National University of Singapore
TCX2002 Introduction to Business Analytics
Tutorial 1

Lesson 3 - Understanding Distributions, Sampling, and Estimation

1. Probability Distribution,
2. The Normal Curve;
3. Population vs Sample,
4. Sampling & Estimation (simple random, systematic, stratified, and convenience sampling),
5. Sampling error vs bias,
6. Point estimation
7. STD Error (SE) & Confidence Intervals

1. Simulating Randomness and Outcome Space

Lab Questions:

1. Simulate 100 coin tosses in R using sample(c("Heads", "Tails"), 100, replace = TRUE).
2. Count how many times "Heads" appears. What if you increase to 1000 coin tosses or more?
3. Is the outcome approximately 50-50? What does this tell you about randomness?

2. Exploring Sampling Techniques

Scenario: Your company wants to survey customers.

Lab Questions:

1. Create a dummy dataset with 500 customers and assign them to regions (North, South, East, West).
2. Perform:
    - Simple random sampling
    - Stratified sampling (by region)
    - Systematic sampling (every 10th customer)
3. Compare the results: Are all regions equally represented in each technique?

National University of Singapore
TCX2002 Introduction to Business Analytics
Tutorial 1

3. Confidence Intervals and Interpretation

Scenario: A manager wants to know the average time to resolve customer complaints.

Lab Questions:

1. Use a dataset with complaint resolution times. **We use a distribution that is skewed to the right (more common in real data)**
   set.seed(123)
   complaint_times <- rgamma(100, shape = 2, scale = 5)

2. Calculate the mean and standard deviation.
3. Compute a 95% confidence interval for average resolution time.

Scenario: An e-commerce site tracks conversion rates. In 1,200 visitors:

- 84 made purchases
- Sample conversion rate: $\hat{p}$ = 84/1,200 = 7%

Lab Questions:

1. Check if conditions are met for normal approximation
2. Calculate standard error for the proportion
3. Build 95% confidence interval for true conversion rate
4. How would you communicate this to the marketing team?

National University of Singapore
TCX2002 Introduction to Business Analytics
Tutorial 1

Tutorial 1 Learning Outcomes	
1.	**Ability to differentiate between key statistical concepts.** This includes distinguishing between a population and a sample, understanding the difference between sampling error and bias, and being able to explain how the Normal Curve and probability distributions relate to the data.
2.	**Able to apply various sampling techniques and calculate key statistics,** perform and explain different methods like **simple random, systematic, stratified, and convenience sampling**. Able to compute **point estimates**, such as the mean, and determine the **standard error** and **confidence intervals** to quantify the uncertainty of their estimates.
3.	**Able to interpret statistical results and their implications for data analysis.** This involves understanding what a **confidence interval** means in a practical context and how factors like **sample size** and **data variability** affect the precision of an estimate. They will also be able to handle **outliers** in a dataset, which is a crucial step in preparing data for analysis.