MEASURES OF CENTRALITY AND LOCATION

OUTLINE

1. Measures of Central Tendency for Interval Data

a. Mean
   1. Mean for Ungrouped Data
   2. Mean for Grouped Data

a. Median
   1. Median for Ungrouped Data
   2. Median for Grouped Data

a. Mode
   1. Mode for Ungrouped Data
   2. Mode for Grouped Data

2. Percentile as a Measure of Location

a. Percentile (Centile)
   1. Percentile for Ungrouped Data
   2. Percentile for Grouped

3. Comparison of the Three Measures

MEASURES OF CENTRAL TENDENCY FOR INTERVAL DATA

Mean

- Mean for Ungrouped Data

The mean is another term for arithmetic average. If you have already computed an "average", you have computed a mean. Suppose you have six scores: 12, 10, 18, 16, 20, 14. If you let $X_1 = 12$, $X_2 = 10$, $X_3 = 18$, $X_4 = 16$, $X_5 = 20$, and $X_6 = 14$, the mean as represented by $\bar{X}$ (read "bar x") is:

$$X = \frac{X_1 + X_2 + X_3 + X_4 + X_5 + X_6}{N}$$

$$= \frac{12 + 10 + 18 + 16 + 20 + 14}{6}$$

$$= \frac{90}{6}$$

$$= 15$$

Instead of writing the equation for the mean as shown above you can shorten it to:

$$\overline{X} = \frac{\Sigma X}{N}$$

Where:

$\overline{X}$ = the mean

ΣX = the sum of all scores

N = the total number of cases

Example: Find the mean of 10 + 20 + 30 + 40 + 50

Solution:

$\Sigma X = 10 + 20 + 30 + 40 + 50 = 150$

$N = 5$

$$\overline{X} = \frac{\Sigma X}{N}$$

$$= \frac{150}{5}$$

$$= 30$$

MEAN FOR GROUPED DATA

Suppose that you have a list of ratings of 50 students in a statistics class as shown in Table 4.1, wherein, in this school, grades are quantified by making an A equal to 6, B equal to 5, C equal to 4, D equal to 3, and E equal to 2 as shown in column 1 of the table. Your task is to get the average of the grades of the students. The values in column 4 are the product of the row values in each of columns 2 and 3, that is 6 X 10 = 60. Then column 4 is summed up to apply the given formula which is:

$$\overline{X} = \frac{\Sigma f M}{N}$$

Where:

$\overline{X}$ = is the mean

M = the midpoint

fM = the product of the frequency and each midpoint

N = the total number of cases or scores

TABLE 4-1

Class interval	(M)	f	fM
(A) 6	6	10	60
(B) 5	5	13	65
(C) 4	4	12	48
(D) 3	3	5	15
(E) 2	2	10	20
		N= 150	= 208

SOLUTION:

The computation on the left may be used when the class interval size is equal to 1.

When the interval size is greater than 1, the method used is given as follows.

$$\overline{X} = \frac{\Sigma f M}{N}$$

$= 208/50$

$= 4.16$

Class interval	f	(M)	fM
25 – 29	3	27	81
30 – 34	2	32	64
35 – 39	5	37	185
40 – 44	8	42	336
45 – 49	8	47	376
50 – 54	8	52	416

Class interval	f	(M)	fM
55 – 59	9	57	513
60 – 64	6	62	372
65 – 69	6	67	402
70 – 74	3	72	216
75 – 79	3	77	231
80 – 84	3	82	246

$$\overline{X} = \frac{\Sigma fM}{N}$$

$$= \frac{3438}{64}$$

$$= 53.72$$

The foregoing computation has been made easy following the steps below.

1. Determine the midpoint of each class interval.
2. Get the product of each midpoint and the corresponding frequency within its interval to obtain ΣfM.
3. Apply the formula by substituting the values ΣfM and N.

MEDIAN

Median for Ungrouped Data

The median of ungrouped data is the centermost score in a distribution. The formula in finding the median is:

$$\text{Mdn} = \frac{\frac{X_N}{2} + \frac{X_N+2}{2}}{2}$$

and

$$\text{Mdn} = \frac{X_1 + N}{2}$$

Example: Find the median of the following sets of scores:

A	B
12	18
15	22
19	31
21	12
6	3
4	9
2	11
	8

Solution: First, you arrange the scores from the highest to the lowest, or from the lowest to the highest.

A	B
2 – X_1	3 – X_1
4 – X_2	8 – X_2
6 – X_3	9 – X_3
10 – X_4	11 – X_4
15 – X_5	12 – X_5
19 – X_6	18 – X_6
21 – X_7	22 – X_7
	31 – X_8

For Group a, N = 7 (odd)

Mdn =

=

=

$= X_4$

= 12

For Group B, N = 8 (even)

Mdn =

=

=

$= \frac{X_4 + \frac{X_{10}}{2}}{2}$

$= \frac{X_4 + X_5}{2}$

$= \frac{X_9}{2}$

$= X_{4.5}$

The answer is 11.5 because $X_{4.5}$ is between 11 and 12. This is computed by adding 11 and 12, and dividing the sum of 23 by 2.

Median for Grouped Data

Example: Find the median of the following distribution

Class interval	F	Cum f<
25 – 29	3	3
30 – 34	2	5
35 – 39	5	10
40 – 44	8	18
45 – 49	8	26
50 – 54	8	34
55 – 59	9	43
60 – 64	6	49
65 – 69	6	55
70 – 74	3	58
75 – 79	3	61
80 – 84	3	64
	N = 64	

The formula for finding the median of grouped data is given as follows:

$$\text{Mdn} = \text{L} + \frac{\frac{N}{2} - F}{fm}\ (i)$$

Where:

L = exact lower limit of interval containing the median class

F = the sum of all the frequencies below L

fm = frequency interval containing the median class

N = total number of cases

i = class interval

The median class is the class interval where N/2 interval is found.

To solve for the median of the data above, the following steps are used.

1. Compute the cumulative frequencies less than.
2. Find N/2, one half the total number of cases
3. Locate the class interval in which the middle class falls, and determine the exact limits of this interval.
4. Apply the formula by substituting the given rules.

In our example,

$L = 49.5$

$F = 26$

$fm = 8$

$N = 64$

$i = 5$

$$Mdn = 49.5 + \frac{\frac{64}{2} - 26}{8}(5)$$

$$= 49.5 + \frac{32 - 26}{8}(5)$$

$$= 49.5 + \frac{6}{8}(5)$$

$$= 49.5 + \frac{30}{8}$$

$$= 49.5 + 3.75$$

$$= 53.25$$

MODE

Mode for Ungrouped Data

For ungrouped data, the mode is defined as that datum value or specific score which has the highest frequency. The most frequently occurring score is the mode.

Example: Find the mode of the following data.

X = 9,9,11,11,11,13,13,13,14,15,15,15,17,17,17,17,17,18

By inspection, the mode (Mo) is 17 because it appears the most number of times.

Mode for Grouped Data

For grouped data, the mode is defined us the midpoint of the interval containing the largest number of cases.

Example: Find the mode for the data below.

Class interval	f
25 – 29	3
30 – 34	2
35 – 39	5
40 – 44	8
45 – 49	8
50 – 54	8
55 – 59	9
60 – 64	6
65 – 69	6
70 – 74	3
75 – 79	3
80 – 84	3
	N = 64

The mode (Mo) is *57* because the *55 – 59* has the largest frequency and *57* is the midpoint of this class interval.

MEASURE OF CENTRAL TENDENCY FOR NOMINAL AND ORDINAL DATA

For nominal data the measure of centrality is just the mode. The median is used for ordinal data just as the mean is used for interval data. Therefore, to solve for the modal category is to determine the class having the largest frequency.

Example: For the following courses offered at a certain college, what is the modal category?

Courses Offered	Number of Students
Education	3,000
AB	1,000
BS	700
Graduate Courses	300

Answer: The modal category is Education, because it contains the biggest number of population.

For ranked data otherwise known as ordinal, the most appropriate measure is the median. However, the mode can also be used to express centrality for ordinal data.

Example: In a frequency distribution of attitudes toward the nuclear plant in Bataan, what is its median class? The modal class?

Attitude Toward Nuclear Plant	f	Cf (cumulative frequency)
Strongly favorable	1	1
Somewhat favorable	2	3
Slightly favorable	3	6
Slightly unfavorable	5	11
Somewhat unfavorable	13	24
Strongly unfavorable	20	44
	N=44	

SOLUTION:

$$\text{Mdn} = \frac{\frac{X_{44}}{2} + \frac{X_{44}+2}{2}}{2}$$

$$= \frac{X_{22} + X_{23}}{2}$$

$$= \frac{X_{45}}{2}$$

$$= X_{22.5}$$

Answer: The median class of attitude is somewhat unfavorable, since the 22.5th score lies in this category.

The modal class is strongly unfavorable because it contains the highest frequency of 22.

PERCENTILE AS A MEASURE OF LOCATION

Percentile (Centile)

This is an important measure which divides the distribution into one hundred parts. It is also called the centile. Other measures similar to the percentile are the quartile and decile. The quartile measure, divides the distribution into four parts; the decile into ten parts.

Percentile for Ungrouped Data

If you are asked to find P_{50}, having an N of 500, simply multiply N by 50 and divide it by 100.

Solution:

$$P_{50} = \frac{50\,(500)}{100}$$

$$= \frac{25{,}000}{100}$$

$$= 250$$

Note: Two hundred fifty is the number obtained after arranging the scores from highest to lowest or vice versa. If 100 is the 250^{th} number, this number 100 is your P_{50}.

The illustration above can easily be followed by using the given formula below:

$$P = \frac{N}{100}$$

If you are solving for quartile, the formula to use is Q = N/4, and D = N/10 for decile.

Example: Solve P_{60} for the following list of scores resulting from a statistics examination administered to 50 students.

48	85	91	54	62	72	68	70	79	90	59	43	52	61
94	98	62	76	99	64	78	49	88	73	51	69	100	93
89	68	98	66	96	55	77	57	61	70	92	46	73	83
91	79	53	62	59	82	93							

Solution: First arrange the scores from the lowest to the highest as indicated below:

43	48	68	78	91
46	59	68	79	92
48	60	69	79	93
49	61	70	82	93
51	61	70	83	94
52	62	72	85	96
53	62	73	88	98
54	62	73	89	98
55	64	76	90	99
57	66	77	91	100

$P_{60} = 60(N)/100$

$= 60(50)/100$

$= 3000/100$

$=$ the 30th score

The 30th score is *77*, which is your $P_{60.}$

Percentile for Grouped Data

The computation of the percentile for grouped data is the same as the computation of the median.

The formula is as follows:

$$P_1 = {}^{L}P_1 + \frac{\frac{N}{100} - {}^{F}P_i}{{}^{f}P_1}(i)$$

Solution:

$$P_{50} = \frac{50\ (N)}{100}$$

$= 50\ (100)/\ 100$

$= 5000/100$

$= 50^{TH}$ item

$$P_{50} = L_{50} + \frac{50(N)/100 - F_{50}}{{}^{f}P_{50}}$$

$$= 44.5 + \frac{50 - 44\ (5)}{20}$$

$$= 44.5 + \frac{6}{20}\ (5)$$

$= 44.5 + 33/20$

$= 44.5 + 1.5$

$= 46$

What is the percentile rank of 42? In answering this question, the formula to apply is as follows:

$$\text{Percentile rank for } X_1 = \frac{F_1 + [(X - L_1)/i]\ [{}^{f}X_1]}{N}\ (100)$$

Solution: When $X_1 = 42$

$$42 = \frac{26 + [42 - 39.5/5]\,[18]}{100}(100)$$

$$= \frac{26 + (2.5/5)\,(18)}{100}\,(100)$$

$$= \frac{26 + 45/5}{100}(100)$$

$$= \frac{26 + 9}{100}(100)$$

$$= \frac{35\,(100)}{100}$$

$$= 3500/100$$

$$= 35$$

COMPARISON OF THE THREE MEASURES

When the three measures are compared with regard to the level of measurement used, the *mean* is used for interval data, the *median* is used for ordinal or interval data, and the *mode* is used for nominal, ordinal or interval data.

If the three measures are compared with regard to the shape or form of their distribution, the *mean* is most appropriate for unimodal symmetrical distribution; the *median* is most appropriate for highly skewed distribution; and the *mode* is most appropriate for bimodal distribution.

Finally, regarding the objective or purpose of the computation, the *mean* is used when the purpose is to consider the value of each score and when further or more advanced statistical computation is needed. The *median* is used when the center score is needed to avoid the influence of extreme values. It can also be used for more advanced statistical operations. The *mode* is used when you want a fast approximation of central tendency or when you want the most frequently occurring score.

EXERCISE SET

1. What are the measures of central tendency for interval data? For nominal data? And ordinal data?
2. Is the underground formula better than the grouped formula when you compute the three measures of central tendency and measures of location? Why?
3. Is percentile the same in meaning as quartile and decile? Why?
4. What is the best measure of centrality? Why?

Example: Solve P_{50} , D6, Median and Q3 in the following distribution of scores a

Class interval	f	cf
20 – 24	3	3
25 – 29	6	9
30 – 34	7	16
35 – 39	10	26
40 – 44	18	44
45 – 49	20	64
50 – 54	17	81
55 – 59	10	91
60 - 64	5	96
65 - 69	4	100
	N = 100	

EXERCISE SET

1. What are the measures of central tendency for interval data? For nominal data? And ordinal data?
2. Is the underground formula better than the grouped formula when you compute the three measures of central tendency and measures of location? Why?
3. Is percentile the same in meaning as quartile and decile? Why?
4. What is the best measure of centrality? Why?

For the set of raw data, find the value of P_{48} and the percentile rank of 126 using the ungrouped data formula.

98	100	104	108	109	110	110	113	117	118
118	119	120	122	122	122	123	126	127	127
128	128	129	129	130	130	131	132	133	136
138	139	141	141	145	147	148	159	155	160