

⚠ 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: Two-Way Tables

🔖 Bookmark this page

Case C→C: Two-Way Tables

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.

Case C→C: Two Categorical Variables

Recall the role-type classification table for framing our discussion about the relationship between two variables:

		Response	
		Categorical	Quantitative
Explanatory	Categorical	C→C	✓ C→Q
	Quantitative	Q→C	Q→Q

We are done with case C→Q, and will now move on to case C→C, where we examine the relationship between two categorical variables.

Earlier in the course, (when we discussed the distribution of a **single** categorical variable) we examined the data obtained when a random sample of 1,200 U.S. college students were asked about their body image (underweight, overweight, or about right.) We are now returning to this example, to address the


following question:

If we had separated our sample of 1,200 U.S. college students by gender and looked at **males and females separately**, would we have found a similar distribution across body-image categories? More specifically, are men and women just as likely to think their weight is about right? Among those students who do not think their weight is about right, is there a difference between the genders in feelings about body image?

Answering these questions requires us to **examine the relationship between two categorical variables**, gender and body image. Because the question of interest is whether there is a gender effect on body image,

- the **explanatory** variable is **gender**, and
- the **response** variable is **body image**.

Here is what the raw data look like when we include the gender of each student:



Student	Gender	Body Image
.	.	.
.	.	.
student 25	M	overweight
student 26	M	about right
student 27	F	underweight
student 28	F	about right
student 29	M	about right
.	.	.
.	.	.

Once again the raw data is a long list of 1,200 genders and responses, and thus not very useful in that form. To start our exploration of how body image is related to gender, we need an informative display that summarizes the data. In order to summarize the relationship between two categorical variables, we create a display called a **two-way table**.

Here is the two-way table for our example:

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

The table has the possible genders in the rows, and the possible responses regarding body image in the columns. At each intersection between row and column, we put the counts for how many times that combination of gender and body image occurred in the data. We sum across the rows to fill in the Total column, and we sum across the columns to fill in the Total row.

Learn By Doing

1/1 point (graded)

How many **females** in our sample feel that they are **underweight**?

☒ 37 ✓

☐ 110

☐ 163

☐ 760

Answer

Correct:

Since we want those students who are both **female** and responded **underweight**, we are looking for the intersection between the **female** row and the **underweight** column.

Submit

Learn By Doing

1/1 point (graded)

How many **males** in our sample feel that they are **about right**?

☒ 295 ✓

☐ 440

☐ 560

☐ 855

Answer

Correct:

Since we want the number who are both **male** and responded **about right**, we are looking for the intersection between the male row and the about right column.

Submit

Learn By Doing

1/1 point (graded)

What is the **total** number of **females** in our sample?

☐ 440

☐ 560

☒ 760 ✓

☐ 1200

Answer

Correct: The number in the **total** column that is in the **female** row is 760.

Submit

Learn By Doing

1/1 point (graded)

What is the **total** number of students who feel they are **overweight**?

☐ 110

☐ 163

☒ 235 ✓

☐ 1200

Answer

Correct: The total of the **overweight** column is 235.

Submit

Comment

Note that from the way the two-way table is constructed, the Total row or column is a summary of one of the two categorical variables, ignoring the other. In our example:

- The Total row gives the summary of the categorical variable body image:

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

(These are the same counts we got earlier in the course when we looked at the single categorical variable body image, and did not consider gender.)

- The Total column gives the summary of the categorical variable gender:

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

Open Learning Initiative



Unless otherwise noted this work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

© All Rights Reserved