|  |  |  |
| --- | --- | --- |
| Related image | **KONERU LAKSHMAIAH EDUCATION FOUNDATION**  (Deemed to be University estd, u/s, 3 of the UGC Act, 1956) (NAAC Accredited “A++” Grade University)  Green Fields, Guntur District, A.P., India – 522502  **Department of Computer Science and Engineering**  (DST - FIST Sponsored Department) |  |

**B.Tech. II CSE(H) PROGRAM**

**A.Y. 2023-24 ODD, Semester-II**

**Course Code: 22MT2005**

**PROBABILITY, STATISTICS AND QUEUING THEORY**

**Course Outcome-2**

**Session 15: MEASURES OF CENTRAL TENDENCY**

1. **Course Description (Description about the subject)**

Measures of central tendency are summary statistics that describe the "center" of a data set. They are used to identify the most typical value in a set of data. The three most common measures of central tendency are Mean, Median and Mode in a data set.

1. **Aim**

To Explain different measures of central tendency

1. **Instructional** **Objectives (Course Objectives)**

Identify the relationship between variables using correlation and regression techniques

1. **Learning** **Outcomes (Course Outcome)**

**CO2**: Students will be able to Explain different measures of central tendency, and dispersion

1. **Module** **Description** **(CO-2 Description)**

Mean, median and mode.

1. **Session** **Introduction**

Measures of central tendency are summary statistics that describe the "center" of a data set. They are used to identify the most typical value in a set of data. The choice of which measure of central tendency to use depends on the shape of the data distribution and the purpose of the analysis. For example, the mean is often the most appropriate measure of central tendency when the data is normally distributed. However, the median may be a better choice when the data is skewed.

1. **Session description**

Measures of central tendency, i.e. condensing the mass of data in one single value, enable us to get an idea of the entire data.

For example, it is impossible to remember the individual incomes of millions of earning people of India. But if the average income is obtained, we get one single value that represents the entire population.

Measures of central tendency also enable us to compare two or more sets of data to facilitate comparison.

For example, the average sales figures of April may be compared with the sales figures of previous months.

**Properties of a Good Measure of Central Tendency**

[1] It should be easy to understand and calculate.

[2] It should be rigidly defined.

[3] It should be based on all observations.

[4] It should be least affected by sampling fluctuation.

[5] It should be capable of further algebraic treatment.

[6] It should be least affected by extreme values.

[7] It should be calculated in case of open end interval.

Followingaresomeoftheimportantmeasuresofcentraltendencywhicharecommonlyusedinbusinessandindustry.

• Mean

• Median

• Mode

The arithmetic mean (or mean or average) is the most commonly used and readily understood measure of central tendency. In statistics, the term average refers to any of the measures of central tendency.

**MEAN**

**Ungrouped data/Raw data**

The arithmetic mean is defined as being equal to the sum of the numerical values of each and every observation divided by the total number of observations. Symbolically, it can be represented as:

Let xi is the variable which takes values x1, x2, x3, ………, xn­ over ‘n’ items, then arithmetic mean is given by 

**Mean**

**Advantages:**

➢ It is simple and easy to compute.

➢ It is rigidly defined.

➢ It is amenable for algebraic manipulations.

➢ It is based on all observations in the series.

➢ It helps for direct comparison.

**Disadvantages:**

➢ It is unduly affected by extreme items.

➢ It is sometimes un-realistic.

➢ It may leads to confusion.

➢ Suitable only for quantitative data (for variables).

➢ It cannot be located by graphical method or by observations.

**MEDIAN**

Median is the value of the middle item of a series arranged in ascending or descending order of magnitude. Hence it is the “Middle

The number is that value of the variable which divides most” or “Most central” value of a set of number. It divide the series into two equal parts, one part containing values greater and the other with values less than the median.the group into two equal parts, one part comprising all values greater and the other, all values less than median.

**Merits of Median**

* It is easy to compute and understand
* It eliminates the effect of extreme item
* The value of median can be located graphically

**Demerits of Median**

* The calculating media, it is necessary to arrange the data other averages do not need an arrangement.
* It is affected more by fluctuations of sampling than the arithmetic mean.
* It’s not based on all the items in the series

**Individual** **Series**

**Case (I) : Arrange the data either ascending or descending order**

If the number of observations is odd, then the median is the *middle value* after the observations have been arranged in ascending or descending order of magnitude. For example, the median of 5 observations 35, 12, 40, 8, 60*i.e.,*8, 12, 35, 40, 60, is 35.

In case of even number of observations median is obtained as the arithmetic mean of the two middle observations after they are arranged in ascending or descending order of magnitude. Thus, if one more observation, say, 50 is added to the above five observations then the six observations in ascending order of magnitude are : 8, 12, 35, 40, 50, 60. Thus,

Median=Arithmetic mean of two middle terms = 

2

**Remark.** It should be clearly understood that in case over numbered observations, in fact, any value lying between the two middle values can serve as a median but it is a convention to estimate median by taking the arithmetic mean of the two middle values.

**Case (II) : Frequency Distribution.** In case of frequency distribution where the variable takes the values *X*1, *X*2,…, *Xn* with respective frequencies *f*1, *f*2,…, *fn* with ∑*f* = *N*, total frequency, median is the size of the (*N* + 1)/2th item or observation. In this case the use of cumulative frequency (*c.f.*) distribution facilitates the calculations.

The steps involved are :

1. Prepare the ‘*less than*’ cumulative frequency(*c.f.*) distribution.
2. Find *N*/2.
3. See the *c.f., just greater than N*/2.
4. The corresponding value of the variable gives median.

**Case (III) : Continuous Frequency Distribution.**

As before, median is the size (value) of the (*N*+ 1)/2thobservation.Stepsinvolved forits computation are :

1. Prepare ‘*less than*’ cumulative frequency(*c.f.*) distribution.
2. Find *N*/2.
3. See *c.f.* just greater than *N*/2.
4. Thecorrespondingclasscontainsthemedianvalueandiscalledthe*medianclass*.The value of median is now obtained by using the interpolation formula:



Where *l* is the lower limit of the median class, *f* is the frequency of the median class, his the magnitude or width of median class,

 is the total frequency,

And *C* is the cumulative frequency of the class *precedin*g the median class.

**Remarks1.** The interpolation formula(5·13)is based on the following assumptions:

* 1. The distribution of the variable under consideration is continuous with exclusive type classes without any gaps.
  2. There is an orderly and even distribution of observations within each class.

However, if the data are given as a grouped frequency distribution where classes are not continuous, then it must be converted into a continuous frequency distribution before applying the formula. This adjustment will affect only the value ofin (5·13).

1. Median will be abbreviated by the symbol *Md.*

**Mode**

Mode is the modal value in the value of the variable which occurs more number of times or most frequently is a distribution. Mode is the value which occurs with the greatest number of frequency in a series.

**Types** **of** **modal**

1. **Uni-model**

If there is only one mode in series is called uni-model

1. **Bi**-**Modal**

If there are two modes in the series, it is called bi-model

1. **Tri**-**Modal**

If they are three modes in the series, it is Relationship between different Averages Symmetrical is called Tri-model

**IV. Multimodal**

If there are more than three modes in the series it is called multi-mode.

**Relationship among mean, median and mode**

The three averages are identical, when the distribution is symmetrical. In an asymmetrical distribution, the values of mean, median and mode are not equal.

Median = 1/3 (Mean-Mode)

Mode = 2(Median-Mode)

Median=Mode\* 2/3(Mean –Mode)

**Advantages**:

* Its result will not be affected by extreme values and open end classes
* If data are not grouped, it can be determined easily

**Disadvantages**:

* It has to be supplemented by other statistics
* It is difficult to obtain an accurate estimate of the mode if the values are classified into a frequency distribution.

1. **Activities/ Case studies/related to the session.**

**NA**

1. **Examples & contemporary extracts of articles/ practices to convey the idea of the Session**

**Example 1:** Marks obtained by 65 students are given below: 20, 15, 23, 22, 25, 20. Calculate the mean.

**Solution:** Mean marks 



= 20.83

**Individual Series:**

Find out mean from the following data.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Roll No | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Marks | 21 | 30 | 28 | 40 | 26 | 34 | 40 | 9 | 15 | 17 |

## **Solution**

|  |  |
| --- | --- |
| Roll No | Marks(X) |
| 1  2  3  4  5  6  7  8  9  10 | 21  30  28  40  26  34  40  9  15  17 |
| N=10 | ∑X=300 |

**Mean =** =30

## **Discrete Series**

**Calculate the arithmetic mean for the wages of workers in Factory.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Wages in Rs. | 4 | 6 | 8 | 10 | 15 | 16 |
| Workers | 5 | 15 | 6 | 7 | 8 | 2 |

**Solution:**

|  |  |  |
| --- | --- | --- |
| Wages in Rs. | Workers f | fx |
| 4 | 5 | 4x5=20 |
| 6 | 15 | 6x15=96 |
| 8 | 6 | 8x6=48 |
| 10 | 7 | 10x7=70 |
| 15 | 8 | 15x8=120 |
| 16 | 2 | 16x2=32 |
|  | N=∑f=43 | ∑fx=380 |

∑fx=380, N=43

Average =380/43=8.837

The average wage of workers=Rs.8.84

**Continuous Series:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Class**  **Intervals** | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 |
| **Frequency** | 6 | 5 | 8 | 15 | 7 |

|  |  |  |  |
| --- | --- | --- | --- |
| Class Intervals | Mid-point m | Frequency f | fm |
| 0-10 | 5 | 6 | 30 |
| 10-20 | 15 | 5 | 75 |
| 20-30 | 25 | 8 | 200 |
| 30-40 | 35 | 15 | 525 |
| 40-50 | 45 | 7 | 315 |
|  |  | N=∑f=41 | N=∑fm=1145 |

Arithmetic Mean =∑fm /N

The Arithmetic mean=27.92

**Example 2: Eight coins were tossed together and the number of heads(X)resulting was noted. The operation was repeated256times and the frequency distribution of the number of heads is given below:**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **No. of heads (X):** | **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** |
| **Frequency (f) :** | **1** | **9** | **26** | **59** | **72** | **52** | **29** | **7** | **1** |

**Calculate Median.**

**Solution:**

|  |  |  |
| --- | --- | --- |
| *X* | *f* | *Less than c.f.* |
| 0 | 1 | 1 |
| 1 | 9 | 1 + 9 = 10 |
| 2 | 26 | 10 + 26 = 36 |
| 3 | 59 | 36 + 59 = 95 |
| 4 | 72 | 95 + 72 = 167 |
| 5 | 52 | 167 + 52 = 219 |
| 6 | 29 | 219 + 29 = 248 |
| 7 | 7 | 248 + 7 = 255 |
| 8 | 1 | 255 + 1 = 256 |

Here *N*=∑*f* = 256, N/2=128

**—**

The cumulative frequency (*c.f.*) just greater than 128is 167 and the value ofX corresponding to 167 is 4. Hence, median number of heads is 4.

**Example 3 : Calculate the median from the following data**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Marks | 0-20 | 20-40 | 40-60 | 60-80 | 80-100 |
| No of Students | 5 | 15 | 30 | 8 | 2 |

Solution:

|  |  |  |
| --- | --- | --- |
| Marks | No of students | Cumulative frequency |
| 0-20  20-40  40-60  60-80  80-100 | 5  15  30  8  2 | 5  20  50  58  60 |

By Applying the formula Median =46.676

**Individual** **Series**

Calculate the mode from the following data of the marks obtain by 10 students

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Serial No | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Marks  obtained | 60 | 77 | 74 | 62 | 77 | 77 | 70 | 68 | 65 | 80 |

**Solution**

Marks obtained by 10 students is here 77 is repeated three times There for the Mode mark is 77

**Discrete Series:**

Calculate the mode form the following data of the wages of workers of are establishment. Find the modal wages

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Daily Wages in Rs | 3 | 4 | 6 | 7 | 9 | 10 | 12 | 13 | 15 |
| No of wage  earners | 2 | 3 | 2 | 6 | 10 | 11 | 12 | 5 | 1 |

**Solution :**

**Grouping Table:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Daily Wagesis  Rs. | Frequency of Wage Earners | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 |
| 3 | 2 | 5 |  | 7 |  |  |
| 4 | 3 | 5 | 11 |  |
| 6 | 2 | 8 | 18 |
| 7 | 6 | 16 | **27** |
| 9 | 10 | **21** | **33** |
| 10 | 11 | **23** | **28** |
| 12 | **12** | 17 | 18 |
| 13 | 5 | 6 |  |
| 15 | 1 |  |  |  |

**Analysis Table :**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Column** | **Size of Item** | | | | | | |
|  | **4** | **6** | **7** | **9** | **10** | **12** | **15** |
| **1** |  |  |  |  |  | **I** |  |
| **2** |  |  |  | **I** | **I** | **I** |  |
| **3** |  |  |  |  | **I** | **I** |  |
| **4** |  |  | **I** | **I** | **I** |  |  |
| **5** |  |  |  | **I** | **I** | **I** |  |
| **6** |  |  |  |  | **I** | **I** | **I** |
|  |  |  | **I** | **3** | **5** | **4** | **1** |

From the analysis table it is known that size10 has been repeated the maximum number of times, thus is, so the modal wages Rs 10

**Continuous series**

**Find out the mode from the** **following series**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **X** | **0-5** | **5-10** | **10-15** | **15-20** | **20-25** | **25-30** | **30-35** |
| **frequency** | **1** | **2** | **5** | **14** | **10** | **9** | **2** |

**Grouping Table:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| X | Frequency | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 |
| 0-5 | 1 | 3 |  | 8 |  |  |
| 5-10 | 2 | 7 | 21 |  |
| 10-15 | 5 | 19 | 29 |
| 15-20 | 14 | 24 | **33** |
| 20-25 | 10 | **19** | **21** |
| 25-30 | 9 | **11** |  |
| 30-35 | **2** |  |  |

**Analysis Table:**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Column** | **Size of Item** | | | | | | |
|  | **0-5** | **5-10** | **10-15** | **15-20** | **20-25** | **25-30** | **30-35** |
| **1** |  |  |  | **I** |  |  |  |
| **2** |  |  | **I** | **I** | **I** |  |  |
| **3** |  |  |  | **I** | **I** | **I** |  |
| **4** |  |  |  | **I** | **I** | **I** |  |
| **5** |  | **I** | **I** | **I** | **I** | **I** | **I** |
| **6** |  |  | **I** |  |  |  |  |
|  |  | **1** | **3** | **6** | **5** | **3** | **1** |

**Modal value lies in 15-20 as it occurs most frequently.**

**Mode (z) =**

Mode (z) =18.46

1. **SAQ's-Self Assessment Questions**
2. **Which measure of central tendency includes the magnitude of scores?**
   * 1. Mean
     2. Mode
     3. Median
     4. Range

**Answer: a**

1. **Which of the following is not a disadvantage of using mean?**
   * 1. It is affected by extreme values
     2. It cannot be computed in grouped data with open-ended class intervals
     3. It does not possess the desired algebraic property
     4. None of the above

**Answer: c**

1. **The two methods of finding mode in a discrete series are \_\_\_\_\_\_\_\_.**
   * 1. Grouping method and ascending method
     2. Table method and midpoint method
     3. Grouping method and inspecting method
     4. None of the above

**Answer: c**

1. **To calculate the median, all the items of a series have to be arranged in a/an \_\_\_\_\_\_\_\_.**
   * 1. Descending order
     2. Ascending order
     3. Ascending or descending order
     4. None of the above

**Answer: c**

1. **Mode refers to the value within a series that occurs \_\_\_\_\_\_\_\_ number of times.**
   * 1. Maximum
     2. Minimum
     3. Zero
     4. Infinite

**Answer: a**

1. **Summary**

The choice of which measure of central tendency to use depends on the shape of the data distribution and the purpose of the analysis. Measures of central tendency are a valuable tool for describing data sets. They can be used to compare different data sets, to identify outliers, and to make predictions about future data. The mean, median, and mode are all measures of central tendency, but they each have different strengths and weaknesses. The mean is the most sensitive to outliers, the median is the most resistant to outliers, and the mode is the least affected by the shape of the distribution. The mean, median, and mode can all be used to compare different data sets. However, it is important to use the same measure of central tendency for all data sets being compared.

1. **Terminal Questions**
2. Explain the difference between mean, median, and mode.
3. What are the advantages and disadvantages of using mean, median, and mode?
4. A random sample of ten students is taken from the student body of a college and their GPAs are recorded as follows:

1.90, 3.00, 2.53, 3.71, 2.12, 1.76, 2.71, 1.39, 4.00, 3.33

Find the mean.

1. Compute the sample median for the following data:

1.39 1.76 1.90 2.12 2.53 2.71 3.00 3.33 3.71 4.00

1. Compute the mode of the following data:

0, 0, 0, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 3, 3, 3, 4

1. **Case Studies (CO Wise)**

**NA**

1. **Answer Key**

**NA**

1. **Glossary**

**NA**

1. **References of books, sites, links Text Books:**

**Textbooks:**

1. Probability and Statistics Rukmangad Achari E. and E. Keshava Reddy
2. Probability and Statistics for Engineers and Scientists” Ronald E. Walpole, Sharon L. Myers and Keying Ye 8th Edition Pearson pub
3. Probability & Statistics for Engineers Dr. J. Ravichandran first Edition Wiley-India

**Reference books:**

1. Hossein Pishro-Nik, Introduction to Probability, Statistics, and Random Processes, 2014, by Kappa Research LLC; ISBN-13: 978-0990637202

**Web Resources**

1. https://ncert.nic.in/textbook.php?kemh1=0- 16
2. https://ncert.nic.in/textbook.php?jemh1=ps-15
3. **Keywords**

Mean, median, mode, central tendency, Interval, continuous data, discrete data.