IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Vol. 04, Issue 10 (October. 2014), ||V3|| PP 33-38

On Analysis Of Essential Causes Of Hypertension: A Statistical Approach

Adeniji, J. O., Bello, K. M., Alfa, M. S., Akoh, D. and Hassan, A. S.
Department Of Mathematics And Statistics The Federal Polytechnic, Bida.

Abstract: - The paper focuses on the identification of the most contributory causes of hypertension. The data used is a secondary data collected from the records of the University College Hospital (UCH), Ibadan. Principal Component Analysis is applied to the data on essential causes of hypertension. The nine research variables under consideration are Genetics, Smoking, Stress, Old age, Obesity, Excess Salt, Excess Alcohol, Kidney Diseases and Lack of Exercise which were observed over twenty years. The results of the Analysis strongly revealed that only four (4) variables, obesity, Smoking, Excess Salt and Excess Alcohol accounted for 76.2%, 8.6%, 5.7% and 4.4% respectively (a cumulative of 94.8%) of the total variance. This clearly shows that the most contributory factor to causes of hypertension is Obesity which accounted for 76.2% of the total variance.

Key-words: *Principal Component Analysis (PCA), Hypertension, Obesity, Correlation Matrix, Contributory Causes, Eigenvalue.*

I. INTRODUCTION

Generally in all nations of the world, the essential purpose of all governmental bodies or functions is to promote the health condition of the entire citizen. It is obvious that a healthy nation is a wealthy nation. Since the human resources of a country accounts to a large extent its productivity, it therefore becomes imperative for its standard of health to be adequately addressed.

Certain factors are known to be hampering and militating against the attainment of sustainable health standard in a country. These include diseases, improper health policies among others. There are some diseases that have been ravaging mankind over the past 3 to 4 decades in which Hypertension is one of them. Statistics have revealed a disturbing trend in this direction.

According to Mulatero and Bertello (2009) hypertension (HTN) or High Blood Pressure (HBP) is a chronic medical condition in which the blood pressure in the arteries is elevated. It is classified as either primary (essential) or secondary. About 90-95% or the cases are termed “primary Hypertension” which refers to high blood pressure for which no medical cure can be found. The remaining 5-10% of the cases are termed secondary hypertensions which are caused by another conditions that affect the kidneys, arteries, heart or endocrine system.

Persistence hypertension is one of the risk factors for stroke, heart attack, heart failure and arterial aneurysm and is a leading cause of chronic kidney failure. Moderate elevation of arterial blood pressure leads to short life. Both dietary and life style changes as well as medicines can improve blood pressure control and disease risk of associated health complications (Sacks and Svetkey, 2008).

II. TYPES OF HYPERTENSION

Primary (Essential) Hypertension

According to Ucheyama (2008) and Kite (2008) primary hypertension is the most prevalent hypertension type, affecting 90-95% of hypertension patients. This type of hypertension is a risk factor for hardening of the arteries (atherosclerosis). It also predisposes individuals to heart disease and obstruction of death in industrialized countries and increases the risk of stroke and heart failure.

Secondary hypertension

Secondary hypertension results from an indefinable causes. This type is important to recognize since it is treated differently from the Essential hypertension by treating the underlying causes of the elevated blood pressure. Many conditions cause hypertension; some are common and well recognized secondary causes such as causing’s syndrome which is a condition where the adrenal glands over produce the hormone cortical (Ucheyama, 2008, Kite, 2008).

III. CAUSES OF HIGH BLOOD PRESSURE

According to Segnella (2009) essential (primary) hypertension is the most prevalent hypertension type. Although no direct cause has identified itself, there are many factors such as sedentary lifestyle, stress, visceral obesity, potassium deficiency (hypokalemia), obesity (more than 85% of cases occur in those with a body mass index greater than 25), salt (sodium) sensitivity, alcohol intake and vitamin D deficiency that increase the risk of developing hypertension. Risks also increases with aging, some inherited genetic mutation and having a family history of hypertension. An elevation of rennin, an enzyme secreted by the kidney is another risk factor as in synthetic nervous system over activity. Insulin resistance which is a component of syndrome X on the metabolic syndrome is also thought to contribute to hypertension.

Consuming food that contains high fructose, corn syrup may increase one's risk of developing hypertension (Piller and Davis, 2007). According to Mulatero and Bertello (2009) recent studies have equally implicated low birth weight as a risk factor for adult essential hypertension.

Statistics have it that between 2000 and 2008, there has been about forty percent (40%) increase in the number of people having high blood pressure. It further shows that in 2005, sixty percent (60%) of people suffering from high blood pressure were suffering from kidney failure and heart disease and that high blood pressure was identified as the remote cause of the disease that later led to the death of a greater percentage of the patients (Mulatero, 2009).

IV. PRINCIPAL COMPONENT ANALYSIS

According to Jollife (2002) Principal Component Analysis (PCA) is probably the oldest and the best known of the techniques of multivariate analysis. The central idea of PCA is to reduce the dimensionality of data set in which there are large numbers of interrelated (correlated) variables while retaining as much as possible of the variation present in the data set. This transformation is possible by transforming to a new set of variables, the principal components which are uncorrelated and which are ordered so that the first "few" retain most of the variation present in "all" of the original variables.

PCA is a standard tool in modern data analysis in diverse field because it is a simple parametric method for extracting relevant information from confusing and complex data set. With minimal efforts, PCA produces a roadmap for how to reduce a complex data set to a lower dimension to reveal the sometimes hidden, simplified structures that often underline it (Jonathan, 2009).

According to Oyebande (2011) PCA generally entails transforming a set of data into a fewer number of new set of data. It is one of the methods in handling the problem of multi-colinearity in which independent variables are highly correlated with each other. Also in multiple regression where we have a large number of independent variables relating to the sample size, the PCA is used to reduce the dimensionality of the original data to a fewer number of independent variables that could be used for the regression.

Edokpayi et al (2010) cited in Oyebande (2011) used the PCA to study the data on soil properties in the oil palm belt in Nigeria in which 12 variables (soil properties) from 58 locations were reduced to 4 components-soil PH, Organic matter, H^+ (Hydrogen Ions) and the Fe (Iron).

In this study therefore, the technique is applied to the data on essential causes of hypertension in order determine which among the essential causes of hypertension retained form of variability in the smallest possible causes. The data used is a secondary data collected from University College Hospital (UCH) Ibadan Record Department. The data gives 20 years of cases of hypertension incidence within the period of 1992 to 2011 with nine possible essential causes of hypertension.

The basic idea of the method is to describe the variation of a set of multivariate data in terms of a set of uncorrelated variables each of which is a particular linear combination of the original variables. The usual objective of this type of analysis is to see whether the first few components account for most of the variation in the data. It is argue that if they do, then they can be used

The first principal component of the observations is that linear combination, y_1 of the original variables,

$$y_1 = a_{11}x_1 + a_{12}x_2 + \ldots + a_{1p}x_p$$

$$i.e. \quad y_1 = a_1'X \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

whose sample variance is greatest for all coefficients, $a_{11},...,a_{1p}$ (which we may write as the vector a_1). Since the variance of y_1 could be increased without limit simply by increasing the elements of a_1, a restriction must be placed on these coefficients; a sensible constraint is to require that the sum of squares of the coefficients, $a_1'a_1$, should be set at a value of unity.

If, for example, 80 percent of the variation in an investigation involving six variables could be accounted for by a simple weighted average of the variable values, then almost all the variation could be expressed along a single continuum rather than in six-dimensional space. This would provide a highly parsimonious summary of the data that might be useful in later analysis.

The second principal component, y_2, is that linear combination

$$y_2 = a_{21}x_1 + a_{22}x_2 + \ldots + a_{2p}x_p$$

$$i.e. \quad y_2 = a_2' X \ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

which has the greatest variance subject to the two conditions,

$$a_2' a_2 = 1 \quad (for\ the\ reason\ \ indicated\ \ previously)$$

$$and\ \ a_2' a_1 = 0\ (so\ that\ y_1\ \ and\ y_2\ \ are\ \ uncorrelated)$$

Similarly the j^{th} *principal component is that linear combination*

$$y_j = a_j' X \ldots\ldots\ldots\ \ldots\ldots\ldots\ \ldots\ldots\ldots (3)$$

which has greatest variance subject to?

$$a_j' a_j = 1$$

$$a_j' a_i = 0 \quad (i < j).$$

To find the coefficients defining the first principal component we need to choose the elements of a_1 so as to maximize the variance of y_1 subject to the constraint, $a_1' a_1 = 1$.

The variance of y_1 is given by

$$\mathrm{var}(y_1) = \mathrm{var}(a_1' X)$$

$$= a_1' S a_1 \ldots\ldots\ldots\ldots\ldots (4)$$

where S is the variance –covariance matrix of the original variables.

The standard procedure for maximizing a function of several variables, subject to one or more constraints, is the method of Lagrange multiplier. Applying this technique to maximise the variance of y_1, as given by $\mathrm{var}(y_1) = a_1' S a_1$, subject to the constraint, $a_1' a_1 = 1$, leads to the solution that a_1 is the eigenvector of S corresponding to the largest eigenvalue.

To determine the second component, the Lagrange multiplier technique is again used to maximize the variance of y_2 i.e.

$$\mathrm{var}(y_2) = a_2' S a_2,$$

Subject to the two constraints $a_2' a_2 = 1\ \ and\ \ a_2' a_1 = 0.$

This leads to the solution that a_2 is the eigenvector of S corresponding to its second largest eigenvalue.

Similarly the j^{th} principal component is defined by the eigenvector associated with the j^{th} largest eigenvalue. If the eigenvalues of S are $\lambda_1, \lambda_2, \ldots, \lambda_p$ then it is easy to show that by choosing $a_1' a_1 = 1$ the variance of the i^{th} principal component is given by λ_i.

For example y_1 has variance given by $a_1' S a_1$, now since a_1 is an eigenvector of S.

We know that $Sa_1 = \lambda_1 a_1 \ldots\ldots\ldots\ \ldots\ldots\ldots\ \ldots\ldots\ldots\ \ldots\ldots (5)$

Therefore (4) .i.e. $\mathrm{var}(y_1) = \mathrm{var}(a_1' X)\ = a_1' S a_1$ may be rewritten as

$$\mathrm{var}(y_1) = a_1' \lambda_1 a_1 = \lambda_1 a_1' a_1$$

$$= \lambda_1 \quad Since\ \ a_1' a_1 = 1$$

The total variance of the p principal components will equal the total variance of the original variables so that

$$\sum_{i=1}^{p} \lambda_i = trace(S) \ldots\ldots\ldots\ \ldots\ldots\ldots\ \ldots (6)$$

Consequently the j^{th} principal component accounts for a proportion

$$t = \frac{\lambda_i}{trace(S)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (7)$$

Of the total variation in the original data and the first, say p_i *components* $(p_i < p)$ account

for a proportion $T = \dfrac{\sum_{i=1}^{p} \lambda_i}{trace(S)}$ ……………………………(8)

of the total variation.

Although the derivation of principal components given above has been in terms of the eigenvalues and eigenvectors of the covariance matrix S, it is much more usual in practice to derive them from the corresponding quantities of the correlation matrix, R.

The reasons are not difficult to appreciate if we imagine a set of multivariate data where the variables $x_1, x_2, ..., x_p$, are of completely different types, for example length, temperature, blood pressures, anxiety ratings, e.t.c. In such a case the structure of the principal components derived from the covariance matrix will depend upon the essential arbitrary choice of units of measurement; additionally if there are large differences between the variances of

$x_1, x_2, ..., x_p$, these variables whose variances are largest will tend to dominate the first few principal components.

Extracting the components as the eigenvectors of R which is equivalent to calculating the principal components from the original variables after each has been standardised to realise, however, that the eigenvalues and eigenvectors of R will generally not be the same as those of S; indeed there is rarely any simple correspondence between the two and choosing to analyse R rather than S involves a definite but possibly arbitrary decision to make the variables "equally important".

V. PRINCIPAL COMPONENTS FROM THE CORRELATION MATRIX

Generally, extracting components from S rather than R remains closer to the spirit and intent of principal component analysis, especially if the components are to be used in further computations.

However, in some cases, the principal components will be more interpretable if R is used. For example, if the variances differ widely or if the measurement units are not commensurate, the components of S will be dominated by the variables with large variances. The other variable(s) will contribute very little. For a more balanced representation in such cases, components of R may be used.

VI. COMPARISONS OF PRINCIPAL COMPONENTS FROM R WITH THOSE FROM S

We now list some general comparisons of principal components from R with those from those from S.

(1) The percent of variance accounted for, by the components of R will differ from the percent from S.

(2) The coefficients of the principal components from R differ from those obtained from S.

(3) If we express the components from R in terms of the original variables, they still will not agree with the components from S.

(4) The principal components from R are scale invariant, because R itself is scale invariant.

(5) The components from a given matrix R are not unique to that R.

VII. DECIDING HOW MANY COMPONENTS TO RETAIN

In every application, a decision must be made on how many principal components should be retained in order to effectively summarize the data. The following guidelines have been proposed.

(1) Retain sufficient components to account for a specified percentage of the total variance, say 80%.

(2) Retain the components whose eigenvalues are greater than the average of the eigenvalues, $\dfrac{\sum_{i=1}^{p} \lambda_i}{p}$, For a correlation matrix, this average is 1.

(3) Use the scree graph, a scree graph is a plot of λ_i *versus* i, and look for a natural break between the large and the small eigenvalues.

(4) Test the significance of the larger component, that is, the components corresponding to the larger eigenvalues.

VIII. RESEARCH VARIABLES

The essential causes of hypertension are the variables under consideration in this paper are (1) Genetics (2) Smoking (3) Stress (4) Old Age (5) Obesity (6) Excess Salt (7) Excess Alcohol (8) Kidney Disease (9) Lack of Exercise.

IX. METHOD OF ANALYSIS

The method of analysis utilized in this work is Principal Component Analysis. Stata statistical software was used for the analysis and the results are presented below.

X. RESULTS OF PRINCIPAL COMPONENT ANALYSIS FROM CORRELATION MATRIX

The results of the analysis are as shown below;

Eigenvalue	410.69	46.20	30.63	23.53	11.06	8.19
Proprotion	0.762	0.086	0.057	0.044	0.021	0.015
Cumulative	0.762	0.847	0.904	0.948	0.968	0.983

Eigenvalue	5.29	2.58	1.08
Proprotion	0.010	0.005	0.002
Cumulative	0.993	0.998	1.000

Variables	PC1	PC2	PC3	PC4	PC5	PC6
Obesity	-0.729	0.101	-0.317	0.120	-0.312	0.172
Smoking	-0.334	-0.261	-0.004	0.020	0.046	0.114
Excess Salt	-0.076	-0.305	-0.408	0.107	-0.404	-0.446
Excess Alcohol	-0.161	-0.206	-0.510	-0.175	0.762	-0.036
Stress	-0.364	-0.104	0.502	0.611	0.338	-0.152
Old age	-0.306	0.715	0.095	-0.320	0.097	-0.455
Genetics	-0.253	-0.061	0.237	-0.406	-0.076	0.591
Kidney Disease	-0.154	-0.427	0.353	-0.302	-0.142	-0.369
Lack of Exercise	-0.109	-0.279	0.176	-0.460	0.079	-0.204

Variables	PC7	PC8	PC9
Obesity	0.335	0,092	-0.309
Smoking	0.113	0.101	0.884
Excess Salt	-0.503	-0.323	0.057
Excess Alcohol	-0.149	0.120	-0.149
Stress	-0.173	-0.215	-0.131
Old age	-0.186	-0.002	0.181
Genetics	-0.566	-0.171	-0.083
Kidney Disease	-0.047	0.621	-0.185
Lack of Exercise	0.461	-0.634	-0.076

XI. DISCUSSION OF RESULTS

From the table, the first eigenvalue λ_1 is 410.69 and its proportion of variance is 0.762. it indicates that the factor having the first eigenvalue (obesity) accounted for 76.2% of the total variance. The second eigenvalue λ_2 is 46.20 and its proportion of variance is 0.086, the factor having the second value is smoking which accounted for 8.6% of the total variance. The third eigenvalue λ_3 is 30.63 and its proportion of variance is 0.057, the factor having the third value is Excess salt which accounted for 5.7% of the total variance. The fourth eigenvalue λ_4 is 23.53 and its proportion of variance is 0.044, the factor having the fourth value is Excess Alcohol which accounted for 4.4% of the total variance. The fifth eigenvalues to the ninth $(\lambda_5 \, to \, \lambda_9)$ were discarded since all their proportion of variance accounted for are less than 10%. The cumulative proportion of the first four eigenvalues $(\lambda_1 \, to \, \lambda_4)$ is 94.8% which is a sufficient percentage component accounted for.

XII. CONCLUSSION

The analysis shows that among the nine essential causes of hypertension considered only four; obesity, smoking, Excess Salt and Excess Alcohol constitute the major causes of hypertension. The contributions of others are also there but not a serious treat as the four major ones identified by the PCA.

XIII. RECOMMENDATION

The seriousness of hypertension is not just that it can lead to more serious illness or complications but raises the risk of stroke, kidney failure, heart disease, and heart attack. The matter is made worse with the existence of too much weight or fat in the body as this tends to make the conditions severe. It is on the basis of these that the following recommendations are made.

(1) It is the duty of all stakeholders in the health sectors to keep people informed about the implications of the causes of hypertension by providing them with accurate, timely and up to date information regarding this all important disease that is threatening human existence.

(2) It is also the duty of health personnel to warn and educate adults who are more prone to high blood pressure about the dangers in management of high blood pressure.

(3) More awareness is expected to be given on the feeding habit of the people, smoking habit, health care habit and general way and pattern of life as well as to avoid poor combination of the classes of food.

(4) Campaign on “stay healthy and fit” and “health is wealth” should be intensified and sustained.

REFERENCES

[1] Jollife, I.T. (2002) Principal Component Analysis. (2nd Edition) Springer series in statistics.Jonathan, S. (2009) A tutorial on Principal Component Analysis. Centre for Neural science, New York University, New York NY10003-6603 and system Neurobiology Laboratory, Salk Institute for Biological Studies La Jolla, LA 92037

[2] Kite, S. (2008) Trends in the prevalent Awareness, Treatment and Control of Hypertension in Adult US population. Data from the Health Examination Surveys 1996-1998, Hypertension(1): 60-69.

[3] Mulatero, P. and Bertello, C. (2009) Differential Diagnosis of primary Aldosteronism Subtypes. Current hypertension reports. Web MD http:www.webmd.com/content/article/73/88927.htm.

[4] Mulatero, P. (2009) Diagnosis of Pre-hypertension: Early stage High blood pressure. Web MD http:www.webmd.com/content/article/73/88927.htm.

[5] Oyebande, B. L. (2011) An application of Principal Component Analysis: Regression Approach to predict storm runoff, *Nigeria geographical Journal*, 8(2), 65-73.

[6] Piller, L. and Davis, B. (2007) The 2007 Canadian Hypertension Education Programme Recommendation for the management of hypertension: Part 1-Blood Pressure Measurement, Diagnosis and Assessment of Risk. *The Canadian Journal of Cardiology*, 51, 125-128.

[7] Sacks, F. and Svetkey L. (2008) Joint National Committee on Prevention, Detection, Evaluation and Treatment of High Blood Pressure. Hypertension (6): 1206-52.

[8] Sagnella, G. (2009) The renal Epithelial Sodium Channel, Genetic heterogeneity and Implication for the treatment of High Blood Pressure. *Current Pharmaceutical Design* (14): 2221-2234.

[9] Ucheyama, M. (2008) The effect of Biofeedback for the treatment and control of Essential Hypertension: A Systematic Review. *Health Technology (3)*: 57-66.