ISSN 0304-9892 (Print) ISSN 2455-7463 (Online)
www.vijnanaparishadofindia.org/jnanabha
***Jñānābha*, Vol. 53(2) (2023), 113-125**
(Dedicated to Professor V. P. Saxena on His 80th Birth Anniversary Celebrations)

A MULTIPLE REGRESSION MODEL FOR IDENTIFYING SOME RISK FACTORS AFFECTING THE CARDIOVASCULAR HEALTH ISSUES IN ADULTS

Mohammad Shakil[1], Mohammad Ahsanullah[2], B. M. G. Kibria[3], J. N. Singh[4], Rakhshinda Jabeen[5], Aneeqa Khadim[6] and Musaddiq Sirajo[7]

[1]Department of Mathematics, Miami Dade College, Hialeah, FL, USA
[2]Department of Management Sciences, Professor Emeritus, Rider University, NJ, USA
[3]Department of Mathematics & Statistics, Florida International University, Miami, FL, USA
[4]Department of Mathematics & Computer Sciences, Barry University, Miami Shores, FL, USA
[5]Department of Medicine, Dow University of Health Sciences, Karachi, Pakistan
[6]Department of Mathematics, Mirpur University of Science & Technology, Mirpur, Pakistan
[7]Department of Statistics, Ahmadu Bello University, Zaria, Nigeria
Email: mshakil@mdc.edu,ahsan@rider.edu,kibriag@fiu.edu, jsingh@barry.edu,
rakhshinda.jabeen@duhs.edu.pk,Aneeqa89@gmail.com,musaddi,musaddiqsirajo@gmail.com

(Received: June 20, 2023; In format: August 26, 2023, Revised: September 29, 2023; Accepted: October 03, 2023)

DOI: https://doi.org/10.58250/jnanabha.2023.53214

Abstract

Multiple Regression analysis is one of the most critical and widely used statistical techniques in medical and applied research. It is defined as a multivariate technique for determining the correlation between a response variable and some combination of two or more predictor variables. Moreover, it is well-known in medical sciences that the obesity, high blood pressure and high cholesterol are major risk factors for cardiovascular health issues. The body mass index is a measure of body size, and combines a person's weight with their height, and therefore can affect their obesity, high blood pressure, high cholesterol and type 2 diabetes mellitus significantly, which are major risk factors for cardiovascular health issues in adults. Motivated by these facts, in this paper, a multiple linear regression model is developed to analyze the obesity in adults, based on a sample data of adult's age, height, weight, waist, diastolic blood pressure, systolic blood pressure, pulse, cholesterol, and the body mass index measurements. The use of multiple linear regression is illustrated in the prediction study of adult's obesity based on their body mass index. It is observed that in the presence of adult's age, weight, waist, diastolic blood pressure, systolic blood pressure, pulse, and cholesterol levels, height is a good predictor of the body mass index. Moreover, in the presence of age, height, waist, diastolic blood pressure, systolic blood pressure, pulse, and cholesterol levels, weight is a good predictor of the body mass index. Some concluding remarks are given in the end.

2020 Mathematical Sciences Classification: 65F359, 15A12, 15A04, 62J05.

Keywords and Phrases: Cardiovascular, high cholesterol levels, high blood pressure, multiple regression, obesity.

1 Introduction

Multiple linear regression is one of the most widely used statistical techniques in medical and other applied research. It is defined as a multivariate technique for determining the correlation between a response variable Y and some combination of two or more predictor variables, X. For example, it can be used to analyze data from causal-comparative, correlational, or experimental research. It can handle interval, ordinal, or categorical data. In addition, multiple regression provides estimates both of the magnitude and statistical significance of relationships between variables. For details on regression analysis and its applications, the interested readers are referred to Neter et al. [19], Draper and Smith [5], Tamhane and Dunlop [25], Mendenhall and Sincich [16], Chatterjee and Hadi [2], Montgomery [17], Surez et al. [23], Cleophas and Zwinderman [3], Guzman and Kibria [7], Johnson and Wichern [9], among others. For recent developments on linear and non-linear regression models, we refer to Kibria [12].

The purpose of the present study is to contribute to the body of knowledge pertaining to the use of multiple linear regression in medical and applied research, and, in particular, in identifying some risk factors affecting the cardiovascular health issues in adults. It appears from the literature that not much attention has been paid to this kind of studies in the multiple regression analysis of the cardiovascular health issues and problems in adults. Motivated by these facts, in this paper, a multiple linear regression model is developed to analyze the obesity in adults, based on their body mass index *(BMI)* by taking a sample data of adult's age, height, weight, waist, diastolic blood pressure, systolic blood pressure, pulse, cholesterol, and BMI measurements. The use of multiple linear regression is illustrated in the prediction study of adult's obesity based on their body mass index, along with these risk indicators.

1.1 Body Mass Index *(BMI)*

In what follows, we first present some basic ideas about the body mass index *(BMI)*, and the review of the literature relevant to the cardiovascular health issues.

Definition 1.1. *The body mass index (BMI) is defined as a measure of body size and for weight-related health risk. It combines a person's weight with their height. It can be calculated using the following formulas:*

$$BMI = Weight(kg)/[height(m)]2, \tag{1.1}$$

$$BMI = Weight(lb)/[height(in)]2 \times 703. \tag{1.2}$$

Thus, the results of a BMI measurement can give an idea about whether a person's weight is correct with respect to their height. Moreover, the BMI of a person can indicate whether they are underweight or if they have a healthy weight, or excess weight, or obesity. If a person's BMI is outside of the healthy range, their health risks may increase significantly. According to the US Centers for Disease Control and Prevention and the World Health Organization, "BMI represents the relationship between weight and height to estimate the amount of fat in the body" (Global Health Observatory. from http://www.who.int/gho/ncd/risk_factors/bmi_text/en/). Moreover, as observed by Young et al. [29], Nguyen et al. [20], and Keum et al. [13], "A higher percentage of body fat is proven to be associated with increased risk for developing certain diseases such as heart disease, high blood pressure, type 2 diabetes, breathing problems, certain cancers, and death". Furthermore, as reported by https://www.weightwatchers.com/us/science-center/bmi-calculator, there appears to be an exponential relationship between BMI and mortality rate which is illustrated in the following Figure 1.1.

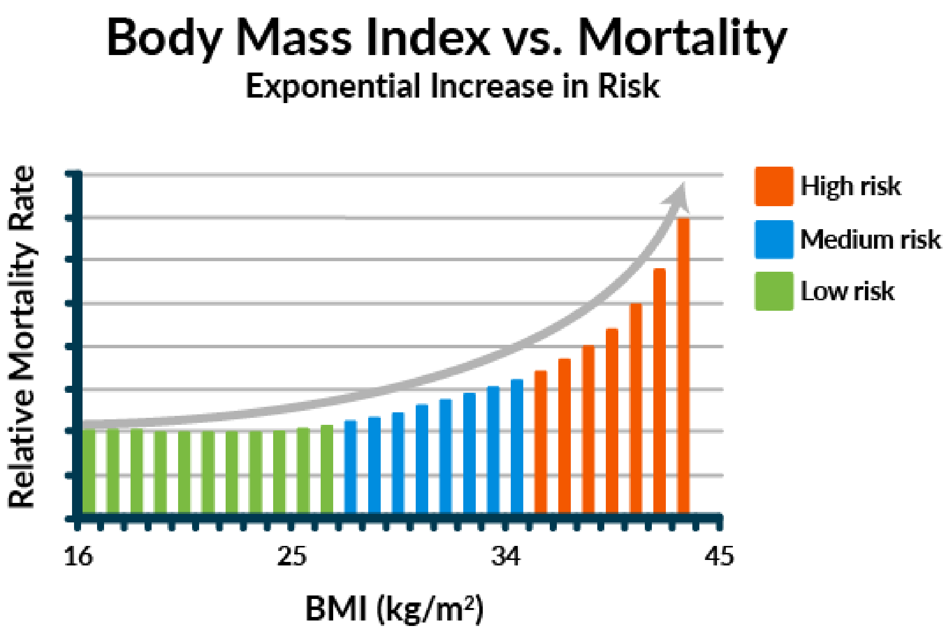

***Figure* 1.1**
(Source: https://www.weightwatchers.com/us/science-center/bmi-calculator)

According to Narkiewicz [22], "Obesity and in particular central obesity have been consistently associated with hypertension and increased cardiovascular risk. Based on population studies, risk estimates indicate that at least two-thirds of the prevalence of hypertension can be directly attributed to obesity". Further, as pointed out by Hall et al. [18], "Major consequences of being overweight or obese include higher prevalence of hypertension and a cascade of associated cardiorenal and metabolic disorders. Studies in diverse populations throughout the world have shown that the relationship between BMI and systolic and diastolic blood pressure (BP) is nearly linear. Risk estimates from the Framingham Heart Study, for example, suggest that 78% of primary (essential) hypertension in men and 65% in women can be ascribed to excess weight gain. Clinical studies indicate that maintenance of a *BMI* <25 kg/m^2 is effective in primary prevention of hypertension and that weight loss reduces *BP* in most hypertensive subjects". Also, according to Jiang *et al.* [10], "Obesity can result in serious health issues that are potentially life-threatening, including hypertension, type II diabetes mellitus, increased risk for coronary disease, increased unexplained heart failure, hyperlipidemia, infertility, higher prevalence of colon, prostate, endometrial, and breast cancer. Although the relationship between obesity and hypertension is well established in children and adults, the mechanism by which obesity directly causes hypertension is under investigation".

"Having obesity puts a strain on our heart and can lead to serious health cardiovascular problems, namely, arthritis in our knees and hips, heart disease, high blood pressure, sleep apnea, type 2 diabetes, and varicose veins" (https://medlineplus.gov/ency/article/007196.htm). Moreover, a person's *BMI* can be categorized (Table 1.1), along with the three classes of obesity (Table 1.2), as given below:

***Table* 1.1**

(https://medlineplus.gov/ency/article/007196.htm)

BMI	**CATEGORY**
Below 18.5	Underweight
18.5 to 24.9	Healthy
25.0 to 29.9	Overweight
30.0 to 39.9	Obese
Over 40	Extreme of high-risk obesity

***Table* 1.2**

(https://medlineplus.gov/ency/article/007196.htm)

CLASS	**OBESITY**
1	*BMI* of 30 to less thank 35
1	*BMI* of 35 to less than 40
3	*BMI* of 40 or higher. Class 3 is considered "severe obesity".

Thus, it is obvious from the Tables 1.1 and 1.2 that a person's obesity can be significantly affected by their body mass index (*BMI*), high blood pressure and high cholesterol, which are all major risk factors for cardiovascular health issues. For further details on cardiovascular diseases and related issues, the interested readers are referred to Mertens and Van Gaal [18], Akil and Ahmad [1], Klop et al. [14], Vach [27], Leggio et al. [15], Seravalle and Grassi [24], Feng et al. [6], Jabeen et al. [11], Rajeshwari and Laishram [22], and references therein.

The organization of this paper is as follows. In Section 2, the proposed multiple linear regression model, and the problem and objective of this study are presented. Section 3 provides the data analysis, justification and adequacy of the multiple regression model developed. Some concluding remarks are given in Section 4.

2 Multiple Linear Regression Model

2.1 A Multiple Linear Regression Model based on a Number of Predictors

Consider following multiple linear regression model

$$Y = X\beta + \epsilon, \tag{2.1}$$

where Y is an $n \times 1$ vector of response variable (observations), β is a $k \times 1$ vector of unknown regression coefficients, X is an $n \times k (n > k)$ observed matrix of the regression, and ϵ is an $n \times 1$ vector of random errors, which is distributed as multivariate normal with mean 0 and covariance matrix $\sigma^2 I_n$, and I_n is an identity matrix of order n. The OLS estimator of β is obtained as $\widehat{\beta} = (X'X)^{-1}X'y$, and covariance matrix of $\widehat{\beta}$ is obtained as Cov $(\widehat{\beta}) = \sigma^2(X'X)^{-1}$.

2.2 Problem and Objective of Study

It is well-known in medical sciences that the obesity, high blood pressure and high cholesterol are major risk factors for cardiovascular health issues. For example, high cholesterol can affect anyone, regardless of their weight. Moreover, high blood pressure, also called hypertension, is a major risk factor for heart disease, kidney disease, stroke, and heart failure. Having excess body weight can lead to increased high blood pressure and cholesterol levels. The body mass index is a measure of body size, and combines a person's weight with their height, the results of a body mass index measurement can indicate whether a person has excess weight, and thus can affect their obesity, high blood pressure and high cholesterol significantly, which are all risk factors for cardiovascular health issues.

Thus, in view of the above facts, the objective of our present investigation would be to develop an appropriate multiple linear regression model to relate the adult's obesity, based on their body mass index (*BMI*) (considered as the dependent or response variable Y) to the adult's age, height, weight, waist, diastolic blood pressure, systolic blood pressure, pulse, cholesterol, *BMI* measurements (considered as the independent or predictor variables X). It will be examined how well the adult's age, height, weight, waist, pulse, diastolic blood pressure, systolic blood pressure, cholesterol, and *BMI* measurements could be used to predict the adult's body mass index (*BMI*), as it affects a person's obesity, high blood pressure and high cholesterol significantly, which are all risk factors for cardiovascular health issues in adults.

To pursue our studies, the data were collected from Triola [26] on the adult's age, height, weight, waist, pulse, diastolic blood pressure, systolic blood pressure, cholesterol, and *BMI* measurements, for a sample of 40 adults, (which we have provided in Appendix 1 for the sake of completeness). Using these variables and the Equation (2.1), the following eight-predictor multiple linear regression model (or the least squares prediction equation) was developed:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8 + \varepsilon, \tag{2.2}$$

where β' s denote the population regression coefficients, ε is a random error, the response variable is the adult's *BMI* (Y), and the respective eight predictors are the adult's age (X_1), height (X_2), weight (X_3), waist ($\boldsymbol{X_4}$), pulse ($\boldsymbol{X}_5$), diastolic blood pressure ($\boldsymbol{X_6}$), systolic blood pressure ($\boldsymbol{X}_7$), and cholesterol ($\boldsymbol{X_8}$).

3 Data Analysis

The Minitab Version 17.0 regression computer programs were used to determine the regression coefficients and analyze the data. The adequacy of the multiple linear regression model for predicting the adult's body mass index (*BMI*) was conducted using the F-test for the significance of regression.

The Minitab regression computer program outputs are given below. The paragraphs that follow explain the computer program outputs.

3.1 Minitab Regression Computer Program Output: Analysis of Variance

3.1.1 Regression Analysis: *BMI* versus Age, $Ht, \ldots$

The regression equation is:

$$BMI = 52.1 + 0.00134 \text{ Age } - 0.772 Ht + 0.147 \text{ Wt } + 0.0125 \text{ Waist } + 0.00710 \text{ Pulse}$$
$$- 0.00229 \text{ Systolic } - 0.00195 \text{ Diastolic } + 0.000211 \text{ Cholesterol }.$$

***Table* 3.1**

Predictor	**Coef**	***SE* Coef**	***T***	***P***	***VIF***
Constant	52.1200000	1.8800000	27.72	0.000	
Age	0.0013420	0.0049270	0.27	0.787	2.0
Ht	-0.7721100	0.0248400	-31.08	0.000	2.4
Wt	0.1465580	0.0063350	23.13	0.000	11.7
Waist	0.0125100	0.0167500	0.75	0.461	11.5
Pulse	0.0070950	0.0047400	1.50	0.145	1.2
Systolic	-0.0022870	0.0059550	-0.38	0.704	1.6
Diastolic	-0.0019480	0.0075320	-0.26	0.798	2.0
Cholesterol	0.0002106	0.0001749	1.20	0.238	1.1

***Table* 3.2**

$\boldsymbol{S}=0.304262$	$R-Sq=99.4\%$	$R-Sq(\text{ adj })=99.2\%$
PRESS = 5.60841	R-Sq(pred) = 98.78%	
Durbin-Watson statistic = 2.80903		

***Table* 3.3**

Analysis of Variance					
Source	*DF*	*SS*	*MS*	*F*	*P*
Regression	8	456.160	57.020	615.93	0.000
Residual Error		2.870	0.093		
Total	39	459.030			

***Table* 3.4**

Unusual Observations						
Obs	Age	*BMI*	Fit	*SE* Fit	Residual	*St* Resid
17	41.0	33.2000	32.3881	0.1767	0.8119	3.28R
36	34.0	20.7000	21.4631	0.1542	-0.7631	−2.91R

Note: Here, in Table 4.4, R denotes an observation with a large standardized residual.

3.1.2 Interpreting the Results

I. From the Analysis of Variance Table 3.3, we observe that the p-value is (0.000). This implies that that the model estimated by the regression procedure is significant at an α-level of 0.05 . Thus at least one of the regression coefficients is different from zero.

II. From the Table 3.1, we observe that the p-values for the estimated coefficients of height (X_2) and weight (X_3) are respectively 0.000 and 0.000 , indicating that they are significantly related to the response variable is *BMI* (Y) at an α-level of 0.05. From the Table 3.1, we also observe that the p-values for the adult's age (X_1), waist (X_4), pulse (X_5), diastolic blood pressure (X_6), systolic blood pressure (X_7), and cholesterol (X_8), are relatively high, indicating that these are probably not related to the response variable *BMI* (Y) at an α-level of 0.05 .

III. **The R^2 and Adjusted R^2 Statistic:** There are several useful criteria for measuring the goodness of fit of the multiple regression model. One such criterion is to determine the square of the multiple correlation coefficient R^2 (also called the coefficient of multiple determination), (see, for example, Draper and Smith [5], and Mendenhall and Sincich [16], among others). The R^2 value in the regression output (Table 3.2) indicates that 99.4% of the total variation of the response variable $BMI(Y)$ values about their mean can be explained by the predictor variables used in the model. The adjusted R^2 value (or ${R_a}^2$) indicates that 99.2% of the total variation of the response variable $BMI(Y)$ values about their mean can be explained by the predictor variables used in the model. As the values of R^2 and R_a^2 are not very different, it appears that at least one of the predictor variables contributes information for the prediction of Y. Thus, both values indicate that the model fits the data well.

IV. **Predicted $\mathbf{R}^2$ Statistic:** Further from Table 3.2, we observe that the predicted R^2 value is 98.78%. Because the predicted R^2 value is close to the R^2 and adjusted R^2 values, the model does not appear to be overfit and has adequate predictive ability.

V. **Estimate of Variance:** The variance about the regression σ^2 of the Y values for any given set of the independent variables $X_1, X_2, \ldots, X_k$ is estimated by the residual mean square s^2, which is equal to SS (residual) divided by an appropriate number of degrees of freedom, and the standard error s is given by

$$s=\sqrt{\text{residual meansquare } s^2}.$$

For our problem, we have

$$\boldsymbol{s^2=0.093} \text{ and } \boldsymbol{s=0.30496}$$

Examination of this statistic indicates that the smaller it is the better, that is, the more precise will be the predictions. A useful way of looking at the decrease in S is to consider it in relation to response, (see, for example, Draper and Smith (1998), among others, for details). In our example, s as a percentage of mean $\bar{Y}$ (of the response variable *BMI*, $\boldsymbol{Y}$), that is, the coefficient of variation (CV), is given by

$$CV = \frac{0.30496}{25.9975} \times 100\% = 1.17303\%.$$

This means that the standard deviation of the adult's *BMI* (Y), is only **1.17303**% of their mean, which means considerably less variation.

VI. **Unusual Observations:** We also note from the Table 3.4 that the observations 17 and 36 (see Appendix 1) are identified as unusual because the absolute value of the standardized residuals is greater than 2 . This may indicate they are outliers.

VII. **Multicollinearity:** By multicollinearity, we mean that some predictor variables are correlated with other predictors. Various techniques have been developed to identify predictor variables that are highly collinear, and for possible solutions to the problem of multicollinearity, (see, for example, Draper and Smith [5], Tamhane and Dunlop [25], Mendenhall and Sincich [16], Chatterjee and Hadi [2], Montgomery et al. [17], Chatterjee and Simonoff [4], and Vittinghoff et al. [28], among others, for details). For example, we can examine the variance inflation factors (*VIF*), which measure how much the variance of an estimated regression coefficient increases if the predictor variables are correlated. Following Montgomery et al. [17], if the *VIF* is 5 - 10, the regression coefficients are poorly estimated. However, it has been observed by many researchers that for a large sample size, multicollinearity is not a big problem when compared to a small sample size. Since the variance inflation factors (VIF) for each of the estimated regression coefficient in our calculations are less than 5 for the adult's age (X_1), height (X_2), pulse (X_5), diastolic blood pressure (X_6), systolic blood pressure (X_7), and cholesterol (X_8), there does not seem to be multicollinearity for these predictors in our model. However, we observe that the VIF are fairly large for the predictor weight (X_3) and waist (X_4), implying that these are highly correlated with at least one of the other predictors in the model. In order to deal with the said multicollinearity is to remove some of the violating predictors from the model, that is, for assessing the predictive ability of a multiple linear regression model, is to examine the associated C_p-statistic. The best subsets regression method is used to choose a subset of predictor variables so that the corresponding fitted regression model optimizes the C_p-statistic, which is described in Sub-Section 3.2 below.

VIII. **Predicted Values for New Observations:** Using the model developed, some values are provided in Table 3.5 .

3.2 Best Subsets Regression:

Another important criterion function for assessing the predictive ability of a multiple linear regression model is to examine the associated Mallows' C_p-statistic, including R-Sq $\left(R^2\right)$, the percentage of variation in the response that is explained by the model, Adjusted R^2 (that is, $R\ Sq(adj)$, the percentage of the variation in the response that is explained by t for the number of predictors in the model relative to the number of observations), and s, the standard error of the estimate. The best subsets regression method is used to choose a subset of predictor variables so that the corresponding fitted regression model optimizes the Mallows' C_p-statistic, which may be interpreted as follows:

(1) A Mallows' Cp value that is close to the number of predictors plus the constant model produces relatively precise and unbiased estimates.

(2) A Mallows' Cp value that is greater than the number of predictors plus the constant model is biased and does not fit the data well.

The model with all the predictor variables should have the highest adjusted R^2, a low Mallows' Cp value, and the lowest s value. Based on these criteria, the following (Table 3.6) are the possible predictor models (X_2, X_3) or (X_1, X_2) with respective highest adjusted R^2, a low Mallows Cp value, and the lowest S value.

Note that three other predictor models, namely, [Height (X_2), Weight (X_3), Waist (X_4), Cholesterol (X_8)], or [Age (X_1), Height (X_2), Weight (X_3), Pulse (X_5)], or [Height (X_2), Weight (X_3), Cholesterol (X_8)] also exist here with respective highest adjusted R^2, a low Mallows Cp value, and the lowest S value (see the output above).

***Table* 3.5:** Predicted Values for New Observations

New				
Obs	**Fit**	***SE* Fit**	**95% *CI***	**95% *PI***
1	23.6038	0.1107	(23.3781,23.8296)	(22.9435,24.2641)
2	23.2779	0.1253	(23.0224,23.5333)	(22.6068,23.9490)
3	24.6224	0.1587	(24.2988,24.9460)	(23.9225,25.3223)
4	26.1172	0.1024	(25.9083, 26.3261)	(25.4624, 26.7720)
5	23.5401	0.1086	(23.3186,23.7616)	(22.8812,24.1990)
6	24.5249	0.1388	(24.2418,24.8081)	(23.8428,25.2070)
7	21.7545	0.1078	(21.5346,21.9744)	(21.0961,22.4128)
8	31.4276	0.1646	(31.0918,31.7634)	(30.7220,32.1331)
9	26.2895	0.1641	(25.9548,26.6243)	(25.5845,26.9946)
10	23.103	70.1407	(22.8168,23.3906)	(22.4200,23.7873)
11	27.813	60.1749	(27.4568,28.1703)	(27.0978,28.5294)
12	28.170	50.1981	(27.7665,28.5745)	(27.4301,28.9110)
13	24.948	40.1353	(24.6724,25.2244)	(24.2693,25.6276)
14	23.159	30.1732	(22.8060,23.5126)	(22.4452,23.8733)
15	31.729	90.1432	(31.4378,32.0220)	(31.0440,32.4157)
16	33.509	50.1753	(33.1521,33.8670)	(32.7934,34.2257)
17	32.388	10.1767	(32.0278,32.7485)	(31.6705,33.1057)
18	27.1068	80.1573	(26.7860,27.4276)	(26.4083,27.8054)
19	26.623	30.1234	(26.3715,26.8750)	(25.9536,27.2930)
20	19.7208	80.2088	(19.2950,20.1467)	(18.9682,20.4734)
21	27.055	10.1043	(26.8422,27.2679)	(26.3990,27.7111)
22	23.012	40.1609	(22.6842,23.3406)	(22.3104,23.7144)
23	27.202	40.1591	(26.8780,27.5268)	(26.5022,27.9026)
24	21.510	60.0911	(21.3248,21.6963)	(20.8628,22.1583)
25	30.904	70.1416	(30.6159,31.1936)	(30.2202,31.5892)
26	28.344	60.1159	(28.1083,28.5809)	(27.6806,29.0086)
27	25.344	10.1196	(25.1002,25.5881)	(24.6774,26.0109)
28	24.662	60.1623	(24.3315,24.9937)	(23.9593,25.3659)
29	23.4573	30.1171	(23.2184, 23.6961)	(22.7923, 24.1222)
30	27.437	40.1302	(27.1718,27.7030)	(26.7624,28.1124)
31	28.9268	80.1154	(28.6916,29.1621)	(28.2632,29.5905)
32	26.281	60.1592	(25.9570,26.6063)	(25.5813,26.9820)
33	26.752	50.1992	(26.3463,27.1587)	(26.0108,27.4942)
34	31.937 !	50.1318	(31.6688,32.2063)	(31.2613,32.6138)
35	19.088	30.1539	(18.7745,19.4022)	(18.3930,19.7837)
36	21.463	10.1542	(21.1486,21.7776)	(20.7674,22.1588)
37	26.280	20.1130	(26.0498,26.5106)	(25.6183,26.9421)
38	26.819	10.1417	(26.5300,27.1081)	(26.1345,27.5036)
39	25.744	20.0920	(25.5566,25.9318)	(25.0959,26.3925)
40	24.243	60.0960	(24.0478,24.4395)	(23.5929,24.8943)

Table **3.6**

Vars	$R-Sq$	$R-Sq(adj)$	$C-p$	S	Possible Predictor Models
(i) 4	99.4	99.3	2.2	0.29191	Height (X_2), Weight (X_3), Pulse (X_5), Cholesterol (X_8)
(ii) 5	99.4	99.3	3.4	0.29222	Height (X_2), Weight (X_3), Waist (X_4), Pulse (X_5), Cholesterol (X_8)
(iii) 5	99.4	99.3	3.8	0.29440	Age (X_1), Height (X_2), Weight (X_3), Pulse (X_5), Cholesterol (X_8)
(iv) 4	99.3	99.3	2.8	0.29458	Height (X_2), Weight (X_3), Waist (X_4), Pulse (X_5)
(iv) 3	99.3	99.3	2.2	0.29677	Height (X_2), Weight (X_3), Pulse (X_5)

3.3 Residual Plots for BMI

The Minitab Version 17.0 regression computer program outputs for residual plots of are given in Figure 3.1 below. The paragraphs that follow examine the goodness of fit model based on residual plots.

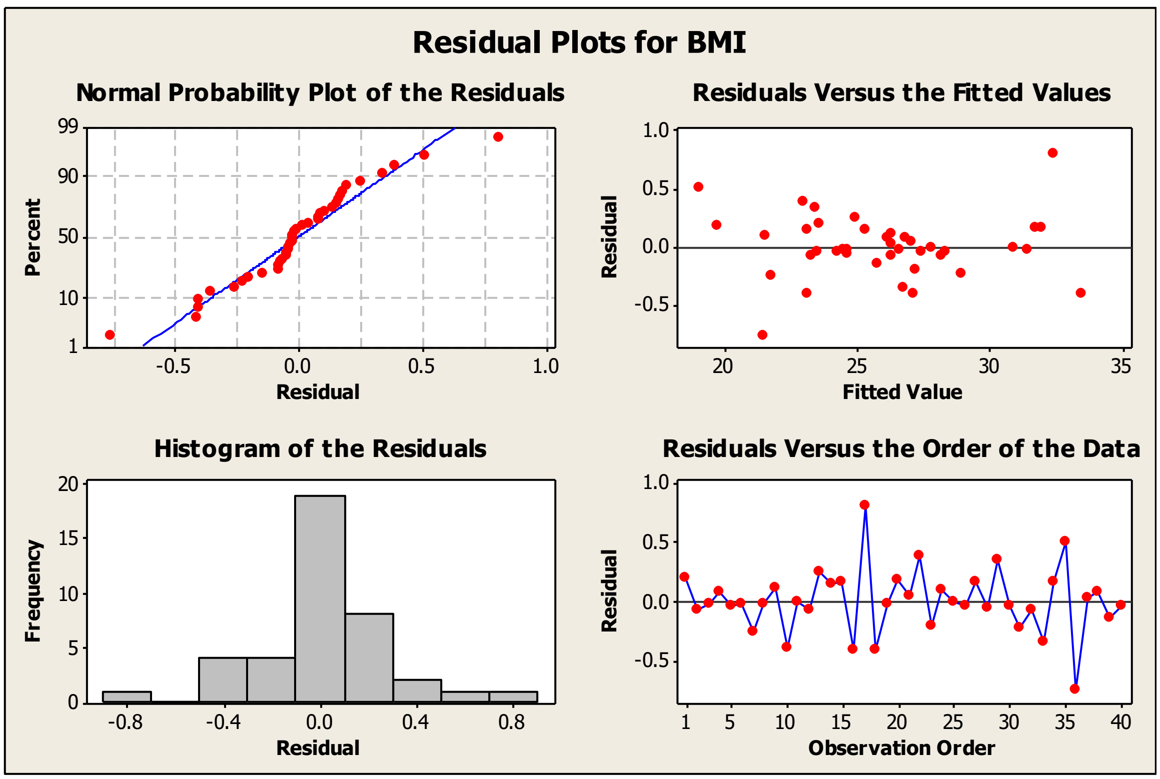

Figure **3.1**

3.3.1 Interpreting the Graphs (Figure 3.1)

A. From the normal probability plot, we observe that there exists an approximately linear pattern. This indicates the consistency of the data with a normal distribution. The outliers are indicated by the points in the upper-right and left-bottom corners of the plot.

B. From the plot of residuals versus the fitted values, it is evident that the residuals get smaller, that is, closer to the reference line, as the fitted values increase. This may indicate that the residuals have non-constant variance, (see, for example, Draper and Smith [2], among others, for details).

C. The histogram of the residuals indicates that no outliers exist in the data.

D. The plot for residuals versus order is also provided in Figure 3.1. It is defined as a plot of all residuals in the order that the data was collected. It is used to find non-random errors, especially of time-related effects. A clustering of residuals with the same sign indicates a positive correlation, whereas a negative correlation is indicated by rapid changes in the signs of consecutive residuals.

3.4 Testing the Adequacy of Multiple Regression Model for Predicting the Adults Body Mass Index *(BMI)*

This section discusses the usefulness and adequacy of the above-developed multiple regression model developed for predicting the adults body mass index (BMI), Y.

3.4.1 Confidence Interval for the Parameters β_i

If we assume that the variation of observations about the line is normal, that is, the error terms ϵ are all from the same normal distribution, $N(0,\sigma^2)$, it can be shown that we can assign $(1-\alpha)100\%$ confidence limits for β_i by calculating

$$\hat{\beta}_i \pm t(n-2, 1-\frac{\alpha}{2}), se(\hat{\beta}_i),$$

where $t(n-2, 1-\frac{\alpha}{2}))$ is the $(1-\alpha)100\%$ percentage point of a t- distribution, with $(n-2)$ degrees of freedom (the number of degrees of freedom on which the estimate s^2 is based). Suppose $\alpha = 0.05$. For $t(38, 0.975)$, we can use $t(40, 0975) = 2.021$, or interpolate in the t table. Thus, we have confidence limits for :

1. 95%; confidence limits for β_1: (-0.00862, 0.011299)
2. 95%; confidence limits for β_2: (-0.82231, -0.72191);
3. 95%; confidence limits for β_3: (0.133755, 0.159361);
4. 95%; confidence limits for β_4: (-0.02134, 0.046362);
5. 95%; confidence limits for β_5: (-0.00248, 0.016675);
6. 95%; confidence limits for β_6: (-0.01432, 0.009748);
7. 95%; confidence limits for β_7: (-0.01717, 0.013274);
8. 95%; confidence limits for β_8: (-0.00014, 0.000564).

3.4.2 Tests of Significance for Individual Parameters

$$H_0 : \beta_i = 0 \text{ versus } H_\alpha : \beta_i \neq 0$$

A test of hypothesis that a particular parameter, say, β_i equals zero, can be conducted by using a t - statistic given by $t = \frac{\hat{\beta}_i - 0}{\text{se}(\hat{\beta}_i)}$. The test can also be conducted by using the F-statistic since the square of a t-statistic (with v degrees of freedom) is equal to an F-statistic with 1 degree of freedom in the numerator and v degrees of freedom in the denominator. That is, $t^2 = F$. Decision Rule: Reject H_0 if $|t| > t\left(n-2, 1-\frac{\alpha}{2}\right)$. Using the Minitab Version 17.0 multiple linear regression computer outputs, the analysis of t statistic values for different β_i 's is given in Table 3.7 below

***Table* 3.7**

Null Hypothesis	$\boldsymbol{t(38, 0.975)^*}$	$\lvert t\rvert$	Inference	Conclusion
$H_0 : \beta_1 = 0$	2.021	0.27	Fail to reject H_0	In the presence of $\boldsymbol{X}_2$, $\boldsymbol{X_3}$, $\boldsymbol{X_4}$, $\boldsymbol{X_5}$, $\boldsymbol{X_6}$, $\boldsymbol{X_7}$, and $\boldsymbol{X_8}$, X_1 is a poor predictor of Y .
$H_0 : \beta_2 = 0$	2.021	31.08	Reject H_0	In the presence of $\boldsymbol{X_1}$, $\boldsymbol{X_3}$, $\boldsymbol{X_4}$, $\boldsymbol{X_5}$, $\boldsymbol{X_6}$, $\boldsymbol{X_7}$, and $\boldsymbol{X_8}, X_2$ is a good predictor of Y .
$H_0 : \beta_3 = 0$	2.021	23.13	Reject H_0	In the presence of $\boldsymbol{X_1}$, $\boldsymbol{X_2}$, $\boldsymbol{X_4}$, $\boldsymbol{X_5}$, $\boldsymbol{X_6}$, $\boldsymbol{X_7}$, $\boldsymbol{X_8}$, X_3 is a good predictor of Y.
$\boldsymbol{H_0 : \beta_4 = 0}$	2.021	0.75	Fail to reject H_0	In the presence of $\boldsymbol{X_1}$, $\boldsymbol{X_2}$, $\boldsymbol{X_3}$, $\boldsymbol{X_5}$, $\boldsymbol{X_6}$, $\boldsymbol{X_7}$, $\boldsymbol{X_8}$, $\boldsymbol{X_4}$ is a poor predictor of Y.
$\boldsymbol{H_0 : \beta_5 = 0}$	2.021	1.50	Fail to reject H_0	In the presence of $\boldsymbol{X_1}$, $\boldsymbol{X_2}$, $\boldsymbol{X_3}$, $\boldsymbol{X_4}$, $\boldsymbol{X_6}$, $\boldsymbol{X_7}$, $\boldsymbol{X_8}$, $\boldsymbol{X_5}$ is a poor predictor of Y.
$\boldsymbol{H_0 : \beta_6 = 0}$	2.021	0.38	Fail to reject H_0	In the presence of $\boldsymbol{X_1}, \boldsymbol{X_2}, \boldsymbol{X_3}, \boldsymbol{X_4}$, $\boldsymbol{X_5}$, $\boldsymbol{X_7}$, $\boldsymbol{X_8}$, $\boldsymbol{X_6}$ is a poor predictor of Y .
$\boldsymbol{H_0 : \beta_7 = 0}$	2.021	0.26	Fail to reject H_0	In the presence of $\boldsymbol{X_1}$, $\boldsymbol{X_2}$, $\boldsymbol{X_3}$, $\boldsymbol{X_4}$, $\boldsymbol{X_5}$, $\boldsymbol{X_6}$, $\boldsymbol{X_8}$, $\boldsymbol{X_7}$ is a poor predictor of Y .
$\boldsymbol{H_0 : \beta_8 = 0}$	2.021	1.20	Fail to reject H_0	In the presence of $\boldsymbol{X_1}, \boldsymbol{X_2}$, $\boldsymbol{X_3}$, $\boldsymbol{X_4}$, $\boldsymbol{X_5}$, $\boldsymbol{X_6}$, $\boldsymbol{X_7}$, $\boldsymbol{X_8}$ is a poor predictor of Y .

*For $\boldsymbol{t(38, 0.975)}$, we can use $t(40, 0.975) = 2.021$ or interpolate in the $t-$ table.

3.4.3 F-Test for Significance of Regression

For details on it, see, for example, Draper and Smith [5], Tamhane and Dunlop [25], and Mendenhall and Sincich [16], Chatterjee and Hadi [2], Montgomery et al. [17], among others. For our proposed multiple regression model, we have

Null Hypothesis: $\boldsymbol{H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = 0}$ (The regression is not significant) versus

Alternate Hypothesis: H_a : at least one of $\beta_i's \neq 0$ (The regression is significant).

Test Statistic: $F = \frac{MS_{\text{reg}}}{s^2}$.

Decision Rule: Reject H_0 if $\boldsymbol{F > F_{\alpha}\,(v_1 = k, v_2 = n - (k+1), 1 - \alpha)}$,

where n = number of values in the sample data = 40,

$\boldsymbol{k}$ = number of estimated β regression coefficients = 8,

$\boldsymbol{k+1} = 8 + 1 = 9$ = number of estimated $\boldsymbol{\beta}$ parameter,

$\boldsymbol{v_1 = k = df}$ in the numerator = 8,

and $\boldsymbol{v}_2 = \boldsymbol{n - (k+1) = df}$ in the denominator = 31

In the decision rule, we compare the calculated $\boldsymbol{F}$ test statistic to a tabulated $\boldsymbol{F_{\alpha}}$ value based on $\boldsymbol{v_1 = kdf}$ in the numerator and $\boldsymbol{v_2 = n-(k+1)df}$ in the denominator for the considered value of $\boldsymbol{\alpha}$, using $\boldsymbol{F}$ distribution.

Thus, for our proposed multiple regression model, the decision rule is given by

Decision Rule: Reject H_0 if $\boldsymbol{F > F_{0.05}\,(v_1 = 8, v_2 = 31, 0.95)}$, for $\boldsymbol{\alpha = 0.05}$.

The value of F - statistic for testing the hypothesis is that at least one of the predictor variables contributes significant information for the prediction of the adult's body mass index (BMI), Y. In the computer output 17 (Table 4.3), it is calculated as $\boldsymbol{F = 615.93}$. Comparing this with the critical value of $\boldsymbol{F_{0.05}\,(v_1 = 8, v_2 = 31, 0.95) = 2.18}$ at $\alpha = 0.05$, we reject the null hypothesis: $\boldsymbol{H_0 : \beta_1 = \beta_2 =} \boldsymbol{\beta}_3 = \boldsymbol{\beta}_4 = \boldsymbol{\beta}_5 = \boldsymbol{\beta}_6 = \boldsymbol{\beta_7} = \boldsymbol{\beta_8} = \boldsymbol{0}$, that is, the regression is not significant. Thus, the overall regression is statistically significant. In fact, $\boldsymbol{F = 615}.93$ exceeds $\boldsymbol{F_{0.05}\,(v_1 = 8, v_2 = 31, 0.95)} = 2.18$, and is significant at a p-value $(= 0.000) < 0.005$. It appears that at least one of the predictor variables contributes information for the prediction of Y.

4 Concluding Remarks

From the above analysis, it appears that our multiple regression model for predicting the adult's body mass index $(BMI), Y$, is useful and adequate. In the presence of $\boldsymbol{X_1, X_3, X_4, X_5, X_6, X_7}$, and $\boldsymbol{X_8}$, X_2 is a good predictor of Y. In the presence of $\boldsymbol{X_1, X_2, X_4, X_5, X_6, X_7, X_8}, X_3$ is a good predictor of Y. As the values of R^2 and R_a^2 are not very different, it appears that at least one of the predictor variables contributes information for the prediction of Y. The coefficient of variation $\boldsymbol{CV = 1.17303\%}$ also tells us that the standard deviation of the adult's body mass index (BMI), Y, is only $\boldsymbol{1.17303\%}$ of their mean. Also, since the test statistic value of F calculated from the data, $\boldsymbol{F = 615.93}$, exceeds the critical value of $\boldsymbol{F_{0.05}\,(v_1 = 8, v_2 = 31, 0.95) = 2.18}$, at $\alpha = 0.05$, we reject the null hypothesis: $\boldsymbol{H_0 : \beta_1 = \beta_2 =} \boldsymbol{\beta}_3 = \boldsymbol{\beta}_4 = \boldsymbol{\beta}_5 = \boldsymbol{\beta_6} = \boldsymbol{\beta_7} = \boldsymbol{\beta_8} = \boldsymbol{0}$, that is, the regression is not significant. Hence, our multiple regression model for predicting the adult's body mass index (BMI), Y, seems to be useful and adequate, and the overall regression is statistically significant. The C_p-statistic criterion and residual plots of Y (Figure 3.1) as discussed above also confirm the adequacy of our model. For future work, one can consider to develop and study similar models for other issues and problems associated with the fields of medical, biological, behavioral, and other applied sciences. One can also develop similar models by adding other variables, for example, the gender, marital status, employment status, race and ethnicity of the adults, as well as the squares, cubes, and, cross products of $\boldsymbol{X_1}, \boldsymbol{X}_2, X_3, \boldsymbol{X_4}, \boldsymbol{X_5}, \boldsymbol{X_6}, \boldsymbol{X_7}$, and $\boldsymbol{X_8}$. In addition, one could also study the effect of some data transformations. We believe that the present study would be useful for researchers in the fields of medical and other applied sciences.

Authors' Contributions. All authors have equally made contributions. All authors read and approved the final manuscript.

Funding. The authors state that they have no funding source for this paper.

Availability of Data and Materials. Not applicable.

Declarations Conflict of interest. The authors declare that they have no competing interests. The authors state that no funding source or sponsor has participated in the realization of this work.

Acknowledgement. The author are thankful to the Editor in chief and the anonymous reviewers whose constructive comments and suggestions have improved the quality and presentation of the paper.

References

[1] L. Akil and H. A. Ahmad, Relationships between Obesity and Cardiovascular Diseases in Four Southern States and Colorado, *Journal of Health Care for the Poor and Underserved,* **22** (4 Suppl), (2011), 61-72. doi: 10.1353/hpu.2011.0166.

[2] S. Chatterjee and A. S. Hadi, *Regression Analysis by Example, 5th Edition.* John Wiley & Sons, New York, USA, 2012.

[3] T. J. Cleophas and A. H. Zwinderman, *Regression Analysis in Medical Research for Starters and 2nd Levelers.* Springer, Basel, Switzerland, 2018.

[4] S. Chatterjee and J. S. Simonoff, *Handbook of Regression Analysis.* John Wiley & Sons, New York, USA, 2013.

[5] N. R. Draper and H. Smith, *Applied Regression Analysis (3rd edition).* New York: John Wiley & Sons, INC., USA, 1998.

[6] G. Feng, G. Qin, T. Zhang, Z. Chen and Y. Zhao, Common Statistical Methods and Reporting of Results in Medical Research, *Cardiovascular Innovations and Applications,* **6**(3), (2022), 117-125. DOI 10.15212/CVIA.2022.0001.

[7] C. I. Guzman and B. M. G. Kibria, Developing Multiple Linear Regression Models for the Number of Citations: A Case Study of Florida International University Professors, *International Journal of Statistics and Reliability Engineering,* **6**(2) (2019), 75-81.

[8] J. E. Hall, J. M. do Carmo, A. A. da Silva, Z. Wang, and M. E. Hall, Obesity-induced hypertension: interaction of neurohumoral and renal mechanisms, *Circulation Research,* **116** (6), (2015), 991-1006. doi: 10.1161/CIRCRESAHA.116.305697.

[9] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis, 9th Edition.* PearsonPrentice Hall, Upper Saddle River, NJ, USA, 2023.

[10] S. Jiang, W. Lu, X. Zong, H. Ruan and Y. Liu, Obesity and hypertension (Review), *Experimental and Therapeutic Medicine,* **12**, (2016), 2395-2399. https://doi.org/10.3892/etm.2016.3667. 19

[11] R. Jabeen, T. Rasheed and H. Talat, Co-relation of Acanthosis Nigricans with Insulin Resistance and Type 2 Diabetes Mellitus in a Tertiary Care Hospital, *Journal of Research in Medical and Dental Science,* **11** (01), (2023), 265-269.

[12] B. M. G. Kibria, Linear and Non-linear Regression Models: Theory, Simulation and Applications, *24th Annual Conference of Vijñāna Parishad of India and National Seminar on New Thrust Areas in Mathematics, Mathematical Sciences and Engineering, India,* April 2023, 28-30.

[13] N. Keum, D. C. Greenwood, D. H. Lee, R. Kim, D. Aune, W. Ju, F. B. Hu and E. L. Giovannucci, Adult weight gain and adiposity-related cancers: a dose-response meta-analysis of prospective observational studies, *Journal of the National Cancer Institute,* **107**(2) (2015). djv088. https://doi.org/10.1093/jnci/djv088.

[14] B. Klop, J. W. Elte and M. C. Cabezas, Dyslipidemia in obesity: mechanisms and potential targets. *Nutrients,* **5**(4) (2013), 1218-1240. doi: 10.3390/nu5041218.

[15] M. Leggio, M. Lombardi, E. Caldarone, P. Severi, S. D'Emidio, M. Armeni, V. Bravi, M. G. Bendini and A. Mazza, The relationship between obesity and hypertension: an updated comprehensive overview on vicious twins. *Hypertension Research,* **40** (2017), 947-963. https://doi.org/10.1038/hr.2017.75.

[16] W. Mendenhall, and T. Sincich, *A Second Course in Statistics: Regression* Analysis. PearsonPrentice Hall, Upper Saddle River, NJ, USA, 2011.

[17] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to Linear Regression Analysis, Fourth Edition.* John Wiley & Sons, New York, USA, 2013.

[18] I. L. Mertens and L. F. Van Gaal, Overweight, obesity, and blood pressure: the effects of modest weight reduction. *Obesity Research,* **8**(3) (2000), 270-278. doi: 10.1038/oby.2000.32.

[19] J. Neter, M. H. Kutner, C. J. Nachtsheim and W. Wasserman, *Applied Linear Statistical Models, 4th Edition.* WCB McGraw-Hill, New York, USA, 1996.

[20] N. T. Nguyen, C. P. Magno, K. T. Lane, M. W. Hinojosa and J. S. Lane, Association of hypertension, diabetes, dyslipidemia, and metabolic syndrome with obesity: findings from the National Health and Nutrition Examination Survey, 1999 to 2004, *Journal of the American College of Surgeons,* **207**(6) (2008), 928-934. doi: 10.1016/j.jamcollsurg.2008.08.022.

[21] K. Narkiewicz, Obesity and hypertension - the issue is more complex than we thought. *Nephrology Dialysis Transplantation*, **21**(2) (2006), 264-267. https://doi.org/10.1093/ndt/gfi290. 20

[22] B. Rajeshwari and G. Laishram, Review Article on Obesity as a Contributor to Type-2 Diabetes Mellitus among Adolescents. *Journal of Research in Medical and Dental Science,* **11**(01) (2023), 108-114.

[23] E. Surez, C. M. Perez, R. Rivera, and M. N. Martinez, *Applications of Regression Models in Epidemiology.* John Wiley & Sons, New York, USA, 2017.

[24] G. Seravalle and G. Grassi, Obesity and hypertension. *Pharmacological Research,* **122** (2017), 1-7. https://doi.org/10.1016/j.phrs.2017.05.013.

[25] A. C. Tamhane, and D. D. Dunlop, *Statistics and Data Analysis: From Elementary to Intermediate (1st edition),* Pearson Prentice Hall, Upper Saddle River, NJ, USA, 2000.

[26] M. F. Triola, Elementary Statistics Using Excel, Fourth Edition, Addision-Wesley, Boston, USA, 2010.

[27] W. Vach, *Regression Models as a Tool in Medical Research, 1st Edition*, Chapman and Hall/CRC, New York, USA, 2013. DOI: https://doi.org/10.1201/b12925.

[28] E. Vittinghoff, D. V. Glidden, S. C. Shiboski and C. E. McCulloch, *Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models, Second Edition,* Springer New York, NY, USA, 2014.

[29] T. Young, P. E. Peppard and D. J. Gottlieb, Epidemiology of obstructive sleep apnea: a population health perspective. *The American Journal of Respiratory and Critical Care Medicine.* **65**(9) (2002), 1217-1239. doi: 10.1164/rccm.2109080.

APPENDIX 1
(Adult's Body Mass Index (BMI) Data, $n = 40$)
(Source: Triola [26])

Age	Ht	Wt	Waist	Pulse	Systolic	Diastolic	Cholesterol	BMI
58	70.8	169.1	90.6	68	125	78	522	23.8
22	66.2	144.2	78.1	64	107	54	127	23.2
32	71.7	179.3	96.5	88	126	81	740	24.6
31	68.7	175.8	87.7	72	110	68	49	26.2
28	67.6	152.6	87.1	64	110	66	230	23.5
46	69.2	166.8	92.4	72	107	83	316	24.5
41	66.5	135	78.8	60	113	71	590	21.5
56	67.2	201.5	103.3	88	126	72	466	31.4
20	68.3	175.2	89.1	76	137	85	121	26.4
54	65.6	139	82.5	60	110	71	578	22.7
17	63	156.3	86.7	96	109	65	78	27.8
73	68.3	186.6	103.3	72	153	87	265	28.1
52	73.1	191.1	91.8	56	112	77	250	25.2
25	67.6	151.3	75.6	64	119	81	265	23.3
29	68	209.4	105.5	60	113	82	273	31.9
17	71	237.1	108.7	64	125	76	272	33.1
41	61.3	176.7	104	84	131	80	972	33.2
52	76.2	220.6	103	76	121	75	75	26.7
32	66.3	166.1	91.3	84	132	81	138	26.6
20	69.7	137.4	75.2	88	112	44	139	19.9
20	65.4	164.2	87.7	72	121	65	638	27.1
29	70	162.4	77	56	116	64	613	23.4
18	62.9	151.8	85	68	95	58	762	27
26	68.5	144.1	79.6	64	110	70	303	21.6
33	68.3	204.6	103.8	60	110	66	690	30.9
55	69.4	193.8	103	68	125	82	31	28.3
53	69.2	172.9	97.1	60	124	79	189	25.5
28	68	161.9	86.9	60	131	69	957	24.6
28	71.9	174.8	88	56	109	64	339	23.8
37	66.1	169.8	91.5	84	112	79	416	27.4
40	72.4	213.3	102.9	72	127	72	120	28.7
33	73	198	93.1	84	132	74	702	26.2
26	68	173.3	98.9	88	116	81	1252	26.4
53	68.7	214.5	107.5	56	125	84	288	32.1
36	70.3	137.1	81.6	64	112	77	176	19.6
34	63.7	119.5	75.7	56	125	77	277	20.7
42	71.1	189.1	95	56	120	83	649	26.3
18	65.6	164.7	91.1	60	118	68	113	26.9
44	68.3	170.1	94.9	64	115	75	656	25.6
20	66.3	151	79.9	72	115	65	172	24.2