**Mukesh Patel School of Technology Management and Engineering**

**Artificial Intelligence Department**

**Course Policy**

|  |  |  |  |
| --- | --- | --- | --- |
| **Program/Branch/Semester** | **:** | B.Tech AI/ AI DS/ AI ML/ and MBA Tech AI/ Sem IV | |
| **Academic Year** | **:** | 2023-24 | |
| **Course Code & Name** | **:** | 702BS0C027\_ Statistical Methods | |
| **Credit Details** | **:** | |  |  |  |  | | --- | --- | --- | --- | | L | T | P | C | | 2 | 0 | 2 | 3 | | |
| **Course Coordinator**  **Faculty** | **:** | Dr Nidhi Asthana  Prof Sameeta Saraf | |
| **Contact No. & Email** | **:** | [nidhi.asthana@nmims.edu.in](mailto:nidhi.asthana@nmims.edu.in)  [sameetaniteshsaraf@gmail.com](mailto:sameetaniteshsaraf@gmail.com)  8097083108 | |
| **Office** | **:** | NMIMS, STME, Indore/Navi Mumbai | |
| **Student Contact hours** | **:** |  | |
| **Other Course Faculty members teaching this course** | **:** |  | |
|  | | |  |
| ***Queries by Emails are encouraged.*** | | | |
| **Course link** | **:** | Portal Link  MS Teams Link | |

# Introduction to the Course

## Importance of the course

Statistical methods are mathematical formulas, models, and techniques that are used in statistical analysis of raw research data. The application of statistical methods extracts information from research data and provides different ways to assess the robustness of research outputs.

## Objective of the Course

This course aims to develop sound knowledge and skills in theoretical and application oriented statistics. It will also help students to equip with intermediate to advanced level concepts and tools in statistics that help them tackle relevant problems within engineering domain.

## Pre-requisite

Random Processes and Estimation, Applied Vector and Linear Algebra

# Course Outcomes (CO) and mapping with Program Outcomes (PO)

## Course Outcomes

After completion of the course, students would be able to:

1. Know the applications of statistics and sample the population using various sampling techniques.

2. Classify, tabulate and represent data and calculate the descriptive statistics.

3. Explain the concepts of multivariate regression models, principal component analysis and discriminant analysis.

4. Analyze and implement a simple and multiple linear regression model

# Syllabus, Pre-class activity and References

## Teaching and evaluation scheme

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Teaching Scheme** | | | | **Evaluation Scheme** | |
| **Lecture**  **Hours per week** | **Practical**  **Hours per week** | **Tutorial**  **Hours per week** | **Credit** | **Internal Continuous Assessment (ICA)**  **As per Institute Norms**  **(50 Marks)** | **Theory**  **(3 Hrs,**  **100 Marks)** |
| 2 | 2 | 0 | 3 | Marks Scaled to 50 | Marks Scaled to 50 |

## Syllabus

|  |  |  |
| --- | --- | --- |
| **Unit** | **Description** | **Duration** |
| 1 | **Introduction to Statistics:**  Definition of Statistics. Basic objectives. Applications in various branches of science with examples. Collection of Data: Internal and external data, Primary and secondary Data. Population and sample, Representative sample. | **02** |
| 2 | **Sampling Techniques:** Random sampling. Sampling from finite and infinite populations. Estimates and standard error (sampling with replacement and sampling without replacement), Sampling distribution of sample mean, stratified random sampling | **03** |
| 3 | **Classification and tabulation of Data**: Meaning and objective of  classification, Types of classification, formation of discrete and  continuous distribution.  **Data Classification and Data Presentation:** Histogram, Frequency distribution, Quantitative Data Graphs (Histograms, Frequency Polygons, Ogives, Dot Plots, Stem-and-Leaf Plots); Qualitative Data Graphs (Pie Charts, Bar Graphs, Pareto Charts); Graphical Depiction of Two-Variable Numerical Data: Scatter Plots.  **Descriptive Statistics:** Measures of Central Tendencies – Grouped and Ungrouped Data; Mean, Sample Mean– Weighted mean, Geometric Mean, Harmonic Mean; Median – Quartiles, Deciles and Percentiles; Mode, Box Plot; Measures of Variability– Dispersion, Range, Standard deviation, Chebyshev’s theorem; Population v/s sample variance and standard deviation, Skewness; Kurtosis. | **07** |
| 4 | **Simple and Multiple Linear Regression Model:**  Least squares and linear regression: Introduction; Notation; Ordinary least squares; Regression to the mean; Simple and Multiple Linear regression; Residuals; Regression inference | **05** |
| 5 | **Multivariate Regression:**  Assumptions of Multivariate Regression Models, Parameter estimation, Multivariate Analysis of variance. | **05** |
| 6 | **Discriminant Analysis:**  Statistical background, linear discriminant function analysis, Estimating linear discriminant functions and their properties. | **04** |
| 7 | **Principal Component Analysis:**  Principal components, Algorithm for conducting principal component analysis, deciding on how many principal components to retain, H-plot. |  |
|  | **Total hours** | **30** |

## Pre-class activity

Outline for preliminary study to be done for each unit will be provided prior to commencement of each unit. Preliminary study material (video links, presentation, notes etc) will be made available on the student portal. Students are expected to go through this material before attending the upcoming session. It is expected that the students put in at least two hours of self-study for every one hour of classroom teaching. During the lecture session, more emphasis will be given on in-depth topics, applications and doubt solving.

## References

|  |
| --- |
| **Text Books:**  1. Richard A. Johnson, Dean W.Wichern, (2008), ”Applied Multivariate Statistical Analysis”,  Pearson.  2. Gupta, S. P (2012). Statistical Methods. Sultan Chand & Sons  3. J.D. Jobson,” Applied Multivariate Data Analysis”, Vol I & II, Springer Publication. |
| **Reference Books:**  1. An Introduction to Categorical Data Analysis. Agresti, A. (2012). John Wiley & sons.  2. The Element of Statistical Learning, Data mining, Inference and Prediction. Hastie, T,  Tibshirani, R, & Friedman, J. (2011). New York: Springer Series in Statistics.  3. Hair, Black, Babin, Anderson and Tatham (2009). Multivariate Data Analysis, Pearson. |

# Laboratory details

The following 13 programming exercises, One lab exercise test, and one design of lab problem statement will form the submission for laboratory coursework.

|  |  |  |  |
| --- | --- | --- | --- |
| **Sr. No.** | **Week No.#** | **List of Lab Exercises** | **Mapped CO** |
| 1 |  | **Introduction to statistics and R programming/Python**  Calculate the following for raw data :   * 1. Mean, Median, Mode   2. Mean Deviation, Variance | CO2 |
| 2 | 2 | **Descriptive Statistics:**  Tabulate and calculate the following for grouped and ungrouped data :   * 1. Mean, Median, Mode   2. Range, Variance, Standard Deviation | CO2 |
| 3 | 3 | **Sampling Techniques**  Sample the data using various sampling techniques:   1. Calculate standard error (sampling with 2. Replacement, without replacement) 3. Estimates | CO1 |
| 4 | 4 | **Classification and tabulation:**  Tabulate and calculate the following for grouped and ungrouped data :   1. Mean, Weighted Mean, 2. Geometric mean, Harmonic mean 3. Composite mean, Variance, SD | CO1 |
| 5 | 5 | **Data Classification and Data Presentation**  Classify, tabulate and represent data Using   1. Histograms, Frequency Polygons, Data Plots 2. Pie Charts, Bar Graphs, Pareto Charts 3. Graphical Depiction of Two-Variable Numerical Data 4. Scatter Plots 5. Skewness, Kurtosis | CO2 |
| 6 | 6 | **Simple and Multiple Linear Regression Model**  Analyze and implement a simple linear regression model:   1. Least squares 2. Residuals 3. Inference | CO4 |
| 7 | 7 | **Simple and Multiple Linear Regression Model**  Analyze and implement a multiple linear regression model:   1. Least squares, Regression to mean 2. Residuals 3. Regression Inference | CO4 |
| 8 | 8 | **Lab Exercise Test** |  |
| 9 | 9 | **Multivariate Regression:**  Concepts of Multivariate Regression Models, Parameter estimation | CO3 |
| 10 | 10 | **Multivariate Regression:**  Concepts of Multivariate Regression Models, Multivariate Analysis of variance. | CO3 |
| 11 | 11 | **Discriminant Analysis:**  Linear discriminant function analysis,  Estimating linear discriminant functions | CO3 |
| 12 | 12 | **Discriminant Analysis:**  Analyse linear discriminant function  Estimating linear discriminant functions, etc | CO3 |
| 13 | 13 | **Principal Component Analysis**  Algorithm for conducting principal component Analysis | CO3 |
| 14 | 14 | **Principal Component Analysis**  Analysing and deciding on how many principal components to retain | CO3 |
| 15 | 15 | Design of lab problem statement |  |

# Assessment Policy

## Component wise Continuous Evaluation Internal Continuous Assessment (ICA) and Term End Examination (TEE)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Assessment Component** | **ICA (100 Marks)**  **(Marks scaled to 50)** | | | | | **TEE (100 marks)**  **(Marks scaled to 50)** |
|  | **Class Test 1 and Class Test 2** | **Lab exercises** | **Assignment and Viva** | **Project Presentation/**  **report/poster** | **Class Participation** |  |
| **Weightage** | 20% | 10% | 10% | 5% | 5% | 50% |
| **Marks** | 20+20 | 20 | 20 | 10 | 10 | 100 |
|  | As per academic calender | Weekly | Week 7 and Week 12 | Week 8 | Monthly |  |

## Assessment Policy for Internal Continuous Assessment (ICA)

Assessment of ICA comprises of the following components.

1. **Class test 1 and 2** 
   1. Two class tests will be conducted as per the academic calendar.
   2. It may be conducted online/ offline for 20 marks each

1. **Laboratory performance evaluation (20 marks)**

Tutorial (20 marks)

* Continuous assessment for tutorials will be conducted. There are 13 lab exercises, Lab Exercise test and one design of lab problem statement, each carrying weightage of 20 marks.
  + Discussion of your work with your peers is allowed. However, each student is expected to submit his/her original work. Assessment of the work will be carried out based on parameters like understanding of the exercise, submission on paper and programmatically, originality in the work, involvement of the student, regularity, discipline etc. during the session.

At the end of the course, average of total marks will be taken to obtain marks out of 20.

The rubrics followed for the evaluation is:

|  |  |  |  |
| --- | --- | --- | --- |
| **Sr. No.** | **Parameter** | **Description** | **Marks** |
| 1 | Completion and Timeliness | Completed and submitted on time on paper and programmatically | 17 |
| Completed and submitted on time on paper but programming results delayed | 15 |
| Not completed on time on paper but completed on time programmatically. | 15 |
| Submissions within a week, before the start of the next tutorial | 12 |
| 2 | Punctuality | Present for the lab exercise on time | 3 |
| **Note** | Absent and submission on paper done | The student is absent but he submits the on paper and programmatically within a week. | 10 |
|  | Absent and no submission done | The student is absent and he also does not submit the lab exercise | ZERO |

1. **Assignment and viva (20 marks) –** There will be two home assignments of 10 marks each. The students will have 8 days time to submit each assignment. Viva will be based on the assignment and will be conducted after submission of the assignment.
2. **Project Presentation/report on real life applications/poster on advanced topics submission (10 marks)**
   1. Project Presentation will be a group activity. Faculty will make group of 2-3 students.
   2. Every group will get a unique topic to present
   3. Report/poster submission will be an individual activity, selected by the student.
   4. Assessment will be based on the content, quality, understanding and originality.
3. **Class Participation (10 marks)-** The faculty will ask some questions in every class based on the content being taught. The question could be asked to a chosen student or a student group (which is formed at the beginning of the semester). One mark can be given to the correct answer. The idea is to encourage students to pay attention in class and actively participate. These marks will be added in ICA class participation component.

## Assessment Policy for Term End Examination (TEE)

A written examination of ------- hours will be conducted for the course as per the academic calendar.

# Lesson Plan

|  |  |  |  |
| --- | --- | --- | --- |
| **Session**  **No.** | **Topics** | **Mapped CO** | **Textbook** |
| 1 | **Introduction to Statistics:**  Definition of Statistics. Basic objectives. Applications in various branches of science with examples. | CO1 | TB2 |
| 2 | Collection of Data: Internal and external data, Primary and secondary Data. Population and sample, Representative sample |  |  |
| 3 | **Descriptive Statistics:** Measures of Central Tendencies – Grouped and Ungrouped Data; Mean, Sample Mean; Median; Mode, Measures of Variability– Dispersion, Range, Standard deviation. | CO2 | TB2 |
| 4 | **Sampling Techniques:** Random sampling. Sampling from finite and infinite populations. | CO1 | TB2 |
| 5 | **Estimates :**  Estimates and standard error (sampling with replacement and sampling without replacement), |  |  |
| 6 | **Sampling distribution** of sample mean, stratified random sampling. |  |  |
| 7 | **Classification and tabulation of Data**: Meaning and objective of classification, Types of classification, formation of discrete and | CO2 | TB2 |
| 8 | **Data Classification and Data Presentation:** Histogram, Frequency distribution, Quantitative Data Graphs (Histograms, Frequency Polygons, Ogives, Dot Plots, Stem-and-Leaf Plots); |  |  |
| 9 | Qualitative Data  Graphs (Pie Charts, Bar Graphs, Pareto Charts); Graphical Depiction of Two-Variable Numerical Data: Scatter Plots Venn diagrams - Examples |  |  |
| 10 | **Descriptive Statistics:** Measures of Central Tendencies – Grouped and Ungrouped Data; Mean– Weighted mean, Geometric Mean, Harmonic Mean; | CO2 | TB2 |
| 11 | **Class Test 1** |  |  |
| 12 | Quartiles, Deciles and Percentiles; Box Plot | CO2 | TB2 |
| 13 | Chebyshev’s theorem; Population v/s sample variance and standard deviation, Skewness; Kurtosis | CO2 |  |
| 14 | **Simple and Multiple Linear Regression Model:**  Least squares and linear regression: Introduction; Notations | CO4 | TB2 |
| 15 | Ordinary least squares; Regression to the mean | CO4 |  |
| 16 | Simple and Multiple Linear  regression; Residuals; Regression inference |  |  |
| 17 | Simple and Multiple Linear  regression; Residuals; Regression inference |  |  |
| 18 | **Multivariate Regression:**  Assumptions of Multivariate Regression Models, Parameter estimation | CO3 | TB1 |
| 19 | Parameter estimation |  |  |
| 20 | Multivariate Analysis of variance. |  |  |
| 21 | Multivariate Regression |  |  |
| |  |  | | --- | --- | | 22 | Class Test 1 | | Class Test 2 |  |  |
| 23 | **Discriminant Analysis:**  Statistical background, linear discriminant function analysis, Estimating  linear discriminant functions and their properties. | CO3 | TB1 |
| 24 | Linear discriminant function analysis |  |  |
| 25 | Estimating linear discriminant functions. |  |  |
| 26 | Estimating linear discriminant functions,  properties. |  |  |
| 27 | **Principal Component Analysis:**  Principal components. | CO3 | TB3 |
| 28 | Algorithm for conducting principal component |  |  |
| 29 | Analysis, deciding on how many principal components to retain, H- plot. |  |  |
| 30 | Analysis, H- plot |  |  |

# Teaching-learning methodology

Faculty will make a group of 2-3 students for any group based activity such as class participation, project, presentation etc. Lecture and laboratory session will be conducted as follows-

1. **Lectures:** 
   * Outline for preliminary study to be done for each unit will be provided prior to commencement of each unit.
   * Deeper concepts and applications will be explained through presentation and video lectures.
   * Numerical problems based on concept will be solved during the session
   * The solutions will be very detailed, simple and easy to understand.
2. **Lab Exercises:**
   * There will be about 13 Lab Exercises, one Lab exercise test and one Design of Lab problem statement tutorials to be submitted on the student portal.
   * Regular assessment and grading will be done. Students will be marked based on parameters like completion of exercises both on paper and programatically, originality, timely submission, logic developed, interaction during the lab, punctuality and discipline.

# Active learning techniques

Active learning is a method of learning in which students are actively or experientially involved in the learning process. Following active learning techniques will be adopted for the course.

1. **Muddiest topic:** Faculty will find out the least understood point/topic in the session. This topic is then further explained to ensure that it is understood well.
2. **The "One Minute Paper":** The faculty will ask students to take out a blank sheet of paper, pose a question (either specific or open-ended), and give them one or two minute(s) to respond.
3. **Blended Learning:** Students will be introduced to the topic at home while the in-depth topics, applications and numerical problems will be discussed by the faculty in the lecture session. Outline for preliminary study to be done for each unit will be provided prior to commencement of each unit. Preliminary study material (video links, presentation, notes, etc) will be made available on the student portal. The solutions provided will be very detailed, simple and easy to understand.
4. **Frame a question: S**tudent will be asked to design and frame their own questions pertaining to the topic being taught. The idea is to stimulate students’ curiosity, engage the students in collaborative teaching and learning, and motivating students to develop deeper understating of the topic.
   * Frame questions for each unit of the course: At the beginning of each using, the faculty will create a new page in *OneNote Class Notebook* in collaborative section where every student will post his/her question.
5. **Brainstorming: S**tudents will be asked to generate ideas on a certain topic, category or question while the faculty will facilitate and record the answers.

# Course Material

Following course material is uploaded on the student portal: [(give](https://sites.google.com/a/nirmauni.ac.in/2cs101-computer-programming/) student portal link)

* Course Policy
* Lecture Notes
* Lecture Videos
* Lecture Presentations
* Books / Reference Books /video lectures link
* Assignments
* List of Program Outcomes

# Course Outcome Attainment

Following means will be used to assess attainment of course learning outcomes.

* Use of formal evaluation components of continuous evaluation, assignments, Tutorial work, semester end examination
* Informal feedback during course conduction

# Academic Integrity Statement

Students are expected to carry out assigned work under Internal Continuous Assessment (ICA) independently. Copying in any form is not acceptable and will invite strict disciplinary action. Plagiarism detection software will be used to check plagiarism wherever applicable. Academic integrity is expected from students in all components of course assessment.