YANGON UNIVERSITY OF ECONOMICS
DEPARTMENT OF STATISTICS

AWARENESS ON HEALTH HAZARD OF SMOKING AMONG STUDENTS IN YANGON UNIVERSITY OF ECONOMICS

BY

KAY ZIN WIN
M. Econ (Statistics)
Roll No.3

NOVEMBER, 2019

YANGON UNIVERSITY OF ECONOMICS
DEPARTMENT OF STATISTICS

AWARENESS ON HEALTH HAZARD OF SMOKING AMONG STUDENTS IN YANGON UNIVERSITY OF ECONOMICS

This thesis is submitted as a partial fulfillment towards the
Degree of Master of Economics (Statistics).

BY

KAY ZIN WIN
M. Econ (Statistics)
Roll No.3

NOVEMBER, 2019

YANGON UNIVERSITY OF ECONOMICS

DEPARTMENT OF STATISTICS

AWARENESS ON HEALTH HAZARD OF SMOKING AMONG STUDENTS IN YANGON UNIVERSITY OF ECONOMICS

This thesis is submitted as a partial fulfillment towards
the Degree of Master of Economics (Statistics).

Approved by the Board of Examiners

Submitted by:

Ma Kay Zin Win
M. Econ (Statistics)
Roll No.3

Co-Supervised by:

Daw Aye Aye Maw
Associate Professor
Department of Statistics
Yangon University of Economics

Supervised by:

Daw Khin Moh Moh
Lecturer
Department of Statistics
Yangon University of Economics

NOVEMBER, 2019

YANGON UNIVERSITY OF ECONOMICS

DEPARTMENT OF STATISTICS

This is certify that the thesis entitled "**AWARENESS ON HEALTH HAZARD OF SMOKING AMONG STUDENTS IN YANGON UNIVERSITY OF ECONOMICS"** submitted as a partial fulfillment towards the requirements of Master of Economics (Statistics) has been accepted by the Board of Examiners.

BOARD OF EXAMINERS

(Chairman**)**
Professor Dr. Tin Win
Rector
Yangon University of Economics

Professor Dr. Ni Lar Myint Htoo
Pro-Rector
Yangon University of Economics

(Chief Examiner)
Prof. Dr. Maw Maw Khin
Professor and Head
Department of Statistics
Yangon University of Economics

(External Examiner)
Prof. Daw Mya Mya Win
Professor and Head (Retd.)
Department of Statistics
Monywa University of Econimics

(Examiner)
Prof. Dr. Mya Thandar
Professor
Department of Statistics
Yangon University of Economics

(Examiner)
Dr. Aye Thida
Associate Professor
Department of Statistics
Yangon University of Economics

NOVEMBER, 2019

ABSTRACT

In this study, awareness on health hazard of smoking among students in Yangon University of Economics was analyzed. The sample 256 students were selected from total 4337 undergraduate students at Yangon University of Economics by using stratified random sampling. In this study, canonical correlation analysis was used to express the relationship between socio-economic characteristics and awareness on health hazard of smoking. The factor analysis was used to analyze the awareness on health hazard of smoking that contributes significantly to the percentage of variance. In canonical correlation analysis, all of the multivariate tests of significance are statistically significant and six canonical roots are obtained. The results show that the socio-demographic information and the awareness of smoking are positively correlated. The factor analysis shows that the seven principal components are obtained to contribute the awareness on health hazard of smoking among students. It has been found that all principal components support the effects on the awareness on health hazard of smoking among students.

ACKNOWLEDGEMENTS

First of all, I would like to express my deepest gratitude to Professor Dr. Tin Win, Rector of Yangon University of Economics, Professor Dr. Ni Lar Myint Htoo, Pro-Rector of Yangon University of Economics, for their permission on this thesis.

I am greatly indebted to Professor Dr. Maw Maw Khin, Head of the Department of Statistics, Yangon University of Economics, for her permission, valuable suggestions and recommendations to prepare this thesis study. I also thank to Professor Dr. Mya Thandar, Professor Dr. Cing Do Nem, Associate Professor Dr. Aye Thida and Associate Professor Daw Khin Nu Win, Department of Statistics, Yangon University of Economics, for their valuable suggestions recommendations to improve my thesis.

I would like to acknowledge Professor Daw Mya Mya Win (Retd.), Head of Department of Statistics, Monywa University of Economics and Associate Professor, co-supervisor, Daw Aye Aye Maw, Department of Statistics, Yangon University of Economics for their beneficial guidance, in filling encouragements and valuable advices throughout the writing this thesis.

I would like to express my special thanks and high indebtedness to my supervisor, Daw Khin Moh Moh, Lecturer, Department of Statistics, for her valuable guidance, helpful advices and supervision.

I am extremely thankful to my family for their patience, encouragements and support on my study.

TABLE OF CONTENTS

		Page
ABSTRACT		**i**
ACKNOWLEDGEMENTS		**ii**
TABLE OF CONTENTS		**iii**
LIST OF TABLES		**vi**
LIST OF FIGURES		**vii**
LIST OF ABBREVIATIONS		**viii**
Chapter I	**INTRODUCTION**	**1**
	1.1 Rationale of the Study	1
	1.2 Objectives of the Study	2
	1.3 Method of Study	3
	1.4 Scope and Limitations of the Study	3
	1.5 Organization of the Study	3
Chapter II	**LITERATURE REVIEW**	**4**
	2.1 Smoking	4
	2.2 Hazard of Smoking	6
	2.3 Reviews on Related Studies	8
Chapter III	**THEORETICAL BACKGROUND**	**12**
	3.1 Canonical Correlation Analysis	12
	3.2 Canonical Variates and Canonical Correlations	12
	3.3 Assumptions in Canonical Correlation	15
	3.4 Key Terms of Canonical Correlation Analysis	16
	3.5 Significant Test	17
	3.5.1 Wilk's Lambda Test	17
	3.5.2 Hotelling's T-square Test	18
	3.5.3 Pillai's Trace Test	18
	3.6 Level of Significance	18
	3.7 Factor Analysis	18

	3.8	Model for Factor Analysis	19
	3.9	Principal Component Analysis	19
	3.10	The Principal Factor (Principal Component Factor)	20
	3.11	Assumption Testing	23
	3.12	Factor Rotation	24
		3.12.1 Oblique Rotation Method	25
		3.12.2 Varimax Rotation Method	25
	3.13	Testing of Sampling Adequacy of the Approach	26
		3.13.1 Bartlett's Test	27
Chapter IV	**ANALYSIS OF AWARENESS ON HAZARD OF SMOKING**		**28**
	4.1	Survey Design and Determining of the Sample Design	28
	4.2	Determination of the Sample Size	28
	4.3	General Characteristics of the Respondents	30
	4.4	Eigenvalues and Canonical Correlations	33
	4.5	Multivariate Test of Significance	34
	4.6	Dimension Reduction Analysis	34
	4.7	Canonical Loadings	35
	4.8	Factor Analysis	38
	4.9	Kaiser-Meyer-Olkin (KMO) Measure and Bartlett's Test of Sphericity	40
	4.10	The Number of Principal Components	40
	4.11	Principal Components through Component Matrix and Rotated Component Matrix	41
Chapter V	**CONCLUSION**		**44**
	5.1	Findings	44
	5.2	Recommendations	46
	REFERENCES		
	APPENDIX		

LIST OF TABLES

Table No.	Title	Page
4.1	Sample Sizes from Each Stratum	30
4.2	Students by Socio-demographic Information	30
4.3	Students by Awareness on Health Hazard of Smoking	32
4.4	Eigenvalues and Canonical Correlations	34
4.5	Multivariate Tests of Significance	34
4.6	Dimension Reduction Analysis	35
4.7	Canonical Loadings	36
4.8	Total Variance Explained by Each Component	39
4.9	Kaiser-Meyer-Olkin (KMO) and Bartlett's Test	40
4.10	Structure Matrix	42

LIST OF FIGURES

Figure No.	Title	Page
4.1	Screen Plot for Eigenvalues and Component Number	41

LIST OF ABBREVIATIONS

ARMD	-	Age Related Macular Degeneration
ADHD	-	Attention Deficit Disorder
KMO	-	Kaiser-Meyer-Olkin
MANOVA	-	Multivariate Analysis of Variance
OLS	-	Ordinary Least Squares
PCA	-	Principal Component Analysis
PCR	-	Principal Component Regression
SEA	-	South East Asia Region
SIDS	-	Sudden Infant Death Syndrome
WHO	-	World Health Organization

CHAPTER I

INTRODUCTION

1.1 Rationale of the Study

Tobacco use continues to be the leading global cause of preventable deaths, killing millions of people and causing hundreds of billions of dollars of economic damage worldwide each year. Use of Tobacco is one of the greatest public health threats for the 21st century. Globally, 1.3 billion persons use Tobacco in the world and most of them are men. Tobacco is the fourth most common risk factor for disease worldwide, and it continues to be the second major cause of death in the world. At present, 70% of the world's 1.1 billion smokers are in developing countries, with over 50% in Asia alone. Although the prevalence of smoking among most Western countries has been in steady decline for more than a decade, smoking rates continue to rise in many developing countries. (WHO SEA, 2007)

The Global burden of Diseases presented that increasing levels of smoking in many middle and low-income countries, where most of the world's 1.2 billion Tobacco users from cardiovascular diseases, chronic obstructive pulmonary diseases and some cancers. Starting smoking earlier in life and smoking cigarettes higher in tar increases the risk of these diseases. In addition, environmental smoke or second-hand smoke has been shown to cause adverse health effects in people of all ages. (WHO 2004)

In developing countries, many of the poorest smokers spend significant amounts of their income on Tobacco instead of basic human needs such as food, shelters, healthcare and education. The effects depend on the number of years that a person smokes and on how much the person smokes. As a consequence, the health and economic burden of Tobacco use is rapidly shifting from high to low and middle income countries. Tobacco's prominent role as a major health hazard and the likelihood of its health hazards increasing dramatically in the future, make it clear that the regular assessment of Tobacco use and associated disease trends should be an integral part of a country's health information system.

The rise in Tobacco use among younger females in high-population countries is one of the most ominous potential developments of the epidemic's growth. In many

countries, women have traditionally not use Tobacco: women smoke at about one-fourth the rate of men.

According to the report of World Health Organization, Tobacco is the only legal consumer product that can harm everyone exposed to it. Tobacco use is common throughout the world due to low prices, aggressive and widespread marketing, lack of awareness about its dangers, and inconsistent public policies against its use. Not only smoking but use of smokeless Tobacco can cause the negative impact on both general and oral health.

In Myanmar, Tobacco is also famous for good economy leading better business, showing modern lifestyle, stylish/fashionable and sometimes it is claimed to be a stress releasers by using Tobacco. It has been used to honor or as a special treat for a special visitor in ancient and tradition of Myanmar. People begin to smoke at younger age, with the median age of initiation under 15 years. The prevalence of smoking was very high among young people.

Tobacco is a major problem among low income groups in Myanmar. Studies showed that there is increasing trend of Tobacco use especially cheroots and smokeless Tobacco among Myanmar people. Working age group play major role for development of country and using Tobacco among them can damage their health which can lead to loss of productivity. Tobacco poses a major challenge not only to health but also to social and economic development and to environmental sustainability. This study focuses on smoking habits and awareness on health hazard of smoking among students in Yangon University of Economics by using Canonical Correlation Analysis and Factor Analysis.

1.2 Objectives of the Study

The main objectives of the study are:

(i) To determine the relationship between socio-economic characteristics and awareness on smoking among undergraduate students in Yangon University of Economics using Canonical Correlation Analysis.

(ii) To analyze the awareness on health hazard of smoking that contributes significantly to the percentage of variance using Factor Analysis.

1.3 Method of Study

This study was conducted using Canonical Correlation Analysis and the homogeneity of variance among the variables obtained with the use of Wilk's Lambda and Bartlett's test respectively. Moreover, Factor analysis was used to investigate the variability among the subjects and find out the variables that contribute significantly to the percentage of variance obtained.

1.4 Scope and Limitations of the Study

The questionnaires were asked to students who are attending at Yangon University of Economics in 2018-2019 academic years by using stratified random sampling.

1.5 Organization of the Study

This study is composed by five chapters with the available information, data and statistical facts and figures. Chapter I consists of introduction which includes Rationale of the Study, Objectives of the Study, Method of Study, Scope and Limitations of the Study and Organization of the Study. Chapter II describes Smoking, Hazard of Smoking and Literature Review. Chapter III presents Theoretical Background of Canonical Correlation Analysis and Factor Analysis. Chapter IV includes Analysis of Awareness on Hazard of Smoking. Chapter V contains Findings and Recommendations.

CHAPTER II

LITERATURE REVIEW

2.1 Smoking

Smoking is an act of inhaling and exhaling the fumes of burning plant material. A variety of plant materials are smoked, including marijuana and hashish, but the act is most commonly associated with Tobacco as smoked in a cigarette, cigar or pipe. Tobacco contains nicotine, an alkaloid that is addictive and can have both stimulating and tranquilizing psychoactive effects. The smoking of Tobacco, long practiced by American Indians, was introduced to Europe by Christopher Columbus and other explorers. Smoking soon spread to other areas and today is widely practiced around the world despite medical, social and religious arguments against it.

About one third of smokers started smoking because of peer pressure and another one third started smoking because they wanted to experiment. A few started smoking when their parents asked them to light cheroots and quite a few experimented because they thought it was stylish.

Other factors influencing decisions to smoke include working in the Tobacco industry or restaurants where Tobacco was easily or freely available, imitating models and actors from foreign films and movies, releasing anxiety and stress, wanting to kill the feeling of loneliness and trying to show their rebellious nature.

Among the current smokers, more than half of them smoked cheroots, about one in four smoked cigarettes and about one in five smoked both cheroots and cigarettes. Only a little over 1% smoked cigars and less than 1% smoked hand-rolled cheroots. There were less than 0.5% of pipe smokers.

Cheroots are cheap, most commonly used by the low income group and low education group. Among the sampled population, the majority of trishaw drivers and manual laborers who earned less than the others were cheroots smokers; whereas the majority of bus drivers and assistants who earned more than the rest of the sampled population were cigarette smokers.

Smoking both cigarettes and cheroots concurrently was also found to be common; among the sampled population there was a tendency to switch to cigarettes when they had more money to spend. So cheroots and cigarettes are the substituted materials. On the other hand, due to the general belief among the community that cheroots are less harmful than cigarettes, cigarette smokers had a tendency to switch to cheroots when they had symptoms such as cough and tightness of chest.

Among the cigarette smokers, London, Vegas, Duya and Golden Triangle were the most popular brands and among the cheroots *Three Lions Gold Strip* was the most popular. Golden Triangle cigarettes were popular in the coastal region whereas London and Vegas cigarettes were smoked in many part of the country. London is a product of British American Tobacco which is produced locally in a joint venture with Myanmar Economic Enterprise and Vegas is a product of Sympoerna Company of Indonesia, also produced locally as a joint venture. Popular brands of cheroots varied in different regions of the country although the brand "Three Lions" was smoked in many parts of the country.

About 85% of current smokers were daily smokers. The median age of starting to smoke was 18 years. A few started smoking as early as 5 years of age and about 21% started smoking before they were 15 years old. The median age of smoking daily was 20 years although about 15% became daily smokers below the age of 15 years. Daily smokers smoked an average of 5 sticks per day and occasional smokers smoked an average of 10 sticks per week.

Global monitoring of the Tobacco epidemic and comparisons between countries requires a standardization of terms and concepts that must be defined concisely. Key definitions are given below.

Smoker

A smoker is someone who, at the time of the survey, smokes any Tobacco product either daily or occasionally.

Daily Smoker

A daily smoker is someone who smokes any product at least once a day.

An Occasional Smoker

An occasional smoker is someone who smokes, but not every day. Occasional smokers include:

Non-smoker

A non-smoker is someone who, at the time of the survey, does not smoke at all. Non-smokers can be divided into three categories:

Ex-smokers

Ex-smokers are people who were formally daily smokers but currently do not smoke at all.

Never-smokers

Never-smokers are those who either have never smoked at all or have never been daily smokers and have smoked less than 100 cigarettes (or the equivalent amount of tobacco) in their lifetime.

Ex-occasional smokers

Ex-occasional smokers are those who were formerly occasional, but never daily, smokers who smoked 100 or more cigarettes (or their equivalent of tobacco) in their lifetime.

These definitions can be used to classify the population according to their lifetime smoking status. In particular, ever smokers are defined as those who ever smoked at least 100 cigarettes in their lifetime.

2.2 Hazard of Smoking

Cigarette smoking is the greatest cause of preventable deaths. On average, people who smoke die 5 to 10 years earlier than people who don't smoke. Recognizing the increasing and indisputable scientific evidence showing that Tobacco smoking is a major cause of chronic bronchitis, emphysema and lung cancer, as well as a major risk factor for myocardial infarction, certain pregnancy-related and neonatal disorders and a number of other serious health problems, and that it also has harmful effects on those who are involuntarily exposed to Tobacco smoke.

Seriously concerned at the alarming increase in production and consumption of cigarettes during the last two decades in some of the countries, particularly developing countries, in which it was previously not widespread, and at the extensive promotional drive for the sale of cigarettes being carried out on radio and television, in newspapers and other news media, and through association with sporting and cultural events, often inducing young people to smoke Tobacco.

Few countries have so far taken comprehensive action to effectively combat smoking through educational, restrictive and legislative measures for the control of publicity and advertisements in the news media, combined with coherent taxation and price policies for Tobacco cultivation and cigarette production.

Smoking increases the risk of many health problems, such as:

- Lung cancer (most people who have lung cancer are smokers or people who live with smokers).
- Other lung diseases, such as emphysema and asthma; colds and other respiratory infections more often
- Other cancers such as cancer of the esophagus, mouth, cervix or bladder
- Heart or blood vessel disease, high blood pressure and stroke
- Diabetes and high cholesterol
- Ulcers, hip fractures
- In the eye, smoking has been associated with an increases risk for Age Related Macular Degeneration (ARMD)

Smoking affects pregnant women and their unborn children. If a woman smokes while are pregnant, that woman has a greater risk of losing baby during pregnancy. The baby may have low birth weight and have trouble breathing at birth; as a child may have more respiratory infections, ear infections, asthma an attention deficit disorder. Moreover, the baby has a greater risk of dying from SIDS (sudden infant death syndrome).

The more cigarettes smoke each day, the greater risk of disease. Switching form cigarettes to a pipe or cigars may not lessen the risk of disease if a person continues to

inhale the smoke. Cigar and pipe smokers are at the same risk for cancers of the mouth, lip, larynx and esophagus as cigarette smokers.

2.3 Reviews on Related Studies

JANE F. EMMERÉ (2003) studied that "Social Factors of Cigarette Smoking Initiation among Under-graduate College Students". In this thesis, descriptive statistics, multiple logistic regression analyses and chi-square analysis are used to determine the percentage of students who initiated smoking during college and identified at-risk populations. Chi-square analysis confirms an adequate model fit and gives the smoking-related behavior of friends and roommates are statistically significant factor of smoking initiation. Multiple logistic regression analysis shows the smoking-related behavior of friends and roommates is statistically significant explanatory variable. The results can be interpreted as an increase in the odds for each unit increase in the explanatory variable.

Myat Thura Khaing (2005) studied "A Study of Smoking Habit (A Case Study of United K. M. K Co. Ltd.)". In this study, non-interventional exploratory study was used to find out the magnitude of problem and consequences of smoking among workers in Myanmar. According to the results, it was found that the knowledge of workers relating to the consequences of smoking should be expanded and by doing this attitude of worker may be on more positive side. The worker might become more healthiness, more efficient and might contribute better performance for the country. The smoking related disease could affect burden to worker and deficient of job performance and smokers smoked on adolescent period at deficient knowledge and attitude. The result showed that the burden of cost of smoking and cost of smoking related disease was a major problem not only for individual family but also for the whole country.

Lukman AbdulSalam (2006) analyzed "Determination of Students' Academic Performance". The statistical methods are canonical correlation analysis and factor analysis. The aim of this thesis is to evaluate Students' Academic Performance through the use of the relationship between mathematical and less-mathematical subjects. There are two sets of variables such as set 1 consists of Mathematics, Chemistry and Physics are classified as the mathematical subjects and set 2 consists of Economics, English Language, Biology, Geography and Hausa Language are classified as less-mathematical subjects. The results showed less-mathematical subjects have significant impact on determining students' academic performance. From the results, three canonical roots are

obtained and two are statistically significant showing a strong correlation between the two sets. Four factors showed that less-mathematical subjects significantly to the variation among the variables. Mathematical subjects are directly related with less-mathematical subjects.

Nouran Mahmoud (2011) made "Self-related Health Status and Smoking". In this study, statistical techniques have been used such as cross-sectional study, bivariate analyses, multivariate logistic regression analysis and chi-square analyses. A cross-sectional study was conducted to evaluate the relationship between self-rated health status and smoking status. Using the bivariate analyses, there is a statistically significant relationship between self-rated health status and the exposure variables smoking status, age group, marital status, education, race, poverty level and physical activity. And also, chi-square analyses revealed that there is a statistically significant relationship between self-rated health and the following exposure variables: smoking status, age group, marital status, BMI, education, race, poverty level and physical activity. There is a significant relationship between self-rated health and age group and between self-rated health status and marital status and between self-rated health status and race. But, there is no effect of race on the relationship between smoking status and self-rated health.

Isaac Kusi Appau (2011) analyzed "Smoking Habits among Adolescents". In this study, various statistical techniques have been used such as Logistic regression and the Chi-square test. The aim of this study are finding the reasons why adolescent start smoking. As a result, friends are an important influence in the smoking behaviors of adolescents. The study found that adolescents whose friends smoked were more at risk than those whose friends did not smoke. And also, the study found that adolescents with psychological problems have a high chance of initiating smoking.

Mandisa Malinga (2011) analyzed that "Smoking Prevalence, Knowledge, Attitudes and Beliefs about Health Risks of Tobacco Smoking among Female Psychology Students at the University of the Western Cape". In this thesis, descriptive and inferential statistics were used to analyze the data. This study indicated that the smoking and non-smoking students had sufficient knowledge of the effects of smoking and there are some differences between smokers and non-smokers in knowledge, attitudes and beliefs. This means that in order to design effective interventions, the knowledge, attitudes and beliefs of that particular population have to be established. There is a need for interventions at

higher education institutions that will aim at reducing the smoking rates of students in particular.

Kyi Win (2012) studied that "A Study on Youth Smoking and Tobacco Consumption in South Dagon Township". This paper was explored tobacco-related knowledge, view, behavior and environmental influences on youth living in South Dagon Township. Aged 15-24 years, 400 respondents, 96% was male. The proportion of youth was found to increase as they grew older: the percentage ever smoked increased from 23% for those aged 15-18 years to 41% for those aged 21-24 years.

Mark Aaron Thompson (2014) analyzed that "Smokers' Awareness and Perceptions of Electronic Cigarettes". The aim of this study is to assess tobacco smokers' awareness of e-cigarettes and their attitudes toward their use. In this study, the independent variables are the number of cigarettes smoked per day, the desire to quit smoking, and age. Dependent variables are awareness and positive perceptions of e-cigarettes. By using chi-square test, there is a correlation between a higher number of cigarettes smoked per day and increased awareness and knowledge about e-cigarettes. The results showed that there was not a statistically significant association between the number of cigarettes smoked per day and awareness of e-cigarettes. Therefore, the findings showed that there exist statistically significant relationships between positive perceptions of e-cigarette. And also, there is a relationship between the desire to quit smoking and positive perceptions of the effectiveness of e-cigarettes. There was no association between the number of cigarettes smoked per day and awareness of e-cigarettes.

Yan Naing Myint (2015) analyzed that "A Study of Awareness and Practice of Smoking among Youths (Case Study in University Students)". The aim of this study was to assess smoking habit and knowledge concerning with smoking related diseases among university students. The study found that more than 75% of students had normal knowledge and there was a relationship between knowledge level concerning with smoking related diseases and practice of smoking among university students. And also, the influencing factors of youth smoking were types of advertising and promotion, selling cigarette in loose form and easy accessible to cigarette which were the influencing factors of smoking in youths were found as a results.

Xianglong Xu, Doris Yin Ping Leung, Bing Li, Pengfei Wang and Yong Zhao (2015) studied that "Smoking-Related Knowledge, Attitude, Social Pressure and Environmental Constraints among New Undergraduates in Chongqing, China". In this thesis, Chi-square test and Logistic regression were used in data analysis. The aim of this study is to examine the smoking-related behaviors of undergraduates such as knowledge, attitude, social pressure, and environmental constraints. There were 196 males and 193 females participants in this thesis. By using Chi-square test, the smoking-related variables between smokers and non-smokers are compared. When Logistic regression is used, it examines the factors that associated with smoking status in undergraduates. As a result, there is statistically significant the differences between the age of smokers and non-smokers. The comparison between the physical condition of smokers and non-smokers yielded no statistically significant difference. This study showed that smoking-related behavior is associated with smoking-related attitude, social pressure and environmental constraints.

Bhuyan KC and Urmi AF (2018) analyzed "Canonical Correlation Analysis to Study the Impacts of Different Social Factors on Awareness of Health Hazard of Tobacco Smoking and Smoking Habit". In this thesis, Canonical correlation analysis is used to study the complex relationship of awareness and smoking habit with other socioeconomic variables. The analysis indicates that important variables for complex relationship of awareness and smoking habit are sex and marital status. The differentials in smoking habit among males and females are statistically significant but the differentials in awareness among males and females are not significant. And also, the study indicates that smoking is highly prevailed among males compared to females but both males and females are similarly aware of the health hazard of smoking. Awareness and smoking habit are negatively significantly associated. Moreover, the study indicates that awareness and smoking habit are highly interrelated and the analysis indicates that sex of respondents and their smoking habit and marital status and awareness are significantly interrelated.

CHAPTER III

THEORETICAL BACKGROUND

In this chapter, the theoretical background of the statistical techniques such as Canonical Correlation Analysis, Canonical Variates and Canonical Correlations, Assumptions in Canonical Correlation, Key Terms of Canonical Correlation Analysis, Significant Test, Level of Significance, Factor Analysis, Models for Factor Analysis, Principal Component Analysis, The Principal Factor (Principal Component Factor), Assumption Testing, Factor Rotation and Testing of Sampling Adequacy of the Approach are presented.

3.1 Canonical Correlation Analysis

Canonical correlation analysis focuses on the correlation between a linear combination of the variables in one set and a linear combination of the variables in another set. The linear combinations are called the canonical variables, and their correlations are called canonical correlations. The canonical correlations measure the strength of association between the two sets of variables. The maximization aspect of the technique represents an attempt to concentrate a high-dimensional relationship between two sets of variables into a few pairs of canonical variables.

3.2 Canonical Variates and Canonical Correlations

The first group, of p variables, is represented by the $(p \times 1)$ random vector $X^{(1)}$. The second group, of q variables, is represented by the $(q \times 1)$ random vector $X^{(2)}$. $X^{(1)}$ represents the smallest set, so that $p \leq q$.

For the random vectors $X^{(1)}$ and $X^{(2)}$, let

$$\begin{aligned} E(X^{(1)}) &= \mu^{(1)}; \qquad Cov(X^{(1)}) = \Sigma_{11} \\ E(X^{(2)}) &= \mu^{(2)}; \qquad Cov(X^{(2)}) = \Sigma_{22} \\ Cov(X^{(1)}, X^{(2)}) &= \Sigma_{12} = \Sigma_{21}' \end{aligned} \tag{3.2.1}$$

The random vector

$$X = \begin{bmatrix} X^{(1)} \\ \cdots\cdots \\ X^{(2)} \end{bmatrix} = \begin{bmatrix} X_1^{(1)} \\ X_2^{(1)} \\ \vdots \\ X_p^{(1)} \\ \cdots\cdots \\ X_1^{(2)} \\ X_2^{(2)} \\ \vdots \\ X_q^{(2)} \end{bmatrix} \tag{3.2.2}$$

has mean vector

$$\mu = E(X) = \begin{bmatrix} E\left(X^{(1)}\right) \\ \cdots\cdots \\ E\left(X^{(2)}\right) \end{bmatrix} = \begin{bmatrix} \mu^{(1)} \\ \cdots\cdots \\ \mu^{(2)} \end{bmatrix} \tag{3.2.3}$$

and covariance matrix

$$\Sigma = E(X-\mu)(X-\mu)'$$

$$= \begin{bmatrix} E\left(X^{(1)}-\mu^{(1)}\right)\left(X^{(1)}-\mu^{(1)}\right)' & E\left(X^{(1)}-\mu^{(1)}\right)\left(X^{(2)}-\mu^{(2)}\right)' \\ E\left(X^{(2)}-\mu^{(2)}\right)\left(X^{(1)}-\mu^{(1)}\right)' & E\left(X^{(2)}-\mu^{(2)}\right)\left(X^{(2)}-\mu^{(2)}\right)' \end{bmatrix}$$

$$= \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \tag{3.2.4}$$

The covariance between pairs of variables from different sets one variable from $X^{(1)}$, one variable from $X^{(2)}$ are contained in Σ_{12}or, equivalently, in Σ_{21}. That is the pq elements of Σ_{12} measure the association between the two sets. When p and q are relatively large, interpreting the elements of Σ_{12} collectively is ordinarily hopeless. Moreover, it is often linear combinations of variables that are interesting and useful for predictive or comparative purposes. The main task of canonical correlation analysis is to summarize the associations between the $X^{(1)}$ and $X^{(2)}$ sets in terms of a few carefully chosen covariance (or correlations) rather than the pq covariance in Σ_{12}.

Linear combinations provide simple summary measures of a set of variables. Set

$$U = a'X^{(1)}$$

$$V = b'X^{(2)} \tag{3.2.5}$$

for some pair of coefficient vectors a and b.

$$\begin{aligned} Var(U) &= a'Cov\left(X^{(1)}\right)a = a'\Sigma_{11}a \\ Var(V) &= b'Cov\left(X^{(2)}\right)b = b'\Sigma_{22}b \end{aligned}$$

$$Cov(U,V) = a'Cov\left(X^{(1)}, X^{(2)}\right)b = a'\Sigma_{12}b \tag{3.2.6}$$

Coefficient vectors a and b such that

$$Corr(U,V) = \frac{a'\Sigma_{12}b}{\sqrt{a'\Sigma_{11}a}\sqrt{b'\Sigma_{22}b}} \tag{3.2.7}$$

is as large as possible.

It defines the following:

The first pair of canonical variable, or first canonical variate pair, is the pair of linear combinations U_1, V_1 having unit variances, which maximize the correlation.

The second pair of canonical variable, or second canonical variate pair, is the pair of linear combinations U_2, V_2 having unit variances, which maximize the correlation among all choices that are uncorrelated with the first pair of canonical variables.

At the k^{th} step,

The k^{th} pair of canonical variable, or k^{th} canonical variate pair, is the pair of linear combinations U_k, V_k having unit variances, which maximize the correlation among all choices that are uncorrelated with the previous k-1 canonical variable pairs.

The correlation between the k^{th} pair of canonical variables is called the k^{th} canonical correlation.

For coefficient vectors a and b, from the linear combinations $U = a'X^{(1)}$ and $V = b'X^{(2)}$. Then $\max Corr(U,V) = \rho_1^*$ attained by the linear combinations (first canonical variate pair)

$$U_1 = e_1'\Sigma_{11}^{-\frac{1}{2}}X^{(1)} \quad \text{and} \quad V_1 = f_1'\Sigma_{22}^{-\frac{1}{2}}X^{(2)} \tag{3.2.8}$$

The k^{th} pair of canonical variates, $k = 2, 3, \dots, p$,

$\mathrm{U_k} = \mathrm{e_k'}\Sigma_{11}^{-\frac{1}{2}}\mathrm{X}^{(1)}$ and $\mathrm{V_k} = \mathrm{f_k'}\Sigma_{22}^{-\frac{1}{2}}\mathrm{X}^{(2)}$ maximizes $\mathrm{Corr}(\mathrm{U}_k, \mathrm{V}_k) = \rho_{\mathrm{k}}^{*}$ among those linear combinations uncorrelated with the preceding $1, 2, \dots, k-1$ canonical variables.

The canonical variates have the properties

$$\mathrm{Var}(\mathrm{U_k}) = \mathrm{Var}(\mathrm{V_k}) = 1$$

$$\mathrm{Cov}(\mathrm{U_k}, \mathrm{U}_\ell) = \mathrm{Corr}(\mathrm{U_k}, \mathrm{U}_\ell) = 0, \qquad \mathrm{k} \neq \ell$$

$$\mathrm{Cov}(\mathrm{U_k}, \mathrm{V}_\ell) = \mathrm{Corr}(\mathrm{U_k}, \mathrm{V}_\ell) = 0, \qquad \mathrm{k} \neq \ell \text{ for } k, \ell = 1, 2, \dots, p. \tag{3.2.9}$$

If the original variables are standardized with $\mathrm{Z}^{(1)} = \left[\mathrm{Z_1}^{(1)}, \mathrm{Z_2}^{(1)}, \dots, \mathrm{Z_p}^{(1)}\right]'$ and $\mathrm{Z}^{(2)} = \left[\mathrm{Z_1}^{(2)}, \mathrm{Z_2}^{(2)}, \dots, \mathrm{Z_q}^{(2)}\right]'$ from first principles, the canonical variates are of the form

$$\mathrm{U_k} = \mathrm{a_k'}\mathrm{Z}^{(1)} = \mathrm{e_k'}\rho_{11}^{-\frac{1}{2}}\mathrm{Z}^{(1)}$$

$$\mathrm{V_k} = \mathrm{b_k'}\mathrm{Z}^{(2)} = \mathrm{f_k'}\rho_{22}^{-\frac{1}{2}}\mathrm{Z}^{(2)} \tag{3.2.10}$$

Here, $\mathrm{Cov}\left(\mathrm{Z}^{(1)}\right) = \rho_{11}, \mathrm{Cov}\left(\mathrm{Z}^{(2)}\right) = \rho_{22}, \mathrm{Cov}\left(\mathrm{Z}^{(1)}, \mathrm{Z}^{(2)}\right) = \rho_{12} = \rho_{21}'$ and e_k and f_k are the eigenvectors of $\rho_{11}^{-\frac{1}{2}}\rho_{12}\rho_{22}^{-1}\rho_{21}\rho_{11}^{-\frac{1}{2}}$ and $\rho_{22}^{-\frac{1}{2}}\rho_{21}\rho_{11}^{-1}\rho_{12}\rho_{22}^{-\frac{1}{2}}$, respectively. The canonical correlations, ρ_{k}^{*}, satisfy

$$\mathrm{Corr}(U_k, V_k) = \rho_{\mathrm{k}}^{*}, \quad k = 1, 2, \dots, p \tag{3.2.11}$$

Where; $\rho_1^{*2} \geq \rho_2^{*2} \geq \cdots \geq \rho_p^{*2}$ are the nonzero eigenvalues of the matrix $\rho_{11}^{-\frac{1}{2}}\rho_{12}\rho_{22}^{-1}\rho_{21}\rho_{11}^{-\frac{1}{2}}$ (or, equivalently, the largest eigenvalues of $\rho_{22}^{-\frac{1}{2}}\rho_{21}\rho_{11}^{-1}\rho_{12}\rho_{22}^{-\frac{1}{2}}$).

3.3 Assumptions in Canonical Correlation

The generality of canonical correlation analysis also extends to its underlying statistical assumptions. Some essential assumptions and their impacts on canonical correlation analysis are discussed.

(1) Normality

Canonical correlation analysis can accommodate any metric variable without the strict assumption of normality. However, normality is desirable because it allows for the highest correlation among the variables. Indeed, canonical correlation analysis can accommodate non-normal variables if the distributional form does not decrease the correlation with other variables. This allows for transformed nonmetric data (in the form of dummy variables) to be used as well. However, multivariate normality is required for the statistical inference test of the significance of each canonical function. Because tests for multivariate normality are not readily available, the prevailing guideline is to ensure that each variable has univariate normality. Thus, although normality is not strictly required, it is highly recommended that all variables be evaluated for normality and transformed if possible.

(2) Homoscedasticity and Multicollinearity

Canonical correlation analysis best portrays the relationships when they are homoscedastic. Homoscedasticity is important because the opposite, heteroscedasticity, decreases the correlation between variables. Similarly, multicollinearity should be dealt with as well. Multicollinearity occurs when two or more variables are highly correlated. Multicollinearity among either variable set will confound the ability of the technique to isolate the impact of any single variable, making interpretation less reliable.

3.4 Key Terms of Canonical Correlation Analysis

Canonical Correlation Coefficient

Canonical correlation coefficient measures the strength of the overall relationship between the two linear composites (canonical variates), one variate for the independent variables and one for the dependent variables. In effect, it represents the bivariate correlation between the two canonical variates in a canonical function.

Canonical Function

Canonical function is the relationship (correlational) between two linear composites (canonical variates). Each canonical function has two canonical variates, one for the set of dependent variables and one for the set of independent variables. The strength of the relationship is given by the canonical correlation coefficient.

Structure Correlation Coefficients (Canonical Loading)

This can also be called canonical factor loading. A structure correlation is the correlation of a canonical variable with the standardized scores of an original input variable. The table of structure correlation is sometimes called the factor structure. The squared structure correlation indicates the contribution made by a given variable to the explanatory power of the canonical variate based on the set of variables to which it belongs. Canonical loading is the simple linear correlation between the independent variables and their respective canonical variates. These can be interpreted like factor loadings; they are also known as canonical structure correlations. Each independent variable has different canonical loadings for each canonical function.

Canonical Roots

Canonical roots are the squared canonical correlation coefficients, which provide an estimate of the amount of shared variance between the respective canonical variates of dependent and independent variables. It is also known as eigenvalues.

Eigenvalue and Eigenvector

Let A be a symmetric matrix and if there exists a scalar λ such that $|A - \lambda I| = 0$, then the roots of the determinant equation are called the eigenvalues (or characteristic values or latent roots) of the matrix A. If also there is a vector X such that $(A - \lambda I) = 0$, such that $X \neq 0$, then the columns of X which satisfy the equation are called the eigenvectors (or characteristic root) of A.

3.5 Significant Test

The significant test of the canonical correlation is straightforward in principle. Simply stated, the different canonical correlations are related, one by one, beginning with the largest one. Only those roots that are statistically significant are then retained for subsequent interpretation. First, evaluate the significant of all roots combined, then of the roots remaining after removing the first roots, the second root and so on.

3.5.1 Wilk's Lambda Test

Wilk's Lambda test is a test statistics used in multivariate analysis of variance (MANOVA) to test whether there are differences between the means of identified groups

of subjects on a combination of dependent variables that are unaccounted for by the independent variables. Wilk's Lambda statistics can be transformed to a statistic which has approximately on F distribution. This makes it easier to calculate p-value.

3.5.2 Hotelling's T-square Test

The Hotelling-Lawley trace is generally converted to the Hotelling's T-square. Hotelling's T-square is used when the independent variable forms two groups and represents the most significant linear combination of the dependent variables.

3.5.3 Pillai's Trace Test

Pillai's trace is one of several test statistics used in MANOVA. This is a positive valued statistic ranging from 0 to 1.

3.6 Level of Significance

The level of significance of a canonical correlation generally considered to be the minimum acceptable for interpretation is the 0.05 level, which has become the generally accepted level for considering a correlation coefficient statistically significant. This consensus has developed largely because of the availability of tables for these levels. These levels are not necessarily required in all situations.

3.7 Factor Analysis

Factor analysis is a statistical technique used to find a set of unobserved, also known as latent variables or factors that can account for the covariance among a larger set of observed, also known as manifest variables. A factor is an unobservable variable that is assumed to influence observed variables. Factor analysis is also used to assess the validity and reliability of measurement scales. Through factor analysis, the underlying dimensions of the observed variables and the variables corresponding to each of the underlying dimensions can be identified. These underlying dimensions are the continuous latent variables or factors and the observed variables are the factor indicators. There are two types of factor analysis that are exploratory factor analysis and confirmatory analysis.

Factor analysis is a data reduction technique used to reduce a large number of variables to a smaller set of underlying factors that summarize the essential information contained in the variables. More frequently, factor analysis is used as an exploratory

technique when the researcher wishes to summarize the structure of a set of variables. However, for testing a theory about the structure of a particular domain, confirmatory factor analysis is appropriate.

3.8 Models for Factor Analysis

The basic idea underlying factor analysis is that p observed random variables, x, can be expressed, expect for an error term, as linear functions of m $(< p)$ hypothetical (random) variables of common factors, that is if $x_1, x_2, x_3, \dots, x_p$ are the variables and $f_1, f_2, f_3, \dots, f_m$ are the factors, then

$$x_1 = \lambda_{11} f_1 + \lambda_{12} f_2 + \cdots + \lambda_{1m} f_m + e_1$$

$$x_2 = \lambda_{21} f_1 + \lambda_{22} f_2 + \cdots + \lambda_{2m} f_m + e_2$$

$$\vdots$$

$$x_p = \lambda_{p1} f_1 + \lambda_{p2} f_2 + \cdots + \lambda_{pm} f_m + e_p \qquad (3.8.1)$$

Where $\lambda_{jk},\ j = 1, 2, \dots, p; k = 1, 2, \dots, m$ are constant called the factor loadings, and $e_j, j = 1, 2, \dots, p$ are error terms, sometimes called specific factors. Equation (3.8.1) can be rewritten in matrix form, with obvious notation, as

$$\mathbf{x} = \mathbf{\Lambda f} + \mathbf{e} \qquad (3.8.2)$$

One contrast between PCA and factor analysis is immediately apparent. Factor analysis attempts to achieve a reduction from p to m dimensions by invoking a model relating $x_1, x_2, x_3, \dots, x_p$ to m hypothetical or latent variables.

3.9 Principal Component Analysis

Principal Component Analysis (PCA) is probably the most popular multivariate statistical technique and it is used by almost all scientific disciplines. It is also likely to be the oldest multivariate techniques. Principal Components Regression (PCR) is a method for combating multicollinearity and result is estimation and prediction better than ordinary least squares when used successfully. With this method, the original k variables are transformed into a new set of orthogonal or uncorrelated variables called principal components of the correlation matrix. This transformation ranks the new orthogonal variables in order of their importance and the procedure then involves eliminating some

of the principal components to effect a reduction in variance. After eliminating some of the principal components, a multiple regression analysis of the response variable against the reduce set of principal components is performed using Ordinary Least Squares estimation (OLS). Because the principal components are orthogonal, they are pair-wise independent and hence OLS is appropriate. Once the regression coefficients for the reduced set of orthogonal variables have been calculated, they are mathematically transformed into a new set of coefficients that correspond to the original or initial correlated set of variables. These new coefficients are principal component estimators.

3.10 The Principal Factor (Principal Component Factor)

The spectral decomposition provides with one factoring of the covariance matrix **Σ**. Let **Σ** have eigenvalue-eigenvector pairs $(\lambda_i, e_i)\ with\ \lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq 0.$

Then $\Sigma = \lambda_1 e_1 e_1' + \lambda_2 e_2 e_2' + \lambda_3 e_3 e_3' + \cdots + \lambda_p e_p e_p'$

$$\Sigma = \left[\sqrt{\lambda_1 e_1}\ \ \sqrt{\lambda_2 e_2}\ \ \sqrt{\lambda_3 e_3}\ \ \ldots\ \ \sqrt{\lambda_p e_p}\right]\begin{bmatrix}\sqrt{\lambda_1 e_1}' \\ \sqrt{\lambda_2 e_2}' \\ \sqrt{\lambda_3 e_3}' \\ \vdots \\ \sqrt{\lambda_p e_p}'\end{bmatrix} \qquad (3.10.1)$$

The fits the prescribed covariance structure for the factor analysis model having as many factors as variables m = p and specific variances $\psi_i = 0$ for all i. The loading matrix has j^{th} column given by $\sqrt{\lambda_i e_i}$. This can be written

$$\mathbf{\Sigma} = \mathbf{LL}' + \mathbf{0} \qquad (3.10.2)$$

A part from the scale factor$\sqrt{\lambda_i}$, the factor loadings on the j^{th} factor ate the coefficients for the population j^{th} principal component.

Although the factor analysis representation of **Σ** is exact, it is not particularly useful. It employs as many common factors as there are available and does not allow for any variation in the specific factors. These models explain the covariance structure in terms of just a few common factors. One approach, when the last p-m eigenvalues are small, is to neglect the contribution of $\lambda_m e_m e_{m+1}' + \cdots + \lambda_p e_p e_p'$ to **Σ**. Neglecting this contribution, the approximation is obtained.

$$\Sigma = \left[\sqrt{\lambda_1 e_1} \;\; \sqrt{\lambda_2 e_2} \;\; \sqrt{\lambda_3 e_3} \;\ldots\; \sqrt{\lambda_p e_p}\right] \begin{bmatrix} \sqrt{\lambda_1 e_1}' \\ \sqrt{\lambda_2 e_2}' \\ \sqrt{\lambda_3 e_3}' \\ \vdots \\ \sqrt{\lambda_p e_p}' \end{bmatrix} \tag{3.10.3}$$

$$\Sigma = \; L\, L'$$

The approximate representation assumes that the specific factors $\boldsymbol{\varepsilon}$ are minor importance and can also be ignored in the factoring of $\boldsymbol{\Sigma}$.

Allowing for specific factors, the approximation becomes

$$\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}$$

$$\Sigma = \left[\sqrt{\lambda_1 e_1} \;\; \sqrt{\lambda_2 e_2} \;\; \sqrt{\lambda_3 e_3} \;\ldots\; \sqrt{\lambda_p e_p}\right] \begin{bmatrix} \sqrt{\lambda_1 e_1}' \\ \sqrt{\lambda_2 e_2}' \\ \sqrt{\lambda_3 e_3}' \\ \vdots \\ \sqrt{\lambda_p e_p}' \end{bmatrix} + \begin{bmatrix} \psi_1 & 0 & 0 & \ldots & 0 \\ 0 & \psi_2 & 0 & \ldots & 0 \\ 0 & 0 & \psi_3 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & \psi_p \end{bmatrix} \tag{3.10.4}$$

Where $\psi_i = \sigma_{ii} - \sum_{j=1}^{m} \ell_{ij}^2$ for $i = 1,2,3, \ldots, p$

To apply this approach to a data set $x_1, x_2, x_3, \ldots, x_n$ it is customary first to center the observations by subtracting the sample mean $\bar{x}$. The centered observations

$$x_j - \bar{x} = \begin{bmatrix} x_{j1} \\ x_{j2} \\ x_{j3} \\ \vdots \\ x_{jp} \end{bmatrix} - \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_3 \\ \vdots \\ \bar{x}_p \end{bmatrix} = \begin{bmatrix} x_{j1} - \bar{x}_1 \\ x_{j2} - \bar{x}_2 \\ x_{j3} - \bar{x}_3 \\ \vdots \\ x_{jp} - \bar{x}_p \end{bmatrix}, \quad nj = 1,2,3, \ldots, n \tag{3.10.5}$$

have the same sample covariance matrix, **S**, as the original observations.

In cases, where the units of the variables are not commensurate, it is usually describable to work with the standardized variables.

$$Z_j = \begin{bmatrix} \frac{x_{j1}-\bar{x}_1}{\sqrt{s_{11}}} \\ \frac{x_{j2}-\bar{x}_2}{\sqrt{s_{22}}} \\ \frac{x_{j3}-\bar{x}_3}{\sqrt{s_{33}}} \\ \vdots \\ \frac{x_{jp}-\bar{x}_p}{\sqrt{s_{pp}}} \end{bmatrix}, \quad j = 1,2,3,\dots,n \tag{3.10.6}$$

The sample covariance matrix is the sample correlation matrix, **R**, of the observations $x_1, x_2, x_3, \dots, x_n$.

The principal component factor analysis of the sample covariance matrix **S** is specifiedin terms of its eigenvalue-eigenvector pairs$(\hat{\lambda}_1, \hat{e}_1), (\hat{\lambda}_2, \hat{e}_2), (\hat{\lambda}_3, \hat{e}_3), \dots, (\hat{\lambda}_p, \hat{e}_p)$ where $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \hat{\lambda}_3 \geq \cdots \geq \hat{\lambda}_p$.

Let $m < p$ be the number of common factors. Then the matrix of estimated factor loadings $(\tilde{\ell}_{ij})$ is given by

$$\tilde{L} = \left[\sqrt{\hat{\lambda}_1, \hat{e}_1} \;\; \sqrt{\hat{\lambda}_2, \hat{e}_2} \;\; \sqrt{\hat{\lambda}_3, \hat{e}_3} \;\; \dots \;\; \sqrt{\hat{\lambda}_p, \hat{e}_p}\right] \tag{3.10.7}$$

The estimated specific variances are provided by the diagonal elements of the matrix $\mathbf{S} - \tilde{\mathbf{L}}\tilde{\mathbf{L}}'$, so

$$\tilde{\Psi} = \begin{bmatrix} \psi_1 & 0 & 0 & \dots & 0 \\ 0 & \psi_2 & 0 & \dots & 0 \\ 0 & 0 & \psi_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \psi_p \end{bmatrix} \tag{3.10.8}$$

With $\psi_{ij} = s_{ii} - \sum_{i=1}^{m} \tilde{\ell}_{ij}^2$

Communalities are estimated as

$$\tilde{h}_i^2 = \tilde{\ell}_{i1}^2 + \tilde{\ell}_{i2}^2 + \tilde{\ell}_{i3}^2 + \cdots + \tilde{\ell}_{im}^2$$

The principal component factor analysis of the sample correlation matrix is obtained by starting with **R** in place **S**.

Consider the residual matrix,

$$S - \left(\tilde{L}\tilde{L} + \tilde{\Psi}\right)$$

resulting from the approximation of **S** by the principal component solution. The diagonal elements are zero. Analytically,

$$\text{Sum of squared entries of } \left(S - \left(\tilde{L}\tilde{L} + \tilde{\Psi}\right)\right) \leq \hat{\lambda}_{m+1} \leq \cdots \leq \hat{\lambda}_p$$

The contribution to the sample variance s_{ii} from the first common factor is $\tilde{\ell}_{i1}^2$. The contribution to the total sample variances s_{11}, s_{22}, s_{33}, ..., $s_{pp} = \text{tr}(\mathbf{S})$, from the first common factor is then

$$\tilde{\ell}_{11}^2 + \tilde{\ell}_{21}^2 + \tilde{\ell}_{31}^2 + \cdots + \tilde{\ell}_{p1}^2 = \left(\sqrt{\hat{\lambda}_1}\hat{e}_1\right)'\left(\sqrt{\hat{\lambda}_1}\hat{e}_1\right) = \hat{\lambda}_1 \qquad (3.10.9)$$

Since the eigenvector $\hat{e}_1$ has unit length.

In general,

$$\begin{pmatrix}\text{Proportion of the total}\\ \text{sample variance}\\ \text{due to j}^{\text{th}}\text{factor}\end{pmatrix} = \begin{cases}\dfrac{\hat{\lambda}_j}{s_{11} + s_{22} + s_{33} + \cdots + s_{pp}} & \text{for a factor analysis of S}\\ \dfrac{\hat{\lambda}_j}{p} & \text{for a factor analysis of R}\end{cases}$$

Criterion is frequently used as a heuristic device for determining the appropriate number of common factors. The number of common factors retained in the model is increased until a suitable proportion of the total sample variance has been explained.

3.11 Assumption Testing

There are a number of assumptions and practical considerations underlying the application of Principal components and Principal axis factoring.

1. Sample size – a minimum of five subjects per variable is required for factor analysis. A sample of 100 subjects is acceptable, but sample sizes of 200+ are preferable.

2. Normality – factor analysis is robust to assumptions of normality. However, if variables are normally distributed, then the solution is enhanced.

3. Multicollinearity and singularity – this assumption is not relevant for principal components. However, in the case of principal axis factoring, singularity and multicollinearity can be identified if any of the squared multiple correlations are near or equal to 1. If this is the case, the inclusion of the offending variables needs to be reconsidered.
4. Factorability of the correlation matrix – a correlation matrix that is appropriate for factor analysis will have several sizeable correlations. Variables with a measure of sampling accuracy that falls below the acceptable level of 0.5 should be excluded from the analysis. Bartlett's test of sphericity and the Kaiser-Meyer-Olkin measure of sampling adequacy are both tests that can be used to determine the factorability of the matrix as a whole. If Bartlett's test of sphericity is large and significant, and if the Kaiser-Meyer-Olkin measure is greater than 0.6, then factorability is assumed.

3.12 Factor Rotation

All factor loadings obtained from the initial loadings by an orthogonal transformation have the same ability to reproduce the covariance matrix. From matrix algebra, an orthogonal transformation corresponds to a rigid rotation of the coordinate axes. An orthogonal transformation of the factor loadings as well as the implied orthogonal transformation of the factors is called factor rotation.

If $\hat{\mathbf{L}}$ is the p×m matrix of estimated factor loadings obtained by any method then

$$\hat{\mathbf{L}} *= \hat{\mathbf{L}}\mathbf{T} \qquad \text{where} \qquad \mathbf{T}\mathbf{T}' = \mathbf{T}'\mathbf{T} = \mathbf{I}$$

is a p × m matrix of rotated loadings.

The estimated covariance matrix remains unchanged, since

$$\mathrm{L}\hat{\mathrm{L}} + \Psi = \hat{\mathrm{L}}\mathrm{T}'\mathrm{T}\hat{\mathrm{L}} + \Psi = \hat{\mathrm{L}} * \hat{\mathrm{L}} *' + \Psi$$

Indicates that the residual matrix, $\mathrm{S_n} - \mathrm{L}\hat{\mathrm{L}} - \Psi = \mathrm{S_n} - \hat{\mathrm{L}} * \hat{\mathrm{L}} *' - \Psi$ remains unchanged. Moreover, the specific variances Ψ_i and the communalities h_i^2 are unaltered. Thus, it is immaterial whether $\mathbf{L}$ or $\hat{\mathbf{L}} *$ is obtained.

Since the original loadings may not be readily interpretable, it is usual practice to rotate them until a simpler structure is achieved. Each variable loads highly on a single

factor and has small to moderate loadings on the remaining factors. It is possible to get this simple structure and the rotated loadings for the decathlon data provide a clearly pattern.

Rotation method is factor structure more interpretable. Rotation may be orthogonal when factors are uncorrelated with one another or oblique when factors are correlated with one another. The choice of rotation is both empirically and theoretically driven.

3.12.1 Oblique Rotation Method

Oblique rotation method allows for correlated factors instead of maintaining independence between the rotated factors. The oblique rotation process does not require that the reference axes be maintained at 90 degrees angle. This rotation strategy is termed oblique because the angles between the factors becomes greater or less than the 90 degrees angle. Oblique rotation method is more flexible because the axes need not be orthogonal. The two major method of oblique rotation method are direct oblimin and promax. Oblimin rotation is that factors are allowed to be correlated and diminished interpretability.

3.12.2 Varimax Rotation Method

Orthogonal rotation shifts the factors in the factor space maintaining 90 degrees angel of the factors to one another to achieve the best simple structure. This rotation strategy maintains the perfectly uncorrelated nature of the factors after the solution is rotated and often aids in the interpretation process since uncorrelated factors are easier to interpret. In theory, the results of an orthogonal rotation are likely to be replicated in future studies since there is less sampling error in the orthogonal rotation due to less capitalization on chance that would occur if more parameters were estimated, as is the case in oblique rotation.

Varimax rotation method is one of the most popular orthogonal rotation techniques is rotation to the varimax criterion developed by Kaiser. In statistics, a varimax rotation method is used to simplify the expression of a particular sub-space in terms of just a few major items. If the actual coordinate system is unchanged, it is the orthogonal basis that is being rotated to align with those coordinates. In this technique,

the factors are that every observed variable has a large factor pattern/structure coefficient on only one of the factors. Other orthogonal rotation methods are quartimax and equimax.

3.13 Testing of Sampling Adequacy of the Approach

In factor analysis, Kaiser-Meyer-Olkin and Bartlett's Test of Sphericity is an important role for accepting the sample adequacy. Kaiser-Meyer-Olkin and Bartlett's test of Sphericity is a measure of sampling adequacy that is recommended to check the case to variable ratio for the analysis being conducted.

The Bartlett's Test of Sphericity relates to significance of the study and thereby shows the validity and suitability of the response collected to the problem being addressed through the study. For a large sample, Bartlett's Test approximates a chi-square distribution. However, the Bartlett's test compares the observed correlation matrix to identify matrix. This test the null hypothesis that the correlation matrix is an identify matrix. For principal component analysis to be recommended suitable, the significance level of the Bartlett's Test of Sphericity must be less than 0.05. In addition to, very small values of significance (below 0.05) indicate a high probability that there are significant relationships between the variables, whereas higher values (0.1 or above) indicate the data is inappropriate for principal component analysis.

Kaiser-Meyer-Olkin (KMO) test is a measure of how suited the data is for factor analysis. The statistics is a measure of the proportion of variance among variables that might be common variance. The lowest the proportion, the more suited the data is to factor analysis. While the values of KMO ranges between 0 and 1, the world over accept index is over 0.6. Kaiser-Meyer-Olkin (KMO) values between 0.8 and 1 indicate that the sampling is adequate. If this value is less than 0.6 indicate, the sampling is not adequate and remedial action should be taken. Kaiser-Meyer-Olkin (KMO) values close to zero means that the large partial correlations compared to the sum of correlations. In other words, there are widespread correlations which are large problem for factor analysis.

The values of Kaiser-Meyer-Olkin (KMO) are:

KMO>0.9	marvelous
0.8<KMO<0.9	meritorious
0.7<KMO<0.8	middling
0.6<KMO<0.7	mediocre
0.5<KMO<0.6	miserable
KMO<0.5	unacceptable

3.13.1 Bartlett's Test

Bartlett's Test is used to test the homogeneity of variance in the factors. Equal variances across groups or samples are called homogeneity of variance. Bartlett's Test is sensitive to departures from normality. Then Bartlett's Test may be testing for non-normality.

CHAPTER IV

ANALYSIS OF AWARENESS ON HAZARD OF SMOKING

This chapter presents analysis of awareness on health hazard of smoking among students in Yangon University of Economics based on the results of data collected from 256 students. The data for this study were collected by personal interviews based on a survey questionnaire. Socio-economic characteristics and smoking awareness of the students were explored in this study.

4.1 Survey Design and Determining of the Sample Design

To obtain the required information on smoking habits survey on the awareness of smoking was conducted in Yangon University of Economics. The study population was approximately 4337 students whose aged 16 to 24.

The sampling design that has been employed for the data collection is stratified random sampling. This sampling method is very useful for complete sampling frame. This sampling method that has five education levels in University is taken to be the stratum. The education levels are stratified into stratum I (First Year), stratum II (Second Year), stratum III (Third Year or 1st Year Honours), stratum IV (Fourth Year or 2nd Year Honours). Then, the students were selected with simple random sampling method (without replacement) from each stratum.

4.2 Determination of the Sample Size

In this study, the proportion of students who have the awareness on health hazard of smoking is assumed to be 0.5. The appropriate stratified random sampling is chosen with 5% precision and taking 90% confidence level by using the following formula:

$$n = \frac{n_0}{1 + \frac{(n_0 - 1)}{N}}$$

$$n_0 = \frac{z^2 pq}{e^2}$$

Where; n = Sample size

$z =$ Critical value of desired confidence level at 90% = 1.645

$p =$ Proportion of students who have the awareness on health hazard of smoking

= 0.5

$q =$ Proportion of students who don't have the awareness on health hazard of smoking

$= 1 - p = 1 - 0.5 = 0.5$

$e =$ The desire level of precision $= 0.05$

The sample size is

$$n_0 = \frac{(1.645)^2 \times 0.5 \times 0.5}{(0.05)^2}$$

$$= 270.6025$$

$$\cong 271$$

Therefore,

$$n = \frac{n_0}{1 + \frac{(n_0 - 1)}{N}}$$

$$= \frac{271}{1 + \frac{(271 - 1)}{4337}}$$

$$= 255.1176$$

$$\cong 256$$

The required sample size is 256 students. Since the cost of sampling within each education level (stratum) does not vary from stratum to stratum, the sample size of each stratum is determined by using the following proportional allocation method.

$$n_i = n \times \frac{N_i}{N} \quad , i = 1, 2, 3, 4$$

Where; n_i = Sample size of the i^{th} strata

N_i = Population size of the i^{th} strata

Table (4.1) Sample Sizes from Each Stratum

Strata	N_i	Sample Size of Each Strata $n_i = n \times \frac{N_i}{N}$
I	1321	$n_1 = n \times \frac{N_1}{N} = 256 \times \frac{1321}{4337} = 78$
II	1153	$n_2 = n \times \frac{N_2}{N} = 256 \times \frac{1153}{4337} = 68$
III	906	$n_3 = n \times \frac{N_3}{N} = 256 \times \frac{906}{4337} = 53$
IV	957	$n_4 = n \times \frac{N_4}{N} = 256 \times \frac{957}{4337} = 57$
	4337	$n = 256$

The corresponding allocation for each stratum (education level) is presented in Table (4.1). The sample sizes for each stratum are 78, 68, 53 and 57 respectively.

4.3 General Characteristics of the Respondents

In this study, the sample population is 256 students who are attending at Yangon University of Economics. Table (4.2) shows students by socio-demographic information.

Table (4.2) Students by Socio-demographic Information

Variables		**Frequency**	**Percent**
Gender	Female	93	33.2
	Male	182	65.0
Age Group	16-19	145	51.8
	20-23	116	41.4
	24 and above	14	5.0
Marital Status	Married	2	0.7
	Single	273	97.5
Religion	Buddhism	259	92.5
	Christian	8	2.9
	Muslim	4	1.4
	Hindu	3	1.1
	Others	1	0.4

Level of Education	First Year	84	30.0
	Second Year	73	26.1
	Third Year	56	20.0
	Fourth Year	62	22.1
Parents	Both are alive	243	86.8
	Only Father alive	5	1.8
	Only Mother alive	23	8.2
	Both are passed away	4	1.4
Living	Parents	117	41.8
	Relatives	12	4.3
	Spouse	1	0.4
	Friends	99	35.4
	Dormitory	46	16.4
Education of Father	Under graduate	127	45.4
	Graduate	145	51.8
	Post graduate	3	1.1
Education of Mother	Under graduate	142	50.7
	Graduate	133	47.5
Occupation of Father	Business	108	38.6
	Farmer	50	17.9
	Company Staff	6	2.1
	Government Staff	62	22.1
	Others	49	17.5
Occupation of Mother	Business	83	29.6
	Farmer	32	11.4
	Company Staff	3	1.1
	Government Staff	37	13.2
	Others	120	42.9
Monthly Family Income	Below 300,000 Ks	41	14.6
	Ks 300,001-400,000Ks	68	24.3
	Ks 400,001-500,000Ks	57	20.4
	Ks 500,001-600,000Ks	39	13.9
	Ks 600,001 and above	70	25.0

Source: Survey Data

Gender is classified into two groups such as Female and Male. The numbers of male are 167 (59.6%) and the numbers of female are 89 (31.8%). Age group are three groups which is 16-19, 20-23 and 24 and above. There are 133 (47.5%) students in 16-19 group, 111 (39.6%) students in 20-23 group and 12 (4.3%) in 24 and above group, respectively. Married and Single are included in Marital Status. Single students are 254 (90.7%) and Married students are 2 (0.7%). Religion includes Buddhism, Christian, Muslim, Hindu and Others. Buddhism are 243 (86.8%) students, Christian are 8 (2.9%) students, Muslim are 2 (0.7%) students, Hindu are 2 (0.7%) students and others are 1 (0.4%) student.

Levels of Education are First Year, Second Year, Third Year and Fourth Year. There are 78 (27.9%) students in First Year, 68 (24.3%) students in Second Year, 53 (18.9%) students in Third Year and 57 (20.4%) students in Fourth Year. Both are alive, Only Father alive, only Mother alive and both are passed away are included in Parents. There are 224 (80.0%) students who have both of parents are alive. There are 5 (1.8%) students who have only Father alive. There are 23 (8.2%) students who have only Mother alive. There are 4 (1.4%) students who both of parents are passed away.

Living is classified Dormitory, Friends, Parents, Relatives and Spouse. There are 110 (39.3%) students who live with Parents, 11 (3.9%) students who live with Relatives, 1 (0.4%) student who live with Spouse, 91 (32.5%) students who live with Friends and 43 (15.4%) students who live with Dormitory. Education of Father and Education of Mother are under graduate, Graduate and Post graduate, respectively. In Education of Father, under graduate are 118 (42.1%), graduate are 135 (48.2%) and post graduate are 3 (1.1%). In Education of Mother, under graduate are 134 (47.9%), graduate are 122 (43.6%) and post graduate is nothing. Occupation of Father is Business, Farmer, Company Staff, Government Staff and others. In Occupation of Father, Business is 103 (36.8%), Farmer is 46 (16.4%), Company Staff is 6 (2.1%), Government Staff is 58 (20.7%) and others are 43 (15.4%).

Occupation of Mother is Business, Farmer, Company Staff, Government Staff and others. In Occupation of Mother, Business is 78 (27.9%), Farmer is 30 (10.7%), Company Staff is 2 (0.7%), Government Staff is 34 (12.1%) and others are 112 (40.0%). Monthly Family Incomes are Below 300,000 Ks, Ks 300,001-400,000Ks, Ks 400,001-500,000Ks, Ks 500,001-600,000Ks and Ks 600,001 and above. Below 300,000 Ks are 38 (13.6%) students, Ks 300,001-400,000Ks are 63 (22.5%) students, Ks 400,001-500,000Ks are 51 (18.2%) students, Ks 500,001-600,000Ks are 37 (13.2%) students and Ks 600,001 and above are 67 (23.9%) students. Table (4.3) shows students by awareness on health hazard of smoking.

Table (4.3) Students by Awareness on Health Hazard of Smoking

Variables		Frequency	Percent
Smoking-related Knowledge	Low in Awareness	26	9.3
	High in Awareness	249	88.9
Smoking-related Attitude	Low in Awareness	178	63.6
	High in Awareness	97	34.6
Promotional Activities	Low in Awareness	248	88.6
	High in Awareness	26	9.3
Social Pressure	Low in Awareness	177	63.2
	High in Awareness	98	35.0
Knowledgeable Programme	Low in Awareness	58	20.7
	High in Awareness	217	77.5
Environmental Constraint	Low in Awareness	40	14.3
	High in Awareness	235	83.9

Source: Survey Data

Low in awareness and high in awareness are divided into two groups in Smoking-related Knowledge. In Smoking-related Knowledge, low in awareness are 25 (8.9%) and high in awareness are 231 (82.5%). Low in awareness and high in awareness are divided into two groups in Smoking-related Attitude. In Smoking-related Attitude, low in awareness are 168 (60.0%) and high in awareness are 88 (31.4%).

Low in awareness and high in awareness are divided into two groups in Promotional Activities. In Promotional Activities, low in awareness are 231 (82.5%) and high in awareness are 25 (8.9%). Low in awareness and high in awareness are divided into two groups in Social Pressure. In Social Pressure, low in awareness are 163 (58.2%) and high in awareness are 93 (33.2%).

Low in awareness and high in awareness are divided into two groups in Knowledgeable Programme. In Knowledgeable Programme, low in awareness are 57 (20.4%) and high in awareness are 199 (71.1%). Low in awareness and high in awareness are divided into two groups in Environmental Constraint. In Environmental Constraint, low in awareness are 38 (13.6%) and high in awareness are 218 (77.9%), respectively.

4.4 Eigenvalues and Canonical Correlations

Table (4.4) shows the canonical correlations of the six canonical variates and their eigenvalues of the canonical roots. The first canonical correlation coefficient is

0.4946 with an explained variance of the correlation of 55.2447% and an eigenvalue of 0.3239. The second canonical correlation coefficient is 0.3055 with an explained variance of the correlation of 17.5511% and an eigenvalue of 0.1029. The third canonical correlation coefficient is 0.2522 with an explained variance of the correlation of 11.5892% and an eigenvalue of 0.0680. The fourth canonical correlation coefficient is 0.2234 with an explained variance of the correlation of 8.9627% and an eigenvalue of 0.0526. The fifth canonical correlation coefficient is 0.1587 with an explained variance of the correlation of 4.4073% and an eigenvalue of 0.0258. The sixth canonical correlation coefficient is 0.1140 with an explained variance of the correlation of 2.2450% and an eigenvalue of 0.0132.

Table (4.4) Eigenvalues and Canonical Correlations

Canonical Functions	Canonical Correlation	Eigenvalues	% of Variance Explained
First Pair of Canonical Variance	0.4946	0.3239	55.2447
Second Pair of Canonical Variance	0.3055	0.1029	17.5511
Third Pair of Canonical Variance	0.2522	0.0680	11.5892
Fourth Pair of Canonical Variance	0.2234	0.0526	8.9627
Fifth Pair of Canonical Variance	0.1587	0.0258	4.4073
Sixth Pair of Canonical Variance	0.1140	0.0132	2.2450

Source: Survey Data

4.5 Multivariate Tests of Significance

Multivariate Tests of Significance are shown in Table (4.5). The test statistics employed included Wilks' Lambda, Pillai's Criterion, Hotelling's Trace and Roy's multivariate criteria. The commonly used test is Wilk's Lambda test and the overall model is significant. And also, all of these tests are statistically significant.

Table (4.5) Multivariate Tests of Significance

Test Name	Value	Approximate F Statistics	Significance of F Statistics
Pillais	0.4897	1.7996	0.000
Hotellings	0.5863	1.9245	0.000
Wilks	0.5862	1.8633	0.000
Roys	0.2447		

Source: Survey Data

4.6 Dimension Reduction Analysis

The results of the dimension reduction analysis are shown in Table (4.6). Dimension reduction analysis tests the significance of each of the roots. Thus, dimension reduction analysis extracted six canonical roots. It was found that of the six possible roots only the first root is significant. The first test of significance tests all six canonical roots of significance, the second test excluded the first root and tests roots two to six, the third test excluded the first and second roots and tests roots three to six, the fourth test excluded the first, second and third roots and tests roots four to six, the fifth test excluded the first, second, third and fourth roots and tests roots five to six and the last test tests roots six by itself.

The first root explains a large proportion of the variance of the correlation. Thus, it was found that the socio-demographic information and the awareness on health hazard of smoking are positively correlated in the population.

Table (4.6) Dimension Reduction Analysis

Roots	Wilk's Lambda	F Statistics	Hypotheses Degree of Freedom	Error Degree of Freedom	Significance of F Statistics
1 to 6	0.5862	1.8633	72	1300.66	0.000
2 to 6	0.7761	1.1359	55	1109.86	0.235
3 to 6	0.8560	0.9546	40	911.91	0.552
4 to 6	0.9141	0.8148	27	704.49	0.735
5 to 6	0.9622	0.5893	16	484.00	0.893
6 to 6	0.9870	0.4569	7	243.00	0.865

Source: Survey Data

4.7 Canonical Loadings

The canonical loadings for the dependent and independent variates for the canonical functions are shown in Table (4.7).

Table (4.7) Canonical Loadings

Sets	Variables	1	2	3	4	5	6
Set 1	Gender	0.893	0.133	-0.138	-0.169	0.089	0.030
	Age Group	0.008	0.529	-0.368	-0.140	-0.173	0.409
	Marital Status	-0.230	0.370	0.311	-0.116	0.171	-0.373
	Religion	0.048	-0.104	-0.288	0.221	0.545	-0.099
	Level of Education	-0.107	0.583	-0.178	0.015	0.035	0.040
	Parents	0.232	-0.083	-0.014	-0.018	0.031	-0.735
	Living	-0.027	-0.184	-0.242	-0.580	-0.507	-0.292
	Education of Father	-0.194	0.623	-0.004	0.113	-0.234	0.045
	Education of Mother	-0.216	0.151	-0.027	0.006	0.182	-0.174
	Occupation of Father	0.026	-0.056	-0.008	-0.016	-0.075	0.154
	Occupation of Mother	0.093	0.144	-0.078	0.617	0.178	0.280
	Monthly Family Income	-0.306	0.244	-0.648	-0.265	0.101	-0.069
	Smoking-related Knowledge	-0.372	-0.570	-0.047	-0.283	-0.655	-0.163
	Smoking-related Attitude	0.707	-0.496	0.194	0.353	-0.137	-0.270

Set 2	Promotional Activities	0.103	-0.271	0.248	-0.416	0.534	-0.629
	Social Pressure	0.608	0.290	-0.113	-0.478	-0.368	-0.412
	Knowledgeable Programme	-0.123	-0.060	-0.854	0.149	-0.125	-0.464
	Environmental Constraint	-0.333	0.258	0.193	0.378	-0.461	-0.656

Source: Survey Data

For the first canonical function, the variables have loadings ranging from 0.103 to 0.707 in the first dependent variate with three dependent variables Y_1 (Smoking-related Knowledge), Y_5 (Knowledgeable Programme) and Y_6 (Environmental Constraint) having negative loadings. The variables have loadings ranging from 0.008 to 0.893 in the first independent variate with six independent variables X_3 (Marital Status), X_5 (Level of Education), X_7 (Living), X_8 (Education of Father), X_9 (Education of Mother) and X_{12} (Monthly Family Income) having negative loadings. A variable with the highest loadings on the independent variate is X_1 (Gender). The two variables with the highest loadings on the dependent variate are Y_2 (Smoking-related Attitude) and Y_4 (Social Pressure).

For the second canonical function, the variables have loadings ranging from 0.060 to 0.570 in the second dependent variate with four dependent variables Y_1 (Smoking-related Knowledge), Y_2 (Smoking-related Attitude), Y_3 (Promotional Activities) and Y_5 (Knowledgeable Programme) having negative loadings. The variables have loadings ranging from 0.056 to 0.623 in the second independent variate with four independent variables X_4 (Religion), X_6 (Parents), X_7 (Living) and X_{10} (Occupation of Father) having negative loadings. The two variables with the highest loadings on the independent variate are X_2 (Age Group), X_5 (Level of Education) and X_8 (Education of Father). The variable with the highest loadings on the dependent variate is Y_1 (Smoking-related Knowledge).

For the third canonical function, the variables have loadings ranging from 0.047 to 0.854 in the third dependent variate with three dependent variables Y_1 (Smoking-related Knowledge), Y_4 (Social Pressure) and Y_5 (Knowledgeable Programme) having negative loadings. The variables have loadings ranging from 0.004 to 0.648 in the third independent variate with eleven independent variables X_1 (Gender), X_2 (Age Group), X_4 (Religion), X_5 (Level of Education), X_6 (Parents), X_7 (Living), X_8 (Education of Father), X_9 (Education of Mother), X_{10} (Occupation of

Father), X_{11} (Occupation of Mother) and X_{12} (Monthly Family Income) having negative loadings. The variable with the highest loadings on the independent variate is X_{12} (Monthly Family Income). The variable with the highest loadings on the dependent variate is Y_5 (Knowledgeable Programme).

For the fourth canonical function, the variables have loadings ranging from 0.149 to 0.478 in the fourth dependent variate with three dependent variables Y_1 (Smoking-related Knowledge), Y_3 (Promotional Activities) and Y_4 (Social Pressure) having negative loadings. The variables have loadings ranging from 0.006 to 0.617 in the fourth independent variate with seven independent variables X_1 (Gender), X_2 (Age Group), X_3 (Marital Status), X_6 (Parents), X_7 (Living), X_{10} (Occupation of Father) and X_{12} (Monthly Family Income) having negative loadings. The two variables with the highest loadings on the independent variate are X_7 (Living) and X_{11} (Occupation of Mother).

For the fifth canonical function, the variables have loadings ranging from 0.125 to 0.655 in the fifth dependent variate with five dependents variables Y_1 (Smoking-related Knowledge), Y_2 (Smoking-related Attitude), Y_4 (Social Pressure), Y_5 (Knowledgeable Programme) and Y_6 (Environmental Constraint) having negative loadings. The variables have loadings ranging from 0.031 to 0.545 in the fifth independent variate with four independent variables X_2 (Age Group), X_7 (Living), X_8 (Education of Father) and X_{10} (Occupation of Father) having negative loadings. The two variables with the highest loadings on the independent variate are X_4 (Religion) and X_7 (Living). The two variables with the highest loadings on the dependent variate are Y_1 (Smoking-related Knowledge) and Y_3 (Promotional Activities).

For the sixth canonical function, the variables have loadings ranging from 0.163 to 0.656 in the sixth dependent variate with six dependent variables Y_1 (Smoking-related Knowledge), Y_2 (Smoking-related Attitude), Y_3 (Promotional Activities), Y_4 (Social Pressure), Y_5 (Knowledgeable Programme) and Y_6 (Environmental Constraint) having negative loadings. The variables have loadings ranging from 0.030 to 0.735 in the sixth independent variate with six independent variables X_3 (Marital Status), X_4 (Religion), X_6 (Parents), X_7 (Living), X_{11} (Occupation of Mother) and X_{12} (Monthly Family Income) having negative loadings.

A variable with the highest loadings on the independent variate is X_6 (Parents). The two variables with the highest loadings on the dependent variate are Y_3 (Promotional Activities) and Y_6 (Environmental Constraint).

4.8 Factor Analysis

Factor analysis was used to construct the factors that related to smoking awareness of the survey among Yangon University of Economics students. The items are Gender, Age Group, Marital Status, Religion, Level of Education, Parents, Living, Education of Father, Education of Mother, Occupation of Father, Occupation of Mother, Monthly Family Income, Smoking-related Knowledge, Smoking-related Attitude, Promotional Activities, Social Pressure, Knowledgeable Programme and Environmental Constraint. The results of the total variance explained by each component are shown in Table (4.8).

Table (4.8) Total Variance Explained by Each Component

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Eigenvalue	% of Variance	Cumulative %	Eigenvalue	% of Variance	Cumulative %
1	2.351	13.061	13.061	2.351	13.061	13.061
2	1.904	10.579	23.640	1.904	10.579	23.640
3	1.659	9.214	32.854	1.659	9.214	32.854
4	1.509	8.382	41.236	1.509	8.382	41.236
5	1.345	7.474	48.710	1.345	7.474	48.710
6	1.226	6.813	55.523	1.226	6.813	55.523
7	1.115	6.195	61.717	1.115	6.195	61.717
8	0.973	5.406	67.123			
9	0.914	5.077	72.200			
10	0.830	4.613	76.813			
11	0.786	4.369	81.182			
12	0.691	3.836	85.018			
13	0.643	3.573	88.591			
14	0.608	3.378	91.969			

15	0.489	2.716	94.685			
16	0.447	2.481	97.166			
17	0.373	2.074	99.240			
18	0.137	0.760	100.000			

Source: Survey Data

The principal factor method was used to generate the initial solution. The total variance explained at seven factors are 61.717% of the overall variance and their eigenvalues are greater than 1. Factor 1 explains 13.061% of the variance in the data with an eigenvalue of 2.351. This factor consists of Education of Father, Education of Mother, Parents and Monthly Family Income. Factor 2 explains 10.579% of the variance in the data with an eigenvalue of 1.904. This factor consists of Age Group and Level of Education.

Factor 3 explains 9.214% of the variance in the data with an eigenvalue of 1.659. This factor consists of Knowledgeable Programme, Environmental Constraint and Smoking-related Knowledge. Factor 4 explains 8.382% of the variance in the data with an eigenvalue of 1.509. This factor consists of Gender and Living. Factor 5 explains 7.474% of the variance in the data with an eigenvalue of 1.345. This factor consists of Religion and Marital Status. Factor 6 explains 6.813% of the variance in the data with an eigenvalue of 1.226. This factor consists of Occupation of Father and Occupation of Mother. Factor 7 explains 6.195% of the variance in the data with an eigenvalue of 1.115. This factor consists of Social Pressure, Smoking-related Attitude and Promotional Activities.

4.9 Kaiser-Meyer-Olkin (KMO) Measure and Bartlett's Test of Sphericity

The Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy and The Bartlett Test of Sphericity are given in Table (4.9).

Table (4.9) Kaiser-Meyer-Olkin (KMO) and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy	0.522
Bartlett's Test of Sphericity (Approximate Chi-Square)	824.883
Degree of Freedom (df)	153
p-value	0.000

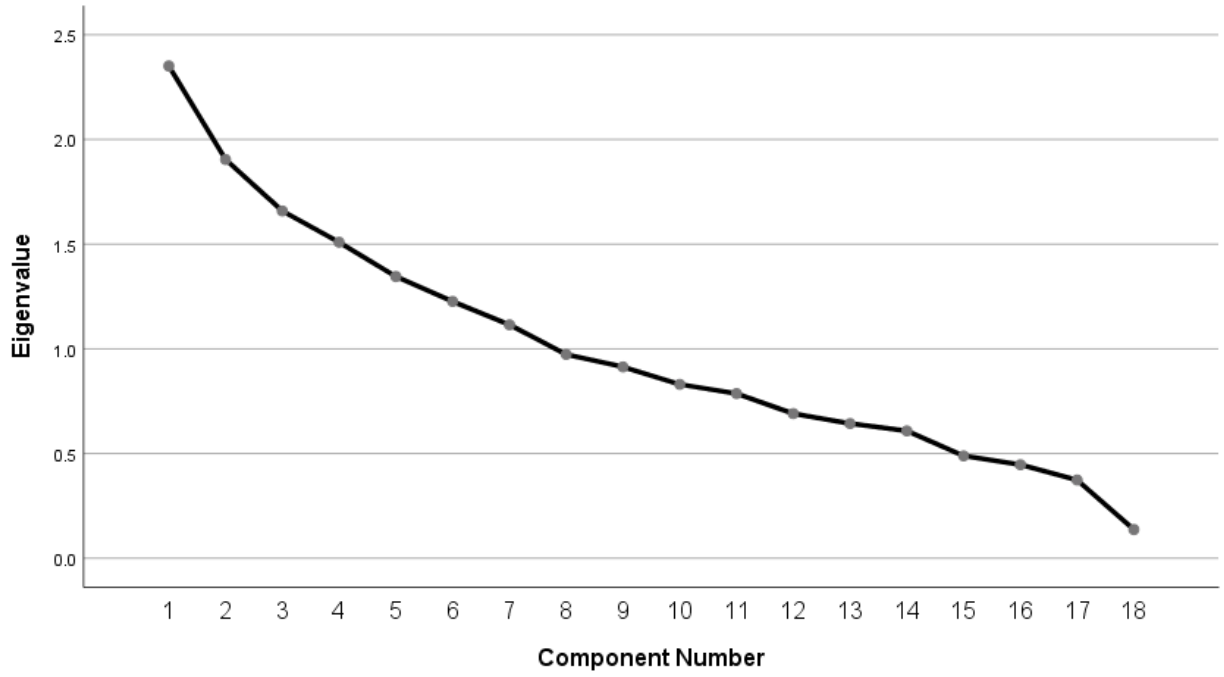
Source: SPSS output

According to the Kaiser-Meyer-Olkin (KMO) and Bartlett's Test table, Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy is 0.522 which is miserable. So, factor analysis is not suitable for this study. Chi-square (Bartlett Test of Sphericity) value is approximately 824.883. The Bartlett Test of Sphericity is significant at 0.000 level.

4.10 The Number of Principal Components

Figure (4.1) shows a screen plot for a situation with eighteen principal components. An elbow occurs in the plot in Figure (4.1) at about $i = 8$. That is, the eigenvalues after $\hat{\lambda}_7$ are all relatively small and about the same size. According to screen plot (eigenvalue graph), the number of factors in the items can be limited to seven. After the seventh point are small and the distances between then are very close and similar in the eigenvalue graph.

Figure (4.1) Screen Plot for Eigenvalues and Component Number



Source: SPSS Output

4.11 Principal Components through Component Matrix and Rotated Component Matrix

The results of the structure matrix are shown in Table (4.10). Component 1 has high correlation for variables Education of Father and Education of Mother and has fairly correlation for variable Monthly Family Income. A negative coefficient of Parents leads to a negative low correlation between component 1 and Parents. Component 2 has high correlation for variables Level of Education and Age Group. Component 3 has high correlation for variables Knowledgeable Programme and has fairly correlation for variables Environmental Constraint and Smoking-related Knowledge. Component 4 has high correlation for Gender and negative fairly correlation between component 4 and Living. Component 5 has high correlation for variable Religion and negative highly correlation between component 4 and Marital Status. Component 6 has high correlation for variables Occupation of Mother and Occupation of Father. Component 7 has high correlation for variable Smoking-related Attitude and fairly correlation for variables Social Pressure and Promotional Activities.

Table (4.10) Structure Matrix

Variables	Component						
	1	2	3	4	5	6	7
Education of Father	0.853						
Education of Mother	0.828						
Monthly Family Income	0.432			0.411		-0.410	-0.347
Parents	-0.360						
Age Group		0.946					
Level of Education		0.941					
Knowledgeable Programme			0.734				
Environmental Constraint			0.696				
Smoking-related Knowledge			0.639				
Gender	-0.315			0.708			0.311
Living				-0.477			
Religion					0.720		

Marital Status					-0.718		
Occupation of Mother						0.834	
Occupation of Father						0.815	
Smoking-related Attitude							0.727
Promotional Activities							0.635
Social Pressure				0.395			0.625

Source: Survey Data

Factor 1 is socio-economic characteristics of family and comprised of four items. This factor consists of Education of Father, Education of Mother, Parents and Monthly Family Income with factor loading ranging from -0.360 to 0.853.

Factor 2 is students' characteristics and comprised of two items such as Age Group and Level of Education with factor loading ranging from 0.941 to 0.946.

Factor 3 is awareness comprised of three items such as Knowledgeable Programme, Environmental Constraint and Smoking-related Knowledge with factor loading ranging from 0.639 to 0.734.

Factor 4 is sexual role and comprised of two items such as Gender and Living with factor loading ranging from -0.477 to 0.708.

Factor 5 is demographic characteristics and comprised of two items such as Marital Status and Religion with factor loading ranging from -0.718 to 0.720.

Factor 6 is job of parents and comprised of two items such as Occupation of Father and Occupation of Mother with factor loading ranging from 0.815 to 0.834.

Factor 7 is social impact and comprised of three items such as Social Pressure, Smoking-related Attitude and Promotional Activities with factor loading ranging from 0.625 to 0.727.

CHAPTER V

CONCLUSION

5.1 Findings

The study aims at analyzing the awareness of health hazard of smoking among students in Yangon University of Economics. In this thesis, Canonical Correlation Analysis and Factor Analysis are combined in determining the relationship between socio-economic characteristics and awareness of smoking of undergraduate students.

The dependent variables are Smoking-related Knowledge, Smoking-related Attitude, Promotional Activities, Social Pressure, Knowledgeable Programme and Environmental Constraint and independent variables are Gender, Age Group, Marital Status, Religion, Level of Education, Parents, Living, Education of Father, Education of Mother, Occupation of Father, Occupation of Mother and Monthly Family Income in the canonical correlation analysis.

Firstly, Canonical Correlation Analysis measured the strength of relationship of the canonical pair. The canonical correlation analysis generated six correlation coefficients. The first pair with a measure of correlation 0.4946 with the proportion of variability of about 55.2447%, the second pair with a measure of correlation 0.3055 with the proportion of variability of about 17.5511%, the third pair with a measure of correlation 0.2522 with the proportion of variability of about 11.5892%, the fourth pair with a measure of correlation 0.2234 with the proportion of variability of about 8.9627%, the fifth pair with a measure of correlation 0.1587 with the proportion of variability of about 4.4073%, the sixth pair with a measure of correlation 0.1140 with the proportion of variability of about 2.2450%.

And then the first canonical pair captured the validity of about 55.2447%, the second canonical pair captured the validity of about 17.5511%, the third canonical pair captured the validity of about 11.5892%, the fourth canonical pair captured the validity of about 8.9627%, the fifth canonical pair captured the validity of about 4.4073% and the sixth canonical pair captured the validity of about 2.2450%. Hence the total validity captured by the six canonical pairs is 100%. According to the Multivariate Tests of Significant Table, Pillais, Hotellings and Wilks Test are

statistically significant. In dimension reduction analysis, there are six canonical roots and only the first root is significant.

Moreover, canonical loadings for the dependent and independent variates for the canonical functions are six canonical functions. In the first canonical function, the variables have loadings ranging from 0.103 to 0.707 in the first dependent variate and loadings ranging from 0.008 to 0.893 in the first independent variate. In the second canonical function, the variables have loadings ranging from 0.060 to 0.570 in the second dependent variate and loadings ranging from 0.056 to 0.623 in the second independent variate. In the third canonical function, the variables have loadings ranging from 0.047 to 0.854 in the third dependent variate and loadings ranging from 0.004 to 0.648 in the third independent variate. In the fourth canonical function, the variables have loadings ranging from 0.149 to 0.478 in the fourth dependent variate and loadings ranging from 0.006 to 0.617 in the fourth independent variate. In the fifth canonical function, the variables have loadings ranging from 0.125 to 0.655 in the fifth dependent variate and loadings ranging from 0.031 to 0.545 in the fifth independent variate. In the sixth canonical function, the variables have loadings ranging from 0.163 to 0.656 in the sixth dependent variate and loadings ranging from 0.030 to 0.735 in the sixth independent variate.

Secondly, Factor Analysis is used to construct the factors. The first factor captured the variability 13.061%, the second factor captured the variability 10.579%, the third factor captured the variability 9.214%, the fourth factor captured the variability 8.382%, the fifth factor captured the variability 7.474%, the sixth factor captured the variability 6.813% and the seventh factor captured the variability 6.195% respectively. The total variability captured due to the seventh factors is 61.717%.

According to the KMO and Bartlett's Test Table, Kaiser-Meyer-Olkin (KMO) Measure was 0.522 and Bartlett's Test of Sphericity was 824.883, which was significant at 0.000 level. It can be seen that correlation matrix is not an identity matrix. From the Eigenvalues and component number plot, there are eighteen principal components and an elbow occurs in the plot at about $i = 8$. After the seventh eigenvalues are small and the distances between then are very close.

The principal component matrix indicates that the component matrix rotated using the Oblimax rotation technique which further provides the rotated component

matrix. It has been found that there are seven components namely first principal component (Education of Father, Education of Mother, Monthly Family Income and Parents), second principal component (Level of Education and Age Group), third principal component (Knowledgeable Programme, Environmental Constriant and Smoking-related Knowledge), fourth principal component (Gender and Living), fifth principal component (Occupation of Mother and Occupation of Father), sixth principal component (Marital Status and Religion) and seventh principal component (Social Pressure, Smoking-related Attitude and Promotional Activities).

The results in this study show that the factors affecting the awareness of health hazard of smoking among students are seven principal components. According to the results, most of the seven components were strong determinants.

5.2 Recommendations

This study provides knowledge base from which to develop targeted cigarette smoking control policies for fresh undergraduates. Especially, the restriction of retail sales outlets near universities will be valuable. Smoking prevention program may focus on modifying attitudes towards smoking and providing a cigarette-free environment near the campus.

It is to be recommended that further studies need to be made with larger sample sizes covering different parts of the country. Knowledge and awareness need to encourage positive behaviors and promoting health. Programs and plans should be conducted to increase the awareness on health hazard of smoking which target mainly at the basic educational level, at the community level and finally at the national level.

REFERENCES

Bhuyan KC and Urmi AF . (2018). Canonical Correlation Analysis to Study the Impacts of Different Social Factors on Awareness of Health Hazard of Tobacco Smoking and Smoking Habit. *Biomedical Journal of Scientific and Technical Research.*

Dhirendra, N.Sinha. (2004). *Report on Oral Tobacco Use and Its Implications in South East Asia. WHO SEARO.*

Isaac Kusi Appau. (2011). *Smoking Habits among Adolescents.*

JANE F. EMMERE. (2003). *Social Factors of Cigarette Smoking Initiation among Under-graduate College Students.*

Jolliffe, I. T. (2002). *Principal Component Analysis, Second Edition.*

Kyi Win . (2012). *A Study on Youth Smoking and Tobacco Consumption in South Dagon Township.*

Lukman AbdulSalam. (2006). *Determination of Students' Academic Performance.*

Mandisa Malinga. (2011). *Smoking Prevalence, Knowledge, Attitudes and Beliefs about Health Risks of Tobacco Smoking among Female Psychology Students at the University of the Western Cape* .

Mark Aaron Thompson. (2014). *Smokers' Awareness and Perceptions of Electric Cigarettes.*

Myat Thura Khaing. (2005). *A Study of Smoking Habit (A Case Study of United K. M. K Co. Ltd.*

Nouran Mahmoud. (2011). *Self-related Health Status and Smoking.*

Swe Swe Win Htet. (2018). *Social Media Impact on Students in Bago University, M. Econ (Statistics) Thesis, Department of Statistics, Yangon University of Economics.*

WICHERN, RICHARD A. JOHNSON AND DEAN W. (1988). *Applied Multivariate Statistical Analysis, Fifth Edition* .

Xianglong Xu, Doris Yin Ping Leung, Bing Li, Pengfei Wang and Yong Zhao . (2015). Smoking-related Knowledge, Attitude, Social Pressure and Environmental Constraints among New Undergraduates in Chongqing, China. *Environmental Research and Public Health*.

Yan Naing Myint. (2015). *A Study of Awareness and Practice of Smoking among Youths (Case Study in University Students.*

Zin Mar Lwin. (2018). *Effects of Major Exports and Imports on the Balance of Trade in Myanmar, M. Econ (Statistics) Thesis, Department of Statistics, Meiktila University of Economics.*

Websites

- https://www.britannica .com
- https://www.randeye.com
- https://www.who.int

APPENDICES

APPENDIX A

Survey Questionnaire

Smoking Habits and Awareness of Health Hazard of Smoking Questionnaires

This survey only relates to thesis component of a Master of Statistics which I am completing at Yangon University of Economics. The information provided will be used only for this research and for no other objectives. Your participation is highly appreciated. Thank you for your time and participation.

I. Socio-demographic information

No.	Questions	Answers	For Coding
1.	Gender	1. Male 2. Female	
2.	Age	1. 16-19 2. 20-23 3. 24 and above	
3.	Marital Status	1. Single 2. Married	
4.	Religion	1. Buddhism 2. Christian 3. Muslim 4. Hindu 5. Others (specified)	
5.	Level of Education	1. 1st year 2. 2nd year 3. 3rd year or 1st year honours 4. 4th year or 2nd year honours	
6.	Parents	1. Both are alive 2. Only Father alive 3. Only Mother alive 4. Both are passed away	
7.	Living with	1. Parents 2. Relatives 3. Spouse 4. Friends 5. Dormitory	
8.	Education of Father		
9.	Education of Mother		
10.	Occupation of Father	1. Business 2. Farmer 3. Company Service 4. Government Service 5. Others (specified)	
11.	Occupation of Mother	1. Business 2. Farmer 3. Company Service 4. Government Service 5. Others (specified)	

12.	Family Income (monthly)	1. Below 300,000 Ks 2. Ks 300,001-400,000 Ks 3. Ks 400,001-500,000 Ks 4. Ks 500,001-600,000 Ks 5. Ks 600,001 and above	

II. Smoking Habits

No.	Questions	Answers	For Coding
1.	Do your parents smoke cigarettes?	1. None 2. Both 3. Father only 4. Mother only 5. Don't know	
2.	Do your closest friends smoke cigarettes?	1. None of them 2. Some of them 3. Most of them 4. All of them 5. Don't know	
3.	Do you smoke cigarette?	1. Yes 2. No	
4.	If you say No, do you desire to smoke cigarette?	1. Yes 2. No	
5.	How old were you when you first tried a cigarette?	1. Under 15 years old 2. 16-19 years old 3. Over 20 years old	
6.	Where do you usually smoke?	1. At home 2. At school 3. At friend's home 4. At social event (e.g. bar or club) 5. In public places	
7.	How many times do you usually smoke per day?		
8.	How many cigarettes do you smoke per day?		
9.	How do you describe your weight?		
10.	Where do you see cigarette advertisements?	1. Newspaper/ Journal/ Magazine 2. TV/ Radio 3. Public transport (bus or train) 4. Billboard/ Poster	
11.	How did you get them?	1. From a store or shop 2. From a street vendor 3. From someone else 4. Others (specified)	
12.	How much money do you spend per day on smoking products?		

13.	Have you ever received help or advice to help you stop smoking?	1. From friend 2. From family member 3. From teacher 4. Others (specified) 5. No	
14.	What is the main reason you decided to stop smoking?	1. I have not stopped smoking 2. To improve my health 3. To save money 4. Because my family doesn't like it 5. Others (specified)	

III. Awareness of Smoking

Please indicate your answer. (Choose one)

1. Strongly Disagree 2. Disagree 3. Neutral

4. Agree 5. Strongly Agree

Serial No.	Items	1	2	3	4	5
	Smoking-related Knowledge					
1.	Smoking is harmful to me.					
2.	Smoking can help one gain or lose weight.					
3.	Smoking can cause heart disease.					
4.	Passive smoking is harmful to children and infants.					
5.	Smoking could cause lung cancer.					
6.	The effects of smoking make my teeth yellow and smell bad.					
7.	I see health warnings on cigarette packages.					
8.	I see any signs stating that adolescents are not allowed to buy any smoking products.					
	Smoking-related Attitude					
9.	Smoking is pleasurable.					
10.	Smoking is a type of self-presentation.					
11.	Smoking relaxes me.					
12.	It clams me down.					
13.	Smoking is an increase in focus or concentration.					
14.	Smoking is a decrease in stress.					
15.	Smoking makes me look strong.					
16.	Smoking makes me look mature.					
17.	Smoking makes me confident.					
18.	Smoking is fashionable.					
19.	Smoking can ease communication.					
20.	Smoking is a waste of money.					
21.	Smoking can help me study better.					
22.	Smoking makes me fit in with other people.					
23.	Cigarette advertising should be banned.					
	Promotional Activities					

24.	I have ever participated in an activity sponsored by a cigarette company.					
25.	I have ever received free cigarettes during promotional activities.					
26.	I have ever received free ticket for an entertainment event sponsored by a cigarette company.					
27.	I have ever exchanged a cigarette case for a prize or on-sale goods.					
	Social Pressure					
28.	Pressure from friends.					
29.	Pressure from family.					
30.	Pressure from teachers.					
31.	Smoking brings comfort during celebrations.					
32.	Smoking brings comfort in social activities.					
	Knowledgeable Programme					
33.	I see anti-smoking media messages on television.					
34.	I see anti-smoking media messages on posters.					
35.	I see anti-smoking media messages on the radio.					
36.	I see anti-smoking media messages at the cinema.					
37.	I see anti-smoking media messages in the newspaper.					
38.	I see smoking people on TV, in videos, in movies.					
	Environmental Constraint					
39.	There is difficult in obtaining cigarettes.					
40.	It is easy to purchase near my university.					
41.	It is easy to buy cigarette from a shop.					
42.	There is difficult in finding an indoor smoking place.					
43.	There is difficult in finding an outdoor smoking place.					
44.	There should not be smoking in workplaces.					
45.	There should not be smoking in restaurants.					
46.	There should not be smoking in all public places.					
47.	The university has a policy or rule specifically prohibiting smoking use among students inside school buildings.					

Thank You for Participating in this Survey!

APPENDIX B

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Gender	256	0	1	0.65	0.477
Age Group	256	1	3	1.53	0.587
Marital Status	256	0	1	0.99	0.088
Religion	256	1	5	1.09	0.434
Level of Education	256	1	4	2.35	1.134
Parents	256	1	4	1.25	0.679
Living	256	1	5	2.79	1.660
Education of Father	256	1	3	1.55	0.521
Education of Mother	256	1	2	1.48	0.500
Occupation of Father	256	1	5	2.58	1.587
Occupation of Mother	256	1	5	3.28	1.775
Monthly Family Income	256	1	5	3.12	1.423
Smoking-related Knowledge	256	0	1	0.90	0.297
Smoking-related Attitude	256	0	1	0.34	0.476
Promotional Activities	256	0	1	0.10	0.297
Social Pressure	256	0	1	0.36	0.482
Knowledgeable Programme	256	0	1	0.78	0.417
Environmental Constraint	256	0	1	0.85	0.356
Valid N (listwise)	256				

Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Female	89	31.8	34.8	34.8
	Male	167	59.6	65.2	100.0
	Total	256	91.4	100.0	
Missing	System	24	8.6		
Total		280	100.0		

Age Group

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	16-19	133	47.5	52.0	52.0
	20-23	111	39.6	43.4	95.3
	24 and above	12	4.3	4.7	100.0
	Total	256	91.4	100.0	
Missing	System	24	8.6		
Total		280	100.0		

Marital Status

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Married	2	0.7	0.8	0.8
	Single	254	90.7	99.2	100.0
	Total	256	91.4	100.0	
Missing	System	24	8.6		
Total		280	100.0		

Religion

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Buddhism	243	86.8	94.9	94.9
	Christian	8	2.9	3.1	98.0
	Muslim	2	0.7	0.8	98.8
	Hindu	2	0.7	0.8	99.6
	Others	1	0.4	0.4	100.0
	Total	256	91.4	100.0	
Missing	System	24	8.6		
Total		280	100.0		

Level of Education

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	First Year	78	27.9	30.5	30.5
	Second Year	68	24.3	26.6	57.0
	Third Year	53	18.9	20.7	77.7
	Fourth Year	57	20.4	22.3	100.0
	Total	256	91.4	100.0	
Missing	System	24	8.6		
Total		280	100.0		

Parents

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Both are alive	224	80.0	87.5	87.5
	Only Father alive	5	1.8	2.0	89.5
	Only Father alive	23	8.2	9.0	98.4
	Both are passed away	4	1.4	1.6	100.0
	Total	256	91.4	100.0	
Missing	System	24	8.6		
Total		280	100.0		

Living

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Parents	110	39.3	43.0	43.0
	Relatives	11	3.9	4.3	47.3
	Spouse	1	0.4	0.4	47.7
	Friends	91	32.5	35.5	83.2
	Dormitory	43	15.4	16.8	100.0
	Total	256	91.4	100.0	
Missing	System	24	8.6		
Total		280	100.0		

Education of Father

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Under graduate	118	42.1	46.1	46.1
	Graduate	135	48.2	52.7	98.8
	Post graduate	3	1.1	1.2	100.0
	Total	256	91.4	100.0	
Missing	System	24	8.6		
Total		280	100.0		

Education of Mother

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Under graduate	134	47.9	52.3	52.3
	Graduate	122	43.6	47.7	100.0
	Total	256	91.4	100.0	
Missing	System	24	8.6		
Total		280	100.0		

Occupation of Father

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Business	103	36.8	40.2	40.2
	Farmer	46	16.4	18.0	58.2
	Company Staff	6	2.1	2.3	60.5
	Government Staff	58	20.7	22.7	83.2
	Others	43	15.4	16.8	100.0
	Total	256	91.4	100.0	
Missing	System	24	8.6		
Total		280	100.0		

Occupation of Mother

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Business	78	27.9	30.5	30.5
	Farmer	30	10.7	11.7	42.2
	Company Staff	2	0.7	0.8	43.0
	Government Staff	34	12.1	13.3	56.3
	Others	112	40.0	43.8	100.0
	Total	256	91.4	100.0	
Missing	System	24	8.6		
Total		280	100.0		

Monthly Family Income

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Below 300,000 Ks	38	13.6	14.8	14.8
	Ks 300,001-400,000 Ks	63	22.5	24.6	39.5
	Ks 400,001-500,000 Ks	51	18.2	19.9	59.4
	Ks 500,001-600,000 Ks	37	13.2	14.5	73.8
	Ks 600,001 and above	67	23.9	26.2	100.0
	Total	256	91.4	100.0	
Missing	System	24	8.6		
Total		280	100.0		

Smoking-related Knowledge

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Low in Awareness	25	8.9	9.8	9.8
	High in Awareness	231	82.5	90.2	100.0
	Total	256	91.4	100.0	
Missing	System	24	8.6		
Total		280	100.0		

Smoking-related Attitude

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Low in Awareness	168	60.0	65.6	65.6
	High in Awareness	88	31.4	34.4	100.0
	Total	256	91.4	100.0	
Missing	System	24	8.6		
Total		280	100.0		

Promotional Activities

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Low in Awareness	231	82.5	90.2	90.2
	High in Awareness	25	8.9	9.8	100.0
	Total	256	91.4	100.0	
Missing	System	24	8.6		
Total		280	100.0		

Social Pressure

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Low in Awareness	163	58.2	63.7	63.7
	High in Awareness	93	33.2	36.3	100.0
	Total	256	91.4	100.0	
Missing	System	24	8.6		
Total		280	100.0		

Knowledgeable Programme

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Low in Awareness	57	20.4	22.3	22.3
	High in Awareness	199	71.1	77.7	100.0
	Total	256	91.4	100.0	
Missing	System	24	8.6		
Total		280	100.0		

Environmental Constraint

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Low in Awareness	38	13.6	14.8	14.8
	High in Awareness	218	77.9	85.2	100.0
	Total	256	91.4	100.0	
Missing	System	24	8.6		
Total		280	100.0		

Multivariate Tests of Significance

Test Name	Value	Approx. F	Hypoth. DF	Error DF	Sig. of F
Piallais	0.48969	1.79957	72.00	1458.00	.000
Hotellings	0.58630	1.92448	72.00	1418.00	.000
Wilks	0.58621	1.86327	72.00	1300.66	.000
Roys	0.24466				

Eigenvalues and Canonical Correlations

Root No.	Eigenvalue	Pct.	Cum. Pct.	Canon Cor.	Sq. Cor
1	.32390	55.24468	55.24468	.49463	.24466
2	.10290	17.55106	72.79574	.30545	.09330
3	.06795	11.58917	84.38492	.25224	.06362
4	.05255	8.96272	93.34763	.22344	.04993
5	.02584	4.40734	97.75497	.15871	.02519
6	.01316	2.24503	100.00000	.11398	.01299

Canonical Correlations

	Correlation	Eigenvalue	Wilks Statistic	F	Num D.F	Denom D.F	Sig.
1	.495	.324	.586	1.863	72.000	1300.664	.000
2	.305	.103	.776	1.136	55.000	1109.864	.235
3	.252	.068	.856	.955	40.000	911.907	.552
4	.223	.053	.914	.815	27.000	704.487	.735
5	.159	.026	.962	.589	16.000	484.000	.893
6	.114	.013	.987	.457	7.000	243.000	.865

Set1 Canonical Loadings

Variable	1	2	3	4	5	6
Q-1	.893	.133	-.138	-.169	.089	.030
Q-2	.008	.529	-.368	-.140	-.173	.049
Q-3	-.230	.370	.311	-.116	.171	-.373
Q-4	.048	-.104	-.288	.221	.545	-.099
Q-5	-.107	.583	-.178	.015	.035	.040
Q-6	.232	-.083	-.014	-.018	.031	-.735
Q-7	-.027	-.184	-.242	.580	-.507	-.292
Q-8	-.194	.623	-.004	.113	-.234	.045
Q-9	-.216	.151	-.027	.006	.182	-.174
Q-10	-.026	-.056	.008	-.016	-.075	.154
Q-11	.093	.144	-.078	.617	.178	.280
Q-12	-.306	.244	-.648	-.265	.101	-.069

Set 2 Canonical Loadings

Variable	1	2	3	4	5	6
SK	-.372	-.570	-.047	-.283	-.655	-.163
SA	.707	-.496	.194	.353	-.137	-.270
PA	.103	-.271	.248	-.416	.534	-.629
SP	.608	.290	-.113	-.478	-.368	-.412
KP	-.123	-.060	-.854	.149	-.125	-.464
EC	-.333	.258	.193	.378	-.461	-.656

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.522
Bartlett's Test of Sphericity	Approx. Chi-Square	824.883
Degree of Freedom	df	153
	Sig.	0.000

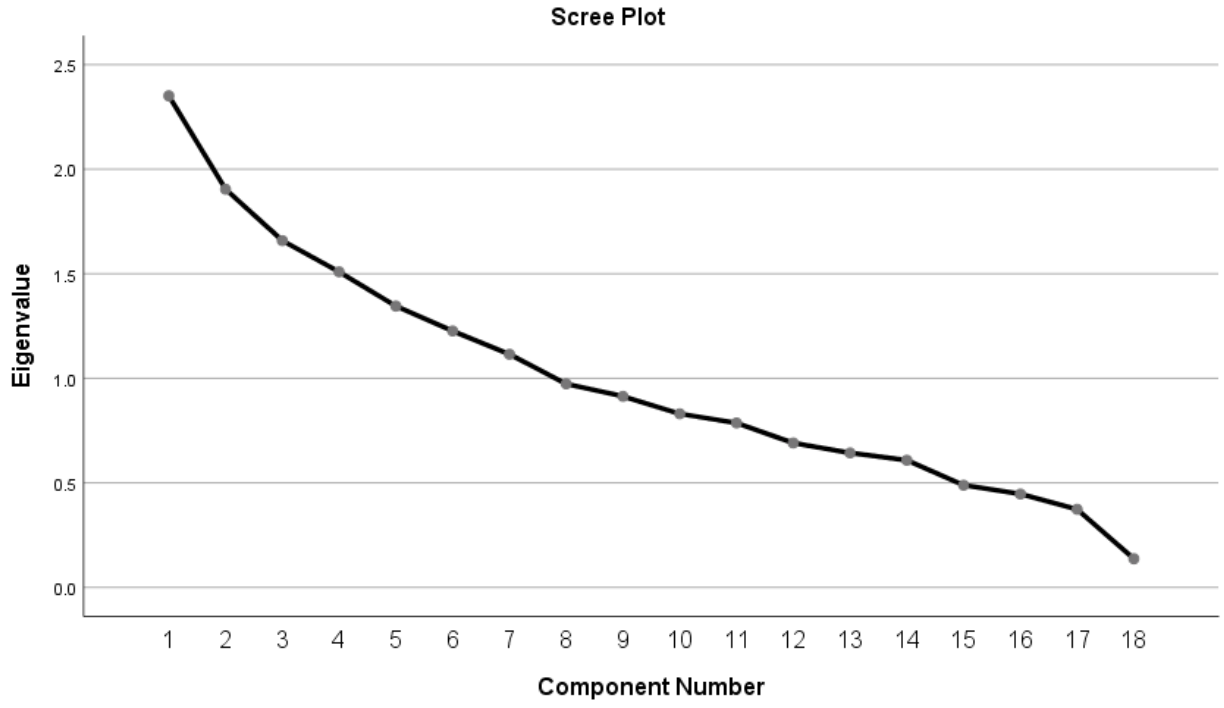
Communalities

	Initial	Extraction
Gender	1.000	.683
Age Group	1.000	.906
Marital Status	1.000	.571
Religion	1.000	.549
Level of Education	1.000	.903
Parents	1.000	.275
Living	1.000	.325
Education of Father	1.000	.767
Education of Mother	1.000	.705
Occupation of Father	1.000	.684
Occupation of Mother	1.000	.733
Monthly Family Income	1.000	.603
Smoking-related Knowledge	1.000	.514
Smoking-related Attitude	1.000	.587
Promotional Activities	1.000	.466
Social Pressure	1.000	.642
Knowledgeable Programme	1.000	.645
Environmental Constraint	1.000	.552

Total Variance Explained

Component	Total	Initial Eigenvalues		Extraction Sums of Squared Loadings			Rotation Sum of Squared Loadings
		% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	2.351	13.061	12.061	2.351	13.061	13.061	1.949
2	1.904	10.579	23.640	1.904	10.579	23.640	1.928
3	1.659	9.214	32.854	1.659	9.214	32.854	1.649
4	1.509	8.382	41.236	1.509	8.382	41.236	1.368
5	1.345	7.474	48.710	1.345	7.474	48.710	1.263
6	1.226	6.813	55.523	1.226	6.813	55.523	1.704
7	1.115	6.195	61.717	1.115	6.195	61.717	1.633
8	0.973	5.406	67.123				
9	0.914	5.077	72.200				
10	0.830	4.613	76.813				
11	0.786	4.369	81.182				
12	0.691	3.836	85.018				
13	0.643	3.573	88.591				
14	0.608	3.378	91.969				
15	0.489	2.716	94.685				
16	0.447	2.481	97.166				
17	0.373	2.074	99.240				
18	0.137	0.760	100.000				

Extraction Method: Principal Component Analysis.



Component Matrix

	1	2	3	4	5	6	7
Education of Father	.627				.479	.319	
Education of Mother	.587				.377		
Gender	-.555			.433			-.312
Smoking-related Attitude	-.484		.307				
Parents	-.416						
Monthly Family Income	.411	-.346				.337	-.365
Age Group	.399	.738		.349			
Level of Education	.532	.709					
Occupation of Father		.587		-.405			
Occupation of Mother		.472		-.384	.345	.435	
Knowledgeable Programme			.649				-.422
Environmental Constraint			.649				
Smoking-related Knowledge			.589	-.308			
Social Pressure			.501	.533			
Marital Status					-.575	.450	
Religion					.493		-.466
Living						-.380	
Promotional Activities				.392			.459

Extraction Method: Principal Component Analysis.
a. 7 components extracted.

Structure Matrix

	1	2	3	4	5	6	7
Education of Father	.853						
Education of Mother	.828						
Monthly Family Income	.432			.411		-.410	-.347
Parents	-.360						
Age Group		.946					
Level of Education		.941					
Knowledgeable Programme			.734				
Environmental Constraint			.696				
Smoking-related Knowledge			.639				
Gender	-.315			.708			.311
Living				-.477			
Religion					.720		
Marital Status					-.718		
Occupation of Mother						.834	
Occupation of Father						.815	
Smoking-related Attitude							.727
Promotional Activities							.635
Social Pressure				.395			.625

Extraction Method: Principal Component Analysis.

Rotation Method: Oblimin with Kaiser Normalization.