Assessing the Consistency of iPhone Releases: A Dual Approach Using Single and Two-Sample T-Tests | Unit 2

Avinash Bunga

Master of Science in Information Systems and Business Analytics

Park University

CIS607HOF1P2024 Applied Business Forecasting

Professor: Dr. Abdelmonaem Jornaz

Aug 25, 2024

Assessing the Consistency of iPhone Releases: A Dual Approach Using Single and Two-Sample T-Tests

Introduction

In this assignment, I set out to analyze the release intervals of various iPhone models across different regions, focusing on the consistency of Apple's global launch strategy. The dataset titled **"iPhone releases of all time,"** sourced from Kaggle, provided comprehensive information on the release dates of iPhone models in numerous countries. This analysis aimed to uncover whether there were significant differences in the time intervals between iPhone releases in the USA and Hong Kong and evaluate the overall consistency of Apple's release schedule (Kaggle, n.d.).

The analysis involved several key steps, including data collection, preparation, and statistical testing. I used a single-sample t-test to compare the release intervals against a hypothesized average and a two-sample t-test to compare the intervals between two different regions. The results of these tests provided insights into Apple's release patterns and whether they differed significantly across markets.

Through careful data preparation and rigorous statistical analysis, this assignment sheds light on how Apple maintains a consistent release schedule across its global markets, offering valuable insights for business strategy and market analysis (Hayes, 2024; ScienceDirect, n.d.).

Data Collection and Preparation

Dataset Description

The dataset used for this analysis, titled "iPhone releases of all time," was sourced from Kaggle. This dataset contains comprehensive information on the release dates of various iPhone models across different countries. The dataset was selected because it provided a

broad and detailed view of Apple's global launch strategy, making it ideal for the analysis of release intervals.

Variables in the Dataset

The dataset includes the following key variables:

- **iPhone model:** The specific model of the iPhone, such as iPhone 3, iPhone 4, etc.
- **Country of Release:** The country where the iPhone model was released.
- **Date of Release:** The official release date of the iPhone model in the specified country.

These variables are essential for understanding the timing and geographical distribution of iPhone releases, which are central to the analyses conducted in this study.

Data Preparation

Step 1: Checking for Missing Values

Before proceeding with the analysis, I first checked the dataset for any missing values. This step was important to ensure the integrity and completeness of the data used in the analysis, As seen in Appendix A and Figure 1 (Datagy, 2022; MachineLearningTutorials, 2023).

Step 2: Handling Missing Data

Upon identifying missing values in the dataset, I handled them appropriately. Given that the number of missing values was low, I opted to drop these rows to maintain the dataset's consistency, As seen in Appendix B and Figure 2 (Datagy, 2022; MachineLearningTutorials, 2023).

Step 3: Filtering USA Data

Next, I filtered the dataset to include only the data for the USA. This step was necessary as the USA was used as a reference point for comparison with other regions,

particularly Hong Kong, in the two-sample t-test, As seen in Appendix C and Figure 3 (Abba, 2023).

Step 4: Calculating Time Differences

Finally, I calculated the time intervals (in days) between consecutive iPhone releases for both the USA and Hong Kong. This involved converting the release dates to a datetime format and then computing the differences, As seen in Appendix D and Figure 4 (Martinez, 2024).

Step 5: Preparing Data for USA and Hong Kong Comparison

To compare the time intervals between iPhone releases in the USA and Hong Kong, I prepared a dataset that included data from both regions. This involved filtering the data for the selected countries and models, calculating the time differences, and saving the prepared data to a new file for further analysis, As seen in Appendix E and Figure 5 (Abba, 2023; Datagy, 2022; MachineLearningTutorials, 2023; Martinez, 2024).

The data preparation steps Appendix A to E and outputs from Figure 1 to 5, including checking for missing values, handling missing data, filtering for USA data, calculating time differences, and preparing data for comparison between the USA and Hong Kong, were important to ensuring the accuracy and reliability of the subsequent analyses. By preparing the data, I was able to perform meaningful statistical tests that provided insights into Apple's iPhone release strategy (Hayes, 2024; ScienceDirect, n.d.).

R Studio Analysis: T-Test on iPhone Release Data in USA

In this analysis, I conducted a one-sample t-test to determine whether the time between iPhone releases is consistent with an average interval of 365 days. The goal was to assess if the release intervals significantly deviate from the expected one-year cycle.

Hypotheses

- **Null Hypothesis (H_0)**: The average time between iPhone releases is equal to 365 days.
- **Alternative Hypothesis (H_1)**: The average time between iPhone releases is not equal to 365 days.

Step 1: Loading the Necessary Library

First, I loaded the readxl library, which is used to read Excel files in R, As seen in Appendix F.

Step 2: Loading the Dataset

Next, I loaded the dataset that contains the time differences between iPhone releases. I used the read_excel function to load the Excel file, As seen in Appendix G.

Step 3: Performing the T-Test

After loading the data, I performed a one-sample t-test to compare the average time between iPhone releases to the hypothesized mean of 365 days, As seen in Appendix H (Hayes, 2024; ScienceDirect, n.d.).

Conclusion for T-Test on iPhone Release Data

As seen in Figure 6, The p-value from the t-test is 0.02836, which is less than the standard significance level of 0.05. This means that there is significant evidence to reject the null hypothesis. Therefore, the average time between iPhone releases is significantly different from the hypothesized 365 days.

The mean time between releases in the dataset is 623 days, with a 95% confidence interval ranging from approximately 406 days to 840 days. This suggests that iPhone releases in USA do not follow a strict annual schedule, and the interval between releases is typically longer than one year (Hayes, 2024; ScienceDirect, n.d.).

R Studio Analysis: Two-Sample T-Test on iPhone Release Data

In this analysis, I compared the time intervals between iPhone releases in the USA and Hong Kong to see if there was a significant difference between the two regions. The goal was to understand whether Apple follows a consistent release schedule across these two markets.

Hypotheses

- **Null Hypothesis (H_0)**: The average time between iPhone releases is the same in the USA and Hong Kong.
- **Alternative Hypothesis (H_1)**: The average time between iPhone releases is different between the USA and Hong Kong.

Step 1: Loading the Data

Since I had already loaded the "readxl" package earlier when performing the previous t-test, I directly loaded the dataset that includes the time intervals between iPhone releases for both the USA and Hong Kong, As seen in Appendix I.

Step 2: Preparing the Data for Analysis

I separated the time intervals for the USA and Hong Kong into two different variables. This step was important to ensure that the data was ready for the two-sample t-test, As seen in Appendix J.

Step 3: Performing the Two-Sample T-Test

Next, I performed the two-sample t-test to compare the average time between iPhone releases in the USA and Hong Kong, As seen in Appendix K (Hayes, 2024; ScienceDirect, n.d.).

Conclusion for Two-Sample T-Test on iPhone Release Data

As seen in Figure 7, After performing the two-sample t-test to compare the average time intervals between iPhone releases in the USA and Hong Kong, the results were as follows:

- **T-statistic**: The t-statistic is 0, indicating that there is no difference in the average time intervals between iPhone releases in the USA and Hong Kong.
- **P-value**: The p-value is 1, which is much higher than the typical significance level of 0.05. This indicates that there is no statistically significant difference between the two regions.
- **Confidence Interval**: The 95% confidence interval for difference in means ranges from approximately -311 days to +311 days. Since this interval includes 0, it further confirms that there is no significant difference between the two means.
- **Sample Means**: Both the USA and Hong Kong have the same average time between iPhone releases, which is 599 days (Hayes, 2024; ScienceDirect, n.d.).

These results suggest that the time intervals between iPhone releases are consistent across the USA and Hong Kong, with no meaningful variation in the timing of releases between these two markets.

Conclusion

This analysis of iPhone release intervals revealed that Apple maintains a consistent global release strategy, with no significant differences in the timing of releases between the USA and Hong Kong. The single-sample t-test showed that the average time between releases significantly differed from the hypothesized annual interval of 365 days, indicating that the actual release intervals are typically longer. However, the two-sample t-test comparing the USA and Hong Kong demonstrated that these intervals do not significantly differ between the two regions, suggesting a uniform release schedule.

The findings from this assignment highlight Apple's commitment to a consistent global launch strategy, with little variation in release timing across major markets. This consistency likely reflects Apple's strategic planning and coordination across its global operations, ensuring that product launches are synchronized across key regions.

Overall, this assignment provided a deeper understanding of Apple's iPhone release strategy, supported by thorough data analysis and statistical testing. The results contribute to a broader comprehension of how global technology companies like Apple manage and execute their product launches on a worldwide scale.

References

Abba, I. (2023, March 13). *Pandas.DataFrame.Sort_values - How To Sort Values in Pandas*. Freecodecamp. https://www.freecodecamp.org/news/how-to-sort-values-in-pandas/

Datagy (n.d.). *All the Ways to Filter Pandas Dataframes*. Retrieved May 31, 2020, from https://datagy.io/filter-pandas/

Datagy (2022, September 7). *Pandas dropna(): Drop Missing Records and Columns in DataFrames*. https://datagy.io/pandas-dropna/

Hayes, A. (2024, August 5). *T-Test: What It Is With Multiple Formulas and When To Use Them*. Investopedia. https://www.investopedia.com/terms/t/t-test.asp

Kaggle (n.d.). *IPhone releases of all time*. Retrieved August 24, 2024, from https://www.kaggle.com/datasets/hanningong/iphone-releases-of-all-time

MachineLearningTutorials (2023, August 23). *Pandas isnull() Function Explained (With Examples)*. https://machinelearningtutorials.org/pandas-isnull-function-explained-with-examples/

Martinez, H. (2024, April 30). *Pandas to_datetime (pd.To_datetime)*. Pyimagesearch. https://pyimagesearch.com/2024/04/30/pandas-to_datetime-pd-to_datetime/

ScienceDirect (n.d.). *Two Sample t-Test*. Retrieved August 24, 2024, from https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/two-sample-t-test#:~:text=The%20two%2Dsample%20t%2Dtest%20is%20a%20way%20of%20assessing,that%20the%20means%20are%20equal.

Appendix A

Python Script 1: Data Preparation

```
import pandas as pd
#Updated file path without spaces
file_path =
'/Users/avinash/Desktop/CIS/CIS607/Unit2/Assignment/py/AppleDa
ta.xlsx'
#Load the dataset
apple_data = pd.read_excel(file_path)
#Check for missing values in each column
missing_values = apple_data.isnull().sum()
#Display the missing values count
print("Missing values in each column:")
print(missing_values)
#Optionally, display the total number of missing values
total_missing = missing_values.sum()
print(f"\nTotal missing values in the dataset:
{total_missing}")
```

Figure 1

Output

```
(myenv) avinash@avinashs-MacBook-Pro py % python3 missing1.py
Missing values in each column:
iPhone model          0
Country of Release    1
Date of Release       0
dtype: int64

Total missing values in the dataset: 1
```

Appendix B

Python Script 2: Handling Missing Data

```
import pandas as pd
#Updated file path without spaces
file_path =
'/Users/avinash/Desktop/CIS/CIS607/Unit2/Assignment/py/AppleDa
ta.xlsx'
#Load the dataset
apple_data = pd.read_excel(file_path)
#Check for missing values in each column
print("Missing values before cleaning:")
print(apple_data.isnull().sum())
#Drop rows with any missing values
apple_data_cleaned = apple_data.dropna()
#Verify if the missing values have been handled
print("\nMissing values after cleaning:")
print(apple_data_cleaned.isnull().sum())
#Save the cleaned dataset to a new file
apple_data_cleaned.to_excel('/Users/avinash/Desktop/CIS/CIS607
/Unit2/Assignment/py/AppleData_Cleaned.xlsx', index=False)
```

Figure 2

Output

```
(myenv) avinash@avinashs-MacBook-Pro py % python3 cleaned.py
Missing values before cleaning:
iPhone model          0
Country of Release    1
Date of Release       0
dtype: int64

Missing values after cleaning:
iPhone model          0
Country of Release    0
Date of Release       0
dtype: int64
```

Appendix C

Python Script 3: Filtering USA Data

```
import pandas as pd
#Load the cleaned dataset
file_path =
'/Users/avinash/Desktop/CIS/CIS607/Unit2/Assignment/py/AppleDa
ta_Cleaned.xlsx'
apple_data_cleaned = pd.read_excel(file_path)
#Filter the dataset for only US release dates
usa_data = apple_data_cleaned[apple_data_cleaned['Country of
Release'] == 'US']
#Drop rows with NaN values in 'Date of Release' if they exist
usa_data = usa_data.dropna(subset=['Date of Release'])
#Sort by 'Date of Release' to ensure correct chronological
order
usa_data.sort_values(by='Date of Release', inplace=True)
#Save the filtered US data
output_path =
'/Users/avinash/Desktop/CIS/CIS607/Unit2/Assignment/py/USA_App
leData.xlsx'
usa_data.to_excel(output_path, index=False)
print(f"Filtered US data saved to: {output_path}")
#Display the first few rows of the US data
print(usa_data.head())
```

Figure 3

Output

```
(myenv) avinash@avinashs-MacBook-Pro py % python3 usa.py
Filtered US data saved to: /Users/avinash/Desktop/CIS/CIS607/Unit2/Assignment/py/USA_AppleData.xlsx
   iPhone model  Country of Release Date of Release
0        iPhone                  US      2007-06-29
3      iPhone 3                  US      2009-07-11
25     iPhone 4                  US      2010-06-24
47     iPhone 5                  US      2012-09-21
78     iPhone 6                  US      2014-09-19
```

Appendix D

Python Script 4: Calculating Time Differences

```
import pandas as pd
#Load the dataset
file_path =
'/Users/avinash/Desktop/CIS/CIS607/Unit2/Assignment/py/USA_App
leData.xlsx'
usa_data = pd.read_excel(file_path)
#Convert 'Date of Release' to datetime format
usa_data['Date of Release'] = pd.to_datetime(usa_data['Date of
Release'])
#Calculate the time differences in days between consecutive
releases
usa_data['Time Between Releases (days)'] = usa_data['Date of
Release'].diff().dt.days
#Drop the first row with NaN value in 'Time Between Releases
(days)'
usa_data_cleaned = usa_data.dropna(subset=['Time Between
Releases (days)'])
#Display the cleaned data with calculated time differences
print("Cleaned Data with Time Differences:")
print(usa_data_cleaned)
#save the cleaned data to a new file
output_path =
'/Users/avinash/Desktop/CIS/CIS607/Unit2/Assignment/py/USA_App
leData_TimeDifferences.xlsx'
usa_data_cleaned.to_excel(output_path, index=False)
print(f"\nData with time differences saved to: {output_path}")
```

Figure 4

Output

```
(myenv) avinash@avinashs-MacBook-Pro py % python3 time.py
Cleaned Data with Time Differences:
  iPhone model  Country of Release Date of Release  Time Between Releases (days)
1      iPhone 3                 US      2009-07-11                         743.0
2      iPhone 4                 US      2010-06-24                         348.0
3      iPhone 5                 US      2012-09-21                         820.0
4      iPhone 6                 US      2014-09-19                         728.0
5      iPhone 7                 US      2016-09-16                         728.0
6      iPhone 8                 US      2017-09-22                         371.0
```

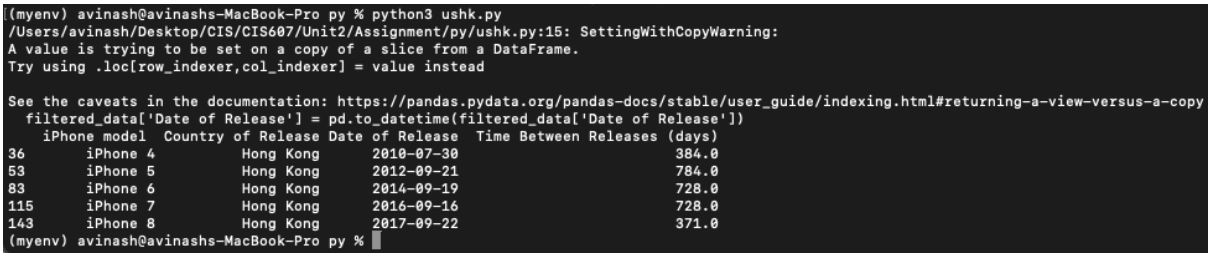
Appendix E

Python Script 5: Preparing Data for USA and Hong Kong Comparison

```
import pandas as pd
#Load the dataset
file_path =
'/Users/avinash/Desktop/CIS/CIS607/Unit2/Assignment/py/AppleDa
ta_Cleaned.xlsx'
apple_data_cleaned = pd.read_excel(file_path)
#Select the countries for comparison (USA and Hong Kong)
selected_countries = ['US', 'Hong Kong']
#Filter the data for the selected countries and iPhone models
iphone_models = ['iPhone 3', 'iPhone 4', 'iPhone 5', 'iPhone
6', 'iPhone 7', 'iPhone 8']
filtered_data = apple_data_cleaned[apple_data_cleaned['Country
of Release'].isin(selected_countries) &
apple_data_cleaned['iPhone model '].isin(iphone_models)]
#Convert 'Date of Release' to datetime format
filtered_data['Date of Release'] =
pd.to_datetime(filtered_data['Date of Release'])
#Sort the data by 'Country of Release' and 'Date of Release'
filtered_data = filtered_data.sort_values(by=['Country of
Release', 'Date of Release'])
#Calculate the time differences in days between consecutive
releases for each country
filtered_data['Time Between Releases (days)'] =
filtered_data.groupby('Country of Release')['Date of
Release'].diff().dt.days
#Drop rows with NaN values resulting from the diff() operation
filtered_data = filtered_data.dropna(subset=['Time Between
Releases (days)'])
#Save the prepared data to a new Excel file
output_path =
'/Users/avinash/Desktop/CIS/CIS607/Unit2/Assignment/py/USA_HK_
AppleData_TimeDifferences.xlsx'
filtered_data.to_excel(output_path, index=False)
#Display the first few rows of the prepared data
print(filtered_data.head())
```

Figure 5

Output



```
(myenv) avinash@avinashs-MacBook-Pro py % python3 ushk.py
/Users/avinash/Desktop/CIS/CIS607/Unit2/Assignment/py/ushk.py:15: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  filtered_data['Date of Release'] = pd.to_datetime(filtered_data['Date of Release'])
    iPhone model  Country of Release Date of Release  Time Between Releases (days)
36       iPhone 4          Hong Kong      2010-07-30                        384.0
53       iPhone 5          Hong Kong      2012-09-21                        784.0
83       iPhone 6          Hong Kong      2014-09-19                        728.0
115      iPhone 7          Hong Kong      2016-09-16                        728.0
143      iPhone 8          Hong Kong      2017-09-22                        371.0
(myenv) avinash@avinashs-MacBook-Pro py %
```

Appendix F

R Script 1: Loading the Necessary Library

```
install.packages("readxl")
library(readxl)
```

Appendix G

R Script 2: Loading the Dataset

```
#Load the necessary library
library(readxl)
#Load the dataset
file_path <-
"/Users/avinash/Desktop/CIS/CIS607/Unit2/Assignment/py/US
A_AppleData_TimeDifferences.xlsx"
usa_data <- read_excel(file_path)
#View the first few rows of the dataset to confirm it's
loaded correctly
head(usa_data)
```

Appendix H

R Script 3: Performing the T-Test

```
#Perform a one-sample t-test
#Hypothesized mean (e.g., 365 days for annual release)
hypothesized_mean <- 365
t_test_result <- t.test(usa_data$`Time Between Releases
(days)`, mu = hypothesized_mean)
#Print the t-test results
print(t_test_result)
```

Figure 6

The t-test results

```
        One Sample t-test

data:  usa_data$`Time Between Releases (days)`
t = 3.052, df = 5, p-value = 0.02836
alternative hypothesis: true mean is not equal to 365
95 percent confidence interval:
 405.6943 840.3057
sample estimates:
mean of x
      623
```

Appendix I

R Script 4: Loading the Data

```
#Load the dataset
file_path <-
"/Users/avinash/Desktop/CIS/CIS607/Unit2/Assignment/py/US
A_HK_AppleData_TimeDifferences.xlsx"
apple_data <- read_excel(file_path)
#View the first few rows to ensure it loaded correctly
head(apple_data)
```

Appendix J

R Script 5: Preparing the Data for Analysis

```
#Separate the data for USA and Hong Kong
us_data <- apple_data[apple_data$`Country of Release` ==
"US", "Time Between Releases (days)"]
hk_data <- apple_data[apple_data$`Country of Release` ==
"Hong Kong", "Time Between Releases (days)"]
#Check the data
print("USA Data:")
print(us_data)
print("Hong Kong Data:")
print(hk_data)
```

Appendix K

R Script 6: Performing the Two-Sample T-Test

```
#Perform the two-sample t-test
t_test_result <- t.test(us_data$`Time Between Releases
(days)`, hk_data$`Time Between Releases (days)`)
#Print the results of the t-test
print(t_test_result)
```

Figure 7

The two sample t-test results