Two Variable Analysis between Height and Heart Rate

by

Andy Yan and Alexandru Darie

A thesis
presented to Mr. Wang
in fulfillment of the
culminating project for
the completion of the Data Management Course
at A.Y Jackson Secondary School

North York, Ontario, Canada, 2022

Author's Declaration

We hereby declare that we are the sole authors of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by our examiners.

We understand that our thesis may be made electronically available to the public.

Abstract

Height and heart rate are two variables that play a significant role in the health industry. Comparing the two can provide useful information that can improve the products used in the industry and our daily lives. This data can also bring helpful contributions to the sports and athletics industry.

Throughout this report, there will be tables and graphs of conducted surveys used to collect the heights and heart rates of 50 male 17-year-old boys and 50 female 17-year-old girls. Using this data, we were able to come up with mathematical relationships that link our data and prove our hypothesis of the assumed relationship.

Acknowledgements

We would like to thank Overleaf for allowing us to create this beautiful document.

Table of Contents

List of Figures

List of Tables

Calculations

Calculating the **Average** of the Data (males):

$$\mu = \frac{\sum_{k=1}^{50} x_k}{50} = 81.44, \quad \boxed{\mu = 81.44} \tag{1}$$

Calculating the **Standard Deviation** of the Data (males):

$$\sigma = \sqrt{\frac{\sum_{k=1}^{N}(x_k - \mu)^2}{N-1}} = \sqrt{\frac{\sum_{k=1}^{50}(x_k - 81.44)^2}{49}} \approx 7.7648, \quad \boxed{\sigma = 7.7648} \tag{2}$$

Calculating the **Variance** of the Data (males):

$$\sigma^2 = 7.7648^2 \approx 60.2921, \quad \boxed{\sigma^2 = 60.2921} \tag{3}$$

Calculating the **Pearson's R** of the Data (males):

$$r = \frac{n\sum_{i=1}^{50} x_i y_i - (\sum_{i=1}^{50} x_i)(\sum_{i=1}^{50} y_i)}{\sqrt{[(n\sum_{i=1}^{50} x_i^2) - (\sum_{i=1}^{50} x_i)^2][(n\sum_{i=1}^{50} y_i^2) - (\sum_{i=1}^{50} y_i)^2]}} = \frac{2261.48}{\sqrt{(4253.22)(2954.32)}} = 0.638 \tag{4}$$

$$\therefore \boxed{\text{r} = 0.638} \tag{5}$$

Calculating the **Line of Best Fit** for the Data (males):

$$a = \frac{n\sum_{i=1}^{50} x_i y_i - (\sum_{i=1}^{50} x_i)(\sum_{i=1}^{50} y_i)}{n(\sum_{i=1}^{50} x_i^2) - (\sum_{i=1}^{50} x_i)^2} = \frac{2261.48}{4253.22} \approx 0.5317, \quad \therefore a = 0.5317 \tag{6}$$

$$b = \overline{y} - a\overline{x} = 81.44 - 0.5317 \cdot 177.66 \approx -13.0236, \quad \therefore b = -13.0236 \tag{7}$$

Therefore the Equation of the line of best fit is:

$$\boxed{\text{y = 0.5317x - 13.0236}} \tag{8}$$

Calculating the **Average** of the Data (females):

$$\mu = \frac{\sum_{k=1}^{50} x_k}{50} = 89.32, \quad \boxed{\mu = 89.32} \tag{9}$$

Calculating the **Standard Deviation** of the Data (females):

$$\sigma = \sqrt{\frac{\sum_{k=1}^{N} (x_k - \mu)^2}{N-1}} = \sqrt{\frac{\sum_{k=1}^{50} (x_k - 89.32)^2}{49}} \approx 7.8101, \quad \boxed{\sigma = 7.8101} \tag{10}$$

Calculating the **Variance** of the Data (females):

$$\sigma^2 = 7.8101^2 \approx 60.9977, \quad \boxed{\sigma^2 = 60.9977} \tag{11}$$

Calculating the **Pearson's R** of the Data (females):

$$r = \frac{n\sum_{i=1}^{50} x_i y_i - (\sum_{i=1}^{50} x_i)(\sum_{i=1}^{50} y_i)}{\sqrt{[(n\sum_{i=1}^{50} x_i^2) - (\sum_{i=1}^{50} x_i)^2][(n\sum_{i=1}^{50} y_i^2) - (\sum_{i=1}^{50} y_i)^2]}} = \frac{1618.08}{\sqrt{(2233.78)(2988.88)}} = 0.6262 \tag{12}$$

$$\therefore \boxed{\text{r = 0.6262}} \tag{13}$$

Calculating the **Line of Best Fit** for the Data (females):

$$a = \frac{n\sum_{i=1}^{50} x_i y_i - (\sum_{i=1}^{50} x_i)(\sum_{i=1}^{50} y_i)}{n(\sum_{i=1}^{50} x_i^2) - (\sum_{i=1}^{50} x_i)^2} = \frac{1618.08}{2233.78} \approx 0.7244, \quad \therefore a = 0.7244 \tag{14}$$

$$b = \overline{y} - a\overline{x} = 89.32 - 0.7244 \cdot 169.62 \approx -33.5474, \quad \therefore b = -33.5474 \tag{15}$$

Therefore the Equation of the line of best fit is:

$$\boxed{\text{y = 0.7244x - 33.5474}} \tag{16}$$

Chapter 1

Introduction

Is there a correlation between one's height and heart rate? And if so would one's heart rate increase with height or the opposite?

1.1 Data Collection

To research this topic, we collected data from primary sources by asking 50 17-year-old boys and girls from our school. We separated the data into boys and girls because girls tend to have varied heart rates from boys. We collected data in a way so that the distribution of height is well spread allowing us to concentrate on our topic.

1.2 Hypothesis

Our hypothesis for this paper is: **the taller one is, the higher their heart rate should be.** We believe this because the larger one's body is, the more blood is needed to be pumped into the body which requires a higher heart rate.

1.3 General Data Analysis

Table 1.1: Main Data Part 1

Males(Age 17)		Females(Age 17)	
Height(cm)	Heartbeat(bpm)	Height(cm)	Heartbeat(bpm)
163	75	160	69
163	60	160	80
164	75	160	80
165	70	161	82
165	72	162	90
165	80	162	76
166	75	162	84
166	87	163	90
167	79	163	85
168	80	163	82
168	76	163	100
170	90	163	80
170	70	163	96
170	80	164	80
171	69	164	86
172	79	164	81
172	92	165	72
173	80	165	98
173	76	165	86
175	78	166	92
175	87	166	82
176	80	167	100
176	85	168	89
177	81	168	96
178	81	168	85
178	74	169	80

Table 1.2: Main Data Part 2

Males(Age 17) Continued

Height(cm)	Heartbeat(bpm)
180	68
180	82
180	84
180	92
181	91
181	72
181	76
182	83
182	76
183	82
183	83
184	83
185	84
185	90
186	88
186	84
188	90
189	84
190	86
190	87
193	92
194	92
195	95
199	97

Females(Age 17) Continued

Height(cm)	Heartbeat(bpm)
169	88
170	90
171	90
172	92
172	88
172	92
172	89
173	91
174	104
174	89
175	88
175	96
176	99
176	91
177	93
178	94
178	88
178	97
179	95
180	97
180	100
181	100
182	101
183	93

To put the data into perspective, we will graph scatter plots with a line of best fit.

Calculated Variables (Males)

$$\mu = 81.44 \quad (1)$$

$$\sigma = 7.7648 \quad (2)$$

$$\sigma^2 = 60.2921 \quad (3)$$

$$r = 0.638 \quad (5)$$

Calculated Variables (Females)

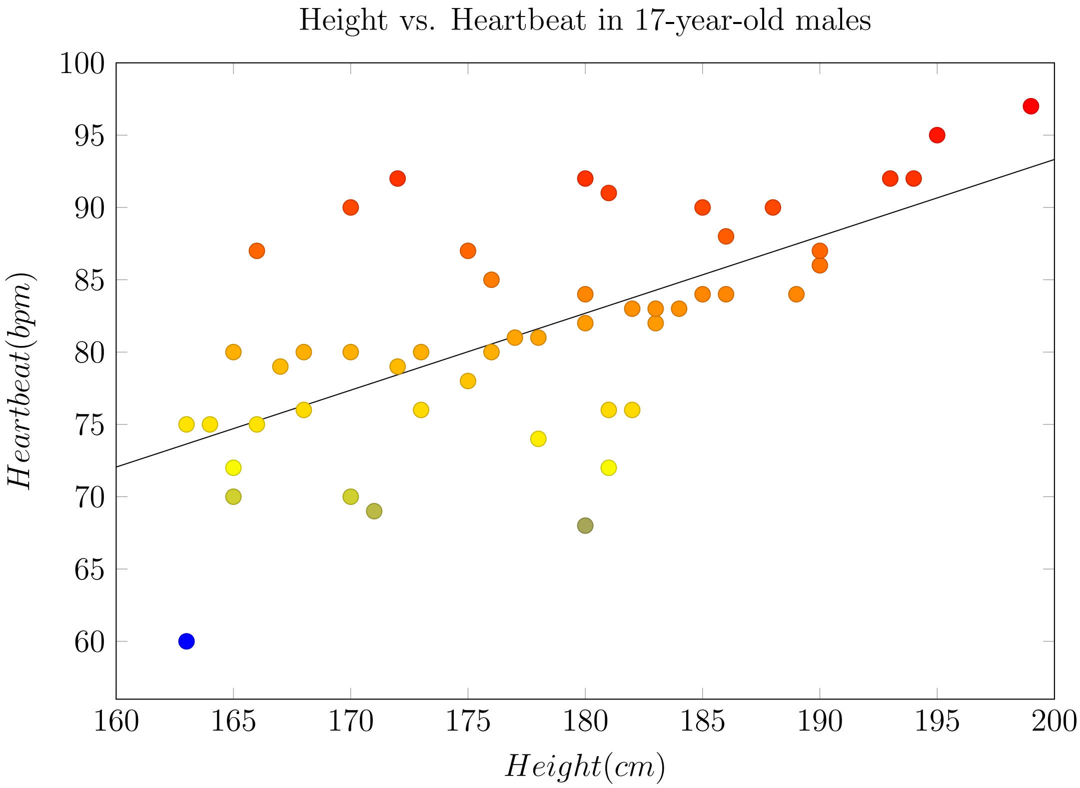
$$\mu = 89.32 \quad (9)$$

$$\sigma = 7.8101 \quad (10)$$

$$\sigma^2 = 60.9977 \quad (11)$$

$$r = 0.6262 \quad (12)$$

Figure 1.1: Scatter Plot Male

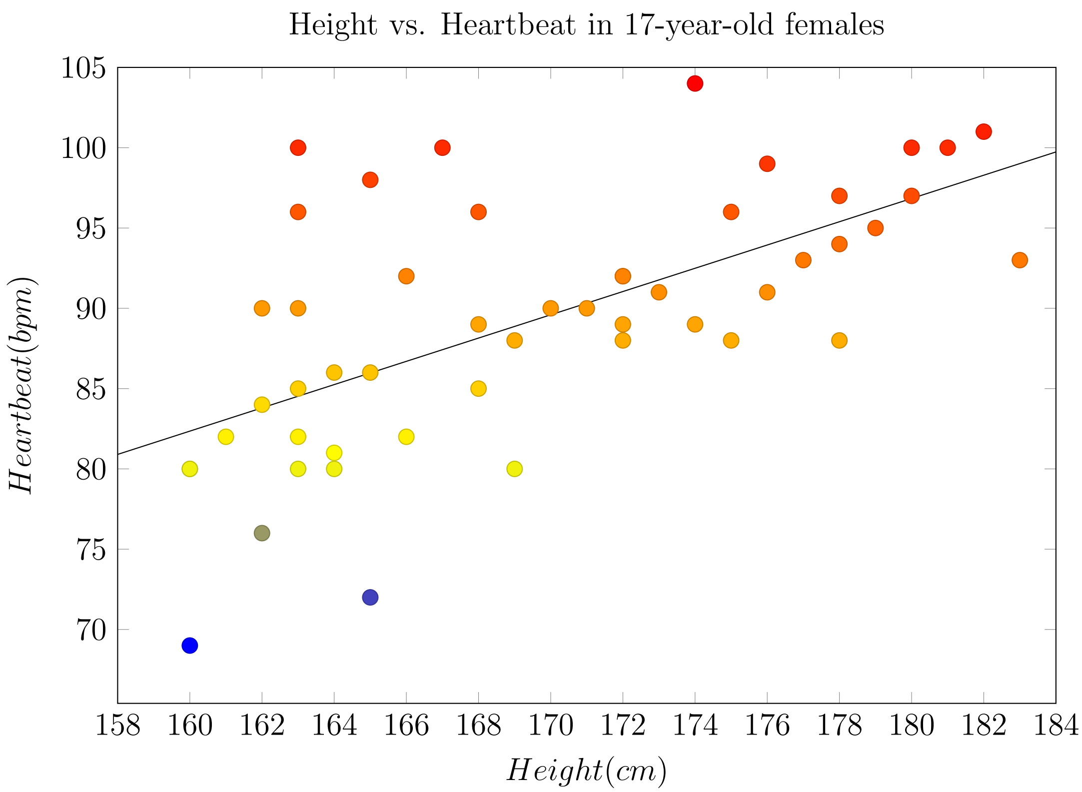


Line of Best Fit (Males):

$$y = 0.5317x - 13.0236 \text{ (8)} \tag{1.1}$$

If we take a closer look at the graph, we can see that there is a moderate-to-strong positive linear correlation between the two variables being plotted. This is evident from the fact that the correlation coefficient (r) is equal to 0.638, which falls within the range of 0.50 to 0.70, which is indicative of a moderate-to-strong positive linear relationship. Upon further examination, we can also see that most of the data points tend to cluster around the line of best fit, which is a good indication that there is a strong relationship between the two variables being plotted. However, there are a few outliers that fall outside of the cluster, which could potentially indicate some variability or inconsistency in the data. Additionally, we can see that the highest heart rate in the data belongs to the tallest boy, and the lowest heart rate belongs to the shortest boy.

Figure 1.2: Scatter Plot Female



Line of Best Fit (Females):

$$y = 0.7244x - 33.5474 \text{ } \mathbf{(16)} \tag{1.2}$$

Upon analyzing the graph for girls, we see that there is also a moderate-to-strong positive linear correlation between the two variables being plotted. This is evident from the fact that the correlation coefficient (r) is equal to 0.6262, which falls within the range of 0.50 to 0.70, indicating a moderate-to-strong positive linear relationship. From the graph, we can see that on the lower half most of the girls' heart-rates fall below the line of best fit, and the girls on the higher side fall on top.

1.4 Grouped Data Analysis

Grouped data analysis allows for more generalized height interval comparisons, helping us investigate the correlation.

Table 1.3: Grouped Heights

Grouped Frequency Distribution Table **Males**	
Class Interval	Frequency
160 - 167	9
168 - 175	12
176 - 183	16
184 - 191	9
192 - 199	4
Total:	**50**

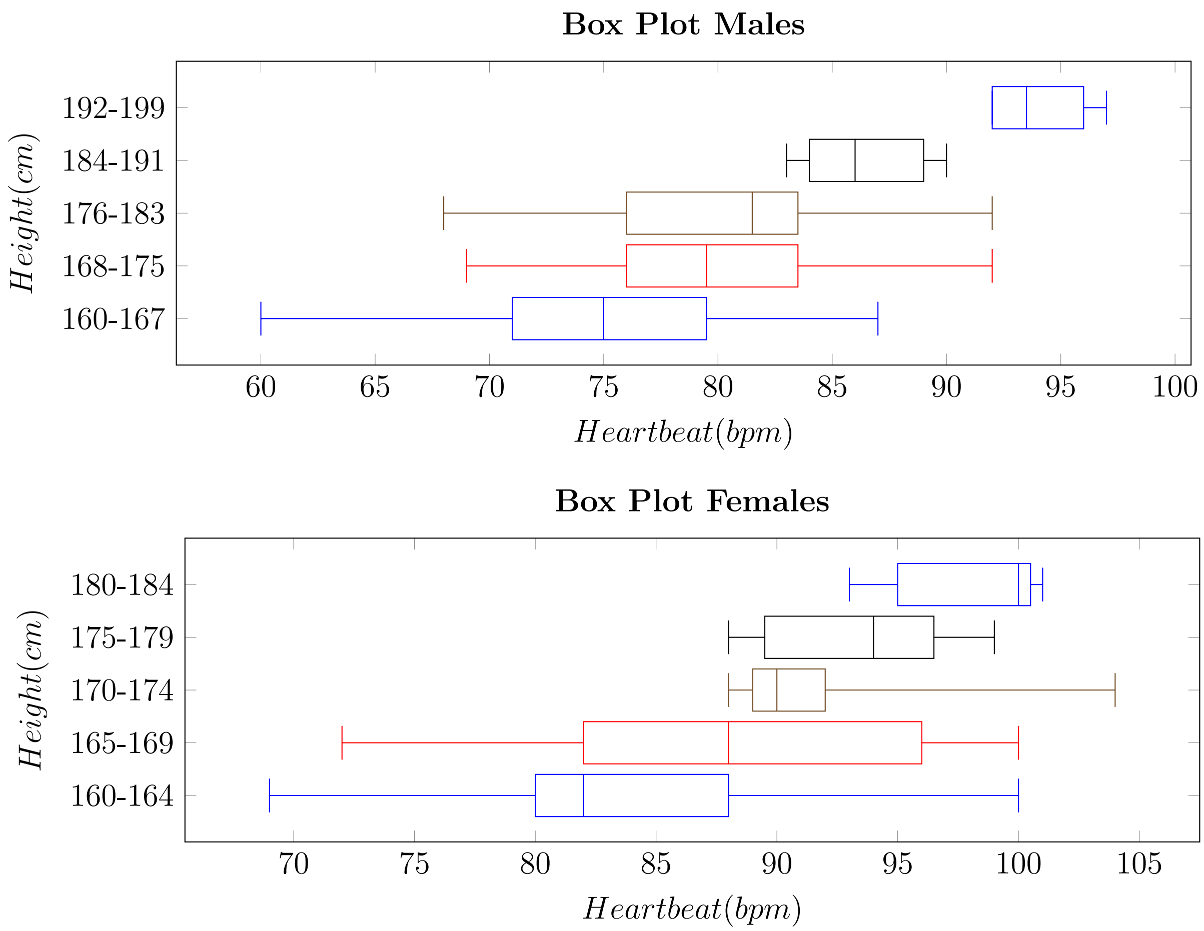
Grouped Frequency Distribution Table **Females**	
Class Interval	Frequency
160 - 164	16
165 - 169	11
170 - 174	9
175 - 179	9
180 - 184	5
Total:	**50**

Table 1.4: Quartile Calculations

Male Quartile Calculations					
Interval	Q_1	Q_2	Q_3	min	max
160-167	71	75	79.5	60	87
168-175	76	79.5	83.5	69	92
176-183	76	81.5	83.5	68	92
184-191	84	86	89	83	90
192-199	92	93.5	96	92	97

Female Quartile Calculations					
Interval	Q_1	Q_2	Q_3	min	max
160-164	80	82	88	69	100
165-169	82	88	96	72	100
170-174	89	90	92	88	104
175-179	89.5	94	96.5	88	99
180-184	95	100	100.5	93	101

Figure 1.3: Box Plots



A box plot is a useful tool for visualizing the distribution of a set of data and identifying any potential outliers. By analyzing our box plot, we can see that the majority of males have a heart rate between 70 beats per minute (bpm) and 85 bpm, while the majority of females have a heart rate between 85 bpm and 95 bpm. This is evident from the fact that the central boxes of the plots for each gender span these ranges, with the median heart rates falling near the center of the boxes. From both graphs, we can see that the majority of the data points are clustered within the mid-range of the heart rates, which suggests that there is a consistent relationship between the two variables being plotted.

Chapter 2

Discussion

As seen throughout the introduction there seems to be a relationship that will be further discussed and analyzed in this section.

2.1 Observations

Firstly, the most evident observation to be made is: the average heart rate is increasing as the average height increases for this particular data set. Both correlations from males and females show a moderate-to-strong linear correlation due to both Pearson's R's being very close to the boundary which shows a strong linear correlation (0.67). This statement is further supported by the box and whisker plots created. Additionally, this proves the hypothesis to be true for this particular sample. In other words, for students in AY Jackson as height increases, heart rate also increases.

Heart rate in this research experiment is labelled as the dependent variable because if it was independent, then it would not make sense for one to be taller just because they have a faster heart rate. Therefore, height is the independent variable and heart rate is the dependent variable. For this reason, this relationship is not a reverse cause-and-effect. Actually, the only possible relationships that can be drawn from this data are cause-and-effect and presumed relationships. The reason this relationship is not just cause-and-effect is because of the size of the data that was dealt with. Since the data size was not very large, this relationship cannot be fully justified as cause-and-effect especially since this is not the size of an entire population. There also could be hidden variables which will be addressed in the reflection tab (section 2.2).

2.2 Reflection

To conduct this experiment, a primary source of data collection was used to gather all the data. With this being said since there were only 2 people working on the collection, the data was obtained from a sample. This sample consisted of 50 seventeen-year-old males, and 50 seventeen-year-old females from the AY Jackson Secondary School. Since this sample does not reflect an entire population of 17-year-old males and females in all of Ontario, a definite conclusion cannot be drawn for a population, but it can be drawn for the sample from which the data originated.

Some very important and crucial pieces of information that were taken into consideration during this statistical experiment are the 3 main hidden variables. The first variable that was also kept constant was the age of all the people who were interviewed. A person's heart rate can change with age, so for this reason, the surveyed group needed to all have the same age. This made data collection slightly more challenging, but the results were more accurate. Another hidden variable would be the elevation of where students live. This is an uncontrolled variable, but since the sample was chosen from AY Jackson Secondary School, all of the students live in a general area where elevation does not play a major role. This made it negligible for the chosen sample but must be considered for national and international applications. Finally, the most important hidden variable would be activity and fitness levels. Students who have high fitness levels will have higher heart rates than others, which can account for the variations seen in the heart rate data. Likewise, health conditions also have a high chance of affecting this data, so that needs to be taken into account as well.

Overall, the results of this investigation prove to be fairly accurate due to the calculated Pearson's R and the careful and well-thought experimenting process. There was not much Bias that partook during the surveying since there were no leading or loaded questions. There were some non-response biases due to the voluntary response method of surveying, but for ethical and personal reasons forced participation cannot be used.

To further improve this analysis, a more universal surveying process of data can be used to gather more data which will better fit the size of a population. This will lead to a definite resolution to this relationship. Another method of improving this research project is to have a more refined and detailed filtering system where activity levels, elevation levels, and health conditions are used as parameters to organize and group the data into sections that reflect these various different populations.

References

[1] Alexandru Darie and Andy Yan. *A Collection of Heart Rates and Height of 17-year-old People*. Alexandru Darie and Andy Yan, Toronto, Ontario, Canada, 2022.

[2] M.D. Edward R. Laskowski. *Factors of Heart Rate*. Phoenix, Arizona, 2022.

[3] K.N. Poornima K. Prabhavathi, K.Tamarai Selvi and A. Sarvanan. *Role of Biological Sex in Normal Cardiac Function*. J Clin Diagn Res, Chennai, Tamil Nadu, India, 2014.

[4] Digital Science UK Limited. *LaTeX — A Document Preparation Library*. Digital Science UK Limited, London, United Kingdom, 2012.

Appendix A

Height Heart Rate Grouper

```
#include <iostream>
#include <vector>
using namespace std;
int main(){
    int a,b;
    vector<int>v[200];
    for(int i = 0;i<50;i++){
        cin>>a>>b;
        v[a].push_back(b);//array index is independent variable
    }
    for(int i = 0;i<=199;i++){
        int sz = v[i].size();
        if(sz>0){
            cout<<i<<":   ";
        }
        for(int j : v[i]){
            cout<<j<<",";
        }
        if(sz>0){
            cout<<"\n";
        }
    }
    return 0;
}
```