

MATH 1501: Elements of Statistics with Applications

Problem Set 1: Cumulative Relative Frequency and the Shape of Data

Due: Wednesday, September 30, 2026, 11:59PM EST

Submit to me via Email: Gregory.Barco@brooklyn.cuny.edu

Directions

Submission: Do this homework on paper or type it up - there will be a graph, take clear photos of your completed work, place the photos or documents into a single document (Word, Google Doc, or similar), and email that document to me by the due date above.

For relative frequency and cumulative relative frequency, round to **three** decimal places — see the rounding guide below. For mean, median, and mode, report one decimal place: since our data are whole numbers, one decimal place is sufficient. Show your work for every calculation. Each question is followed by a link to the exact textbook page it is drawn from — if you get stuck, that is the first place to look.

Math Notation and Rounding

Reading tables and diagrams. Read every table, diagram, and example in this homework left to right, top to bottom, the same way you read a sentence. Later columns and boxes are built directly from the ones before them.

Interval notation. People who work with data describe a range of numbers using a bracket and a parenthesis together.

- A bracket, [or], means that endpoint **is** included.
- A parenthesis, (or), means that endpoint is **not** included.

For example, $[60, 70)$ means every value from 60 up to, but not including, 70: the value 60 belongs to this interval, but 70 belongs to the next one instead. The last interval in a table, such as $[90, 100]$, closes with a bracket on both ends so the largest value in the dataset is included.

How to Round to Three Decimal Places. To round a number to three decimal places (the nearest thousandth), look at the digit in the *fourth* decimal place — the one right after where you want to stop.

- If that digit is 0, 1, 2, 3, or 4, leave the third decimal digit as it is.
- If that digit is 5, 6, 7, 8, or 9, increase the third decimal digit by one.

Example. Round 0.348792 to three decimal places.

- Third decimal digit: 0.34**8**792.
- Fourth decimal digit, the one that decides the rounding: 0.348**7**92.
- 7 is 5 or higher, so the 8 rounds up to 9.

So 0.348792 rounds to **0.349**. This same rule applies no matter which decimal place you're asked to round to.

Cumulative Relative Frequency (CRF)

Definitions.

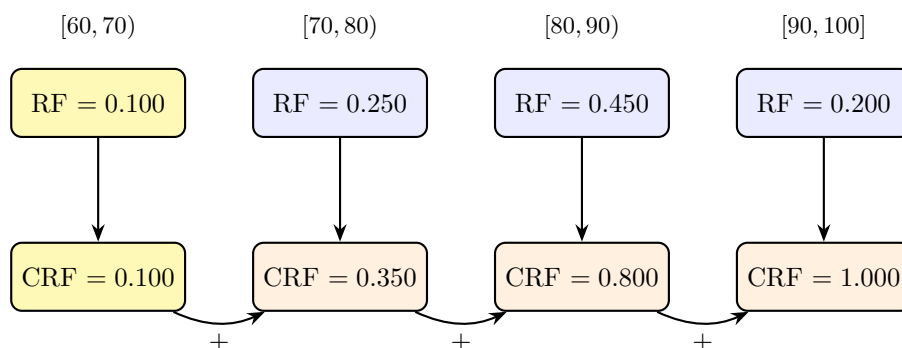
- **Relative frequency (RF)** — the fraction of the data that equals one particular value: how many observed data points fall inside an interval, divided by the total number of data values that you have.
- **Cumulative relative frequency (CRF)** — is a running sum/total of the relative frequencies. It is the fraction of the data that is at or below a given interval. It is calculated by summing up all of the relative frequencies - starting from the first interval up to the current interval. By the time you reach the last interval, the entire dataset's relative frequencies have been counted, so the cumulative relative frequency there always equals 1.

Example. Here is the same test-score frequency table from class [Slide 20/57](#), with two columns added (Relative Frequency and Cumulative Relative Frequency). Start at the top left of the table below and read it from left to right: for the very first interval, [60, 70), the relative frequency and cumulative relative frequency (highlighted below) are the same since the first interval's cumulative relative frequency is the running sum of all relative frequencies which happens to be a running sum of its self (you have nothing to add to the current relative frequency). The second interval, [70, 80)'s cumulative relative frequency is the sum of the first interval [60, 70)'s **RF** and the second interval [70, 80)'s **RF**.

Interval	Frequency	Relative Frequency	Cumulative Relative Frequency
[60, 70)	2	0.100	0.100
[70, 80)	5	0.250	0.350
[80, 90)	9	0.450	0.800
[90, 100]	4	0.200	1.000
Total	20	1.000	—

The relative frequency for [80, 90) is 0.450: 45% of the 20 scores fall in that interval alone. The cumulative relative frequency for that same row is 0.800. The same procedure applies: sum up the previous relative frequencies for all intervals until you get to the third interval's relative frequency. The cumulative relative frequency for [80, 90) means that 80% of all scores are 90 or below. The last row's cumulative relative frequency says that for scores [90, 100], 100% of the data observed has had a score/grade of 100 or below. The final **CRF** always equals 1.000 meaning that the whole dataset is accounted for by the time you reach the last interval.

Here is the same table again, drawn as a flow, so the accumulation itself is visible. Follow it left to right: each Cumulative Relative Frequency box is just the box before it, plus that row's own Relative Frequency box.



Because each relative frequency is rounded separately, the totals in Question 1 may come out to 0.999 or 1.001 instead of exactly 1.000. This comes from rounding, not from an error in your work.

1. Sixty-five randomly selected car salespersons were asked the number of cars they generally sell in one week. Fourteen people answered that they generally sell three cars; nineteen generally sell four cars; twelve generally sell five cars; nine generally sell six cars; eleven generally sell seven cars. Complete the table. (Source: OpenStax, Section 2.2, Question 12, p. 136)

Data Value (# cars)	Frequency	Relative Frequency	Cumulative Relative Frequency
3			
4			
5			
6			
7			
Total			—

2. What is the difference between cumulative relative frequency and relative frequency for each data value? (Source: OpenStax, Section 2.2, Question 16, p. 136)

3. To construct the histogram for the data in Question 1, determine appropriate minimum and maximum x and y values and the scaling. Sketch the histogram. Feel free to use online graphing software. Label the horizontal and vertical axes with words. Include numerical scaling. (Source: OpenStax, Section 2.2, Question 17, p. 136)



Measures of Center

Definitions.

- **Mean** — add every value together, then divide by how many values there are. This is the ordinary average, often written $\bar{x}$.
- **Median** — the middle value of a sorted list. Half the data lies below it, half above it.
- **Mode** — the value that appears most often. A dataset can have one mode, several modes, or no mode at all.
- **Range** — the largest value minus the smallest value. One number describing how spread out the whole dataset is.

Example. Sorted list: 3, 4, 6, 7, 9, 9, 9, 12, 15.

Mean: add all nine values and divide by 9.

$$\frac{3 + 4 + 6 + 7 + 9 + 9 + 9 + 12 + 15}{9} = \frac{74}{9} = 8.2$$

(rounded to one decimal place, since the data are whole numbers)

Median: count in from both ends until you reach the middle.

3, 4, 6, 7, 9, 9, 9, 12, 15

Nine values total, so the middle (5th) value is the median: median = 9.

Mode: look for the value that repeats the most.

3, 4, 6, 7, 9, 9, 9, 12, 15

9 appears three times, more than any other value: mode = 9.

Range: subtract the smallest value from the largest.

$$15 - 3 = 12$$

Summary: mean = 8.2, median = 9, mode = 9, range = 12. Mean below median and mode is the skewed-left pattern: a longer tail on the low end pulls the mean down, while the median and mode stay anchored where most of the data sits. The Shape of the Data section below asks you to apply this same reasoning to new data.

4. The following data show the lengths of boats moored in a marina. The data are ordered from smallest to largest:

16, 17, 19, 20, 20, 21, 23, 24, 25, 25, 25, 26, 26, 27, 27, 27, 28, 29, 30, 32, 33, 33, 34, 35, 37, 39, 40

(Source: *OpenStax, Section 2.5, Questions 43–45, p. 142*)

- (a) Calculate the mean.
 - (b) Identify the median.
 - (c) Identify the mode.
5. Using the same car salesperson data from Question 1 of this document (*OpenStax, Section 2.2, Question 12, p. 136*), find the following: (Source: *OpenStax, Section 2.5, Questions 46–48, p. 143*)
- (a) Calculate the sample mean.
 - (b) Identify the median.
 - (c) Identify the mode.

Shape of the Data

Definitions.

- **Symmetrical** — the data is balanced on both sides of its center. Mean, median, and mode are equivalent to one another, or very close to being equivalent.
- **Skewed left** — a longer tail extends toward the smaller values. The mean is pulled below (less than) the median and mode.
- **Skewed right** — a longer tail extends toward the larger values. The mean is pulled above (greater than) the median and mode.

The sorted list 3, 4, 6, 7, 9, 9, 9, 12, 15 from the Measures of Center section ([Slide 37/57](#)) is a skewed-left example: mean = 8.2 fell below (is less than) the median and the mode, both equal to 9. The same chain of reasoning identifies skewed-right data: the mean is greater than the median and mode. For symmetrical data, the mean, median, and mode are equivalent to one another. The following questions require you to consider the relationships between all three quantities.

6. State whether the data are symmetrical, skewed to the left, or skewed to the right. (Source: OpenStax, Section 2.6, Questions 49–51, p. 143)

(a) 1; 1; 1; 2; 2; 2; 2; 3; 3; 3; 3; 3; 3; 3; 4; 4; 4; 5; 5

(b) 16; 17; 19; 22; 22; 22; 22; 22; 23

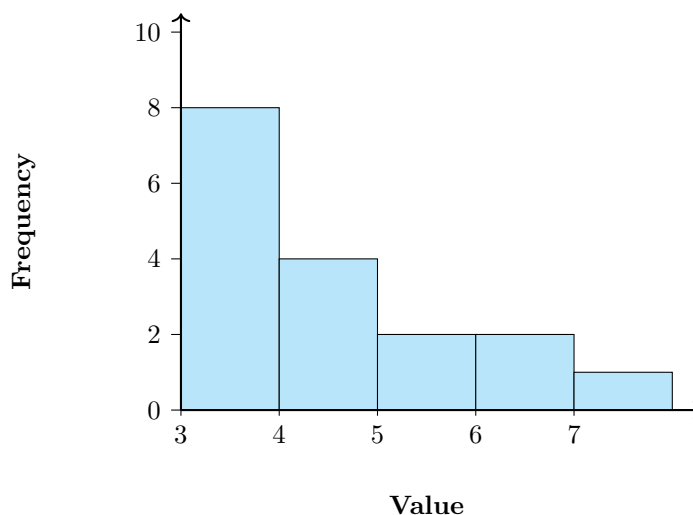
(c) 87; 87; 87; 87; 87; 88; 89; 89; 90; 91

7. Answer the following in your own words. (Source: OpenStax, Section 2.6, Questions 52–53, p. 143)

(a) When the data are skewed left, what is the typical relationship between the mean and median?

(b) When the data are symmetrical, what is the typical relationship between the mean and median?

8. The histogram below shows a distribution. (Source: OpenStax, Section 2.6, Question 55, p. 143)



(a) Describe the shape of this distribution.

(b) Based on that shape, what relationship would you expect between the mean and the median, without calculating either one?

9. In Question 5 you found the mean, median, and mode of the car salesperson data. Using the relationship you identified in Question 7, determine whether that distribution is skewed left, skewed right, or symmetrical. Explain, in your own words, how you can tell just from the three numbers, without looking at a picture of the histogram.