Sindh Univ. Res. Jour. (Sci. Ser.) Vol. 48 (3) 487-490 (2016)

SINDH UNIVERSITY RESEARCH JOURNAL (SCIENCE SERIES)

Comparative Study of Factors Influencing the Performance of Students in Statistics Using Multiple Linear Regression Approach

N. A. QURESHI, W. A. MAGSI, V. SUTHAR, M. H. LAKHO, A. BUKERO* B. K. SOLANGI*

Department of Statistics, Sindh Agriculture University Tandojam, Pakistan

Received 8th May 2015 and Revised 05th October 2015

Abstract: In the present research an attempt is made to explore the factors that significantly influence the student's performance in the Statistics (STAT – 511) course offered by the Faculty of Crop Production (FCPD), Sindh Agriculture University Tandojam. These factors were explored with the help of descriptive and inferential statistics. The results of the present research show that on the basis of t-test and f-test, it was inferred that only two factors were found to be significant i.e. residence (village versus town / city) of students and their marks at intermediate level, measured in grades (A+, A, B, C or D). It was concluded that higher the grade in intermediate, the more percentage of marks in STAT-511 subject had been reported. Likewise, multiple linear regression proposed a model with the significant factors, namely: (1) study hours per day, (2) residence of students (village versus town/city), (3) hostler/ non hostler, and (4) marital status (single versus married).

Keywords: Crop Production (FCPD), Sindh Agriculture University Tandojam

1. INTRODUCTION

Education has long been concern of human beings, which has received more or less automatic attention. It is only recently in human history that education has been regarded as an "institution" which cross all cultural boundaries. Even as late as hundred years ago, education was delivered to some and not to others. Education is more than what we learn at school, home and though our experience of life. Education is a tool to help us to make sense of this experience.

Everyone has the right to get education. Education shall be free at least in the elementary and fundamental stages. Education is a key to development. Through education can: reduce poverty and inequality, give people power to take charge of their own lives, bring about change. Now the question arises that what stops people having access to education? To find the answer for this question, the following factors were found to be responsible for access to education:

Yang and Tang (2004) reported that over 125 million children were not enrolled in school; two third of these are girls; this is through lack of opportunity not through any fault of their own. Education costs money for teachers, building and resources like books and computers. Even when education is provided free by the state, there are still costs such as uniform, books, exam fees and transport. Families living in poverty have to make difficult choices. They have to weight up the benefits of education against other needs such as food and healthcare. If the adult in the family cannot earn enough to live on, children may have to work rather than to go school.

Many of the world's poorer countries are heavily in debt. They have had to cut the money they spend on education, health and social services to pay the interest on their dept. Besides, there are several other factors that are affecting the education. Like mathematics, statistics is considered as a scoring subject for the students whose mathematics is pretty good. On the other hand, getting passing marks for some students is a big challenge. In the real situation, there are many factors that influence on the concentration of the student's study. The purpose of this study was to assess the factors which significantly affecting the student's marks in STAT-511 subject using t-test, f-test, ANOVA and multiple linear regression.

2. MATERIALS AND METHODS

In this study, factors assumed to be affecting the performance of students in STAT-511 subject were examined. A questionnaire was developed to explore the factors affecting student's performance. Many important factors for instance, marital status, residence, domicile, hostler or not, study hours per day, father' education and occupation, mother's education and occupation were included in the questionnaire.

2.1 Data collection

The data were collected from the student of the Faculty of Crop Production (FCPD). Lists of students of B.Sc. (AGRI.)Hons. Part-IV for the academic year 2003-2004 was obtained from Deputy Registrar (Admission),

++Corresponding author email: qureshistat@gmail.com

*Department of Entomology, Sindh Agriculture University Tandojam, Pakistan

Sindh Agriculture University Tandojam. There were total 201 students registered in FCPD. Out of them, 70 students were randomly selected and interviewed. The Statistical Package for Social Science (SPSS) was used to analyze the data.

2.2 Variables created and labeled

Prior to enter the data, variables were created and labeled. The qualitative variables namely marital status, residence, domicile, hostler or non- hostler, father's education, father's occupation, mother's education and mother's occupation etc, were changed into quantitative variables by coding their values. For instance, for marital status, single was coded as 0 and married was coded as 1.

2.3 Statistical analysis

Descriptive and inferential statistics had been used to analyze the data.

2.3.1 Descriptive Statistics Mean

The arithmetic mean is defined as a value obtained by dividing the sum of all the observation by their number (Chaudhry and Kamal, 2012). The sample mean of a set of n observations $x_1, x_2, x_3, \ldots, x_n$ is defined as follows:

$$\overline{X} = \frac{X_i + X_2 + X_3 \ldots\ldots\ldots\ldots\ldots + X_n}{n} = \frac{\sum X_i}{n} \quad (i = 1, 2, 3, \ldots\ldots\ldots, n)$$

It should be noted that the mean represents a point at which the distribution or the set of observations would balance.

Standard Deviation

The standard deviation which measures the average spread around the mean is defined as the positive square root of the variance (Wilcox, 2009). Symbolically,

$$S = \sqrt{\frac{\sum (X_i - \overline{X})^2}{n - 1}}$$

2.3.2 Inferential statistics t-test

t- test is used to test the hypothesis that means of two groups are significantly different (Healy, 2012). The formula for t-test is given as under:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{SE\left(\overline{X}_1 - \overline{X}_2\right)}$$

f-test

f-test is used to test the hypothesis that means of more than two groups are significantly different from each other. In case the f- value is found to be significant, Duncan multiple range test (DMRT) is used to separate means that which mean is significantly different from others (Healy, 2012).

Multiple Regressions

Multiple regressions is a statistical technique in which more than one independent variables are involved in order to predict the dependent variable (Y). Multiple regression analysis is also highly useful in experimental situations where the experiment can control the independent variables. In general, the variables $X_1, \ldots\ldots, X_{p-i}$ in a regression model have to represent different independent variables, therefore the definition of general linear regression model, with normal error terms, simply in terms of X variable is (Groeber *et.al* 1993):

$$Y_i = \beta_o + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots\ldots + \beta_{p-1} X_{i,p-i} + \varepsilon_i$$

Where:

β_o is Y- intercept,

$\beta_1, \beta_2, \beta_3, \beta_p$ are the parameters of the regression equation,

ε_i is a random error in Y for observations i

3. RESULTS AND DISCUSSION

Students data pertaining to the marks in statistics course, STAT-511 and factors assumed to be related to the performance of students in statistics were analyzed, summarized, and discussed as follows. It should be noted that only the results for those factors which were found significant, are reported here.

3.1 Residence

Table 1 reveals that majority (55.72%) of the students who belong to city /town got 62.41 percentage marks in statistics theory while 44.28 percent of students interviewed were from villages and they got 50.23 percent marks. Because the p-value of t-value was calculated to be 0.008, which is less than the level of significance, therefore the difference in means is significant.

Table-1 Descriptive statistics of student's average marks in STAT-511 by residence

Resident of student	N	Percent	Mean (Marks)	Std. Deviation	t-value	p-value
Village	31	44.28	50.23	29.933	-1.75	.008
Town/City	39	55.72	62.41	27.968		
total	70	100.00	57.01	29.283		

3.2 Intermediate grade

Table 2 is evident of the effect of intermediate grade on the performance of student in statistics theory. It can be seen that 10.00 percent of the total students who passed intermediate by securing A grade, scored 69.86 percent marks while 52.86 percent of the student s who got B grade in H.S.C, their average marks in statistics were recorded to be 63.76 percent where as the remaining 37.14 percent of the total students interviewed had C grade in intermediate and their average marks were

calculated to be 43.96 percent. Since the p-value of the f-value (4.690) was calculated to be 0.012, which is less than the level of significance, therefore, it was calculate that differences in means are significant. From this it was concluded that the better grade in intermediate, the more marks in statistics.

Table-2 Descriptive statistics of the students' average marks in statistics by intermediate grade

Grade in intermediate	N	Percent	Mean(Marks)	Sid. Deviation	f-value	p-value
A	7	10.00	69.86	19.735	4.690	0.012
B	37	52.86	63.76	27.771		
C	26	37.14	43.96	29.529		
Total	70	100.00	57.01	29.283		

3.3 Regression model

The multiple linear regression model for the significant factors affecting the performance of students in STAT-511 subject is given as under:

$$Y_i = 24.149 + 5.254\, X_1 + 20.685\, X_2 + 23.30\, X_3 - 29.67 X_4$$

In the above mentioned regression model, Y_i is the dependent variable that stands for marks of students in STAT-511 subject 24.149 is the Y-intercept of the regression equation. Similarly, the estimates for β_1, β_2, β_3, and β_4 are found to be 5.254, 20.685, 23.30, and -29.67 respectively. X_1 stands for study hours per day, X_2 for residence of student, X_3 for student belong to hostel or not, and X_4 stands for the marital status of student. Table 3 shows that 37.9 percent variation in the independent variable was explained by the significant independent variables as reported above while 62.1 percent variation remained unexplained

Table-3 Summary statistics of regression model of students marks in STAT-511

R	R Square	Adjusted R Square	Std. Error of the estimate
0.616(e)	0.379	0.316	24.985

Information regarding co-efficient of the significant factors has been summarized in Table 4. The table shows that p-value of the coefficient is less than 0.05 level of significance; it was concluded that the reported factors have significant effect on student's marks in STAT-511.

Table-4 Coefficient of regression model of statistics, marks in STAT-511

Model	Unstandardized Coefficients		Standardized Coefficients	T	Sig.
	B	Std. Error	Beta		
(Constant)	24.149	24.057		1.004	.320
Study hours per day	5.254	2.040	.304	2.576	.013
Residence of student	20.685	7.052	.339	2.933	.005
Student belongs to hostel or not	23.300	7.697	.354	3.027	.004
Marital status of student	-29.670	12.001	-.285	-2.472	.017

Table-5 ANOVA table for multiple linear regression models for students marks in STAT-511

Source of Variation	Degree of freedom	Sum of Square	Mean square	F-value	Sig
Regression	5	18678.422	3735.684	5.984	.000
Residual	49	30587.287	624.230	---	---
Total	54	49265.709	---	---	---

Likewise, Table 5 shows the Analysis of Variance (ANOVA) results for the fitted regression model. The calculated f-value 5.98 and its corresponding p-value 0.000 shows that the factors (study hours per day, residence of students, hosteller/not, marital status of student and name of faculty) in the linear model have done a good job in explaining the variation in the department variable.

4. CONCULION

In this study, an attempt was made to explore the factors that were significantly affecting the performance of students in statistics course STAT-511. Using averages, it was concluded that the single students got more marks than the married ones. Similarly villagers secured less percent marks while town / city dwellers obtained more percent marks; hostler got 47.50 percent marks whereas non-hostlers' marks were 60.82 percent. Using t-test, it was inferred that only one factor was found to be significant i.e. residence (village versus town / city) of students. Using f-test, grade in intermediate was found to be significant. The higher the grade in intermediate, the more percentage of marks in STAT-511 subject had been reported. Multiple linear regression proposed a model with the significant factors,
namely: (1) study hours per day, (2) residence of students (village versus town/city), (3) hostler/ non hostler, and (4) marital status (single versus married).

REFERENCES:

Chaudhry, S. M., and S. Kamal (2012). Introduction to statistical theory. Part-I. 8th ed. Ilmi Kitab Khana, Urdu bazaar Lahore, Pakistan.

Groebner, D. F., P. W. Shannon, P. C. Fry, and K. D. Smith. (2013). Business Statistics, A decision making approach. 8th ed. Prentice Hall, Upper Saddle River, New Jersey.

Healy, J. F., (2012). The Essentials of Statistics: A Tool for Social Research. Wadsworth Publishing Company, California.

Wilocx, R. R., (2009). Basic Statistics: Understanding Conventional Methods and Modern Insights. 1st ed. Oxford University Press, USA.

Yang, H. L., and J. H. Tang, (2004). World Report on Violence against Children. UNICEF, 29-37.