Advances in Life Science and Technology
ISSN 2224-7181 (Paper) ISSN 2225-062X (Online)
Vol.24, 2014

Application of Wilks' Lambda and Hotelling's T^2 with MANOVA on Drug Addiction and Drug Abuse Data

C. E. Onwukwe
Department of Mathematics/Statistics and Comp. Science
University of Calabar, P.M.B 1115, Calabar, Cross River State, Nigeria

E. N. Ogbonna
Department of Mathematics/Statistics and Comp. Science
University of Calabar, P.M.B 1115, Calabar, Cross River State, Nigeria

Abam Ayeni (Correspondence Author)
Dept. of Maths/Stats. & Computer Science ,Federal University Lafia, Nigeria
E-mail: abamayeni@gmail.com

Abstract
In this study Wilks' Lamda and Hotelling's T^2 were employed as Multivariate Analysis of Variance (MANOVA) technique in determining the significance difference in the causative factor for drug addiction and abuse. The data were primary and were generated by a survey conducted on some patients selected from Neuro-Psychiatric Hospital Enugu-Nigeria. The patients were divided into two groups - those treated of substance abuse before and those treated for the first time. 89 patients were selected and nine factors for substance abuse were considered. They were rated and scored over ten by the patients. The result showed a significant difference in the factors specified for drug abuse. .

Keywords: Wilks' Lamda, Hotelling's T^2, drugs, substance abuse

1.1 Introduction

Drug abuse is one of the most disturbing problems in contemporary Africa and World at large. Recent studies in Nigeria showed that youths are found to constitute high risk group for substance abuse in our society. Obot (1989). Odejide and Olatawura (1977)

Drug is defined as any substance taken into the body which brings about a physiological change or modify one or more of the body's physical or mental function Allan and Vicky (1993) . Alcohol is a potent substance and number one drug problem among youths. A significant number of deaths from accidents, violent crimes, culticism etc have been traced to activities of drugs addicts Adelekan and Adeniran (1991) and Ijeoma (1997).

Wilks' Lamda was used to test the nine factors - peer group, stress, enhance performance, health, unemployment, pride, curiosity, happiness and poverty while Hotelling's T^2 was used to test groups – those treated of substance abuse before and those treated for the first time. Wilks' Lamda compared within sum of squares W and products matrix B to the total sum of squares and products matrix W+B. Thus the multivariate information in W and B about separation of mean vectors is channeled into a single scale on which we can determine if the separation of mean vectors is significant. Alvin (2002). Wilks' Lamda statistic was applied to test the Near Exact distribution for different scenarios independent of two sets of variables Grilo (2005).

Hotelling's T^2 was used to test the difference in mean vectors A_{1m} *and* A_{2m} to form to form D matrix which was used in test statistic.

1.2 Theoretical Section

In multivariate one way ANOVA, observations are obtained for one independent variable. But in multivariate case, P-responses are simultaneously obtained for each subject. Here, interest is centered on differences among g- treatments on several measures. This implies a single set of g-groups is studied. The number of variables is denoted by P and the number of individuals in each g^{th} group by Ng. In general as one sample from a p-variate normal population, X~N(μ, Σ)

$$X = \begin{pmatrix} x_{11} & x_{12} \ldots x_{ij} \ldots x_{p1} \\ x_{21} & x_{22} \ldots x_{i2} \ldots x2 \\ \vdots & \vdots \quad \vdots \quad \vdots \\ x_{1n} & x_{2n} \ldots x_{in} \ldots x_{pn} \end{pmatrix}$$

$$mean \ \bar{x}_1 \quad \bar{x}_2 \ldots \bar{x}_i \ldots \quad \bar{x}_p$$

Where x_{ij} is the i^{th} observation on the j^{th} variance $i = 1,2,\ldots,Ng \quad and \quad j = 1,2,\ldots,P \quad and \quad g$ is the number of groups under investigation $(i.e. \quad g = 2)$

The mean vector U is denoted by

$$U = \begin{pmatrix} u_1 \\ \vdots \\ u_p \end{pmatrix}$$

The variance covariance matrix $\sum$ for the data is given by

$$\sum = \begin{pmatrix} \sigma_{11} & \sigma_{12} \cdots \sigma_{1p} \\ \sigma_{21} & \sigma_{22} \cdots \sigma_{2p} \\ \vdots & \vdots \ \cdots \ \vdots \\ \sigma_{p1} & \sigma_{p2} \cdots \sigma_{pp} \end{pmatrix}$$

The corresponding unbiased estimate are given as

$$U = \begin{pmatrix} x_{11} \\ \vdots \\ x_{1p} \end{pmatrix} \qquad S = \begin{pmatrix} s_{11} & s_{12} \cdots s_{1p} \\ s_{21} & s_{22} \cdots s_{2p} \\ \vdots & \vdots \ \cdots \ \vdots \\ s_{p1} & s_{p2} \cdots s_{pp} \end{pmatrix}$$

We assume that the data are random samples from different populations, independent and the populations have common covariance matrix and multivariate normal.

1.3 Multivariate Analysis of Variance (MANOVA)

The univariate case (1) can be extended to multivariate.

$$X_{ij} = \mu + \tau_i + e_{ij}, \quad i = 1,2,\ldots, p, \quad j = 1,2,\ldots,n \qquad (1)$$

Where

X_{ij} is the observation or response of the ith treatment

μ is the overall mean

τ_i i^{th} Treatment

e_{ij} is the error associated with the observation X_{ij} , $e_{ij} \sim N(0, \sigma^2)$

The model is assumed to be normal, independent a nd homoscedasticity. The MANOVA table is shown below as Table 1.

Multivariate Analysis of Variance (MANOVA) compares mean vectors of g samples for significant differences. It becomes necessary to apply a suitable test statistic to obtain a reliable result. Using table 1:

Hypothesis

$H_0; \tau_1 = \cdots = \tau_9$

$H_1 : At \quad one \quad i \neq 0$

Teast statistic

$$Wilks'\ lamda(\Lambda^*) = \frac{|W|}{|B+W|}$$

where

W = 'within' sum of squares and products matrix
B = 'Between' sum of squares and products matrix
B + W = Total sum of squares and products matrix

Using table 2:
Decision Rule:
(α *level of significance*)

Reject H_0 if Λ^* > F approximation of Wilks' lamda at α level of significance.

1.4 Hotellings: T^2 – Test Criteria

Hypothesis

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

Test statistic

$$T^2 = \frac{n_1 n_2}{n_1 + n_2}\left(\overline{X}_1 - \overline{X}_2\right)' S_{pooled}^{-1}\left(\overline{X}_1 - \overline{X}_2\right)$$

Anderson(1984)

$$T^2 = \frac{(n_1 + n_2 - 2)}{(n_1 + n_2 - p - 1)} F_{p, n_1+n_2-p-1}(\alpha)$$

$$F = \frac{(n_1 + n_2 - p - 1)}{(n_1 + n_2 - 2)} T^2 \sim F_{p, n_1+n_2-p-1}$$

where $F_{p,n_1+n_1-p-1}(\alpha)$ is the value of F with degree of freedom of numerator $n_1 + n_2 - p - 1$.

Decision Rule

Reject H_0 at α level of significance if

$$F > F_{p, n_1+n_2-p-1}(\alpha)$$

1.5 Application Section

The data for this paper is basically primary, the population of study is N = 89 addicts. They were chosen from the Federal Neuro-Psychiatric Hospital, Enugu, Nigeria. The population was divided into two groups and scored over ten–group A_1 of size $N_1 = 43$ consist of those treated of substance abuse before while $N_2 = 46$ in group A_2 are those receiving treatment for the first time.

The selection of the patients was done on a weekly basis for a period of 4 weeks. Only patients medically diagnosed to be on substance abuse or suffering from drug addiction were selected. Each patient was selected through his folder (medical records) and subsequently served a questionnaire. The data was collected from male ward and Out Patient Department (OPD) of the Hospital. See table 3 and 4.

1.6 MANOVA

We are going to form three matrices namely - squares and product for treatment (B), Sum of squares and product for error(W) and total sum of squares (B + W) .

$$A_{1t} = [185.5 \quad 171.5 \quad 191.0 \quad 183.5 \quad 176.5 \quad 191.5 \quad 158.0 \quad 147.0 \quad 157.0]^T$$

$$A_{2t} = [235.5 \quad 190.0 \quad 232.0 \quad 277.0 \quad 224.0 \quad 238.0 \quad 207.0 \quad 205.0 \quad 189.5]^T$$

$$A_{1m} = [4.31 \quad 3.99 \quad 4.44 \quad 4.27 \quad 4.10 \quad 4.45 \quad 3.67 \quad 3.42 \quad 3.65]^T$$

$$A_{2m} = [5.12 \quad 4.13 \quad 5.04 \quad 6.02 \quad 4.87 \quad 5.11 \quad 4.50 \quad 4.46 \quad 5.12]^T$$

$$\bar{x} = \frac{N_1 A_{1m} + N_2 A_{2m}}{N}$$

$$B = \sum_{i=1}^{p} N_1 (A_{1mi} - \bar{x})^2 + \sum_{j=1}^{p} N_2 (A_{2mi} - \bar{x})^2$$

$$B = \begin{pmatrix} 14.58 & 2.85 & 10.80 & 31.50 & 13.86 & 11.88 & 14.94 & 19.31 & 8.46 \\ 2.55 & 0.43 & 1.97 & 5.30 & 2.40 & 2.05 & 2.58 & 3.24 & 1.46 \\ 10.80 & 1.97 & 8.00 & 23.34 & 10.27 & 8.80 & -7.33 & 23.53 & 28.48 \\ 31.50 & 5.30 & 23.34 & 68.07 & 29.95 & 26.94 & 70.98 & 40.45 & 18.28 \\ 13.86 & 2.40 & 10.27 & 29.95 & 13.18 & 11.29 & 14.20 & 17.80 & 8.04 \\ 11.88 & 2.05 & 8.80 & 26.49 & 11.29 & 9.68 & 12.17 & 15.25 & 6.89 \\ 14.94 & 2.58 & -7.33 & 70.98 & 14.20 & 12.17 & 15.31 & 10.18 & 8.87 \\ 19.31 & 3.24 & 23.53 & 40.45 & 17.80 & 15.25 & 10.18 & 24.04 & 10.86 \\ 8.46 & 1.46 & 24.46 & 18.28 & 8.04 & 6.89 & 8.97 & 10.86 & 4.91 \end{pmatrix}$$

$$|B| = 490.62$$

$$B^{-1} = \begin{pmatrix} 4.58E+00 & 2.06E+01 & 1.37E-03 & 2.07E-01 & -1.80E+00 & -9.99E+00 & 2.97E+01 & 8.64E-01 & -2.70E-01 \\ 2.06E+01 & -2.90E+01 & -1.56E-02 & 3.54E-01 & 3.34E+01 & -6.54E+01 & 1.32E+00 & 1.56E+00 & 6.30E-01 \\ 1.37E-01 & -1.56E-02 & -2.00E-05 & -2.38E-04 & -2.58E-02 & -7.23E-03 & 7.94E-05 & 2.91E-04 & 5.49E-02 \\ 2.07E-01 & 3.54E-01 & -2.38E-04 & -6.35E-03 & -8.08E-01 & 6.38E-01 & 1.35E-02 & -2.58E-02 & 2.53E-02 \\ -1.80E+00 & 3.34E+01 & -2.58E-02 & -8.08E-01 & -1.68E+00 & 3.43E+00 & 2.32E-04 & -3.29E+00 & 1.52E+00 \\ -9.99E+00 & -6.54E+01 & -7.23E-03 & 6.38E-01 & 3.43E+00 & 1.82E+01 & -6.51E-01 & 2.76E+00 & -1.77E+00 \\ 2.97E-01 & 1.33E+00 & 7.94E-05 & 1.35E-02 & 2.34E-04 & -6.51E-01 & 1.92E-02 & -5.35E-02 & 4.06E-02 \\ 8.64E-01 & 1.56E+00 & 2.91E-04 & -2.58E-02 & -3.29E+00 & 2.76E+00 & -5.35E-02 & -1.10E-03 & -4.49E-03 \\ -2.70E-01 & 6.35E-01 & 5.49E-02 & 2.53E-02 & -1.452E+00 & -1.77E+00 & 4.06E-02 & 4.49E-03 & 4.02E-02 \end{pmatrix}$$

Within treatment Sum of squares

$$W_1 = \sum_{i=1}^{p} \left(A_{1ti} - A_{1mi}\right)^2 + \sum_{j=1}^{p} \left(A_{2tj} - A_{2mj}\right)^2$$

Sum of cross products

$$W_2 = \sum_{i} \sum_{j} \left(x_{\ell ij} - \bar{x}_{\ell i.}\right)\left(x_{ih} - \bar{x}_{\ell i.}\right)$$

W_1 formed diagonal elements (variances) of matrix W while W_2 are the covariances of W.

$$W = \begin{pmatrix} 85674.53 & 73073.98 & 85908.98 & 114288.23 & 81787.77 & 86890.43 & 74627.39 & 72354.47 & 70508.69 \\ 73073.98 & 62507.66 & 73453.73 & 80015.45 & 65353.53 & 74185.08 & 63644.77 & 61518.76 & 60231.62 \\ 85908.20 & 73435.73 & 86315.48 & 95335.95 & 82071.44 & 87221.68 & 7393.58 & 72536.88 & 70789.50 \\ 114288.23 & 80715.45 & 95335.95 & 106505.05 & 90872.58 & 96402 & 83114.19 & 80710.75 & 78171.79 \\ 81787.787 & 65353.50 & 82071.44 & 90872.58 & 78077.77 & 83022.33 & 71318.76 & 69080.63 & 71531.84 \\ 86890.43 & 74185.08 & 87221.68 & 96402.87 & 83022.33 & 35218.25 & 75745.16 & 73330.91 & 71531.84 \\ 74627.39 & 63644.77 & 74939.58 & 83114.19 & 71318.76 & 75745.16 & 65160.84 & 63145.96 & 61455.39 \\ 72354.47 & 61518.76 & 72536.88 & 80710.75 & 69080.63 & 73330,91 & 63145.96 & 61249.71 & 59481.63 \\ 70508.69 & 60231.06 & 70769.50 & 78171.79 & 71531.89 & 67305.96 & 61455.39 & 59481.63 & 58168.05 \end{pmatrix}$$

$$|W| = -2.82E+27$$

$$|B+W| = \begin{pmatrix} 85689.21 & 73076.52 & 85919.00 & 114319.73 & 81801.63 & 86892.31 & 74612.33 & 72373.78 & 70517.15 \\ 73076.53 & 62507.69 & 73437.70 & 80720.75 & 65355.75 & 74187.13 & 63647.35 & 61522.00 & 60233.00 \\ 85919.00 & 73437.70 & 86323.48 & 95359.29 & 82081.71 & 87230.48 & 74932.25 & 72560.41 & 70813.98 \\ 114319..70 & 80720.75 & 95359.29 & 106573.1 & 90902.53 & 96429.56 & 83185.17 & 80751.20 & 76190.07 \\ 81801.63 & 65355.93 & 82081.71 & 90902.53 & 78090.95 & 83033.62 & 71332.96 & 69098.43 & 67314.00 \\ 86902.31 & 74187.13 & 87230.48 & 96429.36 & 83033.62 & 35227.93 & 75757.33 & 73346.16 & 71538.73 \\ 74642.33 & 63647.35 & 74932.25 & 83185.17 & 71332.96 & 75757.33 & 65176.15 & 63156.14 & 61464.26 \\ 72373.78 & 61522.00 & 72560.41 & 80751.20 & 69098.43 & 73346.16 & 63156.14 & 61273.75 & 59492.49 \\ 70517.15 & 602.0633 & 70813.98 & 78190.07 & 673134.00 & 71538.73 & 61462.26 & 59492.49 & 58172.96 \end{pmatrix}$$

$$|B+W| = -2.08E+31$$

$$\wedge^* = \frac{|W|}{|B+W|} = \frac{-2.82E+27}{-2.08E+31} = 0.000136$$

Decision Rule: Reject H_0 *if* $\Lambda* \geq F_t^{\alpha} = 2.04$.

Computation

$$F_c = \left[\frac{89-9-1}{9}\right]\left[\frac{1-0.000136}{0.000136}\right]$$

$$= (8.77)(7351.94) = 64476.51$$

Critical value:

$$F_P^{\alpha}(\sum n_i - P - 1) = F^{0.05}_{9,(89-9-1)} = F^{0.05}_{9,79} = 2.04$$

Conclusion: Since $F_c = 64476.51 > F_t^{\alpha} = 2.04$, we reject H_0 and conclude that there is a significant difference among the factors that is, the nine factors studied actually affect addicts. .

1.7 HOTELLING'S T[2]

$D = a_{1mi} - a_{2mi} \quad i = 1, \cdots, 9, a_{1mi} \in A_{1m}, a_{2mi} \in A_{2m}$ the differences between the sample mean vectors for groups 1 and 2.

$$D^T B^{-1} D = [9.82E-08, 9.90E-03, -5.66E-04, -5.08E-04, -6.74E-02, -7.33E-04.4.71E-05, -9.73E-04, 5.09E-04]^T$$

$$T_c^{\ 2} = \frac{n_1 n_2}{n_1 + n_2} D^T B^{-1} D \quad = 1.0001$$

Critical values

$$T_{\alpha}^{\ 2} = \frac{P(n_1 n_2 - 2)}{n_1 n_2 - 2 - P + 1} F_{(P, n1+n2-2P+1)\alpha} \quad = 20.22$$

Since $T_c^2 = 1.0001 < T_{\alpha}^2 = 20.22$, we do not reject H_0 and conclude that there is no significant difference in the population means of groups 1 and 2. That is attraction to drugs does not depend on the group.

1.8 Conclusion

The study shows that most addicts especially youth have strong appeal to drug abuse. This resulted in most of them being hospitalized. Many are unemployed and influenced by peer group. Drugs abuse cuts across educational and professional background. The difference in the population mean vectors for the causative factors to drug abuse for groups A and B are not the same. Reasons for addiction and abuse are true. Besides, there is no significant difference in the population means for group A and B respectively. It follows that causative factors for drug abuse for groups A and B are the same. Government and International Organizations should provide adequate funding for drug prevention programme. There should be enlightenment for drug awareness and government should provide job opportunities for the teeming unemployed youths. Since $T_c^2 = 1.0001 < T_{\alpha}^2 = 20.22$, we do not reject Ho we conclude that there is no significant difference in the population means in group A and B.

1.9 References

Adelekan, M and Adeniran, R.A.I.(1991) Rehabilitation and Following Issues In Drug Abusers Managed at Neuropsychiatry Hospital, Abeokuta Nigeria. West African Journal of Medicine 10(1) p 350-360.
Anderson, T. W. (1984) An Introduction to Multivariate Statistical Method. 2nd Edn. John Wiley and Sons. New York.
Bita, K. and Hamid, A. (2012) . The Impact of Teachers Belief on ELF Writing and Instruction. Theory and Practice in Language Studies Vol 2 No2 pg 334-342. Academic Publishers, Finland.
Allan, C. and Vicky, C. (1993) Drugs, Alcohol and Health Cambridge University Press.
Ezekwem, C.C. (1993) (Unpublished) A Multivariate Analysis of Year One Academic Programme of Students Admitted Through Pre-Science and Entrance Examination at the Anambra State College of Education Nsugbe.
Grilo, L.M. (2005) Develpement of Near-Exact Distributions forDifferent Scenarios of Application of the Wilks' Lamda Statistic. Ph.D Thesis, Technical University of Lisbon, Portugal.
Ijeoma, E.O. (1997): Drug Use Among Students. An Introduction to Drug Abuse. Education in Nigerian Schools.
Inyoung, K., Sunbo L., Sang-Cheol K., Byung S. K. (2006) Statistical Analysis for the Feature Subset Selection In a CDNA Microarray Using Hotelling's T^2 Statistics. Bionfomatics and Biosystems, Vol I, No.3 pp. 165-170.
Jin-Ting, Z. (2012). An Approximate Hotelling's-test for Heteroscedastic one way MANOVA. Open Journal of Statistics. 2, 1-11
Obot, L.S. (1989) Treatment Options in Alcohol and Drug Dependence Nigeria Journal Basic and Applied Psychology, 2 (1 & 2) p. 63-79.
Odejide, A. O. and Olatuwura, M.O. (1977) Community African Journal of Psychiatry 3(1&2) p.69-74
Ogum, G.E.O. (2002). Introductions to Multivariate Analysis. Afri Towers Ltd, Aba – Nigeria
Oludare, S. Ariyo, Arogundade,O. and Ayoola M. Abdul- Rafiu(2011) Discriminant and Classification Analysis of Health Status of Bell Peper (Capsium annum) Research Journal of Mathematics and Statistics 3(2):77-81,
Zhindong, B. and Hewa, S.(1996). Effect of High Dimension: By an Example of a Two Sample Problem. Statistica Sinica 6, 311-329.

Table 1 : MANOVA table

Source	**d.f**	**Square**	**Matrix of SS & Cross product**
Treatment	$g-1$	$\sum n_i (x_{L.} - \bar{x}_{..})^2$	B
Error	$\sum n_i - g$	$\sum\sum (x_{ij} - x_{L.})^2$	W
Total	$\sum n_i - 1$	$\sum_{i=1}^{g}\sum_{j=1}^{n} (x_{ij} - \bar{x}_{L.})^2$	$B+W$

Table 2: Lamda distribution table

No of Variables	**No of groups**	**Sampling design for multivariate normal data**	**Degrees of freedom**
$P=1$	$g \ge 2$	$\left(\frac{\sum n_i - g}{g-1}\right)\left(\frac{1-\Lambda^*}{\Lambda^*}\right)$	$F_{(g-1),\sum n_i - g}$
$P=2$	$g \ge 2$	$\left(\frac{\sum n_i - g - 1}{g-1}\right)\left(\frac{1-\sqrt{\wedge^*}}{\sqrt{\wedge^*}}\right)$	$F_{2(g-1),2(\sum n_i - g - 1)}$
$P \ge 1$	$g=2$	$\left(\frac{\sum n_i - P - 1}{P}\right)\left(\frac{1-\wedge^*}{\wedge^*}\right)$	$F_{p,\sum n_i - p - 1}$
$P \ge 1$	$g=3$	$\left(\frac{\sum n_i - P - 2}{P}\right)\left(\frac{1-\wedge^*}{\sqrt{\wedge^*}}\right)$	$F_{2p,2(\sum n_i - p - 2)}$

Table 3 Scores for group A_1

1	2	3	4	5	6	7	8	9	10
S/N	**Peer Group** x_1	**Stress** x_2	**Enhance Performance** x_3	**Health** x_4	**Unemployment** x_5	**People** x_6	**Curiosity** x_7	**Happiness** x_8	**Poverty** x_9
1	7	4.5	5	3.5	20	5	3	2	4
2	5	3.5	4	6	3	3.5	2	4	2
3	3	4	2	4	6	4	6	2	2
:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:	:
43	6	4	3	5	4	5.5	5	4	4
A_{1t}	185.5	171.5	191.0	183.5	176.5	191.5	158.0	147	157
A_{1m}	**4.31**	**3.99**	**4.44**	**4.27**	**4.10**	**4.45**	**3.67**	**3.42**	**3.65**

Source: Enugu Psychiatric Hospital

Table 4 Scores for group A_2

1	2	3	4	5	6	7	8	9	10
S/N	**Peer Group** x_1	**Stress** x_2	**Enhance Performance** x_3	**Health** x_4	**Unemployment** x_5	**Pride** x_6	**Curiosity** x_7	**Happiness** x_8	**Poverty** x_9
1	7	6	8	4	3	6	7	4	3
2	6	5	3	4	5	5	3	9	4
3	3	5	3	5	7	3	5	4	6
:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:	:
46	5	4	7	5	6	5	4	3	5
A_{2t}	235.5	190.0	232.0	277.0	224.0	235.0	207.0	205.0	189.5
A_{2m}	**5.12**	**4.13**	**5.04**	**6.02**	**4.87**	**5.11**	**4.5**	**4.46**	**4.12**

Source: Enugu Psychiatric Hospital

$$g = 2, \quad p = 9, \quad N_1 = 43, \quad N_2 = 46, \quad N = \sum_{i=1}^{2} N_i$$

where
g – the number of groups
p – number of factors under study
N_1 - sample size for group A_1
N_2 - sample size for group A_2