PROBABILIY AND STATISTICS I

LECTURE SIX

Measures of central tendency (2)

Lecturer: Dr. Emily Roche

INTRODUCTION

This lecture will focus on mode as measure of central tendency together with other averages.

Intended learning outcomes

At the end of this lecture, you will be able to compute mode for both simple and grouped data and also compute and apply other averages in real-life scenarios.

References

These lecture notes should be supplemented with relevant topics from the book listed in the Bibliography at the end of the lecture.

MODE

This is the item of data with the highest frequency. It is the most typical or fashionable value of a distribution. It is the value around which the items tend to be most heavily concentrated. Mode is not a unique measure since a distributions can be unimodal (only one mode), bimodal, trimodal or multimodal (two, three or more mode respectively).

Calculation of mode

For individual observations, the mode will be determined by counting the number of times the various values repeat themselves. For example for the data set $10, 27, 24, 10, 27, 15, 30, 15, 27, 18, 27$ the mode is 27.

For grouped data (continuous series) the mode is calculated by applying the formula:

$$M_0 = L_0 + \left(\frac{f_0 - f_1}{2f_0 - f_1 - f_2}\right) h_0$$

Where

M_o − is the estimated mode

L_0 − is the lower class boundary of the modals class

f_0 − is the frequency of the modal class

f_1 − is the pre-modal class frequency

f_2 − is the post-modal class frequency

h_0 − is the modal class interval/size.

Example 1:

Find the mode for the following data

No of absences	$0-4$	$5-9$	$10-14$	$15-19$	$20-24$	$25-29$	$30-34$
No of branches	16	21	12	11	10	8	2

The modal class is $5-9$

$$M_0 = 4.5 + \left(\frac{21-16}{2(21)-16-12}\right) 5 = 6.3$$

Example 2:

Find the mode for the following data

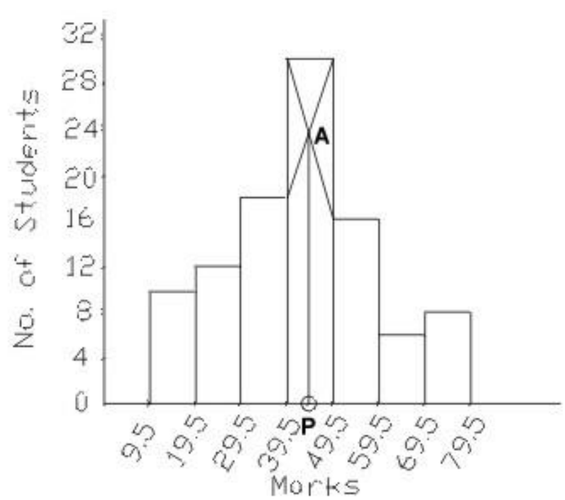
No of km	5.5 – 10.5	10.5 – 15.5	15.5 – 20.5	20.5 – 25.5	25.5 – 30.5	30.5 – 35.5	35.5 – 40.5
No of runners	1	2	3	5	4	3	2

$$M_0 = 20.5 + \left(\frac{5-3}{2(5)-3-4}\right)5 = 21.2$$

Mode can be estimated graphically using a histogram.

To estimate the mode graphically:

1. Construct a histogram
2. Draw two lines diagonally on the internal vertices of the modal class bar to the upper corner of the adjacent bars on either sides.
3. Drop a perpendicular line from the intersection of the two diagonal lines to the horizontal axis which gives us the modal value. In the histogram below the reading at P gives the modal value.



Merits:

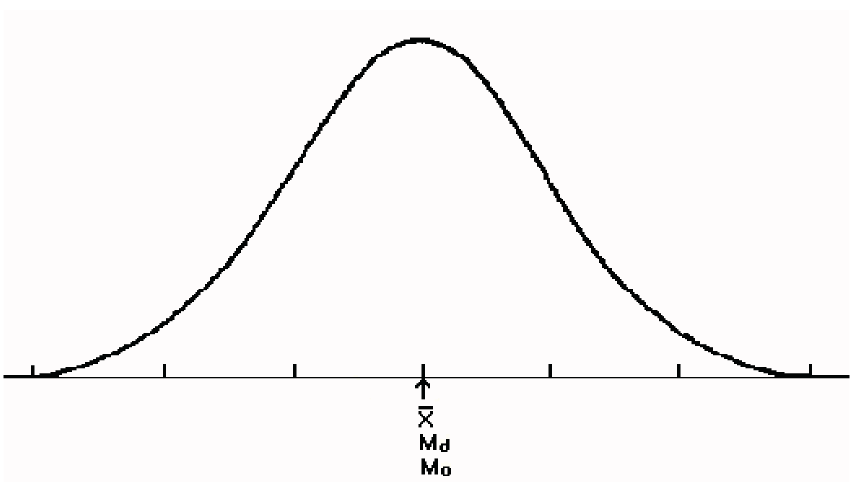
1. It is not affected by extreme values.
2. It can be estimated for open ended distributions.
3. Can be estimated graphically.
4. Can be used to describe qualitative data particularly nominal data.

Limitations:

1. Mode is not a unique measure.
2. It does not lend itself to further algebraic manipulation.
3. It is not based on every item of data in the series.
4. It is not rigidly defined. There are several formulae for calculating the mode all of which usually give somewhat different answers.

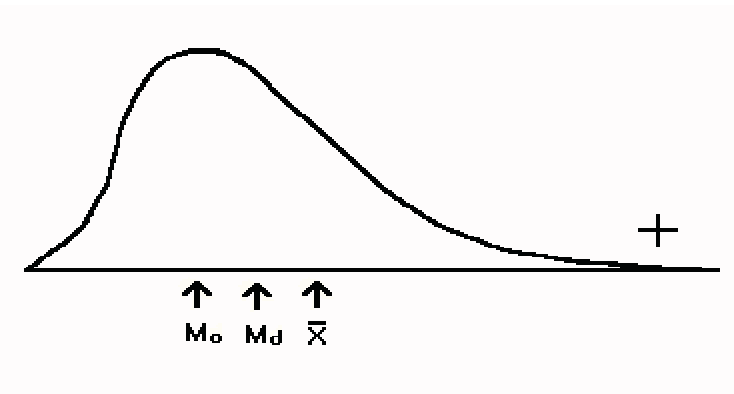
RELATIONSHIP BETWEEN MEAN MODE AND MEDIAN:

In a symmetrical distribution the values of mean, mode and median coincide.

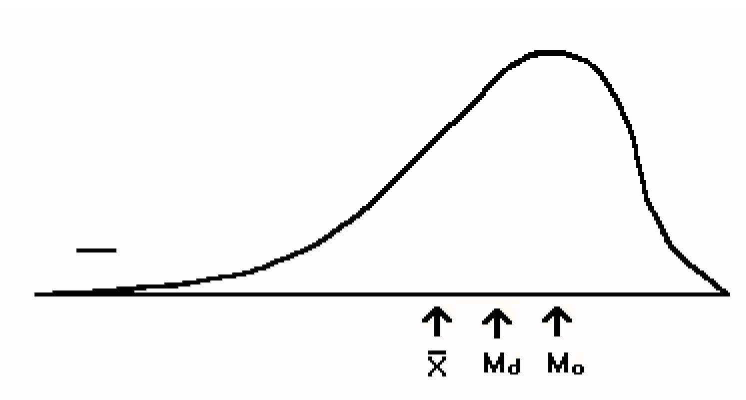


If the mean, mode and median are not equal the distribution is said to be asymmetrical or skewed

A distribution that is skewed to the right or positively skewed has the mean higher than the median.



A distribution that is negatively skewed has the mean is lower than the median



The mean is the balance point of the distribution. Because points further away from the balance point change the center of balance, the mean is pulled in the direction the distribution is skewed. For example, if the distribution is positively skewed, the mean would be pulled in the direction of the skewness, or be pulled toward larger numbers.

One way to remember the order of the mean, median, and mode in a skewed distribution is to remember that the mean is pulled in the direction of the extreme scores. In a positively skewed distribution, the extreme scores are larger, thus the mean is larger than the median.

Summary of when to use mean, median and mode

Type of Variable	Best measure of central tendency
Nominal	Mode
Ordinal	Median
Interval/Ratio (not skewed)	Mean
Interval/Ratio (skewed)	Median

OTHER TYPES OF 'AVERAGES':

a. The Weighted Arithmetic mean

This is applicable in situations where the relative importance of the different items is not the same.

$\bar{x}_w = \frac{\sum wx}{\sum w}$, where $\bar{x}_w$ represents the weighted arithmetic mean, w the weights and x the variable.

Example:

A university student obtained the following grades in a trimester exam. Calculate the corresponding GPA (Grade Point Average);

Course	Calculus I	Linear Algebra	Basic Mathematics	Programming I	Programming II	Assembly language	Communication skills
Credits, w	3	3	2	3	3	3	2
Grade, x	$B-$	A	$A-$	$B+$	B	$B+$	$C+$

Note: A = 12 pts A - = 11pts, B + =10 ……..E =0

Weighted mean

$$\bar{x}_w = \frac{\sum wx}{\sum w} = \frac{3(8)+3(12)+2(11)+3(10)+3(9)+3(10)+2(7)}{3+3+2+3+6+8+2} = \frac{183}{27} = 6.78$$

Exercise:

a. A lecturer grades CATS 20%, term paper 30% and final exam 50%. A student has 83, 72 and 90 respectively for the CATS, term paper and final exam. Find the students final average. (Answer 83.2)

b. Another lecturer gives 4 one hour exams and one final exam which counts as two one hour exams. Find a student's grade if the student received 62, 83, 97 and 90 on the one hour exams and 82 on the final exam. (Ans 82.7)

b. Trimmed mean

A trimmed mean is calculated by discarding a certain percentage of the lowest and the highest scores and then computing the mean of the remaining scores. For example, a mean trimmed 50% is computed by discarding the lower and higher 25% of the scores and taking the mean of the remaining scores.

The median is the mean trimmed 100% and the arithmetic mean is the mean trimmed 0%.

These are commonly used in Olympic scoring to minimize the effects of extreme ratings possibly caused by biased judges.

c. **The Geometric Mean**

This is the nth root of the product of n items of data.

$G.M. = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdot \ldots \cdot x_n}$ where $x_1, \; x_2, \; x_3, \; \ldots, \; x_n$ are the various items

For more than tree data items, these calculations are simplified using logarithms as:

$$\log GM = \log(x_1 \cdot x_2 \cdot x_3 \cdot \ldots \cdot x_n)^{\frac{1}{n}} = \frac{1}{n}(logx_1 + \log x_2 + logx_3 + \cdots + logx_n)$$

$$GM = anti\log \frac{logx_1 + \log x_2 + logx_3 + \cdots + logx_n}{n}$$

Uses of GM

1. Finding average percent increase in sales, production, population or other economic or business series.
2. Construction of index numbers
3. It is the most suitable average when large weights have to be given to small items and small weights to large items

d. **Harmonic Mean**

It is the reciprocal of the arithmetic mean of the reciprocal of the individual observations.

$$HM = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \cdots + \frac{1}{x_n}} = \frac{n}{\sum\left(\frac{f}{x}\right)}$$

Uses

It is useful in situations where the average of rates is required for example finding average speed.

The weighted harmonic mean is used if the items do not have the same weight. For instance, when calculating average speed and distances covered at the different speeds differ.

Example: - If a car covers the first $30km$ at a speed of $40km/h$, the next $45km$ at $65\ km/h$ and the last $20km$ at $45\ km/h$, the average speed will be $\frac{30+45+20}{\left(\frac{1}{40}\times 30\right)+\left(\frac{1}{65}\times 45\right)+\left(\frac{1}{45}\times 20\right)}$ Therefore, $HM_w = \frac{\sum weights}{\left(\frac{1}{x_1}\times w_1\right)+\left(\frac{1}{x_2}\times w_2\right)+\cdots+\left(\frac{1}{x_n}\times w_n\right)}$

Bibliography

Gupta, SP (Dr.), (2014). *Statistical methods* (43rd Ed.). Sultan Chand & Sons.

S. C. Gupta and V. K. Kapoor, (2020). *Fundamentals of mathematical Statistics* (12th Ed). Sultan Chand & Sons.