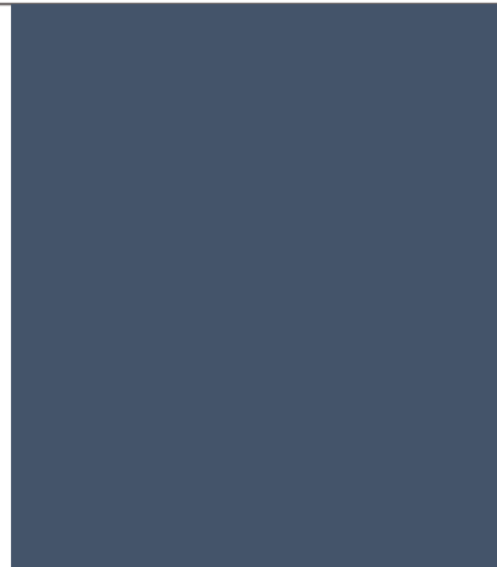


Experimental Data Analysis

by Jamie Gau

FILE	14940_JAMIE_GAU_EXPERIMENTAL_DATA_ANALYSIS_192216_88252999 0.DOCX (46.86K)		
TIME SUBMITTED	17-MAY-2020 11:12PM (UTC+0100)	WORD COUNT	1970
SUBMISSION ID	127990900	CHARACTER COUNT	9282



Experimental Data Analysis

ACBA4002 – Data Analysis and Visualisation

James Gau



Contents

Introduction/Abstract	2
The Data Set in Question	2
A Description of the Data Set.....	3
Developing a Set of Hypotheses/Suitable Hypothesis Tests.....	4
Testing the Hypotheses.....	4
Significance Level: 0.10	4
Significance Level: 0.05	5
Significance Level: 0.01	5
Data Outliers and How they may Affect the Results	6
Changing Factors that may have Affected the Results/Conclusion	6

Introduction/Abstract

This report will contain an analysis of a set of data that has been collected by a company. This analysis will include a description of the data in question (along with descriptive statistics), then, a series of hypotheses with appropriate tests and justifications of why these particular tests were chosen. After that, these tests will be conducted with each hypothesis to varying significance levels and a report on which hypothesis is accepted in each case. Data outliers and anomalies will be discussed also, as well as what may have caused them, and investigating how the results would change if they were to be removed. Finally, other factors that may have affected the results of the experiment will be discussed, and how you might control them.

This report is purely informational and will try to be as unbiased as possible. Its sole purpose is to deliver information about the results of the experiment, not personal opinion.

The Data Set in Question

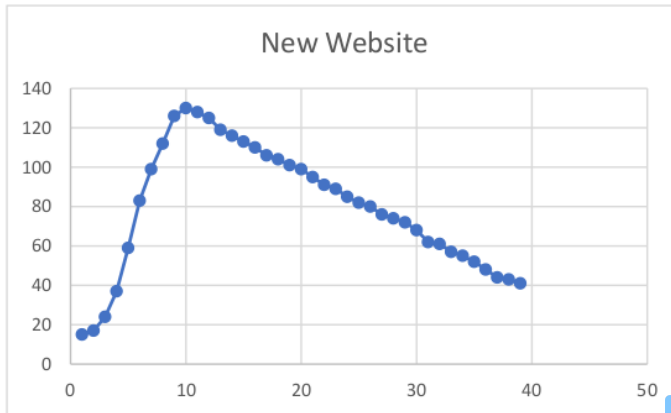
(This is a direct quote from the assignment brief:)

“An online retail company decides to make some major changes to its website and measure the effect that this has upon sales. They arrange for their web server to show each visitor to their website either the original design or the new design (a process known as A/B testing). They then record how many sales were made through the old website and how many were made through the new website during a 40-day test period. In order to encourage people to come to their website the company offers a 15% discount on all orders made through the website during the test period. At the end of the test period they want to know whether they should keep the old website design or move to the new website design. The number of sales for each website each day is shown below:”

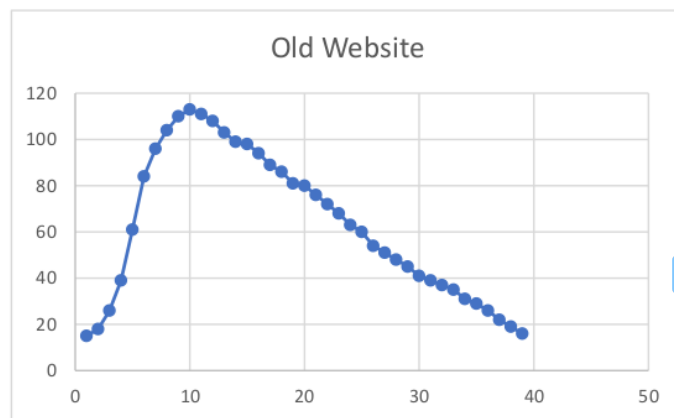
Day	Old website	New website	Day	Old website	New website
1	15	15	21	76	95
2	18	17	22	72	91
3	26	24	23	68	89
4	39	37	24	63	85
5	61	59	25	60	82
6	84	83	26	54	80
7	96	99	27	51	76
8	104	112	28	48	74
9	110	126	29	45	72
10	113	130	30	41	68
11	111	128	31	39	62
12	108	125	32	37	61
13	103	119	33	35	57
14	99	116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

A Description of the Data Set

The data that is represented in this table is the number of users to 2 different websites over the course of 40 days. The data from the old website and the new website were sorted into graphs using Microsoft Excel:



Label



Label

As you can see, the graphs do not differ much in regard to pattern. Both results produce graphs that start rapidly increasing, then slowly start decreasing as the days go by, both also have some strange anomaly at the day 35 mark, a rapid spike in website users (See Data Outliers). Where they do differ, however, is the total amount of visitors each site brought in.

1

Old	2523
New	3201

The old website brought in a total of 2523 users, whilst the new one brought in 3201 users. Whilst the old website brought in more users for the first 6 days, the new website caught up and started surpassing the old website on day 7. It then kept the higher user amount right up until the end of the experiment. This might mean that the new website is more successful, however it is uncertain to say at this point. To find out truly how effective each site is, hypotheses will need to be developed and tested, after which we can give a conclusive answer as to which website produces better results.

2

Developing a Set of Hypotheses/Suitable Hypothesis Tests

Certain criteria must be met in order to make sure that the hypothesis you have developed is testable. One, you must make sure that there is a possibility that your hypothesis can be proved true. Two, similarly, you must make sure that there is a possibility that your hypothesis can be proved false. Third, finally, you must make sure that your hypothesis has repeatable results. This last step is the hardest, as there is a chance that your results aren't repeatable. As well as developing a hypothesis, a null hypothesis must too be developed:

$$H_0 : \mu_1 = \mu_2$$

Good

In other words, the null hypothesis is that there will be no difference between the websites. This hypothesis works because it can be proven either true or false. In fact, it has already been partly proven false in the data description section. For argument's sake, here is the mean of the two sets of data, simply calculated using the Microsoft Excel Average Auto sum:

Old	63.075
New	80.025

So, we can see already that the null hypothesis has been disproven, but there is a slight problem. The null hypothesis hasn't been fully disproven, as we don't know if the results are by chance. To counter this, we use a T Test. This will show the statistical significance of our results, so we can come to a conclusion on whether or not they are valid.

To sum it up, we have 2 hypotheses and 1 null hypothesis:

- Hypothesis 1 – The old website has the highest mean number of visitors.
- Hypothesis 2 – The new website has the highest mean number of visitors.
- Null Hypothesis – The websites are the same.

3

These 3 hypotheses will be tested using T Tests, whilst also assuming unequal variances and having the hypothesized mean difference set to zero, in order to check the statistical significance and find out if the results are by chance or if they are conclusive enough. The hypotheses will be tested using varying significance levels (0.10, 0.05 and 0.01) in order to see if the results are repeatable.

Testing the Hypotheses

Significance Level: 0.10

	Old	New
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	1.292790268	
P(T<=t) two-tail	0.014624447	
t Critical two-tail	1.665151353	

This T Test was done by gathering both sets of data and changing the significance level to 0.10, before displaying them as a table in a separate work sheet. As you can see, there is a clear difference between the means of the old and new websites. In order to see if we can reject the null hypothesis, we have to do what is known as a two-tail test. If t stat is greater than t critical two-tail, then we can reject the null hypothesis. This is not the case, as -2.498691812 is not greater than 1.665151353, so the observed difference between the sets is not convincing enough to say that the websites differ significantly.



Significance Level: 0.05

	<i>Old</i>	<i>New</i>
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	

The same was then done with a significance level of 0.05 this time. The results are much the same, with exception to the t critical one-tail, t critical two-tail, as well as some others. Again, if t stat is greater than t critical two-tail, then we can reject the null hypothesis. This is, yet again, not the case, as -2.498691812 is not greater than 1.99167261, so the observed difference between the sets is not convincing enough to say that the websites differ significantly.

Significance Level: 0.01

	<i>Old</i>	<i>New</i>
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	

Finally, a significance level of 0.01 was tested to see if the null hypothesis was able to be rejected. Once again, if t stat is greater than t critical two-tail, then we can reject the null hypothesis. This is, yet again, not the case, as -2.498691812 is not greater than 2.642078313, so the observed difference between the sets is not convincing enough to say that the websites differ significantly.

Data Outliers and How they may Affect the Results

As mentioned earlier in the data description section, there is an outlier in the data. Specifically, in between days 34 to 35. Both websites massively spike in viewer count, with the old website gaining 45 more viewers and the new website gaining 48 more viewers, before both websites stabilize back down to a more reasonable amount.

One can speculate many reasons why this may have happened. It could just be a simple problem, such as the program recording the number of viewers could have malfunctioned. It could be that a bunch of users saw something on social media, for example, that they like and the website just happened to fit the niche market. There may have been an error in data transcription or transmission, again relating back to the malfunction theory.

There are many different causes of data outliers, but what's most important is how to deal with them. One thing you can do is delete them entirely, making your data be much more average, though this is risky, as the data might not be an accident, therefore your results won't be fully accurate. Another thing you can do with data outliers is to 'winsorize' them, meaning to change the number to a much more reasonable one. Another test is going to be run on the two data sets using a significance level of 0.05. However, this time the anomalous result will be removed from both data sets completely. Here are the results of this test:

	<i>Old</i>	<i>New</i>
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	



The means of the sets have completely changed, as to be expected, and the rest of the results are also different. For the final time, if t stat is greater than t critical two-tail, then we can reject the null hypothesis. This is not the case, as -2.408554612 is not greater than 1.992543495. Therefore, even with the anomalous result removed, the observed difference between the sets is not convincing enough to say that the websites differ significantly.

Changing Factors that may have Affected the Results/Conclusion

The factors that may have affected the results are vague, so it is difficult to say if anything could have been done to change the results of the tests. Perhaps multiple programs could have been run so as to detect any discrepancies with the software early on. They could have chosen a busier time of the year or made their website more accessible so as to get a bigger population sample.

Unfortunately, there is not enough information to know exactly what went wrong and how you would go about fixing it in order to gather different results. In the end, the null hypothesis couldn't be rejected on any of the tests, therefore you can't say for certain which website yielded better results. Maybe if the test was run again with different circumstances, the results may be different.

Experimental Data Analysis

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, but you could have included the data descriptive analysis from Excel. You should have included this, to draw some initial conclusions, mean, median, standard deviation. This helps you develop the hypotheses. You have also used some other visualisation of the data to support your text. Line graph, but it doesn't include the outlier.	10	4
Development of appropriate hypotheses for the analysis. You have included a good statement about the null hypothesis. You have shown a good evaluation of the underlying concepts and principles and that you can interpret qualitative and quantitative data, with minor errors. An emerging ability to use evidence to support the argument. Mostly valid arguments and logical judgements.	10	6.5

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. However, you did not to show if removing the outliers gave a significant difference between the means?	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, and 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%) - I am not convinced that the correct figures were used as the sample size before the outlier was 39 and should have been 40. The graph did not show the anomaly in the data	20	10
Interpretation of the results of the analyses. You have interpreted your T-test results correctly. However, you need to consider the negative side of the two-tails not just the positive. This would have given you more data with which to make your final conclusion. Because I am not convinced that you used the right sample data, I am not sure you have conclusions are correct.	15	7

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

PAGE 2

PAGE 3

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



Comment 1

Your graphs do not show the outlier on day 35!



Comment 2

You should include the descriptive statistics from excel - this would give you means, mode median and standard deviation, which you should also be commenting on

PAGE 5

QM

Good

Good



Comment 3

you should say

Null hypothesis is that they are equal

Alternative hypothesis is that they are different. (it doesn't matter which way, just that there is a significance difference between the means to reject the null hypothesis.



Comment 4

Because there are two tails you should also test

t Stat < -ve critical two-tail

which it is

-2.4986... < -1.9916...



Comment 5

The original sample had 40 observations!