

Experimental Data Analysis (Harry Morris)

by Harry Morris

FILE	11373_HARRY_MORRIS_EXPERIMENTAL_DATA_ANALYSIS__HARRY_MORRIS__192216_1831155439.DOCX (555.79K)		
TIME SUBMITTED	29-MAY-2020 09:42PM (UTC+0100)	WORD COUNT	1944
SUBMISSION ID	129584085	CHARACTER COUNT	8442

EXPERIMENTAL DATA ANALYSIS



25/05/2020

Harry Morris

This document contains the data that I will be using, my hypotheses, my null hypotheses, selecting the testing, testing the data and a conclusion to rap this document up.

TABLE OF CONTENTS

Introduction 2

Description of the Data..... 2

Developing a Hypotheses..... 4

Selecting a Suitable Test..... 4

Testing..... 5

Conclusion..... 8

REFERENCES 8

BIBLIOGRAPHY 8

Introduction

As a part of this assignment I have been given data from an online retail company. This company wants to make some changes to their website, they would like to measure the effect that the changes had on their sales. The company has provided me with data from before they made changes to the website and the data after they made the changes. I have to analyse this data in order to recommend which website provided the company with the most sales. This document shows the process which I took in order to complete this task.

Description of the Data

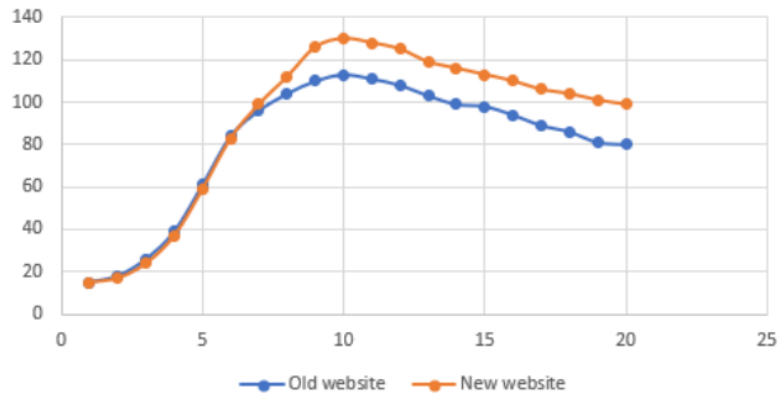
This is the table of data the company has provided me with. As you can see the table shows a period of 40 days and compares both the old and new website together. This table shows that the website reaches its peak between the day 8 and day 13. Between this period of time the old website made over 100 sales each day, but this only lasted until day 13, this means the peak lasted a whole 5 days. While the new website peak lasted a whole 5 days as well but they also had over 100 sales for a whole 11 days. After the peak was reached both websites had a slow decline in sales. Unlike the old website the new website was able to reach over 100 sales after the peak in sales this happened on day 35 the number of visitors then dropped dramatically as the visitor count decreased by 51.

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

After getting my hands on the data I decided to convert the data into graphs in order to make it easier to read. The graph I decided to use was scatter graphs, the reason why I chose this type of graph is because it is easy to see the trends/peak in the data as the line will also peak when the number of sales is high. In total I made two graphs in order to make the data easier to read, I broke the data down into number of days the first graph (which is shown below) shows the number of sales from day 1 to day 20. The reason why I did this was to make the data easier to read.

Good

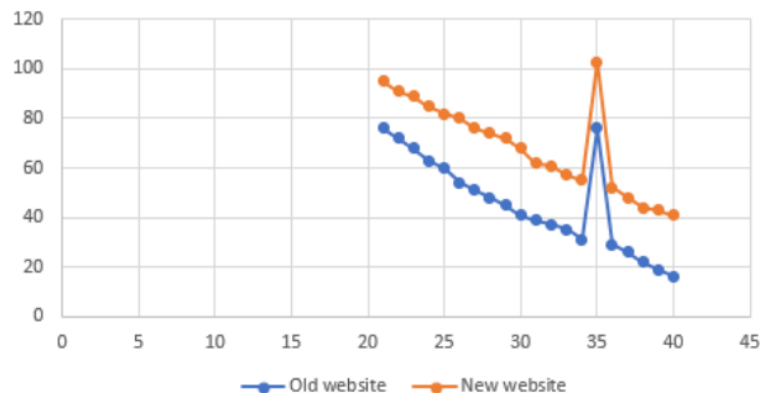
Sale to the both webistes days 1-20



Label

The graph above proves that the peek which happened between the day 8 and day 13. I also show that the new website was able to hold a higher sales count over this period of time, this is proven by the orange line on the graph. Before hitting day 8 both websites had a similar sales count with the older website having more daily sales, but this began to end on day 7 before the peek sales which followed in the next five days. After the peek the newer website had more sales, this lasted until the end of the 20-day period. I also created another scatter graph to show the number of sales between the 21 day and 40 day. The image below is the graph which I created.

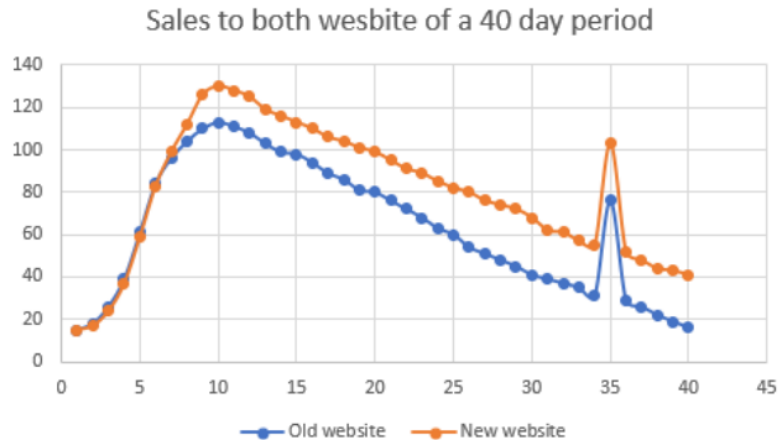
Sales to both webistes days 21-40



Label

When compared to the first half of the graph this part of the graph shows the decline of sales to the website. This decline starts on the 21st day and then continues to decline until day 35, this day had an increase of sales but unlike the spike of sales on day 8 the spike on day 35 only lasted for one day. This is proven by the graph and the number of users recorded on day 36. The new website had 52 sales instead of the 103, this is reduction of 51 sales which is almost half when compared to day 35. The older website had 79 sales on day 35, this number was then decreased on day 36 as only 29 people sales the website on that day.

In order to see the trends and number of visitors in more detail I decided to combine the data from all 40 days together in order to make on graph with all the information on it. Instead of breaking them down into their own small section. Below is the graph which I crated in order to display all of the data.



Developing a Hypotheses

Hypothesis



The hypothesis is made before the testing period. My hypothesis is based around the graphs and the data which I have been provided. Base on these factors it is hypothesised, that the new website attracts more visitors when it's compared to the older website. This is proven by the fact that the sales made by each website over a 40-day period differ as the new website has a higher peak when compared to the old website. The graphs which I have created and the data within the table which I have been provided shows that new website has a higher sales count.

Null Hypothesis

The null hypothesis theorises that before the testing period that both the new and old website have no different between the amount of sales they generate. This means that they would be identical in every single way. Currently there is no evidence or data provided anywhere didn't show that these websites weren't connected.

Selecting a Suitable Test

Currently the data that I have been provided show that the new website is much better than the older one, as it's able to produce more sales over the 40-day period. However, this is not the case without testing as the result could be incorrect. For example the data which I have been provided could be a fluke, maybe in another time the older website might get more views which means it could make more sales or maybe the data which has been provided hasn't been recorded properly meaning that the data for both website is incorrect. Therefore, testing has to be carried out in order to ensure that the data which I have provided isn't a fluke, or hasn't been recorded correctly.

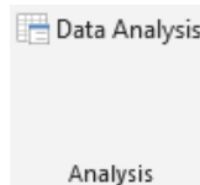
The test which I have selected is a T Test. The reason why I selected a t-test is because it allows me [1] 'to compare the average values of the two data sets and determine if they came from the same population.' This means that the data which I have been provided is correct and the results which I get back from the test are correct. A t-test can also indicate [1] 'that data readings are strong and are probably not due to chance.' But this it only happens if the null hypothesis qualifies to be rejected. Another reason why I chosen a t-test is because it's one of the tests which is used to test the statistics in a hypothesis.

Good

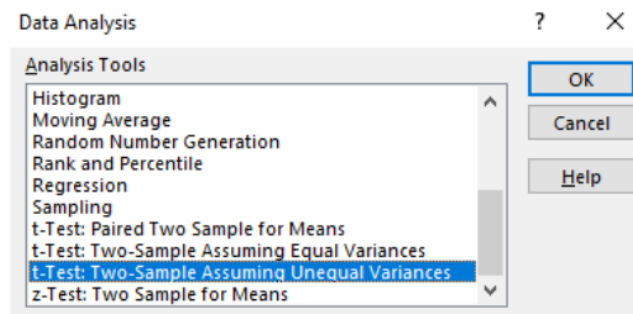
Testing

After deciding on what test to use I then head to Excel in order to test the data which I have been given. The first thing I had to do was add the data analysis section onto Excel, this is done through the options on the program. After added the data analysis section I could start on testing the data.

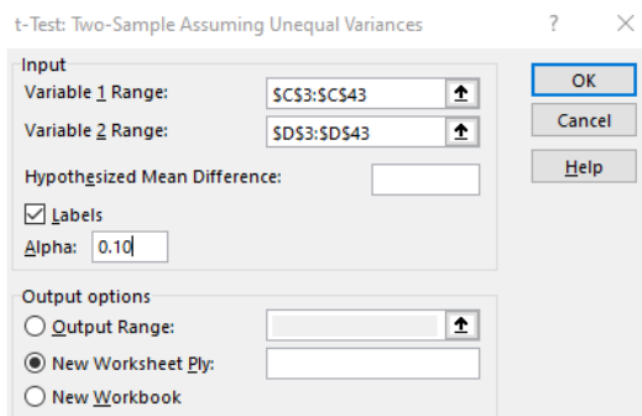
In order to create a t-test I had to click on the data analysis button in Excel.



After clicking the data analysis, a menu popped up, this menu had a list of different data analysis tests. The next thing I had to do was find the correct t-test. The one I needed was called t-Test: Two-Sample Assuming Unequal Variances.



After clicking this option, a new menu popped up asking me to fill in the required data. Once I filled in the range if both variables and change Alpha to 0.10, I could press ok.



The table below is the result of Alpha 0.10

t-Test: Two-Sample Assuming Unequal Variances		
	15	15
Mean	64.30769231	81.69231
Variance	952.0607287	935.7976
Observations	39	39
Hypothesized Mean Difference	0	
df	76	
t Stat	-2.498691812	
P(T<=t) one-tail	0.007312224	
t Critical one-tail	1.292790268	
P(T<=t) two-tail	0.014624447	
t Critical two-tail	1.665151353	

3

This t-test doesn't prove that the new website is better than the older website. This is because the t stat value (-2.498691812) wasn't larger than the t critical two-tail (1.665151353). This doesn't disprove my null hypothesis, in order to make sure this isn't just a fluke with 0.10 I'm going to redo the t-test to ensure that I can't disprove my null hypothesis.

4

Below is a screenshot of the result from another t-test with the Alpha set to 0.05 instead.

t-Test: Two-Sample Assuming Unequal Variances		
	15	15
Mean	64.30769231	81.69230769
Variance	952.0607287	935.7975709
Observations	39	39
Hypothesized Mean Difference	0	
df	76	
t Stat	-2.498691812	
P(T<=t) one-tail	0.007312224	
t Critical one-tail	1.665151353	
P(T<=t) two-tail	0.014624447	
t Critical two-tail	1.99167261	

Again, the t-test doesn't prove that the new website is better than the old website. The reason for this is because t stat value (-2.498691812) is still larger than the t critical two-tail (1.99167261). It also means that I can't disprove my null hypothesis. Therefore, I'll have to carry out another test.

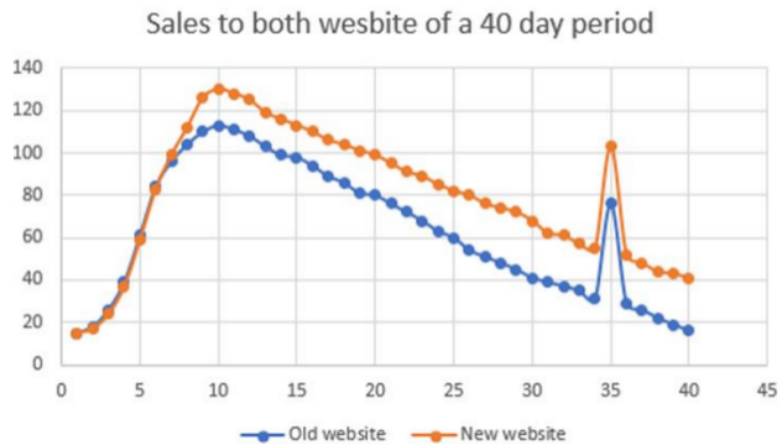
Below is a screenshot of the result from another t-test with the Alpha set to 0.01 instead.

t-Test: Two-Sample Assuming Unequal Variances		
	15	15
Mean	64.30769231	81.69230769
Variance	952.0607287	935.7975709
Observations	39	39
Hypothesized Mean Difference	0	
df	76	
t Stat	-2.498691812	
P(T<=t) one-tail	0.007312224	
t Critical one-tail	2.376420376	
P(T<=t) two-tail	0.014624447	
t Critical two-tail	2.642078313	

Once again, the t-test doesn't prove that the new website is better than the old website. The same reason applies the t stat value (-2.498691812) is still larger than the t critical two-tail (2.642078313). With another result like this mean that I can't disprove my hypothesis again.

Outlier

After completing these 3 test and having t critical two-tail larger than the t stat value. However, there is one thing which I could change which could provide me with the correct result. The thing I could change is the outlier in this data set. On day 35 there was a spike in sales. This could have been a software malfunction, some people being interested in the website or an error within the data transmutation. Whatever, the case I will be removing and redoing the t-test as this could have an effect on the data.



Below is a screenshot of the t-test without the outlier (day 35).

t-Test: Two-Sample Assuming Unequal Variances		
	15	15
Mean	64	81.13157895
Variance	974	948.4957326
Observations	38	38
Hypothesized Mean Difference	0	
df	74	
t Stat	-2.408554612	
P(T<=t) one-tail	0.009252998	
t Critical one-tail	1.665706893	
P(T<=t) two-tail	0.018505997	
t Critical two-tail	1.992543495	

7

Even with the outlier removed the t stat value (-2.408554612) is larger than the t critical two-tail (1.992543495). This mean that I can not disprove my null hypothesis even with the outlier removed.

8

Conclusion

In conclusion I wasn't able to reject my null hypothesis which means that the both websites aren't the same. But I also wasn't able to prove my own hypothesis which means that the new website isn't better than the old website, which is surprising considering the data and the scatter graph show the new website being better. Overall, the data which the company has provide isn't conclusive enough for me to come to a decision. The t-test which I carry out couldn't prove that one website was better than another. Which mean the company would need to make a decision on their own terms rather than using the data which they gave collected.

9

REFERENCES

- [1] W. Kenton, "T-Test," Investopedia, 22 March 2020. [Online]. Available: <https://www.investopedia.com/terms/t/t-test.asp>. [Accessed 29 May 2020].

BIBLIOGRAPHY

- [1] W. Kenton, "T-Test," Investopedia, 22 March 2020. [Online]. Available: <https://www.investopedia.com/terms/t/t-test.asp>. [Accessed 29 May 2020].

Experimental Data Analysis (Harry Morris)

GRADEMARK REPORT

FINAL GRADE

48/0

GENERAL COMMENTS

Instructor

Student FEEDBACK FORM

This section details the *extent* to which the assessment criteria are demonstrated by you, which in turn determines your mark. The marks available for each category of skill are shown. Lecturers will use the space provided to comment on the achievement of the task(s), including those areas in which you have performed well and areas that would benefit from development/improvement.

Generic Assessment Criteria	Marks available	Marks awarded
1. Engagement with Literature Skills		0
	0	0
2. Knowledge and Understanding Skills		0
	0	0
3. Cognitive and Intellectual Skills		0

Descriptions of the dataset. You have included some visualisations to show the population, but you have not included the data descriptive analysis from Excel which should be included to draw some initial conclusions, mean, median, standard deviation. You did not mention that the new website was offering a 15% discount which may have skewed the figures. This should be included to show how you develop the hypothesis.	10	3
Development of appropriate hypotheses for the analysis. You have included a good statement about the null and alternate hypothesis. This shows that you are able to present, evaluate and interpret qualitative and quantitative data, in order to develop lines of argument and make sound judgements in accordance with basic theories and concepts of statistical analysis.	10	6

Selection (and justification) of appropriate techniques to perform the analysis. You have used appropriate testing methods to perform the analysis. You should mention that the tests become – more ‘stringent’, by altering the level of significance. (Alpha) You also do not explain why you chose a two tailed test over a one tailed test? Or why you chose the two sample with unequal variances!	15	8
Consideration of outliers and other factors. You have considered the outliers and how they may have occurred. To improve here you need to expand on how this could skew the overall results. Does removing the outliers give a significant difference between the means?	10	5
4. Practical Application Skills		0

Application of selected analysis techniques. You have performed all the selected tests (T-Test two tailed) accurately, and given consideration the different alpha levels. However, you have not explained what the alpha levels actually signify.	20	11
Interpretation of the results of the analyses. You have not interpreted all your results correctly. There is some confusion with the negative figures and you have not interpreted the lower critical value of the two tails. You have taken into account the various alpha levels and then compared the tests with the outlier data removed. Although you only performed one test with the outlier removed and did not indicate the alpha level used. There was also a question of the number of observations. There should have been 40.	15	6

Presentation of the results in both textual and, where appropriate, diagrammatical form. You have included some scatter charts to support your discussion. To improve you should have included some other graphs to show the means etc – box and whisper for example.	10	4
5. Transferable Skills for Life and Professional Practice		0
Presentation of the paper in an appropriate professional format. Correct use of technical English. Suitable labelling and referencing strategies. You have demonstrated that you can communicate the results of your study/work adequately although not always accurately., You show some structured and coherent arguments. Overall the report was well structured, a few spelling mistakes. To improve, You should proofread your work and you should also give more consideration to the labelling of visualisations and referencing them – Figure 1 for example	10	5
		48

Assessment Mark (Assessment marks are subject to ratification at the Exam Board. These comments and are to give feedback on module work and are for guidance only until they are >

PAGE 1

PAGE 2

PAGE 3

QM

Good

Good

PAGE 4

QM

Label

Please label your figures and use this label as a reference in your text. e.g. Figure 1

QM

Label

Please label your figures and use this label as a reference in your text. e.g. Figure 1

PAGE 5



Comment 1

Spelling -

Singular - hypothesis

Plural - hypotheses

QM

Good

Good

QM

Good

Good

PAGE 6



Comment 2

Why did you choose this test?

PAGE 7



Comment 3

You should indicate that you have removed the outlier in this test!



Comment 4

Because this is a two tail test - you need to ask -

is $T < T_{\text{critical two tail}}$

OR $T > t_{\text{critical two tail}}$

to take in the lower and upper critical values

so IF $--- -2.49..... < 1.66$ Or $-2.49 > 1.66$



Comment 5

Not larger

PAGE 8



Comment 6

NOT

However - from the above comment

if $t < -ve \text{ crit 2 tail}$ Or $t > +ve \text{ crit 2 tail}$ then you cannot reject the null hypothesis - the difference between the 2 means isn't convincing enough.

$-2.49 > -2.64$ so you can reject the null hypothesis

PAGE 9



Comment 7

?? there were 40 observations - why have you gone from 39 to 38?



Comment 8

What value did you use for alpha? 0.1, 0.01 or 0.5



Comment 9

what about the 15% discount?