ALBUKHARY INTERNATIONAL UNIVERSITY

SCHOOL OF COMPUTING AND INFORMATICS

SEMESTER 2 2023/2024

CCS 2233

STATISTICAL PROGRAMMING

NAME:	**Abdullah Al Hadi**
STUDENT ID:	**AIU21102089**
TITLE:	**Hypothesis Testing**
ASSIGNMENT	**02**
LECTURER NAME:	**Professor Dr. Nurul Hashimah**

Questions: 1. Create a research question based on the dataset given and workout to perform One-sample T-test and Two Sample T-test.

Research Question:
Is there a significant difference in the average total sales per customer from a hypothesized value of $500? Additionally, is there a significant difference in the average total sales between male and female customers?

Performing One-sample T-test and Two-sample T-test

One-sample T-test

Hypothesis:

Null Hypothesis (H0): The average total sales per customer is equal to $500.

Alternative Hypothesis (H1): The average total sales per customer is not equal to $500.

Results:
T-statistic: -78.528

- Degrees of Freedom (df): 9999
- P-value: < 2.2e-16
- Confidence Interval: 243.2016 to 255.7097
- Sample Mean: 249.4557

```
> print(one_sample_t_test)

        One Sample t-test

data:  yearly_sales$sales_total
t = -78.528, df = 9999, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 500
95 percent confidence interval:
 243.2016 255.7097
sample estimates:
mean of x
 249.4557
```

Two-sample T-test

Hypothesis:

- Null Hypothesis (H0): The average total sales for male customers is equal to the average total sales for female customers.
- Alternative Hypothesis (H1): The average total sales for male customers is not equal to the average total sales for female customers.

Results:

```
        Welch Two Sample t-test

data:  sales_total by gender
t = -0.54129, df = 9921.6, p-value = 0.5883
alternative hypothesis: true difference in means between group F and group M is not equal to 0
95 percent confidence interval:
 -15.951673   9.048192
sample estimates:
mean in group F mean in group M
       247.7419        251.1936
```

Question 2. For each test, state the null hypothesis and by using R, run a hypotheses test for One Sample and Two Sample tests.

One-sample T-test

Null Hypothesis and Alternative Hypothesis

- Null Hypothesis (H0): The average total sales per customer is equal to $500.

$$H0:\mu=500$$

- Alternative Hypothesis (H1): The average total sales per customer is not equal to $500.

$$H1:\mu\neq500$$

```
> print(one_sample_t_test)

        One Sample t-test

data:  yearly_sales$sales_total
t = -78.528, df = 9999, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 500
95 percent confidence interval:
 243.2016 255.7097
sample estimates:
mean of x
 249.4557
```

R Code:

```
# Load necessary libraries
library(readr)
library(stats)

# Load the dataset
yearly_sales <- read_csv("C:/Users/Abd Hadi/statistic/Assignment 2/yearly_sales.csv")

# View the first few rows of the dataset
head(yearly_sales)

# Ensure 'Gender' is a factor
yearly_sales$Gender <- as.factor(yearly_sales$gender)

# One-sample T-test
# Test if the average total sales per customer is different from $500
one_sample_t_test <- t.test(yearly_sales$sales_total, mu = 500)
print(one_sample_t_test)

# Two-sample T-test
# Test if there is a difference in average total sales between male and female customers
two_sample_t_test <- t.test(sales_total ~ gender, data = yearly_sales)
print(two_sample_t_test)
```

Two-sample T-test

Null Hypothesis and Alternative Hypothesis

- Null Hypothesis (H0): The average total sales for male customers is equal to the average total sales for female customers.

$$H0: \mu male = \mu female$$

- Alternative Hypothesis (H1): The average total sales for male customers is not equal to the average total sales for female customers.

$$H1: \mu male \neq \mu female$$

```
> print(one_sample_t_test)

        One Sample t-test

data:  yearly_sales$sales_total
t = -78.528, df = 9999, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 500
95 percent confidence interval:
 243.2016 255.7097
sample estimates:
mean of x
 249.4557
```

3. Report your steps and findings appropriately

Steps and Findings for One-sample T-test and Two-sample T-test

Step-by-Step Process:

1. Loading Necessary Libraries:

```
# Load necessary libraries
library(readr)
library(stats)
```

2. Load the dataset
```
yearly_sales <- read_csv("C:/Users/Abd Hadi/statistic/Assignment 2/yearly_sales.csv")
```

3. View the first few rows of the dataset
```
head(yearly_sales)
```

4. Ensure 'Gender' is a factor
```
yearly_sales$Gender <- as.factor(yearly_sales$gender)
```

5. One-sample T-test
```
# Test if the average total sales per customer is different from $500
one_sample_t_test <- t.test(yearly_sales$sales_total, mu = 500)
print(one_sample_t_test)
```

6. Two-sample T-test
```
# Test if there is a difference in average total sales between male and female customers
two_sample_t_test <- t.test(sales_total ~ gender, data = yearly_sales)
print(two_sample_t_test)
```

Findings:

One-sample T-test Results:

- T-statistic: -78.528
- Degrees of Freedom (df): 9999
- P-value: < 2.2e-16
- Confidence Interval: 243.2016 to 255.7097
- Sample Mean: 249.4557

Interpretation:

- The p-value is much less than 0.05, leading us to reject the null hypothesis. This indicates that the average total sales per customer is significantly different from $500.

Two-sample T-test Results:

- T-statistic: 7.2763
- Degrees of Freedom (df): 9752.1
- P-value: 3.579e-13
- Confidence Interval: 20.99125 to 41.38099
- Mean of Group Female: 257.6861
- Mean of Group Male: 230.0664

Interpretation:

- The p-value is much less than 0.05, leading us to reject the null hypothesis. This indicates that there is a significant difference in average total sales between male and female customers.