THE PRACTICE OF FACTOR ANALYSING TOURISM SURVEYS

By

RODERICK GUSMAN

A Dissertation Submitted in Partial Fulfillment of the Requirements
For the Degree of Bachelor of Science (Honours)
Statistics & Operations Research as main area

DEPARTMENT OF STATISTICS & OPERATIONS RESEARCH
FACULTY OF SCIENCE
UNIVERSITY OF MALTA

JUNE 2005

To the ones who believed in me

ABSTRACT

Roderick Gusman, B.Sc. (Hons.)
Department of Statistics & Operations Research
April 2005
University of Malta

The objective of this dissertation is to study and apply statistical methods to the Traveller Survey involving Likert scales. A typical approach is through factor analysis but an alternative method is by parceling. In this study, we are considering these two data reduction approaches, together with two extraction methods to obtain a set of factor scores. In conclusion, we finalise this study by considering a relationship of the factor scores with the socio-demographic variables of the tourist.

ACKNOWLEDGEMENTS

A proverb says that 'No man is an island' and this is true because it would not have been possible to produce this work without the help and encouragement of others. In fact, I would like to express my gratitude to the following people for their support and assistance in developing this dissertation.

TO:

Dr. Lino Sant, my tutor: for supporting me, and especially for his patience and dedication towards my work.

Mr. Liberato Camilleri, for helping me on matters related to transformations and also in the last part of my thesis.

Mr. David Suda, for helping me organize the tourists' profile.

Mrs. Marie Louise Mangion, Senior Manager, Strategic Planning and Research, Malta Tourism Authority, for giving me the possibility to access this data.

Mrs. Tania Sultana, Manger and Mr. Oliver Farrugia, Assistant Manager, Strategic Planning and Research, Malta Tourism Authority, for always being available to sort out any queries I had.

Special thanks go to my parents, my sister, my girlfriend Sarah and my friends for their great amount of support and patience.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

Tourism is one of the world's largest industries. Consequently, the practice of tourism is becoming increasingly sophisticated. The development of this industry is related to the wide range of services including transport, accommodation, attractions and infrastructure. Nowadays tourism is becoming more competitive and hence this influences the fact that tourism businesses must constantly evaluate the products and services they offer to their customers. Consequently, these businesses must revisit their commercial goals and objectives to be more up to standard with the modern community. Therefore, it seems that tourism, when planned and managed appropriately can act as a valuable agent in the economic development of the countries.

Hence, for the tourism market to withstand these new challenges, various surveys were planned and analysed with the purpose of understanding better what the tourist requests. Such a survey is the 'Traveller Survey', in which tourists are asked a set of questions. Very often the focus is on questions in which the tourist is asked to rate the above-mentioned services. Then, analysts apply statistical methods to evaluate these ratings so that they can understand better the tourists' response and deduce collective tendencies in attitudes and behaviours. The typical statistical approach is traditionally factor analysis but recent advances in this study showed that statistical sophistications such as bootstrapping, jacknifing, clustering and parceling have been introduced. All these methods are based on the same idea, namely that of data reduction and discovering the latent factors present in the model. Then, after reducing the data into smaller groups, a linear relationship with other variables considered in the study is formulated.

In this dissertation, we are analysing the 'Traveller Survey' published by the Malta Tourism Authority. For this analysis, we are considering a question in which tourists were asked to rate an amount of variables that describe the physical environment and service provided by employees of the tourism industry. But, before we start our statistical analysis, we reduce the data since we have a large amount of missing or non-response results. Then, we apply two types of techniques of data reduction, factor analysis and parceling. Within these two approaches, we apply two extraction methods, maximum likelihood and principal axis factoring. The difference between these two extractions is that the maximum likelihood considers normality distributions while principal axis factoring does not obey the normality conditions. From these two techniques, we obtain a set of factor scores. These factor scores are then transformed to satisfy the normal distribution. This is required to obtain a relationship with the socio-demographic variables, through the application of linear models which need to satisfy the normality condition. The objective of all these steps is to discover the latent factors present in the ratings of various variables and then obtain a relationship of these variables with the background or profile of the respondents.

1.1 Structure of Dissertation

My dissertation consists of seven chapters. The first is an overview of subject and objectives of my dissertation. This is followed by a literature review that gives an idea of the historical development of factor analysis as well as the use of factor analysis for analysing Likert scales and tourism data. The third chapter deals with the theoretical aspects of factor analysis and factor scores. The fourth chapter consists of a detailed description of the case study. The next two chapters are about the application of factor analysis, factors scores and the relationship of the scores with the socio-demographic variables present in the traveller survey conducted by the MTA. Finally, the conclusion summarizes the main outcomes and also confirms the need of more in-depth study of the subject.

2 REVIEW OF THE LITERATURE

2.1 Historical Development of Factor Analysis and Scores

2.1.1 Factor Analysis

Galton (1888) first introduced the original concept of latent factors, but the actual mathematical model originated from Spearman (1904). Spearman assumed that the correlations among a set of intelligence test scores could be generated by a single hidden factor of general intellective ability and a second set of factors reflecting the unique qualities of the individual tests. Later scientists, especially Thurstone, improved this model from two factors to include many common factors. In the new journal, *'Psychometrika'* we find a number of published works that focus on this approach.

There are many writings regarding this topic. Two historical writings are the books '*The Factorial Analysis of Human Ability*' by Thomson (1939) and '*Multiple Factor Analysis*' by Thurstone (1947). Two recent books that provide a psychological perspective are those by Mulaik (1972) and Harman (1976). The one written by Mulaik, has a complete account of the theoretical aspect while that of Harman emphasizes more on the statistical methods and the computational matters.

The major approach of factor analysis within the statistical society is due to Lawley and Maxwell (1971) whose book is the basis as a source of results regarding the normal theory factor model. Recent work, such as that of Bentler, Browne, Joreskog, McDonald and others are considered as a generalization of the factor model. In fact, these generalizations are grouped under the name of the analysis of covariance structures. These generalizations are based on the linear structural equation models, which include not only the basic factor model but also the linear relationships among the factors.

2.1.2 Factor Scores

Henry Thomas Herbert Piaggio (1935) originally considered factor score estimation as a solution to an indeterminancy problem. In fact, if factor scores can be computed, we would have no need of factor score estimation. Yet, during the early 1970's many factor analysis texts did not discuss the basis of the indeterminacy problem and its relation to factor score estimation.

Schonemann and Wang saw the factor score indeterminacy and non-uniqueness as major problems for factor analysis. They pointed out that if these problems are not obscured by misleading terminology, they could lead to an alternative approach of data reduction such as component analysis.

McDonald (1974) did not agree with the conclusions reached by these two scientists and he based his point of view on two major arguments. First, he pointed out that common factors are not subject to indeterminacy since the adopted measure of indeterminacy is not correct. He said that the minimum correlation index promoted by Guttman as a measure of indeterminacy was unreliable. In his second point, he considered the assumption that one of the sets of factor scores fitting the observed data was the true set.

These points of view of McDonald are very similar to those of Spearman (1933) since both of them saw that the measure of indeterminacy is misleading our view and argued against the pessimistic approach.

Mulaik (1976) de-emphasized the importance of this difference by stating that it is of little importance whether we apply one of these correlations, ρ^2 or $2\rho^2-1$, if we consider that these two indices measure different aspects of the same situation. He proved that when different solutions for a factor have equal probability of being considered, then the squared multiple correlation ρ^2 for predicting the factor from the observed variables is the average correlation ρ_{AB} between independently selected alternative solutions A and B.

Green (1976) gave a definition to what he termed as factor score controversy. He started by pointing out the connection between the equations of multiple regression and those of factor analysis. His argument has many similarities with an earlier work of Spearman (1933) and Thomson (1934). He rejected McDonald's second point and instead he pointed out that the use of regression estimates is better since they estimated all of the available sets of factor scores equally well. He also noted that the factor scores are all equally correct since they are all properly estimated by the factor score estimates.

2.2 Influence of Factor Analysis

2.2.1 Likert Scale

It is important to bring to our attention that Likert was not the first to obtain subjective ratings and that the early scale developers used far more sensitive scales than we currently employ. Freyd (1923) discussed the various scale forms available at that time and noted that they tended to be based on 10-point or 100-point formats. This numbering system was definitely the most intuitive and easy to visualize since the traditional counting involved the fingers or toes. It also had the advantage of having a perception of equal psychometric distance between the scale points. This was an essential supposition when such scale was used in combinations with parametric statistics.

Freyd then introduced his 'Graphic rating method' which had the following form:

Does he appear neat or slovenly in his dress?

Extremely neat and clean. Almost a dude.	Appropriately and neatly dressed.	Inconspicuous in dress.	Somewhat careless in his dress.	Very slovenly and unkempt.

The above scale was intended to be used in conjunction with job interviews. He considered a line present on these scales so that the respondents can tick anywhere they wished. Then, he recommended scoring the responses by dividing the line into 10 or 20 equal intervals.

A few years later, Watson (1930) published a similar scale to measure an aspect of subjective quality of life (SQOL) as follows:

Most miserable of all	About three-fourths of the population are happier than you are	The average person of your own age and sex	Happier, on the whole, than three-fourths of the population of similar age and sex	Happiest of all

Then, the scale was scored from 0 to 100.

In 1932, Likert produced his scale which had the following form:

Strongly Approve	Approve	Undecided	Disapprove	Strongly Disapprove

This format is clearly derivative from the previous ones. This scale reduces significantly the number of effective choice-points in two ways. Firstly, the scoring system is no longer continuous and therefore, respondents were now required to tick were necessary. This new format reduces the scoring system to a 1-5 scale. Secondly, he has introduced the bi-dimensional scale with a neutral mid-point.

More than six decades have passed since Likert's formulation was published and until now, it has remained the most popular. The reasons of this popularity include the type of psychometric investigation to which it has been subjected, the difficulty of generating substantially larger numbers of labelled choice points, and the complex nature of alternative scales.

A popular method of obtaining information on human knowledge, attitudes, and behavioural preferences and so on is by applying these types of scales in survey questionnaires. The traditional statistical methods to analyse survey response are frequency analysis, t-test and the measures of central tendency. However, there is a flaw since these methods do not describe the correlation occurring at or between scale level responses, which are the most important features to evaluate unobservable patterns.

From these correlations, we are able to explain the behavioural patterns that are shared within or uniquely associated with some groups of respondents.

Kim and Mueller (1978) argued that factor analysis is an approach in which we can gain insight to survey responses. Factor analysis is a statistical procedure, which extracts a small number of latent variables from a large set of observable variables.

With the advancement of computer knowledge, the issue of tackling Likert scale data is becoming more manageable since modern statistical packages, such as Spss, SAS etc are rendering this analysis easier and reliable.

2.2.2 Tourism Data Analysis

Tourism is one of the largest industries in the world (World Tourism Organization [WTO], 1998) and it continues to grow. Tourism is a multifaceted field and tourism research focuses on a variety of areas. Smith (1989) classifies tourism research into the following categories: (1) tourism as a human experience, (2) tourism as a social behaviour, (3) tourism as a geographic phenomenon, (4) tourism as an economic resource, (5) tourism as an industry, and (6) tourism as a business.

Reading various papers; (1)Sevil Sonmez and Ercan Sirakaya "A Distorted Destination Image? The Case of Turkey", (2) Nick Johns and Szilvia Gyimothy "Market Segmentation and the Prediction of Tourist Behaviour: The Case of Bornholm, Denmark", (3) Metin Kozak and Mike Rimmington "Tourist Satisfaction with Mallorca, Spain, as an Off-Season Holiday Destination" and others, we note that factor analysis is the most appropriate statistical approach. There are various areas in which factor analysis is used, such as in image analysis, where the gathered data is analyzed to understand the tourists' perspectives of that country. The focus of these types of questionnaires is that of supporting promotional exercises. Another issue is the marketing aspect, in which the factors obtained from the different data sets show a behavioural differentiation between specific activities. Marketing strategy consists of the following interrelated tasks: (1) setting marketing goals, (2) segmenting the market and selecting one or more target markets, (3) positioning the product/service, and (4) developing the appropriate marketing mix (Harrell & Frazier, 1999, Perreault & McCarthy, 1999:53). Prior to addressing these tasks, a SWOT (Strengths, Weaknesses, Opportunities, and Threats) analysis should be completed. A typical example is *the traveller survey.*

Most recent traveller surveys are now more elaborate since in addition to the traditional socio – demographic questions, attitudinal questions are added. Although attitudinal surveys are not generally classified as qualitative methods, they provide a means for measuring qualitative factors important in travel behaviour. Most traveller surveys follow the same pattern where a series of attitudinal questions in the form of statements are asked. The respondents are asked whether they agree or disagree on a 5-point or 7-point scale, known as Likert scale. Factor analysis is usually used so that it reduces the questions into a smaller set of factors that are then included as exploratory variables in the travel behavioural models. Analyses of these surveys constantly show that at least some attitudinal factors are significant predictors of travel behaviour.

3 STATISTICAL METHODOLOGY

In this chapter, we have integrated the main theoretical results that are at the centre of the statistical techniques considered in this study. We have two sections in which we focus on the factor analysis and scores estimation respectively.

3.1 Factor Analysis

3.1.1 Introduction

Factor analysis is a mathematical model that attempts to explain the correlation between a large set of variables in terms of a small number of underlying factors. These underlying factors are most of the times unobservable and these are present in subjects such as psychology. In fact, psychologists originated the concept of factor analysis since in psychology it is not possible to measure exactly certain abstract quantities one is studying.

3.1.2 The Normal Factor Model

3.1.2.1 Definition and Properties

Let $\mathbf{x}$ be a $(p \times 1)$ random vector with mean $\boldsymbol{\mu}$ and covariance matrix Σ. Then $\mathbf{x}$ fits the k-factor model if it is represented in the form

$$\mathbf{x} = \Lambda\mathbf{f} + \mathbf{u} + \boldsymbol{\mu}$$

where Λ is a $(p \times k)$ matrix of constants

f is a $(k \times 1)$ random vectors, and

u is a $(p \times 1)$ random vectors.

The elements of **f** are known as common factors while those of **u** are called specific factors. Both **f** and **u** are assumed to be jointly normally distributed. $\Lambda = \begin{bmatrix} \lambda_{11} & \cdots & \lambda_{1k} \\ \cdots & \cdots & \cdots \\ \lambda_{p1} & \cdots & \lambda_{pk} \end{bmatrix}$ is the matrix of factor loadings since, λ_{ij} is a parameter reflecting the importance or in other words the loading of the jth factor with the ith response.

Typical applications of the k-factor model are for instance in psychology, where **x** may represent p results of tests measuring intelligence scores. One common latent factor explaining $\mathbf{x} \in \mathbb{R}^p$ could be the overall level of intelligence. In marketing studies, **x** may consist of p answers to a survey on the levels of satisfaction of the customers. These p measures could be explained by common latent factors like the attraction level of the product or the image of the brand, and so on.

Now, we assume that

$E(\mathbf{f}) = \mathbf{0}$ and $V(\mathbf{f}) = E(\mathbf{f}\mathbf{f}^t) = I$,

$E(\mathbf{u}) = \mathbf{0}, \operatorname{cov}(u_i u_j) = 0, i \neq j$

and that **f** and **u** are independent so,

$$\operatorname{cov}(\mathbf{f},\mathbf{u}) = E\{[\mathbf{u} - E(\mathbf{u})][\mathbf{f} - E(\mathbf{f})]^t\} = E(\mathbf{u}\mathbf{f}^t) = \mathbf{0}.$$

Let us define that $V(\mathbf{u}) = E(\mathbf{u}\mathbf{u}^t) = \Psi = \begin{pmatrix} \psi_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \psi_{pp} \end{pmatrix}$.

These assumptions, together with the factor model compose the orthogonal factor model.

Hence, for the orthogonal factor model we can evaluate the expectation and the covariance matrix of **x** as follows

$E(\mathbf{x}) = E(\Lambda\mathbf{f} + \mathbf{u} + \mu) = \Lambda E(\mathbf{f}) + E(\mathbf{u}) + \mu = \mu$ since $E(\mathbf{f}) = \mathbf{0}$ and $E(\mathbf{u}) = \mathbf{0}$

and

$$\Sigma = \text{cov}(\mathbf{x}\mathbf{x}^t) = \text{var}(\mathbf{x}) = \text{var}(\Lambda\mathbf{f} + \mathbf{u} + \mu) = \Lambda \text{var}(\mathbf{f})\Lambda^t + \text{var}(\mathbf{u}) = \Lambda\Lambda^t + \Psi .$$

Another representation of the factor model is

$$x_i = \sum_{j=1}^{k} \lambda_{ij} f_j + u_i + \mu_i, i = 1,\ldots,p$$

implying that the variance matrix of **x** is

$$\sigma_{ii} = \sum_{i=1}^{k} \lambda_{ij}^2 + \psi_{ii} .$$

Therefore, the variance of **x** is split into two parts. The first part, known as the communality is

$$h_i^2 = \sum_{j=1}^{k} \lambda_{ij}^2 .$$

The communalities represent the part of the variance of x_i that is shared with the other variables through the common factors. The second part ψ_{ii}, known as the specific variance, defines the variability of x_i not shared with the other variables.

To clarify the concepts above we perform a few numerical calculations by considering a one-factor model.

Example: Suppose population Σ is

$$\text{cov}(\mathbf{x}) = \begin{pmatrix} 0.65642 & -0.084483 & 0.088283 \\ -0.084483 & 0.2633 & -0.0225 \\ 0.088283 & -0.0225 & 0.10997 \end{pmatrix} = \Sigma .$$

But $\Sigma = \Lambda\Lambda^t + \Psi = \begin{pmatrix} \lambda_1^2 & \lambda_1\lambda_2 & \lambda_1\lambda_3 \\ \lambda_2\lambda_1 & \lambda_2^2 & \lambda_2\lambda_3 \\ \lambda_3\lambda_1 & \lambda_3\lambda_3 & \lambda_3^2 \end{pmatrix} + \begin{pmatrix} \psi_{11} & 0 & 0 \\ 0 & \psi_{22} & 0 \\ 0 & 0 & \psi_{33} \end{pmatrix}$.

Therefore, $\begin{pmatrix} \lambda_1^2+\psi_1 & \lambda_1\lambda_2 & \lambda_1\lambda_3 \\ \lambda_2\lambda_1 & \lambda_2^2+\psi_2 & \lambda_2\lambda_3 \\ \lambda_3\lambda_1 & \lambda_3\lambda_3 & \lambda_3^2+\psi_3 \end{pmatrix} = \begin{pmatrix} 0.65642 & -0.084483 & 0.088283 \\ -0.084483 & 0.2633 & -0.0225 \\ 0.088283 & -0.0225 & 0.10997 \end{pmatrix}$.

This implies that $\begin{matrix} 0.65642 = \lambda_1^2 + \psi_{11} \\ 0.2633 = \lambda_2^2 + \psi_{22} \\ 0.10997 = \lambda_3^2 + \psi_{33} \end{matrix}$ and $\begin{matrix} -0.084483 = \lambda_1\lambda_2 \\ 0.088283 = \lambda_1\lambda_3 \\ -0.0225 = \lambda_2\lambda_3 \end{matrix}$.

Hence, $\frac{-0.084483}{\lambda_1} = \lambda_2$ and $\frac{0.088283}{\lambda_1} = \lambda_3$

$$\Rightarrow -0.0225 = \lambda_2\lambda_3 = \left(\frac{-0.084483}{\lambda_1}\right)\left(\frac{0.088283}{\lambda_1}\right) = \frac{-0.007458}{\lambda_1^2}$$

$$\Rightarrow \lambda_1^2 = 0.33147$$

$$\Rightarrow \lambda_1 = 0.57573$$

$$0.65642 = \lambda_1^2 + \psi_{11} = 0.3315 + \psi_{11}$$

$$\Rightarrow \psi_{11} = 0.324953$$

Similarly for the other λ's and ψ's. Therefore, we get

$$\lambda_2 = -0.146741$$

$$\lambda_3 = 0.153341$$

$$\psi_{22} = 0.241767$$

$$\psi_{33} = 0.086457$$

$$\therefore \Lambda_x = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} = \begin{pmatrix} 0.57573 \\ -0.146741 \\ 0.153341 \end{pmatrix} \text{ and } \Psi_x = \begin{pmatrix} \psi_{11} & 0 & 0 \\ 0 & \psi_{22} & 0 \\ 0 & 0 & \psi_{33} \end{pmatrix} = \begin{pmatrix} 0.324953 & 0 & 0 \\ 0 & 0.241767 & 0 \\ 0 & 0 & 0.086457 \end{pmatrix}$$

Therefore, we can say that for the covariance matrix given we have the factor analytic representation given above.

□

3.1.2.2 Scale Invariance

It is important to note that factor analysis is unaffected by a re-scaling of the variables. Let $\mathbf{y} = \mathbf{Cx}$ where $\mathbf{C} = \text{diag}(c_i)$. So for the k-factor model, let $\Lambda = \Lambda_x$ and $\Psi = \Psi_x$.

Therefore, we obtain

$$\mathbf{y} = \mathbf{Cx} = \mathbf{C}\Lambda_x\mathbf{f} + \mathbf{Cu} + \mathbf{C}\mu$$

and

$$\text{var}(\mathbf{y}) = \mathbf{C}\Sigma\mathbf{C} = \mathbf{C}\Lambda_x\Lambda_x^t\mathbf{C} + \mathbf{C}\Psi_x\mathbf{C} = \Lambda_y\Lambda_y^t + \Psi_y.$$

Hence, the k-factor model holds for $\mathbf{y}$ with factor loading matrix $\Lambda_y = \mathbf{C}\Lambda_x$ and specific variance $\Psi_y = \mathbf{C}\Psi_x\mathbf{C} = \text{diag}(c_i^2\psi_{ii})$.

Example: Let us take the previous example and let's consider $\mathbf{C}$ to be $\mathbf{C} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 3 \end{pmatrix}$. Therefore

$$\mathbf{C}\Sigma\mathbf{C} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 3 \end{pmatrix}\begin{pmatrix} 0.65642 & -0.084483 & 0.088283 \\ -0.084483 & 0.2633 & -0.0225 \\ 0.088283 & -0.0225 & 0.10997 \end{pmatrix}\begin{pmatrix} 2 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

$$= \begin{pmatrix} 2.62568 & -0.84483 & 0.529698 \\ -0.84483 & 6.5825 & -0.3375 \\ 0.529698 & -0.3375 & 0.98973 \end{pmatrix}.$$

But this is equal to $\mathbf{C}\Lambda_x\Lambda_x^t\mathbf{C} + \mathbf{C}\Psi_x\mathbf{C}$, where

$$\Lambda_x = \begin{pmatrix} 0.57573 \\ -0.146741 \\ 0.153341 \end{pmatrix} \text{ and } \Psi_x = \begin{pmatrix} 0.324953 & 0 & 0 \\ 0 & 0.241767 & 0 \\ 0 & 0 & 0.086457 \end{pmatrix}.$$

Therefore, $\mathbf{C}\Lambda_x = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 3 \end{pmatrix}\begin{pmatrix} 0.57573 \\ -0.146741 \\ 0.153341 \end{pmatrix} = \begin{pmatrix} 1.15146 \\ -0.733705 \\ 0.460023 \end{pmatrix}$

$$\Rightarrow \mathbf{C\Lambda_x\Lambda_x^tC} = \begin{pmatrix} 1.15146 \\ -0.733705 \\ 0.460023 \end{pmatrix} \begin{pmatrix} 1.15146 & -0.733705 & 0.460023 \end{pmatrix}$$

$$= \begin{pmatrix} 1.325860 & -0.844832 & 0.529698 \\ -0.844832 & 0.538323 & -0.337521 \\ 0.529698 & -0.337521 & 0.211621 \end{pmatrix}$$

and

$$\mathbf{C\Psi_xC} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 0.324953 & 0 & 0 \\ 0 & 0.241767 & 0 \\ 0 & 0 & 0.086457 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

$$= \begin{pmatrix} 1.299812 & 0 & 0 \\ 0 & 6.044175 & 0 \\ 0 & 0 & 0.778113 \end{pmatrix}$$

$$\therefore \mathbf{C\Lambda_x\Lambda_x^tC} + \mathbf{C\Psi_xC} = \begin{pmatrix} 2.625672 & -0.844832 & 0.529698 \\ -0.844832 & 6.582498 & -0.337521 \\ 0.529698 & -0.337521 & 0.989734 \end{pmatrix} = \mathbf{C\Sigma C}.$$

□

3.1.2.3 Non-Uniqueness of Factor Loadings

Now, given that the k-factor model for $\mathbf{x}$ holds, then it also holds if the factors are rotated. Therefore introducing an orthogonal $(k \times k)$ matrix **G**, $\mathbf{x}$ is written as

$$\mathbf{x} = (\Lambda\mathbf{G})(\mathbf{G}^t\mathbf{f}) + \mathbf{u} + \boldsymbol{\mu}.$$

Given the presence of the new factors $\mathbf{G}^t\mathbf{f}$ and new factor loadings $\Lambda\mathbf{G}$, the k-factor model is still valid since the assumptions of the random vector **f** are applicable. This implies that the covariance matrix Σ of $\mathbf{x}$ is transformed to

$$\Sigma = (\Lambda\mathbf{G})(\mathbf{G}^t\Lambda^t) + \Psi.$$

If Ψ is fixed, this rotation hinders the decomposition of Σ in terms of Λ and Ψ. This is solved by rotating the factor loadings to satisfy the following constraint

$$\Lambda^t\Psi^{-1}\Lambda \text{ is diagonal}.$$

An interesting observation can be made by comparing the number of parameters of Σ when it is unconstrained with the number of free parameters in the factor model and letting s signify this difference.

$$\therefore s = \text{number of parameters of } \Sigma \text{ - free parameters}$$

$$\Rightarrow s = \frac{1}{2}p(p+1) - \left\{pk + p - \frac{1}{2}k(k-1)\right\}$$

$$\Rightarrow s = \frac{1}{2}(p-k)^2 - \frac{1}{2}(p+k).$$

If $s \geq 0$, Λ, Ψ are known and the rotated factor model holds, then Σ is written in terms of Λ and Ψ subject to the constraint $\Lambda^t\Psi^{-1}\Lambda$ is diagonal on Λ.

In the case of our example, we have that p is equal to 3 while k is equal to 1. Therefore,

$$s = \frac{1}{2}(3-1)^2 - \frac{1}{2}(3+1) = \frac{1}{2}(4) - \frac{1}{2}(4) = 0 \geq 0.$$

3.1.2.4 Use of the Covariance Matrix S

An estimation strategy can be devised by replacing Σ by **S** in the previous equation. Therefore, we have to estimate Λ and Ψ from **S**, that is evaluating $\hat{\Lambda}$ and $\hat{\Psi}$ such that they satisfy the constraint $\Lambda^t\Psi^{-1}\Lambda$ is diagonal for the equation

$$\mathbf{S} \doteq \hat{\Lambda}\hat{\Lambda}^t + \hat{\Psi}.$$

Another representation of the diagonal of **S** is

$$s_{ii} = \sum_{j=1}^{k}\hat{\lambda}_{ij}^2 + \hat{\psi}_{ii},\ i = 1,\ldots,p$$

$$\Rightarrow \hat{\psi}_{ii} = s_{ii} - \sum_{j=1}^{k}\hat{\lambda}_{ij}^2,\ i = 1,\ldots,p.$$

Let us take into consideration those estimates that satisfy the equation of $\hat{\psi}_{ii}$ and that $\hat{\psi}_{ii} \geq 0$. Then, setting $\hat{\Sigma} = \hat{\Lambda}\hat{\Lambda}^t + \hat{\Psi}$ implies that $\hat{\sigma}_{ii} = \sum_{j=1}^{k} \hat{\lambda}_{ij}^2 + \hat{\psi}_{ii}$ and substituting for $\hat{\psi}_{ii}$

$$\Rightarrow \hat{\sigma}_{ii} = \sum_{j=1}^{k} \hat{\lambda}_{ij}^2 + s_{ii} - \sum_{j=1}^{k} \hat{\lambda}_{ij}^2, \ i = 1, \ldots, p$$

$$\Rightarrow \hat{\sigma}_{ii} = s_{ii}, \ i = 1, \ldots, p.$$

Depending on the difference value of s, there are three possibilities that can occur in **S**.

1. If $s < 0$, **S** has more parameters then equations. This implies that there exists an infinity of exact solutions of Λ and Ψ. As a result the factor model is not well defined.
2. If $s = 0$, **S** can be solved for exact solutions of Λ and Ψ. Thus, this model has the same number of parameters as Σ.
3. If $s > 0$, there are more equations than parameters. Hence, to solve **S** we use approximate solutions.

3.1.2.5 Use of the Correlation Matrix R

Note that the factor model is scale invariant, so we shall take into consideration estimates $\Lambda = \Lambda_x$ and $\Psi = \Psi_x$ which are scale invariant.

Let $\mathbf{Y} = \mathbf{HXD}_s^{-1/2}$ where $\mathbf{D}_s = \mathrm{diag}(s_{11}, \ldots, s_{pp})$, denote the standardized variables so that

$$\sum_{r=1}^{n} y_{rj} = 0 \text{ and } \frac{1}{n}\sum_{r=1}^{n} y_{rj}^2 = 1, \ j = 1, \ldots, p.$$

Then the estimated factor loading matrix of **Y** is $\hat{\Lambda}_y = \mathbf{D}_s^{-1/2}\hat{\Lambda}_x$ and the estimated specific variances are $\hat{\Psi}_y = \mathbf{D}_s^{-1}\Psi_x$. Consequently, the correlation matrix of **x** is

$$\mathbf{R} \doteq \hat{\Lambda}_y\hat{\Lambda}_y^t + \hat{\Psi}_y.$$

The correlation matrix **R** has the value of 1 in its diagonal so

$$1=\sum_{j=1}^{k}\hat{\lambda}_{ij}^{2}+\hat{\psi}_{ii},\ i=1,\ldots,p$$

$$\Rightarrow \hat{\psi}_{yii}=1-\sum_{j=1}^{k}\hat{\lambda}_{yij}^{2},\ i=1,\ldots,p$$

implying that Ψ_y is no longer a parameter of the model but a function of Λ_y.

Observe that **R** is made up from p fewer free parameters than **S**. *s*, the difference between the number of equations and the number of free parameters present in **R**, is still calculated by

$$s=\frac{1}{2}(p-k)^2-\frac{1}{2}(p+k)$$

where the p equations for the estimates of the scaling parameters are given by $\hat{\sigma}_{ii}=s_{ii},\ i=1,\ldots,p$.

Example: Using the data of this thesis, we consider a one-factor model and apply the following (4×4) sample correlation matrix **R** from x_1 till x_4.

$$\begin{pmatrix}\hat{\lambda}_1^2+\hat{\psi}_{11} & \hat{\lambda}_1\hat{\lambda}_2 & \hat{\lambda}_1\hat{\lambda}_3 & \hat{\lambda}_1\hat{\lambda}_4\\ \hat{\lambda}_2\hat{\lambda}_1 & \hat{\lambda}_2^2+\hat{\psi}_{22} & \hat{\lambda}_2\hat{\lambda}_3 & \hat{\lambda}_2\hat{\lambda}_4\\ \hat{\lambda}_3\hat{\lambda}_1 & \hat{\lambda}_3\hat{\lambda}_2 & \hat{\lambda}_3^2+\hat{\psi}_{33} & \hat{\lambda}_3\hat{\lambda}_4\\ \hat{\lambda}_4\hat{\lambda}_1 & \hat{\lambda}_4\hat{\lambda}_2 & \hat{\lambda}_4\hat{\lambda}_3 & \hat{\lambda}_4^2+\hat{\psi}_{44}\end{pmatrix}=\begin{pmatrix}1 & 0.26027 & 0.21564 & 0.2454\\ 0.26027 & 1 & 0.10072 & 0.14221\\ 0.21564 & 0.10072 & 1 & 0.3529\\ 0.2454 & 0.14221 & 0.3529 & 1\end{pmatrix}$$

With the given values of **R**, we obtain the following answers

$\hat{\lambda}_1\hat{\lambda}_2=0.26027, \hat{\lambda}_1\hat{\lambda}_3=0.21564$ and $\hat{\lambda}_1\hat{\lambda}_4=0.2454$

$\Rightarrow \hat{\lambda}_3=\hat{\lambda}_2 0.82852$ and $\hat{\lambda}_4=1.138\hat{\lambda}_3$

Also, $\hat{\lambda}_3\hat{\lambda}_2=0.10072$

$\Rightarrow \hat{\lambda}_2^2 0.82852=0.10072$

$\Rightarrow \hat{\lambda}_2^2=0.12157$

$\Rightarrow \hat{\lambda}_2=0.34866$

Therefore $\hat{\lambda}_3 0.34866 = 0.10072$

$$\Rightarrow \hat{\lambda}_3 = 0.28888$$
$$\Rightarrow \hat{\lambda}_4 = 0.32875$$
$$\Rightarrow \hat{\lambda}_1 = 0.74646$$

Now, $\hat{\psi}_{11} = 1 - \lambda_1^2, \hat{\psi}_{22} = 1 - \lambda_2^2,\ \hat{\psi}_{33} = 1 - \lambda_3^2$ and $\hat{\psi}_{44} = 1 - \hat{\lambda}_4^2$

$$\Rightarrow \hat{\psi}_{11} = 0.442797$$
$$\hat{\psi}_{22} = 0.878436$$
$$\hat{\psi}_{33} = 0.916548$$
$$\hat{\psi}_{44} = 0.891923$$

Since $\hat{\psi}_{ii} = 1 - \hat{h}_i^2 = 1 - \hat{\lambda}_i^2$, the model explains a higher proportion of the variance of x_1 than of x_2, x_3 and x_4.

□

Now, we shall consider two methods of estimating the parameters of the factor model when $s > 0$. The first method is the principal factor analysis and the second one is the maximum likelihood factor analysis. The latter is applied when we assume the data to be normally distributed.

3.1.2.6 Methods of Estimation

(i): Principal Factor Method

This method is constructed so as to estimate the k-factor model parameters Λ and Ψ. Let the data yield the correlation matrix **R.**

First, obtain estimates $\tilde{h}_i^2$ of the communalities h_i^2, $i = 1,\ldots,p$. There are various ways to estimate the communalities such as when considering the largest correlation coefficient between the ith variable and one of the other variables, that is, $\max_{j \neq i} |r_{ij}|$. The most frequent method of estimation is by considering the square of the multiple correlation of the ith variable with all the other variables.

Then, consider the matrix $\mathbf{R}-\hat{\Psi}$, which is called the reduced correlation matrix since the 1s on the diagonal are replaced by the estimated communalities $\tilde{h}_i^2 = 1-\tilde{\psi}_{ii}$. Applying the spectral decomposition theorem (view **Appendix A**), it is evaluated by

$$\mathbf{R}-\hat{\Psi} = \sum_{i=1}^{p} a_i \gamma_{(i)} \gamma_{(i)}^t$$

where $a_1 \geq \ldots \geq a_p$ are eigenvalues while $\gamma_{(1)}, \ldots, \gamma_{(p)}$ are the orthonormal eigenvectors.

Now, consider the supposition that the first k eigenvalues of the reduced correlation matrix are positive. Then, the ith column of Λ is estimated by

$$\hat{\lambda}_{(i)} = a_i^{1/2} \gamma_{(i)}, i = 1, \ldots, k.$$

In other words, it means that $\hat{\lambda}_{(i)}$ is proportional to the ith eigenvector of the reduced correlation matrix. In matrix form, it is presented as

$$\hat{\Lambda} = \Gamma_1 \mathbf{A}_1^{1/2}$$

where $\Gamma_1 = \left(\gamma_{(1)}, \ldots, \gamma_{(k)}\right)$ and $\mathbf{A}_1 = \text{diag}\left(a_1, \ldots a_k\right)$. Since the eigenvectors are orthogonal, then, $\hat{\Lambda}^t \hat{\Lambda}$ is diagonal which satisfies the constraint $\Lambda^t \Psi^{-1} \Lambda$ is diagonal. Then, the modified estimates of the specific variances are given by

$$\hat{\psi}_{ii} = 1 - \sum_{j=1}^{k} \hat{\lambda}_{ij}^2, \; i = 1, \ldots, p \text{ which is in terms of } \hat{\Lambda}.$$

Note that, in this method we are performing data reduction and hence considering the estimated value of $\mathbf{R}-\hat{\Psi}$. If the data reduction was to perform well, we will obtain some eigenvalues which are positive and others which are 0 or close to 0. In that case, the principal factor solution would be suitable if all the $\hat{\psi}_{ii}$'s are non-negative. However, due to our estimation process, our data reduction may provide us with a mixed set of eigenvalues, made up of positive, negative and zero eigenvalues. Consequently, our principal factor solution would be suitable depending on the nature of the eigenvalues. In fact, it may be the case that the principal factor solution is suitable if for example all $\hat{\psi}_{ii}$'s are all positive or if all $\hat{\psi}_{ii}$'s are non-negative or also if all the $\hat{\psi}_{ii}$'s are negative.

(ii): Maximum Likelihood Factor Method

In maximum likelihood estimation, under normality, the data considers both **f** and **u** to be jointly normally distributed, where $\mathbf{f} \sim N_k(\mathbf{0},\mathbf{I})$ and $\mathbf{u} \sim N_p(\mathbf{0},\Psi)$. Then **x**, which is a linear function with respect to **f** and **u**, is also normally distributed where $\mathbf{x} \sim N_p(\boldsymbol{\mu},\Sigma)$. Therefore, the joint distribution of the **x**'s is

$$L(\mathbf{x};\boldsymbol{\mu},\Sigma) = |2\pi\Sigma|^{-n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu})^t \Sigma^{-1}(\mathbf{x}_i - \boldsymbol{\mu})\right\}.$$

The log-likelihood function of L is

$$l(\mathbf{x};\boldsymbol{\mu},\Sigma) = -\frac{n}{2}\log|2\pi\Sigma| - \frac{n}{2}\mathrm{tr}\Sigma^{-1}\mathbf{S} - \frac{n}{2}(\bar{\mathbf{x}} - \boldsymbol{\mu})^t \Sigma^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu})$$

where $\mathbf{S} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^t$ is the sample covariance matrix.

Replacing $\boldsymbol{\mu}$ by its maximum likelihood estimator $\bar{\mathbf{x}} = \hat{\boldsymbol{\mu}}$, then the log likelihood function becomes

$$l = -\frac{n}{2}\log|2\pi\Sigma| - \frac{n}{2}\mathrm{tr}\Sigma^{-1}\mathbf{S}.$$

Taking into account that $\Sigma = \Lambda\Lambda^t + \Psi$ is a function of Λ and Ψ, then l can be maximized with respect to these parameters.

Consider the function

$$F(\Lambda,\Psi) = F(\Lambda,\Psi;\mathbf{S}) = \mathrm{tr}\Sigma^{-1}\mathbf{S} - \log|\Sigma^{-1}\mathbf{S}| - p$$

where $\Sigma = \Lambda\Lambda^t + \Psi$.

F is a linear function of the log-likelihood l and a maximum in l is equivalent to a minimum in F. The minimization of F is split into two stages. The first stage is to minimize over Ψ and the second stage is to minimize over Λ.

Let us begin by differentiating with respect to the specific variances. Then, we have that

$$F(\Lambda,\Psi)=F(\Lambda,\Psi;\mathbf{S})=\mathrm{tr}\Sigma^{-1}\mathbf{S}-\log\left|\Sigma^{-1}\mathbf{S}\right|-\mathrm{p}.$$

Note that, $\Psi=\Sigma-\Lambda\Lambda^{t}$ is a diagonal matrix and we would like to estimate the coefficients with respect to Σ. Hence, we can consider

$$\frac{\partial F}{\partial\Psi}=\mathrm{diag}\left(\frac{\partial F}{\partial\Sigma}\right)$$

Now

$$\frac{\partial\left(-\log\left|\Sigma^{-1}\mathbf{S}\right|\right)}{\partial\Sigma}=\frac{\partial\left(\log|\Sigma|-\log|\mathbf{S}|\right)}{\partial\Sigma}$$

$$=\Sigma^{-1}\ \text{since}\ \frac{\partial\log|\mathbf{X}|}{\partial\mathbf{X}}=\mathbf{X}^{-1}.$$

Where $|\mathbf{X}|=\sum_{j=1}^{p}x_{ij}X_{ij}$, which implies that $\dfrac{\partial\log|\mathbf{X}|}{\partial x_{ij}}=\dfrac{\partial\log\left(\sum_{j=1}^{p}x_{ij}X_{ij}\right)}{\partial x_{ij}}=\dfrac{X_{ij}}{|\mathbf{X}|}$

$$\Rightarrow\frac{\partial\log|\mathbf{X}|}{\partial\mathbf{X}}=\mathbf{X}^{-1}.$$

Now $\mathrm{diag}\left(\dfrac{\partial\left(\mathrm{tr}\Sigma^{-1}\mathbf{S}\right)}{\partial\Sigma}\right)$ is a diagonal matrix with entries $\dfrac{\partial\left(\mathrm{tr}\Sigma^{-1}\mathbf{S}\right)}{\partial\sigma_{ii}}$.

$\therefore\dfrac{\partial\left(\mathrm{tr}\Sigma^{-1}\mathbf{S}\right)}{\partial\sigma_{ii}}=\dfrac{\partial\left(\mathrm{tr}\Sigma^{-1}\mathbf{S}\right)}{\partial\omega_{ii}}\dfrac{\partial\omega_{ii}}{\partial\sigma_{ii}}$ where ω_{ii} are the elements on the diagonal of Σ^{-1}

Now, we have that

$$\frac{\partial\Sigma^{-1}}{\partial\sigma_{ii}}=-\Sigma^{-1}J_{ii}\Sigma^{-1}\ \text{and}\ \frac{\partial\mathrm{tr}\left(\Sigma^{-1}\mathbf{S}\right)}{\partial\Sigma^{-1}}=\mathbf{S}$$

$$\Rightarrow\frac{\partial\mathrm{tr}\left(\Sigma^{-1}\mathbf{S}\right)}{\partial\sigma_{ii}}=-\Sigma^{-1}S_{ii}J_{ii}\Sigma^{-1}$$

$$\Rightarrow\frac{\partial\mathrm{tr}\left(\Sigma^{-1}\mathbf{S}\right)}{\partial\Sigma}=-\Sigma^{-1}\mathbf{S}\Sigma^{-1}$$

where J_{ii} denotes a matrix with 1 in the (i, i)th place and zeros elsewhere.

Gathering all these derivatives, we obtain

$$\frac{\partial F}{\partial \Psi} = \text{diag}\left(\Sigma^{-1} - \Sigma^{-1}\mathbf{S}\Sigma^{-1}\right) = \text{diag}\left(\Sigma^{-1}\left(\Sigma - \mathbf{S}\right)\Sigma^{-1}\right).$$

Our derivative is not equal to zero, but we set the derivation equal to zero to solve the likelihood equations. Therefore

$$\text{diag}\left(\hat{\Sigma}^{-1}\left(\hat{\Sigma} - \mathbf{S}\right)\hat{\Sigma}^{-1}\right) = \mathbf{0}$$

$$\Rightarrow \text{diag}\left(\hat{\Sigma}^{-1}\right) = \text{diag}\left(\hat{\Sigma}^{-1}\mathbf{S}\hat{\Sigma}^{-1}\right)$$

Now, let us obtain the other set of equations by differentiating F with respect to the loading parameters. Therefore, we have that

$$\frac{\partial F}{\partial \Lambda} = \left(\frac{\partial F}{\partial \Sigma}\right)\left(\frac{\partial \Sigma}{\partial \Lambda}\right)$$

We already have $\frac{\partial F}{\partial \Sigma}$ derived from the above procedure. Therefore, $\frac{\partial F}{\partial \Sigma} = \Sigma^{-1}\left(\Sigma - \mathbf{S}\right)\Sigma^{-1}$.

Now, $\frac{\partial \Sigma}{\partial \Lambda} = \frac{\partial\left(\Lambda\Lambda^{t} + \Psi\right)}{\partial \Lambda} = \frac{\partial \Lambda\Lambda^{t}}{\partial \Lambda} + \frac{\partial \Psi}{\partial \Lambda} = \Lambda$.

Grouping these two derivatives together, we obtain

$$\frac{\partial F}{\partial \Lambda} = \left(\frac{\partial F}{\partial \Sigma}\right)\left(\frac{\partial \Sigma}{\partial \Lambda}\right) = \left(\Sigma^{-1}\left(\Sigma - \mathbf{S}\right)\Sigma^{-1}\right)\Lambda = \mathbf{0}.$$

Since our derivative is equal to zero, we can simplify the equation to end up with

$$\mathbf{S}\hat{\Sigma}^{-1}\hat{\Lambda} = \hat{\Lambda}.$$

Considering this equation $\text{diag}\left(\hat{\Sigma}^{-1}\right) = \text{diag}\left(\hat{\Sigma}^{-1}\mathbf{S}\hat{\Sigma}^{-1}\right)$, pre and post multiply both sides by $\hat{\Psi} = \hat{\Sigma} - \hat{\Lambda}\hat{\Lambda}^{t}$. Then we have

$$\text{diag}\left[\left(\hat{\Sigma}-\hat{\Lambda}\hat{\Lambda}^t\right)\left(\hat{\Sigma}^{-1}\right)\left(\hat{\Sigma}-\hat{\Lambda}\hat{\Lambda}^t\right)\right]=\text{diag}\left[\left(\hat{\Sigma}-\hat{\Lambda}\hat{\Lambda}^t\right)\left(\hat{\Sigma}^{-1}\mathbf{S}\hat{\Sigma}^{-1}\right)\left(\hat{\Sigma}-\hat{\Lambda}\hat{\Lambda}^t\right)\right].$$

$$\text{diag}\left[\hat{\Sigma}-2\hat{\Lambda}\hat{\Lambda}^t+\hat{\Lambda}\hat{\Lambda}^t\hat{\Sigma}^{-1}\hat{\Lambda}\hat{\Lambda}^t\right]=\text{diag}\left[\mathbf{S}-\mathbf{S}\hat{\Sigma}^{-1}\hat{\Lambda}\hat{\Lambda}^t-\hat{\Lambda}\hat{\Lambda}^t\hat{\Sigma}^{-1}\mathbf{S}+\hat{\Lambda}\hat{\Lambda}^t\hat{\Sigma}^{-1}\mathbf{S}\hat{\Sigma}^{-1}\hat{\Lambda}\hat{\Lambda}^t\right].$$

Now, using the second derived equation, $\mathbf{S}\hat{\Sigma}^{-1}\hat{\Lambda}=\hat{\Lambda}$, we can simplify

$$\text{diag}\left[\hat{\Sigma}-2\hat{\Lambda}\hat{\Lambda}^t+\hat{\Lambda}\hat{\Lambda}^t\hat{\Sigma}^{-1}\hat{\Lambda}\hat{\Lambda}^t\right]=\text{diag}\left[\mathbf{S}-2\hat{\Lambda}\hat{\Lambda}^t+\hat{\Lambda}\hat{\Lambda}^t\hat{\Sigma}^{-1}\hat{\Lambda}\hat{\Lambda}^t\right]$$

$$\Rightarrow \text{diag}\left(\hat{\Sigma}\right)=\text{diag}\left(\mathbf{S}\right)$$

or that the estimate of the variance is equal to the sample variance.

However, there is a problem, since we need to invert a $(p \times p)$ matrix. Therefore, $\mathbf{S}\hat{\Sigma}^{-1}\hat{\Lambda}=\hat{\Lambda}$ is written in an alternative way

$$\mathbf{S}\left(\hat{\Lambda}\hat{\Lambda}^t+\hat{\Psi}\right)^{-1}\hat{\Lambda}=\hat{\Lambda}$$

$$\mathbf{S}\hat{\Psi}^{-1}\left(\hat{\Lambda}^t\hat{\Psi}^{-1}\hat{\Lambda}+\mathbf{I}\right)^{-1}\hat{\Lambda}=\hat{\Lambda}$$

$$\mathbf{S}\hat{\Psi}^{-1}\hat{\Lambda}=\hat{\Lambda}\left(\hat{\Lambda}^t\hat{\Psi}^{-1}\hat{\Lambda}+\mathbf{I}\right).$$

A solution of the maximum likelihood equations is by representing the above equation as

$$\left(\mathbf{S}-\hat{\Psi}\right)\hat{\Psi}^{-1}\hat{\Lambda}=\hat{\Lambda}\left(\hat{\Lambda}^t\hat{\Psi}^{-1}\hat{\Lambda}\right).$$

Pre-multiplying both sides by $\hat{\Psi}^{-1/2}$ we obtain

$$\hat{\Psi}^{-1/2}\left(\mathbf{S}-\hat{\Psi}\right)\hat{\Psi}^{-1}\hat{\Lambda}=\hat{\Psi}^{-1/2}\hat{\Lambda}\left(\hat{\Lambda}^t\hat{\Psi}^{-1}\hat{\Lambda}\right)$$

$\left[\hat{\Psi}^{-1/2}\left(\mathbf{S}-\hat{\Psi}\right)\hat{\Psi}^{-1/2}\right]\hat{\Psi}^{-1/2}\hat{\Lambda}=\hat{\Psi}^{-1/2}\hat{\Lambda}\mathbf{J}$ where $\mathbf{J}=\hat{\Lambda}^t\hat{\Psi}^{-1}\hat{\Lambda}$ is required to be diagonal.

The successive diagonal elements of $\mathbf{J}$ are the first m characteristic roots of $\hat{\Psi}^{-1/2}\left(\mathbf{S}-\hat{\Psi}\right)\hat{\Psi}^{-1/2}$ and the ith column of $\hat{\Psi}^{-1/2}\hat{\Lambda}$ is the characteristic vector of $\hat{\Psi}^{-1/2}\left(\mathbf{S}-\hat{\Psi}\right)\hat{\Psi}^{-1/2}$ corresponding to the ith largest characteristic root. However, the solution for the roots and vectors must be made iteratively, for the elements of

$$\text{diag}\left(\hat{\Sigma}-\mathbf{S}\right)=\text{diag}\left(\hat{\Lambda}\hat{\Lambda}^{t}+\hat{\Psi}-\mathbf{S}\right)=\mathbf{0}$$
$$\Rightarrow \hat{\Psi}=\text{diag}\left(\mathbf{S}-\hat{\Lambda}\hat{\Lambda}^{t}\right)$$

are unknown. Hence, we begin with an initial approximation of $\hat{\Psi}$ and extract the characteristic roots and vectors of $\hat{\Psi}^{-1/2}\left(\mathbf{S}-\hat{\Psi}\right)\hat{\Psi}^{-1/2}$ to obtain the first approximation to $\hat{\Lambda}$. Then, we compute a second approximation of $\hat{\Psi}$ by $\text{diag}\left(\mathbf{S}-\hat{\Lambda}\hat{\Lambda}^{t}\right)$ and then estimate the characteristic roots. Then, we calculate a second approximation of $\hat{\Lambda}$. This iterative process is continued until the elements of $\hat{\Psi}$ and $\hat{\Lambda}$ matrices have converged to a satisfactory degree.

3.1.2.7 Goodness of Fit Test

For the maximum likelihood method, it is possible to check the adequacy of the k-factor model for generating the observed covariances or correlations. In this case, $\Sigma=\Lambda\Lambda^{t}+\Psi$ and testing the adequacy of the k-factor model is equivalent to testing the null hypothesis

$$H_0:\Sigma=\Psi+\Lambda\Lambda^{t}$$

versus the alternative hypothesis (H_1), which states that Σ is any other positive definite matrix.

Now, to test the null hypothesis H_0 against the alternative hypothesis H_1, we use the likelihood ratio statistic, which states that asymptotically, as $n\rightarrow\infty$

$$-2\ln\left(\frac{\max L_0}{\max L_1}\right)\sim\chi_s^2$$

where $s=\frac{1}{2}(p-k)^2-\frac{1}{2}(p+k)$ degrees of freedom and L_i is the largest value which the likelihood function can take; $i=0,1$.

Now, we have that $\ln(\max L_0) = -\frac{n}{2}\log|2\pi\Sigma| - \frac{n}{2}\text{tr}\Sigma^{-1}\mathbf{S}$ and $\ln(\max L_1) = -\frac{n}{2}\log|2\pi\mathbf{S}| - \frac{n}{2}\text{p}$. Therefore,

$$-2\ln\left(\frac{\max L_0}{\max L_1}\right) = -2\left[\ln(\max L_0) - \ln(\max L_1)\right]$$

$$= -2\left[-\frac{n}{2}\log|2\pi\Sigma| - \frac{n}{2}\text{tr}\Sigma^{-1}\mathbf{S} + \frac{n}{2}\log|2\pi\mathbf{S}| + \frac{n}{2}\text{p}\right]$$

$$= \left[n\log|2\pi| + n\log|\Sigma| + n\text{tr}\Sigma^{-1}\mathbf{S} - n\log|2\pi| - n\log|\mathbf{S}| - n\text{p}\right]$$

$$= n\left[\text{tr}\Sigma^{-1}\mathbf{S} - \log|\Sigma^{-1}| - \log|\mathbf{S}| - \text{p}\right]$$

$$= n\left[\text{tr}\Sigma^{-1}\mathbf{S} - \log|\Sigma^{-1}\mathbf{S}| - \text{p}\right]$$

However, we know that $F(\Lambda, \Psi) = \text{tr}\Sigma^{-1}\mathbf{S} - \log|\Sigma^{-1}\mathbf{S}| - \text{p}$, therefore

$$-2\ln\left(\frac{\max L_0}{\max L_1}\right) = nF(\Lambda, \Psi)$$

which is asymptotically chi-squared distributed with s degrees of freedom since it is equal to our statistic.

Bartlett, showed that the chi-squared approximation of $-2\ln\left(\frac{\max L_0}{\max L_1}\right)$ improves if n is replaced by

$$n' = n - 1 - \frac{1}{6}(2p+5) - \frac{2}{3}k.$$

3.1.3 Fitting Without Normality Assumptions

Until now, we have considered that our model obeys the normality conditions. Nevertheless, data rarely satisfy these assumptions. In fact, it is more probable to have non-normal distributed data. In particular, our data under test does not satisfy the normality assumptions. Hence, now, we will consider the theoretical approach when we fit without normality assumptions.

Let us consider the case when nothing is assumed about the distributions of **f** and **u**. We can still consider the covariance matrix of the k-factor model to be equated by $\Sigma = \Lambda\Lambda^t + \Psi$.

Hence, we will estimate the values of Λ and Ψ by the sample covariance matrix **S**. For this to be possible, we need to obtain a scale measure of distance between Σ and S, which is minimized with respect to the parameters. The idea of a distance is that it separates as much as possible the entities, which are not similar so that they give us clarity. Let l be maximized so that we obtain a distance function

$$\Delta(\Sigma, S) = -\mathrm{tr}\Sigma^{-1}S + \log\left|\Sigma^{-1}S\right|.$$

We note that there are other possible methods to measure this distance. For example, a simple least squares criterion is

$$\Delta_1 = \mathrm{tr}(S - \Sigma)^2.$$

Another method is that of using the matrix $S^* = \Psi^{-1/2}S\Psi^{-1/2}$ as used in maximum likelihood estimator

$$\Delta_2 = \mathrm{tr}\left\{\left(S^* - \Sigma^*\right)^2\right\} = \mathrm{tr}\left\{\Psi^{-1}(S - \Sigma)^2\Psi^{-1}\right\} = \mathrm{tr}\left\{(S - \Sigma)\Psi^{-1}\right\}^2.$$

These two equations are special cases of the general class of measures and the general formula is

$$\Delta = \mathrm{tr}\left\{(S - \Sigma)V\right\}^2.$$

An important aspect of Δ_1 and Δ_2 is that the optimization requires only a solution of a simple eigenvalues problem. For the case of Δ_1 the function that requires to be minimized is

$\Delta_1 = \mathrm{tr}(S - \Sigma)^2$, where $\Sigma = \Lambda\Lambda^t + \Psi$

$\Rightarrow \Delta_1 = \mathrm{tr}\left(S - \Lambda\Lambda^t - \Psi\right)^2.$

Differentiating with respect to Λ

$$\frac{\partial \Delta_1}{\partial \Lambda} = -2\left(\mathbf{S} - \Lambda\Lambda^t - \Psi\right)\Lambda = \left\{\Lambda\left(\Lambda^t\Lambda\right) - \left(\mathbf{S} - \Psi\right)\Lambda\right\} = 0$$

$$\Rightarrow \left(\mathbf{S} - \Psi\right)\Lambda = \Lambda\left(\Lambda^t\Lambda\right).$$

Now, differentiating with respect to Ψ

$$\frac{\partial \Delta_1}{\partial \Psi} = -2\left(\mathbf{S} - \Lambda\Lambda^t - \Psi\right) = 0$$

$$\Rightarrow \text{diag}\frac{\partial \Delta_1}{\partial \Psi} = -\text{diag}\mathbf{S} + \Psi + \text{diag}\Lambda\Lambda^t = 0$$

$$\Rightarrow \Psi = \text{diag}\left(\mathbf{S} - \Lambda\Lambda^t\right).$$

Assume that Ψ is known, then $\left(\mathbf{S} - \Psi\right)\Lambda = \Lambda\left(\Lambda^t\Lambda\right)$ is satisfied if:

a) the columns of Λ consist of any q eigenvalues of $\mathbf{S} - \Psi$

b) the diagonal matrix $\Lambda\Lambda^t$ has elements equal to eigenvalues of $\mathbf{S} - \Psi$ corresponding with the vectors in Λ.

Therefore, if we have a starting value of Ψ in $\left(\mathbf{S} - \Psi\right)\Lambda = \Lambda\left(\Lambda^t\Lambda\right)$, we will generate a first approximation of Λ which will then be inserted into $\Psi = \text{diag}\left(\mathbf{S} - \Lambda\Lambda^t\right)$ to produce a second estimate of Ψ. This process continues until convergence to a satisfactory degree is obtained.

Example: To clarify the process above, let us consider a one-factor model, where the covariance matrix **C** is a representation of **S.** From **C** we obtain the column vector **D**, representing Λ and let us consider an initial value for **B**, which is equivalent to Ψ. Therefore, we have

$$\mathbf{C} \equiv \mathbf{S} = \begin{pmatrix} 2.5833 & -0.16667 & -1.635 \\ -0.16667 & 0.72333 & -0.457 \\ -1.635 & -0.457 & 1.4788 \end{pmatrix} \text{ and } \mathbf{B} \equiv \Psi = \begin{pmatrix} 0.5 & 0 & 0 \\ 0 & 0.35 & 0 \\ 0 & 0 & 0.41 \end{pmatrix}$$

Let $\mathbf{F}=(\mathbf{C}-\mathbf{B})\equiv(\mathbf{S}-\Psi)=\begin{pmatrix}2.0833 & -0.16667 & -1.635\\ -0.16667 & 0.37333 & -0.457\\ -1.635 & -0.457 & 1.0688\end{pmatrix}$ and take $\mathbf{D}\equiv\Lambda=\begin{pmatrix}d_1\\ d_2\\ d_3\end{pmatrix}$.

Therefore, we have that $\mathbf{F}\equiv\mathbf{DD}^t$ from $(\mathbf{S}-\Psi)\Lambda=\Lambda(\Lambda^t\Lambda)$, which implies that

$$\begin{pmatrix}2.0833 & -0.16667 & -1.635\\ -0.16667 & 0.37333 & -0.457\\ -1.635 & -0.457 & 1.0688\end{pmatrix}=\begin{pmatrix}d_1^2 & d_1d_2 & d_1d_3\\ d_2d_1 & d_2^2 & d_2d_3\\ d_3d_1 & d_3d_2 & d_3^2\end{pmatrix}$$

$$\Rightarrow d_1^2=2.0833 \qquad\qquad d_1=1.4437$$
$$d_2^2=0.37333 \text{ and then we get } d_2=0.611$$
$$d_3^2=1.0688 \qquad\qquad d_3=1.0338$$

Now, using iteration $\Psi=\text{diag}(\mathbf{S}-\Lambda\Lambda^t)$, we obtain a new value for **B**.

$$\therefore \mathbf{B}=\text{diag}(\mathbf{C}-\mathbf{DD}^t)=\begin{pmatrix}0.4991\\ 0.3500\\ 0.4101\end{pmatrix}$$

So we can say that for the given sample covariance matrix **S** and initial value of Ψ we can generate a value of Λ. Then, we can continue until we reach the required convergence.

□

Now, to find the eigenvectors of $\mathbf{S}-\Psi$ that form part of Λ, we use the following equation

$$\Delta_1=\text{tr}(\mathbf{S}-\Sigma)^2=\text{tr}(\mathbf{S}-\Psi-\Lambda\Lambda^t)^2$$
$$=\text{tr}(\mathbf{S}-\Psi)^2-2\text{tr}(\mathbf{S}-\Psi)\Lambda\Lambda^t+\text{tr}(\Lambda\Lambda^t)^2$$

but $(\mathbf{S}-\Psi)\Lambda=\Lambda(\Lambda^t\Lambda)$

$$\Rightarrow \Delta_1=\text{tr}(\mathbf{S}-\Psi)^2-2\text{tr}\Lambda(\Lambda^t\Lambda)\Lambda^t+\text{tr}(\Lambda\Lambda^t)^2$$
$$=\text{tr}(\mathbf{S}-\Psi)^2-2\text{tr}(\Lambda\Lambda^t)^2+\text{tr}(\Lambda\Lambda^t)^2$$
$$\Rightarrow \Delta_1=\text{tr}(\mathbf{S}-\Psi)^2-\text{tr}(\Lambda\Lambda^t)^2.$$

Now $\Lambda\Lambda^t$ has *(p-q)* zero eigenvalues since it is of rank *q*. The other eigenvalues still form part of $\mathbf{S}-\Psi$, so $\mathbf{S}-\Psi$ is of rank *p*. Assume that the eigenvalues common between the two matrices are $\theta_1,\ldots,\theta_q$ and the remaining eigenvalues are $\theta_{q+1},\ldots,\theta_p$. Therefore Δ_1 is expressed as

$$\Delta_1 = \sum_{i=1}^{p}\theta_i^2 - \sum_{i=1}^{q}\theta_i^2 = \sum_{i=q+1}^{p}\theta_i^2 .$$

For this expression to be a minimum $\theta_{q+1},\ldots,\theta_p$ must be the smallest eigenvalues. This implies that Λ is composed of eigenvectors corresponding to the *q* largest eigenvalues. This particular procedure is known as the principal factor (or axis) method because it is quite similar to the principal components analysis where $\Psi = 0$.

Now, we consider the estimation process of Δ_2 by differentiating with respect to Λ.

$$\frac{\partial\Delta_2}{\partial\Lambda} = -2\left[\Psi^{-1/2}\mathbf{S}\Psi^{-1/2} - \Psi^{-1/2}\Lambda\Lambda^t\Psi^{-1/2} - \mathbf{I}\right]\Psi^{-1/2}\Lambda = \left\{\Lambda^*\left(\Lambda^*\right)^t\Lambda^* - \left(\mathbf{S}^* - \mathbf{I}\right)\Lambda^*\right\} = 0$$

since $\Lambda^* = \Psi^{-1/2}\Lambda$ and $\mathbf{S}^* = \Psi^{-1/2}\mathbf{S}\Psi^{-1/2}$

$$\Rightarrow \left(\mathbf{S}^* - \mathbf{I}\right)\Lambda^* = \Lambda^*\left(\left(\Lambda^*\right)^t\Lambda^*\right).$$

Let us consider differentiating Δ_2 with respect to Ψ. We note that this process is very complicated. In fact, the above-mentioned methods are a lot easier if we know Ψ. A method of how to avoid this problem is by eliminating Ψ since it only forms part of the diagonal elements of Σ. This approach is known as the minres method. In this case, If we consider the case for Δ_1, we would then minimize

$$\Delta_1' = \mathrm{tr}(\mathbf{S} - \Lambda\Lambda^t).$$

3.1.3.1 Estimability

A necessary condition for the parameters to be estimated is that we need to have at least as many sample statistics as there are parameters. In the k-factor model, there are $(\mathrm{pk}+\mathrm{p})$ parameters but to have a unique solution, we need $\frac{1}{2}\mathrm{k}(\mathrm{k}-1)$ constraints. Therefore the number of free parameters is

$$pk+p-\frac{1}{2}k(k-1).$$

S has $\frac{1}{2}p(p+1)$ distinct elements, so for a consistent estimation, we need

$$\frac{1}{2}p(p+1)-pk-p+\frac{1}{2}k(k-1)=\frac{1}{2}\left[(p-k)^2-(p+k)\right]\geq 0.$$

3.1.3.2 Goodness of Fit and Choice of q

Amemiya and Anderson (1985) showed that if the elements of **f** are independent and that both **f** and **u** have finite second moments, then, if $-2\ln\left(\frac{\max H_0}{\max H_1}\right)=nF(\Lambda,\Psi)$ is evaluated by the maximum likelihood estimators, it has the same distribution as in the normal case. This result also holds for another goodness of fit statistic

$$\frac{n}{2}\mathrm{tr}\left\{\left(S-\hat{\Sigma}\right)\hat{\Sigma}^{-1}\right\}^2.$$

There are also two other methods which do not depend on distributional assumptions. They are based on the role of the eigenvalues of the sample correlation matrix within the principal component analysis.

The first method is the Kaiser – Guttman criterion that chooses q equal to the number of eigenvalues greater than one. The underlying principle is that the average contribution of the evident variable to the total variation is one. Also, the principal component which did not contribute at least as much variation as a single variable represents no advantage.

The second method, presented by Cattell, is known as the 'scree test'. The concept is that the eigenvalues are plotted on a decreasing curve against their rank order. Then, we search for a prod in the curve that will indicate the point from which further addition of factors shows diminishing yield in terms of variation explained.

3.2 Factor Scores

3.2.1 Introduction

So far, we have concentrated on the point of how the observed variables are functions of the unknown factors. Now, let us consider the other side of the coin, where we want to know how the factors depend upon the observed variables. This second approach is the concept of factor scores.

3.2.2 Estimation Method

The most commonly used methods of estimation are the Bartlett's also known as least squares method and Thomson's known as the regression estimate. Now, we will focus on the theoretical approach of these two methods.

3.2.2.1 Bartlett's Method

Let $\mathbf{x}$ be a multinormal random vector of the model $\mathbf{x} = \mathbf{\Lambda f} + \mathbf{u} + \boldsymbol{\mu}$ and assume that Λ, Ψ and $\boldsymbol{\mu} = \mathbf{0}$ are known. Let $\mathbf{f}$ be a $(k \times 1)$ vector formed from the common factor scores and let $\mathbf{x}$ have a $N_p(\mathbf{\Lambda f}, \Psi)$ distribution, i.e.

$$f(x) = (2\pi)^{-1/2} \,|\, \Psi \,|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{x} - \mathbf{\Lambda f})^t \Psi^{-1} (\mathbf{x} - \mathbf{\Lambda f}) \right].$$

Therefore the log likelihood of $\mathbf{x}$ is:

$$l(\mathbf{x}, \mathbf{f}) = -\frac{1}{2}(\mathbf{x} - \mathbf{\Lambda f})^t \Psi^{-1} (\mathbf{x} - \mathbf{\Lambda f}) - \frac{1}{2}\log|2\pi\Psi|.$$

Taking the derivative of l with respect to $\mathbf{f}$ and setting equal to $\mathbf{0}$ gives:

$$\frac{\partial l}{\partial \mathbf{f}} = \Lambda^t \Psi^{-1}(\mathbf{x} - \mathbf{\Lambda f}) = \Lambda^t \Psi^{-1}\mathbf{x} - \Lambda^t \Psi^{-1} \mathbf{\Lambda f} = \mathbf{0}$$

$$\Rightarrow \Lambda^t \Psi^{-1}\mathbf{x} = \Lambda^t \Psi^{-1} \mathbf{\Lambda f}$$

$$\Rightarrow \hat{\mathbf{f}} = \left(\Lambda^t \Psi^{-1} \Lambda\right)^{-1} \Lambda^t \Psi^{-1} \mathbf{x}.$$

Bartlett's procedure is intended to keep the non-common factors fixed so that they are used only to explain the discrepancies between the observed scores and those reproduced from the common factors. From this, one can estimate the specific factor scores through the equation:

$$\hat{\mathbf{u}} = \mathbf{x} - \Lambda\hat{\mathbf{f}}.$$

Let us cvaluate the expected value and the predicted error. Therefore,

Bartlett's expectation is:

$$E\left(\hat{\mathbf{f}}|\mathbf{f}\right) = E\left(\left(\Lambda^t\Psi^{-1}\Lambda\right)^{-1}\Lambda^t\Psi^{-1}\mathbf{x}\right)$$

$$\Rightarrow E\left(\hat{\mathbf{f}}|\mathbf{f}\right) = \left(\Lambda^t\Psi^{-1}\Lambda\right)^{-1}\Lambda^t\Psi^{-1}E\left(\mathbf{x}\right)$$

$$\Rightarrow E\left(\hat{\mathbf{f}}|\mathbf{f}\right) = \left(\Lambda^t\Psi^{-1}\Lambda\right)^{-1}\Lambda^t\Psi^{-1}\Lambda\mathbf{f} \text{ since } E\left(\mathbf{x}\right) = \Lambda\mathbf{f}$$

$$\Rightarrow E\left(\hat{\mathbf{f}}|\mathbf{f}\right) = \mathbf{f}.$$

and Bartlett's predicted error is:

$$\Lambda E\left(\left(\hat{\mathbf{f}}-\mathbf{f}\right)\left(\hat{\mathbf{f}}-\mathbf{f}\right)^t\right)\Lambda^t$$

$$= E\left(\Lambda\left(\hat{\mathbf{f}}-\mathbf{f}\right)\left(\hat{\mathbf{f}}-\mathbf{f}\right)^t\Lambda^t\right)$$

$$= E\left(\left(\mathbf{x}-\mathbf{x}+\mathbf{u}\right)\left(\mathbf{x}-\mathbf{x}+\mathbf{u}\right)^t\right)$$

$$= E\left(\mathbf{uu}^t\right)$$

but $E\left(\mathbf{uu}^t\right) = \Psi$.

So $\Lambda E\left(\left(\hat{\mathbf{f}}-\mathbf{f}\right)\left(\hat{\mathbf{f}}-\mathbf{f}\right)^t\right)\Lambda^t = \Psi$

$$\therefore E\left(\left(\hat{\mathbf{f}}-\mathbf{f}\right)\left(\hat{\mathbf{f}}-\mathbf{f}\right)^t\right) = \left(\Lambda^t\Psi\Lambda\right)^{-1}$$

□

3.2.2.2 Thomson's Method

For the alternative estimate, let **f** be a random vector with a $N_p(\mathbf{0},\mathbf{I})$ distribution. First, we consider the Bayesian approach for the common factor **f**:

$$\mathrm{P}(\mathbf{f}|\mathbf{x}) \propto \exp\left(-\frac{1}{2}(\mathbf{x}-\Lambda\mathbf{f})^t \Psi^{-1}(\mathbf{x}-\Lambda\mathbf{f})\right)\times\exp\left(-\frac{1}{2}\mathbf{f}^t\mathbf{f}\right)$$

$$\Rightarrow \mathbf{f} \propto \exp\left[-\frac{1}{2}(\mathbf{x}-\Lambda\mathbf{f})^t \Psi^{-1}(\mathbf{x}-\Lambda\mathbf{f})-\frac{1}{2}\mathbf{f}^t\mathbf{f}\right].$$

Taking logs and then differentiating with respect to **f**, implies:

$$l = \ln\mathbf{f} = -\frac{1}{2}(\mathbf{x}-\Lambda\mathbf{f})^t \Psi^{-1}(\mathbf{x}-\Lambda\mathbf{f})-\frac{1}{2}\mathbf{f}^t\mathbf{f}$$

$$\Rightarrow \frac{\partial l}{\partial \mathbf{f}} = \Lambda^t\Psi^{-1}(\mathbf{x}-\Lambda\mathbf{f})-\mathbf{f} = \Lambda^t\Psi^{-1}\mathbf{x}-\Lambda^t\Psi^{-1}\Lambda\mathbf{f}-\mathbf{f} = \mathbf{0}$$

$$\Rightarrow \Lambda^t\Psi^{-1}\mathbf{x} = \Lambda^t\Psi^{-1}\Lambda\mathbf{f}+\mathbf{f}$$

$$\Rightarrow \Lambda^t\Psi^{-1}\mathbf{x} = \left(\mathbf{I}+\Lambda^t\Psi^{-1}\Lambda\right)\mathbf{f}$$

$$\Rightarrow \mathbf{f}^* = \left(\mathbf{I}+\Lambda^t\Psi^{-1}\Lambda\right)^{-1}\Lambda^t\Psi^{-1}\mathbf{x}.$$

If we evaluate the expected value, we obtain the following

$$\mathrm{E}(\mathbf{f}^*|\mathbf{f}) = \mathrm{E}\left(\left(\mathbf{I}+\Lambda^t\Psi^{-1}\Lambda\right)^{-1}\Lambda^t\Psi^{-1}\mathbf{x}\right)$$

$$\Rightarrow \mathrm{E}(\mathbf{f}^*|\mathbf{f}) = \left(\mathbf{I}+\Lambda^t\Psi^{-1}\Lambda\right)^{-1}\Lambda^t\Psi^{-1}\mathrm{E}(\mathbf{x})$$

$$\Rightarrow \mathrm{E}(\mathbf{f}^*|\mathbf{f}) = \left(\mathbf{I}+\Lambda^t\Psi^{-1}\Lambda\right)^{-1}\Lambda^t\Psi^{-1}\Lambda\mathbf{f} \text{ since } \mathrm{E}(\mathbf{x}) = \Lambda\mathbf{f}.$$

Similarly, we evaluate the predicted error to obtain

$$\mathrm{E}\left((\mathbf{f}^*-\mathbf{f})(\mathbf{f}^*-\mathbf{f})^t\right) = \mathrm{E}\left(\mathbf{f}^*\mathbf{f}^{*t}\right)\text{-}\mathrm{E}\left(\mathbf{f}^*\mathbf{f}^t\right)\text{-}\mathrm{E}\left(\mathbf{f}\mathbf{f}^{*t}\right)+\mathrm{E}\left(\mathbf{f}\mathbf{f}^t\right)$$

Now $\mathrm{E}(\mathbf{f}\mathbf{f}^t) = \mathbf{I}$.

Let $\mathbf{f}^* = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$ and $\mathbf{f} = \begin{pmatrix} d \\ e \\ f \end{pmatrix}$, therefore $\mathbf{f}^*\mathbf{f}^t = \begin{pmatrix} a \\ b \\ c \end{pmatrix}\begin{pmatrix} d & e & f \end{pmatrix} = \begin{pmatrix} ad & ae & af \\ bd & be & bf \\ cd & ce & cf \end{pmatrix}$

Also, $\mathbf{ff}^{*t} = \begin{pmatrix} d \\ e \\ f \end{pmatrix} \begin{pmatrix} a & b & c \end{pmatrix} = \begin{pmatrix} da & ea & fa \\ db & eb & fb \\ dc & ec & fc \end{pmatrix} = \left(\mathbf{f}^{*}\mathbf{f}^{t}\right)^{t}$.

Hence $E\left(\mathbf{ff}^{*t}\right) = \left\{E\left(\mathbf{f}^{*}\mathbf{f}^{t}\right)\right\}^{t}$.

Now, we require to show that $E\left(\mathbf{f}^{*}\mathbf{f}^{*t}\right) = E\left(\mathbf{f}^{*}\mathbf{f}^{t}\right)$. Note that $1 - \frac{1}{1+x} = \frac{x}{1+x} \Rightarrow (1+x)\left(1 - \frac{1}{1+x}\right) = x$.

Let $x = \Lambda^{t}\Psi^{-1}\Lambda$.

$$\therefore \left(I + \Lambda^{t}\Psi^{-1}\Lambda\right)\left[I - \left(I + \Lambda^{t}\Psi^{-1}\Lambda\right)^{-1}\right] = \Lambda^{t}\Psi^{-1}\Lambda$$

So $\left[I - \left(I + \Lambda^{t}\Psi^{-1}\Lambda\right)^{-1}\right] = \left(I + \Lambda^{t}\Psi^{-1}\Lambda\right)^{-1}\Lambda^{t}\Psi^{-1}\Lambda$.

Now, $E\left(\mathbf{f}^{*}\mathbf{f}^{t}\right) = E\left[\left(I + \Lambda^{t}\Psi^{-1}\Lambda\right)^{-1}\Lambda^{t}\Psi^{-1}\mathbf{xf}^{t}\right] = \left(I + \Lambda^{t}\Psi^{-1}\Lambda\right)^{-1}\Lambda^{t}\Psi^{-1}E\left(\mathbf{xf}^{t}\right)$

but $\mathrm{Cov}\left(\mathbf{xf}^{t}\right) = E\left(\mathbf{xf}^{t}\right) - E(\mathbf{x})E\left(\mathbf{f}^{t}\right) = \Lambda$ since $E\left(\mathbf{f}^{t}\right) = 0$ so $E\left(\mathbf{xf}^{t}\right) = \Lambda$.

$$\therefore E\left(\mathbf{f}^{*}\mathbf{f}^{t}\right) = \left(I + \Lambda^{t}\Psi^{-1}\Lambda\right)^{-1}\Lambda^{t}\Psi^{-1}\Lambda$$

Now we consider $E\left(\mathbf{f}^{*}\mathbf{f}^{*t}\right)$

$$= E\left[\left(I + \Lambda^{t}\Psi^{-1}\Lambda\right)^{-1}\Lambda^{t}\Psi^{-1}\mathbf{x}\left(\left(I + \Lambda^{t}\Psi^{-1}\Lambda\right)^{-1}\Lambda^{t}\Psi^{-1}\mathbf{x}\right)^{t}\right]$$

$$= E\left[\left(I + \Lambda^{t}\Psi^{-1}\Lambda\right)^{-1}\Lambda^{t}\Psi^{-1}\mathbf{xx}^{t}\Psi^{-1}\Lambda\left(I + \Lambda^{t}\Psi^{-1}\Lambda\right)^{-1}\right]$$

$$= \left(I + \Lambda^{t}\Psi^{\ 1}\Lambda\right)^{-1}\Lambda^{t}\Psi^{\ 1}E\left(\mathbf{xx}^{t}\right)\Psi^{\ 1}\Lambda\left(I + \Lambda^{t}\Psi^{\ 1}\Lambda\right)^{-1}$$

$$= \left(I + \Lambda^{t}\Psi^{-1}\Lambda\right)^{-1}\Lambda^{t}\Psi^{-1}\Sigma\Psi^{-1}\Lambda\left(I + \Lambda^{t}\Psi^{-1}\Lambda\right)^{-1}$$

$$= \left(I + \Lambda^{t}\Psi^{-1}\Lambda\right)^{-1}\Lambda^{t}\Psi^{-1}\left(\Psi + \Lambda\Lambda^{t}\right)\Psi^{-1}\Lambda\left(I + \Lambda^{t}\Psi^{-1}\Lambda\right)^{-1}$$

$$= \left(I + \Lambda^{t}\Psi^{-1}\Lambda\right)^{-1}\left(\Lambda^{t} + \Lambda^{t}\Psi^{-1}\Lambda\Lambda^{t}\right)\Psi^{-1}\Lambda\left(I + \Lambda^{t}\Psi^{-1}\Lambda\right)^{-1}$$

$$= \left\{\left(I + \Lambda^{t}\Psi^{-1}\Lambda\right)^{-1}\Lambda^{t} + \left(I + \Lambda^{t}\Psi^{-1}\Lambda\right)^{-1}\Lambda^{t}\Psi^{-1}\Lambda\Lambda^{t}\right\}\left(\Psi^{-1}\Lambda\left(I + \Lambda^{t}\Psi^{-1}\Lambda\right)^{-1}\right).$$

But $\left(\mathrm{I}+\Lambda^t\Psi^{-1}\Lambda\right)^{-1}\Lambda^t\Psi^{-1}\Lambda=\mathrm{I}-\left(\mathrm{I}+\Lambda^t\Psi^{-1}\Lambda\right)^{-1}$.

So $\mathrm{E}\left(\mathbf{f}^*\mathbf{f}^{*t}\right)=\left\{\left(\mathrm{I}+\Lambda^t\Psi^{-1}\Lambda\right)^{-1}\Lambda^t+\left(\mathrm{I}-\left(\mathrm{I}+\Lambda^t\Psi^{-1}\Lambda\right)^{-1}\right)\Lambda^t\right\}\left(\Psi^{-1}\Lambda\left(\mathrm{I}+\Lambda^t\Psi^{-1}\Lambda\right)^{-1}\right)$

$$=\left(\mathrm{I}+\Lambda^t\Psi^{-1}\Lambda\right)^{-1}\Lambda^t\Psi^{-1}\Lambda\left(\mathrm{I}+\Lambda^t\Psi^{-1}\Lambda\right)^{-1}+\Lambda^t\Psi^{-1}\Lambda\left(\mathrm{I}+\Lambda^t\Psi^{-1}\Lambda\right)^{-1}$$
$$-\left(\mathrm{I}+\Lambda^t\Psi^{-1}\Lambda\right)^{-1}\Lambda^t\Psi^{-1}\Lambda\left(\mathrm{I}+\Lambda^t\Psi^{-1}\Lambda\right)^{-1}$$
$$=\left(\mathrm{I}+\Lambda^t\Psi^{-1}\Lambda\right)^{-1}\Lambda^t\Psi^{-1}\Lambda\left(\left(\mathrm{I}+\Lambda^t\Psi^{-1}\Lambda\right)^{-1}+\mathrm{I}-\left(\mathrm{I}+\Lambda^t\Psi^{-1}\Lambda\right)^{-1}\right)$$
$$=\left(\mathrm{I}+\Lambda^t\Psi^{-1}\Lambda\right)^{-1}\Lambda^t\Psi^{-1}\Lambda.$$

But $\left(\mathrm{I}+\Lambda^t\Psi^{-1}\Lambda\right)^{-1}\Lambda^t\Psi^{-1}\Lambda=\mathrm{I}-\left(\mathrm{I}+\Lambda^t\Psi^{-1}\Lambda\right)^{-1}$.

$$\therefore \mathrm{E}\left(\left(\mathbf{f}^*-\mathbf{f}\right)\left(\mathbf{f}^*-\mathbf{f}\right)^t\right)=\mathrm{E}\left(\mathbf{f}^*\mathbf{f}^{*t}\right)-\mathrm{E}\left(\mathbf{f}^*\mathbf{f}^t\right)-\mathrm{E}\left(\mathbf{f}\mathbf{f}^{*t}\right)+\mathrm{E}\left(\mathbf{f}\mathbf{f}^t\right)$$
$$=\left[\mathrm{I}-\left(\mathrm{I}+\Lambda^t\Psi^{-1}\Lambda\right)^{-1}\right]-\left[\mathrm{I}-\left(\mathrm{I}+\Lambda^t\Psi^{-1}\Lambda\right)^{-1}\right]-\left[\mathrm{I}-\left(\mathrm{I}+\Lambda^t\Psi^{-1}\Lambda\right)^{-1}\right]+\mathrm{I}$$
$$=-\mathrm{I}+\left(\mathrm{I}+\Lambda^t\Psi^{-1}\Lambda\right)^{-1}+\mathrm{I}$$
$$=\left(\mathrm{I}+\Lambda^t\Psi^{-1}\Lambda\right)^{-1}.$$

Hence, $\mathrm{E}\left(\left(\mathbf{f}^*-\mathbf{f}\right)\left(\mathbf{f}^*-\mathbf{f}\right)^t\right)=\left(\mathrm{I}+\Lambda^t\Psi^{-1}\Lambda\right)^{-1}$

□

Now, if we compare the expected value of these two methods:

Bartlett's: $\mathrm{E}\left(\hat{\mathbf{f}}\,|\,\mathbf{f}\right)=\mathbf{f}$

Thomson's: $\mathrm{E}\left(\mathbf{f}^*\,|\,\mathbf{f}\right)=\left(\mathrm{I}+\Lambda^t\Psi^{-1}\Lambda\right)^{-1}\Lambda^t\Psi^{-1}\Lambda\mathbf{f}$

we see that the Bartlett's score is an unbiased estimate of **f** while the Thompson's score is biased.

In addition, for the predicted errors we have the following formulae:

Bartlett's: $$E\left(\left(\hat{\mathbf{f}}-\mathbf{f}\right)\left(\hat{\mathbf{f}}-\mathbf{f}\right)^{\mathbf{t}}\right)=\left(\Lambda^{t}\Psi\Lambda\right)^{-1}$$

Thomson's: $$E\left(\left(\mathbf{f}^{*}-\mathbf{f}\right)\left(\mathbf{f}^{*}-\mathbf{f}\right)^{\mathbf{t}}\right)=\left(\mathbf{I}+\Lambda^{t}\Psi^{-1}\Lambda\right)^{-1}$$

Having computed and thus comparing the predicted errors for both Bartlett's and Thomson's scores, it is clear that the predicted error is smaller in value for Thomson's score. Therefore, we can conclude that Thomson's score is more accurate.

Now, if the columns of Λ satisfy the condition that $\Lambda^{t}\Psi^{-1}\Lambda$ is diagonal, then the components of both factor scores are uncorrelated with one another. In addition, if the eigenvalues of $\Lambda^{t}\Psi^{-1}\Lambda$ are large, then the predicted errors are small implying that the two factor scores are alike.

4 TOURISM CASE STUDY

4.1 Purpose and Aim of Survey

The traveller survey is a questionnaire in which the Malta Tourism Authority (MTA) gathers information about the tourist's visit to the Maltese Islands. Its main aim is to obtain a wide knowledge of the tourist profile and expenditure information by place of residence. In addition, it is an important source for market research, because it gives essential information for the MTA to plan its strategy and take action to improve the physical environment and the service provided in the Maltese Islands.

The traveller survey was first launched in the early 90's and the information gathered was social demographic. Its focus was on marketing reasons. The new version of the British traveller survey was launched in the beginning of 2003 and was implemented during the summer season of that year. The initiation of the other markets questionnaires took place in the beginning of 2004. This new edition amplified the marketing, promotional and human resources sections. Furthermore, more questions were added to the surveys to obtain a wider knowledge of the tourist perspective. The main aim was to group ideas to reduce the cost of research whilst at the same time amplifying the data obtained. The survey is distributed randomly - every fifth tourist checking in is given a questionnaire in the security area B of the arrivals at the Malta International Airport. In the case when there are few arrivals, a questionnaire is distributed to each tourist. This random distribution of sampling is known as systematic sampling where units are selected from the population at a regular interval. For the British market, around ten to fifteen questionnaires are distributed to each flight. The

Malta Tourism Authority distributes around twenty thousand surveys in a year that is around five thousand per quarter. From these five thousand they receive five hundred. Hence, the response rate is of 12%, which is quite good when compared to the response rate of 10% of the general mail questionnaires.

4.2 Details about the Questionnaire

In this dissertation, I based my study on the British market only, which till now is the highest tourism market for the Maltese islands. I gathered the data of four different seasons. Since this new survey format was launched in Summer of 2003, data for the whole of 2003 was obviously not available. Therefore, so as to cover a one-year span, I considered data ranging from Summer 2003 till Spring 2004.

Various difficulties were encountered during this process. The main problem was that each season had its own codebook. Therefore, for the sake of continuity, I had to build up a new codebook and recode a lot of the information into a new format. In addition, there were some variations in the inputting procedures used for the different seasons and this had to be taken into consideration too. All this was rather time consuming though certainly necessary as this was the first time that an analysis of traveller survey data was to be carried out on annual data. Usually reports about this questionnaire are published by the MTA each quarter.

From the twenty-four questions, I focused my research on question number twelve and on the first two questions. In question twelve, tourists are asked to rate various aspects of their trip in Malta in terms of physical environment and about the service provided by employees. This type of question is known as Likert scale. There are twenty-five aspects or variables grouped under physical environment and seventeen variables under service provided. The main focus of this question is to point out the strong and weak points of the Maltese tourism industry.

The physical environment aspects are divided into three sections. The first section covers a general area of the environment because tourists are asked to rate the Malta International Airport, restaurants, taxi service, beaches etc. In other words, this first section contains a group of different variables that are only related in the sense that they all serve the tourism industry. The second part focuses on attractions of Malta, such as guided tours, museums and historical sites. The third and last part concentrates on the infrastructure such as roads, road signs, level of cleanliness and air quality.

The question on service provided covers the same variables as outlined for the physical environment question. Obviously, for the infrastructure, no employees are involved. Therefore, the service provided by employees is clustered into two sections.

A Likert scale, as the name implies, requires some sort of scale and in our case, this is provided in the form of six columns – the first five columns are numbered from one to five and the last one is labelled as N/A. The following table defines what the scale represents

1	2	3	4	5	N/A
Very Good	Good	Not So Good	Poor	Very Poor	Not Applicable

Table 1: The Likert Scale.

This scale is applicable to both the variables of the physical environment and of the service provided.

The first two questions of the questionnaire ask for personal information. In fact, the tourist is asked where he lives and other personal information such as gender and marital status. These two questions helped me to bring out a general outline of the tourist. This outline is called the profile of the tourist. View **Appendix B** for these questions.

4.3 Profile of Respondents

The socio and demographic aspects compose the profile of the tourist. The majority of the respondents (28%) visited the Maltese Islands during the autumn season and the least amount of respondents amounting to 479 (19%) visited during the winter season. Most of the tourists (85%) resided in England, followed by 8% from Scotland.

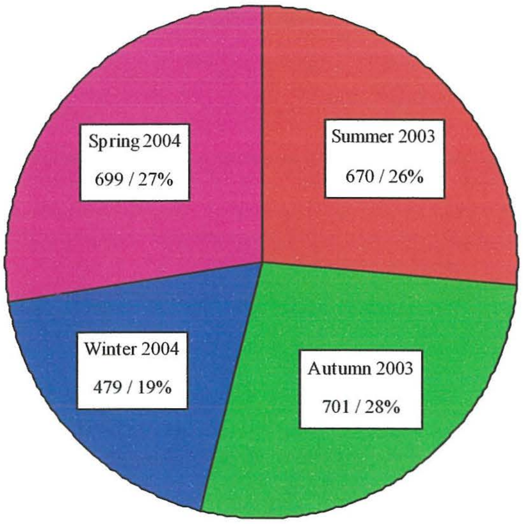

Figure 1: Pie chart showing seasons

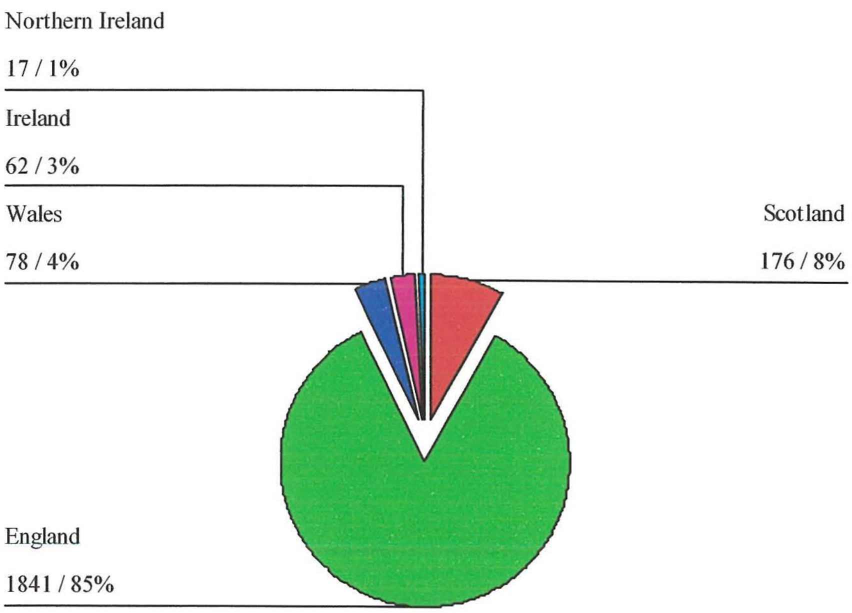

Figure 2: Pie chart showing countries of origin

The highest frequency of respondents (7%) that lived in England came from the north / west or south Yorkshire followed by a 5% form the Lancashire and Kent regions. From the Scottish areas, the highest frequency was of 2% and originated from the Tayside and Strathclyde regions.

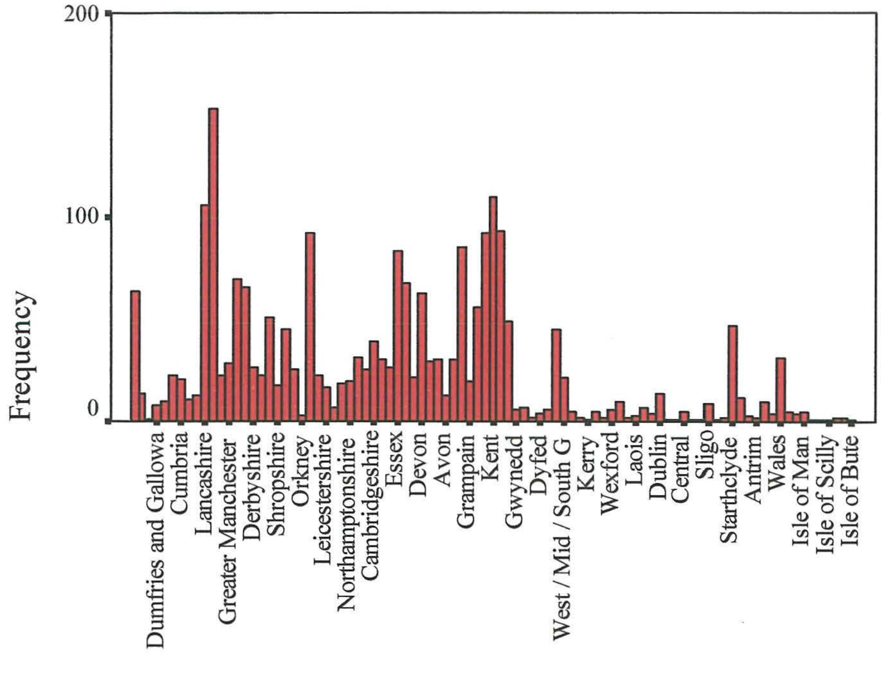

Figure 3: Bar chart of frequency against region

Male respondents amounted to 51 % while female respondents accounted for 49%. As to marital status, the preponderance of respondents were married or cohabiting (80%). Single respondents amounted to 10%, 5% were divorced/separated, and 5% were widowed.

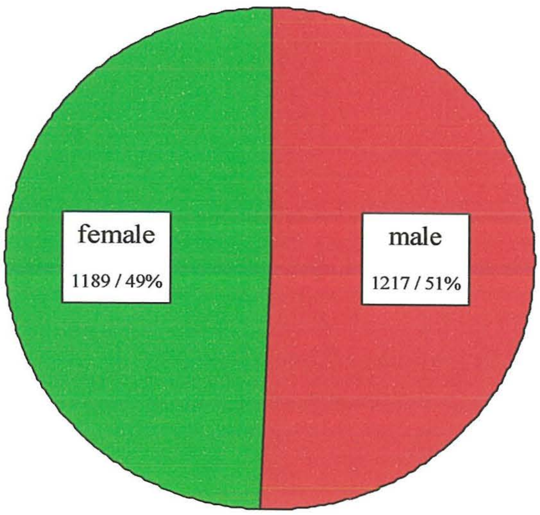

Figure 4: Pie chart showing gender

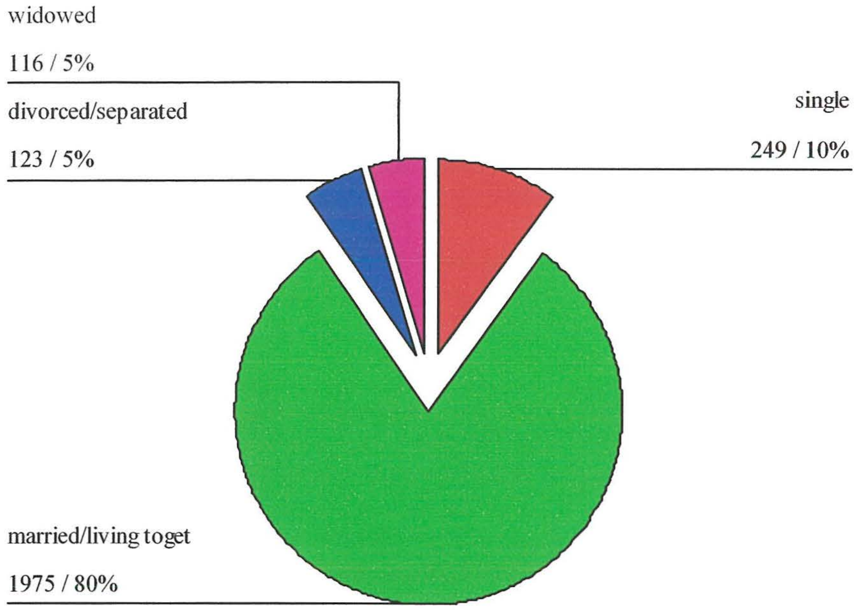

Figure 5: Pie chart showing marital status

The majority of the respondents (34%) belong to the age bracket 56 - 65 years. 23% and 19% in the age brackets 46 - 55 and 66+ respectively followed this main group. The remaining age groups amounted up to 24%.

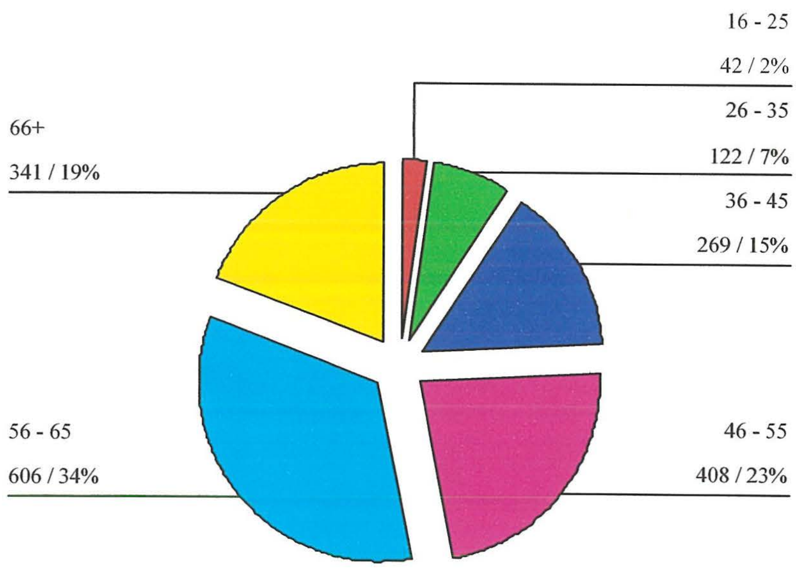

Figure 6: Pie chart showing age ranges

Retired respondents, who had a frequency of 781 (34.7%), formed the largest group while 22.5% had a professional position as their full-time occupation. The third largest segment (10.1%) was made up of respondents in managerial positions. Office/retail workers amounted to 8.2% and skilled workers/tradesmen to 4.6%.

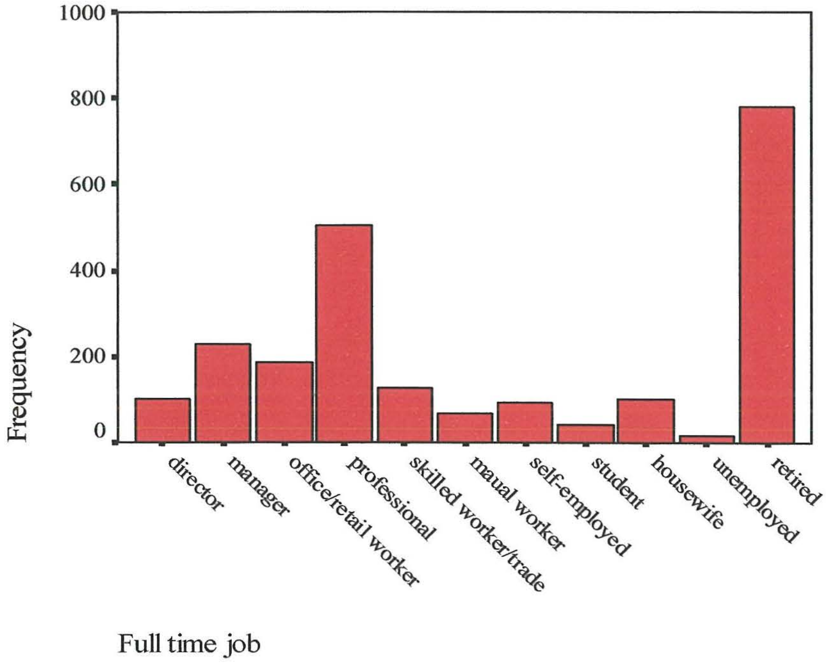

Figure 7: Bar chart of frequency against full time job

Of the total 2549 respondents, 436 said that they did part-time work. The majority of these (34.9%) worked as professionals. Another 22.7% did office or retail work whilst 7.8% were self-employed.

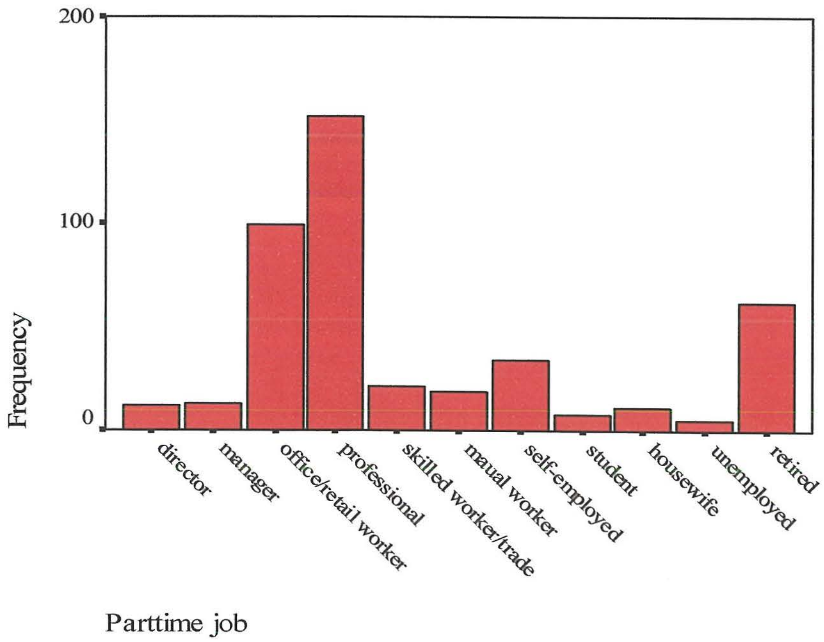

Figure 8: Bar chart of frequency against part time job

66.3% of respondents earned between £0 and £1000 per month whilst another 21.4% had an income within the range of £1001 and £2000. Respondents earning more than £2000 totalled 12.3%. Overall, respondents' mean income per month stood at £895.72.

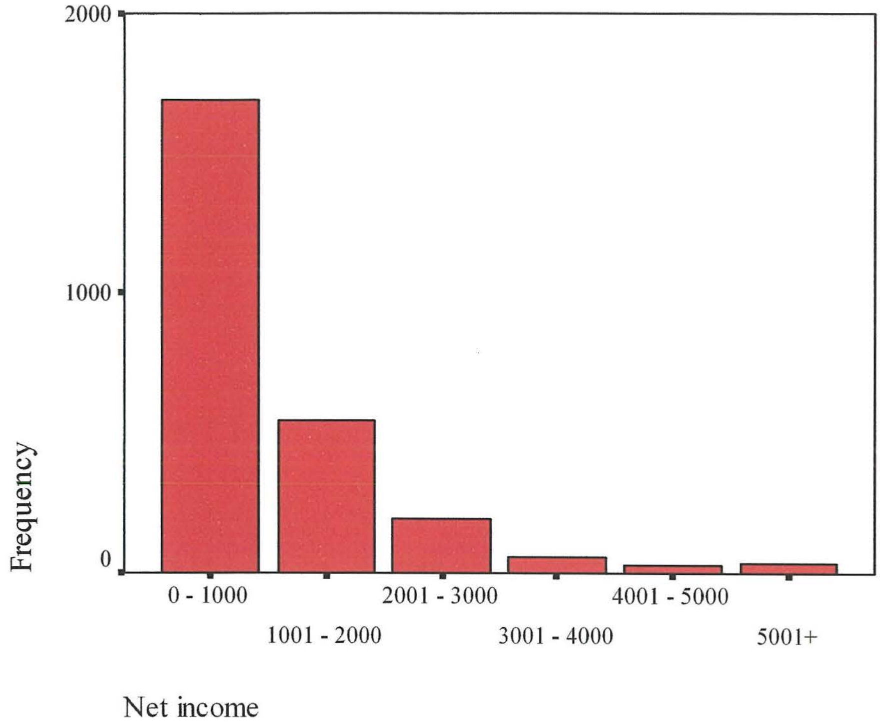

Figure 9: Bar chart of frequency against net income

Now, we compare where possible these profile variables with the United Kingdom demographics. The UK data was obtained from the website of the official UK statistics and is based on the 2001 census.

Here, we can observe variables that feature both in the traveller survey as well as in the UK official statistics. The variables are gender, age and marital status. For the other variables, either the data was not available or it was structured differently from ours. The following are the tables and graphs of these three variables.

Age Range	Total	Males	Females
0 -24	18314618	9323786	8990832
25 - 34	8360547	4095236	4265311
35 - 44	8777390	4334223	4443167
45 - 54	7776562	3854549	3922013
55 - 64	6219078	3061080	3157998
65+	9340999	3910995	5430004
Total	58789194	28579869	30209325

Table 2: Gender total grouped according to age range.

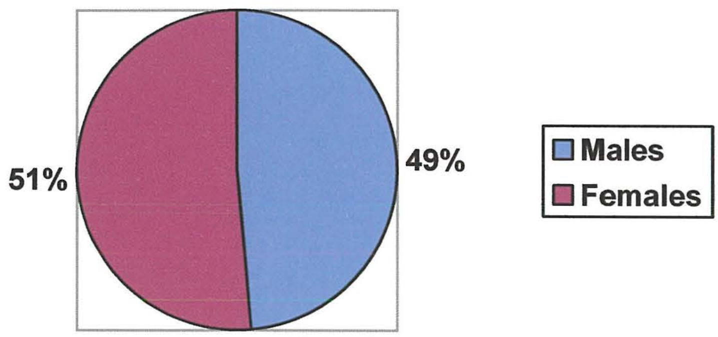

Figure 10: Pie chart showing gender.

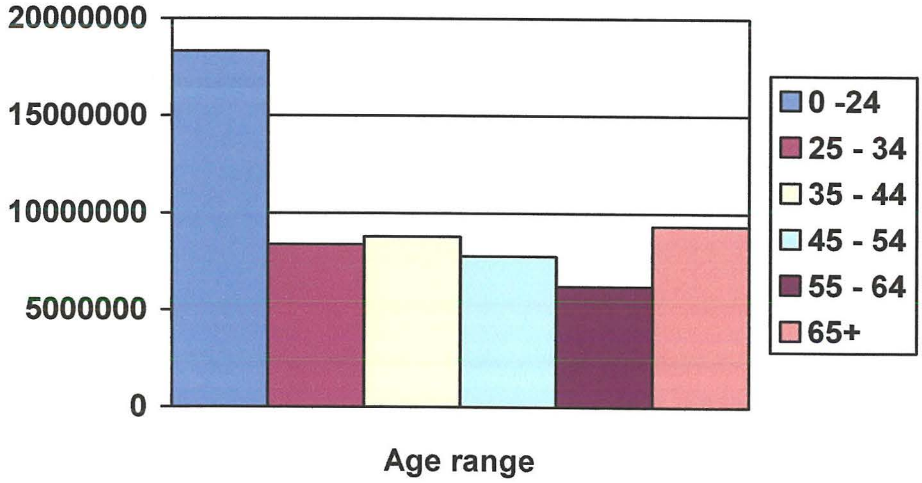

Figure 11: Bar chart of frequency against age range.

Single people (never married)	14186988
Married or re-married people	23853128
Separated or divorced	4942512
Widowed	3947709

Table 3: Marital status.

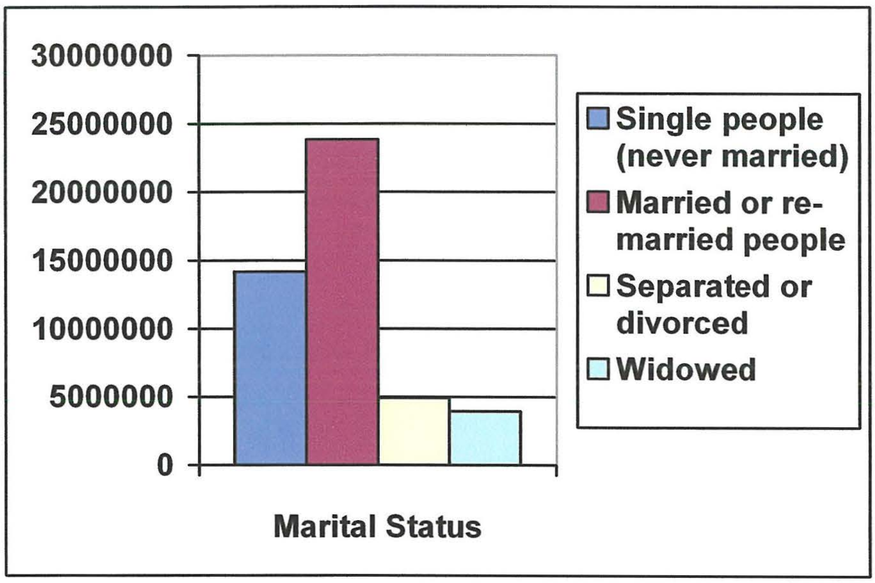

Figure 12: Bar chart showing frequency against marital status.

Comparing these plots, we can conclude that where gender and age are concerned, our survey does not really describe the population of United Kingdom. On the other hand, the pattern of our respondents' marital status is parallel to that of the UK population.

Focussing our attention to the traveller survey once again, the following tables show the variables of the two aspects, physical environment and service provided by employees, on which we are performing our analysis. Together with these variables, we are going to display the percentages of each of the points present in the Likert scale, where 0 signifies no answer.

Physical Environment	Percentages obtained.						
Variable	0	1	2	3	4	5	6
MIA	0.1	30.9	53.4	11.7	1.1	0.4	2.4
Gozo heliport	49.9	1.6	2.4	0.7	0.4	0.1	44.9
Gozo ferry points	32.4	10.9	21.4	10.2	2.5	0.9	21.7
Accommodation	4.1	35.9	38.1	14.1	3.8	2.9	1.1
Restaurants	13.9	22.3	42.8	12	1.8	0.5	6.6
Entertainment	30.7	5.5	18	14.9	5.3	2.9	22.8
Car hire	38.4	6.9	10.4	5.8	2.3	1.6	34.6
Taxi	37.5	8.3	12	5.3	1.3	1.5	34.3
Public transport	13.2	24.8	31.4	11.8	4.9	2.7	11.1
Retail outlets	20.2	8.7	39.3	20.6	4.6	1.5	5.3
Sports facilities	44.8	2	6	4	1.5	0.6	41.1
Beaches	23.5	8.1	20	17.9	6.9	4.2	19.5
Guided tours	28	16.2	21.7	5.8	1.6	0.8	25.9
Historical sites	18.4	25.6	36.4	9.1	2.2	0.5	7.8
Museums	24.8	19.6	28.1	7.6	1.4	0.4	18.2
Cathedrals / Churches	15.3	40.2	30.1	4.8	0.4	0.4	8.9
Theatre / Performing arts	47.6	2.6	3.3	1.1	0.3	0.2	44.9
Road	6.4	0.9	7.3	24	27.5	30.7	3.2
Road signs	12.5	3.9	20.9	23.9	16.5	4.2	8
Traffic	12.1	1.7	13.6	31.7	19.4	17.1	4.4
Parking	23.9	2.2	9.7	15.1	11.3	12.4	25.4
Public conveniences	9.7	6.5	25.5	22.9	15.6	12.1	7.7
Level of cleanliness	6.1	7.9	27.9	26.6	16.6	13.3	1.6
Air quality	7.5	21.2	42.3	18.3	6.2	3.4	1.2
Sea quality	12.3	26.3	33.9	11.2	2.8	1.1	12.4

Table 4: Percentages of the physical environment variables.

Service Provided	Percentages obtained.						
Variable	0	1	2	3	4	5	6
MIA	10.6	26.4	44.7	12.9	2.7	1.3	1.3
Gozo heliport	52.5	1.2	2.2	0.9	0.4	0.1	52.7
Gozo ferry points	38.1	9.9	20.5	7.8	2	0.7	21.1
Accommodation	12.9	38.6	31.5	9.7	3.3	2	2
Restaurants	22.2	25.2	34.1	9.8	1.5	0.7	6.5
Entertainment	38.8	5.3	15.9	10.6	3.3	2.3	23.8
Car hire	42.3	8	9.4	4.7	1.3	1.2	33
Taxi	41.3	9	10.5	4.4	0.8	26.1	32.6
Public transport	21.9	22.5	29.9	9.5	3.3	1.9	11.2
Retail outlets	28.1	12	35.3	15.5	2.6	0.9	5.6
Sports facilities	48.1	2.2	4.9	3.2	0.8	0.5	40.3
Beaches	36.5	6.2	14	11.7	3.9	2.4	25.3
Guided tours	36	16.7	16.8	4.6	1.6	0.9	23.5
Historical sites	30	19.2	27.9	10.2	2.2	1.1	9.5
Museums	34.6	15.8	21.9	7.8	1.4	0.6	17.9
Cathedrals / Churches	28.3	25.2	24.7	8.4	1	0.6	11.9
Theatre / Performing arts	51.3	2.3	2.7	0.9	0.1	0.2	42.6

Table 5: Percentages of the service provided by employees variables.

4.4 General Outline of the Procedure

Considering the above-mentioned variables, we perform two methods of data reduction: factor analysis and parceling. Also, within these two methods, we have applied two extraction methods. The reason of applying two methods is to compare the traditional system of tourism analysis and the statistical improvement through parceling. But before we perform these analyses, we had to reduce our data due to the presence of a large amount of missing data. These missing figures were composed of the values 0 and 6. After this reduction, we performed the analysis of the variables with both methods and extractions to end up with sets of values known as factor scores.

These factor scores, which are values that signify the influence of that particular variable over the set of all variables, were transformed so that they do not depart from the normal behaviour. Then these normal distributed scores were linearly related with our socio-demographic variables, which form our tourist profile. The necessity of normality is required so that we can apply linear models. The scope is to observe whether there were profile variables that were related to our extracted scores.

5 FACTOR ANALYSIS APPLICATION

5.1 Introduction

The aim of this section is to apply techniques discussed earlier to the data collected by the Strategic Planning and Research Division within the Malta Tourism Authority. In this chapter, we will focus on the first part of the applicative method since here we will discuss the steps required to end up with a set of factor scores. These steps form the basis of exploratory factor analysis and will lead us to the result where we have a set of significant variables on which we base our study. This forms confirmatory factor analysis after which we continue to extract the factor scores. Hence, our exploratory factor analysis steps begins by first tackling the problem of which variables we will consider for the analysis due to the fact of a large amount of missing data. Then, we will apply our two data reduction techniques in which we are applying two extraction methods. From these we end up with a set of significant variables. Here we finish exploratory and merge into confirmatory factor analysis. Then we extract the factor scores which will be discussed in the next chapter.

5.2 Exploratory Factor Analysis

5.2.1 Introduction

The primary objectives of exploratory factor analysis are to determine the number of common factors influencing a set of measures and to determine the strength of the relationship between each factor and each observed measure.

Exploratory factor analysis is applied in various applications, for example, when we want to determine which features are most important when classifying a group of items. In addition, it is useful when we want to determine which set of items can be clustered together in a questionnaire or to demonstrate the dimensionality of a measurement scale. Another application is when we want to generate factor scores that represent the values of the underlying constructs for use in other analyses.

5.2.2 Reduction of Variables and Respondents

First, we had to reduce the number of variables because as we can notice from the table below there are eleven variables highlighted in bold for which the number of valid cases (N) is very near to the number of missing values. In fact, we only considered those variables that have a percentage of valid cases higher than 73%. In addition, to this variable reduction, we also viewed the amount of missing values that each respondent had and removed those respondents that have more then four missing values. These missing values are the numbers 0 and 6 which they respectively represent no answer and not applicable. For the case of the physical environment, we have reduced the data from 2549 to 2191 valid cases while for the service provided; we reduced the data to 2240 from 2549 valid cases. For both aspects of the question, physical environment and service provided by employees, the removed variables were the same except for the *parking* variable, since the variables falling under the infrastructure section are not present in the service provided by employees' aspect.

	Cases			
	Valid		Missing	
	N	Percent	N	Percent
MIA	2488	97.6%	61	2.4%
Gozo heliport	**132**	**5.2%**	**2417**	**94.8%**
Gozo ferry points	**1170**	**45.9%**	**1379**	**54.1%**
Accommodation	2416	94.8%	133	5.2%
Restaurants	2027	79.5%	522	20.5%
Entertainment	**1187**	**46.6%**	**1362**	**53.4%**
Car hire	**688**	**27.0%**	**1861**	**73.0%**
Taxi service	**720**	**28.2%**	**1829**	**71.8%**
Public transport	1929	75.7%	620	24.3%
Retail outlets	1900	74.5%	649	25.5%
Sports facilities	**360**	**14.1%**	**2189**	**85.9%**
Beaches	**1453**	**57.0%**	**1096**	**43.0%**
Guided tours	**1177**	**46.2%**	**1372**	**53.8%**
Historical Sites	1881	73.8%	668	26.2%
Museums	**1455**	**57.1%**	**1094**	**42.9%**
Cathedrals / Churches	1933	75.8%	616	24.2%
Theatre / Performing arts	**190**	**7.5%**	**2359**	**92.5%**
Road	2306	90.5%	243	9.5%
Road signs	2026	79.5%	523	20.5%
Traffic	2129	83.5%	420	16.5%
Parking	**1292**	**50.7%**	**1257**	**49.3%**
Public conveniences	2105	82.6%	444	17.4%
Level of cleanliness	2352	92.3%	197	7.7%
Air quality	2328	91.3%	221	8.7%
Sea quality	1920	75.3%	629	24.7%

Table 6: The value (N) and percentage of valid and missing cases of the variables.

5.2.3 Data Reduction Techniques

5.2.3.1 Introduction

In this section, we are going to discuss the two techniques that we have applied. The first method is the factor analysis and the second method is parceling. Factor analysis is the traditional approach of these surveys while parceling is a statistical improvement of factor analysis. Both methods deal with the process of reducing the data from a large group to sets of smaller groups and in the process unmasking the latent factors. Within these two techniques, we applied two extraction methods, maximum likelihood and principal axis factoring. We have chosen these two methods so that we have two different approaches where maximum likelihood assumes the data to be normally distributed while principal axis factoring does not require the data to satisfy this distribution.

Each of these extractions was applied to both the physical environment and to the service provided by employees' variables. Note also that in this process, we have applied a varimax rotation so that it would be more possible to give an interpretation of the factor matrix.

Now we will start each technique with both extraction methods for both aspects. Hence, our first method is factor analysis on the physical environment variables.

5.2.3.2 Method 1: Factor Analysis

(i): Physical Environment

The step involved here is applying factor analysis with a maximum likelihood and principal axis factoring extraction to the physical environment variables. In the analysis, each missing value is replaced with the variable mean. For both extraction methods, we obtained the same following descriptive statistics.

	Mean	Std. Deviation	Analysis N	Missing N
MIA	1.85	.710	2191	18
Accommodation	1.96	.966	2191	57
Restaurants	1.94	.703	2191	351
Public transport	2.08	.917	2191	449
Retail outlets	2.35	.750	2191	395
Historical Sites	1.87	.723	2191	412
Cathedrals / Churches	1.57	.622	2191	397
Road	3.89	.985	2191	67
Road signs	3.21	1.078	2191	272
Traffic	3.43	1.008	2191	179
Public conveniences	3.02	1.117	2191	229
Level of cleanliness	3.00	1.169	2191	52
Air quality	2.22	.985	2191	59
Sea quality	1.92	.792	2191	404

Table 7: The descriptive statistics of the physical environment variables.

Observing the mean, we can notice that *Road* and *Cathedrals / Churches,* written in italics, are the most influential variables rated by the tourists. Since *Road* has the highest mean of 3.89 it implies that the tourists rate the roads very badly. On the other hand, *Cathedrals / Churches* have the lowest mean value of 1.57, which signifies that the tourists are very interested in the architectural richness of our temples.

Now, let us consider the results obtained for both extraction methods. From the descriptive statistics table, we focus our attention on the total variance explained table. For both extraction methods, we have obtained an extraction of four factors, which describe around 56% of the data. From the table we take note of those factors that have an eigenvalue higher than one while the rest are irrelevant since their eigenvalue does not satisfy our criterion. The ones marked in red are the significant factors since they have an eigenvalue greater than one.

	Initial Eigenvalues		
Factor	Total	% of Variance	Cumulative %
1	4.106	29.327	29.327
2	1.457	10.409	39.736
3	1.180	8.429	48.165
4	1.031	7.367	55.532
5	.859	6.134	61.666
6	.852	6.084	67.750
7	.753	5.381	73.131
8	.738	5.272	78.404
9	.603	4.309	82.712
10	.548	3.912	86.625
11	.522	3.731	90.355
12	.494	3.532	93.887
13	.467	3.334	97.221
14	.389	2.779	100.000

Table 8: The total variance explained of the factors using maximum likelihood.

	Initial Eigenvalues		
Factor	Total	% of Variance	Cumulative %
1	4.106	29.327	29.327
2	1.457	10.409	39.736
3	1.180	8.429	48.165
4	1.031	7.367	55.532
5	.859	6.134	61.666
6	.852	6.084	67.750
7	.753	5.381	73.131
8	.738	5.272	78.404
9	.603	4.309	82.712
10	.548	3.912	86.625
11	.522	3.731	90.355
12	.494	3.532	93.887
13	.467	3.334	97.221
14	.389	2.779	100.000

Table 9: Total variance explained of the factors using principal axis factoring.

Now, we have the scree plot, which is a graph of the eigenvalues against all the factors. This helps us to determine the number of factors to retain. Our point of interest is where the curve starts to flatten. This occurs between two factors where the smallest one is considered significant while the other is not. For our case, this occurs between the fourth and fifth factor. This indicates that we can retain four significant factors, which confirms the same decision made earlier. For both extraction methods the scree plot is similar.

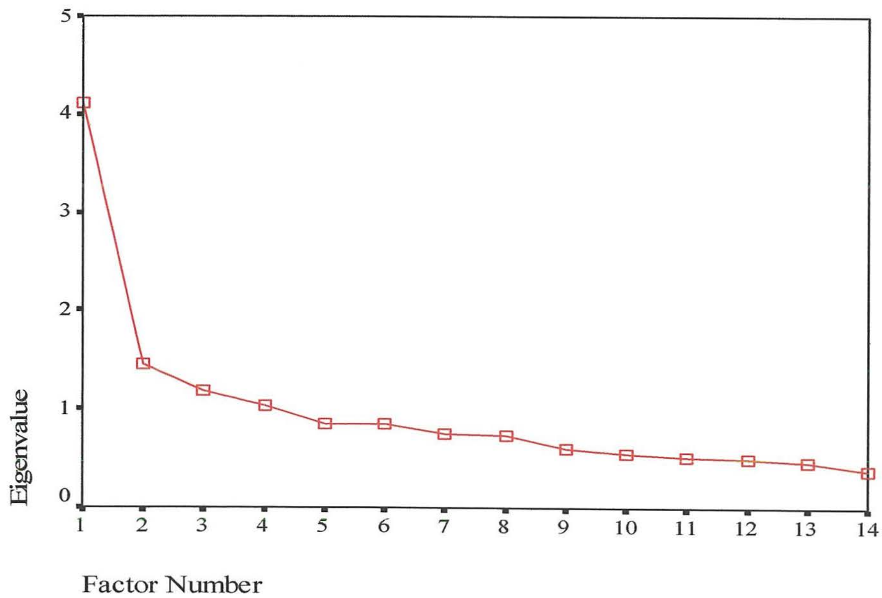

Figure 13: The scree plot of the eigenvalue against the factor number.

Next, we focus on the following table, which is the rotated factor matrix and the concept behind this matrix is to reduce the number of factors on which the variables under investigation have high loadings. In other words, it helps us obtain a better interpretation of the analysis.

	Factor			
	1	**2**	**3**	**4**
MIA	.413	.148		.105
Accommodation	.355			.166
Restaurants	.440		.115	.153
Public transport	.347	.214		
Retail outlets	.478	.179		.127
Historical Sites	.614	.128		
Cathedrals / Churches	.588		.131	
Road	.168	.705		.132
Road signs	.197	.628		.103
Traffic	.120	.599	.190	.204
Public conveniences	.155	.190	.114	.672
Level of cleanliness	.225	.245	.249	.665
Air quality	.198	.198	.859	.184
Sea quality	.315		.466	.186

Table 10: The rotated factor matrix obtained by maximum likelihood extraction..

	Factor			
	1	**2**	**3**	**4**
MIA	.428	.145	.117	
Accommodation	.394		.221	
Restaurants	.495		.219	
Public transport	.342	.214		.123
Retail outlets	.501	.176	.141	
Historical Sites	.557	.154		.199
Cathedrals / Churches	.522			.258
Road	.167	.702	.129	
Road signs	.188	.624	.103	.103
Traffic	.107	.604	.213	.186
Public conveniences	.146	.214	.588	.155
Level of cleanliness	.198	.254	.658	.279
Air quality	.188	.206	.255	.641
Sea quality	.254		.170	.602

Table 11: The rotated factor matrix obtained by principal axis factoring extraction..

From the above tables, we note that the variables are colour-coded. This helps us to relate the different variables to each factor or component. For the maximum likelihood extraction, the first factor represents the highest number of variables. These are described under the general and the attractions section of the analysed question. The second, third and fourth factors describe the infrastructure section. The second refers to the road infrastructure, the third to the environment quality and the last to public conveniences and cleanliness.

Now, if we compare the rotated factor table of the principal axis factoring with the maximum likelihood extraction table, we have that the factors represent the same variables except the last two since the third factor describes public conveniences and level of cleanliness while the fourth describes the air and sea quality.

During this whole procedure, we have obtained our factor scores and here we end this process. Hence, from the above tables we have obtained two sets formed by four factor scores. Our extracted factor scores represent or describe the loadings of these variables.

(ii): Service provided by employees

Considering the other aspect - service provided - we applied the same procedure with the only difference being that we now have a lesser amount of variables. In fact, as we can notice the infrastructure section is not present in this table since no employees are involved in this section.

	Mean	Std. Deviation	Analysis N	Missing N
MIA	1.95	.825	2240	83
Accommodation	1.81	.923	2240	120
Restaurants	1.85	.715	2240	445
Public transport	1.99	.837	2240	550
Retail outlets	2.17	.706	2240	557
Historical sites	1.98	.736	2240	702
Cathedrals / churches	1.78	.673	2240	722

Table 12: The descriptive statistics of the service provided by employees variables.

Retail outlets having the highest mean of 2.17 signify that the services are not appreciated much by the tourists while *Churches / Cathedrals* have a mean of 1.78 signifying that the relevant employees run our places of worship efficiently.

	Initial Eigenvalues		
Factor	Total	% of Variance	Cumulative %
1	2.582	36.885	36.885
2	1.049	14.981	51.866
3	.854	12.207	64.073
4	.784	11.198	75.270
5	.679	9.701	84.971
6	.610	8.711	93.682
7	.442	6.318	100.000

Table 13: The total variance explained by the factors using maximum likelihood.

	Initial Eigenvalues		
Factor	Total	% of Variance	Cumulative %
1	2.582	36.885	36.885
2	1.049	14.981	51.866
3	.854	12.207	64.073
4	.784	11.198	75.270
5	.679	9.701	84.971
6	.610	8.711	93.682
7	.442	6.318	100.000

Table 14: The total variance explained by the factors using principal axis factoring.

Observing the above tables, we note that only two significant factors are extracted, describing 52% of the data. In addition, from the scree plot, we notice that between the second and third factor the curve starts to flatten. This means that the factors after the second one are not significant.

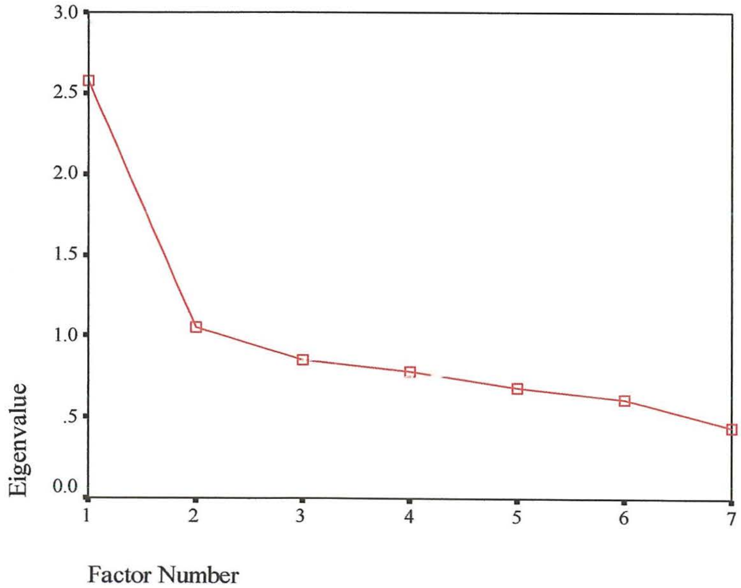

Figure 14: Scree plot.

The rotated matrix helps us to understand better the loadings of these factors. Here we see that for both extractions, we have that factor one is loaded with the general section (highlighted in blue) of the variables while the other factor clusters the attraction variables (highlighted in green).

	Factor	
	1	**2**
MIA	.363	.253
Accommodation	.546	
Restaurants	.570	.167
Public transport	.389	.186
Retail outlets	.488	.266
Historical sites	.219	.737
Cathedrals / churches	.213	.689

Table 15: The rotated matrix of the factors extracted by maximum likelihood.

	Factor	
	1	**2**
MIA	.365	.256
Accommodation	.541	
Restaurants	.564	.164
Public transport	.393	.191
Retail outlets	.496	.266
Historical sites	.224	.719
Cathedrals / churches	.208	.702

Table 16: The rotated matrix of the factors extracted by principal axis factoring.

Here we finish our process of this technique. We ended up with two sets each composed of two factor scores.

5.2.3.3 Method 2: Parceling

(i): Idea

A dictionary's definition of the word 'parcel' is 'to divide into portions' and this is precisely the idea behind the parceling method. Reading various papers, such as *"To Parcel or Not to Parcel: Exploring the Question, Weighing the Merits."* by Little Todd D., William A. Cunningham, Golan Shahar and Keith F. Widaman., helped us identify the main points of the parceling method.

First, we gather all the available variables and then apply factor analysis. From the output, we focus on the rotated factor matrix so that we can group the variables according to the components. Now, the mixed variables are grouped into a number of components, so they are identified into different groups. Let's say that after the first process we ended up with a five-component matrix. Then, we repeat the factor analysis process but this time only with those variables present in the first component. Hence, a new rotated factor matrix is obtained. An important point is that each time we repeat the process we suppress the value for the coefficient to be displayed. We repeat this procedure until we finish with the most significant variables of the first component. When we view the rotated matrix, the coefficients displayed show us whether it is possible to suppress any further. If we reach the limit where the next suppression leads to no values in the matrix, then we stop and consider those results as our main variables. After we obtain the variables, we start again with the next component and so on until the fifth component. At the final step, the most significant variables of each component are gathered together and factor analysis is performed. This time the factor scores are evaluated because they are then linearly related with the respondents' profile.

(ii): Physical Environment

The first step is to gather all the variables and apply factor analysis whilst setting the condition that those values that are less than 0.2 will not be available in the factor and rotated factor matrix. The following table is obtained from both extraction methods.

	Mean	Std. Deviation	Analysis N	Missing N
MIA	1.85	.710	2191	18
Accommodation	1.96	.966	2191	57
Restaurants	1.94	.703	2191	351
Public transport	2.08	.917	2191	449
Retail outlets	2.35	.750	2191	395
Historical Sites	1.87	.723	2191	412
Cathedrals / Churches	1.57	.622	2191	397
Road	3.89	.985	2191	67
Road signs	3.21	1.078	2191	272
Traffic	3.43	1.008	2191	179
Public conveniences	3.02	1.117	2191	229
Level of cleanliness	3.00	1.169	2191	52
Air quality	2.22	.985	2191	59
Sea quality	1.92	.792	2191	404

Table 17: The descriptive statistics of the physical environment variables.

As before, we obtain the total variance table. This gives us how many components the variables are grouped into. In, fact those factors whose eigenvalues are less than one are discarded. Hence, we note that in all there are four factors since from the fifth factor onwards have an eigenvalue less than one. These four factors are highlighted in red and they describe around 56% of the data. Both extraction methods lead to the following table.

	Initial Eigenvalues		
Factor	Total	% of Variance	Cumulative %
1	4.106	29.327	29.327
2	1.457	10.409	39.736
3	1.180	8.429	48.165
4	1.031	7.367	55.532
5	.859	6.134	61.666
6	.852	6.084	67.750
7	.753	5.381	73.131
8	.738	5.272	78.404
9	.603	4.309	82.712
10	.548	3.912	86.625
11	.522	3.731	90.355
12	.494	3.532	93.887
13	.467	3.334	97.221
14	.389	2.779	100.000

Table 18: The total variance explained.

Hence, the same graphical output is obtained. Here, we note that four factors are extracted. In fact, between the fourth and fifth factor it is evident that the curve is flattening, which signifies that we retain the four factors as concluded from the table.

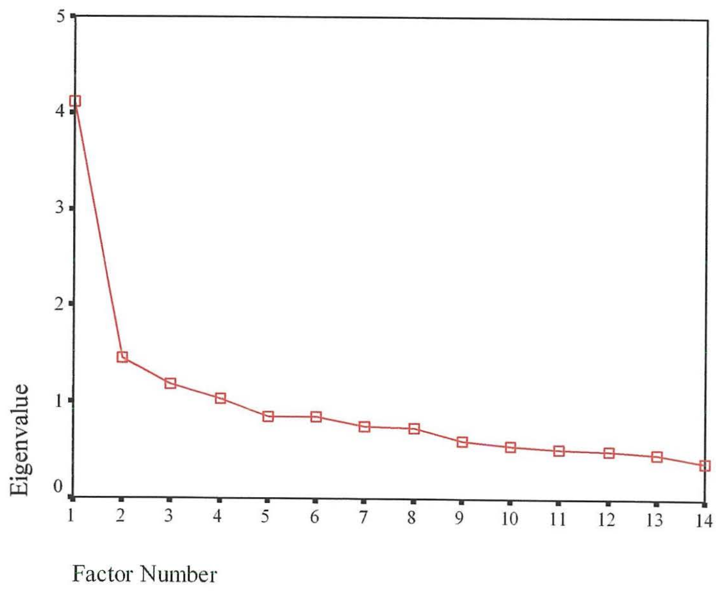

Figure 15: The scree plot.

Now, we have the rotated factor matrix in which we note that the empty cells imply that the coefficient value is less than 0.2. In this case, we obtain two different results for the two extraction methods.

The following is the maximum likelihood table.

	Factor			
	1	**2**	**3**	**4**
MIA	.413			
Accommodation	.355			
Restaurants	.440			
Public transport	.347	.214		
Retail outlets	.478			
Historical Sites	.614			
Cathedrals / Churches	.588			
Road		.705		
Road signs		.628		
Traffic		.599		.204
Public conveniences				.672
Level of cleanliness	.225	.245	.249	.665
Air quality			.859	
Sea quality	.315		.466	

Table 19: The rotated factor matrix obtained from maximum likelihood method.

Now, we observe the principal axis factoring table.

	Factor			
	1	**2**	**3**	**4**
MIA	.428			
Accommodation	.394		.221	
Restaurants	.495		.219	
Public transport	.342	.214		
Retail outlets	.501			
Historical Sites	.557			
Cathedrals / Churches	.522			.258
Road		.702		
Road signs		.624		
Traffic		.604	.213	
Public conveniences		.214	.588	
Level of cleanliness		.254	.658	.279
Air quality		.206	.255	.641
Sea quality	.254			.602

Table 20: The rotated factor matrix obtained from principal axis factoring method.

Each value represents the partial correlation between the variable and the rotated factor. These correlations can help us formulate an interpretation of the factors or components. From this output, we will notice the relations of the factors with the variables. Note that every variable is interpreted by a factor.

Now, we group the variables present in the first factor and apply factor analysis with the same extraction method and rotation. The only difference is that this time we consider only those that have a coefficient value greater than 0.5.

Observing the above tables, we note that the variables of the first factor are similar except for an extra variable present in the maximum likelihood extraction. This variable is *Level of cleanliness*.

	Mean	Std. Deviation	Analysis N	Missing N
MIA	1.85	.710	2191	18
Accommodation	1.96	.966	2191	57
Restaurants	1.94	.703	2191	351
Public transport	2.08	.917	2191	449
Retail outlets	2.35	.750	2191	395
Historical Sites	1.87	.723	2191	412
Cathedrals / Churches	1.57	.622	2191	397
Level of cleanliness	3.00	1.169	2191	52
Sea quality	1.92	.792	2191	404

Table 21: Descriptive statistics of the variables present in the first factor extracted by the maximum likelihood.

	Mean	Std. Deviation	Analysis N	Missing N
MIA	1.85	.710	2191	18
Accommodation	1.96	.966	2191	57
Restaurants	1.94	.703	2191	351
Public transport	2.08	.917	2191	449
Retail outlets	2.35	.750	2191	395
Historical Sites	1.87	.723	2191	412
Cathedrals / Churches	1.57	.622	2191	397
Sea quality	1.92	.792	2191	404

Table 22: Descriptive statistics of the variables present in the first factor extracted by the principal axis factoring.

When considering the maximum likelihood method we have obtained the following output. Observing the following table we note that we have obtained two significant factors since their eigenvalue is greater than one. Here the significant factors describe around 44% of the data.

	Initial Eigenvalues		
Factor	Total	% of Variance	Cumulative %
1	2.923	32.477	32.477
2	1.020	11.329	43.807
3	.924	10.266	54.072
4	.881	9.792	63.864
5	.758	8.424	72.288
6	.734	8.158	80.446
7	.656	7.294	87.740
8	.592	6.576	94.316
9	.512	5.684	100.000

Table 23: Total variance explained by the nine variables when applying maximum likelihood.

Applying the other extraction method, we obtained a two-component factor since only two factors have their eigenvalue greater than 1. For this method, the factors explain around 46% of the data.

	Initial Eigenvalues		
Factor	Total	% of Variance	Cumulative %
1	2.701	33.763	33.763
2	1.012	12.644	46.407
3	.881	11.017	57.424
4	.818	10.223	67.647
5	.752	9.401	77.048
6	.726	9.078	86.126
7	.598	7.474	93.600
8	.512	6.400	100.000

Table 24: Total variance explained by the eight variables when applying principal axis factoring.

A varimax rotation was applied and when the factor coefficient is suppressed under the value of 0.5, we obtained only these three variables. Hence, these variables are our most significant variables for this particular component as obtained for the two extraction methods.

	Factor	
	1	2
MIA		
Accommodation		
Restaurants	.561	
Public transport		
Retail outlets		
Historical Sites		.684
Cathedrals / Churches		.620
Level of cleanliness		
Sea quality		

Table 25: The rotated factor matrix of the first factor variables extracted by maximum likelihood.

	Factor	
	1	2
MIA		
Accommodation		
Restaurants		.602
Public transport		
Retail outlets		
Historical Sites	.662	
Cathedrals / Churches	.640	
Sea quality		

Table 26: The rotated factor matrix of the first factor variables extracted by principal axis factoring.

This process is repeated for the other components until finally we finish with the most significant variables. For the two methods, we end with the same variables. These are gathered in the following table.

	Mean	Std. Deviation	Analysis N	Missing N
Accommodation	1.96	.966	2191	57
Restaurants	1.94	.703	2191	351
Historical Sites	1.87	.723	2191	412
Cathedrals / Churches	1.57	.622	2191	397
Road	3.89	.985	2191	67
Road signs	3.21	1.078	2191	272
Traffic	3.43	1.008	2191	179
Public conveniences	3.02	1.117	2191	229
Level of cleanliness	3.00	1.169	2191	52
Air quality	2.22	.985	2191	59
Sea quality	1.92	.792	2191	404

Table 27: Descriptive statistics of the most significant variables.

We repeat factor analysis on these variables but apply a maximum likelihood extraction. We ended up with three significant eigenvectors having their eigenvalue greater than one.

	Initial Eigenvalues		
Factor	Total	% of Variance	Cumulative %
1	3.396	33.965	33.965
2	1.348	13.481	47.446
3	1.126	11.263	58.709
4	.868	8.678	67.387
5	.813	8.126	75.513
6	.558	5.584	81.097
7	.529	5.285	86.383
8	.501	5.007	91.389
9	.470	4.701	96.091
10	.391	3.909	100.000

Table 28: The total variance explained using maximum likelihood.

We repeated the same procedure as above but applied a principal axis factoring extraction and ended with four factors having eigenvalues greater than 1.

	Initial Eigenvalues		
Factor	Total	% of Variance	Cumulative %
1	3.532	32.110	32.110
2	1.392	12.657	44.768
3	1.127	10.243	55.011
4	1.008	9.167	64.178
5	.850	7.723	71.901
6	.644	5.858	77.759
7	.557	5.062	82.821
8	.528	4.804	87.626
9	.500	4.548	92.174
10	.470	4.273	96.446
11	.391	3.554	100.000

Table 29: The total variance explained using principal axis factoring.

For a maximum likelihood extraction method, we suppressed values less than 0.5 and we obtained the following table. Our ten major factors are grouped into these three factors. Our first factor component describes the infrastructure regarding the environment cleanliness and public conveniences state and the second component represents Malta's road infrastructure. The third component focuses on the attraction sites.

	Factor		
	1	2	3
Restaurants			
Historical Sites			.629
Cathedrals / Churches			.685
Road		.704	
Road signs		.645	
Traffic		.591	
Public conveniences	.623		
Level of cleanliness	.744		
Air quality			
Sea quality			

Table 30: Rotated factor matrix extracted by maximum likelihood.

For *Restaurants, Air quality* and *Sea quality* we do not have a value, which implies that these variables have a factor loading less than 0.5. From the graphical output below, we observe that the variable *Restaurants* is grouped with the third factor while the other two variables are grouped with the first variable.

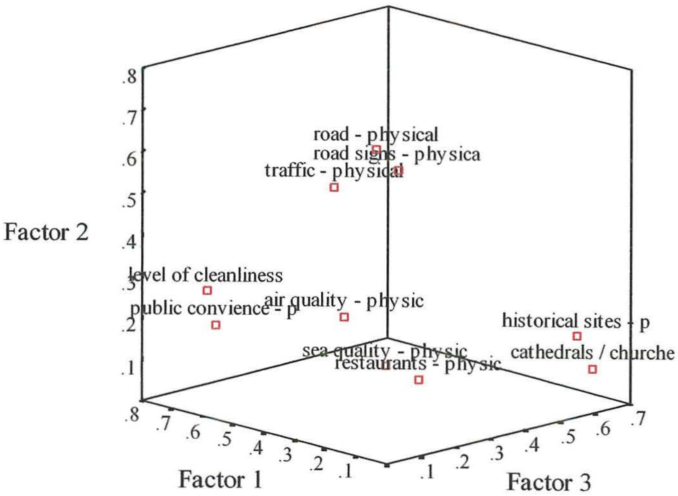

Figure 16: Factor plot of factors 1, 2, 3.

In this process, we suppressed values less than 0.5 and we obtained the following table. Our eleven major factors are grouped into these four factors. Hence, our first factor component represents Malta's road infrastructure and the second component focuses on the attraction sites. The third component describes the infrastructure regarding the environment cleanliness and public conveniences state. While the last component represents the infrastructure but regarding our environment quality.

	Factor			
	1	2	3	4
Accommodation				
Restaurants				
Historical Sites		.644		
Cathedrals / Churches		.638		
Road	.702			
Road signs	.641			
Traffic	.604			
Public conveniences			.628	
Level of cleanliness			.703	
Air quality				.748
Sea quality				.544

Table 31: Rotated factor matrix extracted by principal axis factoring.

For *Accommodation* and *Restaurants*, we do not have a value that signifies that these variables have a factor loading less than 0.5. Since we have a four-factor matrix, it is not possible to obtain a better graphical interpretation.

When we obtain these significant variables and performed factor analysis on these variables, we evaluate the factor scores. Here, the process of parceling finishes since we have gathered the most significant variables and analyzed them.

(iii): Service provided by employees

In the following table, we have those variables within this aspect that were analysed. Here we note that, *Retail Outlets* having the highest mean of 2.17 signifying that the services are not appreciated much by the tourists while *Churches / Cathedrals* have a mean of 1.78 signifying that the relevant employees run our places of worship efficiently.

	Mean	Std. Deviation	Analysis N	Missing N
MIA	1.95	.825	2240	83
Accommodation	1.81	.923	2240	120
Restaurants	1.85	.715	2240	445
Public transport	1.99	.837	2240	550
Retail outlets	2.17	.706	2240	557
Historical sites	1.98	.736	2240	702
Cathedrals / churches	1.78	.673	2240	722

Table 32: The descriptive statistics of the service provided by employees variables.

Applying the same procedure with both extraction methods, we end up with the same number of factors representing approximately 52% of the data. This result is viewed in the following two outputs; the total variance explained table and the scree plot. These outputs are identical for both extraction methods.

	Initial Eigenvalues		
Factor	Total	% of Variance	Cumulative %
1	2.582	36.885	36.885
2	1.049	14.981	51.866
3	.854	12.207	64.073
4	.784	11.198	75.270
5	.679	9.701	84.971
6	.610	8.711	93.682
7	.442	6.318	100.000

Table 33: Total variance explained.

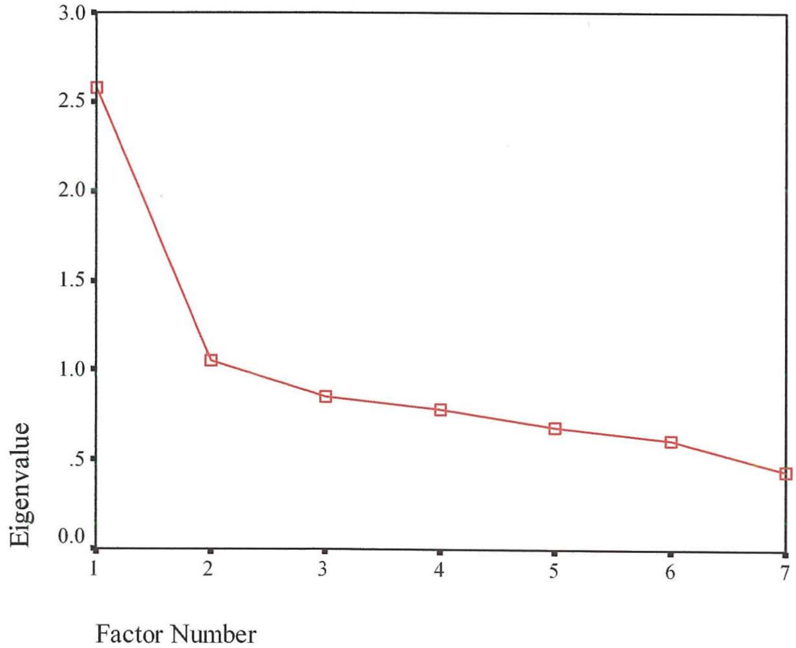

Figure 17: Scree plot of the eigenvalues against the factors.

Then we viewed the rotated factor matrix. For both methods, the variables of each factor component are the same and the only difference is the loading value. In fact, these two tables are the rotated factor matrix of both methods and note the value difference of each variable within each component. The empty values present in these tables represent those values that value their factor loading less than 0.2.

	Factor	
	1	**2**
MIA	.363	.253
Accommodation	.546	
Restaurants	.570	
Public transport	.389	
Retail outlets	.488	.266
Historical sites	.219	.737
Cathedrals / churches	.213	.689

Table 34: Rotated factor matrix obtained by a maximum likelihood extraction.

	Factor	
	1	**2**
MIA	.365	.256
Accommodation	.541	
Restaurants	.564	
Public transport	.393	
Retail outlets	.496	.266
Historical sites	.224	.719
Cathedrals / churches	.208	.702

Table 35: Rotated factor matrix obtained by a principal axis factoring extraction.

Now, the same procedure as in the other aspect was applied until we finished with the most significant variables. In the following table, we have these variables and for both methods, they are the same.

	Mean	Std. Deviation	Analysis N	Missing N
Historical sites	1.98	.736	2240	702
Cathedrals / churches	1.78	.673	2240	722
Accommodation	1.81	.923	2240	120
Restaurants	1.85	.715	2240	445

Table 36: Descriptive statistics of the final significant variables.

Applying factor analysis using both extractions, we ended up with a two-factor model describing 73% of the data.

	Initial Eigenvalues		
Factor	Total	% of Variance	Cumulative %
1	1.893	47.326	47.326
2	1.031	25.768	73.094
3	.632	15.792	88.886
4	.445	11.114	100.000

Table 37: Total variance explained.

Utilizing a varimax rotation and suppressing values under 0.5, we obtain the following two components in which the attractions variables are grouped in the first component while the other two variables are gathered in the other component.

	Factor	
	1	**2**
Historical sites	.726	
Cathedrals / churches	.721	
Accommodation		.576
Restaurants		.592

Table 38: Rotated factor matrix.

Here, the process of parceling for the service provided by employees finishes. We ended with two sets, one for each method, each containing two factor scores.

Now, after all this process, we take a note of those variables that at the end are considered significant for our analysis and from which we draw our factor scores. This leads to confirmatory factor analysis.

5.3 Confirmatory Factor Analysis

The primary objective of confirmatory factor analysis is to determine the ability of a predefined factor model to fit an observed set of data.

Some common uses of confirmatory factor analysis are to establish the validity of a single factor model and to compare the ability of two different models to account for the same set of data. Other uses are to test the significance of a specific factor loading or to test the relationship between two or more factor loadings. These are only few ways in which confirmatory factor analysis is applicable, in fact there are more other procedures where this method is applicable.

Exploratory factor analysis is applied in various applications, for example, when we want to determine which features are most important when classifying a group of items. In addition, it is useful when we want to determine which set of items can be clustered together in a questionnaire or to demonstrate the dimensionality of a measurement scale. Another application is when we want to generate factor scores that represent the values of the underlying constructs for use in other analyses.

Now, we will discuss the significant variables of each process. For the first technique, that is factor analysis, the final sets of variables considered are all those valid variables that we have considered for the analysis. For both extraction methods, the variables are the same. For the physical environment, we have fourteen variables while for the service provided by employees we have seven variables.

For the second technique, we have started with these fourteen and seven variables but were reduced to a lesser amount. For the case of the physical environment applying the maximum likelihood extraction, we ended up with ten significant variables while doing the same procedure but applying the principal axis factoring extraction, we had eleven significant variables. For the case of service provided by employees, for both extraction methods, we have obtained four significant variables.

Now, after all these processes, we consider each technique, method and aspect and evaluate the factor scores. These factor scores are important because these scores are then related with the tourists' profile. Hence, we are trying to find any relationship between the variables forming the tourists' profile and these factor scores derived from these different methods. This application will be discussed in the next chapter.

6 FACTOR SCORES AND LINEAR MODELS

As discussed in the previous chapter we have ended up with different sets of factor scores for each method applied within each technique. The purpose of extracting these factor scores is that we can obtain a linear relationship of these scores with the tourists' profile variables. Before we discuss this relationship, the factor scores were analysed to check whether they satisfy the normality condition. This is required because to obtain linear models, the normality assumption must be satisfied.

6.1 Factor Scores

As we said in the previous paragraph, these scores are analysed to check for normality. In fact, the Kolmogorov-Smirnov test is applied to each factor score and from the p-value (Asymp. Sig) we observe whether they satisfy the normal distribution.

The criterion to accept whether a variable is normally distributed or not, is based on the fact that if the p-value is greater than 0.05, then the variable does not depart significantly from normal behaviour and conversely if the p-value is less than 0.05. When the case that the variable, in our case the factor score, is not normally distributed, we need to apply a transformation to render it adequate to observe the normal behaviour.

For example, consider the following factor score extracted by the maximum likelihood method when applied on the physical environment variables. This is already normally distributed since the p-value is greater than 0.05.

		Factor score 2
N		2191
Normal Parameters	Mean	.0000000
	Std. Deviation	.81979066
Most Extreme Differences	Absolute	.026
	Positive	.026
	Negative	-.026
Kolmogorov-Smirnov Z		1.196
Asymp. Sig. (2-tailed)		.115

Table 39: The One-Sample Kolmogorav Smirnov Test of the second factor score.

In fact, observing the following histogram, we notice that the normal curve fits very smoothly over the bars except for the tail ends. With the help of a Q-Q plot, we can confirm this statement since the majority of the points are on the line except the ends.

A Q-Q plot is a graph where the quantiles of a variable's distribution are plotted against the quantiles of any of a number of test distributions. Probability plots are generally used to determine whether the distribution of a variable matches a given distribution. If the selected variable matches the test distribution, the points cluster around a straight line

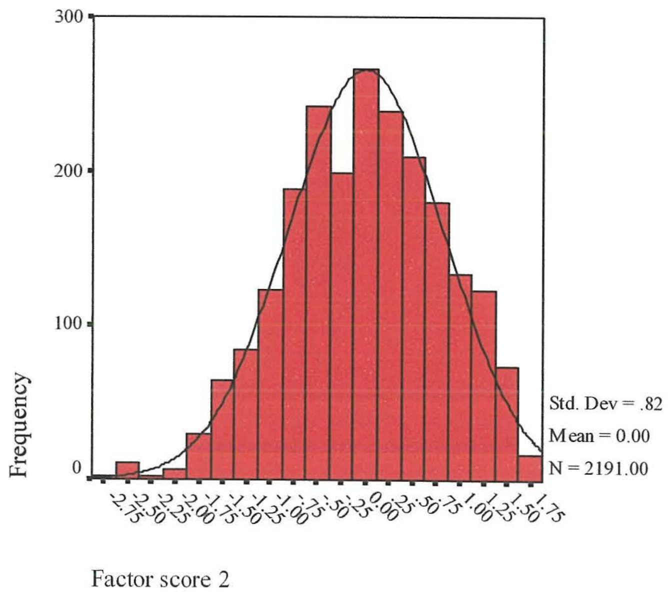

Figure 18: Histogram with normal curve of factor score 2.

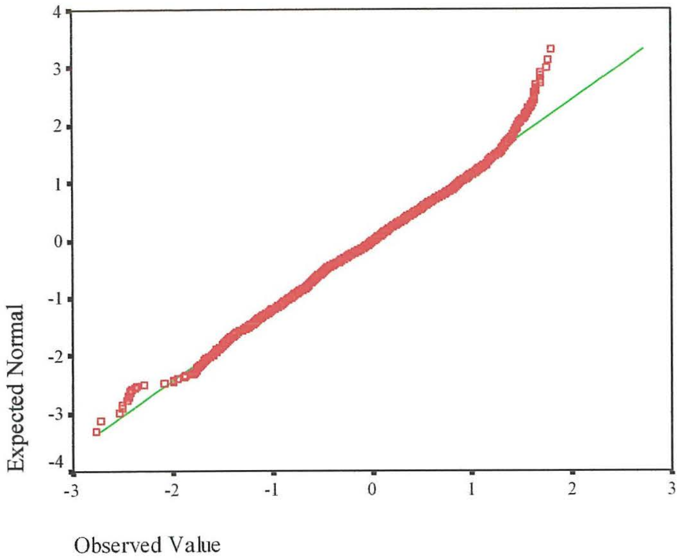

Figure 19: Q-Q plot of the second factor score.

Generally, the distribution of the obtained factor scores departs significantly from the normal behaviour. Hence, in such cases, we applied a transformation to satisfy as much as possible our request for normal behaviour. The following shows is an example of a factor score with a p value of 0.001, which is less than 0.05, hence implying that this score is not normal. The scores of this example are obtained from the physical environment when extracted by the principal axis factoring method.

		Factor score 1
N		2191
Normal Parameters	Mean	.0000000
	Std. Deviation	.79851264
Most Extreme Differences	Absolute	.041
	Positive	.041
	Negative	-.016
Kolmogorov-Smirnov Z		1.903
Asymp. Sig. (2-tailed)		.001

Table 40: One-Sample Kolmogorov-Smirnov Test

Graphically this is noticed with the help of the Q-Q plot since the data points do not fit the green line. The green line present in these plots represents the distribution we wish to fit and in our case, it represents a normal distribution.

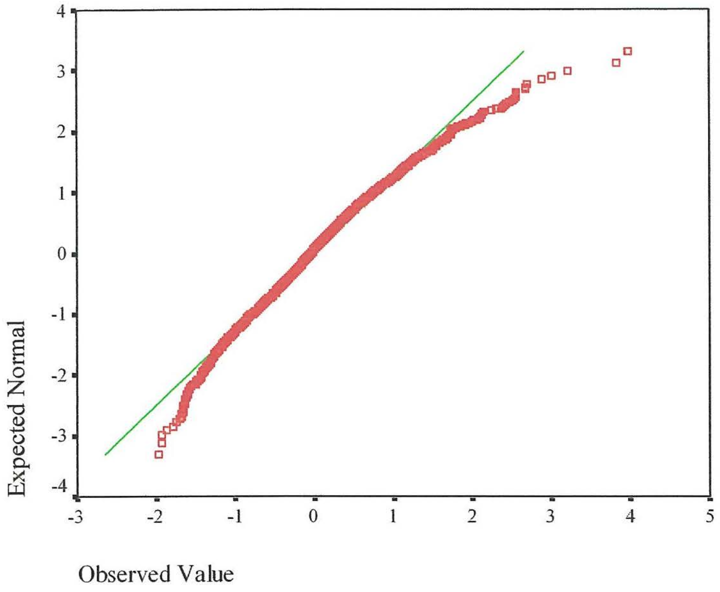

Figure 20: Q-Q plot of factor score 1

Now, examining this example and other similar ones, we require applying a method that renders our data, factor scores, significantly normally distributed. The technique used is the Box-Cox Transformation. In the Box-Cox, we are transforming a response feature either to correct for non-normality or a heteroscedastic variance structure. A useful class of transformations for this purpose is the power transform $\frac{y^{\lambda}-1}{\lambda}$, where λ is a parameter to be determined.

For this transformation, we have used a software package called GLIM in which we have inputted our data and called the Box-Cox directive. First, we have to recode the data because for the use of the Box-Cox there must be no zeros or negative values present in the data. Therefore, for each score the minimal value is found and then is added to each value so that the data becomes positive. Then we insert the data in this package and run the directive. When the directive responds we have to determine a range for lambda and the increment. After repeating the process for a number of times, we end up with a significant value lambda. Then, this lambda is power by which the data is transformed. For each factor score we have a different value for lambda.

Considering the same factor score as above and applying the Box-Cox transformation, we obtain our lambda value equal to 0.6685.

Now, using the other statistical package Spss, we transform our data by multiplying to the power of 0.6685. From the Kolmogorov-Smirnov test, we verify whether our transformed score does not significantly depart from normal behaviour.

		var1 = (newfac1) ** 0.6685
N		2191
Normal Parameters	Mean	1.5597
	Std. Deviation	.42999
Most Extreme Differences	Absolute	.024
	Positive	.022
	Negative	-.024
Kolmogorov-Smirnov Z		1.136
Asymp. Sig. (2-tailed)		.151

Table 41: The Kolmogorov-Smirnov Test of var1.

From the above table, we note that the p-value is 0.151, which is greater than 0.05 and hence implies that this transformed score satisfies the normality behaviour. Observing this result from a graphical point of view, we note that the histogram fits the normal distribution curve smoothly, except at the edges. This is further verified form the Q-Q plot where the data points fit the green line perfectly except for a few points present at the ends.

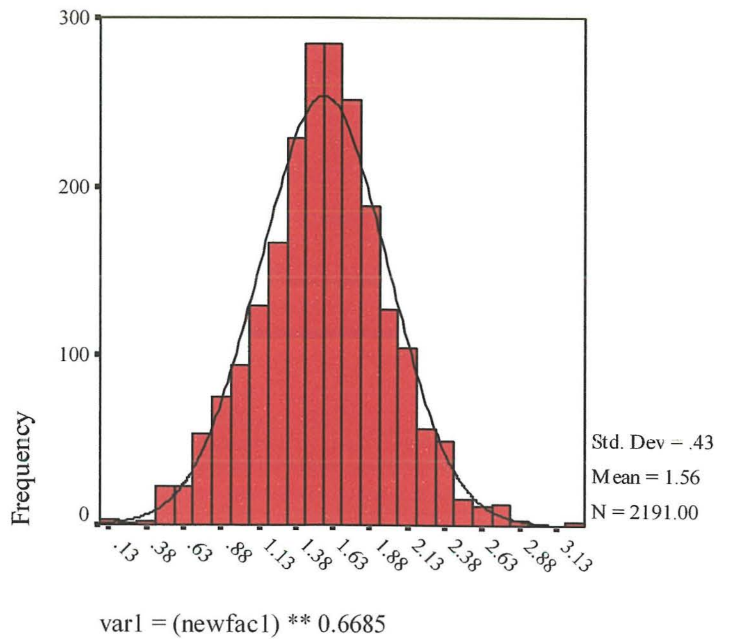

Figure 21: Histogram of var1.

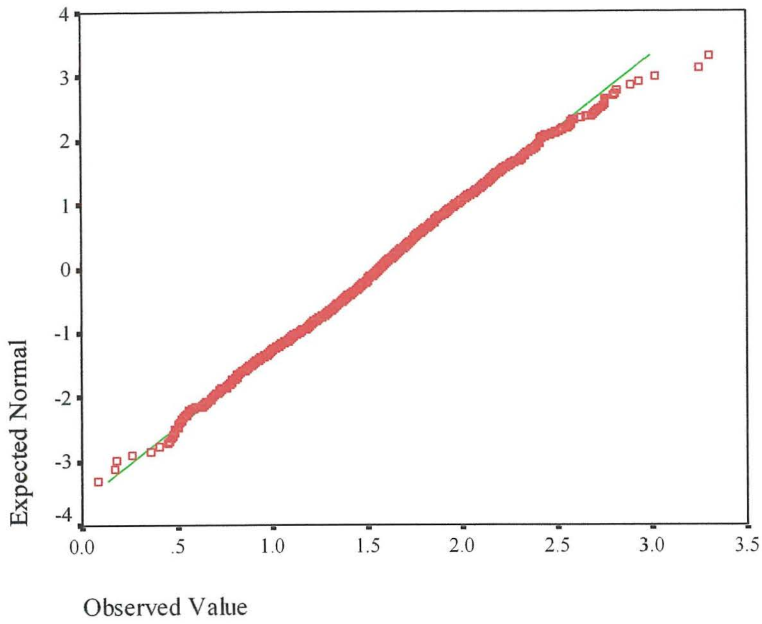

Figure 22: The Q-Q plot of var1.

As we said before, each factor score has its own lambda for the transformation. This table shows the lambda values required for each individual score to be best fitted into a normal distribution.

Technique	Extraction	Aspect	Score	Lambda
Factor Analysis	Maximum Likelihood	Physical	1	0.7
			2	1
			3	0.486
			4	0.6345
		Service	1	0.393
			2	0.3885
	Principal Axis Factoring	Physical	1	0.6685
			2	1
			3	0.6725
			4	0.2545
		Service	1	0.42
			2	0.381
Parceling	Maximum Likelihood	Physical	1	0.6515
			2	1
			3	0.2985
		Service	1	0.3935
			2	0.2045
	Principal Axis Factoring	Physical	1	1
			2	0.3295
			3	0.627
			4	0.42015
		Service	1	0.363
			2	0.477

Table 42: The lambda required for each factor score to be transformed.

Although we have performed this Box-Cox transformation, some factor scores still depart significantly from the normal behaviour. A typical example is the following factor score composed of the service provided variables and extracted from a maximum likelihood method. In the following table, we have the factor score before the transformation and the factor score after the transformation. From the p-value, we note that the p-value of this factor score is 0 before and after transformation. Hence, it still departs significantly from the normal behaviour.

		Factor score 2	var2 = (newfac2) ** 0.3885
N		2240	2240
Normal Parameters	Mean	.0000000	1.2832
	Std. Deviation	.81614727	.20696
Most Extreme Differences	Absolute	.176	.143
	Positive	.176	.135
	Negative	-.094	-.143
Kolmogorov-Smirnov Z		8.347	6.762
Asymp. Sig. (2-tailed)		.000	.000

Table 43: The Kolmogorov-Smirnov test.

Graphically, we note the presence of multi-peaks, which reflect that there are several different processes with different centres. In this case, we have five peaks so we need five processes so that we can obtain a clearer view of what is really happening in either individual process. To do this process we need to separate each peak to obtain separate distributions and then analyse each distribution to see whether each distribution does or does not depart significantly from normal behaviour. Due to lack of time, these factor scores are not tackled in this dissertation.

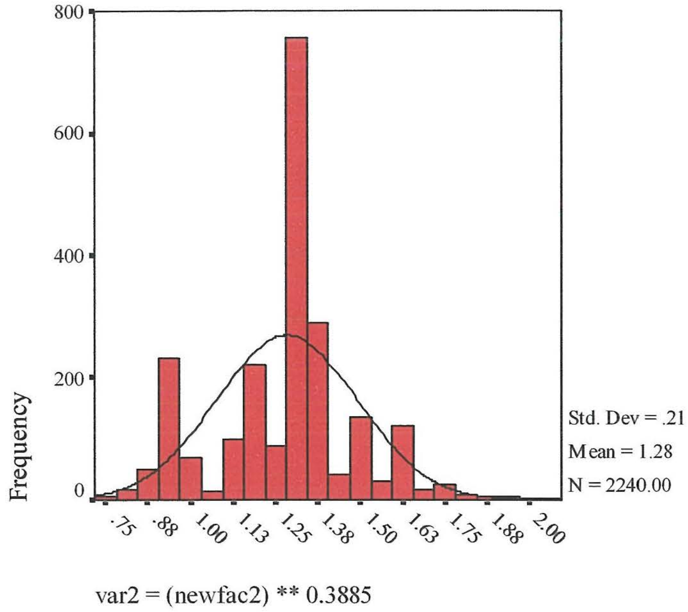

Figure 23: Histogram of the factor score.

Now, considering the histogram of the above-mentioned variable, it should be noted that the normal distribution curve seems to fit the distribution of the 'transformed' scores well. However, since the resulting p-value is less than 0.05 the underlying distribution of the data is not the normal distribution.

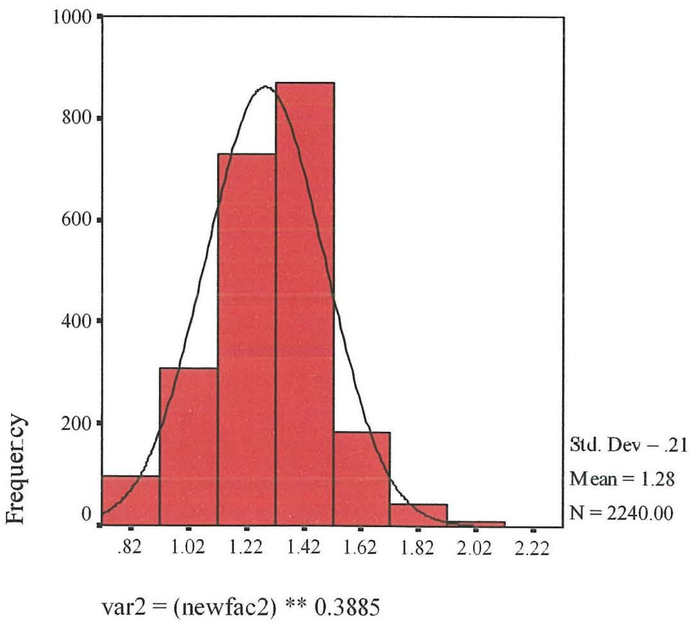

Figure 24: Histogram of the newly categorised data.

Until now, we have analysed these factor scores extracted from the different methods and we have encountered three types of possible outcomes. The first is when the factor score obtain is normally distributed. Secondly is when we apply a transformation and obtain a factor score that does not significantly depart from normal distribution. Finally is when we end up with factor scores that even though they are transformed, they still do not satisfy the normal behaviour. Now, for the first two types we will continue our analysis since to obtain linear models we need to satisfy the normal distribution. For the third type, we need further investigation to identify the different processes hidden in each factor score.

Our next step is to consider those normally distributed factor scores and obtain a relation ship of these scores with the socio-demographic variables of the tourists'. This leads us to the next section, where we are going to discuss these results.

6.2 Linear Models

To obtain these linear relationships, we utilized the statistical package Spss and applied the general Univariate linear model function. In this section, we will discuss the outputs that we acquired from this analysis. Here we have considered the best relationship results obtained by each technique and each extraction method. First, we will go through a quick look at the variables forming the tourists' profile.

6.2.1 Tourists' Profile

The tourist profile is composed of eight variables, which are season, regions grouped into country, gender, marital status, full time and part time job, net income and age-groups. Each variable has a number of parameters. In the following table, we have grouped the variables and their categories.

Season	**Country**	**Gender**	**Marital Status**	**Full time job**	**Part time job**	**Income**	**Age**
Summer	Scotland	Male	Single	Director	Director	0 - 1000	16 – 25
Autumn	England	Female	Married / Living together	Manager	Manager	1001 – 2000	26 – 35
Winter	Wales		Divorced / Separated	Office / Retail worker	Office / Retail worker	2001 – 3000	36 – 45
Spring	Ireland		Widowed	Professional	Professional	3001 – 4000	46 – 55
	Northern Ireland			Skilled worker / tradesman	Skilled worker / tradesman	4001 - 5000	56 – 65
				Manual worker	Manual worker	5001+	66+
				Self-employed	Self-employed		
				Student	Student		
				Housewife	Housewife		
				Unemployed	Unemployed		
				Retired	Retired		

Table 44: Profile variables and their respective categories.

6.2.2 Factor Analysis Technique

When applying the factor analysis technique with a maximum likelihood extraction method, we obtained a factor score which, when analyzed with the profile variables, gave the following results. The profile variables considered for the analysis are highlighted in red.

Source	Type III Sum of Squares	Degrees of freefom	Mean Square	F	Sig.
Corrected Model	22.132	37	.598	.945	.568
Intercept	5.232E-02	1	5.232E-02	.083	.775
Season	1.201	3	.400	.632	.599
Grouped region into country	2.013	4	.503	.795	.536
Gender	.626	1	.626	.989	.327
Marital Status	1.226	3	.409	.645	.591
Fulltime job	5.746	8	.718	1.134	.365
Part time job	3.205	10	.320	.506	.874
Net new income	2.481	3	.827	1.306	.287
Age group	2.138	4	.535	.844	.506
Error	22.791	36	.633		
Total	47.096	74			
Corrected Total	44.924	73			

Table 45: Test of between-subjects effects of the dependent variable factor score 2.

This is the first output, where we considered all the variables without interactions. The purpose of not considering the interaction terms is because we have a high number of levels present in some of the variables such as full time job and part time job. In fact, these two variables contain eleven different jobs. From this model, we obtained an R Squared value of 0.493. This signifies that this model represents 49.3% of the variability of the data.

Now, observing the table, we take note of the significant values (**Sig**). From these values, we choose the one with the highest significance and discard it. In this particular case, we have part time job with the highest p-value equal to 0.874. Then we repeat the process until we end up with a number of variables that have a p-value less than 0.05. Hence, the values that have a p-value less than 0.05 are significant ones for this factor score.

For this regression factor score, we ended up with five significant variables (highlighted in red) with a p-value (**Sig**) less than 0.05. From the R Squared value, we conclude that this linear model is explaining 4.5% of the variability of the data.

Source	Type III Sum of Squares	Degrees of freedom	Mean Square	F	Sig.
Corrected Model	36.149	16	2.259	3.618	.000
Intercept	1.145	1	1.145	1.833	.176
Season	6.176	3	2.059	3.297	.020
Grouped region into country	6.112	4	1.528	2.447	.045
Gender	5.289	1	5.289	8.471	.004
Marital status	7.063	3	2.354	3.771	.010
Age group	8.401	5	1.680	2.691	.020
Error	758.637	1215	.624		
Total	794.879	1232			
Corrected Total	794.786	1231			

Table 46: Tests of Between-Subjects Effects of the second factor score extracted by maximum likelihood method.

In addition to these results, we also obtained the parameter estimates in the column labelled **B**. Observing the parameter estimates table, we note that the last level or parameter of each variable is zero, which signifies that this parameter is redundant. This is so since the Spss is programmed to alias the last level of each variable, hence the β – value of these levels is zero.

From the t statistics (**t**), we can determine the relative importance of each variable in the relationship. The t statistic is **B** divided by the standard error (**Std. Error**). As a guide regarding useful predictors, we look for t values well below -2 or above +2.

	B	Std. Error	t	Sig.	95% Confidence Interval	
Parameter					Lower Bound	Upper Bound
Intercept	-.549	.292	-1.880	.060	-1.121	2.376E-02
[SEASON=1]	1.541E-02	.071	.216	.829	-.125	.156
[SEASON=2]	.162	.059	2.764	.006	4.705E-02	.277
[SEASON=3]	.133	.065	2.037	.042	4.883E-03	.260
[SEASON=4]	0	.	.	.	.	.
[REGGROUP=1]	-.199	.278	-.715	.475	-.744	.347
[REGGROUP=2]	-.117	.269	-.434	.664	-.645	.411
[REGGROUP=3]	.112	.294	.381	.703	-.465	.689
[REGGROUP=4]	-.409	.295	-1.387	.166	-.988	.170
[REGGROUP=5]	0	.	.	.	.	.
[GENDER=1]	.138	.047	2.910	.004	4.488E-02	.231
[GENDER=2]	0	.	.	.	.	.
[MARTST=1]	.358	.137	2.609	.009	8.874E-02	.626
[MARTST=2]	.379	.113	3.341	.001	.156	.601
[MARTST=3]	.318	.155	2.049	.041	1.349E-02	.622
[MARTST=4]	0	.	.	.	.	.
[AGEGROUP=1]	3.938E-02	.166	.238	.812	-.285	.364
[AGEGROUP=2]	.309	.107	2.892	.004	9.951E-02	.519
[AGEGROUP=3]	.171	.083	2.059	.040	8.055E-03	.334
[AGEGROUP=4]	.226	.075	3.027	.003	7.941E-02	.372
[AGEGROUP=5]	.186	.070	2.672	.008	4.935E-02	.322
[AGEGROUP=6]	0	.	.	.	.	.

Table 47: Parameter Estimates of the dependent variable factor score 2 extracted by maximum likelihood method.

In this table we have ten useful predictors since their value is higher than 2. These predictors are the ones highlighted in blue where each parameter signifies a label. For each variable, we have an amount of levels. By levels, we understand the different values inputted for each variable. For the season variable, we have four levels that indicate the four seasons. From the above table, we have that season has two important values that are most influential in this relationship. Hence, season = 2 signifies autumn while season = 3 represents winter. Gender has two levels and gender = 1 represents male population. For the marital status, we have three levels 1, 2 and 3 which respectively mean single, married / living together and divorced / separated. Regarding the last variable, age group we have six levels but only four are the most useful. These are from the second to the fifth level. Each level represents an age bracket. Hence, our most significant age brackets are 26 – 35, 36 – 45, 46 – 55 and 56 – 65.

Now using the same data reduction technique, we analyzed another factor score extracted from the other method, principal axis factoring. Here we have the result of the most significant variables for this factor score. The same process was performed, where we started with all the variables and each time reduced the variable with the highest p-value until we finished with those having a p-value less than 0.05.

Source	Type III Sum of Squares	Degrees of freedom	Mean Square	F	Sig.
Corrected Model	30.368	13	2.336	3.687	.000
Intercept	.811	1	.811	1.281	.258
Regrouped region into country	6.083	4	1.521	2.400	.048
Gender	5.749	1	5.749	9.075	.003
Marital Status	6.452	3	2.151	3.395	.017
Age grouped	8.153	5	1.631	2.574	.025
Error	771.646	1218	.634		
Total	802.164	1232			
Corrected Total	802.013	1231			

Table 48: The test between subjects effect of the second factor score.

From these variables, we obtained the parameter estimates table, where we will observe the **t** value. From this **t** value, we identify the most important variables for this relationship.

	B	Std. Error	t	Sig.	95% Confidence Interval	
Parameter					**Lower Bound**	**Upper Bound**
Intercept	-.417	.286	-1.458	.145	-.978	.144
[REGGROUP=1]	-.201	.278	-.724	.469	-.747	.344
[REGGROUP=2]	-.144	.268	-.538	.590	-.669	.381
[REGGROUP=3]	9.483E-02	.294	.322	.747	-.483	.672
[REGGROUP=4]	-.436	.295	-1.476	.140	-1.016	.143
[REGGROUP=5]	0	.	.	.	.	.
[GENDER=1]	.143	.048	3.012	.003	5.004E-02	.237
[GENDER=2]	0	.	.	.	.	.
[MARTST=1]	.352	.138	2.550	.011	8.114E-02	.623
[MARTST=2]	.364	.114	3.186	.001	.140	.588
[MARTST=3]	.325	.156	2.082	.038	1.873E-02	.630
[MARTST=4]	0	.	.	.	.	.
[AGEGROUP=1]	-1.778E-02	.165	-.108	.914	-.341	.306
[AGEGROUP=2]	.272	.106	2.569	.010	6.434E-02	.480
[AGEGROUP=3]	.147	.082	1.785	.074	-1.455E-02	.308
[AGEGROUP=4]	.216	.074	2.906	.004	7.021E-02	.362
[AGEGROUP=5]	.190	.070	2.708	.007	5.222E-02	.327
[AGEGROUP=6]	0	.	.	.	.	.

Table 49: Parameter estimates of the dependent factor score.

These useful predictors are highlighted in blue and are very similar to the previous table except that in this relationship, the season variable is not significant.

6.2.3 Parceling Technique

Applying the other technique of data reduction, parceling, we have the following well-defined relationships. Let us start with the first extraction method, maximum likelihood, when applied on the physical environment variables. Here we have the following significant variables of the socio-demographic variables for this factor score.

Note that these variables are significant since their significance value is less than 0.05. From the following table we have five variables, which are identical to those obtained in the previous technique when applying the same extraction method.

Source	Type III Sum of Squares	Degrees of freedom	Mean Square	F	Sig.
Corrected Model	36.852	16	2.303	3.696	.000
Intercept	1.100	1	1.100	1.765	.184
Season	6.268	3	2.089	3.353	.018
Regrouped region into country	6.404	4	1.601	2.569	.037
Gender	5.358	1	5.358	8.598	.003
Marital Status	6.938	3	2.313	3.712	.011
Age grouped	8.862	5	1.772	2.844	.015
Error	757.062	1215	.623		
Total	794.013	1232			
Corrected Total	793.914	1231			

Table 50: The test between subjects effect of the second factor score.

From the parameter estimates table we note the following results. These parameters contribute to the relationship of the above variables with this factor score. The ones highlighted in blue signify the ones that are of a greater influence in this relationship.

	B	Std. Error	t	Sig.	95% Confidence Interval	
Parameter					Lower Bound	Upper Bound
Intercept	-.560	.291	-1.921	.055	-1.132	1.180E-02
[SEASON=1]	4.666E-03	.071	.065	.948	-.135	.145
[SEASON=2]	.159	.059	2.706	.007	4.358E-02	.274
[SEASON=3]	.133	.065	2.048	.041	5.616E-03	.261
[SEASON=4]	0	.	.	.	.	.
[REGGROUP=1]	-.186	.278	-.669	.504	-.730	.359
[REGGROUP=2]	-.109	.269	-.407	.684	-.637	.418
[REGGROUP=3]	.125	.294	.424	.671	-.451	.701
[REGGROUP=4]	-.415	.295	-1.407	.160	-.993	.164
[REGGROUP=5]	0	.	.	.	.	.
[GENDER=1]	.139	.047	2.932	.003	4.586E-02	.231
[GENDER=2]	0	.	.	.	.	.
[MARTST=1]	.347	.137	2.537	.011	7.870E-02	.616
[MARTST=2]	.375	.113	3.314	.001	.153	.598
[MARTST=3]	.317	.155	2.050	.041	1.358E-02	.621
[MARTST=4]	0	.	.	.	.	.
[AGEGROUP=1]	8.149E-02	.165	.493	.622	-.243	.406
[AGEGROUP=2]	.320	.107	2.994	.003	.110	.530
[AGEGROUP=3]	.182	.083	2.197	.028	1.950E-02	.345
[AGEGROUP=4]	.237	.074	3.179	.002	9.067E-02	.383
[AGEGROUP=5]	.198	.069	2.843	.005	6.121E-02	.334
[AGEGROUP=6]	0	.	.	.	.	.

Table 51: Parameter estimates of the significant variables of this relationship.

Applying the same process but considering a principal axis factoring extraction, we obtain the following five significant variables.

Source	Type III Sum of Squares	Degrees of freedom	Mean Square	F	Sig.
Corrected Model	37.155	16	2.322	3.695	.000
Intercept	1.184	1	1.184	1.884	.170
Season	6.707	3	2.236	3.557	.014
Regrouped region into country	6.368	4	1.592	2.533	.039
Gender	5.320	1	5.320	8.465	.004
Marital Status	6.852	3	2.284	3.634	.013
Age grouped	8.797	5	1.759	2.800	.016
Error	763.595	1215	.628		
Total	800.875	1232			
Corrected Total	800.750	1231			

Table 52: The test between subjects effect of the first factor score.

From these five variables, we obtain the following parameter estimates.

	B	Std. Error	t	Sig.	95% Confidence Interval	
Parameter					Lower Bound	Upper Bound
Intercept	-.562	.293	-1.921	.055	-1.137	1.186E-02
[SEASON=1]	-9.265E-04	.072	-.013	.990	-.142	.140
[SEASON=2]	.162	.059	2.747	.006	4.620E-02	.277
[SEASON=3]	.136	.065	2.092	.037	8.460E-03	.265
[SEASON=4]	0	.	.	.	.	.
[REGGROUP=1]	-.174	.279	-.624	.533	-.721	.373
[REGGROUP=2]	-.102	.270	-.377	.706	-.632	.428
[REGGROUP=3]	.132	.295	.448	.654	-.446	.711
[REGGROUP=4]	-.409	.296	-1.381	.168	-.990	.172
[REGGROUP=5]	0	.	.	.	.	.
[Q2GENDER=1]	.138	.047	2.909	.004	4.497E-02	.231
[Q2GENDER=2]	0	.	.	.	.	.
[Q2MARTST=1]	.346	.137	2.519	.012	7.665E-02	.616
[Q2MARTST=2]	.373	.114	3.278	.001	.150	.596
[Q2MARTST=3]	.313	.155	2.011	.045	7.566E-03	.618
[Q2MARTST=4]	0	.	.	.	.	.
[AGEGROUP=1]	5.962E-02	.166	.359	.720	-.266	.385
[AGEGROUP=2]	.312	.107	2.910	.004	.102	.523
[AGEGROUP=3]	.178	.083	2.141	.032	1.491E-02	.341
[AGEGROUP=4]	.236	.075	3.153	.002	8.914E-02	.383
[AGEGROUP=5]	.195	.070	2.802	.005	5.858E-02	.332
[AGEGROUP=6]	0	.	.	.	.	.

Table 53: Parameter estimates of the significant variables.

In this section, we focused on the most prominent relationships. In fact, we had other relationships but were not as prominent as these were. There were also situations where we ended with no relation. These cases could have occurred due to the fact that we only considered these eight variables grouped into two questions of the questionnaire. There can be other variables which could result in a relationship with the factor scores. Unfortunately, due to lack of time and due to the complexity of the data it was not possible to consider all the variables for our relationship. From various papers, we note that these socio-demographic variables are the most utilised variables for relationship analysis.

As a concluding note, we realize that from these two techniques, we ended with the same results. In fact, only one factor score was different. The difference was in the amount of variables but not the presence of an other variable which is substituting one of the others.

7 CONCLUSIONS AND RECOMMENDATIONS

7.1 Conclusions

In this dissertation, we have put into use statistical methods to solve problems arising in tourism data. In particular, it tackles problems of analysing Likert scale questions. The application concerns real tourism data collected by the Malta Tourism Authority (MTA). The objective was to apply data reduction techniques to evaluate these types of ratings, thus rendering it valid for future studies. Also, a relationship of the factor scores with the respondents' profile was studied.

Now we will outline the problems encountered during the whole process and discuss the reasons why and how they were tackled if possible.

Starting with the theoretical aspect, one of the main problems was that most theoretical results assume normality conditions. However, in practice it is more likely to deal with non-normal populations. Hence, the concept of non-normal factor analysis had to be introduced since our scale was not normally distributed.

On the practical side, I decided to use yearly data composed of four seasons but collected over a span of two years since this survey was launched in Summer 2003. The question considered for this analysis was the one in which the tourist was asked to rate a number of physical aspects and services provided by the tourism industry. One of the problems encountered in the use of these variables was the fact that the MTA analyse this survey on quarterly basis. Therefore, the idea of analysing yearly data required a lot of recoding since for each quarter a different person used to input the data. This fact resulted in the problem that for each quarter there was a different codebook. Hence, considerable work had to be taken to ensure that the data of these four quarters was appropriately linked. This difficulty was particularly present in the recoding of the socio-demographic variables that were later used to obtain our relationship with the factor scores.

Another difficulty was the decision of which variables and respondents to consider for the analysis. This was so since in our data we had an amount of variables that had a large amount of missing values. By missing values, we understand that either the tourist did not respond or they did not use or have any opinion regarding that variable. After deciding on which variables to consider for our analysis, we considered each respondents response and decided that those respondents that had more than four missing values were removed.

The final stage of factor analysis and parceling was factor scores. Here we met with three possibilities. We had factor scores that when extracted were normally distributed. We also had factor scores that when transformed with some sort of transformation satisfied the normality conditions and lastly we had those factor scores that remained non-normal throughout. For the first two types, we continued our process but for the third type, further study was required.

The last part of the application focused on the use of general univariate linear models to obtain a relationship of these factor scores with the profile, namely the socio-demographic variables. We concluded that some factor scores were in fact linearly related to these variables. However, other factor scores lacked this linear relationship since there may have been other variables that could have been considered for our relationship.

7.2 Recommendations

From this study, I have acquired knowledge of the procedures involved when applying statistical theory to real life. In fact, the combination of theory and practice applied together is now more appreciated and absorbed for future reference. Without doubt, my work needs improvements and further readings would convince me to consider the following points

1. the application of other estimation techniques;
2. the possibilities of other linear relation methods;
3. the study of further analysis of non-normal scores;
4. the use of more variables to obtain a more knowledgeable model of the linear relationship.

My dissertation is based on four quarters collected over two years and it would be a motivating idea if future studies of this data were also considered on a yearly span rather than quarterly. In addition, it can be proposed as a new approach for the Malta Tourism Authority to analyse this data.

APPENDIX A

Theorem: Spectral Decomposition Theorem.

Any symmetric matrix $\mathbf{A}(p \times p)$ can be written as

$$\mathbf{A} = \Gamma\Lambda\Gamma^t = \sum \lambda_i \gamma_{(i)}\gamma_{(i)}^t$$

where Λ is a diagonal matrix of eigenvalues of A, and Γ is an orthogonal matrix whose columns are standardized eigenvectors.

Proof:

Let us assume that we can obtain the orthonormal vectors $\gamma_{(1)},\ldots,\gamma_{(p)}$ such that $\mathbf{A}\gamma_{(i)} = \lambda_i\gamma_{(i)}$ for some λ_i. Then

$$\gamma_{(i)}^t\mathbf{A}\gamma_{(i)} = \gamma_{(i)}^t\gamma_{(i)} = \begin{cases} \lambda_i, & i = j \\ 0, & i \neq j \end{cases}$$

or in matrix form

$$\Gamma^t\mathbf{A}\Gamma = \Lambda\Gamma^t\Gamma = \Lambda$$

Now, pre- and post-multiplying by Γ and Γ^t respectively gives

$$\Gamma\Gamma^t\mathbf{A}\Gamma\Gamma^t = \mathbf{A} = \Gamma\Lambda\Gamma^t$$

Therefore, the elements of Λ are the same as the eigenvalues of $\mathbf{A}$ with the same multiplicities.

Hence, we need to find an orthonormal basis of eigenvectors. If $\lambda_i \neq \lambda_j$ are distinct eigenvalues with eigenvectors $\mathbf{x}+\mathbf{y}$, respectively, then

$$\lambda_i \mathbf{x}^t\mathbf{y} = \mathbf{x}^t\mathbf{A}\mathbf{y} = \mathbf{y}^t\mathbf{A}\mathbf{x} = \lambda_i \mathbf{y}^t\mathbf{x}$$
$$\Rightarrow \mathbf{y}^t\mathbf{x} = 0$$

Therefore, for a symmetric matrix, eigenvectors corresponding to distinct eigenvalues are orthogonal to one another.

Assume that there exist k distinct eigenvalues of $\mathbf{A}$ with $H_1,\ldots,H_k$ corresponding eigenvectors of dimensions $r_1,\ldots,r_k$, hence, let $r = \sum_{j=1}^{k} r_j$.

Given that separate eigenspaces are orthogonal, there exists an orthonormal set of vectors $\mathbf{e}_1,\ldots,\mathbf{e}_r$ such that the vectors

$$\sum_{i=1}^{j-1} r_i + 1,\ldots,\sum_{i=1}^{j} r_i$$

form a basis of H_j. But r_j is less than or equal to the multiplicity of the corresponding eigenvalues. Thus, restructuring, if necessary, the eigenvalues λ_i, we may assume that

$$\mathbf{A}\mathbf{e}_i = \lambda_i \mathbf{e}_i,\ i = 1,\ldots,r \text{ and } r \leq p.$$

Now, consider when r = p, then substitute $\gamma_{(i)}$ by $\mathbf{e}_i$ and the proof is obtained.

Consider the case when $r < p$.
In this case, we obtain a contradiction and so this cannot be possible. Assuming that all the eigenvalues of $\mathbf{A}$ are strictly positive and setting

$$\mathbf{B} = \mathbf{A} - \sum_{i=1}^{r} \lambda_i \mathbf{e}_i \mathbf{e}_i^t$$

$$\Rightarrow \mathrm{tr}\mathbf{B} = \mathrm{tr}\mathbf{A} - \sum_{i=1}^{r} \lambda_i \left(\mathbf{e}_i^t \mathbf{e}_i\right) = \sum_{i=r+1}^{p} \lambda_i > 0$$

since r < p. Therefore **B** has at least one non-zero eigenvalue θ. Let $\mathbf{x} \neq \mathbf{0}$ be the corresponding eigenvector, then for $1 \leq j \leq r$

$$\theta \mathbf{e}_j^t \mathbf{x} = \mathbf{e}_j^t \mathbf{B} \mathbf{x} = \left\{ \lambda_j \mathbf{e}_j^t - \sum_{i=1}^{r} \lambda_i \left(\mathbf{e}_j^t \mathbf{e}_i \right) \mathbf{e}_i^t \right\} \mathbf{x} = 0$$

so that **x** is orthogonal to $\mathbf{e}_j$ for $j = 1, \ldots, r$.

$$\theta \mathbf{x} = \mathbf{B}\mathbf{x} = \left(\mathbf{A} - \sum \lambda_i \mathbf{e}_i \mathbf{e}_i^t \right) \mathbf{x} = \mathbf{A}\mathbf{x} - \sum \lambda_i \left(\mathbf{e}_i^t \mathbf{x} \right) \mathbf{e}_i = \mathbf{A}\mathbf{x}$$

implying that **x** is also an eigenvector of **A**. Consequently, $\theta = \lambda_i$ for some i and **x** is a linear combination of a number of the $\mathbf{e}_i$. This, contradicts the orthogonality between **x** and $\mathbf{e}_i$.

□

APPENDIX B

The following are the questions from MTA's traveller survey that have been considered in my thesis. The first two questions were used to obtain the tourist profile. The emphasis of my thesis was on the last question (question 12) which is measured on a Likert scale.

1. Where do you live?

City / Town: ____

Region / Province / State: ________________________________

Country: ____

2. Where are you in each of the following groups?

Gender

Male	☐
Female	☐

Marital Status

Single	☐
Married / Living together	☐
Divorced / Separated	☐
Widowed	☐

Occupation	Full time	Part time
Director	☐	☐
Manager	☐	☐
Office / Retail worker	☐	☐
Professional	☐	☐
Skilled worker !Tradesman	☐	☐
Manual Worker	☐	☐
Self-employed	☐	☐
Student	☐	☐
Housewife	☐	☐
Unemployed	☐	☐
Retired	☐	☐

What is your net income per month?

Amount ____________________

Currency ____________________

12. How would you rate the following aspects of your trip in Malta in terms of physical environment and service provided by employees?

1= Very Good

2= Good

3= Not So Good

4= Poor

5= Very Poor

N/A= Not Applicable

	Physical Aspect						Service Provided					
	Very good			Very poor			Very good			Very poor		
	1	2	3	4	5	N/A	1	2	3	4	5	N/A
Malta International Airport	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Gozo heliport	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Gozo ferry points	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Accommodation	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Restaurants	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Entertainment	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Car hire	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Taxi service	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Public Transport	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Retail outlets	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Sports facilities	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Beaches	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Attractions												
Guided Tours	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Historical Sites	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Museums	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Cathedrals / Churches	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Theatre / Performing arts	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Infrastructure

Road	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Road signs	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Traffic	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Parking	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Public Conveniences	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Level of cleanliness	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Air Quality	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Sea Quality	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

BIBLIOGRAPHY

Books

Anderson, T. W., *An Introduction To Multivariate Statistical Analysis.* New York: John Wiley & Sons, 1984.

Bartholomew, David J., *Latent Variable Models and Factor Analysis*. London: Charles Griffin & Company Limited, 1987.

Gorsuch, Richard L., *Factor Analysis.* New Jersey: Lawrence Erlbaum Associates, 1983.

Johnson, Richard A., and Dean W. Wichern, *Applied Multivariate Statistical Analysis.* New Jersey: Prentice Hall, 1998.

Krzanowski, W. J., *Principles of Multivariate Analysis.* New York: Oxford University, 1988.

Mardia, K. V., J. T. Kent, and J. M. Bibby. *Multivariate Analysis.* London: Academic Press, 1979.

Morrison, Donald F., *Multivariate Statistical Methods.* Singapore: McGraw-Hill Book Co., 1990.

Papers and Publications

Clason, Dennis L. and Thomas J. Dormody. "Analyzing Data measured by Individual Likert – Type Items." *Journal of Agricultural Education*, Vol. 35, No. 4.

Johns, Nick and Szilvia Gyimothy (2002). "Market Segmentation and the Prediction of Tourist Behaviour: The Case of Bornholm, Denmark." *Journal of Travel Research,* Vol. 40: 316 – 327.

Kozak, Metin and Mike Rimmington (2000). "Tourist Satisfaction with Mallorca, Spain, as an Off-Season Holiday Destination." *Journal of Travel Research,* Vol. 38: 260 – 269.

Lise, Wietze and Richard S. J. Tol (2002). "Impact of Climate on Tourism Demand." *Climatic Change,* 55: 429 – 449.

Little Todd D., William A. Cunningham, Golan Shahar and Keith F. Widaman (2002). "To Parcel or Not to Parcel: Exploring the Question, Weighing the Merits." *STRUCTURAL EQUATION MODELING, 9*(2): 151–173. Lawrence Erlbaum Associates, Inc.

Lubke Gitta and Bengt Muthen. "Factor-analyzing Likert-scale data under the assumption of multivariate normality complicates a meaningful comparison of observed groups or latent classes"

Mangion, Marie Louise, Ramwsh Durbarry, and M. Thea Sinclair. "Aids and Hedonic Pricing Models: Their Relevance to Tourism Practitioners."

Montfort, Kees van (2004). "Factor analyses for non-normal variables by fitting characteristic functions." *Kwantitatieve methoden,* 69: 27-42.

Nassar, Fadia and Joseph Wisenbaker (2003). "A Monte Carlo Study Investigating the Impact of Item Parceling on Measures of Fit in Confirmatory Factor Analysis." *Educational and Psychological Measurement,* Vol. 63 No. 5: 729-757.

Sonmex, Sevil and Ercan Sirakaya (2002). "A Distorted Destination Image? The Case of Turkey." *Journal of Travel Research,* Vol. 41: 185 – 196.

Steiger, James H. (1979). "Factor Indeterminacy in the 1930's and the 1970's Some Interesting Parallels." Psychometrika, Vol. 44 No. 1: 157 – 167.

Steiger James H. and Peter H. Schonemann. "A History of Factor Indeterminacy."

Web sites

www.statistics.gov.uk

(Home of official UK Statistics)