|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **XXXXX Statistics for Engineers L,T,P, J,C**  **Pre-Req: Calculus for engineers 2, 1, 2,0,4** | | | | |
| **Module** | **Topics** | **L Hrs** | **SLO** | |
| **1** | **Introduction to Statistics:**  Introduction to statistics and data analysis-Measures of central tendency –Measures of variability-Moments-Skewness-Kurtosis | 3 | 1,2,7,9 | |
| **2** | **Random variables:**  Introduction -random vectors-Probability mass Function, distribution and density functions - joint distribution and joint density functions- Marginal, conditional distribution and density functions- moment generating function – characteristic function | 6 | 1,2,7,9 | |
| **3** | **Correlation and regression:**  Mathematical expectation, and its properties Covariance, regression and correlation – partial and multiple correlation- multiple regression | 3 | 1,2,7,9 | |
| **4** | **Special Distributions :**  Binomial and Poisson distributions – Normal distribution – gamma and Beta-exponential distributions – Weibull distribution | 4 | 1,2,7,9 | |
| **5** | **Sampling Techniques I :**  Testing of hypothesis – Introduction-Types of errors, critical region, procedure of testing hypothesis- Large sample tests- Z test for Single Proportion, Difference of Proportion, mean and difference of means, | 4 | 1,2,7,9 | |
| **6** | **Sampling Techniques II :**  Small sample tests- Student’s t-test, F-test- chi-square test- goodness of fit - independence of attributes- Design of Experiments - Analysis of variance – one and two way classifications - CRD- RBD- LSD - | 5 | 1,2,7,9 | |
| **7** | **Reliability :**  Basic concepts- Hazard function-Reliabilities of series and parallel systems- System Reliability - Maintainability-Preventive and repair maintenance- Availability | 5 | 1,2,7,9 | |
| **Total Lecture Hours**  **Mode:**  **Lectures –** With thrust on applications, demos using software such as R, Excel etc | | **30** | | |
| **Tutorial**  # Minimum of 10 problems per module | | 15 | | 1,2,7,9 |
| **Laboratory exercises**   * Understand concepts through 12 lab exercises(R/Excel software) * Analyse and interpret the results * Lab exercises:  1. **Understanding Data types; importing/exporting data** 2. **Computing/plotting and visualising the following probability distributions**    1. **Binomial distribution**    2. **Poisson distribution**    3. **Normal distribution** 3. **Computing/plotting and visualizing the following distributions:**    1. **Student t-distribution**    2. **F distribution** 4. **Applying simple linear regression model to real dataset; computing and interpreting the coefficient of determination** 5. **Applying multiple linear regression model to real dataset; computing and interpreting the multiple coefficient of determination** 6. **Testing of hypothesis for large samples from real datasets** 7. **Testing of hypothesis for small samples from real datasets** 8. **Applying the goodness of fit test to real dataset** 9. **Applying Chi-squared independence test to real dataset** 10. **Performing ANOVA for real dataset for Completely randomized design** 11. **Performing ANOVA for real dataset for Randomized block design** 12. **Computing reliabilities for series, parallel systems** | | 30 | | 1, 2, 7,9,12,  18, 20 |
| **Text Books**   1. **Probability and Statistics for engineers and scientists by R.E.Walpole, R.H.Mayers, S.L.Mayers and K.Ye, 9th Edition, Pearson Education (2012).**   **Reference Books**   1. **Reliability Engineering by E.Balagurusamy, Tata McGraw Hill, Tenth reprint 2010.** 2. **Probability and Statistics by J.L.Devore, 8th Edition, Brooks/Cole, Cengage Learning (2012).** 3. **Probability and Statistics for Engineers by R.A.Johnson, Miller & Freund’s, 8th edition, Prentice Hall India (2010)** 4. **Probability, Statistics and Reliability for Engineers and Scientists by Bilal M. Ayub and Richard H. McCuen, 3rd edition, CRC press (2011).** | | | | |

1. Recognize the importance of data collection, identify limitations in data collection methods, and determine how they affect the scope of inference.
2. Use statistical software (R) to summarize data numerically and visually, and to perform data analysis.
3. Have a conceptual understanding of the unified nature of statistical inference.
4. Apply estimation and testing methods (confidence intervals and hypothesis tests) to analyze single variables and the relationship between two variables in order to understand natural phenomena and make data-based decisions.
5. Model and investigate relationships between two or more variables within a regression framework.
6. Interpret results correctly, effectively, and in context without relying on statistical jargon.
7. Critique data-based claims and evaluate data-based decisions.
8. Complete a research project that employs simple statistical inference and modeling techniques.

Course Syllabus

**Week 1: Unit 1 - Introduction to data**

* Part 1 – Designing studies
* Part 2 – Exploratory data analysis
* Part 3 – Introduction to inference via simulation

**Week 2: Unit 2 - Probability and distributions**

* Part 1 – Defining probability
* Part 2 – Conditional probability
* Part 3 – Normal distribution
* Part 4 – Binomial distribution

**Week 3: Unit 3 - Foundations for inference**

* Part 1 – Variability in estimates and the Central Limit Theorem
* Part 2 – Confidence intervals
* Part 3 – Hypothesis tests

**Week 4: Finish up Unit 3 + Midterm**

* Part 4 – Inference for other estimators
* Part 5 - Decision errors, significance, and confidence

**Week 5: Unit 4 - Inference for numerical variables**

* Part 1 – t-inference
* Part 2 – Power
* Part 3 – Comparing three or more means (ANOVA)
* Part 4 – Simulation based inference for means

**Week 6: Unit 5 - Inference for categorical variables**

* Part 1 – Single proportion
* Part 2 – Comparing two proportions
* Part 3 – Inference for proportions via simulation
* Part 4 – Comparing three or more proportions (Chi-square)

**Week 7: Unit 6 - Introduction to linear regression**

* Part 1 – Relationship between two numerical variables
* Part 2 – Linear regression with a single predictor
* Part 3 – Outliers in linear regression
* Part 4 – Inference for linear regression

**Week 8: Unit 7 - Multiple linear regression**

* Part 1 – Regression with multiple predictors
* Part 2 – Inference for multiple linear regression
* Part 3 – Model selection
* Part 4 – Model diagnostics

**Unit 1**

Exploratory Data Analysis. This is organized into two modules – Examining Distributions and Examining Relationships. The general approach is to provide students with a framework that will help them choose the appropriate descriptive methods in various data analysis situations.

**Unit 2**

 Producing Data. This unit is organized into two modules – Sampling and Study Design.

**Unit 3**

 Probability. As stated above, this is the unit where the two versions of the course differ. In the Probability and Statistics course the unit is a classical treatment of probability and includes basic probability principles, conditional probability, discrete random variables (including the Binomial distribution) and continuous random variables (with emphasis on the normal distribution). The probability unit of the Statistical Reasoning version of the course essentially acts as a “bridge” to the inference section and includes only those concepts necessary to support a conceptual understanding of the role of probability as the “machinery” behind inference. Both probability parts culminate in a discussion of sampling distributions that is grounded in simulation.

**Unit 4**

 Inference. This unit introduces students to the logic as well as the technical side of the main forms of inference: point estimation, interval estimation and hypothesis testing. The unit covers inferential methods for the population mean and population proportion, Inferential methods for comparing the means of two groups and of more than two groups (ANOVA), the Chi-Square test for independence and linear regression. The unit reinforces the framework that the students were introduced to in the Exploratory Data Analysis for choosing the appropriate, in this case, inferential method in various data analysis scenarios.