

**ROYAL CIVIL SERVICE COMMISSION
BHUTAN CIVIL SERVICE EXAMINATION (BCSE) 2019
EXAMINATION CATEGORY: TECHNICAL**

PAPER II: GENERAL SUBJECT KNOWLEDGE PAPER FOR STATISTICS

Date	: October 12, 2019
Total Marks	: 100
Writing Time	: 90 minutes (1.5 hours)
Reading Time	: 15 Minutes (prior to writing time)

GENERAL INSTRUCTIONS:

1. Write your Registration Number clearly and correctly on the Answer Booklet.
2. The first 15 minutes is to check the number of pages of Question Paper, printing errors, clarify doubts and to read the instructions. You are NOT permitted to write during this time.
3. This paper consists of **TWO Parts: Part I & Part II**

Part I consists of 70 multiple choice questions of 1 (one) mark each, and
Part II consists of 10 short answer questions of 3 (three) marks each.
4. All questions are **COMPULSORY**.
5. All answers should be written on the Answer Booklet provided to you. Candidates are not allowed to write anything on the question paper. If required, ask for additional Answer Booklet.
6. All answers should be written with correct numbering of Part and Question Number in the Answer Booklet provided to you. Note that any answer written without indicating the correct Part and Question Number will NOT be evaluated and no marks will be awarded.
7. Begin each Part in a fresh page of the Answer Booklet.
8. You are not permitted to tear off any sheet(s) of the Answer Booklet as well as the Question Paper.
9. Use of any other paper including paper for rough work is not permitted.
10. **You are required to hand over the Answer Booklet to the Invigilator before leaving the examination hall.**
11. This paper has **16 printed pages**, including this instruction page.

GOOD LUCK

Part I

Multiple Choice Questions (70 marks)

Choose the correct answer and write down the letter of your chosen answer in the Answer Booklet against the question number e.g. 71 (a). Each question carries ONE mark. Any double writing, smudgy answers or writing more than one choice shall not be evaluated.

1. Which of the following is **NOT** the subject of matter in statistics?
 - a) Appointment
 - b) Presentation of data
 - c) Analysis of data
 - d) Interpretation of data

2. The statistical data in the original form before any statistical techniques are applied to refine, process or summarize is called
 - a) Secondary data.
 - b) Primary data.
 - c) Statistical information.
 - d) Data in general.

3. What statistics can we obtain by adding all numbers in a distribution and then total is divided by number of scores in that distribution?
 - a) Geometric mean
 - b) Arithmetic mean
 - c) Standard deviation
 - d) Variance

4. The most frequently occurring score in a data set is
 - a) Mean
 - b) Median
 - c) Mode
 - d) Midpoint

5. The median is always
 - a) the most frequently occurring score in a data set.
 - b) the middle score when results are ranked in order of magnitude.
 - c) the same as the mean.
 - d) the difference between the maximum and minimum scores.

6. The half of the observations in a data set are greater than the
 - a) Mean
 - b) Median
 - c) Mode
 - d) Standard deviation

7. The mean, median and mode of the following set of five observations (4, 4, 5, 7, 10) are
 - a) Mean=6, median=5 and mode=4
 - b) Mean=6, median=4 and mode=5
 - c) Mean=6, median=5 and mode=7
 - d) Mean=6, median=4 and mode=4

8. The mean, median and mode are part of
 - a) Measures of central tendency.
 - b) Measures of dispersion.
 - c) Logarithms.
 - d) Proportions.

9. The sum of deviations of values from their mean is always equal to
 - a) 1
 - b) 0
 - c) 2
 - d) 3

10. Which of the following is the measure of the variation?
 - a) Mode
 - b) Mean
 - c) Hypothesis
 - d) Dispersion

11. Summary statistics for two distributions A and B are as follows:
A: mean=8 and variance=22.8
B: mean=8 and variance=17.2
Which distribution has the larger spread of observations?
 - a) Distribution A.
 - b) Distribution B.
 - c) Neither; they have same spread.
 - d) There is not enough information to answer the question.

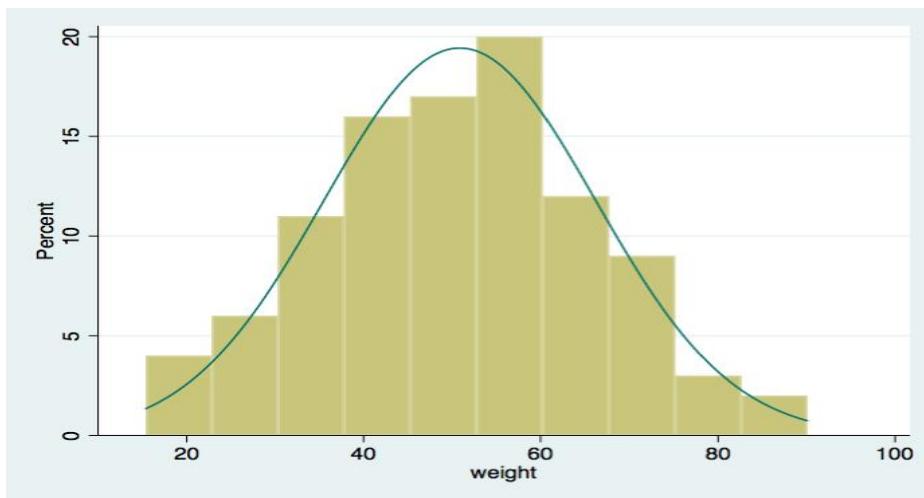
12. Which one of the following is a measure of dispersion?
 - a) Standard deviation
 - b) Variance
 - c) Co-efficient of variation
 - d) None of the above.

13. The mean of a distribution is 14 and the standard deviation is 5. What is the coefficient of variation?
 - a) 60.4%
 - b) 48.3%
 - c) 35.7%
 - d) 27.8%

14. Suppose, the RCSC wants to know the number of students who appeared for the BCSE Main Exam. What kind of graph would be appropriate to present the frequency distributions of these data?
 - a) Histogram
 - b) Scatterplot
 - c) Bar chart
 - d) Box plot

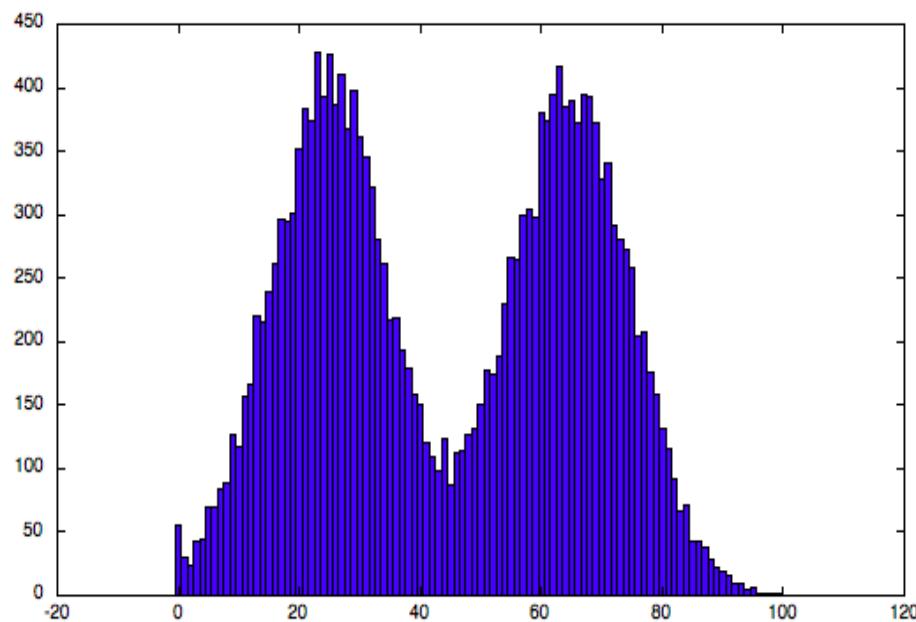
15. Normally distributed data are usually referred to as
- Asymmetrical
 - Skewed
 - Bell-shaped
 - Peaked
16. Which of the following would indicate that a dataset is **NOT** bell-shaped?
- The range is equal to 5 standard deviation.
 - The range is larger than the interquartile range.
 - The mean is much smaller than the median.
 - There are no outliers.

Use the histogram below to answer Questions 17 to 19. It shows the distribution of the weights of students in Bhutan. Note that the percent is given on the vertical axis.



17. What is the approximate shape of the distribution?
- Nearly symmetric.
 - Skewed to the left.
 - Skewed to the right.
 - Bimodal (has more than one peak).
18. The median of the distribution is approximately
- 50 kg
 - 60 kg
 - 70 kg
 - 80 kg
19. A bar chart constructed in which area of each bar is proportional to number of items in each group is known as
- Pie chart
 - Histogram
 - Frequency distribution
 - Polygon

20. The following is a diagram of a histogram.



Which of the following is true about the above histogram?

- a) The above histogram is unimodal.
- b) The above histogram is bimodal.
- c) The above is not a histogram.
- d) None of the above.

21. Which of the following is not based on all the observations?

- a) Arithmetic mean
- b) Geometric mean
- c) Harmonic mean
- d) Weighted mean

22. A researcher studies the factors that determine the number of children by married couples in future. The variable '**number of children**' is a

- a) Discrete variable
- b) Continuous variable
- c) Categorical variable
- d) Ordinal variable

23. A researcher asks respondents, '**Do you believe the Government should introduce consumption tax?** Yes No '. The data produced is an example of:

- a) Quantitative data
- b) Time series data
- c) Nominal data
- d) Ordinal data

24. Your height and weight are examples of which level of measurement?

- a) Nominal
- b) Ordinal
- c) Discrete
- d) Continuous

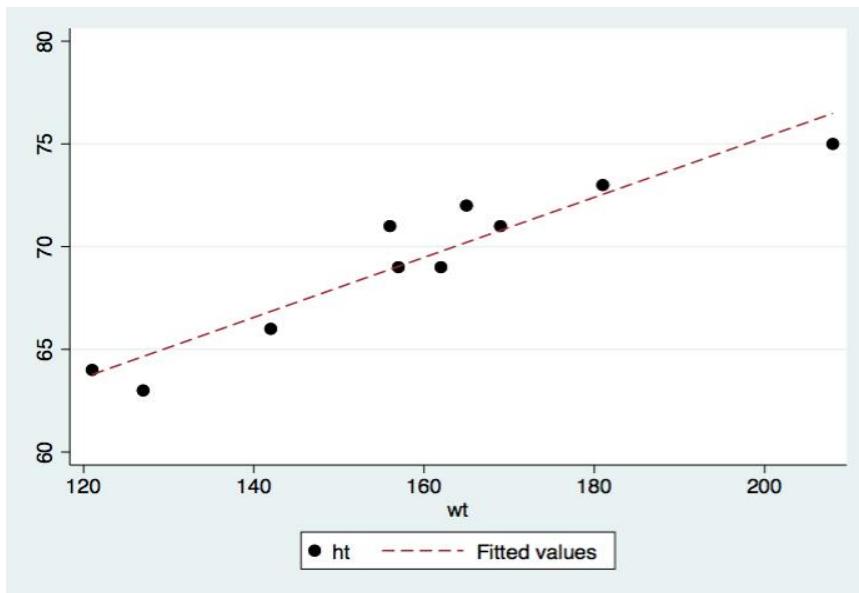
25. Three points that divides the distribution into 4 equal groups are called
a) Deciles
b) Percentiles
c) Quartiles
d) None of the above.
26. The value of $[(n+1)/4]^{\text{th}}$ item is the formula of
a) 1st quartile.
b) 2nd quartile.
c) Mode.
d) 3rd quartile.
27. The 1st quartile is also known as
a) Lower quartile
b) Median
c) Upper quartile
d) Geometric mean
28. The 2nd quartile is also known as
a) Mode
b) Harmonic mean
c) Arithmetic mean
d) Median
29. The N^{th} root of product of values $x_1, x_2, x_3, \dots, x_n$ is called
a) Arithmetic mean
b) Geometric mean
c) Variance
d) Harmonic mean
30. In a week, the prices of a kilogram of rice from different stores are Nu.350, 280, 340, 290, 320, 310, 300. The range is
a) 60
b) 70
c) 80
d) 100
31. Which one of these statistics is unaffected by outliers?
a) Mean
b) Interquartile range
c) Standard deviation
d) Range
32. The square of standard deviation is called
a) Harmonic mean
b) Variance
c) Mode
d) 2nd quartile

33. The table which shows frequency of each score is called
- Polygon
 - Pie chart
 - Histogram
 - Frequency distribution table
34. In statistics, a population consists of
- all people living in a country.
 - all people living in the area under the study.
 - all subjects or objects whose characteristics are being studied.
 - None of the above.
35. A summary measure calculated from population is called
- a Sample
 - a Statistic
 - a Census
 - a Parameter
36. A statistic is
- a sample characteristic.
 - a population characteristic.
 - unknown.
 - normal normally distributed.
37. A parameter is
- a sample characteristic.
 - a population characteristic.
 - unknown.
 - normal normally distributed.
38. Observational studies allow
- Population inference.
 - Causal inference.
 - Both types of inference.
 - Neither type of inference.
39. What is the definition of the standard error?
- Standard deviation of the sample.
 - Squared standard deviation.
 - Standard deviation of the sample mean.
 - Standard deviation of the population mean.
40. Failing to reject the null hypothesis when it is false is
- Alpha
 - type I error
 - Beta
 - type II error
41. Which one of the following is correct?
- The probability of a type II error is $(1-\beta)$.
 - The probability of a type II error is α .
 - The probability of a type I error is $(1-\alpha)$.
 - None of the above.

42. A researcher selects a probability of 100 out of the total population. The process is called
- a cluster sample
 - a systematic sample
 - a random sample
 - a stratified sample
43. When asked questions concerning personal income, people normally tend to lie. This is an example of
- sampling bias
 - confounding
 - non-response bias
 - response bias
44. Select the order of sampling schemes from best to worst:
- Simple random, stratified, convenience.
 - Simple random, convenience, stratified.
 - Stratified, simple random, convenience.
 - Stratified, convenience, simple random.
45. The graph of variables having linear relationship will be
- curved
 - hyperbola
 - straight line
 - None of the above.
46. The coefficient of correlation will be always
- more than 0
 - more than -1
 - less than -1
 - between -1 and +1.
47. When the correlation coefficient, r , is close to one:
- There is no relationship between the two variables.
 - There is a strong linear relationship between the two variables.
 - It is impossible to tell if there is a relationship between the two variables.
 - The slope of the regression line will be close to one.
48. Suppose there is a correlation of $r=0.9$ between the number of hours per day a student study and GPA score. Which of the following is a reasonable conclusion?
- 90% of students who study receive high grades
 - 90 of students who receive high grades study a lot.
 - 90% of the variation in GPA scores can be explained by variation in number of study hours.
 - 81% of the variation in GPA scores can be explained by variation in number of study hours per day.

49. What is the effect of an outlier on the value of a correlation coefficient?
- a) An outlier will always decrease a correlation coefficient.
 - b) An outlier will always increase a correlation coefficient.
 - c) An outlier might either decrease or increase a correlation coefficient, depending on where it is in relation to the other points.
 - d) An outlier will have no effect on a correlation coefficient.
50. What does ANOVA stand for?
- a) Analysis of values and averages.
 - b) Analysis of variability.
 - c) Analysis of non-ordinal values.
 - d) Analysis of variance.
51. IQ tests are standardized so that the mean score is 100 for the entire group of people who take the test. However, if you select a group of 50 who took the test, you probably would not get 100. What statistical concept explains the difference between the two means?
- a) Statistical error
 - b) Inferential error
 - c) Residual error
 - d) Sampling error
52. Sampling error can be reduced by
- a) non-probability sampling.
 - b) increasing the population size.
 - c) decreasing the sample size.
 - d) increasing the sample size.
53. Which statistical test is used to identify whether there is a relationship between two categorial variables?
- a) Student's t-test.
 - b) Spearman's correlation test.
 - c) Pearson's chi-square test.
 - d) Mann-Whitney test.
54. A scatterplot shows:
- a) The frequency with which values appear in the data.
 - b) The average value of groups of data.
 - c) Scores on one variable plotted against scores on a second variable.
 - d) The proportion of data falling into different categories.

55. The scatter plot between the heights and weights of ten students are shown below.



The data in the scatter plot above would have a correlation that is close to:

- a) -1.0
- b) -0.5
- c) +1.0
- d) +0.5

56. A Pearson correlation of $r=-0.6$ indicates

- a) an increase in X is accompanied by an increase in Y; the relationship is strong.
- b) an increase in X is accompanied by an increase in Y; the relationship is moderate.
- c) an increase in X is accompanied by a decrease in Y; the relationship is strong.
- d) an increase in X is accompanied by a decrease in Y; the relationship is moderate.

57. The r^2 is the notation for

- a) the coefficient of correlation.
- b) the coefficient of determination.
- c) the coefficient of variation.
- d) the coefficient of regression.

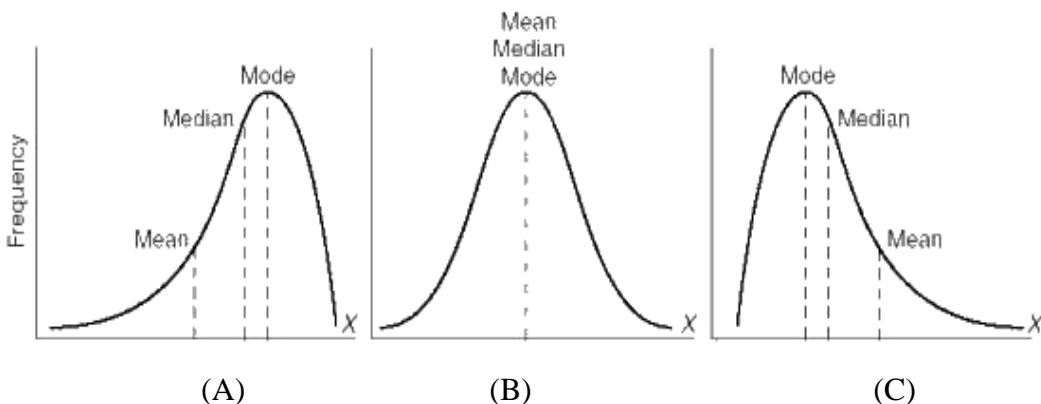
58. Suppose the correlation between the height and weight for adults is +0.80. What proportion of the variability in weight can be explained by the relationship with height?

- a) 20%
- b) 36%
- c) 64%
- d) 80%

59. If a relationship between height and weight of a student is statistically significant, it means that the variables are

- a) related in the sample due to chance alone.
- b) not related in the population represented by the sample.
- c) related in the population represented by the sample.
- d) very important.

60. Simpson's Paradox occurs when
- no baseline risk is given, so it is not known whether or not a high relative risk has practical importance.
 - a confounding variable rather than the explanatory variable is responsible for a change in the response variable.
 - the direction of the relationship between two variables changes when the categories of a confounding variable are considered.
 - the results of a test are statistically significant but are really due to chance.
61. If three letters are to be put in three addressed envelopes randomly, the probability that none of the letters are in the correct envelope is
- 0
 - $1/6$
 - $1/3$
 - $1/2$
62. For which of the following distributions, mean and variance are equal?
- Normal distribution
 - Poisson distribution
 - Binomial distribution
 - Negative Binomial distribution
63. Use the following diagram to answer the question.



- Which of the above diagram is positively skewed distribution?
- Diagram A
 - Diagram B
 - Diagram C
 - None of the above.
64. A coin is tossed five times in succession. What is the probability of getting at least four heads?
- $1/4$
 - $3/4$
 - $1/16$
 - $3/16$

PAPER II: GENERAL SUBJECT KNOWLEDGE PAPER FOR STATISTICS

65. According to the empirical rule, the bell or mound shaped distribution will have approximately 68% of the data within what number of standard deviations of the mean?
- a) one standard deviation.
 - b) two standard deviations.
 - c) three standard deviations.
 - d) four standard deviations.
66. If a population A has a larger standard deviation than population B, which of the following is **NOT** true?
- a) Population B has a smaller variance than population A.
 - b) The mean of a sample of 20 from population A has a larger standard deviation than the mean of a sample of 20 from population B.
 - c) A typical observation from population A will be farther from the mean of population A than a typical observation from B will be from the mean of population B.
 - d) The mean of a sample from population A will be on average be larger than the mean of a sample from population B.
67. A researcher is studying students of Bhutan. She takes a sample of 150 students from 10 schools. The average age of all students in Bhutan is
- a) a statistic
 - b) a parameter
 - c) the median
 - d) a population
68. We can measure the cause and effect relationship by the help of
- a) Time series analysis
 - b) Regression analysis
 - c) Cross-sectional analysis
 - d) Correlation analysis
69. The goodness of fit of a distribution is tested by
- a) t-test.
 - b) chi-square test.
 - c) f-test.
 - d) None of the above.
70. F-test is used to test the significance of the difference between/among
- a) two sample mean.
 - b) more than two sample mean.
 - c) variable of two sample.
 - d) Both (b) and (C).

PART II – Short Answer Questions (30 marks)

Answer ALL 10 short answer questions. Each question carries 3 marks.

1. Find the arithmetic means of the following distributions:

Distribution A:

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

Distribution B:

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

2. Consider the following sample dataset: 8, 2, 6, x, 5.

You are told that the value of the sample mean is 6. Find the following:

- a) The value of x.
- b) The sample variance.

3. The average salary of male employees in a firm is Nu 5200 and that of female is Nu 4200. The mean salary of all the employees is Nu 5000. Find the percentage of male and female employees.
4. In a factory employing 3,000 persons, 5 percent earn less than Nu 3 per hour, 580 earn from Nu 3.01-4.50 per hour, 30 percent earn from Nu 4.51-6.00 per hour, 500 earn from Nu 6.01-7.50 per hour, 20 percent earn from Nu 7.51-9.00 per hour, and the rest earn Nu 9.01 or more per hour. What is the median wage?

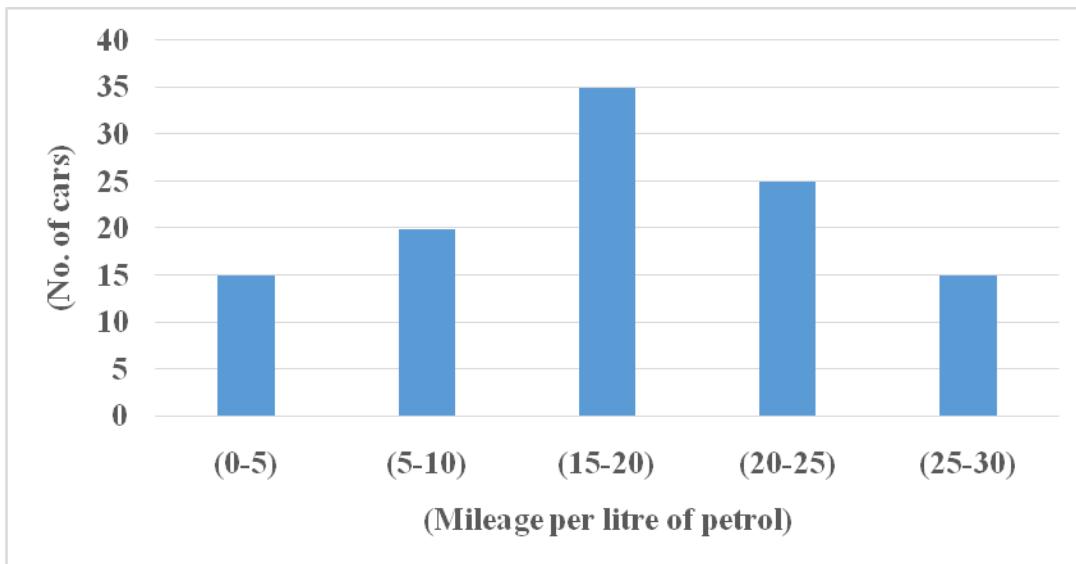
5. 24 patients had a blood test and the results are shown below:

A, B, B, AB, AB, B, O, O, AB, O, B, A, AB, A, O, O, AB, B, O, A, AB, O, B, A

Answer the following:

- a) Construct a frequency distribution for the data and find the mode of the distribution.
- b) If a person is selected randomly from the group of 24 patients, what is the probability that his/her blood type is not O? (Your answer should be rounded to 2 decimal places).

6. The bar graph below shows the efficiency level (in km per litre of petrol) of 110 cars.



Using the diagram above, answer the following:

- a) How many cars have an efficiency between 15 and 20 miles per litre of petrol?
 - b) How many cars have an efficiency more than 20 miles per litre of petrol?
 - c) What percentage of cars have an efficiency less than 20 miles per litres of petrol?
7. Two coins are tossed simultaneously. Find the following:
- a) Mean; and
 - b) Variance.
8. The variable X takes the values 1, 2, 3 and 5 according to the following distribution:

X=x	1	2	3	5
P(X=x)	0.1	0.3	0.4	0.2

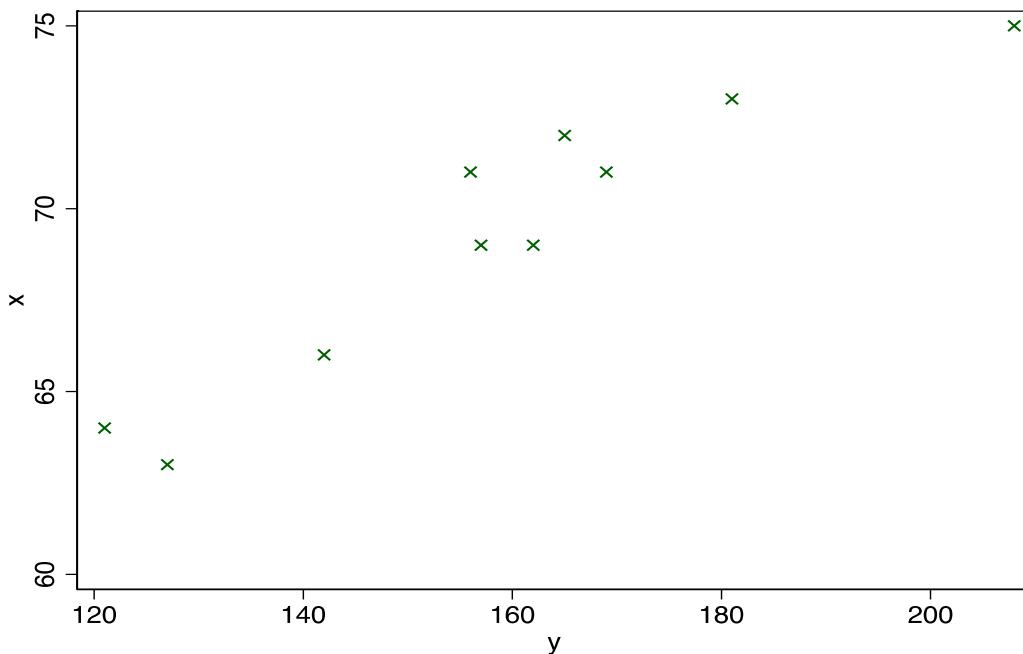
- a) What is the probability that X is negative?
- b) Find $E(X)$, the expected value of X .
- c) Find the probability that $X^2 > 8$.

Use the following table to answer Questions 9 and 10.

A statistical officer in National Statistics Bureau wants to study the relationship between the height (x) and the weight (y) of 10 employees. The results were recorded as follows:

X	63	64	66	69	69	71	71	72	73	75
Y	127	121	142	157	162	156	169	165	181	208

Based on the above information, the scatterplot was drawn as follows:



9. Find the simple correlation co-efficient between heights and weights of the 10 employees.
10. How do you interpret your result of the correlation co-efficient between the heights and weights?

LIST OF STATISTICAL FORMULA

Measures of Central Location and Dispersion:

1. $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$
2. $s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{(n-1)}$
3. $\mu = \frac{\sum_{i=1}^n x_i}{N}$
4. $\sigma^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{N}$
5. $\mu = E(x) = \sum_{i=1}^N x_i P(x_i)$
6. $\sigma^2 = V(x) = \sum_{i=1}^N (x_i - \mu)^2 P(x_i) = \sum x_i^2 P(x_i) - \mu^2$
7. Coefficient of variation (CV) = $\frac{SD}{\bar{x}} * 100$

Normal Distribution:

8. $x \sim \text{Normally } (\mu, \sigma^2)$
9. $z = \frac{x - \mu}{\sigma}$

Confidence Intervals:

10. $\bar{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$
11. $\bar{x} \pm t_{\frac{\alpha}{2}, (n-1)} \frac{s}{\sqrt{n}}$
12. $n = \frac{z^2 \alpha \sigma^2}{B^2}$

Correlation Co-efficient:

13. $r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{(x - \bar{x})^2(y - \bar{y})^2}}$