

Section 3B – Marginal and Joint Percentages from Contingency Tables

Analyzing two categorical data sets involves not only creating contingency tables, bar charts and pie charts, but also being able to find and analyze proportions and percentages.

Remember that a proportion is found by taking the amount (frequency) and dividing by the total (sample size).

$$\text{Proportion} = \frac{\text{Amount (Frequency)}}{\text{Total}}$$

To convert that proportion into a percentage, simply multiply the proportion by 100%.

Marginal Percentages

Let us start with looking at basic marginal proportions. These are proportions where the amount involves only a single variable and the total is everyone in the data (grand total).

Look at the following contingency table created with StatKey from the Fall 2015 Math 075 Survey data. This table describes the relationship between smoking and political party for Math 075 pre-stat students.

Counts Table [Switch Variables](#)

Smoke cigarettes? \ Political Party	Democratic	Other	Republican	Independent	Total
No	176	82	90	78	426
Yes	9	10	7	7	33
Total	185	92	97	85	459

Remember, analyzing data involves asking questions and finding the answers to those questions.

For example. Here are a few questions that came to mind when I looked at this table.

Example 1

What percentage of the pre-stat students smoke cigarettes?

Notice we are looking at all of the students (not just democrats), so we should use the grand total as our total. Where do we find the amount of pre-stat students that smoke cigarettes? Smoking cigarettes (yes) is a row, so we should look in the margin at the total for that row.

Counts Table [Switch Variables](#)

Smoke cigarettes? \ Political Party	Democratic	Other	Republican	Independent	Total
No	176	82	90	78	426
Yes	9	10	7	7	33
Total	185	92	97	85	459

Notice the amount and the grand total are found in the margins where the totals are. This is why this is often called a “marginal proportion” or a “marginal percentage”. Notice the marginal percentage only involves one variable (smoking) and does not include political party.



Proportion of students that smoke = Amount of Smokers \div Grand Total = $33 \div 459 = 0.07189524 \approx 0.072$

Percentage of students that smoke $\approx 0.072 \times 100\% = 7.2\%$

Example 2

What percentage of the pre-stat students identified as other political party?

Notice we are looking at all of the pre-stat students, so we should use the grand total again as our total.

Where will we find the amount of pre-stat students that support “other” political party? Other political party is a column so we will have to look at the total for that column.

Counts Table [Switch Variables](#)

Smoke cigarettes? \ Political Party	Democratic	Other	Republican	Independent	Total
No	176	82	90	78	426
Yes	9	10	7	7	33
Total	185	92	97	85	459

Proportion of students that support other political party = Amount of other political party \div Grand Total = $92 \div 459 = 0.200435729 \approx 0.200$

Percentage of students that are other political party $\approx 0.200 \times 100\% = 20.0\%$

Notice we only looked at one variable (other political party), and the amount of students that identified as other political party and the grand total were both found in the margins. So this is again a “marginal percentage”.

Note: Some students may ask why we did not write the answer as 0.2 or 20%. These are equivalent to 0.200 and 20.0%, but these answers tell us that the answer was rounded to three significant figures.

Formula

Single Variable Marginal Proportion = Total for Row or Column \div Grand Total

Joint Percentages

Sometimes we want to find a proportion or percentages where the amount (frequency) involves more than one variable. These are often called “joint proportions” or “joint percentages”.

There are two types of joint proportions.

AND: This is when we want to know the proportion or percentage involving two things being true about a person or object.

OR: This is when we want to know the proportion of percentage involving either one variable or another variable being true about the person or object.

Let us look at the political party and cigarette data again.

Example 3 (“AND Joint %”)

What percentage of all the pre-stat students both smoked cigarettes and were Republican?



Notice there are two variables involved, republican and smoking. The key though is that we want the proportion for both things being true about the person. We cannot look at only smokers and we cannot look at only republicans. We need the amount of smoking republicans. This is a classic “AND” proportion since both things need to be true about the student.

Notice also we are picking from all pre-stat students, so our total should be the grand total again.

Counts Table [Switch Variables](#)

Smoke cigarettes? \ Political Party	Democratic	Other	Republican	Independent	Total
No	176	82	90	78	426
Yes	9	10	7	7	33
Total	185	92	97	85	459

Notice that to find the smoking republicans, we need to look where the republican column meets the yes smoking row. This is why “AND” proportions are often referred to as an intersection.

Proportion of pre-stat students that both smoke cigarettes and are republican =

$$\text{amount of smoking republicans} \div \text{grand total} = 7 \div 459 \approx 0.015250544 \approx 0.015$$

Percentage of pre-stat students that both smoke and are republican $\approx 0.015 \times 100\% \approx 1.5\%$

Formula

“AND” Intersecting Proportion = Amount where row and column intersect \div Grand Total

Example 4 (“OR” Joint %)

Suppose we only wanted to know the percentage of students that either smoke or are republican. (Not both)

This would be a classic “OR” joint proportion. The key is that we will now need to include everyone that smokes, as well as everyone that is republican. This is why an OR joint proportions are often referred to as a union. When calculating an “OR” joint proportion, you will need to do some adding to find the amount.

Counts Table [Switch Variables](#)

Smoke cigarettes? \ Political Party	Democratic	Other	Republican	Independent	Total
No	176	82	90	78	426
Yes	9	10	7	7	33
Total	185	92	97	85	459

Proportion of students that either smoke or are republican =

amount of students that either smoke or are republican / grand total

$$= (90 + 9 + 10 + 7 + 7) \div 459 = 123 \div 459 \approx 0.267973856 \approx 0.268$$

Percentage of students that either smoke or are republican $\approx 0.268 \times 100\% \approx 26.8\%$



Important Note: Notice that we did not use the row and column totals when calculating an “OR” joint proportion. If we added the total for smokers (33) plus the total for republicans (97), we would have gotten 130 as our amount. This would be wrong. The correct amount was 123. Adding the row and column totals gives you the wrong answer because we would have added the 7 smoking republicans twice.

Here are some other formulas that may be used to calculate an OR (union) proportion.

Formulas

“OR” Union Proportion = Add up all of the values in the row or column without using totals ÷ Grand Total

“OR” Union Proportion = (Row Total + Column Total – Intersection amount) ÷ Grand Total

“OR” Union Proportion = 1st Variable Proportion + 2nd Variable Proportion – Intersecting “AND” Proportion

In the previous example here is how we could have used the other formulas to get the same answer.

What proportion of the pre-stat students either smoke cigarettes or are republican?

Counts Table Switch Variables

Smoke cigarettes? \ Political Party	Democratic	Other	Republican	Independent	Total
No	176	82	90	78	426
Yes	9	10	7	7	33
Total	185	92	97	85	459

“OR” Union Proportion = (Row Total + Column Total – Intersection amount) ÷ Grand Total

$$= (97 + 33 - 7) \div 459 = 123 \div 459 = 123 \div 459 \approx 0.267973856 \approx 0.268$$

Percentage of students that either smoke or are republican $\approx 0.268 \times 100\% \approx 26.8\%$

Notice we got the same answer as before.

“OR” Union Proportion = 1st Variable Proportion + 2nd Variable Proportion – Intersecting “AND” Proportion

= Proportion Smoke + Proportion Republican – Proportion that smoke and are Republican

$$= \frac{33}{459} + \frac{97}{459} - \frac{7}{459} \approx 0.072 + 0.211 - 0.015 = 0.268 = 26.8\%$$

Notice we got the same answer as before. This formula is particularly useful, especially when a statistics program calculates the marginal and intersecting proportions for you.

Calculating Marginal and Joint Proportions with StatKey

StatKey can calculate the marginal and intersecting proportions for you. Under the “Counts table” (Contingency Table) you will see a “Proportions” menu. Click the button that says “Overall”. We put the smoking and political party columns from the Math 075 Summary Data Fall 2015 into StatKey.



Counts Table Switch Variables

Smoke cigarettes? \ Political Party	Democratic	Other	Republican	Independent	Total
No	176	82	90	78	426
Yes	9	10	7	7	33
Total	185	92	97	85	459

Proportions Row Column Overall

Smoke cigarettes? \ Political Party	Democratic	Other	Republican	Independent	Total
No	0.383	0.179	0.196	0.17	0.928
Yes	0.02	0.022	0.015	0.015	0.072
Total	0.403	0.2	0.211	0.185	1

When you click the Overall button, StatKey calculates the marginal proportions and the “AND” intersecting proportions. However, it does not calculate the “OR” union proportions.

Our first example in this section asked what percentage of the pre-stat students smoke cigarettes. Yes (smoking) is a row in this table so we just need to look at the margin (end of the row) to find the answer. Notice it says) 0.072 or 7.2%. This is the same answer we calculated earlier in the section.

Counts Table Switch Variables

Smoke cigarettes? \ Political Party	Democratic	Other	Republican	Independent	Total
No	176	82	90	78	426
Yes	9	10	7	7	33
Total	185	92	97	85	459

Proportions Row Column Overall

Smoke cigarettes? \ Political Party	Democratic	Other	Republican	Independent	Total
No	0.383	0.179	0.196	0.17	0.928
Yes	0.02	0.022	0.015	0.015	0.072
Total	0.403	0.2	0.211	0.185	1

Earlier in this section we asked what percentage of pre-stat students both smoke cigarettes and are republican. To find this answer we just need to go to where Yes (smoking) and Republican intersect. Notice the answer is given as 0.015 or 1.5%. This is again the same answer we calculated earlier in the section.



Counts Table Switch Variables

Smoke cigarettes? \ Political Party	Democratic	Other	Republican	Independent	Total
No	176	82	90	78	426
Yes	9	10	7	7	33
Total	185	92	97	85	459

Proportions Row Column Overall

Smoke cigarettes? \ Political Party	Democratic	Other	Republican	Independent	Total
No	0.383	0.179	0.196	0.17	0.928
Yes	0.02	0.022	0.015	0.015	0.072
Total	0.403	0.2	0.211	0.185	1

Earlier in the section we wanted to find out what percentage of pre-stat students either smoke cigarettes or are republican. StatKey does not calculate “OR” (union) proportions, but we can use the proportions calculated and the following formula.

“OR” Union Proportion = 1st Variable Proportion + 2nd Variable Proportion – Intersecting “AND” Proportion

= Proportion Smoke + Proportion Republican – Proportion that smoke and are Republican

Counts Table Switch Variables

Smoke cigarettes? \ Political Party	Democratic	Other	Republican	Independent	Total
No	176	82	90	78	426
Yes	9	10	7	7	33
Total	185	92	97	85	459

Proportions Row Column Overall

Smoke cigarettes? \ Political Party	Democratic	Other	Republican	Independent	Total
No	0.383	0.179	0.196	0.17	0.928
Yes	0.02	0.022	0.015	0.015	0.072
Total	0.403	0.2	0.211	0.185	1

The proportion that smoke will be at the end of the Yes (smoke) row. The proportion of republicans will be at the bottom of the “Republican” column. The AND proportion will be where the smoking row and republican column intersect. Notice we got the same answer as before.

$$= 0.072 + 0.211 - 0.015 = 0.268 = 26.8\%$$

Note: Categorical data is often given to a data scientist as a contingency table with summary counts. Most data scientists do not calculate things by hand. Recall that in section 3A, we learned we can type in an existing contingency table into StatKey using commas. Typing the table into StatKey allows us to not only have access to the stacked bar chart, but also the proportion button that can calculate proportions automatically for us.



Formulas

Single Variable Marginal Proportion = Total for Row or Column \div Grand Total

“AND” Intersecting Proportion = Amount where row and column intersect \div Grand Total

“OR” Union Proportion = Add up all of the values in the row or column without using totals \div Grand Total

“OR” Union Proportion = (Row Total + Column Total – Intersection amount) \div Grand Total

“OR” Union Proportion = 1st Variable Proportion + 2nd Variable Proportion – Intersecting “AND” Proportion



This material is from *Introduction to Data Analysis*, 2nd edition by Matt Teachout, College of the Canyons, Santa Clarita, CA, USA, and is licensed under a “CC-BY” [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/) – 3/17/2021