SAMPLE SURVEY ON THE ACADEMIC PERFORMANCE OF ADDIS KETEMA SENIOR SECONDARY SCHOOL IN THE YEAR 1990 E.C.

BY

DAWIT ASSEFA KEBEDE

ADDIS ABABA UNIVERSITY
FACULTY OF SCIENCE
DEPARTMENT OF STATISTICS

JUNE 1999

SAMPLE SURVEY ON THE ACADEMIC PERFORMANCE OF ADDIS KETEMA SENIOR SECONDARY SCHOOL STUDENTS IN THE YEAR 1990 E.C.

A SENIOR RESEARCH PROJECT
BY
DAWIT ASSEFA KEBEDE

UNDER THE SUPERVISION OF
ATO MULUGETA LIBSIE

SUBMITTED TO THE DEPARTMENT OF STATISTICS AS A PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR B.Sc. DEGREE IN STATISTICS

ADDIS ABABA
JUNE 1999

ABSTRACT

Statistics is a theory of information with inference as its goal. The branch of statistics that best describes the above fact is sample survey.

In sample survey, we often deal with the collection, organization and analysis of data to make inference.

Sample survey makes it easier to study effects of various variables on the subject matter of interest and making inferences based on the data which have been obtained under the study.

In this paper, data have been collected on the academic performance of students of Addis Ketema Senior Secondary School on the basis of grades scored in the Ethiopian School Leaving Certificate Examination (ESLCE) by both the regular and extension (evening) program students. This could be done by simple random sampling and stratified random sampling.

In the analysis of the data, we will be dealing with estimation, comparisons and hypothesis testing of different parameters. Main and combined effects of various variables such as effects of field of study and sex on the performance of students will also be analyzed.

Finally, conclusions and recommendations will be made.

ACKNOWLEDGMENT

I am profoundly indebted, more than I can say, to the senior project advisor, Ato Mulugeta Libsie, for his continuous assistance and constructive encouragement.

I am pretty much grateful to my brother Ato Alemyehu Assefa and Ato Eskinder Assefa for their strong material assistance.

Also due a special mention is, W/ro Aregash Mengistie, my mother, for her material and moral supporting.

Finally, I would like to extend my warmest appreciation and thanks to all individuals in the record office of Addis Ketema Senior Secondary School.

CONTENTS

CHAPTER I

INTRODUCTION

1.1. Background for the study

The Addis Ketema Senior Secondary School was established in 1952 E.C. It was named Prince Mekonnen Memorial School in memory of the then deceased son of the former emperor.

Initially, it began with grades 3-8 and the enrollment was 510 students. In 1954 E.C., the grades were up to 11 and the intake capacity of the school reached 1,481 students. In 1955 E.C. grades up to 12 were established and for the first time students appeared for the ESLCE.

The School was established to fulfill the educational needs of the vast central business area of Addis Abeba known as Mercato and its hinterland. This area is densely populated with people from all ethnic and linguistic groups in the country and of varying economic status. The City's bus terminal too is situated here. All these factors together generated a growing need for educational facilities. Now the school is implementing its teaching - learning process with regular and extension programs of study.

1.2. Objectives of the study

The main objective of this research is to study the effects of program of study. (That is, regular and extension), sex and field of study (That is, Natural Science and Social Science) on the academic performance of students. In this research project, program of study will be our major variable of interest to determine academic performance of students.

1.3 Limitations of the Study

The data obtained from the record office of the school were compiled manually and hence there could be some possible errors in their work. There is no a well organized scheme of data collection in the school. Also, only GPA of 2.0 and above in the ESLCE are registered or collected in the record office of the school. This limits the scope of the research to a certain extent (that is, it does not include those students who scored bellow 2.0).

At any rate, this research can serve as a source of pertainable information for further studies.

CHAPTER II

COLLECTION OF DATA AND SAMPLE SELECTION

2.1. Concept of data

Data is a pre-requisite for any statistical analysis. Data could be things known or assumed, facts or figures, raw material or information from which conclusions can be inferred. So as to make all the statistical analysis and inferences reliable and consistent, the data are preferred to be timely, accurate, etc.

Before the commencement of data collection, there are some factors that have to be considered. These include preliminary budget estimation, the statement of the purpose of the inquiry in clear and precise terms and the organization of a well-defined scheme of data collection. Failure in considering the above mentioned procedures may lead to confusing and inconsistent results.

Accordingly, to make our study as good as possible, it is important to make the process of data collection as cautious as possible.

2.2. Sources, type of data and data collection methods

Obtaining the right kind of statistical data is one of the problems that have been encountered in sample surveys. The type of data which is going to be analyzed depends on its source. There are two sources of data. These are called primary and secondary sources of data. Primary source of data is a source of data which has been obtained on the first hand, direct. And Secondary source of data is one that has been derived or resulting from primary source of data. The choice of the source of data depends on the purpose of the inquiry, budget or fund availability, time allocation, degree of precision desired, etc. We call data that have been obtained from primary and secondary sources as primary and secondary data respectively. Primary data are data collected by the investigator from the sources. In general, there are various methods of primary data collection which depend on the nature of the subject of study. These include physical observation, personal interviewing questionnaire, etc.

When questionnaires are employed as a means of collecting primary data, they can be filled by the respondents themselves or enumerators can ask question in the questionnaire and fill the questionnaires. There is frequently a tendency to ask too many questions, some of which are never subsequently analyzed. A long questionnaire lowers the quality of answers to important as well as unimportant questions. In designing a good questionnaire the objectives of the study and the nature of respondents must be considered. Its design must minimize potential errors from respondents and it must be covering essential points. Preparation of time table, pilot surveys and pretests, and tabulation programs are some of the activities that should be performed at the planning stage.

On the other hand, secondary data are data obtained from data collecting and publishing agencies or records. In secondary data we find data already collected and published by others.

In this study secondary data have been employed for analysis. Data on ESLCE results (only GPA of 2.0 and above), sex, field of study and program of study are taken from the record office of the school.

2.3. Fundamental concepts of sampling

In dealing with sampling, we should be aware of some basic concepts and definitions that are involved in statistical data collection, organization and analysis methods from which inferences are made.

Population : is the collection of all units under consideration or investigation. It is used to denote the aggregate (totality) from which the sample is to be chosen.

Sample : is any subset of a population selected using statistical methods so as to estimate the characteristics of the population. It should be selected in such a way that it will represent the population under investigation.

Statistical survey : an investigation of a specific population to assess its quantitative and qualitative characteristics. Characteristics of a population are meant any quality or relationship to the population.

Census : refers to the collection of information about characteristics of interest from all units under investigation.

There are some factors that will make a census impractical. These are:

- cost
- labor
- timeliness and speed
- etc.

Census is more efficient in providing information on rare events.

Sampling : is the process of taking a sample and making inferences to the whole population.

Sample survey : refers to the collection of information about the characteristics of interest from only a part of the population which are selected by some statistical techniques.

Frame : Before selecting the sample, the population must be divided into parts that are called sampling units, or simply units. These units must cover the whole population and they must not overlap, in the sense that every element in the population belongs to one and only one unit. Sometimes, the appropriate unit is obvious, as in a population of students who took ESLCE in the Year 1990 E.C. in Addis Ketema Senior Secondary School, in which the unit is a single student.

Target population : is the population defined at the planning stage, for which the results are expected.

Sampled population : is the population actually covered. As much as possible, the population to be sampled should coincide with the population about which information is wanted. Sometimes, for reasons of practicability or convenience, the sampled population is more restricted than the target population. If so, it should be remembered that conclusions drawn from the sample apply to the sampled population. Judgment about the extent to which these conclusions will also apply to the target population must depend on other sources of information. Any supplementary information that can be gathered about the nature of the differences between the sampled and the target populations may be helpful.

Probability and Non-Probability Sampling

Basically there are two methods of sampling. These are called probability and non-probability sampling.

<u>Probability Sampling :</u> is also called random sampling. In probability sampling, every unit of the population has a calculable (non-zero) probability of being selected into the sample. A probability sample provides valid estimates of unknown population values with their measures of reliability by indicating the extent of the error due to sampling and can also set scopes or limits on the possible or likely effects of imperfections.

The following are types of probability sampling methods.

- simple random sampling
- stratified sampling
- cluster sampling
- systematic sampling
- multi-stage sampling
- sampling with varying probabilities

<u>Non-probability sampling :</u> is a kind of sampling in which the selection is non-random. In non-probability sampling, experts select samples purposely by deciding "representative" units which are subject to personal biases. With non-probability sampling, it is not possible both to estimate sampling variability from the sample and to know about the possible biases involved. Non-probability sampling includes the following type of methods.

- Haphazard and samples of volunteer subjects
- Quota sampling

- Sampling of mobile populations "capture-tag-recapture"
- Expert choice

Since in this research project, simple random sampling and stratified random sampling have been employed, we will discuss these two methods briefly.

Simple Random Sampling(SRS)

Simple random sampling is a selection procedure in which a predetermined number of units from a population is drawn. In SRS, each sampling unit has an equal chance of being drawn. In practice a simple random sample is drawn unit by unit. The units in the population are numbered from 1 to N where N is the total number of units in a population. A series of random numbers between 1 and N are ten drawn, either by means of a table of random numbers or by means of a computer program that produces such a table. The other method is a lottery method where each member of the population (students who took ESLCE both in the regular and extension program) is represented by identifiable piece of paper, the papers are placed in an urn/bowl and well mixed, and a sample of specified size is selected. In this research project, lottery method has been employed to select elements.

SRS can be done with or without replacement.

SRS without replacement(WOR) : means that a unit that has been drawn is removed from the population under investigation, for the rest of draws.

That is a unit can not appear more than once in the sample (with the restriction that if there are the same units in the population). The probability of any element being selected is equal to n/N, where n is the sample size and N is the population size.

SRS with replacement : here a sampling unit can be selected more than once. At any draw, all members of the population under investigation are given an equal chance of being drawn, no matter how often they have already been drawn. Accordingly, all selection elements are independent. In the case of SRS with replacement, there are N^n ways of selecting n units out of the total N units where N refers to number of units in the population (population size) and n refers to number of units in the sample (Sample size).

Stratified Random Sampling

Stratified random sampling is a sampling scheme that involves division or stratification of a population into mutually exclusive or non-overlapping sub populations called strata. These strata comprise of the whole population. Then a simple random sample is taken from each stratum. The selection of sampling units from each stratum is done independently of one another.

Let N be number of units in the population under investigation and suppose these N units are divided into K sub population or strata. That means,

$$N = N_1 + N_2 + \ldots + N_k$$

and

Suppose n1, n2, ..., nk represent the number of units which have been randomly selected from each stratum and n the total sample size.
Then

$$n = n_1 + n_2 + \ldots + n_k$$

The total number of possible stratified random samples is equal to :

$$\binom{N_1}{n_1} \times \binom{N_2}{n_2} \times \ldots \times \binom{N_k}{n_k} \leq \binom{N}{n}$$

2.4. Sample size Determination

In this study, a statistical method that has been employed for the purpose of data collection is simple random sampling and elements to be included were chosen with the lottery method.

We also have stratification with respect to program of study into regular and extension. Each sub population or stratum, program of study is further stratified with respect to field of study. In this sub section, we will see how the sample size is determined and allocated to each stratum.

In dealing with sample surveys, one of the most important tasks is the determination of the number of sample units (sample size) to be considered for investigation. There are some procedures to be followed so as to determine the sample size.

When the main objective is to estimate population means, the following formula for determining the sample size is used.

$$n_0 = \frac{Z^2_{\alpha/2} \, S^2}{d^2} \qquad 2.1$$

where

α	is the risk probability
d	is the permissible level of error or absolute margin of error
S^2	is the estimated variance

If $\frac{n_0}{N}$ is appreciable (very small, often less than 5% of the population), then we compute n as in equation (2.1).

Otherwise,

$$n = \frac{n_0}{1+\frac{n_0}{N}} \qquad 2.2$$

In this study, we have assumed $\alpha = 0.05$ and d = 0.1 and hence,

$$Z_{\alpha/2} = Z0.025 = 1.96$$

The variance was estimated as follows:

The total number of students who took ESLCE in 1990 E.C. is 1845. It is the overall sum of students in the regular and extension programs of study. From the 1845 students, a sample of 20 students was randomly selected in order to estimate the variance and it was calculated to be 0.16. The data for the sampled students is available in appendix A. Therefore, using equation (2.1)

$$n_o = \frac{(1.96)^2\ (0.16)}{(0.1)^2} = 61$$

and

$$\frac{n_0}{N} = \frac{61}{1845} = 3.3\%$$ which is less than 5% of the population

Since the ratio $\frac{n_o}{N}$ is very small, we do not need to use equation (2.2)

In this study, we do have stratification of the total students (1845 students) into those who belong to the regular program of study and those who belong to the extension program of study. From the collected data which have been available by the record office of the school. N1, number of students in the extension program, is 895 of which 500 are male students and 395 are female students.

And N2, Number of students in the regular program is 950 of which 620 are male students and 330 are female students.

The total sample size must be allocated to these two groups. In this allocation, three factors may be considered. These are:

- the size of each stratum
- the cost of selecting observations per sampling unit in the stratum
- variability within each stratum

The commonly used methods of sample allocation include equal, proportional and optimum allocation. Since the size of a stratum is the only available information, we employ proportional allocation. In this method, the number of units are drawn in proportion to the size of the stratum.

That is,

$$n_h = \frac{n}{N} N_h$$

Consequently, sample sizes for the extension and regular strata respectively are:

$$n_1 = \left(\frac{n}{N}\right) N_1 = \left(\frac{61}{1845}\right)(895) \approx 26$$

$$n_2 = \left(\frac{n}{N}\right) N_2 = \left(\frac{61}{1845}\right)(950) \approx 31$$

From the observed data we have 555 Natural Science students under the extension stratum and 696 Natural Science students under the regular stratum. Besides, there are 340 Social Science students under extension stratum and 254 Social Science students under regular stratum.

Let m_1 = sample size for natural science students under extension stratum

m_2 = sample size for social science students under extension stratum

m_1' = sample size for natural science students under regular stratum

m_2' = sample size for Social science students under regular stratum

Then

$$m_1 = \frac{555(26)}{895} \approx 16$$

$$m_2 = \frac{340(26)}{895} \approx 10$$

$$m_1' = \frac{696(31)}{950} \approx 23$$

$$m_2' = \frac{254(31)}{950} \approx 8$$

CHAPTER III

METHOD AND ANALYSIS OF THE DATA

As stated in the previous chapters, the objective of statistics is to make an inference about a population based on information contained in a sample. Since populations are characterized by numerical descriptive measures called parameters, the objective of many statistical investigations including sample survey is to make an inference about one or more population parameters, most statistical inference procedures involve either estimation or hypothesis testing.

In its simplest terms, estimation is making an educative guess about the unknown quantity or parameter. Any statistic (known function of observable random variables) whose values are used to estimate the unknown quantity for parameter is defined to be an estimator of the parameter. An estimator is both a function and a random variable. Estimation can be done in two ways.

i) Determine a number from the given observations which can be taken to be the value of the parameter. This is called point estimation.

ii) Determine a range of numbers which can be taken to include the parameter. And this is called interval estimation.

In the next subsections we will employ the above two ways of estimation so as to estimate the population parameters, the mean and proportion.

3.1 Estimation of the mean for subpopulations

Before making estimation of stratum mean, it is important to note the population characteristics for strata.

Let Nh= total number of non-overlapping units in stratum h.

$$N = \sum_{h=1}^{k} Nh = \text{the population size}$$

Yhi = the value of the study variable for the i^{th} unit in stratum h, i = 1,2, ... Nh

$$Yh = \sum_{i=1}^{Nn} Yhi = \text{ total amount of study variable within stratum h.}$$

$$Y = \sum_{h=1}^{K} Yh = \sum_{h=1}^{k} \sum_{i=1}^{Nh} Yhi = \text{total for the entire population.}$$

$$\overline{Y}_h = \frac{yh}{Nh} = \frac{\sum_{i=1}^{Nh} Yhi}{Nh} = \text{ the true mean fo stratum h}$$

$$S_h^2 = \sum_{i=1}^{Nh} \left(\frac{(Y_{hi} = \overline{Y}_h}{N_h - 1} \right)^2 \text{ is the variance of stratum h.}$$

and

The sample characteristics for strata are as follows:

n_h = number of sample units in strata h.

$$n = \sum_{h=1}^{k} nh = \text{ total sample size}$$

y_{hi} = the value of the variable in the sample for the i^{th} unit is stratum h(i = 1,2, ..., nh)

$$y_h = \sum_{i=1}^{nh} y_{hi} = \text{ sample total for stratum h}$$

$$\overline{y}_h = \frac{y_h}{n_h} = \frac{\sum_{i=1}^{nh} y_{hn}}{n_h} \text{ is sample mean forstratum h}$$

$$y = \sum_{h=1}^{k} y_h \text{ is the total for the entire sample}$$

$$f_h = \frac{n_h}{N_h} \text{ is sampling fraction instratum h}$$

$$s_h^2 = \sum_{i=1}^{n_h} \frac{(y_{hi} - \bar{y}_h)^2}{n_h - 1}$$ is sample variance for stratum h.

$\bar{y}_h$ and s_h^2 are unbiassed estimators of Y_h and S_h^2 respectively for a simple random sample of size n_h from stratum h

In this study, we consider the two strata based or the program of study (that is ,regular and extension). We have

$n_1 = 30$ and $n_2 = 31$

$$\bar{y}_1 = \sum_{i=1}^{n1} y_{1i} = \frac{\sum_{i=1}^{30} y_{1i}}{30} = 2.41$$ is a point estimate of the mean ESLCE result for those students belonging to the extension program of study

$$\bar{y}_2 = \sum_{i=f1}^{n_2} y_{2i} = \frac{\sum_{i=1}^{31} y_{2i}}{31} = 2.51$$ is a point estimate of the mean ESLCE result for those students belonging to the regular program of study.

$$S_1^2 = \sum_{i=1}^{30} \frac{(y_{1i} - \bar{y}_1)^2}{30-1} = 0.168$$

$$S_2^2 = \sum_{i=1}^{31} \frac{(y_{2i} - \bar{y}_2)^2}{31-1} = 0.256$$

Var $(\bar{y}_h) = \frac{s_h^2}{n_h}\left(1 - \frac{n_h}{N_h}\right)$ is the variance of the stratum mean $\bar{y}$ for a simple random sample of n_h.

In estimating Var $(\bar{y}_h)$; we replace S_h^2 by s_h^2

Therefore ; Var $(\bar{y}_h) = \frac{s_h^2}{n_h}\left(1 - \frac{n_h}{N_h}\right)$

$$\text{Var}\ (\bar{y}_1) = \frac{s_1^2}{n1}(1-f1) = \frac{0.168}{30}\left(1-\frac{30}{895}\right) = 0.00541$$

$$\text{Var}\ (\bar{y}_2) = \frac{s_2^2}{n2}(1-f2) = \frac{0.256}{31}\left(1-\frac{31}{950}\right) = 0.00799$$

Thus, the corresponding standard errors of the point estimates are

$$\text{S.e.}(\bar{y}_1) = \sqrt{\text{Var}}(\bar{y}_1) = \sqrt{0.00541} = 0.0736$$

$$\text{S.e.}(\bar{y}_2) = \sqrt{\text{Var}}(\bar{y}_2) = \sqrt{0.00799} = 0.0894$$

Consequently, it is possible to construct a 100(1- α) % confidence interval for the true strata means.

In general a 100 (1- α) % confidence level for population mean is given by:

$$\bar{Y} = \bar{y} \pm Z\alpha/_2\ \text{S.e.}\ (\bar{y})$$

Thus, the lower and upper confidence limits for the mean of the extension program of study stratum are.

Lower C.L. = $\bar{y}_1 - Z\alpha/2$ S.e. $(\bar{y}_1)$

Upper C.L = $\bar{y}_1 + Z\alpha/2$ S.e. $(\bar{y}_1)$

Let $\alpha = 5\% = 0.05$ then $Z\alpha/2 = Z0.025 = 1.96$.

Therefore;

Lower = 2.41 - 1.96 S.e $(\bar{y}_1)$ = 2.266

Upper = 2.41 + 1.96 S.e. $(\bar{y}_1)$ = 2.554

Therefore, the 95% confidence interval for $\bar{Y}_1$ is (2.266, 2.554)

Similarly, since n_2=31 is large, we apply the central limit theorem to construct the confidence interval, so, the lower and upper confidence limits for regular program of study stratum would be

Lower Confidence limit = $\bar{y}_2 - Z\alpha/2.$ S.e $(\bar{y}_2)$

Upper confidence limit $\bar{y}_2 + Z\alpha/2.$ S.e $(\bar{y}_2)$

Setting the appropriate values:

Lower confidence limit = 2.51 – 1.96 s.e $(\bar{y}_2) = 2.335$

Upper Confidence limit = 2.51 + 1.96 s.e $(\bar{y}_2) = 2.685$

Accordingly, the 95% confidence interval for $\bar{Y}_2$ is (2.335, 2.685)

3.2 Estimation of stratum Proportion

If we wish to estimate the proportion of units in the population that fall into some defined class C, the ideal stratification is attained if are can place in the first stratum every unit that falls in C, and in the second every unit that does not.

In other words, suppose every unit in the population falls into some classes C or C'. Define the value of Y_{hi} as 0 or 1 depending on the classes of interest.

That is ;

$$Y_{hi} = \begin{cases} 1, & \text{if the } i^{th} \text{ unit in stratum } h \text{ belongs to C} \\ 0, & \text{Otherwise} \end{cases}$$

Let

$P_h = \frac{A_h}{N_h} = \frac{Y_h}{N_h} = \overline{Y}_h$, the proportion of units in the stratum that belong to C.

$$Q_h = 1 - P_h$$

$$S_h^2 = \sum_{i=1}^{Nh} \frac{(Y_{hi} - \overline{Y}_h)^2}{N_h - 1} = \frac{N_h P_h Q_h}{N_h - 1}$$

$$A = \sum_{h=1}^{k} Ah = \sum_{h=1}^{k} \sum_{i=1}^{Nh} Y_{hi} = \sum_{h=1}^{k} Y_h = Y$$

$$P = \frac{A}{N}$$

And for the sample, we have

$$ah = \sum_{i=1}^{nh} y_{hi}$$

$P_h = \frac{ah}{nh} = \bar{y}_h$, sample proportion in h^{th} stratum.

$$s_h^2 = \sum_{i=1}^{nh} \frac{(y_{hi} - \bar{y}_h)^2}{nh-1} = \frac{n_h\ p_h\ q_h}{n_h - 1}\text{; where } q_h = 1 - p_h$$

Both p_h and s_h are unbiased estimtors of P_h and S_h respectively for a stratified randam sample of size n_h from each sratum.

Let

C= those students who have passed the ESLCE (i.e, those who scored 2.8 and above in ESLCE)

C'= those who failed the ESLCE

(i.e, those who scored below 2.8 in ESLCE).

$$\hat{P}_1 = \frac{a1}{n1} = \text{ and } \hat{P}_2 = \frac{a_2}{n_2}$$

Where $n_1 = 30$ and $n_2 = 31$

$a_1 = 5$ and $a_2 = 8$

Accordingly:

$$b_1 = \frac{5}{30} \approx 0.167$$

from which $\hat{q}_1 = 1 - \hat{P}_1 = 0.833$

$$\hat{P}_2 = \frac{8}{31} \approx 0.258$$

from which $\hat{q}_2 = 1 - \hat{p}_2 = 0.742$

Next;

$$S_1^2 = \frac{n_1 p_1 q_1}{n_1 - 1} = \frac{30(0.167)(0.833)}{29} \cong 0.144$$

$$S_1^2 = \frac{n_2 p_2 q_2}{n_2 - 1} = \frac{31(0.258)(0.742)}{30} \cong 0.198$$

and as we note

$$Var(P_h) = \frac{(S_h)^2}{n_h}(1 - f_h)$$

Thus; $Var(\hat{P}_h) = \frac{s_1^2}{n_h}(1-f_1) = \frac{0.144}{30}\left(1-\frac{30}{895}\right) = 0.00464$

$S.e(\hat{p}_1) = \sqrt{Var(\hat{p}_1)} = \sqrt{0.00464} = 0.0681$

$Var(\hat{p}_2) = \frac{(s_2)^2}{n_2}(1-f_2) = \frac{0.198}{31}\left(1-\frac{31}{950}\right) = 0.00618$

$S.e(\hat{p}_2) = \sqrt{Var(\hat{p}_2)} = \sqrt{0.00618} = 0.0786$

As can be seen from the above results, the proportion of students who have scored 2.8 and above and belong to the regular program of study is a bit larger than students who have scored 2.8 and above and belong to the extension program of study.

$\hat{p}_1$ = 16.7% estimates P_1 with astandard error of 0.0681.

and $\hat{p}_2$=25.8% estimates ,p_2 with a standard error of 0.0786. We then compute a 100(1-α)% confidence interval for P_1 and P_2 as follows:

Lower confidence limit for $P_1 = \hat{p}_1 - Z\alpha/2\,S.e.(\hat{p}_1)$

$= 0.167 - 1.96\ (0.0681)$

$= 0.0335$

Upper confidence limit for $P_1 = \hat{p}_1 + Z\alpha/2\,S.e.(\hat{p}_1)$

$= 0.167 + 1.96\ (0.0681)$

$= 0.3005$

Therefore, the 95% confidence interval for P_1 is (0.0335, 0.3005)

Next,

Lower confidence limit for $P_2 = \hat{p}_2 - Z\alpha/2\,S.e.(\hat{p}_2)$

$= 0.2580\ 1.96\ (0.0786)$

$= 0.1039$

Upper Confidence limited or $P_2 = \hat{p}_2 + Z\alpha/2\ \text{S.e.}(\hat{p}_2)$

$$= 0.258+1.96\ (0.0786)$$

$$= 0.4121$$

Thus, the 95% confidence interval for P_2 is (0.1039, 0.4121)

3.3 Estimation of Population mean

The estimator of population mean in stratified random sampling is given by

$$\bar{y}_{st} = \sum_{h=1}^{k} \frac{Nh\bar{y}_h}{N}$$

$\bar{y}_{st}$ is unbiassed estimator ofthe poplation mean -

$$\bar{y}_{st} = \sum_{h=1}^{k} \frac{N_h\bar{y}_h}{N} = \frac{N_1\bar{y}_1}{N} + \frac{N_2\bar{y}_2}{N}$$

$$= \frac{895 \times 2.41}{1845} + \frac{950 \times 2.51}{1845} \approx 2.46$$

the variance of the estimate $\bar{y}_{st}$ is

$$\text{Var}\left(\bar{y}_{st}\right) = \frac{1}{N^2} \sum_{h=1}^{k} Nh\left(N_h - n_h\right)\frac{S_h^2}{n_h}$$

For our case

$$\text{Var}\left(\bar{y}_{st}\right) = \frac{1}{N^2} \sum_{h=1}^{2} Nh\left(N_h - n_h\right)\frac{S_h^2}{n_h}$$

Setting the appropriate values;

$\text{Var}\left(\bar{y}_{st}\right) = 0.00339$

Thus, S.e $\left(\bar{y}_{st}\right) = \sqrt{\text{Var}\ (\bar{y}_{st})} = 0.0582$

We then construct a 100(1-α)% confidence interval for the population mean, μ, as follows

Lower confidence limit : $\bar{y}_{st} - Z_{\alpha/2}\, s.e(\bar{y}_{st})$

Upper confidence limit : $\bar{y}_{st} + Z_{\alpha/2}\, s.e(\bar{y}_{st})$

Making use of the previous assumption

(that is, $\alpha = 0.05$)

Lower confidence limit = 2.46 – (1.96) (0.0582)

= 2.3459

Upper confidence limit = 2.46 +(1.96) (0.0582)

= 2.5741

Accordingly, the 95% confidence interval for the population mean (μ) is (2,3459, 2.5741).

3.4 Estimation of Population Proportion

With stratified random sampling, the estimate of the population proportion, P is

$$P_{st} = \sum_{h=1}^{k} \frac{N_h}{N} \hat{p}_h$$

For our observed data,

$$P_{st} = \sum_{h=1}^{2} \frac{N_h \hat{P}_h}{N} \approx \frac{895\,(0.167) + 950\,(0.258)}{1845}$$

$$= 0.21386$$

and the variance of P_{st} is given by

$$Var\,(P_{st}) = \frac{1}{N^2} \sum_{h=1}^{k} N_h^2 \left(\frac{N_h - n_h}{N_h - 1} \right) \frac{P_h Q_h}{n_h}$$

Setting the appropriate values,

$Var\,(P_{st}) = 0.00264$

from which, s,e $(P_{st}) = \sqrt{Var\,(P_{st})} = 0.0514$

Then we construct a $(1-\alpha)100\%$ confidence interval for the population proportion, P

$$P_{st} \pm Z\alpha/_2 \, s.e(P_{st})$$

assuming $\alpha = 0.05$

Lower Confidence limit = 0.21386-(1.96) (0.05414)

= 0.11312

Upper confidence limit = 0.21386+(1.96) (0.05414)

= 0.3146

Accordingly, the 95% confidence interval for population proportion is (0.11312, 0.3146). Besides P_{st} = 21.386% is an estimate of the proportion of students who passed ESLCE during the year 1990 E.C with a standard error of 0.0514.

3.5 Hypothesis Testing

Most of the time the objective of a statistical test is to test a hypothesis concerning the values of one or more population parameters. The elements of a statistical test are

- null hypothesis, H_0
- alternative hypothesis, H_2
- test statistic
- rejection region

In our case, we want to test whether the mean ESLCE result of the students is 2.8 or not. That is, we test that

$H_0 : \mu = 2.8$

$Ha : \mu < 2.8$ assume $\alpha = 0.05$

The test statistic is

$$Z_{cal} = \frac{\overline{X} - \mu}{\delta\sqrt{n}} = \frac{\overline{Y}_{st} - \mu}{\delta / \sqrt{n}}$$

Under the assumption of the null hypothesis;

$$Z_{cal} = \frac{2.46 - 2.8}{\sqrt{\frac{0.00264}{61}}} = -51.68$$

and

$$-Z\alpha = -Z0.05 = -1.65$$

Since $Z_{cal} < -Z\alpha$, we reject H_0 and conclude that the mean ESLCE result of the students is less than 2.8 (accept H_a). From this conclusion, we merely understand that most of the students did not passed in ESLCE as 2.8 is the passing mark.

3.6 Comparison of Two Means

In this subsection, we will compare the mean ESLCE results of those students who belong to extension program of study and those students who belong to regular program of study. Suppose M_1 represent the mean ESLCE result of those students who belong to extension program and M_2 represent the mean ESLCE result of those students who belong to regular program So, we need to test that

$$H_o : M_1 = M_2$$

$$H_a : M_1 \neq M_2 \qquad \text{, assume } \alpha = 0.05$$

Note: In this case, we have two-sided test.

The test statistic is

$$Z_{cal} = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$= \frac{2.41 - 2.51}{\sqrt{\frac{0.168}{30} + \frac{0.256}{31}}} = -0.8495$$

Since $|Z_{cal}| < Z\alpha/2 = 1.96$; we accept H_0 and conclude that the mean ESLCE result of those students who belong to extension program is almost the same as the same as the mean ESLCE result of those students who belong to regular program of study.

3.7 Comparison of Stratum Proportions

In this subsection we make comparison of proportions of students who passed ESLCE in the extension and regular programs of study.

- We need to test that

$H_0 : P_1 = P_2$

$H_a : P_1 = P_2$ assume $\alpha = 0.05$

The test statistic is

$$Z_{cal} = \frac{|\hat{p}_1 - \hat{p}_2|}{s.e(\hat{p}_1 - \hat{p}_2)}$$

Where s.e $(\hat{p}_1 - \hat{p}_2) = \sqrt{p(1-p)} \quad \sqrt{\frac{1}{n1} + \frac{1}{n2}}$

Where $p = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$

Recall that

$\hat{P}_1 = 0.167$ and $\hat{P}_2 = 0.258$

$n_1 = 30$ and $n_2 = 31$

Therefore $P = \frac{30(0.167) + 31(0.258}{30 + 31}$

$= 0.21325$

Accordingly, s.e $(\hat{P}_1 - \hat{P}_2) = 0.1049$

Therefore , $Zcal = \frac{|0.167 - 0.258|}{0.1049} = 0.8675$

Since $Zcal < Z\alpha/2 = 1.96$ we accept H_0 and infer that the proportion of the students who passed ESLCE in the extension program were not different from those who belong to regular program of study.

3.8 Two Way Analysis with Main and Combined Effects

In this part we want to study the main effects and combined effects of field of study and sex on academic performance of the students.

To evaluate the combined effect of two or more experimental variables when they are used simultaneously, we use factorial experiments. Factorial experiment permits the evaluation of interaction effects besides main effects.

The observed ESLCE results stratified or classified with sex and field of study is tabulated as follows:

		Field of Study	
		Natural Science	Social Science
Sex	Male	2.2, 2.8, 2.4, 3.4, 2.6, 2.0,2.0,2.2,2.6,2.4,2.0,2.4, 2.2,2.0,2.2,3.2,2.2,2.8,2.6, 2.4,4.0,2.0,3.4,2.8,2.6,2.2, 2.0,2.6,2.2	2.8,2.2,2.4,2.0,2.4, 2.0,2.8,2.2,2.2,2.2
	Female	2.6,3.2,2.0,3.0,3.8,2.4,2.6, 2.0,2.0,2.0,2.4	2.0,2.4,2.0,2.6,2.6, 2.2,2.0,2.6,2.6,2.4, 3.2

Let y represent an Observation. Then we have

		Natural Science	Social Science
Male	n_{ij}	29	10
	$\sum y$	72.4	23.2
	$\sum y^2$	187.6	54.56
	SS_{ij}	6.85	0.736
Female	n_{ij}	11	11
	$\sum y$	28	26.6
	$\sum y^2$	71.36	65.64
	SS_{ij}	0.087	1.316

Where n_{ij} is cell frequency

$$SS_{ij} = \Sigma y^2 - \frac{(\Sigma y)^2}{n_{ij}}$$

Σy is sum of cell observations

SS with in cell $= \sum_i \sum_j S_{ij} = 6.85 + 0.736 + 0.87 + 1.316$

$= 8.989$

The harmonic mean of the cell frequencies is computed as

$$n_h = \frac{RC}{\sum_i \sum_j \left(\frac{1}{n_{ij}}\right)}, \textit{where } \text{R refers to number of rows an C refers to number of columns}$$

For the observed data;

$$n_h = \frac{2 \times 2}{\frac{1}{29} + \frac{1}{11} + \frac{1}{10} + \frac{1}{10}} \approx 12.646$$

Cell means	Natural Science	Social Science	$n_{i\cdot}$
Male	2.49	2.32	4.81
Female	2.55	2.42	4.97
n.j	5.04	4.74	9.78

Then using the above table;

$$\frac{\sum n_{.j}^{2}}{r} \approx 23.93; \qquad \sum n_{ij}^{2} \approx 23.94$$

$$\frac{(n..)^2}{RC} \approx 23.91; \qquad \frac{\Sigma n_i^2}{c} \approx 23.92$$

then it follows that

SS field of study = $n_h[23.92 - 23.91] = 0.12646$

$SS_{Sex} = n_h[23.93 - 23.91] = 0.25292$

$SS_{interaction} = \bar{n}_h \left[23.94 - 23.92 - 23.93 + 23.91\right] = 0$

Assume $\alpha = 5\% = 0.05$

ANOVA table

Source of variation	ss	Df	Ms	F
Field of study	0.12646	C-1=1	0.12646	0.802
Sex	0.25292	R-1=1	0.25292	1.604*
Interaction	0	(C-1)(r-1)=1	0	0
Within cell	8.989	$\sum\sum_{nij} - rc = 57$	0.1577	
Total	9.3684	60		

As can be seen from the ANOVA table, there is no interaction effect, Besides, the main effect of field of study is not that much considerable or significant. And main effects of sex are significant.

CHAPTER IV

CONCLUSIONS

The main objective of this research project, as it has been stated in Chapter I, is to study the effects of program of study, sex and field of study on the academic performance of students. In this chapter, making use of the statistical results from the data analyzed, we will make summary and conclusion.

4.1 Inferences of the basis of Estimation of Parameters

In estimation of parameters, we estimated strata means and proposition. The mean ESLCE result of the students (both in the regular and extension programs of study) was found to be 2.46, with a standard error of 0.0582. The 95% confidence interval for the population mean is (2.3459, 2.5741).

We also recall that the proportion of students scoring 2.8 and above was 0.21386 with a standard error of 0.0514. In other words, almost 21.4% of the students (students that belong to the regular and the extension programs of study) passed in ESLCE during the academic year 1990 E.C.

4.2 Inferences on the basis of Hypothesis Testing

As it has been said in the previous chapter, the objective of a statistical test is to test a hypothesis concerning the values of one or more population parameter.

We employed one-sided test in order to conclude whether the mean ESLCE result of the students is 2.8 or not. On the basis of the results obtained, we have rejected the null hypothesis (i.e $H_0 : \mu = 2.8$) and accepted the alternative hypothesis, that is $H_a : \mu < 2.8$ (that is, the mean ESLCE result of the total students is much less than the desired passing result).

4.3 Conclusion on the Basis of Comparison of Means

We have also made comparison of the mean ESLCE results of those students who belong to the extension program of study and those students who belong to regular program of study. We employed the Z-test statistic and accept the null hypothesis and concluded that the mean ESLCE result of those students who belong to extension program is almost the same as the mean ESLCE result of those students who belong to the regular program of study.

4.4 Conclusion on the Basis of Comparison of Proportions

We have made comparison of proportion of students who passed ESLCE in the extension and regular programs of study by testing the hypothesis that the proportion of students who passed ESLCE in the extension program were equal to the proportion of those students who belong to the regular program of study against the alternative hypothesis that the proportion of students who passed ESLCE in the extension program were different to the proportion of those students who belong to the extension program of study.

According to the above subsections, we can simply infer that program of study has not significant effect on the academic performance of the students.

In the case of two-way ANOVA we have studied the main and combined effects of field of study and sex on academic performance of students.

Using the F-test, the main effect of field of study is not significant, while, main effect of sex is significant. We also inferred that there is no interaction or combined effect.

This research project can get rid of the mis-understood idea that program of study will affect academic performance of students which actually will not matter. Also effect of sex on academic performance can be further studied on the basis of this research project with the additional statistical data and methods.

At any rate, this research project can be used as a source of information for further study on the academic performance of students of Addis Ketema Senior Secondary School.

This research project can get rid of the mis-understood idea that program of study will affect academic performance of students which actually will not matter. Also effect of sex on academic performance can be further studied on the basis of this research project with the additional statistical data and methods.

At any rate, this research project can be used as a source of information for further study on the academic performance of students of Addis Ketema Senior Secondary School.

REFERENCE

1. Bulletin (1974 E.C), yearly bulletin of Addis Ketema Senior Secondary School

2. Mendenhall, Scheaffer, Wackerly (2^{nd}) edition), Mathematical Statistics with Applications.

3. Papoulis (International Student edition), Probability, Random variables, and stochastic Processes.

4. William G, Cochran (3^{rd} edition); Sampling Techniques

5. Gupta (2^{nd} edition); Introduction to Statistical Methods.

Appendix A

Data layout
The following data is the observed data of Sex, field of study, ESLCE result, and proforma of study.

No	Sex	Field of study	ESLCE Results	Proframs of study
1	M	N.S	2.2	Regular
2.	M	N.S	2.8	Regular
3.	M	N.S	2.4	Regular
4.	F	N.S	2.6	Extension
5.	M	N.S	3.4	Regular
6.	F	S.S	2.0	Extension
7.	M	S.S	2.8	Regular
8	M	N.S	2.6	Regular
9.	F	S.S	2.4	Extension
10.	F	N.S	3.2	Extension
11.	M	N.S	2.0	Regular
12.	M	S.S	2.2	Regular
13.	M	S.S	2.4	Regular
14.	F	S.S	2.0	Extension
15.	F	N.S	2.0	Extension
16.	M	N.S	2.0	Extension
17.	M	N.S	2.2	Extension
18.	M	N.S	2.6	Extension
19.	M	N.S	2.4	Regular
20.	F	N.S	3.0	Regular
21	M	S.S	2.0	Regular
22	F	S.S	2.6	Extension
23.	M	N.S	2.0	Regular
24.	M	N.S	2.4	Regular
25.	M	N.S	2.2	Regular
26.	M	N.S	2.0	Extension
27.	M	N.S	2.2	Regular
28.	F	N.S	3.8	Regular
29.	M	S.S	2.4	Extension
30.	F	S.S	2.6	Extension
31.	M	N.S	3.2	Extension
32.	M	S.S	2.0	Regular
33.	M	N.S	2.2	Extension
34	F	N.S	2.4	Extension
35	F	S.S	2.2	Regular
36.	F	N.S	2.6	Regular
37.	M	N.S	2.8	Extension

38.	F	N.S	2.0	Extension
39	M	S.S	2.8	Regular
40.	M	S.S	2.2	Extension
41	F	S.S	2.0	Regular
42	F	N.S	2.0	Regular
43	M	S.S	2.2	Extension
44	M	N.S	2.6	Extension
45	F	N.S	2.0	Regular
46	M	N.S	2.4	Regular
47	M	N.S	4.0	Regular
48	M	N.S	2.0	Extension
49	M	N.S	3.4	Extension
50	F	S.S	2.6	Extension
51	F	S.S	2.6	Extension
52	M	N.S	2.8	Regular
53	F	S.S	2.4	Regular
54	M	S.S	2.2	Extension
55.	M	N.S	2.6	Regular
56.	F	S.S	3.2	Extension
57.	F	N.S	2.4	Extension
58	M	N.S	2.2	Extension
59	M	N.S	2.0	Extension
60.	M	N.S	2.6	Regular
61	M	N.S	2.2	Extension

Appendix B

This is the observed data of 20 randomly selected students to estimate the variance of population

No	ESLCE Result
1.	2.2
2.	2.0
3.	2.6
4.	2.0
5.	3.0
6.	2.4
7.	3.4
8.	2.8
9.	2.2
10.	2.0
11.	2.0
12.	2.4
13.	2.8
14.	2.0
15.	2.4
16.	2.6
17.	3.2
18.	2.2
19.	2.6
20.	2.4