Statistics Roadmap for Professional Data Analysts

Md. Rabbi Ali
B.Sc in Statistics , Islamic University, Kushtia
rabbi.stat.iu@gmail.com
https://github.com/RabbiTheAnalyst
https://www.linkedin.com/in/rabbi-the-analyst/

1. Foundations of Statistics

- **Descriptive Statistics**
  - Measures of Central Tendency: Mean, Median, Mode
  - Measures of Dispersion: Variance, Standard Deviation, Range, Interquartile Range (IQR)
  - Data Distributions: Normal, Skewed, Uniform, etc.
  - Percentiles and Quartiles
  - Data Visualization: Histograms, Box Plots, Scatter Plots, Bar Charts
- **Probability Basics**
  - Basic Probability Concepts: Sample Space, Events, Conditional Probability
  - Bayes' Theorem
  - Probability Distributions: Binomial, Poisson, Normal, Exponential
  - Expected Value and Variance

2. Inferential Statistics

- **Sampling Methods**
  - Random Sampling, Stratified Sampling, Cluster Sampling
  - Sampling Distributions
- **Hypothesis Testing**
  - Null and Alternative Hypotheses
  - Type I and Type II Errors
  - p-values and Significance Levels (α)
  - One-tailed and Two-tailed Tests
- **Confidence Intervals**
  - Interpretation and Calculation
  - Z-tests and T-tests
- **ANOVA (Analysis of Variance)**
  - One-way and Two-way ANOVA
  - F-tests
- **Chi-Square Tests**
  - Goodness-of-Fit Test
  - Test of Independence

3. Regression Analysis

- **Linear Regression**
  - Simple Linear Regression
  - Multiple Linear Regression

- Assumptions of Linear Regression
- R-squared and Adjusted R-squared
- Residual Analysis

- **Logistic Regression**
  - Binary and Multinomial Logistic Regression
  - Odds Ratio and Logit Function

- **Model Evaluation**
  - Mean Squared Error (MSE), Root Mean Squared Error (RMSE)
  - Mean Absolute Error (MAE)
  - ROC Curve and AUC

4. Advanced Statistical Techniques

- **Time Series Analysis**
  - Trend, Seasonality, Cyclical Patterns
  - Moving Averages, Exponential Smoothing
  - ARIMA Models

- **Multivariate Analysis**
  - Principal Component Analysis (PCA)
  - Factor Analysis
  - Cluster Analysis (K-means, Hierarchical Clustering)

- **Experimental Design**
  - A/B Testing
  - Control vs. Treatment Groups
  - Randomized Controlled Trials (RCTs)

5. Statistical Software and Tools

- **Programming Languages**
  - Python: Libraries like NumPy, Pandas, SciPy, StatsModels, Scikit-learn
  - R: Libraries like dplyr, ggplot2, caret

- **Data Visualization Tools**
  - Tableau, Power BI, Matplotlib, Seaborn

- **Statistical Software**
  - SPSS, SAS, STATA (optional, depending on industry)

6. Practical Applications

- **Real-world Data Analysis**
  - Cleaning and Preprocessing Data
  - Exploratory Data Analysis (EDA)
  - Feature Engineering
  - Interpreting Results and Communicating Insights
- **Case Studies**
  - Analyzing Business Data
  - Customer Segmentation
  - Predictive Modeling

7. Optional but Useful Topics

- **Bayesian Statistics**
  - Bayesian Inference
  - Prior and Posterior Distributions
- **Non-parametric Tests**
  - Wilcoxon Signed-Rank Test, Mann-Whitney U Test, Kruskal-Wallis Test
- **Machine Learning Basics**
  - Supervised vs. Unsupervised Learning
  - Overfitting and Underfitting
  - Cross-validation

Recommended Resources

- **Books:**
  - *Naked Statistics* by Charles Wheelan
  - *Practical Statistics for Data Scientists* by Peter Bruce and Andrew Bruce
  - *Think Stats* by Allen B. Downey
  - *Introduction to Statistical Learning* by Gareth James et al.
- **Online Courses:**
  - Statistics with Python Specialization (Coursera)
  - Data Science: Statistics and Machine Learning (edX)
  - Khan Academy Statistics and Probability
- **Practice Platforms:**
  - Kaggle (for datasets and competitions)
  - Datacamp (for hands-on practice)

"A good statistician is like a detective—uncovering the truth hidden in data!"
"May this journey of learning and discovery lead to success. Keep me in your prayers!"