

2014825_DAV_A2

by Carys Newing

Submission date: 18-Jan-2022 10:46PM (UTC+0000)

Submission ID: 169533343

File name: 24850_Carys_Rebeca_Newing_2014825_DAV_A2_394034_9301436.docx (169.25K)

Word count: 2475

Character count: 10816

Experimental Data Analysis – Standard Teaching vs Learning Applications

Module Code: ACCB4002

Carys Newing

Student ID: 2014825

Abstract

This report aims to identify whether the mathematics learning app provided by an online learning application development company, when used to supplement a student's studies, results in an improvement in the scores achieved on tests. Two sets of data have been provided to identify if there are any overall improvements in the scores achieved, the sets of data provided are: test score before using the app and test score after using the app.

The methods that will be used to identify if there are any changes to the before and after scores will be: using the sample means to identify the average scores, using the Chi-Squared test to find the p-value of the sample so that they can be compared to the set significance levels of 0.10, 0.05, 0.01.

Any anomalous data will be identified, and the tests repeated to check whether they have any impacts on the results.

The null hypothesis is that there is no change to the data results and the mean of the data will remain the same. The alternate hypothesis is that there is a change in the results, and that the mean of sample 2 will be greater than the mean of sample 1.

Based on the findings and the results from using the means of the samples alongside the Chi-Squared values and the p-value generated from these figures, we fail to reject the null hypothesis despite evidence for the alternate hypothesis being present.

Should the anomalous results be removed, the p-value changes and alters the findings when compared to the proposed significance levels. When the significance level is set to 0.10 or 0.05 the null hypothesis is rejected as there is evidence for the alternate however for a significance level of 0.01, we fail to reject the null hypothesis.

Contents

Abstract	2
Experimental Data Analysis – Standard Teaching vs Learning Applications.....	4
Hypotheses	4
Testing	5
Chi Squared.....	6
Significance Levels	7
Outliers and Anomalies	8
Experiment Variables	11
Conclusion	11
Appendices.....	13
<i>Appendix 1 – Class Test Results</i>	13
<i>Appendix 2 – Class Test Results – Percentage (Decimal)</i>	14
<i>Appendix 3 – Sample Means</i>	15
<i>Appendix 4 – Outliers</i>	16

Experimental Data Analysis – Standard Teaching vs Learning Applications

The aim of this analysis is to identify whether the math learning app that was provided to the school as a learning tool has resulted in an improvement of test scores and thus the improvement of the students understanding of the materials. For this sample 40 students have had their scores recorded before using the app and after using the app to see if there are any improvements in the result.

Hypotheses

The main element of the analysis is to identify whether there has been any changes in scores between the standard teaching and the online math application. The figures for the mean test scores, as shown in appendix 3, will be the baseline of the null hypothesis.

The mean test scores have been gathered from comparing all the results from before the app and after the app. As the test prior to the application is the baseline for the analysis, this is the mean test score that will be used to identify whether the achievements of the students are greater prior to the application then after.

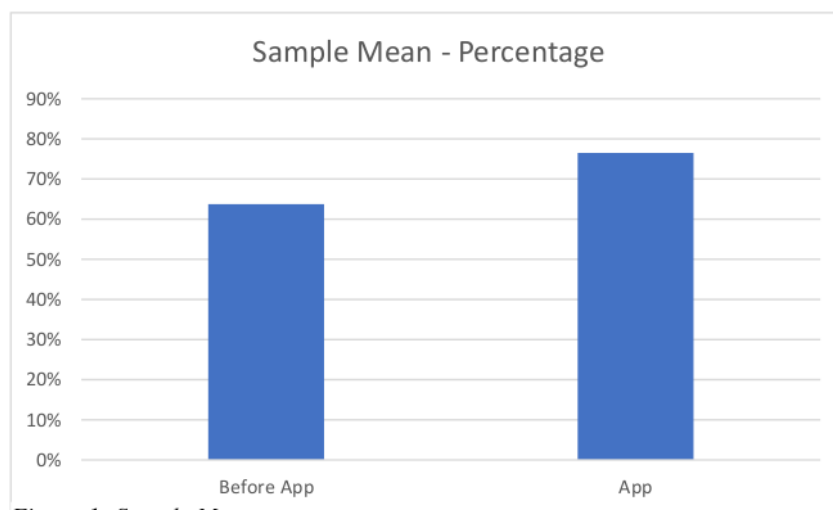


Figure 1- Sample Means

As shown in the image above the sample mean for both sets of data, the mean of the students before the app will be considered as the base line as they have received the standard teaching material.

Where μ is the mean of the test scores of the students, the hypotheses are as follows:

Null Hypothesis (H_0): $\mu = 64\%$

Alternate Hypothesis (H_a): $\mu \geq 64\%$

The significance testing which will be referred to as α is used to identify whether the H_0 is rejected, and the H_a is accepted or whether the significant level is less than or equal to the p-value resulting in the failure to reject the H_0 although this does not mean that the H_0 is accepted, more so the H_0 or the H_a cannot be proven with the data available. The α values that will be compared to the null hypothesis and alternate hypothesis are 0.10, 0.05 and 0.01.

Testing

Due to the quantity of the data gathered, the t-test has been ruled out and the data has been quantified into the following categories to use in the Chi Squared test. The Countifs function in excel was used to categorize the data appropriately.

Test Results	50% or less	51 -75%	76-100%
Before App	12	16	12
After App	6	14	20

Figure 2- Test 1

Chi Squared

The test results data, as shown in Figure 2, can be used within the Chi Squared (X^2) Method as the minimum value of 5 has been exceeded.

$$x^2 = \sum \frac{(O - E)^2}{E}$$

From using the X^2 method, the expected distribution of the results shows that there should be a larger proportion of students achieving 51% or greater.

	50% or less	51 -75%	76-100%
before	9	15	16
after	9	15	16

Figure 3 - Expected Distribution

Between the Observed Data and the Expected Data, the X^2 method can be used to get the X^2

value by utilizing $\frac{(O-E)^2}{E}$

The data below has been created from the observed data minus the expected data, squared then divided by the expected data value.

	50% or less	51 -75%	76-100%
before	1	0.066667	1
after	1	0.066667	1

Figure 4 - Chi Squared

The total sum of the figures from the before and the after rows equals the X^2 value, therefore from the values above, the X^2 value is 4.133

As the X^2 value has been worked out the Degree of Freedom and the p-value can now be obtained.

The Degree of Freedom is rows – 1 * columns – 1, which from the given figures can be worked out as:

$$\text{Degree of Freedom} = (2 - 1) * (3 - 1) = 2$$

Using either a Chi squared calculator or the Chi Squared Table the p-value of 0.127 is achieved.

Significance Levels

As a significance level of 0.05 is typically used in statistical testing and this is not a conservative test, 0.05 would be a suitable level to compare to the p-value as it would indicate whether the null hypothesis is statistically significant. However, it is worth noting that depending on the criteria required by the business the significance level can be adjusted to suit the requirements of the test therefore with these tests the Significance Levels (α) that the P-value will be tested against are 0.10, 0.05 and 0.01

With the Hypotheses of:

Null Hypothesis (H_0): $\mu = 64\%$

Alternate Hypothesis (H_a): $\mu \geq 64\%$

The α can be used to accept the H_0 or reject it with evidence of the H_a .

$$\text{if } \alpha = 0.10$$

$$\text{then } p - \text{value} < \alpha$$

$$0.127 < 0.10$$

As the p-value is greater than the α of 0.10, we fail to reject the null hypothesis.

$$\text{if } \alpha = 0.05$$

$$\text{then } p - \text{value} < \alpha$$

$$0.127 < 0.05$$

As the p-value is greater than the α of 0.05, we fail to reject the null hypothesis.

$$\text{if } \alpha = 0.01$$

$$\text{then } p - \text{value} < \alpha$$

$$0.127 < 0.01$$

As the p-value is greater than the α of 0.01, we fail to reject the null hypothesis.

Outliers and Anomalies

From reviewing the data there are four entries that could be considered outliers (as shown in Appendix 4), in which the students scored less after using the app than with the traditional materials. There is the potential that these results could have affected the overall result and resulting in a Type 1 or a Type 2 error occurring.

If the 4 anomalous results, where students scored less after using the learning application are removed from the samples bringing the total of each sample to 36 students. By altering the data used in the χ^2 method, the Chi-squared value and the p-value will be changed.

	50% or less	51 -75%	76-100%
before	11	16	9
after	4	13	19

Figure 5- Outliers Removed

This then changes the output values when comparing them to the expected values.

	50% or less	51 -75%	76-100%
before	1.633333333	0.155172	1.785714
after	1.633333333	0.155172	1.785714

Figure 6- Chi Squared - Outliers

The sum of the above items gives a Chi-Squared value of 7.148, a Degree of Freedom of 2 and a p-value of 0.028.

Therefore, with the Hypotheses of:

Null Hypothesis (H_0): $\mu = 64\%$

Alternate Hypothesis (H_a): $\mu \geq 64\%$

The α can be used to accept the null hypothesis or reject it, so if the significance limit is set to 0.10:

$$\text{if } \alpha = 0.10$$

$$\text{then } p - \text{value} < \alpha$$

$$0.028 < 0.10$$

As the p-value is less than the α of 0.10, we fail to accept the null hypothesis as there is evidence for the alternate hypothesis, if the significance limit is reduced to 0.05 then:

$$\text{if } \alpha = 0.05$$

$$\text{then } p - \text{value} < \alpha$$

$$0.028 < 0.05$$

As the p-value is less than the α of 0.05, we fail to accept the null hypothesis as there is evidence for the alternate hypothesis, if the significance limit is reduced to 0.01 then:

$$\text{if } \alpha = 0.01$$

$$\text{then } p - \text{value} < \alpha$$

$$0.028 < 0.01$$

As the p-value is greater than the α of 0.01, we fail to reject the null hypothesis.

Experiment Variables

There are several variables that could affect the results of the experiment, as the students are in effect doing the same test twice, covering the same subject. The improvements could be seen as students learning from their mistakes. To prevent this from affecting the results, it would be appropriate to have a control group and a test group – both with students with similar capabilities.

A control that could be put in place is to limit the access to the web app as students that receive the standard teaching materials or alternatively, the learning application could be used instead of the standard teaching methods rather than as a supplementary material. As there is also no indication on whether the test and the material were in relation to a specific subject within the topic of math, e.g., algebra or parabolas. As some aspects of mathematics can be easier to grasp than others.

A larger data set could also change the outcome as any improvements with either of the above variable changes and could mitigate some of the anomalous data that is found within the data sets.

Conclusion

From the evidence provided and tested against the significance levels of 0.01, 0.05 and 0.10, it is prudent to conclude that the null hypothesis is failed to be rejected. However as shown in Figure 1, the mean result for after the application is approximately 10% higher than the original results.

There is a possibility that a type 2 error could have occurred where the null hypothesis was failed to be rejected despite there being evidence to potentially support the alternate

hypothesis. As there was no control group, as it was the same set of students taking both exams – this could also indicate that the learning app was beneficial despite the failure to reject the null hypothesis. It would be recommended that additional tests be done with a control group and a test group to best convey whether the learning application was of benefit.

It is evident that there are anomalies within the data set that, should they be removed the Chi-Squared value changes as does the p-value. These changes have a significant impact on the results when compared with the original results in which all of the findings result in the null hypothesis being failed to be rejected. Whilst with the anomalies removed and the significance level is set to 0.10 or 0.05 the null hypothesis is rejected as there is evidence for the alternate however for a significance level of 0.01, we fail to reject the null hypothesis.

Overall, for a business case, there would be recommendations that the anomalies are removed as it would indicate that there is evidence for the alternate hypothesis in comparison with the original data sets where, regardless of the significance level set, the null hypothesis was failed to be rejected in each case.

Appendices***Appendix 1 – Class Test Results***

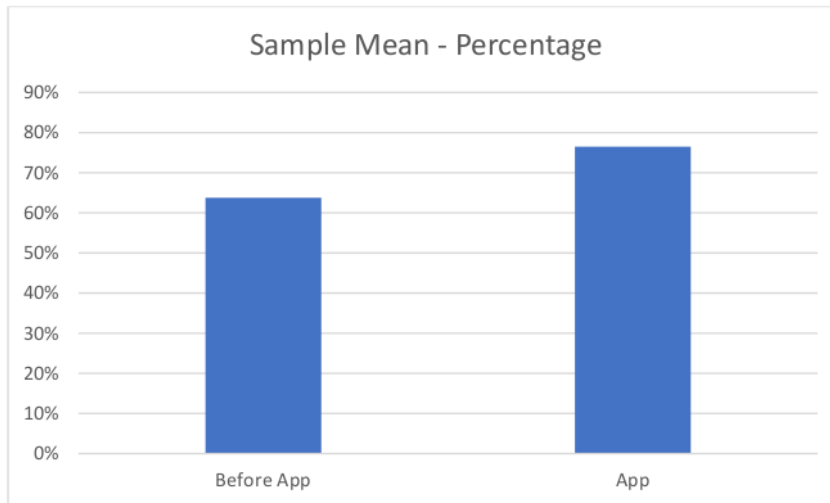
Student	Test score before using app	Test score after using app
1	12	14
2	9	9
3	8	10
4	9	11
5	12	15
6	14	15
7	15	15
8	17	10
9	18	20
10	15	20
11	14	18
12	14	16
13	14	17
14	12	14
15	10	20
16	11	15
17	8	4
18	7	10
19	6	11
20	6	19
21	5	16
22	17	17
23	19	19
24	20	18
25	17	20
26	16	19
27	17	20
28	12	14
29	13	14
30	11	13
31	12	12
32	6	9
33	8	20
34	10	17
35	15	17
36	17	17
37	17	19
38	18	20
39	12	13
40	17	15

Appendix 2 – Class Test Results – Percentage (Decimal)

Student	Test score before using app	Test score after using app
1	0.60	0.70
2	0.45	0.45
3	0.40	0.50
4	0.45	0.55
5	0.60	0.75
6	0.70	0.75
7	0.75	0.75
8	0.85	0.50
9	0.90	1.00
10	0.75	1.00
11	0.70	0.90
12	0.70	0.80
13	0.70	0.85
14	0.60	0.70
15	0.50	1.00
16	0.55	0.75
17	0.40	0.20
18	0.35	0.50
19	0.30	0.55
20	0.30	0.95
21	0.25	0.80
22	0.85	0.85
23	0.95	0.95
24	1.00	0.90
25	0.85	1.00
26	0.80	0.95
27	0.85	1.00
28	0.60	0.70
29	0.65	0.70
30	0.55	0.65
31	0.60	0.60
32	0.30	0.45
33	0.40	1.00
34	0.50	0.85
35	0.75	0.85
36	0.85	0.85
37	0.85	0.95
38	0.90	1.00
39	0.60	0.65
40	0.85	0.75

Appendix 3 – Sample Means

	Before App	App
Sample Mean - Percentage	64%	77%
Sample Mean - Test Results	13	15



Appendix 4 – Outliers

Student	Test score before using app	Test score after using app
1	12	14
2	9	9
3	8	10
4	9	11
5	12	15
6	14	15
7	15	15
8	17	10
9	18	20
10	15	20
11	14	18
12	14	16
13	14	17
14	12	14
15	10	20
16	11	15
17	8	4
18	7	10
19	6	11
20	6	19
21	5	16
22	17	17
23	19	19
24	20	18
25	17	20
26	16	19
27	17	20
28	12	14
29	13	14
30	11	13
31	12	12
32	6	9
33	8	20
34	10	17
35	15	17
36	17	17
37	17	19
38	18	20
39	12	13
40	17	15

Outliers removed, Chi-Squared count

	50% or less	51 -75%	76-100%
before	11	16	9
after	4	13	19

The expected results

	50% or less	51 -75%	76-100%
before	7.5	14.5	14
after	7.5	14.5	14

	50% or less	51 -75%	76-100%
before	1.633333333	0.155172	1.785714
after	1.633333333	0.155172	1.785714

Chi-Squared:	7.148440066
Degree of Freedom	2
p-value:	0.028

FINAL GRADE

GENERAL COMMENTS

60/100

Instructor

Although you have demonstrated a good understanding of statistics by providing some evidence of your analysis of descriptive to statistics you decided to choose the chi squared test. To do this you had to manipulate the data into ranges and this gave some incorrect results although you did use the test appropriately. A better test would have been the T test as we have data before and after an event with the same set of users. Your report however was well written and you do demonstrate an understanding of building and assessing the data via chi squared tests and it was interesting to see the difference in results when you took out the outlying data.

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9

PAGE 10

PAGE 11

PAGE 12

PAGE 13

