

Data Analysis Report

by Daniel Jones

Submission date: 12-Jan-2026 12:37AM (UTC+0000)

Submission ID: 272877387

File name: 57246_Daniel_Jones_Data_Analysis_Report_709067_1250475055.docx (444.99K)

Word count: 2056

Character count: 10949



Data Analysis Report (ACCB4002)

Daniel Jones

Daniel Jones (2404095)
2404095@student.uwtsd.ac.uk

Contents

1. Introduction	3
2. Hypothesis	4
1. Null Hypothesis (H_0)	4
2. Alternative Hypothesis (H_A)	4
3. Statistics & Significance Testing	5
4. Outlier and Impacts on Results	7
5. Additional Patterns and External Factors	8
6. Conclusion & Evaluation	9
7. References	10

1. Introduction

This report will look at analysing a sample comparison between two servers and their latency between data packets recorded in milliseconds to determine if the new algorithm within server 2 would improve network response time over the existing algorithm currently running within server 1. The investigation will determine the mean latency of both servers to see if the average latency is lower with the newer algorithm.

The report will cover the hypothesis, describe the tests that were made from the sample data and identify any outliers present within the data set. It will also discuss any key decisions made during the analysis, highlight any limitations or external factors not accounted for and assess how they could influence the results throughout the testing. Finally, the report will present a conclusion that summarises all the findings throughout this report and determine whether the new algorithm does or does not improve network response time.

2. Hypothesis

For this investigation a set of two hypothesis were created to appropriately determine the effectiveness of the newly implemented algorithm setup within server 2. A null hypothesis and an alternative hypothesis were formulated to represent the purpose of this experiment and measure its success. By using a null hypothesis for this investigation, it will provide a benchmark against which the alternative hypothesis can be tested. It also assists with the control of any effects that may occur due to random variations [1]. Below is the set of hypotheses that will be used for this investigation:

1. Null Hypothesis (H_0)

There is no significant difference between the mean latency of server 1 (existing routing algorithm) and server 2 (new routing algorithm). Any difference between the observed latency is simply due to random variation and not due to the difference in algorithm.

2. Alternative Hypothesis (H_A)

There is a significant difference between the mean latency of server 1 (existing routing algorithm) and server 2 (new routing algorithm). Indicating that the new routing algorithm does indeed have a significant effect on the network response time.

3. Statistics & Significance Testing

Now that the set of hypotheses has been determined it is now time to test and analyse the data provided and present all the findings within this section. To determine if the null hypothesis is indeed false several tests were made on two independent samples of latency measurements from server 1 and server 2. Each server provided a sample list of 42 observations of data packet latency measured in milliseconds to help ascertain the mean latency values between server 1 and 2.

The results of the tests were evaluated against the significance levels of $p=0.10$, $p=0.05$ and $p=0.01$ to assess the overall strength of the evidence against the null hypothesis. The reason for using multiple different significance levels while testing the hypothesis were to implement the Bonferroni correction and hopefully reduce or at the very least present the likelihood of a type 1 error [2].

Presented below are the screenshots of tests conducted and graphs created within Microsoft Excel that calculated the sample mean latency between the two servers as well as the standard deviation and even the significance t-tests to help provide detailed evidence for this investigation:

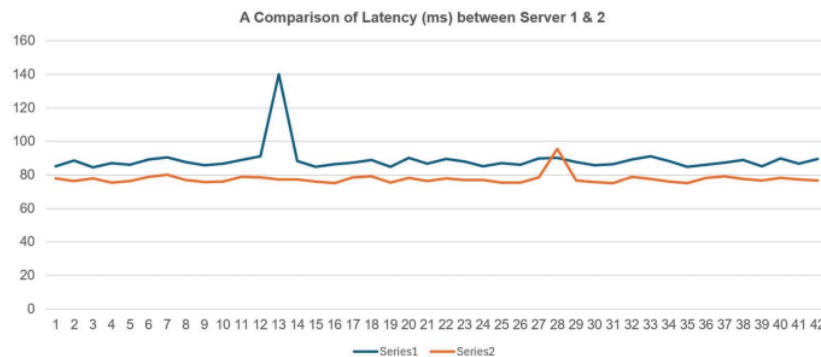


Figure 1: Comparison of Latency - Line Graph

Server 1 Latency (ms)	
Mean	88.87
Standard Error	1.28
Median	87.60
Mode	N/A
Standard Deviation	8.30
Sample Variance	68.89
Kurtosis	37.44
Skewness	5.96
Range	55.60
Minimum	84.40
Maximum	140.00
Sum	3732.60
Count	42.00

Figure 2: Server 1 Latency - Descriptive Statistics

Server 2 Latency (ms)	
Mean	77.62
Standard Error	0.49
Median	77.15
Mode	78.90
Standard Deviation	3.14
Sample Variance	9.88
Kurtosis	27.10
Skewness	4.72
Range	20.60
Minimum	75.00
Maximum	95.60
Sum	3260.00
Count	42.00

Figure 3: Server 2 Latency - Descriptive Statistics

t-Test: Two-Sample Assuming Unequal Variances (0.10 alpha)		
	Variable 1	Variable 2
Mean	88.87142857	77.61904762
Variance	68.88550523	9.879628339
Observations	42	42
Hypothesized Mean Difference	0	
df	53	
t Stat	8.216787683	
P(T<=t) one-tail	2.52469E-11	
t Critical one-tail	1.297729843	
P(T<=t) two-tail	5.04938E-11	
t Critical two-tail	1.674116237	

Figure 4: t-Test $p=0.10$

t-Test: Two-Sample Assuming Unequal Variances (0.05 alpha)		
	Variable 1	Variable 2
Mean	88.87142857	77.61904762
Variance	68.88550523	9.879628339
Observations	42	42
Hypothesized Mean Difference	0	
df	53	
t Stat	8.216787683	
P(T<=t) one-tail	2.52469E-11	
t Critical one-tail	1.674116237	
P(T<=t) two-tail	5.04938E-11	
t Critical two-tail	2.005745995	

Figure 5: t-Test $p=0.05$

t-Test: Two-Sample Assuming Unequal Variances 0.01 alpha		
	Variable 1	Variable 2
Mean	88.87142857	77.61904762
Variance	68.88550523	9.879628339
Observations	42	42
Hypothesized Mean Difference	0	
df	53	
t Stat	8.216787683	
P(T<=t) one-tail	2.52469E-11	
t Critical one-tail	2.398789836	
P(T<=t) two-tail	5.04938E-11	
t Critical two-tail	2.671822636	

Figure 6: t-Test $p=0.01$

Data Packet	Server 1 Latency (ms)	Server 2 Latency (ms)	Data Packet	Server 1 Latency (ms)	Server 2 Latency (ms)
1	89.2	78.1	22	89.6	78
2	89.6	79.3	23	88.1	79.9
3	84.4	77.9	24	85.3	77
4	87	79.4	25	87.2	75.3
5	86.1	76.5	26	86	75.5
6	89.3	79	27	89.9	78.6
7	90.5	80.2	28	90.3	85.6
8	87.8	77.1	29	87.7	76.6
9	85.7	75.8	30	85.8	75.7
10	86.9	76.2	31	86.5	75.2
11	89	78.5	32	89.2	78.9
12	91.2	78.5	33	91	77.5
13	140	77.4	34	88.3	76.1
14	88.2	77.3	35	84.8	75
15	85	76	36	86.2	78.2
16	86.4	75.1	37	87.3	79.2
17	87.5	78.7	38	88.9	77.8
18	88.8	79.1	39	85.1	76.8
19	84.9	75.6	40	90	78.4
20	90.1	78.3	41	86.7	77.2
21	86.6	76.4	42	89.5	76.7

Figure 7: Data Latency Sample Table

As you can see from the figures presented above there does indeed seem to be a difference between the latency mean of server 1 and server 2. It seems that from the sample data provided the mean latency in the server that is using the new algorithm is around 11ms lower than the server running the original algorithm which might show that the null hypothesis presented earlier in this report may be false.

To back this up further three t-tests were performed with different significance levels to assist with comparing the mean values of the two sample and determine if there is or isn't a statistically significant difference between the two samples [3]. Presented in figure 4-6 above are the t-tests done to analyse the data. For this investigation the t-tests used were the two-sample test that assume there is an unequal variance known as the Welch's t-test. This version of the t-test was used after the working out the difference between the variance of both samples by dividing both variances and checking if the ratio is less than 4 which in this case the ratio between the two samples as shown in figure 2-3 is 6.97 which shows that the variance between the samples is most likely not equal [4].

The t-tests displayed in figure 4-6 above show that t statistic or the T_{opt} are higher compared to the t critical value which helps indicate that there is potentially a significant difference between the two sample sets of server 1 and server 2 which seems to disprove the null hypothesis of there not being a difference between the servers.

4. Outlier and Impacts on Results

During the investigation of the latency data between server 1 and server 2, potential outliers were discovered that might have influenced the results of the significance testing demonstrated in the previous section. One potential outlier was found within server 1 where a notably high latency value of 140ms was found for data packet 13 which seems substantially larger than the majority of clustered between 84ms and 91ms as shown in the graph in figure 1. Additionally, the graph presents another possible outlier within server 2 regarding the value found within data packet 28 which in relation to the typical range for that server seems abnormally high as most values within server 2 fall between 75ms and 80ms.

The obvious impact of these outliers is their potential to inflate the mean latency of the samples and increase the variance of the samples. This is definitely a threat for sample as the outlier value of 140ms is so much higher than the rest of the dataset that could seriously effect both the mean and variance of the sample which as shown in the previous section are both highly important with disproving the null hypothesis.

For this analysis, the impact of these identified outliers was considered by evaluating the influence they have on the statistics and hypothesis test outcomes. While the removal of such datapoint could potentially lead to a better comparison of the performance between the two servers, excluding them could also potentially underrepresent the real-world performances of the servers and lead to miss information. As a result, the primary analysis has retained these values to ensure the integrity of the dataset, while also acknowledging that these inclusions could potentially affect the statistics presented in the previous section of this report.

5. Additional Patterns and External Factors

Beyond the presence of outliers, there have been several additional patterns that can be observed throughout the dataset provided. It seems overall server 2 has had consistently lower latency values throughout the majority of the dataset compared to server 1 and it seems that range of values recorded within the server 2 sample is quite a lot smaller as can be seen by the range value recorded within the figures 2-3 within section 3. This reduced range within server 2 could potentially suggest that it has a more stable and predictable performance due to the new algorithm as opposed to the previous algorithm within server 1 that displays a much wider range of network latency values.

Due to the limited information surrounding the data samples there could be a number of external factors that might have influenced the network latency within the datasets. After some careful research a number of potential external factors have been found that could have an effect on the latency. One such potential factor could have been a difference between the hardware of server 1 and server 2 such as the cabling, routers, switches or Wi-Fi access points any of these differences could have the potential to affect the latency of the network. Another possible external factor could be the distance the data packet would have to travel for both server locations. A final potential factor could be the method the data packet is transmitted over to the two different servers such as server 1 could be using a wireless network medium while server 2 could be using a Fiber optic medium to transmit the data packet [5].

6. Conclusion & Evaluation

This investigation aimed to determine whether or not the newly implemented algorithm used for server 2 makes any significant difference to the network response time when compared to the current algorithm used within server 1. This was achieved by analysing a sample of forty-two data packets sent to both server 1 and server 2 and the network latency measured in microseconds for each data packet within each server, supported by a hypothesis and statistical evaluation.

It seems the results of the analysis do in fact indicate that there is a difference between the latency and therefore the network response time between server 1 and server 2 and so the new algorithm seems to have had an effect on the network latency in what seems to a positive way due to its reduced network latency on average compared to the previous algorithm. This was discovered throughout testing when server 2 consistently presented a lower mean network latency than server 1. The difference between the mean latency of server 1 and server 2 was around 11 milliseconds, which certainly shows a noticeable improvement on the network performance.

Furthermore, the Welch's T-test was implemented at multiple significance levels to provide evidence on the statistical significance of improvement between the two algorithms by showing that t value exceeds the critical value for all three tests. Once all of these tests were completed it led to the rejection of the null hypothesis which stated that there would be no significant difference between the two algorithms which these tests seem to prove is not the case.

Though the findings within this report strongly suggest that there is a significant difference and even improvement with the network response time of the new algorithm, the presence of outliers within both datasets does complicate matters. In particular, the giant latency value observed in server 1 could potentially inflate the variance of the dataset and in doing so affect the statistical outcome. However, these potential outliers were retained for this analysis to ensure the data is as accurate as can be for realistic operating conditions which acknowledges that there will be an occasional spike in latency from time to time.

Additionally, several external factors were identified within this report that could have affected the dataset such as hardware differences, transmission medium and routing distance. As these factors were not controlled for this experiment it's hard to tell for definite how much the new algorithm is responsible for the difference between network response time or one of these potential factors. As a result, though the conclusion is that there is a significant difference between the algorithms it should be interpreted with some caution and hopefully further tests could be carried out with these factors controlled to ensure a more accurate result.

Overall, based on the statistical evidence and observed trend in performance, the new routing algorithm used within server 2 does appear to have a significant difference on the network response time and is a potential improvement over the existing algorithm. Future testing under more controlled conditions would be preferable to assist with validating the results within this investigation and strengthen the confidence that the new algorithm provides a significant improvement on the network response time.

7. References

- [1] N. S. Sumalatha Gangadhar, "Importance of Null Hypothesis in Research," 01 January 2024. [Online]. Available: <https://scispace.com/resources/null-hypothesis-in-research/>. [Accessed 16 December 2025].
- [2] Statsig, "Bonferroni Correction," [Online]. Available: <https://www.statsig.com/glossary/bonferroni-correction>. [Accessed 16 December 2025].
- [3] A. Hayes, "T-Test: What It Is With Multiple Formulas and When to Use Them," 18 December 2025. [Online]. Available: <https://www.investopedia.com/terms/t/t-test.asp#toc-using-the-t-test>. [Accessed 18 December 2025].
- [4] Z. Bobbitt, "How to Determine Equal or Unequal Variance in t-tests," 11 April 2021. [Online]. Available: <https://www.statology.org/determine-equal-or-unequal-variance/>. [Accessed 18 December 2025].
- [5] IR Team, "Network Latency - Common Causes and Best Solutions," [Online]. Available: <https://www.ir.com/guides/what-is-network-latency>. [Accessed 22 December 2025].

Data Analysis Report

GRADEMARK REPORT

FINAL GRADE

GENERAL COMMENTS

56/100

Generic Assessment Criteria	Marks available	Marks awarded
1. Engagement with Literature Skills	10	
Bibliography and referencing present and good. Limited referencing through the text.		6
2. Knowledge and Understanding Skills	30	
Reasonable knowledge demonstrated of the outliers and descriptive stats. Reasonable visualisations however more comparison may have been advantageous through discussion of correlation but overall a good knowledge demonstrated through research.		16

3. Cognitive and Intellectual Skills	30	
Reasonable interpretation of results and discussion of outliers. More detail to explicit interpretation of results in the discussion would have been beneficial. Some reasonable justification to choices made, however was the two tailed test the right one?		16
4. Practical Application Skills	20	
A reasonably well written H0 and H1, may have been better with H0 being $\text{Server2} \leq \text{Server1}$ and H1 being $\text{Server2} > \text{Server1}$. Well demonstrated results leading to rejection of null hypotheses to significance levels. The rejection at the explicit p values as well as acceptance or rejection of H1 as well as the significance levels could have been written more explicitly. Reasonable visualisations and tables.		12

5. Transferable Skills for Life and Professional Practice	1 0	
A reasonably well laid out report. Attention to detail in arts would have been beneficial (i.e. comparing server 2 to server 2 in two sections).		6

PAGE 1
PAGE 2
PAGE 3
PAGE 4
PAGE 5
PAGE 6
PAGE 7
PAGE 8
PAGE 9
PAGE 10