Available online at http://ajol.info/index.php/njbas/index ISSN 0794-5698
Nigerian Journal of Basic and Applied Science (2010), 18(1):120-125

The Use of Hotelling T^2 Statistic for Tracking the Normalcy of Systolic and Diastolic Blood Pressure

A. Usman[1], I.O Oshungade[2] and H. G. Dikko[3]

[1]Department of Mathematics, Statistics and Computer Science, Kaduna Polytechnic.
[2]Department of Statistics, University of Ilorin.
[3]Department of Mathematics, Ahmadu Bello University, Zaria.

ABSTRACT: This paper has introduced a direct statistical inference technique for checking the normalcy of systolic blood pressure (SBP) and diastolic blood pressure (DBP) as a simplified technique for cardio-vascular prevalence survey. The use of multivariate test of hypothesis was particularly used to track the normalcy of both the SBP and DBP so as to isolate cases of hypertension and hypotension. Multivariate test of hypothesis was to monitor the SBP and DBP. Samples of SBP and DBP from 12 diabetic patients were collected using systematic sampling. The Hotelling T^2 statistic was computed for each patient and appropriate conclusions were drawn. From the analysis, ten of the twelve diabetic have some level of abnormality in their blood pressure.

INTRODUCTION

Blood pressure is the pressure of blood against the walls of the main arteries. Pressure is highest during ***systole*** when the ventricles are contracting (***systolic blood pressure, SBP***) and lowest during diastole, when the ventricles are relaxing and refilling (***diastolic blood pressure, DBP***). Blood pressure is measured in millimeters of mercury (mmHg) by means of a sphygmomanometer at the brachial artery of the arm, where the pressure is most similar to that of blood leaving the heart. The normal range varies with age, but a young adult would be expected to have a SBP of around 120mmHg and a DBP of 80mmHg. These are recorded as 120/80 (Rao, 2007).

Individual variations are common. Muscular exertion and emotional factors such as fear, stress and excitement all raise SBP which could lead to ***hypertension***. SBP is normally at its lowest during sleep. Severe shock may lead to an abnormally low blood pressure and possible circulatory failure, ***hypotension***. Where necessary, blood pressure is adjusted to its normal level by the sympathetic nervous system and hormonal controls (Preben & Jogen, 2003)

In this paper we used a direct statistical inference technique for checking the normalcy of systolic blood pressure (SBP) and diastolic blood pressure (DBP). The results herein could also serve as a foundation as well as a simplified technique for cardio-vascular prevalence survey. The use of multivariate test of hypothesis was particularly used to track the normalcy of both the SBP and DBP so as to isolate cases of hypertension and hypotension.

According to Rencher (2002), the motivation for testing *p* variables multivariately rather than (or in addition to) univariately, as, for example, in hypotheses about $\mu_1, \mu_2, \ldots, \mu_p$ in $\boldsymbol{\mu}$ lies in at least four arguments for a multivariate approach to hypothesis testing:

1. The use of *p* univariate tests inflates the Type I error rate, α, whereas the multivariate test preserves the exact α level.
2. The univariate tests completely ignore the pair-wise correlations among the variables, whereas the multivariate tests make direct use of the pair-wise correlations.
3. The multivariate test is more powerful in many cases. The *power* of the test has been the probability of rejecting H_0 when it is indeed false.
4. Many multivariate tests involving means have as a by-product the construction of a linear combination of variables that reveals more about how the variables unite to reject the hypothesis.

The Hotelling T^2 test is a generalization of the Student's t-test for multivariate data. The distribution is indexed by two parameters, the dimension of *p* and the degrees of freedom

$v = n-1$. We reject H_0 if $T^2 > T^2_{\alpha,p,n-1}$ and accept H_0 otherwise (Anderson, 1971)

The T^2 statistic can be viewed as the sample standardized distance between the observed sample mean vector and the hypothetical mean vector.

In our case, $p = 2$, with 2 means, 2 variances and ${}^2C_2 = 1$ covariances. Let X_1 and X_2 denote SBP and DBP respectively. The variables X_1 and X_2 follow the bivariate normal distribution. Let μ_1 and μ_2 be the normal values of the SBP and DBP and the variances of X_1 and X_2 be estimated by the sample variances S_1^2 and S_2^2 respectively. Also, let the sample covariance is denoted by S_{12}. If $\overline{X}_1$ and $\overline{X}_2$ are the sample means of the SBP and DBP computed from a sample of sizes n , from an individual, then the statistic is distributed according to the Hotelling T^2 distribution with $p = 2$ and $n-1$ degrees of freedom is given by:

$$T^2 = \frac{n}{S_1^2 S_2^2 - S_{12}^2}\left[S_1^2(\overline{X}_1 - \mu_1)^2 + S_2^2(\overline{X}_2 - \mu_2)^2 - 2S_{12}(\overline{X}_1 - \mu_1)(\overline{X}_2 - \mu_2)\right] \quad (1)$$

If the statistic $T^2 > T^2_{\alpha,(2,n-1)}$, then either of SBP and DBP of that individual is abnormal; where $T^2_{\alpha,(2,n-1)}$ is the upper $\alpha\%$ point of the Hotelling T^2 distribution with $p = 2$ and $n-1$ degrees of freedom.

Consider the case where $\overline{X}_1$ and $\overline{X}_2$ are independent so that $S_{12} = 0$; geometrically, the expression will then define an ellipse central at (μ_1, μ_2) with principal axis parallel to $\overline{X}_1$ and $\overline{X}_2$ axes. The test statistic is given by:

$$T^2 = n(\overline{X} - \mu_0)' S^{-1}(\overline{X} - \mu_0) \quad (2)$$

Where;

$\overline{X}' = (\overline{X}_1 \ \overline{X}_2)$, $\mu' = (\mu_1 \ \mu_2)$ and S is the sample variance-covariance matrix.

Both the SBP and DBP have an upper control limit of $T^2_{\alpha,p,n-1}$; we may obtain the percentage points of the T^2 as $T^2_{\alpha,p,n-1} = \frac{p(n-1)}{n-p} F_{\alpha,p,n-p}$.

The procedure is to obtain $\overline{X}$ *and* S from the measurements of an individual's blood pressure.

This paper has used inferential multivariate statistical methods - Hotelling T^2 statistic to track the normalcy of both the SBP and DBP so as to identify existing or potential cases of hypertension and hypotension. More specifically, the paper has achieved the following objectives:

1. Applied the Hotelling T^2 statistic to track cases of hypertension and hypotension among Nigerians.
2. Provided a unified approach in survey sampling methods of cardio-vascular diseases where there are no complete medical records for the patients.
3. Used sample data to infer about the normalcy of both the SBP and DBP of an individual.

The Hotelling T^2 quality statistic used in this paper will particularly produce an efficient, alternative procedure for checking both the SBP and DBP of individuals from sample data. The method used herein can also serve as a sample survey method for patients with hypertension and hypotension.

METHODOLOGY

Assume sampling from the multivariate normal density, $N_p(\mu, \Sigma)$ where Σ is positive definite.

The pair of hypothesis whose conclusion will signify the presence of either hypertension or hypotension is given as:

$H_0 : \mu = \mu_0$

$H_1 : \mu \neq \mu_0$

I f Σ is known and $X_1, X_2, ..., X_n$ is a random sample of size n from $N_p(\mu, \Sigma)$ for $n > p$. Then;

- $\overline{X}$ is unbiased for μ
- $\sqrt{n}(\overline{X} - \mu) \sim N_p(0, \Sigma)$

- $Y = n(\overline{X} - \mu)'\Sigma^{-1}(\overline{X} - \mu) \sim \chi_p^2$ under H_0.

The decision rule then is to reject H_0 if $Y \geq \chi_{p,\alpha}^2$ (Normal, 1999)

When Σ is unknown, under H_0, $\hat{\Sigma}_0 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_0)(X_i - \mu_0)'$

The likelihood under H_0

$$\hat{L}_0 = L(\mu_0, \Sigma_0) = (2\pi)^{-\frac{np}{2}} |\Sigma_0|^{-\frac{n}{2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(X_i - \mu_0)'\Sigma_0^{-1}(X_i - \mu_0)\right]$$

$$\Rightarrow \hat{L}_0 = L(\mu_0, \Sigma_0) = (2\pi)^{-\frac{np}{2}} |\Sigma_0|^{-\frac{n}{2}} \exp\left[-\frac{1}{2}\Sigma_0^{-1}\sum_{i=1}^{n}(X_i - \mu_0)(X_i - \mu_0)'\right]$$

$$\Rightarrow \hat{L}_0 = L(\mu_0, \Sigma_0) = (2\pi)^{-\frac{np}{2}} |\Sigma_0|^{-\frac{n}{2}} \exp\left[-\frac{1}{2}\Sigma_0^{-1}n\Sigma_0\right] = (2\pi)^{-\frac{np}{2}} |\Sigma_0|^{-\frac{n}{2}} \exp\left[-\frac{n}{2}\right] \quad (3)$$

The likelihood under H_1

Here $\hat{\mu} = \overline{X}$

$$\hat{\Sigma} = S = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})(X_i - \overline{X})'$$

$$\hat{L}_1 = L(\overline{X}, S) = (2\pi)^{-\frac{np}{2}} |S|^{-\frac{n}{2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(X_i - \overline{X})'S^{-1}(X_i - \overline{X})\right]$$

$$\Rightarrow \hat{L}_1 = L(\overline{X}, S) = (2\pi)^{-\frac{np}{2}} |S|^{-\frac{n}{2}} \exp\left[-\frac{1}{2}S^{-1}\sum_{i=1}^{n}(X_i - \overline{X})(X_i - \overline{X})'\right]$$

$$\Rightarrow \hat{L}_1 = L(\overline{X}, S) = (2\pi)^{-\frac{np}{2}} |S|^{-\frac{n}{2}} \exp\left[-\frac{1}{2}S^{-1}nS\right] = (2\pi)^{-\frac{np}{2}} |S|^{-\frac{n}{2}} \exp\left[-\frac{n}{2}\right] \quad (4)$$

Hence, the likelihood ratio, also called the Wilk's lambda in this case, is given by equation (5) below:

$$\lambda = \frac{\hat{L}_1}{\hat{L}_0} = \left[\frac{|\Sigma_0|}{|S|}\right]^{\frac{n}{2}} \quad (5)$$

Decision rule

For a test of size α of H_0 *Vs* H_1, reject H_0 if $\lambda^{\frac{2}{n}} \geq K_\alpha$. Noting that Σ_0 can be split as follows:

$$\hat{\Sigma}_0 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_0)(X_i - \mu_0)' = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X} + \overline{X} - \mu_0)(X_i - \overline{X} + \overline{X} - \mu_0)'$$

$$\Rightarrow \hat{\Sigma}_0 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})(X_i - \overline{X})' + n(\overline{X} - \mu_0)(\overline{X} - \mu_0)'$$

$$\Rightarrow \hat{\Sigma}_0 = S + n(\overline{X} - \mu_0)(\overline{X} - \mu_0)'$$

Hence, the Wilk's lambda on simplification now gives equation (6) below:

$$\lambda^{\frac{2}{n}} = \left[\frac{\left|S + n(\overline{X} - \mu_0)(\overline{X} - \mu_0)'\right|}{|S|}\right] = 1 + n(\overline{X} - \mu_0)' S^{-1} (\overline{X} - \mu_0) \qquad (6)$$

Using the Wilk's lambda, we are to reject H_0 if equation (6) above satisfied the following inequalities:

$1 + n(\overline{X} - \mu_0) S^{-1} (\overline{X} - \mu_0)' \geq K_\alpha$

or $n(\overline{X} - \mu_0) S^{-1} (\overline{X} - \mu_0)' \geq K'_\alpha$

We now have the multivariate quality control model in form of the Hotelling T^2 statistic as follows:

$T^2 = n(\overline{X} - \mu_0)' S^{-1} (\overline{X} - \mu_0) \sim T^2_{\alpha,p,n-1}$

In this particular case for controlling both the SBP and DBP, the test statistic is more properly written as:

$$T^2 = n(\overline{X} - \mu_0)' S^{-1} (\overline{X} - \mu_0) \qquad (7)$$

Where μ_0 is the vector of standard SBP and DBP; while $\overline{X}$ and S are the sample mean vector and variance-covariance matrix respectively.

Reject H_0 if

$T^2 = n(\overline{X} - \mu_0)' S^{-1} (\overline{X} - \mu_0) \geq T^2_{\alpha,p,n-1}$ or

if $T^2 \geq \frac{p(n-1)}{n-p} F_{\alpha,p,n-p}$.

ANALYSIS AND RESULTS

Samples of SBP and DBP from 12 diabetic patients were collected using the systematic sampling at every other day ($k = 2$) in Kaduna, Nigeria. Therefore, the data were collected in twenty days; with $n = 10$ for each patient, across the 12 patients. Both the SBP and DBP are two health indices that are to be simultaneously controlled and be kept normal. The issue here is to use the Hotelling T^2 statistic and the data collected on patient's SBP and DBP to infer about their blood pressure, we define the following parameters:

$\mu_1 = 120mmHg$ (Standard SBP)

$\mu_2 = 80mmHg$ (Standard DBP)

$n = 10$ (SBP and DBP sample size for each patient)

$p = 2$ (Number of variables i.e. SBP and DBP)

The mean vector and the sample dispersion matrix for the first patient, in Table 1are given by:

$\overline{X}' = \begin{pmatrix} 124 & 82 \end{pmatrix}$

$S = \begin{pmatrix} 6.85 & 2.00 \\ 2.00 & 1.56 \end{pmatrix}$

Therefore, we derive the following statistics:

$(\overline{X} - \mu_0)' = \begin{pmatrix} 4 & 2 \end{pmatrix}$

$$T^2 = n(\overline{X} - \mu_0)' S^{-1} (\overline{X} - \mu_0) = 10\begin{pmatrix} 4 & 2 \end{pmatrix}\begin{pmatrix} 6.85 & 2.00 \\ 2.00 & 1.56 \end{pmatrix}^{-1}\begin{pmatrix} 4 \\ 2 \end{pmatrix} = 30.45$$

Hypothesis:

$H_0 : \mu' = \begin{pmatrix} 120 & 80 \end{pmatrix}$ (The patient's BP is normal)

$H_1 : \mu' \neq \begin{pmatrix} 120 & 80 \end{pmatrix}$ (The patient's BP is not normal)

Level of significance:

$\alpha = 0.05$

Test statistic:

$T^2 = n(\overline{X} - \mu_0)' S^{-1} (\overline{X} - \mu_0)$

Decision criterion:

Reject H_0 if $T^2 = n(\bar{X} - \bar{\bar{X}})S^{-1}(\bar{X} - \bar{\bar{X}})' \geq T^2_{\alpha,p,n-1}$ or if $T^2 \geq \frac{p(n-1)}{n-p} F_{\alpha,p,n-p} = 10.035$

Computations:

For each of the 12 patients, the values of the sample mean vector, dispersion matrix and T^2 were obtained from sample SBP and DBP of size $n = 10$ and the results are summarized in Table 1:

Table 1: Computational Summary and worksheet for T_k^2

Sample No	Means		Variance - Covariance			T_k^2
	SBP ($\bar{X}_{1k}$)	DBP ($\bar{X}_{2k}$)	S_{1k}^2	S_{2k}^2	S_{12k}	
1	124	82	6.85	1.56	2.00	30.45*
2	147	83	20.25	36.00	10.50	39.96*
3	121	81	0.41	0.16	-0.06	4.50
4	129	82	2.36	1.44	1.52	44.12*
5	127	84	37.01	8.84	11.62	19.50*
6	126	83	44.36	9.41	17.34	9.74
7	127	84	34.24	10.16	15.42	16.54*
8	122	82	9.76	3.96	5.84	17.76*
9	125	84	21.41	9.00	10.10	18.03*
10	127	85	18.24	7.25	7.90	36.98*
11	126	84	20.96	9.00	10.30	19.98*
12	125	83	18.29	5.01	1.77	26.75*

DISCUSSION

Since all the $T_k^2 < 10.035$, except patients number 3 and 6, as indicated with asterisk, we accept H_0 and conclude that ten of the twelve diabetic patients considered in the sample proved to have some level of abnormality in their blood pressure. Either or both of the SBP and DBP are not normal for those diabetic patients. From the result, it was established that the multivariate statistical inference method used herein can tract the status of the SBP and DBP of an individual.

From the results we can now conclude that ten of the twelve diabetic patients considered in the sample proved having some level of abnormality in their blood pressure. Either or both of the SBP and DBP are not normal for those diabetic patients. If samples size is large enough and the samples are randomly selected, this technique can serve the purpose of health survey of individuals with BP problem.

This approach is yet another improvement in the field of biostatistics, medical statistics and by extension sample survey methods. In addition, the technique can be extended to several related health characteristics simultaneously. It is highly recommended that this method be used in the area of biostatistics and medical statistics.

REFERENCES

Anderson, T.W. (1971). *An introduction to multivariate statistical analysis*: New York: John Wiley and Sons.

Bryan, F.M. (1992). *The design and analysis of research studies*. London: Cambridge University Press.

Jerrold, H.Z. (1999). *Biostatistical analysis* (4th. ed). New Delhi: Pearson.

Lindeman, H.R, Morenda, F.P, & Gold, R.Z. (1980). *Introduction to bivariate and multivariate analysis*. London: Scott, Foresman and Company.

Marcello, P. & Kimberlee, G. (2000). *Principles of biostatistics* (2nd ed). New York: Duxbury.

Neil, H.T. (1990). *Multivariate analysis with application in education and psychology* (2nd ed.). New York: John Wiley & Sons.

Norman, C. (1999). *Analyzing multivariate data* (2nd ed.). New York: Academic Press.

Preben, B. & Jorgen, G. (2003). *Statistics with application in biology and geology*. New York: Chapman and Hall.

Rao, K.V. (2007). *Biostatistics a manual of statistical methods for use in health, nutrition and anthropology* (2nd ed.). Jaypee Brothers, Medical Publishers Ltd.

Rencher, A.C. (2002). *Methods of multivariate analysis* (2nd ed.). New York: John Wiley & Sons.