Vidyodaya J. Arts. Sci., Lett., Vol. 7 Nos. 1 & 2, 1979 *pp.* 111–122

A Statistical Analysis of Admission and Performance of Science Students at Vidyodaya University

by

G. Y. L. FERNANDO
Dept. of Census and Statistics

and

S. WEERAHANDI
Dept. of Mathematics, University of Sri Jayawardenepura

1. Introduction

This study is based on the findings of a survey carried out by us in December 1977 on pre-university and university educational background of science undergraduates studying at Vidyodaya campus and Colombo campus. The subjects of the survey comprised (i) first year science students of Vidyodaya campus (ii) third year science students of Vidyodaya campus (iii) first year science students of Colombo campus, who were all studying in the academic year 1977/78. The relevant information was collected by means of a questionnaire specially designed for the purpose. The results of the survey are reported in Fernando (1978) and the relevant particulars will be reproduced here when their need arises.

The objectives of this study are basically two fold. Firstly, in Section 2 we shall make an attempt to characterise the composition of students admitted to Vidyodaya by means of number of classifications. Such classifications will in particular allow us to test the validity of some hypotheses of interest.

Secondly we wish to identify within the limitations allowed by the available data, the factors affecting the performance of students at university examinations. We shall do this, in Section 3, by designing a suitable regression model.

We have made a number of very interesting and important observations and conclusions throughout this analysis and it is hoped that they will be of immense use to decision-making bodies of our universities here in Sri Lanka and Vidyodaya in particular.

Except for some comparisons done in Section 2, we were compelled to confine our study to Vidydoaya Campus due to fack of facitities to obtain the access to similar data in respect of otner campuses. Nevertheless it is argued that, almost all the observations and conclusions made in Section 3 should have a general validity and thus may well apply to other campuses as well. In any case, the analysis done in this study will illustrate the type of approach one should take in drawing conclusions systematically in similar problems.

2. University Admission

In this section we shall characterize the composition of students who entered the Applied Science faculty of Vidyodaya Campus in the sample year 1977, and then undertake a statistical study of the factors affecting the university admission in respect of science students. We hope that the observations derived from this sample year will serve as a typical representation of the general structure as long as the course structure of the faculty and the criterion of selection for this campus as compared with other campuses prevail unchanged. We wish to use the data to test certain statistical hypotheses of interest.

2.1. Classifications

SEX : It is a generally accepted hypothesis that the proportion of female students in Bio-Science courses tend to be greater than that in Physical Science courses. As the number of students that will be enrolled to different courses are likely to increase at different rates, in view of the above fact, the following two way classification is more informative while it is useful in the testing of a hypothesis of independence.

Stream of Subjects \ *Sex*	*Male*	*Female*	
Physical Science ..	25 (23)	8 (10)	33
Bio Science ..	32 (34)	17 (15)	49
	57	25	82

Hence in this sample year, of the physical science students (33) 75% are male students while of the bio-science students (49), 65% are male students.

Now in order to test the hypothesis that the above two criteria of classifications are independent we compute the respective expected frequencies assuming that the hypothesis is true. These frequencies are indicated in Figure 2.1 within brackets Hence under the desired hypothesis, we have $\psi^2 = .96$. As the corresponding table value is $\psi_1^2(.05) = 3.84$, here we do not have sufficient evidence to reject the independence hypothesis at .05 level of significance. In fact it is seen that proportion of male students in physical science group is not significantly different from that in bio-science group at .05 level.

RACE AND RELIGION : Normally Tamil students are not admitted to the science faculty of Vidyodaya Campus (Sri Jayewardenapura University). However non-Sinhalese students may enter the faculty provided that they have done their advanced level studies in Sinhala medium. Accordingly our samples

consist of 81 Sinhala students and 1 Muslim student. The Sinhala student population comprises of Buddhist students and catholics having representations more or less in the same percentages as in the whole island. The following table summarizes the religion wise classification of the first year Science student population of this campus in 1977.

	Males	*Females*	*Total*
Buddhists	49	22	71
Catholics	7	2	9
Non-respondents ..			2

Because of its religious background, there was a time when Vidyodaya was open and enjoyed the admission of Buddhist students only. This situation has now changed since the introduction of the G.C.E. (A. Level) examination as the basis for entrance requirement. It is therefore of interest to compare the current religious composition of this campus with another campus which admits Sinhala students. For this purpose we set up the following table which gives a two way classification of first year students in 1977 at Colombo campus and Vidyodaya campus.

	Colombo	*Vidyodaya*	
Buddhists	113 (116)	71 (68)	184
Non Buddhists	24 (21)	9 (12)	33
	137	80	217

In order to test the interested hypothesis that the two underlying criteria of classification, namely 'campus' and 'religion', are independent we compute the corresponding expected frequencies as shown within brackets in Table 2.3. It is easily checked that the computed value of $\psi^2 = 1.55$ under the desired hypothesis is less than the table value $\psi_1^2(.05) = 3.84$. Hence there is no evidence against the above hypothesis at .05 level of significance. We therefore have some evidence to assert that, when the students express their preference over campuses or courses of study, the religion is not taken into consideration.

AGE : The age of first-year University students in our sample ranges from 18 to 23. In the following classification each age reported by the students was rounded off to the nearest integer.

TABLE 2.4

Age			18	19	20	21	22	23
Male	Bio	..	2	5	9	6	6	3
	Physical	..	2	3	6	7	6	1
Female	Bio	..	9	1	3	0	0	0
	Physical	..	4	1	0	3	0	0

In the calculation of mean values ($\bar{Y}$) and their standard deviations ($S_{\bar{y}}$) for each group we make use of the formulae $\bar{Y} = \Sigma f_i\, y_i \,/\, \Sigma\, f_i$,

$$S_y^2 = (\Sigma f_i y_i^2 \,/\, \Sigma\, f_i - \bar{y}^2) \quad \text{and} \quad S_{\bar{y}}^2 = S_y^2 \,/\, \Sigma f_i.$$

In the following table abbreviations are usea for each of four groups indicated in Table 2.4 in an obvious sense.

TABLE 2.5

	Mean Age ($\bar{y}$)	S_y^2	$S_{\bar{y}}$	$S_{\bar{y}}^2$	Sample size N
M.B.	20.58	1.99	.25	.065	31
M.P.	20.60	1.75	.27	.070	25
F.B.	20.12	1.29	.28	.076	13
F.P.	21.00	2.5	.53	.277	8

where for instance M.B. stands for male bio science students. It is deduced from table 2.5 that in respect of first year science students,

(i) Sample mean age of male students : $y_m = 20.59$ years

$S^2_{y_m} = .034, \quad N_m = 56$

(ii) Sample mean age of female students : $y_f = 20.56$ years

$S^2_{y_f} = .088, \quad N_f = 21$

(iii) sample mean age of bio science students : $y_b = 20.35$ years

$S^2_{y_b} = .035 \quad , \quad N_b = 44$

(iv) sample mean age of Physical science students · $y_p = 20.80$ years

$S^2_{y_p} = .087 \quad , \quad N_p = 33$

Let us denote the corresponding population means of above quantities by $\mu m'$ $\mu f'$ μb and μp respectivelv, the population being the set of all students in the relevant category who will ever take first vear undergiaduate studies in this campus so long as the appropriate conditions do not change. As the standard deviations of above sample means are quite small, they are good estimates of population means. Hypotheses such as $H_1 : \mu m = \mu f$ and $H_2 : \mu b = \mu p$ are of some interest. For , falsity of such a hypothesis will lead us to search for the barriers, that exist for a certain group of students to lag behind another using the direction of some correct hypothesis as a guideline. To facilitate our tests, we assume that (i) $y_m \sim N(\mu_m, .034)$ (ii) $y_f \sim N(\mu_f, .088)$, (iii) $y_b \sim N(\mu_b), .035)$ and (iv) $y_p \sim N(\mu p, .087)$ approximately. As $y_m - y_f \sim N(\mu_m - \mu_f , .35^2)$, it is easily tested that the observed difference $y_m - y_f = .03\,\%$ is insignificant, thus accepting the hypothesis H_1 at .05 level for instance. It is deduced similarly from $y_p - y_b \sim N(\mu_p - \mu_b , .35^2)$ whose observed value is .052, that even H_2 is highly significant. Hence it seems that the mean age at which a student enters the university is more or less independent of students sex and course of study.

2.2. Factors of a District Governing University Admission

Of 22 districts of Sri Lanka, in 1977, 13 districts enjoyed the admission of a total of 82 students to the science faculty of Vidyodaya. The table below shows the number of students who were reported to have done their G.C.E. advanced level studies in each district.

TABLE 2.5

Colombo	30
Kalutara	11
Matara	8
Galle	5
Kandy	5
Ratnapura	4
Kurunegala ..	3
Badulla	3
Kegalle	3
Hambantota ..	3
Anuradhapura ..	3
Trincomalee ..	2
Matale	2

In view of the controversy over the fairness of the standardization system of G.C.E. (A. Level) scores obtained by students of various districts, it is tempting to make use of the above data to test whether the students come from rural areas are at an advantage. To accomplish this task, only by some readily available data we propose to test whether the following factors of a district have any significant impact on the number of students who get a chance to enter the University.

X_1 — Total number of students in the district who sat for the G.C.E. (A/L) science examination in 1976.

X_2 — Total number of schools in the district with G.C.E. (A/L) science facilities.

X_3 — Literacy rate of the district.

Total number of students who registered at a particular campus from each district should adequately reflect the pattern in which students got university admission in general. To improve the validity of the conclusions that we wish to draw from the study however, in addition to the data given in table 2.5, we make use of some similar data obtained from Colombo campus in the same year (for the results of the survey see Fernando 1978). Hence we treat the total number of students who registered at the university campuses of Vidyodaya and Colombo to pursue science courses from each district as our

dependent variable Y. The necessary data for the explanatory variables X_1 and X_2 were obtained from the Ministry of Education and as values for X_3, estimates based on 1971 population census available at the Department of Census and Statistics were utilized. These data estimates the assumed linear relationship between Y and X_1, X_2, X_3 to be

$$Y = 24.34 + 0.04X_1 - 1.33X_2 - 0.03X_3$$
$$(49.76)\ (0.01)\ (1.07)\ (0.07)$$

the numbers indicated within brackets being respective standard errors of estimates. It is now found applying the t-test that the two variables X_2 and X_3 are statistically insignificant at .01 level and the variables in Y is almost completely attributable to X_1 namely the total number of students who sat for the examination from the relevant district. In absence of insignificant variables the relationship between Y and X_1 is best represented by

$$Y = -9.27 + .025X_1$$
$$(4.09)\quad (0.003)$$

which account for a correlation coefficient $r^2 = 0.9$. As X_2 is highly significant we conclude that the number of students who got university admission from a given district is roughly proportional to the number of students who sat for the G.C.E. (A/L) examination from the same district. Hence the standardizing system has pretty well offset the advantageous position of urban students (Colombo students in particular) who enjoy better facilities including private tuition compared with rural students.

3. Students' Performance at University

It is argued by some people that the standardizing systems of G.C.E. marks keep some good students from getting into university. On the other hand it is the belief of the majority of the educated community that in view of the different levels of facilities enjoyed and various amounts of private tuition obtained those students who obtain best marks at the G.C.E.(A/L) are not really the best students. The real issue here boils down to the question "are the students who obtain comparatively high scores at the G.C.E. manage to keep up their position at the university as well. Then there are interesting questions such as "Does a student's family income have any bearing on his performance at the University? It is our objective of this section to answer questions of such nature. To achieve this goal we make use of some information we obtained from third year science students of this campus in 1978. In this study the average score (to be defined later) obtained by a student in his first year examination at the University was taken as a measure of students performance. Treating this as the dependent variable, we shall test whether each of the following have any impact on Y.

1. The facilities that were available at the student's pre-University school
2. Sex
3. Amount of tuition obtained by the student at G.C.E. (A level)
4. Family income of the student
5. Student's performance at G.C.E. (A level)
6. Student's course of studies

To facilitate this type of analysis it is customary to formulate a linear regression model of the form,

$$Y' = a_0 + a_1 X_1 + a_2 X_2 + a_3 X_3 + a_4 X_4 + a_5 X_5 + a_6 X_6 + a_7 X_7 \quad (3.1)$$

Where a's are parameters to be estim ted and X's are the following variables of interest

3.1. Definition of Variables

SCHOOL GRADES : A student is categorized to have done his Pre-University studies in a Grade 1, Grade 2 school or Grade 3 school according as in the opinion of the student, facilities available in his school were excellent, sufficient or insufficient. This classification leads us to treat X_1 and X_2 as dummy variables defined by

$$X_1 = \begin{cases} 1 \text{ if the student is from a Grade 1 school} \\ 0 \text{ otherwise} \end{cases}$$

$$\text{and } X_2 = \begin{cases} 1 \text{ if the student is from a Grade 2 school} \\ 0 \text{ otherwise} \end{cases}$$

so that fractional attributions of Y corresponding to Grade 1, Grade 2 and Grade 3 schools are of the form $X_1 + G$, $X_2 + G$ and G respectively, where G is an arbitrary constant which can set to be zero for comparing purposes.

SEX : X_3 is the sex dummy defined as

$$X_3 = \begin{cases} 1 \text{ if the student is a male} \\ 0 \text{ if the student is a female} \end{cases}$$

Then the incremental attribution to a male relative to a female is X_3.

TUITION : To minimise the work involved we do not wish to carry out separate analyses for separate subjects. So we shall try to give a meaning to the whole problem in terms of 'averages'. Accordingly the underlying variable here we define to be,

X_4—average number of hours spent on tuition per subject during the period of G.C.E. (A. level) studies.

This definition gives rise to the formula

$$\times_4 = \sum_{i=1}^{4} {}^{4}r_i \, n_i / 4 = \sum_{i=1}^{4} r_i \, n_i$$

where r_i — number of tuition hours obtained for i^{th} subject per week
n_i — total number of months for which the student sought tuition for i^{th} subject

INCOME : Define the income variable as

X_5 = student's family income in rupees

Here family income is the sum of all incomes received by student's father, mother, all unmarried brothers and sisters and himself.

GCE SCORE : The exact marks obtained by students for each subject was not accessible and only the grades were available. Anyhow what we really seek is the use of a quantitative variable and so we propose to give a mark to each of the grades A, B, C, S and F on basis of the frequency distribution of the available classified data in which class limits are defined by the known ranges of A, B, C, S and F. Then the median of each class evaluated using the smooth ed curve of the histogram is taken as the mark of the corresponding grade. Moreover, in order to be consistent with our plan of confining to averages, here we take the average of the marks that we have given to a student's four grades he obtained at G.C.E. (A. Level) examination as his GCE score. Hence

X_6 — score obtained by the student at G.C.E. (A. Level).

COURSE : X_7 is the subjects stream dummy defined by

$$X_7 = \begin{cases} 1 & \text{for a Bio science student} \\ 0 & \text{for a Physical science student} \end{cases}$$

This dummy variable will enable us to test whether, one of these two groups of students are at an advantage, by examining the sign of the coefficient of X_7.

PERFORMANCE : This is the dependent variable, the variation of which we are trying attribute to the explanatory variables outlined above. The performance of a student at the University examinations are measured in the same manner as the GCE score was defined. Although one should really take all three year ending examinations into account to accomplish our task owing to lack of facilities that were available to conduct a survey of the students who have passed out as the subjects, we were compelled to base our analysis on the findings of the said survey which revealed in particular the grades obtained by the students at the 1978 first year examination alone. The university grading system A, B, C, D and E which characterizes the classification of marks in this context were treated with scores by the same criterion as before. To simplify the analysis again university scores were computed in respect of each student by taking the average of the marks obtained in the three subjects. Hence,

Y — score obtained by the student at university.

3.2. Estimation of Model and Conclusions

The proposed linear model (3.1) was estimated by the method of least squares using the data given in Table (3.1). The estimated model with respective standard errors appearing within brackets is as follows,

$$\begin{aligned} Y = {} & 48.9105 + 4.4421\,X_1 - .1008\,X_2 + 1.1226X_3 \\ & (10.3765) \quad (2.7924) \quad (2.1580) \quad (1.8048) \\ & +.0120\,X_4 - .0022\,X_5 - .0123\,X_6 - 3.1186X_7 \quad (3.2) \\ & (.0122) \quad (.0012) \quad (.2235) \quad (1.8911) \end{aligned}$$

TABLE 3.1

Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7
47.67	0	1	0	216	0800	45.38	1
47.67	0	1	0	000	0250	39.00	1
51.00	0	1	0	003	0400	42.25	1
51.83	0	1	1	012	1600	44.88	1
35.50	0	1	0	006	0900	48.36	1
44.00	0	1	1	000	0500	42.25	1
39.83	0	0	1	120	0940	48.25	1
43.50	0	1	0	036	9000	39.75	1
39.50	0	1	0	048	1650	42.24	1
35.33	0	1	0	048	3720	42.24	1
43.50	0	1	0	000	0500	42.38	1
43.83	0	0	1	216	0650	37.13	1
47.67	0	1	0	132	3000	44.88	1
43.33	0	0	0	090	1600	44.88	1
51.83	0	1	1	015	0100	41.50	1
45.50	0	1	0	000	0750	46.50	1
47.67	0	1	0	000	0600	45.38	1
47.67	0	0	0	432	0600	42.88	1
47.67	0	1	0	006	0500	40.25	1
47.67	0	1	1	024	0200	44.88	1
47.67	0	1	0	000	0800	42.25	1
44.00	0	0	1	000	0900	42.00	1
51.67	0	1	0	036	0300	44.88	1
39.50	1	0	0	036	1250	44.88	1
51.00	1	0	1	000	1000	44.88	1
39.33	0	0	1	096	0100	44.88	1
47.67	0	0	0	144	0850	37.13	1
43.50	1	0	1	000	1500	42.25	1
35.67	0	1	0	000	0600	42.25	1
46.83	0	1	0	032	0500	40.25	1
31.67	0	1	1	000	2000	37.13	1
43.83	0	1	0	068	0450	40.25	1
51.67	0	1	0	024	0850	31.13	1
39.33	0	1	1	000	0200	37.13	1
55.83	1	0	0	072	200	44.88	1
55.53	0	0	1	072	0225	53.00	0
43.67	1	0	0	024	0900	45.75	0
47.33	0	0	0	192	0300	38.38	0
51.83	0	1	0	120	0750	42.25	0

Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7
47.33	0	0	0	048	0450	42.25	0
55.67	0	0	1	084	1000	41.63	1
40.00	0	0	1	000	0700	48.63	0
43.67	0	1	1	054	0800	45.75	0
40.00	0	0	1	096	0350	45.25	0
51.83	0	1	1	000	0200	50.00	0
37.17	0	1	0	063	0500	45.75	0
51.83	0	0	0	192	0300	56.63	0
51.83	0	1	1	024	1100	50.00	0
48.17	0	0	1	018	2100	50.00	0
47.33	0	0	0	056	1800	42.25	0
40.00	0	1	0	153	1500	45.63	0
55.67	1	0	0	144	1000	42.25	0
56.83	0	0	1	120	0800	49.63	0
43.77	0	1	1	136	0300	42.25	0
46.60	1	0	1	000	0400	48.63	0
50.30	0	0	1	120	1100	49.00	0
57.67	0	0	1	144	0900	38.38	0
41.67	0	1	0	072	0850	35.75	0
51.83	0	1	1	052	0250	52.38	0
67 83	1	0	1	048	0500	46.63	0

Multiple correlation coefficient of the fit is .47 thus indicating the regression explains only 22% of the Y variation. It is easily seen that none of the explanatory variables are significant at .05 level implying that factors outlined in Section 3.1 exert no significant influence on Y. However tests at .1 level of significance leads us to make the following conclusions positively.

1. Students family income has an adverse effect on their performances at the university, i.e. on average students from poor families usually are better in university studies.

2. At Vidyodaya the physical science students compared with Bio-science students, enjoy an advantageous position induced by their combination of subjects; i.e. physical science students usually do better than Bio science students at this university provided that they are identical in every other respect.

3. The factors such as sex, standard of pre-university school, pre-university tuition and in particular students' performance at G.C.E (A. Level) do not contribute to his or her performance at the university.

4. Discussion

In this study mainly we have made three important observations that may be of particular use to decision makers of the university governing body. It had been noticed even in the past by some university lecturers that some of those students who had done exceptionally well at the G.C.E. (A. Level) fail to keep up their position when they come to the university. When this downfall cannot be attributed to any conceivable cause, it is argued that he or she was not really a best student even at G.C.E. (A. Level) in spite of his or her outstanding

performance tnen. Of course no bad student can pass the university entrance barriei. The point is that non-identical facilities that students enjoy at high school and various amounts of tuition they seek, upset the proper rank the students' reallv deserve. As far as we know, this is tne first time that the interested hypothesis was tested systematically. At this point one may argue that the variation of X_6 variable, which ranges only from 31.13 to 53.00, is not quite enough for Y to pick its influence. Yet it must be pointed out that, there is no sign of X_6 exerting a positive effect on Y, for with tne prevailing variation the coefficient of X_6 is negative. Hence the objections, to the standardizing systems, based on no systematic study but mere imaginations and speculations has to be rejected outright. The prevailing university admission criteria are further justified when we found out in Section 2.2 that number of students getting university admission from each district is roughly proportional to the number of students who sat for the examination from the same district.

The second interesting observation that we have made is that, at .1 level of significance, tne family income of a student has an adverse effect on his or her performance at the university. The hypothesis of no family income effect was rejected even at .05 level when the Y variable was redefined as the average score of just Physics and Chemistry, the two common subjects done by both Bio and Physical groups at the G.C.E. (A. Level) (see Fernando (1978)).

Thirdly our survey has revealed that at Vidyodaya, on average physical science students score more at examinations compared with Bio science students (even at .05 level of significance when Y is redefined as in the last paragraph). Tne way we have redefined Y, interestingly enough, one cannot attribute this effect to the deficiency of Bio science students in Mathematics. In fact Fernando (1978) has shown that, as is our next best guess, they suffer this disadvantageous position because of their deficiency in Physics rather. It is shown that the mean score obtained by a physical science student in physics is 57.9 (variance 3.5) whereas that of a Bio science student is only 45.9 (variance = 4.9) which is significantly smaller than the former.

Reference

Fernando (1978)—"*A Study on University and Pre-University Educational Background of the Science Undergraduates of Sri Lanka*", Diploma in Statistics Dissertation, Dept. of Mathematics, University of Sri Jayawardenapura, Nugegoda.