# **Drills With R On Point Estimates And Confidence Intervals**

Suraj Eswaran

University Of The Cumberlands

2024 Spring – Statistics For Data Science (MSDS -531 – A01) – First Bi-Term

Professor Dr. Danny T. Barnes

01/28/2024

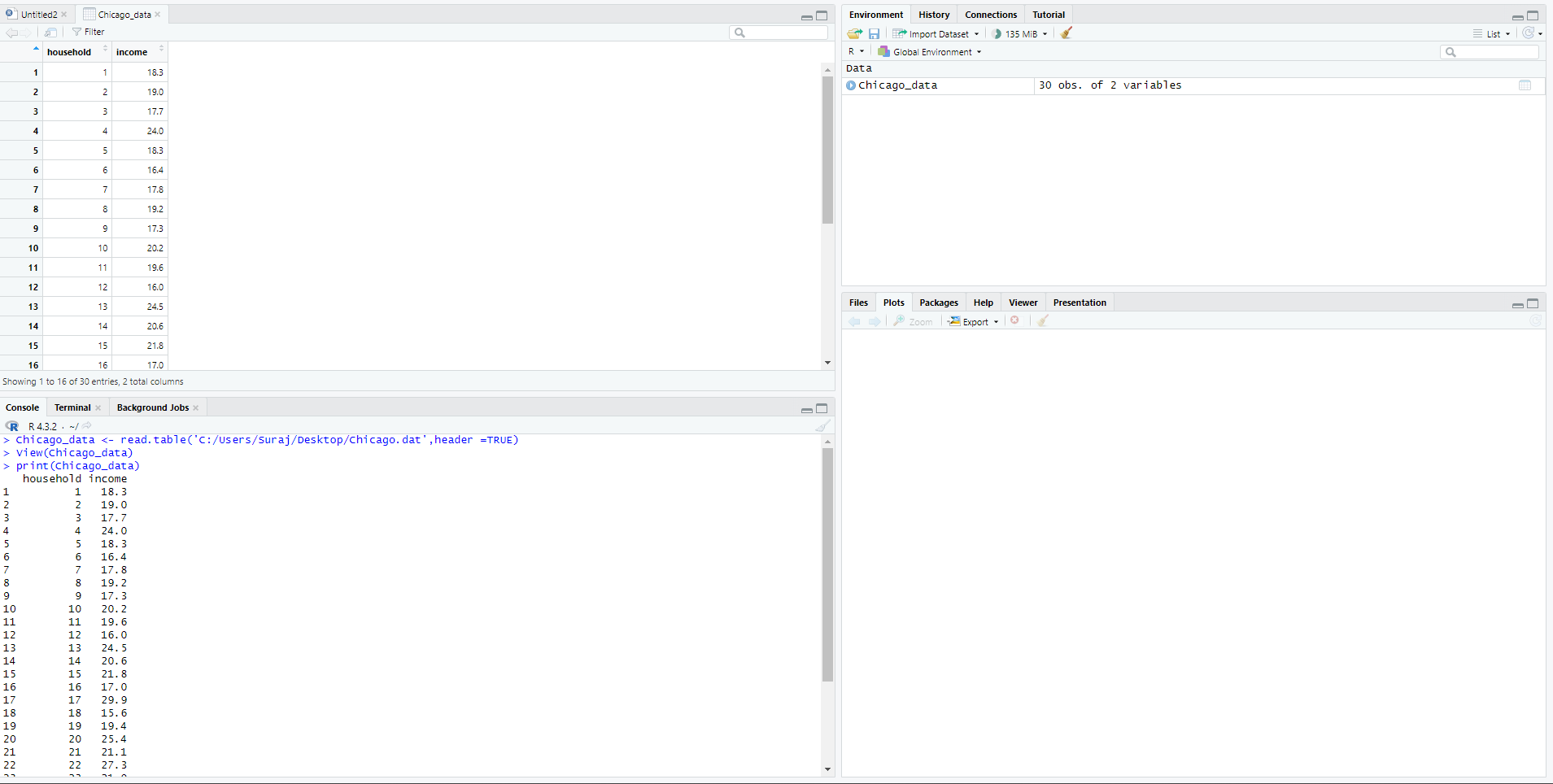
## Provide in the plain text R commands that finds/solves the following:

**A study investigates the distribution of annual income for heads of households living in public housing in Chicago. For a random sample of size 30, the annual incomes (in thousands of dollars) are in the Chicago data file.**

In Chicago data file, we have a sample size of 30 with list of annual income for heads of households who are living in Chicago. We are reading that data file using “read\_table” method where we are mentioning header as “TRUE” (Thulin, 2021, p. 120).



In order to validating the directory chicago\_data, we will use “view” and “print” where we will get to see annual income for various households in Chicago (Thulin, 2021, p. 38).

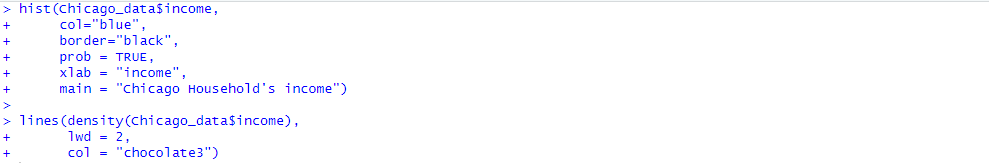


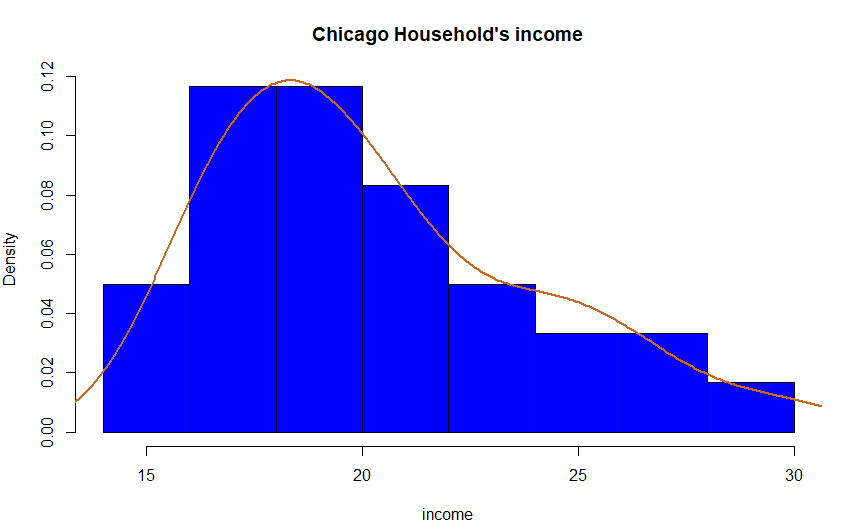
**Based on a descriptive graphic, describe the shape of the sample data distribution. Find and interpret point estimates of the population mean and standard deviation.**

From the summary function, we can find out that our minimum value as 15.60, median as 19.30, mean is equal to 20.33, maximum value is 29.90, first quartile and third quartile as 17.73 and 22.25 respectively.

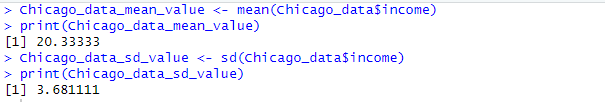


We have plotted a histogram graph along with line density with the help of “hist” and “line” , setting lwd equal to 2. From the density line, we can infer that the density curve is right skewed since the right-side tail is longer than left side tail, thus the value mean has to be greater than median.





For finding out population mean and standard deviation, we have used “mean” and “sd” function, where we have received 20.333 and 3.681 respectively.

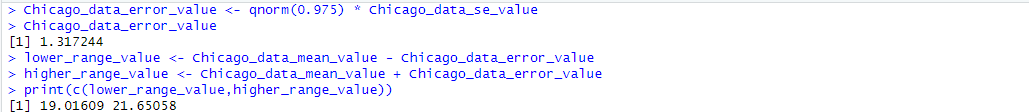


**Construct a 95% confidence interval for μ, using R software.**

Standard error is measured as standard deviation of a distribution which is formed by the sample means (Mbatha & Gustafsson, 2013). It can be measured as standard deviation by square root of number of samples obtained in this experiment.

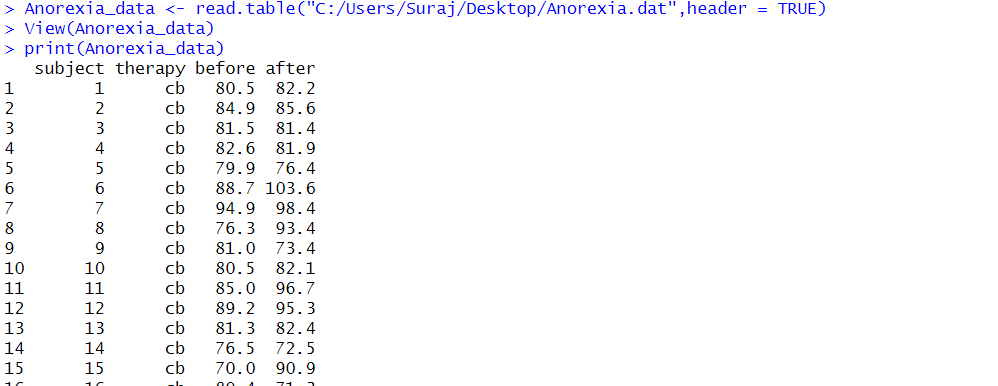


For calculating margin of error, we will calculate the product Standard error with normal score with the help of qnorm() function in R(Thulin, 2021, p. 249). Using this formula, we have obtained the margin of error as 1.31722. Now, over objective is to find lower and higher range value, where we have to add and subtract the resultant margin of error value with population mean. Thus, from the results we have infer that 95% confident is obtained for the Chicago households with their income is computed between the values 19.01609 and 21.65058.

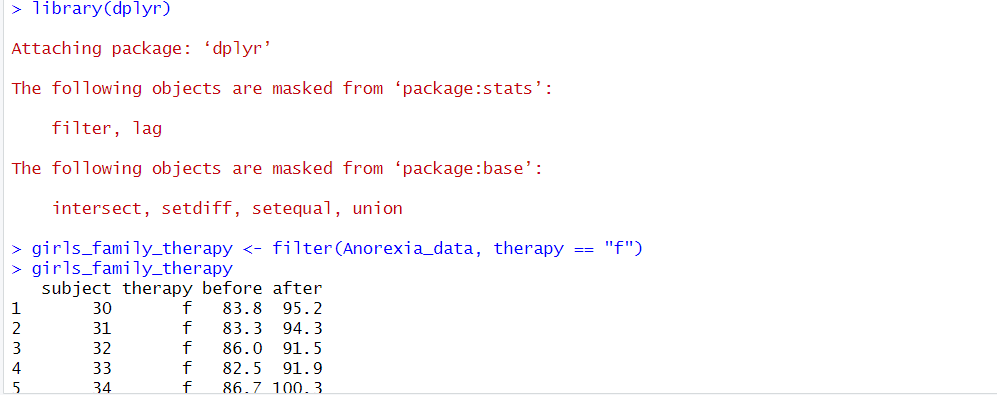


## The Anorexia data file contains results for the cognitive behavioral and family therapies and the control group. Using data for the 17 girls who received the family therapy:

In Anorexia data file, we have a list of 72 girls results who have received cognitive behavioral, and family therapies and the control group with their before and after values. We are storing that data file using “read\_table” method where we are mentioning header as “TRUE” (Thulin, 2021, p. 120).

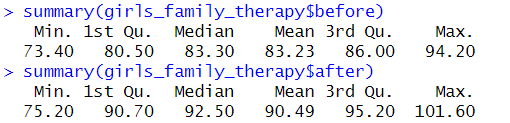


Our objective here is to list those list of girls who have received family therapy with their before and after values. In order to get that, we have utilized “filter” function where we filter function based on a column value(‘f’). Before, using filter it is necessary to have the library “dplyr” in our system by entering “library(dplyr)” (Thulin, 2021, p. 114).



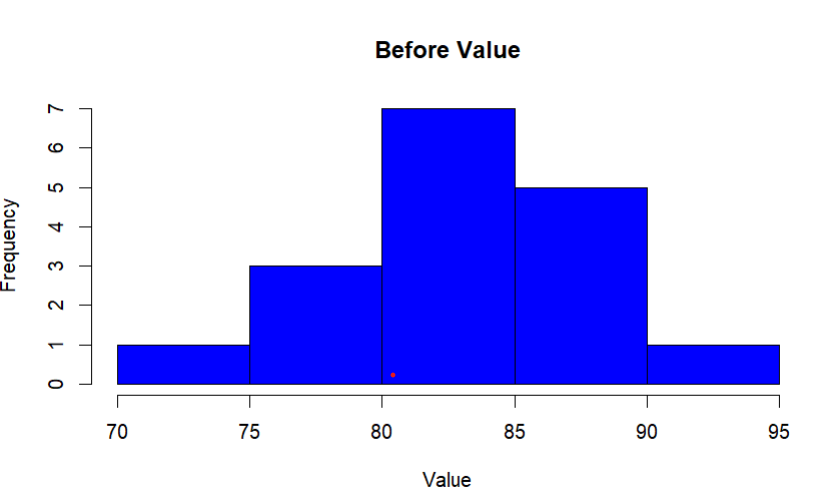
**Conduct a descriptive statistical analysis using graphs and numerical summaries.**

For numerical summaries, we have used the method “summary” for both before and after value for 17 girls who have received family therapy (Thulin, 2021, p. 39).



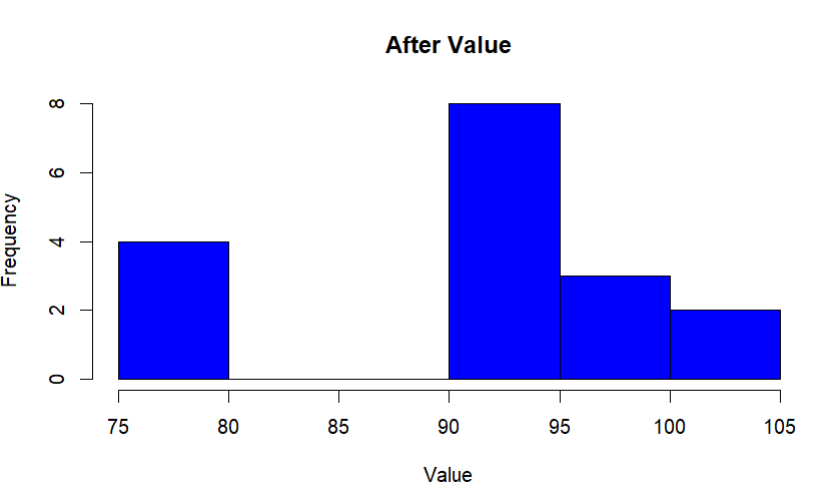
From the above two summaries, we have obtained mean value of before family therapy is lesser than after family therapy, which infers us that there is an increase in the trend of the value from before value and after value. Also, there has been an increase in value of median, first quartile and third quartile value, thus implies that there is an increase in trend of those values as well. We have plotted histogram for both before and after family therapy with the help of hist function(Thulin, 2021, p. 50), where we could find out that the trend of peak for before is different from peak value for after value, thus implies an increase in trend.





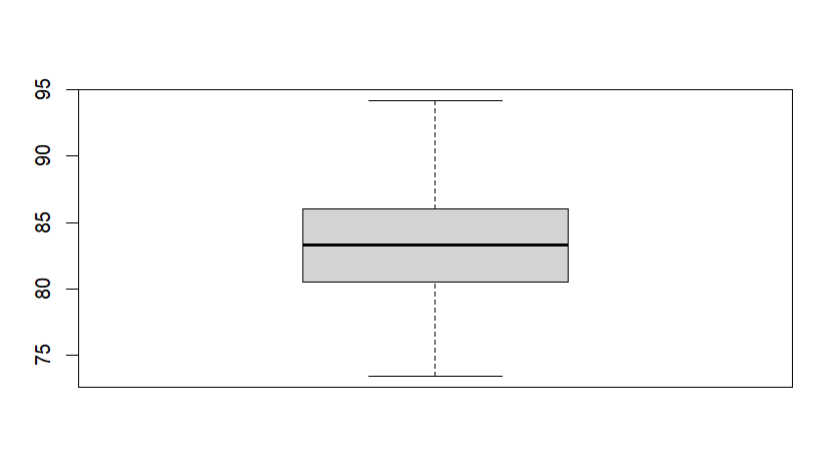
From the histogram of after value, we could notice an empty bin between the range 80 and 90 value, which shows that there has not been any observations or data points between those range.



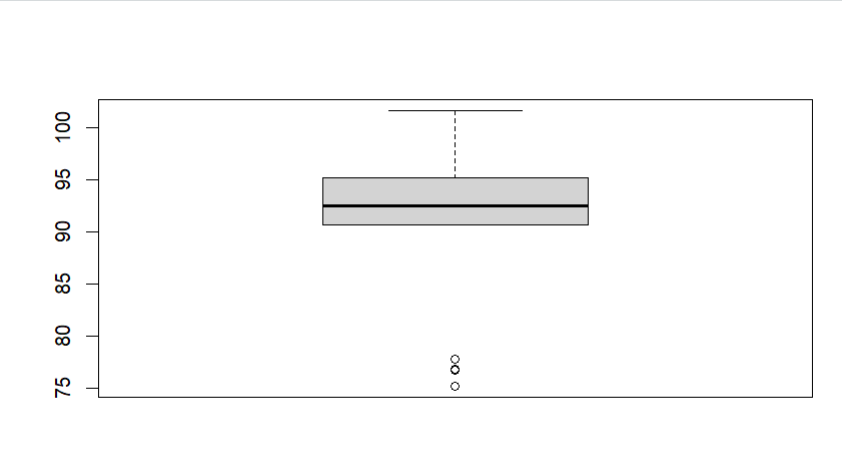


From the histogram of after value, we could notice an empty bin between the range 80 and 90 value, which shows that there has not been any observations or data points between those range. Along with histogram, we have also utilized box plot for graphical representation with the help of “boxplot” function (Thulin, 2021, p. 48).





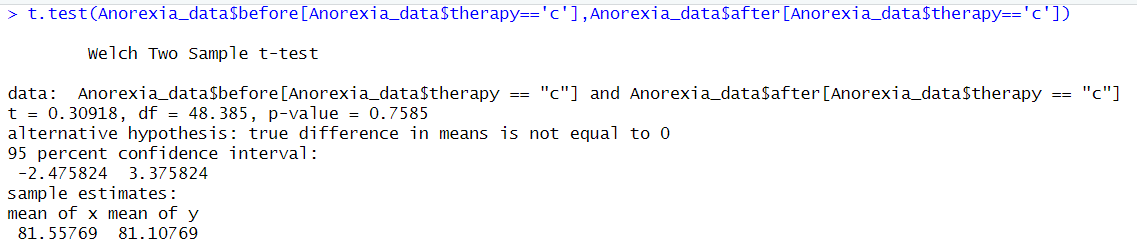




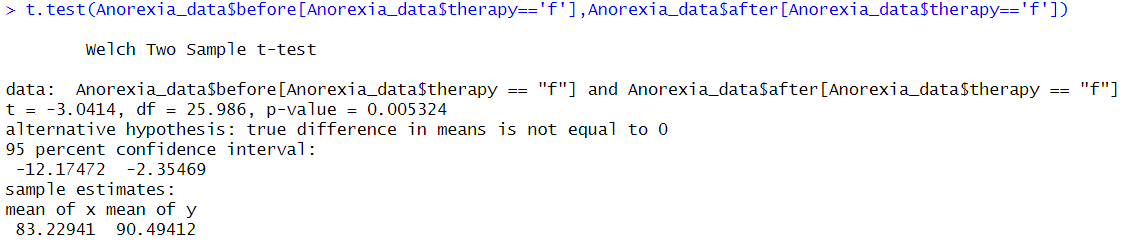
From the graph, we could notice the horizontal line(i.e.) median for before value is 83 whereas at after value close to 90, which depicts that an increase in trend from before and after value. Also, we could notice that there are lower outliers obtained since the range between 80 and 90 was not recorded.

**Construct a 95% confidence interval for the difference between the population mean weight changes for the family therapy and the control. summaries.**

In order to calculate 95% confidence interval between family therapy and control summaries, we have used t.test() function(Thulin, 2021, p. 74), which calculates one-sample t-test showing t value as 0.30918, df( degree of freedom) equal to 48.385, p-value as 0.7585.

****

From the above t.test() function for before value, we have obtained 95% confidence interval for value change between before and after control to be between -2.375824 and 3.375824 with mean value as 81.55769 and 81.10769 respectively whereas the 95% confidence interval for value change between before and after family therapy lies between -12.17472 and -2.35469 having mean value as 83.22941 and 90.49412 respectively. From these, we could infer that value increases drastically for family therapy than control method, although the samples are not dependent to each other, and they are normally distributed.



## References

Thulin, M. (2021). Modern Statistics with R: From wrangling and exploring data to inference and predictive modelling. BoD-Books on Demand.

Mbatha, C. N., & Gustafsson, M. A. (2013). The standard error of regressions: a note on new evidence of significance misuse. Agrekon, 52(1), 28-39.