# IS Project Report

Prepared By: Parthasarathi Behura

# CONTENTS:

# LIST OF FIGURES:



# LIST OF TABLES:

## IS Project

### Problem 1

### Context

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).

### Objective

Based on the given data, answer the following questions.

1. What is the probability that a randomly selected CMSU student will be male?
2. What is the probability that a randomly selected CMSU student will be female?
3. What is the conditional probability of different majors among male students in CMSU?
4. What is the conditional probability of different majors among the female students of CMSU?
5. What is the probability That a randomly chosen student is a male and intends to graduate?
6. What is the probability that a randomly selected student is a female and does NOT have a laptop?
7. What is the probability that a randomly chosen student is a male or has full-time employment?
8. What is the conditional probability that given a female student is randomly chosen, she is majoring in international business or management?
9. If a student is chosen randomly, what is the probability that his/her GPA is less than 3?
10. What is the conditional probability that a randomly selected male earns 50 or more?
11. What is the conditional probability that a randomly selected female earns 50 or more?
12. Are the continuous variables in the data normally distributed? Write a note summarizing your conclusions.

Note: Assume that the sample is representative of the population of CMSU.

### Data Description

- ID: A unique identifier for each undergraduate student.
- Gender: The gender of the student.
- Age: The age of the student. It is a numeric value representing the student's age in years.
- Class: The student's current academic class or year.
- Major: The student's declared major field of study.
- Grad Intention: The student's intention regarding graduation. It can have categorical values such as "Yes" or "No" indicating whether the student intends to graduate or not.
- GPA: The student's grade point average. It is a numeric value representing the student's academic performance.

- Employment: The employment status of the student. It can have categorical values such as "Full-time," "Part-time," or "Unemployed."
- Salary: The student's monthly salary. It is a numeric value representing the amount in dollars.
- Social Networking: The amount of time the student spends on social networking per day. It is a numeric value representing the time in hours.
- Satisfaction: The student's satisfaction level with their college experience. It is a numeric value representing satisfaction on a scale.
- Spending: The amount of money the student spends per semester. It is a numeric value representing the amount in dollars.
- Computer: The type of device of the student.
- Text Messages: The number of text messages the student sends per day. It is a numeric value.

## Question Solutions:

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

Ans: -

```
The probability that a randomly selected CMSU student will be a male is: 46.774193548387096
```

The probability of a randomly selected CMSU student be a male is 46.77% .

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

Ans: -

```
The probability that a randomly selected CMSU student will be a female is: 53.2258064516129
```

The probability of a randomly selected CMSU student be a female is 53.23% .

### *1.3 What is the conditional probability of different majors among male students in CMSU?*

Ans: -

Conditional Probability of Different Majors among Male Students in CMSU:

| col_0 | count |
|---|---|
| Major | |
| Management | 0.206897 |
| Retailing/Marketing | 0.172414 |
| Accounting | 0.137931 |
| Economics/Finance | 0.137931 |
| Other | 0.137931 |
| Undecided | 0.103448 |
| International Business | 0.068966 |
| CIS | 0.034483 |

Table 1: Conditional Probability of different majors among MALES

From the table 1 we can see that,

Conditional probability of majors among male students are as follows:

- Management: 20.69%
- Retailing/Marketing: 17.24%
- Accounting, Economics/Finance & Others are 13.79% each
- International Business: 6.89%
- CIS: 3.44%

### *1.4 What is the conditional probability of different majors among the female students of CMSU?*

Ans: -

```
Conditional Probability of Different Majors among Female Students in CMSU:
col_0                     count
Major
Retailing/Marketing     0.272727
Economics/Finance       0.212121
International Business  0.121212
Management              0.121212
Accounting              0.090909
CIS                     0.090909
Other                   0.090909
```

Table 2: Conditional Probability of different majors among FEMALES

From the table 2 we can see that,

Conditional probability of majors among female students are as follows:

- Retailing/Marketing: 27.27%
- Economics/Finance: 21.21%
- International Business, Management are 12.12% each
- Accounting, CIS & Other are 9.09% each

### *1.5 What is the probability that a randomly chosen student is a male and intends to graduate?*

Ans: -

```
Probability that a randomly chosen student is a male and intends to graduate: 58.620689655172406
```

We can say that the probability of a randomly chosen male student intends to graduate is 58.62% .

### *1.6 What is the probability that a randomly selected student is a female and does NOT have a laptop?*

Ans: -

```
Probability that a randomly selected student is a female and does NOT have a laptop: 12.121212121212121
```

We can say that the probability of a randomly selected female student doesn't have a laptop is 12.12% .

### *1.7 What is the probability that a randomly chosen student is a male or has full-time employment?*

Ans: -

```
Probability that a randomly chosen student is a male or has full-time employment: 51.61290322580645
```

We can say that the probability of a randomly chosen student being male or having full-time employment: 51.61% .

### *1.8 What is the conditional probability that given a female student is randomly chosen, she is majoring in international business or management?*

Ans: -

```
Conditional probability that given a female student is randomly chosen, she is majoring in internat
ional business or management: 0.24242424242424243
```

We can say that the conditional probability of a randomly chosen female majoring in international business or management is 24.24% .

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

Ans: -

```
Probability that a randomly chosen student has a GPA less than 3: 27.419354838709676
```

We can say that the probability of a randomly chosen student having a GPA less than 3 is 27.42% .

### *1.10 What is the conditional probability that a randomly selected male earns 50 or more?*

Ans: -

```
Conditional probability that a randomly selected male earns 50 or more: 48.275862068965516
```

We can say that the conditional probability of a randomly selected male earning 50 or more is 48.28% .

### *1.11 What is the conditional probability that a randomly selected female earns 50 or more?*

Ans: -

```
Conditional probability that a randomly selected female earns 50 or more: 54.54545454545454
```

We can say that the conditional probability of a randomly selected female earning 50 or more is 54.55% .

### *1.12 Are the continuous variables in the data normally distributed? Write a note summarizing your conclusions.*

Ans: -

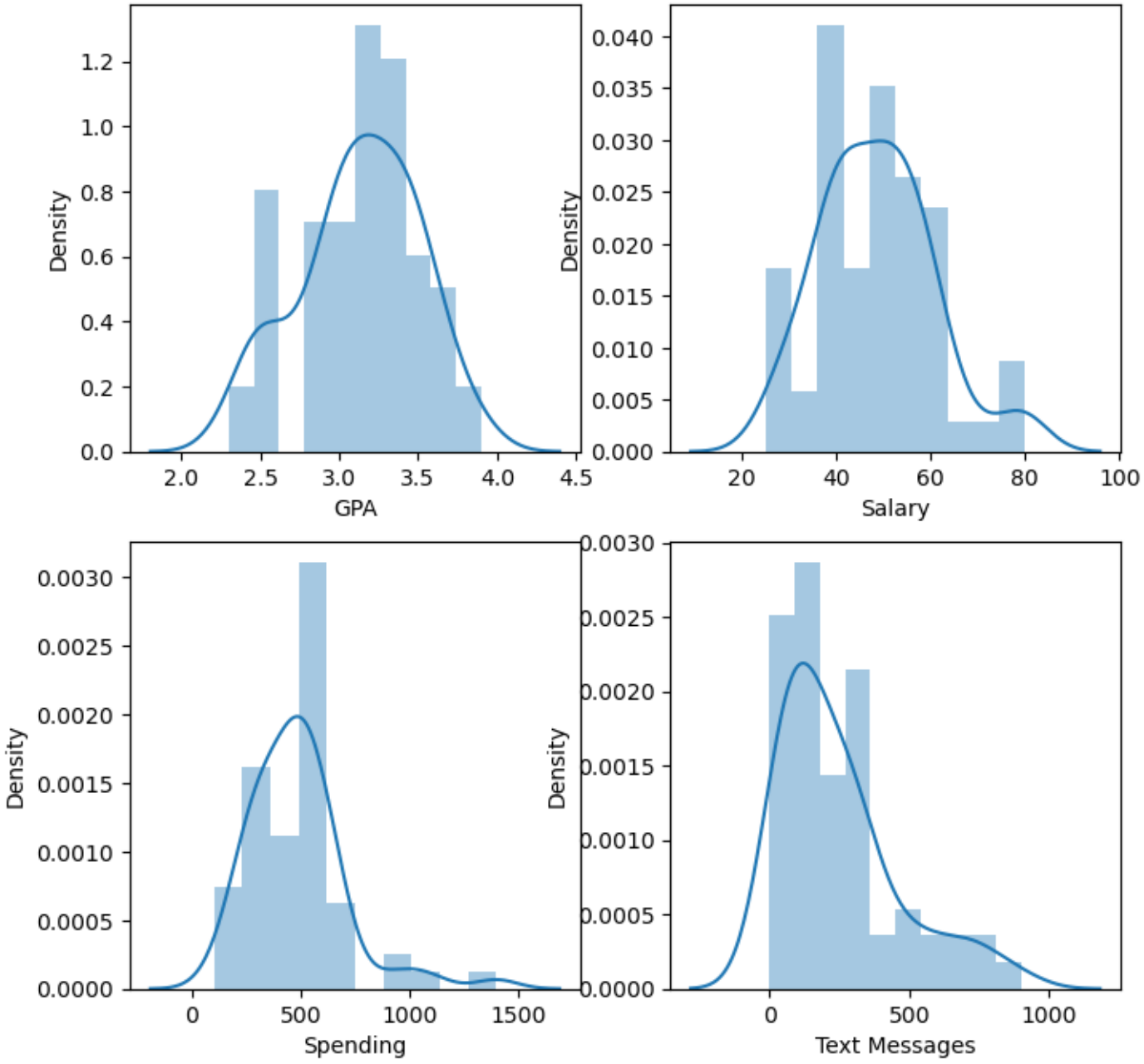


Figure-1 : Checking Skewness

Distribution and Skewness:

1. GPA:
    - Skewness: -0.31 (negative skew)
    - Distribution: Slightly left-skewed but relatively close to normal.
2. Salary:
    - Skewness: 0.53 (positive skew)
    - Distribution: Slightly right skewed but not severely deviating from normality.
3. Spending:
    - Skewness: 1.59 (positive skew)
    - Distribution: Noticeably right skewed, indicating a longer tail on the right.
4. Text Messages:
    - Skewness: 1.30 (positive skew)
    - Distribution: Right skewed with a longer tail on the right.

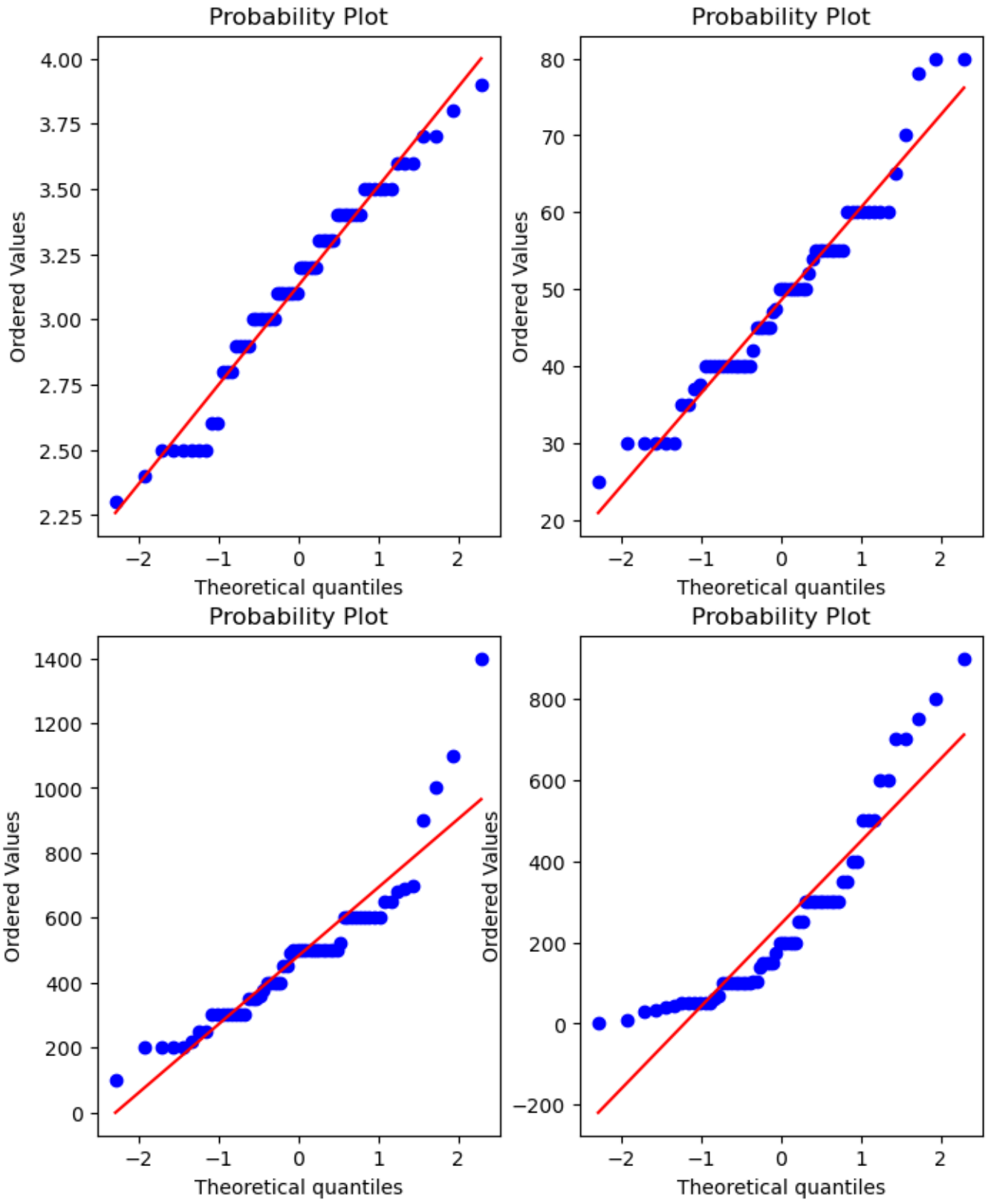


Figure-2 : Probability Plot

Probability Plots:

- The probability plots for GPA and Salary show points relatively close to the diagonal line, suggesting that these variables are closer to normal distribution.
- The probability plots for Spending and Text Messages show points that deviate more from the diagonal line, indicating these variables are less normally distributed.

Summary

- GPA and Salary are approximately normally distributed with slight skewness.
- Spending and Text Messages are not normally distributed, exhibiting significant positive skewness.

## Problem 2

**Context**

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 would like to show that the mean moisture content is less than 0.35 pounds per 100 square feet.

**Objective**

Based on the above context, the manufacturer wants to understand the following:

1. Is there any evidence that the mean moisture content in both types of shingles is within the permissible limits?
2. Is the population mean for shingles A and B are equal?

## Question Solutions: -

### *2.1 Is there any evidence that mean moisture content in both types of shingles are within the permissible limits?*

- State the null and alternate hypotheses - Conduct the hypothesis test and compute the p-value - Write down conclusions from the test results Note: Consider the level of significance as 5%.

Ans: -

**Null and alternate hypotheses:**

- For Shingle A: H0: $\mu A \geq 0.35$ Ha: $\mu A < 0.35$
- For Shingle B: H0: $\mu B \geq 0.35$ Ha: $\mu B < 0.35$

**Hypothesis test and p-values:**

```
tstat -1.4735046253382782
P Value 0.07477633144907513
```

We found that,

- Shingle A: t-statistic = -1.4735, p-value = 0.0748

**Null and alternate hypotheses:**

H0: μA = μB Ha: μA ≠ μB

**Hypothesis test and p-value:**

```
tstat -3.1003313069986995
P Value 0.0020904774003191813
```

We found that,

- Shingle B: t-statistic = -3.1003, p-value = 0.0021

**Conclusion:**

```
one-sample t-test p-value= 0.07477633144907513
We do not have enough evidence to reject the null hypothesis in favour of alternative hypothesis
We conclude that the moisture content is greater than permissible limit in sample A.
one-sample t-test p-value= 0.0020904774003191813
We have enough evidence to reject the null hypothesis in favour of alternative hypothesis
We conclude that the moisture content is less than permissible limit in sample B.
```

- For Shingle A: p-value (0.0748) > 0.05, fail to reject H0. There's not enough evidence to conclude that the mean moisture content is within the permissible limit.
- For Shingle B: p-value (0.0021) < 0.05, reject H0. There's sufficient evidence to conclude that the mean moisture content is within the permissible limit.

## *2.2 Is the population mean for shingles A and B are equal?*

- State the null and alternate hypotheses - Conduct the hypothesis test and compute the p-value - Write down conclusions from the test results Note: Consider the level of significance as 5%.

Ans: -

**Null and alternate hypotheses:**

H0: μA = μB Ha: μA ≠ μBNull and alternate hypotheses: H0: μA = μB Ha: μA ≠ μB

We have two samples and we do not know the population standard deviation. Sample sizes for both samples are not the same. The sample size is , n > 30. So we use the t distribution and the $tSTAT$ test statistic for two sample test.

**Hypothesis test and p-value:**

```
tstat 1.289628271966112
P Value 0.2017496571835328
```

We found that,

t-statistic = 1.2896, p-value = 0.2017

**Conclusion:**

```
two-sample t-test p-value= 0.2017496571835328
We do not have enough evidence to reject the null hypothesis in favour of alternative hypothesis
We conclude that mean for shingles A and singles B are not the same
```

- p-value (0.2017) > 0.05, fail to reject H0. There's not enough evidence to conclude that the population means for shingles A and B are different.

These results suggest that while Shingle B meets the moisture content requirement, Shingle A may not. However, there's no significant difference between the mean moisture contents of the two types of shingles.

## Problem 3

**Context**

Salary is hypothesized to depend on educational qualification and occupation. To understand the dependency, the salaries of 40 individuals are collected and each person's educational qualification and occupation are noted. Educational qualification is at three levels, High school graduate, Bachelor's, and Doctorate. Occupation is at four levels, Administrative and clerical, Sales, Professional or specialty, and Executive or managerial. A different number of observations are in each level of education–occupation combination.

**Objective**

Based on the above context, we want to understand the following:

1. Is there any significant difference in salaries among different levels of education?
2. Is there any significant difference in salaries among different levels of different occupations?
3. Is there a significant interaction between Education and Occupation on Salary?

Use the relevant statistical tests to answer the above questions and state your conclusions along with all necessary steps.

**Note**: Assume that the data follows a normal distribution. In reality, the normality assumption may not always hold if the sample size is small.

## *3.1 Is there any significant difference in salaries among different levels of education?*

- State the null and alternate hypotheses - Check the assumptions of the hypothesis test. - Conduct the hypothesis test and compute the p-value - Write down conclusions from the test results Note: Consider the level of significance as 5%.

Ans: -

**State the null and alternate hypotheses:**

- **Null Hypothesis (H0)**: There is no significant difference in salaries among different levels of education.
- **Alternate Hypothesis (H1)**: There is a significant difference in salaries among different levels of education.

**Assumptions:**

1. **Independence**: The samples are independent of each other.
2. **Normality**: The salary distribution within each education level is approximately normal.
3. **Homogeneity of Variances**: The variance of salaries is similar across the different education levels.

**Conducting the hypothesis test: ANOVA**

```
Fail to reject the null hypothesis. There is no significant difference in salaries based on education.
```

**Conclusion:**

As there is no significance difference in salaries based on education, the test fails to reject the null hypothesis.

## *3.2 Is there any significant difference in salaries among different levels of different occupations?*

- State the null and alternate hypotheses - Check the assumptions of the hypothesis test. - Conduct the hypothesis test and compute the p-value - Write down conclusions from the test results Note: Consider the level of significance as 5%.

Ans: -

**State the null and alternate hypotheses:**

- **Null Hypothesis (H0)**: There is no significant difference in salaries among different occupations.
- **Alternate Hypothesis (H1)**: There is a significant difference in salaries among different occupations.

**Check the assumptions of the hypothesis test:**

1. Independence: The samples are independent of each other.
2. Normality: The salary distribution within each education level is approximately normal.
3. Homogeneity of Variances: The variance of salaries is similar across the different education levels.

```
Fail to reject the null hypothesis. There is no significant difference in salaries based on occupation.
```

**Conclusion:**

As there is no significance difference in salaries based on occupation, the test fails to reject the null hypothesis.

## *3.3 Is there a significant interaction between Education and Occupation on Salary?*

- State the null and alternate hypotheses - Check the assumptions of the hypothesis test. - Create Interaction Plot and draw inferences - Conduct the hypothesis test and compute the p-value - Write down conclusions from the test results Note: Consider the level of significance as 5%.

Ans: -

**State the null and alternate hypotheses:**

- **Null Hypothesis (H0)**: There is no significant interaction between education and occupation on salary.
- **Alternate Hypothesis (H1)**: There is a significant interaction between education and occupation on salary.

**Check the assumptions of the hypothesis test:**

Ensure normality and homogeneity of variances.

```
Reject the null hypothesis. There is a significant interaction between Education and Occupation on Salary.
```

**Conclusion:**

Fail to Reject the Null Hypothesis: As the p-value is greater than or equal to 0.05, we fail to reject the null hypothesis and conclude that there is no significant interaction between education and occupation on salary.

- If a significant interaction is found, it means that the effect of education on salary depends on the occupation and vice versa. This implies a complex relationship where both factors together influence salary in a way that is not predictable by either factor alone.
- Human resource policies, salary structures, and career development programs should consider the interaction between education and occupation to ensure fair and effective compensation strategies.
- The results indicate that there is a significant interaction between education and occupation on salary. Therefore, both factors together significantly influence salary, and the effect of one factor depends on the level of the other.
- To design effective compensation policies, it is important to consider this interaction, ensuring that both education and occupation are factored into salary decisions. This approach will help in creating equitable and motivating compensation structures that align with the diverse qualifications and roles within the organization.

END