

Experimental Data Analysis - ACCB4002 - K Sumner

by Kes Sumner

Submission date: 11-Jan-2022 04:47PM (UTC+0000)

Submission ID: 168471055

File name: 25515_Kes_Anthony_Sumner_Experimental_Data_Analysis_-_ACCB4002_-_K_Sumner_394034_1914672869.docx (81.2K)

Word count: 2707

Character count: 13128

Experimental Data Analysis

Author: Kes Sumner (2015409)

Data Analysis and Visualisation (ACCB4002)

School of Applied Computing

Wales Institute of Science and Art

Date and Time of submission: 11/01/2022 1700

Contents

1	Data Analysis	3
1.1	The Data	3
1.2	Descriptive Statistics	3
1.2.1	Frequency.....	3
1.2.2	Measures of Central Tendency	4
1.2.3	Measures of Variability	5
1.3	Analysis	5
1.3.1	Hypotheses.....	6
1.3.2	t Test Analysis	6
1.3.3	Recommendation.....	6
1.3.4	Outliers.....	7
1.3.5	Other Factors	9
2	References	10
3	Appendices.....	11
3.1	Test score data sets.....	11
3.2	Full Excel t Test outputs	12
3.2.1	Data set t Test results	12
3.2.2	Amended data set t Test results.....	13

1 Data Analysis

1.1 The Data

The data provided for this analysis exercise consists of a list of forty students each with a pair of mathematics test scores. Each student has one score from a test taken prior to using a trial online learning application, and a corresponding score from a second test taken after they had used the application. Each of the two tests was marked out of a total of twenty.

The data from the assignment description was input into Excel in a table format as shown in Table 6 in [Appendix 3.1](#). This table contains the source data for all the following data analysis.

1.2 Descriptive Statistics

1.2.1 Frequency

Initially the frequency of the scores was investigated by creating a pivot table of the two sets of marks and creating a frequency count of each score. The following frequency chart was created from the table of values.

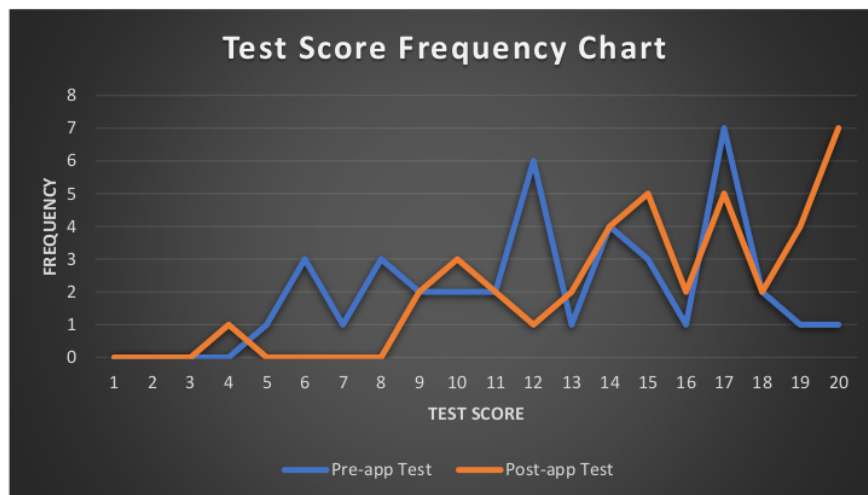


Figure 1: Test score frequency chart

In Figure 1 above, the chart shows a shift to the right (higher scores) for the frequency of the post application usage test scores (coloured orange). This can be seen more clearly in the cumulative frequency polygon chart, shown in Figure 2.

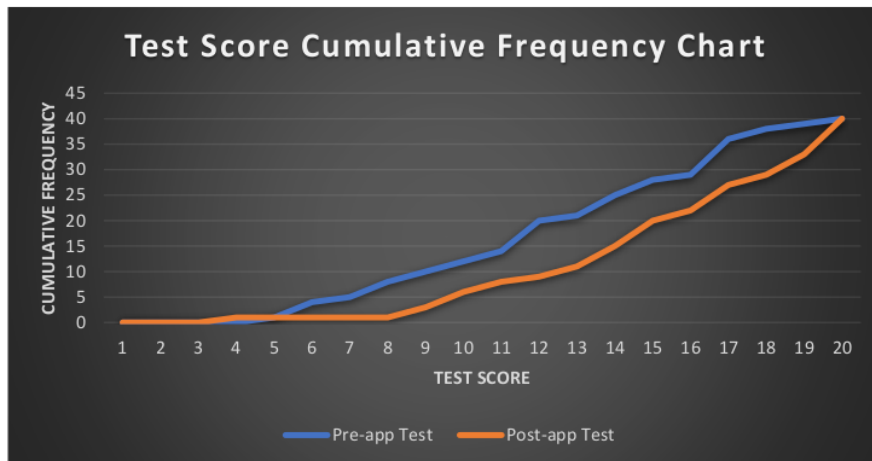


Figure 2: Test score cumulative frequency chart

The frequency of the scores in the post application test do not start to rise significantly until several points higher than the pre-application usage test scores in the cumulative frequency chart.

1.2.2 Measures of Central Tendency

The next method of statistical investigation was to calculate and evaluate the more common measures of central tendency. Measures of central tendency can be particularly well suited to comparing groups or sets of figures against each other, such as different groups of individuals or in a 'before and after' scenario, such as this example of test scores.[\[1\]](#)

The three main measures are:

- Mean: The sum of all the numbers in the data set divided by the number of numbers in the set. In the example, this would be the total sum (for each test) divided by the number of students, 40.
- Median: With the data sorted in ascending order, the median will be the centre value in the list (or the centre two numbers added and divided by 2 if there are an even number of data items).
- Mode: The mode of a set of numbers will be the most frequently occurring number. If only one mode exists in the data set, the distribution is unimodal, if more than one exists it is bimodal.[\[2\]](#) Both data sets in this analysis are unimodal.

The three measures of central tendency were calculated for each of the two examples test score case. The results are shown in the table below.

Measure of central tendency	Test score before using app	Test score after using app
Mean	12.75	15.3
Median	12.5	15.5
Mode	17	20

Table 1: Measures of central tendency

In Table 1 each of the measures has increased between the two tests. While this indicates a positive shift in the test scores after using the application, note that there is a significant difference between the mode scores and the other two measures, mean and median scores. The mode indicates that there are a greater number of higher scores than any one other score, although this could be marginal.

The mode will also indicate the skew of the distribution (as it is the most frequent or common value), in this analysis the data is negatively skewed with the highest frequency being towards the right tail of the distribution.[\[3\]](#)

The three measures are affected differently by the variability (or spread) of the scores, particularly for skewed distributions. Without taking in to account the variability of the data, and only relying on the mean, median or mode to represent the level of the data can lead to inaccurate estimates.

1.2.3 Measures of Variability

There are several measures of statistical variability. One simplistic measure is range, which is the smallest data number subtracted from the largest data number. As the range is based on the two extreme test scores and does not consider any other values in between, it can be deceptive if the values in between are not distributed evenly, which is likely.

The most common measure of variability is variance, which is the standard deviation squared. The variance is the mean of the squared deviation scores, where each deviation score is each score minus the mean score. The variance was calculated in Excel using the VAR.S function

Table 2 shows the ranges and variances for the pre and post application usage test results.

Measure of variability	Test score before using app	Test score after using app
Range	15	16
Variance	16.81	15.29

Table 2: Measures of variability

It is notable that the range scores and the variance scores are in opposite orders in that the range shows the variability to be higher in the post app test results whereas the variance shows the variability to be lower from the same test results. This demonstrates the lack of granularity (and potentially accuracy) of using the range as a measure of variability.

1.3 Analysis

The requirement from the analysis of the pre and post application usage test score data is to provide a recommendation to the application developer if the trial of the application is successful. To be able to provide appropriate guidance this analysis will need to show whether there is a statistically significant directional difference between the two sets of test results, i.e. an improvement in test scores.

To achieve this for these two sets of dependent (or matched) data then a *t* Test analysis can be performed.

There are several steps required to perform a *t* Test analysis:

1. State the hypotheses.
2. Define the significance level.
3. Calculate the obtained value, t_{obt}
4. Calculate the critical value, t_{crit}
5. Either reject the null hypothesis or not.

1.3.1 Hypotheses

The first step in this process is to define suitable hypotheses. As the request is to find if the use of the application makes a significant difference to test results, then the alternative hypothesis (H_1) will be:

- 'That there is a significant statistical difference in the test scores achieved before and after using the online learning application.'

From the above hypothesis, the null hypothesis (H_0) will therefore be:

- 'That there is no significant statistical difference in the test scores achieved before and after using the online learning application i.e., they are the same.'

1.3.2 t Test Analysis

Excel was used to perform the t Test analysis for dependent samples using the two data sets and the three requested levels of significance. As the interest for the application is in the post test results being improved compared to the pre-test results, a directional or one tail test result was used. The full Excel t Test results for the three levels of significance are shown in [Appendix 3.2.1](#). The summary results are shown in Table 3, below.

Level of significance (α)	t_{obt} (absolute)	t_{crit}	Difference $t_{obt} - t_{crit}$	Accepted hypothesis
0.01	4.154	2.426	1.728	H_1
0.05	4.154	1.685	2.469	H_1
0.1	4.154	1.304	2.850	H_1

Table 3: t Test results for test score data sets

The results show that there is considerable statistical difference using all three levels of significance, thus the null hypothesis is rejected in all 3 cases and the alternative hypothesis is accepted at each level of significance.

1.3.3 Recommendation

Based on the rejection of the null hypothesis, the acceptance of the alternative hypothesis and the significant improvement in the test scores between the two tests then the recommendation to the company would be to continue development of the application.

1.3.4 Outliers

To investigate whether any outliers exist in either data set, Excel was used to independently sort both sets of test scores in to ascending order and the top and bottom values were examined. Table 4 shows the values, with the rest of the values in between being omitted for conciseness.

Test score before using app	Test score after using app
5	4
6	9
6	9
6	10
.....	
18	20
19	20
20	20

Table 4: Top and bottom values of test score results

The lowest score of 4 in the post application usage test appears to lie considerably outside the rest of the range for that data set.

To visually clarify, a scatterplot chart was created for each data set, shown in Figures 3 and 4 below.

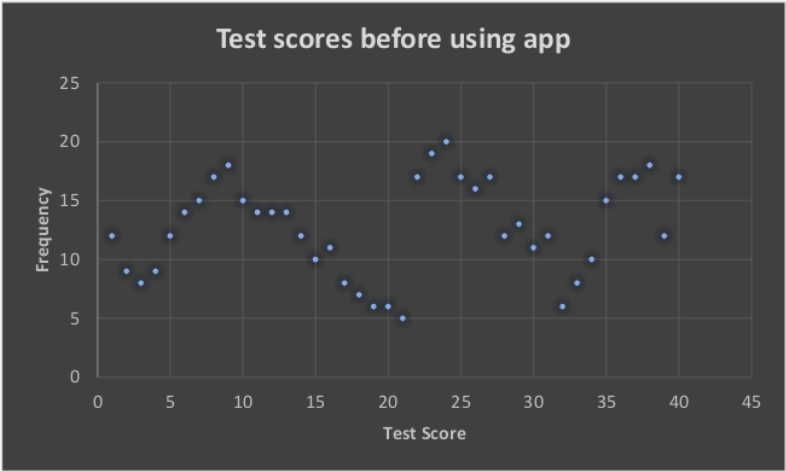


Figure 3: Test scores before using app

Figure 3 shows that the pre application usage test scores are all within a clearly defined range.

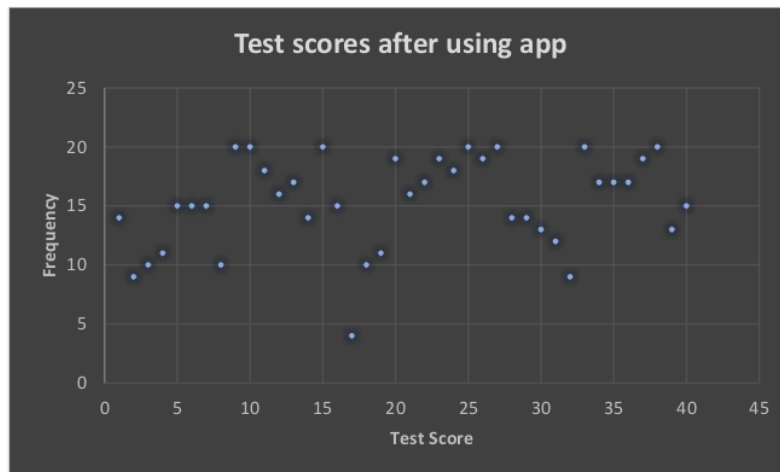


Figure 4: Test scores after using app

In Table 4 and Figure 4, the test score of 4 is significantly outside the main range of test results and can be considered an outlier.

As it is a single outlier and a reduction in test score after the use of the application there could be many reasons as to the cause. If there was more than one outlier, then it would potentially be easier to identify causes via common factors. Reasons for the single outlier could include:

- Personal factors affecting use of the application between tests such as illness, commitments, time available etc.
- Difficulty experienced by the individual in using or understanding the application.
- Difference in the pre and post application test questions maybe more difficult for some individuals.

Student 17 was removed from both sets of test score results to remove the outlier in the post app test results and their corresponding pre app test score and the t Test was repeated using the adjusted data sets. The full Excel t Test results for the amended data sets using the three levels of significance are shown in [Appendix 3.2.2](#). The summary results are shown in Table 5 below.

Level of significance (α)	t_{obt} (absolute)	t_{crit}	Difference $t_{obt} - t_{crit}$	Accepted hypothesis
0.01	4.486	2.429	2.058	H_1
0.05	4.486	1.686	2.801	H_1
0.1	4.486	1.304	3.182	H_1

Table 5: t Test results for adjusted test score data sets

Removing the outlier makes an expected difference to the t Test results, increasing the difference between the t_{obt} and t_{crit} results. The accepted hypothesis does not change and is reinforced by the amendment to the data set and the recommendation stands.

1.3.5 Other Factors

In addition to outliers there are many external factors that could affect the results of the experiment. Some potential influences are:

- A key factor that could affect the results is the size of the data set. Less subjects would potentially lead to less accurate results; more subjects could demonstrate more accurate results.
- The description of the experiment gives little indication of the level of control over the whole process. Some students could potentially spend more time using the application and hence improve their scores more than others. Some students could be using other tools between the two tests to help their learning after the first test.
- Accessibility and support of the use of application. If a student is struggling to use the application, what levels of support are in place to help? Some students may find the application easier to use than others.

2 References

IEEE format references are used in this report.

- [1] J. H. Kranzler, *Statistics for the terrified*, Sixth edition. Lanham, Maryland: Rowman & Littlefield, 2018, ch.5, p50.
- [2] Spiegel M., L.Stephens. *Schaum's Outline of Statistics*, Fifth edition. New York: McGraw-Hill, 2014, ch.3, pp. 59-62.
- [3] J. H. Kranzler, *Statistics for the terrified*, Sixth edition. Lanham, Maryland: Rowman & Littlefield, 2018, ch.4, pp.45-46.

3 Appendices

3.1 Test score data sets

Student number	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

Table 6: Test score data sets

3.2 Full Excel t Test outputs

The appendix contains the full Excel t Test output table, for reference. Note that 't Stat' in the tables is the t_{obt} value and 't Critical one-tail' is the t_{crit} value that was used for directional t Test results.

3.2.1 Data set t Test results

The full output tables for the t Test calculated using Excel for the three difference levels of significance (0.1, 0.05 and 0.01) are shown below.

t-Test: Paired Two Sample for Means Level of significance (α) = 0.01		
	Test score before using app	Test score after using app
Mean	12.750	15.300
Variance	16.808	15.292
Observations	40.000	40.000
Pearson Correlation	0.530985814	
Hypothesized Mean Difference	0.000	
df	39.000	
t Stat	-4.154	
P(T<=t) one-tail	8.62016E-05	
t Critical one-tail	2.426	
P(T<=t) two-tail	0.000	
t Critical two-tail	2.708	

Table 7: Full t Test results for test score data sets ($\alpha=0.01$)

t-Test: Paired Two Sample for Means Level of significance (α) = 0.05		
	Test score before using app	Test score after using app
Mean	12.750	15.300
Variance	16.808	15.292
Observations	40.000	40.000
Pearson Correlation	0.530985814	
Hypothesized Mean Difference	0.000	
df	39.000	
t Stat	-4.154	
P(T<=t) one-tail	8.62016E-05	
t Critical one-tail	1.685	
P(T<=t) two-tail	0.000	
t Critical two-tail	2.023	

Table 8: Full t Test results for test score data sets ($\alpha=0.05$)

t-Test: Paired Two Sample for Means		Level of significance (α) = 0.1	
		Test score before using app	Test score after using app
Mean		12.750	15.300
Variance		16.808	15.292
Observations		40.000	40.000
Pearson Correlation		0.530985814	
Hypothesized Mean Difference		0.000	
df		39.000	
t Stat		-4.154	
P(T<=t) one-tail		8.62016E-05	
t Critical one-tail		1.304	
P(T<=t) two-tail		0.000	
t Critical two-tail		1.685	

Table 9: Full t Test results for test score data sets ($\alpha = 0.1$)

3.2.2 Amended data set t Test results

The full output tables for the t Test calculated using Excel with the amended data sets after the outlier has been removed for the three difference levels of significance (0.1, 0.05 and 0.01) are shown below.

t-Test: Paired Two Sample for Means		Level of significance (α) = 0.01	
		Test score before using app	Test score after using app
Mean		12.872	15.590
Variance		16.641	12.248
Observations		39.000	39.000
Pearson Correlation		0.510490398	
Hypothesized Mean Difference		0.000	
df		38.000	
t Stat		-4.486	
P(T<=t) one-tail		3.25341E-05	
t Critical one-tail		2.429	
P(T<=t) two-tail		0.000	
t Critical two-tail		2.712	

Table 10: Full t Test results for amended test score data sets ($\alpha = 0.01$)

t-Test: Paired Two Sample for Means		Level of significance (α) = 0.05	
		Test score before using app	Test score after using app
Mean		12.872	15.590
Variance		16.641	12.248
Observations		39.000	39.000
Pearson Correlation		0.510490398	
Hypothesized Mean Difference		0.000	
df		38.000	
t Stat		-4.486	
P(T<=t) one-tail		3.25341E-05	
t Critical one-tail		1.686	
P(T<=t) two-tail		0.000	
t Critical two-tail		2.024	

Table 11: Full t Test results for amended test score data sets ($\alpha=0.05$)

t-Test: Paired Two Sample for Means		Level of significance (α) = 0.1	
		Test score before using app	Test score after using app
Mean		12.872	15.590
Variance		16.641	12.248
Observations		39.000	39.000
Pearson Correlation		0.510490398	
Hypothesized Mean Difference		0.000	
df		38.000	
t Stat		-4.486	
P(T<=t) one-tail		3.25341E-05	
t Critical one-tail		1.304	
P(T<=t) two-tail		0.000	
t Critical two-tail		1.686	

Table 12: Full t Test results for amended test score data sets ($\alpha=0.1$)

Experimental Data Analysis - ACCB4002 - K Sumner

GRADEMARK REPORT

FINAL GRADE

GENERAL COMMENTS

96/100

Instructor

As reflected in the mark this was an excellent submission. You've covered each aspect of the assignment remit in detail looking at descriptive statistics and then discussing the reasons behind your choice of test and your hypothesis. Use of tables to present the statistical data findings allowed the reader to capture the information quite easily and your graphs helped to provide visual impact to the data were discussing. Additional information provide in the appendices added to the overall readability of your analysis.

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

PAGE 14