Applied Statistics

Course Code: MAT1011

Dr. Sukanta Nayak

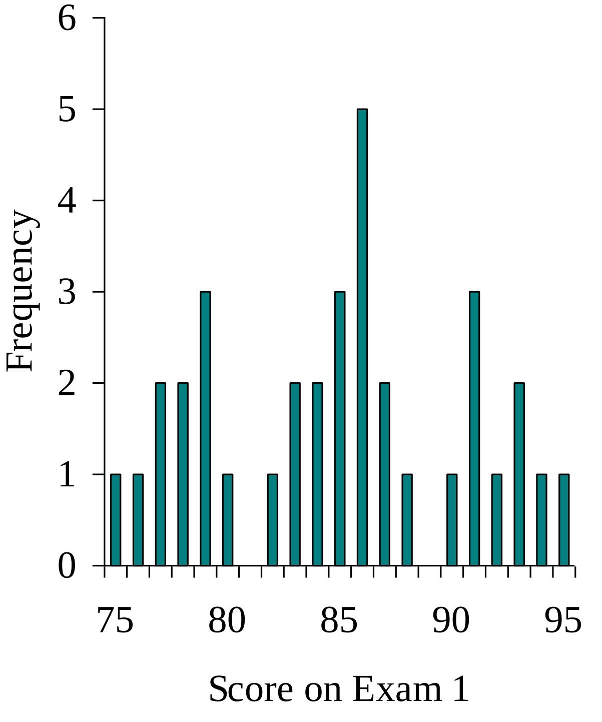
Department of Mathematics
School of Advanced Sciences
VIT-AP University, Amaravati
Andhra Pradesh

Measures of Central Tendency

- A *measure of central tendency* is a descriptive statistic that describes the average, or typical value of a set of scores
- There are three common measures of central tendency:
  - the mode
  - the median
  - the mean

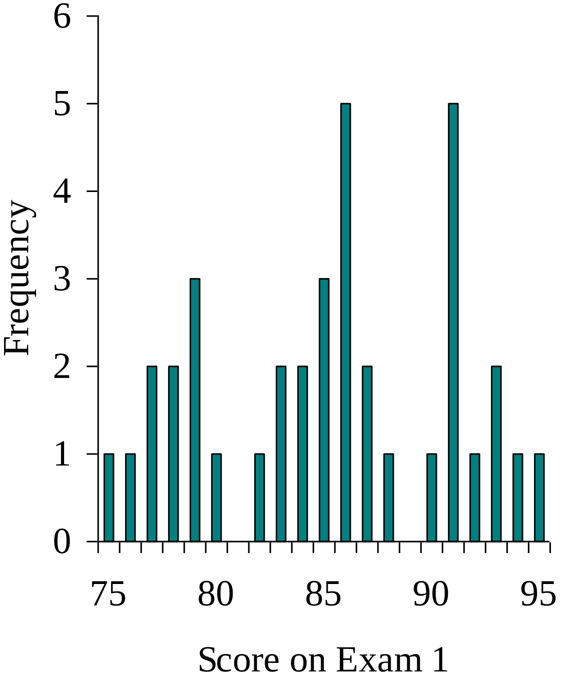
The Mode

- The *mode* is the score that occurs most frequently in a set of data



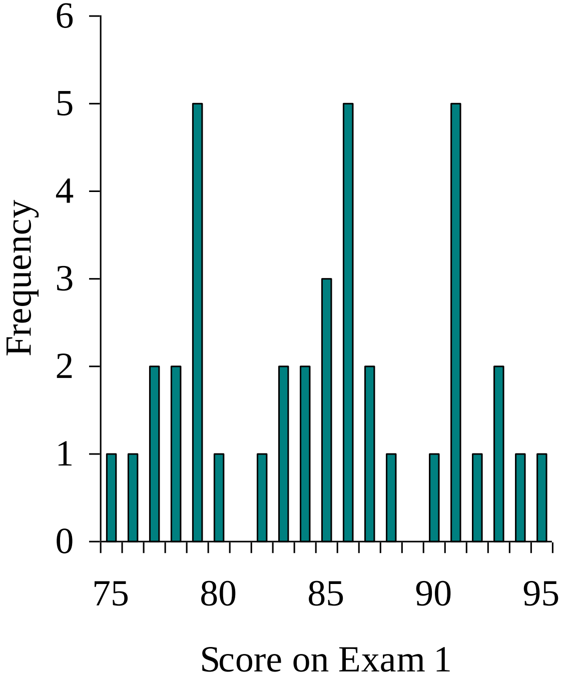
Bimodal Distributions

- When a distribution has two “modes,” it is called *bimodal*



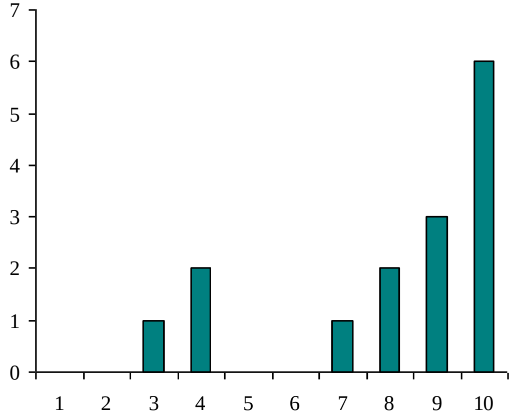
Multimodal Distributions

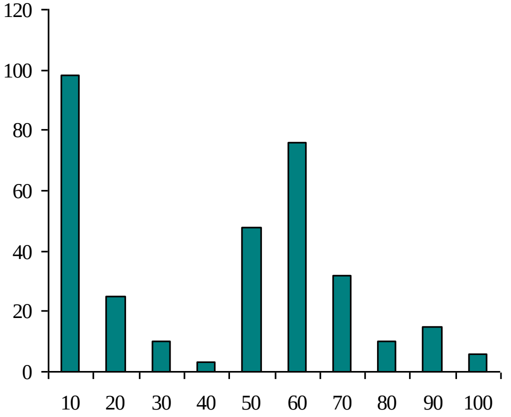
- If a distribution has more than 2 "modes," it is called *multimodal*



When To Use the Mode

- The mode is not a very useful measure of central tendency
  - It is insensitive to large changes in the data set
    - That is, two data sets that are very different from each other can have the same mode





When To Use the Mode

- The mode is primarily used with nominally scaled data
  - It is the only measure of central tendency that is appropriate for nominally scaled data

The Median

- The *median* is simply another name for the 50th percentile
  - It is the score in the middle; half of the scores are larger than the median and half of the scores are smaller than the median

How To Calculate the Median

- Conceptually, it is easy to calculate the median
  - There are many minor problems that can occur; it is best to let a computer do it
- Sort the data from highest to lowest
- Find the score in the middle
  - middle = (N + 1) / 2
  - If N, the number of scores, is even the median is the average of the middle two scores

Median Example

- What is the median of the following scores:
  10 8 14 15 7 3 3 8 12 10 9
- Sort the scores:
  15 14 12 10 10 9 8 8 7 3 3
- Determine the middle score:
  middle = (N + 1) / 2 = (11 + 1) / 2 = 6
- Middle score = median = 9

Median Example

- What is the median of the following scores:
  24 18 19 42 16 12
- Sort the scores:
  42 24 19 18 16 12
- Determine the middle score:
  middle = (N + 1) / 2 = (6 + 1) / 2 = 3.5
- Median = average of 3^{rd} and 4^{th} scores:
  (19 + 18) / 2 = 18.5

When To Use the Median

- The median is often used when the distribution of scores is either positively or negatively skewed
  - The few really large scores (positively skewed) or really small scores (negatively skewed) will not overly influence the median

The Mean

- The *mean* is:
  - the arithmetic average of all the scores
    $(\Sigma X)/N$
  - the number, m, that makes $\Sigma(X - m)$ equal to 0
  - the number, m, that makes $\Sigma(X - m)^2$ a minimum
- The mean of a population is represented by the Greek letter μ; the mean of a sample is represented by X

Calculating the Mean

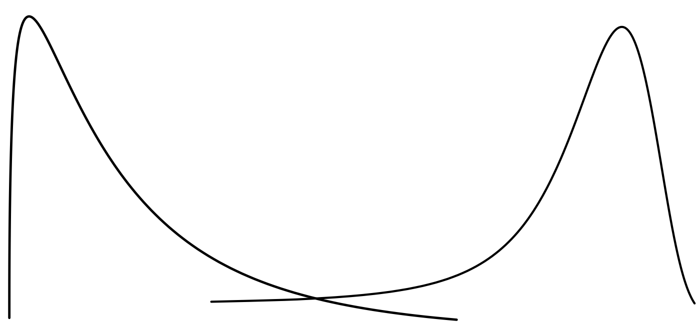
- Calculate the mean of the following data:
  1   5   4   3   2
- Sum the scores (ΣX):
  $1 + 5 + 4 + 3 + 2 = 15$
- Divide the sum ($\Sigma X = 15$) by the number of scores ($N = 5$):
  $15 / 5 = 3$
- Mean = X = 3

When To Use the Mean

- You should use the mean when
  - the data are interval or ratio scaled
    - Many people will use the mean with ordinally scaled data too
  - and the data are not skewed
- The mean is preferred because it is sensitive to every score
  - If you change one score in the data set, the mean will change

Relations Between the Measures of Central Tendency

- In symmetrical distributions, the median and mean are equal
  - For normal distributions, mean = median = mode
- In positively skewed distributions, the mean is greater than the median
- In negatively skewed distributions, the mean is smaller than the median



Examples

Example 2·1. *(a) Find the arithmetic mean of the following frequency distribution:*

x :	1	2	3	4	5	6	7
f :	5	9	12	17	14	10	6

(b) Calculate the arithmetic mean of the marks from the following table :

Marks :	0–10	10–20	20–30	30–40	40–50	50–60
No. of students :	12	18	27	20	17	6

Solution. *(a)*

x	f	fx
1	5	5
2	9	18
3	12	36
4	17	68
5	14	70
6	10	60
7	6	42
	73	299

$$\therefore \qquad \bar{x} = \frac{1}{N} \Sigma f x = \frac{299}{73} = 4{\cdot}09$$

(b)

Marks	*No. of students (f)*	*Mid - point (x)*	*fx*
0–10	12	5	60
10–20	18	15	270
20–30	27	25	675
30–40	20	35	700
40–50	17	45	765
50–60	6	55	330
Total	100		2,800

Arithmetic mean or $\bar{x} = \dfrac{1}{N} \Sigma f x = \dfrac{1}{100} \times 2{,}800 = 28$

Example 2·2. *Calculate the mean for the following frequency distribution.*

Class-interval :	*0–8*	*8–16*	*16–24*	*24–32*	*32–40*	*40–48*
Frequency :	*8*	*7*	*16*	*24*	*15*	*7*

Solution.

Class–interval	*mid–value* (x)	*Frequency* (f)	$d = (x-A)/h$	fd
0–8	4	8	–3	–24
8–16	12	7	–2	–14
16–24	20	16	–1	–16
24–32	28	24	0	0
32–40	36	15	1	15
40–48	44	7	2	14
		77		–25

Here we take $A = 28$ and $h = 8$.

$$\therefore \qquad \bar{x} = A + \frac{h\,\Sigma f d}{N} = 28 + \frac{8 \times (-\ 25)}{77} = 28 - \frac{200}{77} = 25{\cdot}404$$

Example 2·3. *The average salary of male employees in a firm was Rs.520 and that of females was Rs.420. The mean salary of all the employees was Rs.500. Find the percentage of male and female employees.*

Solution. Let n_1 and n_2 denote respectively the number of male and female employees in the concern and $\bar{x}_1$ and $\bar{x}_2$ denote respectively their average salary (in rupees). Let $\bar{x}$ denote the avarage salary of all the workers in the firm.

We are given that :

$$\bar{x}_1 = 520, \quad \bar{x}_2 = 420 \quad \text{and} \quad \bar{x} = 500$$

Also we know

$$\bar{x} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2}$$

$$\Rightarrow \quad 500\,(n_1 + n_2) = 520\,n_1 + 420\,n_2$$

$$\Rightarrow \quad (520 - 500)\,n_1 = (500 - 420)\,n_2$$

$$\Rightarrow \quad 20\,n_1 = 80\,n_2$$

$$\Rightarrow \quad \frac{n_1}{n_2} = \frac{4}{1}$$

Hence the percentage of male employees in the firm

$$= \frac{4}{4+1} \times 100 = 80$$

and percentage of female employees in the firm

$$= \frac{1}{4+1} \times 100 = 20$$

Example 2·5. *Obtain the median for the following frequency distribution:*

x :	1	2	3	4	5	6	7	8	9
f :	8	10	11	16	20	25	15	9	6

Solution.

x	f	$c.f.$
1	8	8
2	10	18
3	11	29
4	16	45
5	20	65
6	25	90
7	15	105
8	9	114
9	6	120
	120	

Hence $N = 120 \quad \Rightarrow \quad N/2 = 60$

Cumulative frequency ($c.f.$) just greater than $N/2$, is 65 and the value of x corresponding to 65 is 5. Therefore, median is 5.

Example 2·6. *Find the median wage of the following distribution :*

Wages (in Rs.) :	20—30	30—40	40—50	50—60	60—70
No. of labourers :	3	5	20	10	5

[Gorakhpur Univ. B. Sc. 1989]

Solution.

Wages (in Rs.)	*No. of labourers*	*c.f.*
20—30	3	3
30—40	5	8
40—50	20	28
50—60	10	38
60—70	5	43

Here $N/2 = 43/2 = 21{\cdot}5$. Cumulative frequency just greater than 21·5 is 28 and the corresponding class is 40–50. Thus median class is 40–50. Hence using (2·6), we get

$$\text{Median} = 40 + \frac{10}{20}(21{\cdot}5 - 8) = 40 + 6{\cdot}75 = 46{\cdot}75$$

Thus median wage is Rs. 46·75.

References

1. Norean R. Sharpe, Richard D. De Veaux, Paul F. Velleman, Business Statistics, Fourth Edition, Pearson Education. 2019.
2. R.E. Walpole, R.H. Mayers, S.L. Mayers and K.Ye, Probability and Statistics for engineers and scientists, 9th Edition, Pearson Education, 2018.
3. S. C. Gupta and V. K. Kapoor, Fundamentals of Applied Statistics, S. Chand, 2006.
4. Douglas C. Montgomery, George C. Runger, Applied Statistics and Probability for Engineers, Sixth Edition, John Wiley and Sons, 2014.
5. Miller & Freund's, Probability and statistics for engineers, 8th edition, Pearson publication, 2018.

THE END