It is not necessary to reduce most fractions in this course. Especially in **Probability Topics**, the chapter on probability, it is more helpful to leave an answer as an unreduced fraction.

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

frequency table

DATA VALUE	FREQUENCY
2	3

Table 1.9 Frequency Table of Student Work Hours

how many times that value shows up

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

Table 1.9 Frequency Table of Student Work Hours

A **frequency** is the number of times a value of the data occurs. According to **Table 1.9**, 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 1.10 Frequency Table of Student Work Hours with Relative Frequencies

The sum of the values in the relative frequency column of **Table 1.10** 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 1.11**.

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

Table 1.11 Frequency Table of Student Work Hours with Relative and Cumulative Relative Frequencies

Example 1.17

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 1.14 was produced:

DATA	FREQUENCY	RELATIVE FREQUENCY	CUMULATIVE RELATIVE FREQUENCY
2	2		
3	1		
4	1		
5	3		
6	0		
7	2		
8	0		
9	0		
10	3		

Table 1.14 Frequency of Commuting Distances

a. Is the table correct? If it is not correct, what is wrong?

b. 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.

c. What fraction of the people surveyed commute five or seven miles?

d. 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)?

Solution 1.17

a. No. The frequency column sums to 18, not 19. Not all cumulative relative frequencies are correct.

b. 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.

c. $\frac{5}{19}$

d. $\frac{7}{19}, \frac{12}{19}, \frac{7}{19}$

Example 1.17

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 1.14 was produced:

Values must be in increasing order

frequency/n

relative frequency of the current row & all above

DATA	FREQUENCY	RELATIVE FREQUENCY	CUMULATIVE RELATIVE FREQUENCY	
2	2	2/19	2/19	
3	1	1/19	3/19	$\frac{2}{19}+\frac{1}{19}$
4	1	1/19	4/19	$\frac{2}{19}+\frac{1}{19}+\frac{1}{19}$
5	3	3/19	7/19	$\frac{2}{19}+\frac{1}{19}+\frac{1}{19}+\frac{3}{19}$
6	0	0	7/19	$\frac{2}{19}+\frac{1}{19}+\frac{1}{19}+\frac{3}{19}+0$
7	2	2/19	9/19	$\frac{2}{19}+\frac{1}{19}+\frac{1}{19}+\frac{3+0}{19}+\frac{2}{19}$
8	0	0	9/19	
9	0	0	9/19	
10	3	3/19	12/19	$\frac{2}{19}+\frac{1}{19}+\frac{1}{19}+\frac{3+0}{19}+\frac{2}{19}+0+0+\frac{3}{19}$
11	0	0	12/19	
12	2	2/19	14/19	
13	1	1/19	15/19	
14	0	0	15/19	
15	1	1/19	16/19	
16	0	0	16/19	
17	0	0	16/19	
18	2	2/19	18/19	
19	0	0	18/19	
20	1	1/19	$\boxed{19/19 = 1}$	
	+ 19	1		

Table 1.14 Frequency of Commuting Distances

sample size = n = 19

this last cumulative relative frequency always 1

Q What percentage of this sample commute between 14 and 20 miles?

Method 1

How many people: 4

Percentage of the whole sample: $4/19$

$= \boxed{21\%}$

Method 2

add the relative frequencies.

$$0+\frac{1}{19}+0+0+\frac{2}{19}+0+\frac{1}{19} = \frac{4}{19} = \boxed{21\%}$$

Method 3

Cumulative relative frequency of 20 is $19/19$

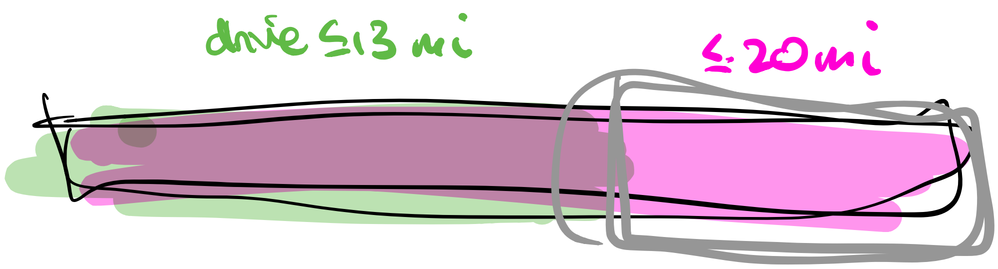
That means $\frac{19}{19} = 100\%$ is sum of all relative frequencies ≤ 20

Cumulative relative frequency of 13 is $15/19$
$= 79\%$

The percentage that drive ≤ 13 miles is $\frac{15}{19} = 79\%$

The percentage that drive ≤ 20 miles is $\frac{19}{19} = 100\%$

So the percentage that drive ≤ 20 miles but > 13 miles is...



$100\% - 79\%$

$= \boxed{21\%}$

why use crf (13) instead of 14?

cumulative relative frequency of 13

Answer: less than or equal to 20

but not less than or equal to 13

means

greater than 13 and less than or equal to 20

In math: $x \leq 20$ but x is not ≤ 13

so $x \leq 20$ and $x > 13$

so $13 < x \leq 20$

not including