

2014801_Assignment2_ACCB40 02

by Dylan Thomas

Submission date: 09-Jan-2022 08:31PM (UTC+0000)

Submission ID: 168063253

File name: 24885_Dylan_Emyr_Thomas_2014801_Assignment2_ACCB4002_394034_243059721.docx (4.12M)

Word count: 1989

Character count: 9367

Experimental Data Analysis Assignment

Author: Dylan Thomas (2014801)

Data Analysis and Visualisation (ACCA4002)

School of Applied Computing

Wales Institute of Science and Art

09/01/22 Date and Time of Submission

Abstract

...

Table of Contents

Contents

1. Introduction 4

2. Testing..... 4

 2.1 Descriptive Statistics 4

 2.2 Testing the Hypotheses..... 6

 2.3 Outliers in the Data 7

 2.4 Other Factors 8

3. References 9

1. Introduction

The purpose of this assignment is to showcase the knowledge and skills that I have gained during my Data Analysis and Visualisation studies. The assignment is focussed entirely on analysing a set of data for a specific scenario using the most appropriate formulas and tests to prove and disprove hypotheses.

The scenario I have been provided with is a set of marks from a class test marked out of 20 for 40 pupils, all children are from the Year 5 age group. The scenario tests school children using traditional teaching methods, and again after the children have used the online learning app which has been provided to them by the company. I must now investigate whether the app has improved or hindered the children's ability, and whether the company should proceed with the development.

2. Testing

2.1 Descriptive Statistics

The data set provided shows the class test results of 40 Year 5 age grade children, the test scores are marked out of 40. The test is based on mathematical concepts, and the two sets of scores provided for each child shows their score using traditional teaching methods, and another after using an app developed by a local online learning application development company.

Having looked at the measures of frequency, there are 2 scores provided for each child, one before and one after use of the app, totalling 80 scores in total.

As can be seen from Figure 1, there are 29 children in total who scored higher on the test after using the learning application, meaning 72.5% of the children within the survey scored better having sat the test after using the learning app. There were 7 children

Result	Total	Percentage
Better Score	29	72.50%
Identical Score	7	17.50%
Worse Score	4	10%

who had identical scores having sat the test before and after using the app, meaning 17.5% of the children had identical scores. This then left 4 children who scored less on the test after using the learning app, meaning 10% of the children scored worse after using the learning app. Of those 4

children who scored less after using the app, 2 children scored 10% less (2 points), however the other two children scored 35% less (7 points) on the test after using the app.

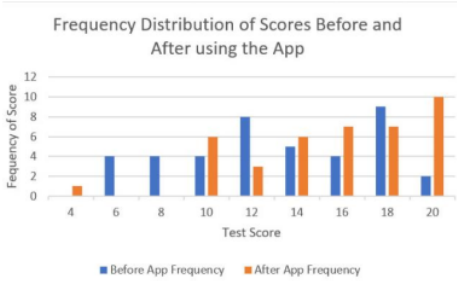
When looking at the measures of Central Tendency I was able to pull together figures showing the Mean, Median and Mode values (Figure 2, below), the mean value also known as the average figure,

	Before App	After App
Mean	12.75	15.05
Median	12.5	15
Mode	17	20

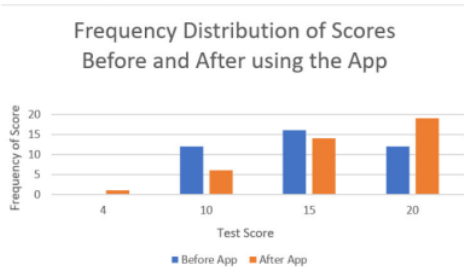
involves adding up all the values of the dataset and dividing by the number of values for said dataset. The mean value can be a whole number and/or include decimals. The

median value refers to the value that's defined at the middle value of the dataset, like the mean value it isn't always a whole number. The Mode value refers to the value that appears the most often within a dataset. Looking at Figure 2, the mode value is the most interesting as the maximum score of 20/20 appears to be the most common value for the test scores after using the app. The mean scores also show an increase of % between the before and after app test scores.

Continuing the measures of frequency, I've used the data to create frequency distribution diagrams for a visual aid of how the results stack up. For this I've used a couple of different bin ranges to show some differences between the datasets, and that using a different bin range affects how we perceive the data.



Firstly, I decided to use a bin range of 2 (Figure 3, above), from 0 to the highest possibility of 20. What this chart shows me is there is only one score below 10 after using the app, but it is the lowest score of the set. There's a lot of scores for both before and after of 16+, but there is a significant amount of 20's scored after using the app, 10 in fact which is 25% of the sample. It also shows a lot of students scoring 18 prior to using the app.



I have created another frequency distribution chart, starting at 4 for the lowest score, moving up in increments of 5 (Figure 4). What's interesting is it shows that almost 50% of the students are scoring between 15-20 after using the app, but the figures look far closer than in the previous chart.

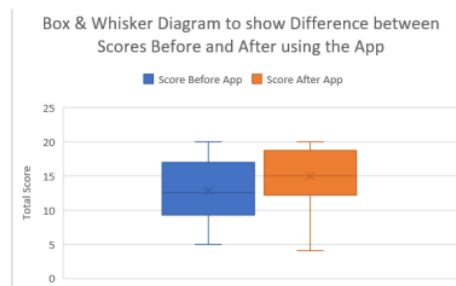
This doesn't show the significance difference at the top end score due to the difference in grouping and a high number of 16-18 scores. This chart does

show the spread of scores, with almost three quarters of the after-app scores sitting at the 15+ scores. Though after using the app also holds the only score below 5, which seems to be an anomaly in the data.

I've also investigated the measures of dispersion and/or variation. This includes the figures for the range, variance, and standard deviation. I've collated these figures to the right (Figure 5).

	Before App	After App
Range	14	16
Variance	16.8076923	15.3820513
Standard Deviation	4.09971856	3.92199583

Due to the abnormally low score of 4 from student 16 the range of scores after the app is greater



than before. The Standard Deviation is slightly lower after using the App. Due to the low standard deviation for both before and after app scores, we can assume that the data clusters closely around the mean. This can be seen from the Box & Whisker diagram (Figure 6) to the left, though there's a wide range both scores cluster closely to the respective

means. The diagram also shows the difference in the mean values with the x, along with the closeness of the clusters to the upper range. We can see that the scores after using the app average higher than prior to using the app.

2.2 Testing the Hypotheses

The Hypothesis I have decided to test is the following,

H_0 : There is no increase in scores after using the App

H_1 : There is an increase in scores after using the App

The null hypothesis is the opposite result to the truth I am expecting/hoping for, the purpose of the

t-Test: Paired Two Sample for Means		
	Variable 1	Variable 2
Mean	12.75	15.05
Variance	16.80769231	15.38205128
Observations	40	40
Pearson Correlation	0.514285098	
Hypothesized Mean Difference	0	
df	39	
t Stat	-3.676908925	
P(T<=t) one-tail	0.00035493	
t Critical one-tail	2.42584141	
P(T<=t) two-tail	0.000709861	
t Critical two-tail	2.707913184	

null hypothesis is to disprove this so that I can then demonstrate the H_1 to be true.

To test these hypotheses, I am going to use the following tests, t-Test. For this test I will be testing to various significance levels, the levels I will test are 0.01, 0.05 and 0.10. By testing the results to various

significance levels, it provides further confidence to the acceptance of the correct hypothesis.

Beginning with the t-Test, having used the Data Tooling pack within Excel I was provided the figures above (Figure 7). It's provided a measure of variance and central tendency, along with the t Stat, P and t Critical figures which is of interest. The difference between the one-tail and two-tail figures is the one tail provides the differences in means to one direction, two-tail showing differences in means in both directions.

Variable 1 is the before app scores, Variable 2 after app, as the scores for both variables correspond to a specific individual, I have used the Paired Two Sample for Means T-test. Though the t-Stat figure is showing as a minus value we'll use the absolute value, as changing the variable order would've provided an identical positive figure.

For this test, I will use the $P(T \leq t)$ two-tail value of 0.000709861 as my p-value as I would like to test the difference in both directions. If I were to use significance levels of 0.01, 0.05 and 0.1, I would be able to reject the null hypothesis due to the p-value being lower than the significance level.

Even if I were to use the one-tail value of 0.00035493 as my p-value, I would again reject the null hypothesis for all significance levels of 0.01, 0.05 and 0.1 as the p-value is again lower than all significance level values. Thus, meaning the sample data collected support the H1 hypothesis that there is a difference between scores after using the app.

2.3 Outliers in the Data

As I'd mentioned earlier, there were 4 children who scored less on the test after using the learning app, meaning 10% of the children scored worse after using the learning app. Of those 4 children who scored less after using the app, 2 children scored 10% less (2 points), however the other two children scored 35% less (7 points) on the test after using the app. One of those two children scoring 35% less on the second occasion was student 16, who's score dropped from 11 to just 4 marks. Student 16 is the main outlier of the data as it's the only score below 5. This could be down to any number of factors, of which I explore in section 2.4 of the report.

When attempting to run a t-Test for the sample without including student 16 the figures did change somewhat, unsurprisingly the Mean score for the after app scores increased from 15.05 to 15.3, and the variance dropped from 15.3 to 12.5 given the dramatically lower score.

What's interesting from the results of (Figure 8, below) is the interesting score provided in the $P(T \leq t)$ one-tail value of 6.12834E-05, however this is just a shorthand way for Excel to present the value of 0.0000612833859878324 to the end user. The results without the abnormal scoring of Student 16 are similar in results to the original test results in the test including all students. However the results are more dramatic, in that the one-tail and two-tail values are much lower than the significance values of 0.01, 0.05 and 0.1 meaning that the H_0 can again be rejected.

t-Test: Paired Two Sample for Means		
	Variable 1	Variable 2
Mean	12.79487179	15.33333333
Variance	17.16734143	12.49122807
Observations	39	39
Pearson Correlation	0.543909828	
Hypothesized Mean Difference	0	
df	38	
t Stat	-4.278459968	
$P(T \leq t)$ one-tail	6.12834E-05	
t Critical one-tail	1.68595446	
$P(T \leq t)$ two-tail	0.000122567	
t Critical two-tail	2.024394164	

2.4 Other Factors

Though I have investigated the data thoroughly, and the sample dataset seems straight forward there are still some other factors with this experiment which should be taken into consideration before we view the results as concrete.

Are the tests undertaken identical? If so, in my opinion it's a test of memory rather than ability. The practice of sitting the test first time around makes the child more accustomed to what's ahead, and a sense of what's to come rather the sense of unknown in sitting the first test. The test could be done over a longer period where the children sit the test using the traditional teaching methods, then having used the app for 6 months or so then re-sitting a similar test with subtle differences, thus removing the test predictability element. It also removes the need to master techniques to cram information for the test.

Does the app assist with all learning types? Some children learn in different ways, this could affect their ability in the test. This could be managed by offering different variations of the app which benefit the different learning types. Linked to this topic, is there a user guide provided to the children to ensure the children can effectively navigate the app, if not is there suitable guidance/assistance provided to the children by teachers or the company staff to ensure the children are maximising the use of the app.

The specific topic of the quiz isn't disclosed either, the mathematics app will cover a broad range of

topics within maths, but is the topic for the quiz covered in detail or briefly touched upon? This could affect the after-app learning scores.

To conclude, looking at the data having calculated a variety of statistics, accompanied by diagrams in certain cases and considering the other factors which could affect the results of the tests for the random sample, I'm confident that the app does improve the test results of the children and that the company should proceed with further development of the app.

3. References

FINAL GRADE

GENERAL COMMENTS

83/100

Instructor

An excellent report well laid out with consideration for many different techniques to analyse the data. Using a frequency analysis provided you with some great insight into the data set. If correctly used the T test to analyse the data and looked at the significant values asset and required which helped you to reflect upon and reject the correct hypothesis. A comprehensive discussion of data that could be classed as outliers was presented and duly analysed. Be mindful this is an academic document and therefore you should always write in third person also your report suggested references but there were no references listed

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9