

## 2.2: Tabular Displays

### Frequency Tables for Quantitative Data

To create many of these graphs, you must first create the frequency distribution. The idea of a frequency distribution is to take the interval that the data spans and divide it into equal sized subintervals called classes. The grouped frequency distribution gives either the frequency (count) or the relative frequency (usually expressed as a percent) of individuals who fall into each class. When creating frequency distributions, it is important to note that the number of classes that are used and the value of the first class boundary will change the shape of, and hence the impression given by, the distribution. There is no one correct width to the class boundaries. It usually takes several tries to create a frequency distribution that looks just the way you want. As a reader and interpreter of such tables, you should be aware that such features are selected so that the table looks the way it does to show a particular point of view. For small samples, we usually have between four and seven classes, and as you get more data you will need more classes.

We will start with an example of a random sample of 35 ages from credit card applications given below in no particular order. Organize the data in a frequency distribution table.

46 47 49 25 46 22 42 32 39 24 46 40 39 27 25 30 31 29 33 27 46 21 29 20 26 39 26 25 25 26 35 49 33 26 30

#### Solution

As you can see, the data is hard to interpret in this format. If we were to peruse through the data we could find that minimum age is 20 and the maximum age is 49. If we cut the age groups up in to 10-year intervals, we get only three classes 20-29, 30-39 and 40-49. Although this would work, if we had more classes we can sometimes see trends within the data at a more granular level. If we split the ages up in to 5-year intervals to 20-24, 25-29, 30-34, 35-39, 40-44, and 45-49 we would get six classes.

Note that the class limits should never overlap; we call these mutually exclusive classes. For example, if your class went from 20-25, 25-30 etc., the 25-year-olds would fall within both classes. Also, make sure each of the classes have the same width. A more formal way to pick your classes uses the following process.

#### Steps involved in making a frequency distribution table:

1. Find the range = largest value – smallest value.
2. Pick the number of classes to use. Usually the number of classes is between five and twenty. Five classes are used if there are a small number of data points and twenty classes if there are a large number of data points (over 1,000 data points).
3. Class width =  $\frac{\text{range}}{\# \text{ of classes}}$ . Always round up to the next integer (if the answer is already a whole number, go to the next integer).  
If you do not round up, your last class will not contain your largest data value, and you would have to add another class just for it. If you round up, then your largest data value will fall in the last class, and there are no issues.
4. Create the classes. Each class has limits that determine which values fall in each class. To find the **class limits**, set the smallest value in the data set as the lower class limit for the first class. Then add the class width to the lower class limit to get the next lower class limit. Repeat until you get all the classes. The upper class limit for a class is one less than the lower limit for the next class.
5. If your data value has decimal places, then round up the class width to the nearest value with the same number of decimal places as the original data. As an example, if your data was out to two decimal places, and you divided your range by the number of classes to get 4.8333, then the class width would round the second decimal place up and end on 4.84.
6. The frequency for a class is the number of data values that fall in the class.

For the age data let us use 6 classes. Find the range by taking  $49 - 20 = 29$  and divide this by the number of classes  $29/6 = 4.8333$ . Round this number up to 5 and use 5 for the class width. Once you determine your class width and class limits, place each of these classes in a table and then tally up the ages that fall within each class. Count the tally marks and record the number in the frequency table.

The total of the frequency column should be the number of observations in the data. You may want to total the frequencies to make sure you did not leave any of the numbers out of the table.

Class Limits	Tally		Class	Frequency
20-24	4		20-24	4
25-29	12		25-29	12
30-34	6		30-34	6
35-39	4		35-39	4
40-44	2		40-44	2
45-49	7		45-49	7
			<b>Total</b>	<b>35</b>

Figure 2-1

Using the frequency table, we can now see that there are more people in the 25- 29 year-old class, followed by the 45-49 year-old class.

We call this most frequent category the **modal class**. There may be no mode at all or more than one mode.

## Frequency Tables for Qualitative Data

A frequency distribution can also be made for qualitative data.

Suppose you have the following data for which type of car students at a college drive. Make a frequency table to summarize the data.

Ford Honda Nissan Chevy Chevy Chevy Chevy Toyota Chevy Saturn Honda Toyota Nissan Honda Toyota Toyota Nissan Ford Toyota Chevy Toyota Ford Chevy Chevy Chevy Nissan Toyota Toyota Ford Nissan Kia Nissan Nissan Nissan Honda Nissan Mercedes Honda Toyota Toyota Chevy Chevy Porsche Chevy Toyota Toyota Ford Hyundai Honda Nissan

### Solution

The list of data is hard to analyze, so you need to summarize it. The classes in this case are the car brands. However, several car brands only have one car in the list. In that case, it is easier to make a category called “other” for the categories with low frequencies.

Count how many of each type of cars there are: there are 12 Chevys, 5 Fords, 6 Hondas, 10 Nissans, 12 Toyotas, and 5 other brands (Hyundai, Kia, Mercedes, Porsche, and Saturn). Place the other brands into a frequency distribution table alphabetically:

Category	Frequency
Chevy	12
Ford	5
Honda	6
Nissan	10
Toyota	12
Other	5
<b>Total</b>	<b>50</b>

For nominal data, either alphabetize the classes or arrange the classes from most frequent to least frequent, with the “other” category always at the end. For ordinal data put the classes in their order with the “other” category at the end.

## Relative Frequency Tables

Frequencies by themselves are not as useful to tell other people what is going on in the data. If you want to know what percentage the category is of the total sample then we can use the relative frequency of each category. The relative frequency is just the frequency divided by the total. The relative frequency is the proportion in each category and may be given as a decimal, percentage, or fraction.

Using the car data's frequency distribution, we will create a third column labeled relative frequency. Take each frequency and divide by the sample size, see Figure 2-2. The relative frequencies should add up to one (ignoring rounding error).

Type of Car	Frequency	Relative Frequency
Chevy	12	$12/50 = 0.24$
Ford	5	$5/50 = 0.1$
Honda	6	$6/50 = 0.12$
Nissan	10	$10/50 = 0.2$
Toyota	12	$12/50 = 0.24$
Other	5	$5/50 = 0.1$
<b>Total</b>	<b>50</b>	<b>1</b>

Figure 2-2

Many people understand percentages better than proportions so we may want to multiply each of these decimals by 100% to get the following relative frequency percent table.

Type of Car	Percent
Chevy	24
Ford	10
Honda	12%
Nissan	20%
Toyota	24%
Other	10%
<b>Total</b>	<b>100%</b>

Figure 2-3

We can summarize the car data and see that for college students Chevy and Toyota make up 48% of the car models.

### Excel

Recall the frequency table for the credit card applicants. The relative frequency table for the random sample of 35 ages from credit card applications follows.

**Class** **Frequency** **Relative Frequency** 20-24 4  $4/35 = 0.1143$  25-29 12  $12/35 = 0.3429$  30-34 6  $6/35 = 0.1714$  35-39 4  $4/35 = 0.1143$  40-44 2  $2/35 = 0.0571$  45-49 7  $7/35 = 0.2$  **Total** **35** **1**

Making a relative frequency table using Excel

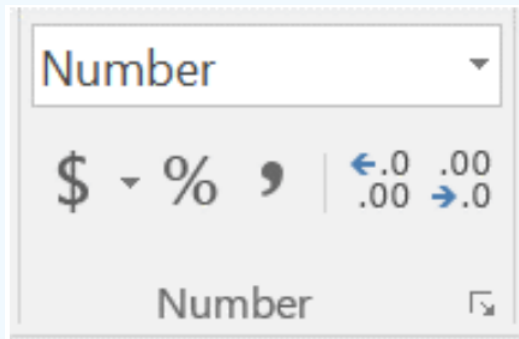
#### Solution

In Excel, type your frequencies into a column and then in the next column type in  $=(\text{cell reference number})/35$ . Then copy and paste the formula in cell B2 down the page. If you used the cell reference number, Excel will automatically change the copied cells to the next row down.

You get the following relative frequency table. The sum of the relative frequencies will be one (the sum may add to 0.9999 or 1.0001 if you are doing the calculations by hand due to rounding).

	A	B
1	Frequency	Relative Frequency
2	4	=A2/35
3	12	=A3/35
4	6	=A4/35
5	4	=A5/35
6	2	=A6/35
7	7	=A7/35
8	=SUM(A2:A7)	=SUM(B2:B7)

To get Excel to show the percentage instead of the proportion by highlighting the relative frequencies and selecting the percent % button on the Home tab.



Class	Frequency	Relative Frequency	Relative Frequency Percent
20-24	4	0.1143	11%
25-29	12	0.3429	34%
30-34	6	0.1714	17%
35-39	4	0.1143	11%
40-44	2	0.0571	6%
45-49	7	0.2	20%
<b>Total</b>	<b>35</b>	<b>1</b>	<b>100%</b>

The relative frequency table lets us quickly see that a little more than half  $34\% + 20\% = 54\%$  of the ages of the credit card holders are between the ages of 25 – 29 and 45 – 49 years-old.

### Cumulative & Cumulative Relative Frequency Tables

Another useful piece of information is how many data points fall below a particular class. As an example, a teacher may want to know how many students received below a 70%, a doctor may want to know how many adults have cholesterol above 160, or a manager may want to know how many stores gross less than \$2,000 per day. This calculation is known as a **cumulative frequency** and is used for ordinal or quantitative data. If you want to know what percent of the data falls below a certain class, then this fact would be a **cumulative relative frequency**.

To create a **cumulative frequency distribution**, count the number of data points that are below the upper class limit, starting with the first class and working up to the top class. The last upper class should have all of the data points below it.

Recall the credit card applicants. Make a cumulative frequency table.

**Class** **Frequency** **Relative Frequency** **Percent** 20-24 4 11% 25-29 12 34% 30-34 6 17% 35-39 4 11% 40-44 2 6% 45-49 7 20%  
**Total** 35 100%

### Solution

To find the cumulative frequency, carry over the first frequency of 4 to the first row of the cumulative frequency column. Then take this 4 and add it to the next frequency of 12 to get 16 for the second cumulative frequency value. For the third cumulative frequency, take 16 and add to the third frequency of 6 to get 22. Keep doing this additive process until you finish the column.

The cumulative frequency in the last class should have the last number as the sample size. The cumulative relative frequency is just the cumulative frequency divided by the sample size. The cumulative relative frequencies should have one as the last number.

Class	Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency
20-24	4	0.1143	4	$4/35 = 0.1143$
25-29	12	0.3429	$4 + 12 = 16$	$16/35 = 0.4571$
30-34	6	0.1714	$16 + 6 = 22$	$22/35 = 0.6286$
35-39	4	0.1143	$22 + 4 = 26$	$26/35 = 0.7429$
40-44	2	0.0571	$26 + 2 = 28$	$28/35 = 0.8$
45-49	7	0.2	$28 + 7 = 35$	$35/35 = 1$
<b>Total</b>	<b>35</b>	<b>1</b>		

Use Excel to add up the values for you.

	A	B	C	D
1	Frequency	Relative Frequency	Excel Formula	Cumulative Frequency
2	4	$=A2/35$	4	4
3	12	$=A3/35$	$=C2+A3$	16
4	6	$=A4/35$	$=C3+A4$	22
5	4	$=A5/35$	$=C4+A5$	26
6	2	$=A6/35$	$=C5+A6$	28
7	7	$=A7/35$	$=C6+A7$	35
8	$=SUM(A2:A7)$	$=SUM(B2:B7)$		

You can also express the cumulative relative frequencies as a percentage instead of a proportion.

Class	Cumulative Frequency	Cumulative Relative Frequency
20-24	4	11
25-29	16	46
30-34	22	63
35-39	26	74
40-44	28	80
45-49	35	100

If a manager wanted to know how many applicants were under the age of 40 we could look across the 35-39 year-old class to see that there were 26 applicants that were under 40 (39 years old or younger), or if we used the cumulative relative frequency, 74% of the applicants were under 40 years old. If the manager wants to know how many applicants were over 44, we could subtract  $100\% - 80\% = 20\%$ .

## Contingency Tables

A **contingency table** provides a way of portraying data that can facilitate calculating probabilities. A contingency table summarizes the frequency of two qualitative variables. The table displays sample values in relation to two different variables that may be dependent or contingent on one another. There are other names for contingency tables. Excel calls the contingency table a pivot table. Other common names are two-way table, cross-tabulation, or cross-tab for short.

A fitness center coach kept track of members over the last year. They recorded if the person stretched before they exercised, and whether they sustained an injury. The following contingency table shows their results. Find the relative frequency for each value in the table.

Injury No Injury Stretched 52 270 Did Not Stretch 21 57

### Solution

Each value in the table represents the number of times a particular combination of variable outcomes occurred, for example, there were 57 members that did not stretch and did not sustain an injury. It is helpful to total up the categories. The row totals provide the total counts across each row (i.e.  $52 + 270 = 322$ ), and column totals are total counts down each column (i.e.  $52 + 21 = 73$ ). See Figure 2-4.

	Injury	No Injury	Total
Stretched	52	270	322
Did Not Stretch	21	57	78
Total	73	327	400

Figure 2-4

We can quickly summarize the number of athletes for each category. The bottom right-hand number in Figure 2-4 is the grand total and represents the total number of 400 people. There were 322 people that stretched before exercising. There were 73 people that sustained an injury while exercising, etc.

If we find the relative frequency for each value in the table, we can find the proportion of the 400 people for each category. To find a relative frequency we divide each value in the table by the grand total. See Figure 2-5.

	Injury	No Injury	Total
Stretched	$52/400 = 0.13$	$270/400 = 0.675$	$322/400 = 0.805$ or $0.13 + 0.675 = 0.805$
Did Not Stretch	$21/400 = 0.0525$	$57/400 = 0.1425$	$78/400 = 0.195$ or $0.0525 + 0.1425 = 0.195$
Total	$73/400 = 0.1825$ or $0.13 + 0.0525 = 0.1825$	$327/400 = 0.8175$ or $0.675 + 0.1425 = 0.8175$	$400/400 = 1$ or the sum of either the row or column totals

Figure 2-5

When data is collected, it is usually presented in a spreadsheet where each row represents the responses from an individual or case.

*"Of course, one never has the slightest notion what size or shape different species are going to turn out to be, but if you were to take the findings of the latest Mid-Galactic Census report as any kind of accurate Guide to statistical averages you would probably guess that the craft would hold about six people, and you would be right. You'd probably guessed that anyway. The Census report, like most such surveys, had cost an awful lot of money and didn't tell anybody*

*anything they didn't already know - except that every single person in the Galaxy had 2.4 legs and owned a hyena. Since this was clearly not true the whole thing had eventually to be scrapped.” (Adams, 2002)*

Make a pivot table using Excel. A random sample of 500 records from the 2010 United States Census were downloaded to Excel. Below is an image of just the first 20 people.

There are seven variables:

- State
- Total family income (in United States dollars)
- Age
- Biological Sex (with reported categories female and male)
- Race (with reported categories American Indian or Alaska Native, Black, Chinese, Japanese, Other Asian or Pacific Islander, Two major races, White, Other)
- Marital status (with reported categories Divorced, Married/spouse absent, Married/spouse present, Never married/single, Separated, Widowed)
- Total personal income (in United States dollars).

	A	B	C	D	E	F	G
1	stateFIPScod	totalFamilyInc	age	sex	raceGeneral	maritalStatus	totalPersonalIncome
2	Alabama	5500	73	Female	Black	Divorced	3000
3	Alabama	63820	40	Male	White	Married/spouse present	47950
4	Alabama	11200	60	Female	Black	Divorced	11200
5	Alabama	34500	43	Female	Other	Married/spouse present	15300
6	Alabama	33600	7	Male	White	Never married/single	NA
7	Arizona	32500	9	Female	White	Never married/single	NA
8	Arizona	46800	53	Female	White	Married/spouse absent	30000
9	Arizona	30000	60	Female	White	Widowed	30000
10	Arizona	0	67	Female	White	Widowed	0
11	Arizona	51000	27	Male	White	Married/spouse present	14000
12	Arizona	6000	10	Male	White	Never married/single	NA
13	Arizona	45000	48	Male	Japanese	Married/spouse present	45000
14	Arkansas	10940	6	Female	Black	Never married/single	NA
15	California	52000	40	Female	White	Married/spouse present	27000
16	California	156000	65	Female	White	Married/spouse present	10000
17	California	59000	80	Male	White	Married/spouse present	52100
18	California	37500	46	Female	White	Married/spouse present	17500
19	California	67500	3	Female	White	Never married/single	NA
20	California	21800	6	Male	Other Asian	Never married/single	NA
21	California	79000	26	Female	White	Never married/single	25000

### Solution

In Excel, select the Insert tab, then select Pivot Table.

Excel should automatically select all 500 rows in the Table/Range cell, if not then use your mouse and highlight all the data including the labels. Then select OK.

Each version of Excel may look different at this point. One common area, though, is the bottom right-hand drag and drop area of the Pivot Table dialogue box.

Drag the sex variable to the COLUMNS box, and marital status variable to the ROWS box.

You will see the contingency table column and row headers appear as you drop the variables in the boxes.

To get the counts to appear, drag and drop marital status into the Values box and the default will usually say, “Count of maritalStatus,” if not then change it to count in the drop-down menu. A contingency table should appear on your spreadsheet as you fill in the pivot table dialogue box. Contingency tables go by many names including pivot tables, two-way tables, cross tabulations, or cross-tabs.

Count of Marital Status	Column Labels		
Row Labels	Female	Male	Grand Total
Divorced	21	17	38



Married/spouse absent	5	9	14
Married/spouse present	92	100	192
Never married/single	93	129	222
Separated	1	2	3
Widowed	20	11	31
<b>Grand Total</b>	<b>232</b>	<b>268</b>	<b>500</b>

This page titled [2.2: Tabular Displays](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Rachel Webb](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.