

## Section 3C – Conditional Percentages from Contingency Tables and Categorical Relationships

### Conditional Proportions and Percentages

Conditional proportions and percentages are the key to understanding categorical relationships. A condition is thought of as prior knowledge about the person or situation that may change the percentage. Let us say that the Los Angeles Lakers have a 75% chance of beating the Phoenix Suns. If the Lakers best player LeBron James does not play, will that change the percentage? Of course. Knowing that LeBron James will not play is called a condition.

In contingency tables, a condition involves restricting to one particular group before you calculate the percentage.

#### Example:

What percentage of the Canyon Country campus Math 075 pre-stat students prefer Twitter as their favorite social media?

First notice that this is not a joint proportion. It does NOT ask for the percentage of all students both prefer Twitter and go to the Canyon Country campus.

The key is to identify which group we are restricting ourselves to. In other words, what is the condition? Look for words that say “if” or “given this is true” or “out of”. This designates the condition. In this example, notice that the problem said “of the Canyon Country students”. That means that we are supposed to only look at the Canyon Country students when we find our amount (frequency) and total. A commonly used method for calculating conditional percentages from a contingency table is to circle the row or column that has your condition (Canyon Country). Then only use numbers in that row or column.

#### Counts Table Switch Variables

Campus \ Social Media favorite?	Twitter	Instagram	Facebook	Snapchat	Other	Total
Canyon Country Campus	35	52	43	36	14	180
Valencia Campus	56	85	56	58	24	279
Total	91	137	99	94	38	459

Notice that the Canyon Country Campus counts are in the first row. So we should highlight or circle the first row and only use numbers in the first row when we calculate. We should not use the grand total anymore. We need the total number of students that attend the Canyon Country campus. In other words, the total from our condition. The amount will be the number of students that prefer Twitter in the Canyon Country row. In other words the intersection cell frequency.

#### Counts Table Switch Variables

Campus \ Social Media favorite?	Twitter	Instagram	Facebook	Snapchat	Other	Total
Canyon Country Campus	35	52	43	36	14	180
Valencia Campus	56	85	56	58	24	279
Total	91	137	99	94	38	459

$$\text{Conditional Proportion} = \frac{\text{Amount in Intersection Cell (Canyon Country meets Twitter)}}{\text{Row or Column Total (Row total Canyon Country)}} = \frac{35}{180} \approx 0.19444444 \approx 0.194$$

$$\text{Conditional Percentage} = \text{Conditional Proportion} \times 100\% = 0.194 \times 100\% = 19.4\%$$



This material is from *Introduction to Data Analysis*, 2<sup>nd</sup> 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

So 19.4% of the Canyon Country pre-stat students prefer Twitter as their favorite social media.

We can also have StatKey calculate conditional proportions for us by using the “row” and “column” proportion buttons. We need to ask ourselves if the condition is a row or a column? In the last question we were restricting ourselves to only Canyon Country students. This is a row. Since the condition is a row, we should click the “row” proportion button. If the condition had been a column, we would have clicked on the “Column” proportion button.

### Counts Table Switch Variables

Campus \ Social Media favorite?	Twitter	Instagram	Facebook	Snapchat	Other	Total
Canyon Country Campus	35	52	43	36	14	180
Valencia Campus	56	85	56	58	24	279
Total	91	137	99	94	38	459

### Proportions Row Column Overall

Campus \ Social Media favorite?	Twitter	Instagram	Facebook	Snapchat	Other	Total
Canyon Country Campus	0.194	0.289	0.239	0.2	0.078	1
Valencia Campus	0.201	0.305	0.201	0.208	0.086	1
Total	0.198	0.298	0.216	0.205	0.083	1

Notice that all of the rows add up to 1 (100%). This confirms that the computer is calculating the conditional proportions for the rows. We are looking for the proportion of Canyon Country pre-stat students that prefer Twitter. Notice the answer we are looking for is given in the intersecting cell. If we restrict ourselves to considering only the Canyon Country students, 0.194 or 19.4% of them prefer twitter. This is the same answer we got earlier in the section.

Example:

What proportion of the Snapchat math 075 pre-stat students attend the Valencia campus?

To answer this we need to recognize that we are no longer considering all the students. We are restricting our proportion to considering only the Snapchat students (“out of”). Notice that the student that prefer Snapchat are in a column. Since the condition is preferring Snapchat, we should only use numbers in the Snapchat column to calculate the proportion. Notice we highlighted the numbers in Snapchat column. The total will now be the total number of Snapchat students and the amount will be the amount of Snapchat students that attend the Valencia campus.

### Counts Table Switch Variables

Campus \ Social Media favorite?	Twitter	Instagram	Facebook	Snapchat	Other	Total
Canyon Country Campus	35	52	43	36	14	180
Valencia Campus	56	85	56	58	24	279
Total	91	137	99	94	38	459

$$\text{Conditional Proportion} = \frac{\text{Amount in Intersection Cell (Snapchat meets Valencia)}}{\text{Row or Column Total (column total Snapchat)}} = \frac{58}{94} \approx 0.617021276 \approx 0.617$$

$$\text{Conditional Percentage} = \text{Conditional Proportion} \times 100\% = 0.617 \times 100\% = 61.7\%$$



We can also use StatKey to find the proportion of the Snapchat math 075 pre-stat students attend the Valencia campus. Notice our condition is now Snapchat (“out of”). This is a column so I will click the “column” proportion button in StatKey.

**Counts Table** Switch Variables

Campus \ Social Media favorite?	Twitter	Instagram	Facebook	Snapchat	Other	Total
Canyon Country Campus	35	52	43	36	14	180
Valencia Campus	56	85	56	58	24	279
Total	91	137	99	94	38	459

**Proportions** Row Column Overall

Campus \ Social Media favorite?	Twitter	Instagram	Facebook	Snapchat	Other	Total
Canyon Country Campus	0.385	0.38	0.434	0.383	0.368	0.392
Valencia Campus	0.615	0.62	0.566	0.617	0.632	0.608
Total	1	1	1	1	1	1

Notice that when we click the “Column” proportion button, all of the columns add up to 1 (100%). This lets us know that StatKey has calculated all of the conditional proportions for the columns.

The conditional proportion we are looking for is where Snapchat and Valencia intersect. 0.617 or 61.7%. Notice this is the same answer as our earlier calculation.

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.

Relationship Principle

Let us go back to the LeBron James example. The key to understanding categorical relationships is to judge how close or far apart conditional percentages are.

- Chances of Los Angeles Lakers beating the Phoenix Suns if LeBron James plays  $\approx 75\%$
- Chances of Los Angeles Lakers beating the Phoenix Suns if LeBron James does not play  $\approx 25\%$

These percentages are significantly different, so it tells us that the condition of LeBron James playing in the game is related to the Lakers winning.

Note: Does this mean that LeBron playing is the only factor that CAUSES the Lakers to win? No. Remember related (associated) does NOT prove cause and effect. There are many confounding variables that go into the Lakers winning or losing. (Health of LeBron, Health of other players, the team the Lakers are playing, home game or away game, Number of games played, etc...) We can say that LeBron James playing is related to the Lakers winning, but the data does not prove that LeBron James playing is the only factor that causes the Lakers to win.



Let us look at another example using the Lakers chances of beating the Phoenix Suns.

Chances of Lakers winning if it snows in Nebraska  $\approx 75\%$   
 Chances of Lakers winning if it does not snow in Nebraska  $\approx 75\%$

These percentages are not significantly different, so it tells us that the condition of snowing in Nebraska is not related to the Lakers winning. The condition does not matter.

**Relationship Principle:**

**Close Conditional Percentages from same variable = Condition is NOT related to the categorical variable**

**Significantly Different Conditional Percentages from same variable = Condition IS related to the categorical variable**

*Note: You cannot compare any conditional percentages you want. They must be the same variable for the percentage and from different groups (different condition). You cannot compare the percentage of Snapchat students from the Canyon Country campus to the percentage of Twitter from the Valencia campus. They are not the same thing and will likely have very different percentages regardless of the relationship. Compare the percentage of Snapchat students from the Canyon Country campus to the percentage of Snapchat students from the Valencia campus. That will give us information about the relationship. Conditional percentage analysis is the basis behind the Chi-Square test statistics in more advanced statistics classes.*

Example:

Look at the following conditional proportions StatKey calculated based on the rows (Canyon Country campus and Valencia campus).

**Counts Table** Switch Variables

Campus \ Social Media favorite?	Twitter	Instagram	Facebook	Snapchat	Other	Total
Canyon Country Campus	35	52	43	36	14	180
Valencia Campus	56	85	56	58	24	279
Total	91	137	99	94	38	459

**Proportions** Row Column Overall

Campus \ Social Media favorite?	Twitter	Instagram	Facebook	Snapchat	Other	Total
Canyon Country Campus	0.194	0.289	0.239	0.2	0.078	1
Valencia Campus	0.201	0.305	0.201	0.208	0.086	1
Total	0.198	0.298	0.216	0.205	0.083	1

We can only compare Twitter to Twitter, Instagram to Instagram, Facebook to Facebook, Snapchat to Snapchat, Other to Other. Notice the proportions look very close. This gives us the idea that a pre-stat students favorite social media may not be related to the campus they go to. If the proportions were significantly different, that may indicate a relationship.

