Statistical Analysis of Height Differences by Gender: A Hypothesis Testing Approach

Data Description: The objective of this project is to test the hypothesis concerning the difference between the mean heights of two populations, males and females. The raw data includes heights (in centimeters) for 30 males and 30 females.

Data: The sample of 30 males and 30 females was randomly selected from a dataset obtained from Kaggle[1] and is based on the research statistics in an Indian study[2].

	Height (cm)	Sex
1	166.95	Male
2	181.04	Male
3	183.98	Male
4	179.00	Male
5	175.43	Male
6	155.17	Male
7	167.36	Male
8	169.67	Male
9	179.44	Male
10	167.16	Male
11	185.28	Male
12	173.00	Male
13	167.81	Male
14	162.55	Male
15	176.67	Male
16	179.29	Male
17	167.94	Male
18	187.13	Male
19	186.59	Male
20	154.26	Male
21	171.81	Male
22	174.73	Male
23	163.52	Male
24	164.22	Male
25	173.55	Male
26	160.12	Male
27	176.69	Male
28	160.04	Male
29	177.03	Male
30	152.35	Male

	Height (cm)	Sex
1	131.81	Female
2	141.16	Female
3	135.57	Female
4	130.47	Female
5	158.99	Female
6	172.00	Female
7	152.30	Female
8	148.14	Female
9	169.15	Female
10	135.70	Female
11	140.52	Female
12	120.80	Female
13	140.89	Female
14	163.57	Female
15	137.95	Female
16	136.10	Female
17	159.68	Female
18	155.00	Female
19	140.30	Female
20	156.14	Female
21	144.32	Female
22	142.56	Female
23	149.57	Female
24	147.57	Female
25	153.92	Female
26	150.91	Female
27	142.13	Female
28	147.79	Female
29	163.96	Female
30	149.58	Female

Normal Probability Plots: To test if the data is normally distributed, we plot the data on normal probability plots as seen in Figure 1.

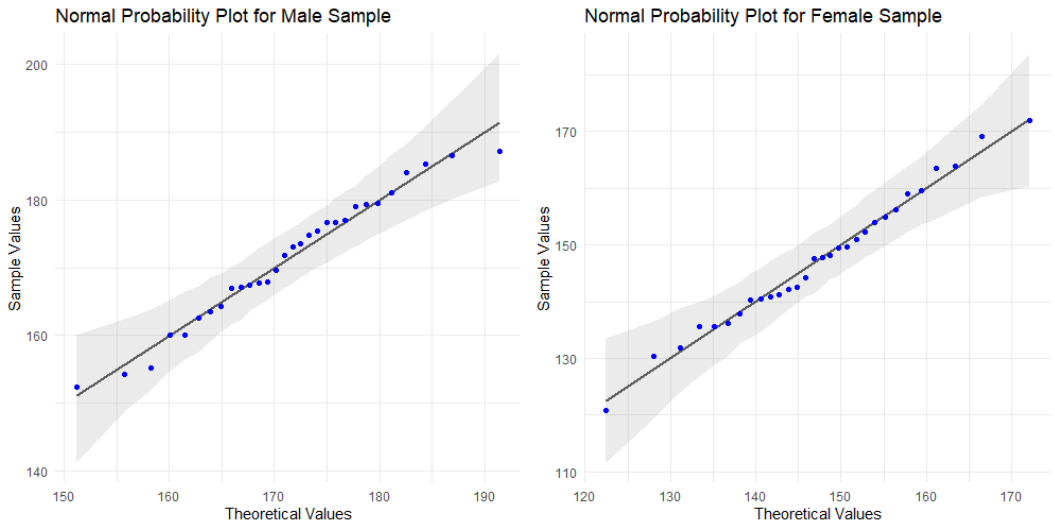

Figure 1: Normal probability plots for height data split by gender, bands indicate 95% confidence interval.

The normal probability plots in Figure 1 indicate that the data for both genders closely follows a normal distribution, with points lining up well along the theoretical line.

Sample Means and Sample Variances: To calculate the sample means of the male sample $m_1, \dots, m_N$ and female sample $f_1, \dots, f_N$ we use the equations:

$$\bar{m} = \frac{1}{N}\sum_{i=1}^{N} m_i \text{ and } \bar{f} = \sum_{i=1}^{N} f_i$$

The equation to calculate the sample variance is given by:

$$s_M^2 = \frac{1}{N-1}\sum_{i=1}^{N}(m_i - \bar{m})^2 \text{ and } s_F^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(f_i - \bar{f}\right)^2$$

The sample mean height for the male population is approximately 171.33 cm with a sample standard deviation of approximately 9.61 cm. The sample mean height for the female population is approximately 147.29 cm with a sample standard deviation of approximately 11.88 cm.

Test of Hypothesis on Equality of Population Variances:

Null Hypothesis H_0: The male and female populations have equal variances, that is $s_M^2 = s_F^2$.

Alternative Hypothesis H_1: The male and female populations have unequal variances, that is $s_M^2 \neq s_F^2$.

We use the F-test of equality of variances with the test statistic:

$$F = \frac{s_F^2}{s_M^2}$$

The F-test statistic for comparing the variances of the two samples is approximately 1.5277, and the two-tailed p-value is approximately 0.2597. This p-value is greater than 0.05, suggesting that there is

no statistically significant difference in variances between the sampled male and female height populations.

Test of Hypothesis on Equality of Population Means:

Null Hypothesis H_0: The male and female populations have equal means, that is $\bar{m} = \bar{f}$.

Alternative Hypothesis H_1: The male and female populations have unequal means, that is $\bar{m} \neq \bar{f}$.

The hypothesis test for equality of means was conducted using Student's t-test due to the lack of evidence to reject the hypothesis for equal variances. The test statistic is

$$t = \frac{\bar{m} - \bar{f}}{\sqrt{\frac{s_M^2 + s_F^2}{N}}}$$

The test statistic t = -8.6172 yielded a statistically significant difference (p-value = 5.709×10^{-12}).

Conclusion: The analysis strongly suggests that there is a significant difference in the mean heights between male and female populations, corroborating the hypothesis that gender is associated with height differences within this dataset. However, there was a lack of evidence to reject the hypothesis that the male and female populations have the same height variances.

References

[1] https://www.kaggle.com/datasets/saranpannasuriyaporn/male-female-height-and-weight/data

[2] *Saravanam, Prasanna & Ravikumar, Arunachalam. (2014). Biometric Study of the Internal Dimensions of Subglottis and Upper Trachea in Adult Indian Population. Indian Journal of Otolaryngology and Head & Neck Surgery. 66. 10.1007/s12070-012-0477-x. URL:*https://www.researchgate.net/publication/257769120_Biometric_Study_of_the_Internal_Dimensions_of_Subglottis_and_Upper_Trachea_in_Adult_Indian_Population.