****

**Problem 1**

1.1.1 Use methods of descriptive statistics to summarize data……………………………………………………………….…………………...…...1

1.1.2 Which Region and which Channel spent the most?.....................................................................................................................1.5

1.1.3 Which Region and which Channel spent the least?......................................................................................................................1.5

1.2. There are 6 different varieties of items that are considered. Describe and comment/explain all the varieties across Region and Channel? Provide a detailed justification for your answer………………………………………………………………………………………….…..3

1.3 On the basis of the descriptive measure of variability, which item shows the most inconsistent behaviour? Which items shows the least inconsistent behaviour?.......................4

1.4 Are there any outliers in the data? Back up your answer with a suitable plot/technique with the help of detailed comments………………………………………………………………3

1.5 On the basis of your analysis, what are your recommendations for the business? How can your analysis help the business to solve its problem? Answer from the business perspective………………………………………………………………………………………….2

**Problem 2**

2.1. For this data, construct the following contingency tables (Keep Gender as row variable)0 2.1.1.Gender & Major…………………………………………………………………………..1 2.1.2.Gender and Grad Intention……………………………………………………………...1

2.1.3. Gender and Employment………………………………………………………………...1

2.1.4. Gender and Computer……………………………………………………………………1

2.2. Assume that the sample is a representative of the population of CMSU. Based on the data, answer the following questions: ................................................................................ 0

2.2.1 What is the probability that a randomly selected CMSU student will be male?........1

2.2.2 What is the probability that a randomly selected CMSU student will be female?.....1 2.3.Assume that the sample is a representative of the population of CMSU. Based on the data, answer the following question …………………………………………………………….0

2.3.1 Find the conditional probability of different majors among the male students in CMSU………………………………………………………………………………………2

2.3.2 Find the conditional probability of different majors among the female students of CMSU………………………………………………………………………………………2

2.4 Assume that the sample is a representative of the population of CMSU. Based on the data, answer the following question: ……………………………………………………………0

2.4.1 Find the probability that a randomly chosen student is a male and intends to graduate…………………………………………………………………………………...2

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

2.5 Assume that the sample is a representative of the population of CMSU. Based on the data, answer the following question: ……………………………………………………………0

2.5.1 Find the probability that a randomly chosen student is a male or has a full-time Employment……………………………………………………………………………2

2.5.2 Find the conditional probability that given a female student is randomly chosen, she is majoring in international business or management……………………… 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 graduate intention and being female are independent events?.............................4

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……0

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

2.7.2 Find conditional probability that a randomly selected male earns 50 or more. Find conditional probability that a randomly selected female earns 50 or more.3

2.8.1 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…………………………………………………………………………..3

2.8.2 Write a note summarizing your conclusions……………………………………………...1

**Problem 3**

3.1 Do you think there is evidence that mean moisture contents in both types of shingles are within the permissible limits? State your conclusions clearly showing all steps……………4

3.2 Do you think that the population means for shingles A and B are equal? Form the hypothesis and conduct the test of the hypothesis. What assumption do you need to check before the test for equality of means is performed?.............................................4

Quality of Business Report………………………………………………………………………..6

Please reflect on all that you learnt and fill this reflection report - https://forms.gle/G8Q6dgVEiuXj5Uoo6................................................................................0

**Problem 1**

* 1. **Use methods of descriptive statistics to summarize data. Which Region and which Channel spent the most? Which Region and which Channel spent the least?**

**Channel**

Hotel 7999569

Retail 6619931

**Region**

Lisbon 2386813

Oporto 1555088

Other 10677599

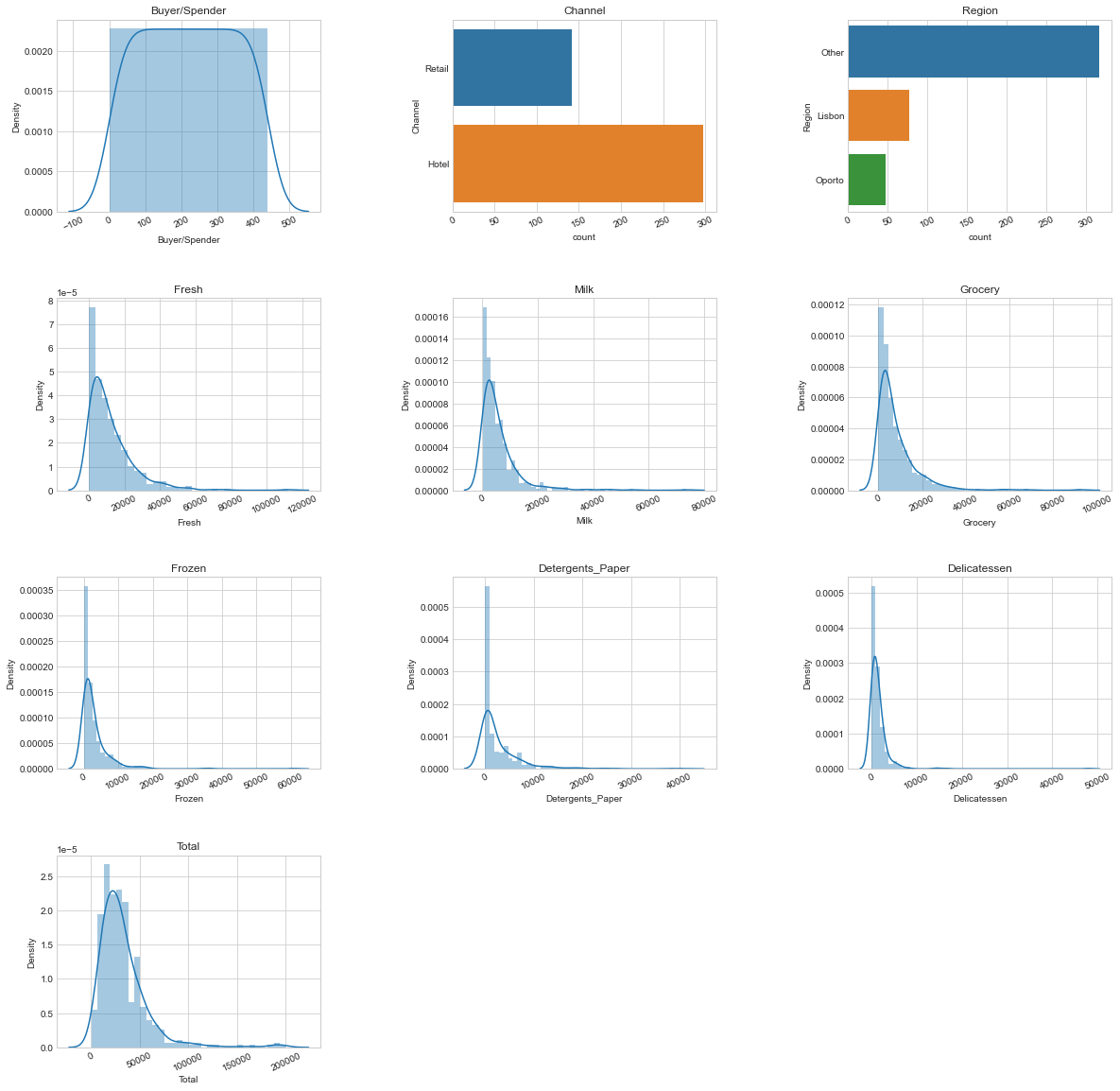
**Which Region and which Channel spent the most?**

**Answer- Other region and Hotel channel spent the most.**

**Which Region and which Channel spent the least?**

**Answer- Oporto region and Retail channel spent the least.**

* 1. **There are 6 different varieties of items that are considered. Describe and comment/explain all the varieties across Region and Channel? Provide a detailed justification for your answer.**

****

**Distribution of groceries has to be equal in both the channels if we look at above snap because in all the snaps it is showing as un-equal distribution so it is suggested to keep it even.**

* 1. **On the basis of a descriptive measure of variability, which item shows the most inconsistent behaviour? Which items show the least inconsistent behaviour?**

Buyer/Spender 127.161315

Fresh 12647.328865

Milk 7380.377175

Grocery 9503.162829

Frozen 4854.673333

Detergents\_Paper 4767.854448

Delicatessen 2820.105937

Total 26356.301730

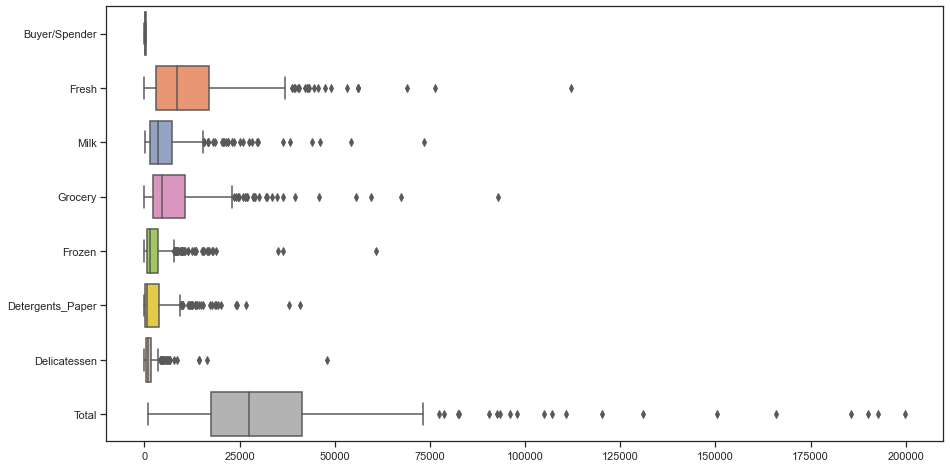
**Which item shows the most inconsistent behaviour?**

**Answer-** Fresh item has highest Standard Deviation So that is Inconsistent.

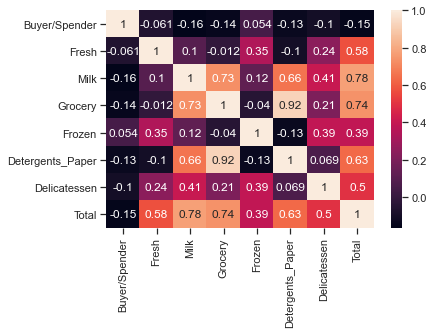
**Which items show the least inconsistent behaviour?**

**Answer-** Delicatessen item have smallest Standard deviation, so that is consistent.

* 1. **Are there any outliers in the data? Back up your answer with a suitable plot/technique with the help of detailed comments.**



* Yes there are outliers in all the items across the product range (Fresh, Milk, Grocery, Frozen, Detergents\_Paper & Delicatessen)
* Outliers are detected but not necessarily removed, it depends on the situation. Here I will assume that the wholesale distributor provided us a dataset with correct data, so I will keep them as it is.
  1. **On the basis of your analysis, what are your recommendations for the business? How can your analysis help the business to solve its problem? Answer from the business perspective**

****

**Answer** –

As per the analysis, I find out that there are inconsistencies in spending of different items (by calculating Coefficient of Variation), which should be minimized. The spending of Hotel and Retail channel are different which should be more or less equal. And also spent should equal for different regions. Need to focus on other items also than “Fresh” and “Grocery”.

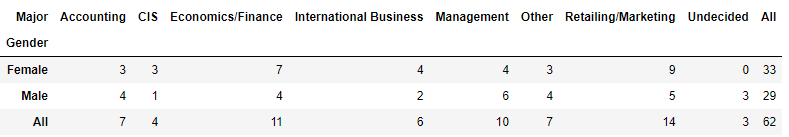
**Problem 2**

**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).**

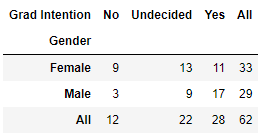
**2.1. For this data, construct the following contingency tables (Keep Gender as row variable).**

****

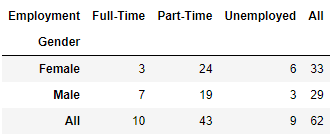
**2.1.1. Gender and Major**

****

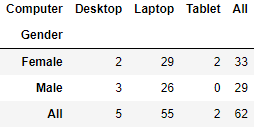
**2.1.2. Gender and Grad Intention**



**2.1.3. Gender and Employment**



**2.1.4. Gender and Computer**



**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?**

**Answer-**

P(male) = 29/62=46.77%.

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

**Answer-**

P(female) = 32/62=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.**

**Answer-**

• Probability of Males opting for Accounting is = 4/29 = 13.79%.

• Probability of Males opting for CIS is = 1/29=3.45%.

• Probability of Males opting for Economics/Finance is = 4/29 = 13.79%.

• Probability of Males opting for International Business is = 2/29= 6.90%.

• Probability of Males opting for Management is =6/29=20.69%.

• Probability of Males opting for other is = 4/29=13.79%.

• Probability of Males opting for Retailing/Marketing is = 5/29= 17.24%.

• Probability of Males opting for Undecided is 3/29=10.34%.

**And from this output we can easily say that most of the males students prefer Management as Majors and CIS is the least preferred one.**

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

**Answer-**

* Probability of Females opting for Accounting is =3/33=9.09%.
* Probability of Females opting for CIS is =3/33=9.09%.
* Probability of Females opting for Economics/Finance is =7/33=21.21%.
* Probability of Females opting for International Business is =4/33=12.12%.
* Probability of Females opting for Management is =4/33=12.12%.
* Probability of Females opting for other is 3/33=9.09%.
* Probability of Females opting for Retailing/Marketing is =9/33=27.27%.
* Probability of Females opting for Undecided is 0.00%.

**And from this output we can easily say that most of the Females students prefer Retailing/Marketing as Majors.**

**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.**

**Answer- The probability that a randomly chosen student is a male and intends to graduate is =17/62=0.274.**

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

**Answer-**

* The probability that a randomly selected student is a female and does NOT have a laptop is =4/62=0.06.

**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?**

**Answer-** The **probability that a randomly chosen student is a male or has full-time employment is 0.518.**

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

**Answer-**

* Female students in sample are 33.
* Female students majoring in International Business or Management are 8.

• The probability that when a female student is randomly chosen, she is majoring in I.B or management 24.242 %

**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?**

**Answer-**

* P(A)=11/28=0.39

• P(B)=11/20=0.55

**Since, P(A) & P(B) are not equal to each other hence these 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?**

**Answer-** The probability that his/her GPA is less than 3 is =17/62=0.27.

**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.**

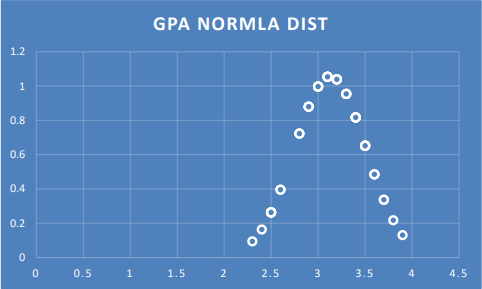
**Answer-**

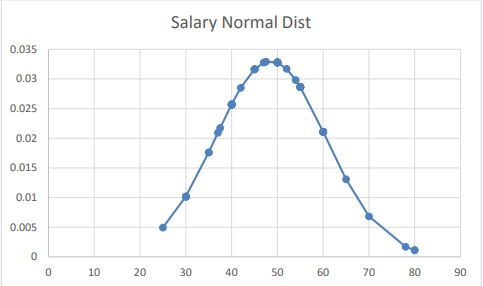
The conditional probability that a randomly selected male earns 50 or more is =14/29=0.482

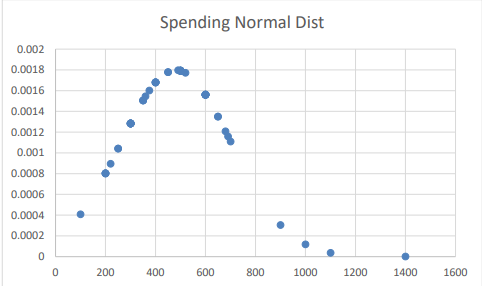
The conditional probability that a randomly selected female earns 50 or more is =18/33=0.545.

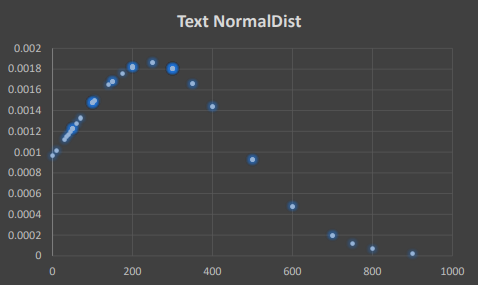
**2.8. Note that there are four numerical (continuous) variables in the data set, GPA, Salary, Spending, and Text Messages. For each of them comment whether they follow a normal distribution.**

* When checked in the excel with the normal distribution curve, I found out that all four variables are normally distributed.





****



**2.8. Write a note summarizing your conclusions.**

* Using contingency tables of Gender and Majors we got the total numbers of males and number of males opting for different majors.
* And from this output we can easily say that most of the males students prefer Management as Majors and CIS is the least preferred one.
* Using contingency tables of Gender and Majors we got the total numbers of females and number of females opting for different majors.
* And from this output we can easily say that most of the females students prefer Retailing/Marketing as Majors.
* When checked in the excel with the normal distribution curve, I found out that all four variables (Text, spending, GPA, Salary) are normally distributed.

**Problem 3**

**An important quality characteristic used by the manufacturers of ABC asphalt shingles is the amount of moisture the shingles contain when they are packaged. Customers may feel that they have purchased a product lacking in quality if they find moisture and wet shingles inside the packaging.   In some cases, excessive moisture can cause the granules attached to the shingles for texture and coloring purposes to fall off the shingles resulting in appearance problems. To monitor the amount of moisture present, the company conducts moisture tests. A shingle is weighed and then dried. The shingle is then reweighed, and based on the amount of moisture taken out of the product, the pounds of moisture per 100 square feet are calculated. The company would like to show that the mean moisture content is less than 0.35 pounds per 100 square feet.**

**The file (**[**A & B shingles**](https://olympus.greatlearning.in/courses/58768/files/3586310/download?verifier=SrmYQApsdg4JJCQaut8LI3F3MHq8RsDSNiw40eCT&wrap=1) **) includes 36 measurements (in pounds per 100 square feet) for A shingles and 31 for B shingles.**

For the A shingles, the null and alternative hypothesis to test whether the population mean moisture content is less than 0.35 pound per 100 square feet is given:

𝐻0≤ 0.35

𝐻1 > 0.35

For the B shingles, the null and alternative hypothesis to test whether the population mean moisture content is less than 0.35 pound per 100 square feet is given:

𝐻0≤ 0.35

𝐻1 > 0.35

**3.1 Do you think there is evidence that means moisture contents in both types of shingles are within the permissible limits? State your conclusions clearly showing all steps.**

**One sample t test**

t statistic: [-1.47350463 nan] p value: [0.07477633 nan]

* In this, p value is 0.07477633 and it is greater than 5% level of significance.

• So the statistical is failing to reject the null hypothesis at 5% level of significance.

• So at 95% confidence level, there is sufficient evidence to prove that mean moisture content is less than 0.35 pound per 100 square feet.

**One sample t test**

t statistic: [-1.4735046253382782 -3.1003313069986995] p value: [0.07477633144907499 0.0 020904774003191826]

Since pvalue < 0.05, reject H0 . There is enough evidence to conclude that the mean moisture content for Sample B shingles is not less than 0.35 pounds per 100 square feet. p-value = 0.00 21.

**3.2 Do you think that the population mean for shingles A and B are equal? Form the hypothesis and conduct the test of the hypothesis. What assumption do you need to check before the test for equality of means is performed?**

t\_statistic=1.29 and pvalue=0.202

As the pvalue > α , do not reject H0; and we can say that population mean for shingles A and B are equal Test Assumptions When running a two-sample t-test, the basic assumptions are the distributions of the two populations are normal, and that the variances of the two distributions are the same.