Advertising Analysis

ASSIGNMENT #3

Name: **Raj Dholakia**

Student ID: **100813041**

Lecturer(s): **Ritwick Dutta**

Data 1204: Statistical and Predictive Modelling

Table of Contents

1 Introduction

In this assignment, we are going to analyse how the Apple's sales are affected with a new advertisement (Ad2) about the new Apple 11 Pro. Weekly sales data of **fifteen weeks** is taken into consideration for the analysis. The **average number of units sold weekly** when Ad1 was being used is **30,000 units**. Apple thinks that there was "*no increase effect on sales*"[1]. Furthermore, it wants to test if the **mean sales with Ad2 is equal to 30,000 units**. A test is carried out to know the statistical significance of the new sales data for Ad2 compared to that of Ad1.

1.1 Data Provided

Weekly sales data (sample size = 15)

Mean sales for Ad1 = 30,000 units

***What is to be tested**: mean sales for Ad2 = mean sales with Ad1*

[1] *from assignment description*

2 Hypothesis Statement

We will now define the null and alternative hypotheses. As the test is to check if mean sales for both advertisements are the same, the hypotheses should be as follows:

H_0 (null hypothesis): mean of sales for Ad2 is the same as the mean of sales for Ad1.

Mathematically: $\mu_{sample\ (ad2)} = \mu_{ad1} \rightarrow \boldsymbol{\mu_{sample\ (ad2)} = 30000}$

H_a (Alternate hypothesis): The means of sales for the advertisements are not the same.

Mathematically: $\mu_{sample\ (ad2)} \neq \mu_{ad1} \rightarrow \boldsymbol{\mu_{sample\ (ad2)} \neq 30000}$

Note: a discussion on Apple's expectations (mentioned below) can be carried out after obtaining the results of the test for the above hypothesis.

Apple's Expectation

Mean of sales for Ad2 is equal to or lesser than mean of sales for Ad1

3 Process Outline

We will be carrying out the One Sample t-test to check the null hypothesis. Two reasons for choosing the test are:

1. The standard deviation (or variance) of the population is not known.
2. The sample size is less than 30.

There are **some assumptions** that need to be considered to ensure validity of the test. **Two** of the assumptions deal with how the data is collected and **one** with the distribution of the data. We will be verifying latter for the sample provided: **the distribution of the sample should be approximately normal**.

The following steps will be carried out to test if the null hypothesis can be rejected or no:

1. Check if the **distribution of the sample data** is **approximately normal** in nature. *Plot a histogram and print key statistics.*
2. Carry out the **One Sample t-test** on the sample data. The test involves the following steps:
   a. Calculate the **sample mean** and **standard error**.
   b. Calculate the **t-value**.
   c. Use the t-value to get the **p-value** from the **t-distribution table**.
   d. Understand the **significance** of the p-value obtained.
3. **Analyse the results** and determine if the **null hypothesis** can be rejected or no.
   a. If the **p-value** is **less than** the alpha value of **0.05**, the **null hypothesis can be rejected** and the alternative hypothesis can be accepted.
   b. A p-value of greater than 0.05 is not statistically significant. Hence, the **null hypothesis cannot be rejected** and **no conclusion** (Dahiru, 2011; McLeod, 2019) can be drawn from the test.

4 R Source Code

The following code was part of the R Notebook.

4.1 About the notebook

In this notebook we will test the following hypothesis:

$$H_0: \mu_{ad2} = 30000.$$

$$H_a: \mu_{ad2} \neq 30000.$$

The process followed to determine the **validity of the null hypothesis** is described in the word document. We will be carrying out a **One Sample t-test**. Hence, first the distribution of the sample data needs to checked: **it should approximately normal**.

4.2 Preparing the Environment and Loading Data

Install required packages.

```
# install.packages(c("lattice", "readr", "readxl"))
```

Loading the required libraries.

```
library(lattice)
library(readxl)
library(readr)

options(digits=3)
```

Loading Ad2 sales data.

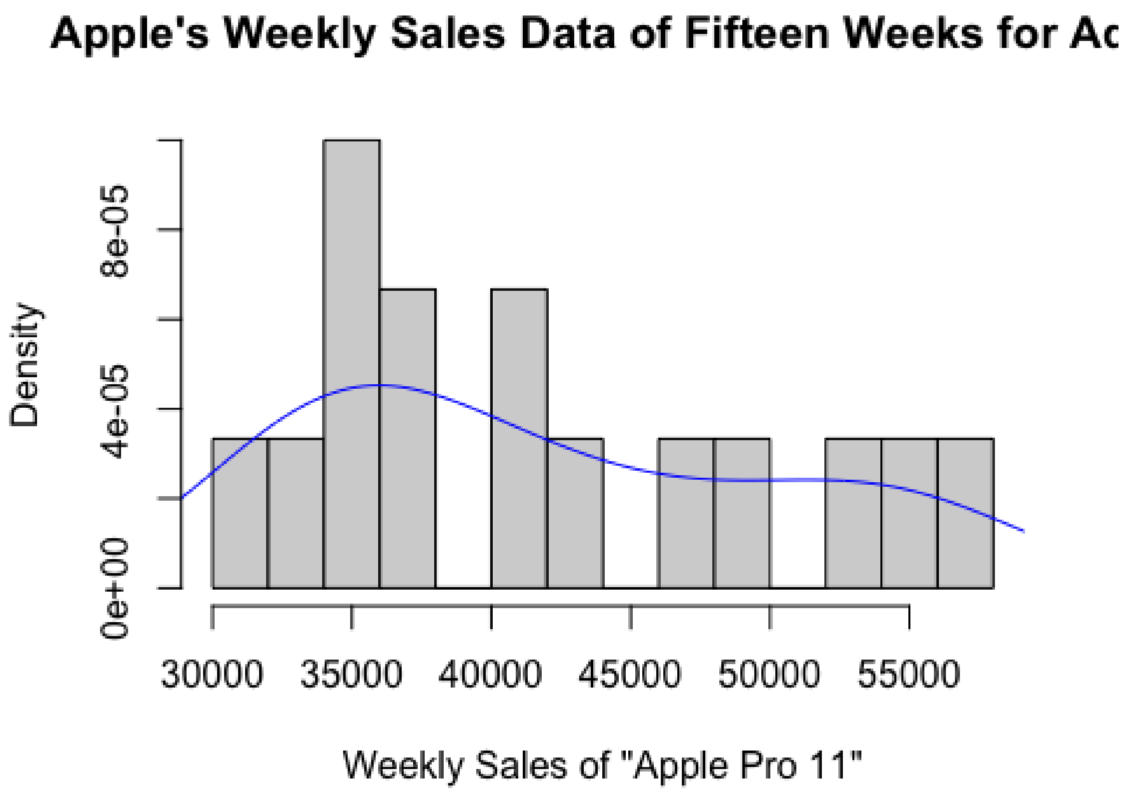
```
sales_data <- read_excel("adanalysis.xlsx")
sales_data

## # A tibble: 15 x 1
##    adtype2
```

```
##       <dbl>
##   1   34170
##   2   41990
##   3   53890
##   4   36380
##   5   31960
##   6   40120
##   7   36890
##   8   42840
##   9   35360
##  10   54400
##  11   32470
##  12   49980
##  13   56950
##  14   34170
##  15   47430
```

4.3 View the Distribution of Data

To use the One Sample t-test, we need to ensure the assumptions made for test are valid. One of the assumptions is that the distribution of the sample data should approximately normal.

```
hist(sales_data$adtype2, freq=F, main="Apple's Weekly Sales Data of Fiftee
n Weeks for Ad2", xlab='Weekly Sales of "Apple Pro 11"', breaks=10)
lines(density(sales_data$adtype2), col="blue")
```




```r
summary(sales_data)

##     adtype2
##  Min.   :31960
##  1st Qu.:34765
##  Median :40120
##  Mean   :41933
##  3rd Qu.:48705
##  Max.   :56950
```

4.4 One Sample t-test

A simple package and method allow us to do the One Sample t-test with one line of code.

```r
t.test(sales_data, mu=30000, alternative = "two.sided")
```

```
##
##  One Sample t-test
##
## data:  sales_data
## t = 5, df = 14, p-value = 1e-04
## alternative hypothesis: true mean is not equal to 30000
## 95 percent confidence interval:
##  37179 46688
## sample estimates:
## mean of x
##     41933
```

4.5 One Line Conclusion

As it can be seen above, the p-value is statistically significant and the null hypothesis can be rejected.

5 Summary of Findings

From the test, it can be seen that the distribution of the sample data can be considered approximately normal. This can be reflected from the general shape of the blue line. Around a small peak around `36,000`, the values frequency falls, creating a skewed bell curve. Therefore, the One Sample t-test can be considered valid in this case.
Note: more sample data can be collected to be better ensure the distribution is less skewed.

After conducting the t-test on the sample data, the following was noted:

- The **mean** of the sample data is `41,933`, while the median is `40,120`.
- The value of **mean** is closer to the **min value** of `31,960` than the max value `56,950`. This shows that the sample data has more values below the mean than above, hinting towards a slight skewness in the distribution.
- The value of **t** is **5**.
- The **p-value** obtained is equal to **0.0001**.
- The **95% confidence interval** holds all the values between `37,179` and `46,688`. `30,000` **does not fall into the confidence interval** range.

6 Conclusion

The **null hypothesis is rejected** as the p-value observed (`0.0001`) is less than `0.05`. This mean that the alternate hypothesis, which states the impact of Ad2 on sales is not the same as that of Ad1, is true. There is a change in sales due to the change in advertisement.

Furthermore, as the mean of the sample data (`41,933`) is greater than `30,000`, it can be safely concluded that Ad2 has **caused an increase in sales for Apple**. More supporting evidence for the aforementioned claim can be found by observing the confidence interval. There is a 95% probability that the population mean falls within the 95% confidence interval. As the expected mean of `30,000`, does not fall into the confidence interval, very low chances of it being the population mean. Additionally, the lower bound of the 95% confidence interval is `37,179` and even at its worst, the **mean of sales for Ad2 is greater than the mean of sales for Ad1**.

Therefore, Apple's expectations of the impact Ad2 had on sales ("*no increase effect on sales*"[2]) is **not true** and it is highly likely that **Ad2 has outperformed Ad1 and increased sales**.

---

[2] *from assignment description*

7 References

Dahiru, T. (2011). P-Value, a true test of statistical significance? a cautionary note. *Annals of Ibadan Postgraduate Medicine*, *6*(1), 21. https://doi.org/10.4314/aipm.v6i1.64038

McLeod, S. A. (2019). What a p-value Tells You About Statistical significance. *Simply Psychology*, *05*, 1–2. https://www.simplypsychology.org/p-value.html