Liberian Mathematics Teacher Training Program 2023–2024

Andrew Obus[1]

CUNY Baruch College

October 27, 2023

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

HW Exercise 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?

5 games. $\frac{x}{5} = \frac{18.6}{1}$ (x = total score in 1st 5 games).

$\rightsquigarrow x = 5 \cdot 18.6 = 93.$ ← total points in 1st 5 games.

6 games: Total score $= 20 \times 6 = 120$ ← total points needed in 6 games.

So in Game 6, needs to score

$120 - 93 = \boxed{27}$ points.

HW Exercise 2

Can you construct a data set with mean 10, median 20, and mode 30?

Example: $\boxed{-20, -10, 20, 30, 30}$

Strategy: Start with 20 in middle.
Add in 2 30's so mode is 30.
Add in 2 #'s below 20 to make total sum 50, so mean is 10.

Mode = 30
Median = 20
Mean $= \frac{-20-10+20+30+30}{5}$
$= 10.$ ✓

Recall: Mean, median, and mode

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.
- 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.

Exact averages from frequency tables: Example

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

Score	Frequency
4	2
5	4
6	3
7	3
8	2
9	3
10	3

Could write:

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

↳ calculate average.

To calculate the mean of this data, we need to use a *weighted average*.

Calculate total # of data entries: $2+4+3+3+2+3+3=20$.

Add up each data entry, weighted by its frequency, divide by total.

$$(4\times 2+5\times 4+6\times 3+7\times 3+8\times 2+9\times 3+10\times 3)/20=\boxed{7}$$

Approximate averages from frequency tables

- Suppose we want to calculate a mean from a frequency table where each row represents a class range. (or "group data")
- There is not one correct way to do this, but a reasonable way to proceed is to use a weighted average as in the previous example, but to use the midpoint of each class as the input into the weighted average.

Approximate averages: Example

Let us calculate the approximate mean for the data from the following frequency table:

Time	Frequency
0 to $<$ 5	5200
5 to $<$ 10	18200
10 to $<$ 15	19600
15 to $<$ 20	15400
20 to $<$ 25	13800
25 to $<$ 30	5700
30 to $<$ 35	10200
35 to $<$ 40	2000
40 to $<$ 45	2000

$\frac{0+5}{2} = 2.5$

$\frac{5+10}{2} = 7.5$

$\frac{10+15}{2} = 12.5$ ...

midpoints.

Total # of entries:

$5200+18200+ \ldots + 2000+2000$

$= 92100$

Midpoints: $2.5, 7.5, 12.5, \ldots$

$\ldots, 37.5, 42.5.$

Weighted average:

$$\Big(2.5\times5200 + 7.5\times18200 + 12.5\times19600 + 17.5\times15400 + 22.5\times13800 + 27.5\times5700 + 32.5\times10200 + 37.5\times2000 + 42.5\times2000\Big)/92100$$

$$\approx \boxed{17.62}$$

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*. (GM)

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

One always has $HM \leq GM \leq AM$, with equality only when all the data entries are equal.

An example of harmonic means

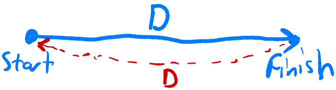
RATE (SPEED)

Suppose a car drives a certain distance at (30 kilometers per hour) and returns at (60 kilometers per hour). What is the car's average speed for the trip?

RATE (SPEED)

in km

Let D = one-way distance.



$$\text{Average speed} = \frac{\text{Total Distance}}{\text{Total time}} = \frac{2D}{\text{time out} + \text{time back}}$$

$$= \frac{2\cancel{D}}{\frac{\cancel{D}}{30} + \frac{\cancel{D}}{60}} = \frac{2}{\frac{1}{30} + \frac{1}{60}}.$$

This is the definition of the harmonic mean of 30 and 60.

$$\frac{2}{\frac{1}{30}+\frac{1}{60}} = \frac{2}{\frac{2}{60}+\frac{1}{60}} = \frac{2}{\frac{3}{60}} = 2 \cdot \frac{60}{3} = \frac{120}{3} = \boxed{40 \text{ km/h}}$$

A car drives at 30 km/h for 3 hours and then 60 km/h for 2 hours. What is its average speed?

$$\text{Average Speed} = \frac{\text{Total distance}}{\text{Total time}}$$

$$= \frac{30\times3 + 60\times2}{3+2} = \frac{90+120}{5} = 42$$

$$\boxed{42\ \text{km/h}}$$

Weighted average of 30 and 60 with weights 3 and 2.

Proof that GM $\leq$ AM for two data points

We will prove $GM \leq AM$ for the case of two data points:

Data: x_1, x_2.

Ex: $x_1 = 4 \quad x_2 = 9$

$GM = \sqrt{4 \cdot 9} = \sqrt{36} = 6$

$AM = \frac{4+9}{2} = 6.5$.

GM: $\sqrt{x_1 x_2}$ AM: $\frac{x_1+x_2}{2}$

Want $\sqrt{x_1 x_2} \leq \frac{x_1+x_2}{2}$. Equivalent to show

$$\left(\sqrt{x_1 x_2}\right)^2 \leq \left(\frac{x_1+x_2}{2}\right)^2, \text{ or } \quad x_1 x_2 \leq \frac{x_1^2+2x_1x_2+x_2^2}{4}$$

Equivalent to show: $4x_1x_2 \leq x_1^2+2x_1x_2+x_2^2$

$$\Leftrightarrow 0 \leq x_1^2 - 2x_1x_2 + x_2^2 = (x_1-x_2)^2$$

TRUE because squares are always ≥ 0. $\square$

Homework Exercises

1. Two numbers x and y have a geometric mean of 12 and an arithmetic mean of 12.5. Find $x^2 + y^2$. ← challenging.
2. Calculate the mean absolute deviation of the data set $1, 2, 3, 4, 5, 6$.
3. What happens to the mean absolute deviation of a data set if all the entries are increased by 5? What about if all the entries are multiplied by 5?

$\boxed{4:30}$

Thank you for your attention. Next week we will discuss standard deviation and box and whisker plots, other measures of dispersion, and median from group data.