Assignment3 Report

Sean Henner

02/11/2021

STAT1378 Project Report

Investigating Relationship Between Height and Weight Between Human Males and Females

Introduction

The purpose of this report is to display my knowledge and skills of creating a report of proper structure with the right amount of data and analysis.

The purpose of this analysis is to find if a correlation exists between height and weight of people who do different amounts of physical activity. If a correlation is found, we can determine that there is some biological reasoning for this relationship and be able to accurately predict the healthy weight of individuals from their height.

Data

The data, project.csv, was provided by the unit. The data was sampled out of a random selection of men and women aged between 26-45.

The data contains 1000 entries that contain the columns as follows:

- ID - A unique identifier for the entry
- gender - Gender of the entry, between 'Male' and 'Female'.
- height - Height of the entry, recorded in cm of two decimal places.
- weight - Weight of the entry, recorded in kg of two decimal places.
- phys - The level of physical activity the entry does, between 'None', 'Moderate', and 'Intense'.

Summary of the data set:

```
##       ID              gender             height          weight      
##  Length:1000        Length:1000        Min.   :158.7   Min.   : 37.54  
##  Class :character   Class :character   1st Qu.:167.9   1st Qu.: 61.08  
##  Mode  :character   Mode  :character   Median :173.2   Median : 68.17  
##                                        Mean   :173.1   Mean   : 68.56  
##                                        3rd Qu.:178.2   3rd Qu.: 75.75  
##                                        Max.   :190.0   Max.   :108.08  
##      phys          
##  Length:1000
```

```
##  Class :character
##  Mode  :character
##
##
##
```

Summary of Females only from data set:

```
##       ID                gender             height          weight
##  Length:493         Length:493         Min.   :158.7   Min.   :37.54
##  Class :character   Class :character   1st Qu.:165.3   1st Qu.:58.19
##  Mode  :character   Mode  :character   Median :167.9   Median :64.14
##                                        Mean   :168.1   Mean   :64.60
##                                        3rd Qu.:170.6   3rd Qu.:71.04
##                                        Max.   :181.2   Max.   :99.87
##      phys
##  Length:493
##  Class :character
##  Mode  :character
##
##
##
```

Summary of Males only from data set:

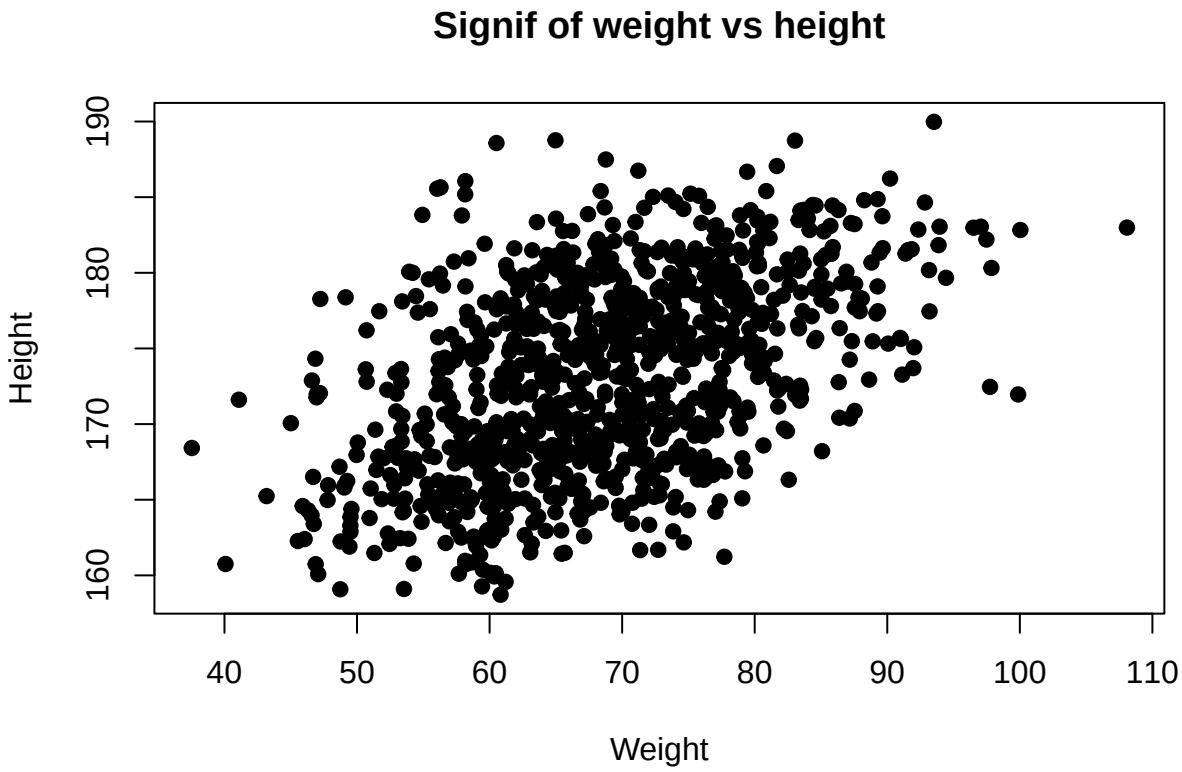
```
##       ID                gender             height          weight
##  Length:507         Length:507         Min.   :165.0   Min.   : 46.60
##  Class :character   Class :character   1st Qu.:175.3   1st Qu.: 65.55
##  Mode  :character   Mode  :character   Median :178.1   Median : 72.10
##                                        Mean   :178.0   Mean   : 72.41
##                                        3rd Qu.:180.5   3rd Qu.: 78.82
##                                        Max.   :190.0   Max.   :108.08
##      phys
##  Length:507
##  Class :character
##  Mode  :character
##
##
##
```

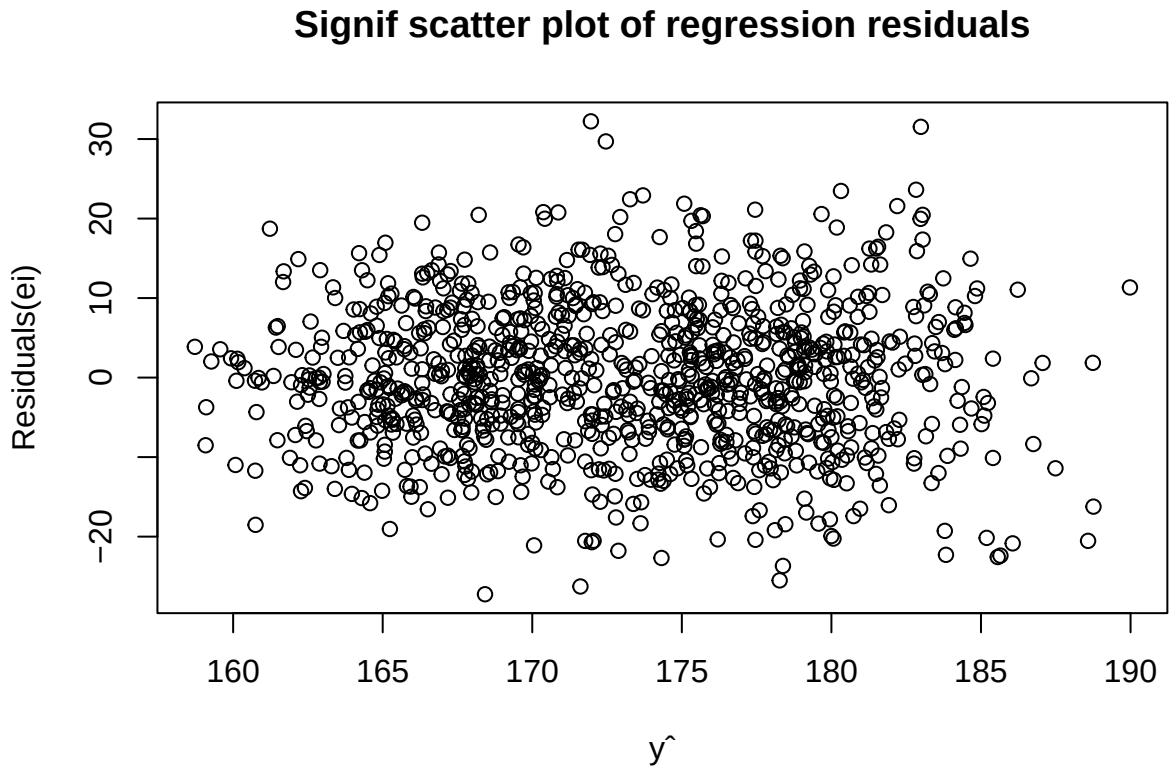
Chi-squared tests for independence

Null and alternate hypothesis: H_0 = There is no correlation between weight and height

H_1 = There appears to be a correlation between weight and height

Plots involving weight and height



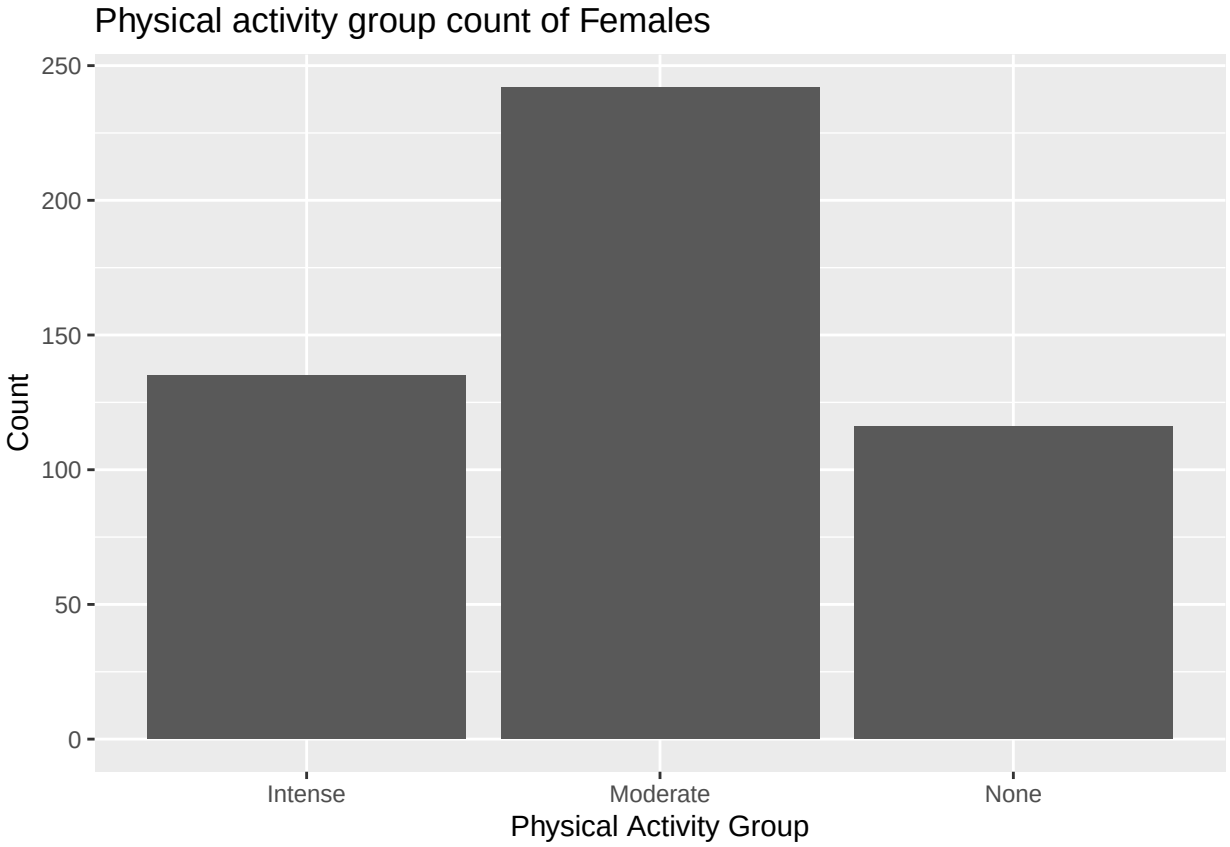


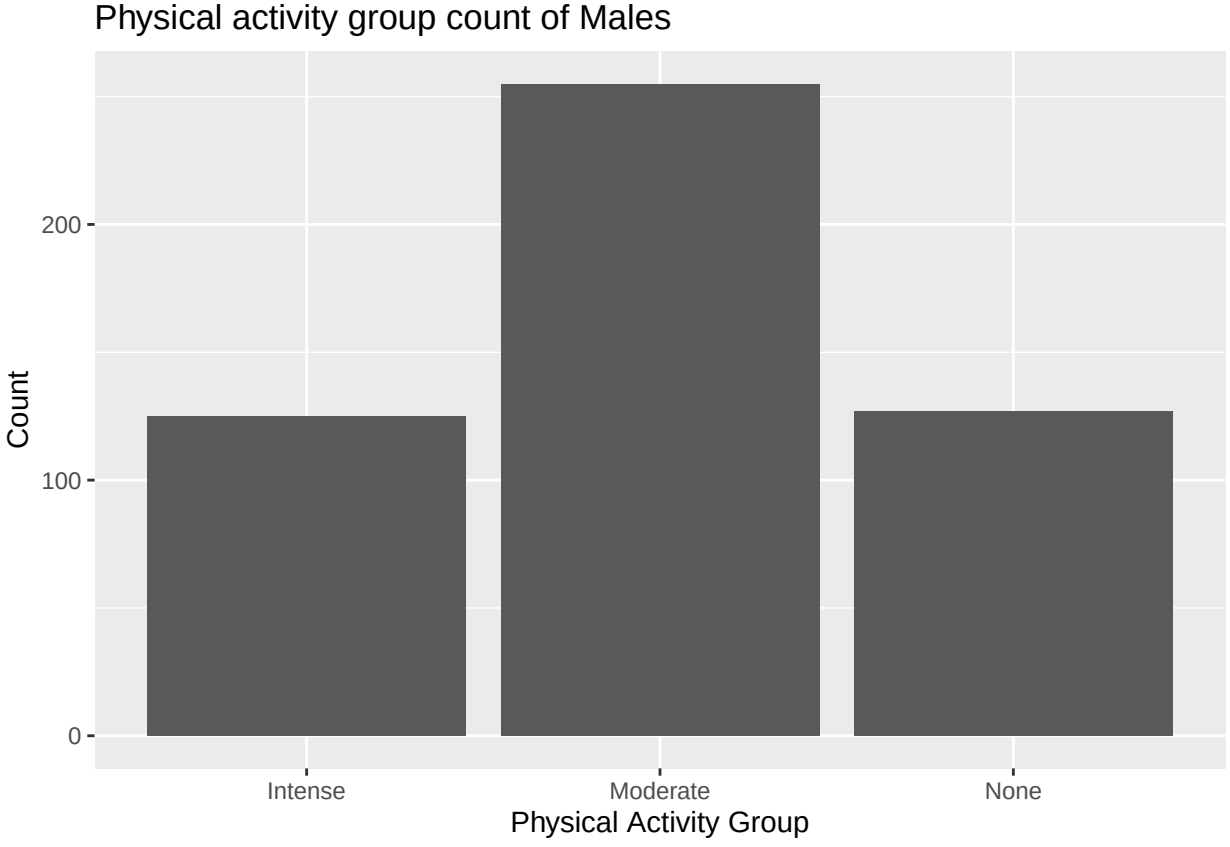
Calculations and test involving weights and height

```
##
##  Pearson's product-moment correlation
##
## data:  data$weight and data$height
## t = 17.586, df = 998, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.4376015 0.5323406
## sample estimates:
##       cor
## 0.4863994
```

```
##
##  Pearson's Chi-squared test
##
## data:  data$weight and data$height
## X-squared = 705090, df = 704576, p-value = 0.3323
```

Plots involving gender and physical activity





Calculations and tests for gender and physical activity

```
##
##  Pearson's product-moment correlation
##
## data:  data.phys$females and data.phys$males
## t = 6.4532, df = 1, p-value = 0.09787
## alternative hypothesis: true correlation is not equal to 0
## sample estimates:
##       cor
## 0.9882056
```

```
##
##  Pearson's Chi-squared test
##
## data:  data.phys$females and data.phys$males
## X-squared = 6, df = 4, p-value = 0.1991
```

Methods

Packages used include:

- tidyverse
- ggplot
- dplyr
- stats

Pearson's product-moment correlation and Pearson's chi-squared tests were used throughout the calculations to determine p values, correlation scores, and other values.

Results

Relationship Between Height and Weight

```
## Some points from height and weight are close to the line but other points are far from it,
## which indicates only a moderate linear relationship between the variables.
```

Mean Height Between Male and Female

```
## [1] "The mean height of males and females are not the same."
```

Association Between Gender and the Amount of Physical Activity

```
##
##  Pearson's Chi-squared test
##
## data:  data.phys$females and data.phys$males
## X-squared = 6, df = 4, p-value = 0.1991
```

```
## The p value is greater than the significance value of 0.05, therefore the null
## hypothesis cannot be rejected. This means we can support the null hypothesis and that
## there is no correlation between gender and amount of physical activity.
```

Conclusion/Discussion

Judging from the results of the analysis, we can conclude from this specific study that there is a slight correlation between height and weight with a few outliers. We can determine that the average male is taller than the average female from this specific study. There also appears to be no correlation whatsoever between gender and the amount of exercise an individual does.

From this study, a larger sample size could be used in the future to test for the correlation between weight and height since a correlation does seem to exist according to this specific study.

In conclusion we can say we met out aim of trying to find a correlation between height and weight as a correlation was found, and that further investigation and sampling is required to prove the correlation.