# 18CS203: Introduction to Data Science
# Detailed Syllabus for ESA

| Class # | Chapter Title/Reference Literature | Topics to be covered | |
|---|---|---|---|
| 1. | **Unit: 1**<br><br>**Introduction to Data Science,Statistics and Visualizing data**<br><br>**T1: Chapter 1 1.2, 1.3;** | **Introduction to Data Science:** Motivating Examples and Scope. Refer DS-Intro.pdf | |
| 2. | | **Sampling : T1:1.1**<br>**Brief Introduction to statistics,Types of statistics**<br>Descriptive and Inferential Statistics<br>population<br>sample<br>simple random sample<br>simple random sampling<br>sample of convenience<br>sampling variation.<br>tangible populations<br>conceptual population,<br>Independence<br>sampling with replacement.<br>Sampling methods.(1.1)(Excluding Types of experiments)<br>Sampling Methods:weighted sampling<br>stratified random sampling<br>cluster sampling<br>Types of Data<br>numerical or quantitative<br>categorical or qualitative<br>Controlled Experiments and Observational Studies | |
| 3. | | Sampling methods.(1.1) | |
| 4. | | Sampling errors.(Handout1) | |
| 5. | | **Getting and Analyzing Data:** Scraping the Web,Reading Files, (Handout2 &3)(Demo) | |
| 6. | | Need for Data Cleaning,Basics of Data Cleaning.(Handout 4)(Demo) | |
| 7. | | Summary Statistics(1.2) | |
| 8. | | **Summary Statistics (cont) (1.2)**(Demo)<br>The Sample Mean<br>The Standard Deviation<br>the sample variance<br>Outliers | |

# 18CS203: Introduction to Data Science
# Detailed Syllabus for ESA

| | | |
|---|---|---|
| | | The Sample Median<br>The Trimmed Mean<br>The Mode and the Range<br>Quartiles<br>Percentiles<br>Summary Statistics for Categorical Data<br>Sample Statistics and Population Parameters |
| 9. | | **Data Visualization and Interpretation : Graphical summaries**-Histogram.(1.3) ,Unequal Class Widths,Symmetry and Skewness,Unimodal and Bimodal Histograms<br>(Demo) |
| 10. | | Visualizing Data: Bar Charts(1.3)(Demo)(Handout 5),Box plots(1.3) |
| 11. | | Visualizing Data: Multivariate Data,two variables (scatter plots)(1.3)(Demo) |
| 12. | | Good vs. Bad Visualization.(Handout 6) |
| 13. | **Unit: 2**<br>**Random Variables and Probability Distributions**<br>**T1: Chapter 2 2.4 – 2.5, Chapter 4 4.1 – 4.3, 4.5** | Brief overview of Probability Basics.(Handout of definitions)(Self Learning-2.1 and 2.3)<br>**Random Variables** : Introduction,Discrete Random Variables(2.4)The Cumulative Distribution Function of a Discrete<br>Random Variable,Mean and Variance for Discrete Random Variables,The Probability Histogram |
| 14. | | Continuous Random Variables(2.4),Computing Probabilities with the Probability<br>Density Function,The Cumulative Distribution Function<br>of a Continuous Random Variable |
| 15. | | Continuous Random Variables(2.4) **Contd.**<br>**Mean and Variance for Continuous Random Variables,** |

# 18CS203: Introduction to Data Science
# Detailed Syllabus for ESA

| | | |
|---|---|---|
| | | **The Population Median and Percentiles,** |
| 16. | | Linear Functions of Random Variables.(2.5)<br>Adding a Constant<br>Multiplying by a Constant<br>Means of Linear Combinations of Random Variables |
| 17. | | Linear Functions of Random Variables.(2.5)<br>Independent Random Variables<br>Variances of Linear Combinations of Independent<br>Random Variables<br>Independence and Simple Random Samples<br>The Mean and Variance of a Sample Mean |
| 18. | | **Probability Distributions: The** Bernoulli Distribution(4.1),Mean and Variance of a Bernoulli Random Variable<br>The Binomial Distribution(4.2)(Demo)<br>Probability Mass Function of a Binomial Random Variable<br>A Binomial Random Variable Is a Sum of Bernoulli<br>Random Variables<br>The Mean and Variance of a Binomial Random Variable<br>Using a Sample Proportion to Estimate a Success Probability<br>Uncertainty in the Sample Proportion |
| 19. | | The Poisson Distribution(4.3)(Demo)<br>The Mean and Variance of a Poisson Random Variable<br>Using the Poisson Distribution to Estimate a Rate<br>Uncertainty in the Estimated Rate |
| 20. | | The Normal Distribution(4.5),(Demo)<br>Estimating the Parameters of a Normal Distribution<br>Linear Functions of Normal Random Variables<br>Linear Combinations of Independent Normal Random Variables |

# 18CS203: Introduction to Data Science
# Detailed Syllabus for ESA

| | | |
|---|---|---|
| | | How Can I Tell Whether My Data Come from a Normal Population?<br><br>Chebyshev's inequality(2.4) |
| 21. | | Derivation of Distributions.:Bernoulli ,Binomial Distribution(Handout) |
| 22. | | Derivation of Distributions:Poisson Distribution,The Normal Distribution.(Handout) |
| **23.** | **Unit: 3**<br><br>**Sampling and Estimation**<br><br>**T1: Chapter 4 4.9 – 4.11; Chapter 5 5.1-5.4, 5.7** | Principles of Point Estimation :<br><br>**Measuring the Goodness of an Estimator**<br><br>Mean squared error(4.9)+ (Handout) |
| **24.** | | Maximum likelihood estimate (4.9)+(Handout) |
| **25.** | | Maximum likelihood estimate ( 4.9)+(Handout)<br>**Desirable Properties of Maximum Likelihood Estimators** |
| **26.** | | Normal Probability Plot (4.10)(Demo)<br>**Interpreting Probability Plots** |
| **27.** | | **Sampling concepts :** The Central Limit Theorem and its applications(4.11) |
| **28.** | | The Central Limit Theorem Applications.(4.11)<br>**Normal Approximation to the Binomial**<br>**The Continuity Correction**<br>**Accuracy of the Continuity Correction**<br>**Normal Approximation to the Poisson**<br>**Continuity Correction for the Poisson Distribution** |
| **29.** | | **Confidence Intervals :** Introduction,<br>Interval estimates for mean of large samples.(5.1)<br>**More About Confidence Levels**<br>**Probability versus Confidence** |
| **30.** | | Interval estimates for mean of large samples.(5.1)(Demo)<br>**Determining the Sample Size Needed for a Confidence Interval of Specified Width**<br>**One-Sided Confidence Intervals**<br>**Confidence Intervals Must Be Based on Random Samples** |
| **31.** | | Interval estimates for proportion of large samples. (5.2)(Demo)<br>(Excluding The Traditional Method) |
| **32.** | | Confidence intervals for mean of Small Samples.(5.3)<br>Student's t Distribution(Demo) |

# 18CS203: Introduction to Data Science
## Detailed Syllabus for ESA

| | | |
|---|---|---|
| | | **Don't Use the Student's *t* Statistic If the Sample Contains Outliers**<br>**Confidence Intervals Using the Student's *t* Distribution**<br>**How Do I Determine Whether the Student's *t* Distribution Is Appropriate?**<br>**Use *z*, Not *t*, If $\sigma$ Is Known** |
| **33.** | | Confidence Intervals for the Difference Between Two Means for large samples(5.4) |
| **34.** | | Confidence Interval estimates for paired data.(5.7) |
| **35.** | | Factors affecting Margin of Error.(Handout)(Demo) |
| 36. | **Unit: 4**<br><br>**Hypothesis and Inference.**<br><br>**T1: Chapter 6**<br>**6.1 – 6.3, 6.5,6.9, 6.10** | **Hypothesis Testing:Introduction,** mean (6.1)Large sample tests for a Population Mean |
| 37. | | Large sample tests for a Population mean (6.1)**(contd.)**(Demo) |
| 38. | | Drawing conclusions from the results of Hypothesis tests(6.2) |
| 39. | | Drawing conclusions from the results of Hypothesis tests(6.2)**contd.**<br>**Statistical Significance**<br>**The *P*-value Is Not the Probability That *H*0 Is True**<br>**Choose *H*0 to Answer the Right Question**<br>**Statistical Significance Is Not the Same as Practical Significance**<br>**The Relationship Between Hypothesis Tests and Confidence Intervals** |
| 40. | | Large sample tests for a Population proportion (6.3)<br>**The Sample Size Must Be Large**<br>**Relationship with Confidence Intervals for a Proportion** |
| 41. | | Large -Sample tests for Difference between two means(6.5) |
| 42. | | Distribution Free Tests.(6.9)<br>**The Wilcoxon Signed-Rank Test**<br>**Ties**<br>**Differences of Zero**<br>**Large-Sample Approximation**<br>**The Wilcoxon Rank-Sum Test-Large-Sample Approximation**<br>**Distribution-Free Methods Are Not Assumption-Free** |
| 43. | | Chi-squared Test.(6.10)(Demo) |

# 18CS203: Introduction to Data Science
# Detailed Syllabus for ESA

| | | |
|---|---|---|
| | | **The Chi-Square Test for Homogeneity**<br>**The Chi-Square Test for Independence** |
| 44. | **Unit: 5**<br>**Errors of Hypothesis Testing**<br>**T1: Chapter 6**<br>**6.12,6.13**<br><br>**Simple Linear Regression.**<br><br>**T1: Chapter 7**<br>**7.1 – 7.4;** | Fixed Level Testing(6.12)<br>**Critical Points and Rejection Regions** |
| 45. | | Fixed Level Testing.(6.12)<br>**Type I and Type II Errors** |
| 46. | | Power of a Test.(6.13) |
| 47. | | Factors affecting Power of a Test.(Handout)(Demo) |
| 48. | | **Simple Linear Regression: Introduction,**Correlation.(7.1)<br>**How the Correlation Coefficient Works**<br>**The Correlation Coefficient Is Unitless**<br>**The Correlation Coefficient Measures Only *Linear* Association** |
| 49. | | **The Correlation Coefficient can be Misleading when Outliers are Present**<br>**Correlation Is Not Causation**<br>**Controlled Experiments Reduce the Risk of Confounding** |
| 50. | | The Least squares Line.(7.2)<br>**Computing the Equation of the Least-Squares Line**<br>**Computing Formulas**<br>**The Estimates Are Not the Same as the True Values** |
| 51. | | **The Residuals Are Not the Same as the Errors**<br>**Don't Extrapolate Outside the Range of the Data**<br>**Don't Use the Least-Squares Line When the Data Aren't Linear**<br>**Another Look at the Least-Squares Line**<br>**Measuring Goodness-of-Fit** |
| 52. | | Predictions using regression models - Uncertainties in Regression Coefficients.(7.3)<br>**The More Spread in the *x* Values, the Better (Within Reason)** |
| 53. | | Checking Assumptions and transforming data.(7.4)<br>**The Plot of Residuals versus Fitted Values** |
| 54. | | **The Plot of Residuals versus Fitted Values** |
| 55. | | Checking Assumptions and transforming data.(7.4)**contd.**<br>**Transforming the Variables** |
| 56. | | Checking Assumptions and transforming data.(7.4)**contd.Transforming the Variables** |

# 18CS203: Introduction to Data Science
# Detailed Syllabus for ESA

| | | |
|---|---|---|
| | | |