SMDM PROJECT

ANALYSIS OF WHOLESALE CUSTOMER,SURVEY,A&B SHINGLES

APRIL 24, 2022

S NITIN KUMAR
PGP-DSBA ONLINE

Contents

Problem 1

A wholesale distributor operating in different regions of Portugal has information on annual spending of several items in their stores across different regions and channels. The data consists of 440 large retailers' annual spending on 6 different varieties of products in 3 different regions (Lisbon, Oporto, Other) and across different sales channel (Hotel, Retail).

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?

1.1.1 Use methods of descriptive statistics to summarize data.

Below is the descriptive statistics we of the data

	count	unique	top	freq	mean	std	min	25%	50%	75%	max
Buyer/Spender	440.0	NaN	NaN	NaN	220.5	127.161315	1.0	110.75	220.5	330.25	440.0
Channel	440	2	Hotel	298	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Region	440	3	Other	316	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Fresh	440.0	NaN	NaN	NaN	12000.297727	12647.328865	3.0	3127.75	8504.0	16933.75	112151.0
Milk	440.0	NaN	NaN	NaN	5796.265909	7380.377175	55.0	1533.0	3627.0	7190.25	73498.0
Grocery	440.0	NaN	NaN	NaN	7951.277273	9503.162829	3.0	2153.0	4755.5	10655.75	92780.0
Frozen	440.0	NaN	NaN	NaN	3071.931818	4854.673333	25.0	742.25	1526.0	3554.25	60869.0
Detergents_Paper	440.0	NaN	NaN	NaN	2881.493182	4767.854448	3.0	256.75	816.5	3922.0	40827.0
Delicatessen	440.0	NaN	NaN	NaN	1524.870455	2820.105937	3.0	408.25	965.5	1820.25	47943.0

From the descriptive statistics, we can see that there are 6 unique types of items available in the dataset. Which are distributed in stores across different regions and channels and the annual spending of all items in different region and channels

Fig1-total expender of region and channel

1.1.2 Which Region and which Channel spent the most?

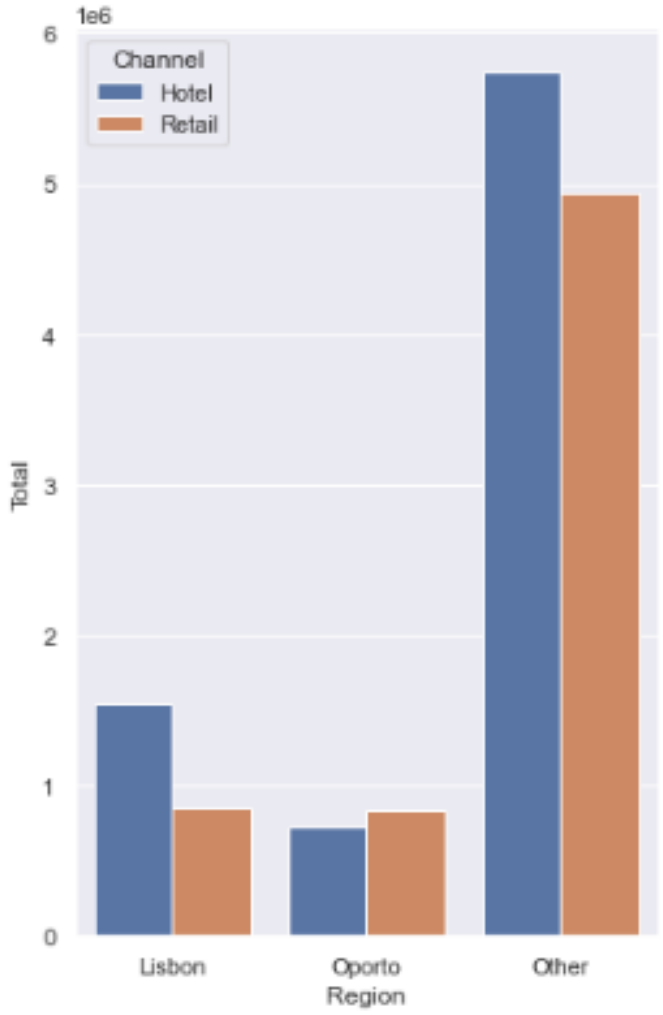
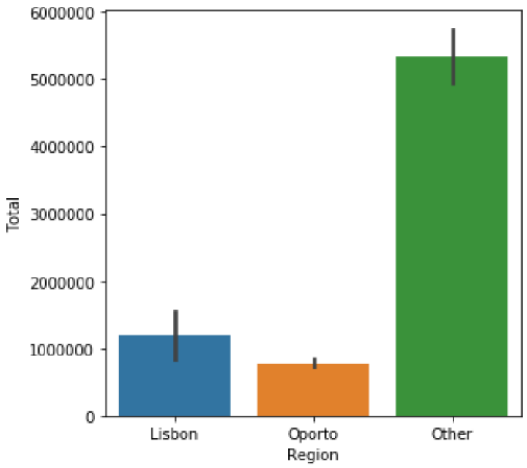


Fig:-1.1

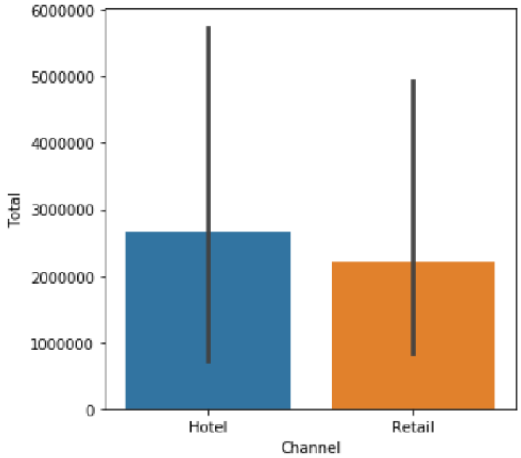
Region	Channel	Buyer/Spender	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicatessen
Lisbon	**Hotel**	14026	761233	228342	237542	184512	56081	70632
	Retail	4069	93600	194112	332495	46514	148055	33695
Oporto	**Hotel**	8988	326215	64519	123074	160861	13516	30965
	Retail	5911	138506	174625	310200	29271	159795	23541
Other	**Hotel**	48020	2928269	735753	820101	771606	165990	320358
	Retail	16006	1032308	1153006	1675150	158886	724420	191752

Fig1.1-total expenditure



From the figure we can see that in region others have spent the most and Oporto have spent the least

1.1.3 Which Region and which Channel spent the least?



From the figure 1.1 and figure we can say that in channel hotel have spent the most and retail have spent the least.

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.

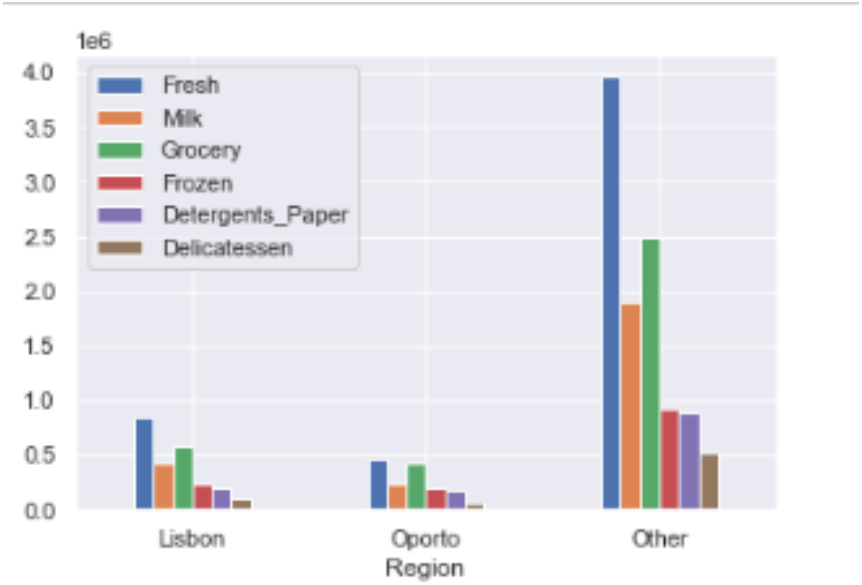


Fig1.2-different available items

There are 6 different varieties of items that are distributed in stores across different regions and channels they are

- Fresh
- Milk
- Grocery
- Frozen
- Detergents Paper
- Delicatessen

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

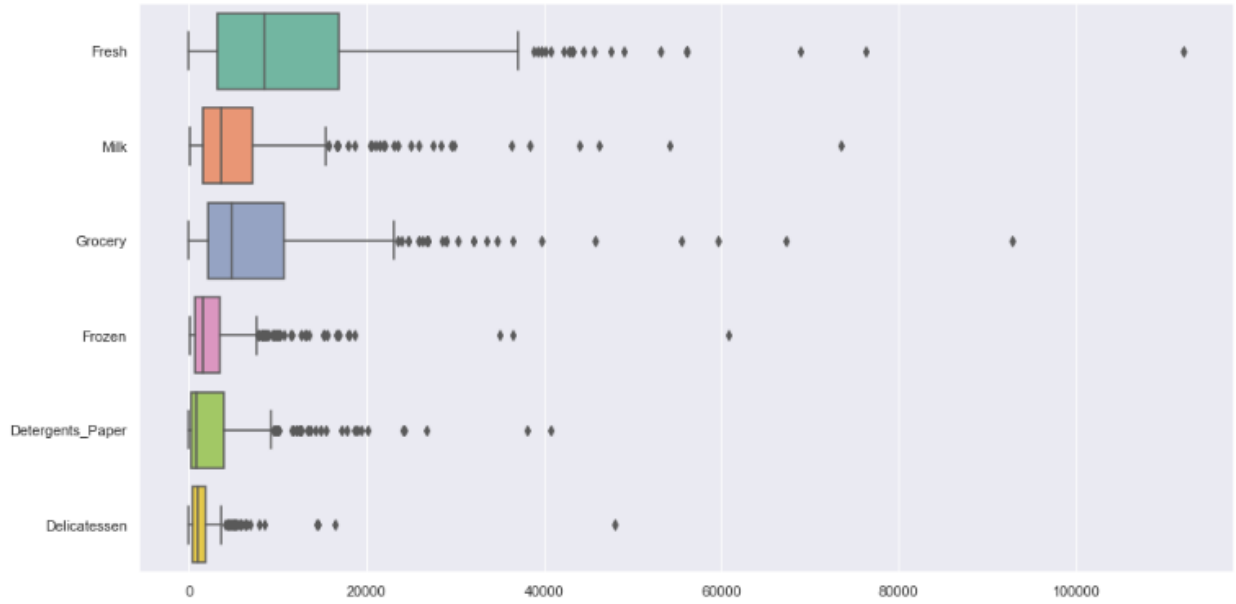
```
Fresh              12647.328865
Milk                7380.377175
Grocery             9503.162829
Frozen              4854.673333
Detergents_Paper    4767.854448
Delicatessen        2820.105937
dtype: float64
```

```
Fresh              1.599549e+08
Milk               5.446997e+07
Grocery            9.031010e+07
Frozen             2.356785e+07
Detergents_Paper   2.273244e+07
Delicatessen       7.952997e+06
dtype: float64
```

Fresh item has highest Standard deviation So that is Inconsistent.

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

1.3 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 the given data set of has an out liner as it is not necessary to remove so we will assume that the given data set is true

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

From the figure we can see that in region others have spent the most and Oporto have spent the least

- Based on the analysis we can say that there are many buyers in other region so we could develop more our retailer base in Lisbon and Orto region
- We can find that fresh items are wildly used by many retailers among all the region
- Since based on the analysis the items like frozen detergents paper and delicatessen are not popular we can maximize the production where demand is more

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

2.1.1. Gender and Major

Major Gender	Accounting	CIS	Economics/Finance	International Business	Management	Other	Retailing/Marketing	Undecided
Female	3	3	7	4	4	3	9	0
Male	4	1	4	2	6	4	5	3

2.1.2. Gender and Grad Intention

Grad Intention	No	Undecided	Yes
Gender			
Female	9	13	11
Male	3	9	17

2.1.3. Gender and Employment

Employment	Full-Time	Part-Time	Unemployed
Gender			
Female	3	24	6
Male	7	19	3

2.1.4. Gender and Computer

Computer	Desktop	Laptop	Tablet
Gender			
Female	2	29	2
Male	3	26	0

2.2. Assume that the sample is a 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?

Probability of male= (total number of male)/(total number of male and female)

Probability of male =29/62= 0.4677

probability that a randomly selected CMSU male student will be 48%

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

Probability of female= (total number of female)/(total number of male and female)

Probability of female =33/62= 0.5322

probability that a randomly selected CMSU female student will be 53%

2.3. Assume that the sample is a 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.

Major Gender	Accounting	CIS	Economics/Finance	International Business	Management	Other	Retailing/Marketing	Undecided
Female	0.090909	0.090909	0.212121	0.121212	0.121212	0.090909	0.272727	0.000000
Male	0.137931	0.034483	0.137931	0.068966	0.206897	0.137931	0.172414	0.103448
All	0.112903	0.064516	0.177419	0.096774	0.161290	0.112903	0.225806	0.048387

The conditional probability of event A given that B has occurred is given by

P(majors/males) = P(majors ∩ males)/ P (males)

conditional probability of different majors among the male students in accounting is 0.137= 13%

conditional probability of different majors among the male students in CIS is 0.034=3%

conditional probability of different majors among the male students in Economics/Finance is 0.13=13%

conditional probability of different majors among the male students in International Busines s is 0.068=6%

conditional probability of different majors among the male students in Management is 0.20 =20%

conditional probability of different majors among the male students in Other is 0.137=13%

conditional probability of different majors among the male students in Retailing/Marketing i s 0.48=48%

conditional probability of different majors among the male students in Undecided is 0.1034= 10%

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

The conditional probability of event A given that B has occurred is given by

P(majors/females) = P(majors ∩ females)/ P (females)

conditional probability of different majors among the female students in accounting is 9.090909090909092=9%

conditional probability of different majors among the female students in CIS is 9.090909090909092=9%

conditional probability of different majors among the female students in Economics/Finance is 21.21212121212121=21%

conditional probability of different majors among the female students in International Business is 12.121212121212121=12%

conditional probability of different majors among the female students in Management is 12.121212121212121=12%

conditional probability of different majors among the female students in Other is 9.090909090909092=9%

conditional probability of different majors among the female students in Retailing/Marketing is 27.27272727272727=27%

conditional probability of different majors among the female students in Undecided is 0.0=0%

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

Probity of the male students intends to graduate=

(total number of male that intends to graduate)/(total number of males=17/29=0.5862

probability That a randomly chosen student is a male and intends to graduate is 58%

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

Probability of the female student that does not have a laptop =(total female student that does not have a laptop)/(total number of female students)

=4/33=0.12112

Probability of the female student that does not have a laptop is 12%

2.5. Assume that the sample is a 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 a full-time employment

probability that a randomly chosen student is a male P(A)=46%

probability that a randomly chosen student has a full time job P(b)=16.13

probability of male having fulltime job P(A&B)=11.28%

P=Probability of male students+ probability of total full time -total probability of male full time =51.61%

probability that a randomly chosen student is a male or has a full-time employment is 51%

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

conditional probability that given a female student in is majoring in international business or management. =(Total number of female students majoring in international business + management.)/(total number of female students)=8/33=0.242

conditional probability that given a female student in is majoring in international business or management =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 graduate intention and being female are independent events?

Grad Intention	No	Yes
Gender		
Female	9	11
Male	3	17

Hear is the 2x2 matrix gender and Intent to Graduate at 2 levels (Yes/No).

P(female n yes)=P(female) *P(yes)

11/40 = 20/40*28/40

0.275=0.35

From the above solved equation, we can say that graduate intention and being female are not independent events

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?

If a student is chosen randomly probability that his/her GPA is less than 3 is 27.41%

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.

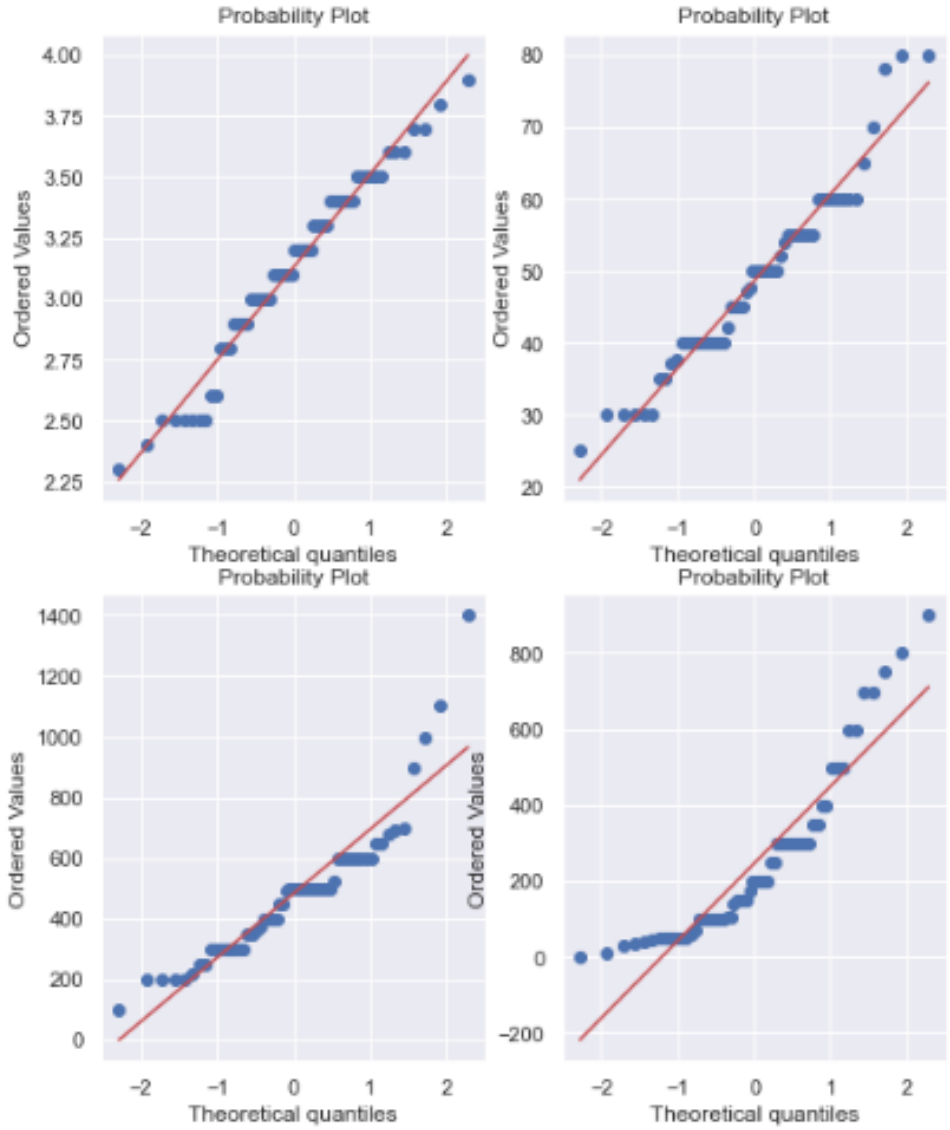
Salary	False	True
Gender		
False	0.454545	0.545455
True	0.517241	0.482759

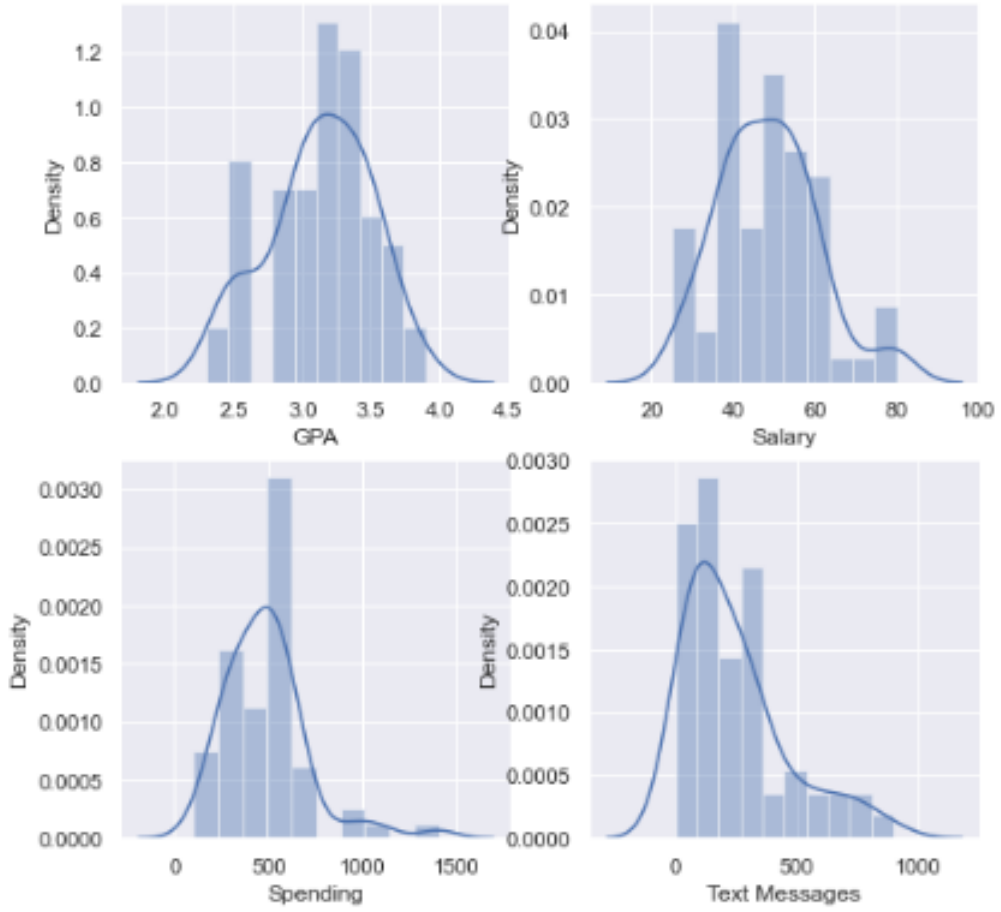
conditional probability that a randomly selected male earns 50 or more is 48.2%

Salary	False	True
Gender		
False	0.517241	0.482759
True	0.454545	0.545455

conditional probability that a randomly selected Female earns 50 or more is 54.5%

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.





The probability plot follows a normal distribution as the points follow the straight line we can say that variables in the data set, GPA, Salary, Spending and Text Messages follow normal distribution

2.8.2 Write a note summarizing your conclusions

In the given data set we have a survey of 62 response from the students. Among which most of the students intends to graduate the retailing and marketing and also from the given data set most of the students are looking for a part time job

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.

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:

H_0 >= 0.35

H_1 < 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:

H_0 >=0.35

H_1 < 0.35

Step 2

Decide the significant limit

Alpha value is not given in the question so we assume that the alpha value in 0.05

Step 3

From the sample A we do not have the population standard deviation so we use tstat test statistics for one sample t-test

Step 4

Calculate the p value and t statistic

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

By evaluation the above formula we get p value=0.0747

Since pvalue > 0.05, do not reject H0 . We conclude thart the moisture content is greater then permissible limit in sample A

Similarly for the B shingles p value is 0.0020

Since p value is lesser then the alpha value we reject ho and conclude that the moisture content of B is lesser then the permissible value

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?

Step1 :
Define null and alternate hypothesis

The null hypothesis states that the mean of shingle A is equal to the mean of shingle b
Ho=uA=uB=0

H1=uA!=uB

Step2
Decide the significance limit
Since alpha is not given we assume that the alpha value is 0.05

Step3
Identify the test static

Given we have 2 samples and we do not know the standard deviation
The sample size is n>30. So we use t distribution and tSTAT for 2 sample test

Two tail test

Step4

Calculating the p value and Test statistic

$$t = (M_1 - M_2)/\sqrt{(s^2_{M_1} + s^2_{M_2})}$$

We get Tstat=1.289
P value=0.201

Step 5

Deciding to reject or accept the null hypo thesis

Since p value is greater than the alpha value, we can say that the mean of shingles A and shingles B are not the same .