A Comparison of Students' Performance: An Application of Hotelling's T^2

Onyedikachi O. John[1], Joseph O. Onu[2]

[1]Department of Physical Sciences, Rhema University, Nigeria
[2]Department of Statistics, Abia State Polytechnic, Nigeria

*Corresponding Author:
Onyedikachi O. John
Email: johnkady@yahoo.com

Abstract: This paper examined the equality of the performance of two sets of students: those who graduated in 2012/2013 and 2013/2014 academic sessions from the Department of Statistics, Abia State Polytechnic, Aba, Nigeria, based on their scores in the same eight courses which both sets offered during their second semester HND II. Hotelling's T^2 was applied to the data. The analysis and tests showed there is a significant difference in performance between the two sets. A Roy-Bose simultaneous confidence interval was used to identify the mean vectors that contributed to the rejection of the null hypothesis. The result showed in general that students who graduated in 2012/2013 academic session performed better than those who graduated in 2013/2014 academic session.

Keywords: Hotelling's T^2 , mean vectors, simultaneous confidence interval, students' performance

INTRODUCTION

Studies have shown that the mean performance of both pupils and students either from the same school, class or department vary from one set to another, even when the different sets of pupils or students involved are taught the same subjects or courses, though in different academic sessions. For those in primary and secondary schools, this variation in the mean performances may be between one term and the other, or between one academic session and the other, while for those in tertiary institutions, this difference in the mean performance could be between one semester and another or between one academic session and the other.

It has been observed among all other that it takes a good teacher (lecturer), a good student and a good environment (facilities and technical know-how) to produce a good result. Mkpa [1], observed that over the years, teachers have been trained to adopt various methods of teaching. However, some methods are obviously more productive than others. For students in tertiary institutions, the foundations they have in their secondary education and their commitment to the day to day lectures and studies in general go a long way to determining their performances. These and other factors constitute why the mean performances of students that graduated in one academic session may differ from those who graduated in another academic session, even when both groups are subjected to the same courses. Hence the performance of students in our various tertiary institutions is of great interest and concern to many researchers.

Ugbe and Nyiam [2], applied discriminant analysis to compare the performance of students who gained admission into the university system through pre-degree programme and those who passed the University Matriculation Examination (UME). It was observed that there is no significant difference in the performance of UME and pre-degree students at 5 percent level of significance.

Elem-Uche *et al* [3], applied the Hotelling's T-square statistic in testing equality of export and import commodities price index in Nigeria using some selected export and import commodities price index in Nigeria. The test showed that there is significant difference between the price index of export and import commodities.

In this study, our major task is to compare the mean performances of two different sets of HND II students in Statistics Department, Abia State Polytechnic, Aba, that graduated during the 2012/2013 and 2013/2014 academic sessions respectively, based on the scores they made on the same set of eight courses offered during each of the second semesters of 2012/2013 and 2013/2014 academic sessions. Hence to answer the question (i) whether there is equality between the HND II students' mean performance for those who graduated in 2012/13 and those of 2013/14 academic sessions based on the same eight courses offered by both sets during their final year second semester (ii) if no equality, which of the courses led to the difference existing between the performances in the two academic sessions.

MATERIALS AND METHODS

The data for this work are the results of the second semester for the Higher National Diploma (HND) II students in Department of Statistics, Abia State Polytechnic, Aba. The results are for two different sets of students: those who graduated in 2012/2013 and 2013/2014 academic sessions respectively. The result involve eight courses which both sets offered during their second semester final year, namely, STA 421 (Operations Research), STA 422 (Demography), STA 423 (Non-parametric), STA 424 (Statistical Computing), STA 425 (Time Series Analysis), STA 426 (Multivariate and Stochastic), STA 427 (Project), and STA 428 (Seminar).

The description of vectors as used in the analysis is as follows: $x_1, x_2, ..., x_8$ represent the scores made by students in STA 421, STA 422,…, STA 428, respectively, for the second semester 2012/2013 academic session. Similarly, $y_1, y_2, ..., y_8$ represent the scores made by students in the same courses for the second semester 2013/2014 academic session.

We used the mean vector of the students' scores from both academic sessions and their variance-covariance matrices. The variance-covariance matrix for a particular set can be represented as follows:

$$S = \begin{bmatrix} \text{var}(x_1) & \text{cov}(x_1 x_2) & . & . & . & \text{cov}(x_1 x_8) \\ & \text{var}(x_2) & . & . & . & \text{cov}(x_2 x_8) \\ & & . & & & . \\ & & & . & & . \\ & & & & . & . \\ & & & & & \text{var}(x_8) \end{bmatrix}$$

Their pooled variance-covariance matrix is given by

$$S_{pooled} = \frac{S_X + S_Y}{n_X + n_Y - 2} \qquad (1)$$

Where, S_X and S_Y are the variance-covariance matrices for the 2012/13 and 2013/14 academic sessions respectively. n_X and n_Y are the number of sample results selected in each of the 2012/13 and 2013/14 academic sessions. Here, $n_X = n_Y = 100$, which refers to the first 100 students that registered from each session, and also offered the courses.

Hotelling's T^2

The Hotelling's T^2 as a test statistic is the generalization of the student's t-statistic that is used in multivariate hypothesis testing, Hotelling [4]. This test statistic which measures the difference in mean vector of two populations is given by

$$T^2 = \frac{n_X n_Y}{n_X + n_Y} \left(\bar{x}_1 - \bar{y}_1, .., \bar{x}_8 - \bar{y}_8\right)^T S_{pooled}^{-1} \left(\bar{x}_1 - \bar{y}_1, ..., \bar{x}_8 - \bar{y}_8\right) \qquad (2)$$

And the relationship between T^2 - statistic and the Fisher's F - distribution with degrees of freedom p and $n_X + n_Y - p - 1$ is given by the quantity

$$F = \frac{n_X + n_Y - p - 1}{(n_X + n_Y - 2) p} T^2 \qquad (3)$$

Hence,

$$T^2_{\alpha; p, n_X + n_Y - p - 1} = \frac{(n_X + n_Y - 2) p}{n_X + n_Y - p - 1} F_{\alpha; p, n_X + n_Y - p - 1} \qquad (4)$$

This is applied to test the equality of the two mean vectors based on the students' performances from two independent and identical populations.

The hypotheses are stated as follows:

$H_0: \mu_X = \mu_Y$; there is no significant difference between the HND II students' performance in 2012/2013 session and that of 2013/2014 session

$H_1: \mu_X \neq \mu_Y$; there is significant difference between the HND II students' performance in 2012/13 session and that of 2013/14 session.

H_0 is rejected if $T^2 > T^2_{\alpha;p,n_X+n_Y-p-1}$

Simultaneous Confidence Interval

Roy and Bose [5] gave confidence interval which is used to obtain which response vector appear to contribute to the strong decision to reject the null hypothesis. The expression for the Roy-Bose simultaneous confidence interval is as follows:

$$\mu_X - \mu_Y : \left(\bar{x}_i - \bar{y}_i\right) \pm \sqrt{T^2_{\alpha;p,n_X+n_Y-p-1}\left(\frac{n_X+n_Y}{n_X n_Y}\right)S_{ii.pooled}} \qquad (5)$$

If the $100(1-\alpha)\%$ simultaneous confidence interval of the responses includes zero value, the null hypothesis is evident; otherwise the alternative hypothesis is evident.

RESULTS AND DISCUSSION

The mean values and mean differences for the students' performances are given in table 1:

Table 1: Mean values and mean differences for the students' performances

Course title/code (vectors)	2012/2013 session	2013/2014 session	Mean vector difference
Operations Research (STA 421)	63.13	57.63	5.50
Demography (STA 422)	66.40	60.35	6.05
Non-parametric (STA 423)	64.51	61.62	2.89
Statistical Computing (STA 424)	72.22	63.45	8.77
Time Series Analysis (STA 425)	62.32	59.36	2.96
Multivariate & Stochastic (STA 426)	54.14	62.21	-8.07
Project (STA 427)	76.42	75.80	0.62
Seminar (STA 428)	72.50	77.05	-4.55

The pooled variance-covariance matrix and its inverse are

$$S_{pooled} = \begin{bmatrix} 86.7335 & & & & & & & \\ 24.3510 & 161.9066 & & & & & & \\ 30.0032 & 34.9384 & 234.1672 & & & & & \\ 23.8564 & 13.2167 & 2.8776 & 162.5648 & & & & \\ 40.5572 & 30.7233 & 27.1047 & 49.5734 & 125.7455 & & & \\ 24.3631 & 15.8086 & -0.5875 & 48.2270 & 33.6460 & 250.1276 & & \\ 11.7892 & 2.8605 & -0.9497 & 7.6036 & 5.3107 & 9.0941 & 87.3370 & \\ 4.2460 & -6.3409 & 14.3070 & -9.1945 & 13.1636 & -14.2025 & 36.6363 & 86.7955 \end{bmatrix}$$

$$S_{pooled}^{-1} = \begin{bmatrix} 1.3572 \\ -0.1759 & 0.6273 \\ -0.1687 & -0.1065 & 0.4536 \\ -0.1270 & -0.03853 & -0.006513 & 0.6965 \\ -0.3924 & -0.1379 & -0.004801 & -0.2792 & 1.09832 \\ -0.1072 & -0.04148 & -0.001825 & -0.1205 & -0.07691 & 0.4236 \\ -0.2867 & -0.1174 & 0.2830 & -0.1596 & 0.1031 & -0.1098 & 1.3584 \\ -0.09757 & -0.008216 & -0.1497 & 0.03566 & -0.2667 & 0.03443 & -0.8145 & 1.2694 \end{bmatrix}$$

$T^2 = 8347.682$ and $T_{\alpha;p,n_X+n_Y-p-1}^2 = 16.076$

This showed that there is a significant difference between the mean performance of students who graduated in 2012/2013 academic session and those who graduated in 2013/2014 academic session at a 0.05 level of significance, base on the same eight courses both sets offered during their final year second semester.

Table 2: Roy-Bose simultaneous confidence interval

Course Title/code	Lower bound	Upper bound
Operations Research (STA 421)	0.2192	10.7808
Demography (STA 422)	-1.1650	13.2650
Non-parametric (STA 423)	-5.7869	11.5669
Statistical Computing (STA 424)	1.5404	15.9996
Time Series Analysis (STA 425)	-3.3984	9.3184
Multivariate & Stochastic (STA 426)	-17.0378	0.8978
Project (STA 427)	-4.679	5.9190
Seminar (STA 428)	-9.8327	0.7327

Although the null hypothesis has been rejected, it still remains unknown which of the eight courses led to the conclusion that the mean performance of students who graduated in 2012/2013 session and those who graduated in 2013/2014 session are statistically unequal. The Roy-Bose simultaneous confidence interval results as given in table II reveals that the courses which led to the rejection of the null hypothesis are Operations Research (STA 421) and Statistical Computing (STA 424).

CONCLUSION

We have compared the performance of students who graduated from the Department of Statistics, Abia State Polytechnic, Aba, in 2012/2013 academic session and 2013/2014 academic session respectively based on their HND II second semester result. The result of the analysis reveals that statistically there exists a significant difference between the mean performance of students who graduated in 2012/2013 session and those who graduated in 2013/2014 session based on the eight courses offered during their HND II second semester. The courses that contributed to the rejection of the null hypothesis are Operations Research and Statistical Computing. In general, the students who graduated in 2012/2013 academic session performed better than those who graduated in 2013/2014 academic session.

REFERENCES
1. Mkpa AM. The Learner-Centred Teaching Methods for Successful Universal Basic Education in Nigeria. Barloz Publishers Inc. Owerri, Nigeria, 2003.
2. Ugbe TA, Nyiam PB. A Comparison of the Performance of Students in Pre-degree an and University Matriculation Examination (UME) Classes in a University System: A Discriminant Analysis Approach. International Journal of Adv. in Research and Technology. 2013;2(9).
3. Elem-Uche O, Onu JO, Orji GO. Multivariate Approach to Parametric Analysis of selected Export and Import Commodities Price Index in Nigeria. A paper presented to a conference held at Akanu Ibiam Polytechnic , Uwana-Afikpo, 8th -10th June, 2016.
4. Hotelling H. A generalized T test and Measure of Multivariate Dispersion. Proceedings the second Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley. 1951;23–41.
5. Roy SN, Bose RC. Simultaneous Confidence Interval Estimation. Annals of Mathematical Statistics. 1953;24:513-536.