

**Model Questions**

SKR/KW/24/2677/2684/2690

**Faculty of Science & Technology**  
**Eighth Semester B.Tech. (Computer Engineering/IT/CT) (CBCS) Examination**  
**DATA ANALYTICS USING PYTHON**  
**PROG ELE-VII**

Time : Three Hours]

[Maximum Marks : 70

**INSTRUCTIONS TO CANDIDATES**

- (1) All questions carry marks as indicated.
  - (2) Solve Question **1 OR** Question No. **2**.
  - (3) Solve Question **3 OR** Question No. **4**.
  - (4) Solve Question **5 OR** Question No. **6**.
  - (5) Solve Question **7 OR** Question No. **8**.
  - (6) Solve Question **9 OR** Question No. **10**.
  - (7) Due credit will be given to neatness and adequate dimensions.
  - (8) Assume suitable data wherever necessary.
  - (9) Diagrams should be given wherever necessary.
  - (10) Illustrate your answers wherever necessary with the help of neat sketches.
  - (11) Use of non-programmable calculator is permitted.
1. (a) Define Data Analytics and briefly explain its types. 8
  - (b) Explain the measures of central tendencies. Given the data set {2, 4, 4, 4, 5, 5, 6, 8, 10}. Calculate the mean, median and mode. 6
- OR**
2. (a) What do we mean by events ? Explain with suitable example an Independent Events, Collectively Exhaustive Events and Complementary Events. 4
  - (b) Explain the difference between systematic sampling and cluster sampling, and provides an example of each. 6
  - (c) How do you manipulate and aggregate data using Pandas in Python ? 4
  3. (a) Explain the difference between a null hypothesis and an alternative hypothesis and provide an example of each. 6
  - (b) Explain the difference between one-tailed and two-tailed tests in hypothesis testing and provide examples of when each might be used. 8

**OR**



## Model Questions

4. (a) Discuss the assumptions that must be met for the results of an ANOVA test to be valid. 5
  - (b) Explain the concept of linear regression and how it is used to model the relationship between two variables. 6
  - (c) (i) State TRUE or FALSE : Statement : The variance of error is same for all values of the independent variable.
    - (a) TRUE (b) FALSE
  - (ii) Null hypothesis,  $H_0: \mu_1 - \mu_2 = 0$ 
    - (a) Upper tail test (b) Lower tail test
    - (c) Two tail test (d) F test
  - (iii) The 'F' ratio in a completely randomized ANOVA is the ratio of :
    - (a) MSTR/MSE (b) MST/MSE
    - (c) MSE/MSTR (d) MSE/MST 3
  5. (a) How important is it to validate the assumptions of a regression model before interpreting the results ? 5
  - (b) What is Residual Analysis ? Explain different types of residual analysis. 6
  - (c) Define the terms : 3
    - (i) Confidence interval
    - (ii) Prediction interval
    - (iii) Point estimate.
- OR**
6. (a) Explain Maximum Likelihood Estimation (MLE). 7
  - (b) What do you mean by Logistic Regression and explain when it is used ? 7
  7. (a) Describe various accuracy measures for classification. 7
  - (b) What is ROC ? How to select Threshold value on ROC curve. 7
- OR**
8. (a) Describe the Chi-square test of independence. What are limitations of Chi-square test ? 5
  - (b) What do you mean by Cluster Analysis ? State its purpose. 5
  - (c) Differentiate between t-test and Chi-square test. 4
  9. (a) What do you mean by categorical variable ? How to find dissimilarity between categorical variables ? 4
  - (b) Explain k-means clustering algorithm. 8
  - (c) What is ratio-scaled variable ? 2
- OR**
10. (a) Explain how a decision tree works and how it is constructed ? 7
  - (b) Explain the concept of pruning in decision trees. Why is it important ? 5
  - (c) What is CART ? 2



## Model Questions

PRS/KS/24/2954/2960/2967

**Faculty of Science and Technology**  
**B.Tech. (Computer Engineering / IT/CT) Eighth Semester (C.B.C.S.) Examination**  
**DATA ANALYTICS USING PYTHON**  
**Prog. Elective-VII**

Time : Three Hours]

[Maximum Marks : 70

**INSTRUCTIONS TO CANDIDATES**

- (1) All questions carry marks as indicated.
  - (2) Solve Question **1 OR** Question No. **2**.
  - (3) Solve Question **3 OR** Question No. **4**.
  - (4) Solve Question **5 OR** Question No. **6**.
  - (5) Solve Question **7 OR** Question No. **8**.
  - (6) Solve Question **9 OR** Question No. **10**.
  - (7) Due credit will be given to neatness and adequate dimensions.
  - (8) Assume suitable data wherever necessary.
  - (9) Illustrate your answers wherever necessary with the help of neat sketches.
1. (a) Define data analytics and explain its types. 6  
 (b) What do you mean by Sample Space ? What are different methods for describing a sample space ? 4  
 (c) Write the python code to creat an employee's data frame from the emp.CV file and display head and summary. 4
- OR**
2. (a) Describe various measures of central tendency and dispersion. 7  
 (b) Define the concept of a sampling distribution and give the types of sampling distributions. 7
  3. (a) What is Hypothesis Testing ? Define Null hypothesis and Alternate hypothesis. 7  
 (b) Differentiate between : 7
    - (i) Type I and Type II error
    - (ii) One way ANOVA and Two-way ANOVA
- OR**
4. (a) Mean marks obtained by male and female students of schools ABCD in the first unit test are shown below : 7  
 What is the standard error for the difference between the two means ?  

	Male	Female
Sample Size	64	36
Sample Mean Marks	44	41
Population Variance ( )	128	72
  - (b) What is regression analysis ? In regression analysis, which are the assumptions about the error term  $\varepsilon$  ? 7

**Model Questions**

5. (a) What is residual analysis ? Explain different types of residual analysis. 7
- (b) Differentiate between : 7
  - (i) Simple and multiple regression.
  - (ii) Confidence intervals vs prediction intervals.
- OR**
6. (a) What do you mean by logistic regression ? Explain its objectives. 7
- (b) Compare : 7
  - (i) Linear Regression model and Logistic Regression model.
  - (ii) Logistic Regression and Discriminant analysis.
7. (a) Describe various Accuracy Measures for Classification. 7
- (b) What is ROC ? How to select a Threshold using ROC ? 7
- OR**
8. (a) Describe the Chi-square test of independence. 6
- (b) Calculate the Euclidean distance and Manhattan distance between points (2,-1) and (-2,2). 4
- (c) What are different types of data ? 4
9. (a) What do you mean by categorical variable ? How to find dissimilarity between categorical variables ? 7
- (b) Give and explain the classification of clustering methods. 7
- OR**
10. (a) What are different attribute selection measures in decision tree ? 7
- (b) Explain the following : 7
  - (i) Decision Tree
  - (ii) CART
  - (iii) Dendrogram.