Statistics for Data Science 1

Number of Questions : 14

Section Marks : 50

Mark As Answered Required? : Yes

Question Number : 65

Correct Marks : 3

Question Label : Multiple Choice Question

The mean annual salary paid to all employees in a company is ₹66 lakhs. The mean annual salaries paid to male and female employees of the company are ₹50 lakhs and ₹70 lakhs respectively. Then, the percentage of males employed by the company is:

Options :

1. ✖ 60%

2. ✖ 80%

3. ✔ 20%

4. ✖ 40%

5. ✖ 50%

Question Number : 66

Correct Marks : 3

Question Label : Multiple Choice Question

If first quartile (Q_1) = 30 and third quartile (Q_3) = 40, which of the following must be true?

I. The variance is at most 100.

II. The median is 35.

III. The mean is between 30 and 40.

Options :

1. ✖ I only

2. ✖ II only

3. ✖ III only

4. ✖ I and II

5. ✖ All I,II and III are true.

6. ✔ None is true.

Question Number : 67

Correct Marks : 3

Question Label : Multiple Choice Question

Suppose the correlation coefficient between two variables *x* and *y* is 0.47. What will be the new correlation coefficient if 0.15 is added to all values of the *x* variable, every value of *y* variable is doubled, and the two variables are interchanged?

Options :

1. ✔ 0.47

2. ✖ 0.32

3. ✖ 0.94

4. ✖ 0.62

5. ✖ 0.97

Question Number : 68

Correct Marks : 3

Question Label : Multiple Choice Question

Which of the following statements about correlation coefficient, *r*, is true?

Options :

1. ✖ A correlation coefficient of 0.80 indicates a linear relationship whose slope is 4 times that of data whose correlation is 0.20.

2. ✖ A correlation coefficient of 0.4 means that 40% of the points are highly correlated.

3. ✔ Correlation coefficient is affected by outliers.

4. ✖ Correlation coefficient of variable *x* with variable *y* need not be the same as correlation coefficient of *y* with *x*.

Question Number : 69

Correct Marks : 1

Question Label : Multiple Choice Question

If the standard deviation of a set of non-zero observations is zero, you can conclude

Options :

1. ✖ that the mean (average) value is zero.

2. ✖ that the observations have same number of positive and negative data points.

3. ✔ that all observations have the same value.

4. ✖ that a mistake in calculation has been made.

5. ✖ none of these

Question Number : 70

Correct Marks : 1

Question Label : Multiple Select Question

Consider various variables that describe a used Royal Enfield motorcycle. These variables include price(INR lakhs), mileage(km/litre), model type, and model year. Their values are collected for 250 used motorcycles which are then organised in a data table. Based on this information, choose the correct option(s) from below.

Options :

1. ✔ The number of cases/observations in the data table is 250.

2. ✖ The number of cases/observations in the data table is 4.

3. ✔ Model type is a categorical variable.

4. ✖ Mileage is a categorical variable.

5. ✔ Price is a numerical variable.

Question Number : 71

Correct Marks : 3

Question Label : Multiple Select Question

The bar chart given in Figure Q.1 shows the shoe sizes of a group of 50 children. Based on this information, which of the following statements is(are) true?

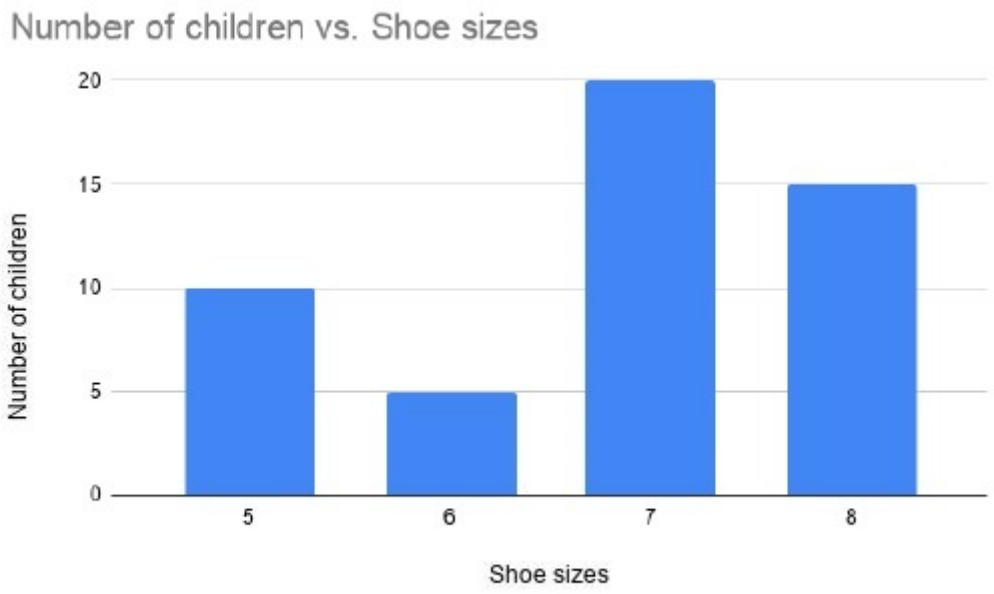

Figure Q.1: Shoe size dataset

Options :

1. ✖ 15 children wear size 7 shoes.

2. ✖ 30 children wear shoes of size less than 8.

3. ✔ 7 is the modal shoe size.

4. ✖ 6 is the median shoe size.

5. ✖ Range of the shoe size is 4.

6. ✖ The value of the first quartile (Q_1) for the shoe size is 5.

7. ✔ 35 children wear shoes size larger than 6.

Question Number : 72

Correct Marks : 3

Question Label : Multiple Select Question

In a call center, there are 100 employees and the number of calls they receive vary over the length of the day. The working hours are 9 AM to 6 PM with lunch break from 1 PM to 2 PM. The average number of calls received from 9 AM to 1 PM by an employee is 6 per hour, and the average number of calls received by the employee from 2 PM to 6 PM is 10 per hour. Based on this data, choose the correct options from below.

Options :

1. ✖ Average number of calls received by an employee in working hours is 10 *calls/hour*.

2. ✔ Average number of calls received by an employee in working hours is 8 *calls/hour*.

3. ✔ The correlation coefficient of time and calls received is positive.

4. ✖ The correlation coefficient of time and calls received is negative.

5. ✖ The standard deviation of the calls received is equal to zero.

6. ✖ The slope of the trend line is negative.

Question Number : 73

Correct Marks : 3

Question Label : Multiple Select Question

The correlation was found to be r = - 0.86 between price (x) and demand of mobile phones (y). Which of the following options could be true?

Options :

1. ✔ Given two points from the scatter plot of price and demand of mobile phones, one point has a smaller x value and a larger y value than another point.

2. ✔ Given two points from the scatter plot of price and demand of mobile phones, one point has a larger x value and a smaller y value than another point.

3. ✔ The covariance of price and demand of mobile phones is negative.

4. ✖ The covariance of price and demand of mobile phones is positive.

Question Number : 74

Correct Marks : 3

Question Label : Short Answer Question

Annual summary of five employees in an insurance company is given in Table Q.2. If the ratio of insurance policies sold by B and D are in the ratio 5:12, then how many insurance policies did employee D sell?

Insurance employee	Frequency	Relative frequency
A	175	0.21875
B		
C	100	0.125
D		
E	100	0.125

Table Q.2: Insurance dataset

NOTE: Enter your answer to the nearest integer.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

300

Question Numbers : (75 to 77)

Question Label : Comprehension

Use the following information and data given in Figure Q.2 and Figure Q.3 to answer the given subquestions

The stem and leaf plot diagrams given in Figure Q.2 and Figure Q.3 show the results of Statistics and Mathematics exams conducted in a school respectively. In Figure Q.3, a is an unknown value.

4	2 4 8
5	1 3 6 9
6	2 4 6 8 9
7	3 5 7 8 8 8 9
8	2 5 8
9	5 7

Figure Q.2: Stem and leaf plot of scores of Statistics paper, Key: $4|2 = 42$

4	0
5	2 3 7
6	1 2 4 9
7	7
8	1 3 5 7 7 7
9	a

Figure Q.3: Stem and leaf plot of scores of Mathematics paper, Key: $4|0 = 40$

Sub questions

Question Number : 75

Correct Marks : 1

Question Label : Short Answer Question

What is the difference between the modal scores of Mathematics and Statistics?

NOTE: Enter your answer to the nearest integer.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

9

Question Number : 76

Correct Marks : 3

Question Label : Short Answer Question

If the range of Mathematics scores is greater than the range of Statistics scores by 3, then the value of α is

NOTE: Enter your answer to the nearest integer.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

8

Question Number : 77

Correct Marks : 1

Question Label : Short Answer Question

What is the difference between the medians of the two scores?

NOTE: Enter your answer to the nearest integer.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

2

Question Numbers : (78 to 81)

Question Label : Comprehension

Use the following information and data given in Table Q.1 to answer the given subquestions

The placement statistics for the year 2020 of an engineering college that grants degrees in Computer Engineering and Mechanical Engineering is given in Table Q.1.

Roll No	Gender	Score Percentage	Specialisation	Placement Status	Salary (INR lakhs)
CS18B001	M	85%	CSE	Placed	12.00
ME18B001	M	95%	ME	Placed	8.00
ME18B002	F	75%	ME	Placed	9.00
CS18B002	F	78%	CSE	Placed	8.00
CS18B003	M	85%	CSE	Not placed	
CS18B004	F	88%	CSE	Placed	9.00
ME18B003	M	85%	ME	Not placed	
CS18B005	F	75%	CSE	Placed	12.00
CS18B006	F	65%	CSE	Placed	6.00
CS18B007	M	92%	CSE	Placed	25.00
CS18B008	F	55%	CSE	Not placed	
ME18B004	M	95%	ME	Placed	9.00
CS18B009	M	82%	CSE	Placed	15.00
CS18B010	F	87%	CSE	Placed	8.00
ME18B005	M	50%	ME	Not placed	

Table Q.1: Placements dataset

Sub questions

Question Number : 78

Correct Marks : 1

Question Label : Multiple Select Question

Which of the following is (are) case(s)?

Options :

1. ✔ CS18B001

2. ✔ ME18B004

3. ✖ F

4. ✔ CS18B009

5. ✖ CSE

6. ✖ ME

Question Number : 79

Correct Marks : 1

Question Label : Multiple Select Question

Which of the following is (are) numerical variable(s)?

Options :

1. ✖ Roll No

2. ✔ Score Percentage

3. ✖ Placement Status

4. ✖ Specialisation

5. ✔ Salary (INR lakhs)

Question Number : 80

Correct Marks : 3

Question Label : Short Answer Question

What is the population standard deviation of the salary in INR lakhs of the students? (Ignore the cases of students who are not placed.) Enter the answer up to 3 decimal points accuracy.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

4.8 to 5.2

Question Number : 81

Correct Marks : 5

Question Label : Short Answer Question

What is the absolute value of the point bi-serial correlation coefficient of association between gender and salary among the students? (Ignore the cases of students who are not placed.) Enter the answer up to 3 decimal points accuracy.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

0.48 to 0.53

Question Numbers : (82 to 84)

Question Label : Comprehension

Use the following information and data given in Figure Q.4 to answer the given subquestions

The stacked bar chart given in Figure Q.4 represents the repayment status of 2000 loans sanctioned by a prominent bank in the month of July.

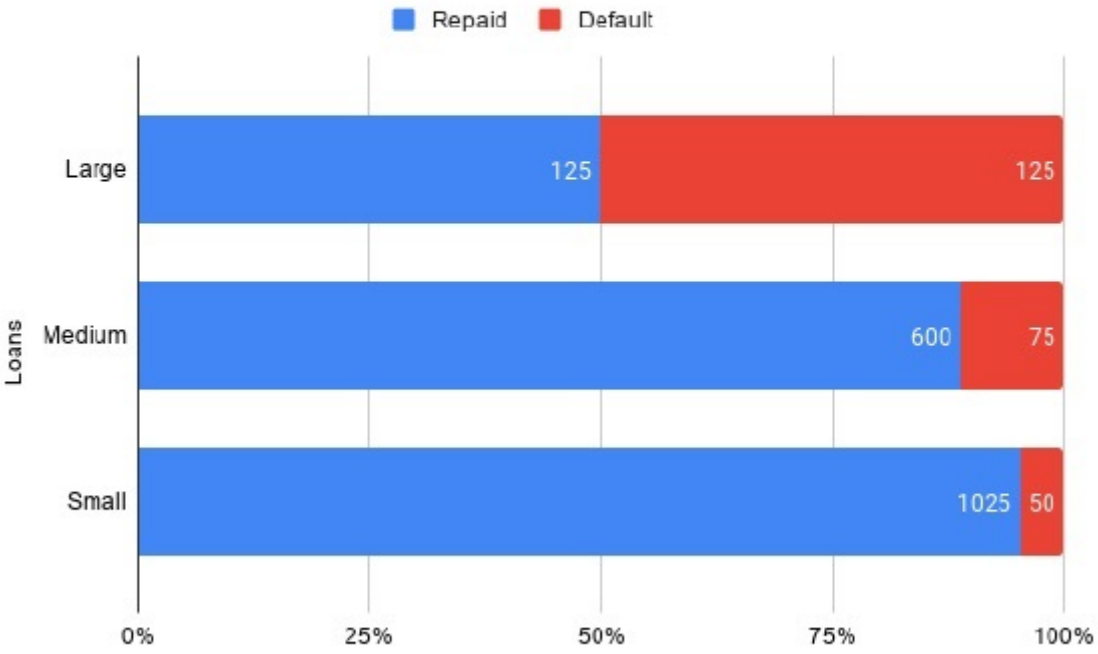

Figure Q.4: Loan dataset

Sub questions

Question Number : 82

Correct Marks : 1

Question Label : Multiple Choice Question

What is the median of the loan variable?

Options :

1. ✖ Large

2. ✔ Small

3. ✖ Medium

4. ✖ Median is not defined for the loans.

Question Number : 83

Correct Marks : 1

Question Label : Short Answer Question

What is the relative frequency of repaid loans in the overall 2000 loans lent by the bank? Enter the answer up to three decimals accuracy.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

0.875

Question Number : 84

Correct Marks : 1

Question Label : Multiple Choice Question

What is the mode of the loans lent by the bank?

Options :

1. ✖ Large

2. ✔ Small

3. ✖ Medium

4. ✖ Mode is not defined for the loans.

Question Numbers : (85 to 86)

Question Label : Comprehension

Use the data given in Figure Q.5 to answer the given subquestions

The histogram of runs scored by a batsman in his career is given in Figure Q.5.

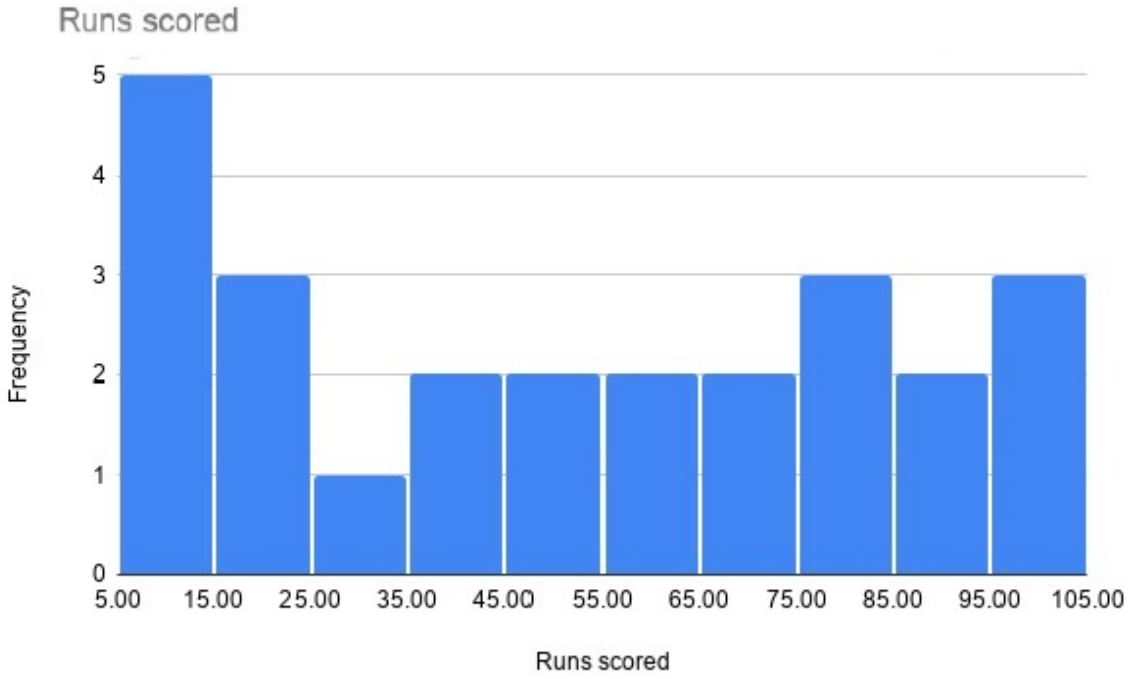

Figure Q.5: Runs dataset

Sub questions

Question Number : 85

Correct Marks : 1

Question Label : Short Answer Question

What is the approximate mean of the runs scored by the batsman?

NOTE: Enter your answer to the nearest integer.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Question Number : 86

Correct Marks : 5

Question Label : Short Answer Question

What is the approximate sample standard deviation of the runs scored by the batsman? Enter the answer up to 3 decimals accuracy. Hint: Use the class mark and frequency to solve for standard deviation.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Range

Text Areas : PlainText

32.4 to 32.9