

**HIGHER COLLEGES OF TECHNOLOGY**

**Computer and Information Science**

**Non-Exam Based Assessment**

**Cover Sheet**

|  |  |  |  |
| --- | --- | --- | --- |
| **Course Name** | **Statistics and Probability** | **Course Code** | **CIS 2003** |
| **Assessment** | **Statistics Project** | **Handing out Week** | **Week 9** |
| **To be submitted in:** | **Week 14** |
| **Maximum Marks** | 85 Marks | **Percentage of Final Grade.** | **25%** |

**This assessment will assess the following Course learning outcomes:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **CLO1** | **CLO2** | **CLO3** | **CLO4** | **CLO5** |  |
| Question No. | √ | √ | √ | √ | √ |  |
| * The entire project/case study/poster is designed and developed by me (and my team members). * The proper citation has been used when I (and my team members) used other sources. * No part of this project has been designed, developed or written for me (and my team members) by a third party. * I have a copy of this project in case the submitted copy is lost or damaged. * None of the music/graphics/animation/video/images used in this project have violated the Copy Right/Patent/Intellectual Property rights of an individual, company or an Institution.   **Student Signature: Date:**  **Student Signature: Date:**  **Student Signature: Date:** | | | | | | |
|  | | | | | | |

**For Examiner’s Use Only**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Question No.** | **A**  **Project Report** | **B**  **50% Weighted marks of Project Report** | **C**  **Oral/viva** | **D**  **50% weighted marks of Oral/viva** | **Total Marks=B +D** |
| Marks Allocated | 70 | 50% | 15 | **50%** |  |
| Student 1 |  |  |  |  |  |
| Student 2 |  |  |  |  |  |
| Student 3 |  |  |  |  |  |

**Project Objectives:**

1. To summarize data using descriptive statistics.
2. To check for normality of data
3. To estimate the population parameters from sample statistics using Confidence Levels for the normal data set.
4. To conduct hypothesis testing about the Mean.

**Project Description:**

* The practical part of the project is described in sections I, II, III and IV.
* MS-Excel, Megastat Add-in program can be used to complete the tasks.
* An appropriately formatted MS-Word report describing all the project tasks should be submitted at the end of the project.
* This will be followed by an individual Viva/Oral defense.

**Section I: Descriptive Statistics- Population Parameters and Sample Statistics (23 Marks)**

1. A short introduction. 2 Marks
2. Select a Data set (No Two Groups can choose the Same Dataset) **2 Marks**

The following links may be used for obtaining data sets:

1. Data.gov [http://data.gov](http://data.gov/)
2. US Census Bureau <http://www.census.gov/data.html>
3. Data.gov.uk <http://data.gov.uk/>
4. Dubai Statistics Center <https://www.dsc.gov.ae/en-us>
5. Summarize and present the Population using Descriptive Statistics (Mean, Median, Mode, Variance and Standard Deviation). 4 Marks
6. Present Population data graphically using box plot and histogram. 3 Marks
7. Compare and comment on the shape of the histogram obtained for the above data with general shapes of histograms. (Note: You may refer online to find out and compare the various shapes of a histogram.) 3 Marks
8. Take a Sample of 15 from the above Population and use Descriptive Statistics (Mean, Median, Mode, Variance, and Standard Deviation) to summarize the sample. 4 Marks
9. State which method of sampling was used by you for selecting the above sample.

Explain clearly how you carried out sampling using this method. 5 Marks

**PART II: Checking the Normality of Population (6 Marks)**

1. Read pages 322-324 of the textbook (provided in BBL). 0 Mark

b. Using the histogram constructed in Section Id, calculate Pearson’s coefficient of skew-ness and check for outliers. 5 Marks

c. Make a final comment (using the values/findings from the above step) about the normality of the population. 1 Mark

**PART III: Estimation of Population Mean (20 Marks)**

1. For the Random sample of size 15 taken above in **PART 1,** make both the Point and Interval estimates for the Population mean at a Confidence level of 95%.

5 Marks

1. Repeat (**Part II a)** for samples of sizes 31 and 45. 10 Marks
2. Do the widths of the above three Confidence Intervals differ?

Comment on this. 2 Marks

1. From the value of the Population mean calculated in **PART I** check if all of your confidence intervals contain the . Elaborate your answer. 3 Marks

**PART IV: Hypothesis testing (16 marks)**

1. Write down a clear ‘*Research Problem’* (This should be a clear statement mentioning the variable and the Hypothesis Test you intend to perform on it.)

3 Marks

1. Perform Hypothesis Testing for the research problem. Use your own confidence/significance level, type of test and any other data you need to conduct the hypothesis test.

Show all your work clearly, including calculations, sketch of the normal curve, the chosen test and the values of level of significance, critical value etc.

10 Marks

1. Summarize the results (What do the test results indicate – Reject the Null hypothesis OR Fail to reject the Null Hypothesis?).Write a Final Conclusion connecting your results obtained above with the “Research Problem” stated in **Part a** of this section. 3 Marks

**Project Report Format (5 marks)**

* Cover page, page numbers, table of contents and appendices if needed. 3 Marks
* Include a Conclusions Section at the end summarizing the results of the two Inferential Statistics methods used by you in this study. 2 Marks

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_