Session 7: Frequency and Probability

Objective: Familiarize students with key probability concepts and their application in business decision-making, including:

- Probability
- Sample space
- Events
- Independence

Pre-Class Work (PCW)

Task: Use the provided dataset (Market Research Study) to calculate the frequency of different survey responses and derive probabilities.

Concepts Covered:

- Frequency distribution
- Empirical probability
- Probability of independent and dependent events

Step 1: Understanding Frequency and Probability

Question 1: *Why is understanding probability important in business decision-making?*

Excel Instructions (Detailed):

1. Open the dataset in Excel.
2. Use **Pivot Tables** to calculate the frequency of different responses (e.g., product category preferences).
3. Calculate probabilities using:
    - **Relative Frequency (Empirical Probability):** `=COUNTIF(Range, Criteria)/COUNTA(Range)`

Python Instructions:

python

```python
import pandas as pd
df = pd.read_excel("Market_Research_Dataset.xlsx")

# Frequency count of Product Category
category_counts = df["Product_Category"].value_counts()
```

```
total_responses = len(df)

# Empirical probability
category_probabilities = category_counts / total_responses
print(category_probabilities)
```

Question 2: *What is the probability that a randomly selected customer prefers online shopping?*

Step 2: Conditional Probability and Independence

Question 3: *What is the probability that a customer who prefers online shopping is also very satisfied?*

- **Excel Instructions:**

    1. Use a **Pivot Table** to filter for customers who prefer online shopping.
    2. Calculate the probability of being very satisfied within this subgroup.

Python Instructions:

 python

```
online_shoppers = df[df["Preferred_Shopping_Channel"] == "Online"]
prob_very_satisfied =
len(online_shoppers[online_shoppers["Satisfaction_Level"] == "Very
Satisfied"]) / len(online_shoppers)
print(prob_very_satisfied)
```

-

 Question 4: *Are shopping preferences and satisfaction levels independent? Why or why not?*

Step 3: Visualizing Frequency and Probability

Question 5: *Create a bar chart showing the distribution of preferred shopping channels. What trends do you notice?*

- **Excel Instructions:**

    1. Select the **Preferred Shopping Channel** column.
```

2. Click **Insert > Charts > Bar Chart**.

**Python Instructions:**

python

```python
import matplotlib.pyplot as plt
category_counts.plot(kind="bar")
plt.xlabel("Product Category")
plt.ylabel("Frequency")
plt.title("Product Category Preferences")
plt.show()
```

-

◆ **Save and submit your calculations before class.**

Polls (Verifying PCW Completion)

Poll A

"If a product category has a probability of 0.3, what does this mean?"

1. 30% of respondents prefer this category.
2. There is a 30% chance every customer will buy this product.
3. The product is bought in 30% of stores.
4. The product is the most popular one.

(Correct answer: **1. 30% of respondents prefer this category.**)

Poll B

"If two events are independent, what does that mean?"

1. One event affects the probability of the other.
2. The probability of both happening is always zero.
3. The occurrence of one does not change the probability of the other.
4. They always have the same probability.

(Correct answer: **3. The occurrence of one does not change the probability of the other.**)

Poll C

"You found that 60% of respondents prefer online shopping and 40% prefer in-store shopping. What is the probability of randomly selecting two people who both prefer online shopping?"

1. 0.60
2. 0.24
3. 0.36
4. 0.40

(Correct answer: **3. 0.36** → 0.60 × 0.60)

Breakout Activity: Post-Sale Survey Analysis

Scenario:
Your company conducted a post-sale survey to understand customer satisfaction and repurchase likelihood. Your task is to analyze the probability of certain customer behaviors.

Dataset: Post-Sale Survey Responses

Step 1: Determining Event Probabilities

Question 1: *What is the probability that a randomly selected customer is "Very Satisfied"?*

Question 2: *What is the probability that a customer who purchased online is also a repeat customer?*

- **Excel Instructions:**

  1. Use `=COUNTIF(Satisfaction_Level, "Very Satisfied")/COUNTA(Satisfaction_Level)`.
  2. Use **Pivot Tables** to filter and calculate conditional probabilities.

Python Instructions:

python

```python
df = pd.read_excel("Post_Sale_Survey_Dataset.xlsx")

prob_very_satisfied = len(df[df["Customer_Satisfaction"] == "Very Satisfied"]) / len(df)
```

```python
prob_repeat_given_online = len(df[(df["Purchase_Location"] ==
"Online") & (df["Repeat_Customer"] == "Yes")]) /
len(df[df["Purchase_Location"] == "Online"])

print(f"Probability of Very Satisfied: {prob_very_satisfied}")
print(f"Probability of Repeat Customer given Online Purchase:
{prob_repeat_given_online}")
```

-

Step 2: Conditional and Joint Probabilities

Question 3: *What is the probability that a randomly selected customer both recommends the product and is a repeat customer?*

 Question 4: *What is the probability that a customer recommends the product given that they are "Satisfied" or "Very Satisfied"?*

Step 3: Probability of Independent vs. Dependent Events

 Question 5: *Are satisfaction level and likelihood to recommend independent events? Why or why not?*

- **Excel Instructions:**

    1. Compare `P(A and B)` vs. `P(A) * P(B)`.

Python Instructions:

 python

```python
prob_recommend = len(df[df["Recommend_to_Friend"] == "Yes"]) /
len(df)
prob_satisfied =
len(df[df["Customer_Satisfaction"].isin(["Satisfied", "Very
Satisfied"])]) / len(df)
prob_both = len(df[(df["Recommend_to_Friend"] == "Yes") &
(df["Customer_Satisfaction"].isin(["Satisfied", "Very Satisfied"])]
) / len(df)
```

```
is_independent = prob_both == (prob_recommend * prob_satisfied)
print(f"Are events independent? {is_independent}")
```

-

Step 4: Business Decision

Final Strategic Question:
 "Based on your probability analysis, which customer segment is most likely to be repeat customers? What marketing strategy would you suggest?"