Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Probability and Statistics (01CT1401)	**Aim: Term Work – Tutorial – 2 (Statistics)**	
Tutorial - 2	**Date:-** 20-04-2024	**Enrollment No:-** 92200133030

Problem Statement:-

The Student News Service at Marwadi University (MU) has decided to gather data about the undergraduate students that attend MU. MU creates and distributes a survey of 14 questions and receives responses from 62 undergraduates (stored in the Survey data set).

Importing Necessary Libraries and Dataset :-

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

Dataset = pd.read_excel("./Tutorial 2 Tk1 A batch_MU_Students.xlsx", sheet_name="Q.2")
```

2.1. For this data, construct the following contingency tables (Keep Gender as row variable.

1. **Gender and Major**

   Code :-

```
contingency_table_1 = pd.crosstab(Dataset["Major"], Dataset["Gender"])
print("Contingency Table for Gender and Major:")
print(contingency_table_1)
```

   Output :-

```
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_2> & "C:/Program Files/Python312/python.e
xe" "d:/Aryan Data/Usefull Data/Semester - 4/Probability and Statistics/Tutorials/Tutorial_2/Tutorial_2.py"
Contingency Table for Gender and Major:
Gender                Female  Male
Major
Accounting                 3     4
CIS                        3     1
Economics/Finance          7     4
International Business     4     2
Management                 4     6
Other                      3     4
Retailing/Marketing        9     5
Undecided                  0     3
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_2>
```

2. **Gender and Grad Intention**

   Code :-

```
contingency_table_2 = pd.crosstab(Dataset["Grad Intention"], Dataset["Gender"])
print("Contingency Table for Gender and Grand Intention:- ")
print(contingency_table_2)
```

Output :-

```
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_2> & "C:/Program Files/Python312/python.exe" "d
:/Aryan Data/Usefull Data/Semester - 4/Probability and Statistics/Tutorials/Tutorial_2/Tutorial_2.py"
Contingency Table for Gender and Grand Intention:-
Gender          Female  Male
Grad Intention
No                   9     3
Undecided           13     9
Yes                 11    17
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_2>
```

3. Gender and Employment

Code :-

```
contingency_table_3 = pd.crosstab(Dataset["Employment"], Dataset["Gender"])
print("Contingency Table for Gender and Employment:- ")
print(contingency_table_3)
```

Output :-

```
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_2> & "C:/Program Files/Python312/python.exe" "d
:/Aryan Data/Usefull Data/Semester - 4/Probability and Statistics/Tutorials/Tutorial_2/Tutorial_2.py"
Contingency Table for Gender and Employment:-
Gender      Female  Male
Employment
Full-Time        3     7
Part-Time       24    19
Unemployed       6     3
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_2>
```

4. Gender and Computer

Code :-

```
contingency_table_4 = pd.crosstab(Dataset["Computer"], Dataset["Gender"])
print("Contingency Table for Gender and Computer:- ")
print(contingency_table_4)
```

Output :-

```
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_2> & "C:/Program Files/Python31
2/python.exe" "d:/Aryan Data/Usefull Data/Semester - 4/Probability and Statistics/Tutorials/Tutorial_2/Tutorial_2.py"
Contingency Table for Gender and Computer:-
Gender    Female  Male
Computer
Desktop        2     3
Laptop        29    26
Tablet         2     0
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_2>
```

Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Probability and Statistics (01CT1401)	**Aim: Term Work – Tutorial – 2 (Statistics)**	
Tutorial - 2	**Date:-** 20-04-2024	**Enrollment No:-** 92200133030

2.2. Assume that the sample is representative of the population of MU. Based on the data, answer the following question:

1. **What is the probability that a randomly selected MU student will be male?**

   Code :-

```
print(f"The Probability that a randomly selected MU student will be male is
{"{:.3f}".format((Dataset[(Dataset['Gender'] == "Male")].shape[0] / Dataset.shape[0]) * 100)} %")
```

   Output :-

```
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_2> & "C:/Program Files/Python31
2/python.exe" "d:/Aryan Data/Usefull Data/Semester - 4/Probability and Statistics/Tutorials/Tutorial_2/Tutorial_2.py"
The Probability that a randomly selected MU student will be male is 46.774 %
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_2>
```

2. **What is the probability that a randomly selected MU student will be female?**

   Code :-

```
print(f"The Probability that a randomly selected MU student will be female is
{"{:.3f}".format((Dataset[(Dataset['Gender'] == "Female")].shape[0] / Dataset.shape[0]) * 100)} %")
```

   Output :-

```
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_2> & "C:/Program Files/Python31
2/python.exe" "d:/Aryan Data/Usefull Data/Semester - 4/Probability and Statistics/Tutorials/Tutorial_2/Tutorial_2.py"
The Probability that a randomly selected MU student will be female is 53.226 %
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_2>
```

2.3. Assume that the sample is representative of the population of MU. Based on the data, answer the following question:

1. **Find the conditional probability of different majors among the male students in MU.**

   Code :-

```
Males_Data = Dataset[(Dataset['Gender'] == "Male")]
print("Conditional Probability of different majors among the male students in MU :-")
print(Males_Data["Major"].value_counts(normalize=True))
```

Output :-

```
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_2> & "C:/Program Files/Python312/python.exe" "d:/Aryan Data/Usefull Data/Semester - 4/Probability and Statistics/Tutorials/Tutorial_2/Tutorial_2.py"
Conditional Probability of different majors among the male students in MU :-
Major
Management               0.206897
Retailing/Marketing      0.172414
Other                    0.137931
Economics/Finance        0.137931
Accounting               0.137931
Undecided                0.103448
International Business   0.068966
CIS                      0.034483
Name: proportion, dtype: float64
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_2>
```

2. Find the conditional probability of different majors among the female students of MU.

Code :-

```
Females_Data = Dataset[(Dataset["Gender"] == "Female")]
print("Conditional Probability of different majors among the female students in MU :-")
print(Females_Data["Major"].value_counts(normalize=True))
```

Output :-

```
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_2> & "C:/Program Files/Python312/python.exe" "d:/Aryan Data/Usefull Data/Semester - 4/Probability and Statistics/Tutorials/Tutorial_2/Tutorial_2.py"
Conditional Probability of different majors among the female students in MU :-
Major
Retailing/Marketing      0.272727
Economics/Finance        0.212121
Management               0.121212
International Business   0.121212
Other                    0.090909
CIS                      0.090909
Accounting               0.090909
Name: proportion, dtype: float64
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_2>
```

Marwadi University Faculty of Technology Department of Information and Communication Technology		
Subject: Probability and Statistics (01CT1401)	**Aim: Term Work – Tutorial – 2 (Statistics)**	
Tutorial - 2	**Date:-** 20-04-2024	**Enrollment No:-** 92200133030

2.4 Assume that the sample is a representative of the population of MU. Based on the data, answer the following question:

1. **Find the probability That a randomly chosen student is a male and intends to graduate.**

Code :-

```
Graduate_Male = Dataset[(Dataset["Gender"] == "Male") & (Dataset["Grad Intention"] == "Yes")]
probability = '{:.3f}' .format(Graduate_Male.shape[0] / Dataset.shape[0])
print(f"The Probability That a randomly chosen student is a male and intends to graduate is {probability} %")
```

Output :-

```
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_2> & "C:/Program Files/Python31
2/python.exe" "d:/Aryan Data/Usefull Data/Semester - 4/Probability and Statistics/Tutorials/Tutorial_2/Tutorial_2.py"
The Probability That a randomly chosen student is a male and intends to graduate is 0.274 %
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_2>
```

2. **Find the probability that a randomly selected student is a female and does NOT have a laptop.**

Code :-

```
Not_Laptop_Feale = Dataset[(Dataset["Gender"] == "Female") & (Dataset["Computer"] != "Laptop")]
probability = "{:.3f}".format(Not_Laptop_Feale.shape[0] / Dataset.shape[0])
print(f"The Probability That a randomly chosen student is a Female and not having laptop is {probability} %")
```

Output :-

```
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_2> & "C:/Program Files/Python31
2/python.exe" "d:/Aryan Data/Usefull Data/Semester - 4/Probability and Statistics/Tutorials/Tutorial_2/Tutorial_2.py"
The Probability That a randomly chosen student is a Female and not having laptop is 0.065 %
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_2>
```

2.5 Assume that the sample is representative of the population of MU. Based on the data, answer the following question:

1. **Find the probability that a randomly chosen student is either a male or has full-time employment?**

Code :-

```
Male_or_full_time = Dataset[(Dataset["Gender"] == "Male") | (Dataset["Employment"] == "Full-Time")]
probability = "{:.3f}".format(Male_or_full_time.shape[0] / Dataset.shape[0])
print(f"The Probability that a randomly chosen student is either a male or has full-time employment is
{probability}")
```

Output :-

```
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_2> & "C:/Program Files/Python31
2/python.exe" "d:/Aryan Data/Usefull Data/Semester - 4/Probability and Statistics/Tutorials/Tutorial_2/Tutorial_2.py"
The Probability that a randomly chosen student is either a male or has full-time employment is 0.516
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_2>
```

2. **Find the conditional probability that given a female student is randomly chosen, she is majoring in international business or management.**

Code :-

```
Female_or_full_time = Dataset[(Dataset["Gender"] == "Female")& ((Dataset["Major"] == "Management")|
(Dataset["Major"] == "International Business"))]
probability = "{:.3f}".format(Female_or_full_time.shape[0] / Dataset.shape[0])
print(f"The Probability that given a female student is randomly chosen, she is majoring in international business or
management. is {probability}")
```

Output :-

```
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_2> & "C:/Program Files/Python31
2/python.exe" "d:/Aryan Data/Usefull Data/Semester - 4/Probability and Statistics/Tutorials/Tutorial_2/Tutorial_2.py"
The Probability that given a female student is randomly chosen, she is majoring in international business or management.
 is 0.129
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_2>
```

2.6 Construct a contingency table of Gender and Intent to Graduate at 2 levels (Yes/No). The Undecided students are not considered now and the table is a 2x2 table. Do you think the graduate intention and being female are independent events?

Code :-

```
subset = Dataset[Dataset["Grad Intention"].isin(["Yes", "No"])]
cont_table = pd.crosstab(subset["Gender"], subset["Grad Intention"])
print(cont_table)
sum = 0

for i in range(0,2) :
    for j in range(0,2) :
        sum = sum + cont_table.iloc[i, j]

for i in range(0, 2):
    for j in range(0, 2):
        cont_table.iloc[i, j] = cont_table.iloc[i, j] / sum
```

```
print(cont_table)

p_female = 0
graduate = 0

for i in range(0, 2):
    p_female = p_female + cont_table.iloc[0, i]

for i in range(0, 2):
    graduate = graduate + cont_table.iloc[i, 1]

if (cont_table.iloc[0, 1] == p_female * graduate) :
    print("The Events are Independent")

else :
    print("The Events are Dependent")
```

Output :-

```
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_2> & "C:/Program Files/Python312/python.exe" "d
:/Aryan Data/Usefull Data/Semester - 4/Probability and Statistics/Tutorials/Tutorial_2/Tutorial_2.py"
Grad Intention  No  Yes
Gender
Female           9   11
Male             3   17
d:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_2\Tutorial_2.py:64: FutureWarning: Setting an item
 of incompatible dtype is deprecated and will raise an error in a future version of pandas. Value '0.225' has dtype incompatible with in
t64, please explicitly cast to a compatible dtype first.
  cont_table.iloc[i, j] = cont_table.iloc[i, j] / sum
d:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_2\Tutorial_2.py:64: FutureWarning: Setting an item
 of incompatible dtype is deprecated and will raise an error in a future version of pandas. Value '0.275' has dtype incompatible with in
t64, please explicitly cast to a compatible dtype first.
  cont_table.iloc[i, j] = cont_table.iloc[i, j] / sum
Grad Intention     No    Yes
Gender
Female          0.225  0.275
Male            0.075  0.425
The Events are Dependent
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_2> & "C:/Program Files/Python312/python.exe" "d
:/Aryan Data/Usefull Data/Semester - 4/Probability and Statistics/Tutorials/Tutorial_2/Tutorial_2.py"
```

Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Probability and Statistics (01CT1401)	**Aim: Term Work – Tutorial – 2 (Statistics)**	
Tutorial - 2	**Date:-** 20-04-2024	**Enrollment No:-** 92200133030

2.7. Note that there are four numerical (continuous) variables in the data set, GPA, Salary, Spending, and Text Messages. Answer the following questions based on the data

1. **If a student is chosen randomly, what is the probability that his/her GPA is less than 3?**

   Code :-

```
num_low_gpa = Dataset[Dataset["GPA"] < 3]["GPA"].count()
total_students = Dataset["GPA"].count()
prob_low_gpa = num_low_gpa / total_students
print(f"The probability that a student has a GPA less than 3 is {prob_low_gpa:.2f}")
```

   Output :-

```
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_2> & "C:/Program Files/Python312/python.exe" "d
:/Aryan Data/Usefull Data/Semester - 4/Probability and Statistics/Tutorials/Tutorial_2/Tutorial_2.py"
The probability that a student has a GPA less than 3 is 0.27
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_2>
```

2. **Find the conditional probability that a randomly selected male earns 50 or more. Find the conditional probability that a randomly selected female earns 50 or more.**

   Code :-

```
table = pd.crosstab(index=Dataset["Gender"],columns=pd.cut(Dataset["Wealth"], bins=[-1, 49, 100], labels=["<50",
">=50"]),)
total_males = table.loc["Male"].sum()
total_females = table.loc["Female"].sum()
prob_male_50_or_more = table.loc["Male", ">=50"] / total_males
prob_female_50_or_more = table.loc["Female", ">=50"] / total_females
print("Conditional probability that a randomly selected male earns 50 or more:
{:.2%}".format(prob_male_50_or_more))
print("Conditional probability that a randomly selected female earns 50 or more:
{:.2%}".format(prob_female_50_or_more))
```

   Output :-‘

```
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_2> & "C:/Program Files/Python312/python.exe" "d
:/Aryan Data/Usefull Data/Semester - 4/Probability and Statistics/Tutorials/Tutorial_2/Tutorial_2.py"
Conditional probability that a randomly selected male earns 50 or more: 10.34%
Conditional probability that a randomly selected female earns 50 or more: 0.00%
PS D:\Aryan Data\Usefull Data\Semester - 4\Probability and Statistics\Tutorials\Tutorial_2>
```

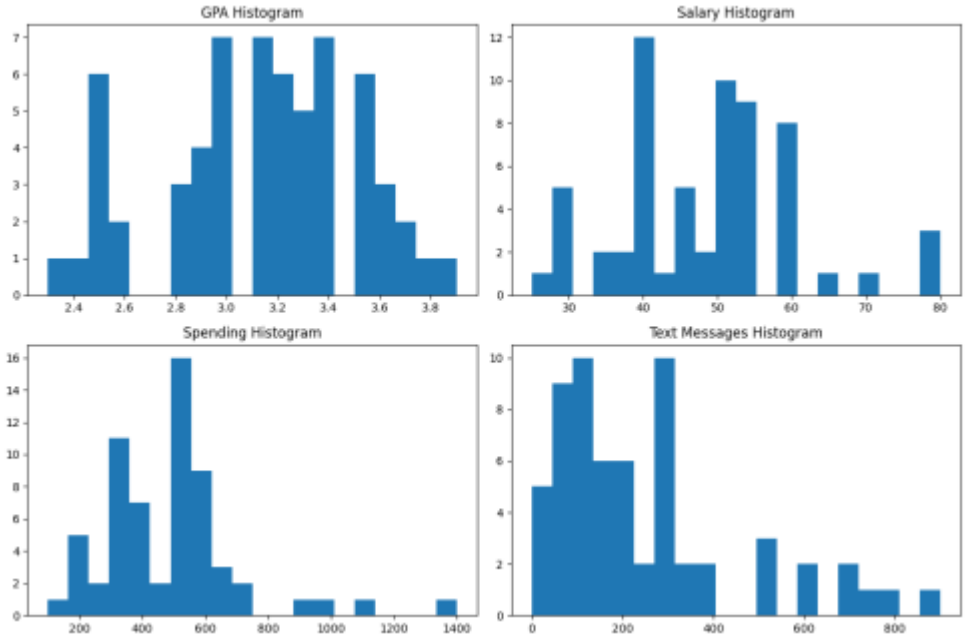
Marwadi University Faculty of Technology Department of Information and Communication Technology		
Subject: Probability and Statistics (01CT1401)	**Aim: Term Work – Tutorial – 2 (Statistics)**	
Tutorial - 2	**Date:-** 20-04-2024	**Enrollment No:-** 92200133030

2.8. Note that there are four numerical (continuous) variables in the data set, GPA, Salary, Spending, and Text Messages. For each of them comment whether they follow a normal distribution. Write a note summarizing your conclusions.

Code :-

```
plt.figure(figsize=(12, 8))
plt.subplot(2, 2, 1)
plt.hist(Dataset["GPA"], bins=20)
plt.title("GPA Histogram")

plt.subplot(2, 2, 2)
plt.hist(Dataset["Salary"], bins=20)
plt.title("Salary Histogram")

plt.subplot(2, 2, 3)
plt.hist(Dataset["Spending"], bins=20)
plt.title("Spending Histogram")

plt.subplot(2, 2, 4)
plt.hist(Dataset["Text Messages"], bins=20)
plt.title("Text Messages Histogram")

plt.tight_layout()
plt.show()
```

Output :-



Marwadi University Marwadi Chandarana Group NAAC A+	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Probability and Statistics (01CT1401)	**Aim: Term Work – Tutorial – 2 (Statistics)**	
Tutorial - 2	**Date:-** 20-04-2024	**Enrollment No:-** 92200133030

In conclusion, GPA and Salary appears to be approximately normally distributed, while Spending, and Text Messages are not normally distributed.