Liberian Mathematics Teacher Training Program 2023–2024

Andrew Obus[1]

CUNY Baruch College

October 20, 2023

[1] This program is partially supported by NSF CAREER Grant DMS-2047638

HW Exercises

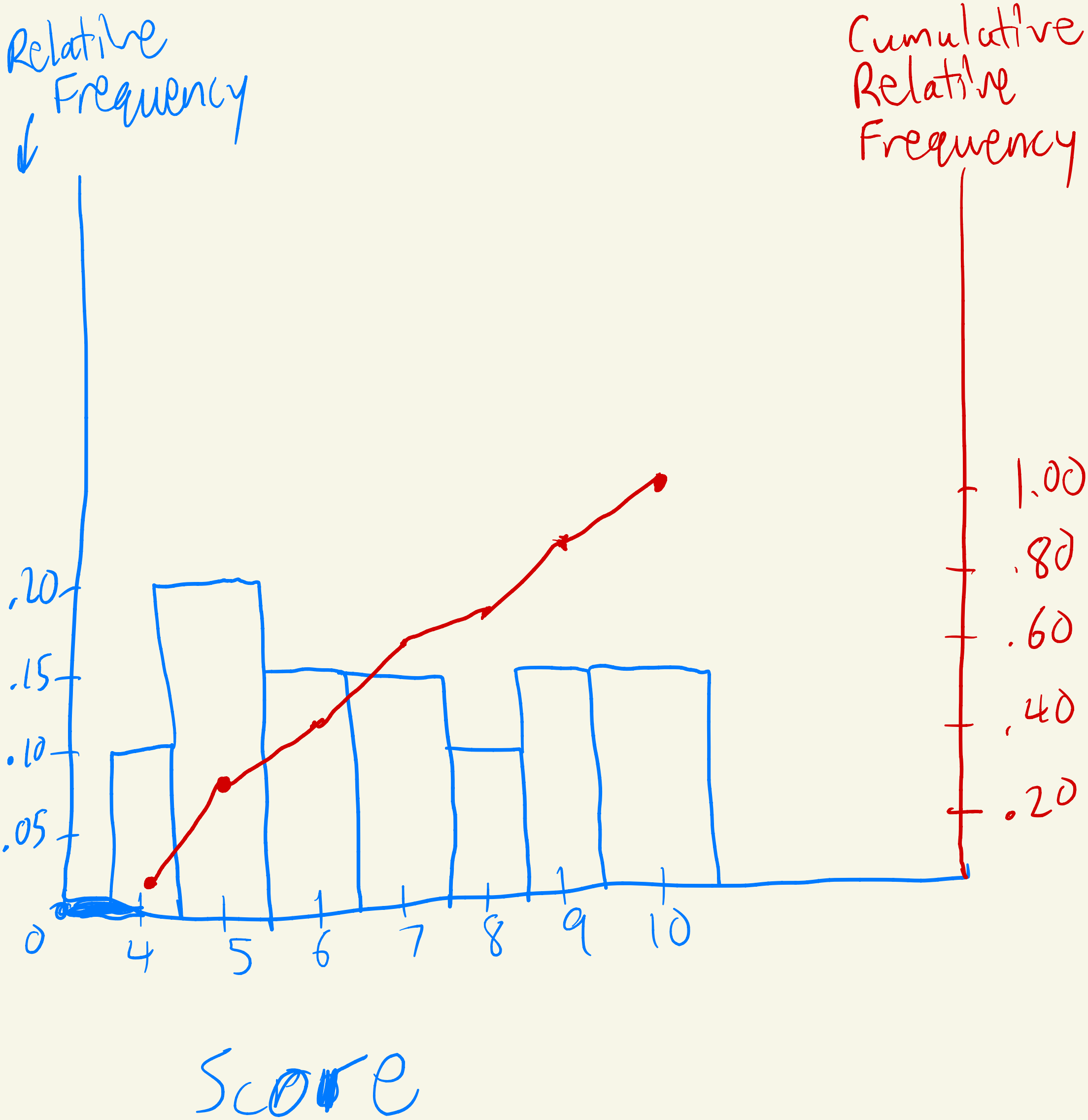
1. Using the frequency table from slide number 9 (the quiz scores), add columns for relative frequency and cumulative relative frequency.
2. Construct a histogram for this data, as well as a cumulative relative frequency plot.
3. Calculate the 60th and 70th percentiles for the data.

Example

Consider a class whose scores on a quiz (graded out of 10 points) are given by the following frequency table:

Score	Frequency	Relative Frequency	Cumulative Relative Frequency
4	2	.10 (2/20)	.10
5	4	.20 (4/20)	.30 (.10+.20)
6	3	.15 (3/20)	.45 (.30+.15)
7	3	.15 (3/20)	.60 (.45+.15)
8	2	.10 (2/20)	.70 (.60+.10)
9	3	.15 (3/20)	.85 (.70+.15)
10	3	.15 (3/20)	1.00 (.85+.15)

Total Number of Measurements: 2+4+3+3+2+3+3 = 20


Relative Frequency
Cumulative Relative Frequency
.20
.15
.10
.05
0
1.00
.80
.60
.40
.20
4
5
6
7
8
9
10
Score

60th Percentile

$N = 20$

$P = 60$

$R = \frac{60}{100}(20+1) = \frac{3}{5} \cdot 21 = \frac{63}{5} = 12.6$

Data: 4, 4, 5, 5, 5, 5, 6, 6, 6, 7, 7, 7, 8, 8, 9, 9, 9, 10, 10, 10

$x = 12^{th}$ entry $= 7$ $\qquad F = .6$

$y = 13^{th}$ entry $= 8$

60th %ile: $x + F(y-x) = 7 + .6(8-7)$

$= 7 + .6 \cdot 1$

$= \boxed{7.6}$

70th %ile: $P = 70$.

$R = \frac{70}{100}(20+1) = \frac{7}{10} \cdot 21 = 14.7$

$F = .7 \qquad x = 8,\ y = 9 \qquad 8 + .7(9-8) = \boxed{8.7}$

Review of Percentile Calculation

- We will use what is called the "interpolation method"
- Suppose we have a data set with N entries, and we want to calculate the Pth percentile. 20
- First we arrange the data in ascending order.
- Then, we calculate the *rank* R of the Pth percentile. This is done with the formula
$$R = \frac{P}{100}(N+1).$$
- If R is an integer, the Pth percentile is simply the Rth data entry.
- If R is not an integer, we look at the Rth and the $R+1$st data entries, call these x and y. rounded down.
- Let F be the fractional part of R. Then the P percentile is
$$x + F(y-x).$$

Percentiles and cumulative relative frequency

- Percentiles are related to cumulative relative frequency.
- In particular, if the cumulative relative frequency of a class is p, then the $100p$th percentile of the data should be roughly the maximum of that class.
- Let's try this with the second example from the homework. We will consider the cumulative relative frequency .6, corresponding to the 60th percentile.
- Without doing any calculation, what do we expect to be a good approximation to the 60th percentile?

Mean, median, and mode - Measures of Central Tendency

Suppose we have a finite set of data entries $x_1, x_2, \ldots, x_n$.

- The *(arithmetic) mean* or *average* of the data is given by

$$\frac{x_1 + x_2 + \cdots + x_n}{n}$$

- The *median* is the data entry that occurs in the middle of the sequence of data, if they are arranged from lowest to highest. If n is even, then the median is the average of the two middle entries.
- In fact, this is exactly the same as the 50th percentile of the data using our interpolation method.
- The *mode* of the data is the entry that occurs most often. If there are multiple such entries, they are all considered the mode. If every entry occurs equally, we say there is **no mode.**

All of the terms above are called "measures of central tendency". The idea is that they are all reasonable numbers to represent what a "standard" instance of the data is likely to be.

Example 1

Suppose my data set consists of $\{1, 2, 3, 4, 5, 6\}$

Mean: $\dfrac{1+2+3+4+5+6}{6} = \dfrac{21}{6} = 3.5.$

Median: $\dfrac{3+4}{2} = \dfrac{7}{2} = 3.5$

Mode: No Mode. (OR Mode is 1, 2, 3, 4, 5, and 6)

Example 2

Now, suppose my data set consists of $\{1, 2, 3, 4, 5, 600\}$

Mean: $\frac{1+2+3+4+5+600}{6} = \frac{615}{6} = 102.5$

Median: $\frac{3+4}{2} = \frac{7}{2} = 3.5$

Mode: No Mode.

Observation: Mean is much bigger than the median.
This happens when you have large outliers.
Median is less sensitive to outliers than the mean is.

Even # of entries: 2 in the middle.
Odd # of entries: 1 in the middle
e.g. 1, 3, 5, 7, 9

Example problem 1

Can you come up with a data distribution with median 10 but mean 3?

$1+2+1+10+1=15$

$\frac{1+2+1+10+1}{5}=3$ = Mean.

Doesn't work.

1, 2, 10, 1, 1

1, 1, 1, 2, 10 → Median = 1.

Try Again!

$\boxed{-11, 10, 10}$

median = 10.

Mean = $\frac{-11+10+10}{3} = \frac{9}{3} = 3$

Try to have only 3 data entries. For the mean to be 3, need the sum to be 9.

Example problem 2

25 students have a combined average of 70 on a test, whereas another 20 students have a combined average of 80. Find the total average for all 45 students.

Idea: If we can find the total sum of the scores of all 45 students, then we can answer the question.

Sum of first 25 students: Average $= 70 = \frac{\text{sum}}{25}$

$\rightsquigarrow$ sum $= 70 \cdot 25 = 1750$.

$80 \cdot 20 = 1600$

Sum of next 20 students:

Total sum $= 1750 + 1600 = 3350$.

Average: $\frac{3350}{45} = \boxed{74.\overline{4}}$

Example problem 3

The arithmetic mean of 12 scores is 82. When the highest and lowest scores are removed, the new mean becomes 84. If the highest score is 98, what is the lowest score?

Sum of total scores: $12 \times 82 = 984$.

After removing highest and lowest, new sum $= 10 \times 84 = 840$.

So highest + lowest $= 984 - 840 = 144$.

Highest $= 98$. So lowest $= 144 - 98 = \boxed{46}$

Other measures of central tendency

Two other measures of central tendency are the *geometric mean* and the *harmonic mean*. These are used only when the data entries are *positive*.

- Geometric mean: $\sqrt[n]{x_1 x_2 \cdots x_n}$
- Harmonic mean: $\frac{1}{1/x_1 + 1/x_2 + \cdots + 1/x_n}$

One always has $HM \le GM \le AM$, with equality only when all the data entries are equal. We will prove $GM \le AM$ for the case of two entries:

NEXT WEEK

Homework Exercises

1. A basketball player scores an average of 18.6 points per game for his first 5 games. How many points must he score in the sixth game to raise his average to 20 points per game?
2. Can you construct a data set with mean 10, median 20, and mode 30? → Need negative numbers.
3. ~~Two numbers~~ $\cancel{x}$ ~~and~~ $\cancel{y}$ ~~have a geometric mean of 12 and an arithmetic mean of 12.5. Find~~ $\cancel{x^2 + y^2}$~~.~~

Next week: More on measures of central tendency.
Begin measures of dispersion.
Meet at 4:30.