University of Benghazi
Faculty of Education Almarj

جامعة بنغازي
كلية التربية – المرج
ISSN 2518-5845

Global Libyan Journal

المجلة الليبية العالمية

العدد السبعون / ديسمبر / 2023

Use Of Factor Analysis And Discriminant Analysis To Determine The Factors Affecting Climate Type And Building Shape In Libya

Naeimah A. ABDULNABI
Computer Department
Faculty Of Arts and Science Salouq
University Of Benghazi

YASMINA B. EL – FIGIH
Department Of Statistics
Faculty Of Science
University Of Benghazi

naemaali1982@gmail.com

Use Of Factor Analysis And Discriminant Analysis To Determine The Factors Affecting Climate Type And Building Shape In Libya

Abstract

Climate in simple concept commonly known as the rate of the weather , there are many observed elements of the atmosphere that are subject to constant change. These elements generally are: Maximum temperature(X_1), Minimum temperature (X_2), Relative humidity(X_3), Wind speed(X_4), Wind direction(X_5), Duration of sun shine(X_6), Clouds amount(X_7) and Rainfall amount(X_8).The main objectives of the study are: (i) To define the factors influencing the Climate, and to determine the most important factors of the climatic variables.(ii) To discriminate the Climate by region and define the factors which responsible for that discrimination. (iii) To determine the factors that influencing in the building form.

To achieve these goals, a simple random sample of ten cities were selected from the Libyan Meteorological Department in Tripoli. These cities are divided into (6) cities as a coastal climate and (4) cities as a desert climate. The data of this study were collected as a series from each city from January of the first year 1970 to December of the last year 2000. Each series consists of 12×30×30×10 monthly observations, and the total data points in the study will be (108,000) from Derna, Benghazi, Jalu, Agedabia, Tripoli, Misurata, Sebha, EL-Kufra, Ghadames and Shahat. To study the variability of climate system of these places needs to explain the observed correlations between the elements and the regions.

Two different statistical techniques are used in this study to analyze the data to study the first objective, factor analysis is used to determine the importance of the variables and their effect on the climate in both coastal and desert climate in general, and for each city separately. To study the second and third objectives discriminant analysis is used. This study provides a number of variables with significant positive effect on the Type of Climate variable(X_9), in the coastal cities, (Derna, Benghazi, Agedabia, Tripoli, Misurata, Shahat) Maximum Temperature(X_1), Duration Of Sun Shine(X_6),Clouds Amount(X_7) and Rainfall Amount (X_8) are good representative for all eight original variables. But in the desert cities (Sebha, Jalu, EL-Kufra, Ghadames) the variables Maximum temperature(X_1), Minimum temperature(X_2), Relative humidity(X_3), Wind speed(X_4), Duration of sun shine(X_6), Clouds amount(X_7) and

Rainfall amount(X_8) all these variables have significant positive effect on the Type of Climate variable(X_9).

The results from discriminant and classification analysis for the dependent variable Type of Climate (X_9) show that: There is significant difference between the climate of ten regions. Where, Maximum Temperature(X_1), Relative humidity(X_3), Duration of Sun Shine(X_6), Clouds Amount(X_7) and Rainfall Amount(X_8) are significant and important, and the Relative humidity(X_3) have high correlation with the discriminant function. The results from discriminant and classification analysis for the dependent variable Type of Building(X_{10}) show that: There is significant difference between the building form of the ten regions. Where, Maximum Temperature(X_1),Relative humidity(X_3), Duration of Sun Shine(X_6) and Clouds Amount(X_7) have high correlation with the discriminant function or these variables are important and significant and have an effect on the building form in each city.

Key words : Discriminate Analysis, Classification, Climate, Factor Analysis.

استخدام التحليل العاملي والتحليل التصنيفي لتحديد العوامل المؤثرة في نوع المناخ و شكل المبنى في ليبيا

نعيمة علي عبدالنبي
قسم الحاسوب– كلية الآداب و العلوم
جامعة بنغازي فرع سلوق

ياسمينة بوزيد الفقيه
قسم الإحصاء – كلية العلوم
جامعة بنغازي

الملخص :

المناخ بمفهوم بسيط يعرف عادة باسم معدل الطقس ، هناك العديد من العناصر الملحوظة في الغلاف الجوي والتي تخضع لتغير مستمر. هذه العناصر بشكل عام هي: درجة الحرارة القصوى X_1، ودرجة الحرارة الدنيا X_2، والرطوبة النسبية X_3، وسرعة الرياح X_4، واتجاه الرياح X_5، ومدة سطوع الشمس X_6، وكمية السحب X_7وهطول الأمطار الكمية X_8. الأهداف الرئيسية للدراسة هي: (1) تحديد العوامل التي تؤثر على المناخ ، وتحديد أهم عوامل المتغيرات المناخية. (2) تصنيف المناخ حسب المنطقة وتحديد العوامل المسئولة عن هذا التصنيف. (3) تحديد العوامل التي تؤثر على شكل المبنى.

ولتحقيق هذه الأهداف ، تم اختيار عينة عشوائية بسيطة من عشر مدن من دائرة الأرصاد الليبية بطرابلس. هذه المدن مقسمة إلى (6) مدن ذات مناخ ساحلي و (4) مدن ذات مناخ صحراوي. جمعت بيانات هذه الدراسة كسلسلة من كل مدينة من يناير من العام الأول1970 إلى ديسمبر من العام الأخير 2000. تتكون كل سلسلة من 12×30 ملاحظة شهرية ، على مدى 30 عام لعشر مدن وسيكون مجموع نقاط البيانات في الدراسة (108,000) من : درنة ، بنغازي ، جالو ، أجدابيا ، طرابلس

، مصراته ، سبها ، الكفرة ، غدامس ، شحات. لدراسة التغير في النظام المناخي لهذه المناطق ، نحتاج توضيح الارتباطات الملحوظة بين العناصر والمناطق.

تم استخدام طريقتين إحصائيتين مختلفتين في هذه الدراسة لتحليل البيانات : لدراسة الهدف الأول تم استخدام التحليل العاملي لتحديد أهمية المتغيرات وتأثيرها على المناخ في كل من المناخ الساحلي والصحراوي بشكل عام ، ولكل مدينة على حدة. لدراسة الهدف الثاني والثالث تم استخدام التحليل التمييزي. تقدم هذه الدراسة عددا من المتغيرات ذات التأثير الإيجابي المعنوي على نوع المناخ المتغير X_9 في المدن الساحلية (درنة ، بنغازي ، أجدابيا ، طرابلس ، مصراتة ، شحات) حيث أظهرت نتائج التحليل العاملي أن: درجة الحرارة القصوى X_1، مدة سطوع الشمس X_6كمية السحب X_7 و كمية الأمطارX_8 تمثل جيدًا المتغيرات الثمانية الأصلية. أما في المدن الصحراوية (سبها ، جالو ، الكفرة ، غدامس) متغيرات : درجة الحرارة العظمى X_1، درجة الحرارة الصغرى X_2، الرطوبة النسبية X_3، سرعة الرياح X_4، مدة سطوع الشمس X_6، كمية السحب X_7 وكمية الأمطار X_8 كل هذه المتغيرات لها تأثير إيجابي كبير على نوع المناخ المتغيرX_9 .

تظهر نتائج التحليل التمييزي والتصنيفي للمتغير التابع نوع المناخ X_9 أن: هناك فرق كبير بين مناخ العشرة مناطق. حيث تكون درجة الحرارة القصوى X_1، والرطوبة النسبية X_3، ومدة سطوع الشمس X_6، و كمية السحب X_7 ومقدار هطول الأمطار X_8 ذات دلالة معنوية ومهمة ، والرطوبة النسبية X_3 لها ارتباط كبير مع دالة التمييز. تظهر نتائج التحليل التمييزي والتصنيفي للمتغير التابع نوع المبنى X_{10} أن: هناك فرق معنوي بين شكل المبنى للعشر مناطق. حيث أن درجة الحرارة القصوى X_1والرطوبة النسبية X_3 ومدة سطوع الشمس X_6 وكمية السحب X_7 لها ارتباط كبير بالدالة التمييزية أي أن هذه المتغيرات مهمة وذات دلالة معنوية ولها تأثير على شكل المبنى في كل مدينة.

الكلمات المفتاحية: التحليل التمييزي ، التصنيف ، المناخ ، التحليل العاملي.

INTRODUCTION

The climates prevailing around the globe vary greatly, ranging from the polar extreme to tropical climates. These are primarily influenced by the sun's energy heating up the land and water masses. At the regional level, the climate is influenced by altitude, topography, patterns of wind and ocean currents, the relation of land to water masses, the geomorphology, and by the vegetation pattern. Accordingly, the tropical and subtropical regions can be divided into many different climatic zones, but for practical reasons, in this publication three main climate zones are considered:

1. The hot-arid zone, including the desert or semi desert climate and the hot-dry maritime climate.
2. The warm-humid zone, including the equatorial climate and the warm-humid island climate.
3. The temperate zone, including the monsoon climate and the tropical upland zone.

The main climatic factors relevant to construction are those affecting human comfort:

1. Air temperature, its extremes and the difference between day and night, and between summer and winter temperatures.
2. Humidity and precipitation.
3. Incoming and outgoing radiation and the influence of the sky condition.
4. Air movements and winds.[1]

The architecture started since ancient times to comply one of the basic needs of the human being, it represented the shelter for him, was spontaneous constantly changing in order to provide space appropriate to exercise its activity where away from what could be troublesome or harmful than surrounded by the environment; therefore inherent in the development rights of vacuum which adapts him to exercise the activity with dealing with the environmental conditions surrounding access to the most comfort-able space.From here began architecture, which was carrying apart styles according to the region in which they appear (depending on the privacy environmental of the area). Then began the circumstances and social needs, ideas, and ideological and cultural needs of human. After the evolution of architecture methods ; became what simulates human nature and respect his conditions,

thoughts and beliefs, and conformity with what surrounded environmental conditions provide a more comfortable vacuum.[2]

Climate is the long-term weather zone or region for more than 30 years old or so, quite simply is the average temperature and precipitation for a period of time, and this includes the amount of sun in the region, and the rate of wind speed, amount of rainfall each year, and the state of extreme weather with, and local geography of the region.

Climate simple concept commonly known as "the rate of the weather," or more precisely that the statistical description for the average and fluctuation appropriate amounts through a period of time ranging from months to thousands or millions of years The traditional term is 30years old, as defined by the World Meteorological Organization (WMO), that these statistics are often in a superficial variables such as the degree of heat and rain and wind. A wider range of climate is a case containing a description of the statistical system of the climate.

The building must be adapted with the climate and its different components, in the moment that ends the construction becomes part of the environment, like a tree or a stone, and it becomes an exposed to the same effects of the sun or rain or wind like any something in the environment, if the building be able to face the climate problems and at the same time using all climatic and natural resources available in order that to achieve human comfort inside the building, this building can be called " a climatic balanced building " .

Many contemporary buildings ignored climate and its factors and dominated crust glass on this buildings, and the home to go to the outside instead of inside and exposed openings to direct sunlight, the glass openings and surfaces regard as the main source to entry of the heat into the building, thus glass increases the force of the heat to the interior by far exceeds the force that occurs during the dark surfaces , Libya, generally considered a country with a desert climate and sun protection inside and outside of the building is desirable all the year. The expected population increases calls for architects and construction workers to interest by application of the process of designing buildings in a manner consistent with the prevailing climate in each area and taking into account to reduce the consumption of energy , with minimizing the effects of the construction and the use on the environment and maximizing

the harmony with nature. This study try to introduce the factors which have a significance effect on the climate and on the managed of the building design in general or the factors that influencing the building form.

Objectives of The Study

The main objectives of this study are :

1- To define the factors influencing the Climate and to determine the most important factors of the climatic variables.

2- To discriminate the Climate by region and define the factors which responsible for that discrimination.

3- To determine the factors that influencing the building form.

Hypotheses of The Study

The main hypotheses of this study are :

i. There is two nonsingular groups covariance matrices $\sum_1 \neq \sum_2$.
ii. There is no relationship between the Climatic Factors and the Type of Climate .
iii. There is no relationship between the Climatic Factors and the Type of Building .

Source Of Data and Sample Selection

The data of this study selected from the Libyan Meteorological Department, Tripoli(1970-2000), National Geophysical Data Center(2004), and World Meteorological Organization (2005). [3]

A simple random sample of ten Libyan cities has been selected. These cities are : Derna, Benghazi, Jalu, Agedabia, Tripoli, Misurata, Sebha, EL-Kufra, Ghadames and Shahat. The data of this study collected series of each city from January of the first year1970 to December of the last year2000. Each series consists of 12×30×30×10 monthly observations, and the total data points in the study will be (108,000).

Description Of Variables

The number of variables in this study is 10 [Maximum temperature(X_1), Minimum Temperature (X_2), Relative Humidity (X_3), Wind Speed (X_4), Wind Direction (X_5), Duration

Of Sun Shine (X_6), Clouds Amount (X_7) and Rainfall Amount (X_8), Type of Climate(X_9), Type of Building (X_{10})]. Some of these variables are qualitative and most of them are quantitative. These variables are descript in brief below :

Maximum temperature(X_1), Minimum temperature(X_2), Relative humidity (X_3), Wind speed(X_4), Duration of Sunshine (X_6), Clouds Amount (X_7), Rainfall Amount(X_8): These variables are quantitative so they are included in the analysis directly .

Wind direction(X_5):This variable is qualitative and has been included in the analysis by coding, one for north, and two for east, and three for south, and four for west, and fife for northeast, and six for northwest, and seven for southeast, and eight for southwest .

Type of Climate(X_9):This variable is qualitative and has been included in the analysis by coding, zero for Desert Climate, and one for otherwise .

Type of Building(X_{10}):This variable is qualitative and has been included in the analysis by coding , zero for Prevalent Building, and one for Distinctive Building.

METHODOLOGY

To achieve the objectives of the study, two different statistical techniques were used, namely, factor analysis " data or variables reduction ", and discriminant analysis "classification of variables ".

In factor analysis the dependent variables represented as linear combinations of a few independent random variables called factors, where the number of the factors to be less than from the number of independent variables. The factors are underlying Constructs or Latent variables that “generate” the dependent variables. Like the original variables, the factors vary from individual to individual; but unlike the variables, the factors cannot be measured or observed. The existence of these hypothetical variables is therefore open to question. [4]

Discriminant analysis is used to identify: 1) the factors which have effect on the Type of Climate, (i.e., classifying Climate as to determined region) , 2) the factors which have effect on the Type of buildings, (i.e., classifying building as to climate of determined region).

Discriminant analysis involves deriving a variate, the discriminant variate is the linear combination of the two (or more) independent variables that will discriminant best between the variables(climatic factors) in the groups defined a priori Discrimination is achieved by calculating the variety's weights for each independent variable to maximize the differences between the groups (i.e., the between– group variance relative to the within – group variance). The variate for a discriminant analysis, also known as the discriminant function. [5]

1. Factor Analysis

Factor analysis is a multivariate technique which attempts to account for the correlation pattern present in the distribution of an observable random vector in terms $X = (X_1,...,X_p)^T$ of a minimal number of unobservable random variables, called factors. In this approach each component X is examined to see if it could be generated by a linear function involving a mini-mum number of unobservable random variables, called common factor variates, and a single variable, called the specific factor variate.[6]

A frequently applied paradigm in analyzing data from multivariate observations is to model the relevant information (represented in a multivariate variable X) as coming from a limited number of latent factors. we assume that there is a model (it will be called the "Factor Model") stating that most of the covariance's between the P elements of X can be explained by a limited number of latent factors. Factor analysis is interest in many fields such as psychology, marketing, economics , politic sciences, etc .

The aim of factor analysis is to explain the outcome of p variables in the data matrix X using fewer variables, the so-called factors. Ideally all the information in X can be reproduced by a smaller number of factors. These factors are interpreted as latent (unobserved) common characteristics of the observed $X \in R^p$. The case just described occurs when every observed $X = (X_1,...,X_p)^T$ can be written as

$$X_i = \sum_{j=1}^{m} \ell_{ij} f_j + \mu_i \quad , \; i = 1,...,p \qquad (1.1)$$

Here f_j , j = 1,...,p denotes the factors. The number of factors, m should always be much smaller than p . For instance, in psychology X may represent p results of a test measuring intelligence scores. One common latent factor explaining $X \in R^p$ could be the overall level of "intelligence" . [7]

1.1. The Orthogonal Factor Model

The observable random vector X ; with p components has mean μ and covariance matrix Σ the factor model postulates that X is linearly dependent upon a few unobservable random variables $F_1,...,F_m$, called common factors and p additional sources of variation $\varepsilon_1,...,\varepsilon_p$, called errors , the factor analysis model is

$$\begin{aligned} X_1 - \mu_1 &= \ell_{11}F_1 + \ell_{12}F_2 + ... + \ell_{1m}F_m + \varepsilon_1 \\ X_2 - \mu_2 &= \ell_{21}F_1 + \ell_{22}F_2 + ... + \ell_{2m}F_m + \varepsilon_2 \qquad (1.2) \\ &\vdots \qquad\qquad\qquad \vdots \\ X_p - \mu_p &= \ell_{p1}F_1 + \ell_{p2}F_2 + ... + \ell_{Pm}F_m + \varepsilon_p \end{aligned}$$

Or in matrix notation

$$\underset{(p\times 1)}{X - \mu} = \underset{(p\times m)}{L}\ \underset{(m\times 1)}{F} + \underset{(p\times 1)}{\varepsilon} \qquad (1.3)$$

The coefficient ℓ_{ij} is called the loading of the i-th variable on the j-th factor, so the matrix L is the matrix of factor loadings. Note that the i-th specific factor ε_i is associated only with the i-th response X_i . The p deviations $X_1 - \mu_1, X_2 - \mu_2, ..., X_p - \mu_p$ are expressed in terms of p+m random variables $F_1, F_2, ..., F_m, \varepsilon_1, \varepsilon_2, ..., \varepsilon_p$ which are unobservable. This distinguishes the factor model of (1.3) in which the independent variables [whose position is occupied by F in (1.3)] can be observed.

With so many unobservable quantities, a direct verification of the factor model from observations on X_1 , X_2 ,...,X_p is hopeless. However, with some additional assumptions about the random vectors F and ε , the model in (1.3) implies certain covariance relationships, can be checked. Assume that

$$E(F) = 0, \quad E(\varepsilon) = 0, \quad \text{cov}(F) = E[FF^T] = I_{(m\times m)}$$

And

$$\text{cov}(\varepsilon) = E[\varepsilon\varepsilon^T] = \psi_{(p\times p)} = \begin{bmatrix} \psi_1 & 0 & \dots & 0 \\ 0 & \psi_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \psi_p \end{bmatrix} \qquad (1.4)$$

These assumptions and the relation in (1.3) constitute the orthogonal factor model, the orthogonal factor model implies a covariance structure for X from the model in (1.3)

$$\begin{aligned}(X-\mu)(X-\mu)^T &= (LF+\varepsilon)(LF+\varepsilon)^T \\ &= (LF+\varepsilon)((LF)^T+\varepsilon^T) \\ &= LF(LF)^T+\varepsilon(LF)^T+LF\varepsilon^T+\varepsilon\varepsilon^T\end{aligned}$$

Also

$$\begin{aligned}\Sigma = \text{cov}(X) &= E(X-\mu)(X-\mu)^T \\ &= LE(FF^T)L^T+E(\varepsilon F^T)L^T+LE(F\varepsilon^T)+E(\varepsilon\varepsilon^T) \\ &= LL^T+\psi\end{aligned}$$

According to (1.4). Also by independence $\text{cov}(\varepsilon, F) = E(\varepsilon, F^T) = 0$ then

$$(X-\mu)F^T = (LF+\varepsilon)F^T = LFF^T+\varepsilon F^T$$
$$\text{cov}(X,F) = E(X-\mu)F^T = LE(FF^T)+E(\varepsilon F^T) = L$$

1.2. Covariance Structure for the Orthogonal Factor Model

1. $\text{cov}(X) = LL^T+\psi$

or $\quad \text{var}(X_i) = \ell_{i1}^2+\dots+\ell_{im}^2+\psi_i$

$$\text{cov}(X_i, X_k) = \ell_{i1}\ell_{k1}+\dots+\ell_{im}\ell_{km} \qquad (1.5)$$

2. $\text{cov}(X,F) = L$

or $\quad \text{cov}(X_i, F_j) = \ell_{ij}$

That portion of the variance of the i-th variable contributed by the m common factors is called the i-th communality. That portion of $\text{var}(X_i) = \sigma_{ii}$ due to the specific factor is often called the specific variance. Denoting the i-th communality by h_i^2, from (1.5)

$$\underset{\text{var}(X_i)}{\sigma_{ii}} = \underbrace{\ell_{i1}^2 + \ell_{i2}^2 + \ldots + \ell_{im}^2}_{\text{communality}} + \underset{\text{specific variance}}{\psi_i}$$

or
$$h_i^2 = \ell_{i1}^2 + \ell_{i2}^2 + \ldots + \ell_{im}^2 \qquad (1.6)$$

and

$$\sigma_{ii} = h_i^2 + \psi_i \qquad ,i = 1,2,\ldots,p$$

The i-th communality is the sum of squares of the loadings of the i-th variable on the m common factors. [8]

1.3. Oblique Factor Model

This is obtained from the orthogonal factor model by replacing $\text{cov}(F) = I$ by $\text{cov}(F) = R$, where R is a positive definite correlation matrix; that is, all its diagonal elements are equal to unity. In other words, all factors in the oblique factor model are assumed to have mean **0** and variance **1** but are correlated, in this case $\Sigma = LRL^T + \psi$. [6]

To choosing the number of factors (m), several criteria have been proposed. consider four criteria for choosing the number of principal components to retain.

1. Choose m equal to the number of factors necessary for the variance accounted for to achieve a predetermined percentage, say 80%, of the total variance tr(**S**) or tr(**R**).

2. Choose *m* equal to the number of eigenvalues greater than the average For **R** the average is **1**; for **S** it is $\sum_{j=1}^{p} h_i^2 / p$.

3. Use the scree test based on a plot of the eigenvalues of **S** or **R**. If the graph drops sharply, followed by a straight line with much smaller slope, choose m equal to the number of eigenvalues before the straight line begins.

4. Test the hypothesis that m is the correct number of factors, $H_0 : \Sigma = LL^T + \Psi$ where L is (p×m). [4]

1.3.1. Test Of Hypothesis In Factor Models

Let $X^{\alpha} = (X_{\alpha 1},...,X_{\alpha p})^T$, $\alpha = 1,...,N$ be a sample of size **N** from a p-variate normal population with positive definite covariance matrix Σ . On the basis of these observations we are interested in testing, with the orthogonal factor model. The null hypothesis $H_0 : \Sigma = LL^T + \psi$ against the alternatives H_1 that Σ is a symmetric positive definite matrix (The corresponding hypothesis in the oblique factor model is $H_0 : \Sigma = LRL^T + \psi$). Rejects H_0 whenever , with N – 1 = n

$$\lambda = \left[\frac{\det(s/N)}{\det(LL^T + \psi)}\right]^{-n/2} \exp\left\{\frac{1}{2}\mathrm{tr}(LL^T + \psi)^{-1}s - \frac{1}{2}np\right\} \geq C \qquad (1.7)$$

Where $\mathrm{dig}(LL^T + \psi) = \mathrm{dig}(\frac{s}{N})$, $(\frac{s}{N}) = \Sigma = \sum_{\alpha=1}^{N}(X^{\alpha} - \bar{X})(X^{\alpha} - \bar{X})^T N$

and **C** depends on the level of significance α of the test. In large samples under H_0 , using

$$P\{-2\log\lambda \leq Z\} = P\{X_f^2 \leq Z\}$$

where

$$f = \frac{1}{2}p(p+1) - [mp + p - \frac{1}{2}m(m+1) + m] \qquad (1.8)$$

The modification needed for the oblique factor model is obvious and the value of degrees of freedom f for the chi-square approximation in this case is

$$f = \frac{1}{2}p(p - 2m + 1) \qquad (1.9)$$

has pointed out that if N-1= n is replaced by n_0, where

$$n_0 = n - \frac{1}{6}(2p+5) - \frac{2}{3}m \qquad (1.10)$$

then under H_0, the convergence of $-2\log\lambda$ to chi-square distribution is more rapid. [6]

1.3.2. Method of Estimation

Let $\sum$ have eigenvalue-eigenvector pairs (λ_i, e_i) with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$.

Then

$$\sum = \lambda_1 e_1 e_1^T + \lambda_2 e_2 e_2^T + \ldots + \lambda_p e_p e_p^T$$

$$= \left[\sqrt{\lambda_1}e_1, \sqrt{\lambda_2}e_2, \ldots, \sqrt{\lambda_p}e_p\right] \begin{bmatrix} \sqrt{\lambda_1}e_1 \\ , \\ \sqrt{\lambda_2}e_2 \\ , \\ \vdots \\ , \\ \sqrt{\lambda_p}e_p \end{bmatrix} = \underset{(p\times m)}{L}\ \underset{(m\times p)}{L^T}$$

This approximate assumes that the specific factors ε in (1.3) are of minor importance and can also be ignored in the factoring of $\sum$, If specific factors are included in the model, their variances may be taken to be the diagonal elements, so we find this approximation becomes

$$\sum = LL^T + \psi$$

$$= \left[\sqrt{\lambda_1}e_1, \sqrt{\lambda_2}e_2, \ldots, \sqrt{\lambda_p}e_p\right] \begin{bmatrix} \sqrt{\lambda_1}e_1 \\ , \\ \sqrt{\lambda_2}e_2 \\ , \\ \vdots \\ , \\ \sqrt{\lambda_p}e_p \end{bmatrix} + \begin{bmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi p \end{bmatrix}$$

Where $\psi_i = \sigma_{ii} - \sum_{j=1}^{m} \ell_{ij}^2$ for $i = 1, 2, \ldots, p.$ [8]

1.3.3. Estimation Of Loadings And Communalities

By use an initial estimate ψ and factors $S - \psi$ or $R - \psi$ to obtain

$$S - \psi \cong LL^T \qquad (1.11)$$

$$R - \psi \cong LL^T \qquad (1.12)$$

Where L is (p×m) and is calculated as

$$L = (\sqrt{\lambda_1}\hat{e}_1, \sqrt{\lambda_2}\hat{e}_2, \ldots, \sqrt{\lambda_p}\hat{e}_p) \qquad (1.13)$$

Therefore $(\lambda_i, \hat{e}_i), i = 1, 2, \ldots, p$ define as the (largest) eigenvalue eigenvector pairs determined from R. The i-th diagonal element of $S - \psi$ is given by $s_{ii} - \psi_i$, which is the i-th communality $h_i^2 = s_{ii} - \psi_i$. Likewise, the diagonal elements of $R - \psi$ are the Communalities $h_i^2 = 1 - \psi_i$ (clearly , ψ_i and h_i^2 have different values for S than for R). With these diagonal values, $S - \psi$ and $R - \psi$ have the form

$$S - \psi = \begin{pmatrix} h_1^2 & s_{12} & \cdots & s_{1p} \\ s_{21} & h_2^2 & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & h_p^2 \end{pmatrix}, \quad R - \psi = \begin{pmatrix} h_1^2 & r_{12} & \cdots & r_{1p} \\ r_{21} & h_2^2 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & h_p^2 \end{pmatrix} \qquad (1.14)$$

A popular initial estimate for a communality in $R - \psi$ is $h_i^2 = R_i^2$ the squared multiple correlation between F_i and the other p-1 variables. This can be found as

$$h_i^2 = R_i^2 = 1 - \frac{1}{r^{ii}} \qquad (1.15)$$

Where r^{ii} is the i-th diagonal element of R^{-1}.

For $S-\psi$, an initial estimate of communality analogous to (1.15) is

$$h_i^2 = s_{ii} - \frac{1}{s^{ii}} \qquad (1.16)$$

Where S_{ii} is the i-th diagonal element of S and S^{ii} is the i-th diagonal element of S^{-1}. It can be shown that (1.16) is equivalent to

$$h_i^2 = s_{ii} - \frac{1}{s^{ii}} = s_{ii}R_i^2 \qquad (1.17)$$

which is a reasonable estimate of the amount of variance that F_i has in common with the other F's.

To use (1.15) or (1.16), R or S must be nonsingular. If R is singular, we can use the absolute value or the square of the largest correlation in the i-th row of R as an estimate of communality. After obtaining communality estimates, we calculate eigenvalues and eigenvectors of $S-\psi$ or $R-\psi$ and use (1.13) to obtain estimates of factor loadings L. Then the columns and rows of L can be used to obtain new eigenvalues (variance explained) and communalities, respectively. The sum of squares of the j-th column of L is the j-th eigenvalue of $S-\psi$ or $R-\psi$ and the sum of squares of the i-th row of L is the communality of F_i. The proportion of variance explained by the j-th factor is

$$\frac{\lambda_j}{tr(S-\psi)} = \frac{\lambda_j}{tr(R-\psi)} = \frac{\lambda_j}{\sum_{i=1}^{p}\lambda_i}$$

Where λ_j is the j-th eigenvalue of $S-\psi$ or $R-\psi$. The matrices $S-\psi$ and $R-\psi$ are not necessarily positive semi definite and will often have some small negative eigenvalues. In such a case, the cumulative proportion of variance will exceed 1, and then decline to 1 as the negative eigenvalues are added. (Note that loadings cannot be obtained by(1.13) for the negative eigenvalues). [4]

2. Discriminant Analysis

1- The basic idea of discriminant analysis consists of assigning an individual or a group of individuals to one of several known or unknown distinct populations, on the basis of observations on several characters of the individual or the group and a sample of observations on these characters from the populations if these are unknown. In scientific literature, discriminant analysis has many synonyms, such as classification, pattern recognition, character recognition, identification, prediction, and selection, depending on the type of scientific area in which it is used. [6]

2- Discriminant analysis is used in situations where the clusters are known **a priori**. The aim of discriminant analysis is to classify an observation, or several observations, into these known groups. [7]

2.1. Some Applications Of Discriminant Analysis

I. On a patient with a diagnosis of myocardial infarction, Observations on his systolic blood pressure (X_1), diastolic blood pressure (X_2), heart rate, (X_3) stroke index (X_4), and mean arterial pressure (X_5) are taken. On the basis of these observations it is possible to predict whether or not the patient will survive.

II. In developing a certain rural area a question arises regarding the best strategy for this area to follow in its development. This problem can be considered as one of the problems of discriminant analysis. For example, the area can be grouped as catering to recreation users or attractive to industry by means of variables such as distance to the nearest city (X_1), distance to the nearest major airport (X_2), percentage of land under lakes (X_3) , and percentage of land under forests (X_4) .

III. Admission of students to the state-supported medical program on the basis of examination marks in mathematics (X_1), physics (X_2), Chemistry (X_3), English (X_4), and bioscience (X_5)is another example of discriminant analysis. [6]

2.2. Fisher's Linear Discriminante Function

Fisher's idea was to base the discriminant rule on a projection a^TX such that a good separation was achieved. This Linear Discrimination analysis projection method is called Fisher's Linear Discrimination Function if $Y = X a$

denotes a linear combination of observations, then the total sum of squares of $Y, \sum_{i=1}^{n}(Y_i - \overline{Y})^2$, is equal to

$$Y^T Y = a^T X X a = a^T T a \qquad (2.1)$$

and $T = X^T X$

Suppose there are samples X_j , $j = 1, \ldots, J$ from J populations.

Fisher's suggestion was to find the linear combination $a^T X$ which maximizes the ratio of the between-group-sum of squares to the within-group-sum of squares. The within-group-sum of squares is given by

$$\sum_{j=1}^{J} Y_j^T Y_j = \sum_{j=1}^{J} a^T X_j^T X_j a = a^T W a \qquad (2.2)$$

Where Y_j denotes the j-th sub-matrix of Y corresponding to observations of group j and the coefficient vector $\hat{a} = S_{pooled}^{-1}(\overline{X}_1 - \overline{X}_2)$. The within-group-sum of squares measures the sum of variations within each group. The between-group-sum of squares is

$$\sum_{j=1}^{J} n_j(\overline{Y}_j - \overline{Y})^2 = \sum_{j=1}^{J} n_j\{a^T(\overline{X}_j - \overline{X})\}^2 = a^T B a \qquad (2.3)$$

Where $\overline{Y}_j$, $\overline{X}_j$ denote the means of Y_j and X_j and $\overline{Y}$, $\overline{X}$ denote the sample means of X and Y . The between-group-sum of squares measures the variation of the means across groups. The total sum of squares (2.1) is the sum of the within-group-sum of squares and the between-group-sum of squares, i.e.,

$$a^T T a = a^T W a + a^T B a$$

Fisher's idea was to select a projection vector *a* that maximizes the ratio

$$\frac{a^T B a}{a^T W a} \qquad (2.4)$$

The vector a that maximizes (2.4) is the eigenvector of $W^{-1}B$ that corresponds to the largest eigenvalue. Now a discrimination rule is easy to obtain:

Classify X into group j where $a^T \overline{X}_j$ is closest to $a^T X$, i.e.,

$$X \to \Pi_j \quad \text{where} \quad j = \min_j \left| a^T (X - \bar{X}_j) \right|$$

When $J = 2$ groups, the discriminant rule is easy to compute.

Suppose that group 1 has n_1 elements and group 2 has n_2 elements. In this case

$$B = \left(\frac{n_1 n_2}{n_1} \right) dd^T, \quad d = (\bar{x}_1 - \bar{x}_2).$$

$W^{-1}B$ has only one eigenvalue which equals

$$tr\left(W^{-1}B\right) = \left(\frac{n_1 n_2}{n_1} \right) d^T W^{-1} d$$

and the corresponding eigenvector is $a = W^{-1}d$. The corresponding discriminant rule is

$$X \to \Pi_1 \qquad \text{if} \quad a^T \{ X - \frac{1}{2}(\bar{X}_1 + \bar{X}_2) \} > 0$$

(2.5)

$$X \to \Pi_2 \qquad \text{if} \quad a^T \{ X - \frac{1}{2}(\bar{X}_1 + \bar{X}_2) \} \le 0$$

2.3. Allocation Rule for Known Distributions

In general we have populations Π_j, $j = 1, 2, ..., J$ and we have to allocate an observation X to one of these groups. A discriminant rule is a separation of the sample space (in general R^p) into sets R_j such that if $X \in R_j$, it is identified as a member of population Π_j. The main task of discriminant analysis is to find "good" regions R_j such that the error of misclassification is small, suppose the densities of each population Π_j by $f_j(x)$. Then to allocating X to Π_j maximizing $L_j(x) = f_j(x) = \max_i f_i(x)$.

If several f_i give the same maximum then any of them may be selected. Mathematically, the sets R_j are defined as:

$$R_j = \{ x : L_j(x) > L_i(x) \text{ for } i = 1, ..., J, i \neq j \}. \quad (2.6)$$

By classifying the observation into a certain group we may encounter a misclassification error.

For $J=2$ groups the probability of putting X into group 2 although it is from population 1 can be calculated as

$$p_{21} = P(X \in R_2 \setminus \Pi_1) = \int_{R_2} f_1(x)dx. \qquad (2.7)$$

Similarly the conditional probability of classifying an object as belonging to the first population Π_1 although it actually comes from Π_2 is

$$p_{12} = P(X \in R_1 \setminus \Pi_2) = \int_{R_1} f_2(x)dx. \qquad (2.8)$$

The misclassified observations create a cost $C(i/j)$ when a Π_j observation is assigned to R_i . The cost structure can be pinned down in a cost matrix:

	Classified	Population
	Π_1	Π_2
True population Π_1	0	$C(2\|1)$
True population Π_2	$C(1\|2)$	0

Let π_j be the prior probability of population Π_j, where "prior" means the a prior probability that an individual selected at random belongs to Π_j (i.e., before looking to the value X).

The Expected Cost of Misclassification (ECM) is given by

$$ECM = C(2\setminus 1)p_{21}\pi_1 + C(1\setminus 2)p_{12}\pi_2. \qquad (2.9)$$

For two given populations, the rule minimizing the ECM is given by

$$R_1 = \left\{ x : \frac{f_1(x)}{f_2(x)} \geq \left(\frac{C(1\setminus 2)}{C(2\setminus 1)}\right)\left(\frac{\pi_2}{\pi_1}\right) \right\} \qquad (2.10)$$

$$R_2 = \left\{ x : \frac{f_1(x)}{f_2(x)} < \left(\frac{C(1\setminus 2)}{C(2\setminus 1)}\right)\left(\frac{\pi_2}{\pi_1}\right) \right\}$$

Thus a special case of the ECM rule for equal misclassification costs and equal prior probabilities. For simplicity the unity cost case, C(1|2) = C(2|1) = 1 , and equal prior probabilities , $\pi_1 = \pi_2$

$$R_1 = \left\{ x : \frac{f_1(x)}{f_2(x)} \geq 1 \right\} \quad , \quad R_2 = \left\{ x : \frac{f_1(x)}{f_2(x)} < 1 \right\} \quad . [7]$$

Now Estimate the Actual Error Rate :

	Π_1	Π_2	
Π_1	n_1c	$n_1m=n_1-n_1c$	n_1
Π_2	$n_2m=n_2-n_2c$	n_2c	n_2

(2.11)

Where

n_1c = number of Π_1 items correctly classified as Π_1 items.

n_1m = number of Π_1 items misclassified as Π_2 items.

n_2c = number of Π_2 items correctly classified as Π_2 items.

n_2m = number of Π_2 items misclassified as Π_1 items.

The **Actual Error Rate** is $$AER = \frac{n_1m + n_2m}{n_1 + n_2}$$

Which is recognized as the proportion of items in the training set that are misclassified. and can be Estimates of the conditional misclassification probabilities as:

$$p_{(2/1)} = \frac{n_1m}{n_1} \quad , \quad p_{(1/2)} = \frac{n_2m}{n_2}$$

Where

$p_{(2/1)}$ = probability of Π_1 items misclassified as Π_2 items.

$p_{(1/2)}$ = probability of $\prod_2$ items misclassified as $\prod_1$ items. [8]

Assumptions Underlying The Discriminant Function

i. The **p** independent variable must have multivariate normal distribution.

ii. The **p×p** variance - covariance matrix of the independent variables in each of the two groups must be the same .

2.3.1. Classification For Two Normal Populations

Assume that $f_1(x)$ is $N_p(\mu_1, \Sigma_1)$, and $f_2(x)$ is $N_p(\mu_2, \Sigma_2)$ if $\Sigma_1 = \Sigma_2 = \Sigma$ then the classification rule is allocate X to $\prod_1$ if

$$(\mu_1 - \mu_2)^T \Sigma^{-1} X - \frac{1}{2}(\mu_1 - \mu_2)^T \Sigma^{-1}(\mu_1 + \mu_2) \geq \log\left[\frac{c(1\backslash 2)\pi_2}{c(2\backslash 1)\pi_1}\right] \qquad (2.12)$$

and allocate X to $\prod_2$ otherwise .

In most practical situations, μ_1, μ_2 & Σ are unknown, suppose n_1 measurements of $X = (X_1, X_2, ..., X_p)$ from $\prod_1$ and n_2 measurements of X from $\prod_2$. Then can be estimate μ_1, μ_2 & Σ by

$$\bar{X}_1 = \frac{1}{n_1}\sum_{j=1}^{n_1} X_{1j} \quad ; \quad \bar{X}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2} X_{2j}$$

$$S_1 = \frac{1}{n_1 - 1}\sum_{j=1}^{n_1}(X_{1j} - \bar{X}_1)(X_{1j} - \bar{X}_1)^T \; ; \; S_2 = \frac{1}{n_2 - 1}\sum_{j=1}^{n_2}(X_{2j} - \bar{X}_2)(X_{2j} - \bar{X}_2)^T \qquad (2.13)$$

Where

$$S_p = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{n_1 + n_2 - 2} \qquad (2.14)$$

So the classification rule given by (2.10) can be reduced as allocate X to $\prod_1$ if

$$(\bar{X}_1 - \bar{X}_2)^T S_p^{-1} X - \frac{1}{2}(\bar{X}_1 - \bar{X}_2)^T S_p^{-1}(\bar{X}_1 + \bar{X}_2) \geq \log\left[\frac{c(1\backslash 2)\pi_2}{c(2\backslash 1)\pi_1}\right] \qquad (2.15)$$

And allocate X to $\prod_2$ otherwise . If $\Sigma_1 \neq \Sigma_2$ then allocate X to $\prod_1$ if

$$-\frac{1}{2}X^T\left(\Sigma_1^{-1} - \Sigma_2^{-1}\right)X + \left(\mu_1^T\Sigma_1^{-1} + \mu_2^T\Sigma_2^{-1}\right)X - K \geq \log\left[\frac{c(1\backslash 2)\pi_2}{c(2\backslash 1)\pi_1}\right] \qquad (2.16)$$

And allocate X to $\prod_2$ otherwise . Where K is given by

$$K = \frac{1}{2}\log\left(\frac{|\Sigma_1|}{|\Sigma_2|}\right) + \frac{1}{2}\left(\mu_1^T\Sigma_1^{-1}\mu_1 - \mu_2^T\Sigma_2^{-1}\mu_2\right)$$

The corresponding sample version is allocate X to $\prod_1$ if

$$-\frac{1}{2}X^T(S_1^{-1} - S_2^{-1})X + (\bar{X}_1^T S_1^{-1} + \bar{X}_2^T S_2^{-1})X - K \geq \log\left[\frac{c(1\backslash 2)\pi_2}{c(2\backslash 1)\pi_1}\right] \qquad (2.17)$$

And allocate X to $\prod_2$ otherwise. This classification regions are known as quadratic functions. [9]

2.3.2. Test Of Assumption

An assumption of discriminant analysis is the null hypothesis that the covariances for i-th group (i = 1,2) do not differ between groups formed by the dependent .

$$H_0 : \Sigma_1 = \Sigma_2 \quad \text{vs} \quad H_1 : \Sigma_1 \neq \Sigma_2$$

The statistical test for H_0 is (Box's M-test statistical) .

$$M = N \log|S| - \sum_{i=1}^{k} \nu_i \log|S_i|$$

This test statistic asymptotically distributed as χ^2 with degrees of freedom $f_1 = \frac{1}{2}(k-1)p(p+1)$. The adjusted test is $M \sim \chi^2_{(f_1)} / (1 - D_1)$

Where $D_1 = \frac{2p^2 + 3p - 1}{6(p+1)(k-1)}\left[\sum_{i=1}^{k}\frac{1}{\nu_i} - \frac{1}{N}\right]$

Amor accurate approximation is given by $M \sim bF_{\alpha,(f_1,f_2)}$

$$\text{And} \qquad f_1 = \frac{1}{2}p(p+1)(k-1), \qquad f_2 = \frac{f_1+2}{D_2 - D_1^2}$$

$$D_2 = \frac{(p-1)(p+2)}{6(k-1)}\left\{\sum_{i=1}^{k}\frac{1}{\nu_i^2} - \frac{1}{N^2}\right\}, \qquad b = \frac{f_1}{1 - D_1 - f_1/f_2}$$

Where

k : number of groups which is equal to 2.

S : pooled sample dispersion matrix .

S_i : the dispersion matrix for the i-th sample drawn from the i-th group.

n_i : the number of data points drawn from i-th group .

$N = \sum_{i=1}^{2} \nu_i$ is the total number of observations , $\nu_i = n_i - 1.$

p : the number of independent variable considered for discriminant analysis.

This statistic test is very sensitive to lack of normality. the hypothesis H_0 maybe rejected due to lack of normality rather than non-homogeneity .

2.3.3. Test Of Significance

To test significance of the discriminant function where

$$H_0 : \mu_1 = \mu_2 \qquad \text{vs} \qquad H_1 : \mu_1 \neq \mu_2$$

μ_1, μ_2 are the means vectors of the groups from which the i-th sample is drawn , this hypothesis was tested by using the univariate analysis of variance Willk's lambda(Λ) , also called U-statistic . When variable are considered individually , Where

$$\Lambda = \frac{BSS}{WSS}$$

BSS and WSS are the between and within groups sum of squares respectively. In this situations the smaller value for Λ greater the probability that the null hypothesis will be

rejected vice versa thus, small value of Λ indicate that groups means do appear to be different, while large value of Λ indicate that groups means do appear to be equally. To assess the statistical significance of the Willk's lambda, it can converted into an F-ratio by using the following transformation :

$$F = \left(\frac{1-\Lambda}{\Lambda}\right)\left(\frac{n_1+n_2-p-1}{p}\right), \quad f_{\alpha,p,n_1+n_2-p-1}$$

Bartlett has shown that if H_0 is true and $\sum n_i = n_1 + n_2 =$ is large, H_0 rejected at significance level α if

$$-\left(n-1-\frac{(p+g)}{2}\right)\ln\left(\frac{|W|}{|B+W|}\right) > \chi^2_{p(g-1)},(\alpha) \quad .[8]$$

2.3.4. Summary Of Canonical Discriminant function

I. The **Eigen value** of each discriminant function reflects the ratio of importance of the discriminant classify cases of the dependent variable .

II. **Standardized canonical discriminant function coefficients** are used to compare the relative importance of independent variables .

Structure matrix it is shows the correlations of each variable with each discriminant function .

ANALYSIS AND RESULTS

Estimated Factor Analysis Model

In this study requiring to determine the relationship between the type of climate (X_9), where X_9 represent the model dependent variable which takes two values X_9=0 for desert climate , and X_9=1 for costal climate, and the remaining variables (X_1, X_2, X_3, X_4, X_5, X_6, X_7 and X_8) as independent variables. On the other hand, the purpose of data reduction is to remove highly correlated variables from the data file , and replacing the entire data file with a smaller number of uncorrelated factors. The estimation of factor analysis model for each city in coastal or desert climate are shown as follows :

Table(1) Comparison Between Factor Analysis Results

Cities		Variables have high correlation with the factors							
		Max. T. (X_1)	Min. T. (X_2)	R. H. (X_3)	W. S. (X_4)	W. D. (X_5)	D. S. S. (X_6)	C. A. (X_7)	R. A. (X_8)
Derna	F	-0.917				0.870	-0.986	0.952	0.956
Benghazi	F	0.978	0.947		0.575	-0.947	0.986	-0.961	-0.952
Agedabia	F	-0.976	-0.957	0.737		0.833	-0.956	0.907	0.952
Tripoli	F	0.977	0.938	-0.907			0.982	-0.894	-0.954
Misurata	F_1'	-0.728	-0.730			0.894	-0.883	0.746	0.966
	F_2'				-0.993				
Shahat	F	0.975	0.942		-0.968		0.959	-0.935	-0.889
Sebha	F_1'	0.774	0.747	-0.903	0.976				
	F_2'						-0.888	0.980	0.775
Jalu	F_1'	-0.960	-0.976	0.733		0.709	-0.939	0.954	0.895
	F_2'				0.959				
El-Kufra	F	0.983	0.983	-0.912	0.853	-0.614	0.951	-0.856	-0.459
Ghadames	F_1'	0.881	0.834	-0.728			0.918	-0.914	-0.877
	F_2'				0.877	-0.765			

Discriminant Analysis Results For The Dependent Variable Type Of Climate (X_9)

The second aim in this study is to classify cases into the value of a categorical dependent variable Type Of Climate(X_9). The set of independent variables are Maximum Temperature(X_1), Minimum Temperature(X_2), Relative Humidity(X_3), Wind Speed(X_4), Wind Direction(X_5), Duration Of Sun Shine(X_6), Clouds Amount(X_7) and Rainfall Amount(X_8) .

Test Of Assumptions

The first group [population π_1] is Desert Climate and the second group [population π_2] is Coastal Climate . Box's M to test the null hypothesis of equal population covariance matrixes. Table (2) shows Box's M test.

Table (2) The Test Results

Box's M	396.297
F approx.	10.169
p- value	0.000

From table (2) that is $\sum_1 \neq \sum_2$.

Tests of Equality of Group Means

The tests of equality of group means measure each independent variable's potential before the model is created. Table (3) shows that

Table (3) The Test Results

Variables	Wilks' Lambda	F	p- value
Maximum Temperature (X_1)	0.865	18.470	0.000
Minimum Temperature (X_2)	1.000	0.005	0.943
Relative Humidity (X_3)	0.242	369.900	0.000
Wind Speed (X_4)	0.987	1.533	0.218
Wind Direction (X_5)	0.970	3.683	0.057
Duration Of Sun Shine (X_6)	0.928	9.146	0.003
Clouds Amount (X_7)	0.661	60.514	0.000
Rainfall Amount (X_8)	0.782	32.910	0.000

Table (3) shows that Maximum temperature (X_1), Relative Humidity (X_3), Hours Of Sun Shine (X_6), Clouds Amount (X_7) and Rainfall Amount (X_8) are significant. But Minimum Temperature (X_2), Wind Speed (X_4), Wind Direction (X_5) are not significant.

Wilk's Lambda is used to test the significant of the discriminant function as a whole . Table (4) shows wilk's lambda

Table (4) Wilks' Lambda

Test of Function	Wilks' Lambda	Chi-square	df	p- value
1	0.119	243.055	8	0.000

The discriminant function is significant, wilk's lambda is 0.119 and chi-square is 243.055 with degree of freedom 8 . Then H_0 will be rejected, and H_1 accepted or $H_1 : \mu_0 \neq \mu_1$.

Standardized Canonical Discriminant function Coefficients are used to compare the relative importance of the independent variables :

Table(5) Standardized Canonical Discriminant Function Coefficients .

Variables	Function
Maximum Temperature (X_1)	1.658
Minimum Temperature (X_2)	-0.711
Relative Humidity (X_3)	1.473
Wind Speed (X_4)	0.034
Wind Direction (X_5)	0.140
Duration Of Sun Shine (X_6)	1.458
Clouds Amount (X_7)	1.482
Rainfall Amount (X_8)	0.180

From table (5) it is see that Maximum Temperature (X_1) , Relative Humidity (X_3) , Duration Of Sun Shine (X_6) and Clouds Amount (X_7) are most important variables.

Structure matrix shows the correlations of each variable with the discriminant function

Table (6) Structure Matrix

Variables	Function
Relative Humidity (X_3)	0.649
Clouds Amount (X_7)	0.263
Rainfall Amount (X_8)	0.194
Maximum Temperature (X_1)	-0.145
Duration Of Sun Shine (X_6)	-0.102
Wind Direction (X_5)	0.065
Wind Speed (X_4)	0.042
Minimum Temperature (X_2)	0.002

Table (6), it can be observed that the Relative Humidity (X_3) have highest correlation with discriminant function.

Classification Statistics

Fisher's linear discriminant function. The classification method of discriminant classification is show in table (7) :

Table (7) Classification Function Coefficients

Variables	Type Of Climate (X_9)	
	Desert	Coastal
Maximum Temperature (X_1)	16.331	17.701
Minimum Temperature (X_2)	-14.336	-15.024
Relative Humidity (X_3)	4.975	5.897
Wind Speed (X_4)	4.087	4.176
Wind Direction (X_5)	5.624	5.976
Duration Of Sun Shine (X_6)	31.139	35.444
Clouds Amount (X_7)	48.038	56.515

Rainfall Amount (X_8)	0.396	0.441
(Constant)	-438.364	-572.539

The Classification table is shown in table (8) :

Table (8) The Classification Table

Type Of Climate (X_9)	Predicted group membership		Total
	Desert Climate	Coastal Climate	
Desert Climate **Count**	48	0	48
Coastal Climate	0	72	72
Desert Climate **Percent**	100.0	0	100.0
Coastal Climate	0	100.0	100.0
100.0 % of original grouped cases correctly classified			

From classification table (8)which shows that 48 Desert Climate cases are correctly predicted by the function which formed 100.0 % , and 72 Coastal Climate cases are correctly predicted by the function which formed 100.0 % . Where 100.0 % of original grouped cases correctly classified. so can say that the discriminant and classification analysis has good predictive validity.

The Apparent Error Rate

$$[\text{APER}] = \frac{0+0}{120} \times 100 = 0.00\,\%$$

Which is recognized as the proportion of items in the training set that are misclassified .

MANOVA Table

The F test for comparing k= 2 means

$H_0 : [\mu_{i1}\ \mu_{i2}\ \dots\ \mu_{i8}] = [\mu_{j1}\ \mu_{j2}\ \dots\ \mu_{j8}]$ Against $H_1 : [\mu_{i1}\ \mu_{i2}\ \dots\ \mu_{i8}] \neq [\mu_{j1}\ \mu_{j2}\ \dots\ \mu_{j8}]$

Where i=1,2,...,72 , j=1,2,...,48 for first and second group respectively.

Table (9) MANOVA Table

Effect	Value Wilk's Lambda	F	Hypothesis d.f	p- value
Intercept	0.001	13571.560	8.000	0.000
Type Of Climate(X_9)	0.119	103.122	8.000	0.000

From table (9) F= 103.122 and p- value = 0.000 then H_0 will be rejected [the means of tow group are not equally], and the value of Wilk's Lambda (0.119) it's very good result.

Discriminant Analysis Results For The Dependent Variable Type Of Building (X_{10})

Where the Type Of Building variable is categorical dependent variable(X_{10}). The set of data defined by Maximum Temperature(X_1), Minimum Temperature (X_2), Relative Humidity (X_3), Wind Speed (X_4), Wind Direction (X_5), Duration Of Sun Shine (X_6), Clouds Amount (X_7), Rainfall Amount(X_8).

Test Of Assumption

The first group [population π_1] is Distinctive Building and the second group [population π_2] is Prevalent Building. Box's M test the null hypothesis of equal population covariance matrixes. Table (10) shows Box's M test.

Table (10) The Test Results

Box's	305.529
F approx.	7.706
p- value	0.000

From table (10) that is $\sum_1 \neq \sum_2$.

Tests of Equality of Group Means

The tests of equality of group means measure each independent variable's potential before the model is created. Table (11) shows that

Table (11) The Test Results

Variables	Wilks' Lambda	F	p- value
Maximum Temperature (X_1)	0.915	10.939	0.001

Minimum Temperature (X_2)	1.000	0.049	0.825
Relative Humidity (X_3)	0.533	103.196	0.000
Wind Speed (X_4)	0.944	7.018	0.009
Wind Direction (X_5)	0.955	5.525	0.020
Duration Of Sun Shine (X_6)	0.945	6.862	0.010
Clouds Amount (X_7)	0.769	35.466	0.000
Rainfall Amount (X_8)	0.862	18.829	0.000

Table (11) shows that just Minimum Temperature (X_2) is not significant, and all the remaining variables are significant.

Wilk's Lambda is used to test the significant of the discriminant function as a whole. Table (12) shows wilk's lambda

Table (12) Wilks' Lambda

Test of Function	Wilks' Lambda	Chi-square	df	p- value
1	0.427	97.094	8	0.000

The discriminant function is significant , wilk's lambda is 0.427 and chi-square is 97.094 with degree of freedom 8 . Then H_0 will be rejected, and that is $H_1 : \mu_0 \neq \mu_1$ and H_1 is accepted .

Standardized Canonical Discriminant function Coefficients are used to compare the relative importance of the independent variable which shows in table (13).

Table(13) Standardized Canonical Discriminant Function Coefficients

variables	Function
Maximum Temperature (X_1)	1.142
Minimum Temperature (X_2)	-0.559
Relative Humidity (X_3)	1.256
Wind Speed (X_4)	0.353

Wind Direction (X_5)	0.154
Duration Of Sun Shine (X_6)	0.513
Clouds Amount (X_7)	0.618
Rainfall Amount (X_8)	0.024

From above table (13) Maximum Temperature (X_1) and Relative Humidity (X_3) are important.

Structure matrix shows the correlations of the variable with discriminant function.

Table (14) Structure Matrix

Variables	Function
Relative Humidity (X_3)	0.807
Clouds Amount (X_7)	0.473
Rainfall Amount (X_8)	0.345
Maximum Temperature (X_1)	-0.263
Wind Speed (X_4)	0.210
Duration Of Sun Shine (X_6)	-0.208
Wind Direction (X_5)	0.187
Minimum Temperature (X_2)	0.018

From table (14) note that Relative Humidity (X_3) have highly correlation with discriminant function.

Classification Statistics

Fisher's linear discriminant function. The classification method of discriminant classification in table (15).

Table (15) Classification Function Coefficients

Variables	Type Of Building (X_{10})	
	Prevalent	Distinctive
Maximum Temperature (X_1)	13.144	12.727
Minimum Temperature (X_2)	-12.860	-12.614
Relative Humidity (X_3)	2.692	2.451
Wind Speed (X_4)	5.254	4.826
Wind Direction (X_5)	5.044	4.867
Duration Of Sun Shine (X_6)	18.965	18.282
Clouds Amount (X_7)	24.562	23.073
Rainfall Amount (X_8)	0.255	0.252
(Constant)	-307.077	-273.980

The Classification table is shown in table (16)

Table (16) The Classification Table

Type Of Building (X_{10})	Predicted group membership		Total
	Prevalent Building	Distinctive Building	
Prevalent Building **Count**	72	12	84
Distinctive Building	1	35	36
Prevalent Building **Percent**	85.7	14.3	100.0
Distinctive Building	2.8	97.2	100.0
89.2% of original grouped cases correctly classified			

From classification table (16) which shows that 35 Distinctive Building cases are correctly predicted by the function which formed 100.0 percent , and 27 Prevalent Building cases are correctly formed 100.0 percent . Where 89.2% of original grouped cases correctly classified. So we can say that the discriminant and classification analysis has good predictive validity.

The Apparent Error Rate

$$[\text{APER}] = \frac{1+12}{120} \times 100 = 10.83\%$$

Which is recognized as the proportion of items in the training set that are misclassified .

MANOVA Table

The F test for comparing k= 2 means

H_0 : $[\mu_{i1}\ \mu_{i2}\ \dots\ \mu_{i8}] = [\mu_{j1}\ \mu_{j2}\ \dots\ \mu_{j8}]$ Against H_1 : $[\mu_{i1}\ \mu_{i2}\ \dots\ \mu_{i8}] \neq [\mu_{j1}\ \mu_{j2}\ \dots\ \mu_{j8}]$

Where i=1,2,...,84 , j=1,2,...,36 for first and second group respectively.

Table (17) MANOVA Table

Effect	Value Wilk's Lambda	F	Hypothesis d.f	p- value
Intercept	0.002	6851.928	8.000	0.000
Type Of Building (X_{10})	0.427	18.643	8.000	0.000

From table (17) F= 18.643 and p- value = 0.000 then H_0 will be rejected [the means of tow group are not equally], and the value of Wilk's Lambda (0.427) it's very good result.

Summary And Conclusions

This study is to investigate the climatic elements that affect the design of buildings in ten Libyan cities , and that to try to adapt to human comfort inside the building , where one of the most important design goals.

The main objectives of the study are (i) To define the factors influencing the Climate and to determine the most important factors of the climatic variables. (ii) To discriminate the Climate by region and define the factors which responsible for that discrimination. (iii) To determine the factors that influencing the building form.

A sample consist of 10 Libyan cities(Derna, Benghazi, Jalu, Agedabia, Tripoli, Misurata, Sebha, EL-Kufra, Ghadames and Shahat), and the total number of variables in this study

was 10, some of these variables are qualitative and most of them were quantitative ,the qualitative variables included in the analysis as dummy variables while the quantitative variables included in the analysis directly, and the data of this study collected as average monthly values for each city for 30 years (108000 data points).

Thus, to studying the variability of climate system of these places needs to explain the observed correlations between elements and situations.

To achieve the first goal of this study ,to define the factors influencing the Climate, the factor analysis is used and the following results are obtained :

1- Kaiser - Meyer – Olkin (KMO) , $0.6 <$ KMO < 0.8 , that is mean the factor analysis have a good results.

2- Maximum Temperature (X_1), Hours Of Sun Shine (X_6), Clouds Amount (X_7) and Rainfall Amount (X_8) are significant and good representative for all eight original variables in the coastal cities (Derna , Benina , Agedabia , Tripoli , Misurata, Shahat). While , the variables Maximum temperature (X_1), Minimum temperature (X_2), Relative humidity (X_3), Wind speed (X_4), Duration of sun shine (X_6), Clouds amount (X_7) and Rainfall amount (X_8) all these variables have high correlations with them factors in the desert cities (Sebha, Jalu, EL-Kufra, Ghadames). Bartlett's Test is referred to the test is significant in each case.

To achieve the second goal for this study , to classify the Type of Climate for all the cities to two type (coastal or desert) , the discriminant and classification analysis results revealed that :

3- Maximum temperature (X_1), Relative Humidity (X_3), Duration Of Sun Shine (X_6), Clouds Amount (X_7) and Rainfall Amount (X_8) are significant. But Minimum Temperature (X_2), Wind Speed (X_4), Wind Direction (X_5) are not significant.

4- The Maximum Temperature (X_1), Relative Humidity (X_3), Duration Of Sun Shine (X_6) and Clouds Amount (X_7) are most important variables, but just Relative Humidity (X_3) have high correlation with discriminant function.

To achieve the third goal for this study, to classify the Type of Building for all the cities to two type (prevalent or distinctive) , the discriminant and classification analysis results revealed that:

5- Minimum Temperature(X_2) is not significant, and all the remaining variables are significant.

6- The Maximum Temperature (X_1) and Relative Humidity (X_3) are important variables. But just Relative Humidity (X_3) have highly correlation with discriminant function.

REFERENCES

[1] Gut P. "*Climate Responsive Building –Appropriate Building Construction In Tropical And Subtropical Regions*".Swiss centre for development cooperation in technology and management . Skat, Niedermann AG, St. Gallen, Switzerland. 1993; 11,49,63,70,86,136.

[2] Alkaabi H. "*Structural Planning And Architecture In The Desert*". archhanan87@hotmail.com. 2003; 20.

[3] World Weather Meteorological. https://worldweather.wmo.int/ar/home.html. (2013).

[4] Rencher A. "*Methods Of Multivariate Analysis*", Brigham young University, United stated of America. Wiley interscience. Publication. 2002; 415,423,426.

[5] Joseph F , William C , Barry J , Rolph E. "*Multivariate Data Analysis*". New Jersey , Pearson prentice Hall. 2010 ; 236.

[6] Giri N. "*Multivariate Statistical Analysis*". University of Montreal , Quebec Canada, Marcel dekker . Inc. New York . Basel. 1991; 435-438, 517, 519

[7] Härdle W , Simar L . "*Applied Multivariate Statistical Analysis*". Springer Berlin Heidelberg. 2007; 254,292,293,300.

[8] Johnson and Wichern ."*Applied Multivariate Statistical Analysis*". New Jersey , Pearson prentice Hall. 1998; 482-496.

[9] El geziri N. "*Statistical Analysis For Mental Retardation And Down Syndrome Phenomenon*". M. SC. Thesis , Garyonis University Benghazi Libya. 2005; 34,35.

[10] National Geophysical Data Center. https://serc.carleton.edu/resources/22195.html. 2004.

[11] World Meteorological Organization. https://public.wmo.int/en. 2005.

[12] Libyan Meteorological Department, Tripoli. https://www.meteoblue.com/en/weather/historyclimate/weatherarchive/tripoli_libya_2210247 . 2000.