Difference in Perceived Body Weight Between Males and Females

Danny Kim

3/25/2020

REPORT

INTRODUCTION

Health is the center of everyone's lives, and one of the ways to measure an individual's health is through body weight. With a body weight that is not too high or not too low, it is a great indicator that a person is healthy. However, people tend to have the wrong idea at times, where they have different expectations about their weight. For example, some individuals may believe they need to lose more weight than the expected weight. This is very common in society, especially among certain genders. There is a belief that women tend to be more conscience more about their body weight than men. As a result, we want to ask the question: do men and women think differently about their body weight? To address this question, we will use data from the Behavioral Risk Factor Surveillance System (BRFSS), a survey of 350,000 United States citizens collected by the Centers for Disease Control. There are variables, such as gender, weight, desired weight, and general health status, that will help us answer the question. Before the data analysis, the hypothesis is that men and women will have different views about their body weight. We will attempt to see the difference of actual weight and desired weight for each gender. Hopefully, we will see that there is a significant difference for the two variables in one gender and not in the other, as this would reveal that they have different expectations for their body weight.

METHODS

To conduct the hypothesis testing, we have taken a random sample size of 300 from the 20000 observations. From the sample, we will use the variables of gender, weight, and desired weight. We want to subset the sample into two using the gender variable: males and females. The weight variable will tell us their actual weight, and the desired weight will tell us how much they want to weigh. We will be able to see whether they are content with their body weight or not by seeing the difference between the two variables. If the difference is not significantly different from 0, it would provide evidence that they are content with their weight, and if the difference is significantly different from 0, it would suggest that they are not content with their weight. To carry out the actual data analysis portion, we would need to conduct a paired t-test, where we compare each individuals actual weight and desired weight. This would give us a better estimate of how different genders view their body weight, as we are comparing the data on an individual by individual basis. We would find the average discrepancy between their weight and desired weight for each gender,

rather than finding the average of the difference between the average weight and average desired weight. The hypotheses for our testing are as follows:

$H_0 : \mu_m = 0$ and $\mu_f = 0$ $H_A : \mu_m \neq 0$ and $\mu_f \neq 0$

However, if they both reject the null hypothesis or both fail to reject the null, we want to compare the two subsets with each other to see if there are any differences. Since the number of female surveyees and male surveyees are not equal in the sample, we will conduct an two sample t-test comparing the difference of actual weight and desired weight in the two genders. In this case, the hypotheses will be:

$H_0 : \mu_m = \mu_f$ $H_A : \mu_m \neq \mu_f$

In all cases, the significance value of α will be .05., so the confidence intervals will be calculated at a 95% level.

RESULTS

Descriptive Results 1

From the data summary for the male weight and desired weight, it is evident that the mean and the median of the male weight is slightly greater than for the male desired weight. This can also be observed in the boxplot labeled "Actual vs. Desired Weight in Males." Similar to the males, the mean and the median of the female weight is slightly greater than the desired weight as well, which can be observed from its respective boxplot. One thing that is noticeable is that the spread of the actual weights is greater than the spread of the desired weights for both genders. From these observations, there are hints that both genders may want to weigh less than they actually do. In addition, it seems that the desired weight varies less, which suggest that many people do not want to weigh more or less than the healthy range.

Analytic Results 1

To analyze the data, we ran a paired t-test for both subsets of gender to see whether there was significant evidence that actual weight differed from desired weight. For males, the p-value was 1.109×10^{-5}, which is extremely lower than the significance value of .05. Therefore, we are able to reject the null hypothesis that there is no difference in actual weight and desired weight for males, which provides evidence that there is a significant difference between the two variables. Furthermore, since the 95% confidence interval for the difference is (5.63, 14.29), which contains all positive numbers, there is evidence that males want to weigh less than their actual weight.

For females, the p-value was even lower at 6.778×10^{-15}. Since the p-value is less than the alpha, this shows that there is significant evidence for us to reject the null, and supports the idea that there is a significant difference between the actual weight and desired weight in females. Similar to males, their 95% confidence interval, (16.27,25.81) contains only positive numbers, which suggest that females also want to weigh less than their actual weight.

Since both subsets turned out to have evidence that supports wanting to weigh less than their actual weight, we cannot conclude that men and women think differently about their body weight. As a result, we need to test if there is a difference between the discrepancy of actual vs. desired weight in males and females directly. However, since the confidence interval for females was higher, we can predict that females will want a slightly lower weight than their actual weight than males.

Descriptive Results 2

If we see the numerical summary for the difference in weight and desired weight for males, it can be observed that the mean of the data 9.96. This means that males, on average, want to weight 9.96 lb less than their actual weight. On the other hand, females have a mean of 21.04, which is much higher than the male average. This can also be observed through the boxplot, where you can see that the spread of females is generally higher on the boxplot than males. This suggests that females generally want to weigh less than their actual weight than males.

Analytic Results 2

To analyze this phenomena further, we ran a t-test to see if the difference between actual weight and desired weight in males and females significantly differed from each other. The p-value that was calculated was .000778, which is lower than the significance level of .05. As a result, there is significant evidence that the difference of actual weight and desired weight in males differs from females. Also, since the 95% confidence interval is (-17.5,-4.65), which contain only negative values, there is evidence that females have a lower relative desired weight than males.

DISCUSSION

From the results that were collected from this testing, males and females both wanted to weigh less than their actual weight. After analyzing this further, it was found that females wanted to weigh less than their actual weight with a greater magnitude than males. Going back to our initial question of, "Do men and women think differently about their body weight," there are evidence that supports that they do think differently. Even though they both want to weigh less, women want to decrease their weight more than males. This observation is reasonable, as it may be attributed to certain societal factors. For example, there are misleading ideas in society where people expect women to be "skinny." Consequently, it makes sense that our results showed that women had a greater discrepancy of actual weight and desired weight than males. One of the limitations that we had in this study was that the data used for this test was a survey from United States citizens only. There could have been different results if we incorporated data from other countries or regions, as they might have different societal factors. Also, there are other variables that need to be taken into account such as family dynamics, culture, and exposure to social media, to name a few, to truly figure out if gender is the only variable that affects how people view body weight. To tackle this problem, one of the future studies that can be done to further analyze this finding is to take in the societal factors into account. By controlling for these variables, we can see a more accurate relationship between viewpoint on body weight and gender.

APPENDIX

METHODS

```
# load the data
source("http://www.openintro.org/stat/data/cdc.R")

# create cdc.sample
sample.size = 300
set.seed(3502)
sample.rows = sample(1:nrow(cdc), sample.size)
cdc.sample = cdc[sample.rows, ]

# subset data
cdc.male = subset(cdc.sample, cdc.sample$gender == "m")
cdc.female = subset(cdc.sample, cdc.sample$gender == "f")
```

RESULTS

Descriptive Results 1 (Data Exploration)

```
# male weight
summary(cdc.male$weight)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    97.0   160.0   179.0   184.9   200.5   350.0
```

```
# male desired weight
summary(cdc.male$wtdesire)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   105.0   160.0   172.5   175.0   190.0   280.0
```

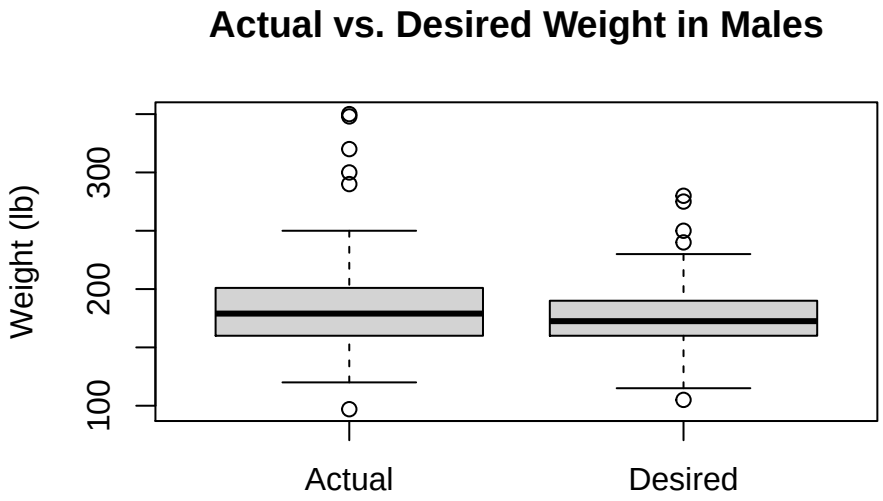
```
# female weight
summary(cdc.female$weight)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    85.0   124.8   147.5   153.4   175.0   320.0
```

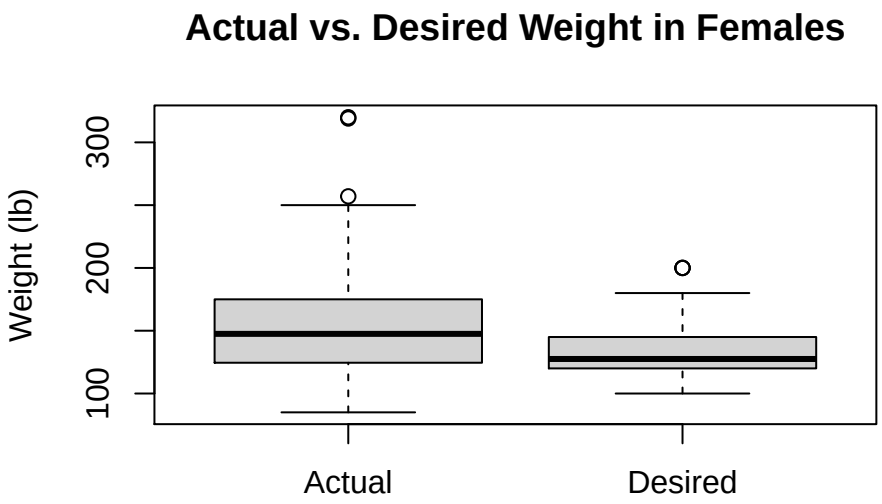
```
# female desired weight
summary(cdc.female$wtdesire)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   100.0   120.0   127.5   132.3   145.0   200.0
```

```r
# males
boxplot(cdc.male$weight, cdc.male$wtdesire,
        main = "Actual vs. Desired Weight in Males",
        ylab="Weight (lb)",
        names = c("Actual","Desired"))
```




```r
# females
boxplot(cdc.female$weight, cdc.female$wtdesire,
        main = "Actual vs. Desired Weight in Females",
        ylab="Weight (lb)",
        names = c("Actual","Desired"))
```



Descriptive Results 2 (Data Exploration)

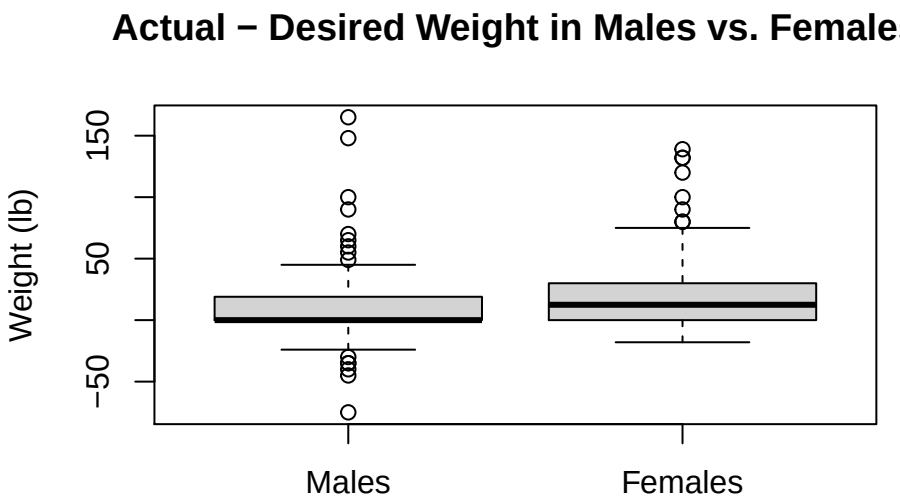
```
# weight - desired weight for males
summary(cdc.male$weight - cdc.male$wtdesire)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -75.000   0.000   0.000   9.962  18.500 165.000
```

```
# weight - desired weight for females
summary(cdc.female$weight - cdc.female$wtdesire)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -18.00    0.00   12.50   21.04   30.00  139.00
```

```
boxplot(cdc.male$weight - cdc.male$wtdesire, cdc.female$weight - cdc.female$wtdesire,
        main = "Actual - Desired Weight in Males vs. Females",
        ylab="Weight (lb)",
        names = c("Males","Females"))
```



Analytic Results 1 (Data Analysis)

```
# t-test for males
t.test(cdc.male$weight, cdc.male$wtdesire,
       alternative = "two.sided", paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  cdc.male$weight and cdc.male$wtdesire
## t = 4.543, df = 155, p-value = 1.109e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   5.630034 14.293043
## sample estimates:
## mean of the differences
##                9.961538
```

```
# t-test for females
t.test(cdc.female$weight, cdc.female$wtdesire,
       alternative = "two.sided", paired = TRUE)
```

```
##
##  Paired t-test
##
```

```
## data:  cdc.female$weight and cdc.female$wtdesire
## t = 8.7104, df = 143, p-value = 6.778e-15
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  16.26659 25.81674
## sample estimates:
## mean of the differences
##                21.04167
```

Analytic Results 2 (Data Analysis)

```r
# t-test for males and females
t.test(cdc.male$weight - cdc.male$wtdesire,
       cdc.female$weight - cdc.female$wtdesire,
       alternative = "two.sided", paired = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  cdc.male$weight - cdc.male$wtdesire and cdc.female$weight - cdc.female$wtdesire
## t = -3.3962, df = 292.52, p-value = 0.0007777
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -17.501000  -4.659256
## sample estimates:
## mean of x mean of y
##  9.961538 21.041667
```