Unit 6: One-Variable Data Analysis

Measures of Central Tendency

Greek "mu"

Mean:
(commonly referred to as "average)

Population Mean $$\mu = \frac{x_1 + x_2 + ... + x_N}{N}$$

$\bar{x}$

Sample Mean $$\bar{x} = \frac{x_1 + x_2 + ... + x_n}{n}$$

"x bar"

where N is the population size, and n is the sample size.

These calculations are identical, but different symbols are used to distinguish between a population or sample.

Nov 13-8:13 PM

Recall: Sigma notation, which can be used to express a mathematical series.

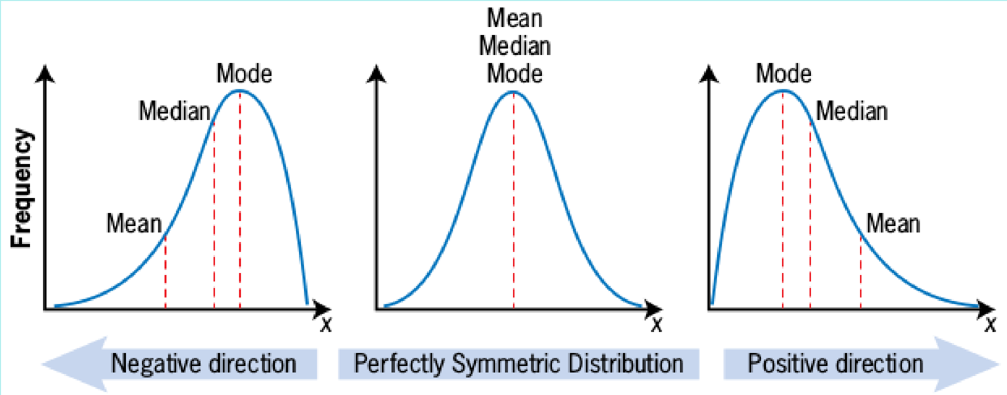
$$\sum_{i=1}^{4} i = 1 + 2 + 3 + 4$$

$$\underbrace{\sum_{i=1}^{4} x_i = x_1 + x_2 + x_3 + x_4}$$

	1	2	3	4
x	5	7	13	-4

Nov 13-8:07 PM

Outliers are data that don't fit well with the other data in the sample. Outliers can skew the distribution of data.



Mean is most affected by outliers.

Mode is least affected by outliers.

Nov 13-8:19 PM

Median:

The middle value from a sorted list of data. If the middle is actually between two values, which occurs for an even number of data points, take the midpoint of those two values.

Mode:

The most frequently occurring value(s) in the data. A frequency table can be useful in determining mode.

Nov 14-9:50 AM

Weighted Mean:

In some data sets, certain data is more important than other data. It can be useful to assign a weight to each datum.

$$\bar{x} = \bar{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} = \frac{\sum w_i x_i}{\sum w_i}$$

where 'w' is the weight given to each data value, 'x'.

Nov 13-8:25 PM

Mean of Grouped Data (frequency table):

Similar to the weighted mean, we use the midpoint of each category, along with the frequency of each category, to calculate the overall mean for the sample.

$$\bar{x} = \frac{\sum f_i m_i}{\sum f_i}$$

f_i is the frequency of each interval

m_i is the midpoint of each interval

Nov 14-9:40 AM

Assigned Work:

Pg 263 #1 - 8, 13, 15, 16

A group of children were asked how many hours a day they spend playing video games. The table shows the data.

a) Determine the estimated mean, median number of hours, and modal interval for the above distribution.

b) Discuss any skewing of the data with respect to the measures of central tendency.

Number of Hours	Number of Children
0–2	3
2–4	11
4–6	7
6–8	2
8–10	1

$$\bar{x} = \frac{\sum f_i m_i}{\sum f_i}$$

$$= \frac{3(1) + 11(3) + 7(5) + 2(7) + 1(9)}{3 + 11 + 7 + 2 + 1}$$

Nov 14-9:54 AM

5. The mean of Daniel's marks on five tests was 77.4. His marks on the first four tests were 88, 77, 70, and 72. Calculate Daniel's mark on the fifth test.

$$\bar{x} = \frac{x_1 + x_2 + x_3 + x_4 + x_5}{5}$$

$$77.4 = \frac{88 + 77 + 70 + 72 + x_5}{5}$$

$$(77.4)(5) = 307 + x_5$$

Nov 15-12:35 PM

13. Karen's term mark is 82%. The term counts for 70% of the final mark. What mark must Karen achieve on the exam to earn a final mark of

a) 80%? **b)** 85%?

c) at least 75%?

d) Can Karen achieve 88%? Explain.

$$\bar{x} = \frac{\sum w_i x_i}{\sum w_i}$$

(a) $80 = \frac{(0.7)(82) + (0.3)x_2}{0.7 + 0.3}$

(b) $85 = (\text{ same })$

(c) at least 75%

$75 = (\text{ Same })$

(d) $88 = (\quad)$

Nov 15-12:36 PM

15. The table shows student absences from Lakeside High School during the first semester. Assume that the absences located exactly on the endpoints of an interval were placed in the lower interval.

MP	Student Absences	Number of Students	
1.5	0–3	47	47
4.5	3–6	89	136
7.5	6–9	33	169
⋮	9–12	102	
	12–15	24	
	15–18	19	
	18–21	6	
	21–24	8	
	24–27	0	
	27–30	2	

me: [0, 3) [3, 6) ⋮ [27, 30)

book: [0, 3] (3, 6] (27, 30]

a) Calculate the estimated mean, median, and modal interval of student absences.

b) Does there appear to be an outlier? If so, how does it affect the mean and median of the data set?

(a) $\bar{x} = \frac{\sum f_i M_i}{\sum f_i}$

$= \frac{47(1.5) + 89(4.5) + \ldots + 2(28.5)}{47 + 89 + \ldots + 2}$

$\sum f_i = 330 \rightarrow$ # of students

$\rightarrow$ halfway student at 165

found student #165 and #166 in (6-9] group

∴ median is 7.5

2.5

Nov 15-12:35 PM