THE EFFECT OF TYPE OF SCHOOL ON ACADEMIC PERFORMANCE OF SCIENCE FRESHMAN STUDENTS OF AAU

BY

AWRARIS AGIDEW

SCIENCE FACULTY

ADDIS ABABA UNIVERSITY

ADDIS ABABA

JUNE, 1998

THE EFFECT OF TYPE OF SCHOOL ON ACADEMIC PERFORMANCE OF SCIENCE FRESHMAN STUDENTS OF AAU

BY
Awraris Agidew

ADVISOR
Ato Gemechis Dilba

A dissertation submitted to the Addis Ababa University in partial ful filment of the requirement for the award of degree of:
BACHELOR OF SCIENCE
in
STATISTICS

Department of Statistics
Science Faculty
Addis Ababa University
Addis Ababa

June, 1998

TABLE OF CONTENTS

ACKNOWLEDGMENT

I am heartily indebted to my advisor Ato Gemechis Dilba, for his invaluable, honest and warm guidance in the accomplishment of this paper. It is through his careful assistance and constructive suggestions that I succeeding in completing my work.

I would like to express my gratitude to Ato Mebrahtu G/Mariam (freshman advisor) and Ato Yared Seyoum (worker of Ministry of Education) for providing me the data and all necessary information for my work.

My special appreciation also goes to Yerukmisrake Afework who patiently and carefully typed this paper.

Last but not least I would like to express my deepest gratitude to all my family, specially to my brother Ato Tadesse Bamnew, with out his continuos assistance and encouragement not only this paper but also my academic life would not be possible.

Awraris Agidew
Thank you!

CHAPTER I

INTRODUCTION

1.1 General

Higher education is the base for economic, social and political development of a country. This can be seen from the experience of developed countries. One can say that all developed countries reach the present stage of their development because they give high attention for the development and improvement of higher education. In other words this is to say, with out higher education a country can only be under dream of development.

Any developing country expect more from those who acquire higher learning to play vital roles in designing and formulation of appropriate economic policies, in avoiding various social problems, in the establishment of stable political situations etc. To this effect, problems which are thought to affect higher learning should be identified and appropriate solutions must be sought.

Research in education, as in other fields, is a search for knowledge. It is not a search that yields infallible truth, but it is rather a search that provides knowledge for the solution of problems in the field of education. In our country various educational researches have been done in various educational problems. Hence this paper can be considered as a further attempt which is aimed at assessing the difference in academic performance between students who complete their high School studies in Non-Governmental Schools and that of Governmental Schools. It is designed not only for the understanding of the significance of the above stated difference, but also it tries to recommend appropriate solutions that can be used to tackle the problem.

The ministry of Education distinguishes Non-Government Schools to government Schools as:

Government Schools are Schools which are operated by the ministry of Education.

Non-Government Schools are Schools which are operated by authorities other than the ministry of Education.

According to the above definition Non-Governmental schools include community schools, missionary schools, schools operated by churches, private schools, etc.

1.2 Distribution of Governmental and Non-Governmental Schools in Ethiopia

In our country there are 329 senior secondary schools. Out of these 31 schools are Non-Governmental and the remaining 298 schools are Governmental schools. The following table gives the distribution and number of Governmental and Non-Governmental schools by region.

TABLE 1

NUMBER AND DISTRIBUTION OF GOVERNMENTAL AND NON-GOVERNMENTAL SCHOOLS IN ETHIOPIA

REGION	NUMBER OF SCHOOLS		
	N.G*	G**	Total
Tigray	3	18	21
Afar	-	3	3
Amhara	-	76	76
Oromiya	7	102	109
Somalie	-	3	3
Benushangul-Gumuz	-	5	5
SNNP	1	61	62
Gambella	-	4	4
Harari	-	2	2
Addis Ababa	19	23	42
Dire Dawa	1	1	2
Total	31	298	329

Source:- The Federal Democratic Republic Of Ethiopia Education Statistics Annual Abstract (1987 E.C)

* N.G - Non-Governmental

** G - Governmental

As can be seen from table 1 the number of Non -Governmental schools is very limited as compared to Governmental schools. Also Non- Governmental schools are characterized by regional concentration .i.e. Non-Governmental schools are almost concentrated in Addis Ababa and Oromiya regions.

It is believed that students who complete their high schools studies in Non-Governmental schools perform better than those from governmental schools

The concentration of the Non-Governmental schools, which are thought to be the origin of students with better performance create unbalanced distribution of students acquired higher learning. This unbalanced distribution also creates a huge gap in the development of the regions.

1.3 Objective of the study

There are different factors that affect the academic performance of students in universities and colleges. One amongst these is the type of school, that is, the impact of the difference in type of schools from which the students come. The main objective of this paper is to compare the academic performance of Science Freshman students who completed their high school in Governmental schools and Non-Governmental school. To meet this the students are classified as Governmental and Non-Governmental.

1.4 Overview of the study

The data collection methods and the procedure of the sample selection is presented in Chapter 2. The collected data is also summarized using frequency distribution tables and histograms.

The major part of any research is the analysis of the data. In this undertaking, before analyzing the data the method of analyzing the data is discussed in Chapter 3. i.e the method of estimating population parameters, the method of testing various hypothesises etc. are discussed.

Based on the methods discussed in Chapter 3, analysis of the data collected for this research is done in Chapter 4. Various estimates of population paramete are given. Moreover hypothesises are tested based on these estimates. Finally, conclusions and recommendations are given, in the last chapter, Chapter 5.

CHAPTER II

DATA COLLECTION METHODS AND SAMPLE SELECTION

2.1 Introduction

Data can be defined as facts or information used in deciding or discussing something. In any statistical investigation data is the base to reach at conclusion and to say something about the investigation.

Data collection is essentially an important part of any research process. It is only from the collected data that inferences, hypothesis or generalizations held may be identified as valid, verified as correct, or rejected as unattainable. In order to collect the requisite data the researcher has either to entertain the entire population or to sample the population concerned. In this research we sample from the population under consideration.

2.2 Type and Source of the Data

Depending on the source from which the data is obtained, data can be classified in to two, namely, primary data and secondary data.

- **Primary Data**:- is defined as data collected by the investigator himself for the first time.

- **Secondary Data**:- is referred to the data collected by another investigator or data obtained from already compiled files.

The data to be analysed in this undertaking is obtained from the records of Science Freshman Program Office, hence the data is a secondary data.

2.3 Variables of Interest

Most of the educational phenomena consists of a large number of factors. It would be impracticable to test, or observe the variables of interest under controlled condtions. This is to say that, it is not possible to control each and every factory that can affect the variables of interest, inorder to study the variables of interest alone. Hence it is necessary to assume factors, other than we are interested, that can affect the variables of interest are held constant.

In this undertaking we assess the overall academic performance of the students under consideration based on their GPA. Further, to assess in which particular subjects difference in performance between the two group students is observed, we consider the five subjects, namely Biology, Chemistry, English, Mathematics and Physics.

Therefore, the variables of interest of the paper are the GPA and grades of the five subjects. In this undertaking other factors other than the difference in type of school such as age, sex, region etc. are assumed constant.

2.4 The Sampling Procedure

A sample is defined as any subset of a population selected to obtain some information concerning some characteristics of the population. Sample should be selected

in such a way that it will represent the population i.e. it should be selected using appropriate method of sampling.

Sampling may be defined as the process of taking sample and making inferences to the whole population. Sampling procedures provide generalization on the basis a relatively small portion of the population. To obtain a representative sample, the researcher should select each unit in a specified way under controlled conditions. Usually four steps are involved in the process:

A. Defining the population
B. Listing the population
C. Selecting a representative sample
D. Obtaining an adequate sample

A. Defining the population

The population is properly defined so that there is no ambiguity as to whether a given unit belongs to the population. The population of interest in this undertaking is the Freshman Students of the Science Faculty. The population of Freshman Students divided in to two sub populations, namely, students from Non-Governmental and students from Governmental schools. For this classification, information is obtained from the ministry of education.

B. Listing the population

Inorder to select a sample from a given population, it is also necessary to have a complete, accurate, and uptodate list of all the units in the population. Such a list is known as sampling frame. If a sampling frame is not readily available at the time of investigation,

in such a way that it will represent the population i.e. it should be selected using appropriate method of sampling.

Sampling may be defined as the process of taking sample and making inferences to the whole population. Sampling is a process of

The population is properly defined so that there is no ambiguity as to whether a given unit belongs to the population. The population of interest in this undertaking is the Freshman Students of the Science Faculty. The population of Freshman Students divided in to two sub populations, namely, students from Non-Governmental and students from Governmental schools. For this classification, information is obtained from the ministry of education.

B. Listing the population

Inorder to select a sample from a given population, it is also necessary to have a complete, accurate, and uptodate list of all the units in the population. Such a list is known as sampling frame. If a sampling frame is not readily available at the time of investigation,

it is essential to prepare a frame before the samples are drown. For this undertaking appropriate sampling frames are prepared for the two populations.

C. Selecting a representative sample

After defining a population and listing all the units, a researcher should select a sample of units from the sampling frame. In this investigation to select representative samples from both populations, we make use of a computer program which is designed to generate random numbers with out repetition.

D. Obtaining an adequate sample

If the population under study is homogeneous, a small sample is sufficient. On the other hand, a much larger sample is necessary if there is greater variability in the units of the population. But increasing the size of the sample is of little value, if units are not chosen in a way that ensures representativeness of the population. In this investigation the sizes of the samples are determined using the variability in each population. Details will be given later.

2. 4. 1 The method of sampling

Based on the procedure in which the sample is selected, the sampling plan used for the selection of the sample can be classified in to two:

- Probability sampling, and
- Non probability sampling.

A sample is called a *probability sample* when the method of its selection is based on the theory of probability i.e. in probability sampling every unit of the population has a calculable (known non-zero) probability of being selected in the sample. Probability

samples provide valid estimates of unknown population values with their measures of reliability by indicating the extent of error due to sampling. But with non -probability *sampling*, it is not possible to measure the degree of reliability of the sample results.

Examples of *probability sampling* are:-

- simple random sampling
- stratified sampling
- systematic sampling
- cluster sampling

Examples of *non-probability sampling* are:-

- sample of volunter subjects
- expert choice (purposive or judgment sampling)
- quota sampling
- sampling of mobile population.

In this undertaking we use probability sampling.

2.5 The Sampling Technique Used

Simple random sampling is a basic probability selection scheme in which a predetermined number of units from a population list is selected so that each unit on that list has an equal chance of being included in the sample i.e. each possible sample of n units from a population of N units has an equal chance of being selected which also implies that every member of the population has an equal chance of selection in to the sample.

As stated clearly in the objective, in this paper there are two population to be compared, namely, students from Non-Government and Government schools. The sampling technique we are going to use is simple random sampling . First the entire population is

classified in to the two sub-populations mentioned in the pervious section of this chapter, and then a random sample of appropriate size will be drawn from each population.

2.6 Sample Size Determination

The question " What should be the sample size? " confronts investigators frequently. Which means in any sample survey a researcher reaches at a stage which requires a decision on the sample size. This is because the total cost of the survey and the precision of the main estimates are related to the sample size.

The first step in the decision of the sample size is to state the permissible margin of error (or limit of tolerance) in the estimates and a confidence limit in which the estimates to lie.

To put the above paragraph in another way, if we are interested to estimate the universe mean , μ , the acceptable risk that the universe mean will lie outside the limits $\bar{y} \pm d$ is α, where $\bar{y}$ is the sample mean.

The standard error of y is

$$s.e\,(\bar{y}) = \frac{s}{\sqrt{n}}\sqrt{\frac{N-n}{N}}$$

where s - Advance estimate of the population standard deviation

N - The population size

n - The sample size

The (1 -α) percent probability limits of the universe mean , μ , are

$$\bar{y} \pm Z\,\alpha/_2 \;\; s.e\,(\bar{y})$$

where $Z\,\alpha/_2$ is the value of the standard normal deviate that cuts off an area of $\alpha/_2$ on the right tail.

Setting the permissible margin of error:

$d = Z\,\alpha/_2$ s.e $(\bar{y})$ and solving for n, we get

$$n = \frac{\left(\frac{z_{\alpha/2}S}{d}\right)^2}{1+\frac{1}{N}\left(\frac{z_{\alpha/2}S}{d}\right)^2}$$

$n_0 = \left(\frac{z_{\alpha/2}S}{d}\right)^2$ will be taken as the first approximation

If n_o/N is negligible (i.e. if $n_o/N \leq 0.05$) then n_o can be taken as the required sample size. If not n is calculated as

$$n = \frac{n_0}{1+\frac{n_0}{N}}$$

Another important thing one has to do in the determination of the sample size is to obtain an advance estimate of the population standard deviation. This can be done in four possible ways.

i. From the results of pilot survey

ii. By previous sampling of the same or similar population.

iii. By taking sample in two steps, the first being a simple random sample of size n_1 from which s^2, estimate of S^2, will be obtained.

iv. By guess work about the structure of the population, assisted by some mathematical results.

The method used to estimate S^2, in this survey is the third method, taking 10 percent of each of the two populations under consideration.

For this survey, a confidence limit and the permissible margin of error ,d , are decided to be 95% and 10%, respectively, for both populations. For 95% confidence limit the normal table provides the $Z\,\alpha/_2$ value to be 1.96.

let:- N_1 = size of the first population (students from Non-Government schools)

N_2 = size of the second population (students from Government schools)

n_1 = size of sample from the first population

n_2 = size of sample from the second population

s^2_1 = variance estimate for the first population

s^2_2 = variance estimate for the second population

The total number of freshman students registered in the Science Faculty in the 1997/98 academic year and took the first semester final examination is 1088, of this 113 are from Non -Governmental schools. That is ,N_1=113 and N_2=975.

Taking 10 percent of the students from both populations, s^2_1 and s^2_2 are calculated to be 0.454 and 0.446, respectively. The variances are computed from the Grade Point Average (GPA) for the first semester. In this semester Science students take five basic courses: Biology, Chemistry, English, Mathematics and Physics.

The sample sizes are determined as follows

◆ **For the first population**

$$n_1'(initial\_estimate\_of\_n_1) = \left(\frac{(1.96)(0.674)}{0.1}\right)^2 = 175$$

$n'_1/N_1 = 175/113 = 1.55 > 0.$

$$\therefore \quad n_1 = \frac{n'_1}{1+\frac{n'_1}{N_1}} = \frac{175}{1+\frac{175}{113}} = 69$$

Therefore, we take 69 students from the Non-Government schools.

◆ **For the second population** (students from Government schools)

$$n'_2(initial\_estimate\_of\_n_2) = \left(\frac{(1.96)(0.669)}{0.1}\right)^2 = 172$$

$n'_2/N_2 = 172/975 = 0.176 > 0.05$

$$\therefore n_2 = \frac{n'_2}{1+\frac{n'_2}{N_2}} = \frac{172}{1+\frac{172}{975}} = 146$$

And we take 146 students from the Government schools. The raw data is tabulated in the Appendix.

2.7 Presentation of the Data

In this section the data obtained from the samples drawn from the two populations is summarized using frequency distribution tables and histograms.

Data of the first population

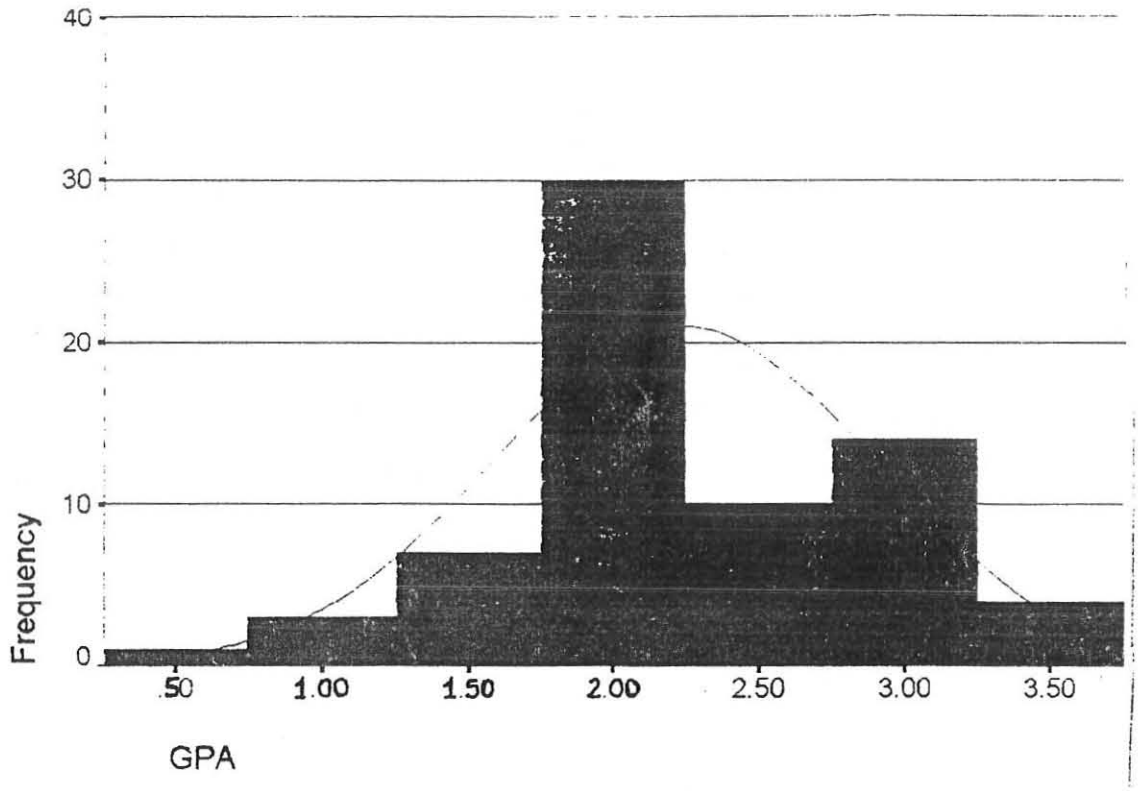
TABLE 2

Frequency Distribution Table

GPA	FREQUENCY	PERCENT	CUM.PERCENT*
0.40	1	1.4	1.4
1.00	1	1.4	2.9
1.20	2	2.9	5.8
1.40	3	4.3	10.1
1.60	4	5.8	15.9
1.80	12	17.4	33.3
2.00	7	10.1	43.5
2.20	11	15.9	59.4
2.40	6	8.7	68.1
2.60	4	5.8	73.9
2.80	5	7.2	81.2
3.00	7	10.1	91.3
3.20	2	2.9	94.2
3.60	4	5.8	100.0
TOTAL	69	100.0	

*** Cum.percent - Cumulative percent**

Histogram of GPA of Non-Governmental School Students.



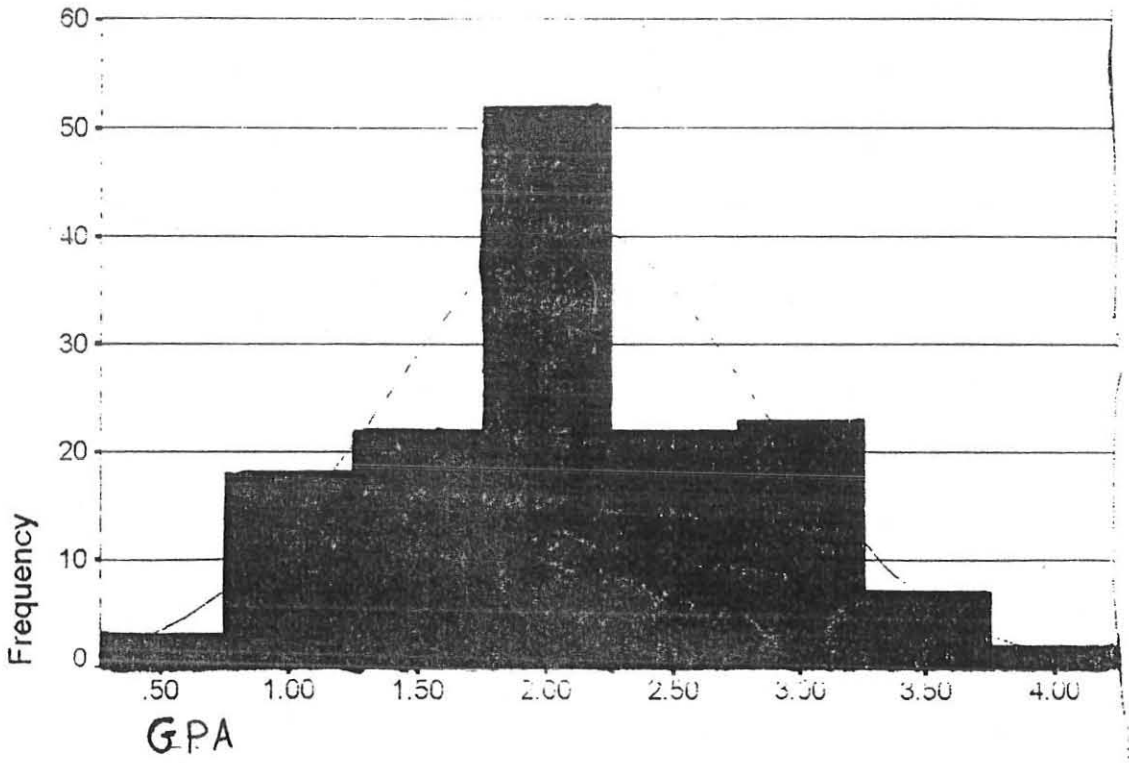
Data of the second population

TABLE 3

Frequency Distribution Table

GPA	FREQUENCY	PERCENT	CUM.PERCENT
0.40	2	1.4	1.4
0.60	1	0.7	2.1
0.80	1	0.7	2.7
1.00	8	5.5	8.2
1.20	9	6.2	14.4
1.40	9	6.2	20.5
1.60	13	8.9	29.5
1.80	20	13.7	43.2
2.00	13	8.9	52.1
2.20	18	12.3	64.4
2.40	11	7.5	71.9
2.60	11	7.5	79.5
2.80	11	7.5	87.0
3.00	7	4.8	91.8
3.20	4	2.7	94.5
3.40	4	2.7	97.3
3.60	2	1.4	98.6
3.80	1	0.7	99.3
4.00	1	0.7	100.0
TOTAL	146	100.0	

Histogram of GPA of Governmental School Students.



CHAPTER III

METHODS OF ANALYSIS

3.1 Parameter Estimation

One of the major objectives of any sample survey is the estimation of unknown population parameter. Estimation can be defined as a procedure of estimating unknown population parameters based on sample observations.

3. 1. 1 Basic Concepts and Definitions

Definitions

Parameter:- a numerical value of some characteristic of the population.

Estimator:- is a function of sample observations that is used to estimate unknown population parameter.

Estimate:- is a value of the estimator which is obtained from a specific random sample.

In estimation of unknown population parameters, one should select best estimators based on certain properties. Among various properties of estimators, the following are widely used for the selection

- unbiasedness
- minimum variance
- consistency

Therefore, the selection should be " select the unbiased estimator which has the minimum variance and consistent."

We can estimate unknown population parameters in two different ways, these are point and interval estimation. In point estimation we look for a specific numerical value of the estimator which is obtained from a random sample. But in interval estimation we look for two functions of a random sample, say L and U, such that (L,U) constructs the confidence interval for the population parameter we want to estimate. L and U, are said to be lower and upper confidence limits respectively.

3. 1. 2 Estimation of Means and Variances

Let $\mu = \frac{\sum_{i=1}^{N} Y_i}{N}$ is the mean of the population.

$\sigma^2 = \frac{\sum_{i=1}^{N} (Y_i - \mu)^2}{N}$ is the variance of the population.

$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$ is the mean of the sample drawn from the population.

$s^2 = \frac{\sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2}{n-1}$ is the variance of the sample drawn from the population.

where N and n are population and sample size respectively.

For a simple random sample, one can show that s^2 and $\bar{y}$ are unbiased point estimates of σ^2 and μ respectively.

The unbiased estimator of the variance of $\bar{y}$ is:

$$V\left(\bar{y}\right) = s^2_{\bar{y}} = \frac{s^2}{n}\left(\frac{N-n}{n}\right) = \frac{s^2}{n}(1-f)$$ where f=n/N is the sampling fraction.

There fore, the standard error of $\bar{y}$ is:

$$S_{\bar{y}} = \sqrt{V\left(\bar{y}\right)} = \frac{s}{\sqrt{n}}\sqrt{1-f}$$

The standard error of $\bar{y}$ supplies information about the amount of error in $\bar{y}$ when used to estimate μ.

Since our sample sizes are large (see sample size determination, chapter two), the central limit theorem tells us than the sample mean is normally distributed with mean μ and variance σ^2 /n. Hence, the lower and upper confidence limits for the population mean are given by:

LCL = $\bar{y}$ -Zα/$_2$ SE ($\bar{y}$) and
UCL= $\bar{y}$ + Zα/$_2$ SE ($\bar{y}$)

Therefore, (LCL, UCL) is the interval estimate of μ.

In this preceding chapter point and interval estimates of μ and point estimate of σ^2 of the two populations under consideration, students from Non-Governmental and Governmental schools, will be calculated using the above methods.

3.2 Hypothesis Testing

A statistical hypothesis is a statement about a population. The procedure or the rule used for checking whether a hypothesis is true or not is called hypothesis testing. There are two types of hypothesis, null hypothesis, denoted by H_o and alternative hypothesis, denoted by H_A.

A null hypothesis is a specific hypothesis about a population that is being tested. A null hypothesis may be accepted or rejected based on the sample results. The hypothesis we accept when we reject the null hypothesis is said to be an alternative hypothesis.

In carrying out any hypothesis testing the following errors can be committed.

i) The null hypothesis, H_o, is true but test concludes it should be rejected: *Type I error.*

ii) The test fails to reject H_o when it is false: *Type II error*.

In general, we wish to minimize the probability of both types of error. It is not possible to completely avoid the error as long as the decision is based on a sample.

Steps in hypothesis testing :

1) State the hypothesis

2) Define the test statistic

3) Specify the level of significance, α

p(Type I error) = α

4) Determine the rejection region. The rejection region is a region in which we reject the null hypothesis. The null hypothesis is rejected if the calculated value of the test statistic is in the rejection region.

3. 2. 1 Comparison of Means

In this section we see the method of comparing the means of two populations. For this procedure we assume that the samples drawn from the two populations are independent.

Let μ_1 - The population mean of the first population

μ_2 - The population mean of the second population

$\overline{X_1}$ - The mean of the sample drawn from the first population

$\overline{X_2}$ - The mean of the sample drawn from the second population

The null hypothesis is that there is no difference between the means of the two populations. we write as

$$H_o: \mu_1 = \mu_2$$

There are three possible alternative hypothesis that can be set based on the objective of the test:

Alternative 1: $H_A : \mu_1 > \mu_2$, the mean of the first population is greater than mean of the second population.

Alternative 2: $H_A: \mu_1 < \mu_2$, the mean of the first population is less than the mean of the second population.

Alternative 3: $H_A: \mu_1 \neq \mu_2$, the mean of the first population is different from the mean of the second population.

In this undertaking we consider the first alternative.

i.e. $H_A: \mu_1 > \mu_2$

The test statistic is

$$Z_{cal} = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{s^2_1}{n_1} + \frac{s^2_2}{n_2}}}$$

Where s^2_1 and s^2_2 and are the variance of the samples from the first and the second populations respectively. The rejection region is:

Reject H_o if $Z_{cal} > Z_\alpha$

3. 2. 2 Comparison of Variability

In this section we see the method of comparing the variability of two independent populations.

Let σ_1^2 - the variance of the first population

σ_2^2 - the variance of the second population

s_1^2 - the variance of the sample drawn from the first population

s_2^2 - the variance of the sample drawn from the second population.

The null hypothesis: the variability in the first population is equal to the variability in the second population.

i.e. $H_o: \sigma_1^2 = \sigma_2^2$

The alternative hypothesis: the variability in the first population is less than the variability in the second population.

i.e. $H_A: \sigma_1^2 < \sigma_2^2$

Under H_o s_1^1/s_2^2 has an F distribution with n_1-1 and n_2-2 degrees of freedom.

Assumptions.

1. $x_{11}, x_{12},..., x_1n_1$ is a random sample from N (μ_1, σ_1^2)
2. $x_{21}, x_{22},..., x_2n_2$ is a random sample from N (μ_2, σ_2^2)
3. The two samples are independent.

The rejection region is : Reject H_o if and only if $F = s_1^2/s_2^2$ is less than F_α (n_1-1, n_2-2).

3.3 Multivariate Analysis

3.3.1 Comparison of Multivariate Means

We want to make inferences about (mean vector of population 1) - (mean vector of population 2). With the following assumptions we can make the desired inference.

Assumptions

1. The sample $X_{11}, X_{12}, \ldots, X_1 n_1$ is a random sample of size n_1 from a p-variate population with mean vector μ_1 and covariance matrix Σ_1.
2. The sample $X_{21}, X_{22}, \ldots, X_2 n_2$ is a random sample of size n_2 from a p-variate population with mean vector μ_2 and covariance matrix Σ_2.
3. The two sample are independent .

We desire a test of the null hypothesis

$H_o: \mu_1 = \mu_2$ Versus

$H_A: \mu_1 \neq \mu_2$

The test statistic is Hotellings, T^2, test

$$T^2 = \frac{n_1 n_2}{n_1 + n_2}\left(\bar{X}_1 - \bar{X}_2\right)'\left(\frac{1}{n_1}\Sigma_1 + \frac{1}{n_2}\Sigma_2\right)^{-1}\left(\bar{X}_1 - \bar{X}_2\right)$$

Since all elements of Σ_i are unknown, we need to estimate these values from the sample drawn in order to test the hypothesis. Hence the test statistic becomes.

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

where $S_{pooled} = [\,(n_1-1)\,S_1 + (n_2-1)S_2]/(n_1+n_2-2)$

The following is the lay out of the computed estimates.

$$\bar{\mathbf{X}}_1 = \begin{bmatrix} \bar{X}_{11} \\ \bar{X}_{12} \\ \bar{X}_{13} \\ \bar{X}_{14} \\ \bar{X}_{15} \end{bmatrix} \qquad \bar{\mathbf{X}}_2 = \begin{bmatrix} \bar{X}_{21} \\ \bar{X}_{22} \\ \bar{X}_{23} \\ \bar{X}_{24} \\ \bar{X}_{25} \end{bmatrix}$$

$$\mathbf{S}_1 = \begin{bmatrix} S_{111} & S_{112} \cdots & S_{115} \\ S_{121} & S_{122} \cdots & S_{125} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ S_{151} & S_{152} \cdots & S_{155} \end{bmatrix} \qquad \mathbf{S}_2 = \begin{bmatrix} S_{211} & S_{212} \cdots & S_{215} \\ S_{221} & S_{222} \cdots & S_{225} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ S_{251} & S_{252} \cdots & S_{255} \end{bmatrix}$$

where

$\mathbf{S}_1$ - Covariance matrix obtained from the sample drawn from population 1.

$\mathbf{S}_2$ - Covariance matrix obtained from the sample drawn from population 2.

$\bar{\mathbf{X}}_1$- Vector of means obtained from the sample drawn from population 1.

$\bar{\mathbf{X}}_2$- Vector of means obtained from the sample drawn from population 2.

and

$$s_{hik} = \frac{1}{n_h} \sum_{j=1}^{n_h} \left(x_{ij} - \bar{x}_{hi} \right) \left(x_{kj} - \bar{x}_{hk} \right)$$

Under H_o, the quantity

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

has an F distribution with degrees of freedom p and n_1+n_2 -p-1, where p is the number of variables.

The rejection region is : Reject H_o if

CHAPTER IV

ANALYSIS OF THE DATA AND RESULTS

In the pervious chapter we have seen how to find estimates of unknown population parameters from sample values and the method of hypothesis testing. In this chapter we use these methods to estimate unknown population parameters and to test certain hypothesises.

Definitions

Population 1 - Freshman students of the Science Faculty of AAU who come from Non-Governmental schools.

Population 2 - Freshman students of the Science Faculty of AAU who come from Governmental schools.

4.1 Estimation of Average Performance of Students from Non-Governmental Schools

Point Estimation

$$\bar{x}_1 = \frac{\sum_{i=1}^{n_1} x_{1i}}{n_1} = 2.269$$

$$s_1^2 = \frac{\sum_{i=1}^{n_1} (x_{1i} - \bar{x}_1)^2}{n_1 - 1} = 0.446$$

$$SE(\bar{x}_1) = \frac{s_1}{\sqrt{n_1}}\sqrt{1-\frac{n_1}{N_1}} = 0.050$$

Therefore, the mean GPA of Non-Governmental school students is 2.269 with standard error of 0.050

Interval Estimation

$$LCL = \bar{x}_1 - Z\alpha/_2 \; SE(\bar{x}_1) = (2.269) - (1.96)(0.050)$$
$$= 2.171$$
$$UCL = \bar{x}_1 + Z\alpha/_2 \; SE(\bar{x}_1) = (2.269) + (1.96)(0.050)$$
$$= 2.367$$

Therefore, the 95% Confidence Interval for μ_1 is (2.171, 2.367), where μ_1 is the mean GPA of population 1.

4.2 Estimation of average Performance of Students from Governmental Schools

Point Estimation

$$\bar{x}_2 = \frac{\sum_{i=1}^{n_2} x_{2i}}{n_2} = 2.083$$

$$s_2^2 = \frac{\sum_{i=1}^{n_2}(x_{2i} - \bar{x}_2)^2}{n_2 - 1}$$

$$SE(\bar{x}_2) = \frac{s_2}{\sqrt{n_2}}\sqrt{1-\frac{n_2}{N_2}} = 0.055$$

There fore the mean GPA of Governmental school students is 2.083 standard error of 0.055.

Interval Estimation

$LCL = \bar{x}_2 - Z\alpha/_2 \ SE(\bar{x}_2) = 2.083 - (1.96)(0.055)$

$= 1.975$

$UCL = \bar{x}_2 + Z\alpha/_2 \ SE(\bar{x}_2) = 2.083 + (1.96)(0.055)$

$= 2.191$

Therefore, the 95% confidence interval for μ_2 is (1.975, 2.191), where μ_2 is the mean GPA of population 2.

4.3 Hypothesis Testing

4. 3. 1 Comparison of Means

In this section we test the hypothesis: H_o: The academic performance of students in the two populations is equal versus H_A: The academic performance of students from Non-Governmental schools is better than that of Governmental school students.

i.e. $H_o: \mu_1 - \mu_2 = 0$

$H_A: \mu_1 - \mu_2 > 0$

$\alpha : 0.05$

For this our test statistics is:

$$Z_{cal} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{2.269 - 2.083}{\sqrt{\frac{0.446}{69} + \frac{0.512}{146}}}$$

$= 1.863$

$Z_{0.05} = 1.65$

Since $Z_{cal} > Z_{0.05}$, we reject H_o: Which implies the average GPA of students from Non-Governmental schools is significantly better than those from Governmental schools.

4. 3. 2 Comparison of Variability

In this section attempts will be on the assessment of variability in the two population. Before proceeding to the analysis there is a need to check the normality assumption made in section 3.2.2. From the histograms drawn in chapter 2, it's reasonable to assume that the two samples are drawn from respective normal populations. We want to test

$H_o: \sigma^2_1 = \sigma^2_2$

$H_A: \sigma^2_1 > \sigma^2_2$

α: 0.05

To carry out this test the appropriate test statistic is

$F_{cal} = s^1_2/s^2_2 = 0.446/0.512 = 0.871$

$F_{0.05} = (68,145) > F_{cal} = 0.875$

Since $F_{cal} < F_{0.05}$ (68,145) we reject H_o. Which implies the variability in academic performance among students from Non-Governmental schools is significantly less than the variability in academic performance among students from Governmental schools. In other words, there is more variability in the mean GPA of students from Governmental schools as compared to those from Non-Governmental schools.

4.4 Multivariate Analysis

In this section we test the equality of the two mean vectors using the methods discussed in chapter 3. The vector of means we want to compare are

$\boldsymbol{\mu}'_1 = (\mu_{11}, \mu_{12}, \mu_{13}, \mu_{14}, \mu_{15})$ and

$\boldsymbol{\mu}'_2 = (\mu_{21}, \mu_{22}, \mu_{23}, \mu_{24}, \mu_{25})$

where μ_{i1} is the population average of Biology in population i

μ_{i2} is the population average of Chemistry in population i

μ_{i3} is the population average of English in population i

μ_{i4} is the population average of Mathematics in population i

and μ_{i5} is the population average of Physics in population i

$i = 1,2$

The following results are computed from the samples drawn

$\bar{\mathbf{X}}'_1 = (2.00 , 2.29 , 3.01 , 2.20 , 1.65)$

$\bar{\mathbf{X}}'_2 = (2.08 , 2.40 , 2.09 , 2.14 , 1.69)$

$$\mathbf{S}_{pooled} = \begin{bmatrix} 0.83 & 0.54 & 0.22 & 0.45 & 0.51 \\ 0.54 & 0.80 & 0.23 & 0.47 & 0.56 \\ 0.22 & 0.23 & 0.64 & 0.19 & 0.20 \\ 0.45 & 0.47 & 0.19 & 0.89 & 0.56 \\ 0.51 & 0.56 & 0.20 & 0.56 & 1.13 \end{bmatrix}$$

$$\mathbf{S}^{-1}{}_{pooled} = \begin{bmatrix} 2.39 & -1.16 & -0.19 & -0.39 & -0.27 \\ -1.16 & 2.82 & -0.31 & -0.48 & -0.59 \\ -0.19 & -0.31 & 1.77 & -0.10 & -0.03 \\ -0.39 & -0.48 & -0.10 & 1.95 & -0.54 \\ -0.27 & -0.59 & -0.03 & -0.54 & 1.57 \end{bmatrix}$$

Using the above values, the observed Hotelling's T^2 value is found to be 76.0027.

But $\dfrac{(n_1+n_2-2)p}{(n_1+n_2-p-1)} F_\alpha(p, n_1+n_2-p-1)$

$$= \frac{(69+146-2)5}{(69+146-5-1)}(2.2582)$$
$$= 11.507$$

Since the value of $F_{0.05}$ (5,209) is 2.2582 after interpolation.

As can be seen the observed T^2 value is greater than 11.507 we reject H_o at 5 percent level of significance. Therefore, we conclude that there is significant difference in the performance of the two student populations.

Although we have rejected the null hypothesis , we still do not know which of the five mean differences may have contributed to the significant T^2 or for which it might be reasonable to conclude that their population means are equal. The Hotelling's simultaneous confidence intervals can be used to test the individual difference. The resulting confidence intervals are given below.

$-0.534 \leq \mu_{11} - \mu_{12} \leq 0.369$ (for Biology)

$-0.570 \leq \mu_{12} - \mu_{22} \leq 0.314$ (for Chemistry)

$0.528 \leq \mu_{13} - \mu_{23} \leq 1.323$ (for English)

$-0.342 \leq \mu_{14} - \mu_{24} \leq 0.590$ (for Maths)

$-0.566 \leq \mu_{15} - \mu_{25} \leq 0.487$ (for Physics)

Since zero is not included in the third interval we conclude, at five percent significance level, that the performance of students from Non-Governmental schools is better than students from Governmental schools in English. But no significant difference is observed in the remaining subjects.

One important point we have to note here is that the overall difference in performance is occurred as a result of the difference in English performance.

CHAPTER V

CONCLUSIONS AND RECOMMENDATIONS

As stated in the beginning of this undertaking, our principal objective is to assess the effect of type of school on academic performance of freshman students of the Science Faculty of AAU.

In chapter III and IV we have discussed the method of analysis and the analysis of the data respectively . Now in this chapter we try to discuss the results and the reasons for these results.

5.1 Conclusions about the Difference in Mean

We have made a test on the overall mean difference of students from the two sub-populations. This test shows that the performance of students from Non-Governmental schools is better than those from Governmental schools. This result is also supported by the multivariate analysis we made in section four of the pervious chapter.

To assess which mean difference contributed to the significance of the overall performance, we have constructed Hoteelling's 95 percent simultaneous confidence intervals. From these confidence intervals it is observed that the significant difference comes only from the difference in the performance of English. No significant difference is observed in the remaining subjects.

The following are the researcher's reasoning in the observed difference in English:

- Most Non -Governmental schools in our country are established by foreigners or foreign institutes, hence they are English oriented.

- Almost all Non-Governmental schools are concentrated in few urban centers (see table 1) where communication and other infrastructures are relatively in abundant. Hence students from these schools get better access to better facilities that can help them to improve their English. These facilities may include radios, televisions, films, videos, etc.

5.2 Conclusions about the Difference in Variability

In chapter 4 the comparison in variability between the two populations is made. As we have seen the performance of students from Governmental schools has more variability than those from Non-Governmental schools. The researcher gives the following reason as the cause for the observed difference in variability.

- As stated in section 1 of this chapter almost all Non-Governmental schools are concentrated in urban centers which is, more or less, uniform environment. But Governmental schools are diversified through out the country in both urban and rural centers. Hence the uniformity in environment among Non-Governmental school students and the difference in environment among Governmental school students is the cause for the difference in variability.

5.3 Recommendations

1) To create more or less equal (some how balanced) performance in students from the two populations, there has to be experience sharing between the two types of schools.

2) The performance at freshman level is the decisive factor in future career of students. Hence students from governmental schools should do their best to perform better in English.

3) By extending this study in such a way that it includes the performance in the two semesters and it includes observations from other faculties and colleges, more reliable result can be obtained.

REFERENCES

ASKALE MOSISSA (1997), Sample Survey On Academic Performance Of Freshman Student of AAU, Science Faculty of 1987 And 1988 E.C as Affected By Methods Of ESLCE Preparation, Addis Ababa, Science Faculty (Unpublished)

COCHRAN, W.G (1977), Sampling Techniques, 3rd Edition, New Delhi, Wiley Eastern Limited.

EMIS (1996) , Federal Democratic Republic of Ethiopia Educational Statistics Annual Abstract, Addis Ababa, Ministry of Education.

JESA (1996) , Journal of the Ethiopian Statistical Association, Volume VI, Addis Ababa, Ethiopian Statistical Association.

JOHNSON R.A And Wichern, D.W (1982) Applied Multivariate Statistical Analysis, New Jersey, Prentice -Hall, Inc.

KISH L. (1965) , Survey Sampling, New York, Wiley and Sons.

KOUL L. (1984) , Methodology of Educational Research, 2nd edition, New Delhi, Vikas Publishing House Pvt. Ltd.

MORRISON, D.F (1967) Multivariate Statistical Methods, 2nd Edition, New York, Mc Graw-Hill.

SNEDECOR, G.W AND COCHRAN, W.G (1967) , Statistical Methods, 6th Edition, Iowa, Iowa State University

SOM R.K (1973) , A manual of Sampling Techniques London, Heinemann Education Books Ltd.

APPENDIX

DATA LAYOUT

No	BIO	CHEM	ENGLISH	MATH	PHY	GPA	SCHOOL_T*
1	2	3	3	0	1	1.8	1
2	4	3	4	3	2	3.2	1
3	3	3	3	3	2	2.8	1
4	1	2	4	3	1	2.2	1
5	1	1	4	1	0	1.4	1
6	3	3	4	3	2	3.0	1
7	1	2	3	0	0	1.2	1
8	4	4	3	4	3	3.6	1
9	3	3	2	4	4	3.2	1
10	2	2	3	3	2	2.4	1
11	2	2	4	2	1	2.2	1
12	2	2	3	2	1	2.0	1
13	2	3	2	2	0	1.8	1
14	2	3	2	2	2	2.2	1
15	2	2	4	2	2	2.4	1
16	2	2	4	2	2	2.4	1
17	1	2	4	0	1	1.6	1
18	2	1	2	3	1	1.8	1
19	1	2	3	2	0	1.6	1
20	3	3	4	3	1	2.8	1
21	2	3	4	3	3	3.0	1
22	2	2	3	2	1	2.0	1
23	3	2	2	2	2	2.2	1
24	2	3	2	3	3	2.6	1
25	3	4	4	2	2	3.0	1
26	1	2	2	4	4	2.6	1
27	2	2	2	2	3	2.2	1
28	2	2	2	2	1	1.8	1
29	1	1	3	1	0	1.2	1
30	2	2	4	2	1	2.2	1
31	3	3	4	2	3	3.0	1
32	2	2	2	2	0	1.6	1
33	2	2	3	2	1	2.0	1
34	2	3	4	2	3	2.8	1
35	2	2	2	2	2	2.0	1

continued

No	BIO	CHEM	ENGLISH	MATH	PHY	GPA	SCHOOL_T*
36	2	3	4	3	3	3.0	1
37	3	3	4	3	2	3.0	1
38	2	3	2	2	2	2.2	1
39	2	2	3	2	1	2.0	1
40	0	2	2	1	0	1.0	1
41	1	2	3	2	1	1.8	1
42	3	3	4	3	2	3.0	1
43	2	3	4	4	1	2.8	1
44	3	3	4	4	4	3.6	1
45	2	2	4	2	0	2.0	1
46	2	2	2	2	1	1.8	1
47	0	0	0	2	0	0.4	1
48	3	2	3	2	2	2.4	1
49	1	1	3	1	2	1.6	1
50	2	1	3	4	1	2.2	1
51	2	2	2	2	2	2.0	1
52	2	1	4	1	1	1.8	1
53	2	3	2	3	2	2.4	1
54	0	2	4	2	1	1.8	1
55	2	1	3	2	1	1.8	1
56	3	4	4	3	4	3.6	1
57	2	3	2	2	2	2.2	1
58	2	2	4	3	2	2.6	1
59	1	2	2	0	2	1.4	1
60	2	3	2	3	3	2.6	1
61	2	2	1	2	2	1.8	1
62	2	2	4	2	1	2.2	1
63	1	2	2	2	2	1.8	1
64	2	2	4	2	1	2.2	1
65	2	2	4	2	2	2.4	1
66	0	2	2	2	1	1.4	1
67	2	1	3	2	1	1.8	1
68	4	4	3	4	3	3.6	1
69	3	3	3	3	2	2.8	1
70	2	2	3	1	1	1.8	0

Continued

No	BIO	CHEM	ENGLISH	MATH	PHY	GPA	SCHOOL_T*
71	0	1	2	2	1	1.2	0
72	2	2	2	2	0	1.6	0
73	4	4	4	3	4	3.8	0
74	4	3	2	2	2	2.6	0
75	2	2	2	2	2	2.0	0
76	2	2	2	2	1	1.8	0
77	2	2	2	3	2	2.2	0
78	2	2	1	2	2	1.8	0
79	2	3	3	2	2	2.4	0
80	2	3	2	2	2	2.2	0
81	4	4	2	2	3	3.0	0
82	3	3	2	3	2	2.6	0
83	2	2	2	2	2	2.0	0
84	2	1	3	1	0	1.4	0
85	2	2	2	2	1	1.8	0
86	2	2	3	2	2	2.2	0
87	2	2	2	2	2	2.0	0
88	2	2	2	1	1	1.6	0
89	2	2	2	3	2	2.2	0
90	2	2	2	1	1	1.6	0
91	2	2	2	1	0	1.4	0
92	0	1	1	0	0	0.4	0
93	2	2	2	4	1	2.2	0
94	2	3	2	2	2	2.2	0
95	3	3	2	2	2	2.4	0
96	2	2	2	2	3	2.2	0
97	1	2	2	1	0	1.2	0
98	2	2	2	2	1	1.8	0
99	3	3	2	3	3	2.8	0
100	1	2	2	2	0	1.4	0
101	1	1	1	1	1	1.0	0
102	1	2	2	2	1	1.6	0
103	2	1	2	1	0	1.2	0
104	2	1	2	2	1	1.6	0
105	2	3	2	2	2	2.2	0

Continued

No	BIO	CHEM	ENGLISH	MATH	PHY	GPA	SCHOOL_T*
106	1	1	2	2	2	1.6	0
107	3	4	4	3	2	3.2	0
108	2	3	2	2	1	2.0	0
109	3	2	0	2	2	1.8	0
110	3	3	2	2	2	2.4	0
111	2	3	2	2	0	1.8	0
112	2	4	3	4	2	3.0	0
113	2	2	2	3	3	2.4	0
114	1	2	2	2	2	1.8	0
115	2	4	2	4	4	3.2	0
116	2	4	3	4	4	3.4	0
117	2	2	2	2	1	1.8	0
118	2	2	2	2	1	1.8	0
119	2	2	1	1	1	1.4	0
120	2	3	2	2	2	2.2	0
121	2	3	2	4	1	2.4	0
122	0	1	3	1	0	1.0	0
123	0	0	2	2	1	1.0	0
124	2	2	2	3	4	2.6	0
125	3	3	2	4	2	2.8	0
126	2	3	2	2	1	2.0	0
127	2	2	2	2	1	1.8	0
128	2	2	2	3	2	2.2	0
129	2	2	2	2	2	2.0	0
130	2	3	2	2	4	2.6	0
131	1	2	1	2	0	1.2	0
132	3	4	3	4	4	3.6	0
133	0	1	2	1	1	1.0	0
134	4	4	2	4	3	3.4	0
135	2	2	1	2	1	1.6	0
136	4	3	2	2	4	3.0	0
137	3	4	2	2	2	2.6	0
138	3	3	2	4	2	2.8	0
139	2	4	1	3	2	2.4	0
140	3	4	3	2	2	2.8	0

Continued

No	BIO	CHEM	ENGLISH	MATH	PHY	GPA	SCHOOL_T*
141	4	4	2	3	2	3.0	0
142	2	3	2	2	3	2.4	0
143	4	4	3	4	3	3.6	0
144	3	3	2	3	2	2.6	0
145	2	3	3	2	2	2.4	0
146	0	2	1	0	0	0.6	0
147	2	1	2	2	0	1.4	0
148	3	4	3	3	4	3.4	0
149	2	2	2	2	3	2.2	0
150	2	2	0	1	1	1.2	0
151	2	2	2	2	1	1.8	0
152	2	2	1	2	1	1.6	0
153	4	4	2	3	4	3.4	0
154	2	3	2	2	2	2.2	0
155	2	2	1	0	0	1.0	0
156	4	3	2	3	2	2.8	0
157	1	2	2	1	1	1.4	0
158	3	4	3	3	2	3.0	0
159	4	3	2	3	1	2.6	0
160	2	2	2	1	2	1.8	0
161	2	2	2	2	1	1.8	0
162	1	2	2	1	0	1.2	0
163	2	2	2	2	1	1.8	0
164	4	3	2	2	3	2.8	0
165	2	2	2	1	1	1.6	0
166	2	3	2	2	2	2.2	0
167	1	3	2	2	2	2.0	0
168	2	1	2	0	1	1.2	0
169	2	3	2	2	1	2.0	0
170	2	1	2	1	1	1.4	0
171	4	4	4	4	4	4.0	0
172	2	1	2	1	0	1.2	0
173	2	2	2	1	2	1.8	0
174	0	2	2	1	2	1.4	0
175	2	2	2	2	1	1.8	0

Continued

NO	BIO	CHEM	ENGLISH	MATH	PHY	GPA	SCHOOL_T*
211	3	3	4	3	3	3.2	0
212	2	2	2	2	4	2.4	0
213	2	3	4	3	2	2.8	0
214	1	1	2	1	0	1.0	0
215	2	2	2	2	2	2.0	0

* **School-t**: 1- Non-Government
0-Government