



Title: Spread Locator: A Statistical Distribution Analysis Model

Duration: 6 Hours

Type: Theory + Practical

🎯 Objective:

To help students understand and apply the concepts of probability distributions and spread analysis on a given dataset. This will test their theoretical understanding and practical application of distribution types, Q-Q plots, statistical transformations, and probability functions.

Project Scenario:

You are a data analyst working for an e-commerce platform that wants to understand customer purchase behaviors by analyzing daily transaction amounts of customers. Management is particularly interested in knowing whether the transactions follow certain statistical distributions, how to handle skewed data, and what probability insights can be derived.

Your task is to apply statistical distribution concepts and transformations to derive insights on transaction behaviors.

✳️ Dataset Structure ([Download Dataset](#))

Field Name	Data Type	Description
transaction_id	UUID/String	Unique identifier for each transaction
customer_id	UUID/String	Unique identifier for each customer
transaction_amount	Float	Total amount of the transaction (₹)

<code>transaction_date</code>	Date	Date of the transaction
<code>transaction_count</code>	Int	Number of transactions made by a customer in a given week
<code>region</code>	String	Customer's geographic region (North, South, East, West)
<code>transaction_status</code>	String	Whether transaction was successful or failed (Success/Fail)

★ Task Checklist

- **Part A - Theoretical Foundation (Short Notes & Explanation):**

Write a detailed note covering:

1. What is Statistical Distributions?
2. What is a Q-Q Plot and why is it used?
3. Difference between Discrete and Continuous Distributions.
4. What is Bernoulli Distribution?
5. What is Binomial Distribution?
6. Explain Log-Normal Distribution.
7. Explain Power Law Distribution.
8. What is Box-Cox Transform?
9. Explain Poisson Distribution with an example.
10. What is Z-score Probability?
11. Differentiate Probability Density Function (PDF) and Cumulative Distribution Function (CDF).

- **Part B - Data Analysis & Testing Tasks:**

Perform the following using Python (NumPy, Pandas, SciPy, Statsmodels, Matplotlib, Seaborn):

1. Fit the data to Bernoulli and Binomial distributions (transaction occurrence & weekly count).
2. Fit the data to Poisson distribution (number of transactions per day).
3. Model transaction amounts using Log-Normal and Power Law distributions.
4. Generate and interpret a Q-Q Plot to test normality.
5. Apply Box-Cox Transform to stabilize variance.

6. Calculate Z-scores for transaction amounts and compute probability of transactions exceeding ₹5000.
 7. Plot and interpret PDF and CDF for transaction amounts.
 8. Conclude which distribution best fits the dataset and justify insights for decision-making.
-

 **Submission Guidelines**

- Include practical implementation in Jupyter Notebook, Excel/Sheets, or screenshots.
- Label all charts clearly and write short interpretations under each result.
- GitHub Repository:
 - Create a GitHub repository to host your project.
 - Upload your project files, including source code and documentation.
 - Add a document (PDF) explaining theory concepts with definitions.
 - Ensure that you provide a clear and descriptive README.md file.

Spread Locator
Mathematics & Advanced Statistics

BRING ON YOUR CODING ATTITUDE