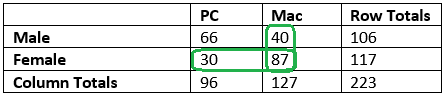
Learn more about [Marginal Distributions](https://statisticsbyjim.com/basics/marginal-distribution/).

Advertisement

Conditional Distribution

For these distributions, you specify the value for one of the variables in the contingency table and then assess the distribution of frequencies for the other variable. In other words, you *condition* the frequency distribution for one variable by setting a value of the other variable. That might sound complicated, but it’s easy using a contingency table. Just look across one row or down one column.

The following conditional distribution examples correspond to the green highlights.



Advertisement

For example, the conditional distribution of computer type for females is the following:

* **PC**: 30
* **Mac**: 87

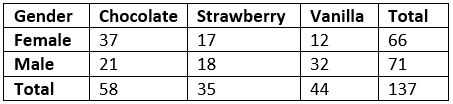
Alternatively, the conditional distribution of gender for Macs is the following:

* **Males**: 40
* **Females**: 87

Learn more about [Conditional Distributions](https://statisticsbyjim.com/basics/conditional-distribution/).

Finding Relationships in a Contingency Table

In the contingency table below, the two categorical variables are gender and ice cream flavor preference. This is a two-way table (2 X 3) where each cell represents the number of times males and females prefer a particular ice cream flavor. The CSV datasheet shows one format you can use to enter the data into your software: [Flavor Preference](https://statisticsbyjim.com/wp-content/uploads/2019/07/flavor_preference.csv).



How do we go about identifying a relationship between gender and flavor preference?

If there is a relationship between ice cream preference and gender, we’d expect the conditional distribution of flavors in the two gender rows to differ. From the contingency table, females are more likely to prefer chocolate (37 vs. 21), while males prefer vanilla (32 vs. 12). Both genders have an equal preference for strawberry. Overall, the two-way table suggests that males and females have different ice cream preferences.

The Total column indicates the researchers surveyed 66 females and 71 males. Because we have roughly equal numbers, we can compare the raw counts directly. However, when you have unequal groups, use percentages to compare them.

Row and Column Percentages in Contingency Tables

Row and column percentages help you draw conclusions when you have unequal numbers in the margins. In the contingency table example above, more women than men prefer chocolate, but how do we know that’s not due to the [sample](https://statisticsbyjim.com/glossary/sample/) having more women? Use percentages to adjust for unequal group sizes. Percentages are relative frequencies. Learn more about [Relative Frequencies and their Distributions](https://statisticsbyjim.com/basics/relative-frequency/).

Here’s how to calculate row and column percentages in a two-way table.

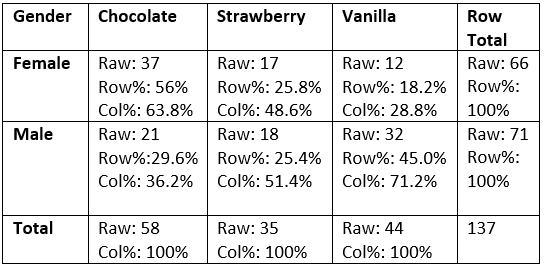
* **Row Percentage**: Take a cell value and divide by the cell’s row total.
* **Column Percentage**: Take a cell value and divide by the cell’s column total.

For example, the row percentage of females who prefer chocolate is simply the number of observations in the Female/Chocolate cell divided by the row total for women: 37 / 66 = 56%.

The column percentage for the same cell is the frequency of the Female/Chocolate cell divided by the column total for chocolate: 37 / 58 = 63.8%.

Interpreting Percentages in a Contingency Table

The contingency table below uses the same raw data as the previous table and displays both row and column percentages. Note how the row percentages sum to 100% in the right margin while the column percentages sum to 100% at the bottom.

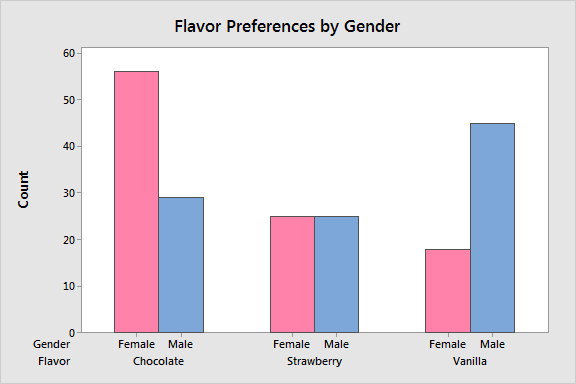


Whether you focus on row percentages or column percentages in a contingency table depends on the question you’re answering. In our case, we want to know whether flavor preference depends on gender. Because the two genders display in separate rows, we’ll look for differences in the row percentages.

56% of females prefer chocolate versus only 29.6% of males. Conversely, 45% of males prefer vanilla, while only 18.2% of females prefer it. These results reconfirm our previous findings using the raw counts.

How to Graph a Contingency Table

You can use bar charts to display a contingency table. The following clustered bar chart shows the row percentages for the previous two-way table. I’ve set the graph to cluster the female and male pairs of bars together for each flavor, making comparisons easier. I think it gives a nice oomph to the tabular results.



This bar chart reiterates our conclusions from the contingency table. Women in this sample prefer chocolate, men favor vanilla, and both genders have an equal preference for strawberry.

Learn more about [Bar Charts: Using, Examples and Interpreting](https://statisticsbyjim.com/graphs/bar-charts/).

How to Analyze a Contingency Table

We’ve already looked at various ways to analyze a contingency table. Here are two more methods that take it to another level.

Contingency tables are a fantastic way to display and find various types of probabilities. Use these tables to calculate joint, marginal, and conditional probabilities. I’ve written an article about calculating probabilities using two-way tables, and it includes all the definitions, notation, and formulas you need. Read about [Using Contingency Tables to Calculate Probabilities](https://statisticsbyjim.com/probability/contingency-tables-probabilities/).

In this post, we looked for a relationship between gender and ice cream preference by noting the differences between counts and row percentages in the contingency table. If we’re using this sample to draw inferences about the entire [population](https://statisticsbyjim.com/glossary/population/) of ice cream consumers, we’ll need to use a [hypothesis test](https://statisticsbyjim.com/glossary/hypothesis-tests/) to evaluate the relationship.

In other words, are the differences we noticed in the sample large enough to support the notion that a relationship exists in the population? Or can we chalk up the differences to [random sampling error](https://statisticsbyjim.com/hypothesis-testing/sampling-error/)? Learn how the [chi-square test of independence](https://statisticsbyjim.com/hypothesis-testing/chi-square-test-independence-example/) can help us out by analyzing contingency tables!

Naming Conventions for Contingency Tables

Contingency tables come in a variety of flavors. The key considerations for naming the types are the number of categorical variables and the number of values for each categorical variable.

Number of Categorical Variables

You must have at least two categorical variables to create a contingency table. When you have two variables, it’s a two-way table. If you have three, it’s a three-way table, and so on.

For example, suppose we run a computer store and record the sales using the two categorical variables of gender and computer type. Those variables create a two-way contingency table. If we add a third categorical variable for store location, it becomes a three-way table.

How do you present a three-way table?

Because contingency tables display in two dimensions, you need multiple tables to represent anything more than a two-way table.

If there are four store locations in our three-way example, we’ll need to use four two-way tables. Each table displays gender and computer type for one location.

Number of Rows and Columns

The rows represent values of one categorical variable, while the columns denote the values of another. Analysts indicate the number of values for each variable by describing these tables as an A X B contingency table, where A represents the number of rows and B signifies the number of columns.

For example, at our computer store, Gender has two possible values and computer type has two values (PC and Mac). Hence, we have a 2 X 2 contingency table. If we add a third type of computer as a new column, it becomes a 2 X 3 table.