# Problem 2

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

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

### *2.1.1. Gender and Major*

| Major<br>Gender | Accounting | CIS | Economics/Finance | International Business | Management | Other | Retailing/Marketing | Undecided | All |
|---|---|---|---|---|---|---|---|---|---|
| **Female** | 3 | 3 | 7 | 4 | 4 | 3 | 9 | 0 | 33 |
| **Male** | 4 | 1 | 4 | 2 | 6 | 4 | 5 | 3 | 29 |
| **All** | 7 | 4 | 11 | 6 | 10 | 7 | 14 | 3 | 62 |

### *2.1.2. Gender and Grad Intention*

| Grad Intention<br>Gender | No | Undecided | Yes | All |
|---|---|---|---|---|
| **Female** | 9 | 13 | 11 | 33 |
| **Male** | 3 | 9 | 17 | 29 |
| **All** | 12 | 22 | 28 | 62 |

### *2.1.3. Gender and Employment*

| Employment<br>Gender | Full-Time | Part-Time | Unemployed | All |
|---|---|---|---|---|
| **Female** | 3 | 24 | 6 | 33 |
| **Male** | 7 | 19 | 3 | 29 |
| **All** | 10 | 43 | 9 | 62 |

### *2.1.4. Gender and Computer*

| Computer<br>Gender | Desktop | Laptop | Tablet | All |
|---|---|---|---|---|
| **Female** | 2 | 29 | 2 | 33 |
| **Male** | 3 | 26 | 0 | 29 |
| **All** | 5 | 55 | 2 | 62 |

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

### *2.2.1. What is the probability that a randomly selected CMSU student will be male?*

Total students = 62

Total male students = 29

P(Male) = 29/62

= 0.4677

**The probability of males is 46.77%.**

### *2.2.2. What is the probability that a randomly selected CMSU student will be female?*

Total students = 62

Total female students = 33

P(Female) = 33/62

= 0.5322

**The probability of females is 53.23%.**

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

### *2.3.1. Find the conditional probability of different majors among the male students in CMSU.*

Using the contingency table in 2.1.1, we got the total numbers of males and females opting for different majors.

| Major<br>Gender | Accounting | CIS | Economics/Finance | International Business | Management | Other | Retailing/Marketing | Undecided | All |
|---|---|---|---|---|---|---|---|---|---|
| Female | 3 | 3 | 7 | 4 | 4 | 3 | 9 | 0 | 33 |
| Male | 4 | 1 | 4 | 2 | 6 | 4 | 5 | 3 | 29 |
| All | 7 | 4 | 11 | 6 | 10 | 7 | 14 | 3 | 62 |

1. Accounting Major

Males who have taken accounting as major = 4

Total males = 29

P(Male_accounting) = 4/29

= 0.1379

**Probability of males opting for accounting is 13.79%.**

2. CIS Major

Males who have taken CIS as major = 1

Total males = 29

P(Male_CIS) = 1/29

= 0.0344

**Probability of males opting for CIS is 3.45%.**

3. Economics/Finance Major

Males who have taken economics/finance as major = 4

Total males = 29

P(Male_ economics_finance) = 4/29

= 0.1379

**Probability of males opting for economics/finance is 13.79%.**

4. International Business Major

Males who have taken international business as major = 2

Total males = 29

P(Male_ international_business) = 2/29

= 0.0689

**Probability of males opting for international business is 6.90%.**

5. Management Major

Males who have taken management as major = 6

Total males = 29

P(Male_ management) = 6/29

= 0.2068

**Probability of males opting for management is 20.69%.**

6. Other Major

Males who have taken other as major = 4

Total males = 29

P(Male_ other) = 4/29

= 0.1379

**Probability of males opting for other is 13.79%.**

7. Retailing/Marketing Major

Males who have taken retailing/marketing as major = 5

Total males = 29

P(Male_ retailing/marketing) = 5/29

= 0.1724

**Probability of males opting for retailing /marketing is 17.24%.**

8. Undecided

Males who have not decided the major (undecided) = 3

Total males = 29

P(Male_ undecided) = 3/29

= 0.1034

**Probability of males who have not decided the major is 10.34%.**

### *2.3.2 Find the conditional probability of different majors among the female students of CMSU.*

Using the contingency table in 2.1.1, we got the total numbers of males and females opting for different majors.

| Major<br>Gender | Accounting | CIS | Economics/Finance | International Business | Management | Other | Retailing/Marketing | Undecided | All |
|---|---|---|---|---|---|---|---|---|---|
| **Female** | 3 | 3 | 7 | 4 | 4 | 3 | 9 | 0 | 33 |
| **Male** | 4 | 1 | 4 | 2 | 6 | 4 | 5 | 3 | 29 |
| **All** | 7 | 4 | 11 | 6 | 10 | 7 | 14 | 3 | 62 |

1. Accounting Major

Females who have taken accounting as major = 3

Total females = 33

P(Female_accounting) = 3/33

= 0.0909

**Probability of females opting for accounting is 9.09%.**

2. CIS Major

Females who have taken CIS as major = 3

Total females = 33

P(Female_CIS) = 3/33

= 0.0909

**Probability of females opting for CIS is 9.09%.**

3. Economics/Finance Major

Females who have taken economics/finance as major = 7

Total females = 33

P(Female_ economics_finance) = 7/33

= 0.2121

**Probability of females opting for economics/finance is 21.21%.**

4. International Business Major

Females who have taken international business as major = 4

Total females = 33

P(Female_ international_business) = 4/33

= 0.1212

**Probability of females opting for international business is 12.12%.**

5. Management Major

Females who have taken management as major = 4

Total females = 33

P(Female_ management) = 4/33

= 0.1212

**Probability of females opting for management is 12.12%.**

6. Other Major

Females who have taken other as major = 3

Total females = 33

P(Female_ other) = 3/33

= 0.0909

**Probability of females opting for other is 9.09%.**

7. Retailing/Marketing Major

Females who have taken retailing/marketing as major = 9

Total females = 33

P(Female_ retailing/marketing) = 9/33

= 0.2727

**Probability of females opting for retailing /marketing is 27.27%.**

8. Undecided

Females who have not decided the major (undecided) = 0

Total females = 33

P(Female_ undecided) = 0/33

= 0

**Probability of females who have not decided the major is 0.00%.**

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

### *2.4.1. Find the probability that a randomly chosen student is a male and intends to graduate.*

Using the contingency table in 2.1.2, we got the total numbers of males and females who intend to graduate.

| Grad Intention | No | Undecided | Yes | All |
|---|---|---|---|---|
| Gender | | | | |
| Female | 9 | 13 | 11 | 33 |
| Male | 3 | 9 | 17 | 29 |
| All | 12 | 22 | 28 | 62 |

Males who have the intention to graduate = 17

Total students = 62

P(Male_gradintention) = 17/62

= 0.2742

**Probability of males who intend to graduate is 27.42%.**

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

Using the contingency table in 2.1.4, we got the total numbers of males and females who have a desktop, laptop or a tablet

| Computer | Desktop | Laptop | Tablet | All |
|---|---|---|---|---|
| Gender | | | | |
| Female | 2 | 29 | 2 | 33 |
| Male | 3 | 26 | 0 | 29 |
| All | 5 | 55 | 2 | 62 |

Females who have desktop = 2

Females who have tablet = 2

Total students = 62

P(Female_no_laptop) = (2+2)/62

= 0.645

**Probability of females who do not have a laptop is 6.45%.**

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

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

Using the contingency table in 2.1.3, we got the total numbers of males and females who are employed.

| Employment / Gender | Full-Time | Part-Time | Unemployed | All |
|---|---|---|---|---|
| **Female** | 3 | 24 | 6 | 33 |
| **Male** | 7 | 19 | 3 | 29 |
| **All** | 10 | 43 | 9 | 62 |

Total males = 29

Total full time employment = 10

Total number of males with full time employment = 7

Total number of students = 62

P(Male_or_fulltimeemployment) = (29+10-7)/62

= 0.5161

**Probability that a randomly chosen student is a male of has full time employment is 51.61%.**

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

Using the contingency table in 2.1.1, we got the total numbers of males and females opting for different majors.

| Major / Gender | Accounting | CIS | Economics/Finance | International Business | Management | Other | Retailing/Marketing | Undecided | All |
|---|---|---|---|---|---|---|---|---|---|
| **Female** | 3 | 3 | 7 | 4 | 4 | 3 | 9 | 0 | 33 |
| **Male** | 4 | 1 | 4 | 2 | 6 | 4 | 5 | 3 | 29 |
| **All** | 7 | 4 | 11 | 6 | 10 | 7 | 14 | 3 | 62 |

Probability (International Business/Management | Female)

Total students = 33

Females opting for international business = 4

Females opting for management = 4

P(Female_internationalbusiness_management) = (4+4)/33

= 0.2424

**Probability that a female student randomly chosen, she is majoring in international business or management is 24.24%.**

## 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?

| Grad Intention | No | Yes |
|---|---|---|
| Gender | | |
| Female | 9 | 11 |
| Male | 3 | 17 |

Events A and B are independent if the equation $P(A \cap B) = P(A) \cdot P(B)$ holds true.

| Grad Intention | No | Yes | All |
|---|---|---|---|
| Gender | | | |
| Female | 9 | 11 | 20 |
| Male | 3 | 17 | 20 |
| All | 12 | 28 | 40 |

Let us consider P(A) as the probability of a student with graduate intention, P(B) as the probability of the student being a female. Therefore $P(A \cap B)$ will be the probability of a female graduate.

Total number of students = 40

Number of students with graduate intention = 28

Total number of female students = 20

Number of female students with graduate intention = 11

P(A) = 28/40

= 0.7

P(B) = 20/40

= 0.5

$P(A \cap B) = 11/40$

= 0.275

$P(A) \cdot P(B) = 0.35$

**From the above calculated values it is seen that $P(A \cap B)$ is not equal to $P(A) \cdot P(B)$. Therefore the events are not independent.**

## 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

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

Total students = 62

Students with GPA less than 3 = 17

P(GPA_lessthan3) = 17/62

= 0.2741

**Probability of a student that his/her GPA is less than 3 is 27.42%.**

### *2.7.2. 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.*

**Probability (Male|50 or more):**

Total males = 29

Males with salary 50 or more = 14

P(Male_50ormore) = 14/29

= 0.4827

**Probability that a randomly selected male earns 50 or more is 48.28%.**

**Probability (Female|50 or more)**

Total females = 33

Females with salary 50 or more = 18

P(Female_50ormore) = 18/33

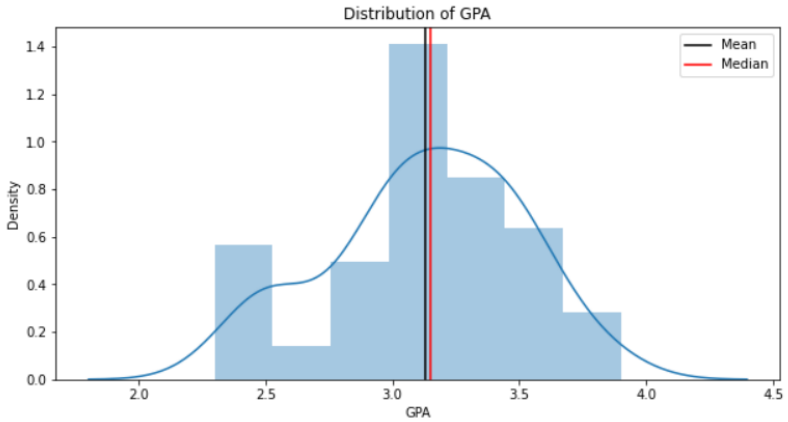
= 0.5454

**Probability that a randomly selected female earns 50 or more is 54.55%.**

## 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.

| | ID | Age | GPA | Salary | Social Networking | Satisfaction | Spending | Text Messages |
|---|---|---|---|---|---|---|---|---|
| **count** | 62.000000 | 62.000000 | 62.000000 | 62.000000 | 62.000000 | 62.000000 | 62.000000 | 62.000000 |
| **mean** | 31.500000 | 21.129032 | 3.129032 | 48.548387 | 1.516129 | 3.741935 | 482.016129 | 246.209677 |
| **std** | 18.041619 | 1.431311 | 0.377388 | 12.080912 | 0.844305 | 1.213793 | 221.953805 | 214.465950 |
| **min** | 1.000000 | 18.000000 | 2.300000 | 25.000000 | 0.000000 | 1.000000 | 100.000000 | 0.000000 |
| **25%** | 16.250000 | 20.000000 | 2.900000 | 40.000000 | 1.000000 | 3.000000 | 312.500000 | 100.000000 |
| **50%** | 31.500000 | 21.000000 | 3.150000 | 50.000000 | 1.000000 | 4.000000 | 500.000000 | 200.000000 |
| **75%** | 46.750000 | 22.000000 | 3.400000 | 55.000000 | 2.000000 | 4.000000 | 600.000000 | 300.000000 |
| **max** | 62.000000 | 26.000000 | 3.900000 | 80.000000 | 4.000000 | 6.000000 | 1400.000000 | 900.000000 |

1. GPA



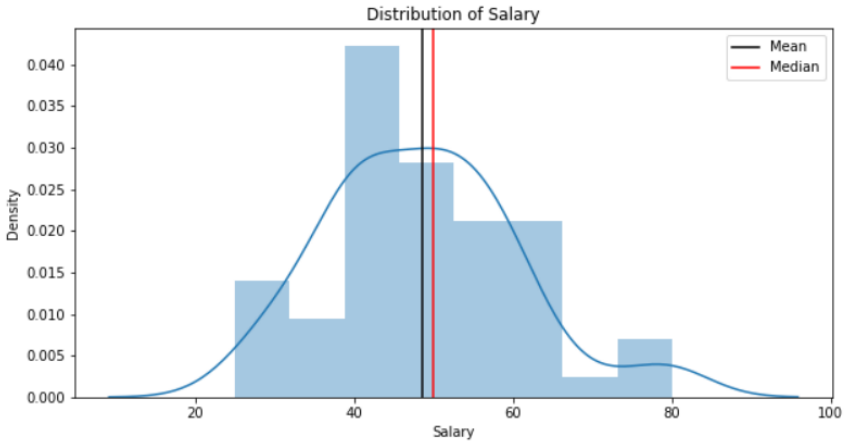

```
ShapiroResult(statistic=0.9685361981391907, pvalue=0.11204058676958084)
```

| | |
|---|---|
| Mean | 3.13 |
| Median (50%) | 3.15 |

The mean and median of a normal distribution are equal. From the above table, we can conclude that 'GPA' follows a normal distribution as the values for mean and median are almost equal.

A Shapiro test was done to further confirm the distribution. The p value of Shapiro test done on GPA is 0.1120. The pvalue is greater than 0.05. Therefore '**GPA' follows a normal distribution.**

2. Salary



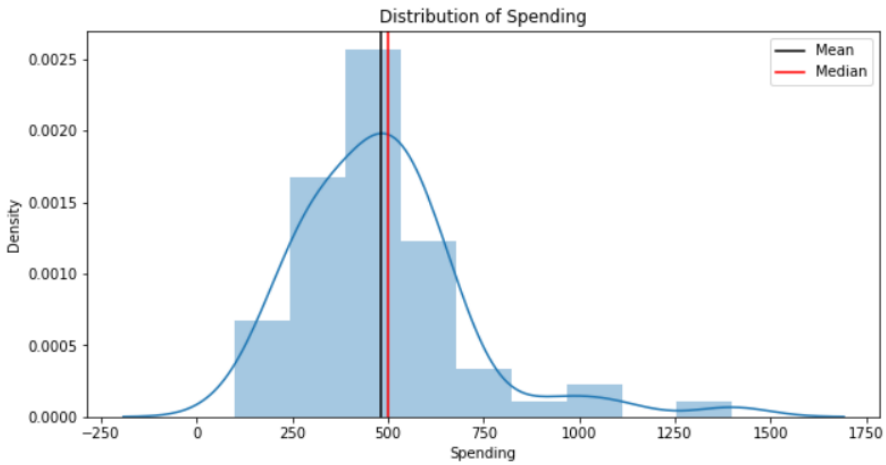

```
ShapiroResult(statistic=0.9565856456756592, pvalue=0.028000956401228905)
```

| Mean | 48.55 |
|---|---|
| Median (50%) | 50 |

The mean and median of a normal distribution are equal. From the above table, we can conclude that 'Salary' does not follow normal distribution as the values for mean and median are different from each other.

A Shapiro test was done to further confirm the distribution. The p value of Shapiro test done on Salary is 0.0280. The pvalue is lesser than 0.05. Therefore **'Salary' does not follow a normal distribution.**

3. Spending




```
ShapiroResult(statistic=0.8777452111244202, pvalue=1.6854661225806922e-05)
```

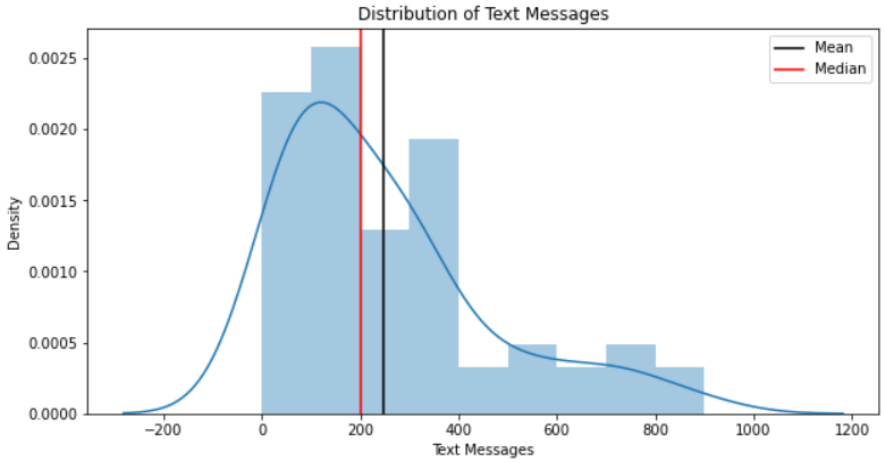
| Mean | 482.02 |
|---|---|

| | |
|---|---|
| Median (50%) | 500 |

The mean and median of a normal distribution are equal. From the above table, we can conclude that 'Spending' does not follow normal distribution as the values for mean and median are different from each other.

A Shapiro test was done to further confirm the distribution. The p value of Shapiro test done on Spending is 0.00001685. The pvalue is lesser than 0.05. Therefore **'Spending' does not follow a normal distribution.**

4. Text Messages




```
ShapiroResult(statistic=0.8594191074371338, pvalue=4.324040673964191e-06)
```

| | |
|---|---|
| Mean | 246.21 |
| Median (50%) | 200 |

The mean and median of a normal distribution are equal. From the above table, we can conclude that 'Text Messages' does not follow normal distribution as the values for mean and median are different from each other.

A Shapiro test was done to further confirm the distribution. The p value of Shapiro test done on Text Messages is 0.000004324. The pvalue is lesser than 0.05. Therefore **'Text Messages' does not follow a normal distribution.**