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 from 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 week linear correlation i.e. $r$ is close to zero, a value near zero means there is a random, nonlinear relationship between the two variables.

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

**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 $R^2 = \frac{\text{Explained variance}}{\text{Total variance}}$

**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:-

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

In matrix notation regression line can be written as:

$$E(Y) = X \beta$$
Where $Y$ is $1 \times n$ vector of observation, $X$ is $n \times (p+1)$ matrix given by

$$X = [1 \ x_1 \ x_2 \ \ldots \ x_n]$$

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:

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y$</td>
<td>$X_1$</td>
<td>$X_2$</td>
<td>$X_n$</td>
</tr>
<tr>
<td>$y_1$</td>
<td>$X_{11}$</td>
<td>$X_{12}$</td>
<td>$X_{1n}$</td>
</tr>
<tr>
<td>$y_2$</td>
<td>$X_{21}$</td>
<td>$X_{22}$</td>
<td>$X_{2n}$</td>
</tr>
<tr>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
</tr>
<tr>
<td>$y_n$</td>
<td>$X_{n1}$</td>
<td>$X_{n2}$</td>
<td>$X_{mn}$</td>
</tr>
</tbody>
</table>

Hence $y_i$ denotes the $i^{th}$ observation and $X_i$ represent the $i^{th}$ level of independent variable $x_i$. The model equation for the above data can be written as follows:

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

Method of least square is used to estimate the $\beta$’s. The method of least square chooses $\beta$’s in the above equation. So that the sum square $e_i$ are minimized. The function of the least square is

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

The function $L$ is to be minimized with respect to $\beta_0, \beta_1, \beta_2, \ldots, \beta_n$. The least square estimate must therefore be satisfy

$$\frac{\partial L}{\partial \beta_0} \{\beta_0, \beta_1, \beta_2, \ldots, \beta_n\} = -2\sum_{i=1}^{n} (y_i - \beta_0 - \sum_j \beta_j x_{ij}) = 0$$

$$\frac{\partial L}{\partial \beta_j} \{\beta_0, \beta_1, \beta_2, \ldots, \beta_n\} = -2\sum_{i=1}^{n} (y_i - \beta_0 - \sum_j \beta_j x_{ij}) x_{ij} = 0$$

By simplifying the above equation, we have:

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

$$\beta_0 \sum_{i=1}^{n} x_{i1} + \beta_k \sum_{i=1}^{n} x_{ik} + \beta_1 \sum_{i=1}^{n} x_{i1}x_{i2} + \ldots + \beta_k \sum_{i=1}^{n} x_{i1}x_{ik} = \sum_{i=1}^{n} x_{i1}y_i$$

$$\beta_0 \sum_{i=1}^{n} x_{ik} + \beta_k \sum_{2=1}^{n} x_{ik} + \beta_1 \sum_{2=1}^{n} x_{ik}x_{i2} + \ldots + \sum_{1=i}^{n} x_{ih}^2 = \sum_{2=1}^{n} x_{ik}^2$$

These equation are $p=k+1$ equation for each regression coefficient solution of these equation be the least square estimation of $\beta_0, \ldots, \beta_k$. The above equation are known as normal equations.

The model can be written in a matrix notation as follows:

$$y = X\beta + e,$$ where

$$1 \ x_{i1} \ x_{i2} \ x_{ik}$$

$$y = (y_1, y_2, \ldots, y_n)',$$ $X= 1 \ x_{21} \ x_{22} \ x_{2k}$
\[ \beta = [\beta_0, \beta_1, ..., \beta_k]' \]

\[ e = (e_1, e_2, ..., e_n)' \]

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 significant 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] = \[ b'X'y - \frac{\left( \sum_{i=1}^{n} y_i \right)^2}{n} \]
- Sum of squares due to residuals = [SSE] = \[ y'y - b'X'y \]
- Total sum of squares = \[ y'y - \frac{\left( \sum_{i=1}^{n} y_i \right)^2}{n} \]

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 \( x_1, ..., x_j \) we set up the following hypothesis:

\[ H_0: \beta_1 = \beta_2 = ... = \beta_k = 0 \]

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

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

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Degree of Freedom (d.f)</th>
<th>Sum squares (SS)</th>
<th>Mean squares (MS)</th>
<th>F-statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>k</td>
<td>SS_R</td>
<td>MS_R</td>
<td>MS_R / MS_E</td>
</tr>
<tr>
<td>Error</td>
<td>n-k-1</td>
<td>SS_E</td>
<td>MS_E</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>n-1</td>
<td>SS_y</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The coefficient of determination $R^2$ is given by

$$R^2 = \frac{SS_R}{SS_y} = 1 - \frac{SS_E}{SS_y}$$

$R^2$ measure the reduction in variability of $y$ obtained using regressors $x_1, \ldots, x_k$ 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:

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

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

<table>
<thead>
<tr>
<th></th>
<th>Age</th>
<th>BMI</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMI</td>
<td>0.498341</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Blood Pressure</td>
<td>0.598482</td>
<td>0.736728</td>
</tr>
</tbody>
</table>

Cell Contents: Pearson correlation
P-Value

Multiple Regression: Blood Pressure versus Age, BMI

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Adj SS</th>
<th>Adj MS</th>
<th>F-Value</th>
<th>P-Value</th>
</tr>
</thead>
</table>

Regression

<table>
<thead>
<tr>
<th>Term</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T-Value</th>
<th>P-Value</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-1.20</td>
<td>12.32</td>
<td>-0.10</td>
<td>0.9224</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.26729</td>
<td>0.06319</td>
<td>4.23</td>
<td>&lt;0.0001</td>
<td>1.33</td>
</tr>
<tr>
<td>BMI</td>
<td>4.1434</td>
<td>0.5168</td>
<td>8.02</td>
<td>&lt;0.0001</td>
<td>1.33</td>
</tr>
</tbody>
</table>

Regression Equation

Blood Pressure = -1.20 + 0.26729 Age + 4.1434 BMI

Fits and Diagnostics for Unusual Observations

<table>
<thead>
<tr>
<th>Obs</th>
<th>Blood Pressure</th>
<th>Fit</th>
<th>Resid</th>
<th>Std Resid</th>
</tr>
</thead>
<tbody>
<tr>
<td>41</td>
<td>163</td>
<td>138.469</td>
<td>24.5312</td>
<td>2.72</td>
</tr>
<tr>
<td>58</td>
<td>149.5</td>
<td>129.247</td>
<td>20.2532</td>
<td>2.23</td>
</tr>
<tr>
<td>61</td>
<td>158.5</td>
<td>153.305</td>
<td>5.1952</td>
<td>0.59</td>
</tr>
<tr>
<td>75</td>
<td>148</td>
<td>113.074</td>
<td>34.9256</td>
<td>3.83</td>
</tr>
<tr>
<td>76</td>
<td>116.5</td>
<td>96.902</td>
<td>19.5981</td>
<td>2.19</td>
</tr>
<tr>
<td>79</td>
<td>111.5</td>
<td>140.608</td>
<td>-29.1080</td>
<td>-3.31</td>
</tr>
<tr>
<td>82</td>
<td>114.5</td>
<td>133.658</td>
<td>-19.1576</td>
<td>-2.12</td>
</tr>
</tbody>
</table>

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