

**⚠** Lagunita is retiring and will shut down at 12 noon Pacific Time on March 31, 2020. A few courses may be open for self-enrollment for a limited time. We will continue to offer courses on other online learning platforms; visit <http://online.stanford.edu>.

Course > EDA: Examining Relationships > Case C→C > Case C→C: Conditional Percents

🔖 Bookmark this page

## Case C→C: Conditional Percents

**Learning Objective: Produce a two-way table, and interpret the information stored in it about the association between two categorical variables by comparing conditional percentages.**

So far, we have organized the raw data in a much more informative display—the two-way table:

		Body Image			
		About Right	Overweight	Underweight	Total
Gender	Female	560	163	37	760
	Male	295	72	73	440
	Total	855	235	110	1200

Remember, though, that our primary goal is to explore how body image is related to gender. Exploring the relationship between two categorical variables (in this case, body image and gender) amounts to comparing the distributions of the response variable (in this case, body image) across the different values of the explanatory variable (in this case, males and females):

Compare these distributions! →

		Body Image			
		About Right	Overweight	Underweight	Total
Gender	Female	560	163	37	760
	Male	295	72	73	440
	Total	855	235	110	1200

Note that it doesn't make sense to compare raw counts, because there are more females than males overall. So, for example, it is not very informative to say, "There are 560 females who responded 'about right' compared to only 295 males," since the 560 females are out of a total of 760, and the 295 males are out of a total of only 440.

We need to supplement our display, the two-way table, with some numerical summaries that will allow us to compare the distributions. These numerical summaries are found by simply **converting the counts to percentages within (or restricted to) each value of the explanatory variable separately**.

In our example, we look at each gender separately and convert the counts to percentages **within that gender**. Let's start with females:

		Body Image			
		About Right	Overweight	Underweight	Total
Gender	Female	$560/760 = 73.7\%$	$163/760 = 21.5\%$	$37/760 = 4.9\%$	$760/760 = 100\%$
	Male	%	%	%	%

Note that each count is converted to a percentage by dividing by the total number of females, 760. These numerical summaries are called **conditional percentages**, since we find them by "conditioning" on one of the genders.

## Learn By Doing

1/1 point (graded)

Now complete the table by computing the conditional percentages for the males.

What is the correct number for the cell indicated by the question mark?

		Body Image			
		About Right	Overweight	Underweight	Total
Gender	Female	$\frac{560}{760} = 73.7\%$	$\frac{163}{760} = 21.5\%$	$\frac{37}{760} = 4.9\%$	$\frac{760}{760} = 100\%$
	Male	?			100%

☐  $\frac{295}{560} = 52.7\%$

☐  $\frac{295}{855} = 34.5\%$

☒  $\frac{295}{440} = 67\%$  ✓

☐  $\frac{295}{1200} = 24.6\%$

Answer

Correct: 295 out of the 440 males responded "about right," which is 67%.

Submit

Learn By Doing

1/1 point (graded)

Complete the table by computing the conditional percentages for the males.

What is the correct number for the cell indicated by the question mark?

		Body Image			
		About Right	Overweight	Underweight	Total
Gender	Female	$\frac{560}{760} = 73.7\%$	$\frac{163}{760} = 21.5\%$	$\frac{37}{760} = 4.9\%$	$\frac{760}{760} = 100\%$
	Male	67%	?		100%

☐  $72/235 = 30.6\%$

☒  $72/440 = 16.4\%$  ✓

☐  $72/1200 = 6\%$

☐  $72/163 = 44.2\%$

**Answer**  
Correct: 72 out of the 440 males responded "overweight," which is 16.4%.

Submit

Learn By Doing

1/1 point (graded)  
Complete the table by computing the conditional percentages for the males.

What is the correct number for the cell indicated by the question mark?

		Body Image			
		About Right	Overweight	Underweight	Total
Gender	Female	$560/760 = 73.7\%$	$163/760 = 21.5\%$	$37/760 = 4.9\%$	$760/760 = 100\%$
	Male	67%	16.4%	?	100%

☒  $73/440 = 16.6\%$  ✓

☐  $73/110 = 66.4\%$

☐  $73/760 = 9.6\%$

☐  $73/1200 = 6.1\%$

**Answer**

Correct: Indeed, 73 out of the 440 males responded "underweight," which is 16.6%.

Submit

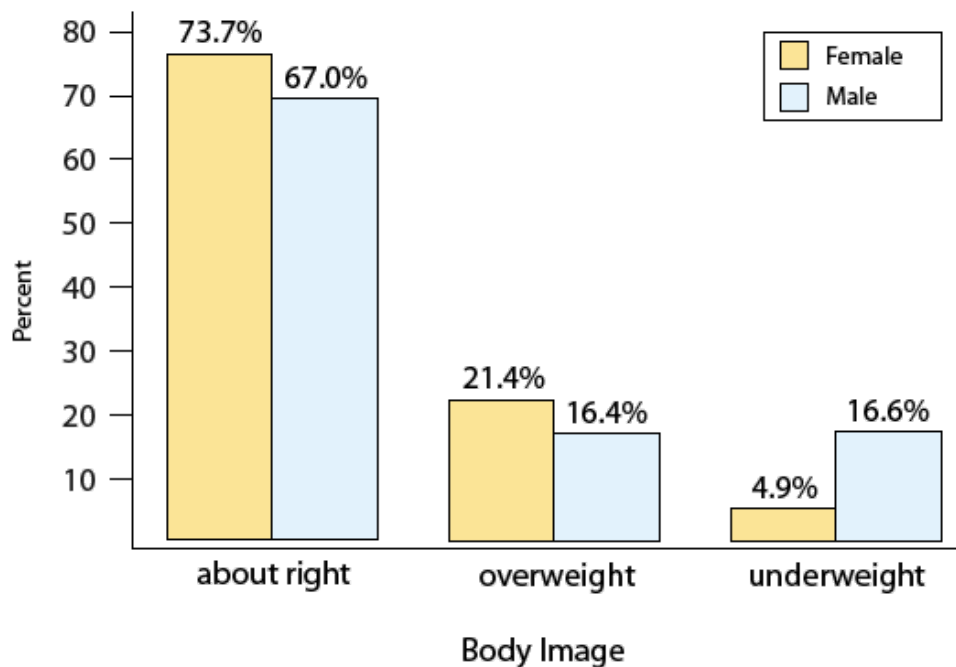
Good work! Here is the completed table:

		Body Image			
		About Right	Overweight	Underweight	Total
Gender	Female	$560/760 = 73.7\%$	$163/760 = 21.5\%$	$37/760 = 4.9\%$	$760/760 = 100\%$
	Male	67%	16.4%	16.6%	100%

## Comments

1. In our example, we chose to organize the data with the explanatory variable gender in rows and the response variable body image in columns, and thus our conditional percentages were *row percentages*, calculated within each row separately. Similarly, if the explanatory variable happens to sit in columns and the response variable in rows, our conditional percentages will be *column percentages*, calculated within each column separately. For an example, see the "Did I Get This?" exercises below.
2. Another way to visualize the conditional percentages, instead of in a table, is to use a *double bar chart*. This display is quite common in newspapers.

		Body Image			Total
		About Right	Overweight	Underweight	
Gender	Female	73.7%	21.4%	4.9%	100%
	Male	67.0%	16.4%	16.6%	100%



Now that we have summarized the relationship between the categorical variables gender and body image, let's go back and interpret the results in the context of the questions that we posed.

## Learn By Doing

1/1 point (graded)

The results suggest that the proportion of males who are happy with their body weight is \_\_\_\_\_ than among female students.

☐ slightly greater

☒ slightly less ✓

### Answer

Correct:

Students who are happy with their body weight responded "about right." 73.7% of females responded "about right," compared to only 67% of males.

Submit

## Learn By Doing

1/1 point (graded)

Females who are not happy with their body weight more often feel they are \_\_\_\_\_ .

☐ about right

☒ overweight ✓

☐ underweight

### Answer

Correct:

Indeed, 21.4% of females responded "overweight" compared to only 4.9% of females who responded "underweight."

Submit

## Learn By Doing

1/1 point (graded)

Males who are not happy with their body weight feel they are overweight \_\_\_\_\_ they feel they are underweight.

☒ about as often as ✓

☐ much more often than

☐ much less often than

### Answer

Correct:

Indeed, 16.4% of males responded "overweight," which is roughly the same as the percentage of males who responded "underweight" (16.6%).

Submit

### Scenario: Smoking Habits

Suppose a study were done to answer the question: "Is the smoking of students related to their parents' smoking habits?" in which data were collected from 5,375 students and organized in the following two-way table:

		Parents' Smoking		
		Parents Do Not Smoke	Parents Smoke	Total
Student Smoking	Student Does Not Smoke	1168	3203	4371
	Student Smokes	188	816	1004
	Total	1356	4019	5375

### Did I Get This

1/1 point (graded)

We would like to learn about the relationship between the two categorical variables "parents' smoking" and "students' smoking." More specifically, we want to explore whether parents' smoking habits affect their children's smoking habits

In this situation, what is the explanatory variable?

☐ students' smoking

☒ parents' smoking ✓

### Answer

Correct:

We are exploring the impact of parents' smoking habits on their children and therefore parents' smoking is the explanatory variable.

Submit

### Did I Get This

1/1 point (graded)



What is the response variable?

☒ students' smoking ✓

☐ parents' smoking

**Answer**  
Correct:  
Since we are exploring the impact of parents' smoking habits on their children, students' smoking is the response variable.

Submit

Did I Get This

1/1 point (graded)  
Based on your answer to the question above, which of the following three tables is the appropriate table of conditional percentages with which to supplement the above contingency table?

		Table A		
		Parents' Smoking		
		Parents Do Not Smoke	Parents Smoke	Total
Student Smoking	Student Does Not Smoke	26.7%	73.3%	100%
	Student Smokes	18.7%	81.3%	100%

		Table B	
		Parents' Smoking	
		Parents Do Not Smoke	Parents Smoke
Student Smoking	Student Does Not Smoke	86.1%	79.7%
	Student Smokes	13.9%	20.3%
	Total	100%	100%

		Table C		
		Parents' Smoking		
		Parents Do Not Smoke	Parents Smoke	Total
Student Smoking	Student Does Not Smoke	21.73%	59.59%	81.32%
	Student Smokes	3.50%	15.18%	18.68%
	Total	25.23%	74.77%	100%

☐ Table A

☒ Table B ✓

☐ Table C

### Answer

Correct:

Since the explanatory variable is in the columns, we find conditional column percentages. For example, in the **Parents Do Not Smoke** column, we find  $1,168/1,356 = 86.1\%$  and  $188/1,356 = 13.9\%$ .

### Scenario: Night Lights

An Associated Press article captured the attention of readers with the headline "Night lights bad for kids?" The article was based on a 1999 study at the University of Pennsylvania and Children's Hospital of Philadelphia, in which parents were surveyed about the lighting conditions under which their children slept between birth and age 2 (lamp, night-light, or no light) and whether or not their children developed nearsightedness (myopia). The purpose of the study was to explore the effect of a young child's nighttime exposure to light on later nearsightedness. There were 479 subjects included in the study. A summary of the two-way table is below.

Light	Nearsightedness	
	No	Yes
lamp	34	41
night light	153	79
no light	155	17

The table of conditional proportions showing which proportion of children in each “treatment group” became nearsighted is:

Light	Nearsightedness	
	No	Yes
lamp	45.333333	54.666667
night light	65.948276	34.051724
no light	90.116279	9.883721

The table of conditional proportions showing which proportion of children who are nearsighted or not were in each “treatment group” is:

Light	Nearsightedness	
	No	Yes
lamp	9.94152	29.92701
night light	44.73684	57.66423
no light	45.32164	12.40876

## Learn By Doing

1/1 point (graded)

How many children were exposed to no light at all?

☐ 137

☒ 172 ✓

☐ 232

☐ 342

### Answer

Correct: The number of children who were exposed to no light at all is calculated by  $155 + 17 = 172$ .

Submit

## Learn By Doing

1/1 point (graded)

Among the children who were exposed to no light at all, what percentage developed nearsightedness?

☒ 9.88 ✓

☐ 34.05

☐ 45.33

☐ 90.12

**Answer**

Correct:

Using the table of conditional percentages showing which percentage of children in each “treatment group” became nearsighted, the percentage of children in the "No light" condition who also had nearsightedness (= Yes) is 9.88%.

Submit

## Learn By Doing

1/1 point (graded)

Among the children who slept with a lamp, what percentage developed nearsightedness?

☐ 9.88

☐ 45.33

☒ 54.67 ✓

☐ 65.95

**Answer**

Correct:

Using the table of conditional percentages showing which percentage of children in each "treatment group" became nearsighted, the percentage of children in the "Lamp" condition who also had nearsightedness (= Yes) is 54.67%.

Submit

## Let's Summarize

- The relationship between two categorical variables is summarized using:
  - **Data display:** two-way table, supplemented by
  - **Numerical summaries:** conditional percentages.
- Conditional percentages are calculated for each value of the explanatory variable separately. They can be row percents, if the explanatory variable "sits" in the rows, or column percents, if the explanatory variable "sits" in the columns.
- When we try to understand the relationship between two categorical variables, we compare the distributions of the response variable for values of the explanatory variable. In particular, we look at how the pattern of conditional percentages differs between the values of the explanatory variable.

---

Open Learning Initiative [🔗](#)



[🔗](#) Unless otherwise noted this work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License [🔗](#).

© All Rights Reserved