Rasheeqa Isaacs

Student number: 3696142

Biostatistics Assignment

The average heights and weights of American women

Abstract

North America makes up good percentage of the world's population, as such, many Americans are overweight with an average weight being 80.7 kilograms and make-up 1.2 million tons of biomass. As a result, the majority of women overweight occur between the ages of 20 to 39 years of age. A series of coding was done in R which included testing for normality, visualizing the data using graph and doing statistical analyses using the analyses of variance (ANOVA) to test for great significance in the data. The results showed that there was great significance in the data in which the hypothesis was accepted. The alternative hypothesis states that the average weight of women increase with height.

Keywords: women, height, weight, BMI, ANOVA.

Introduction

North America is home to 6% of the world's population and contributes to 34% of the world's biomass, this is because of the occurrence of obesity (Walpole et al., 2012). Individuals in North America generally have the highest average body mass in the world. This was tested in 2012, as results showed that they had an average body mass of 80.7 kilograms. This showed that approximately 70% of American population was overweight, this was due to the biomass being 1.2 million tons (Walpole et al., 2012). The average weight of women in America over 20 years of age is 166 pounds, the average height is 5 feet and 4 inches. It was tested using the body mass index (BMI) as well as other body weight tests showed that the woman with the height of 5 feet 4 inches weigh more than there supposed weight, which is between 108 and 132 pounds. This may increase their risk of diseases and illnesses such as heart disease, diabetes and cholesterol (Colemen, 2010).

The importance of this study is to make sure women are healthy by having the correct weight that correlates with their height. The BMI is used to calculate whether they are overweight or not, this is to avoid women becoming overweight which later on leads to obesity (Healthy weight, overweight, and obesity among U.S. adults, 1960).

The limitations of the study is that the BMI is not accurate because it only includes the weight and height of an individual and do not consider factors such as bone density, muscle mass and the overall structure of the body (Nordqvist, 2013).

The aim of this experiment is to test whether height has an influence on the weight of women between 30 to 39 years of age in America.

The alternative hypothesis states that the average weight of women increase with height. The null hypothesis states that the average weight of women do not increase with height.

Methods

Samples of data was collected from American women between the ages of 30 to 39. Their height and weight was measured, these individuals was average American women who had ordinary clothing and shoes which did not affect the weight component and shoes alone did not affect the height of an individual. A statistical analyses was run using R, this included multiple packages being used such as Tidyverse, ggplot2 for creating the graphs. The Shapiro-test was used to determine whether the data is normal or not, this included the following code:

```
#Testing if the data is normal using the Shapiro Test
shapiro.test(women_america$weight)
```

```
        Shapiro-Wilk normality test

data:  women_america$weight
W = 0.88457, p-value = 0.05554
```

The data was visualized in a density graph using the code:

```
#Created a density plot to visualize the data
ggplot(data = women_america, aes(x = weight)) +
  geom_density(colour = "green", fill = "green", alpha = 0.6) +
  ggtitle("The average weight of American women between ages 30-39") +
  labs(x = "weight", y = "") +
  theme_new
```

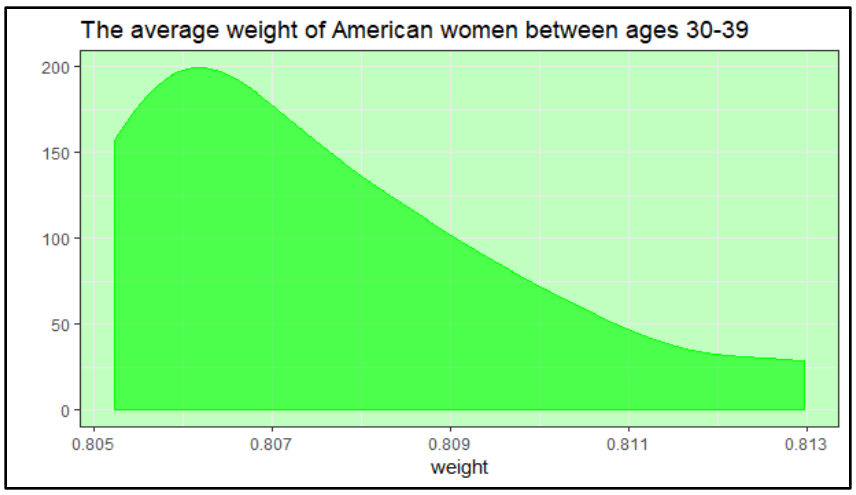

Figure 1: The density graph representing the average weight of women aged between 30-39 years

However, for more accuracy on normal distribution, a Cullen and Frey graph was produced:

```
#Testing if it is normal data using the Cullen and Frey graph
descdist(women_america$weight)
```

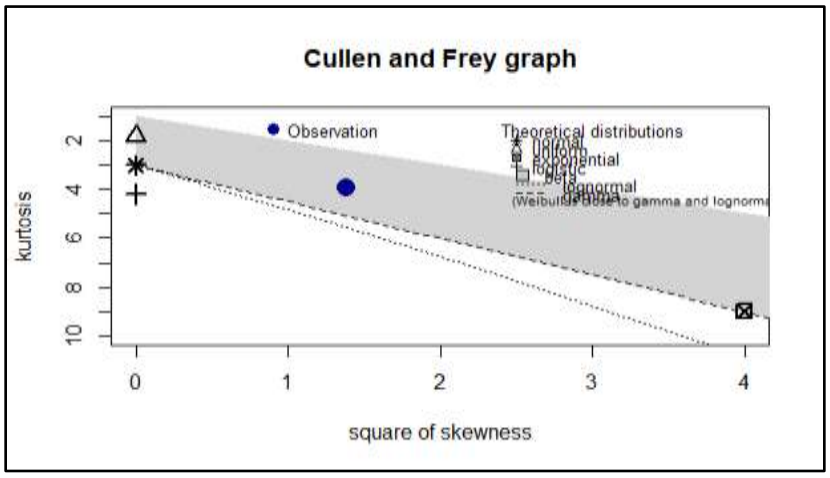

Figure 2: The Cullen and Frey graph showing the type of data distribution occurring

The statistical analyses that was used was the analysis of variance (ANOVA). This was used to test for significant difference within the dataset.

```
#ANOVA Test
#Hypothesis
#H0: The average weight of women do not increase with height.
#H1: The average weight of women increase with height.
women.aov <- aov(height~weight, data = women_america)
summary(women.aov)
```

```
             Df   Sum Sq   Mean Sq F value  Pr(>F)
weight        1 0.001021 0.0010207   9.264 0.00941 **
Residuals    13 0.001432 0.0001102
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

A single panel visual was used to view the statistical data:

```
#One panel visual would be done to view the statistical data
#Relationship between height and weight of women
#calculating Pearson r
r_print <- paste0("r = ",
                  round(cor(x = women_america$height, women_america$weight),2))

#Create a single panel visual
ggplot(data = women_america, aes(x = height, y = weight)) +
  geom_smooth(method = "lm", colour = "maroon2", se = F) +
  geom_point(colour = "darkorchid1") +
  geom_label(x = 0.54, y = 0.812, label = r_print) +
  labs(x = "height (in)", y = "weight (lbs)") +
  theme_new +
  ggtitle("One panel visual")
```

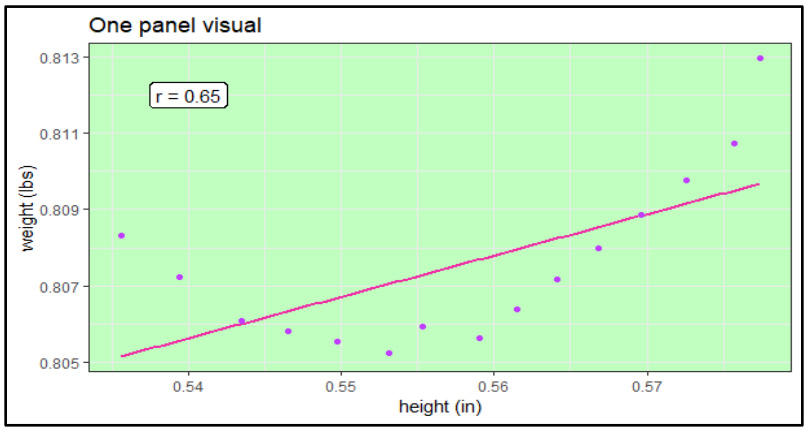

Figure 3: A one panel visual showing the correlation of the data

Results

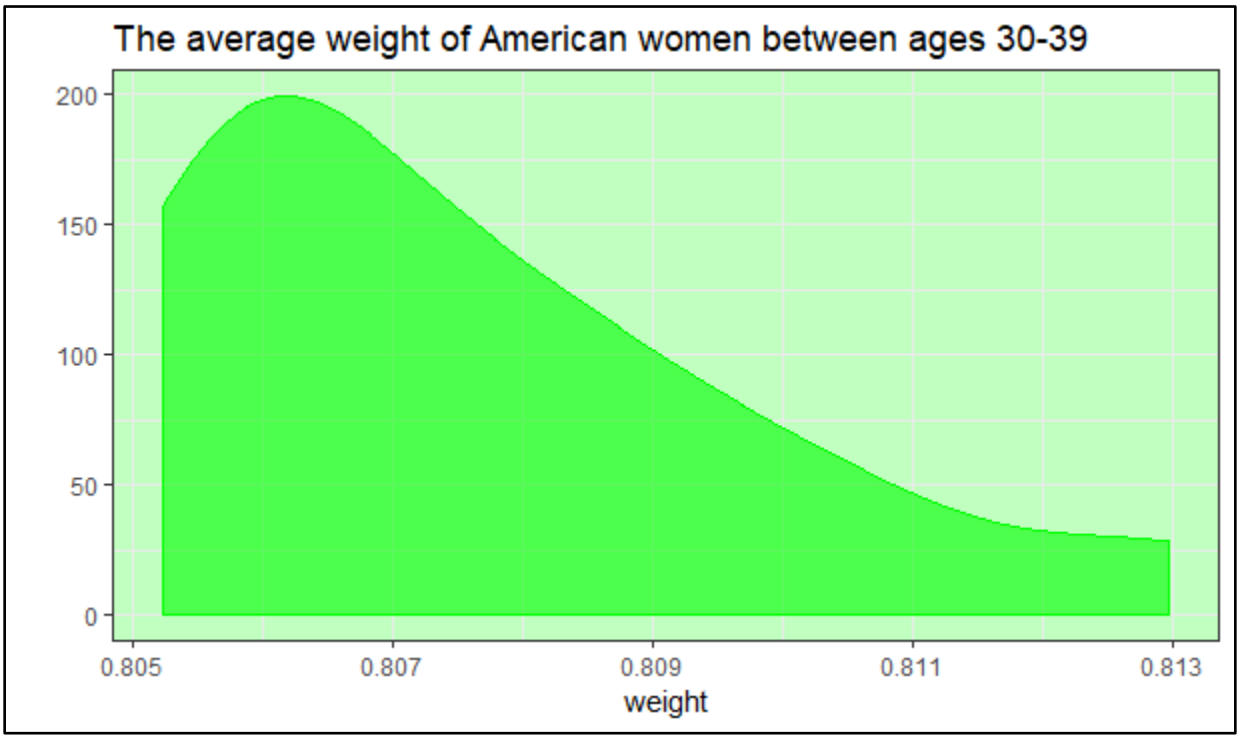

Figure 1: The density graph representing the average weight of women aged between 30-39 years

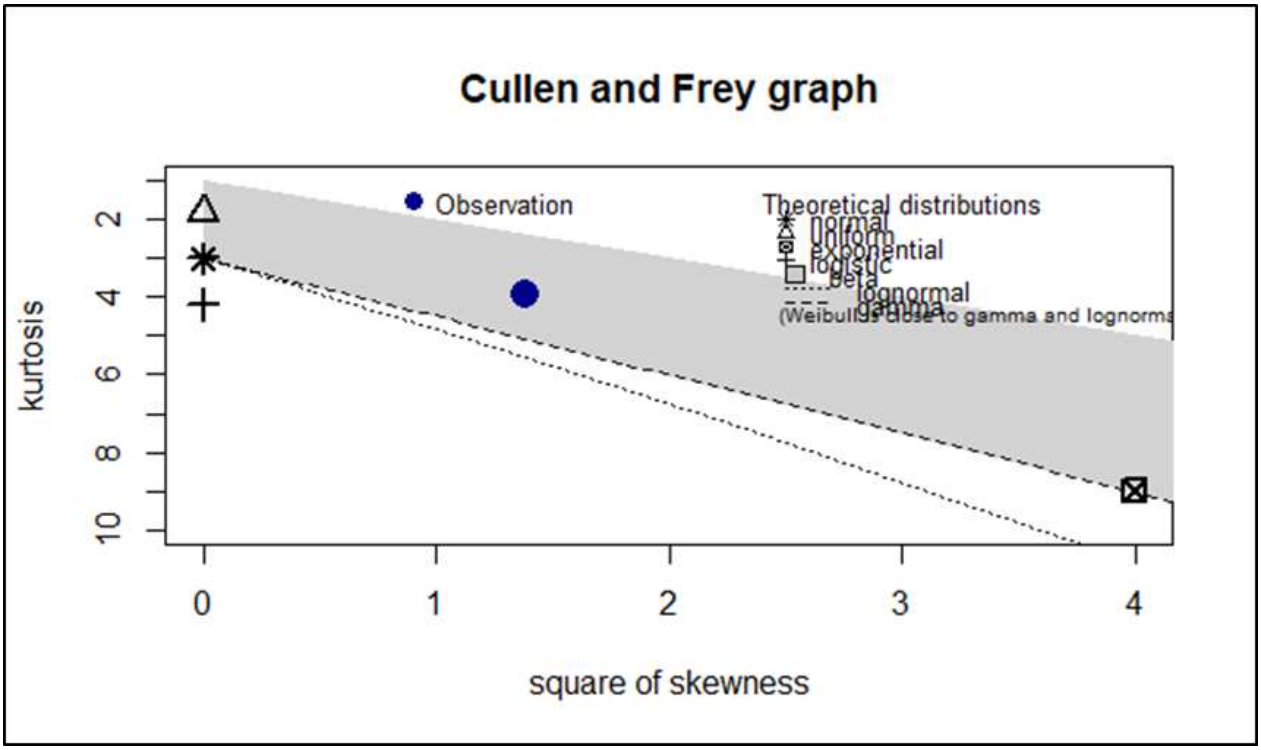

Figure 2: The Cullen and Frey graph showing the type of data distribution occurring

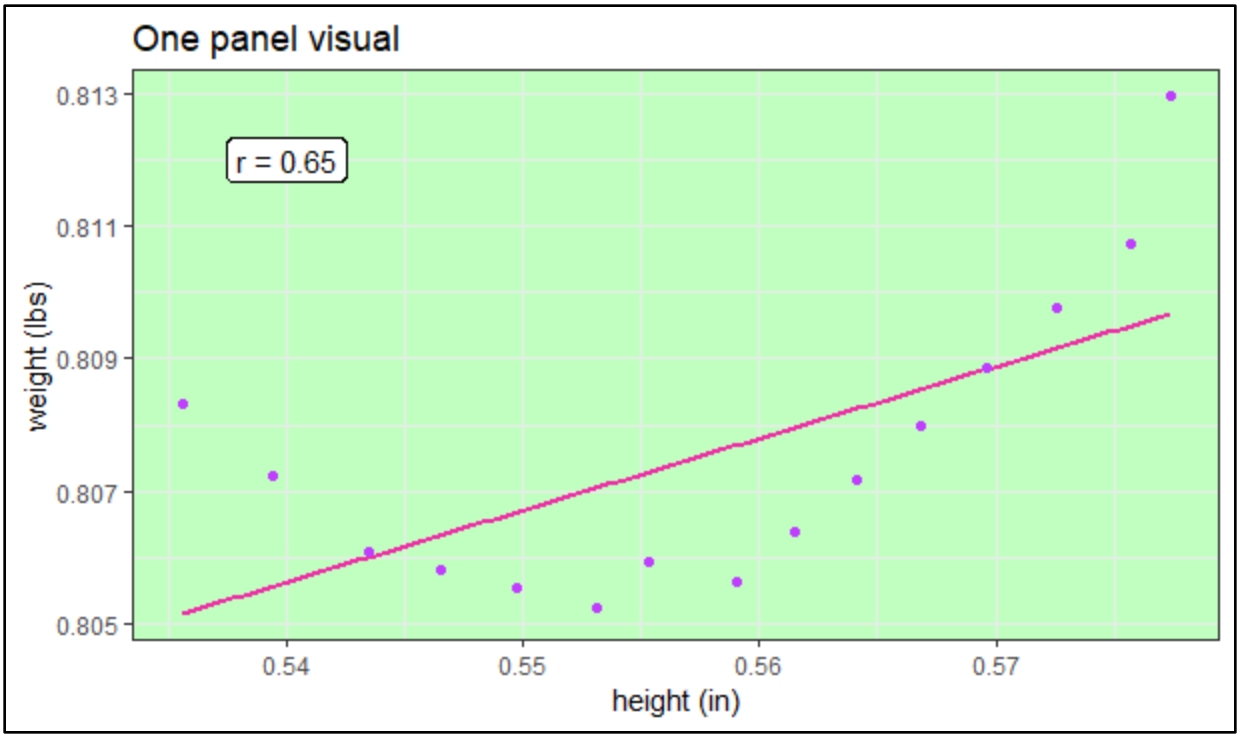

Figure 3: A one panel visual showing the correlation of the data

Discussion

The data was tested for normal distribution using the Shapiro's test and the results showed that the p-value was more than 0.05. This made it of normal distribution, because the value was 0.6986. A density graph was made to confirm it, however the graph [figure 1] did not show normal distribution. It was then converted from a non-normal to normal distribution using the function **decostand**. For more accuracy a Cullen and Frey graph was made and the results also confirmed that the data was of normal distribution [figure 2]. A statistical analysis was done using ANOVA, the results showed that there was a significant difference between the average height and weight of women in America. This is because the p-value was less than 0.05, resulting in the acceptance of the alternative hypothesis and the rejection of the null hypothesis. However, the TUKEYHSD model could not run and show where the exact significance in the data occurs, the reason for this was that the dataset is too small. To visualize the relationship between height and weight data and show how correlated the data is we did a one panel visual, this graph [figure3] showed that there was a dip [weight loss] as the height increased, however after a brief period of time the weight gradually increased with height.

These results are important as it shows the relationship between the individual's height and weight and allows the individual to check their BMI based on the values of their weight and height, which shows whether they are overweight, underweight or the correct weight.

A similar study was done in America, between 1960-1962, it included a health examination survey of men and women between 18-79 years of age. The results for women showed that the average height and weight of women increased from 129 pounds at 58 inches to 157 pounds at 67 inches (Washington *et al.*, 1966). These results is similar to this study in that the weight increases with height as well.

This study's limitations was that the dataset was too small, as a result more data could have been collected for the study, instead of only having 15 observations.

Conclusion

The alternative hypothesis is accepted as results showed that an increase in the average weight, increases the average height. The ANOVA test showed significant difference in the dataset, however, the exact place of significance could not be done using TUKEYHSD as it could not run, one of the reasons being that the dataset was too small.

References

1. Colemen, E. (2010). *The Average Body Weight for Women | Livestrong.com*. [online] LIVESTRONG.COM. Available at: https://www.livestrong.com/article/264142-the-average-body-weight-for-women/ [Accessed 10 May 2019].
2. Healthy weight, overweight, and obesity among U.S. adults. (1960). [online] Available at: https://www.cdc.gov/nchs/data/nhanes/databriefs/adultweight.pdf [Accessed 10 May 2019].
3. Nordqvist, C. (2013). *Why BMI is inaccurate and misleading*. [online] Medical News Today. Available at: https://www.medicalnewstoday.com/articles/265215.php [Accessed 10 May 2019].
4. Walpole, S.C., Prieto-Merino, D., Edwards, P., Cleland, J., Stevens, G. and Roberts, I. (2012). The weight of nations: an estimation of adult human biomass. *BMC Public Health*, 12(1).
5. Washington, D., Stewart, W. and Gardner, J. (1966). *Weight by Height and Age of Adults United States*. [online] Available at: https://www.cdc.gov/nchs/data/series/sr_11/sr11_014acc.pdf [Accessed 12 May 2019].