1.4: Levels of Measurement

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 needed for your research and presentation purposes.

Levels of Measurement

The way a set of data is measured is called its **level of measurement**. Correct statistical procedures depend on a researcher being familiar with levels of measurement. Not every statistical operation can be used with every set of data. Data can be classified into four levels of measurement. They are (from lowest to highest level):

- **Nominal scale level**
- **Ordinal scale level**
- **Interval scale level**
- **Ratio scale level**

Data that is measured using a **nominal scale** is **qualitative (categorical)**. Categories, colors, names, labels and favorite foods along with yes or no responses are examples of nominal level data. Nominal scale data are not ordered. For example, trying to classify people according to their favorite food does not make any sense. Putting pizza first and sushi second is not meaningful.

Smartphone companies are another example of nominal scale data. The data are the names of the companies that make smartphones, but there is no agreed upon order of these brands, even though people may have personal preferences. Nominal scale data cannot be used in calculations.

Data that is measured using an **ordinal scale** is similar to nominal scale data but there is a big difference. The ordinal scale data can be ordered. An example of ordinal scale data is a list of the top five national parks in the United States. The top five national parks in the United States can be ranked from one to five but we cannot measure differences between the data.

Another example of using the ordinal scale is a cruise survey where the responses to questions about the cruise are “excellent,” “good,” “satisfactory,” and “unsatisfactory.” These responses are ordered from the most desired response to the least desired. But the differences between two pieces of data cannot be measured. Like the nominal scale data, ordinal scale data cannot be used in most calculations.

Data that is measured using the **interval scale** is similar to ordinal level data because it has a definite ordering but there is a difference between data. The differences between interval scale data can be measured though the data does not have a starting point.

Temperature scales like Celsius (C) and Fahrenheit (F) are measured by using the interval scale. In both temperature measurements, 40° is equal to 100° minus 60°. Differences make sense. But 0 degrees does not because, in both scales, 0 is not the absolute lowest temperature. Temperatures like -10° F and -15° C exist and are colder than 0.

Interval level data can be used in calculations, but one type of comparison cannot be done. 80° C is not four times as hot as 20° C (nor is 80° F four times as hot as 20° F). There is no meaning to the ratio of 80 to 20 (or four to one).

Data that is measured using the **ratio scale** takes care of the ratio problem and gives you the most information. Ratio scale data is like interval scale data, but it has a 0 point and ratios can be calculated. For example, four multiple choice statistics final exam scores are 80, 68, 20 and 92 (out of a possible 100 points). The exams are machine-graded.

The data can be put in order from lowest to highest: 20, 68, 80, 92.

The differences between the data have meaning. The score 92 is more than the score 68 by 24 points. Ratios can be calculated. The smallest score is 0. So 80 is four times 20. The score of 80 is four times better than the score of 20.

Frequency

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 $1.4.1$ lists the different data values in ascending order and their frequencies.

Data value	Frequency
2	3
3	5
4	3
5	6
6	2
7	1

Table $1.4.1$ Frequency Table of Student Work Hours

A **frequency** is the number of times a value of the data occurs. According to Table $1.4.1$, 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.

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.

Data value	Frequency	Relative frequency
2	3	$\frac{3}{20}$ or 0.15
3	5	$\frac{5}{20}$ or 0.25
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

Table $1.4.2$ Frequency Table of Student Work Hours with Relative Frequencies

The sum of the values in the relative frequency column of Table $1.4.2$ is $\frac{20}{20}$, or 1.

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 $1.4.3$.

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

Table $1.4.3$ Frequency Table of Student Work Hours with Relative and Cumulative Relative Frequencies

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

> **Note**
>
> 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.

Table 1.4.4 represents the heights, in inches, of a sample of 100 male semiprofessional soccer players.

Heights (inches)	Frequency	Relative frequency	Cumulative relative frequency
59.96–61.95	5	$\frac{5}{10} = 0.05$	0.05
61.96–63.95	3	$\frac{3}{100} = 0.03$	0.05 + 0.03 = 0.08
63.96–65.95	15	$\frac{15}{100} = 0.15$	0.08 + 0.15 = 0.23
65.96–67.95	40	$\frac{40}{100} = 0.40$	0.23 + 0.40 = 0.63
67.96–69.95	17	$\frac{17}{100} = 0.17$	0.63 + 0.17 = 0.80
69.96–71.95	12	$\frac{12}{100} = 0.12$	0.80 + 0.12 = 0.92
71.96–73.95	7	$\frac{7}{100} = 0.07$	0.92 + 0.07 = 0.99
73.96–75.95	1	$\frac{1}{100} = 0.01$	0.99 + 0.01 = 1.00
	Total = 100	**Total = 1.00**	

Table 1.4.4 Frequency Table of Soccer Player Height

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

- 59.96 to 61.95 inches
- 61.96 to 63.95 inches
- 63.96 to 65.95 inches
- 65.96 to 67.95 inches
- 67.96 to 69.95 inches
- 69.96 to 71.95 inches
- 71.96 to 73.95 inches
- 73.96 to 75.95 inches

In this sample, there are **five** players whose heights fall within the interval 59.96–61.95 inches, **three** players whose heights fall within the interval 61.96–63.95 inches, **15** players whose heights fall within the interval 63.96–65.95 inches, **40** players whose heights fall within the interval 65.96–67.95 inches, **17** players whose heights fall within the interval 67.96–69.95 inches, **12** players whose heights fall within the interval 69.96–71.95, **seven** players whose heights fall within the interval 71.96–73.95, and **one** player whose heights fall within the interval 73.96–75.95.

> **Example 1.4.1**
>
> From Table 1.4.4, find the percentage of heights that are up to 65.95 inches.

> **Exercise 1.4.1**
>
> Table 1.4.5 shows the amount, in inches, of annual rainfall in a sample of towns.

Table 1.4.5

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

Rainfall (inches)	Frequency	Relative frequency	Cumulative relative frequency
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	

From Table $1.4.5$, find the percentage of rainfall that is less than 9.01 inches.

Example $1.4.2$

From Table $1.4.4$, find the percentage of heights that fall between 61.96 and 65.95 inches.

Answer

Add the relative frequencies in the second and third rows: $0.03 + 0.15 = 0.18$or 18%.

Exercise $1.4.2$

From Table $1.4.5$, find the percentage of rainfall that is between 6.99 and 13.05 inches.

Example $1.4.3$

Use the heights of the 100 male semiprofessional soccer players in Table $1.4.4$. Fill in the blanks and check your answers.

1. The percentage of heights that are from 67.96 to 71.95 inches is: _____.
2. The percentage of heights that are from 67.96 to 73.95 inches is: _____.
3. The percentage of heights that are more than 65.95 inches is: _____.
4. The number of players in the sample who are between 61.96 and 71.95 inches tall is: _____.
5. What kind of data are the heights?
6. 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

1. 29%
2. 36%
3. 77%
4. 87
5. quantitative continuous
6. get rosters from each team and choose a simple random sample from each

Example $1.4.4$

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 $1.4.6$ was produced:

Table $1.4.6$ Frequency of Commuting Distances

Data	Frequency	Relative frequency	Cumulative relative frequency

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	$\frac{4}{19}$	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

1. Is the table correct? If it is not correct, what is wrong?
2. 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.
3. What fraction of the people surveyed commute five or seven miles?
4. 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

1. No. The frequency column sums to 18, not 19. Not all cumulative relative frequencies are correct.
2. 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.
3. $\frac{5}{19}$
4. $\frac{7}{19}, \frac{12}{19}, \frac{7}{19}$

Exercise 1.4.3

Table $1.4.5$ 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?

Example 1.4.5

Table $1.4.7$ contains the total number of deaths worldwide as a result of earthquakes for the period from 2000 to 2012.

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

Year	Total number of deaths
2007	712
2008	88,011
2009	1,790
2010	320,120
2011	21,953
2012	768
Total	823,856

Table 1.4.7

Answer the following questions.

1. What is the frequency of deaths measured from 2006 through 2009?
2. What percentage of deaths occurred after 2009?
3. What is the relative frequency of deaths that occurred in 2003 or earlier?
4. What is the percentage of deaths that occurred in 2004?
5. What kind of data are the numbers of deaths?
6. 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

1. 97,118 (11.8%)
2. 41.6%
3. 67,092/823,356 or 0.081 or 8.1 %
4. 27.8%
5. Quantitative discrete
6. Quantitative continuous

Exercise 1.4.4

Table 1.4.8 contains the total number of fatal motor vehicle traffic crashes in the United States for the period from 1994 to 2011.

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		

Table 1.4.8

Answer the following questions.

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