**DATA 1204 Statistical and Predictive Modeling for Analytics DATA 1204**

**Assignment 3-Advertising Analysis**

**Srilasya Garigipaty Student #100822953**

**1.State the Hypothesis Statement that helps solve the research question:**

**For the One Sample T-test the hypothesis statement is :**

H₀: μs = 30000(null hypothesis)

Ha: μs >30000 (alternative hypothesis)

The **null hypothesis** indicates that the new Ad2 Apple has been trying for the purpose of increasing the sales in the new iPhone 11 Pro is the same as average sales of 30,000 units per week from their previous Ad1.

The **alternative hypothesis** indicates that the new Ad2 Apple has been trying for the purpose of increasing the sales in the new iPhone 11 Pro has increased from the average sales of 30,000 units per week from Apple’s previous Ad1.

**2.Provide a written *step-by-step outline* on how you would prove (or dis-prove) your hypothesis statements that you developed above.**

For this advertising analysis, a **one sample t-test** is performed. A **t*-*test** is a tool for evaluating the means of one or two populations using hypothesis testing. T-tests are a way of testing a hypothesis when population variance is not known, and sample size *n* is less than 30.

A **one-sample t-test** may be used to evaluate whether a single group differs from a known value.

An **independent two-sample t-test** is used to determined whether two groups differ from each other.

A **paired, or dependent samples t-test** is used to determine whether there is a significant difference in paired measurements.

In this Advertising Analysis situation, a **one sample t-test** was chosen because a t-test is being done to determine if a single group sample of Ad2 differs from a known value of the mean of 30000 units per week of sales from Ad1.

Assumptions of the test (data should meet these requirements for the test to be valid):

• Data is independent.

• Data is collected randomly.

• The data is approximately normally distributed

To evaluate the statistical significance of the t-test, **p-value** must be calculated.

The **p-value** ranges from 0 to 1, and is interpreted as follow:

• A p-value lower than 0.05 - strongly confident to reject the null hypothesis

• A p-value higher than 0.05 -do not have enough evidence to reject the null hypothesis

**Steps in Conducting a One Sample T- Test:**

**Step 1:** Write null hypothesis statement .

**Step 2:** Write your [alternate hypothesis](https://www.statisticshowto.com/what-is-an-alternate-hypothesis/). This is the hypothesis being tested.

**Step 3**:  Indicate the alpha level confidence interval and degrees of freedom. **The confidence interval** represents the amount of uncertainty expected while determining the sample population estimate of the sales or mean of a true population of total sales for Ad1. The use of confidence intervals makes the estimation of the sample population of Ad2 estimate more manageable. The degrees of freedom is the sample size of 15 minus the number of parameters needed to calculate for analysis. Degrees of freedom for the advertising analysis is 15-1=14.

**Step 4:** Run the t-test on R program.

**Step 5**: Compare the results of t-test to the confidence interval and make a conclusion. The **P-value** is themeasure that quantifies the strength of the evidence made about a population parameter. It is the strength of the evidence against null hypothesis that the mean of sales per week for Ad2 is 30000 units per week, and in favor of the alternative hypothesis that the mean of sales for Ad2 is greater than 30000 units per week and there is an increase effect on the sales.

**3.Conduct the analysis you outlined in #2 in R**

**Input:**

#Load data

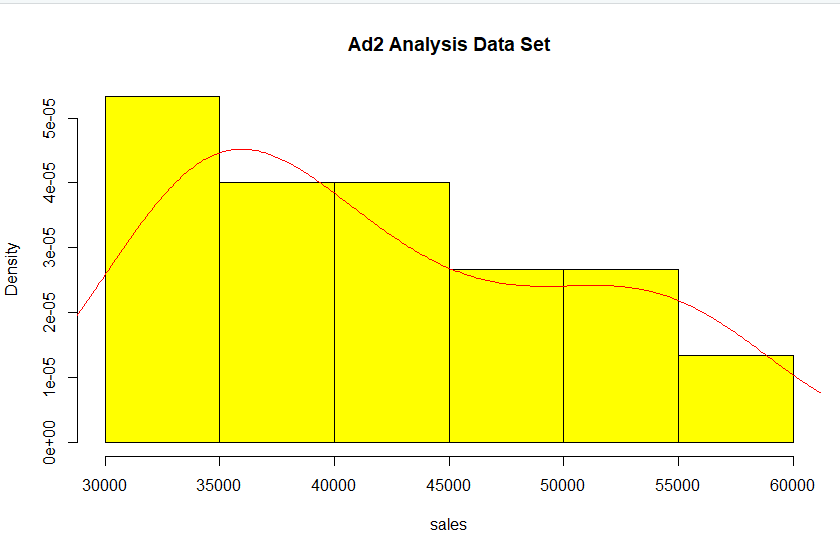
library(ggplot2)

#Histogram

hist(sample, freq=F,col='yellow',main="Ad2 Analysis Data Set", xlab='sales')

lines(density(sample), col="red")

**Output:**



**Input:**

#Create Dataset

sample = adanalysis$adtype2

#Summary Statistics

summary(sample)

**Output:**



**Input:**

#Hypothesis

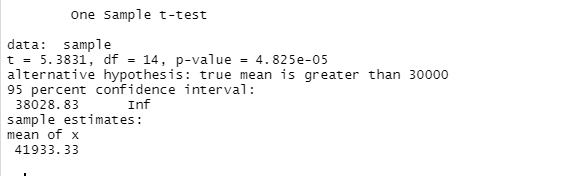
#H₀: μs = 30000(null hypothesis)

#Ha: μs >30000 (alternative hypothesis)

#One Sample t-test - Greater than 30000

t.test(sample, mu=30000, alternative = "greater")

**Output:**



**4.Summary of findings and conclusion**

Upon conducting the analysis in R program, the following findings were determined:

From the histogram, it was found that the number of units of sales per week for Ad2 sample were as following:

4 of the samples have 30000 to 35000 sales per week.

3 of the samples have 35000 to 40000 sales per week.

3 of the samples have 40000 to 45000 sales per week.

2 of the samples have 45000 to 50000 sales per week.

2 of the samples have 50000 to 55000 sales per week.

1 of the samples have 55000 to 60000 sales per week.

From the summary statistics of the sample of Ad2 sales, the following findings were determined:

The minimum number of sales per week for Ad2 was 31960 units per week.

The maximum number of sales per week for Ad2 was 56950 units per week.

The average number of sales per week for Ad2 was 40120 units per week.

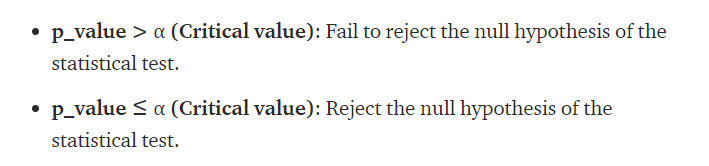
The 25% or first quartile is the middle value of the lower half of the sample of Ad2 sales and was 34765 sales per week.

The 75 % or third quartile is the middle value of the upper half of the sample of Ad2 sales and was 48705 sales per week.

After conducting the One-Sample T-test, the following insights were determined:

A t-value of 5.381 was found. The greater the magnitude of the t-value, the greater there is evidence to reject the null hypothesis that the mean average sales for Ad2 are 30000 units per week and the same as Ad1. The larger the absolute value of the t-value, the smaller the p-value, and the greater the evidence against the null hypothesis.

A p -value of 4.825\*10^-5 was found. The p-value or **probability value** is the probability that when the null hypothesis that the mean average sales for Ad2 are 30000 units per week and the same as Ad1 is true, the sample mean difference between Ad1 and Ad2 would be the same as or of greater magnitude than the observed results.



The critical value (confidence interval) used for this advertising analysis is 0.05. This means that if there were 1000 Ad2 samples of 15 collected sales readings, 5% of the time the null hypothesis that the mean average sales for Ad2 are 30000 units per week and the same as Ad1 can be rejected, and 95% of the time it cannot be rejected.

**Conclusion:**

In this case, the **p-value is 4.825\*10^-5** which is less than critical value of 0.05. Apple should reject the null hypothesis Ho that mean average sales for Ad2 are 30000 units per week and the same as Ad1, and accept the alternative hypothesis Ha that Ad2 has an increase effect on the sales for the I-phone 11 Pro. This is proven by the T-test performed, which indicated the mean of the collected sample of Ad2 sales is 41933.33 units per week and greater than mean of 30000 units per week of Ad1. In conclusion, the Ad2 has made significant improvement to the sales of the new Apple of I-phone 11 Pro and Apple should continue to use it.

**References:**

DATA 1204 Week 7 Class notes

The statistical analysis t-test explained for beginners and experts. Retrieved March 8, 2021, from <https://towardsdatascience.com/the-statistical-analysis-t-test-explained-for-beginners-and-experts-fd0e358bbb62>.

Bevans, R. (2020, December 14). An introduction to t-tests: Definitions, formula and examples. Retrieved March 08, 2021, from https://www.scribbr.com/statistics/t-test/#:~:text=A%20t-test%20is%20a%20statistical%20test%20that%20is,whether%20two%20groups%20are%20different%20from%20one%20another.