
SMDM PROJECT SAMPLE REPORT

DSBA



Disclaimer: This document will act just as a reference for the learners on how the format of business report for submission is expected. The questions in this sample report might differ from the actual questions in the project.

Contents

Problem 1.....	2
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?	2
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.....	2
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?	2
1.4. Are there any outliers in the data? Back up your answer with a suitable plot/technique with the help of detailed comments.....	2
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.	3
Problem 2.....	3
2.1. For this data, construct the following contingency tables (Keep Gender as row variable).....	3
2.1.2. Gender and Grad Intention.....	3
2.1.3. Gender and Employment.....	3
2.1.4. Gender and Computer	3
2.2. Assume that the sample is a representative of the population of CMSU. Based on the data, answer the following questions:.....	3
2.2.1. What is the probability that a randomly selected CMSU student will be male? What is the probability that a randomly selected CMSU student will be female?	3
2.3. Based on the above probabilities, do you think that the column variable in each case is independent of Gender? Justify your comment in each case.	4
2.4. Note that there are three numerical (continuous) variables in the data set, Salary, Spending and Text Messages. For each of them comment whether they follow a normal distribution. Write a note summarizing your conclusions. [Recall that symmetric histogram does not necessarily mean that the underlying distribution is symmetric]	4
Problem 3.....	5
3.1 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?	5
3.2 What assumption about the population distribution is needed in order to conduct the hypothesis tests above?.....	5

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 (Wholesale Customer.csv) 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?

Add code outputs and plots (if any) but avoid adding code. Mention what do you understand of the values are putting here.

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.

Add code outputs and plots but avoid adding code. Write your understanding of the outputs

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?

Add code outputs and plots (if any) but avoid adding code and your understanding of output

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

Add code outputs and plots but avoid adding code

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.

Add your business insights here from above analysis. Summaries what you have done above in a business perspective.

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.csv file).

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

Add the contingency tables for each

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

Calculate the probabilities as asked in python or excel and paste the results here and try to display in mathematical format for each.

Example: $P(\text{Head}) = 1/2 = 0.5$

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

2.2.2. Find the conditional probability of different majors among the male students in CMSU.
Find the conditional probability of different majors among the female students of CMSU.

2.2.3. Find the conditional probability of intent to graduate, given that the student is a male.

Find the conditional probability of intent to graduate, given that the student is a female.

2.2.4. Find the conditional probability of employment status for the male students as well as for the female students.

2.2.5. Find the conditional probability of laptop preference among the male students as well as among the female students.

2.3. Based on the above probabilities, do you think that the column variable in each case is independent of Gender? Justify your comment in each case.

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

[Recall that symmetric histogram does not necessarily mean that the underlying distribution is symmetric]

Add code outputs and plots but avoid adding code.

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 colouring 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 is calculated. The company claims that the mean moisture content cannot be greater than 0.35 pound per 100 square feet. The file (A & B shingles.csv) 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:

$$H_0 \leq 0.35$$

$$H_1 > 0.35$$

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:

$$H_0 \leq 0.35$$

$$H_1 > 0.35$$

3.1 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?

Write the hypothesis as in given format and paste the test results along with the interpretations from the results.

3.2 What assumption about the population distribution is needed in order to conduct the hypothesis tests above?

SAMPLE REPORT DO NOT COPY