ELE 503

Advanced Computer Programming and Statistics

Week#3: Advanced Statistical Applications in Engineering

By Kingsley E. Erhabor

Week 03.

Advanced Statistical Applications in Engineering

Introduction to Advanced Statistics in Engineering

Analyzing Variability and Central Tendency

Identifying Distribution Types

Skewness and Kurtosis

Hypothesis Testing

Null and Alternative Hypotheses

Type I and Type II Errors

Common Tests (t-test, ANOVA)

Confidence Intervals

Construction and Interpretation

Applications in Engineering

Q&A

Closing Take away

Efosa's Introduction

Engineer | Programmer | Innovator

Technical Authority

Shell Nigeria

Subject Mater Expert (EMEA) for Process Automation & Control (PACO)-Subsea control systems and Subsea Distribution

Containerization 85 %

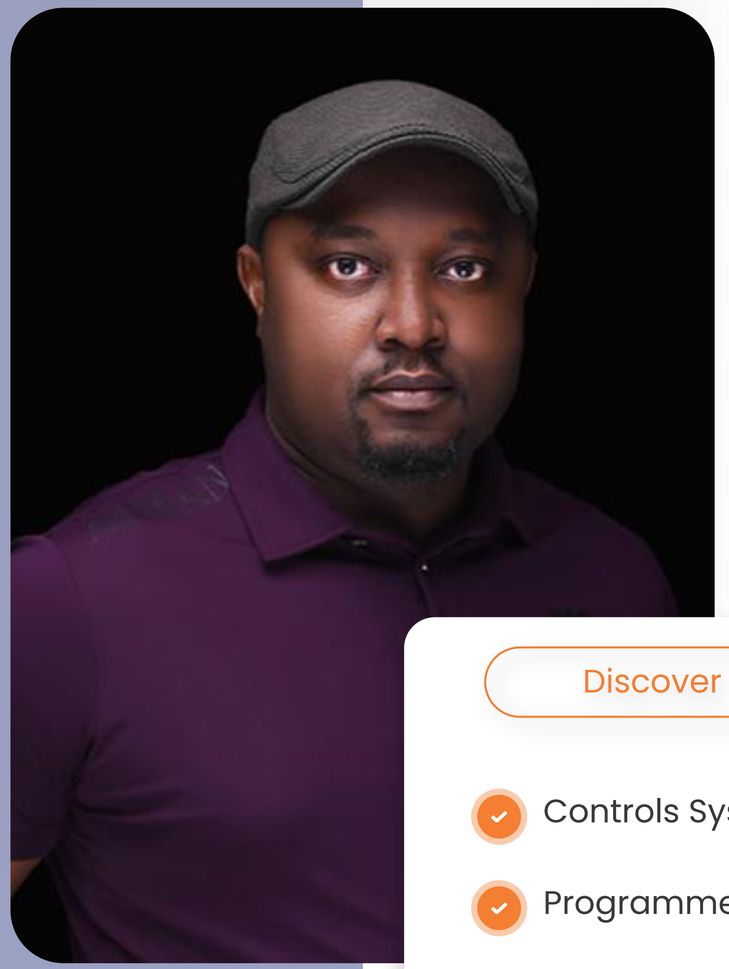
AI & ML 80 %

Renewable Energy 90 %

Innovator, VC

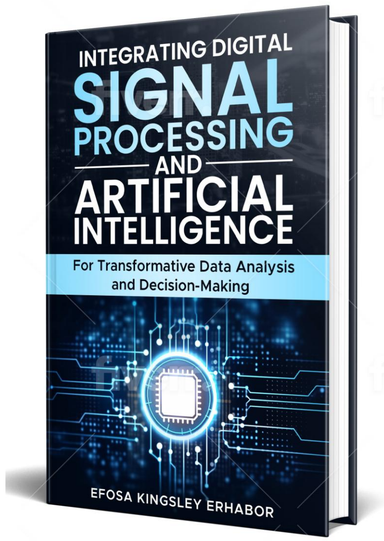
Katharos Technologies

Linux, Devops, AI and Software SME, Innovator and enterprenur



Discover

- Controls System
- Programmer
- Subsea Engineer
- AI and ML



X or Twitter

Linkedin

Part 1:

Introduction to Advanced Statistics in Engineering

Learning Objectives

ELE 503: Advanced Computer Programming and Statistics

- **Analyze** engineering data to determine variability, central tendency, and distribution patterns.

- **Perform** hypothesis testing and construct confidence intervals.

- **Utilize** statistical software tools for data analysis and interpretation

Importance of Advanced Statistics in Engineering

ELE 503: Advanced Computer Programming and Statistics

- **Enhanced Data Interpretation:** Gain deeper insights from complex datasets.

- **Improved Decision Making:** Make informed choices based on rigorous analysis.

- **Quality Assurance:** Ensure products and processes meet high standards.

- **Innovation:** Drive engineering solutions through data-driven strategies

Measures of Central Tendency

ELE 503: Advanced Computer Programming and Statistics

- **Mean:** Average value of the dataset.

- **Median:** Middle value when data is ordered.

- **Mode:** Most frequently occurring value.

- **Advanced Metrics:**
  - **Geometric Mean:** Suitable for multiplicative processes.
  - **Harmonic Mean:** Useful for rates and ratios.

Measures of Variability

ELE 503: Advanced Computer Programming and Statistics

- **Range:** Difference between maximum and minimum values.

- **Variance (σ2 - sigma^2):** Average squared deviation from the mean.

- **Standard Deviation (σ -sigma):** Square root of variance.

- **Coefficient of Variation (CV):** Ratio of standard deviation to mean.

Identifying Distribution Types

ELE 503: Advanced Computer Programming and Statistics

- **Normal Distribution:** Symmetrical, bell-shaped.

- **Skewed Distributions:** Asymmetric, data tails.

- **Bimodal and Multimodal Distributions:** Multiple peaks.

- **Application:** Selecting appropriate statistical methods.

Skewness and Kurtosis

ELE 503: Advanced Computer Programming and Statistics

•**Skewness:** Measure of asymmetry in data distribution.

- **Positive Skew:** Tail on the right.
- **Negative Skew:** Tail on the left.

•**Kurtosis:** Measure of the "tailedness" of the distribution.

- **Leptokurtic:** Heavy tails.
- **Platykurtic:** Light tails.

•**Importance:** Understanding data characteristics for analysis.

Part 2:

Introduction to Hypothesis Testing

Introduction to Hypothesis Testing

ELE 503: Advanced Computer Programming and Statistics

- **Purpose:** Determine if there is enough evidence to reject a hypothesis.

- **Components:**
  - **Null Hypothesis (Ho):** No effect or difference.
  - **Alternative Hypothesis (Ha):** Effect or difference exists.

- **Process:**
  1. Formulate hypotheses.
  2. Choose significance level (α - alphaα).
  3. Select appropriate test.
  4. Calculate test statistic and p-value.
  5. Make decision based on p-value.

Type I and Type II Errors

ELE 503: Advanced Computer Programming and Statistics

- **Type I Error (α - alpha):** Rejecting a true null hypothesis.

- **Type II Error (β - beta):** Failing to reject a false null hypothesis.

- **Trade-Off:** Balancing the risks of both error types.

- **Implications in Engineering:** Ensuring reliability and safety

Common Hypothesis Tests

ELE 503: Advanced Computer Programming and Statistics

- **t-Test:** Comparing means between two groups.

- **ANOVA (Analysis of Variance):** Comparing means among three or more groups.

- **Chi-Square Test:** Assessing relationships between categorical variables.

- **Regression Analysis:** Exploring relationships between variables.

Confidence Intervals (CI)

ELE 503: Advanced Computer Programming and Statistics

- **Definition:** Range of values within which the population parameter lies with a certain probability.

- **Components:**
  - **Point Estimate:** Sample statistic (e.g., mean).
  - **Margin of Error:** Reflects uncertainty.

- **Formula:**

  $$CI = \text{Point Estimate} \pm (\text{Critical Value} \times \text{Standard Error})$$

- **Interpretation:** 95% CI means 95% confidence that the parameter lies within the interval.

Constructing Confidence Intervals

ELE 503: Advanced Computer Programming and Statistics

- **For Means:**

$$\bar{x} \pm t_{\alpha/2,df} \times \frac{s}{\sqrt{n}}$$

- **For Proportions:**

$$\hat{p} \pm z_{\alpha/2} \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

- **Choosing the Right Interval:** Based on data type and distribution.

Applications of Confidence Intervals in Engineering

ELE 503: Advanced Computer Programming and Statistics

- **Quality Control:** Estimating mean product measurements.

- **Reliability Testing:** Determining confidence in product lifespan estimates.

- **Process Improvement:** Assessing the effectiveness of changes.

- **Risk Assessment:** Quantifying uncertainty in safety measures.

Introduction to Statistical Software Tools

ELE 503: Advanced Computer Programming and Statistics

- **R:**
  - **Strengths:** Extensive statistical packages, data visualization.
  - **Usage:** Widely used in academia and research.

- **Python:**
  - **Strengths:** Versatile, integrates with other engineering tools.
  - **Key Libraries:** NumPy, pandas, SciPy, matplotlib, seaborn.

- **Choosing the Right Tool:** Based on project requirements and personal preference

Key Libraries and Packages

ELE 503: Advanced Computer Programming and Statistics

- **R:**
  - **ggplot2:** Data visualization.
  - **dplyr:** Data manipulation.
  - **tidyr:** Data tidying.
  - **stats:** Statistical tests and models.
- **Python:**
  - **NumPy:** Numerical operations.
  - **pandas:** Data manipulation.
  - **SciPy:** Scientific computations.
  - **matplotlib & seaborn:** Data visualization

Part 3:

Python and R code Examples

Example Slide 1 – Hypothesis Testing in Python

ELE 503: Advanced Computer Programming and Statistics

•**Range:** Difference between maximum and minimum values.

Example Slide 2 - Confidence Interval Calculation in

ELE 503: Advanced Computer Programming and Statistics

•**Range:** Difference between maximum and minimum values.

Example Slide 4 - Data Visualization with ggplot2 in R

ELE 503: Advanced Computer Programming and Statistics

•**Range:** Difference between maximum and minimum values.

Example Slide 4 - Data Visualization with ggplot2 in R

ELE 503: Advanced Computer Programming and Statistics

•**Range:** Difference between maximum and minimum values.

Example Slide 3 – ANOVA in Python

ELE 503: Advanced Computer Programming and Statistics

•**Range:** Difference between maximum and minimum values.

Part 4:

Hands-On Exercise & Case Studies

Case Study 1 - Quality Control Improvement

ELE 503: Advanced Computer Programming and Statistics

- **Problem:** High defect rates in manufacturing.

- **Approach:**
  - Collect defect data.
  - Analyze variability and identify patterns.
  - Perform hypothesis testing to determine factors affecting defects.

- **Outcome:** Identified key process parameters; implemented changes reducing defects by 30%.

Case Study 2 – Reliability Engineering

ELE 503: Advanced Computer Programming and Statistics

- **Problem:** Predicting product lifespan.
- **Approach:**
  - Gather lifespan data from testing.
  - Construct confidence intervals for mean lifespan.
  - Use regression analysis to identify influencing factors.
- **Outcome:** Enhanced product design leading to increased reliability and customer satisfaction.

Hands-On Exercise 1 – Hypothesis Testing with Python

ELE 503: Advanced Computer Programming and Statistics

- **Task:** Compare the mean performance of two engineering materials.

- **Steps:**
  1. Load provided dataset.
  2. Perform an independent t-test.
  3. Interpret the results.

- **Tools:** Python, SciPy

Hands-On Exercise 2 - Confidence Interval Construction in R

ELE 503: Advanced Computer Programming and Statistics

- **Task:** Estimate the mean efficiency of a new engine design.

- **Steps:**
  1. Load the efficiency data.
  2. Calculate the 95% confidence interval for the mean.
  3. Discuss the implications.

- **Tools:** R, ggplot2

Hands-On Exercise 3 - Data Visualization with Python

ELE 503: Advanced Computer Programming and Statistics

- **Task:** Create a histogram and box plot of stress test results.

- **Steps:**
  - Load the stress test dataset.
  - Generate visualizations using matplotlib and seaborn.
  - Analyze the distribution and identify outliers.

- **Tools:** Python, matplotlib, seaborn

Summary of Key Concepts

ELE 503: Advanced Computer Programming and Statistics

- **Variability and Central Tendency:** Comprehensive understanding of data distribution.

- **Hypothesis Testing:** Making informed decisions based on data.

- **Confidence Intervals:** Quantifying uncertainty in estimates.

- **Statistical Software:** Leveraging R and Python for advanced analysis.

- **Real-World Applications:** Enhancing engineering solutions through data-driven approaches.

Importance of Statistical Software in Engineering

ELE 503: Advanced Computer Programming and Statistics

- **Efficiency:** Automate complex calculations and analyses.

- **Accuracy:** Minimize human errors in data processing.

- **Visualization:** Create insightful and interpretable visual representations.

- **Reproducibility:** Ensure analyses can be replicated and validated.

Q&A Session

ELE 503: Advanced Computer Programming and Statistics

- **Open Floor:** Address any questions or clarifications.

- **Discussion Points:**
  - Challenges faced during hands-on exercises.
  - Insights from case studies.
  - Further exploration of advanced statistical methods.

Homework Assignment

ELE 503: Advanced Computer Programming and Statistics

- **Programming Task:** 1
  - **Objective:** Perform hypothesis testing and construct confidence intervals.
  - **Tasks:**
    - Use Python or R to analyze a provided engineering dataset.
    - Conduct a t-test to compare two groups.
    - Calculate and interpret a 95% confidence interval for the mean.

- **Data Analysis Project:** 2
  - **Objective:** Apply advanced statistical methods to a real-world dataset.
  - **Tasks:**
    - Select an engineering-related dataset.
    - Analyze variability and distribution patterns.
    - Perform regression analysis to identify key factors.
    - Present findings with visualizations.

Homework Assignment

ELE 503: Advanced Computer Programming and Statistics

- **Case Study Reflection:** 3
  - **Objective:** Understand the application of statistics in engineering solutions.
  - **Tasks:**
    - Read the provided case studies.
    - Write a summary on how statistical analysis improved engineering outcomes.
    - Suggest additional statistical methods that could enhance the study.

Closing Remarks

ELE 503: Advanced Computer Programming and Statistics

- **Mastering Advanced Statistics:** Essential for tackling complex engineering challenges.

- **Programming Proficiency:** Enhances ability to perform sophisticated data analyses.

- **Real-World Impact:** Apply statistical insights to drive engineering innovations and improvements.

- **Support:** Utilize office hours, online forums, and resources for continued learning.

- **Encouragement:** Stay engaged, practice regularly, and explore advanced topics to deepen your expertise.