University of Wales Trinity St David

School of Engineering

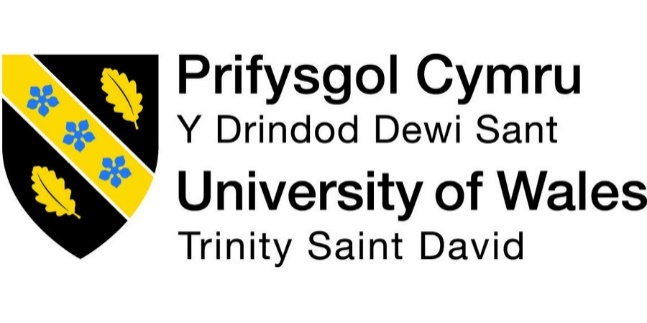
Experimental data analysis

Data Analysis & Visualisation  
ACBA4002

Richard Lamnea

1802749

24/04/2020



# Abstract

An online retail company created a new website design and decided to measure the effect this has upon sales. A/B testing was applied, and t-testing was conducted on the data. The results are discussed in this report.

# Table of Contents

[Abstract 1](#_Toc39752404)

[Table of Contents 1](#_Toc39752405)

[List of Figures 1](#_Toc39752406)

[List of Tables 2](#_Toc39752407)

[1 Introduction 3](#_Toc39752408)

[2 Methodology 3](#_Toc39752409)

[2.1 Hypothesis 4](#_Toc39752410)

[2.2 Data Testing 4](#_Toc39752411)

[2.2.1 T-Test 1: 4](#_Toc39752412)

[2.2.2 T-Test 2: 5](#_Toc39752413)

[2.2.3 Test 3: 5](#_Toc39752414)

[2.3 Investigation of results 6](#_Toc39752415)

[2.3.1 Outliers 6](#_Toc39752416)

[2.4 Discussion of results 9](#_Toc39752417)

[3 Conclusion 9](#_Toc39752418)

[4 References 10](#_Toc39752419)

[5 Appendices 11](#_Toc39752420)

[Appendix A – Website design test data 11](#_Toc39752421)

# List of Figures

[Figure 1 - Outlier test 7](#_Toc39752422)

# List of Tables

[Table 1 - Statistics of website design sales data 2](#_Toc39752423)

[Table 2 - t-Test result for 0.10 significance value 4](#_Toc39752424)

[Table 3 - t-Test result for 0.05 significance value 4](#_Toc39752425)

[Table 4 - t-Test result for 0.01 significance value 5](#_Toc39752426)

[Table 5 - t-Test retest result for 0.10 significance value 6](#_Toc39752427)

[Table 6 - t-Test retest result for 0.05 significance value 7](#_Toc39752428)

[Table 7 - t-Test retest result for 0.05 significance value 7](#_Toc39752429)

# Introduction

This report discusses a statistical analysis of the sales of an e-commerce retail company which has recently created a new website design and wishes to measure the effects the design has upon sales. A web server was arranged to display either the original website or the new website design through the process of A/B Testing.[1] Sales data for each website design was recorded each day over a 40-day testing period and this data was analyzed to determine if there was a statistically significant difference between both designs. The company had also offered a 15% discount over this period.

# Methodology

A statistical analysis of the data was conducted on the dataset given which can be seen in appendix A; Using this data a set of descriptive statistics were created to help provide a better understanding of said data which can be seen in table 1.

Table - Statistics of website design sales data

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Descriptive Statistics | | |  | |  |  | |  | |  |
| *Old Website Design* | | |  |  | | *New Website Design* | | | | |
|  |  | |  |  | |  | | |  | |
| Mean | 63.075 | |  |  | | Mean | | | 80.025 | |
| Standard Error | 4.970991 | |  |  | | Standard Error | | | 5.057171539 | |
| Median | 62 | |  |  | | Median | | | 82.5 | |
| Mode | 26 | |  |  | | Mode | | | 99 | |
| Standard Deviation | 31.43931 | |  |  | | Standard Deviation | | | 31.98436116 | |
| Sample Variance | 988.4301 | |  |  | | Sample Variance | | | 1022.999359 | |
| Kurtosis | -1.3617 | |  |  | | Kurtosis | | | -0.835646087 | |
| Skewness | 0.036911 | |  |  | | Skewness | | | -0.287496805 | |
| Range | 98 | |  |  | | Range | | | 115 | |
| Minimum | 15 | |  |  | | Minimum | | | 15 | |
| Maximum | 113 | |  |  | | Maximum | | | 130 | |
| Sum | 2523 | |  |  | | Sum | | | 3201 | |
| Count | 40 | |  |  | | Count | | | 40 | |

At first comparison of the data, the statistics show the mean (average) sales over the period are greater for the new website design than of the old website design. They also show us the median (the middle) value is greater for the website design and that the number that appears most is greater for the new design. There are also other indications that the new design is better for sales as the maximum sales are larger and the sum (total) of sales was greater however this is just an assumption. To understand if these values have weight, we need to test the means against each other to measure the effect of sales when changing the design of the website by using a t-test to show if the difference is statistically significant over the period allowing us to reject the hypothesis and accept the alternative.  
  
The t-test is a statistical hypothesis test in which the test statistic follows the t-distribution under the null hypothesis. The t-test is commonly used when the test statistic would follow a normal distribution, when the scaling value of the test is known. When this value is unknown and is replaced by an estimate based on the data, the test stat will follow the t distribution. The t-test can be used to determine if the average (means) of two sets of data are significantly different from each other and allow us to calculate a probability.[1]

## Hypothesis

The experiment is designed to measure whether the output results (DV, Dependant variable) of a condition differ from another condition which is known as the independent variable (IV). In this case:  
 The output variable (DV): Number of sales  
 The independent variable (IV): Design  
 conditions: Old website design, New website design.

As it is incredibly difficult to prove the truth, instead we prove false and accept the alternative hypothesis which is known as the null hypothesis.[2] A testable hypothesis was created based on the variables above which can be seen below:

H0: μOldDesign = μNewDesign  
Hα: μOldDesign ≠ μNewDesign

The null hypothesis in this test is that the mean of the old website design is the same as the mean of the new design, and our alternative hypothesis is that there is a difference between the old design and the new design.[2]

## Data Testing

A two-tail t-test was chosen to be appropriate as we want to determine whether there is a difference between the means of both groups.[2]  
The significance level, alpha (α) used will be 0.10, 0.05, and 0.01 for cross-referencing the t obtained value against the t critical value.[3]

### T-Test 1:

The results for the t-test for 0.10 alpha value can be seen in Table 2 below.

Table - t-Test result for 0.10 significance value

|  |  |  |
| --- | --- | --- |
| t-Test: Two-Sample Assuming Unequal Variances | | |
|  |  |  |
|  | Old Website Design | New Website Design |
| Mean | 63.075 | 80.025 |
| Variance | 988.4301282 | 1022.999359 |
| Observations | 40 | 40 |
| Hypothesized Mean Difference | 0 |  |
| df | 78 |  |
| t Stat | -2.390271823 |  |
| P(T<=t) one-tail | 0.009624153 |  |
| t Critical one-tail | 1.292499597 |  |
| P(T<=t) two-tail | 0.019248306 |  |
| t Critical two-tail | 1.664624645 |  |

α: 0.10

Calculated tobt: -2.390271823

Found tcrit: 1.664624645

tobt >tcrit: therefore, we can reject the null hypothesis and accept the alternative.

### T-Test 2:

The results for the t-test for 0.05 alpha value can be seen in Table 3 below.

Table - t-Test result for 0.05 significance value

|  |  |  |
| --- | --- | --- |
| t-Test: Two-Sample Assuming Unequal Variances | | |
|  |  |  |
|  | *Old Website Design* | *New Website Design* |
| Mean | 63.075 | 80.025 |
| Variance | 988.4301282 | 1022.999359 |
| Observations | 40 | 40 |
| Hypothesized Mean Difference | 0 |  |
| df | 78 |  |
| t Stat | -2.390271823 |  |
| P(T<=t) one-tail | 0.009624153 |  |
| t Critical one-tail | 1.292499597 |  |
| P(T<=t) two-tail | 0.019248306 |  |
| t Critical two-tail | 1.990847069 |  |

α: 0.05

Calculated tobt: -2.390271823

Found tcrit: 1.990847069

tobt >tcrit: therefore, we can reject the null hypothesis and accept the alternative.

### Test 3:

The results for the t-test for 0.01 alpha value can be seen in Table 4 below.

Table - t-Test result for 0.01 significance value

|  |  |  |
| --- | --- | --- |
| t-Test: Two-Sample Assuming Unequal Variances | | |
|  |  |  |
|  | *Old Website Design* | *New Website Design* |
| Mean | 63.075 | 80.025 |
| Variance | 988.4301282 | 1022.999359 |
| Observations | 40 | 40 |
| Hypothesized Mean Difference | 0 |  |
| df | 78 |  |
| t Stat | -2.390271823 |  |
| P(T<=t) one-tail | 0.009624153 |  |
| t Critical one-tail | 2.375110958 |  |
| P(T<=t) two-tail | 0.019248306 |  |
| t Critical two-tail | 2.640340015 |  |

α: 0.01

Calculated tobt: -2.390271823

Found tcrit: 2.640340015

tobt >tcrit: therefore, we will accept the null hypothesis and reject the alternative at a significance level of 0.01.

## Investigation of results

The results from the t-test experiment showed that at a significance value of 0.10, and 0.05 rejected the hypothesis and accepted the alternative but at a level of 0.01 accepted the null hypothesis and rejected the alternative. These results would suggest that overall, there is less than a 5% probability that the difference between the means occurred by chance alone suggesting that the new website design performs better than the old website design. However, at a significance level of 0.01, the null hypothesis was accepted, meaning there was a 1% chance that this difference occurred by chance.

### Outliers

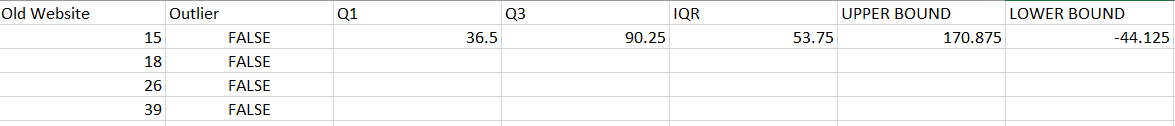
The data was investigated for outliers and then they were removed from the dataset as they can skew the mean and reduce the accuracy of the prediction and cause misleading assumptions. These outliers are observations that appear abnormal from the other values in a random sample of the population.[4] The data was split into quartiles which is a dividing point in which the data is separated. There are 3 quartiles: Q1, Q2, and Q3. The first(Q1) is the lower quarter of the data where the smallest values are located. Q2 are the middle values and Q3 are the largest values of the data. The range of values from Q1 to Q3 are called the inter-quartile range (IQR).[5] Once the calculations had been made the values in the dataset were tested against the upper and lower bounds to detect if there had been any outliers which can be seen in Fig 1.  


Figure - Outlier test

One outlier was detected within the new website design data and an anomaly was noticed when reading through the data, so the tests were conducted again with the outlier data and anomalies removed to allow for better distribution.

#### Significance 0.10 retest results

The results for the t-test for 0.10 alpha value can be seen in Table 5 below.

Table - t-Test retest result for 0.10 significance value

|  |  |  |
| --- | --- | --- |
| t-Test: Two-Sample Assuming Unequal Variances | |  |
|  |  |  |
|  | *Old Website* | *New Website* |
| Mean | 60.08108108 | 76.75675676 |
| Variance | 923.9099099 | 949.3003003 |
| Observations | 37 | 37 |
| Hypothesized Mean Difference | 0 |  |
| df | 72 |  |
| t Stat | -2.343640767 |  |
| P(T<=t) one-tail | 0.010930468 |  |
| t Critical one-tail | 1.293420507 |  |
| P(T<=t) two-tail | 0.021860937 |  |
| t Critical two-tail | 1.666293696 |  |

α: 0.10

Calculated tobt: -2.343640767

Found tcrit: 1.666293696

tobt >tcrit: therefore, we can reject the null hypothesis and accept the alternative.

#### Significance 0.05 retest results

The results for the t-test for 0.05 alpha value can be seen in Table 6 below.

Table - t-Test retest result for 0.05 significance value

|  |  |  |
| --- | --- | --- |
| t-Test: Two-Sample Assuming Unequal Variances | |  |
|  |  |  |
|  | *Old Website* | *New Website* |
| Mean | 60.08108108 | 76.75675676 |
| Variance | 923.9099099 | 949.3003003 |
| Observations | 37 | 37 |
| Hypothesized Mean Difference | 0 |  |
| df | 72 |  |
| t Stat | -2.343640767 |  |
| P(T<=t) one-tail | 0.010930468 |  |
| t Critical one-tail | 1.666293696 |  |
| P(T<=t) two-tail | 0.021860937 |  |
| t Critical two-tail | 1.993463567 |  |

α: 0.05

Calculated tobt: -2.343640767

Found tcrit: 1.993463567

tobt >tcrit: therefore, we can reject the null hypothesis and accept the alternative.

#### Significance 0.01 retest results

The results for the t-test for 0.01 alpha value can be seen in Table 7 below.

Table - t-Test retest result for 0.05 significance value

|  |  |  |
| --- | --- | --- |
| t-Test: Two-Sample Assuming Unequal Variances | |  |
|  |  |  |
|  | *Old Website* | *New Website* |
| Mean | 60.08108108 | 76.75675676 |
| Variance | 923.9099099 | 949.3003003 |
| Observations | 37 | 37 |
| Hypothesized Mean Difference | 0 |  |
| df | 72 |  |
| t Stat | -2.343640767 |  |
| P(T<=t) one-tail | 0.010930468 |  |
| t Critical one-tail | 2.379262129 |  |
| P(T<=t) two-tail | 0.021860937 |  |
| t Critical two-tail | 2.645851913 |  |

α: 0.01

Calculated tobt: -2.343640767

Found tcrit: 2.645851913

tobt >tcrit: therefore, we will accept the null hypothesis and reject the alternative at a significance level of 0.01.

## Discussion of results

The t-test results show us that at a significance level of 0.10 and 0.05, or 90%, 95% probability that the new design made a difference to the overall sales. However, at a 0.01 significance level, the t-test accepted the null hypothesis. These results are similar to the original test results even with the removal of anomalies and the IQR outliers.

# Conclusion

The overall results of the tests conducted produce similar results, and the evidence suggests that the new website design is overall a better candidate, however, there are factors within the study that could impede the accuracy of the results, such as the 15% discount applied over the testing period as the test results would not be reflective of the average sale period. The data should be recorded without a discount applied to get a better reflection of the normal distribution.

# References

[1] M. B. Editor, ‘Understanding t-Tests: t-values and t-distributions’. https://blog.minitab.com/blog/adventures-in-statistics-2/understanding-t-tests-t-values-and-t-distributions (accessed Apr. 30, 2020).

[2] J. W. Schneider, ‘Null hypothesis significance tests. A mix-up of two different theories: the basis for widespread confusion and numerous misinterpretations’, *Scientometrics*, vol. 102, no. 1, pp. 411–432, Jan. 2015, doi: 10.1007/s11192-014-1251-5.

[3] ‘t-Tables’. https://faculty.washington.edu/heagerty/Books/Biostatistics/TABLES/t-Tables/ (accessed Apr. 30, 2020).

[4] ‘7.1.6. What are outliers in the data?’ https://www.itl.nist.gov/div898/handbook/prc/section1/prc16.htm (accessed Apr. 30, 2020).

[5] D. Clark‐Carter, ‘Interquartile Range’, in *Encyclopedia of Statistics in Behavioral Science*, American Cancer Society, 2005.

# Appendices

## Appendix A – Website design test data

|  |  |  |
| --- | --- | --- |
| Day | Old Website Design | New Website Design |
| 1 | 15 | 15 |
| 2 | 18 | 17 |
| 3 | 26 | 24 |
| 4 | 39 | 37 |
| 5 | 61 | 59 |
| 6 | 84 | 83 |
| 7 | 96 | 99 |
| 8 | 104 | 112 |
| 9 | 110 | 126 |
| 10 | 113 | 130 |
| 11 | 111 | 128 |
| 12 | 108 | 125 |
| 13 | 103 | 119 |
| 14 | 99 | 116 |
| 15 | 98 | 113 |
| 16 | 94 | 110 |
| 17 | 89 | 106 |
| 18 | 86 | 104 |
| 19 | 81 | 101 |
| 20 | 80 | 99 |
| 21 | 76 | 95 |
| 22 | 72 | 91 |
| 23 | 68 | 89 |
| 24 | 63 | 85 |
| 25 | 60 | 82 |
| 26 | 54 | 80 |
| 27 | 51 | 76 |
| 28 | 48 | 74 |
| 29 | 45 | 72 |
| 30 | 41 | 68 |
| 31 | 39 | 62 |
| 32 | 37 | 61 |
| 33 | 35 | 57 |
| 34 | 31 | 55 |
| 35 | 76 | 103 |
| 36 | 29 | 52 |
| 37 | 26 | 48 |
| 38 | 22 | 44 |
| 39 | 19 | 43 |
| 40 | 16 | 41 |