



**INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH  
TECHNOLOGY**

**STUDY OF THE RELATIONSHIP BETWEEN THE AGE, BODY MASS INDEX AND  
HIGH BLOOD PRESSURE**

**Aminu Suleiman\*, Abubakar Usman, Aliyu Ismail Ishaq**

\* Statistics Branch, Mathematics Department, School of Basic Science and Research, Sharda University,  
Greater Noida-201306.

---

**ABSTRACT**

The relationship between the level of blood pressure and BMI has been perceived to be linear and strong. To lay a ground we need to have an insight idea or knowledge on the issues of Hypertension and high body mass index including age as well as its going to be of high importance for us in tackling and handling the blood pressure level. This research carried out in Sharda University, examined the relationship of blood pressure, body mass index (BMI) and age among 100 patient by using a statistical tools regression and correlation, we also test the significance of the coefficient of regression. It found that age, BMI, systolic and diastolic are positively correlated and the regression model and the parameters are statistically significant. Which proved to us that there is a relationship between body mass index with blood pressure, so also age has a relationship with blood pressure but not as strong as that of body mass index.

**KEYWORDS:**

---

**INTRODUCTION**

Body mass index is currently one of the primary focus in the treatment recommended for obesity, the risk of high blood pressure is generally associated with the increase in body fat. It is very critical when it comes to understanding pathophysiological processes which lead to excessive fat-related metabolic disease that when we are looking for the relationship between body weight and metabolic diseases. The processes in which blood pressure occurs is when your heart beats, it pumps blood round your body to give it the energy and oxygen it needs. As the blood moves, it pushes against the sides of the blood vessels. The strength of this pushing is your blood pressure. If your blood pressure is too high, it puts extra strain on your arteries (and your heart) and this may lead to heart attacks and strokes. That's why it becomes very essential for us to study both the incidence, prevalence and consequences of elevated blood pressure. Elevated blood pressure greatly increases the risk of cardiovascular disease, the leading cause of death in almost all part of the world. This research has been discovered widely across many age groups and amongst people of different racial backgrounds for many years though the main cause of hypertension (high blood pressure) remains unknown, but body fat is thought to be a major casual factor of increasing the blood pressure, the true relationship between high blood pressure and body mass remains inconspicuous.

Body mass index which is abbreviated (BMI) is happens to be the key index of relating weight to height. BMI is a person's weight in kilograms (kg) divided by his or her height in meters squared. And age is the whole duration of a being.

In this research our problem is to find the relationship between the variables i.e Age, BMI, and blood pressure, at the end we would like to know that does the increase in body weight and age of a person lead to the increase in blood pressure. Is body mass index also related to age? We also test the effect of some medication on patients with high blood pressure

**MATERIAL AND METHODOLOGY**

In this research work we are going to use a data collected on 100 patients form Sharda hospital the data contains the age, body mass index and blood pressure measurement both systolic and diastolic. The data will be used to determine the extent and direction of the relationship between the variables Age, body mass index BMI and blood pressure (both

systolic and diastolic). We also estimate the regression model between the variables based on the sample collected and test the significance of the coefficient of regression. -

The validity as well as the relevance of the statistical investigation depends on the statistical tool or model used in data analysis. In this research work, I am going to use statistical models in the analysis of the data which are: - Regression, correlation coefficient R, coefficient of determination  $R^2$ .

### CORRELATION

Correlation measures the strength and the direction of linear relationship between two variables. The linear correlation coefficient is some time referred as the Pearson product moment correlation coefficient in honor of its developer Karl Pearson. Correlation is a statistical measure that indicates the extent to which two or more variables fluctuate together. A positive correlation indicates the extent to which those variables increase or decrease in parallel, while a negative correlation indicates the extent to which one variable increases as the other decreases. No correlation means there is no linear correlation or a weak linear correlation i.e.  $r$  is close to zero, a value near zero means there is a random, nonlinear relationship between the two variables.

$$r = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{\left( n \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2 \right) \left( n \sum_{i=1}^n y_i^2 - \left( \sum_{i=1}^n y_i \right)^2 \right)}}$$

### COEFFICIENT OF DETERMINATION

Convenient and useful way of interpreting the value of coefficient of correlation between two variables is to use the square of the coefficient of correlation which is called coefficient of determination. Thus, the coefficient of determination equals to  $R^2$  which is defined as the ratio of the explained variance to the total variance. The coefficient of determination is such that  $0 \leq r^2 \leq 1$ , and denotes the strength of the linear association between  $x$  and  $y$ . the coefficient of determination represent the percent of the data that is closest to the line of the best fit. Coefficient of determination  $R^2 = \text{Explained variance} / \text{Total variance}$

### REGRESSION

Regression is the techniques used to study the relationship between variables. Linear regression is used for a special class of relationship, those that can be described by straight line. The method of least square can be used to estimate the values of the unknown parameters of the regression model in such way that this least square estimate when fitted to the model provides a line of best of fit to the data. Since the first step in regression is to assume that the relationship between the dependent and independent variable can be fitted by a linear model.

### MULTIPLE REGRESSION

Multiple regression analysis is a powerful technique used for predicting the unknown value of a variable from the known value of two or more variables- also called the predictors. By multiple regression, we mean models with just one dependent which is systolic or diastolic measures of blood pressure and two independent (exploratory) variables i.e. age and BMI. The variable whose value is to be predicted is known as the dependent variable and the ones whose known values are used for prediction are known independent (exploratory) variables.

### FITTING A REGRESSION LINE

Consider fitting a regression line equation between the independent variable  $X$  and dependent variable  $Y$ . let this be represented by:-

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n$$

In matrix notation regression line can be written as:

$$E(Y) = X \beta$$

Where Y is 1 x n vector of observation, X is n x (p+1) matrix given by

$$X = [ 1 \quad x_1 \quad x_2 \quad \dots \quad x_n ]$$

And is a vector of order (p+1) unknown parameters. 1 is a vector of unities.

The method of fitting any type of regression are similar either simple, multiple or polynomial regression. Data of multiple linear regression will be of the type given in the following table:

Y	X <sub>1</sub>	X <sub>2</sub>	...	X <sub>n</sub>
y <sub>1</sub>	X <sub>11</sub>	X <sub>12</sub>	...	X <sub>1n</sub>
y <sub>2</sub>	X <sub>21</sub>	X <sub>22</sub>	...	X <sub>2n</sub>
⋮	⋮	⋮	⋮	⋮
y <sub>n</sub>	X <sub>n1</sub>	X <sub>n2</sub>	...	X <sub>nn</sub>

Hence y<sub>i</sub> denotes the i<sup>th</sup> observation and X<sub>i</sub> represent the i<sup>th</sup> level of independent variable x<sub>j</sub>. The model equation for the above data can be written as follows:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + e_i \quad i=1, \dots, n.$$

$$y_i = \beta_0 + \sum_{j=1}^n \beta_j x_{ij} + e_i$$

Method of least square is used to estimate the β's. the method of least square chooses β's in the above equation. So that the sum square e<sub>i</sub> are minimized. The function of the least square is

$$L = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \beta_0 - \sum \beta_j x_{ij})^2$$

The function L is to be minimized with respect to β<sub>0</sub>, β<sub>1</sub>, β<sub>2</sub>, ..., β<sub>n</sub>. The least square estimate must therefore be satisfy

$$\frac{\delta L}{\delta \beta_0} \{ \beta_0, \beta_1, \beta_2, \dots, \beta_n \} = -2 \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^n b_j x_{ij}) = 0$$

$$\frac{\delta L}{\delta \beta_j} \{ \beta_0, \beta_1, \beta_2, \dots, \beta_n \} = -2 \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^n b_j x_{ij}) x_{ij} = 0$$

By simplifying the above equation, we have:

$$n\beta_0 + \beta_1 \sum_{i=1}^n x_{i1} + \dots + \beta_k \sum_{i=1}^n x_{in} = \sum_{i=1}^n y_i$$

$$\beta_0 \sum_{i=1}^n x_{i1} + \beta_k \sum_{i=1}^n x_{i1}^2 + \beta_2 \sum_{i=1}^n x_{i1} x_{i2} + \dots + \beta_k \sum_{i=1}^n x_{i1} x_{ik} = \sum_{i=1}^n x_{i1} y_i$$

$$\beta_0 \sum_{i=1}^n x_{ik} + \beta_k \sum_{i=1}^n x_{ik}^2 + \beta_2 \sum_{i=1}^n x_{ik} x_{i2} + \dots + \sum_{i=1}^n x_{ik}^2 = \sum_{i=1}^n x_{ik} y_i$$

These equation are p=k+1 equation for each regression coefficient solution of these equation be the least square estimation of β<sub>0</sub>, ..., β<sub>k</sub>. The above equation are known as normal equations.

The model can be written in a matrix notation as follows:  
y = Xβ + e, where

$$y = (y_1, y_2, \dots, y_n) ', X = \begin{matrix} & 1 & x_{11} & x_{12} & x_{1k} \\ 1 & x_{21} & x_{22} & x_{2k} \end{matrix}$$

$$\begin{matrix} \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & x_{nk} \end{matrix}$$

$\beta = [\beta_0, \beta_1, \dots, \beta_k]'$ ,  $e = (e_1, e_2, \dots, e_n)'$ , and  $D(e) = \delta^2 I$ ,  $E(e) = 0$

The main reason for doing this work is to find the vector of least square estimators which minimizes

$$L = e'e = (y - X\beta)'(y - X\beta)$$

The normal equation in this case are given by

$$X'X\beta = X'Y$$

Or 
$$\hat{\beta} = b = (X'X)^{-1} X'Y$$

And the fitted line is: 
$$\hat{y} = Xb$$

### TEST OF SIGNIFICANCE OF REGRESSION

Whenever you fitted your regression, there is a need to test the significance of the fit. To check the significance of the fit we need the following sum of squares in the analysis of variance.

✓ Sum of squares due to regression = [SSR] = 
$$b'X'y - \frac{(\sum_{i=1}^n y_i)^2}{n}$$

✓ Sum of squares due to residuals = [SSE] = 
$$y'y - b'X'y$$

✓ Total sum of squares = 
$$y'y - \frac{(\sum_{i=1}^n y_i)^2}{n}$$

The covariance of matrix b is given by

$$Cov(b) = \delta^2 (X'X)^{-1}$$

It is necessary to estimate  $\delta^2$ . An unbiased estimator of  $\delta^2$  is given by

$$\hat{\delta}^2 = s^2 = \frac{SSE}{n - p}$$

In order to determine if there is a linear relationship between the response variable y and a subset of the regressor  $x_1, \dots, x_j$  we set up the following hypothesis:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1 : \beta_j \neq 0, \text{ for at least one } j.$$

When the null hypothesis is rejected, it implies that at least one if  $x_1 \dots x_k$  contributes significantly to the model. The test statistics for  $H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$  is

$$F = \frac{SS_R / R}{SS_E / n - k - 1} = \frac{MS_R}{MS_E}$$

We reject  $H_0$  if  $F_0$  is greater than  $F_{\alpha, k, n-k-1}$

**ANOVA TABLE**

Source of variation	Degree of Freedom (d.f)	Sum squares (SS)	Mean squares (MS)	F-statistics
Regression	k	SS <sub>R</sub>	$\frac{SS_R}{k} = MS_R$	$\frac{MS_R}{MS_E}$
Error	n-k-1	SS <sub>E</sub>	$\frac{SS_E}{n-k-1} = MS_E$	
Total	n-1	SS <sub>Y</sub>		

The coefficient of determination  $R^2$  is given by  $R^2 = \frac{SS_R}{SS_y} = 1 - \frac{SS_E}{SS_{yy}}$

$R^2$  measure the reduction in variability of y obtained using regressors  $x_1, \dots, x_k$ , also  $0 \leq R^2 \leq 1$ . Large value of  $R^2$  do not necessarily imply that the model is good as addition of the variables will always increase  $R^2$ . For this reason many a times  $R^2$  adjusted is used as statistics given below:

$$R^2_{adj} = 1 - \frac{SS_E / (n-p)}{SS_{YY} / (n-1)} = 1 - \left(\frac{n-1}{n-p}\right)(1 - R^2)$$

Generally the adjusted  $R^2$  will not always increase as variables are added to the model. In fact, if unnecessary variables are added to the model  $R^2$  adjusted will often decrease.

**RESULT AND NUMERICAL APPLICATION**  
**CORREALTAION AND REGRESSION ANALYSIS OF DATA**

In finding the regression we use variable age and body mass index as predictors that’s independent variables while the average of systolic and diastolic values is use to represent blood pressure because here in our analysis our blood pressure values i.e. average number of systolic and diastolic value will serve as our responses (independent variables). In correlation, we take the correlation between all the variables age, body mass index and blood pressure (average of systolic and diastolic) to see the existence of the relationship among them.

Correlation: Age, BMI, Blood Pressure

Correlations

	Age	BMI
BMI	0.498341 <0.0001	
Blood Pressure	0.598482 <0.0001	0.736728 <0.0001

Cell Contents: *Pearson correlation*  
*P-Value*

Multiple Regression: Blood Pressure versus Age, BMI

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
--------	----	--------	--------	---------	---------

Regression	2	13009.5	6504.73	77.14	<0.0001
Age	1	1508.7	1508.68	17.89	<0.0001
BMI	1	5419.9	5419.92	64.27	<0.0001
Error	97	8179.7	84.33		
Lack-of-Fit	84	7739.5	92.14	2.72	0.0236
Pure Error	13	440.2	33.86		
Total	99	21189.1			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
9.18295	61.40%	60.60%	58.36%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-1.20	12.32	-0.10	0.9224	
Age	0.26729	0.06319	4.23	<0.0001	1.33
BMI	4.1434	0.5168	8.02	<0.0001	1.33

Regression Equation

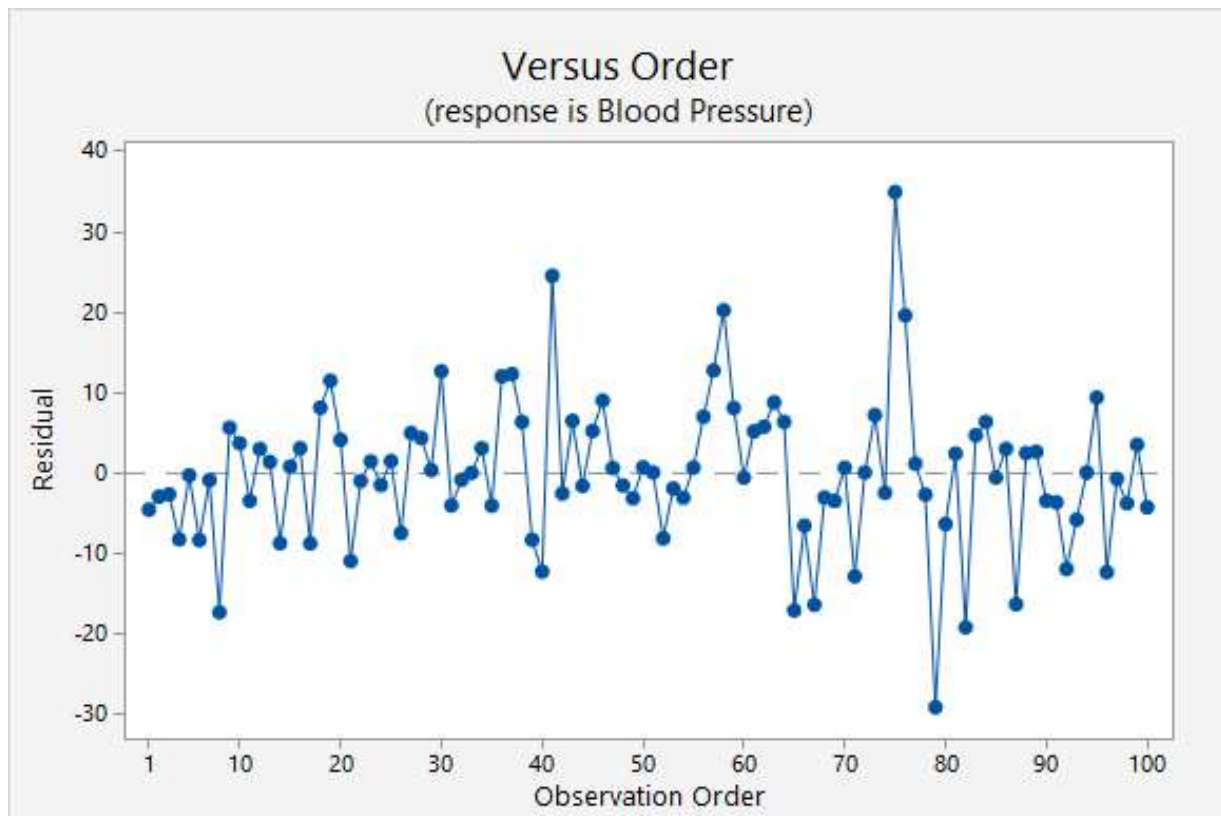
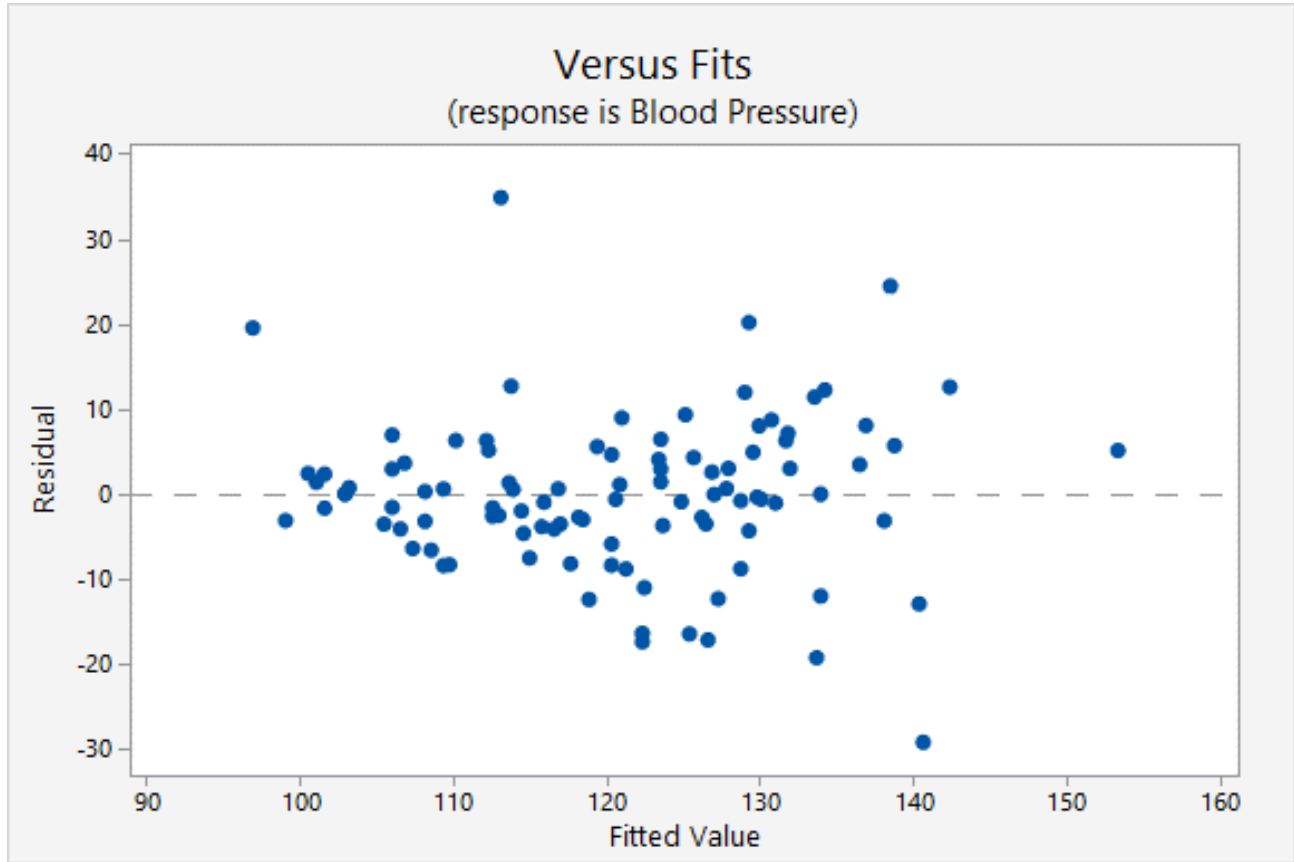
$$\text{Blood Pressure} = -1.20 + 0.26729 \text{ Age} + 4.1434 \text{ BMI}$$

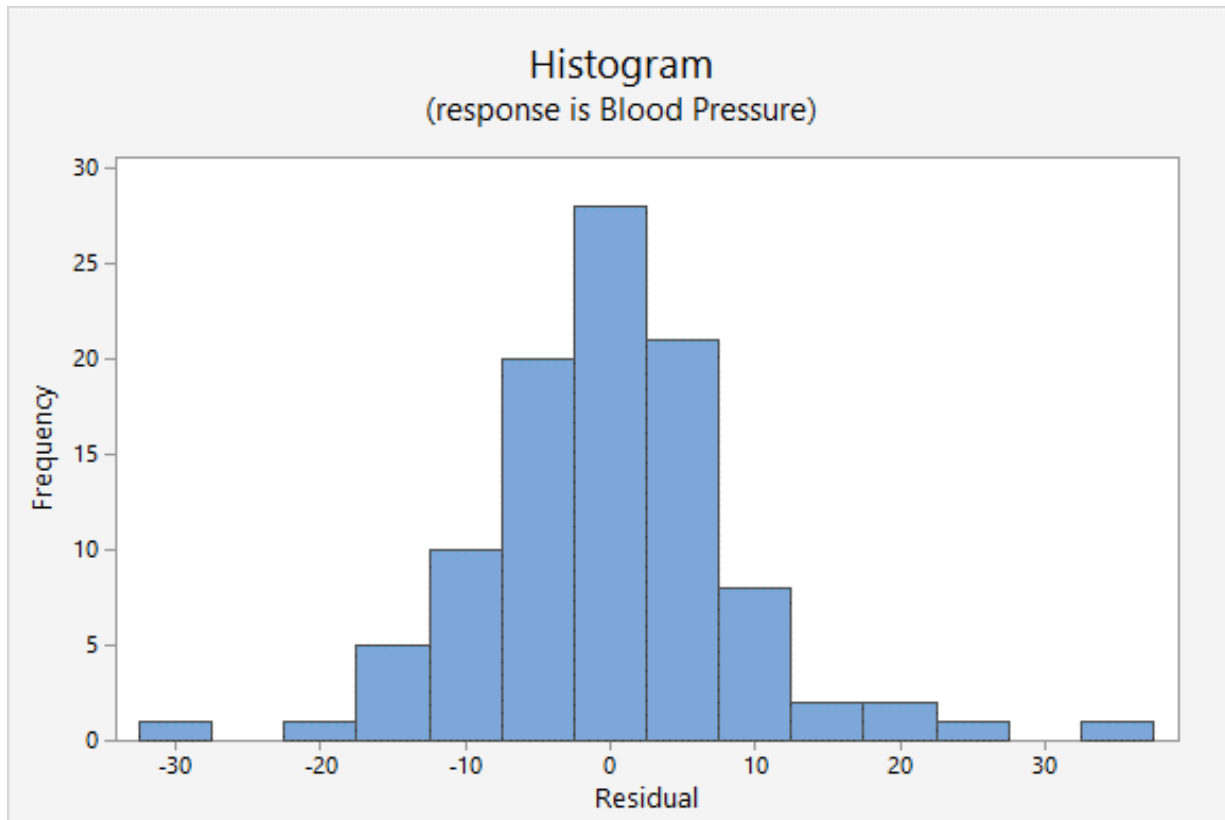
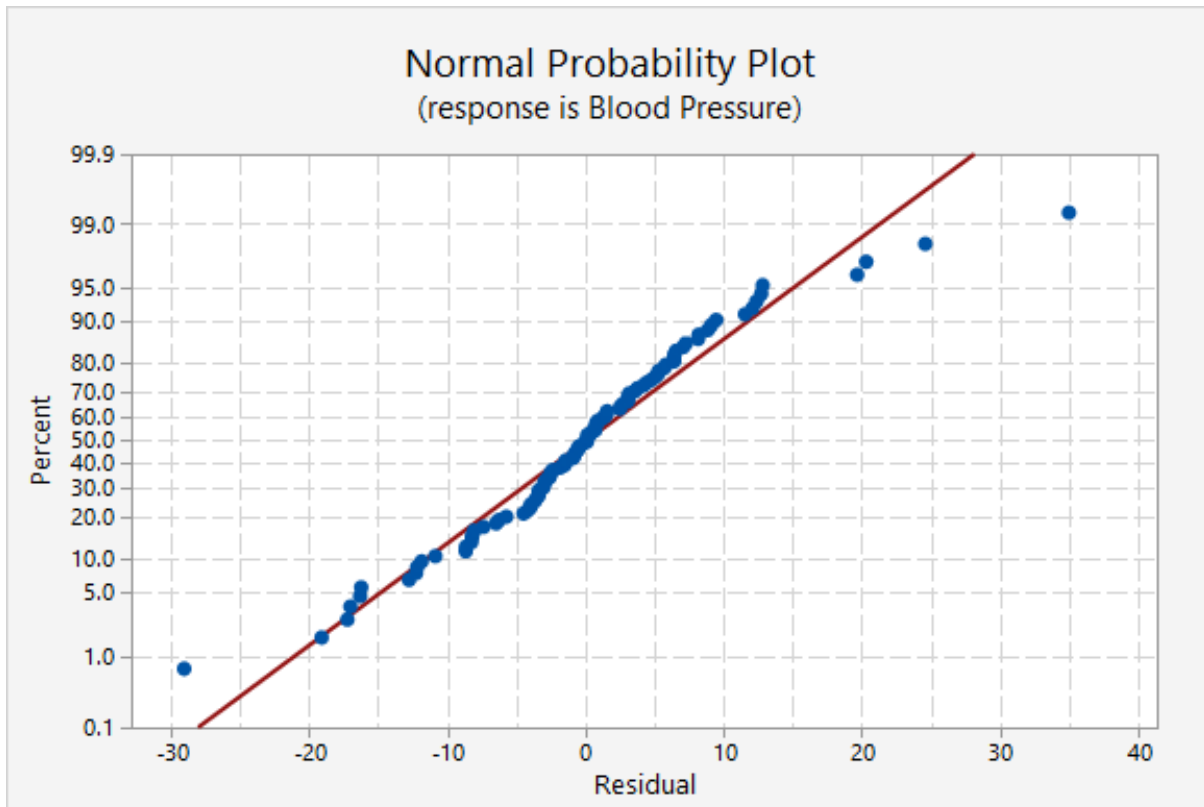
Fits and Diagnostics for Unusual Observations

Obs	Blood Pressure	Fit	Resid	Std Resid	
41	163	138.469	24.5312	2.72	R
58	149.5	129.247	20.2532	2.23	R
61	158.5	153.305	5.1952	0.59	X
75	148	113.074	34.9256	3.83	R
76	116.5	96.902	19.5981	2.19	R
79	111.5	140.608	-29.1080	-3.31	R
82	114.5	133.658	-19.1576	-2.12	R

R Large residual

X Unusual X







## DISCUSSION

This table shows that Age and BMI are positively correlated but the correlation is not very strong, since the approximated value of correlation ( $r$ ) 0.5, but the correlation analysis between age and blood pressure is positively correlated with strong correlation but not as strong as that of BMI and blood pressure, it is also shown that there is a strong linear relationship between blood pressure and BMI as the correlation value ( $r$ ) is 0.74. The regression analysis show that the model is statistically significant since the P-value of all in the analysis of variance table is less than ( $\alpha$ ) level of significance i.e. 0.05 as well as the parameter age and BMI but the parameter constant is not significant since its P-value is greater than ( $\alpha$ ) level of significance which is 0.9224. And the r-square adjusted indicate that there is 60.6% of variability between the variables.

## CONCLUSION

In conclusion, we have done two separate analysis that is correlation and regression. The result showed positive correlation between the variables age, BMI and blood pressure. We found that weight loss may contribute positively in managing the problem of hypertension and age may also contribute to blood pressure level. The regression model shows that both the models and parameters are significant except constant which has the P-value less than  $\alpha$ .

We therefore conclude:

1. Higher body weight (BMI) contributes to high blood pressure.
2. As age goes up it has an escalating effect on blood pressure.

## REFERENCES

- [1] Murray CJL, Lopez AD. Mortality by cause for eight regions of the world: Global Burden of Disease Study. *Lancet* 1997; **349**: 1269–1276. | Article | PubMed | ISI | ChemPort |
- [2] Tesfaye F, Nawi NG, Van Minh H, Byass P, Berhane Y, Bonita R *et al.* Association between body mass index and blood pressure across three countries in Africa and Asia. *J Human Hypertens* 2007; 21(1): 28–37 (this issue).
- [3] Mufunda J, Scott LJ, Chifamba J, Matenga J, Sparks B, Cooper R *et al.* Correlates of blood pressure in an urban Zimbabwean population and comparison to other populations of African origins. *J Hum Hypertens* 2000;**14**: 65–73. | Article | PubMed | ISI | ChemPort |
- [4] Mufunda J, Chatora R, Ndambakuwa Y, Nyarango P, Kosia A, Chifamba J *et al.* Emerging non communicable disease epidemic in Africa: preventive measures from the WHO Regional Office for Africa. *Ethn Dis* 2006; **16**: 521–526.
- [5] M'buyamba Kabangu JR, Fagard R, Stassen J, Lijnen P, Amery A. Correlates of blood pressure in rural and urban Zaire. *J Hypertens* 1987; **5**: 371–375. | Article | PubMed | ChemPort |
- [6] Mufunda J, Fink GD, Sparks HV. Blood pressure responses to dietary salt in rural and urban African men. *Ethn Dis* 1993; **3**: s46–s58. | PubMed |
- [7] Lebovitz HE. Insulin resistance a common link between type 2 diabetes and cardiovascular disease. *Diabetes Obes Metab* 2006; **8**: 237–249. | Article | PubMed | ChemPort |
- [8] Mufunda J, Sigola LB, Chifamba J, Vengesa PM. Hyperinsulinemia: possible cause of high blood pressure in unemployed urban black women. *High Blood Pressure* 1995; **4**: 137–140.
- [9] Baldinger B, Schwarz C, Jaggy C. Cardiovascular risk factors, BMI and mortality in a cohort of Swiss males (1976–2001). With high sum assured life insurance cover. *J Insur Med* 2006; **38**: 44–53. | PubMed |
- [10] Kaufman JS, Asuzu MC, Mufunda J, Forrester T, Wilks R, Luke A *et al.* Relationship between blood pressure and body mass index. *Hypertension* 1997; **30**: 1511–1516. | PubMed | ISI | ChemPort |
- [11] Thompson and J. Higgins. How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine*, 21:1559{1573, 2002.
- [12] R. Johnson and D. Wichern. *Applied Multivariate Statistical Analysis*. Prentice-Hall, Inc., 5 edition, 2002.
- [13] M. Woodward. *Epidemiology: Study Design and Data Analysis*. Chapman and Hall, 2 edition, 2005.
- [14] P. Royston and D. Altman. *Stata Technical Bullention*, chapter sg26: Using fractional polynomials to model curved regression relationships, pages 11{23. Stata Corporation, September 1994.
- [15] S. Weisberg. *Applied Linear Regression*. John Wiley and Sons Inc., 1 edition, 2005.
- [16] Irfan Ali Khan and Atiya Khanum. *Biostatistics for pharmacy*. Ukaaz publication Hyderabad, second revised edition, 2008.