Programme Name

B.Sc. III Sem. V – SEC

Course Name

Statistics in Research Methodology

**Practical Problem Sheets
With Solution**

**Prepared By
Dr. Sanjay Supanekar
Balasaheb Desai College, Patan**

Balasaheb Desai College, Patan
Statistics in Research Methodology
B.SC. III (SEC Semester V)
Practical No. 1
Design a questionnaire and collect data from small sample

Problem 1: Design a questionnaire and collect data for the following problem.
A local gym wants to improve its services and member satisfaction. They need to understand members' opinions on current services and potential improvements.

Problem 2: Design a questionnaire and collect data for the following problem.
A local coffee shop wants to gather feedback from its customers to enhance the quality of its products and services.

Problem 3:   Design a questionnaire and collect data for the following problem.
A college wants to evaluate student satisfaction with a specific course to improve its curriculum and teaching methods.

Problem 4:   Design a questionnaire and collect data for the following problem.
A public health center wants to gather patient feedback to improve service quality and patient experience.

Solution

Solution 1:

Questionnaire

Q.1   Age:

Under 18 ☐   18-24 ☐   25-34 ☐   35-44 ☐   45 and over ☐

Q.2   Gender

Male ☐   Female ☐

Q.3   Membership Duration

Less than 6 months ☐   6 months - 1 year ☐   1-2 years ☐ More than 2 years ☐

Q.4   How satisfied are you with the following services?

	Very Dissatisfied	Dissatisfied	Neutral	Satisfied	Very Satisfied
Equipment quality					
Cleanliness					
Staff professionalism					
Personal training					

Q.5   What aspects of the gym do you like the most? (Open-ended)

Q.6   What aspects of the gym do you think need improvement?

Q.7   What new services or amenities would you like to see at the gym?

Solution 2:

Questionnaire

Q.1   How often do you visit our coffee shop?

Daily ☐   Weekly ☐   Monthly ☐   Rarely ☐

Q.2   How satisfied are you with the following aspects?

	Very Dissatisfied	Dissatisfied	Neutral	Satisfied	Very Satisfied
Quality of coffee					
Cleanliness of the shop					
Friendliness of staff					
Speed of service					

Q.3   What do you like most about our coffee shop? (Open-ended)

Q.4   What areas need improvement?

Q.4   Are there any new products or services you would like us to offer?

Solution 3:
Questionnaire

Q.1 **Year of Study?**
First ☐   Second ☐   Third ☐

Q.2 **How clear was the course objectives?**

	Very Dissatisfied	Dissatisfied	Neutral	Satisfied	Very Satisfied
Course objectives					
Engagement of lectures					
the usefulness of the teaching methods and course materials					

Q.3 **What aspects of the course did you find most valuable?** (Open-ended)

Q.4 **What improvements would you suggest for the course?**

Q.4 **What improvements would you suggest for the course?**

**Balasaheb Desai College, Patan
Statistics in Research Methodology
B.SC. III (SEC Semester V)
Practical No. 2
Classification, Tabulation, and frequency distribution of qualitative and quantitative data**

Problem 1: A library wants to categorize and analyze the types of books borrowed over a month.
Raw Data of books borrowed are listed with their genres:
Fiction , Non-Fiction, Science Fiction, Fiction, Biography, Fiction, Non-Fiction, Science Fiction, Biography, Science Fiction, Science Fiction, Fiction, Biography, Fiction, Non-Fiction, Science Fiction, Biography, Science Fiction, Fiction , Non-Fiction.
Classify each book into one of the genres: Fiction, Non-Fiction, Science Fiction, Biography and Create a table to count the number of books in each genre.

Problem 2: A company wants to analyze customer feedback on a new product to improve it. Customer feedback categories are as follows:
Excellent, Good, Good, Poor, Excellent, Fair, Poor, Fair, Good, Excellent, Poor, Excellent, Fair, Poor, Fair, Good, Excellent, Good, Good, Poor.
Classify each feedback into one of the categories: Excellent, Good, Fair, Poor and Create a frequency distribution of feedback category appears.

Problem 3: A fitness center wants to determine which types of exercises are most popular among its members. Members' favorite exercise types are recorded as follows:
Yoga, Cardio, Strength Training, Yoga, Cardio, Strength Training, Yoga, Pilates, Pilates, Cardio, Cardio, Yoga, Cardio, Yoga, Cardio, Strength Training, Cardio
Classify each response into one of the exercise types: Yoga, Cardio, Strength Training, Pilates and create a frequency distribution of responses for each exercise type

Problem 4: A company wants to analyze the number of complaints received each month over a year to identify patterns or trends. Number of complaints received each month is as follows:
5, 7, 8, 6, 9, 10, 7, 6, 8, 9, 7, 5, 8, 6, 9, 10, 7, 6, 8,
Classify the data and create a ungrouped frequency distribution.

Problem 5: A market research study aims to understand the age distribution of customers visiting a store.
A sample of 30 customers' ages is recorded as follows:
22, 25, 30, 45, 50, 32, 29, 40, 55, 43, 27, 34, 29, 38, 42, 31, 28, 46, 49, 53, 24, 31, 29, 35, 41, 52, 26, 37, 45, 48, 40, 50, 28, 29, 31, 22, 39, 47, 42, 54
Group ages into ranges (e.g., 20-30, 30-40, 40-50, 50-60) and create a grouped frequency distribution.

Solution

Solution 1:

Book Genre: Classification, tabulation and frequency distribution

Genre	Tally Mark	Frequency
Fiction		
Non-Fiction		
Science Fiction		
Biography		

Solution 2:

Customer Feedback: Classification, tabulation and frequency distribution

Feedback Category	Tally Mark	Frequency
Excellent		
Good		
Fair		
Poor		

Solution 3:

Favorite Exercise Type: Classification, tabulation and frequency distribution

Exercise Type	Tally Mark	Frequency
Yoga,		
Cardio		
Strength Training		
Pilates		

Solution 4:

Number of Complaints: Classification, tabulation and frequency distribution

Number of Complaints	Tally Mark	Frequency
5		
6		
7		
8		
9		
10		

Solution 5:

Age Range: Classification, tabulation and frequency distribution

Age Range	Tally Mark	Frequency
20-30		
30-40		
40-50		
50-60		

Balasaheb Desai College, Patan
Statistics in Research Methodology
B.SC. III (SEC Semester V)
Practical No. 3
Diagrammatic Presentation of data by using bar diagrams and pie diagrams

Problem 1: A meteorologist wants to present the total rainfall in inches for each month of the year. Rainfall in inches for each month is recorded as follows:

Month	Jan.	Feb.	March	April	May	June	July	Aug.	Sep.	Oct.	Nov.	Dec.
Rainfall	0.2	0.1	0.2	0.3	2.4	6.1	7.5	6.7	6.2	4.1	2.6	0.4

Draw simple bar diagram.

Problem 2: A university wants to display the number of students enrolled in various departments and the distribution among different levels (Undergraduate, Master's, Doctoral) and data given below

Department	Undergraduate	Master's	Doctoral
Engineering	500	200	50
Business	400	180	40
Arts	300	150	30
Sciences	350	170	60

Draw sub divided bar diagram.

Problem 3: A teacher wants to compare test scores across three subjects for five students and data given below

Student	Math	Science	English
A	85	78	88
B	90	82	84
C	78	85	80
D	88	90	86
E	84	79	82

Draw multiple bar diagram.
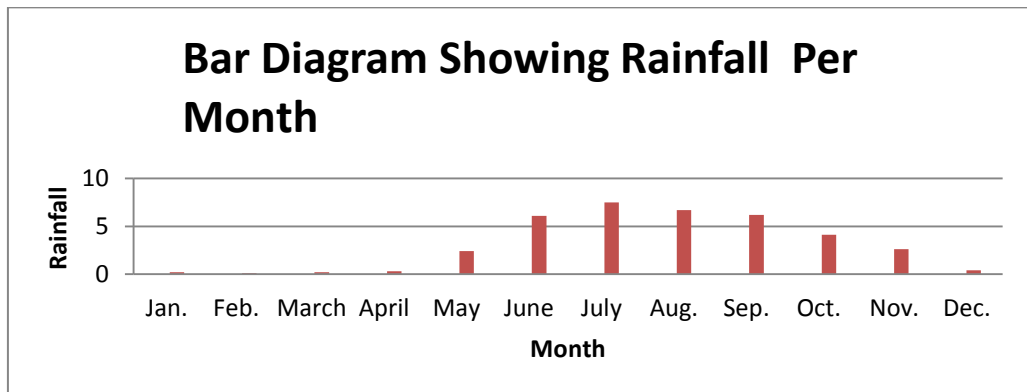
Problem 4: A person wants to visualize how they spend their time during a typical day and data given below

Activity	Time Spent (%)
Sleeping	30
Work	40
Meals	10
Leisure	10
Exercise	5
Other	5

Draw Pie diagram.

Solution

Solution 1:
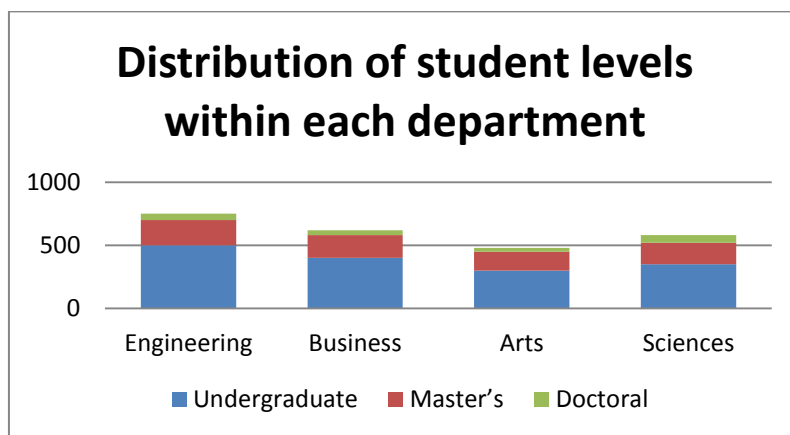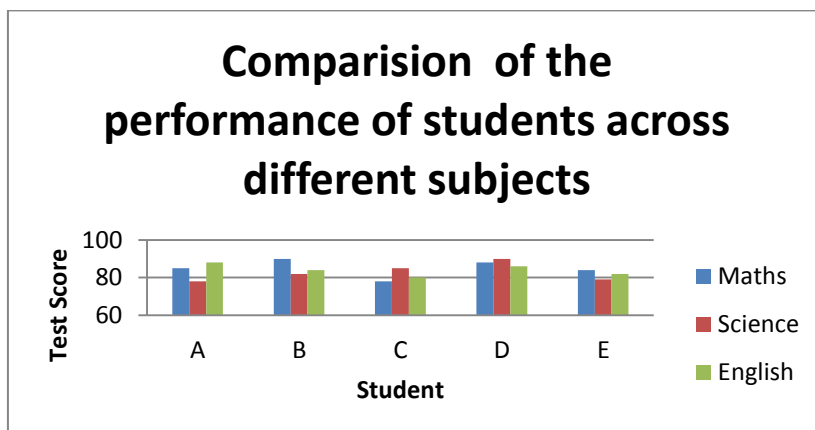
Solution 2:

Solution 3:

Solution 4:
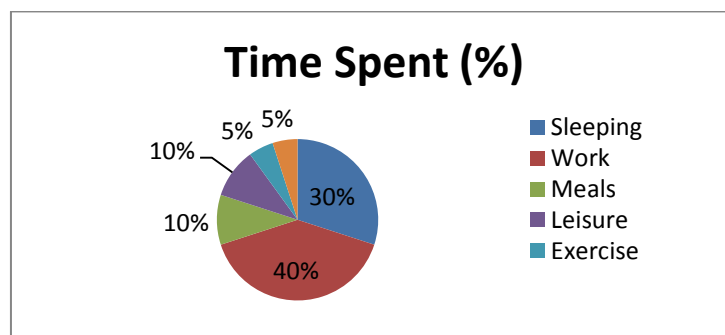
Balasaheb Desai College, Patan
Statistics in Research Methodology
B.SC. III (SEC Semester V)
Practical No. 4
Graphical representation of data by using histogram and frequency polygon

Problem 1: The marks in Statistics of 100 students are as given below:

Marks:	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
Students:	4	6	10	15	22	18	12	8	3	2

Draw the histogram & find the value of mode.

Problem 2: A community center wants to understand the age distribution of its members. Age distributions of member are as given below:

Ages:	20-30	30-40	40-50	50-60	60-70
No. of members:	5	12	20	10	02

Draw the frequency polygon

Problem 3: The 100 students are classified according to class are as given below:

Class :	0-35	35-45	45-60	60-70	70-100
Students:	10	25	40	17	8

Draw the histogram & find the value of mode.

Problem 4: A fitness coach wants to analyze how clients' daily exercise times are distributed. Exercise time distribution are given below

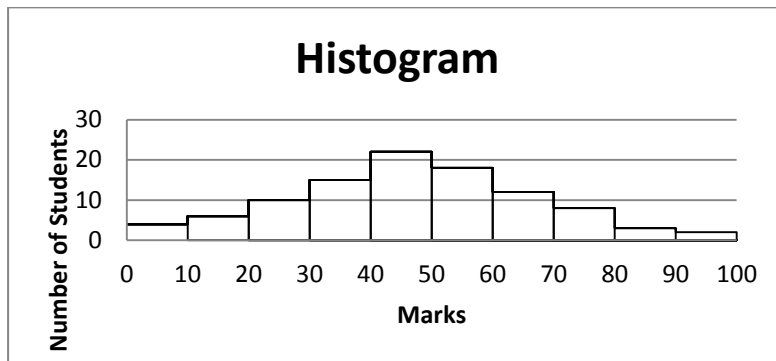| Time of exercise (in minutes) :	30-40	40-50	50-60	60-70
No. of clients:	5	12	20	10

Draw histogram and frequency polygon

Solution

Solution 1:
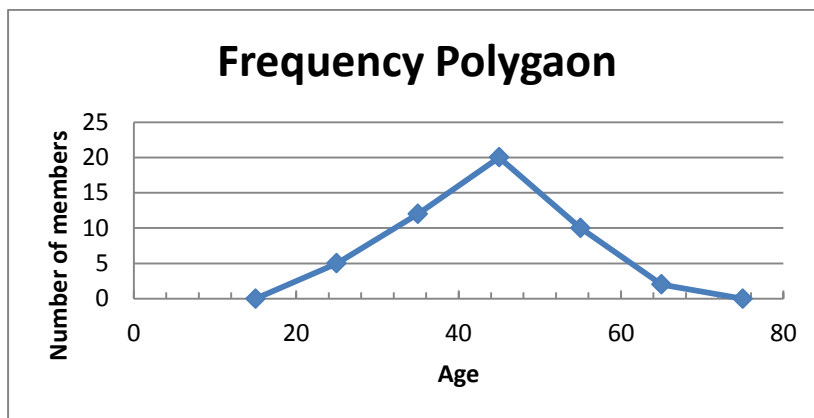
Solution 2:

Solution 3:

Marks	Class width	Students	Frequency Density
0-35	35	10	0.29
35-45	10	25	2.5
45-60	15	40	2.67
60-70	10	17	1.7
70-100	30	8	0.27
		100	
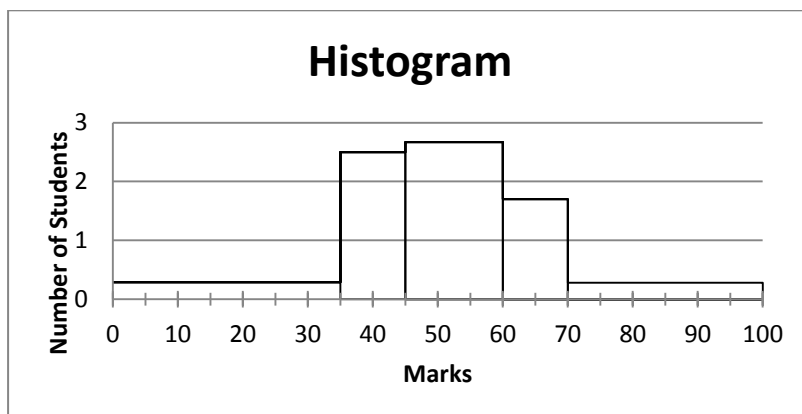
Balasaheb Desai College, Patan
Statistics in Research Methodology
B.SC. III (SEC Semester V)
Practical No. 5
Sampling I : Simple random sampling (with and without replacement)

Problem 1: Suppose you have a population of 5 items: {A, B, C, D, E }. You want to take a sample of 3 items without replacement. How many different possible samples can you draw?

Problem 2: Suppose you have a population of 3 items: {A, B, C}. You want to take a sample of 2 items with replacement. How many different possible samples can you draw?

Problem 3: Consider the following sampling frame:

Population Unit	1	2	3	4	5	6	7	8	9	10
Value of Y	48	40	35	30	25	45	32	22	44	18

Draw random sample of size five from above population by using Simple random sampling with replacement (SRSWR) method

Prob. 4: Consider the following sampling frame:

Population Unit	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Value of Y	100	250	150	300	125	225	122	235	345	220	280	230	150	175	200

Draw random sample of size five from above population by using Simple random sampling without replacement (SRSWOR) method

Solution

Solution 1: When sampling without replacement, the number of different possible samples is given by the combination formula

$$n_{Ck} = \frac{n!}{(n-k)!\, k!}$$

Here, n=5 and k=3

$$5_{C3} = \frac{5!}{2!\, 3!} = \frac{5 \times 4}{2 \times 1} = 10$$

(ABC, ABD, ABE, ACD, ACE, ADE, BCD, BCE, BDE, CDE)

So, there are 10 different possible samples.

Solution 2: When sampling with replacement, each of the 2 positions in the sample can be filled by any of the 3 items. Thus, the total number of possible samples is given by:

$n^k = 3^2 = 9$

(AA, AB, AC, BA, BB, BC, CA, CB. CC)

So, there are 9 different possible samples.

Solution 3: Here Population size = 10 and Sample size = 5

Simple random sampling with replacement (SRSWR) method
Using lottery method, we draw a 5 random numbers between (1, 10) as 8, 2, 7, 2, 5

Sr. No.	1	2	3	4	5
Random No. (Population Unit no.)	8	2	7	2	5
Sample (Value of Y)	22	40	32	40	25

Solution 4: Here Population size = 15 and Sample size = 5

Simple random sampling with replacement (SRSWOR) method
Using Random number table, we draw a 5 random numbers between (1, 15) as 4, 5, 7, 2, 12

Sr. No.	1	2	3	4	5
Random No. (Population Unit no.)	4	5	7	2	12
Sample (Value of Y)	300	125	122	250	230

**Balasaheb Desai College, Patan
Statistics in Research Methodology
B.SC. III (SEC Semester V)
Practical No. 6
Sampling II: Stratified Random Sampling**

Problem 1: A university has 1,200 students divided into three faculties:

- Faculty A: 500 students

- Faculty B: 300 students

- Faculty C: 400 students

You want to draw a stratified random sample of 120 students, with the sample size proportional to the size of each faculty. How many students should you sample from each faculty?

Problem 2: Suppose we have 3 types of flowers Rose, Lotus and Lily with number 4, 6 and 10 respectively. Researcher wants to choose a sample of size 10 from an entire population.

(i) Obtain how many flowers are to be selected from each type of flowers

(ii) Draw random samples of above sizes from different types of flowers

Problem 3: A workshop has 200 participants out of which 150 are males and 50 females. We want to draw a stratified random sample of 20 participants, with the sample size proportional to the size of each male and female participant's.

(i) Obtain how many participants are to be selected from each category

(ii) Draw random samples of above sizes from different categories

---

Solution

Solution 1: Here Population size = 1200 and Sample size = 120

To allocate the sample size proportionally, use the following formula:

$$\text{Sample size from stratum} = \frac{\text{Size of stratum}}{\text{Total population size}} \times \text{Total sample size}$$

Faculty A: Sample size from stratum $= \frac{500}{1200} \times 120 = 50$

Faculty B: Sample size from stratum $= \frac{300}{1200} \times 120 = 30$

Faculty C: Sample size from stratum $= \frac{400}{1200} \times 120 = 40$

So, you should sample 50 students from Faculty A, 30 students from Faculty B, and 40 students from Faculty C.

Solution 2: Here Population size = 20 and Sample size = 10

By Stratified random sampling

Number of flowers are to be selected from each type of flowers

Strata	Population Size	Sample size = (Ni/N)*n
Rose flowers	4	=(4/20)*10 = 2
Lotus flowers	6	=(6/20)*10 = 3
Lily flowers	10	=(10/20)*10 = 5
Total	20	10

Strata 1: Rose flower : 4   (Let R1, R2,….R4) ➔ n_1 =2

By lottery method (SRSWOR), we draw a 2 random numbers as 3, 2 i.e. R3 and R2

Strata 2: Rose flowers: 6   (Let Lo1, Lo2,….Lo6) ➔ n_2 = 3

By lottery method (SRSWOR), we draw a 3 random numbers as 1, 5, 2 i.e. Lo1, Lo5, Lo2

Strata 3: Rose flowers: 6   (Let Li1, Li2, ……Li10) ➔ n_3 = 5

By lottery method (SRSWOR), we draw a 3 random numbers as 4, 2, 6, 7, 3

i.e. Li4, Li2, Li6, Li7, Li3

Solution 3: Here Population size = 200 and Sample size = 20

By Stratified random sampling, Number of flowers is to be selected from each type of flowers

Strata	Population Size	Sample size = (Ni/N)*n
Male	150	=(150/200)*20 = 15
Female	50	=(50/200)*20 = 5
Total	200	20

Strata 1: Male Participants: 150   (Let M1, M2,….M150) ➔ n_1 =15

By lottery method (SRSWOR), we draw a 15 random numbers and corresponding population units as

Sr. No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Random Number	20	54	72	5	8	123	143	38	73	16	82	140	102	136	3
Population Unit	M20	M54	M72	M5	M8	M123	M143	M38	M73	M16	M82	M140	M102	M136	M3

Strata 2: Female Participants: 50   (Let F1,F2,….F50) ➔ n_2 = 5

By lottery method (SRSWOR), we draw a 5 random numbers and corresponding population units as

Sr. No.	1	2	3	4	5
Random Number	23	18	5	46	32
Population Unit	F23	F18	F5	F46	F32

Balasaheb Desai College, Patan
Statistics in Research Methodology
B.SC. III (SEC Semester V)
Practical No. 7
Measure of Central tendency: Mean, Median and Mode

Problem 1: The following data set gives the number of children for 20 families selected at random.

3	5	1	2	3	5	5	4	4	2
3	4	1	2	3	2	4	5	3	3

Find the mean, median and mode of the number of children

Problem 2: Following are the marks obtained by 12 students in two papers

Paper I	36	56	41	46	54	59	55	51	52	44	50	80
Paper II	58	54	21	51	59	46	55	31	66	41	70	36

Find (i) The A.M. of the marks of each paper ii) Combined mean of the marks of both papers

Problem 3: The survey report on the effectiveness of the sleeping drug gives the following data for

additional sleep :

Additional Sleep	1	2	3	4	5	6	7	8
No. of Persons	10	13	18	23	28	24	16	8

Find Mean, Median and Mode

Problem 4: Marks obtained by 125 students in Mathematics are as follows

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
No. of Students	1	9	15	19	26	23	17	10	3	2

Find Mean, Median and Mode

Solution

Solution 1: Let X = number of children's in family and here n = 20

First we arrange the data in ascending order

1	1	2	2	2	2	3	3	3	3	3	3	4	4	4	4	5	5	5	5

$Mean = \bar{x} = \frac{\sum x_i}{n} = \frac{64}{20} = 3.2$

Median = Size of $(\frac{n+1}{2})th$ observation

= Size of $(10.5)^{th}$ observation $= (\frac{10^{th}+11^{th}}{2}) = (\frac{3+3}{2}) = 3$

Mode = most repeated observation = 3

Solution 2: Let X : Marks of paper I and Y : Marks of Paper II and Here $n_1 = n_2 = 12$

(i)   The A.M. of the marks of paper I = $\bar{x} = \frac{\sum x_i}{n} = \frac{624}{12} = 52$

The A.M. of the marks of paper II = $\bar{y} = \frac{\sum y_i}{n} = \frac{588}{12} = 49$

(ii)   Combined mean of the marks of both papers = $\bar{x}_c = \frac{n_1\bar{x}+n_2\bar{y}}{n_1+n_2} = \frac{12\times52+12\times49}{24} = 50.5$

Solution 3: Let X : Additional Sleep and N = $\sum f_i = 140$

x_i	f_i	$f_i x_i$	l.c.f.
1	10	10	10
2	13	26	23
3	18	54	41
4	23	92	64
5	28	140	92 ←
6	24	144	116
7	16	112	132
8	8	64	140
	140	642	

V $Mean = \bar{x} = \frac{\sum f_i x_i}{N} = \frac{642}{140} = 4.5857$

Median = Size of $(\frac{N+1}{2})th$ observation

= Size of $(70.5)^{th}$ observation

= 5 (from table)

Mode = Observation which has maximum frequency = 5

Solution 4: Let X : Marks in Statistics and $N = \sum f_i = 125$

xi	fi	mi	fimi	l.c.f.
0-10	1	5	5	1
10-20	9	15	135	10
20-30	15	25	375	25
30-40	19	35	665	44
40-50	**26**	**45**	**1170**	**70**
50-60	23	55	1265	93
60-70	17	65	1105	110
70-80	10	75	750	120
80-90	3	85	255	123
90-100	2	95	190	125
	125		5915	

i) $Mean = \bar{x} = \frac{\sum f_i m_i}{N} = \frac{5915}{125} = 47.37$

ii) Median = Size of $(\frac{N}{2})th$ observation

$\qquad$ = Size of (62.5) th observation

$\qquad$ = 40-50 (from table)

Median = $L_1 + \left(\dfrac{N/2 - c.f.}{f} \right) \times h$

Where, L_1 = Lower limit of the median class. = 40

$\qquad$ N = total frequency = $\sum f_i$ =125

$\qquad$ c.f. = Less than cumulative frequency of previous class of median class = 44

$\qquad$ f = frequency of median class = 26

$\qquad$ h = Class width of median class. 10

Median = $40 + \left(\dfrac{62.5 - 44.}{26} \right) \times 10 = 47.1153$

iii) Here, Maximum frequency is 26 $\therefore$ Modal class = 40 – 50

Mode = $L_1 + \left(\dfrac{f_1 - f_0}{2f_1 - f_0 - f_2} \right) \times h$

Here, $\quad L_1$ = Lower limit of the modal class = 40

$\qquad f_0$ = Frequency of previous class of modal class = 19

$\qquad f_1$ = Frequency of modal class = 26

$\qquad f_2$ = Frequency of next class of modal class = 23

$\qquad$ h = Class width of modal class = 10

$\therefore$ Mode = $40 + \left(\dfrac{26 - 19}{52 - 19 - 23} \right) \times 10 = 47$

Balasaheb Desai College, Patan
Statistics in Research Methodology
B.SC. III (SEC Semester V)
Practical No. 8
Measure of Dispersion: Range, Variance and Standard Deviation

Problem 1: The weights (in kg.) of college students are as follows :

    49   63   46   59   65   52   60   54

    Find Range, Variance and S.D

Problem 2: The printing errors in book are given below

No. of error :	0	1	2	3	4	5	6
No. of pages:	3	10	17	25	15	12	8

    Find Range, Variance and S.D

Problem 3: Age wise workers in a factory are given below

Age in year :	10-20	20-30	30-40	40-50	50-60
No. of workers:	2	10	28	20	12

    Find Range, Variance and S.D

Problem 4: Calculate Mean and variance from the following data

Value X	: 7	8	9	10	11	12	13
Frequency (f)	: 4	6	9	12	9	6	4

Solution

Solution 1: Let X = weights (in kg.) of college students and here n = 8

Here, Largest value =L = 65 and   Smallest value = S = 46

$\therefore$ Range = L – S = 19

xi	49	63	46	59	65	52	60	54	**448**
xi^2	2401	3969	2116	3481	4225	2704	3600	2916	**25412**

Now mean = $\bar{x} = \dfrac{\sum x_i}{n} = \dfrac{448}{8} = 56$

$Variance = \dfrac{\sum_{i=1}^{n} x_i^{2}}{n} - \bar{x}^2 = \dfrac{25412}{8} - (56)^2 = 40.4999$

S.D = $\sigma = \sqrt{Variance} = \sqrt{40.4999} = 6.3639$

Solution 2: Let X = printing errors in book and N = $\sum f_i = 90$

Here, Largest value =L = 6 and   Smallest value = S = 0

$\therefore$ Range = L – S = 6

x_i	f_i	$f_i x_i$	$f_i x_i^2$
0	3	0	0
1	10	10	10
2	17	34	68
3	25	75	225
4	15	60	240
5	12	60	300
6	8	48	288
	90	287	1131

$\bar{x} = \dfrac{\sum f_i x_i}{N} = \dfrac{287}{90} = 3.1889$

$Variance = \dfrac{\sum_{i=1}^{n} f_i . x_i^{2}}{N} - \bar{x}^2 = \dfrac{1131}{90} - (3.1889)^2$

$= 0.2358$

S.D = $\sigma = \sqrt{Variance} = \sqrt{0.2358} = 1.5484$

Solution 3: Let X = Age of workers and N = $\sum f_i = 72$

Here, Largest value =L = 60 and   Smallest value = S = 10

$\therefore$ Range = L – S = 50

x_i	f_i	m_i	$f_i m_i$	$f_i m_i^2$
10-20	2	15	30	450
20-30	10	25	250	6250
30-40	28	35	980	34300
40-50	20	45	900	40500
50-60	12	55	660	36300
	72		2820	117800

Now mean = $\bar{x} = \dfrac{\sum f_i m_i}{N} = \dfrac{2820}{72} = 39.17$

$Variance = \dfrac{\sum_{i=1}^{n} f_i . m_i^{2}}{N} - \bar{x}^2 = \dfrac{117800}{72} - (39.17)^2$

$= 101.8081$

S.D = $\sigma = \sqrt{Variance} = \sqrt{101.8081} = 10.09$

Balasaheb Desai College, Patan
Statistics in Research Methodology
B.SC. III (SEC Semester V)
Practical No. 9
Scatter Plots

Problem 1: Draw the scatter of the following data and comment on it.

X	5	10	15	20	5	10	15	20
Y	10	10	5	5	15	5	10	15

Problem 2: The marks in Mathematics and Physics are given blow.

Marks in Mathematics	12	22	35	47	59	60	52	40
Marks in Physics	15	25	42	50	65	57	48	38

Draw the scatter of the following data and comment on it.

Problem 3: The supply and price of commodity are given blow.

Supply (Kg.)	38	40	33	32	30	34	26	23	37	29
Price (Rs./kg)	17	19	20	21	22	23	25	26	18	24

Draw the scatter of the following data and comment on it.

Problem 4: The data on **Length of side of square and Perimeter of square** are given blow.

Length of side of square (in cm.)	5	10	12	15	20
Perimeter of square (in cm.)	20	40	48	60	80

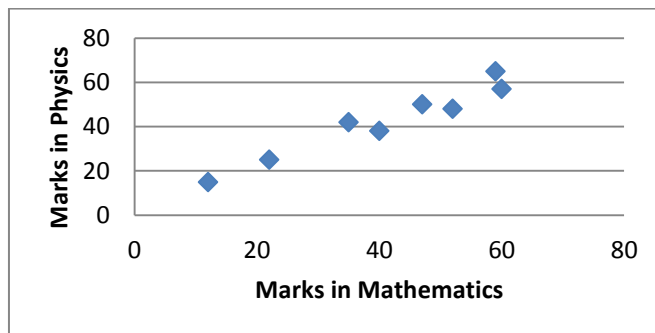Draw the scatter of the following data and comment on it.
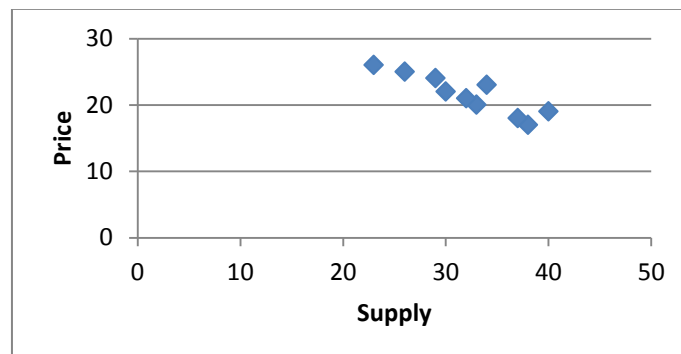
Solution
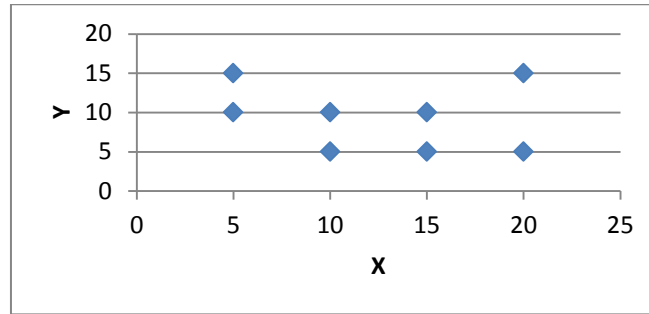
Solution 1

Comment: There is no correlation.
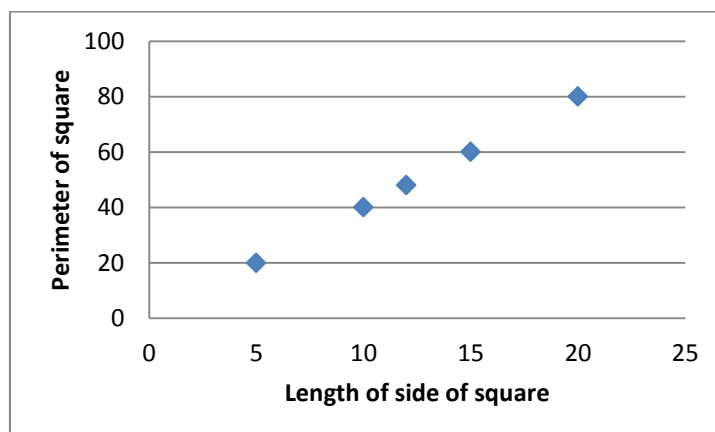
Solution 2

Comment: There is high degree positive correlation.

Solution 3

Comment: There is high degree negative correlation

Solution 4

Comment: There is Perfect positive correlation

Balasaheb Desai College, Patan
Statistics in Research Methodology
B.SC. III (SEC Semester V)
Practical No. 10
Karl Pearson's correlation coefficient

Problem 1: Calculate correlation coefficient between price and demand. Comment on your result.

        Price   :   2   3   4   7   6

        Demand: 10   7   3   1   2

Problem 2: Calculate correlation coefficient between marks of two subjects A and B:

        Sub. A : 3   5   8   4   9   6   7   0

        Sub. B : 0   3   6   5   9   4   8   2

Problem 3: A sample of size 10 from a bivariate distribution gives the following values

$\Sigma X = 80, \quad \Sigma Y = 100, \quad \Sigma X^2 = 700, \Sigma Y^2 = 1200, \Sigma XY = 870.$

find correlation coefficient between X and Y.

Problem 4: Find correlation coefficient between X and Y. from following data :

$n = 10, \Sigma X = 100, \quad \Sigma Y = 220, \quad \Sigma X^2 = 1024, \Sigma Y^2 = 4980, \Sigma XY = 2147.$

Solution

Solution 1 Let X : Price and Y : Demand

X	Y	X^2	Y^2	X.Y
2	10	4	100	20
3	7	9	49	21
4	3	16	9	12
7	1	49	1	7
6	2	36	4	12
22	23	114	163	72

$\bar{x} = 22/5 = 4.4$

$\bar{y} = 23/5 = 4.6$

$$r = \frac{\sum XY - n.\bar{X}.\bar{Y}}{\sqrt{\sum X^2 - n.\bar{X}^2}.\sqrt{\sum Y^2 - n.\bar{Y}^2}}$$

$$= \frac{72 - 5(4.4)(4.6)}{\sqrt{114 - 5(19.36)}.\sqrt{163 - 5(21.16)}} = \frac{-29.2}{\sqrt{17.2}\sqrt{57.2}}$$

$$= -29.2/31.3662 = -\mathbf{0.93093}$$

Comment: There is high degree negative correlation.

Solution 2: Let X : Marks of subject A and Y : Marks of subject B

X	Y	X^2	Y^2	XY
3	0	9	0	0
5	3	25	9	15
8	6	64	36	48
4	5	16	25	20
9	9	81	81	81
6	4	36	16	24
7	8	49	64	56
0	2	0	4	0
42	37	280	235	244

$\bar{x} = 42/8 = 5.25$ and $\bar{y} = 37/8 = 4.625$

$$V(X) = \frac{\sum_{i=1}^{n} x_i^2}{n} - \bar{x}^2 = \frac{280}{8} - (5.25)^2 = 7.4375$$

$$\sigma x = \sqrt{Variance} = \sqrt{7.4375} = 2.7271$$

$$V(Y) = \frac{\sum_{i=1}^{n} y_i^2}{n} - \bar{y}^2 = \frac{235}{8} - (4.625)^2 = 7.9843$$

$$\sigma x = \sqrt{Variance} = \sqrt{7.9843} = 2.8256$$

$Cov(X,Y) = \frac{\sum xy}{n} - \bar{x}\bar{y} = \frac{244}{n8} - (5.25)(4.625) = 6.2187$

Correlation coefficient (r) $= \frac{Cov(x,y)}{\sigma_x \sigma_y} = \frac{6.2187}{(2.7271)(2.8256)} = 0.8069$

Comment: There is high degree positive correlation.

Solution 3: Given n=10, $\Sigma X = 80$, $\Sigma Y = 100$, $\Sigma X^2 = 700$, $\Sigma Y^2 = 1200$, $\Sigma XY = 870$

$$r = \frac{\sum XY - n.\bar{X}.\bar{Y}}{\sqrt{\sum X^2 - n.\bar{X}^2}.\sqrt{\sum Y^2 - n.\bar{Y}^2}} = \frac{870 - 10(8)(10)}{\sqrt{700 - 10(64)}.\sqrt{1200 - 10(100)}} = \frac{70}{\sqrt{60}\sqrt{200}} = 70/109.5445 = \mathbf{0.6390}$$

Comment: There is low degree positive correlation.

Solution 4: Given n = 10, $\Sigma X = 100$, $\Sigma Y = 220$, $\Sigma X^2 = 1024$, $\Sigma Y^2 = 4980$, $\Sigma XY = 2147$

$\bar{x} = 100/10 = 10$, $\bar{y} = 220/10 = 22$, $V(X) = \frac{\sum_{i=1}^{n} x_i^2}{n} - \bar{x}^2 = \frac{1024}{10} - 100 = 2.4$ and $\sigma x = \sqrt{Variance} = 1.5491$

$$V(Y) = \frac{\sum_{i=1}^{n} y_i^2}{n} - \bar{y}^2 = \frac{4980}{10} - 484 = 14$$ and $\sigma x = \sqrt{Variance} = 3.7416$

$Cov(X,Y) = \frac{\sum xy}{n} - \bar{x}\bar{y} = \frac{2147}{10} - (10)(22) = -5.3$

Correlation coefficient (r) $= \frac{Cov(x,y)}{\sigma_x \sigma_y} = \frac{-5.3}{(1.5491)(3.7416)} = -\mathbf{0.9143}$

Comment: There is high degree negative correlation.