

Anca-24601

by Student Ex 1

Submission date: 19-Aug-2021 12:34PM (UTC-0500)

Submission ID: 1623246154

File name: Data_Analysis_and_Visualization_1.docx (1.04M)

Word count: 1915

Character count: 9918

Data Analysis and Visualization



STUDENT NAME:
DATE:

All the analysis given below and also results are attached below, let us take the null hypothesis that there is no difference between the infection rates to the two groups: white and the BAME in UK, as the alternative that infection rate is raised between white and BAME actually it is to test $\mu_0, \mu_1 = \mu_2$ as μ_A : μ_1 is not equal to μ_2 , where μ_1 and μ_2 are the mean infection rates in the population of white and BAME respectively [1].

$\bar{x}_1 = 36.18$, $\bar{x}_2 = 36.74$, $s_1 = 9.4192$, and $s_2 = 9.3171$ are the descriptive statistics computed from the samples white and BAME each of the size is 50. I say that a "urgent study" should be carried out on how hospitals contained Covid's spread during the first epidemic. An analysis of 314 hospitals in the United Kingdom indicated that one in four infections was acquired in some circumstances [2]. In all, one in 10 cases was acquired in the hospital. One of the authors of the article said, "People who had one problem in hospital caught Covid and died horribly." The research finds a number of problems that affected infection control, including the availability of personal protective gear, the design of hospitals and the amount of testing "[3]. The academy, which includes medical groups responsible for training and development in different fields, said in a statement that when limits were legally removed, people had to take a responsive approach. He cautioned that "things will become worse before they get better." There was little question. The academy claimed instances of regular care being delayed because health staff are being isolated under increasing strain as people who remained away from the NHS during the pandemic are again seeking treatment [4]. It noted that the rise of Covid-19 long history and the backlog of elected work during the crisis meant that "it is the worst thing about a hard winter." "If this risky and irregular sickness is largely controlled by the whole community, we will certainly be able to 'learn how to live' Covid-19." "We're not yet in that position, however, and we need to sadly wait for things to get worse again." The two datasets have been checked for their normality, white in perfectly normal (normal probability plot) [5]. BAME is showed out the sample mean can be assumed to be normally distributed [central limit theorem] since $n=50$ in large. Boxplot for the two datasets have been constructed for any potential outliers there is nothing unusually seen in the two datasets. A z test for testing the above hypothesis has the test statistics [4].

$$z = \frac{x_1 - x_2}{\sqrt{s_1^2/n_1 - s_2^2/n_2}}$$

$$z = \frac{36.18 - 36.74}{\sqrt{9.4192^2/50 - 9.3171^2/50}}$$

$$z = \frac{-0.56}{\sqrt{3.5106}}$$

$$z = -0.2989$$

P value for z score is -0.2989 is 0.765016, hence the null hypothesis cannot be rejected at 10%, 1% and 8.2% as value of $p > 0.19$, > 0.01 , > 0.02 respectively, hence it can be concluded that there is no difference between the infection rates in the two population: white and BAME. A hypothesis is a statement that presents a research topic and offers an anticipated conclusion. It is a fundamental aspect of the scientific method on which scientific experiments are based. Therefore, you have to build your hypothesis carefully and thoroughly. A little mistake in your theory could have a negative impact on our experiment [6]. A good hypothesis of study involves more than just a guess. In particular, a question that can be further examined through background study may start with your idea. The validity and results of your experiment depend on a robust testing hypothesis [7]. Developing a good testable hypothesis has few benefits, it forces us to reflect carefully and particularly on the results of a study. It therefore allows us to comprehend the implications of the topic and the various factors involved in the study.

X<-

C

(20,22,25,26,27,48,43,55,43,27,34,44,50,37,35,33,40,42,48,49,30,33,43,50,50,32,20,30,40,44,35,28,40,43,38,33,34,50,22,37,45,20,44,33,20,38,35,25,40)

Y<-

C

(23,34,44,45,29,27,35,45,44,30,39,50,50,33,45,47,20,32,22,28,30,34,44,47,30,44,29,40,30,40,24,50,46,30,27,50)

3

Summary(X); Summary (Y)

sd (X); sd (Y)

qqnorm (X, main="white"); qqline(X); boxplot(X)

qqnorm (Y, main="BAME"); qqline(Y); boxplot(Y)

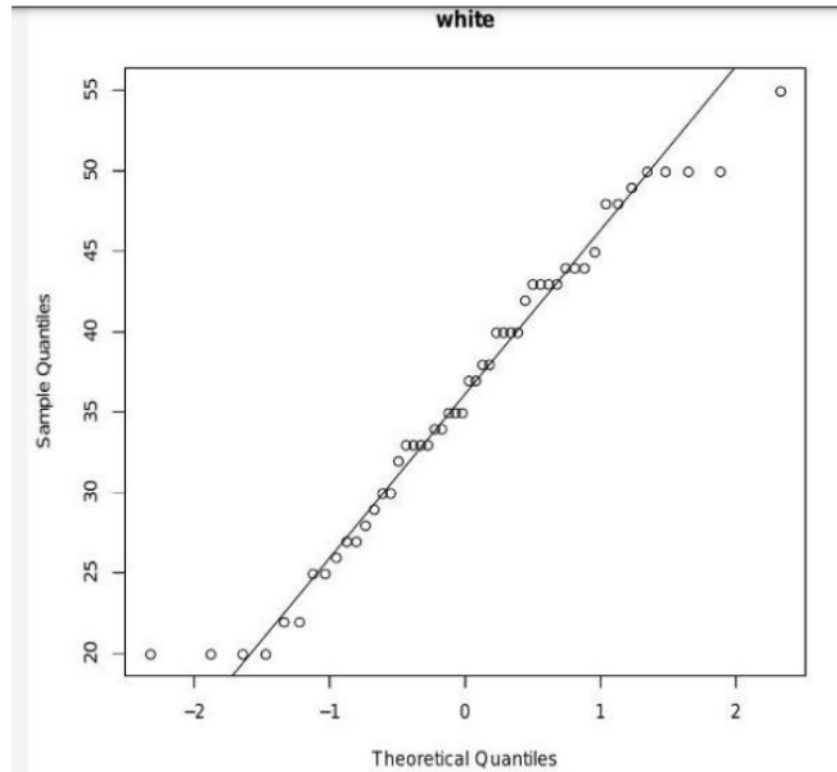
	Minimum.	1 st Qu.	Median	Mean	3rd Qu.	Maximum.
	20.00	29.25	36.00	36.18	43.00	55.00
Max.	20.00	30.00	34.50	36.74	45.00	50.00

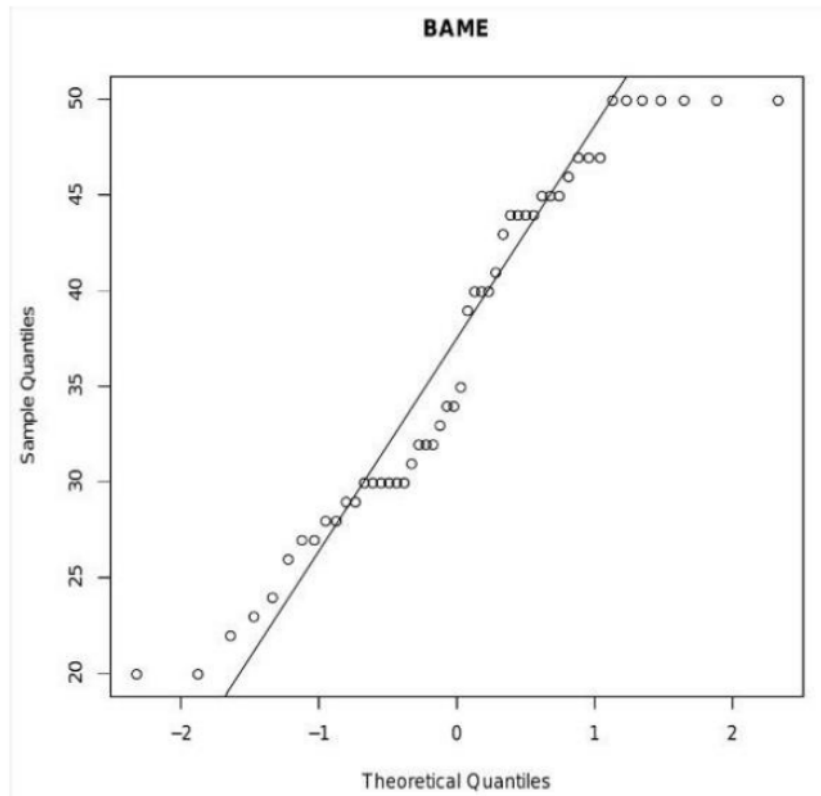
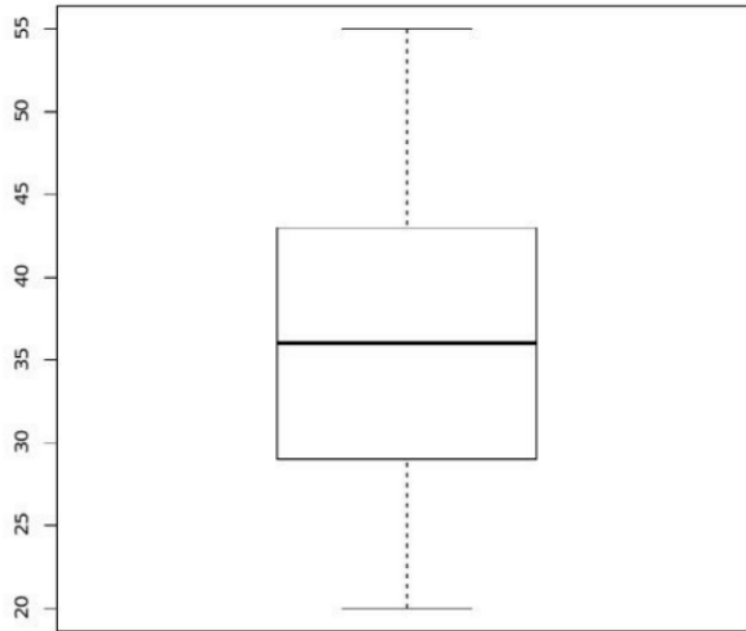
[1] 9.419238

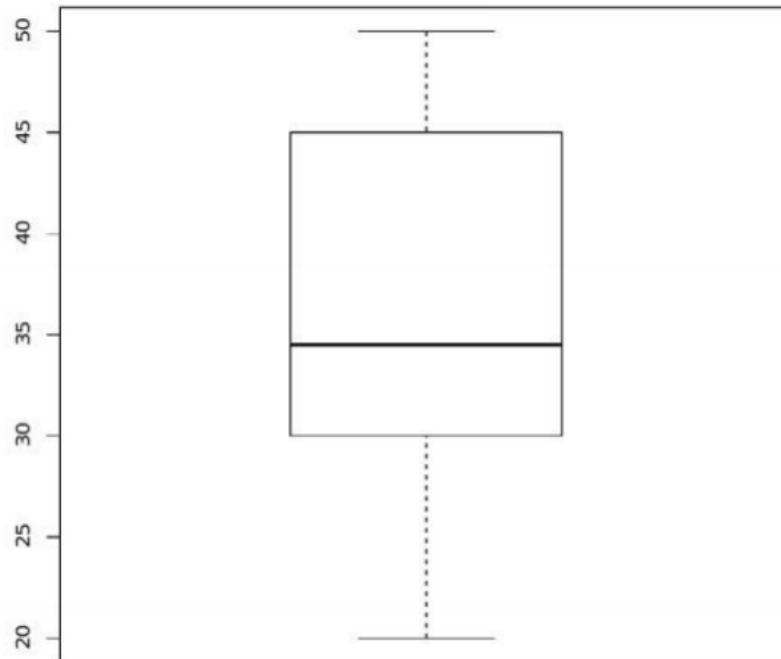
[1] 9.317112

Significant testing developed in the 1930s from the notion and practice of the great statistician, R.A. Fisher. His ideas are straightforward: imagine we have identified a link between poverty and malnutrition among children under the age of five. This is a discovery, but could it be an opportunity? Or perhaps we might analyses if a novel nutrition therapy improves malnourished children's nutritional status. We analyses a new treatment group for malnourished children and a comparable group of traditional nutritional therapies and discover a 2-units better nutritional status in the new therapy group compared with the old therapy group [6]. This finding is clearly welcome, however this finding can potentially be solely due to chance. Fisher therefore saw P as an index that measures the strength of evidence against the zero hypothesis (in our examples, the hypothesis that there is no association between poverty level and malnutrition or the new therapy does not improve nutritional status). To assess the strength of the evidence against the null hypothesis "He defended $P < 0, 05$ (5% importance) as the usual level to find evidence against the test hypothesis but not as an absolute rule" [5]. And the fisher never stopped but gave certain observations and results against a Blank or null

hypothesis. He suggested that "if value of P is 0.1 to 0.9 there is no cause to mistrust the testing hypothesis.







All the assumptions are validated hence there is no such factor that can be projected which would affect the results. Some of the observations of data are as follows, normal probability plots are drawn, for a data to test whether the data is normally distributed. In a normal probability plots of all the points lie on the normal line then we conclude that the data fits well with the normal probability distributions and hence these plots can be used for testing the normality of a sample the normal probability plot for white shows that the data is perfectly normal distributed [8]. Our data set may even be skewed by extremities in a controlled online A/B test. How do you handle them? Do you cut them out, or is there a different way? How can you even detect the presence and extent of outliers? If you optimize your revenue site, you should be concerned about outliers. This post examines the characteristics of outliers, how they are detected and popular approaches. Outliers may include significant data [3]. Or be meaningless aberrations induced by mistakes in measuring and documenting. In any event, they can produce problems with repeatable A/B test results, therefore outliers must be questioned and analyzed. While outliers appear in several analyses, you should be mostly worried with tests for optimizing revenue indicators such as average order value or income per visitor for conversion optimization [4].

If there are any outliers in the data, this graphs helps us to find out that too, outliers are found too, outliers are found to tie for the line (theoretical) line around which the points are around. No such observation in found in any of the normal probability plots that means both the samples are free from outliers, the plot for white is more normal than the plot for sample BAME, the normality or non-normality of data will not affect the decision of applying z test due to the large sample size [plots are mainly constructed to free whether there are any other outliers] for the data white Mean is very near to Median which again supports symmetry of the sample of whites. Where for "BAME" Mean is far from Median, so the data cannot be considered to be normally distributed, the boxplot gives five numbers summary of the data. We have already seen that just a small amount of information about a distribution is available through a measure of the central trend itself. To finish the description, you must have some understanding of how the scores of the central value are distributed. if they are widespread, as with the pass. Plays, we say high variability. If the core value is compactly distributed, as with we refer to the variability as low for the running plays. But high and low are terms without description. Accurate quantitative significance. Just as we needed to assess centrality quantitatively, so did us. Requires a variability quantitative index. This approach is used to display the value displacement along a distinct central time line. It shows gains over time in data from several categories. The size of each stream form in a stream chart is equal to the values of each category. Ideal for a large dataset. Data visualization technologies immediately become aware of a quantity of information. People can find things they don't know by means of the perfect instrument to view data. These devices can also be utilized to move data sets fast. The primary features are aimed towards huge data visualization applications. Another example is Space Titans 2.0, which aids with the exploration of the solar system. The goal was to use current spatial awareness to take a fresh look at how our world seems. [7].

Summary is (20,29,25,36,43,55) and for the data "BAME" it is (20,30,34,50,45,50).

The standard derivation of the data are almost equal for white $sd = 9.419$, the standard derivation of the data are almost equal to BAME $sd = 9.317$, hence they are eligible for testing equality of mean assuming equal variances. Some investors searching for a great profit possibility will participate in speculative undertakings with a large risk, but also

the possible payout. Others adopt an investing strategy in blue chip inventories, in which the revenue does not fluctuate like a yo-yo. Many other real-life decisions are based on extremes. Transportation systems and public services (e.g. gas, electricity, and water) must be prepared to satisfy the highest demand rather than the average demand, so as to avoid shortages and interruptions [9].

References

- [1] Abdullah, P. Y., Zeebaree, S. R., Jacksi, K., & Zeabri, R. R. (2020). AN HRM SYSTEM FOR SMALL AND MEDIUM ENTERPRISES (SME) S BASED ON CLOUD COMPUTING TECHNOLOGY. *International Journal of Research GRANTHAALAYAH*, 8(8), 56–64.
- [2] Ali, S. M., Gupta, N., Nayak, G. K., & Lenka, R. K. (2016). Big data visualization: Tools and challenges. 656–660.
- [3] Alzakholi, O., Haji, L., Shukur, H., Zebari, R., Abas, S., & Sadeeq, M. (2020). Comparison among Cloud Technologies and Cloud Performance. *Journal of Applied Science and Technology Trends*, 1(2), 40– 47.
- [4] Byron, L., & Wattenberg, M. (2008). Stacked graphs–geometry & aesthetics. *IEEE Transactions on Visualization and Computer Graphics*, 14(6), 1245–1252.
- [5] Caldarola, E. G., & Rinaldi, A. M. (2017). Big Data Visualization Tools: A Survey.
- [6] Caldarola, E. G., Sacco, M., & Terkaj, W. (2014). Big data: The current wave front of the tsunami. *Applied Computer Science*, 10(4), Article 4.
- [7] Chawla, G., Bamal, S., & Khatana, R. (2018). Big Data Analytics for Data Visualization: Review of Techniques'. *International Journal of Computer Applications*, 182(21), 37–40.

- [8] Chen, C. P., & Zhang, C.-Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences*, 275, 314-347.
- [9] Dino, H., Abdulrazzaq, M. B., Zeebaree, S. R., Sallow, A. B., Zebari, R. R., Shukur, H. M., & Haji, L. M. (2020). Facial Expression Recognition based on Hybrid Feature Extraction Techniques with Different Classifiers. *TEST Engineering & Management*, 83, 22319-22329

ORIGINALITY REPORT

6%

SIMILARITY INDEX

6%

INTERNET SOURCES

0%

PUBLICATIONS

1%

STUDENT PAPERS

PRIMARY SOURCES

1

www.ncbi.nlm.nih.gov

Internet Source

4%

2

ijsab.com

Internet Source

1%

3

dione.lib.unipi.gr

Internet Source

<1%

Exclude quotes On

Exclude matches Off

Exclude bibliography On