

Reg. No. :

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Question Paper Code : 90042

B.E. / B.Tech. DEGREE EXAMINATIONS, NOVEMBER/DECEMBER 2022.

Fourth Semester

Artificial Intelligence and Data Science

AD 8403 – DATA ANALYTICS

(Regulations 2017)

Time : Three hours

Maximum : 100 marks

Answer ALL questions.

PART A — (10 × 2 = 20 marks)

1. Define data analytics. What are the 5 types of data analytics?
2. What is the standard error of the mean?
3. Define null hypothesis H_0 and an alternative hypothesis H_1 ?
4. What is the difference between the z-test Vs. t-test?
5. State the difference between a two-tailed and one-tailed test?
6. Define degree of freedom with an example.
7. What is the p-value in statistics?
8. Why is the ANOVA test used?
9. What is the chi-square test used for?
10. Differentiate logistic regression and linear regression.

PART B — (5 × 13 = 65 marks)

11. (a) (i) What are the differences between Population and Sample? (4)
- (ii) {3,5,6,9,10} are the values in a dataset. Find out the mean, variance, Coefficient of Variation, and standard deviation. (4)
- (iii) A population has a mean of 128 and a standard deviation of 22. (5)

Find the mean and standard deviation for samples of size 36.

Find the probability that the mean of a sample of size 36 will be within 10 units of the population mean that is, between 118 and 138

Or

- (b) (i) Define decision rule. (3)
- (ii) A bag contains 6 white and 4 red balls, Three balls are drawn at random. What is the probability that one ball is red and the other two are white? (5)
- (iii) A teacher claims that the mean score of students in his class is greater than 82 with a standard deviation of 20. If a sample of 81 students was selected with a mean score of 90, then check if there is enough evidence to support this claim at a 0.05 significance level. (6)
12. (a) (i) Why hypothesis test? (2)
- (ii) What is the relation between sample size and the power of the test? (5)
- (iii) A sample of 100 clients of ABC is taken, and brokerage charges are calculated with the new rates of XYZ broker. If the mean of the model is \$18.75 and the sample standard deviation is \$6, can any inference be made about the difference in the average brokerage bill between ABC and XYZ broker? (two-tailed test) (6)
- Or
- (b) (i) What is level of confidence? (2)
- (ii) What is the difference between a 95% confidence interval and a 95% confidence level? (5)
- (iii) From a random sample of 400 citizens in Ottawa, 136 indicated that the city's transportation system is adequate. Construct a 99% confidence interval for the population proportion who feel the transportation system is sufficient. (6)
13. (a) (i) What is the sample size for the t-test? (4)
- (ii) Which distribution does the t-test rely on? (4)
- (iii) Find the t-test value for the following two sets of values: 7, 2, 9, 8 and 1, 2, 3, 4? (5)
- Or
- (b) (i) What is effect size? What Does an R-Squared Value of 0.9 Mean? (4)
- (ii) How Two-Sample t-tests Calculate t-Values? (4)
- (iii) State the purpose of meta-analysis. (5)

14. (a) (i) What is F-test? (2)
- (ii) When two-sample T-test is used instead of a two sample Z-test? (5)
- (iii) Conduct an F-Test on the following samples. (6)
- Sample – 1 having variance = 109.63, sample size = 41.
- Sample – 2 having variance = 65.99, sample size = 21.
- Or
- (b) (i) Explain two – way ANOVA, with an example. (4)
- (ii) What is the difference between a single – factor and two – factor Anova? (4)
- (iii) What is a 2-factor experiment? Give an example. (5)
15. (a) (i) Find the linear line (7)
- | | | | | | |
|---|---|---|---|---|---|
| x | 1 | 2 | 3 | 4 | 5 |
| y | 2 | 5 | 3 | 8 | 7 |
- (ii) Consider the set of points : (1,1), (-2,-1), and (3,2) : plot these points and the least – squares regression line in the same graph. (6)
- Or
- (b) (i) Elaborate time series analysis with an example. (4)
- (ii) What is the purpose of calculating moving averages for time series data? (4)
- (iii) How does data analysis deal with missing values? Which method is more appropriate in dealing with missing data in time series? (5)
- PART C — (1 × 15 = 15 marks)
16. (a) (i) What are correlation and autocorrelation? (5)
- (ii) Case study : Singaporean customers are unhappy with high-priced products. (10)
- (1) Analyze a Dataset : Question, Expectations
- (2) Product Mindset
- (3) Business Sense
- (4) Metric – driven, Conclusion
- Or
- (b) (i) What is logistic regression? What is the purpose of logistic regression? (4)
- (ii) Discuss three types of logistic regression. (4)
- (iii) Define Sigmoid Function? (3)
- (iv) What are the steps of logistic regression? (4)