

Project

Van Whiteman

2024-10-21

PROJECT

Will be testing 3 different questions:

1. Is there a linear relationship between height and weight?
2. Is the mean height of male and female the same?
3. Is there any association between gender and the amount of physical activity?

The data set used for the statistical analysis was project.csv which contains 1000 observations on:

- Gender
- Height(in cm)
- Weight (in kg)
- Physical activity (None, Moderate, Intense)

To test the relationship between height and weight, we used simple linear regression to find that a linear relationship between weight and height exists, the relationship is positive with a slope of: 0.774930348387773

The mean heights of female and males were shown, through a t test the mean height of males and females are not equal. The mean height of males is 178.13 cm and the mean height of females is 167.98 cm.

Through a Chi squared test, it was deduced that there is no association between gender and physical activity.

Is there a linear relationship between height and weight?

We wish to determine whether a linear relationship exists between the height of a person (in cm) and their weight (kg). This is of interest to many fields of research, such as the health sector, as someone that is very tall may be a lot heavier than someone who is short, but that doesn't necessarily mean they are more "overweight" than the shorter person.

The dataset contains a large pool of data with varying genders, ages and physicality to create an unbiased test. The method to test the linear relationship between the two will be simple linear regression, as it tests if one variable has a linear relationship with another using the `lm()` function. Our linear model will be:

$$Weight = \beta_0 + \beta * Height$$

where β is the slope of the line of weight regressed on height and β_0 is the intercept of the line.

we will have:

Null hypothesis: $H_0 : \beta = 0$

Alternative Hypothesis $H_1 : \beta \neq 0$

using `summary()` of the created linear model, we get the following results:

p-value of the regression is equal to: 4.80717124081435e-59, which is less than 0.05. Hence reject the null hypothesis in favour of the alternative hypothesis at the 5% significance level. Hence, a linear relationship between weight and height exists. The relationship is positive with a slope of: 0.774930348387773

Is the mean height of male and female the same?

Human height, throughout the history of humanity has been correlated with many different variables, including economic status, health, living standards and many other factors. (“Body Height in Adult Women and Men — Pmc.ncbi.nlm.nih.gov”, “Height, Body Mass Index, and Socioeconomic Status — Bmj.com”) In this study we will investigate if there is a difference in the average height of males in comparison to the average height of females. This study could be of interest to sports analysts, for example does the mean height of athletes affect performance, hence can this be seen in each gender’s ability in a particular sport.

The data set contains 525 males and 475 females with the average male height being 178.13cm and average female height 167.98cm.

gender	frequency	mean
Female	475	167.9800
Male	525	178.1299

Assuming equal variances of male and female, we can use a t test to test if the means of both genders are the same: `t.test(height~gender,data=project,var.equal=TRUE)`

With:

Null hypothesis $H_0 : \mu_f = \mu_m$

Alternative Hypothesis: $H_0 : \mu_f \neq \mu_m$

where μ_f is average height of females and μ_m is average height of males

p-value is equal to: 1.71197212128797e-33. p-value is less than 0.05, hence reject the null hypothesis in favour of the alternative hypothesis at the 5% significance level. The mean height of males and females are not equal. The mean height of males is 178.13cm and the mean height of females is 167.98cm.

Is there any association between gender and the amount of physical activity?

Many differences between males and females are often speculated to gender stereotypes, one of those speculations is whether males or females are more physically active. Previous studies have shown that males are more physically active than females starting from a young age and that trend continues through their lives (Health (2019)). This information is necessary for measuring the gender gaps in professional sports.

The data set has three levels of physical activity assigned: None, Moderate and Intense. We can hence segregate the and compare the frequency of men and women over the three levels:

	Female	Male
Intense	109	141
Moderate	233	259
None	133	125

```
We will use a Chi squared test to test the association between physicality and gender using chisq.test():
cont_tab <- table(data$phys,data$gender)
chisq.test(cont_tab)
```

With: Null Hypothesis: H_0 : no association between gender and physicality

Alternative Hypothesis: H_1 : there is association between gender and physicality

the p-value is: 0.199277786323652 . p-value is not less than 0.05, hence do not reject the null hypothesis at the 5% significance level. There is no association between gender and physical activity.

References

“Body Height in Adult Women and Men — Pmc.ncbi.nlm.nih.gov.” <https://pmc.ncbi.nlm.nih.gov/articles/PMC9272122/>.

Health, The Lancet Public. 2019. “Time to Tackle the Physical Activity Gender Gap.” *The Lancet Public Health* 4 (8): e360. [https://doi.org/10.1016/s2468-2667\(19\)30135-5](https://doi.org/10.1016/s2468-2667(19)30135-5).

“Height, Body Mass Index, and Socioeconomic Status — Bmj.com.” <https://www.bmj.com/content/352/bmj.i582#:~:text=The%20causal%20effect%2C%20as%20estimated,in%20men%20than%20in%20women.>