Statistics for Engineers

Overview

Overview

Overview

Syllabus

1. Introduction to Statistics

Syllabus

1. Introduction to Statistics
2. Random variables

Syllabus

1. Introduction to Statistics
2. Random variables
3. Correlation and Regression

Syllabus

1. Introduction to Statistics
2. Random variables
3. Correlation and Regression
4. Probability Distributions

Syllabus

1. Introduction to Statistics
2. Random variables
3. Correlation and Regression
4. Probability Distributions
5. Hypothesis Testing - I

Syllabus

1. Introduction to Statistics
2. Random variables
3. Correlation and Regression
4. Probability Distributions
5. Hypothesis Testing - I
6. Hypothesis Testing - II

Syllabus

1. Introduction to Statistics
2. Random variables
3. Correlation and Regression
4. Probability Distributions
5. Hypothesis Testing - I
6. Hypothesis Testing - II
7. Reliability

Books

Text books

1. R. E. Walpole, R. H. Mayers, S.Probability and Statistics for Engineers and Scientists, 9th Edition, Pearson Education (2012)
2. Douglas C. Montgomery, George C. Runger, Applied Statistics and Probability for Engineers, John Wiley & Sons, 5th Edition (2010).

Books

Text books

1. R. E. Walpole, R. H. Mayers, S.Probability and Statistics for Engineers and Scientists, 9th Edition, Pearson Education (2012)
2. Douglas C. Montgomery, George C. Runger, Applied Statistics and Probability for Engineers, John Wiley & Sons, 5th Edition (2010).

Reference Books

1. E. Balagurusamy, Reliability Engineering, Tata McGraw Hill, Tenth reprint 2010.
2. J.L Devore, Probability and Statistics , 8th Edition, Brooks/Cole, Cengage Learning (2012)
3. R. A. Johnson, Miller & Freund's, Probability and Statistics for Engineers, 8th Edition, PHI (2010).

Introduction

- Data

Introduction

- Data - collection of any number of related observations

Introduction

- Data - collection of any number of related observations
- Statistics deals with collection, presentation, analysis, and interpretation of the numerical data

Introduction

- Data - collection of any number of related observations
- Statistics deals with collection, presentation, analysis, and interpretation of the numerical data
- There are two major types of statistics: (a) Descriptive Statistics, (b) Inferential Statistics
- Descriptive Statistics - consist of methods of organizing and summarizing information. (includes construction of graphs, charts, tables and calculation of mean, median, mode, measures of variation.
- Inferential Statistics - consist of methods of drawing conclusions based on the information. (includes estimation, hypothesis testing)

Illustration

Example

Consider the event of tossing dice. The dice is rolled 100 times. The results are 5,4,5,6,6,6,1,1,1,1,1,2,5,5,3,2,1,4,4,4,4,2,2,6,6,6,6,2,...

Illustration

Example

Consider the event of tossing dice. The dice is rolled 100 times. The results are 5,4,5,6,6,6,1,1,1,1,1,2,5,5,3,2,1,4,4,4,4,2,2,6,6,6,6,2,...
Descriptive statistics is used for grouping the data to the following table

Outcome of the roll	Frequencies
1	10
2	20
3	18
4	16
5	11
6	25

Illustration

Example

Consider the event of tossing dice. The dice is rolled 100 times. The results are 5,4,5,6,6,6,1,1,1,1,1,2,5,5,3,2,1,4,4,4,4,2,2,6,6,6,6,2,...
Descriptive statistics is used for grouping the data to the following table

Outcome of the roll	Frequencies
1	10
2	20
3	18
4	16
5	11
6	25

Inferential statistics can be used to verify whether the dice is a fair or not.

Population

is the set of observations or individuals corresponding to an entire collection of units for which inferences are to be made.
Example: The books in our library, The students of VIT.

Population

is the set of observations or individuals corresponding to an entire collection of units for which inferences are to be made.
Example: The books in our library, The students of VIT.

Sample

is a subset of a population that are actually collected in the course of an investigation. Example: The collection of Statistics books in our library, The students of our class.

Example

Let the blood types of 40 persons are as follows: O, O, A, B, A, O, A, A, A, O, B, O, B, O, O, A, O, O, A, A, A, A, AB, A, B, A, A, O, O, A, O, O, A, A, A, O, A, O, O, AB.

Example

Let the blood types of 40 persons are as follows: O, O, A, B, A, O, A, A, A, O, B, O, B, O, O, A, O, O, A, A, A, A, AB, A, B, A, A, O, O, A, O, O, A, A, A, O, A, O, O, AB.

Blood type	Frequencies
O	16
A	18
B	4
AB	2
Total	16

Example

Let the blood types of 40 persons are as follows: O, O, A, B, A, O, A, A, A, O, B, O, B, O, O, A, O, O, A, A, A, A, AB, A, B, A, A, O, O, A, O, O, A, A, A, O, A, O, O, AB.

Blood type	Frequencies
O	16
A	18
B	4
AB	2
Total	16

Try to graph this data (pie chart and bar graph)

Measures of Central Tendancy

Data are classified as

- Individual observations or raw data
- Discrete data
- Continuous data

Measures of Central Tendancy

Data are classified as

- Individual observations or raw data
- Discrete data
- Continuous data

The Central measures are

- Mean
- Median
- Mode
- Geometric mean
- Harmonic mean

Mean (Individual observations or raw data)

Mean (Direct method)

$\bar{X} = \frac{X_1+X_2+...+X_N}{N} = \frac{\sum X_i}{N}$ where N is the total number of observations.

Example

Given the monthly income of 10 employees in an office,
1780,1760,1690,1750,1840,1920,1100,1810,1050,1950
Find the mean of their monthly income (or Average).

Mean (Individual observations or raw data)

Mean (Direct method)

$\bar{X} = \frac{X_1+X_2+...+X_N}{N} = \frac{\sum X_i}{N}$ where N is the total number of observations.

Example

Given the monthly income of 10 employees in an office,
1780,1760,1690,1750,1840,1920,1100,1810,1050,1950
Find the mean of their monthly income (or Average). Mean=1665

Mean (Individual observations or raw data)

Mean (Direct method)

$\bar{X} = \frac{X_1+X_2+...+X_N}{N} = \frac{\sum X_i}{N}$ where N is the total number of observations.

Example

Given the monthly income of 10 employees in an office,
1780,1760,1690,1750,1840,1920,1100,1810,1050,1950
Find the mean of their monthly income (or Average). Mean=1665

Mean (Short-cut method)

- Taken an assumed mean A
- Find the deviations: $d = X - A$
- $\bar{X} = A + \frac{\sum d}{N}$

Mean (Discrete data)

- Direct method

$$\bar{X} = \frac{\sum fX}{N}$$

- Short-cut method

$$\bar{X} = A + \frac{\sum fd}{N}$$

where A is the assumed mean, $d = X - A$, N is the total number of observations.

Problem

1. From the following data of the marks obtained by 60 students of a class. Calculate the arithmetic mean.

Marks	20	30	40	50	60	70
No. of students	8	12	20	10	6	4

Mean (Discrete data)

- Direct method

$$\bar{X} = \frac{\sum fX}{N}$$

- Short-cut method

$$\bar{X} = A + \frac{\sum fd}{N}$$

where A is the assumed mean, $d = X - A$, N is the total number of observations.

Problem

1. From the following data of the marks obtained by 60 students of a class. Calculate the arithmetic mean.

Marks	20	30	40	50	60	70
No. of students	8	12	20	10	6	4

Arithmetic mean=41

Mean (Continuous data)

- Direct method

$$\bar{X} = \left(\frac{\sum fX}{N}\right)$$

- Short-cut method

$$\bar{X} = A + \left(\frac{\sum fd}{N}\right) * h$$

where A is the assumed mean, $d = \frac{X-A}{h}$, N is the total number of observations, h is the class size, X is the mid value of each class interval.

Example

From the following data, compute arithmetic mean

Marks	0-10	10-20	20-30	30-40	40-50	50-60
No. of students	5	10	25	30	20	10

Mean (Continuous data)

- Direct method

$$\bar{X} = \left(\frac{\sum fX}{N}\right)$$

- Short-cut method

$$\bar{X} = A + \left(\frac{\sum fd}{N}\right) * h$$

where A is the assumed mean, $d = \frac{X-A}{h}$, N is the total number of observations, h is the class size, X is the mid value of each class interval.

Example

From the following data, compute arithmetic mean

Marks	0-10	10-20	20-30	30-40	40-50	50-60
No. of students	5	10	25	30	20	10

$\bar{X} = 33$

Problem

Compute mean for the following data:

Marks	0-10	10-30	30-60	60-100
No. of students	5	12	25	8

Median

Arrange the data in ascending or descending order.

- For raw data:
  - If N is odd, then Median is $X_{\frac{N+1}{2}}$
  - If N is even, then Median is $\frac{X_{\frac{N}{2}}+X_{\frac{N}{2}+1}}{2}$

Median

Arrange the data in ascending or descending order.

- For raw data:
  - If N is odd, then Median is $X_{\frac{N+1}{2}}$
  - If N is even, then Median is $\frac{X_{\frac{N}{2}}+X_{\frac{N}{2}+1}}{2}$
- Discrete data:
  - Find the cummulative frequencies (c.f.)
  - Identify the Median as average of $\frac{N}{2}$ th observation and $\frac{N}{2}+1$ th observation (if N is even) or $\frac{N+1}{2}$ th observation (if N is odd) using c.f.

Median

Arrange the data in ascending or descending order.

- For raw data:
  - If N is odd, then Median is $X_{\frac{N+1}{2}}$
  - If N is even, then Median is $\frac{X_{\frac{N}{2}}+X_{\frac{N}{2}+1}}{2}$
- Discrete data:
  - Find the cummulative frequencies (c.f.)
  - Identify the Median as average of $\frac{N}{2}$ th observation and $\frac{N}{2}+1$ th observation (if N is even) or $\frac{N+1}{2}$ th observation (if N is odd) using c.f.
- Continuous data:
  - Find the cummulative frequencies (c.f.)
  - Identify the Median class as the class that contains $\frac{N}{2}$ th observation
  - Median $= L + (\frac{\frac{N}{2} - c.f.}{f}) * h$ where L is the lower limit of the median class, c.f. is the cummulative frequency of the class preceding to the median class, f is the frequency of the median class, and h is the median class size.

Problems

1. Compute median for the following data:
$1100, 1150, 1080, 1120, 1200, 1160, 1400$

Problems

1. Compute median for the following data:
$1100, 1150, 1080, 1120, 1200, 1160, 1400$
Median$= 1150$

Problems

1. Compute median for the following data:
$1100, 1150, 1080, 1120, 1200, 1160, 1400$
Median$= 1150$
2. Compute median for the following data:

Income (Rs.)	1000	1500	800	2000	2500	1800
No. of persons	24	26	16	20	6	30

Problems

1. Compute median for the following data:
$1100, 1150, 1080, 1120, 1200, 1160, 1400$
Median$= 1150$
2. Compute median for the following data:

Income (Rs.)	1000	1500	800	2000	2500	1800
No. of persons	24	26	16	20	6	30

Median $= 1500$

Problems

1. Compute median for the following data:
$1100, 1150, 1080, 1120, 1200, 1160, 1400$
Median$= 1150$
2. Compute median for the following data:

Income (Rs.)	1000	1500	800	2000	2500	1800
No. of persons	24	26	16	20	6	30

Median $= 1500$
3. Calculate the median for the following data:

Marks	0-10	10-30	30-60	60-100
No. of students	5	12	25	8

Problems

1. Compute median for the following data:
$1100, 1150, 1080, 1120, 1200, 1160, 1400$
Median$=$ 1150
2. Compute median for the following data:

Income (Rs.)	1000	1500	800	2000	2500	1800
No. of persons	24	26	16	20	6	30

Median $=$ 1500
3. Calculate the median for the following data:

Marks	0-10	10-30	30-60	60-100
No. of students	5	12	25	8

Median $=$ 39.6

4. An incomplete frequency distribution is given below.

Variable	10-20	20-30	30-40	40-50	50-60	60-70	70-80	Total
frequency	12	30	f_1	65	f_2	25	18	229

Given that the median value is 46, determine the missing frequencies.

$f_1 = 34, f_2 = 45$

Mode

The observation with maximum frequency

Mode

Mode

The observation with maximum frequency

- Discrete data:
  - If the frequency distribution is regular, then the observation with maximum frequency is the mode.
  - For irregular frequency distribution, mode is calculated by the method of grouping.

Mode

Mode

The observation with maximum frequency

- Discrete data:
  - If the frequency distribution is regular, then the observation with maximum frequency is the mode.
  - For irregular frequency distribution, mode is calculated by the method of grouping.
- Continuous data:
  - If the frequency distribution is regular, then the Class interval with maximum frequency is the modal class. Then Mode $= L + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right) * h$ where L is the lower limit of the modal class, f_1 is the frequency of the modal class, f_0 is the frequency of the class preceding the modal class, f_2 is the frequency of the class succeeding the modal class, and h is the modal class size.
  - For irregular frequency distribution, mode is calculated by the method of grouping.

Problems

1. Compute mode for the following data:

Income (Rs.)	1000	1500	800	2000	2500	1800
No. of persons	24	26	16	20	6	30

Problems

1. Compute mode for the following data:

Income (Rs.)	1000	1500	800	2000	2500	1800
No. of persons	24	26	16	20	6	30

Mode $= 1800$

1. Compute mode for the following data:

Income (Rs.)	1000	1500	800	2000	2500	1800
No. of persons	24	26	16	20	6	30

Mode $= 1800$

2. Calculate the mode for the following data:

Marks	0-10	10-30	30-60	60-100
No. of students	5	12	25	8

Problems

1. Compute mode for the following data:

Income (Rs.)	1000	1500	800	2000	2500	1800
No. of persons	24	26	16	20	6	30

Mode = 1800

2. Calculate the mode for the following data:

Marks	0-10	10-30	30-60	60-100
No. of students	5	12	25	8

Modal class is 30-60. From the problem, $f_1 = 25, f_0 = 12, f_2 = 8$

Mode = 43

4. The median and mode are Rs.33.50 and Rs.34 respectively for the following data.

Wages	0-10	10-20	20-30	30-40	40-50	50-60	60-70	Total
frequency	4	16	f_1	f_2	f_3	6	4	230

5. Find mean, median, and mode for the following data:

CI	20-40	40-60	60-80	80-100	100-120	120-140	140-160
f	8	12	20	30	40	35	18

CI	160-180	180-200
f	7	5

Mean=108.5, Median=108.75, Mode=118.3

Problems on Inclusive intervals

Exclusive intervals: Lower limit included but upper limit excluded

Inclusive intervals: Lower and upper limits are included

6. Calculate the mean, median, and mode for the following data:

CI	0-9	10-19	20-29	30-39	40-49	50-59
f	3	15	10	8	3	1

Problems on Inclusive intervals

Exclusive intervals: Lower limit included but upper limit excluded

Inclusive intervals: Lower and upper limits are included

6. Calculate the mean, median, and mode for the following data:

CI	0-9	10-19	20-29	30-39	40-49	50-59
f	3	15	10	8	3	1

7. Compute mean and median for the following data

Height (in cm)	< 140	< 145	< 150	< 155	< 160	< 165
No. of students	4	11	29	40	46	51

Exclusive intervals: Lower limit included but upper limit excluded

Inclusive intervals: Lower and upper limits are included

6. Calculate the mean, median, and mode for the following data:

CI	0-9	10-19	20-29	30-39	40-49	50-59
f	3	15	10	8	3	1

7. Compute mean and median for the following data

Height (in cm)	< 140	< 145	< 150	< 155	< 160	< 165
No. of students	4	11	29	40	46	51

The empirical relationship is Mode$=$ 3 Median - 2 Mean

Geometric Mean

Geometric mean is the nth root of product of n observations.
G.M. $=$ Antilog $\left(\frac{\sum log(x_i)}{N}\right)$ (raw data) where N is the total number of observations.
G.M. $=$ Antilog $\left(\frac{\sum f_i log(x_i)}{N}\right)$ (Discrete or Continuous data) where N is the total frequency.

Problem

Calculate G.M. for 125, 1462, 38, 7, 0.22, 0.08, 12.75, 0.5
G.M.$=$6.952

Harmonic Mean

Harmonic mean is the reciprocal of the arithmetic mean of the reciprocal of observations.

H.M. $= \frac{N}{\sum \frac{1}{x_i}}$ (raw data)

H.M. $= \frac{N}{\sum f_i \frac{1}{x_i}}$ (Discrete or Continuous data)

Problem

Find H.M. of 2574, 475, 75, 5, 0.8, 0.08, 0.005, 0.0009.

H.M. $= 0.006$

Measures of dispersion

- Range
- Quartile Deviation
- Mean Deviation
- Standard Deviation

Range

Range is the difference between max observation and min observation

Problems

1. Find the range of $1100, 1150, 1080, 1120, 1200, 1160, 1400$.

Range= 1400-1080=320

2. Find the range of the following data:

X	1	2	3	4	5	6	7	8	9	10	11	12
f	3	8	15	23	35	40	32	28	20	45	14	6

Range=12-1=11

3. Find the range of the following data:

Marks	0-10	10-30	30-60	60-100
No. of students	5	12	25	8

Range=100-0=100

Quartile Deviation

Order the given data and find the cummulative frequencies.

- First Quartile $= Q_1 = L + (\frac{\frac{N}{4} - c.f.}{f}) * h$ where L is the lower limit of the first quartile class.
- Second Quartile (Q_2) coincides with Median
- Third Quartile $= Q_3 = L + (\frac{\frac{3N}{4} - c.f.}{f}) * h$ where L is the lower limit of the third quartile class
- Quartile Deviation $= \frac{Q_3 - Q_1}{2}$
- Inter-quartile range $= Q_3 - Q_1$
- Coefficient of Quartile Deviation $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$

Problems

1. Find all the quartiles, quartile deviation, and coefficient of quartile deviation for the following data:

Marks	0-10	10-30	30-60	60-100
No. of students	5	12	25	8

Mean Deviation

- For raw data, Mean deviation $= \frac{\sum |D_i|}{N}$ where N is the total number of observations and $D_i = X - A$, A is a central measure (Mean or Median or Mode)
- For discrete or continuous data, Mean deviation $= \frac{\sum f_i |D_i|}{N}$ where N is the total frequency.

Mean Deviation

- For raw data, Mean deviation $= \frac{\sum |D_i|}{N}$ where N is the total number of observations and $D_i = X - A$, A is a central measure (Mean or Median or Mode)
- For discrete or continuous data, Mean deviation $= \frac{\sum f_i |D_i|}{N}$ where N is the total frequency.

Problems

1. Calculate the mean deviation about median for the following data:

Income (Rs.)	1000	1500	800	2000	2500	1800
No. of persons	24	26	16	20	6	30

Mean Deviation

- For raw data, Mean deviation $= \frac{\sum |D_i|}{N}$ where N is the total number of observations and $D_i = X - A$, A is a central measure (Mean or Median or Mode)
- For discrete or continuous data, Mean deviation $= \frac{\sum f_i |D_i|}{N}$ where N is the total frequency.

Problems

1. Calculate the mean deviation about median for the following data:

Income (Rs.)	1000	1500	800	2000	2500	1800
No. of persons	24	26	16	20	6	30

2. Calculate the mean deviation about median for the following data:

Marks	0-10	10-30	30-60	60-100
No. of students	5	12	25	8

Standard Deviation

- Raw data: $\sigma = \sqrt{\frac{\sum (x_i)^2}{N} - (\frac{\sum x_i}{N})^2}$ where N is the total number of observations.
- Discrete or Continuous data: $\sigma = \sqrt{\frac{\sum f_i(x_i)^2}{N} - (\frac{\sum f_i x_i}{N})^2}$ where N is the total frequency.
- Discrete or Continuous data: $\sigma = (\sqrt{\frac{\sum f_i(d_i)^2}{N} - (\frac{\sum f_i d_i}{N})^2}) * h$ where N is the total frequency and $d_i = \frac{x_i - A}{h}$

Standard Deviation

- Raw data: $\sigma = \sqrt{\frac{\sum(x_i)^2}{N} - (\frac{\sum x_i}{N})^2}$ where N is the total number of observations.
- Discrete or Continuous data: $\sigma = \sqrt{\frac{\sum f_i(x_i)^2}{N} - (\frac{\sum f_i x_i}{N})^2}$ where N is the total frequency.
- Discrete or Continuous data: $\sigma = (\sqrt{\frac{\sum f_i(d_i)^2}{N} - (\frac{\sum f_i d_i}{N})^2}) * h$ where N is the total frequency and $d_i = \frac{x_i - A}{h}$

Problems

1. Calculate standard deviation for the following data:

X	10	11	12	13	14
f	3	12	18	12	3

Problems

2. From the following data, compute standard deviation

Marks	0-10	10-20	20-30	30-40	40-50	50-60
No. of students	5	10	25	30	20	10

Problems

2. From the following data, compute standard deviation

Marks	0-10	10-20	20-30	30-40	40-50	50-60
No. of students	5	10	25	30	20	10

Coefficient of Variation

$C.V = \frac{\sigma}{\bar{x}} * 100\%$

Problems

1. Goals scored by two teams in a foot ball season were as follows:

No. of goals scored	0	1	2	3	4	5
No. of matches (Team A)	15	10	7	5	3	2
No. of matches (Team B)	20	10	5	4	2	1

State which team is more consistent. (A is more consistent.)

1. Goals scored by two teams in a foot ball season were as follows:

No. of goals scored	0	1	2	3	4	5
No. of matches (Team A)	15	10	7	5	3	2
No. of matches (Team B)	20	10	5	4	2	1

State which team is more consistent. (A is more consistent.)

2. Two brands of tyres are tested with the following results

Life (in '000 miles)	20-25	25-30	30-35	35-40	40-45
No. of tyres (Brand X)	1	22	64	10	3
No. of tyres (Brand Y)	0	24	76	0	0

(a) Which brand of tyres have greater average life?
(b) Compare the variability and state which brand of tyres would you use on your fleet of trucks?
$\bar{X} = 32.1, \sigma_X = 3.441, C.V_X = 10.72\%, \bar{Y} = 31.3, \sigma_Y = 2.136, C.V_Y = 6.824\%$, (a) X has greater average life (b) Y is more consistent

Moments

Central moment

The A.M of various powers of the deviations about mean

$$\mu_r = \frac{\sum(x-\bar{x})^r}{N} \qquad \text{raw data}$$

Moments

Central moment

The A.M of various powers of the deviations about mean

$$\mu_r = \frac{\sum (x - \bar{x})^r}{N} \qquad \text{raw data}$$

$$\mu_r = \frac{\sum (f(x - \bar{x})^r)}{\sum f} \qquad \text{frequency distribution}$$

Moments

Central moment

The A.M of various powers of the deviations about mean

$$\mu_r = \frac{\sum(x-\bar{x})^r}{N} \qquad \text{raw data}$$

$$\mu_r = \frac{\sum(f(x-\bar{x})^r)}{\sum f} \qquad \text{frequency distribution}$$

Moments about Assumed mean or Origin

$$\mu_r^{'} = \frac{\sum(x-A)^r}{N} = \frac{\sum d^r}{N} \qquad \text{raw data}$$

Moments

Central moment

The A.M of various powers of the deviations about mean

$$\mu_r = \frac{\sum(x-\bar{x})^r}{N} \qquad \text{raw data}$$

$$\mu_r = \frac{\sum(f(x-\bar{x})^r)}{\sum f} \qquad \text{frequency distribution}$$

Moments about Assumed mean or Origin

$$\mu_r^{'} = \frac{\sum(x-A)^r}{N} = \frac{\sum d^r}{N} \qquad \text{raw data}$$

$$\mu_r^{'} = \frac{\sum(f(x-A)^r)}{\sum f} = \left(\frac{\sum(fd^r)}{\sum f}\right) * h^r \qquad \text{frequency distribution}$$

Calculation of central moments using moments about origin

$$\mu_1 = 0$$

Calculation of central moments using moments about origin

$$\mu_1 = 0$$

$$\mu_2 = \mu_2^{'} - (\mu_1^{'})^2 = \sigma^2$$

Calculation of central moments using moments about origin

$$\mu_1 = 0$$

$$\mu_2 = \mu_2' - (\mu_1')^2 = \sigma^2$$

$$\mu_3 = \mu_3' - 3\mu_1'\mu_2' + 2(\mu_1')^3$$

Calculation of central moments using moments about origin

$$\mu_1 = 0$$

$$\mu_2 = \mu_2^{'} - (\mu_1^{'})^2 = \sigma^2$$

$$\mu_3 = \mu_3^{'} - 3\mu_1^{'}\mu_2^{'} + 2(\mu_1^{'})^3$$

$$\mu_4 = \mu_4^{'} - 4\mu_1^{'}\mu_3^{'} + 6(\mu_1^{'})^2\mu_2^{'} - 3(\mu_1^{'})^4$$

Calculation of central moments using moments about origin

$$\mu_1 = 0$$

$$\mu_2 = \mu_2^{'} - (\mu_1^{'})^2 = \sigma^2$$

$$\mu_3 = \mu_3^{'} - 3\mu_1^{'}\mu_2^{'} + 2(\mu_1^{'})^3$$

$$\mu_4 = \mu_4^{'} - 4\mu_1^{'}\mu_3^{'} + 6(\mu_1^{'})^2\mu_2^{'} - 3(\mu_1^{'})^4$$

1. Calculate the first four central moments from the following data :

Class Interval	0-10	10-20	20-30	30-40
Frequency	1	3	4	2

Calculation of central moments using moments about origin

$$\mu_1 = 0$$

$$\mu_2 = \mu_2^{'} - (\mu_1^{'})^2 = \sigma^2$$

$$\mu_3 = \mu_3^{'} - 3\mu_1^{'}\mu_2^{'} + 2(\mu_1^{'})^3$$

$$\mu_4 = \mu_4^{'} - 4\mu_1^{'}\mu_3^{'} + 6(\mu_1^{'})^2\mu_2^{'} - 3(\mu_1^{'})^4$$

1. Calculate the first four central moments from the following data :

Class Interval	0-10	10-20	20-30	30-40
Frequency	1	3	4	2

$$\mu_1^{'} = 7, \mu_2^{'} = 130, \mu_3^{'} = 1900, \mu_4^{'} = 37000$$

$$\mu_1 = 0, \mu_2 = 81, \mu_3 = -144, \mu_4 = 14817$$

Problems

2. Calculate the first four central moments from the following data :

Marks (above)	0	10	20	30	40	50	60	70	80
No. of students	150	140	100	80	80	70	30	14	0

Table 1: Mark distribution of students

2. Calculate the first four central moments from the following data :

Marks (above)	0	10	20	30	40	50	60	70	80
No. of students	150	140	100	80	80	70	30	14	0

Table 1: Mark distribution of students

CI	f	x	$d = \frac{x-35}{10}$	fd	fd^2	fd^3	fd^4
0-10	10	5	-3	-30	90	-90	810
10-20	40	15	-2	-80	160	-320	640
20-30	20	25	-1	-20	20	-20	20
30-40	0	35	0	0	0	0	0
40-50	10	45	1	10	10	10	10
50-60	40	55	2	80	160	320	640
60-70	16	65	3	48	144	432	1296
70-80	14	75	4	56	224	896	3584
	$\sum f = 150$			$\sum fd = 64$	$\sum fd^2 = 808$	$\sum fd^3 = 1228$	$\sum fd^4 = 7000$

$$\mu_1^{'} = 4.27, \mu_2^{'} = 538.67, \mu_3^{'} = 8186.67, \mu_4^{'} = 466666.67$$

$$\mu_1 = 0, \mu_2 = 520.44, \mu_3 = 1442.02, \mu_4 = 384770.13$$

Skewness

- Lack of symmetry (measures the assymetry of the distribution)

Skewness

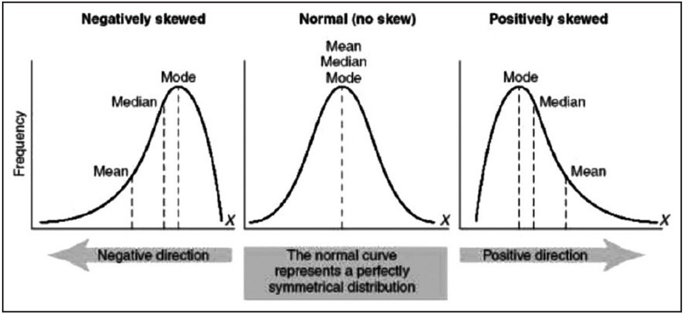
- Lack of symmetry (measures the assymetry of the distribution)
- tells the degree and direction of departure from symmetry

Skewness

- Lack of symmetry (measures the assymetry of the distribution)
- tells the degree and direction of departure from symmetry
- if the distribution has longer tail towards right - positive skewness. (Mean > Median > Mode)

Skewness

- Lack of symmetry (measures the assymetry of the distribution)
- tells the degree and direction of departure from symmetry
- if the distribution has longer tail towards right - positive skewness. (Mean > Median > Mode)
- if the distribution has longer tail towards left - negative skewness. (Mode > Median > Mean)



Skewness

Measure of skewness using moments

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}, \qquad \gamma_1 = \sqrt{\beta_1}$$

If $\mu_3 > 0$, then positive skewness, else negative skewness.

Kurtosis

Kurtosis is a measure of peakness.

measure of Kurtosis

$$\beta_2 = \frac{\mu_4}{(\mu_2)^2}, \qquad \gamma_2 = \beta_2 - 3$$

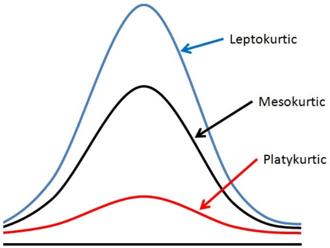
Kurtosis

Kurtosis is a measure of peakness.

measure of Kurtosis

$$\beta_2 = \frac{\mu_4}{(\mu_2)^2}, \qquad \gamma_2 = \beta_2 - 3$$

- If $\gamma_2 = 0$, then the distribution is Normal or Mesokurtic.
- If $\gamma_2 > 0$, then the distribution is Leptokurtic.
- If $\gamma_2 < 0$, then the distribution is Platykurtic.



Problems

1. Calculate the measure of Skewness (using moments) and measure of Kurtosis from the following data :

Class Interval	0-10	10-20	20-30	30-40
Frequency	1	3	4	2

Problems

1. Calculate the measure of Skewness (using moments) and measure of Kurtosis from the following data :

Class Interval	0-10	10-20	20-30	30-40
Frequency	1	3	4	2

$$\mu_1' = 7, \mu_2' = 130, \mu_3' = 1900, \mu_4' = 37000$$

$$\mu_1 = 0, \mu_2 = 81, \mu_3 = -144, \mu_4 = 14817$$

1. Calculate the measure of Skewness (using moments) and measure of Kurtosis from the following data :

Class Interval	0-10	10-20	20-30	30-40
Frequency	1	3	4	2

$$\mu_1^{'} = 7, \mu_2^{'} = 130, \mu_3^{'} = 1900, \mu_4^{'} = 37000$$

$$\mu_1 = 0, \mu_2 = 81, \mu_3 = -144, \mu_4 = 14817$$

Measure of Skewness: $\beta_1 = \frac{\mu_3^2}{\mu_2^3} = 0.039, \qquad \gamma_1 = \sqrt{\beta_1} = 0.1975$

Problems

1. Calculate the measure of Skewness (using moments) and measure of Kurtosis from the following data :

Class Interval	0-10	10-20	20-30	30-40
Frequency	1	3	4	2

$$\mu_1^{'} = 7, \mu_2^{'} = 130, \mu_3^{'} = 1900, \mu_4^{'} = 37000$$

$$\mu_1 = 0, \mu_2 = 81, \mu_3 = -144, \mu_4 = 14817$$

Measure of Skewness: $\beta_1 = \frac{\mu_3^2}{\mu_2^3} = 0.039,\qquad \gamma_1 = \sqrt{\beta_1} = 0.1975$

Measure of Kurtosis: $\beta_2 = \frac{\mu_4}{\mu_2^2} = 2.2583,\qquad \gamma_2 = \beta_2 - 3 = -0.7417$

1. Calculate the measure of Skewness (using moments) and measure of Kurtosis from the following data :

Class Interval	0-10	10-20	20-30	30-40
Frequency	1	3	4	2

$$\mu_1' = 7, \mu_2' = 130, \mu_3' = 1900, \mu_4' = 37000$$

$$\mu_1 = 0, \mu_2 = 81, \mu_3 = -144, \mu_4 = 14817$$

Measure of Skewness: $\beta_1 = \frac{\mu_3^2}{\mu_2^3} = 0.039,\qquad \gamma_1 = \sqrt{\beta_1} = 0.1975$

Measure of Kurtosis: $\beta_2 = \frac{\mu_4}{\mu_2^2} = 2.2583,\qquad \gamma_2 = \beta_2 - 3 = -0.7417$

The distribution is Platykurtic and negative skewed (sign of μ_3).

Problems

2. The first three moments of a distribution about the value 3 of a variable are 2,10 and 30 respectively. Obtain mean,μ_2, μ_3, and hence β_1. Comment upon the nature of skewness. Also determine the first three moments about zero.

Problems

2. The first three moments of a distribution about the value 3 of a variable are 2,10 and 30 respectively. Obtain mean, μ_2, μ_3, and hence β_1. Comment upon the nature of skewness. Also determine the first three moments about zero. ($\bar{x} = 5, \mu_1 = 0, \mu_2 = 6, \mu_3 = -14$, moments about zero are $5, 31, 201$)

Problems

2. The first three moments of a distribution about the value 3 of a variable are 2,10 and 30 respectively. Obtain mean,μ_2, μ_3, and hence β_1. Comment upon the nature of skewness. Also determine the first three moments about zero. ($\bar{x} = 5, \mu_1 = 0, \mu_2 = 6, \mu_3 = -14$, moments about zero are $5, 31, 201$)

3. The standard deviation of a symmetrical distribution is 3. What must be the value of the fourth moment about the mean in order that the distribution is mesokurtic?

Problems

2. The first three moments of a distribution about the value 3 of a variable are 2,10 and 30 respectively. Obtain mean, μ_2, μ_3, and hence β_1. Comment upon the nature of skewness. Also determine the first three moments about zero. ($\bar{x} = 5, \mu_1 = 0, \mu_2 = 6, \mu_3 = -14$, moments about zero are $5, 31, 201$)

3. The standard deviation of a symmetrical distribution is 3. What must be the value of the fourth moment about the mean in order that the distribution is mesokurtic?($\mu_4 = 243$)

Problems

2. The first three moments of a distribution about the value 3 of a variable are 2,10 and 30 respectively. Obtain mean,μ_2, μ_3, and hence β_1. Comment upon the nature of skewness. Also determine the first three moments about zero. ($\bar{x} = 5, \mu_1 = 0, \mu_2 = 6, \mu_3 = -14$, moments about zero are $5, 31, 201$)
3. The standard deviation of a symmetrical distribution is 3. What must be the value of the fourth moment about the mean in order that the distribution is mesokurtic?($\mu_4 = 243$)
4. The mean and SD of a set of 100 observations were worked out as 40 and 5 respectively. But by mistake a value 50 was taken in place of 40 for one observation. Calculate the correct mean and SD.

Problems

2. The first three moments of a distribution about the value 3 of a variable are 2,10 and 30 respectively. Obtain mean, μ_2, μ_3, and hence β_1. Comment upon the nature of skewness. Also determine the first three moments about zero. ($\bar{x} = 5, \mu_1 = 0, \mu_2 = 6, \mu_3 = -14$, moments about zero are $5, 31, 201$)

3. The standard deviation of a symmetrical distribution is 3. What must be the value of the fourth moment about the mean in order that the distribution is mesokurtic?($\mu_4 = 243$)

4. The mean and SD of a set of 100 observations were worked out as 40 and 5 respectively. But by mistake a value 50 was taken in place of 40 for one observation. Calculate the correct mean and SD. ($\bar{x} = 39.9, \sigma = 4.9$)