

2.1: Organizing Data - Frequency Distributions

Once you have a set of data, you will need to organize it so that you can analyze how frequently each datum occurs in the set. However, when calculating the frequency, you may need to round your answers so that they are as precise as possible.

Answers and Rounding Off

A simple way to round off answers is to carry your final answer one more decimal place than was present in the original data. Round off only the final answer. Do not round off any intermediate results, if possible. If it becomes necessary to round off intermediate results, carry them to at least twice as many decimal places as the final answer. For example, the average of the three quiz scores four, six, and nine is 6.3, rounded off to the nearest tenth, because the data are whole numbers. Most answers will be rounded off in this manner.

It is not necessary to reduce most fractions in this course. In [Probability Topics](#), the chapter on probability, it is more helpful to leave an answer as an unreduced fraction. Use your instructor's guidance regarding whether to reduce fractions.

Categorical Frequency Distribution

Twenty students were asked how many hours they worked per day. Their responses, in hours, are as follows:

5; 6; 3; 3; 2; 4; 7; 5; 2; 3; 5; 6; 5; 4; 4; 3; 5; 2; 5; 3.

Table lists the different data values in ascending order and their frequencies.

Table 2.1.1: Frequency Table of Student Work Hours

DATA VALUE	FREQUENCY
2	3
3	5
4	3
5	6
6	2
7	1

A frequency is the number of times a value of the data occurs. According to Table 2.1.1, there are three students who work two hours, five students who work three hours, and so on. The sum of the values in the frequency column, 20, represents the total number of students included in the sample.

Definition: Categorical Frequency Distribution

A categorical frequency distribution is a table to organize data that can be placed in specific categories, such as nominal- or ordinal-level data.

Definition: Relative frequencies

A *relative frequency* is the ratio (fraction or proportion) of the number of times a value of the data occurs in the set of all outcomes to the total number of outcomes. To find the relative frequencies, divide each frequency by the total number of students in the sample—in this case, 20. Relative frequencies can be written as fractions, percents, or decimals.

Table 2.1.2: Frequency Table of Student Work Hours with Relative Frequencies

DATA VALUE	FREQUENCY	RELATIVE FREQUENCY
2	3	$\frac{3}{20}$ or 0.15
3	5	$\frac{5}{20}$ or 0.25

DATA VALUE	FREQUENCY	RELATIVE FREQUENCY
4	3	$\frac{3}{20}$ or 0.15
5	6	$\frac{6}{20}$ or 0.30
6	2	$\frac{2}{20}$ or 0.10
7	1	$\frac{1}{20}$ or 0.05

The sum of the values in the relative frequency column of Table 2.1.2 is $\frac{20}{20}$, or 1.

Definition: Cumulative relative frequency

Cumulative relative frequency is the accumulation of the previous relative frequencies. To find the cumulative relative frequencies, add all the previous relative frequencies to the relative frequency for the current row, as shown in Table 2.1.3.

Table 2.1.3: Frequency Table of Student Work Hours with Relative and Cumulative Relative Frequencies

DATA VALUE	FREQUENCY	RELATIVE FREQUENCY	CUMULATIVE RELATIVE FREQUENCY
2	3	$\frac{3}{20}$ or 0.15	0.15
3	5	$\frac{5}{20}$ or 0.25	$0.15 + 0.25 = 0.40$
4	3	$\frac{3}{20}$ or 0.15	$0.40 + 0.15 = 0.55$
5	6	$\frac{6}{20}$ or 0.30	$0.55 + 0.30 = 0.85$
6	2	$\frac{2}{20}$ or 0.10	$0.85 + 0.10 = 0.95$
7	1	$\frac{1}{20}$ or 0.05	$0.95 + 0.05 = 1.00$

The last entry of the cumulative relative frequency column is one, indicating that one hundred percent of the data has been accumulated.

Because of rounding, the relative frequency column may not always sum to one, and the last entry in the cumulative relative frequency column may not be one. However, they each should be close to one.

Grouped Frequency Distribution

Definition: Grouped Frequency Distribution

A grouped frequency distribution is a table to organize data in which the data are grouped into classes with more than one unit in width. Used when the data is large, or it makes sense to group the data.

Table 2.1.4 represents the heights, in inches, of a sample of 100 male semiprofessional soccer players.

Table 2.1.4: Frequency Table of Soccer Player Height

HEIGHTS (INCHES)	FREQUENCY	RELATIVE FREQUENCY	CUMULATIVE RELATIVE FREQUENCY
59.95–61.95	5	$\frac{5}{100} = 0.05$	0.05
61.95–63.95	3	$\frac{3}{100} = 0.03$	$0.05 + 0.03 = 0.08$
63.95–65.95	15	$\frac{15}{100} = 0.15$	$0.08 + 0.15 = 0.23$
65.95–67.95	40	$\frac{40}{100} = 0.40$	$0.23 + 0.40 = 0.63$

HEIGHTS (INCHES)	FREQUENCY	RELATIVE FREQUENCY	CUMULATIVE RELATIVE FREQUENCY
67.95–69.95	17	$\frac{17}{100} = 0.17$	$0.63 + 0.17 = 0.80$
69.95–71.95	12	$\frac{12}{100} = 0.12$	$0.80 + 0.12 = 0.92$
71.95–73.95	7	$\frac{7}{100} = 0.07$	$0.92 + 0.07 = 0.99$
73.95–75.95	1	$\frac{1}{100} = 0.01$	$0.99 + 0.01 = 1.00$
	Total = 100	Total = 1.00	

The data in this table have been **grouped** into the following intervals:

- 61.95 to 63.95 inches
- 63.95 to 65.95 inches
- 65.95 to 67.95 inches
- 67.95 to 69.95 inches
- 69.95 to 71.95 inches
- 71.95 to 73.95 inches
- 73.95 to 75.95 inches

This example is used again in [Descriptive Statistics](#).

The next section will explain in detail how to create a grouped frequency distribution given a raw data set.

In this sample, there are **five** players whose heights fall within the interval 59.95–61.95 inches, **three** players whose heights fall within the interval 61.95–63.95 inches, **15** players whose heights fall within the interval 63.95–65.95 inches, **40** players whose heights fall within the interval 65.95–67.95 inches, **17** players whose heights fall within the interval 67.95–69.95 inches, **12** players whose heights fall within the interval 69.95–71.95, **seven** players whose heights fall within the interval 71.95–73.95, and **one** player whose heights fall within the interval 73.95–75.95. All heights fall between the endpoints of an interval and not at the endpoints.

Exercise 2.1.1

- From the [Table](#), find the percentage of heights that are less than 65.95 inches.
- Find the percentage of heights that fall between 61.95 and 65.95 inches.

-
-
-
-
-
-

Answer

- If you look at the first, second, and third rows, the heights are all less than 65.95 inches. There are $5 + 3 + 15 = 23$ players whose heights are less than 65.95 inches. The percentage of heights less than 65.95 inches is then $\frac{23}{100}$ or 23%. This percentage is the cumulative relative frequency entry in the third row.
- Add the relative frequencies in the second and third rows: $0.03 + 0.15 = 0.18$ or 18%.

Exercise 2.1.2

Table 2.1.5 shows the amount, in inches, of annual rainfall in a sample of towns.

Table 2.1.5:

Rainfall (Inches)	Frequency	Relative Frequency	Cumulative Relative Frequency
2.95–4.97	6	$\frac{6}{50} = 0.12$	0.12
4.97–6.99	7	$\frac{7}{50} = 0.14$	$0.12 + 0.14 = 0.26$
6.99–9.01	15	$\frac{15}{50} = 0.30$	$0.26 + 0.30 = 0.56$
9.01–11.03	8	$\frac{8}{50} = 0.16$	$0.56 + 0.16 = 0.72$
11.03–13.05	9	$\frac{9}{50} = 0.18$	$0.72 + 0.18 = 0.90$
13.05–15.07	5	$\frac{5}{50} = 0.10$	$0.90 + 0.10 = 1.00$
	Total = 50	Total = 1.00	

- Find the percentage of rainfall that is less than 9.01 inches.
- Find the percentage of rainfall that is between 6.99 and 13.05 inches.

Answer

- 0.56 or 56
- $0.30 + 0.16 + 0.18 = 0.64$ or 64

Exercise 2.1.3

Use the heights of the 100 male semiprofessional soccer players in Table 2.1.4. Fill in the blanks and check your answers.

- The percentage of heights that are from 67.95 to 71.95 inches is: ____.
- The percentage of heights that are from 67.95 to 73.95 inches is: ____.
- The percentage of heights that are more than 65.95 inches is: ____.
- The number of players in the sample who are between 61.95 and 71.95 inches tall is: ____.
- What kind of data are the heights?
- Describe how you could gather this data (the heights) so that the data are characteristic of all male semiprofessional soccer players.

Remember, you **count frequencies**. To find the relative frequency, divide the frequency by the total number of data values. To find the cumulative relative frequency, add all of the previous relative frequencies to the relative frequency for the current row.

Answer

- 29%
- 36%
- 77%
- 87
- quantitative continuous
- get rosters from each team and choose a simple random sample from each

Exercise 2.1.4

From Table 2.1.5, find the number of towns that have rainfall between 2.95 and 9.01 inches.

Answer

$$6 + 7 + 15 = 28 \text{ towns}$$

COLLABORATIVE EXERCISE 2.1.7

In your class, have someone conduct a survey of the number of siblings (brothers and sisters) each student has. Create a frequency table. Add to it a relative frequency column and a cumulative relative frequency column. Answer the following questions:

- What percentage of the students in your class have no siblings?
- What percentage of the students have from one to three siblings?
- What percentage of the students have fewer than three siblings?

Example 2.1.7

Nineteen people were asked how many miles, to the nearest mile, they commute to work each day. The data are as follows: 2; 5; 7; 3; 2; 10; 18; 15; 20; 7; 10; 18; 5; 12; 13; 12; 4; 5; 10. Table was produced:

Table 2.1.6: Frequency of Commuting Distances

DATA	FREQUENCY	RELATIVE FREQUENCY	CUMULATIVE RELATIVE FREQUENCY
3	3	$\frac{3}{19}$	0.1579
4	1	$\frac{1}{19}$	0.2105
5	3	$\frac{3}{19}$	0.1579
7	2	$\frac{2}{19}$	0.2632
10	3	$\frac{3}{19}$	0.4737
12	2	$\frac{2}{19}$	0.7895
13	1	$\frac{1}{19}$	0.8421
15	1	$\frac{1}{19}$	0.8948
18	1	$\frac{1}{19}$	0.9474
20	1	$\frac{1}{19}$	1.0000

- Is the table correct? If it is not correct, what is wrong?
- True or False: Three percent of the people surveyed commute three miles. If the statement is not correct, what should it be? If the table is incorrect, make the corrections.
- What fraction of the people surveyed commute five or seven miles?
- What fraction of the people surveyed commute 12 miles or more? Less than 12 miles? Between five and 13 miles (not including five and 13 miles)?

Answer

- No. The frequency column sums to 18, not 19. Not all cumulative relative frequencies are correct.
- False. The frequency for three miles should be one; for two miles (left out), two. The cumulative relative frequency column should read: 0.1052, 0.1579, 0.2105, 0.3684, 0.4737, 0.6316, 0.7368, 0.7895, 0.8421, 0.9474, 1.0000.
- $\frac{5}{19}$
- $\frac{7}{19}$, $\frac{12}{19}$, $\frac{7}{19}$

Exercise 2.1.8

Table represents the amount, in inches, of annual rainfall in a sample of towns. What fraction of towns surveyed get between 11.03 and 13.05 inches of rainfall each year?

Answer

$$\frac{9}{50}$$

Example 2.1.9

Table contains the total number of deaths worldwide as a result of earthquakes for the period from 2000 to 2012.

Table 2.1.7:

Year	Total Number of Deaths
2000	231
2001	21,357
2002	11,685
2003	33,819
2004	228,802
2005	88,003
2006	6,605
2007	712
2008	88,011
2009	1,790
2010	320,120
2011	21,953
2012	768
Total	823,356

Answer the following questions.

- What is the frequency of deaths measured from 2006 through 2009?
- What percentage of deaths occurred after 2009?
- What is the relative frequency of deaths that occurred in 2003 or earlier?
- What is the percentage of deaths that occurred in 2004?
- What kind of data are the numbers of deaths?
- The Richter scale is used to quantify the energy produced by an earthquake. Examples of Richter scale numbers are 2.3, 4.0, 6.1, and 7.0. What kind of data are these numbers?

Answer

- 97,118 (11.8%)
- 41.6%
- 67,092/823,356 or 0.081 or 8.1 %
- 27.8%
- Quantitative discrete
- Quantitative continuous

Exercise 2.1.10

Table contains the total number of fatal motor vehicle traffic crashes in the United States for the period from 1994 to 2011.

Table 2.1.8:

Year	Total Number of Crashes	Year	Total Number of Crashes
1994	36,254	2004	38,444
1995	37,241	2005	39,252
1996	37,494	2006	38,648
1997	37,324	2007	37,435
1998	37,107	2008	34,172
1999	37,140	2009	30,862
2000	37,526	2010	30,296
2001	37,862	2011	29,757
2002	38,491	Total	653,782
2003	38,477		

Answer the following questions.

- What is the frequency of deaths measured from 2000 through 2004?
- What percentage of deaths occurred after 2006?
- What is the relative frequency of deaths that occurred in 2000 or before?
- What is the percentage of deaths that occurred in 2011?
- What is the cumulative relative frequency for 2006? Explain what this number tells you about the data.

Answer

- 190,800 (29.2%)
- 24.9%
- 260,086/653,782 or 39.8%
- 4.6%
- 75.1% of all fatal traffic crashes for the period from 1994 to 2011 happened from 1994 to 2006.

References

- “State & County QuickFacts,” U.S. Census Bureau. quickfacts.census.gov/qfd/download_data.html (accessed May 1, 2013).
- “State & County QuickFacts: Quick, easy access to facts about people, business, and geography,” U.S. Census Bureau. quickfacts.census.gov/qfd/index.html (accessed May 1, 2013).
- “Table 5: Direct hits by mainland United States Hurricanes (1851-2004),” National Hurricane Center, <http://www.nhc.noaa.gov/gifs/table5.gif> (accessed May 1, 2013).
- “Levels of Measurement,” infinity.cos.edu/faculty/wood...ata_Levels.htm (accessed May 1, 2013).
- Courtney Taylor, “Levels of Measurement,” about.com, <http://statistics.about.com/od/Helpa...easurement.htm> (accessed May 1, 2013).
- David Lane. “Levels of Measurement,” Connexions, <http://cnx.org/content/m10809/latest/> (accessed May 1, 2013).

Glossary

Categorical Frequency Distribution

A table to organize data that can be placed in specific categories, such as nominal- or ordinal-level data.

Cumulative Relative Frequency

The term applies to an ordered set of observations from smallest to largest. The cumulative relative frequency is the sum of the relative frequencies for all values that are less than or equal to the given value.

Frequency

the number of times a value of the data occurs

Relative Frequency

the ratio of the number of times a value of the data occurs in the set of all outcomes to the number of all outcomes to the total number of outcomes

Contributors and Attributions

- Barbara Illowsky and Susan Dean (De Anza College) with many other contributing authors. Content produced by OpenStax College is licensed under a Creative Commons Attribution License 4.0 license. Download for free at <http://cnx.org/contents/30189442-699...b91b9de@18.114>.

This page titled [2.1: Organizing Data - Frequency Distributions](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by [OpenStax](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.