![Microsoft logo]

Essential Math for Data Analysis Using Excel Online

Module 3, Lab 3: Proportions

Learning Objectives
- Calculate proportions for a categorical variable in Excel.
- Create crosstabs for two categorical variables.

Description
Learners will analyze the proportions and create crosstabs for two of the categorical variables in the coffee data set in Excel.

Data set
Mod2Lab.csv

Overview
Hopefully you're not getting tired of coffee talk yet, because in this lab, we'll create a crosstab for two of the categorical variables in the coffee data set we've been using. Using pivot tables, we'll organize these categories into raw counts and proportions in a couple different ways.

What You'll Need
To complete the lab, you will need the online version of Microsoft Excel.

Exercise 1: Coffee Crosstabs
In this exercise, you will compare the "preference" and "additions" variables from the coffee data set, both of which are categorical (i.e. non-numerical). To set up a crosstab, you'll need to use Excel's pivot table feature.

1. Open up the data set in Excel. There should be 100 different rows, with column headings for various coffee preferences.

2. Click on any cell that contains data, then go to the ribbon and click Insert > PivotTable.
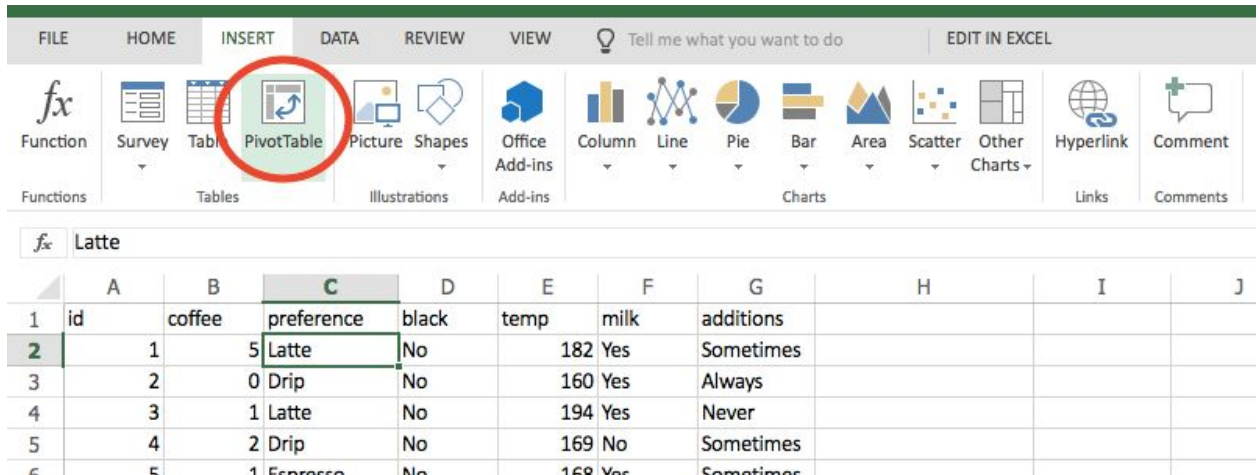
3. The Create PivotTable box will pop up.
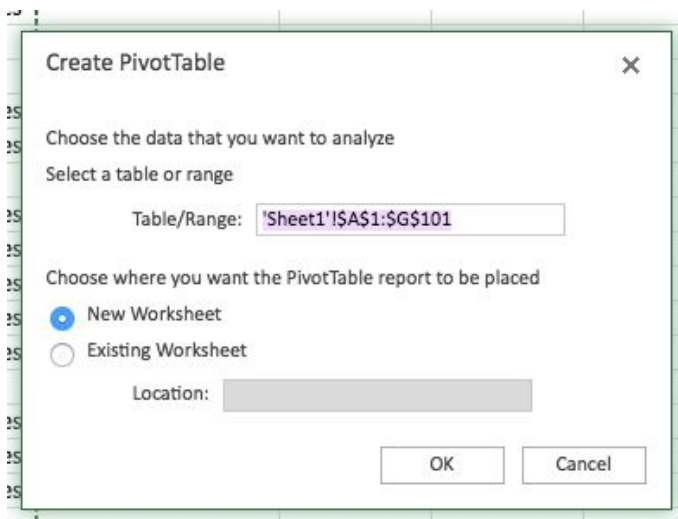
The Table/Range field should already include all the data from the table. Select "New Worksheet" and hit OK.

4. In a new worksheet, you'll see the PivotTable Fields box on the right side of the screen.

PivotTable Fields ✕

Choose fields:

- ☐ id
- ☐ coffee
- ☐ preference
- ☐ black
- ☐ temp
- ☐ milk
- ☐ additions

Drag fields between areas below:

▼ FILTERS	▥ COLUMNS

☰ ROWS	Σ VALUES

You want to compare the "preference" and "additions" variables, so drag "preference" down into the Rows box, and drag "additions" into the Columns box.

PivotTable Fields ✕

Choose fields:

- ☐ id
- ☐ coffee
- ☑ **preference**
- ☐ black
- ☐ temp
- ☐ milk
- ☑ **additions** ▾

Drag fields between areas below:

▼ FILTERS	▥ COLUMNS
	additions ▾

☰ ROWS	Σ VALUES
preference ▾	

This should create an empty pivot table on the left side of the screen that looks like this:

	A	B	C	D	E
1					
2					
3		Column Labels ▾			
4	Row Labels ▾	Always	Never	Sometimes	Grand Total
5	Drip				
6	Espresso				
7	Latte				
8	**Grand Total**				
9					
10					
11					

5. Now you want to add some actual values to your crosstab/pivot table. Back in the PivotTable Fields box on the right, drag "preference" down into the Values box.

Check out the table on the left again. It's been populated with values from your original data.

Count of preference	Column Labels			
Row Labels	Always	Never	Sometimes	Grand Total
Drip	11	2	39	52
Espresso	6	2	18	26
Latte	8	4	10	22
Grand Total	25	8	67	100

6. With this new crosstab, you can see how the "preferences" variable (Drip, Espresso, or Latte) relates to the "additions" variable (Always, Never, or Sometimes, which refers to whether people add things like sugar or flavoring to their coffee). For example, you can see that only 2 of the people who prefer drip coffee said they never add anything (cell C5).

The Grand Total column and row give the total frequency of each individual variable. For example, you can see from the far-right column that 52 people out of 100 prefer drip coffee to anything else. The pivot table gives you both the crosstab *and* the original data counts for each variable.

7. Use the pivot table to find proportions. Excel can give you proportions/percentages in a few different ways here. Head back over to the PivotTable Fields box, click on Count of Preferences in the Values section, and select Value Field Settings from the dropdown menu.
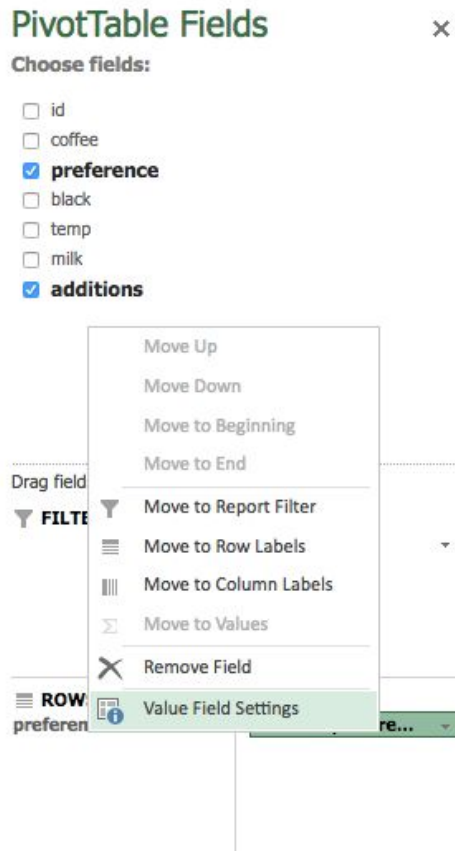
8. In the Value Field Settings box that pops up, click over to Show Value As.

Now click the field that says No Calculation. The dropdown menu gives you all kinds of options.

9. Click on % of Row Total, and the values in the crosstab/pivot table will change to percentages.

	A	B	C	D	E
1					
2					
3	Count of preference	Column Labels ▼			
4	Row Labels ▼	Always	Never	Sometimes	Grand Total
5	Drip	21.15%	3.85%	75.00%	100.00%
6	Espresso	23.08%	7.69%	69.23%	100.00%
7	Latte	36.36%	18.18%	45.45%	100.00%
8	Grand Total	25.00%	8.00%	67.00%	100.00%
9					
10					

Now you have an even handier way of cross-tabulating these two variables. The cells show each "addition" category as a percentage of each "preference" category. For instance, the highlighted cell above shows that 21.15% of drip coffee drinkers said they always add stuff to their drip coffee. Similarly, 3.85% of drip coffee drinkers said they never do, and 75% said they sometimes do. Those are your proportions.

10. You can also flip this idea around. Go back to the PivotTable Field section and click on Count of Preferences in the Values section again. Select Value Field Settings from the dropdown menu again.

This time, in the Show Value As section, try choosing % of Column Total instead.
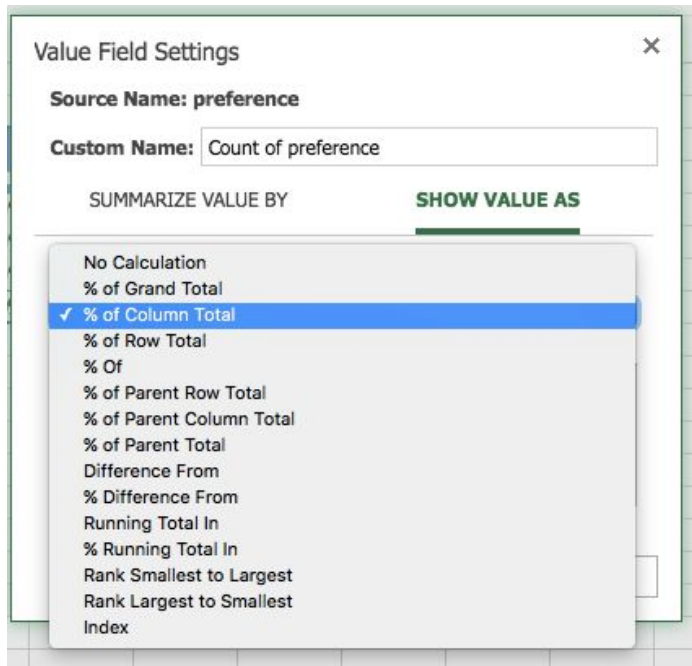
11. The pivot table now displays percentages in a different dimension.

This time, the cells show each "preference" category as a percentage of each "addition" category. For example, the highlighted cell shows that 44% of the people who always add stuff to their coffee said they prefer drip coffee. 24% of the "Always" folks prefer espresso, and 32% of the "Always" folks prefer lattes.