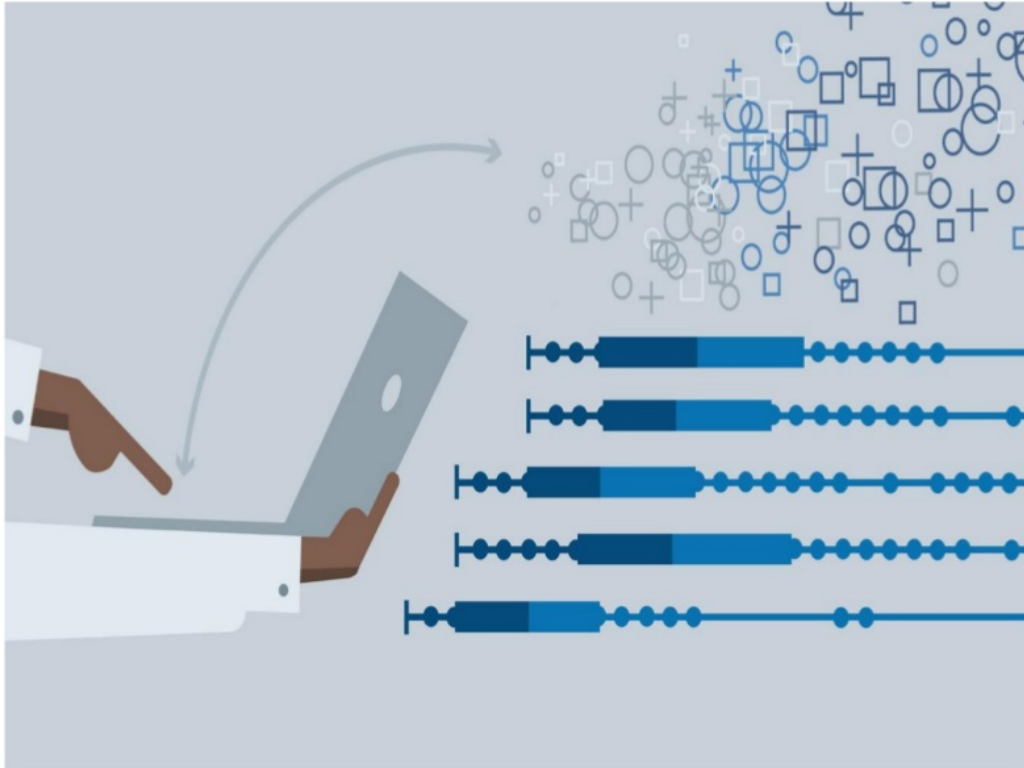


Assignment 2

by Thomas Howells

FILE	8112_THOMAS_HOWELLS_ASSIGNMENT_2_192216_701287817.DOCX (251.74K)		
TIME SUBMITTED	18-MAY-2020 03:13PM (UTC+0100)	WORD COUNT	1456
SUBMISSION ID	128100609	CHARACTER COUNT	6743



Internet Technology and Security



DATA ANALYSIS

Thomas Howells | HND Applied Computing | 17/5/2020

CONTENTS

Introduction 2

Initial Thoughts 3

..... 4

.....5

.....5

Alternative Hypothesis 6

Null Hypothesis 7

..... 8

Tests 9

Conclusion.....11

Appendices and References 12

INTRODUCTION

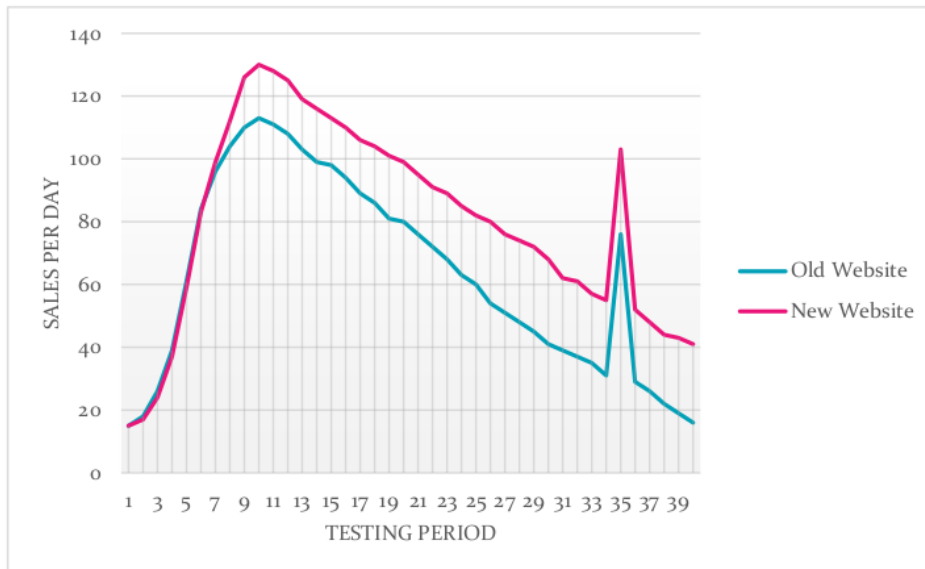
This analysis is going to cover the analysis of the data provided, a hypothesis developed by the student, and a suitable test to test the hypothesis made by the student. An online retail company made some changes to their website design and they measured the impact it had on their sales, they recorded the sales on their new website and their old website and compared them in a table using A/B testing. In order to encourage people to come to their website they have added 15% off all orders if customers buy during the testing phase. This testing phase lasted 40 days. From the information they have given, they would like to know whether it is worth updating their webpages design or leaving it as is. The student will give their input on the matter and provide evidence for why they should keep or change their websites design.

INITIAL THOUGHTS

First of all, the data provided over the 40 days must be added up to give definitive answers on which website sold the most over the testing period.

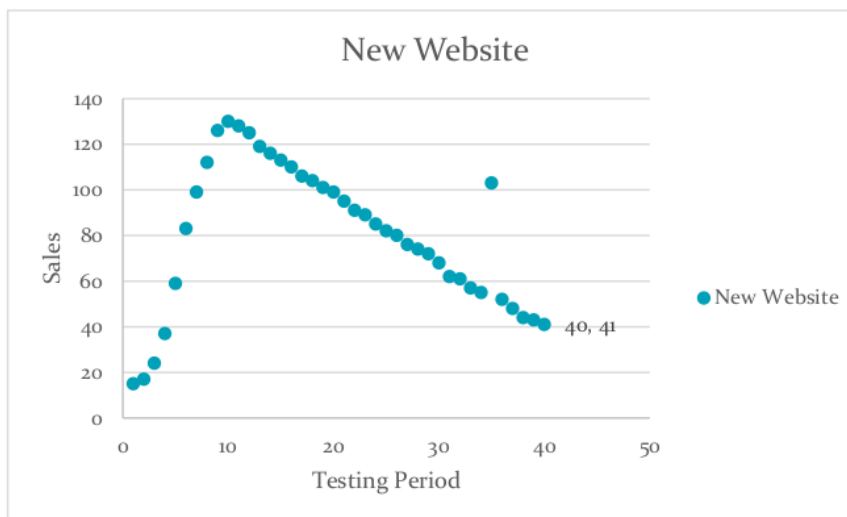
The total amount of old website sales gives an answer of = 1069 Total Sales. This was done by taking the Old Website sales from the 40 days testing period table and putting the information into an excel spreadsheet. Once in excel, the information was put into column and it was highlighted, and the button 'AutoSum' was clicked to add all of the numbers together and find the answer. A calculator could have been used for this, but it would have taken some time and would have become tedious. The same strategy would be used to find the total sum of the sales from the new website. All the sales put into excel and the 'AutoSum' button would tell the student the answer. The answer of the New Website sales = 3201. [2] [3]

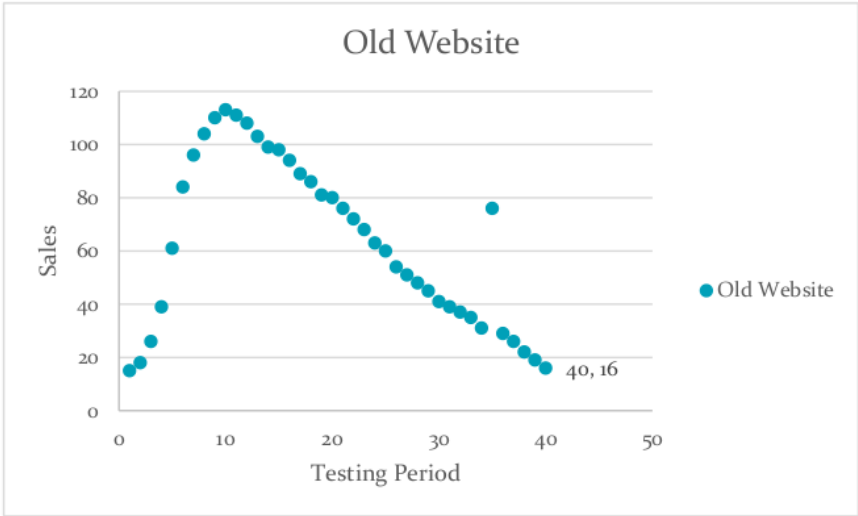
It can be seen from the information developed by the company that at the start of the of the testing there were 15 sales on the first day for both the Old Website and the New Website. It can be assumed that day 1 was that start of a month and not many people had been payed yet that is why the sales were so low. But as this is not stated this will not be spoken about. If the student were to know this, it would make analysis considerably easier.



Label

It can be seen that both the Old and New Websites follow the same increase in sales up until the 7 day mark. From here the New Website starts to sell considerably more at its peak. Both then start to decrease in sales but the New Website is still selling more than the Old Website. When the testing reaches the 35 day mark both the New and Old website have a spike of sales, this is an anomaly in the data. This can be seen in a scatter graph too.





	Old	New
Mean	63.075	80.025
Median	62	82.5
Mode	26	99
Standard Deviation	31.43931	31.98436
Sample Variation	988.4301	1022.999
Count	40	40
Range	98	115
Max	113	130
Min	15	15

Label

It can be seen from the table that there is an increase by 17 within the Mean

3

ALTERNATIVE HYPOTHESIS

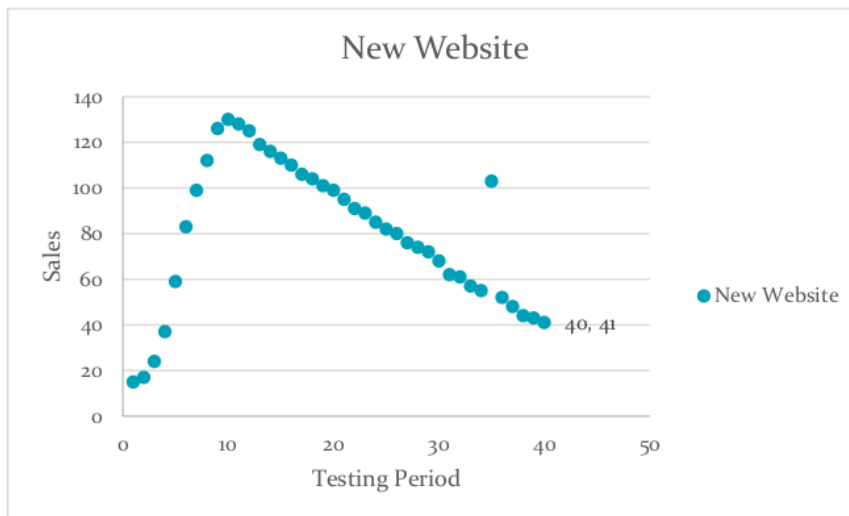
The hypotheses for this set of data made by the student is, "Would the website have doubled the sales if it wasn't for the 15% discount or was the reason the new web design?"

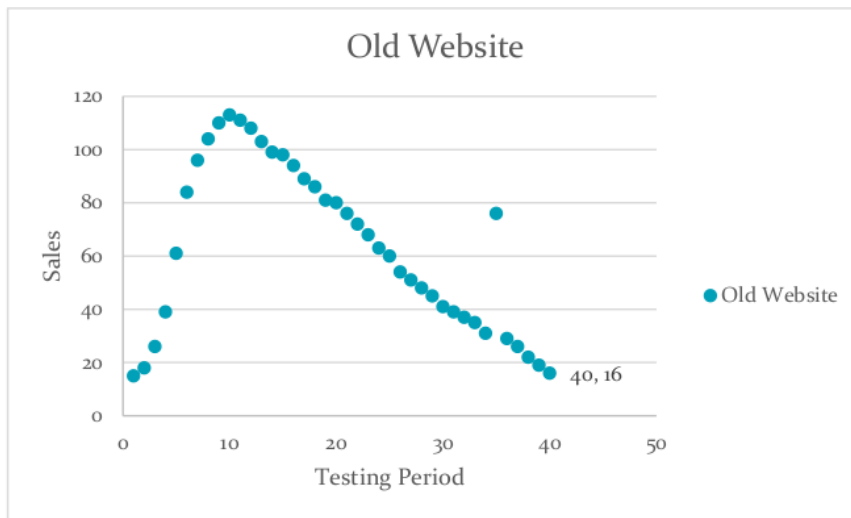
It can be argued from both perspectives that they are the sole reason for the company doubling their sales. It can be seen from the table provided that the New Website had a significant increase in sales everyday over the Old Website. A better way of comparing the traffic and sales of the Websites would have been to eradicate the 15% discount as it makes it harder for analysers to pinpoint what the reason was for the increase in sales and traffic. It would have been better to first add the discount to the Old Website for a certain amount of days then stop the discount, then release the New Website for a certain amount of time without the discount and see what sales are like. After this period of time is up then release the discount on the website. This will then show more valid information and data towards the reason for the sales and traffic. A new Website can look appealing and clean, this will definitely bring traffic to the page as it will give consumers a natural attraction to feel comfortable whilst exploring the page. It states on the productiveinbox Website that "Bad quality visuals, Colour scheme, Mobile devices, Slow load times" can be the reasons for. It will push customers away if the page were to be unattractive and this is not good from a business point of view, this is usually why big companies with plenty of money will have a sleek and attractive look.



NULL HYPOTHESIS

Within the Null Hypothesis the student is going to explain that the Old Website is worth keeping and the New Website is time and effort no better outcome. It could be argued that the main reason for the New Website doing so well is due to the fact that there is a 15% discount on the product they are selling. Maybe from the data provided, the lower the price in a product, the more people are going to be attracted towards this. This is where the Quality vs Quantity matters. There would need to be a balance between both. There is not much information to back this hypothesis, there all only the data the student have received and the tables being created from the data to back the null hypothesis.



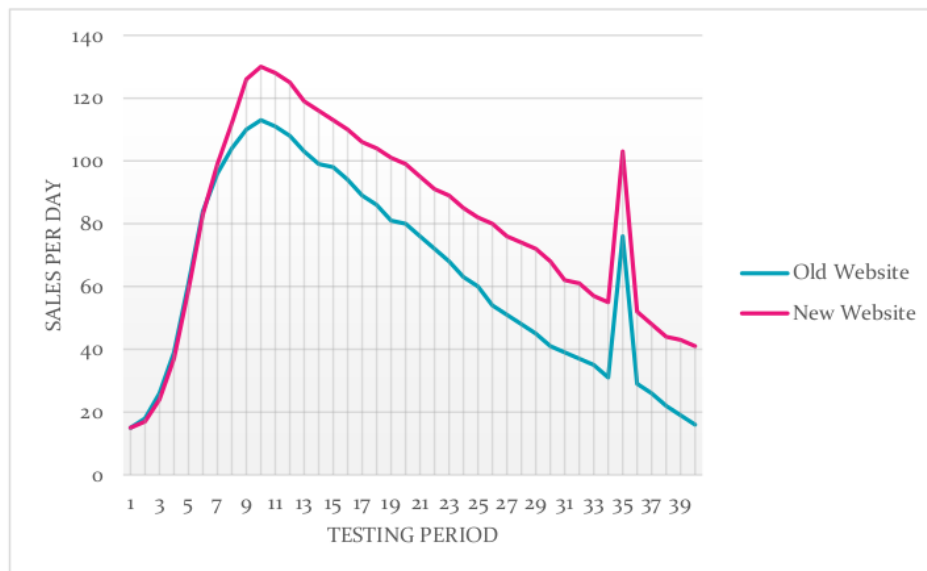


As it can be seen, both Websites follow the same trend. The New Website is higher than the Old Website. From here it can be argued that the number of sales were significantly higher on the New Website only because of the discount.



TESTS

In order to get correct information more tests need to be done. What if the 40 day test period was just coincidence? To prove whether it was or not T tests have to be done. The significance levels mentioned in the assignment brief were 0.10, 0.02 and 0.01, these are known as P values. 0.10 representing 90%, 0.02 representing 95% and 0.01 representing 99%.



There must be a reason for the spike around 33-37 days in the graph. This spike in information could mean there may have been a software malfunction causing this. It could mean that multiple people got payed roughly the same time and decided to buy on the same set days It could be an error in the data transmission. In order to get rid of the outlier in the data, it must be removed from the table provided in the assignment brief. Once the information is deleted the student will complete more T-tests to show the most information valued one.

	Variable 1	Variable 2
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	2.088663468	
P(T<=t) two-tail	0.019248306	Label
t Critical two-tail	2.375110958	

This information was collected from T-testing before removing the outlier at P value 0.02.

	Variable 1	Variable 2
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	2.088663468	
P(T<=t) two-tail	0.019248306	Label
t Critical two-tail	2.375110958	

This information above was done with the same P value of 0.02 but the outlier had been removed. As it can be seen there is absolutely no change from removing the outlier and creating the two variable T-tests. This proves if I were to do with P values of 0.01 and 0.10 they would also not change if I were to get rid of the outlier the information would display the same as before. To stop errors like this occurring, the testing period could be extended. The data could be tested inhouse so there is no mix up with communication and data.

CONCLUSION

The student was unable to reject the null hypothesis due to there being insufficient evidence to help back it. The initial hypothesis was also not proven true as there was insufficient evidence to prove true without assuming, assuming is not a clear indication if something is true or false, therefore it cannot be said that whether it was true or false or the student would be lying. The data provided was not conclusive enough to come to a decision. Had testing period been extended it would give the student more figures to work with and hopefully come to a conclusion of what the reality of the matter is.

APPENDICES AND REFERENCES

[1]- The front cover display picture:

<https://www.lynda.com/Business-Intelligence-tutorials/Data-Visualization-Tips-Tricks/475454-2.html>

[2] The Old Website 40 day period sales:

	94	
	89	
	86	
	81	
	80	
	76	
	72	
	68	
	63	
	60	
	54	
	51	
	48	
	45	
	41	
	39	
	37	
	35	
	31	
	76	
	29	
	26	
	22	
	19	
	16	
Total =	1069	

[3] The New Website 40 day period sales:

80		
76		
74		
72		
68		
62		
61		
57		
55		
103		
52		
48		
44		
43		
41		
3201		

[4] How a Professional website can attract more customers-

Productiveinbox.co.uk. 2020. *Productive Inbox I How A Professional Website Can Attract More Customers*. [online] Available at: <<http://www.productiveinbox.co.uk/news/how-a-professional-website-can-attract-more-customers>> [Accessed 17 May 2020].

Assignment 2

GRADEMARK REPORT

FINAL GRADE

53/0

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 given a good written description of the data, and I'd have included the data descriptive analysis from Excel. This, has enabled you to draw some initial conclusions, mean, median, standard deviation. This helps you develop the hypotheses. You have also used a variety of visualisations to support your text	10	7
Development of appropriate hypotheses for the analysis. You have shown that you have developed hypotheses, but there is some confusion in your text as to what it really means. Your commentary on hypotheses should look at the Null hypothesis first. H_0 - the websites are the same. The alternative hypothesis (H_a) is that they are different, you don't need to add the 15% discount as a factor here - You are performing t-tests to help you determine if there is a significant difference in the means to either reject your null hypothesis or not.	10	5.5

Selection (and justification) of appropriate techniques to perform the analysis. You have selected and justified testing methods to perform the analysis. You should mention that the tests become – more ‘stringent’, by altering the level of significance. (Alpha)	15	8
Consideration of outliers and other factors. You have considered the outliers and how they may have occurred. Although the 15% discount wasn’t really a factor as both websites offered the 15% discount.	10	4.5
4. Practical Application Skills		0

Application of selected analysis techniques. You have performed some of the selected tests (T-Test two tailed) accurately, but not considered the different alpha levels. You should explain what the alpha levels actually signify. In all there should have been 3 tests each for the data and another 3 when the outlier is removed. At the alpha levels 0.1, 0.05, 0.01 (10%,5% and 1%) - You used an alpha level of 0.02 with no explanation as to why.	20	9
Interpretation of the results of the analyses. You have not interpreted your T-test results correctly. The two t tests you performed were obviously the same. It is important to review your work carefully. Because I am not convinced that you used the right sample data, I am not sure you have conclusions are correct.	15	6

Presentation of the results in both textural and, where appropriate, diagrammatical form. You have included some good description of the figures and have supported your text with appropriate visualisations, however, you could have included other visualisations to support your text –box and whisper for example.	10	6
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 very well. You include some structured and coherent arguments. Overall the report was well structured. To improve, you should give more consideration to the labelling of visualisations and referencing them – Figure 1 for example	10	7

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



Comment 1

?

PAGE 2

PAGE 3

PAGE 4



Comment 2

considered

PAGE 5



Label

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

PAGE 6



Label

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

Additional Comment

You should mention that this data is taken from the descriptive statistics in Microsoft Excel



Comment 3

Does this mean anything? is it a significant difference? - you won't know until you have performed the t-test - that is what the test is for.

PAGE 7



Comment 4

both websites had the 15% discount.

Visitors didn't know what website they were looking at - old or new.

PAGE 8

PAGE 9



Comment 5

You commentary on hypotheses should look at the Null hypothesis first. That is the websites are the same. The alternative hypothesis is that they are different, You don't need to add the 15% discount as a factor here - You are performing t-tests to help you determine if there is a significant difference in the means to either reject your null hypothesis or not.

PAGE 10



Comment 6

it isn't a coincidence - it is the sample population you have been given.



Comment 7

0.5



Comment 8

day 35 only

PAGE 11

QM

Label

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



Comment 9

They are exactly the same - 40 observations

QM

Label

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



Comment 10

I disagree - you need to perform the tests at all alpha levels. Also 0.05 not 0.02!

PAGE 12

PAGE 13

PAGE 14
