A Simple Study on Weight and Height of Students

Dr. Mohammad Rafiqul Islam, (Associate Professor)
BRAC University, Dept. of Mathematics and Natural Sciences (MNS), Dhaka, Bangladesh
Md. Ishraque Bin Shafique, (Student)
Md. Kowshikur Rahman, (Student)
Md. Arafat Haque, (Student)
Dept. of Electrical and Electronic Engineering (EEE), BRAC University, Dhaka, Bangladesh

doi: 10.19044/esj.2017.v13n6p63 URL:http://dx.doi.org/10.19044/esj.2017.v13n6p63

Abstract

This study was conducted on a group of male and female students of age range of 18-25 years. In this paper it is tried to find out a correlation between height and weight of male and female students. Then the simple regression equations of weight on height are fitted for both for male and female students. A total of 639 students of different departments of BRAC University, Dhaka, Bangladesh in the spring semester of 2016 are participated in this survey. Body Mass Index (BMI) of the students was calculated to compare the health status of male and females students. It is found that that the most of the students (males and females) have the normal weight. It is interestingly noticed that the higher percentage (34.18%) of males are overweight than the females; whereas the females (13.33%) are more than double in underweight than their male's counterpart (5.93%). The correlation between height and weight of male students is calculated as 0.435 (Pearson's coefficient of correlation). On the other hand the correlation between height and weight of female students was 0.319. From the t tests, it is proved that the both the coefficients of correlation are highly statistically significant (p-value<0.01). From the simple regression equations of weight on height, it is found that the both for male and female students the effect of height on weight is almost same. It is also found that the effect of height on weight both for male and female students is highly significant (p-value<0.01).

Keywords: Weight, height, BMI, correlation and regression

Introduction

This paper illustrates the study on investigation of changes in height versus weight. The survey was conducted in BRAC University among the students of 18-24-year-old in the fall semester of 2015. At first a circular was put up on the university student affairs office notice boards to let people volunteer in the study. A total of 639 students have signed up of which 354 are male and 285 are female students. Both the groups were led by separate members of this study. The ages of the participants were calculated from their date of birth according to the university admission papers. The weight (in kilograms-kg) of each student was calculated using a high precision digital balance scale. The use of digital scale has two advantages. Firstly, it provides high precision data; secondly it reduces parallax errors. The heights have been measured using a wall measure tape of 0.01 inch precision. During this step the maximum stretching of the body was insured without any shoes. The measurements were taken very carefully to minimize the possibilities of any error. The identities of the volunteers were promised not to be disclosed. Body Mass Index (BMI) of the students was calculated to compare the health status of male and females students. In addition, the correlation and simple regression were calculated between heights and weights of male and female students.

Mathematical Models

Correlation

Statistics is referred to a division of mathematics which deals with learning from data of a large magnitude. It primarily consists of accumulating, evaluating and visualizing the numerical data in a suitable manner. Statistics is widely used in engineering, medical, demography, economics, business, politics and many other fields of studies. Although the origin of statistics date back to very old times, it is considered that its real use started around the late 1700's.

Correlation is a commonly used statistical model used to compute the linear relation between two variables. The foundation of correlation is laid by Sir Francis Galton in 1877 (Galton, 1877) and later developed by Karl-Pearson. Karl-Pearson's coefficient of correlation (Pearson, 1896) is a measure of linear correlation between two variables x and y.

The formula is given by:

$$r_{xy} = \frac{1}{n-1} \sum \frac{(x_i - \bar{x})(y_i - y)}{s_x s_y}$$

Other methods of finding correlation include Spearman's rank correlation coefficient (Spearman, 1904) , Kendall's rank coefficient (Kendall & Gibbons, 1990), Brownian correlation etc. The value of the correlation

coefficient can vary from -1 to +1. The magnitude, r represents the strength of the relation of the two variables. The sign represents if the variables are directly or inversely related. The nature of relationship is shown in figure 1:

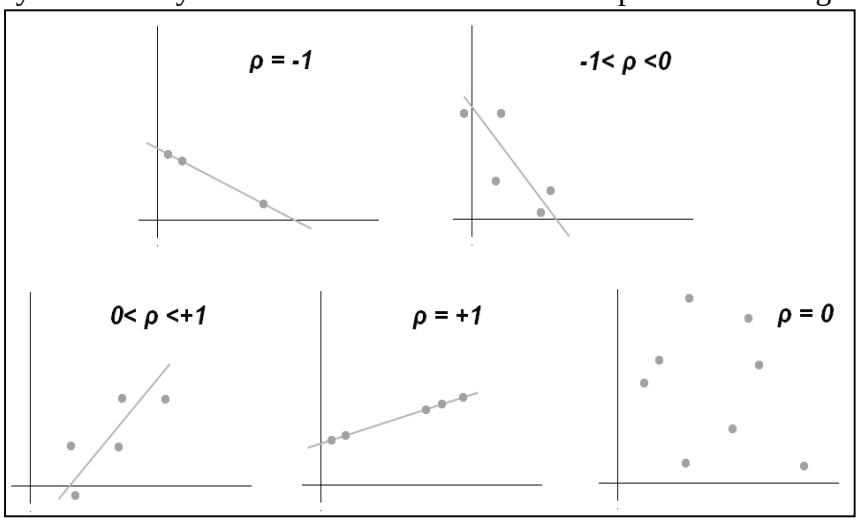

Figure 1. Different types of correlation according to the values of r found

Regression

Regression analysis is another widely used statistical method to estimate the relationship among the variables. By applying regression analysis, we can find the approximate value of the dependent variable when the independent variable is changed. There are several techniques for regression analysis where we need to have at least one dependent variable and one or more independent variables.

The idea of least-squares analysis was independently formulated by the French mathematician Adrien-Marie Legendre in 1805 (Adrien-Marie, 1805). There is also a saying that the method (of least squares) was first described by Carl Friedrich Gauss around 1794. According to Carter (Rice University, 1995 - textbook on Linear Algebra), Gauss developed least squares (Gauss, 1809) to solve a chemistry problem for his friend (when Gauss was about the age of a high-school senior). Gauss did not publish this that early time but he published it on 1809.

There are several models for regression analysis such as linear regression, simple regression, non-linear regression etc.

In this paper, we will work with the linear regression. Linear regression was the very first type of regression analysis which used to study very vastly and it has many practical applications. The vector form of linear regression equation is:

$$y = a + bx + e$$

where,

y = dependent variable, x = independent variable, e = the random error

The regression parameter a is the intercept (on the y axis), and the regression parameter b is the slope of the regression line. The random error term e is assumed to be uncorrelated, with a mean of 0 and constant variance

The regression coefficients 'a' and 'b' are calculated by the least squares method as follows:

$$\widehat{b} = \frac{SS_{xy}}{SS_{xx}}$$

And

$$\hat{a} = \bar{y} - \hat{b}\bar{x}$$

Where $SS_{xy} = \sum xy - \frac{\sum x \sum y}{n}$

And

$$SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{n}$$

So the fitted regression model is

$$\hat{y} = \hat{a} + \hat{b}x$$

It is noted that when slope 'b' is positive, relationship between x and y is positive, indicating that as x increases, y also increases. And when 'b' is negative, relationship between x and y is negative, implies that as x increases, y decreases. The estimate of intercept 'a' is of little significance. It locates the regression line at the point when x = 0. On the other hand, the slope 'b' is of great significance indicating the amount of change in the dependent variable y for unit change in the independent variable x.

Data Analysis

The data were recorded using proper computer spread sheet software. Two different set of data were collected according to the gender of the participants. The mean values of heights and weights of each group were calculated. Furthermore, the values of the standard deviations of each of the parameters were also computed. The heights of the students were recorded in inches which have then converted into meter which was used to calculate the Body Mass Index (BMI) of each student. BMI is given by:

$$Body\ Mass\ Index = \frac{mass_{kg}}{{h_m}^2}$$

BMI is as attempt to estimate the amount of body fat mass of a person. It is an indicator of health risk factor. The summary result of the data for the male and females students is given below table-1:

Table 1. Results of Analysis

	Male	Female
Number of Participants	354	285
Average Height (in)	68	62.9
Average Weight (kg)	71.3	58.5
Average BMI	23.9	23

From the table-1, it is observed that the averages of height and weight of male students are higher than the female students but the average BMI for the both is almost same. The chart below developed by the World Health Organization (WHO) in 1997 describes the ranges of values of BMIs that indicates each class.

Table 2: Classification of people according to BMI ranges

Classification of Weight	BMI ranges	Number of Male	Number of Female
Underweight	<18.5	21 (05.93%)	38 (13.33%)
Normal weight	18.5 to 25	212 (59.88%)	176 (61.75%)
Overweight	≥ 25.0	121 (34.18%)	71 (24.91%)

From table 2, we can see that most of the male and females lie in the normal weight class. However, a significantly higher percentage (34.18%) of males are overweight than the females (24.91). It is also noticed that the females (13.33%) are more than double in underweight than their male's counterpart (5.93%).

Correlation and Regression Analysis

Male

At first the correlation between height and weight of male is calculated by the method of Pearson's coefficient of correlation. The value of coefficient of correlation between height and weight of male students is 0.435. Then a t test is conducted and the value of t statistic is 9.06 with 354-2=352 degree of freedom (d.f). It is found that the relation between the height and weight of male students is highly significant (p-value<0.01).

Secondly, the simple regression model of weight on height for the male students is calculated as:

$$y = 1.85x - 54.85$$

For every inch increase in height, the weight of male students increases by 1.84 kilograms (kgs). The fitted regression line of regression of weight on height for the male is shown in figure 2:

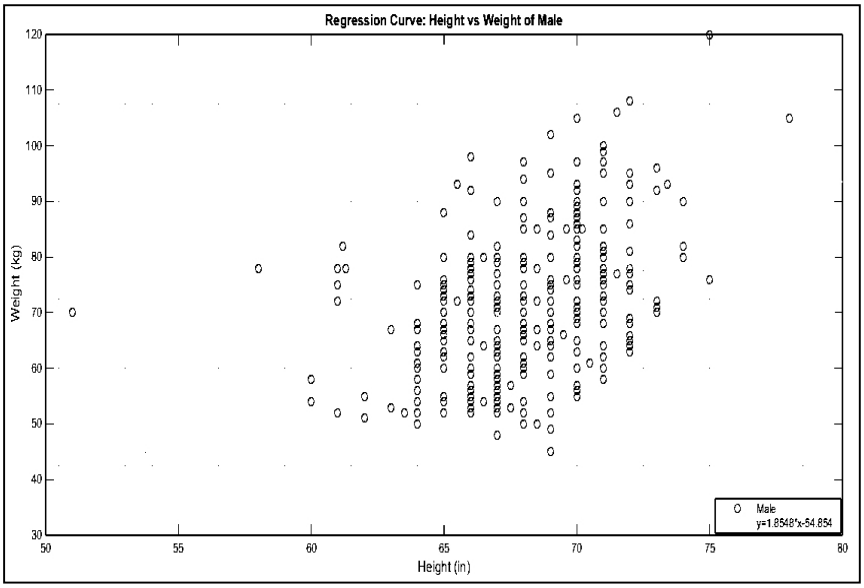

Figure 2. Regression model for height versus weight of male students

The t test is conducted to find out whether the regression coefficient (slope coefficient) 1.85 is statistical significant or not. First of all we calculated standard error of the slope coefficient, which is se(b)=0.242 and then the t statistic as 7.65 with 354-2=352 degree of freedom (d.f). It is found that the effect of height on weight of male students is highly significant (p-value<0.01).

Female

The correlation between height and weight of female students is calculated by the Pearson's coefficient of correlation. The value of coefficient of correlation between height and weight of female students is 0.319. Then a t test is conducted and the value of t statistic is 5.66 with 285-2=283 degree of freedom (d.f). It is found that the relation between the height and weight of female students is highly significant (p-value<0.01).

Secondly, the simple regression model of weight on height for the male students is calculated as:

$$y = 1.17x - 14.84$$

According to the regression line equation, the height of female students increases by 1.17 kilograms (kgs) for rise in every inch of height. The fitted regression line of regression of weight on height for the female is shown in figure 3:

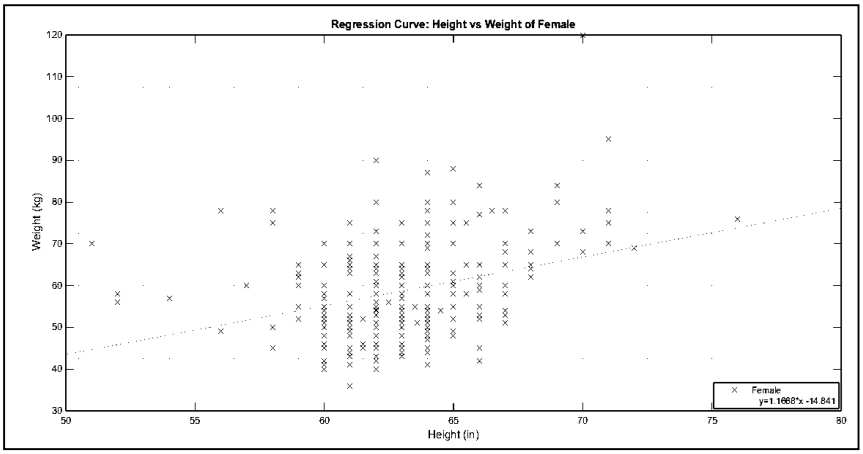

Figure 3. Regression model for height versus weight of female students

The t test is conducted to find the significance of the regression coefficient (slope coefficient) 1. First of all we calculated standard error of the slope coefficient, which is se(b)=0.119 and then the t statistic as 9.83 with 285-2=283 degree of freedom (d.f). It is also found that the effect of height on weight of female students is highly significant (p-value<0.01).

Conclusion

From the study, it is observed that the averages of height and weight of male students are higher than the female students but the average BMI for the both is almost same. The averages of BMI of both male and female students are in the normal weight limit. So it can be said that the students have a healthy lifestyle. The Pearson's correlation coefficient between height and weight of male; and between height and weight female students are calculated. It is observed that the relationship between height and weight of male students (r=0.435) is little bit stronger than the relationship between height and weight of female students (r=0.319). After doing the t test for the both it is found that both the correlations are highly statistical significant (for both p-value<0.01). Furthermore simple regression lines of weight on height is fitted both for male and female students. It is also found that the effect of height on weight both for male and female students is highly significant (p-value<0.01).

References:

1. Kelly H. Zou, PhD, Kemal Tuncali, MD, Stuart G. Silverman, MD. (2003). *Correlation and Simple Linear Regression. Statistical Concepts Series, Radiology* (pp. 227, 617–628). Published online, 10.1148/radiol.2273011499.

BMI Classification (2006). *Global Database on Body Mass Index.* World Health Organization, Retrieved July 27, 2012.

2. McLeod, S. A. (2008). *Correlation*. Retrieved from www.simplypsychology.org/correlation.html.
3. Merry, Brenda. (2012). *Measures of central tendency*, Chapter 5, February 23. Retrieved from: http://www.ck12.org/flexbook/chapter/9079.
4. Dalgaard, P. (2002). *Introductory Statistics with R* (pp. 106,114). Springer Verlag.
5. Ary, D., Jacobs, L. C., & Razavieh, A. (1996). *Introduction to research in education*. Fort Worth, TX: Harcourt Brace College Publishers.
6. Dowdy, S., & Weaden, S. (1991). *Statistics for Research*. 2nd edition. John Willey & sons.
7. Zou, K. H., & Hall, W. J. (2002). *On estimating a transformation correlation coefficient*. Journal of Applied Statistics, 29, 745-760. doi: 10.1080/02664760120098801.
8. Fisher, R. A. (1915). *Frequency distributions of the values of the correlation coefficient in samples from an indefinitely large population* (pp. 507–521). Biometrika, Vol.10.
9. Seber, GAF. (1997). *Linear regression analysis* (pp. 48–51). New York, NY: Wiley.
10. Pearson, K. (1896). *Mathematical contributions to the theory of evolution. III. Regression, heredity and panmixia* (pp. 253–318). Phil Trans R Soc Lond Series A, vol.187.
11. Galton, F. (1888). *Correlations and their measurements* (pp.45:219–247). Chiefly from anthropometric data, Proc R Soc London.
12. Galton, F. (1877). *Typical laws of heredity* (pp. 282–301). Proc R Inst Great Britain, vol.8.
13. Gauss, C.F. (I 809). Theoriu Motus Corporum Coelesrium. English translation by C.H. Davis, reprinted 1963 Dover, New York.
14. Kendall, M., & Gibbons, J.D. (1990). *Rank correlation methods*, 5th edition. New York: Oxford University Press.
15. Neter, J., Wasserman, W., & Kutner, MH. (1990). *Applied linear models: regression, analysis of variance, and experimental designs* (pp.38- 44, 62–104). 3rd ed. Homewood, Ill: Irwin.

16. Spearman, C. (1904). *The proof and measurement of association between two things* (pp.72–101). American Journal of Psychology, vol.15, doi: 10.2307/1412159.
17. Rodriguez, RN. (1982). *Correlation* (193–204). In: Kotz S, Johnson NL, eds. Encyclopedia of statistical sciences, New York, NY: Wiley.
18. Kruskal, WH. (1958). *Ordinal measurement of association* (pp. 814–861). J Am Stat Assoc, vol.53.
19. Daniel, WW. (1999). *Biostatistics: a foundation for analysis in the health sciences*. 7th ed., New York, NY: Wiley.
20. Carlin, J. B., Gurrin, L. C., Sterne, J. AC., Morley, R., & Dwyer, T. (2005). *Regression models for twin studies: a critical review.*. *International Journal of Epidemiol*, 34(5), 1089-1099. doi: 10.1093/ije/dyi153
21. Loker, S., Miglior, F., Koeck, A., Neuenschwander, T. O., Bastin, C., Jamrozik, J., ... & Kelton, D. (2012). Relationship between body condition score and health traits in first-lactation Canadian Holsteins. *Journal of dairy science*, *95*(11), 6770-6780.
22. Long, C. S., Lingyun, Z., Kowang, T. O., Fei, G. C., & Ismail, W. K. W. (2016). The Relationship between Sexual Harassment and Job Satisfaction in the Context of Retailing Industry in China. *International Journal of Human Resource Studies*, *6*(4), 1-19.