Probability and Statistics

Mohit Singh

***Abstract*—This document provides the solution to the problem nos. 40 to 50 of proabability and 41-51 of statistics. The computations are done in python and excel.**

This documentation can be downloaded from

```
svn co https://github.com/mohit−singh−9/Summer
    −2020/tree/master/geometry/
    probability_statistics.git
```

1 PROBABILITY EXERCISE

1.1 Problem No. 40

1.1.1 Question: A die is tossed thrice. Find the probability of getting an odd number at least once.

1.1.2 Solution:

1.1. This is a binomial distribution where success is at getting an odd number.
1.2. No. of trials = $n = 3$.
1.3. Probability of getting an odd number = $p = \frac{3}{6} = \frac{1}{2}$.

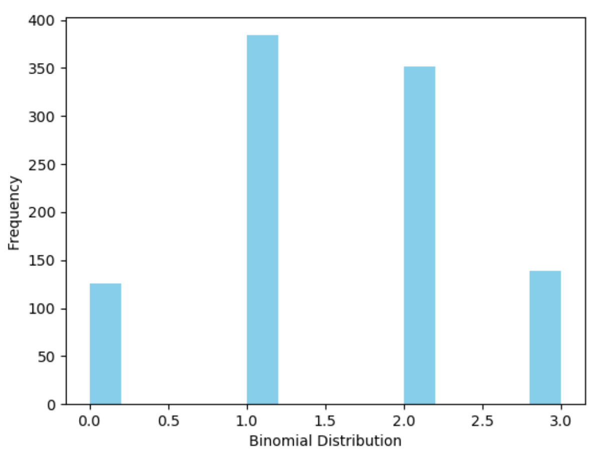

Fig. 1.3: Binomial distribution: n=3, p=0.5 (size =1000)

1.4. Then by binomial distribution

$$P(r \geq 1) = 1 - P(r = 0) \tag{1.4.1}$$
$$= 1 - {}^{n}C_r p^r (1-p)^{n-r} \tag{1.4.2}$$
$$= 0.875 \tag{1.4.3}$$

1.5. From the graph in fig. , we can find the answer.
1.6. Download the python code from

```
probability/codes/Q40binom.py
```

1.2 Problem No. 41

1.2.1 Question:

1.1. Two balls are drawn at random with replacement from a box containing 10 black and 8 red balls. Find the probability that
   a) both balls are red.
   b) first ball is black and second is red.
   c) one of them is black and other is red.

1.2.2 Solution:

1.1. Probability of picking a black ball = $P(B) = \frac{10}{18}$
1.2. Probability of picking a red ball = $P(R) = \frac{8}{18}$
1.3. Two balls are drawn with replacement. So each event is independent of each other.
1.4. Probability that both balls are red

$$= P(R)P(R) \tag{1.4.1}$$
$$= \frac{8}{18} \times \frac{8}{18} \tag{1.4.2}$$
$$= 0.1975 \tag{1.4.3}$$

1.5. Probability that first ball is black and second is red=

$$= P(B)P(R) \tag{1.5.1}$$
$$= \frac{10}{18} \times \frac{8}{18} \tag{1.5.2}$$
$$= 0.2469 \tag{1.5.3}$$

1.6. Probability that one ball is black and other is red =

$$= P(B)P(R) + P(R)P(B) \tag{1.6.1}$$
$$= \frac{10}{18} \times \frac{8}{18} + \frac{8}{18} \times \frac{10}{18} \tag{1.6.2}$$
$$= 0.4938 \tag{1.6.3}$$

1.7. The python code for finding probability using a sample size of 10000 can be downloaded from probability/codes/Q41.py

1.3 Problem No. 42

1.3.1 Question:

1.1. Probability of solving specific problem independently by A and B are $\frac{1}{2}$ and $\frac{1}{3}$ respectively. If both try to solve the problem independently, find the probability that

a) problem is solved
b) exactly one of them solves the problem

1.3.2 Solution:

1.1. Probability that A solves the problem = $P(A) = \frac{1}{2}$
1.2. Probability that B solves the problem = $P(B) = \frac{1}{3}$
1.3. A problem is solved when either A or B solves the problem or both solve the problem. So the probability that problem is solved =

$$= P(A)P(B)' + P(A)'P(B) + P(A)P(B) \tag{1.3.1}$$

$$= \frac{1}{2} \times \frac{2}{3} + \frac{1}{2} \times \frac{1}{3} + \frac{1}{2} \times \frac{1}{3} \tag{1.3.2}$$

$$= 0.667 \tag{1.3.3}$$

1.4. Probability that exactly one of them solves the problem =

$$= P(A)P(B)' + P(A)'P(B) \tag{1.4.1}$$

$$= \frac{1}{2} \times \frac{2}{3} + \frac{1}{2} \times \frac{1}{3} \tag{1.4.2}$$

$$= 0.5 \tag{1.4.3}$$

1.5. The python code for this can be downloaded from

probability/codes/Q42.py

1.4 Problem No. 43

1.4.1 Question:

1.1. One card is drawn at random from a well shuffled deck of 52 cards. In which of the following cases are the events E and F independent?

a) E :the card drawn is a spade, F :the card drawn is an ace.
b) E :the card drawn is black, F :the card drawn is a king.
c) E :the card drawn is a king or queen, F :the card drawn is a queen or jack.

1.4.2 Solution:

1.1. Two events E and F are said to be independent if they satisfy the criterion:

$$P(E \cap F) = P(E)P(F) \tag{1.1.1}$$

a) There are 13 cards of spades, 4 cards of aces and 1 card of ace of spades.

$$P(E) = \frac{13}{52} \tag{1.1.2}$$

$$P(F) = \frac{4}{52} \tag{1.1.3}$$

$$P(E \cap F) = \frac{1}{52} \tag{1.1.4}$$

Clearly, $P(E \cap F) = P(E)P(F)$. Therefore E and F are independent events.

b) There are 26 black cards, 4 king cards and 2 black and king cards.

$$P(E) = \frac{26}{52} \tag{1.1.5}$$

$$P(F) = \frac{4}{52} \tag{1.1.6}$$

$$P(E \cap F) = \frac{2}{52} \tag{1.1.7}$$

Clearly, $P(E \cap F) = P(E)P(F)$. Therefore E and F are independent events.

c) There are 8 kings or queens, 8 queens or jacks. In both of these, common is the quuen cards.

$$P(E) = \frac{8}{52} \tag{1.1.8}$$

$$P(F) = \frac{8}{52} \tag{1.1.9}$$

$$P(E \cap F) = \frac{4}{52} \tag{1.1.10}$$

Clearly, $P(E \cap F) \neq P(E)P(F)$. Therefore E and F are not independent events.

1.5 Problem No. 44

1.5.1 Question:

1.1. In a hostel, 60% of the students read Hindi newspaper, 40% read English newspaper and 20% read both Hindi and English newspapers. A student is selected at random.

a) Find the probability that she reads neither Hindi nor English newspapers.
b) If she reads Hindi newspaper, find the probability that she reads English newspaper.

c) If she reads English newspaper, find the probability that she reads Hindi newspaper.

1.5.2 Solution:

1.1. Let the total no. of students be 100. Then no.of students reading Hindi newspapers = 60, no.of students reading Engish newspaper= 40, no.of students reading both English and Hindi newspapers = 20.

1.2. So, $P(H) = 0.60, P(E) = 0.40, P(H \cap E) = 0.20$

a) Probability that the student reads neither Hindi nor English:

$$P(H \cup E)' = 1 - P(H \cup E) \tag{1.2.1}$$
$$= 1 - (P(E) + P(H) - P(H \cap E)) \tag{1.2.2}$$
$$= 1 - (0.60 + 0.40 - 0.20) \tag{1.2.3}$$
$$= 0.20 \tag{1.2.4}$$

b) Probability that she reads English newspaper if it is known that she reads Hindi newspaper

$$P(E \mid H) = \frac{P(E \cap H)}{P(H)} \tag{1.2.5}$$
$$= \frac{0.20}{0.60} \tag{1.2.6}$$
$$= 0.33 \tag{1.2.7}$$

c) Probability that she reads Hindi newspaper if it is known that she reads English newspaper

$$P(H \mid E) = \frac{P(E \cap H)}{P(E)} \tag{1.2.8}$$
$$= \frac{0.20}{0.40} \tag{1.2.9}$$
$$= 0.50 \tag{1.2.10}$$

1.6 Problem No. 45

1.6.1 Question:

1.1. The probability of obtaining an even prime number on each die when a pair of dice is rolled

a) 0
b) $\frac{1}{3}$
c) $\frac{1}{12}$
d) $\frac{1}{36}$

1.6.2 Solution:

1.1. 2 is the only even prime number which we can get. Therefore, prbability of getting 2 on a die $= P(A) = \frac{1}{6}$.

1.2. The number coming up on one die is independent of what comes on the other die.

1.3. Probability of getting an even prime number (i.e. 2) on both the die $= P(A)P(A) = \frac{1}{6}\times\frac{1}{6} = \frac{1}{36}$.

1.4. Answer = option (d)

1.7 Problem No. 46

1.7.1 Question:

1.1. Two events A and B are independent if

a) A and B are mutually exclusive.
b) $P(A'B') = (1 - P(A))(1 - P(B))$
c) $P(A) = P(B)$
d) $P(A) + P(B) = 1$

1.7.2 Solution:

1.1. A and B are not mutually exclusive because $P(A \cap B) = P(A) \times P(B)$ and it is not zero.

1.2. Also $P(A) = P(B)$ is not necessariy true.

1.3. $P(A) + P(B)$ is not always equal to 1.

1.4. $P(A'B') = P(A')P(B') = (1 - P(A))(1 - P(B))$

1.5. Answer= option(b)

1.8 Problem No. 47

1.8.1 Question:

1.1. An urn contains 5 red and 5 black balls. A ball is drawn at random, its colour is noted and is returned to the urn. Moreover, 2 additional balls of the colour drawn are put in the urn and then a ball is drawn at random. What is the probability that the second ball is red?

1.8.2 Solution:

1.1. Let E_1 be the event of getting a red ball in the first draw and E_2 be the evnt of getting a black ball in the first draw. If E_1 occurs then there will be 7 red balls and 5 black balls. If E_2 occurs then there will be 5 red and 7 black balls.

1.2. Let R be the event of drawing red ball in the second draw.

$$P(E_1) = \frac{5}{10} \tag{1.2.1}$$
$$P(E_2) = \frac{5}{10} \tag{1.2.2}$$
$$P(R \mid E_1) = \frac{7}{12} \tag{1.2.3}$$
$$P(R \mid E_2) = \frac{5}{12} \tag{1.2.4}$$

1.3. The Substituting the above values in (**??**)

$$P(R) = P(E_1)P(R \mid E_1) + P(E_2)P(R \mid E_2) \quad (1.3.1)$$

1.4. . Probability that second ball is red = 0.5.
1.5. The python code for finding probability using a sample size of 10000 can be downloaded from

```
probability/codes/Q47.py
```

1.9 Problem No. 48

1.9.1 Question:

1.1. A bag contains 4 red and 4 black balls, another bag contains 2 red and 6 black balls. One of the two bags is selected at random and a ball is drawn from the bag which is found to be red. Find the probability that the ball is drawn from the first bag.

1.9.2 Solution:

1.1. Let $P(E_1) and P(E_2)$ be the probability of choosing bag1 and bag2 respectively. So $P(E_1) = P(E_2) = \frac{1}{2}$.
1.2. Probability of drawing red ball from bag1= $P(R \mid E_1) = \frac{4}{8}$. Probability of drawing red ball from bag2 = $P(R \mid E_2) = \frac{2}{8}$
1.3. From Bayes theorem, probability that the red ball drawn is from bag1:

$$P(E_1 \mid R) = \frac{P(E_1) \times P(R \mid E_1)}{P(E_1) \times P(R \mid E_1) + P(E_2) \times P(R \mid E_2)} \quad (1.3.1)$$

$$= \frac{2}{3} \quad (1.3.2)$$

1.4. The python code for finding probability using a sample size of 10000 can be downloaded from

```
probability/codes/Q48.py
```

1.10 Problem No. 49

1.10.1 Question:

1.1. Of the students in a college, it is known that 60% reside in hostel and 40% are day scholars (not residing in hostel). Previous year results report that 30% of all students who reside in hostel attain A grade and 20% of day scholars attain A grade in their annual examination. At the end of the year, one student is chosen at random from the college and he has an A grade, what is the probability that the student is a hostlier?

1.10.2 Solution:

1.1. Probability of selecting a hostelier= $P(H) = 0.60$.
1.2. Probability of selecting a day scholar = $P(D) = 0.40$.
1.3. Probability of selecting an A grade hostelier= $P(A \mid H) = 0.30$.
1.4. Probability of selecting an A grade day scholar= $P(A \mid D) = 0.20$.
1.5. Using Bayes Theorem, probability that a selected A grade student is a hostelier is

$$P(H \mid A) = \frac{P(H) \times P(A \mid H)}{P(H) \times P(A \mid H) + P(D) \times P(A \mid D)} \quad (1.5.1)$$

$$= \frac{9}{13} \quad (1.5.2)$$

1.6. The python code for finding probability using a sample size of 10000 can be downloaded from

```
probability/codes/Q49.py
```

1.11 Problem No. 50

1.11.1 Question:

1.1. In answering a question on a multiple choice test, a student either knows the answer or guesses. Let $\frac{3}{4}$ be the probability that he knows the answer and $\frac{1}{4}$ be the probability that he guesses. Assuming that a student who guesses at the answer will be correct with probability $\frac{1}{4}$. What is the probability that the student knows the answer given that he answered it correctly?

1.11.2 Solution:

1.1. Probability that the student knows the answer= $P(X) = \frac{3}{4}$.
1.2. Probability that the student guesses the answer= $P(Y) = \frac{1}{4}$.
1.3. Probability that the student who guesses the answer will be correct= $P(C \mid Y) = \frac{1}{4}$.
1.4. if the student knows the answer then definitely he will be correct. So $P(C \mid X) = 1$.
1.5. using Bayes theorem, probability that the sudent knows the answer given that he answered it correctly:

$$P(X \mid C) = \frac{P(X) \times P(C \mid X)}{P(X) \times P(C \mid X) + P(Y) \times P(C \mid Y)} \quad (1.5.1)$$

$$= \frac{12}{13} \quad (1.5.2)$$

1.6. The python code for finding probability using a sample size of 10000 can be downloaded from

```
probability/codes/Q50.py
```

2 STATISTICS EXERCISE

2.1 Problem No. 41

2.1.1 Question:

2.1. The following table gives the lfetime of 400 neon lamps:

Life time (in hours)	**Number of lamps**
300-400	14
400-500	56
500-600	60
600-700	86
700-800	74
800-900	62
900-1000	48

a) Represent the given information with the help of a histogram.
b) How many lamps have a life time of more than 700 hours?

2.1.2 Solution:

2.1. The histogram for the above data is represented in fig. **??**

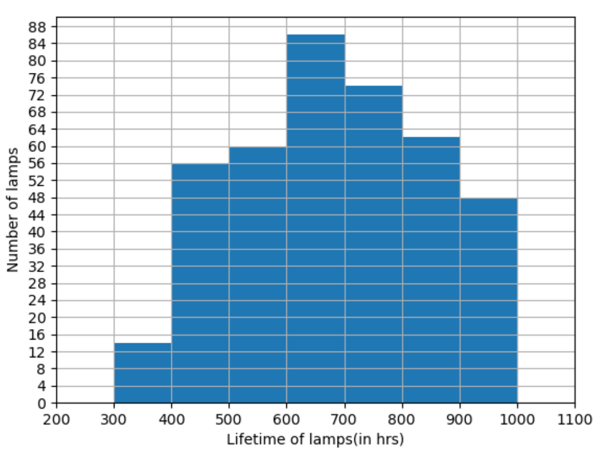

Fig. 2.1: Histogram showing No.of lamps vs. Life-time of lamps

2.2. From the graph, the no.of lamps having life-time grater than 700 hrs = 74+62+48=184.

2.3. Download the python code for the figure from

```
statistics/codes/Q41.py
```

2.2 Problem No. 42

2.2.1 Question:

2.1. The following table gives the distribution of students of two sections according to the marks obtained by them: Represent the marks of the

Section A		**Section B**	
Marks	**Frequency**	**Marks**	**Frequency**
0-10	3	0-10	5
10-20	9	10-20	19
20-30	17	20-30	15
30-40	12	30-40	10
40-50	9	40-50	1

students of both the sections on the same graph by two frequency polygons. From the two polygons compare the performance of the two sections.

2.2.2 Solution:

2.1. The figure **??** represents the frequency polygon of two sections.

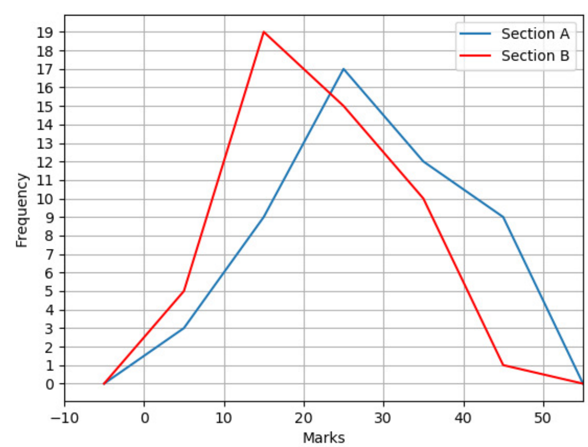

Fig. 2.1: Frequency polygon of Section A and Section B

2.2. From the graph the following can be drawn
a) Section A has more number of students who got more marks in the range 30-50.
b) Section B has more students in the marks range of 0-20 than Section A.
c) Section A has performed better than Section B.

2.3 Problem No. 43

2.3.1 Question:

2.1. The runs scored by two teams A and B on the first 60 balls in a cricket match are given below: Represent the data of both the teams on

No.of balls	Team A	Team B
1? 6	2	5
7 – 12	1	6
13-18	8	2
19-24	9	10
25-30	4	5
31-36	5	6
37-42	6	3
43-48	10	4
49-54	6	8
55-60	2	10

the same graph by frequency polygons.

2.3.2 Solution:

2.1. First we need to make the class intervals continuous. For this we need to subtract and add 0.5 to lower class limit and upper class limit respectively. The continuous class intervals are represented in **??**.

No.of balls	Team A	Team B
0.5- 6.5	2	5
6.5-12.5	1	6
12.5-18.5	8	2
18.5-24.5	9	10
24.5-30.5	4	5
30.5-36.5	5	6
36.5-42.5	6	3
42.5-48.5	10	4
48.5-54.5	6	8
54.5-60.5	2	10

TABLE 2.1: Continuous class intervals of given data

2.2. The frequency polygons of given data set is represented in fig. **??**

2.3. The python code for the plot can be downloaded from

```
statistics/codes/Q43.py
```

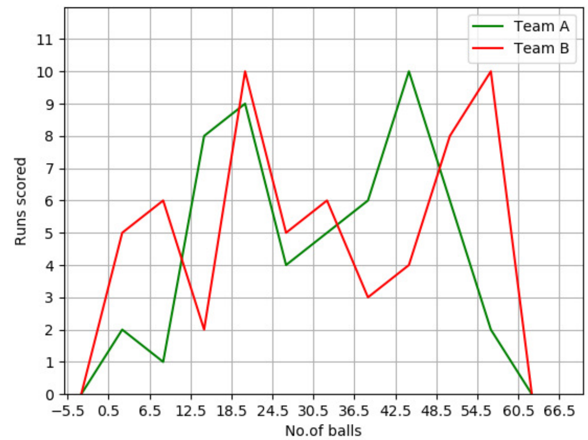

Fig. 2.2: Frequency polygon of Team A and Team B

2.4 Problem No. 44

2.4.1 Question:

2.1. A random survey of the number of children of various age groups playing in a park was found as follows: Draw a histogram to plot the above

Age (in years)	No.of children
1- 2	5
2- 3	3
3- 5	6
5- 7	12
7- 10	9
10- 15	10
15- 17	4

data.

2.4.2 Solution:

2.1. The histogram of given data set is represented in fig. **??**

2.2. The python code for the plot can be downloaded from

```
statistics/codes/Q44.py
```

2.5 Problem No. 45

2.5.1 Question:

2.1. 100 surnames were randomly picked up from a local telephone directory and a frequency distribution of the number of letters in the

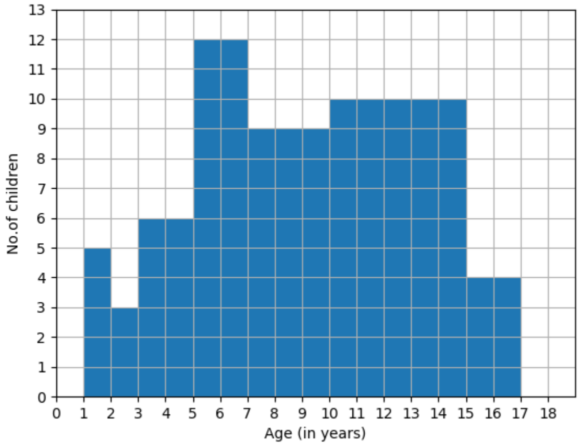

Fig. 2.1: Histogram showing No.of children vs. age(in years)

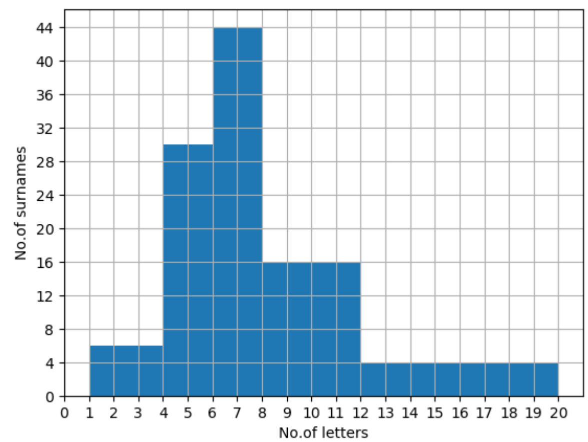

Fig. 2.1: Histogram of no.of surnames vs. no.of letters

No.of letters	**No.of surnames**
1- 4	6
4- 6	30
6- 8	44
8- 12	16
12- 20	4

English alphabet in the surnames was found as follows:

a) Draw a histogram to depict the given information.
b) Write the class interval in which the maximum number of surnames lie.

2.5.2 Solution:

2.1. The histogram for the above data is plotted in the fig.
2.2. From the graph, we can easily say that the maximum number of surnames is 44 which lies in the interval of 6 - 8.
2.3. Download the python code for the figure from

```
statistics/codes/Q45.py
```

2.6 Problem No. 46

2.6.1 Question:

2.1. The following number of goals were scored by a team in a series of 10 matches:
2, 3, 4, 5, 0, 1, 3, 3, 4, 3
Find the mean, median and mode of these scores.

2.6.2 Solution:

2.1. First we will sort the given data in ascending order.
2.2. 0,1,2,3,3,3,3,4,4,5
2.3. Let n= No.of elements in the given data= 10.

$$Mean = \frac{Sumofdatavalues}{n} \tag{2.3.1}$$
$$= \frac{0+1+2+3+3+3+3+4+4+5}{10} \tag{2.3.2}$$
$$= 2.8 \tag{2.3.3}$$
$$Mode = 3 \tag{2.3.4}$$

2.4. Since n is even , the median is given by the average of $\left(\frac{n}{2}\right)^{th}$ and$\left(\frac{n}{2}+1\right)^{th}$ element.

$$Median = \frac{3+3}{2} \tag{2.4.1}$$
$$= 3 \tag{2.4.2}$$

2.5. The python code to calculate mean, median , mode for the given data can be downloaded from

```
statistics/codes/Q46.py
```

2.7 Problem No. 47

2.7.1 Question:

2.1. In a mathematics test given to 15 students, the following marks (out of 100) are recorded:
41, 39, 48, 52, 46, 62, 54, 40, 96, 52, 98, 40,

42, 52, 60

Find the mean, median and mode of this data.

2.7.2 Solution:

2.1. First we will sort the given data in ascending order.

2.2. 39,40,40,41,42,46,48,52,52,52,54,60,62,96,98

2.3. Let n= No.of elements in the given data= 15.

$$Mean = \frac{Sumofdatavalues}{n} \tag{2.3.1}$$
$$= 54.8 \tag{2.3.2}$$
$$Mode = 52 \tag{2.3.3}$$

2.4. Since n is odd , the median is the $\left(\frac{n+1}{2}\right)^{th}$ element.

$$Median = 52 \tag{2.4.1}$$

2.5. The python code to calculate mean, median , mode for the given data can be downloaded from

```
statistics/codes/Q47.py
```

2.8 Problem No. 48

2.8.1 Question:

2.1. The following observations have been arranged in ascending order. If the median of the data is 63, find the value of x.

29, 32, 48, 50, x, x + 2, 72, 78, 84, 95.

2.8.2 Solution:

2.1. The no.of elements in the given data=n= 10

2.2. Since n is even , the median is given by the average of $\left(\frac{n}{2}\right)^{th}$ and$\left(\frac{n}{2}+1\right)^{th}$ element.

$$\frac{x+x+2}{2} = 63 \tag{2.2.1}$$
$$x = 62 \tag{2.2.2}$$

2.9 Problem No. 49

2.9.1 Question:

2.1. Find the mode of

14, 25, 14, 28, 18, 17, 18, 14, 23, 22, 14, 18.

2.9.2 Solution:

2.1. Mode is the value which occurs for the maximum time. Here the mode is 14.

2.2. The python code to find mode can be downloaded from

```
statistics/codes/Q49.py
```

2.10 Problem No. 50

2.10.1 Question:

2.1. Find the mean salary of 60 workers of a factory from the following table:

Salary (in Rs.)	No.of Workers
3000	16
4000	12
5000	10
6000	8
7000	6
8000	4
9000	3
10000	1
Total	60

2.10.2 Solution:

2.1. To find the mean, we need to multiply Salary and No.of workers. This can be seen in given table **??** :

Salary (in Rs.)	No.of Workers	Salary X No.of workers
3000	16	48000
4000	12	48000
5000	10	50000
6000	8	48000
7000	6	42000
8000	4	32000
9000	3	27000
10000	1	10000
Total	60	305000

TABLE 2.1: Finding mean of the given data

$$Mean = \frac{305000}{60} \tag{2.1.1}$$
$$= 5083.33 \tag{2.1.2}$$

2.11 Problem No. 51

2.11.1 Question:

2.1. Give one example of a situation in which

a) the mean is an appropriate measure of central tendency.

b) the mean is not an appropriate measure of central tendency but the median is an appropriate measure of central tendency.

2.11.2 Solution:

2.1. When the data set has normal distribution with no skewness, such that the adjacent values dont show bigger deviation from adjacent values then in that case mean is an appropriate measure of central tendency. Example: Birth weight, Height(over a large population)

2.2. When the skewness increases the difference between mean and median increases. Hence median becomes the appropriate measure of central tendency. Example: Income distribution in a developing country.