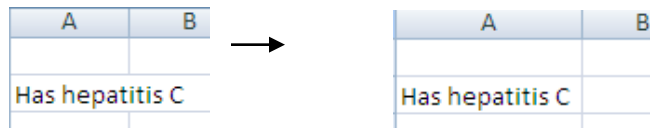


Excel: Two-Way Tables

A study by the University of Texas Southwestern Medical Center examined 626 people to see if an increased risk of contracting hepatitis C was associated with having a tattoo. If the subject had a tattoo, researchers asked whether it had been done in a commercial tattoo parlor or elsewhere. The following two-way table shows the results of the study.

	Tattoo done in		
	commercial parlor	Tattoo done elsewhere	No tattoo
Has hepatitis C	17	8	18
No hepatitis C	35	53	495

1. Create this table yourself in a new *Excel* spreadsheet. In order to have the text wrap around in a cell as shown in the top row, enter the text and then choose **Wrap Text** in the **Alignment** group of the **Home** menu. (If you cannot find it, enter the words “wrap text” in the “Tell me what you want to do” search bar.) Also, you can widen a column by clicking and dragging the top of that column. For example, if you want the words “Has hepatitis C” (shown below) to fit in column A without running into column B, simply click on the line separating the **A** and the **B** at the top of the columns and drag to the right until column A is wide enough to fit your text.



2. Add a Total row at the bottom and a Total column at the right. *Excel* can calculate the totals for you. Starting in cell B4, either enter **=SUM(B2:B3)** or use **AutoSum** in the **Editing** group of the **Home** menu.

	A	B	C	D	E
		Tattoo done in commercial parlor	Tattoo done elsewhere	No tattoo	Total
1					
2	Has hepatitis C	17	8	18	
3	No hepatitis C	35	53	495	
4	Totals	=SUM(B2:B3)			
5		SUM(number1, [number2], ...)			
6					

Continue to find the totals for columns C and D. Next find the row totals. When you have *Excel* calculate the sum of row 2 (which you will want to have in cell E2), the proper command should be **=SUM(B2:D2)** (or use **AutoSum**). Complete your table so that you have the totals for each column and row, including the table total in cell E4 (which should be 626, the total number of people included in the study).

Helpful tip: If you ever need to edit what has already been entered in a cell, it must be done in what is called the “formula bar” located just above the column headings of your spreadsheet just to the right of . For example, there is an error in the formula below; the sum should include D2. To correct the error, you would click in the

formula bar where you see “=SUM(B2:C2)” and change the C to a D. Hit Enter, and the correct sum, 43, will then appear in cell E2.

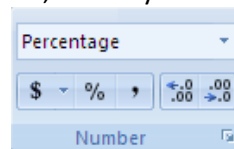
	A	B	C	D	E
1		Tattoo done in commercial parlor	Tattoo done elsewhere	No tattoo	Total
2	Has hepatitis C	17	8	18	25
3	No hepatitis C	35	53	495	

- Now we are ready to investigate the relationship between tattoos and hepatitis C by looking at some *conditional distributions*. Start by copying and pasting the original table (without the totals) so that you have a fresh one that you can work with while keeping the one with the totals unchanged for reference. Do this by highlighting cells A1 through D3 and copying them to cells A6 through D8. Now we will calculate the conditional distribution of tattoo for those with hepatitis C. That means dividing each entry in the “Has hepatitis C” row by 43, the sum of that row. Start by replacing the 17 in cell B7 with =17/43, as shown below. (Alternatively, enter =B2/43 in order to copy the formula in the next steps, as shown in class.)

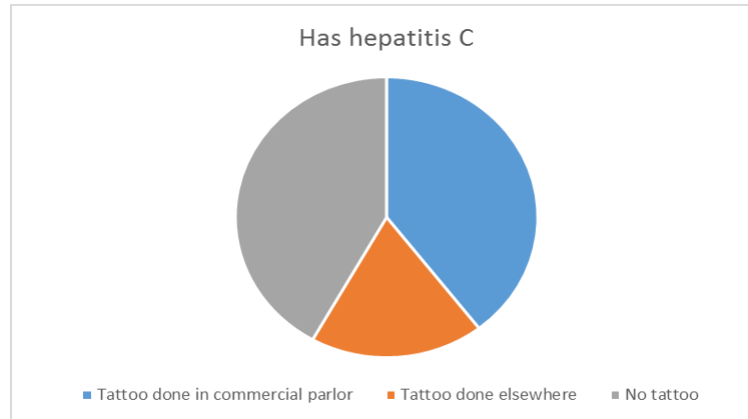
	Tattoo done in commercial parlor	Tattoo done elsewhere	No tattoo
Has hepatitis C	=17/43	8	18
No hepatitis C	35	53	495

	Tattoo done in commercial parlor	Tattoo done elsewhere	No tattoo
Has hepatitis C	0.39534884	8	18
No hepatitis C	35	53	495

That’s a lot of decimal places! Besides, we want to see a percentage, not a proportion. Click back on 0.39534884 and then in the **Number** group of the **Home** menu, click on the % so that “General” changes to “Percentage” as shown below. You should now see 40% in cell B7. Use the buttons with the 0’s and arrows (shown in the bottom right of the image below, but may be below the \$ and %) so that you have 39.53%.

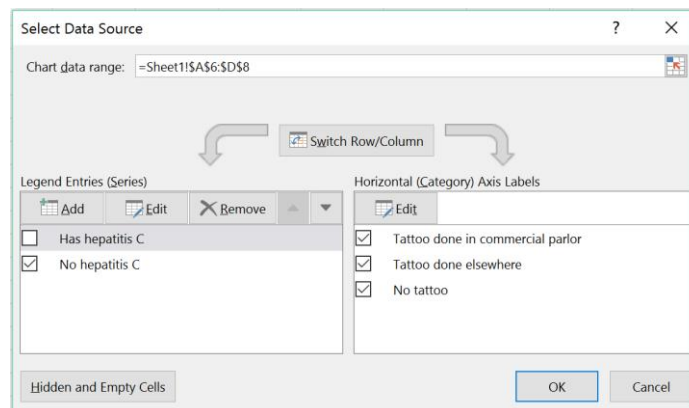


- Now complete the conditional distributions of tattoo for people with hepatitis C as well as for those without. (Of course, for “no hepatitis C” you will divide each number by the total for THAT row.) Make sure you change all proportions to percentages with an extra two decimal places. *The percentages in each row should add up to 100%.* (But if you still have the Total row/column, delete them so that they do not appear in the next graph.)
- Next we will make pie charts for these two conditional distributions. Click on any one of the cells in your table and then go to the **Insert** menu. In **Charts**, you will see an image of a pie chart. Click on it and then choose the simplest 2-D option at the top left. You should see the following graph.



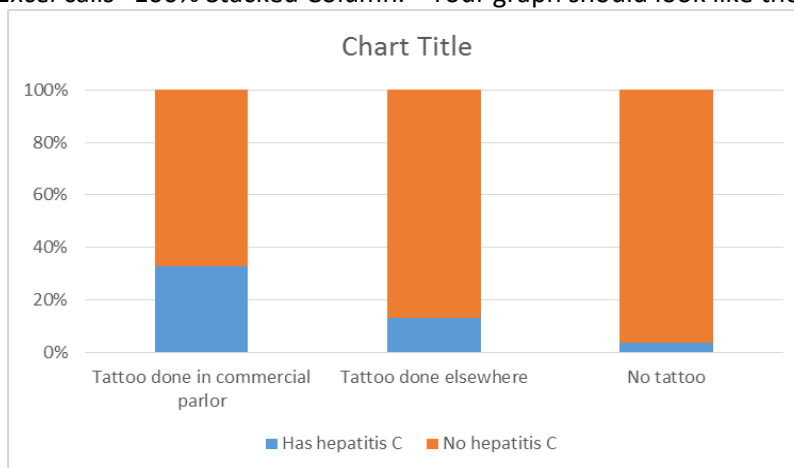
Now we would like it to show the percentages on each piece of the pie. Easy. Click on the graph and you will see some new menu options under **Chart Tools**. In **Design**, you will see **Chart Styles**. Choose whichever one of the options that include % that you like best.

- Good, but what about the conditional distribution for the people who do not have hepatitis C? We need to make another pie chart, this time making sure that we get the data for the “No hepatitis C” row. One way to do this is to highlight just the row with the tattoo information (row 6) and the row with “No hepatitis C” (row 8). If you are having trouble simultaneously highlighting two separated rows, or if you still have trouble getting the graph, then follow these alternative instructions: Make a duplicate copy of your pie chart. Now click on one of the graphs to get the **Chart Tools** menu, then click on **Select Data** in the **Design** group. In the window that appears (see below), **uncheck** the box for “Has hepatitis C” on the left, leaving just “No hepatitis C” checked. Click **OK**. The new pie chart should now display the percentages for “No hepatitis C.”



- Before moving on, it’s a good idea to start paying attention to how you will organize your output so that it will fit nicely on one page for printing. In the **File** menu, go to **Print** so that you can see the print preview. You will probably already see that one of your pie charts is falling off the page. When you exit out of **Print**, you should now see dashed vertical and horizontal lines in your spreadsheet; these are the page breaks. Either move one graph so that it is below the other, resize each graph to make them smaller, or change your page orientation to landscape (found in the **Layout** menu and also in the **Print** window).
- Now write a sentence or two describing how the two conditional distributions differ. Discuss the pie charts in the context of tattoos and hepatitis C. Make sure that everything (two tables, two pie charts, and your paragraph) fits nicely within the page boundaries.

9. Okay. These pie charts showed us some interesting information about whether or not people *with hepatitis C* have tattoos and whether or not people *without hepatitis C* have tattoos. But what percentage of people *with tattoos* has hepatitis C? And do people who had their tattoo done in a commercial parlor have a lower or higher hepatitis C rate than those with tattoos done elsewhere? These questions are not answered by our pie charts because they deal with different conditional distributions. Now we need the percentages calculated by column rather than by row. Make another copy of the original two-way table (without totals) and put it in a new worksheet.
10. Repeat step 3, but this time find the conditional distributions by column. For instance, start by calculating the conditional distribution of hepatitis C for people with tattoos done in commercial parlors by dividing 17 by 52 and 35 by 52. Continue with the remaining columns, each time dividing by the sum of that column. As before, convert the proportions to percentages. Make sure that the percentages in each *column* add up to 100%. (If you still have the Total row/column, delete them so that they do not appear in the next graph.)
11. This time we will make something called a segmented bar graph. Click anywhere in your newest data table, go to the **Insert** menu, and choose what looks like a vertical bar graph in **Charts**. Choose the third option in 2-D Column, which *Excel* calls “100% Stacked Column.” Your graph should look like the following:



Either change “Chart Title” to something appropriate or you may remove it by clicking on it and then hitting the Delete key on your keyboard.

12. Now use this graph to write a brief description of the association between tattooing and hepatitis C. Which group has the highest incidence of hepatitis C? The lowest? Does it seem to matter where the tattoo is done?
13. Finally, print out your two pages of work. Before printing each page, make sure everything is formatted nicely on each page. If you see “Page 1 of 2” at the bottom of the print preview window, then some of your work is spilling onto a second page. Close out of the print window and rearrange your work to fit within the boundaries of a single page by moving and/or resizing graphs and text boxes.

For more information on this actual study, see <https://www.ncbi.nlm.nih.gov/pubmed/11307589>