Indian Statistical Institute

Documentation Research and Training Centre

MS(LIS) (2021-23)

Final term Examination

Paper: 9 Elements of Mathematics and Statistics

Date:11/05/2021 **Max. Marks:60** **Time: 3 hrs**

Note: ***Part A*** and ***Part C*** are compulsory and attempt any three questions for 30 marks in ***Part B.***

PART A

All objective questions are compulsory carries equal marks [1 x 20=20]

1. What is true about sample and population?

a) Population can always be directly observed
b) Parameters from population is always the same as statistics from sample
c) Sample is a subset of population which is randomly draw from population
d) The size of population is always finite

2. You have a DataFrame called 'data' which has only one column 'population'.

```
data = pd.DataFrame()
data['population'] = [47, 48, 85, 20, 19, 13, 72, 1 6, 50, 60]
```

How to draw a sample with sample size =5, from a 'population' with replacement?

a) data['population'].sample(5, replace=True)
b) datarpopulationisample(10)
c) data['population'].sample(5)
d) data['population'].sample(5, replace=False)

3. The measurements of spread or scatter of the individual values around the central point is called:

a) Measures of dispersion
b) Measures of central tendency
c) Measure of skewness
d) Measures of central tendency and Measures of dispersion

4. What is the difference between the maximum and minimum data entries in the set?

a) Mean b) Variance c) Range d) Mode

5. What is the 75th percentile of the following data set; 1, 3, 3, 4, 5, 6, 6, 7, 8, 8

a) 5.5 b) 7 c) 8 d) 3

6. Which of the following measures of central tendency will not be good to use if outliers exist in the dataset?

a) Mean b) Mode c) Median d) Cannot deal with dataset

7. Which of the following data sets has a mean of 10 and standard deviation of 0?

a) 10, 10, 10 b) 0, 10, 20 c) 15, 15, 15 d) 0, 0, 0

8. What's the best way to display median and outliers?

a) A scatter plot b) A box plot c) A bubble chart d) A time series plot

9. What is a suitable way to display the average basketball scores between two teams?

a) A bar chart b) A pie chart c)A histogram d) A scatter plot

10. What is a suitable way to display the relationship between two continuous variables?

a) A bar chart b) A pie chart c)A histogram d) A scatter plot

11. Which of the following will return a plot of age and evaluation scores differentiated by gender?

a) sns.scatterplot(x='age', y='eval', hue='gender', data=ratings_df)
b) sns.boxplot(x='credits', y='beauty', data=ratin gs_df)
c) sns.distplot(ratings_dfrevall, kde = False)
d) ratings_df.groupby('division') [reval'Emean().reset_index()

12. If you got a 75 on a test in a class with a mean score of 85 and a standard deviation of 5, the z-score of your test score would be

a) 2 b) -3 c) 3 d) -2

13. The probability of getting a double by rolling TWO six-sided dice (with sides labelled as 1, 2, 3, 4, 5, 6) is:

a) 1/36 b) 1/6 c) 2/36 d) 1

14. A normal distribution can best be described as which of the following? (Select all that apply)

a) Bell-shaped b) Skewed c) Uniform d) Symmetric

15. Which of the following is a possible alternative hypothesis H1 for a two-tailed test.

a) $p = 85$ b) $p < 85$ c) $p > 85$ d) $p \neq 85$

16. Which additional library should we import when we want to calculate log daily return specifically?

a) Matplotlib b) Statsmodels c) Pandas d) Numpy

17. Suppose you have a DataFrame — data, which contains columns 'Open', 'High', low', 'Close', 'Adj Close' and 'Volume'. What does data[['Open', 'Low']] return?

a) The first row of data which contains only columns 'Open' and 'High'
b) Columns 'Open' and low'
c) No results are shown
d) All columns of data except 'Open' and 'High'

18. The limits of the coefficient of skewness are

a) ± 1 b) ± 2 c) ± 3 d) $\pm \infty$

19. How are the expected values in a Contingency Table calculated?

a) (Column Total - Row Total)/Grand Total
b) ((Column Total) *(Row Total))/Grand Total
c) (Column Total + Row Total)/Grand Total
d) Grand Total / ((Column Total) *(Row Total))

20. In the regression equation Y=mX+b, Y increases with X in all cases

a) If $b > 0$ b) If $b < 0$ c) If $m > 0$ d) If $m < 0$

PART B

Attempt any three questions from this part

1. What do you understand by regression? Critically explain simple linear regression and logistic regression with any dummy data? [10]

2.

a. Three machines I, II, and III manufacture respectively 0.4, 0.5, and 0.1 of the total production. The percentage of defective items produced by I, II, and III is 2, 4, and 1 per cent respectively. For an item chosen at random, what is the probability it is defective? [5]

b. For a group of 20 items, $\Sigma X=1452$, $\Sigma X^2= 144280$ and mode =63.7. Find the Pearsonian coefficient of skewness. [5]

3. A survey of 320 families with 5 children is given below: [10]

Number of boys	5	4	5	2	1	0	Total
Number of girls	0	1	2	3	4	5	
Number of families	14	56	110	88	40	12	320

Is this result consistent with hypothesis i.e.; the male and female births are equally possible.

4. The I. Q. and economic condition of home of 1000 students of an engineering college, Delhi were noted as given in the table : [10]

I. Q. ➡ Economic Con. ⬇	**High**	**Low**	**Total**
Rich	100	300	400
Poor	350	250	600
Total	450	550	1000

Find out whether there is any association between the economic condition of home and I. Q. of the students.

Given the level of significance 0.05.

PART C

1. Define the following: [2x5=10]

a) Equally likely events **OR** Mutually exclusive events

b) Conditional probability **OR** Expected value

c) Confusion matrix **OR** Confidence Interval

d) Test of Significance **OR** Level of Significance

e) Hypothesis Testing **OR** Conditions for χ^2 testing and ANOVA