

İSTANBUL TEKNİK ÜNİVERSİTESİ FEN BİLİMLERİ ENSTİTÜSÜ

Üç Girişli Veri Tablosu Ana Bileşenli
Faktör Analizi Yöntemi

E. Sabri Türkler

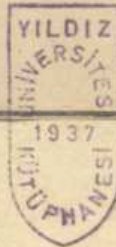
Doktora Tezi

YILDIZ ÜNİVERSİTESİ
GENEL KİTAPLIĞI

R 209

Kot :59.....
Alındığı Yer : ..Fen..Bil..Enst.....
Tarih :26/5/1937.....
Fatura :-.....
Fiatı :1000Tl.....
Ayniyat No :1/6.....
Kayıt No :44822.....
UDC :517.5.....
Ek :378.267.....

X



Comp.



ÜÇ GİRİŞLİ VERİ TABLOLARININ
ANA BİLEŞENLİ FAKTÖR ANALİZİ YÖNTEMİYLE
İNCELENMESİ VE BİR UYGULAMA

"DOKTORA TEZİ"

Yük. Müh. Eyüp Sabri TÜRKER

Tezin Fen Bilimleri Enstitüsüne Verildiği Tarih : 11 Temmuz 1984

Tezin Savunulduğu Tarih : 16 Kasım 1984

Doktorayı Yöneten Öğretim Üyesi

: Prof. Dr. Cevdet KOÇAK

Diğer Jüri Üyeleri

: Prof. Dr. Muzaffer İPEK

: Doç. Dr. Aziz BENER

I.T.Ü.

SAKARYA MÜHENDİSLİK FAKÜLTESİ

MATBAASI - 1984

İÇİNDEKİLER

SAYFA

Ö Z E T	I
S U M M A R Y	IV
S E M B O L L E R	IX
G İ R İ Ş	1
B Ü L Ü M I	
ANA BİLEŞENLİ FAKTÖR ANALİZİ YÖNTEMİ	
1.1. Temel Kavramlar ve Tanımlar	6
1.1.1. Veri Tabloları	6
1.1.2. Çok Boyutlu Vektörlerin Momentleri ...	7
1.1.3. Norm	10
1.1.4. Ana Eksenler	11
1.1.5. Nokta Bulutunun Yeri	12
1.1.6. Varyanslar Matrisi	13
1.1.7. Metrik	14
1.1.8. Bireyler Bulutu	18
1.2. Ana Bileşenlerin Özvektörler Yardımıyla Elde Edilmesi	20
1.2.1. Ana Bileşenlerin Hesabı	20
1.2.2. Değişkenler Bulutunun R^n Uzayında İncelenmesi	31
1.2.3. R^p ve R^n Uzayları Arasındaki İlişki ..	32
1.2.4. Ana Bileşenlere Geçiş Formülleri	33
1.2.5. Veri Tablosunun Ana Bileşenler Yardımla Yeniden Kuruluşu	34
1.2.6. Gösterilimin Kalitesi	36
1.2.7. Katılma Payları	37
1.2.8. Ek Elemanlar	38

1.3. Normlu Ana Bileşenler	40
1.3.1. Normlu Bireylerin Bileşenleri	41
1.3.2. Normlu Değişkenlerin Bileşenleri	42
1.4. Çok Değişkenli Normal Veriler İçin Ana Bileşenli Faktör Analizi	45

B Ü L Ü M 2

UÇ GİRİŞLİ VERİ TABLOLARININ FAKTÖR ANALİZİ

2.1. Uç Girişli Veri Tabloları	49
2.2. Uç Girişli Tablolar İçin Yeniden Kuruluş Formülü	50
2.3. Yeniden Kuruluş Formülündeki Ölçümlerin İncelenmesi	55

B Ü L Ü M 3

U Y G U L A M A

3.1. Analizi Yapılacak Veri Tabloları	57
3.2. Analiz Programının Hazırlanması	59
3.3. Uygulanacak Analizlerde Seçim Kriteri	60
3.4. Grafikselsel Yorum Tekniği	61

B Ü L Ü M 4

ARAŞTIRMA SONUÇLARI

4.1. Veri Tablolarının Analizi	66
4.2. İki Girişli k_{IJ} Tablosunun Analizi	68
4.3. Uç Girişli k_{IJT} Tablosunun Analizi ve Sonuçların Yorumu	70
4.4. Uç Girişli k_{IJT} Tablosunun Yeniden Kuruluş Formülünün Grafikselsel Sağlanması	73

B Ü L Ü M 5

S O N U Ç L A R	75
E K B Ü L Ü M	78
R E F E R A N S L A R	104

T E Ş E K K Ü R

Ö Z G E Ç M İ Ş

Ö Z E T

Çok değişkenli veri analizinde, değişkenler arasındaki bağımlılığın incelenmesi regresyon ve korelasyon teknikleri yardımıyla yapılabilmektedir. Ancak, değişkenler arasındaki bağımlılığın söz konusu olmadığı, yahut değişkenler arası ilişkilerin yapısı hakkında ön bilgilerin bulunmadığı hallerde başka tekniklere gereksinim vardır.

Bu çalışmada, çok boyutlu veri analizinin önemli yöntemlerinden biri olan ana bileşenli faktör analizi çerçevesinde, üç girişli veri tabloları incelenmiştir.

Ana bileşenli faktör analizi yardımıyla üç girişli veri tablolarının incelenmesi oldukça yeni olup, uygulanmasında bazı güçlüklerle karşılaşmaktadır. Bu güçlüklerden en önemlisi, iki girişli veri tabloları için varolan ve ana bileşenleri verilere bağlayan yeniden kuruluş formülünün, üç girişli tablolara genelleştirilmesinde ortaya çıkmaktadır. Bu çalışmada, bu tür tablolarda kullanılmak üzere, bir yeniden kuruluş formülü önerilmektedir. Hem ana bileşenli faktör analizini uygulamak ve hem de formülün geçerliliğini sağlamak amacıyla, üç girişli tablolara model olarak bir ekonomi problemi seçilmiştir.

Giriş Bölümünde: Çok boyutlu veri analizindeki istatistiksel yöntemlerden kısaca söz edilmiş ve ana bileşenli faktör analizi yöntemiyle ilgili yapılmış araştırmalardan bazıları tanıtılmıştır. Ayrıca, iki ve üç girişli veri tablolarında göz önüne alınan yeniden kuruluş formülleriyle ilgili son yıllara kadar yapılan bazı çalışmalara değinilmiştir.

Birinci Bölümde: Verilerin çok boyutlu uzayda oluşturduğu nokta bulutunun, mümkün olduğu kadar az bilgi kaybı ile, daha küçük boyutlu bir alt uzayda gösterilimine çalışılmıştır. Veri tablosunun satırlarını oluşturan bireyler ile, ko-

lonlarını oluşturan değişkenler yardımıyla, araştırılan alt uzaya ait ana eksenlerin bulunması problemi, değişik yönle- riyle ele alınarak incelenmiştir. Ayrıca bu yöntemde; prob- lemin çözümünün Özdeğer-Özvektör araştırılması problemine dönüştürülebileceği gösterilmiş ve Özvektörlerden ana bile- şenlere geçiş formülleri çıkartılarak, bileşenler yardımıy- la asıl veri tablosunun yeniden kuruluşunu sağlayan formül elde edilmiştir.

İkinci Bölümde: İki girişli veri tabloları için var olan yeniden kuruluş formülü, üç girişli tablolar için kul- lanılabilecek daha genel bir halde verilmiş ve bu formülün ana bileşenlerle olan ilişkisi gösterilmiştir. Varlığı ve tekliği son yıllarda ispatlanmış bulunan söz konusu formül, üçüncü basamaktan etkileşimleri kapsamak suretiyle, üç boyut- lu verileri ana bileşenler cinsinden tam olarak ifade edebil- mektedir. Yalnız ikinci basamaktan etkileşimlerin alınması halinde dahi, yeterli bir yaklaşım sağlayabilmektedir.

Üçüncü Bölümde: Ana bileşenli faktör analizi yöntemi- le ilgili, FORTRAN IV dilinde geliştirilen, oldukça genel amaçlı bir programın algoritması tanıtılmış ve program çı- kışlarının yorumunda kullanılan istatistiksel kriterler özetlenmiştir.

Dördüncü Bölümde: Ülkemizin ekonomik yapısını belirli ölçüde yansıtan, son yıllardaki ihracat problemi ele alına- rak, bu problemin istatistiksel analizi yapılmıştır. 1976/80 yılları arasında, ülkemiz ihracatında önemli rol oynayan ürünler "Değişken" ve ihracatta önemli payı olan ülkeler "Birey" olarak seçilmiş ve böylece her yıl için oluşturulan (25 x 12) boyutlu veri tablolarının ana bileşenli faktör analizi yapılmıştır. Ayrıca; iki ve üç girişli karşılaştır- ma tabloları için önerilen yeniden kuruluş formüllerinden elde edilen sonuçlar ile, başlangıç tabloları karşılaştırıl- mıştır.

Son Bölüm sonuçlara ayrılmıştır. Burada, analiz sonuçlarının genel bir değerlendirmesi yapılmıştır.

Analiz sonucunda elde edilen tablo ve grafikler Ek'te verilmiştir.

S U M M A R Y

Regression and Correlation techniques can be used in the investigation of dependence in such situations where a variable of a multivariate system depends on other variables. However, if a differentiation between dependent and independent variables is not a matter of importance, in other words if all the variables carry equal importance from the point of view of the purposes of the investigation, then it is essential that other techniques will be used in the analysis of dependence. The techniques included in this category are useful especially when there is not sufficient information about the structure of the relationship between the variables

Starting with the Principal Component Analysis method and with the introduction of formula maintaining the reestablishment of the starting data table, an analysis of an economical problem closely related to the recent export structure of Turkey was made to prove the conformity of this formula.

In the introduction chapter, the initiation of multi-dimensional data analysis methods and its development has been briefly given and the principal component analysis method and its relation with the other multivariate methods has been explained. Also, the work carried out on the formula which maintain the reestablishment of the starting table with the aid of principal component has been summarised.

In the first section, the principal component analysis method and the basic hypothesis on which it depends has been introduced and the related formula are given, also the analysis has been investigated from the geometrical point of view and the calculations are established more clearly.

The aim of principal component analysis method is to reduce the number of dimension of the problem by using variables related to the problem and by creating a few artificial variables. In the maintaining, some information in the data matrix is naturally lost, still the method depends on the principal of finding and optimum solution in such away to minimize the information loss. In the most general sense, the principal component analysis method consists of looking for the linear combinations of the original variables by summarising the starting data table and with the minimum information loss possible.

After establishing the basis of the method, all the definitions and the calculation methods required for the analysis are given, at the end of the first section the geometrical study of the principal component is made and the principal component under the hypothesis normal distribution is investigated as well.

In the second section, starting with the principal component, the formula which provided the reestablishment of the starting table are introduced. The validity of the two way contingency tables of a formula is shown, followed by an investigation of a more general formula which can be used for the three way tables. Especially in the situation where the number of the selected variables is great, this formula gives results close to the starting data table.

In the third section, the Algorithm of the general purpose program which can make all the required calculations about the principal component analysis method and which is coded in FORTRAN IV language is given. This program is prepared to analyse (50 x 50) data matrix and to give the results as appropriate tables and graphs. The algorithmic structure of the program can be summarised as follows;

- i. It reads the data table with the required FORMAT,
- ii. Starting with the data table it calculates the correlation (or covariance) Matrix,
- iii. It calculates the eigen values and eigen vectors of the correlation (or covariance) matrix with the aid of iteration method,
- iv. Starting with the eigen values and eigen vectors and using the transformation formula it calculates the coefficients of principal components,
- v. Since the calculated eigen values are defined as the variance of principal components, it gives the eigen value percentages, accumulated percentages and the histograms of the eigen values as appropriate tables,
- vi. If required, it prints the data table and its correlation matrix in an appropriate way,
- vii. It gives the quality of the figure, the average and standard deviation of the variables and the calculations of the contribution ratio and thus helps the discussion of the results,
- viii. It prints the graphs of the calculated principal components in the required planes.

In the forth section the results and discussion obtained as a result of application of a program of principal component analysis method to an economical problem which is possible to be investigated statistically is given. The results obtained by the application of the related program to the five (25x12) data matrix created in accordance with the last five years exports of Turkey can be summarized as follows;

i. As a result of analysis made by evaluating the data matrices in various ways, it is found that the first four components are generally meaningful. The possibility of explaining the properties, which can be explained with 12 variables, by four equations which are found as the linear components of these variables shows that this analysis is a very effective dimension reducing method for the used data.

ii. When the five years' export quantities between 1976/80 is analyzed by taking the time dimension into account, it is observed that the types of products that Turkey has exported to the neighbouring countries are accumulated at the same side of the principal components. This can be considered as there is no significant change in the export structure.

iii. When the calculated components of the countries and the relative placement of these countries are observed on the graphs, it is seen that the countries having the same geographical character are accumulated at the same place. Still it is observed that some countries are very closed to the components and some are very far.

iv. Although the results obtained from the analysis of the three way k_{IJT} tables contains the results of two way k_{IJ} tables, it is possible to see more clearly the grouping of both the countries and the products in the analysis of the three way tables.

v. The reestablishment formula proposed for the data tables of correspondance analysis contains density functions. Since it is possible to pass from the correspondance analysis to principal component analysis with the aid of appropriate transformations, a formula which maintains the reestablishment of the three way tables can be given in the principal component analysis. In this study, a reestablishment formula

for three way contingency tables is proposed to be used in the principal component analysis method.

vi. By analysing the result obtained from two and three way contingency tables the similarity with the starting table is observed. This means that, the formula put forward are very useful in the principal component analysis method.

The tables and graphs obtained as a result of the application of the program related to the analysis are given in the appendix.

SEMBOLLER

C	: $(p \times p)$ korelasyon matrisi.
$d(\underline{x}^i, \underline{x}^j)$	: $\underline{x}^i$ ve $\underline{x}^j$ vektörleri arasındaki uzaklık.
d^{ijt}	: k_{ijt} ölçümünün $k_i k_j k_t$ ölçümler çarpımına göre yoğunluğu.
D_λ	: Köşegen üzerindeki elemanları özdeğerler olan $(p \times p)$ köşegen matris.
$\underline{e}_i$	: R^p uzayına ait birim vektör.
$E(\underline{x}_i)$	: $\underline{x}_i$ vektörünün beklenen değeri.
$F(\underline{x}^j)$	: Normlu değişkenlerin dağılım fonksiyonu.
f_{IJ}	: İki girişli frekanslar tablosu.
f_{IJT}	: Üç girişli frekanslar tablosu.
$G(\underline{\Psi})$	: Ana bileşenlerin dağılım fonksiyonu
$\underline{g}_b$	: Bireyler bulutunun ağırlık merkezi.
$\underline{g}_d$	: Değişkenler bulutunun ağırlık merkezi.
I_p	: $(p \times p)$ Birim matris.
J	: Jakobiyen.
k_{IJT}	: Üç girişli veri tablosu.
k_{IJ}	: İki girişli veri tablosu.
k	: Veri tablosunun elemanlar toplamı.
KI	: İzafi katılma payı.
KM	: Mutlak katılma payı.
$Kor(\underline{x}^i, \underline{x}^j)$	: $\underline{x}^i$ ve $\underline{x}^j$ arasındaki korelasyon katsayısı.
$Kov(\underline{x}^i, \underline{x}^j)$	: $\underline{x}^i$ ve $\underline{x}^j$ arasındaki kovaryans.
M	: Seçilen uzaya ait metrik.
$r_{1,1,2,\dots,q}^2$	: Gösterilimin kalitesi.

R^n, R^p	: n ve p boyutlu reel uzaylar.
S^j	: Normlu değişkenlerin bileşenlerine ait varyans
$\underline{U}_\alpha$	: Bireylere ait α . ana eksenin birim vektörü.
U	: (p x p) özvektörler matrisi.
$\underline{V}_\alpha$	: Değişkenlere ait α . ana eksenin birim vektörü.
$\text{Var} (\underline{x}^i)$	: $\underline{x}^i$ değişkeninin varyansı.
X	: (n x p) veri tablosu.
X'	: (p x n) veri tablosunun transpozisi.
X^*	: (n x p) Veri tablosunun yaklaşık değeri.
$\underline{x}_i$	: Bireyler bulutunun elemanı ($i=1, \dots, n$).
$\underline{x}^j$	: Değişkenler bulutunun elemanı ($j=1, \dots, p$).
$\langle \underline{x}^i, \underline{x}^j \rangle_M$	: $\underline{x}^i$ ve $\underline{x}^j$ 'nin M'ye göre skaler çarpımı.
$ \underline{x}^i _M$	: $\underline{x}^i$ vektörünün M normu.
$\underline{x}_0$	: Ek olarak seçilen birey.
$\underline{x}^0$	: Ek olarak seçilen değişken.
Y	: (n x p) Merkezi veri tablosu.
Δ_α	: Bireylere ait α ana eksen.
Δ_r	: Değişkenlere ait r. ana eksen.
L	: Lagrange'yen.
λ_α	: Öz değerler ($\alpha=1, \dots, p$).
$\underline{f}$	: Normlu bireylere ait ana bileşenler vektörü.
$\hat{\underline{f}}$	: Normlu değişkenlere ait ana bileşenler vektörü
ρ_{ij}	: i. ve j. bileşen arasındaki korelasyon katsayı
σ_{ij}	: i. ve j. bileşen arasındaki kovaryans.
Σ	: Varyans-Kovaryans matrisi.
$\underline{\varphi}$	: Ek bireylere ait ana bileşenler vektörü.
$\underline{\hat{\varphi}}$	: Ek değişkenlere ait ana bileşenler vektörü.

- $\underline{\Psi}$: Bireylere ait ana bileşenler vektörü.
 $\underline{\dot{\Psi}}$: Değişkenlere ait ana bileşenler vektörü.
 $\underline{\Psi}^*$: Zaman ile ilgili ana bileşenler vektörü.
 N_p : p değişkenli normal dağılım.

Çok değişkenli dağılışı gösteren herhangi bir sistemin değişkenleri arasındaki birlikte değişimin yorumlanması gereksinimi, çeşitli boyut indirgeme yöntemlerinin kullanılması en önemli nedenlerden biri olmuştur. Bu yöntemlerin pek çoğu, veri tablosundan elde edilen bazı özellikleri kullanarak, bireyler veya değişkenler arasındaki ilişkileri, anlaşılması ve yorumu daha kolay bir şekle sokmayı amaçlarlar.

Günümüzde, Faktör Analizi (1-6), Karşılaştırma Analizi (7-10), Çok Boyutlu Ölçekleme (8, 11), Kanonik Analiz (12, 13) ve Ana Bileşenli Faktör Analizi (14-16) gibi çok değişkenli istatistiksel yöntemler, özellikle uygulamalı bilim dallarında yaygın bir biçimde kullanılmaktadır. Bu yöntemlerin dayandığı temel ilkeler çok eskilere dayanmakla birlikte, gelişmeleri oldukça yenidir.

Faktörel yöntemler içerisinde en eski olarak bilineni, psikologlar tarafından ortaya atılıp geliştirilen klasik Faktör Analizi olup, bu yöntemin ilk uygulamaları psikoloji alanında olmuştur. Faktör Analizi, bir değişken takımı arasındaki ilişkilerin analizi amacıyla geliştirilmiş yöntemlerin, genel anlamda bir adlandırılması olarak kullanılmaktadır. Bu değişkenler, test skorları, anket yanıtları veya istatistiksel kestirmeler olabilirler. Faktör Analizinin bazı yöntemleri, zihni kabiliyetlerin temel yapısını konu alan özel hipotezlerin açıklanması için geliştirilmiştir.

Başlangıçta C. SPEARMAN (17), H. THURSTONE (18) ve C. BURT (19) tarafından temelleri atılan ve geliştirilen bu yöntemler yardımıyla, verilerden elde edilen sonuçlardan başka, doğrudan doğruya gözlenemeyen veya gözlemlerin dışın-

da kalan birtakım değişkenler de ortaya çıkarılmaktadır. Böylece, psikolojik testlere tabi tutulan kişiler üzerinde elde edilen sonuçlar, bellek, zekâ, çalışma gücü, gibi çok az sayıdaki birtakım gizli faktörlerle açıklanabilmektedir (20, 21). Psikologların kurduğu bu yöntemler, bilgisayar olanaklarının da artmasıyla kısa zamanda gelişerek, çeşitli varsayımlara ve matematiksel kurallara dayanan daha değişik analiz yöntemlerinin doğmasına yol açmıştır (22 - 28). Bu nedenle, psikologları çok boyutlu veri analizi dünyasının keşfinde ilk öncüler olarak kabul edebiliriz.

Veri analizlerindeki başarılı çalışmaları ile tanınan J.D. CAROLL, J.B. KRUSKAL, W.S. TORGERSON, G. YOUNG, R.N. SHEPARD gibi isimlere Psychometrika adlı bültende rastlıyoruz. Bu istatistikçiler, psikolojik deney veya gözlemlerden elde edilen veri tablolarının gösterilimleri ile ilgili olarak, çok boyutlu ölçekleme adı altında güçlü ve başarılı teknikler geliştirmişlerdir (29 - 31).

H. HOTELLING, yine C. SPEARMAN (17) ve K. PEARSON (32)'in yaptığı çalışmalara dayanarak ilk olarak 1930'larda Ana Bileşenli Faktör Analizi yöntemini ve 1936'larda da yine ilk olarak Kanonik Analiz yöntemini ortaya koymuştur (33).

Çok değişkenli istatistiksel yöntemler içerisinde, kovaryans yapısının incelenmesinde kullanılan en etkin yöntemlerden birisi Ana Bileşenli Faktör Analizi yöntemidir (46 - 54). Başlangıçta, psikoloji alanında kullanılan bu yöntem, zamanla ekonomi (34, 35), iktisat (36, 37), ziraat ve hayvancılık (38, 39), madencilik (40), ekonometri (41) ve tıp (42 - 45) alanlarında da yaygın bir biçimde kullanılmaya başlanmıştır. Analiz objektif olarak, başlangıç değişkenlerinin maksimum varyanslı lineer kombinezonlarının araştırılması esasına dayanır. Özellikle, araştırmada kullanılacak değişken sayısının çok olduğu durumlarda değişkenlerin çok boyutlu örnek içerisinde teker teker incelenmesi oldukça zordur ve yanlış kararlar vermeye neden olmaktadır. Bu nedenle, başlan-

gıç deęiřkenleri yerine, bunların lineer kombinezonlarından oluřan ve kendi aralarında korelasyonsuz yeni bileřenler incelenmektedir. Metod, büyük varyansı ieren lineer kombinezonların arařtırılması esasına dayandıđından, küçük varyanslı lineer kombinezonlar ihmal edilmektedir. Bulunan yeni deęiřkenler tamamıyla yapay olup, bir anlamda sistemin temelinde var olan esas faktörler olarak kabul edilebilirler. Faktörler her deęiřkeni az veya çok etkiledikleri iindir ki, deęiřkenlerle faktörler arasındaki korelasyonlar ortaya ıkmaktadır.

Faktör Analizi yöntemlerinden olan Karřılařtırma Analizi yöntemi de 1960'lardan sonra J.P. BENZECRI tarafından geliřtirilmiřtir (55 - 58). Aslında yöntemin esaslarını H.D. HIRSCHFELD (29) ve R.A. FISHER (59)'in 1935-1940 yılları arasında tesadüf tabloları üzerine yaptıkları alıřmalara kadar götürmek mümkündür. Ancak, o zaman yapılan alıřmalar, klasik istatistiđin ok dar erevesi ierisinde kalıyordu. Konuyla ilgili olarak, son zamanlarda uygulamaya yönelik bazı alıřmaların da yapıldıđı görölmektedir (60, 61).

J.P. BENZECRI'nin alıřmalarıyla veri analizinin esasını teřkil eden deskriptif yapıya ait cebirsel özellikler de kullanılmaya bařlandı (55, 62). Aslında, uygun deęiřken dönüřümleri yardımıyla, bu yöntemden, Ana Bileřenli Faktör Analizi yöntemine geilebildiđi gibi, onu Diskriminant Analizi'nin özel bir hali saymak, hatta verilerin ayrı kodlanması halinde de Kanonik Analiz'in özel bir haline indirgemek mümkündür (59, 63, 64, 65). W.S. TORGERSON, M.O. HILL ve L. GUTTMAN, Karřılařtırma Analizi'ni bir ölekleme yöntemi olarak yorumlamıřlardır (66).

İki giriřli tabloların Karřılařtırma Analizi'ni, ok giriřli tablolara genelleřtirmek iin birok yöntem vardır. Bunların ilk örnekleri, 1964'te J.P. BENZECRI ve 1965'te de B. ESCOFIER-CORDIER tarafından verilmiřtir (62, 67). Fakat

en önemlisi, 1972 yılında yine J.P. BENZECRI tarafından geliştirilen bir yöntemdir ki, bunun da ana prensipleri, C. BURT'un 1950 yıllarında yaptığı çalışmalara dayanır (19). Bu konu ile ilgili olarak, özellikle J.D. CAROLL (31) ve P. HORST (68)'in yaptığı ve M. MASSON (69)'in bunlara dayanarak gerçekleştirdiği çalışmaları, yakın tarihlere kadar yapılan son çalışmalar olarak söyleyebiliriz. Karşılaştırma Analizi yöntemini, H. HOTELLING'in Kanonik Korelasyon Analizi'nin özel bir hali olarak da düşünebiliriz (66).

C. ECKART ve G. YOUNG verilen $(n \times n)$ boyutlu bir $K_{IJ} = \{k(i, j) \mid i, j \in \{1, n\}\}$ tablosuna karşılık, en küçük kareler anlamında, p. basamaktan yaklaşık bir tablo bulma problemi ile ilgili olarak basit ve tek bir çözüm elde etmişlerdir (70). İki girişli tablolar için verilen bu formülü, J.D. CAROLL ve J.J. CHANG üç girişli tablolara genelleştirmek istemişlerse de, çözümün tekliğini sağlayamamışlardır (30, 31, 55, 71).

İki girişli veri tablolarına uygulanan klasik Faktör Analizi yöntemleri, üç girişli tablolar için yetersiz kalmaktadır. Bu nedenle, daha değişik yöntemler kullanmak gerekir. Bu konuda geliştirilmiş ilk ve formel çalışma, üç girişli karşılaştırma tablolarının, Lojik tablolara dönüştürülmesi esasına dayanır (55).

Verilerin Ana Bileşenler yardımıyla ifade edilebilmesinin zorluğu, üç girişli veri tablolarının analizinde, son zamanlara kadar, en önemli problemlerden biri olarak kabul edilmekteydi. Ancak, bu problem 1982'den beri, üçüncü basamaktan etkileşimleri de içermek suretiyle, üç boyutlu verileri, Ana Bileşenler cinsinden tam olarak ifade eden bir formülün verilmesiyle çözülmüş bulunmaktadır. Sadece ikinci basamaktan etkileşimlerin alınması halinde bile, yeterli bir yaklaşım elde edilebilmektedir (72).

A. BENER tarafından ilk defa 1982'de verilen, üç girişli tablolar için yeniden kuruluş formülü, Karşılaştırma Analizi'nde uygulanarak, başlangıç veri tablosuna oldukça yakın sonuçlar ortaya çıkarmıştır (72).

Bu çalışmada, varlığı ve tekliği teorik olarak kanıtlanan en genel anlamda yeniden kuruluş formülünün, Ana Bileşenli Faktör Analizinde de, gerekli düzeltmeler yapılmak suretiyle kullanılabilme olanakları araştırılmıştır. Formülün geçerliliğini ortaya koymak bakımından, uygulama olarak, Türkiye'nin 1976/80 yılları arasındaki ihracatı seçilmiştir.

Analiz ile ilgili programlar, İ.T.Ü. Bilgi İşlem Merkezi'nde hazırlanmış ve tüm uygulamalar bu merkezdeki IBM 4341 Elektronik Hesap Makinası'nda gerçekleştirilmiştir.

B Ö L Ü M : I

ANA BİLEŞENLİ FAKTÖR ANALİZİ YÖNTEMİ

1.1. TEMEL KAVRAMLAR VE TANIMLAR

Genel olarak p tane değişken arasındaki ilişki, bu değişkenleri ana bileşenler adı verilen ve aralarında korelasyon bulunmayan yeni değişken takımına dönüştürerek incelenir. Gerçekte, değişkenlerden ana bileşenlere yapılan dönüşüm, koordinat eksenlerinin, birçok özellikleri olan yeni bir koordinat sistemine dönüşümünden ibarettir. Eğer, veri tablosundaki informasyonun büyük bir kısmı ilk iki ana bileşen tarafından açıklanabilmişse, analiz gerçek anlamda başarılı olmuş demektir. Bununla birlikte, eğer değişkenlerin tümü kendi aralarında korelasyonsuz ise, Ana Bileşenli Faktör Analizi yöntemini uygulamanın pratik yararı olmaz (14).

Burada, öncelikle çok boyutlu veri analizlerinde kullanılan bazı temel tanım ve kavramları vereceğiz.

1.1.1. VERİ TABLOLARI

Veri analizinde genellikle (n x p) tipinde

$$X = \begin{pmatrix} x_1^1 & x_1^2 & \dots & x_1^p \\ x_2^1 & x_2^2 & \dots & x_2^p \\ \vdots & \vdots & \ddots & \vdots \\ x_n^1 & x_n^2 & \dots & x_n^p \end{pmatrix} \quad (1.1)$$

bir X tablosu göz önüne alınır. Bu tablo, p değişkeninin her biri üzerinde yapılmış n tane ölçümünün sonuçları olarak yorumlanabilir. Buna göre, tablo

$$\underline{x} = (\underline{x}_1, \underline{x}_2, \dots, \underline{x}_p)$$

veya

$$\underline{x} = [\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n]$$

olarak yazılabilir. Birinci halde R^n uzayına ait p değişkenin, ikinci halde de R^p uzayına ait n bireyin, oluşturduğu iki nokta bulutu elde edilir.

1.1.2. ÇOK BOYUTLU VEKTÖRLERİN MOMENTLERİ

Beklenen Değer: Herhangi bir $\underline{x}_i$ şans vektörünün beklenen değeri $E(\underline{x}_i)$ ile gösterilir ve

$$E(\underline{x}_i) = [E(x_1^1), E(x_1^2), \dots, E(x_1^p)] \quad (1.2)$$

olarak tanımlanır.

Benzer şekilde, beklenen değerlerden oluşan matris tanımlanabilir.

Kovaryans: $\underline{x}_i$ vektörünün x_i^j ve x_i^k gibi iki bileşeninin kovaryansı

$$\begin{aligned} \text{Kov}(x_i^j, x_i^k) &= E([x_i^j - E(x_i^j)] [x_i^k - E(x_i^k)]) \\ &= E(x_i^j x_i^k) - [E(x_i^j)] [E(x_i^k)] \\ &= \sigma_{jk} \end{aligned} \quad (1.3)$$

formülü ile verilir.

Eğer $j = k$ ise (1.3) formülü x_i^j elemanının varyansını ifade eder. Yani,

$$\text{Var}(x_i^j) = E([x_i^j - E(x_i^j)]^2)$$

olur.

Genel olarak, p bileşenli $\underline{x}_i$ şans vektörünün varyans kovaryans (veya varyanslar) matrisi tanım olarak,

$$\Sigma = E\{[\underline{x}_i - E(\underline{x}_i)][\underline{x}_i - E(\underline{x}_i)]'\}$$

formülü veya açık olarak,

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ \vdots & \vdots & & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{pmatrix} \quad (1.4)$$

matrisi ile verilir.

Ayrıca, $a_i \in R$ ($i = 1, 2, \dots, p$) (Burada R Reel sayılar kümesidir) ve

$$\underline{a}' \underline{x}_i = a_1 x_i^1 + a_2 x_i^2 + \dots + a_p x_i^p$$

olmak üzere

$$\text{Var}(\underline{a}' \underline{x}_i) = \sum_{i=1}^p \sum_{j=1}^p a_i a_j \sigma_{ij} = \underline{a}' \Sigma \underline{a}$$

şeklinde tanımlanır. Benzer olarak,

$$b_i \in R \quad (i = 1, \dots, p) \quad \text{ve}$$

$$\underline{b}' \underline{x}_i = b_1 x_i^1 + b_2 x_i^2 + \dots + b_p x_i^p$$

için

$$\text{Kov}(\underline{a}' \underline{x}_i, \underline{b}' \underline{x}_i) = \sum_{i=1}^p \sum_{j=1}^p a_i b_j \sigma_{ij} = \underline{a}' \Sigma \underline{b}$$

yazılabilir.

A ve B sırasıyla reel elemanlı $(r \times p)$ ve $(s \times p)$ matrisler olmak üzere yukarıdaki kavramlar;

$$Y = AX \quad , \quad Z = BX \quad (1.5)$$

alınarak

$$\begin{aligned} \text{Kov } (Y, Y) &= A \Sigma A' \\ \text{Kov } (Z, Z) &= B \Sigma B' \\ \text{Kov } (Y, Z) &= A \Sigma B' \end{aligned} \quad (1.6)$$

şeklinde yazılmak suretiyle, matris notasyonuyla gösterilebilirler.

Korelasyon: x_i vektörünün x_i^j ve x_i^k bileşenleri arasındaki korelasyon katsayısı,

$$\rho_{jk} = \frac{\text{Kov } (x_i^j, x_i^k)}{\sqrt{\text{Var}(x_i^j) \text{Var}(x_i^k)}} \quad (1.7)$$

ile verilir.

Eğer x_i^j ve x_i^k bağımsız iseler, Kovaryans ve dolayısıyla korelasyon katsayısı sıfırdır.

Genel olarak x_i vektörünün korelasyon matrisi

$$C = \begin{pmatrix} 1 & \rho_{12} & \dots & \rho_{1p} \\ \rho_{12} & 1 & \dots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1p} & \rho_{2p} & \dots & 1 \end{pmatrix} \quad (1.8)$$

ile tanımlanır. Böylece, varyans-kovaryans matrisi ile korelasyon matrisini birbirine bağlayan,

$$C = D\left(\frac{1}{\sigma_i}\right) \Sigma D\left(\frac{1}{\sigma_i}\right) \quad (1.9)$$

$$\Sigma = D(\sigma_i) C D(\sigma_i)$$

(Burada $D(\sigma_i)$ matrisi, standart sapmalardan oluşan köşegen matristir) bağıntıları yazılabilir.

1.1.3. NORM

M , simetrik pozitif definit bir matris olmak üzere, R^p uzayına ait herhangi $\underline{x}_i$ ve $\underline{x}_j$ elemanlarının M 'ye göre skaler çarpımı tanım olarak,

$$\langle \underline{x}_i, \underline{x}_j \rangle_M = \underline{x}_i^T M \underline{x}_j = M(\underline{x}_i, \underline{x}_j) \quad (1.10)$$

şeklinde verilir.

R^p uzayındaki herhangi bir $\underline{x}_i$ elemanının Normu ise $\|\underline{x}_i\|$ ile gösterilir ve

$$\|\underline{x}_i\|^2 = \underline{x}_i^T M \underline{x}_i \quad (1.11)$$

olarak tanımlanır.

Böylece R^p veya R^n uzayında simetrik pozitif definit bir M matrisi verilmekle, bu uzaya Euklidyen bir yapı verilmiş olur. Burada M 'ye, uzayın metriği denir.

Uzaya ait iki eleman arasındaki açının kosinüsü,

$$\cos \theta = \frac{\langle \underline{x}_i, \underline{x}_j \rangle_M}{\|\underline{x}_i\|_M \|\underline{x}_j\|_M} \quad (1.12)$$

şeklinde tanımlanır. Ayrıca, nokta bulutunun elemanları arasındaki uzaklık M metriğine göre,

$$d^2(\underline{x}_i, \underline{x}_j) = \|\underline{x}_i - \underline{x}_j\|_M^2 = (\underline{x}_i - \underline{x}_j)' M (\underline{x}_i - \underline{x}_j) \quad (1.13)$$

formülü ile verilir.

I Birim matris olmak üzere $M = I$ için,

$$\|\underline{x}_i\|^2 = \underline{x}_i' \underline{x}_i \quad (1.14)$$

$$d^2(\underline{x}_i, \underline{x}_j) = (\underline{x}_i - \underline{x}_j)' (\underline{x}_i - \underline{x}_j)$$

yazılabilir.

1.1.4. ANA EKSENLER

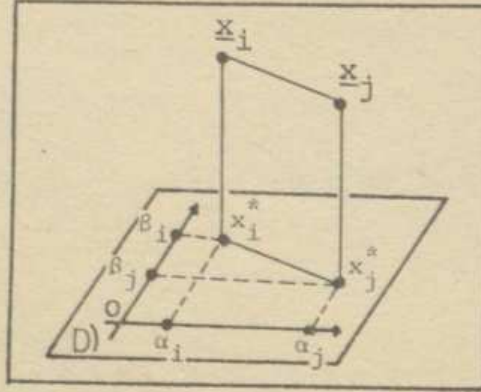
Veri analizinde amaç, veri tablosundan elde edilen nokta bulutunu daha az boyutlu, özellikle iki boyutlu, uzayda göstermektir. Bunu gerçekleştirebilmek için, ele alınan nokta bulutunun, söz konusu alt uzay üzerine izdüşümü alınır. Öyleki, izdüşüm sonucunda noktalar arasındaki uzaklıklar en az düzeyde değişmelidir. O halde, öyle bir indirgenmiş alt uzay (izdüşüm düzlemi) bulmalıyız ki, noktalar arasındaki uzaklıkların, izdüşüm sonucundaki değişimi minimum olsun. Başka bir ifadeyle, noktaların indirgenmiş alt uzay üzerindeki dik izdüşümleri arasındaki uzaklıkların kareleri toplamı maksimum olmalıdır.

Ana düzlem adını vereceğimiz izdüşüm düzlemini belirlemek için, bu düzleme ait Δ_1 ve Δ_2 gibi iki doğrunun bulunması yeterli olacaktır (Şekil: 1.1).

Δ_1 ve Δ_2 doğruları dik iseler, bireyler bulutuna ait herhangi $\underline{x}_i$ ve $\underline{x}_j$ elemanı için,

$$d^2(\underline{x}_i^*, \underline{x}_j^*) = d^2(\alpha_i, \alpha_j) + d^2(\beta_i, \beta_j) \quad (1.15)$$

yazılabilir. Burada $\underline{x}_i^*$ ve $\underline{x}_j^*$ sırasıyla, $\underline{x}_i$ ve $\underline{x}_j$ bireylerinin düzlem üzerindeki



Şekil: 1.1) Ana Düzlem

izdüşümleri ve α_i, α_j ile β_i, β_j ler de sırasıyla x_i^* ve x_j^* nin Δ_1 ve Δ_2 eksenleri üzerindeki izdüşümleridirler.

Açıktır ki, noktalar arasındaki uzaklıkların kareleri, α 'lar arasındaki uzaklıkların kareleri ile β 'lar arasındaki uzaklıkların kareleri toplamına eşittir.

$d^2(\alpha_i, \alpha_j)$ uzaklıklarını maksimum yapan Δ_1 doğrusu ile, Δ_1 doğrusuna dik olup $d^2(\beta_i, \beta_j)$ uzaklıklarını maksimum yapan Δ_2 doğrusu bulunarak, söz konusu izdüşüm düzlemi elde edilir. Ancak, nokta bulutunu üzerine izdüşüreceğimiz alt uzayın boyutu k ise, Δ_1 ve Δ_2 doğrularından başka, $\Delta_3, \Delta_4, \dots, \Delta_k$ gibi $(k-2)$ tane daha ortogonal eksenin benzer yolla bulunması gerekmektedir. Araştırılacak olan bu doğrulara ana eksenler adı verilir.

1.1.5. NOKTA BULUTUNUN YERİ

Analizde, nokta bulutunun orijine göre durumundan çok, uzaydaki şekliyle ilgilendiğimizden, ana bileşenlerin orijinden geçmesi gerekmez. Ancak, pratik birtakım yararlar sağlaması bakımından, söz konusu bileşenlerin ağırlık merkezinden geçmesi sağlanır.

Bulutun ağırlık merkezine karşılık gelen vektör g ile gösterilecek olursa, bireyler ve değişkenler için

$$g_b = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n)$$

$$g_d = (\bar{x}^1, \bar{x}^2, \dots, \bar{x}^p)$$

yazılabilir. Burada $\bar{x}_i$ ve $\bar{x}^j$ ler sırasıyla, birey ve değişkenlere ait ortalamalardır.

Böylece her bir nokta bulutu için

$$Y_i = x_i - g_b \quad (\text{Bireyler için}) \quad (1.16)$$

$$Y^j = x^j - g_d \quad (\text{Değişkenler için})$$

dönüşümleri yapılır. (1.16) formülleri yardımıyla X tablosundan Y ortalanmış tablosuna geçilmiş olur.

Birtakım pratik kolaylıklar getirmesi bakımından, başlangıç noktasını ağırlık merkezine taşıyarak, yani $g = \underline{0}$ konularak,

$$Y = X \quad (1.17)$$

elde edilir.

1.1.6. VARYANSLAR MATRİSİ

Köşegen elemanları ağırlıklardan oluşan $(n \times n)$ köşegen matrisi

$$D_P = \begin{pmatrix} p_1 & & 0 \\ & \ddots & \\ 0 & & p_n \end{pmatrix}$$

olmak üzere, Varyanslar matrisi,

$$V = X' D_P X \quad (1.18)$$

ile verilir.

R^p uzayında $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$ bireylerinin ağırlıkları için ana bileşenli faktör analizinde

$$p_1 = p_2 = \dots = p_n = 1 \quad (1.19)$$

alınır. Bu durumda $D_p = I$ olacağından varyanslar matrisi

$$V = X'X \quad (1.20)$$

olarak elde edilir.

Yukarıdaki V matrisini bir kuvadratik form olarak düşünebiliriz. R^p uzayının $\underline{e}_i$ ve $\underline{e}_j$ birim vektörleri gözönüne alınırsa

$$\begin{aligned} V(\underline{e}_i, \underline{e}_j) &= \underline{e}_i' V \underline{e}_j = \underline{e}_i' X'X \underline{e}_j \\ &= \sum_{k=1}^n x_k^i x_k^j = \text{Var}(\underline{x}^i, \underline{x}^j) \end{aligned} \quad (1.21)$$

olduğu görülür. Bu da göstermektedir ki, V matrisinin elemanları, değişkenler arasındaki varyans-kovaryans olarak ortaya çıkmaktadır.

1.1.7. METRİK

Ana bileşenli faktör analizinde kolaylık sağlaması yönünden metrik olarak

$$D_a = \begin{pmatrix} a_1 & & 0 \\ & \ddots & \\ 0 & & a_p \end{pmatrix}$$

şeklindeki köşegen matrisler kullanılır. Böyle bir metrik verilince, değişken dönüşümleri yardımıyla, adi Euclid metriğine kolayca geçilebilir. Bunun için her bir $\underline{x}^i$ değişkenine karşılık,

$$y^i = \sqrt{a_i} x^i \quad (1.22)$$

dönüşümünü yapmak yeterlidir. Bu sonucun, aşağıdaki (Teorem 1.1) ile, daha genel haller için de geçerli olacağı gösterilebilir.

Teorem 1.1: Cholesky Doolittle Teoremi.

p boyutlu R^p uzayında, M Euklidyen bir metrik olmak üzere, daima

$$M = T'T \quad (1.23)$$

olacak şekilde $(p \times p)$ boyutlu T matrisi vardır.

İspat: p boyutlu R^p uzayı, R (Reel sayılar) cismi üzerinde tanımlanmış ve R^p nin vektörleri, bu uzaydaki Kanonik Baz'a göre ifade edilmiş olsun. Ayrıca R^p de $(p \times p)$ M metriği ile $(p \times p)$ I Kanonik Euclid metriği tanımlanmış olsun. Bu halde M 'ye göre skaler çarpım,

$$\text{Bütün } \underline{V}_1, \underline{V}_2 \in M \text{ için } \langle \underline{V}_1, \underline{V}_2 \rangle_M = M(\underline{V}_1, \underline{V}_2) = \underline{V}_1' M \underline{V}_2$$

şeklinde tanımlanır. Burada $\underline{V}_1'$ ve $\underline{V}_2$ sırasıyla $\underline{V}_1$ ve $\underline{V}_2$ vektörlerine karşılık gelen matrislerdir.

M matrisi için aşağıdaki özellikler yazılabilir.

- i) $M' = M$ (Simetrik)
- ii) $\underline{V}' M \underline{V} \geq 0$ (Pozitif)
- iii) $\underline{V}' M \underline{V} = 0 \Leftrightarrow \underline{V} = '0$ (Definit)

Kanonik Euclid metriğine göre skaler çarpım ise,

$$\langle \underline{V}_1, \underline{V}_2 \rangle_{I_p} = I_p (\underline{V}_1, \underline{V}_2) = \underline{V}_1' I_p \underline{V}_2 = \underline{V}_1' \underline{V}_2$$

olarak tanımlanır.

Diğer taraftan R^p 'nin kanonik baz vektörlerini M 'nin özvektörlerine karşı getiren ortogonal operatör F olsun, böylece

$$F' = F^{-1}$$

yazılır. Ve M 'nin özdeğerlerinden oluşan köşegen matris ise

$$D_\lambda = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_p \end{pmatrix}$$

şeklindedir. M pozitif definit olduğundan tüm özdeğerleri pozitiftir.

M matrisini özdeğerler matrisine bağlayan ifade

$$M = F' D_\lambda F \quad (1.24)$$

şeklindedir. Böylece,

$$M(\underline{V}_1, \underline{V}_2) = \underline{V}_1' M \underline{V}_2 = \underline{V}_1' F' D_\lambda F \underline{V}_2$$

$$M(\underline{V}_1, \underline{V}_2) = (F \underline{V}_1)' D_\lambda (F \underline{V}_2) = D_\lambda (F \underline{V}_1, F \underline{V}_2) \quad (1.25)$$

elde edilir. (1.25) eşitliğinden şu sonuçlar çıkartılabilir;

i) Her $\underline{V} \in R^p$ için $F \underline{V}$ vektörü, M 'nin özvektörlerine göre $\underline{V}$ nin bir ifadesidir,

$$\text{ii) } D_\lambda (F \underline{V}) = \lambda F \underline{V} = M(F \underline{V}) \text{ dir,}$$

iii) (ii) sonucu F^{-1} ile soldan çarpılarak

$$F^{-1} D_\lambda F \underline{V} = \lambda \underline{V} \quad (F' = F^{-1})$$

elde edilir.

Şu halde;

$$\langle \underline{V}_1, \underline{V}_2 \rangle_M = \underline{V}_1' M \underline{V}_2 = \underline{V}_1' F' D_\lambda F \underline{V}_2$$

yazılabilir.

$D_{\sqrt{\lambda}}$, M nin özdeğerlerinin kare köklerinden oluşan köşegen matris olsun. Bu halde

$$D_\lambda = D_{\sqrt{\lambda}}' D_{\sqrt{\lambda}}$$

olacağından

$$M = F' D_{\sqrt{\lambda}}' I_p D_{\sqrt{\lambda}} F$$

yazılabilir.

Γ , kanonik metriğe göre ortogonal bir operatör olmak üzere,

$$T = \Gamma D_{\sqrt{\lambda}} F$$

şeklinde bir matris tanımlayalım. Bu taktirde,

$$T'T = F' D_{\sqrt{\lambda}}' \Gamma' \Gamma D_{\sqrt{\lambda}} F$$

yazılabilir. $\Gamma' \Gamma = I$ konularak,

$$T'T = F' D_{\sqrt{\lambda}}' D_{\sqrt{\lambda}} F$$

elde edilir. Böylece,

$$T'T = F' D_\lambda F = M$$

bulunarak teorem ispatlanmış olur.

Ayrıca bu teoremden

$$\begin{aligned} \langle \underline{x}_i, \underline{x}_j \rangle_M &= \underline{x}_i' M \underline{x}_j = \underline{x}_i' T'T \underline{x}_j \\ &= (T \underline{x}_i)' (T \underline{x}_j) = (T \underline{x}_i)' I_p (T \underline{x}_j) \\ &= \langle T \underline{x}_i, T \underline{x}_j \rangle_{I_p} \end{aligned}$$

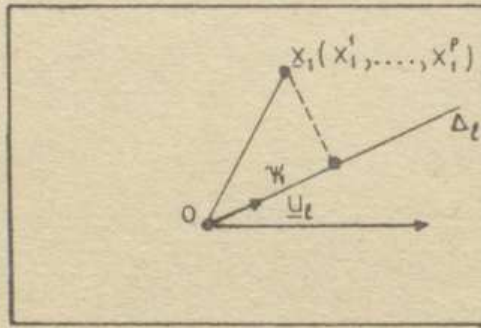
sonucu da yazılabilir. Şu halde I_p metriği ile işlem yapmak istendiğinde X tablosu yerine

$$Y = X' T \quad (1.26)$$

almak yeterlidir.

1.1.8. BİREYLER BULUTU

$\underline{x}^1, \underline{x}^2, \dots, \underline{x}^p$ ler veri tablosundan elde edilen değişkenleri göstermek üzere ilk koordinat sistemi verilmiş olsun. $\underline{x}_1$ bireyinin bileşenleri açık olarak, bu eksen sistemi üzerindeki izdüşümlerdir (Şekil: 1.2).



Şekil: 1.2) Bireyler bulutunun gösterilimi

Şimdi bireyler bulutuna ait $\underline{x}_i$ ($i = 1, \dots, n$) elemanların, başlangıçtan geçen herhangi bir Δ_l doğrusu üzerindeki izdüşümlerini bulalım. Bunun için, $\underline{u}_l$, Δ_l ekseninde M -Normu 1 olan birim vektör seçilirse, bu taktirde herhangi bir $\underline{x}_i$ bireyinin Δ_l doğrusu üzerindeki izdüşümünün cebirsel ölçümü,

$$T_{i,l} = \langle \underline{u}_l, \underline{x}_i \rangle_M = \underline{u}_l' M \underline{x}_i \quad (1.27)$$

şeklindedir. Böylece tüm bireylerin Δ_l üzerindeki izdüşümleri için

$$\begin{aligned}
 \psi_{1,\ell} &= \langle \underline{U}_\ell, \underline{x}_1 \rangle_M = \underline{U}_\ell^T M \underline{x}_1 \\
 \psi_{2,\ell} &= \langle \underline{U}_\ell, \underline{x}_2 \rangle_M = \underline{U}_\ell^T M \underline{x}_2 \\
 &\vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \\
 \psi_{n,\ell} &= \langle \underline{U}_\ell, \underline{x}_n \rangle_M = \underline{U}_\ell^T M \underline{x}_n
 \end{aligned} \tag{1.28}$$

yazılabilir. M simetrik olduğundan (1.27) ifadesi

$$\psi_{1,\ell} = (M \underline{U}_\ell^T)^T \underline{x}_1 \tag{1.29}$$

ve

$$\underline{C}^\ell = M \underline{U}_\ell \tag{1.30}$$

konularak,

$$\psi_{i,\ell} = (\underline{C}^\ell)^T \underline{x}_i = \sum_{j=1}^P C_j^\ell x_i^j \quad (i = 1, \dots, n) \tag{1.31}$$

şeklinde gösterilebilir. Böylece $\psi_{i,\ell}$ bileşenlerinden oluşan vektör $\underline{\psi}_\ell$ olmak üzere,

$$\underline{\psi}_\ell = [\psi_{1,\ell}, \psi_{2,\ell}, \dots, \psi_{n,\ell}]$$

yazılabilir. X veri tablosunu göstermek üzere,

$$\underline{\psi}_\ell = X \underline{C}^\ell \tag{1.32}$$

olduğu açıktır.

$M = I_P$ alınması halinde

$$\underline{C}^\ell = \underline{U}_\ell \tag{1.33}$$

olacağından,

$$\underline{\psi}_\ell = X \underline{U}_\ell \tag{1.34}$$

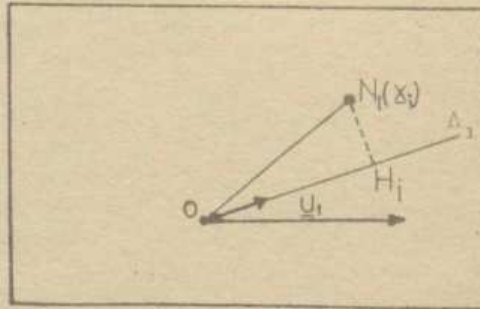
elde edilir.

Böylece $\underline{\Psi}_k$ ler, ilk p değişkenin $\underline{U}_k$ faktörü yardımıyla lineer kombinasyonu olarak ifade edilmiş olur.

1.2. ANA BİLEŞENLERİN ÖZVEKTÖRLER YARDIMIYLA ELDE EDİLMESİ

1.2.1. ANA BİLEŞENLERİN HESABI

Bundan önceki bölümde, $(n \times p)$ X tablosundan, birim vektörü $\underline{U}_k$ olan Δ_k eksenini üzerinde p boyutlu bir $\underline{\Psi}_k$ vektörüne geçiş formülü çıkartılmıştı (1.34). Bu bölümde ana bileşenler ile varyans-kovaryans matrisinin özvektörleri arasındaki ilişki incelenmektedir. Bunun için,



Şekil: 1.3) Herhangi bir bireye ait birinci ana bileşen.

(Şekil: 1.3)te görüldüğü gibi, herhangi bir x_i bireyi-ne ait nokta N_i olmak üzere, bu noktaların uzaydaki yerleri belli olduklarından, orijine olan ON_i uzaklıkları ve dolayısıyla bu uzaklıkların kareleri toplamı sabittir. Yani,

$$\sum_{i=1}^n \overline{ON_i}^2 = \text{Sabit} \quad (1.35)$$

yazılabilir. Diğer taraftan Pithagor teoremine göre,

$$\sum_{i=1}^n \overline{ON}_i^2 = \sum_{i=1}^n \overline{OH}_i^2 + \sum_{i=1}^n \overline{N_i H_i}^2 \quad (1.36)$$

olur.

Amacımız, buluta ait noktaların izdüşümünde, uzaklıkları en az değiştiren bir alt uzay bulmaktır. Başka bir ifadeyle, nokta bulutuna en yakın alt uzayı (özellikle düzlemi) belirlemek istiyoruz. Bunun için buluta ait N_1 noktaları ile, bu noktaların söz konusu alt uzaya ait Δ_1 doğrusu üzerindeki H_1 izdüşümleri arasındaki $\overline{N_1 H_1}$ uzaklıklarının ve böylece bu uzaklıkların kareleri toplamının minimum yapılması gerekir. Diğer taraftan (1.35) ve (1.36) eşitlikleri göz önüne alınırsa, $\overline{N_1 H_1}$ uzaklıklarının kareleri toplamının minimum yapılması, $\overline{OH}_1$ uzaklıklarının kareleri toplamının maksimum yapılması demektir. Yani,

$$\min_{(\underline{U}_1)} \left(\sum_{i=1}^n \overline{M_i H_i}^2 \right) \Leftrightarrow \max_{(\underline{U}_1)} \left(\sum_{i=1}^n \overline{OH}_i^2 \right)$$

dir. Şu halde problem, $\sum_{i=1}^n \overline{OH}_i^2$ toplamını maksimum yapacak şekilde bir alt uzayın belirlenmesi problemine indirgenmiş olur. Böylece, alt uzaya ait Δ_1 gibi bir doğrunun bulunması için,

$$\begin{aligned} \sum_{i=1}^n \overline{OH}_i^2 &= \sum_{i=1}^n \Psi_{i,1}^2 = \underline{\Psi}_1' \underline{\Psi}_1 \\ &= (\underline{X} \underline{U}_1)' (\underline{X} \underline{U}_1) = \underline{U}_1' \underline{X}' \underline{X} \underline{U}_1 \end{aligned} \quad (1.37)$$

yazılarak,

$$\underline{U}_1' \underline{U}_1 = 1 \quad (1.38)$$

koşulu altında $\underline{U}_1' \underline{X}' \underline{X} \underline{U}_1$ ifadesini maksimum yapacak şekilde $\underline{U}_1$ vektörünü belirleyeceğiz. Böylece nokta bulutunu en iyi gösteren alt uzaya ait Δ_1 doğrusunun araştırılması problemi,

(1.38) koşulu altında $\underline{U}'_1 X' X \underline{U}_1$ ifadesini maksimum yapan $\underline{U}_1$ vektörünün araştırılması problemine dönüşmüş oldu. Bunun için λ_1 Lagrange çarpanı olmak üzere

$$L = \underline{U}'_1 X' X \underline{U}_1 - \lambda_1 (\underline{U}'_1 \underline{U}_1 - 1) \quad (1.39)$$

fonksiyonunu maksimum yapan $\underline{U}_1$ vektörünün belirlenmesi.

Extremum için gerekli olan

$$\frac{\partial L}{\partial \underline{U}_1} = 0 \quad (1.40)$$

koşulu gözönüne alınarak,

$$\frac{\partial}{\partial \underline{U}_1} (\underline{U}'_1 X' X \underline{U}_1) = 2 X' X \underline{U}_1$$

ve

$$\frac{\partial}{\partial \underline{U}_1} [\lambda_1 (\underline{U}'_1 \underline{U}_1 - 1)] = 2 \lambda_1 \underline{U}_1$$

konulursa,

$$X' X \underline{U}_1 = \lambda_1 \underline{U}_1 \quad (1.41)$$

elde edilir. (1.41) eşitliğinin ortaya koyduğu sonuç şudur;

Nokta bulutunu en iyi gösteren alt uzaya ait Δ_1 doğrusu, $X' X$ matrisinin λ_1 özdeğerine karşılık gelen $\underline{U}_1$ özvektörü yardımıyla belirlenmektedir.

Diğer taraftan

$$\underline{U}'_1 X' X \underline{U}_1 = \underline{U}'_1 (X' X \underline{U}_1) = \underline{U}'_1 \lambda_1 \underline{U}_1 = \lambda_1 \quad (1.42)$$

olduğundan, gerçekte maksimum yapılması gereken ifade λ_1 özdeğeridir.

Şimdi (1.34) eşitliği ile belirlenen $\underline{\Psi}_1$ bileşeninin varyansını hesaplayalım.

$$\text{Var} (\underline{\Psi}_1) = \underline{\Psi}_1' \underline{\Psi}_1 = (X \underline{U}_1)' (X \underline{U}_1) = \lambda_1 \quad (1.43)$$

Böylece, $X' X$ matrisinin en büyük özdeğeri, birinci ana bileşenin varyansı olarak elde edilir.

Şimdi de, R^D uzayında başlangıçtan geçen ve nokta bulutunun gösterilimi için en uygun olan alt uzayı belirleyelim.

Aranan düzlem, daha önce bulunan Δ_1 eksenini kapsamalıdır. Aksi halde $\underline{U}_1$ vektörünü belirleyen koşullarla çelişir. O halde, aranan düzlem için $\underline{U}_1$ den başka ikinci bir $\underline{U}_2$ vektörünün belirlenmesi gerekir. Öyle ki

$$\begin{aligned} \langle \underline{U}_2, \underline{U}_2 \rangle_M &= 1 & (M\text{-Normallik koşulu}) \\ \langle \underline{U}_2, \underline{U}_1 \rangle_M &= 0 & (M\text{-Ortogonalite koşulu}) \end{aligned} \quad (1.44)$$

koşulları gerçekleşmelidir. Böylece (1.44) koşulları altında $\underline{U}_2' X' X \underline{U}_2$ ifadesi maksimum olacak şekilde, Δ_2 doğrusunun $\underline{U}_2$ birim vektörü belirlenecektir. Bu vektör ise λ_2 ve μ Lagrange çarpanları olmak üzere,

$$L = \underline{U}_2' X' X \underline{U}_2 - \lambda_2 (\underline{U}_2' \underline{U}_2 - 1) - \mu \underline{U}_2' \underline{U}_1 \quad (1.45)$$

fonksiyonunu maksimum yapan vektör olarak belirlenir. Bunun için

$$\frac{\partial L}{\partial \underline{U}_2} = 0$$

koşulundan

$$2 X' X \underline{U}_2 - 2 \lambda_2 \underline{U}_2 - \mu \underline{U}_1 = 0 \quad (1.46)$$

elde edilir. (1.46) eşitliğinin her iki yanı soldan $\underline{U}_1'$ ile çarpılarak

$$2 (\underline{U}_1' X' X) \underline{U}_2 - 2 \lambda_2 \underline{U}_1' \underline{U}_2 - \mu \underline{U}_1' \underline{U}_1 = 0 \quad (1.47)$$

bulunur. (1.41) eşitliğiyle (1.44) koşulları gözönüne alınarak (1.47)den

$$\mu = 0$$

olduğu görülür. Bu sonuç (1.46) ifadesinde yerine yazılarak,

$$X' X \underline{U}_2 = \lambda_2 \underline{U}_2 \quad (1.48)$$

elde edilir. (1.48) eşitliği, $\underline{U}_2' X' X \underline{U}_2$ ifadesini (1.44) koşulları altında maksimum yapan $\underline{U}_2$ vektörünün, $X'X$ matrisinin ikinci özdeğerine karşılık gelen özvektörü olduğunu gösterir. Diğer taraftan

$$\begin{aligned} \text{Var} (\underline{\Psi}_2) &= \underline{\Psi}_2' \underline{\Psi}_2 = (X \underline{U}_2)' (X \underline{U}_2) \\ &= \underline{U}_2' \lambda_2 \underline{U}_2 = \lambda_2 \end{aligned} \quad (1.49)$$

yazılabilir. Buradan maksimum yapılması gereken ifadenin gerçekte $X' X$ matrisinin ikinci en büyük özdeğeri ve bu özdeğerin de ikinci ana bileşenin varyansı olduğu sonucu çıkar. Benzer şekilde $X' X$ matrisinin $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_q$ özdeğerleri ve bu özdeğerlere karşılık gelen $\underline{U}_1, \underline{U}_2, \dots, \underline{U}_q$ öz vektörleri bulunarak, nokta bulutuna en yakın olan $q (q \leq p)$ boyutlu alt uzay elde edilmiş olur. Boyut sayısı $q = 2$ alınarak, bulutun düzlemdeki gösterilimi elde edilmek suretiyle inceleme yapılabilir.

Şu halde problem, $(p \times p)$ boyutlu bir kare matris $A = X' X$ ve $(p \times 1)$ boyutlu bir vektör de $\underline{x}$ olmak üzere

$$A\underline{x} = \lambda\underline{x} \quad (1.50)$$

denklemini gerçekleyen bütün özdeğerlerin belirlenmesi problemine indirgenmiş olur. (1.50) denkleminin çözümünün olabilmesi için,

$$|A - \lambda I| = 0 \quad (1.51)$$

gereklidir. Bu ise p 'inci dereceden

$$(-1)^0 \lambda^p + (-1)^1 C_1 \lambda^{p-1} + (-1)^2 C_2 \lambda^{p-2} + \dots + (-1)^p C_p = 0$$

veya

$$(\lambda - \lambda_1) (\lambda - \lambda_2) \dots (\lambda - \lambda_p) = 0$$

polinomudur. Burada

$$\begin{aligned} C_1 &= \lambda_1 + \lambda_2 + \dots + \lambda_p \\ C_2 &= \lambda_1 \lambda_2 + \lambda_1 \lambda_3 + \dots + \lambda_1 \lambda_p + \lambda_2 \lambda_3 \\ &\vdots \\ &\quad + \dots + \lambda_{p-1} \lambda_p \\ C_p &= \lambda_1 \lambda_2 \dots \lambda_p \end{aligned}$$

dır. Böylece A matrisinin Determinantı ve iz'i için

$$|A| = \lambda_1 \lambda_2 \dots \lambda_p$$

$$\text{iz}(A) = \lambda_1 + \lambda_2 + \dots + \lambda_p$$

yazılabilir.

A matrisi ile buna bağlı olarak elde edilen λ_i ($i = 1, \dots, p$) özdeğerlerinin sağladığı bazı temel özellikler vardır. Bunları teoremler halinde şöyle sıralayabiliriz.

Teorem 1.2: A simetrik bir matris ve $\lambda_1, \lambda_2, \dots, \lambda_p$ ler A matrisinin özdeğerleri olmak üzere,

$$D_\lambda = \begin{pmatrix} \lambda & & & 0 \\ & 1 & & \\ & & \ddots & \\ 0 & & & \lambda_p \end{pmatrix}$$

olsun. Bu taktirde,

$$U^{-1} A U = D_{\lambda} \quad (1.52)$$

olacak şekilde daima bir U ortogonal matrisi vardır.

İspat: U ortogonal matrisi, kolonları A 'nın öz vektörlerinden oluşan

$$U = [\underline{x}_1, \underline{x}_2, \dots, \underline{x}_p]$$

matrisi olsun. Buradan

$$AU = A [\underline{x}_1, \underline{x}_2, \dots, \underline{x}_p]$$

veya,

$$AU = [\lambda \underline{x}_1, \lambda \underline{x}_2, \dots, \lambda \underline{x}_p]$$

yazılabilir. Diğer taraftan,

$$\lambda \underline{x}_1 = \lambda_1 \underline{x}_1, \dots, \lambda \underline{x}_p = \lambda_p \underline{x}_p$$

olduğundan

$$AU = [\lambda_1 \underline{x}_1, \lambda_2 \underline{x}_2, \dots, \lambda_p \underline{x}_p] \quad (1.53)$$

elde edilir. Son eşitliğin ikinci tarafının UD_{λ} olduğu göz-önüne alınarak,

$$AU = UD_{\lambda} \quad (1.54)$$

yazılabilir. U matrisi tekil olmadığından (1.54) denkleminin her iki tarafı soldan U^{-1} ile çarpılarak

$$U^{-1} A U = D_{\lambda} \quad (1.55)$$

elde edilir. Şayet U matrisinin her bir kolonunda bulunan $\underline{x}_i$ ($i = 1, \dots, p$) vektörü normlanırsa

$$U' = U^{-1}$$

olacağından (1.55) eşitliği

$$U' A U = D_{\lambda}$$

yazılarak teorem ispatlanmış olur.

Teorem 1.3: Bir A matrisinin bütün özdeğerleri birbirinden farklı ise, bu özdeğerlere karşılık gelen özvektörler de birbirinden farklıdır.

İspat: A matrisinin λ_r ve λ_s gibi farklı iki özdeğerine karşılık gelen özvektörler sırasıyla $\underline{x}_r$, $\underline{x}_s$ ve $\underline{x}_r = \underline{x}_s$ olsun. Bu durumda (1.50)den

$$A \underline{x}_r = \lambda_r \underline{x}_r, \quad A \underline{x}_s = \lambda_s \underline{x}_s$$

yazılabilir. İkinci denklemden $\underline{x}_s = \underline{x}_r$ konulup birinci denklemden çıkartılırsa,

$$(\lambda_r - \lambda_s) \underline{x}_r = 0$$

elde edilir. $\underline{x}_r \neq 0$ olduğundan $\lambda_r = \lambda_s$ olduğu görülür. Bu ise varsayımla çelişir. O halde $\underline{x}_r \neq \underline{x}_s$ olmalıdır.

Teorem 1.4: Reel simetrik bir matrisin bütün özdeğerleri reeldir.

İspat: $\lambda_1, \lambda_2, \dots, \lambda_p$ ler verilen A matrisinin özdeğerleri ve $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_p$ ler de bu özdeğerlere karşılık gelen özvektörler olsun. Bu durumda (1.50)den

$$A \underline{x}_r = \lambda_r \underline{x}_r \quad (r = 1, \dots, p) \quad (1.56)$$

yazılabilir. Varsayalım ki λ_r ($r = 1, \dots, p$) ler kompleks. Bu durumda herhangi bir $\lambda_r = \alpha + i\beta$ şeklinde yazılabilir.

(1.56) denkleminin her iki yanının kompleks eşleniği alınarak,

$$A \underline{x}_r^* = \lambda_r^* \underline{x}_r^* \quad (1.57)$$

(Burada * lı harfler eşleniği göstermektedir)

elde edilir. (1.56) ve (1.57)'nin her iki yanını sırasıyla $\underline{x}_r^{*'} ve \underline{x}_r'$ ile soldan çarparak

$$\begin{aligned} \underline{x}_r^{*'} A \underline{x}_r &= \lambda_r^* \underline{x}_r^{*'} \underline{x}_r \\ \underline{x}_r' A \underline{x}_r^* &= \lambda_r^* \underline{x}_r' \underline{x}_r^* \end{aligned} \quad (1.58)$$

bulunur. İkinci denklemin her iki yanının transpozesi alınıp $A' = A$ konursa,

$$\underline{x}_r^{*'} A \underline{x}_r = \lambda_r^* \underline{x}_r^{*'} \underline{x}_r \quad (1.59)$$

elde edilir.

(1.59) denkleminde (1.58)'deki birinci eşitlik çıkarılarak,

$$(\lambda_r^* - \lambda_r) \underline{x}_r^{*'} \underline{x}_r = 0 \quad (1.60)$$

bulunur. $\underline{x}_r^{*'} \underline{x}_r > 0$ olduğundan

$$\lambda_r^* - \lambda_r = 2 i\beta = 0$$

buradan da $\beta = 0$ bulunur. Yani λ_r özdeğerleri reeldirler.

Teorem 1.5: A herhangi bir kare matris olmak üzere, A'nın normlu özvektörlerinden oluşan matris U olsun. Bu durumda A ile $U' A U$ matrisinin özdeğer ve özvektörleri aynıdır.

İspat:
 $|U' A U - \lambda I| = 0$

dır. U ortogonal olduğundan $U' U = I$ yazılır. Böylece,

$$|U' A U - \lambda U' U| = 0$$

veya

$$|U' (A - \lambda I) U| = 0 \quad (1.61)$$

olur. Ve bu da

$$|U'| |A - \lambda I| |U| = 0 \quad (1.62)$$

şeklinde yazılabilir. Böylece (1.62) denkleminde

$$|A - \lambda I| = 0$$

elde edilir. Bu ise A matrisinin karakteristik denklemdir. O halde, $U' A U$ ile A matrisinin karakteristik denklemi aynıdır. Bu sonuç, tekil olmayan herhangi bir matris Q olmak üzere $Q A Q^{-1}$ matrisi için de geçerlidir. Çünkü;

$$|Q A Q^{-1} - \lambda I| = |Q (A - \lambda I) Q^{-1}| \quad (1.63)$$

yazılabilir. Ve benzer şekilde

$$|Q A Q^{-1} - \lambda I| = |Q Q^{-1}| |A - \lambda I| ,$$

yazılarak $Q Q^{-1} = I$ den

$$|Q A Q^{-1} - \lambda I| = |A - \lambda I|$$

olduğu görülür.

Teorem 1.6: Pozitif definit reel simetrik bir matrisin bütün özdeğerleri pozitifdir.

İspat: A matrisi pozitif definit reel simetrik bir matris olmak üzere, A'nın özdeğerleri λ_i ($i = 1, \dots, p$) ve bu özdeğerlere karşılık gelen özvektörler de $\underline{x}_i$ olsunlar (1.50)den dolayı

$$A \underline{x}_i = \lambda_i \underline{x}_i \quad (i = 1, \dots, p)$$

yazılabilir. Bu denklemin her iki yanı soldan $\underline{x}_i'$ ile çarpılarak,

$$\underline{x}_i' A \underline{x}_i = \lambda_i \underline{x}_i' \underline{x}_i$$

elde edilir. $\underline{x}_i' A \underline{x}_i$ ve $\underline{x}_i' \underline{x}_i$ skaler olduklarından,

$$\lambda_i = \frac{\underline{x}_i' A \underline{x}_i}{\underline{x}_i' \underline{x}_i}$$

bulunur. $\underline{x}_i' A \underline{x}_i > 0$ ve $\underline{x}_i' \underline{x}_i > 0$ olduğundan her $i \in [1, p]$ için $\lambda_i > 0$ olduğu görülür.

Teorem 1.7: Reel simetrik bir matrisin farklı özvektörleri ortogonaldır.

İspat: $i \neq k$ olmak üzere, λ_i ve λ_k reel simetrik bir matrisin farklı iki özdeğeri ve $\underline{x}_i, \underline{x}_k$ da sırasıyla bu özdeğerlere karşılık gelen özvektörler ise

$$A \underline{x}_i = \lambda_i \underline{x}_i$$

$$A \underline{x}_k = \lambda_k \underline{x}_k$$

yazılabilir. Birinci denklemin her iki yanını $\underline{x}_k'$ ve ikinci denklemin her iki yanını da $\underline{x}_i'$ ile soldan çarparak,

$$\underline{x}_k' A \underline{x}_i = \lambda_i \underline{x}_k' \underline{x}_i \quad (1.64)$$

$$\underline{x}_i' A \underline{x}_k = \lambda_k \underline{x}_i' \underline{x}_k$$

elde edilir. (1.64)deki birinci denklemde transpoze alınıp $A' = A$ konularak,

$$\underline{x}_i' A \underline{x}_k = \lambda_i \underline{x}_i' \underline{x}_k \quad (1.65)$$

ve bundan (1.64)deki ikinci denklem çıkarılarak

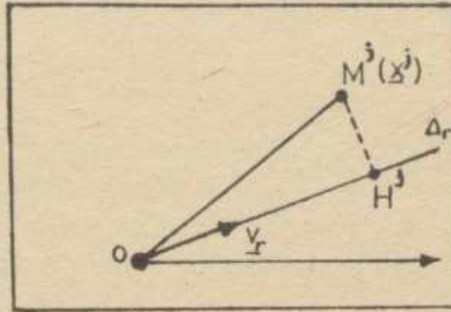
$$(\lambda_i - \lambda_k) \underline{x}_i' \underline{x}_k = 0$$

bulunur. $\lambda_i \neq \lambda_k$ olduğundan, $\underline{x}_i' \underline{x}_k = 0$

olur. Bu da özvektörlerin ortogonal olduğunu gösterir.

1.2.2. DEĞİŞKENLER BULUTUNUN R^n UZAYINDA İNCELENMESİ

X tablosundan elde edilen değişkenler R^n uzayında bir $N = \{ \underline{x}^j \mid j = 1, \dots, p \}$ bulutu oluşturur. Buluta ait herhangi bir $\underline{x}^j$ değişkenini göz önüne alalım (Şekil: 1.4).



Şekil: 1.4) Değişkenler Bulutunun Gösterilimi

$\underline{x}^j$ değişkenine uzayda bir M^j noktası karşılık geleceğinden, $\overline{OM}^j$ uzaklığı sabittir. Böylece,

$$\sum_{j=1}^p \overline{OM}^{j2} = \text{Sabit} \quad (1.66)$$

yazılabilir. Pithagor teoremi kullanılırsa,

$$\sum_{j=1}^p \overline{OM}^{j2} = \sum_{j=1}^p \overline{OH}^{j2} + \sum_{j=1}^p \overline{M^jH^j}^2$$

elde edilir. Burada H^j noktaları, M^j 'nin Δ_r eksenini üzerindeki izdüşümüdür. Amacımız, değişkenlerin oluşturduğu nokta bulutuna en yakın alt uzayı belirlemek olduğundan, $\overline{M^jH^j}^2$ uzaklıkları toplamının minimum yapılması gerekmektedir. Başka bir ifadeyle, $\overline{OH}^{j2}$ uzaklıkları toplamını maksimum yapacak şekilde Δ_r doğrularını belirleyeceğiz.

$\underline{x}^j$ değişkeninin Δ_r üzerindeki bileşenlerinden oluşan vektör $\underline{\psi}$ olmak üzere,

$$\begin{aligned} \sum_{j=1}^p \overline{OH}^j{}^2 &= \underline{\Psi}' \underline{\Psi} = (\underline{X}' \underline{V}_r)' (\underline{X}' \underline{V}_r) \\ &= \underline{V}_r' \underline{X} \underline{X}' \underline{V}_r \end{aligned} \quad (1.67)$$

yazılabilir. Bu durumda

$$\underline{V}_r' \underline{V}_r = 1 \quad (1.68)$$

koşulu altında $\underline{V}_r' \underline{X} \underline{X}' \underline{V}_r$ ifadesinin maksimum yapılması gerekmektedir. Bireylerle ilgili bileşenlerin hesabında daha önce izlenen yol, değişkenler için de aynen tekrarlanırsa $\underline{V}_r$ vektörlerinin bulunması için $\underline{X}\underline{X}'$ matrisinin $\mu_1, \mu_2, \dots, \mu_s$ ($s \leq n$) özdeğerlerinin bulunması problemi ortaya çıkar. Böylece, $\mu_1, \dots, \mu_s$ özdeğerlerine karşılık gelen $\underline{V}_1, \underline{V}_2, \dots, \underline{V}_s$ özvektörleri aranan eksenlere ait birim vektörler olarak elde edilir.

1.2.3. R^p ve R^n UZAYLARI ARASINDAKİ İLİŞKİ

R^p ve R^n uzaylarında, ana eksenleri taşıyan özvektörler sırasıyla $(\underline{U}_1, \underline{U}_2, \dots, \underline{U}_p), (\underline{V}_1, \underline{V}_2, \dots, \underline{V}_n)$ ve bu özvektörlere karşılık gelen özdeğerler de $(\lambda_1, \dots, \lambda_p), (\mu_1, \dots, \mu_n)$ olsun. Bu taktirde

$$\begin{aligned} \underline{X}' \underline{X} \underline{U}_\alpha &= \lambda_\alpha \underline{U}_\alpha & (\alpha = 1, \dots, p) \\ \underline{X}\underline{X}' \underline{V}_\alpha &= \mu_\alpha \underline{V}_\alpha & (\alpha = 1, \dots, n) \end{aligned} \quad (1.69)$$

yazılabilir. (1.69)daki ikinci eşitliğin her iki yanını soldan $\underline{X}'$ ile çarpılırsa,

$$(\underline{X}' \underline{X}) \underline{X}' \underline{V}_\alpha = \mu_\alpha (\underline{X}' \underline{V}_\alpha) \quad (1.70)$$

elde edilir. Buradan $\underline{X}' \underline{V}_\alpha$ vektörünün, $\underline{X}' \underline{X}$ matrisinin özvektörlerinden biri olduğu sonucu çıkar. O halde

$$\begin{aligned} \lambda_{\alpha} &= \mu_{\alpha} \\ \frac{U}{\gamma_{\alpha}} &= c X' \frac{V}{\gamma_{\alpha}} \end{aligned} \quad (1.71)$$

(Burada c bir sabittir)

yazılabilir. Diğer taraftan $\frac{U'}{\gamma_{\alpha}} \frac{U}{\gamma_{\alpha}} = 1$ olduğundan,

$$(c X' \frac{V}{\gamma_{\alpha}})' (c X' \frac{V}{\gamma_{\alpha}}) = 1 \quad (1.72)$$

veya

$$c^2 \frac{V'}{\gamma_{\alpha}} X X' \frac{V}{\gamma_{\alpha}} = c^2 \lambda_{\alpha} \frac{V'}{\gamma_{\alpha}} \frac{V}{\gamma_{\alpha}} = 1 \quad (1.73)$$

yazılabilir. Burada $\frac{V'}{\gamma_{\alpha}} \frac{V}{\gamma_{\alpha}} = 1$ olduğundan

$$c = \frac{1}{\sqrt{\lambda_{\alpha}}} \quad (1.74)$$

elde edilir. Böylece R^D uzayından R^n uzayına geçiş formülleri olarak bilinen aşağıdaki formüller elde edilir.

$$\begin{aligned} \frac{U}{\gamma_{\alpha}} &= c X' \frac{V}{\gamma_{\alpha}} = \frac{1}{\sqrt{\lambda_{\alpha}}} X' \frac{V}{\gamma_{\alpha}} \\ \frac{V}{\gamma_{\alpha}} &= \frac{1}{\sqrt{\lambda_{\alpha}}} X \frac{U}{\gamma_{\alpha}} \end{aligned} \quad (1.75)$$

1.2.4. ANA BİLEŞENLERE GEÇİŞ FORMÜLLERİ

R^D uzayındaki n bireyin $\frac{U}{\gamma_{\alpha}}$ üzerindeki bileşenler vektörü, (1.75)deki ikinci denklem gözönüne alınarak,

$$\frac{V}{\gamma_{\alpha}} = X \frac{U}{\gamma_{\alpha}}$$

yazılır ve

$$\frac{V}{\gamma_{\alpha}} = \sqrt{\lambda_{\alpha}} \frac{V}{\gamma_{\alpha}} \quad (1.76)$$

formülüne göre hesaplanır.

R^n uzayındaki p değişkenin $\frac{V}{\gamma_{\alpha}}$ üzerindeki bileşenler vektörü ise,

$$\hat{\Psi}_{\alpha} = X' \underline{V}_{\alpha} = \sqrt{\lambda_{\alpha}} \underline{U}_{\alpha} \quad (1.77)$$

bağıntısı yardımıyla hesaplanır, Açık olarak $\underline{V}_{\alpha}$ ve $\hat{\Psi}_{\alpha}$ bileşen vektörleri bütün α 'lar için,

$$\underline{\Psi}'_{\alpha} \underline{\Psi}_{\alpha} = \hat{\underline{\Psi}}'_{\alpha} \hat{\underline{\Psi}}_{\alpha} = \lambda_{\alpha} \quad (1.78)$$

yazılabilir. Bu son bağıntıdan bireyler ve değişkenlerin α 'yıncı ana eksen üzerindeki bileşenlerinin varyansları birbirine ve α 'yıncı özdeğere eşittir.

1.2.5. VERİ TABLOSUNUN ANA BİLEŞENLER YARDIMIYLA YENİDEN KURULUŞU

(1.76) formülünün her iki yanı sağdan $\underline{U}'_{\alpha}$ ile çarpılırsa,

$$X \underline{U}_{\alpha} \underline{U}'_{\alpha} = \sqrt{\lambda_{\alpha}} \underline{V}_{\alpha} \underline{U}'_{\alpha} \quad (1.79)$$

elde edilir. Bu son ifadede α üzerinde toplam alınırsa,

$$X \left(\sum_{\alpha=1}^P \underline{U}_{\alpha} \underline{U}'_{\alpha} \right) = \sum_{\alpha=1}^P \sqrt{\lambda_{\alpha}} \underline{V}_{\alpha} \underline{U}'_{\alpha} \quad (1.80)$$

bulunur.

$\underline{U}_{\alpha}$ vektörlerinden ibaret matrisi U ile gösterelim;

$$U = (\underline{U}_1, \underline{U}_2, \dots, \underline{U}_p)$$

Açık olarak,

$$U' = \begin{pmatrix} \underline{U}'_1 \\ \underline{U}'_2 \\ \vdots \\ \underline{U}'_p \end{pmatrix}$$

yazılabilir. Böylece.

$$U U' = \sum_{\alpha=1}^p \underline{u}_{\alpha} \underline{u}_{\alpha}' \quad (1.81)$$

elde edilir. $\underline{u}_{\alpha}$ lar birim vektörler olduklarından

$$U U' = I_p$$

yazılabilir. Buradan (1.80) formülü

$$X U U' = \sum_{\alpha=1}^p \sqrt{\lambda_{\alpha}} \frac{\underline{v}_{\alpha}}{\sqrt{\lambda_{\alpha}}} \underline{u}_{\alpha}' \quad (1.82)$$

veya

$$X = \sum_{\alpha=1}^p \sqrt{\lambda_{\alpha}} \frac{\underline{v}_{\alpha}}{\sqrt{\lambda_{\alpha}}} \underline{u}_{\alpha}' \quad (1.83)$$

olarak bulunur. Bu formül ilk olarak 1936-1939 yıllarında ECKART-YOUNG tarafından verilmiştir (70).

(1.83) formülünü değişik şekillerde de ifade etmek mümkündür. $\underline{v}_{\alpha}$ ve $\underline{u}_{\alpha}'$ 'nin (1.76) ve (1.77) ile verilen değerleri yerlerine konularak,

$$X = \sum_{\alpha=1}^p \sqrt{\lambda_{\alpha}} \frac{\underline{v}_{\alpha}}{\sqrt{\lambda_{\alpha}}} \frac{\underline{\hat{v}}_{\alpha}'}{\sqrt{\lambda_{\alpha}}} = \sum_{\alpha=1}^p \frac{1}{\lambda_{\alpha}^{1/2}} \underline{v}_{\alpha} \underline{\hat{v}}_{\alpha}' \quad (1.84)$$

bulunur. Bilindiği üzere $\underline{v}_{\alpha}$ lar bireylerin λ_{α} ana eksenleri üzerindeki, $\underline{\hat{v}}_{\alpha}'$ lar da değişkenlerin buna karşı gelen ana eksenler üzerindeki bileşenleridir. (1.84) formülüne dikkat edilirse, veri tablosundaki değerler, ana bileşenler yardımıyla tam olarak ifade edilebilmektedir. Eğer toplam ilk q ($q < p$) bileşen için hesaplanacak olursa, o takdirde X tablosunun gerçek değeri yerine X^* ile göstereceğimiz yaklaşık değeri,

$$X = X^* = \sum_{\alpha=1}^q \frac{1}{\lambda_{\alpha}^{1/2}} \underline{v}_{\alpha} \underline{\hat{v}}_{\alpha}' \quad (1.85)$$

bağıntısıyla elde edilir.

1.2.6. GÖSTERİLİMİN KALİTESİ

Euclid metriğinin tanımlandığı R^p uzayındaki Δ_α eksen sistemine göre, bireylerin oluşturduğu nokta bulutunu tekrar gözönüne alalım. Ortalanmış X tablosu n bireyden oluşan bir kolon matris olarak yorumlanabiliyordu (1.1). $\underline{x}_i$ bireyine karşılık gelen N_1 noktasının Δ_α eksen üzerindeki bileşeni $\psi_{i\alpha}$ olmak üzere, bu eksen üzerindeki gösterilimin kalitesi

$$r_{i,\alpha}^2 = \frac{\psi_{i\alpha}^2}{\overline{ON}_1^2} \quad (1.86)$$

olarak tanımlanır. Diğer taraftan

$$\overline{ON}_1^2 = \langle \underline{x}_1, \underline{x}_1 \rangle_{I_p} = \underline{x}_1' \underline{x}_1 = \sum_{j=1}^p (x_1^j)^2 \quad (1.87)$$

olduğundan (1.86) ifadesi

$$r_{i,\alpha}^2 = \frac{\psi_{i\alpha}^2}{\sum_{j=1}^p (x_1^j)^2} \quad (1.88)$$

olur.

Aynı bireyin Δ_α ve Δ_β gibi iki ana eksenin belirlediği ana düzlem üzerindeki gösteriliminin kalitesi de

$$r_{i,\alpha,\beta}^2 = \frac{\psi_{i\alpha}^2 + \psi_{i\beta}^2}{\overline{ON}_1^2} \quad (1.89)$$

şeklindedir. Daha genel olarak, N_1 noktasının $\Delta_1, \Delta_2, \dots, \Delta_q$ ana eksenlerinin belirlediği ana alt uzayındaki gösteriliminin kalitesi

$$r_{i,1,2,\dots,q}^2 = \frac{\sum_{\alpha=1}^q \psi_{i\alpha}^2}{\overline{ON}_i^2} \quad (1.90)$$

veya

$$r_{i,1,\dots,q}^2 = \frac{\sum_{\alpha=1}^q \psi_{i\alpha}^2}{\sum_{j=1}^q (x_{i1}^j)^2} \quad (1.91)$$

formülü ile verilir. Açık olarak $q = p$ olması halinde

$$r_{i,1,\dots,p}^2 = 1 \quad (1.92)$$

yazılabilir. O halde gösterilimin kalitesi 0 ile 1 arasında değişen bir büyüklük olur.

1.2.7. KATILMA PAYLARI

i 'yinci bireyin α 'yinci ana eksen üzerindeki bileşeni $\psi_{i\alpha}$ olmak üzere tüm bireyler için

$$\sum_{i=1}^n \overline{OH}_i^2 = \psi_{\alpha}' \psi_{\alpha} = (X \underline{U}_{\alpha})' (X \underline{U}_{\alpha}) = \lambda_{\alpha}$$

yazılabilir. λ_{α} , α 'yinci bileşene ait varyansı ifade ettiğinden i 'yinci bireyin bu varyansa olan katılma payı $\psi_{i\alpha}^2$ ile ifade edilebilir. Şu halde i 'yinci bireyin α 'yinci ana eksenin kuruluşundaki mutlak katılma payı $KM_{i\alpha}$ ile gösterilerek

$$KM_{i\alpha} = \psi_{i\alpha}^2 \quad (1.93)$$

olarak tanımlanır. Aynı bireyin $KI_{i\alpha}$ ile gösterilen relatif katılma payı da

$$KI_{i\alpha} = \frac{\psi_{i\alpha}^2}{\lambda_{\alpha}} \quad (1.94)$$

formülü ile verilir.

Açık olarak relatif katılma payı 0 ile 1 arasında değişen bir değerdir. Zira tüm bireyler için

$$\sum_{i=1}^n KI_{ia} = 1$$

yazılabilir.

1.2.8. EK ELEMANLAR

Ek elemanlar ana bileşenli faktör analizinde önemli rol oynarlar. Bu elemanlar (birey veya değişkenler) esas elemanlar gibi faktörel eksenlerin belirlenmesine doğrudan katılmazlar. Ancak, bunlarla birlikte ana düzlem üzerine izdüşürülerek, birlikte gösterilimleri sağlanır. Bu durumu gerektiren birçok nedenler arasından en önemlilerini şu şekilde sıralayabiliriz;

i) Veri tablosu analiz edilirken konu ile ilgili yeni bazı bilgiler sağlanmış olabilir. Bu yeni bilgiler yardımıyla elde edilen elemanların doğrudan doğruya analize sokulmaları yerine, tabloya ek satır veya ek sütunlar olarak yerleştirilirler. Bu şekilde, yapılacak analiz sonucunda yeni elemanların eski elemanlarla birlikteki gösterilimi elde edilebilir.

ii) Bazı elemanların ana eksenlerin kuruluşundaki katılma payları çok yüksek olabilir. Bu durumda analiz sonuçları gerçek olayı tam olarak yansıtmaz. Gizli kalmış birtakım bilgiler, söz konusu etkili elemanlar yanında kaybolup gider. Bu durumu düzeltmek için, katılma payları çok yüksek olan elemanlar ek olarak tabloya ilave edilir. Böylece, verilerin sakladığı gerçek olayın bir tasviri kesin bir şekilde ortaya çıkar.

iii) Analiz sonucunda, birtakım elemanların hatalar içerdiği veya gerçek probleme yabancı kaldığı izlenimi doğabilir.

Bu durumda, bu elemanları atmak yerine, tabloda son satır veya son sütuna ek olarak almak daha uygun olur. Böylece bu elemanlar analize girmeseler dahi, diğer elemanlarla karşılaştırılmaları ve birlikte gösterilimleri sağlanır.

Ek olarak alınan elemanın $\underline{x}_0$ bireyi olduğunu varsayalım. Bu bireye karşılık gelen N_0 noktasının Δ_α ana eksenindeki izdüşümü,

$$\underline{U}_\alpha = \underline{x}_0' \underline{U}_\alpha$$

olacaktır. Burada $\underline{U}_\alpha$ vektörü, Δ_α ekseninin birim vektörüdür.

Ek elemanın $\underline{x}^0$ gibi bir değişken olması halinde, buna karşılık gelen T_0 noktasının Δ_r ana eksenindeki izdüşümü ise,

$$\underline{U}_r = \underline{x}^{0'} \underline{V}_r$$

olacaktır. Burada da $\underline{V}_r$ vektörü, Δ_r ana ekseninin birim vektörüdür.

Diğer taraftan $\underline{V}_r$ nin (1.76)daki değeri yerine yazılarak,

$$\underline{U}_r = \frac{1}{\sqrt{\lambda_r}} \underline{x}^{0'} \underline{V}_r \quad (1.95)$$

şeklinde ek değişkenin, bireylerin bileşenleri cinsinden ifadesi elde edilir.

1.3. NORMLU ANA BİLEŞENLER

Veri tabloları oluşturulurken, değişkenlerin bireyler üzerindeki ölçümleri, değişik ölçü birimlerine göre (Yaş, yıl olarak: Ağırlık, kilogram olarak: v.s.) yapılmış olabilir. Bunun sonucunda veri tablosundan elde edilen bireyler veya değişkenler arasındaki uzaklıkların hesabında, birtakım birey ve değişkenlere daha çok önem verilmiş olur. Bu bakımdan, bireyler arasındaki uzaklığın hesaplanmasında bütün değişkenlerin aynı önemi taşıması ve uzaklıkların yapılan ölçü birimlerinden bağımsız olması için ilk veriler üzerinde

$$y_i^j = \frac{x_i^j - \bar{x}^j}{s^j} \quad (1.96)$$

dönüşümü yapılır. Burada

$$\bar{x}^j = \sum_{i=1}^n p_i x_i^j \quad ; \quad s^j = \sqrt{\sum_{i=1}^n p_i (x_i^j - \bar{x}^j)^2}$$

dır.

Böylece yeni Y tablosundan elde edilen y^j ($j=1, \dots, p$) ler ölçü birimlerinden bağımsız olduklarından karşılaştırılabilirler. Ayrıca

$$\text{Var}(y^j) = 1 \quad (1.97)$$

olduğundan her bir y^j nin dağılımı aynı dikliği haizdir.

Metrik olarak D_p ağırlıklar matrisinin alınması halinde y^j ve y^k değişkenleri arasındaki uzaklık

$$d^2(y^j, y^k) = \sum_{i,i'=1}^p p_{ii'} (y_i^j - y_{i'}^k)^2 \quad (1.98)$$

($i = i'$ ise $p_{ii'} = p_i$ aksi halde $p_{ii'} = 0$)

formülü ile verilir. (1.96) dönüşümünden sonra gözönüne alınacak olan tablo, elemanları

$$z_i^j = \sqrt{p_i} \frac{x_i^j - \bar{x}^j}{s^j} = \sqrt{p_i} y_i^j$$

ile verilen

$$Z = D \sqrt{p} Y \quad (1.99)$$

matrisi olacaktır. Buradan, Özdeğerleri ve dolayısıyla özvektörleri bulunacak olan $(p \times p)$ matris

$$C = Z' Z = Y' D_p Y \quad (1.100)$$

olur. Burada $D_p =$ köşegen (p_i) dir. Her bir C_j^k elemanı,

$$C_j^k = \sum_{i=1}^n p_{ii} y_i^j y_i^k$$

ile verilir. Görüldüğü gibi C_j^k lar Y^j ve Y^k arasındaki korelasyon katsayılarıdır. Yani C korelasyon matrisidir. Böylece

$$Z' Z \underline{U}_\alpha = Y' D_p Y \underline{U}_\alpha = \lambda \underline{U}_\alpha \quad (1.101)$$

koşulunu sağlayan $\underline{U}_\alpha$ lar, korelasyon matrisinin özvektörleri ve λ_α lar da bunlara karşılık gelen özdeğerlerdir.

1.3.1. NORMLU BİREYLERİN BİLEŞENLERİ

Bireylerin Δ_α eksenı üzerindeki izdüşümlerinden oluşan vektör $\underline{x}_\alpha$ ile gösterilirse,

$$\underline{x}_\alpha = Y \underline{U}_\alpha \quad (1.102)$$

yazılabilir. Buradan i'yinci bireyin α 'yinci eksen üzerindeki bileşeni

$$\xi_{i\alpha} = \sum_{j=1}^p y_i^j u_{j\alpha} \quad (1.102)$$

olur. Buradan bileşenlerin ortalaması ve varyansı sırasıyla

$$E(\xi_{i\alpha}) = \sum_{i=1}^n p_i \xi_{i\alpha} = 0 \quad (\alpha=1,2,\dots,p) \quad (1.103)$$

$$\text{Var}(\xi_{i\alpha}) = \sum_{i=1}^n p_i \xi_{i\alpha}^2 = \lambda_{\alpha} \quad (\alpha=1,2,\dots,p)$$

olarak bulunur. (1.103)deki ikinci ifade matris gösterilimi ile

$$\xi_{\alpha}' D_p \xi_{\alpha} = \lambda_{\alpha}$$

yazılabilir. D_p pozitif definit karesel bir form olarak alınır

$$D_p (\xi_{\alpha}, \xi_{\alpha}) = \langle \xi_{\alpha}, \xi_{\alpha} \rangle_{D_p} = \lambda_{\alpha} \quad (1.104)$$

bulunur. Buradan da

$$\langle \frac{\xi_{\alpha}}{\sqrt{\lambda_{\alpha}}}, \frac{\xi_{\alpha}}{\sqrt{\lambda_{\alpha}}} \rangle_{D_p} = 1$$

bulunur ki bu da $\xi_{\alpha} / \sqrt{\lambda_{\alpha}}$ nin D_p ye göre normlanmış olduğu anlamına gelir.

1.3.2. NORMLU DEĞİŞKENLERİN BİLEŞENLERİ

Normlu değişkenler için aşağıdaki sonuçlar yazılabilir:

$$1) \text{ Her } j \text{ için; } d^2(G, Y^j) = 1 \quad (1.105)$$

dir. Yani herhangi bir değişkenin $0 \in G$ başlangıç noktasına olan uzaklığı 1'dir. Çünkü:

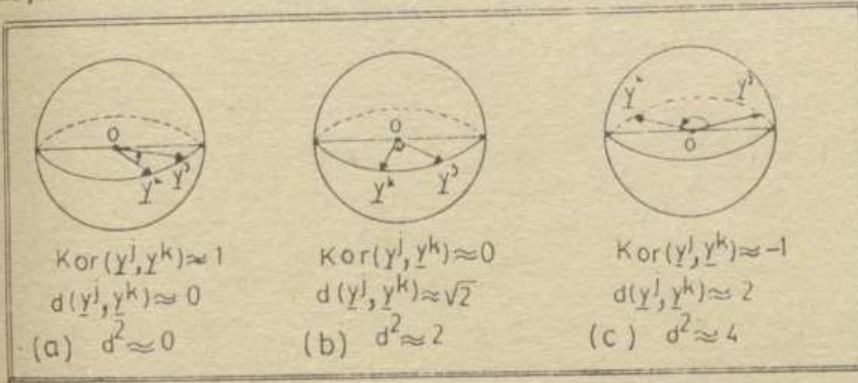
0 halde

$$0 \leq d^2(y^j, y^k) \leq 4 \quad (1.107)$$

veya

$$0 \leq d(y^j, y^k) \leq 2 \quad (1.108)$$

olduğu kolayca görülür. Bu da gösteriyor ki, herhangi iki değişken arasındaki uzaklık 0 ile 2 arasında değişmektedir. Ekstrem hallerde noktalar ya çakışık veya köşegen üzerinde karşılıklı noktalardır (Şekil 1.5.a,b,c).



Şekil: 1.5) Normlu değişkenlerin korelasyon katsayılarına bağlı olarak gösterilimi.

iii) $\hat{\xi}_{i\alpha}$, normlu değişkenlerin α 'yıncı ana eksen üzerindeki bileşenler vektörü olmak üzere

$$\hat{\xi}_{i\alpha} = Z' \cdot \underline{V}_{i\alpha} = Y' \cdot D \sqrt{p} \cdot \underline{V}_{i\alpha} = \sum_{j=1}^n \sqrt{p_j} y_i^j v_{j\alpha} \quad (1.109)$$

yazılabilir. Diğer taraftan

$$\xi_{i\alpha} = \frac{1}{\sqrt{p_1}} \sqrt{\lambda_{i\alpha}} v_{i\alpha}$$

olduğundan

$$\begin{aligned} d^2(G, Y^j) &= \sum_{i=1}^n p_i (y_i^j)^2 = \sum_{i=1}^n p_i \left(\frac{x_i^j - \bar{x}^j}{s^j} \right)^2 \\ &= \frac{1}{(s^j)^2} \sum_{i=1}^n p_i (x_i^j - \bar{x}^j)^2 \end{aligned}$$

yazılabilir. Diğer taraftan

$$\text{Var}(Y^j) = \sum_{i=1}^n p_i (x_i^j - \bar{x}^j)^2 = (s^j)^2$$

olduğundan sonuç elde edilir. Böylece, R^n uzayındaki p değişkenin her biri, yarıçapı 1 olan bir Hiperküre üzerinde bulunur.

11) Herhangi iki değişken arasındaki uzaklık

$$\begin{aligned} d^2(Y^j, Y^k) &= \sum_{i=1}^n p_i (y_i^j - y_i^k)^2 \\ &= \sum_{i=1}^n p_i (y_i^j)^2 - 2 \sum_{i=1}^n p_i y_i^j y_i^k + \sum_{i=1}^n p_i (y_i^k)^2 \end{aligned}$$

yazılabilir. Diğer taraftan

$$\sum_{i=1}^n p_i (y_i^j)^2 = \sum_{i=1}^n p_i (y_i^k)^2 = 1$$

$$\sum_{i=1}^n p_i y_i^j y_i^k = \text{Kor}(Y^j, Y^k)$$

olduğundan

$$d^2(Y^j, Y^k) = 2 [1 - \text{Kor}(Y^j, Y^k)] \quad (1.106)$$

yazılabilir.

$$\begin{aligned}\xi_{j\alpha} &= \sum_{i=1}^n (\sqrt{p_i} y_i^j) (\sqrt{p_i} \frac{\xi_{i\alpha}}{\sqrt{\lambda_\alpha}}) = \sum_{i=1}^n p_i y_i^j \frac{\xi_{i\alpha}}{\sqrt{\lambda_\alpha}} \\ &= D_p (Y^j, \frac{\xi_\alpha}{\sqrt{\lambda_\alpha}}) = \text{Kor} (Y^j, \xi_\alpha)\end{aligned}\quad (1.110)$$

bulunur. Buradan da açık olarak

$$\sum_{\alpha=1}^p \text{Kor}^2 (Y^j, \xi_\alpha) = 1 \quad (1.111)$$

olduğu görülür. Çünkü: $\text{Kor}(Y^j, \xi_\alpha)$ ifadesi değişkenler ile α 'yıncı ana eksen arasındaki açının kosinüsüdür. p tane eksen için bu ifadelerin kareleri toplamı 1 olur.

Sonuç olarak, R^n uzayındaki p değişkenin ana düzlem üzerindeki gösteriliminde, noktalar hep birim çember içerisinde kalır. Bu değişkenlerin ağırlık merkezi bireyler gibi orijinde değildir. Çünkü:

$$\sum_{j=1}^p p_j \xi_{j\alpha} \neq 0 \quad (p_j = 1, j=1, \dots, p)$$

dir. Ana düzlem ile hiperkürenin arakesit çemberi üzerinde veya çembere yakın olan noktalar, gösterilimi iyi olan noktadırlar.

1.4. ÇOK DEĞİŞKENLİ NORMAL VERİLER İÇİN ANA BİLEŞENLİ FAKTÖR ANALİZİ

Genellikle, ana bileşenli faktör analizinde, örneklerin seçildiği topluluklara ait dağılımlar gözönüne alınmaz. Ancak örneklerin normal dağılımlı topluluklardan seçilmesi halinde elde edilecek olan ana bileşenlerin yorumu daha kolay yapılabilir.

$\underline{x}^j \sim N_p(\underline{\mu}, V)$ olmak üzere, V 'nin özvektörlerinden oluşan matris U ise

$$\underline{\hat{\Psi}} = U' (\underline{x}^j - \underline{\mu}) \quad (1.112)$$

yazılabilir. Yani $\underline{\hat{\Psi}}$ nin herbir bileşeni, $\underline{x}^j$ lere ait bileşenlerin lineer kombinasyonu olarak ifade edilebilir. $\underline{x}^j$ ler normal dağılımlı olduklarından, bunların lineer kombinasyonları da normal dağılımı haiz olur. Yani $\underline{\hat{\Psi}} \sim N_p(0, D_p)$ yazılabilir (Burada $D_p = k\ddot{s}egen(\lambda_i)$ dir). V nin tekil olmadığı hallerde (1.112) dönüşümü birebirdir. Böylece $\underline{x}^j$ ve $\underline{\hat{\Psi}}$ nin olasılık dağılım fonksiyonları için,

$$G(\underline{\hat{\Psi}}) = F(\underline{x}^j)_{\underline{\hat{\Psi}}} |J| \quad (1.113)$$

bağıntısı yazılabilir. Burada J , jakobiyendir.

U ortogonal olduğundan (1.112) eşitliği

$$U \underline{\hat{\Psi}} = \underline{x}^j - \underline{\mu}$$

veya,

$$\underline{x}^j = U \underline{\hat{\Psi}} + \underline{\mu} \quad (1.114)$$

şeklinde yazılabilir.

Böylece

$$|J| = |U| \quad (1.115)$$

elde edilir.

$$f(\underline{x}^j) = \frac{1}{(2\pi)^{p/2} |V|^{1/2}} \exp \left[-\frac{1}{2} (\underline{x}^j - \underline{\mu})' V^{-1} (\underline{x}^j - \underline{\mu}) \right] \quad (1.116)$$

olduğundan ve

$$|U| = 1 \quad ; \quad |V| = |D_p| \quad ; \quad V^{-1} = U' D_p^{-1} U \quad (1.117)$$

yazılabildiğinden,

$$g(\underline{\hat{\Psi}}) = \frac{1}{(2\pi)^{p/2} |D|^{1/2}} \exp \left(-\frac{1}{2} \underline{\hat{\Psi}}' D^{-1} \underline{\hat{\Psi}} \right)$$

ya da

$$g(\underline{\hat{\Psi}}) = \frac{1}{(2\pi)^{p/2} \left(\prod_{i=1}^p \lambda_i \right)^{1/2}} \exp \left(-\frac{1}{2} \sum_{i=1}^p \frac{\hat{\Psi}_i^2}{\lambda_i} \right) \quad (1.118)$$

elde edilir. Böylece $\underline{\hat{\Psi}}$ nin olasılık yoğunluk fonksiyonu, p bağımsız normal değişkene ait yoğunluk fonksiyonlarının çarpımı şeklinde yazılabilir.

Çok değişkenli normal dağılımdan elde edilen ana bileşenlerin geometrik yorumu da şu şekilde yapılabilir; C bir sabit olmak üzere,

$$(\underline{x}^j - \underline{\mu})' V^{-1} (\underline{x}^j - \underline{\mu}) = C \quad (1.119)$$

p boyutlu uzayda bir elipsoid ailesi gösterir. (1.114) ve (1.119) eşitlikleri gözönüne alınırsa

$$\underline{\hat{\Psi}}' D_p^{-1} \underline{\hat{\Psi}} = \lambda_\alpha = C \quad (1.120)$$

bulunur. (1.120) bağıntısı, ana eksenlerinin yarı uzunlukları $\sqrt{\lambda_\alpha}$ ($\alpha=1, \dots, p$) olan elipsoid ailesini gösterir. C parametresini değiştirmekle elde edilecek elipsoid ailesi, ortak bir $\underline{\mu}$ merkez noktasını iletir. Her bir elipsoidin en büyük eksen, birinci ana eksen ile çakışır. Yani, yüzey üzerinde $\underline{x}^j$ ile $\underline{\mu}$ arasındaki uzaklığı maksimum yapan doğrultu birinci ana eksen oluşturacaktır. O halde

$$(\underline{x}^j - \underline{\mu})' (\underline{x}^j - \underline{\mu})$$

ifadesi (1.119) koşulu altında maksimum yapılmalıdır. Çünkü (1.119) koşulu, $\underline{x}^j$ nin elipsoid yüzeyi üzerinde olmasını

ifade eder. Böylece, problemin çözümünü verecek olan Lagrange'yenin ifadesi

$$\mathcal{L} = (\underline{x}^j - \underline{\mu})' (\underline{x}^j - \underline{\mu}) - \lambda [(\underline{x}^j - \underline{\mu})' V^{-1} (\underline{x}^j - \underline{\mu}) - C] \quad (1.121)$$

olur. Buradan türev alınıp sıfıra eşitlenerek

$$(\underline{x}^j - \underline{\mu}) - \lambda V^{-1} (\underline{x}^j - \underline{\mu}) = 0 \quad (1.122)$$

veya

$$[I - \lambda V^{-1}] (\underline{x}^j - \underline{\mu}) = 0 \quad (1.123)$$

elde edilir. Son denklem soldan V ile çarpılarak

$$[V - \lambda I] (\underline{x}^j - \underline{\mu}) = 0 \quad (1.124)$$

bulunur. Bu ise, ana eksenleri belirleyen koordinatların V 'nin öz vektörleri olacağını ifade eder. Öz değerlerin tümü farklı ise, V pozitif definit olduğundan,

$$\lambda_1 > \lambda_2 > \dots > \lambda_p > 0$$

Özdeğerlerine karşılık gelen özvektörler birbirine dik olacak ve böylece eksenlerin konumları tek türlü belirlenecektir (Teorem 1.6-1.7).

BÖLÜM 2

ÜÇ GİRİSLİ VERİ TABLOLARININ FAKTÖR ANALIZİ

2.1. ÜÇ GİRİSLİ VERİ TABLOLARI

Bir R kümesi verildiğinde, I, J ve T sonlu sayıda elemanlı üç küme olmak üzere, $I \times J \times T$ kartezyen çarpım kümesi üzerinde tanımlanan pozitif reel değerli bir fonksiyona üç girişli bir tablo adı verilir:

Her $(i, j, t) \in I \times J \times T$ için $(i, j, t) \rightarrow k(i, j, t) \in R$

O halde üç girişli bir tablo,

$$k(I, J, T) = \{k(i, j, t) \mid i \in I, j \in J, t \in T\} \quad (2.1)$$

şeklinde tanımlanır. Kısalık bakımından,

$$k(I, J, T) = k_{IJT} \quad ; \quad k(i, j, t) = k_{ijt}$$

alacağız. Üç girişli bir tablodan aşağıdaki şekilde iki girişli üç alt tablo elde edilebilir.

$$k_{IJ} = \{k_{ij} \mid i \in I, j \in J\} \quad ; \quad k_{ij} = \sum_t k_{ijt} \quad (t \in T)$$

$$k_{IT} = \{k_{it} \mid i \in I, t \in T\} \quad ; \quad k_{it} = \sum_j k_{ijt} \quad (j \in J) \quad (2.2)$$

$$k_{JT} = \{k_{jt} \mid j \in J, t \in T\} \quad ; \quad k_{jt} = \sum_i k_{ijt} \quad (i \in I)$$

Ayrıca üç girişli k_{IJT} tablosu aşağıdaki şekilde iki girişli bir tablo olarak da düşünülebilir:

$$k_{I(JT)} = \{k(i, (j, t)) \mid i \in I, j \in J, t \in T\} \quad (2.3)$$

Burada $k(i, (j, t)) = k(i, j, t)$ dir. Aynı gösterilim $k_{T(IJ)}$ ve $k_{J(IT)}$ tabloları için de düşünülebilir. Üç girişli veri tablolarının faktör analizinin bu gösterilim şekline göre yapılması literatürde (55) önerilmektedir.

2.2. ÜÇ GİRİŞLİ TABLOLAR İÇİN YENİDEN KURULUŞ FORMÜLÜ

Faktör analizi yöntemlerinin teorisi, genellikle iki girişli tablolara uygulanacak şekilde hazırlanmış ve geliştirilmiştir. 1936'lardan bu yana iki girişli tablolar için kullanılan yeniden kuruluş formülü üç girişli tablolar için başarılı olamamıştır. Son yıllara kadar bu konuda yapılan çalışmalardan da olumlu sonuçlar elde edilememiştir.

Üç girişli veri tabloları için yeniden kuruluş formülü ilk olarak karşılaştıрма analizi yönteminde, etkileşimlerin ayrışımı çerçevesinde ele alınarak, tansörel incelemesi yapılmış ve aynı formülün varlığı ve tekliği de ispatlanmıştır (72).

Üç girişli k_{IJT} tablosu yardımıyla f_{IJT} ile gösterilen frekanslar tablosu aşağıdaki şekilde elde edilebilir:

$$f_{IJT} = \{f_{ijt} | i \in I, j \in J, t \in T\} ; \quad f_{ijt} = \frac{k_{ijt}}{k} \quad (2.4)$$

Burada $k = \sum \{k_{ijt} | i \in I, j \in J, t \in T\}$ dir.

Esas tabloda olduğu gibi (2.4) formülü ile verilen frekanslar tablosundan marjinal frekans tabloları aşağıdaki gibi oluşturulabilir.

$$\begin{aligned} f_{IJ} &= \{f_{ij} | i \in I, j \in J\} ; \quad f_{ij} = \sum \{f_{ijt} | t \in T\} \\ f_{IT} &= \{f_{it} | i \in I, t \in T\} ; \quad f_{it} = \sum \{f_{ijt} | j \in J\} \\ f_{JT} &= \{f_{jt} | j \in J, t \in T\} ; \quad f_{jt} = \sum \{f_{ijt} | i \in I\} \end{aligned} \quad (2.5)$$

İki girişli olan bu tablolardan da, bunların profilleri olan tek girişli tablolar elde edilebilir.

$$\begin{aligned} f_I &= (f_i | i \in I) \quad ; \quad f_i = \sum (f_{ij} | j \in J) \\ f_J &= (f_j | j \in J) \quad ; \quad f_j = \sum (f_{ij} | i \in I) \\ f_T &= (f_t | t \in T) \quad ; \quad f_t = \sum (f_{it} | i \in I) \end{aligned} \quad (2.6)$$

Buna göre a_α , b_β , c_γ ve d_δ lar ileride belirlenecek katsayılar olmak üzere, üç girişli tablolar için, karşılaştırma analizinde önerilen yeniden kuruluş formülü,

$$\begin{aligned} f_{ijt} &= f_i f_j f_t [1 + \sum (\varphi_\alpha^i \varphi_\alpha^j a_\alpha | \alpha \in A) \\ &+ \sum (\varphi_\beta^i \varphi_\beta^t b_\beta | \beta \in B) + \sum (\varphi_\gamma^j \varphi_\gamma^t c_\gamma | \gamma \in C) \\ &+ \sum (\psi_\delta^i \psi_\delta^j \psi_\delta^t d_\delta | \delta \in D)] \end{aligned} \quad (2.7)$$

şeklindedir (72). Burada A, B, C, D ler indis kümeleri ve $\varphi_\alpha^i, \varphi_\beta^j, \varphi_\gamma^t$ ler ise, elemanlarına sırasıyla f_i, f_j, f_t kütle sistemi bağlanan I, J ve T üzerinde tanımlanmış, ortalamaları 0 ve varyansları 1'e eşit olan ortogonal yoğunluk fonksiyonlarını göstermektedir. O halde, örneğin φ_α^i ve $\varphi_{\alpha'}^i$ ($\alpha, \alpha' \in A$) için,

$$E(\varphi_\alpha^i) = \sum (f_i \varphi_\alpha^i | i \in I) = 0$$

$$\text{Var}(\varphi_\alpha^i) = \sum (f_i \varphi_\alpha^i \varphi_{\alpha'}^i | i \in I) = \begin{cases} 0, & \alpha \neq \alpha' \text{ ise} \\ 1, & \alpha = \alpha' \text{ ise} \end{cases}$$

yazılabilir. Daha açık bir ifadeyle, $\varphi_\alpha^i, \varphi_\beta^j$ ve φ_γ^t ler sırasıyla U_I^α, U_J^β ve U_T^γ birim eksen vektörlerinin yoğunluklarıdır.

(2.7) formülünün ikinci tarafındaki ilk üç $\sum$ toplam, ikinci basamaktan, sonuncu toplam ise, üçüncü basamaktan etkileşimleri kapsar.

Burada üçüncü basamaktan etkileşimleri ihmal ederek, belirli bir hata riskine karşılık, formülün, yoğunluk fonksiyonları yerine ölçüler cinsinden ifadesini daha basit ve kullanışlı bir şekilde vermeye çalışacağız.

Üçüncü basamaktan etkileşimlerin ihmal edilmesi halinde (2.7) ile verilen kuruluş formülü

$$f_{ijt} = f_i f_j f_t [1 + \sum (\varphi_{\alpha}^i \varphi_{\alpha}^j a_{\alpha} | \alpha \in A) + \sum (\varphi_{\beta}^i \varphi_{\beta}^t b_{\beta} | \beta \in B) + \sum (\varphi_{\gamma}^j \varphi_{\gamma}^t c_{\gamma} | \gamma \in C)] \quad (2.8)$$

şeklini alır.

Bu formülde, her iki taraf φ_{α}^i , ($\alpha' \in A$) ile çarpılıp i ve t üzerinde toplam alınırsa,

$$\sum (f_{ijt} \varphi_{\alpha}^i, | i \in I, t \in T) \cong (\sum (f_i f_j f_t [1 + \sum (\varphi_{\alpha}^i \varphi_{\alpha}^j a_{\alpha} | \alpha \in A) + \sum (\varphi_{\beta}^i \varphi_{\beta}^t b_{\beta} | \beta \in B) + \sum (\varphi_{\gamma}^j \varphi_{\gamma}^t c_{\gamma} | \gamma \in C)] | i \in I, t \in T) \varphi_{\alpha}^i, \quad (2.9)$$

elde edilir. Ayrıca, ortogonal olan φ lerin ortalama değerinin 0 olduğu gözönüne alınarak,

$$\sum (f_{ijt} \varphi_{\alpha}^i, | i \in I) = f_j \varphi_{\alpha}^j, a_{\alpha}, \quad (2.10)$$

bulunur. Bu son eşitliğin her iki tarafı f_j ile bölünerek

$$\sum \left(\frac{f_{ijt}}{f_j} \varphi_{\alpha}^i, | i \in I \right) = \varphi_{\alpha}^j, a_{\alpha}, \quad (2.11)$$

veya

$$\sum (f_{ij}^j \varphi_{\alpha}^i, | i \in I) = \varphi_{\alpha}^j, a_{\alpha}, \quad (2.12)$$

elde edilir. Bu ise karşılaştırma analizinde geçiş formülü olarak bilinir. O halde a_{α} katsayıları için

$$a_{\alpha} = \sqrt{\lambda_{\alpha}} \quad (\alpha \in A) \quad (2.13)$$

bulunur. Böylece, karşılaştırma analizinde verilen bu formülden, ana bileşenli faktör analizine geçerek bir yeniden kuruluş formülü vermek mümkündür. Bunun için (2.12) formülündeki Q_{α}^j , yoğunluk fonksiyonu yerine $U_{j\alpha}$ ölçüsü ve f_1^j yerine de k_{ij} almak yeterlidir. Çünkü, M metrik olmak üzere Q_{α}^j yoğunluk fonksiyonu

$$M(U_{j\alpha}) = Q_{\alpha}^j \quad (2.14)$$

olarak tanımlanmaktadır (63).

Böylece (2.8) formülündeki ilk toplam içinde bulunan a_{α} değerinin bulunması için yapılan işlemler aynen diğer iki toplamdaki b_{β} ve c_{γ} lar için tekrarlanarak

$$b_{\beta} = \sqrt{\lambda_{\beta}} \quad ; \quad c_{\gamma} = \sqrt{\lambda_{\gamma}} \quad (2.15)$$

elde edilir. Böylece üçüncü basamaktan etkileşimlerin ihmal edildiği kuruluş formülü, ana bileşenli faktör analizinde

$$k_{ijt} \approx k_i k_j k_t [1 + \sum (Q_{\alpha}^i Q_{\alpha}^j \sqrt{\lambda_{\alpha}} | \alpha \in A) + \sum (Q_{\beta}^i Q_{\beta}^t \sqrt{\lambda_{\beta}} | \beta \in B) + \sum (Q_{\gamma}^j Q_{\gamma}^t \sqrt{\lambda_{\gamma}} | \gamma \in C)]$$

olarak verilebilir. Eşitliğin ikinci yanındaki her bir toplamın karşılaştırma analizinde verilen kuruluş formülüne göre ifadeleri gözönüne alınarak

$$k_{ijt} \approx k_i k_j k_t \left[1 + \frac{k_{ij} - k_i k_j}{k_i k_j} + \frac{k_{it} - k_i k_t}{k_i k_t} + \frac{k_{jt} - k_j k_t}{k_j k_t} \right] \quad (2.17)$$

yazılabilir (72). Buradan açık olarak

$$k_{ijt} \approx k_{ik_jk_t} \left[\frac{k_{ij}}{k_{ik_j}} + \frac{k_{it}}{k_{ik_t}} + \frac{k_{jt}}{k_{jk_t}} - 2 \right] \quad (2.18)$$

elde edilir. Böyle bir formülün ana bileşenli faktör analizinde de kullanılabilmesi için, eşitliğin ikinci yanına belirli bir dönüşüm uygulanması gerekir. k_{ijt} tablosuna bilgisayar programı uygulanmak suretiyle elde edilen sonuçlardan, bu dönüşümün $1/k^2$ kuvvetinde bir homotesi olduğu görülmüştür.

Buna göre ana bileşenli faktör analizinde başlangıç tablosundaki k_{ijt} lerin yaklaşık değeri olarak,

$$k_{ijt} \approx \frac{k_{ik_jk_t}}{k^2} \left[\frac{k_{ij}}{k_{ik_j}} + \frac{k_{it}}{k_{ik_t}} + \frac{k_{jt}}{k_{jk_t}} - 2 \right] \quad (2.19)$$

alınacaktır. Bu formülün geçerliliğini göstermek için, formül yardımıyla yaklaşık olarak hesaplanan k_{ijt} lerden oluşan tabloya, ana bileşenli faktör analizi uygulanarak elde edilen sonuçlar, başlangıç tablosu için elde edilen sonuçlarla karşılaştırılmış ve bu yaklaşımın yeterliliği saptanmıştır (Şekil 6.9-6.10).

Benzer grafiksel sağlama, ECKART-YOUNG'ın önerdiği yeniden kuruluş formülü için de sağlanmıştır (Şekil 6.2-6.4).

(2.19) formülünden, tablodaki ölçümleri, ana bileşenlere bağlayan ifade, yani üç girişli tablolar için yeniden kuruluş formülü yazılabilir. Bunun için (1.84) göz önüne alınarak,

$$\begin{aligned} k_{ij} &= \sum_{\alpha} \lambda_{\alpha}^{-1/2} \bar{\Psi}_{\alpha} \bar{\Psi}_{\alpha}^* \\ k_{it} &= \sum_{\beta} \lambda_{\beta}^{-1/2} \bar{\Psi}_{\beta} \bar{\Psi}_{\beta}^* \\ k_{jt} &= \sum_{\gamma} \lambda_{\gamma}^{-1/2} \bar{\Psi}_{\gamma} \bar{\Psi}_{\gamma}^* \end{aligned} \quad (2.20)$$

ifadeleri (2.19) formülünde yerine konursa, herhangi bir k_{ijt} nin ana bileşenler cinsinden ifadesi,

$$k_{ijt} \approx \frac{k_t}{k^2} \sum_{\alpha} \lambda_{\alpha}^{-1/2} \frac{\bar{y}_{t\alpha}}{\bar{y}_{t\alpha}^*} + \frac{k_j}{k^2} \sum_{\beta} \lambda_{\beta}^{-1/2} \frac{\bar{y}_{j\beta}}{\bar{y}_{j\beta}^*} + \frac{k_i}{k^2} \sum_{\gamma} \lambda_{\gamma}^{-1/2} \frac{\bar{y}_{i\gamma}}{\bar{y}_{i\gamma}^*} - 2 \frac{k_i k_j k_t}{k^2} \quad (2.21)$$

elde edilir.

2.3. YENİDEN KURULUŞ FORMÜLÜNDEKİ ÖLÇÜMLERİN İNCELENMESİ

k_{ijt} ölçümünün $k_i k_j k_t$ ölçümler çarpımına göre yoğunluğunu d_{ijt} ile göstererek

$$d_{ijt} = \frac{k_{ijt}}{k_i k_j k_t} \quad (2.22)$$

şeklinde tanımlayacağız. Buna göre

$$d^{ij} = \frac{k_{ij}}{k_i k_j} ; d^{it} = \frac{k_{it}}{k_i k_t} ; d^{jt} = \frac{k_{jt}}{k_j k_t} \quad (2.23)$$

yazılabilir.

Böylece (2.19) formülü yoğunluklar cinsinden

$$d^{ijt} = \frac{1}{k^2} (d^{ij} + d^{it} + d^{jt} - 2) \quad (2.24)$$

olarak yazılabilir.

Teorem 2.1: $I \times J$ kümesinin her (i, j) elemanının ölçümü olarak $k_i k_j$ alınırsa, d^{ij} yoğunluğunun ortalama değeri k olur.

İspat: Kısaca,

$$\sum (k_i k_j d^{ij} | i \in I, j \in J) = \sum (k_{ij} | i \in I, j \in J) = k$$

olur.

Teorem 2.2: I kümesinin her i elemanı için d^{ij} , J üzerinde tanımlanmış k_j ölçümlü bir yoğunluk olarak gözönüne alınabilir. O takdirde d^{ij} nin ortalaması 1 dir.

İspat: Her i için

$$E(d^{iJ}) = \sum (k_j d^{ij} | j \in J) = \sum \left(\frac{k_{ij}}{k_i} | j \in J \right)$$

yazılırsa

$$E(d^{iJ}) = \frac{1}{k_i} \sum (k_{ij} | j \in J) = 1$$

bulunur.

BÜ LÜ M 3

U Y G U L A M A

3.1. ANALİZİ YAPILACAK VERİ TABLOLARI

Analizi yapılacak veri tabloları, 1976/80 yıllarını içeren beş yıllık süre içinde, ülkemizin 25 değişik ülkeye ihraç ettiği 12 farklı tür üründen sağladığı gelir miktarlarından oluşmaktadır. Grafik gösterilimlerde kolaylık bakımından, birey ve değişkenler için kullanılan notasyonlar aşağıdaki gibi belirlenmiştir.

Değişkenler:

1. Yaş sebzeler (YSEB)
2. Kuru sebzeler (KSEB)
3. Turunçgiller (LIMO)
4. İncir (INCI)
5. Üzüm (UZUM)
6. Kuru yemişler (KYEM)
7. Yaş meyvalar (YMEY)
8. Hububat (HUBU)
9. Tütün (TUTU)
10. Pamuk (PAMU)
11. Kimyasal ve inorganik bileşikler (KİMY)
12. Maden cevherleri (MADE)

Bireyler:

1. Fed. Alm. Cum. (ALMA)
2. Suudi Arab. (ARAB)
3. Avusturya (AVUS)
4. Belçika (BELC)
5. Danimarka (DANI)
6. İngiltere (INGL)

7. A.B.D. (AMER)
8. Fransa (FRAN)
9. Yunanistan (YUNA)
10. İran (IRAN)
11. Irak (IRAK)
12. İsrail (ISRA)
13. İtalya (ITAL)
14. Ürdün (URDU)
15. Japonya (JAPO)
16. Libya (LIBY)
17. Lübnan (LUBN)
18. Hollanda (HOLL)
19. Kuveyt (KUVE)
20. Romanya (RUMA)
21. Suriye (SURI)
22. İsveç (ISVE)
23. İsviçre (ISVI)
24. Rusya (RUSY)
25. Yugoslavya (YUGO)

Burada 12 üründen (yani değişkenlerden) oluşan küme I ve 25 ülkeden (yani bireylerden) oluşan küme de J ile gösterilirse, her bir yıl için $k_{IJ} = \{k_{ij} | i \in I, j \in J\}$ şeklinde iki girişli bir veri tablosu elde edilmiş olur. Her yıla ait k_{IJ} tablosundaki k_{ij} değeri, i yinci ülkeye j yinci ürünün ihraç edilmesiyle elde edilen gelir miktarını (10^3 dolar olarak) ifade eder. Böylece beş yıllık ihracat için beş adet iki girişli veri tablosu düzenlenmiş olur. Burada esas olarak farklı karakterde üç tablonun analizi yapılacaktır.

1) Ülkelere yapılan yıllık ihracattan elde edilen gelir miktarları gözönüne alınarak oluşturulan beş adet iki girişli tablo,

2) Beş yıllık toplam ihracat gelirlerinin hesaplanmasıyla elde edilen iki girişli bir tablo,

3) 1976'dan 1980'e kadar geçen beş yıllık süreyi üçüncü boyut olarak gözönüne almak suretiyle oluşturulan gerçek üç girişli bir tablo.

Yıllık ihracat gelirlerinin ayrı ayrı analiz edilmesiyle ortaya çıkan sonuçlar, yorumları itibariyle, beş yıllık toplam ihracat gelirlerinin analiz sonuçlarına çok yakın bulunmuştur. Bu nedenle, yıllık ihracat gelirlerinin analiz sonuçları üzerinde durmayacağız.

3.2. ANALİZ PROGRAMININ HAZIRLANMASI .

Bu çalışmada, ana bileşenli faktör analizi yöntemine ilişkin, oldukça genel amaçlı bir program hazırlanarak FORTRAN IV dilinde kodlanmıştır.

Tüm analizler, söz konusu program yardımıyla, İ.T.Ü. Bilgi İşlem merkezindeki IBM 4341 hesap makinalarında gerçekleştirilmiştir.

Burada program algoritması şu adımlardan oluşmaktadır.

- i) Veri tablosunun arzu edilen FORMAT'a göre okunması,
- ii) Korelasyon (veya Kovaryans) matrisinin hesaplanması,
- iii) Korelasyon (veya Kovaryans) matrisinin özdeğer ve özvektörlerinin hesaplanması (Housholder (74) algoritması kullanılarak)
- iv) Bulunan özdeğerlere ait, özdeğer yüzdeleri ile birikimli özdeğer yüzdelerinin hesaplanarak, özdeğerler histogramının çizimi,
- v) Özdeğer ve özvektörlerden hareketle ve ilgili formüller yardımıyla ana bileşenlerin hesaplanması,

vi) Analiz için gerekli olabilecek, katılma payları, gösterilimin kalitesi ve bireylerin ana bileşenlerle yaptıkları açıların kosinüs karelerinin hesaplanması,

vii) Değişkenlere ait maksimum ve minimum değerler ile ortalama ve standart sapmaların hesabı,

viii) Birey ve değişkenlerin ayrı ayrı ve birlikte gösterilimlerinin, birinci ve ikinci eksenlerden oluşan $\Delta(1,2)$ düzleminde (istenilirse ikinci ve üçüncü eksenlerden oluşan $\Delta(2,3)$ düzleminde, vb.) verilmesi.

3.3. UYGULANACAK ANALİZLERDE SEÇİM KRİTERİ

Analizde korelasyon matrisi kullanılabileceği gibi, kovaryans matrisi de kullanılabilmektedir. Ancak bu iki matristen elde edilecek sonuçlar farklı olduğu için, analizi yapılacak olan matris rasgele seçilmez (14). Matris seçiminde göz önüne alınan kriter, bu çalışmada, değişken birimleri arasındaki farklılıktır. Değişkenlerin birbirlerine kıyasla çok farklı değerler aldığı durumlarda, bunlara ilişkin bileşenler fazla bir anlam taşımayabilir. Böyle durumlarda değişkenler arasındaki farklılıkları yok eden, korelasyon matrisi kullanmak daha doğru olur. Değişken değerlerinin karşılaştırılabilir olduğu veri tabloları için kovaryans matrisi seçilmelidir (76).

Çalışmada, değişken birimlerinin aynı olmasına rağmen, miktarlar arasında büyük farklılıklar olması nedeniyle, korelasyon matrisi kullanılmıştır. Ancak, kovaryans matrisi kullanılarak yapılan analizlerden de benzer sonuçlar elde edilmiştir.

Önceden de belirtildiği gibi, hesaplanan ana bileşenlerin, çok değişkenli sisteme ilişkin toplam informasyonu açıklama payı, ilgili özdeğer ile doğru orantılıdır. Infor-

masyonun tamamının açıklanması, ancak tüm özdeğerlerin alınmasıyla gerçekleşebilir. Ancak bu durumda bileşen sayısı ikiden fazla olacağından, sonuçları grafik üzerinde yorumlamak mümkün olmaz. Bu nedenle, tüm analizlerde ilk iki bileşen alınmıştır.

3.4. GRAFİKSEL YORUM TEKNİĞİ

Birey ve değişken bulutlarının düzlemsel gösterilimlerini, yani grafiklerini yorumlayabilmek için, her bir bulutun elemanları arasındaki ortak özellikleri gözönünde bulundurmak gerekir. Bu ortak özellikleri her bir bulut için ayrı ayrı inceleyelim.

Daha önce, $\underline{y}^j$ ve $\underline{y}^k$ gibi iki değişken arasındaki uzaklığın, korelasyon katsayısı cinsinden ifadesi

$$\| \underline{y}^j - \underline{y}^k \|^2 = 2 (1 - r_{jk})$$

olarak verilmişti (1.108). Burada,

$$r_{jk} = \text{Kor} (\underline{y}^j, \underline{y}^k)$$

dır. O halde bu bağıntıya dayanarak şu sonuçları sıralayabiliriz:

i) $r_{jk} = 1$ veya 1'e yakın ise, $\| \underline{y}^j - \underline{y}^k \| = 0$ dir. Yani $\underline{y}^j$ ve $\underline{y}^k$ çakışır veya birbirlerine çok yakındırlar. Şu halde grafik üzerinde birim çembere yakın olan gruplardaki değişkenler arasında pozitif işaretli yüksek bir korelasyon olduğu söylenebilir.

ii) $r_{jk} = 0$ ise yani $\underline{y}^j$ ve $\underline{y}^k$ bağımsız ise, $\underline{y}^j$ ve $\underline{y}^k$ ortogonal olur. Ayrıca $\| \underline{y}^j - \underline{y}^k \| = \sqrt{2}$ elde edilir. Şu halde grafik üzerinde, birbirine dik durumlu değişken veya değişken gruplarının elemanları arasındaki korelasyonun sıfır olduğunu anlarız.

iii) $r_{jk} = -1$ olması halinde ise $\|Y^j - Y^k\| = 2$ olur. Yani noktalar 1 yarıçaplı hiper kürenin çapı üzerinde, karşıt uçlarda bulunur. Böylece, grafik üzerinde birim çemberin çapı üzerinde, karşıt uçlardaki değişken veya değişken grupları arasındaki korelasyon katsayısının -1 olduğunu söyleyebiliriz.

Değişkenler bulutu bilindiği gibi, bireyler bulutu gibi ortalanmamıştır. Yani, bireyler bulutunun grafik gösteriliminde, bireyler başlangıç noktası etrafında dağılmış oldukları halde, değişkenler için aynı durum söz konusu olmaz. Hatta, bütün değişkenlerin başlangıçtan geçen eksenlerden birinin başlangıca göre, yalnız bir tarafında olduğu durumlar da vardır (Şekil: 6.4).

Şimdi, değişkenler için yorumu büyük ölçüde kolaylaştıran yeni bir kavram tanımlayacağız.

Tanım: Y^j değişkeni ile Δ_α ana eksenini arasındaki $r_{j\alpha}$ korelasyonunu, bütün α değerleri için

$$r_{j\alpha} = \frac{\sum_{i=1}^n \xi_{i\alpha} Y_i^j}{\sigma_j \sqrt{\lambda_\alpha}} \quad (3.1)$$

olarak tanımlayacağız. Burada,

$\xi_{i\alpha}$: i yinci bireyin Δ_α eksenini üzerindeki izdüşümüdür (ana bileşen).

σ_j : Y^j değişkeninin standart sapmasıdır ve 1'e eşittir (Normlu analiz hali).

λ_α : $\xi_{i\alpha}$ nin varyansıdır. Aynı zamanda α 'yinci eksene karşılık gelen özdeğerdir.

Teorem 3.1: Y^j ($j=1,2,\dots,p$) değişkeni ile Δ_α ana eksenini arasındaki korelasyon, bu değişkenin eksen üzerindeki

$$r_{j\alpha} = \tilde{\xi}_{j\alpha} \quad (3.2)$$

bileşenine eşittir.

İspat: (1.76) ve (1.77) bağıntılarından açık olarak

$$\begin{aligned} \xi_{i\alpha} &= \sqrt{\lambda_{\alpha}} \quad v_{i\alpha} \\ \tilde{\xi}_{j\alpha} &= \sqrt{\lambda_{\alpha}} \quad u_{j\alpha} \end{aligned} \quad (3.3)$$

yazılabilir. Ayrıca (1.102)den dolayı da

$$\xi_{i\alpha} = \sum_{j=1}^p y_1^j u_{j\alpha} = \langle y_{\alpha}^j, u_{\alpha} \rangle_{I_p} = y_{\alpha}^j \cdot u_{\alpha} \quad (3.4)$$

yazılır. O halde (3.3) ve (3.4) birlikte gözönüne alınarak

$$\begin{aligned} \xi_{i\alpha} &= \lambda_{\alpha}^{-1/2} \sum_{j=1}^p y_1^j \tilde{\xi}_{j\alpha} \\ \tilde{\xi}_{j\alpha} &= \lambda_{\alpha}^{-1/2} \sum_{i=1}^n y_1^j \xi_{i\alpha} \end{aligned} \quad (3.5)$$

bulunur. Diğer taraftan $\sigma^j = 1$ olduğundan ve (3.1) gözönüne alınarak teorem ispat edilmiş olur.

Bu teoremden elde edilen sonuç, y^j nin Δ_{α} üzerindeki izdüşümünün, değişkenin bu eksen ile, yani bireylerin α 'yıncı bileşenleriyle olan korelasyonuna eşit olmasıdır ve burada üç farklı hal ayırdedilebilir.

i) $\tilde{\xi}_{j\alpha} = 1$ ise, $\xi_{i\alpha}$ ile y_1^j değişkenleri arasında

$$\xi_{i\alpha} = a y_1^j \quad (a \neq 0) \quad (3.6)$$

şeklinde lineer bir bağıntı vardır. Bu taktirde $\xi_{i\alpha}$ faktörünü,

bireylerin Δ_{α} eksenini boyunca bir sınıflandırılması olarak yorumlarız. Bu sınıflandırmada bireyler, y^j değişkeninin artan değerlerine göre sıralanır.

ii) $\hat{\epsilon}_{ju} = -1$ ise bireyler Δ_{α} eksenini üzerinde, y^j nin azalan değerlerine göre sıralanır.

iii) $\hat{\epsilon}_{ju} = 0$ olması halinde ise ϵ_{ju} lar ile y^j lar arasında bir ilişki olmayabilir. Yani ϵ_{ju} lar ile y^j ler incelenen topluluk üzerinde bağımsız olabilir.

Özet olarak denilebilir ki, $|\hat{\epsilon}_{ju}|$ değeri 1'e ne kadar yakınsa, y^j ler ile ϵ_{ju} lar arasındaki lineer ilişki o ölçüde büyük olur.

Teorem 3.2: y^j , ana düzlemlerden birinde birim çember üzerinde (veya yakınında) bulunan bir değişken ve bundan farklı herhangi bir y^k değişkeninin Oy^j üzerindeki izdüşümü L ise, bu iki değişken arasındaki korelasyon $\overline{OL}$ uzaklığına eşittir. Yani,

$$r_{jk} = \overline{OL} \quad (3.7)$$

dir.

İspat: Burada

$$(y^j)' y^k = \|y^j\| \|y^k\| \cos \theta = \overline{OL} \|y^k\| \quad (3.8)$$

yazılabilir. Diğer taraftan $\|y^k\| = 1$ (veya -1) olduğundan

$$\overline{OL} = (y^j)' y^k \quad (3.9)$$

elde edilir. Ayrıca

$$\text{Var}(y^k) = \text{Var}(y^j) = 1 \quad (3.10)$$

olmasından dolayı da,

$$r_{jk} = \text{Kov} (\underline{y}^j, \underline{y}^k) = (\underline{y}^j)' \underline{y}^k \quad (3.11)$$

yani

$$r_{jk} = \overline{OL}$$

bulunur. Böylece, grafik düzleminde $\underline{y}^j$ değişkeni birim çember üzerinde veya yakınında gözükmüşse, bunun diğer bir $\underline{y}^k$ değişkeni ile arasındaki r_{jk} korelasyonu grafik üzerinde kolayca hesaplanabilir.

B Ü L Ü M 4

ARAŞTIRMA SONUÇLARI

4.1. VERİ TABLOLARININ ANALİZİ

Faktör analizinde inceleme konusu edilen tablolar, çoğunlukla ya bir karşılaştırma tablosu veya da lojik bir tablodur. Bunların veri analizinden bildiğimiz bazı klasik koşulları sağlaması istenir. Her şeyden önce, verilerin yapıları ve içerikleri itibariyle homojen olması ve gerek satırlar (bireyler) arası ve gerekse sütunlar (değişkenler) arasındaki uzaklıkların bir anlamı olması gerekir. Ayrıca, incelenen olayın yapısını yansıtan veri tablosunun, bu yapının basit istatistiksel yöntemlerin uygulanması halinde bile çözülemeyecek ölçüde büyük olması amaca daha uygun olacaktır.

Çalışmada, değişkenler ile bireylerin ana bileşenli faktör analizi birlikte yapılarak, yorum için gerekli olabilecek tüm istatistiksel büyüklükler uygun tablolar halinde verilmiştir. Ayrıca, birey ve değişkenlerin, birinci ve ikinci bileşenden oluşan (1,2) düzleminde grafik gösterilimleri yapılmıştır.

Grafiklerin yorumu, bir eksenin her iki tarafında gruplanmış bulunan noktalar arasında bir ilgi ve benzerlik araştırılması şeklinde yapılır. Noktaların mutlak katılma payları dikkate alınmalı ve en uzaktaki noktaların her zaman katılma payları en büyük olan noktalar olmayacağı göz önünde bulundurulmalıdır. Ayrıca bir eksenin her iki tarafında, merkezden eşit uzaklıkta bulunan nokta veya nokta grupları için, istatistiksel anlamda "Birbirlerinin zıtlarıdır" şeklinde bir yorum yapılmalıdır.

Bireylere ilişkin sonuçlar tablosunda, bireylerin eksenlere yakınlık derecesini ifade eden KO2 kolonu bulunmaktadır. Bu kolon aynı zamanda, bireylerin bileşenlerle yaptıkları açıların kosinüsünün karesini vermektedir. Dolayısıyla her bir bireye ait KO2 değeri 1'e yaklaştıkça, o bireyin söz konusu eksene yakınlığı daha fazla olacaktır. Aynı tabloda, KLT ile gösterilen ve gösterilimin kalitesini veren bir başka kolon daha bulunmaktadır. Bu kolon, her bireyin bütün eksenlerle yapmış olduğu açıların kosinüs kareleri toplanarak oluşturulmuştur. Bu nedenle, KLT kolonundaki herhangi bir değer 1'e çok yakın olması, o bireyin gösteriliminin iyi olduğu anlamına gelir.

Analiz sonucunda, birey ve değişkenlerin ayrı ayrı gösterilimlerinden başka, birlikte gösterilimleri de yapıldı. Birlikte gösterilimdeki amaç, bireylerle değişkenler arasındaki çok yönlü ilişkileri ortaya çıkarmaktır.

Uygulama olarak seçilen ekonomik problemin analizinde, beş yıllık toplamardan oluşan iki girişli tablonun ortaya koyduğu sonuçlar ile, yeniden kuruluş formülünden elde edilen sonuçlar, grafikler üzerinde karşılaştırılmıştır.

Zaman faktörünü de gözönüne alarak, iki girişli alt tablolar yardımıyla oluşturulan (25 x 60) boyutlu üç girişli veri tablosunun analizinden başka bireylerle değişkenler yer değiştirerek oluşturulan (60 x 25) boyutlu üç girişli tablonun da analizi yapılmıştır.

Ayrıca, analizler sonucunda elde edilen grafikler ile, üç girişli veri tabloları için önerilen formüle ait grafikler karşılaştırılarak, söz konusu formülün geçerliliği gösterilmiştir.

4.2. İki Girişli k_{IJ} TABLOSUNUN ANALİZİ

Burada beş yıllık zaman süresi üzerinde toplam alınarak oluşturulan k_{IJ} tablosunun analizi ile birlikte istatistiksel anlamda yorumu yapılacaktır. Üç girişli k_{IJT} tablosundan elde edilen k_{IJ} tablosu gerçek anlamda iki girişli bir tablodur (Tablo: 6.1).

Önce k_{IJ} tablosundan elde edilen bireyler bulutuna ana bileşenli faktör analizi yöntemi uygulanarak, ekte verilen grafik elde edilmiştir (Şekil: 6.1). Birinci ana düzlem ile ilgili karakteristik değerler aşağıdadır:

$$\lambda_1 = 3.99$$

$$\lambda_2 = 1.91$$

$$\tau_1 = 0.33$$

$$\tau_2 = 0.16$$

Buradan toplam bilgi miktarının % 49 kadarlık bir kısmının 1. ve 2. bileşenden oluşan (1,2) düzlemi tarafından tüketilebileceği anlaşılmaktadır (Tablo: 6.3).

Grafik düzleminde ilk bakışta ikinci ana eksenin, Avrupa ülkeleri ile Ortadoğu ülkelerini ayırdığı göze çarpmaktadır. Sağ tarafta (ISVI, AMER, INGL, FRAN, ITAL, ALMA, RUSY, IRAK) yer alırken sol tarafta IRAK müstesna, tüm Ortadoğu ülkeleri ile birlikte bazı Avrupa ülkelerinin de yer aldığı görülmektedir. Bireylerin grafik düzlemi üzerindeki bu konumu, ihracatta, ürünlerin satıldığı ülkelerin coğrafik konumlarından ileri gelen ortak gereksinimleri yansıttığı ilk yorum olarak söylenebilir. Böylece, özet olarak denilebilir ki, (1,2) düzlemi Ortadoğu-Avrupa ikilemini kesin bir çizgi ile vurgulamaktadır.

Değişkenler bulutunun (1,2) düzlemindeki gösterilimine ait grafik Ek bölümde (Şekil: 6.2) ile verilmiştir. Grafik düzlemi üzerinde, değişkenler arasında birtakım gruplaşmalar göze çarpmaktadır. Bunları şu şekilde sıralayabiliriz:

- i) (KYEM, YMEY, HUBU)
- ii) (UZUM, PAMU, INCI)
- iii) (KIMY, MADE)
- iv) (LIMO, YSEB)

Bunların dışında TUTU ve KSEB başlıbaşına grup oluşturan değişkenler olarak düşünülebilir. Aynı gruptaki değişkenler arasındaki korelasyonların yüksek olduğu (Tablo: 6.2) ile verilen korelasyon tablosundan da sağlanabilir. Bu sonucun böyle olması gerektiği, grafiksel yorum tekniği bölümünde (3.4), birbirine yakın olan değişkenler arasındaki korelasyonun yüksek oluşu gerçeğinden anlaşılabılır.

Burada grup oluşturan, yani aralarındaki pozitif korelasyon katsayısı yüksek olan değişkenlerin, geçen süre içindeki ihracat artışlarında bir paralellik olduğu söylenebilir. Bunun gibi, aralarındaki korelasyon katsayıları yüksek fakat negatif olan ürünler için, "birinin ihracat artışına, diğerinin azalması karşılık gelir" şeklinde bir yorum yapılabilir.

Şimdi bireyler ile değişkenleri aralarındaki çok yönlü ilişkileri gözönüne alarak birlikte yorumlayalım. Değişkenlerin gösterdiği grafik düzleminde, birim çembere yakın olan değişkenlerden PAMU, UZUM ve INCI ürünleri göze çarpmaktadır. Bu ürün grubu, bireylerin birinci ana eksen üzerinde bir sıralanışını sağlamaktadır. Yani, birinci ana eksen üzerinde, pozitif yönde, başlangıçtan en uzakta olan bireyler, bu ürünlerin en çok ihraç edildiği ülkeler olarak görülecektir. Bireylerin gösterildiği grafik düzlemi üzerinde bu özelliği sağlayan ülkeler olarak ALMA, ITAL, FRAN, INGL ve ISVI görülmektedir. Bu sonuç (Tablo: 6.1) de verilen veri tablosundan da sağlanmaktadır. Gerçekten de veri tablosundaki PAMU, UZUM, INCI kolonlarında en yüksek ihracatın yapıldığı ülkeler, yukarıda söz konusu edilenlerdir.

Burada 25 birey ve 60 değişkenden oluşan bir tablonun analizi söz konusudur. Ancak, analizi yapılacak tablolardaki kolon sayısının birey sayısından daha az olması uygun olacağından, (25 x 60) tablosunun transpozese alındıktan sonra, ana bileşenli faktör analizi uygulanmıştır. Böylece, analizde ülkeler, değişken ve ürünler de birey olarak göz önüne alınmıştır.

Önce, değişkenler bulutunun analizi yapılarak Ek bölümde verilen grafik elde edilmiştir (Şekil: 6.5). Burada birinci ana düzlem ile ilgili karakteristik değerler aşağıdaki gibidir:

$$\lambda_1 = 5.95$$

$$\lambda_2 = 3.59$$

$$\tau_1 = 0.24$$

$$\tau_2 = 0.14$$

Görüldüğü gibi birinci ve ikinci bileşenden oluşan ana düzlem tek başına, toplam informasyonun % 38 kadarlık bir kısmını tüketmektedir.

Grafik düzleminde ilk bakışta, Ortadoğu ülkelerinin net bir biçimde diğer Avrupa ülkelerinden ayrıldıkları göze çarpmaktadır. Bu durum, K_{IJ} tablosunun analizinde de ortaya çıkmakla birlikte, IRAK'ın Avrupa ülkeleri ile birlikte görülmesi Avrupa-Ortadoğu ayırımında istisna oluşturmaktaydı. Bu da, zaman boyutunu da dikkate alarak, problemi üç girişli tablo yardımıyla analiz etmenin daha gerçekçi olacağını vurgulamaktadır. Grafik düzleminde oluşan gruplar şu şekilde sıralanabilir:

- i) (AMER, DANI, ISVE, YUGO, AVUS, RUSY, JAPO)
- ii) (ARAB, LIBY, ISRA, IRAN, SURİ, LUBN), (URDU, KUVE)
- iii) (YUNA, RUMA, ITAL, BELC)
- iv) (HOLL, ALMA, FRAN)

Benzeri durum (Şekil: 6.6)da verilen, ikinci ve üçüncü bileşenden oluşan (2,3) düzleminde de belirgin bir şekilde göze çarpmaktadır.

Ortadoğu ülkeleri ile Avrupa ülkelerinin grafik düzlemlerinde birbirlerinden ayrılma nedenlerinin, coğrafik konumlarından kaynaklandığı burada da söylenebilir.

Aynı gruptaki ülkeler arasındaki korelasyonların yüksek oluşu, (Tablo: 6.6) dan da gerçekleştirilebilir.

Gerçekten, bu tabloda pozitif işaretli büyük korelasyonların, söz konusu gruplara giren ülkeler arasında olduğu görülmektedir.

Diğer taraftan, bireyler bulutunun grafik gösterilimi de (Şekil: 6.7)de verilmiştir. Bu grafik incelendiğinde düzlem üzerinde değişik yıllara ait aynı ürün türlerinin birbirlerine çok yakın bulunduğunu görmekteyiz. Bunun yorumu, söz konusu ürünler farklı yıllarda, değişik miktarlarda ihraç edilmekle birlikte, bu ürünlerin ihracatında çok büyük sapmalar meydana gelmediği şeklinde yapılabilir. Ancak KYM, TUTU ve PAMU gibi ürünlerde bu gruplaşmanın olmadığı grafik üzerinde görüldüğü gibi, veri tablosundan da sağlanabilir.

İncelemede, bireyler arasında, birinci ana eksen üzerindeki gösteritlimleri en iyi olan, yani kaliteli yüksek olanlar (Tablo: 6.9)a bakılarak (KY76, KI76, TU77, KI78, PA78, YS79, KI79, LI80) şeklinde belirlenmiştir. Bunlar arasında birinci ana eksene en yakın olan (YS79, LI80, KI79) bireyleri, KO2 kolonu incelenerek ayırt edilebilir.

Son olarak 25 birey ve 60 değişkenden oluşan (25x60) veri tablosunun analizi yapılmıştır. Özdeğerler tablosundan, birinci özdeğerin, diğerlerinden açık bir biçimde ayrıldığı görülmektedir (Tablo: 6.11). Aynı tablodan ilk iki bileşenden

oluşan (1,2) düzleminin tek başına toplam informasyonun % 44 kadarlık bir kısmını açıkladığı da anlaşılır.

Bireyler ile ilgili grafik incelendiğinde (60 x 25) veri tablosunun analizi sonucundaki kadar net bir biçimde, Ortadoğu-Avrupa ülkeleri ayırımı görülmemektedir. Bunun nedeni, daha önce belirtildiği gibi, değişken sayısının birey sayısından az olmasındandır. Bireylere ait grafik gösterilimde, gösterilimi iyi yapılan ülkelerin sırasıyla ALMA, AMER, ITAL ve YUNA olduğu (Tablo: 6.13)'den görülebilir. Aynı ülkelerden birinci eksen ile açısı küçük olan sadece YUNA'yı KO2 sütununa bakmak suretiyle görebiliriz.

Değişkenlerle ilgili grafikte, tüm değişkenlerin ikinci bileşenin pozitif tarafında yer aldığını görüyoruz. Ayrıca, değişik yıllara ait aynı ürün türlerinin bir arada gruplaşması, bu ürünlerin ihracatında bir paralellik olduğu anlamına gelir. Grup oluşturan değişkenlerin aralarında yüksek korelasyonlar bulunduğunu, (Tablo: 6.10)'dan da görmek mümkündür. Aynı grafik düzleminde, yarıçapı 1 olan çembere yakın olarak HU76, HU79 değişkenleri görülmektedir. Bu değişkenler, birinci bileşen üzerinde bireylerin bir sınıflandırılışını vermektedir.

4.4. ÜÇ GİRİŞLİ k_{IJT} TABLOSUNUN YENİDEN KURULUŞ FORMÜLÜNÜN

GRAFİKSEL SAĞLANMASI

Üç girişli veri tabloları için 3. bölümde önerilen yeniden kuruluş formülü, (60x25) ve (25x60) boyutlu veri tablolarına uygulanarak elde edilen grafikler yardımıyla yapmak, formülün geçerliliğini ortaya koymak için yeterli olacaktır.

(60 x 25) boyutlu üç girişli veri tablosunun yeniden kuruluş formülüne ait değişkenler bulutunun grafik gösteri-

limi (Şekil: 6.8) ile verilmiştir. Bu grafik (Şekil: 6.7) ile karşılaştırıldığında, değişkenler arasındaki gruplaşmalarda büyük benzerlikler bulunduğu göze çarpmaktadır. Aynı durum, bireyler bulutuna ait grafikte de, Ortadoğu ülkeleri ile Avrupa ülkelerinin ayrılması şeklinde ortaya çıkmıştır.

Diğer taraftan, (25 x 60) boyutlu üç girişli veri tablosunun yeniden kuruluş formülüne ait değişkenler bulutunun grafik gösterilimi de (Şekil: 6.9)da verilmiştir. Bu grafik, esas tablonun analizi sonucunda elde edilen, değişkenlerle ilgili grafikte karşılaştırıldığında, yine gruplaşmalarda büyük benzerlikler bulunduğu, hatta bazı noktaların üst üste çakıştığı görülmektedir. Bu sonuç önerilen formülün, ana bileşenli faktör analizi için kullanılabilir olduğu anlamına gelmektedir.

BÖLÜM 5

SONUÇLAR

1. Ana bileşenli faktör analizinde kullanılmak üzere, üç girişli tablolar için başlangıç tablosundaki değerleri, marjinal değerler yani ölçümler cinsinden ifade eden bir yeniden kuruluş formülü verilmiştir.

2. Yeniden kuruluş formülünde yer alan ölçümlerin özellikleri incelenmiş ve iki teorem ispatlanmıştır (Teorem: 2.1, Teorem: 2.2).

3. Yeniden kuruluş formülünden hareketle, başlangıç tablosundaki değerler, ana bileşenler cinsinden ifade edilmiştir.

4. Grafikselsel yorum tekniği çerçevesinde iki yeni kriter belirlenmiştir (Teorem: 3.1, Teorem: 3.2).

5. Türkiye'nin beş yıl gibi kısa bir süre içinde ele alınan ihracatının analizlerinden, bu çok kısa süreye karşın önemli sonuçlar elde edilmiştir. Gruplaşmalar, ihracatın gerçek yapısını yansıtacak biçimdedir. Beş yıllık toplam gelir miktarlarından oluşan tablo ile, yıllık gelirlerden oluşan tabloların ana bileşenli faktör analizlerinden elde edilen sonuçlar arasındaki benzerlikler, ihracat yapısının süre konusu süre içinde değişmediği şeklinde yorumlanmıştır. Avrupa ülkeleri ile, Ortadoğu ülkeleri bu ihracatta zıt iki grup oluşturmakta ve doğu-batı ikilemini belirgin bir biçimde yansıtmaktadır. Ayrıca üç girişli veri tablosunun analizinden, beş yıllık süre içinde yapılan ihracatta, satılan ürünlerde, diğerlerine göre büyük ölçüde değişme gösteren ürün olmadığı ve değişimin bütün ürünler için aynı oranda olduğu saptanmıştır.

6. Üç girişli k_{IJT} tablosunun analizlerinden elde edilen sonuçlar, iki girişli k_{IJ} tablosundan elde edilen sonuçları kapsamakla birlikte, üç girişli tablonun analizlerinde gerek bireyler (yani ülkeler) ve gerekse değişkenler (yani ürünler) arasındaki gruplaşmalar, istatistiksel ifadeyle çekmeler ve itmeler daha net bir biçimde gözlenebilmiştir. Böylece, olayın zaman akışı içindeki gelişimini izlemek ve yorumunu bu zaman akışındaki değişimleri gözönüne alarak yapmak için, üç girişli tabloların ve buna bağlı olarak yeniden kuruluş formülünün kullanılmasıyla elde edilen sonuçlar karşılaştırılmıştır.

7. Veri tablosundan elde edilen bireyler (yani ülkeler) bulutunun analizinde, ihraç edilen ürünlere göre, birbirleriyle ilişkili olan ülkelerin şu gruplardan oluştuğu gözlenmiştir:

- i) (AMER, DANI, ISVE, YUGO, AVUS, RUSY, JAPO)
- ii) (HOLL, ALMA, FRAN)
- iii) (YUNA, RUMA, ITAL, BELC)
- iv) (URDU, KUVE, ARAB, LIBY, ISRA, IRAK, IRAN, SURI, LUBN)

8. Veri tablosundan elde edilen değişkenler (yani ürünler) bulutunun analizinde, değişkenlerin ihracat yapılan ülkelere göre gruplaşmaları şu şekilde belirlenmiştir:

- i) (LIMO, YSEB)
- ii) (TUTU, KSEB)
- iii) (KYEM, HUBU, YMEY)
- iv) (KIMY, MADE)
- v) (INCI, UZUM, PAMU)

9. Bireyler ve değişkenlerin grafiksel gösterilimlerinde, grafik düzleminin aynı bölgelerine rastlayan değişken

ve birey grupları arasındaki ilişki, grup oluşturan ülkelere ihraç edilen ürünler olarak yorumlanmıştır. Buna göre, (TUTU, KSEB) ürünlerinin (AMER, DANI, ISVE, YUGO, AVUS, RUSY, JAPO) ülkelere, (UZUM, INCI, PAMU) ürünleriyle birlikte (KIMY, MADE)lerin sırasıyla (ALMA, HOLL, FRAN, YUNA, BELC, ITAL, RUMA) ülkelere ve (LIMO, YSEB) ürünlerinin ise Ortadoğu ülkelere daha fazla ihraç edildiği grafik düzleminden anlaşılmış ve bu sonuç veri tablosundan da sağlanmıştır.

10. Uç girişli veri tablolarının yeniden kuruluş formülü yardımıyla oluşturulan tablo ile, başlangıç tablosunun grafiksel gösterilimlerdeki noktalar, birbirlerine çok yakın bulunmuştur. Bu da ana bileşenli faktör analizinde kullanılmak üzere önerilen yeniden kuruluş formülünün elde edilmesinde, göz önüne alınan hata riskinin küçük olduğunu ve dolayısıyla güvenle kullanılabileceğini göstermektedir.

E K B Ü L Ü M

ANALİZLER SONUCUNDA ELDE EDİLEN

TABLO VE GRAFİKLER

Tablo 6.10: İAİ GİRİŞLİ TABLOLAR İÇİN ÖNERİLEN YERİDEN KURULUS
FORMÜLÜMÜNE GÖRE ÇIKARILMIŞ SONUÇLAR.

..... VERİ TABLOSU												
YSLG	KİB	LİD	İNGİ	UZUM	YMEY	KYEM	HÜDÜ	TOTU	PANU	KİMY	MADE	FJPL
ALBA	236.	30.	198.	201.	795.	7972.	178.	1243.	1300.	439.	336.	123.
ARA	53.	53.	13.	1.	11.	60.	17.	26.	10.	36.	19.	0.
AVUS	11.	5.	323.	56.	29.	501.	0.	37.	14.	313.	57.	1619.
BECC	3.	193.	7.	32.	214.	90.	16.	56.	172.	1303.	10.	75.
CBNI	1.	24.	1.	36.	2.	226.	10.	19.	19.	41.	9.	370.
ENOL	5.	22.	25.	71.	459.	430.	73.	35.	44.	2240.	279.	23.
AMEA	9.	32.	0.	29.	23.	618.	121.	11.	4442.	58.	600.	302.
FAMA	29.	219.	9.	380.	233.	1143.	159.	126.	167.	570.	637.	163.
YUNA	0.	17.	1.	0.	1.	25.	0.	1.	1.	55.	6.	1.
IRAG	12.	33.	175.	11.	28.	238.	139.	442.	80.	117.	32.	2.
IRAK	17.	205.	43.	6.	70.	77.	187.	776.	238.	307.	167.	5.
ISRA	1.	127.	0.	31.	5.	53.	17.	6.	14.	3.	0.	1.
ITAL	5.	433.	1.	153.	666.	151.	53.	617.	184.	4422.	736.	298.
URDU	419.	173.	104.	1.	35.	65.	127.	123.	18.	42.	1.	3.
JAPJ	1.	1.	1.	0.	63.	137.	4.	27.	706.	357.	361.	100.
LİBY	3.	41.	6.	23.	70.	522.	37.	713.	162.	337.	161.	2.
LOBY	12.	295.	7.	3.	18.	1.	3.	62.	26.	74.	16.	0.
HOLL	33.	58.	6.	25.	614.	377.	39.	41.	272.	949.	259.	7.
KUVE	53.	36.	37.	13.	30.	223.	64.	61.	87.	137.	13.	2.
KOMA	198.	51.	370.	7.	33.	261.	9.	104.	69.	1217.	142.	83.
SURI	31.	64.	43.	8.	91.	40.	8.	266.	48.	443.	631.	25.
LSVL	2.	29.	3.	101.	97.	73.	18.	51.	270.	777.	51.	121.
LSVI	19.	79.	7.	91.	136.	464.	3.	167.	77.	2965.	17.	195.
AUSY	1.	1.	551.	126.	307.	2440.	53.	68.	442.	274.	32.	9.
YUGO	0.	17.	211.	8.	12.	31.	3.	19.	326.	272.	35.	47.
FJPL	1170.	2235.	2142.	1419.	4502.	16446.	1338.	5997.	9357.	21430.	5163.	1640.

TABLO.6.2: {25,12} BOYUTLU K(I,J) TABLOSUNA AIT KORELASYONLAR
TABLOSU

YSEB	KSEB	LIMO	INCI	UZUM	YMEY	KYEM	HUBU	TUTU	PAMU	KIMY	MADE
1.00	1.00										
0.05	-0.30	1.00									
0.23	0.30	0.07									
0.01	0.28	0.04	1.00								
0.10	0.14	0.34	0.45								
0.35	-0.19	0.01	0.48	1.00							
0.35	0.25	0.05	0.38	0.27	1.00						
0.24	-0.17	-0.03	0.23	0.36	0.44	1.00					
-0.02	0.39	-0.03	0.08	0.03	0.60	0.61	1.00				
-0.12	0.33	-0.03	0.43	0.67	0.25	0.30	0.03	1.00			
-0.12	0.33	0.08	0.53	0.42	0.50	0.17	0.47	0.30	1.00		
-0.07	0.33	-0.14	0.47	0.25	0.23	0.26	0.25	0.30	0.49	1.00	
YSEB	KSEB	LIMO	INCI	UZUM	YMEY	KYEM	HUBU	TUTU	PAMU	KIMY	MADE
ADJ											

TABLO.6.3: {25,12} BOYUTLU K(I,J) TABLOSUNA AIT ÖZDEĞERLER VE

ÖZDEĞER YÜZDELERİ.

NO	ÖZDEĞERLER	YÜZDELER	BİRİKİMLİ	MATRİSİN ÖZDEĞERLER HİSTOGRAMI
1	3.99749	33.31	33.31	*****
2	1.91462	15.96	49.27	*****
3	1.49240	12.44	61.71	*****
4	1.17050	9.75	71.46	*****
5	0.85405	7.12	78.58	*****
6	0.77923	6.49	85.07	*****
7	0.58435	4.87	89.94	*****
8	0.50257	4.19	94.13	*****
9	0.32056	2.67	96.80	*****
10	0.20267	1.69	98.49	*****
11	0.13678	1.14	99.63	*****
12	0.04462	0.37	100.00	*****

TABLO-6.4: (25,12) BOYUTLU K(1, J) TABLOSU'NUN ANALİZİ SONUCUNDA
BULUNAN DEĞİŞKENLER'E AİT ANA BİLEŞENLER.

DEĞİŞKENLER	ORTALAMA	S. SAPMA	MINIMUM	MAKSİMUM	EKSEN-1	EKSEN-2	EKSEN-3	EKSEN-4
YSEB	46.1	94.8	0.0	419.0	234	617	113	286
KSEB	89.2	109.9	0.0	482.0	412	-446	581	243
LINCI	84.6	140.8	0.0	551.0	477	-571	-276	227
UZUM	54.4	135.8	0.0	386.0	717	-153	-20	-32
YMLY	167.9	280.8	0.0	940.0	716	4	244	-22
KYEM	55.2	160.5	0.0	608.0	707	523	-179	-155
HUTU	192.4	318.1	0.0	1922.0	602	316	174	558
TUTU	343.0	891.7	0.0	1226.0	662	358	-217	301
PAMY	804.0	1302.2	1.0	4501.0	315	-191	302	221
KMADE	205.1	222.2	0.0	4164.0	748	-367	-210	-166
	65.0	89.6	0.0	312.0	603	-553	-344	46

TABLO.6.5: (25+12) BOYUTLU K(I,J) TABLOSU'NUN ANALIZI SONUCUNDA
BULUNAN BIREYLER'E AIT ANA BİLESENLER

BIREYLER	KLT	EKS-1	KO2	EKS-2	KO2	EKS-3	KO2	EKS-4	KO2
ALMA	936	6400	715	3539	212	-440	3	-26	0
AKAB	901	1782	849	215	116	332	30	-194	10
AVUS	720	-1782	102	318	116	-1009	170	-1917	436
BELLC	415	-1835	132	-644	147	-1571	115	-1249	218
DANI	755	-1275	147	-203	0	220	133	-1097	130
INCL	419	1275	149	-403	15	1180	661	-1997	121
AMLR	491	2033	38	-1555	125	-4654	0	-1997	103
FRAN	854	-2033	74	-1555	0	140	4	-1317	125
YURAK	446	-5423	77	1162	250	-127	0	-2250	47
IRAKA	610	1603	278	-3425	360	924	80	-191	197
ISRALU	805	-4411	5	-3341	277	607	94	-2798	227
URPO	501	-9527	285	-1002	252	1014	107	-2143	197
JAEY	340	-3324	253	-1680	252	-9023	203	-1439	180
LUBEN	623	-1112	253	-7603	252	1373	254	-9556	323
HOLLE	175	-1522	70	-2538	252	1437	0	-5556	101
KUMAL	387	-1628	46	-1405	175	-1152	2	-9271	129
SURIE	103	-5972	187	-8969	239	-4708	150	-1226	39
ISVEI	380	-3511	177	-1808	121	-1402	330	-2514	417
ISUSV	203	-150	612	-1179	121	-1402	131	-2514	417
YUGO	891	-150	612	-1179	121	-1402	131	-2514	417

TABLO.6.7: (60,25) BOYUTLU K(I,J,T) TABLOSU'NUN KORELASYON MATRİSİ-NE AİT ÖZDEĞERLER VE ÖZDEĞER YÜZDELERİ.

NO	ÖZDEĞERLER	YÜZDELER	BİRİKLİLİ	MATRİSİN ÖZDEĞERLER HİSTOGRAMI
1	5.95943	23.84	23.84	*****
2	3.59084	14.36	38.20	*****
3	2.90603	11.62	49.83	*****
4	1.38596	7.54	59.37	*****
5	1.40914	7.59	66.96	*****
6	1.02351	4.25	72.21	*****
7	0.82720	3.29	77.10	*****
8	0.73035	2.95	80.05	*****
9	0.65438	2.52	84.00	*****
10	0.55430	2.08	87.00	*****
11	0.45129	1.61	89.13	*****
12	0.35432	1.21	91.30	*****
13	0.25432	0.81	93.50	*****
14	0.16921	0.48	95.66	*****
15	0.08976	0.36	97.79	*****
16	0.05075	0.26	98.08	*****
17	0.03125	0.20	99.40	*****
18	0.01723	0.07	99.70	*****
19	0.00523	0.02	99.97	*****
20	0.00223	0.01	99.99	*****
21	0.00023	0.00	100.00	*****

ABLO.6.8: (60,25) BOYUTLU KILI(JIT) TABLOSUNUN ANALİZİ SONUCUNDA
BULUNAN DEĞİSKENLERE AIT ANA BİLEŞENLER.

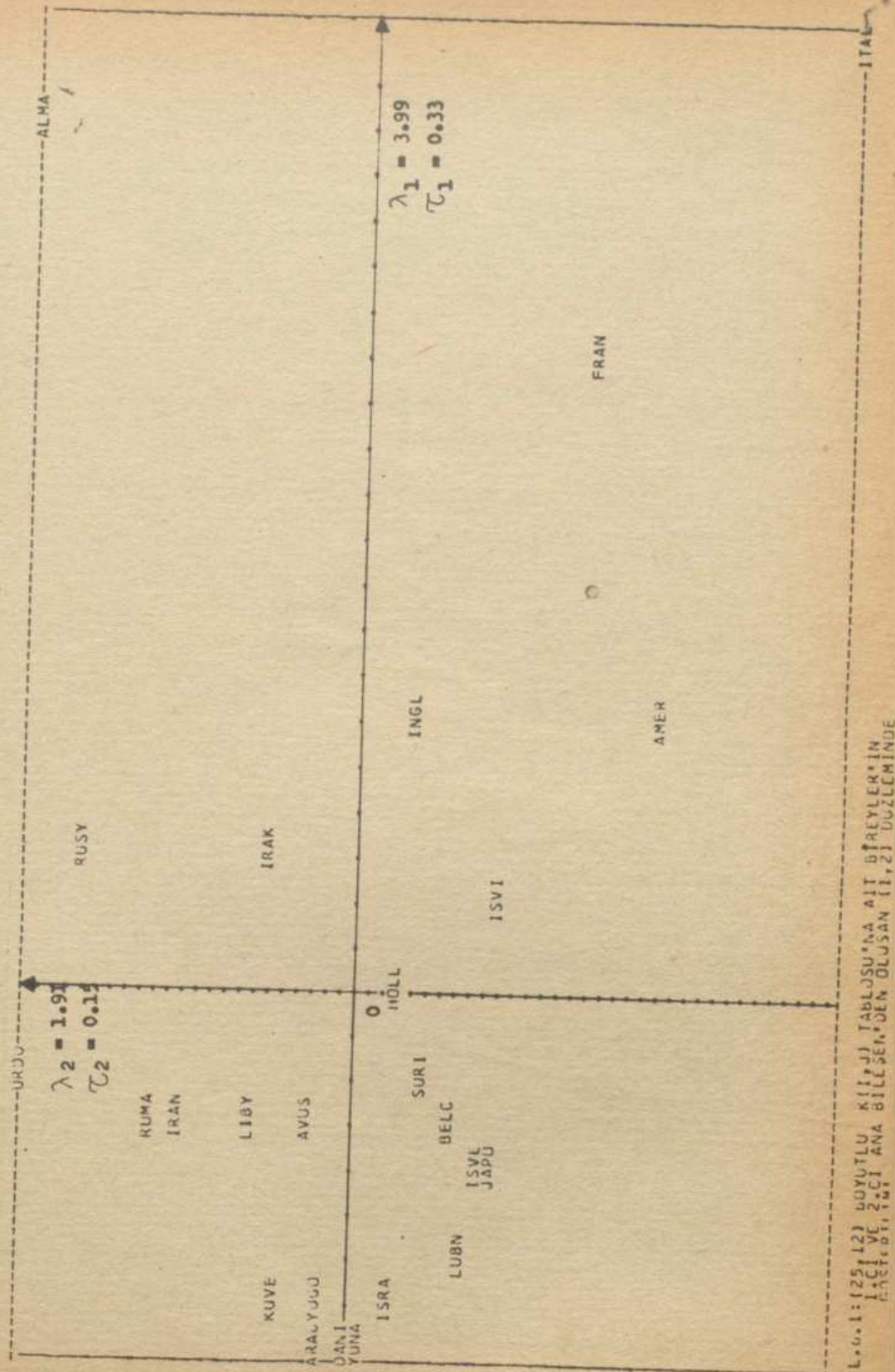
DEĞİSKENLER	JORTALAMA	S. SAPMA	MINIMUM	MAKSİMUM	EKSEN-1	EKSEN-2	EKSEN-3	EKSEN-4
ALMA	21888.9	40386.0	0.0	208945.0	672	-233	-506	136
ARAB	368.3	499.9	0.0	2692.0	54	-254	-353	-357
AVUS	2155.1	3540.4	0.0	17309.0	223	-340	-353	-179
BELC	399.7	11055.0	0.0	56652.0	896	-109	-364	-112
DANGL	269.3	521.6	0.0	30774.0	97	-177	-233	-164
AMERN	345.5	13494.3	0.0	72302.0	902	-113	-233	-97
FRAN	8308.5	24513.5	0.0	139381.0	202	-183	-138	-69
FUNAN	475.5	5903.5	0.0	24138.0	703	-261	-467	-41
YURAN	114.8	331.5	0.0	1707.0	603	-121	-155	-184
IRAKA	257.4	3026.2	0.0	16191.0	158	-280	-192	-160
ITALA	338.5	7739.3	0.0	25243.0	158	-380	-276	-640
ITADU	10338.5	20268.1	0.0	94127.0	903	-254	-276	-501
JAPPO	1915.5	32095.4	0.0	16801.0	-219	-450	-197	-245
LUBN	1233.5	4534.0	0.0	21922.0	213	-472	-308	-769
HOLLE	2834.8	2127.4	0.0	10471.0	90	-472	-170	-387
KUUMA	2421.1	7734.1	0.0	18822.0	259	-438	-170	-128
RUMI	2976.2	6018.1	0.0	31205.0	199	-438	-315	-480
SURI	1382.3	2727.9	0.0	14761.0	120	-290	-215	-152
ISVEI	6772.3	10894.1	0.0	47701.0	891	-191	-465	-480
ISUY	6058.9	20194.1	0.0	124081.0	305	-491	-235	-152
YUGO	1179.9	11655.1	0.0	16685.0	104	-382	-780	-530

TAHLİ.9.9: (60,25) BÖYÜTLÜ Kİİ.J.Tİ TABLOSU'NUN ANALİZİ SONUCUNDA
DULUNAN DİRLİLER'E AIT ANA DİLEŞENLER.

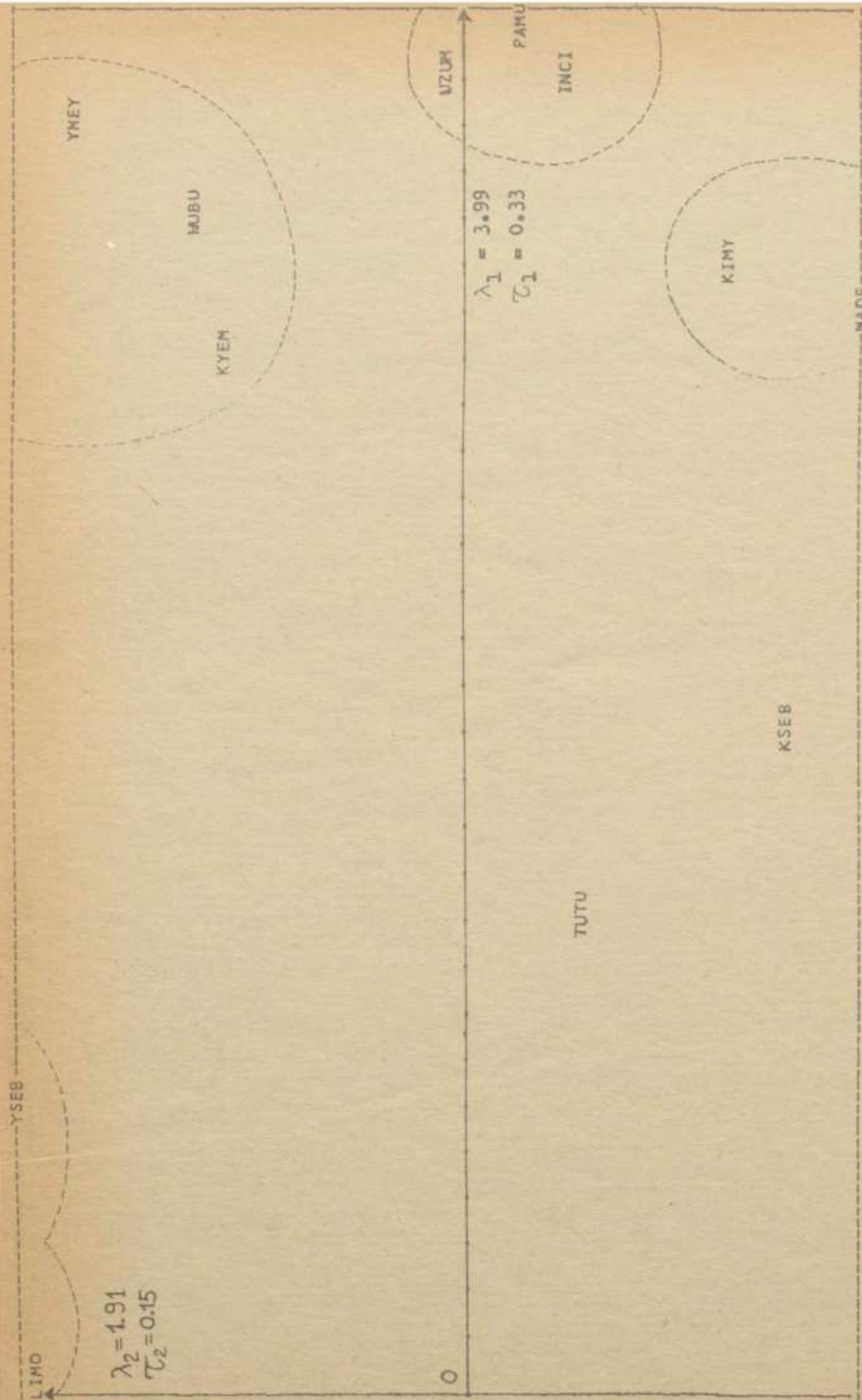
NO	DİRLİLER	KLT	EKS-1	KO2	EKS-2	KO2	EKS-3	KO2	EKS-4	KO2
1	Y576	342	-1601	95	2001	162	-465	8	-1468	80
2	N576	311	-739	20	2144	172	-2	0	-1786	119
3	L176	211	-924	74	-351	10	315	8	-1172	119
4	I776	575	-1246	301	-915	163	604	70	-464	41
5	U776	164	-433	25	-303	12	920	115	-325	14
6	YM76	555	-1493	368	87	0	1024	173	-293	14
7	KY76	805	-1450	111	-1957	203	-3206	547	-298	14
8	HU76	293	-864	90	-711	61	1028	128	-343	14
9	TU76	377	-163	0	-3340	293	1124	33	-216	1
10	PA76	631	-1762	449	-141	2	1111	178	-119	2
11	KI76	889	10159	801	1068	0	3074	74	-908	6
12	MA76	296	-321	8	540	24	-1036	89	1453	175
13	Y577	405	-795	73	-1397	226	924	99	-258	7
14	K577	571	-1525	384	901	134	216	7	-532	46
15	L177	317	-783	43	1293	118	-65	0	-1487	156
16	I777	217	-987	72	180	0	615	28	1260	117
17	U777	503	-1136	249	-917	166	453	18	-335	56
18	YM77	170	-77	1	-414	33	787	123	-359	13
19	KY77	297	-1293	127	384	11	1107	93	934	66
20	HU77	631	-1762	449	-141	2	1111	178	-119	2
21	TU77	846	1074	117	-2300	178	-4054	550	-177	1
22	PA77	658	-477	16	1684	206	906	59	2345	417
23	KI77	500	234	0	-5053	388	2720	112	90	0
24	MA77	773	4371	648	600	14	1782	107	228	1
25	Y578	223	50	0	-951	98	-792	68	730	57
26	K578	427	-914	117	-1369	263	539	40	-226	7
27	L178	626	-1823	430	900	106	489	30	-683	60
28	I778	295	-781	53	1078	101	-558	27	-1144	114
29	U778	115	-951	93	-302	9	26	0	337	11
30	YM78	631	-1762	449	-141	2	1111	178	-119	2
31	KY78	401	-972	129	-1122	172	450	27	-744	75
32	HU78	105	645	30	-462	15	717	37	-565	23
33	TU78	550	-1513	436	700	93	206	8	-297	13
34	PA78	810	3426	160	-2546	48	-6372	554	-789	8
35	KI78	847	829	8	3750	177	15	0	7254	662
36	MA78	584	453	4	-4681	498	1906	82	178	0
37	Y579	903	7177	817	1935	36	2224	48	-524	2
38	K579	359	751	12	517	6	-2755	174	2698	167
39	L179	199	-1191	76	-1367	100	455	11	-481	12
40	I779	631	-1762	449	-141	2	1111	178	-119	2
41	U779	354	-1965	127	2239	165	-146	0	-1378	62
42	YM79	663	-332	2	4017	421	-1720	77	-2687	188
43	KY79	261	-1181	147	-233	5	-1014	109	-50	0
44	HU79	346	-854	67	-1712	269	238	5	-704	45
45	TU79	132	936	75	-422	15	496	21	-503	21
46	PA79	355	-1583	167	1827	221	-267	4	445	3
47	KI79	821	2789	122	-2855	128	-6002	566	647	7
48	MA79	660	-808	12	3373	211	-685	8	4812	428
49	Y580	217	-101	0	-2493	186	904	5	1177	28
50	K580	631	-1762	449	-141	2	1111	178	-119	2
51	L180	832	7529	795	1182	19	1099	16	-450	2
52	I780	229	813	29	-599	16	-1119	25	1705	129
53	U780	332	-1405	192	-1025	102	550	29	-313	2
54	YM80	457	-1742	117	2563	254	-337	4	-1458	82
55	KY80	695	352	2	4619	398	-2074	72	-3959	263
56	HU80	177	-823	42	900	50	-1147	62	237	3
57	TU80	386	-896	105	-1166	178	282	10	-846	93
58	PA80	84	1153	75	10	0	291	4	-323	3
59	KI80	414	-1703	144	2042	216	-962	45	-427	9
60	MA80	531	-1762	449	-141	2	1111	178	-119	2

TABLO.6.11: (25,60) BOYUTLU K(1,J,T) TABLOSU'NUN KORELASYONMATRİSİ'Nİ AIT ÖZDEĞERLER VE ÖZDEĞER YÜZDELERİ.

NO	ÖZDEĞERLER	YÜZDELER	BİRİKLİMİ	MATRİSİN ÖZDEĞERLER HİSTOGRAMI
1	16.61724	30.77	30.77	*****
2	7.26158	13.45	44.21	*****
3	5.29291	11.10	55.31	*****
4	4.55900	8.44	63.76	*****
5	4.06270	7.52	71.28	*****
6	2.91747	5.41	76.69	*****
7	2.68962	4.98	81.67	*****
8	2.56590	4.75	86.42	*****
9	2.03446	3.77	90.19	*****
10	1.16394	2.16	92.34	**
11	1.09303	2.02	94.37	**
12	0.89183	1.65	96.02	**
13	0.60062	1.11	97.13	*
14	0.33919	0.63	97.76	*
15	0.31556	0.58	98.34	*
16	0.25307	0.47	98.81	*
17	0.17148	0.32	99.13	*
18	0.15023	0.28	99.41	*
19	0.10750	0.20	99.61	*
20	0.08112	0.15	99.76	*
21	0.06256	0.12	99.87	*
22	0.03640	0.07	99.94	*
23	0.02797	0.05	99.99	*
24	0.00364	0.01	100.00	*
25	0.00003	0.00	100.00	*
26	0.00001	0.00	100.00	*
27	0.00001	0.00	100.00	*
28	0.00001	0.00	100.00	*
29	0.00001	0.00	100.00	*
30	0.00000	0.00	100.00	*
31	0.00000	0.00	100.00	*
32	0.00000	0.00	100.00	*
33	0.00000	0.00	100.00	*
34	0.00000	0.00	100.00	*
35	0.00000	0.00	100.00	*
36	0.00000	0.00	100.00	*
37	0.00000	0.00	100.00	*
38	0.00000	0.00	100.00	*
39	0.00000	0.00	100.00	*
40	0.00000	0.00	100.00	*
41	0.00000	0.00	100.00	*
42	0.00000	0.00	100.00	*
43	0.00000	0.00	100.00	*
44	0.00000	0.00	100.00	*
45	0.00000	0.00	100.00	*
46	0.00000	0.00	100.00	*
47	0.00000	0.00	100.00	*
48	0.00000	-0.00	100.00	*
49	0.00000	-0.00	100.00	*
50	0.00000	-0.00	100.00	*
51	0.00000	-0.00	100.00	*
52	0.00000	-0.00	100.00	*
53	0.00000	-0.00	100.00	*
54	0.00000	-0.00	100.00	*
55	0.00000	-0.00	100.00	*
56	0.00000	-0.00	100.00	*
57	0.00000	-0.00	100.00	*
58	0.00001	-0.00	100.00	*
59	0.00001	-0.00	100.00	*
60	0.00001	-0.00	100.00	*



SEKIL, G. 1: (25, 12) DUYUTLU KIL, J) TABLOSU'NA AIT GRAFYLER, IN
1. CI VE 2. CI ANA BİLİŞEN DEN OLUSAN (1, 2) DÜZLEMINDE



ŞEKİL-6.2: (25,12) BOYUTLU $K(1,2)$ TABLOSU'NA AİT DEĞİSKENLERİN 1. GİTİ VE 2. GİTİ ANA BİLEŞENLERDEN OLUŞAN (1,2) DÜZLEHİNDE GÖSTERİLMİŞİ

ALMA

URGU

RUSY

RUMA
IRAN

LIBY

IRAK

AVUS

KUVE

ARABYUGO

DANI
YUNA

ISRA

HOLL

INGL

SURI

BELC
ISVE

JAPO

LUBN

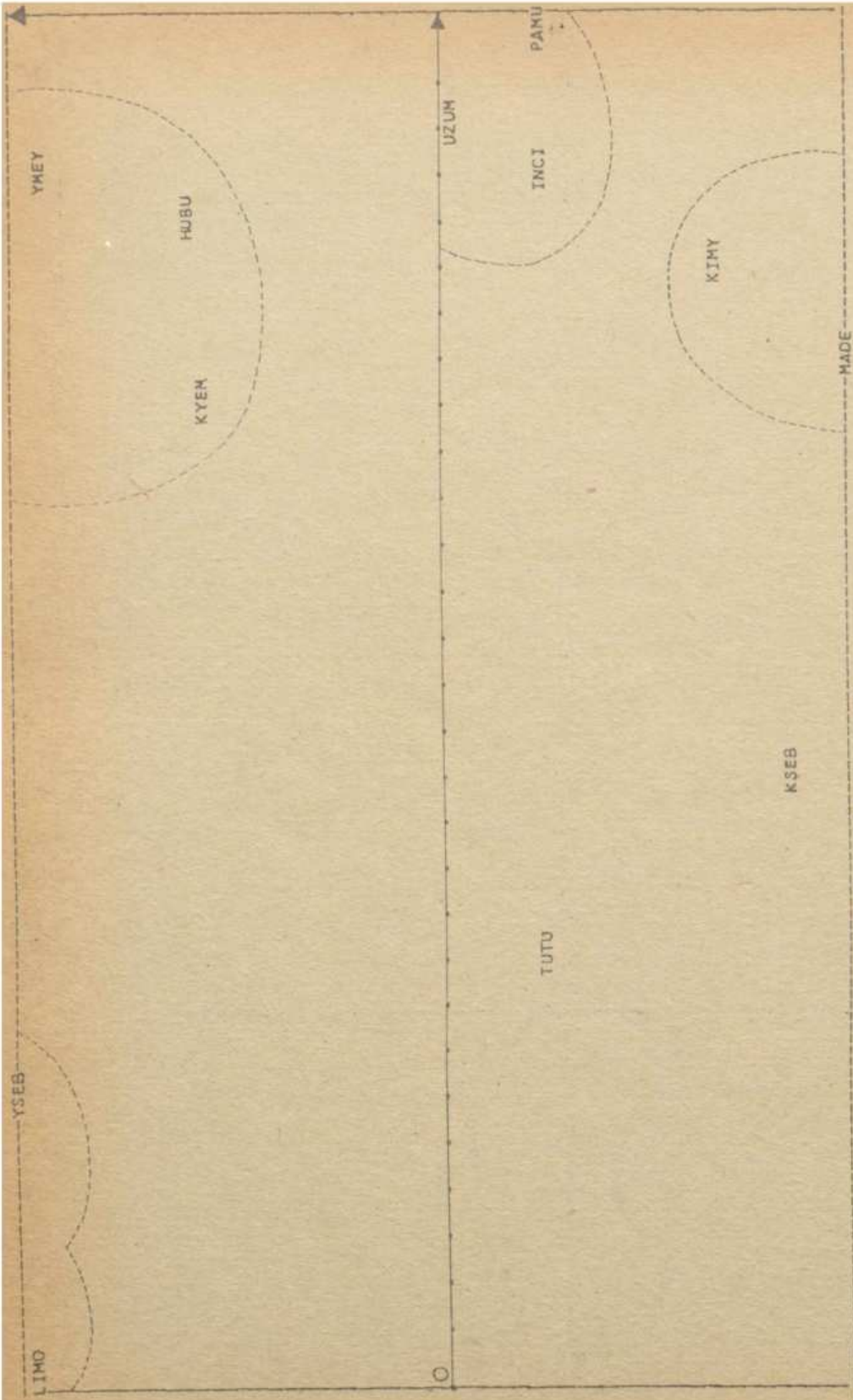
ISVI

FRAN

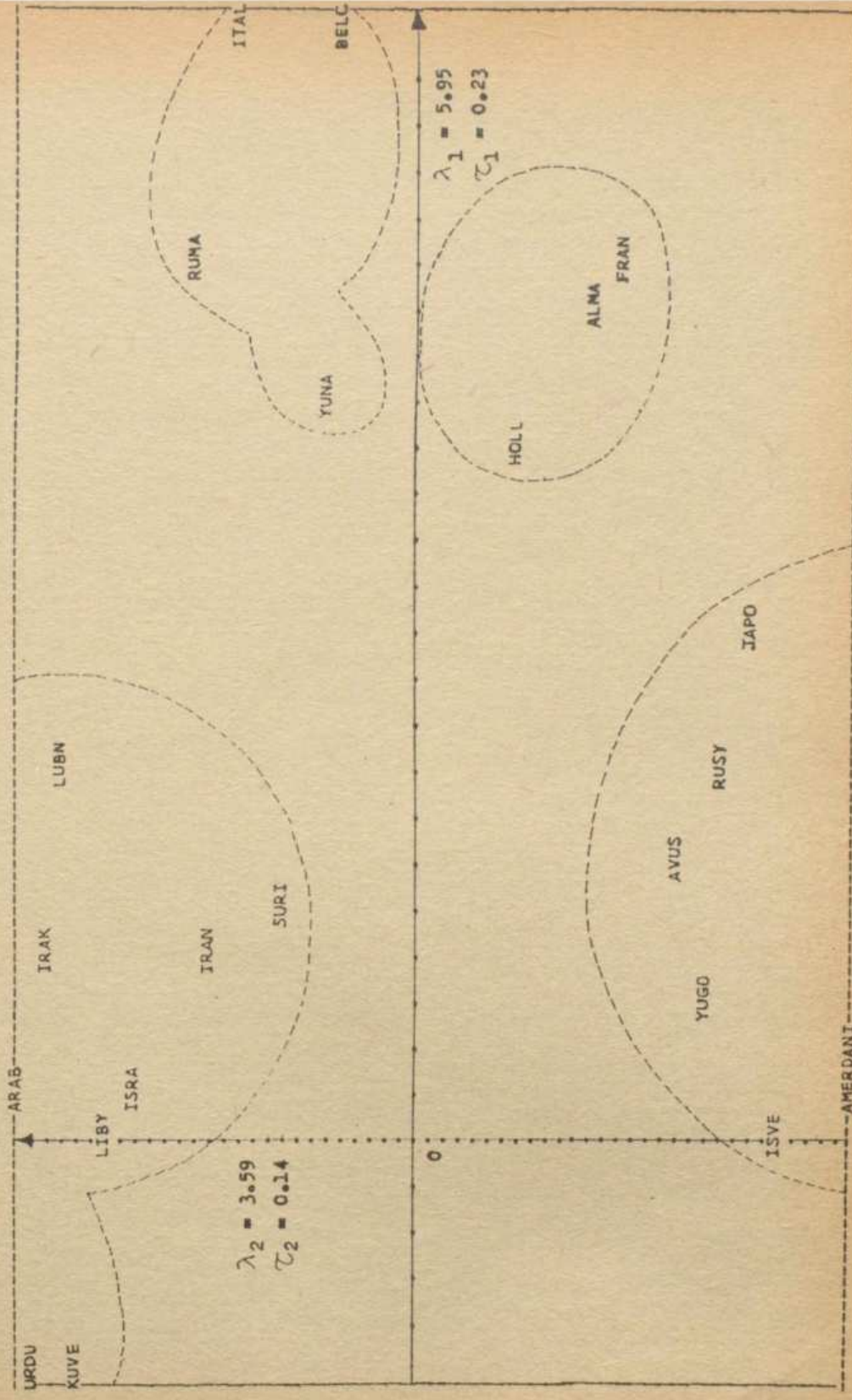
AMER

ITAL

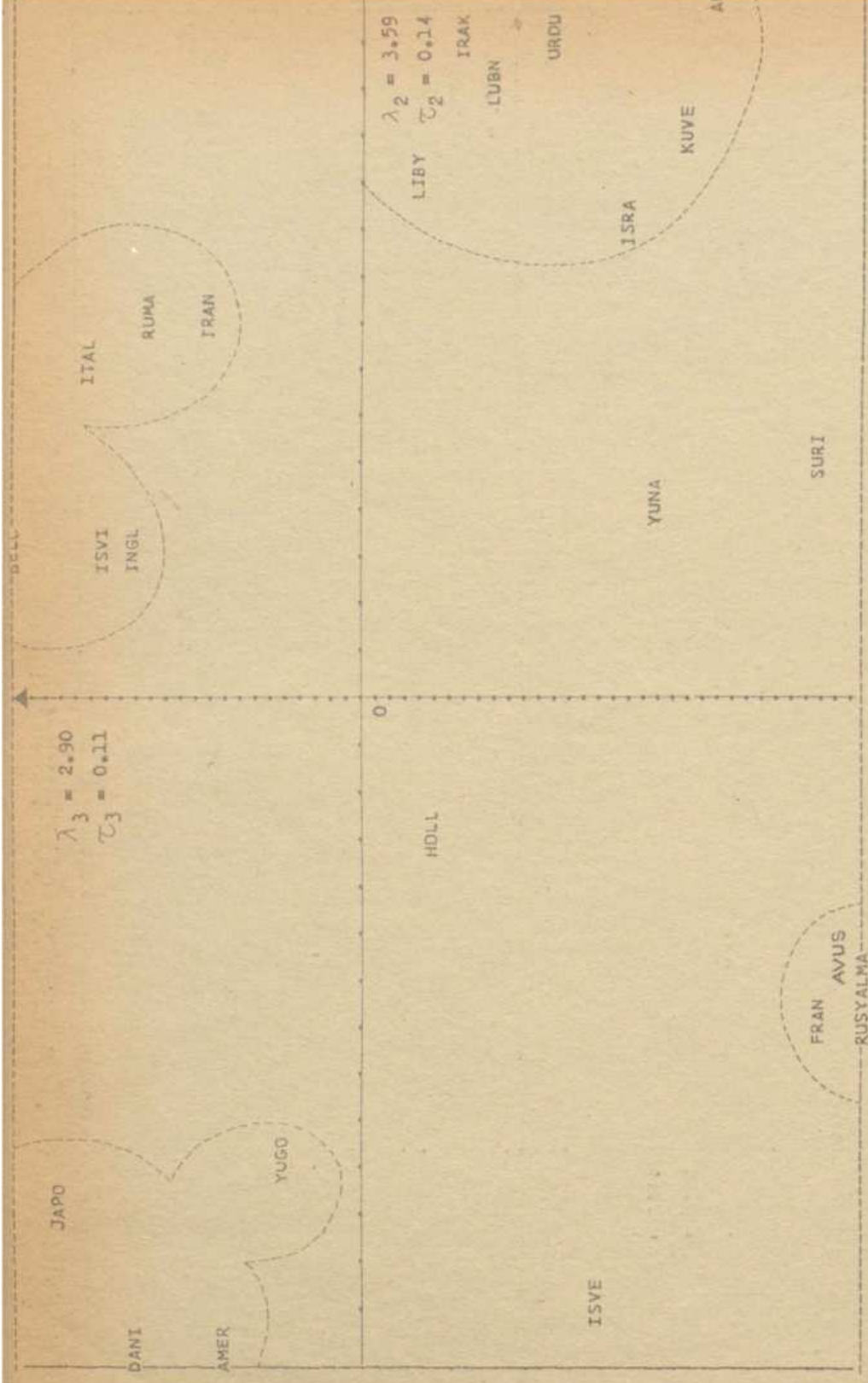
SEKİL-6.3: (25,12) BOYUTLU K(I,7) TABLOSU'NUN YENİDEN KURULUSUNU
VEREN FORMUL'UN ORTAYA KOYDUĞU SONUÇLAR(BİREYLER İÇİN).



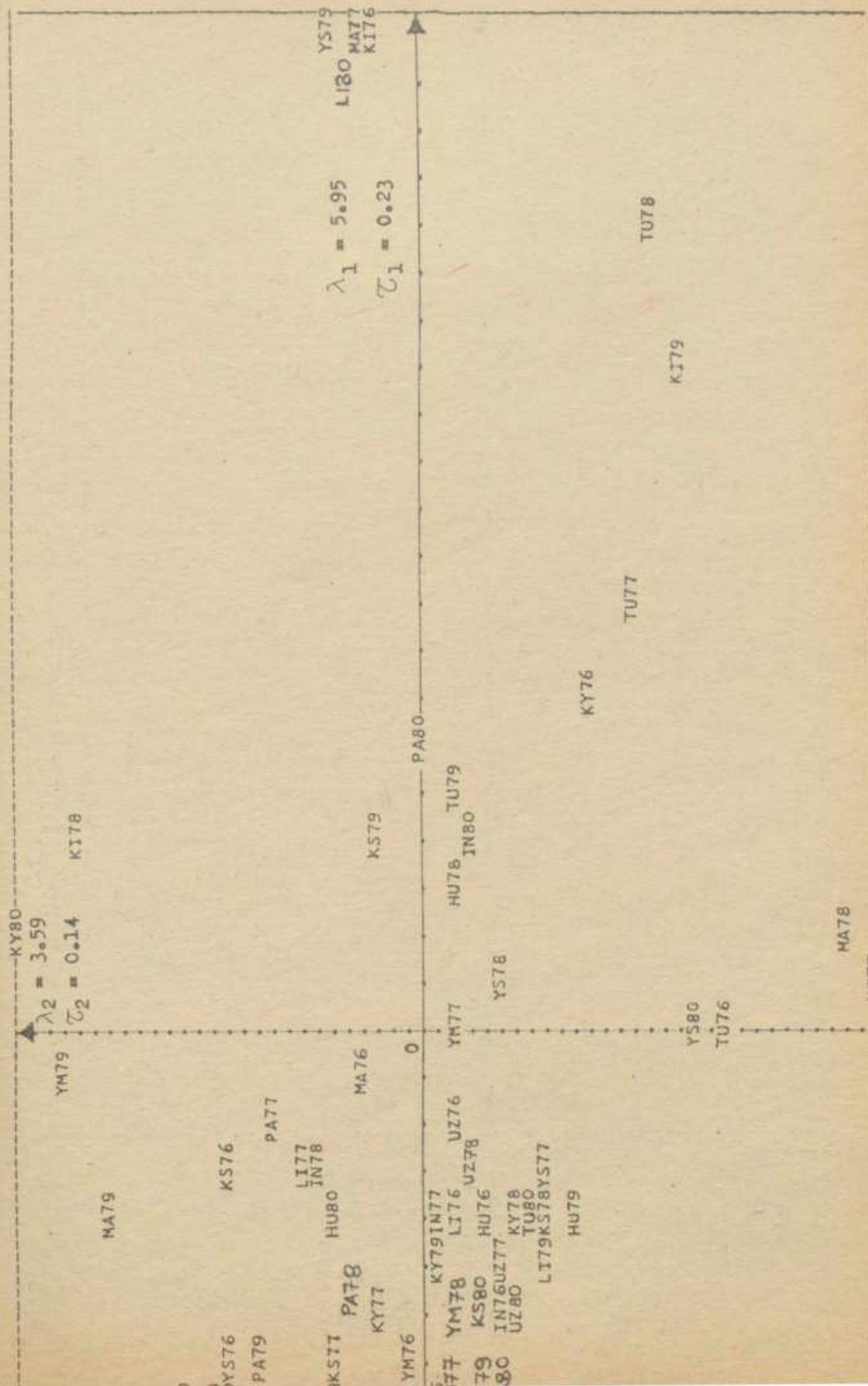
SEKIL-6.4: (25,12) BOYUTLU KILIJI TABLOSU'NUN YENIDEN KURULUSUNU VEREN FORMUL'UN ORTAYA KÖTÜGÜ SONUÇLAR (DEĞİŞKENLER 1-ÇIN)



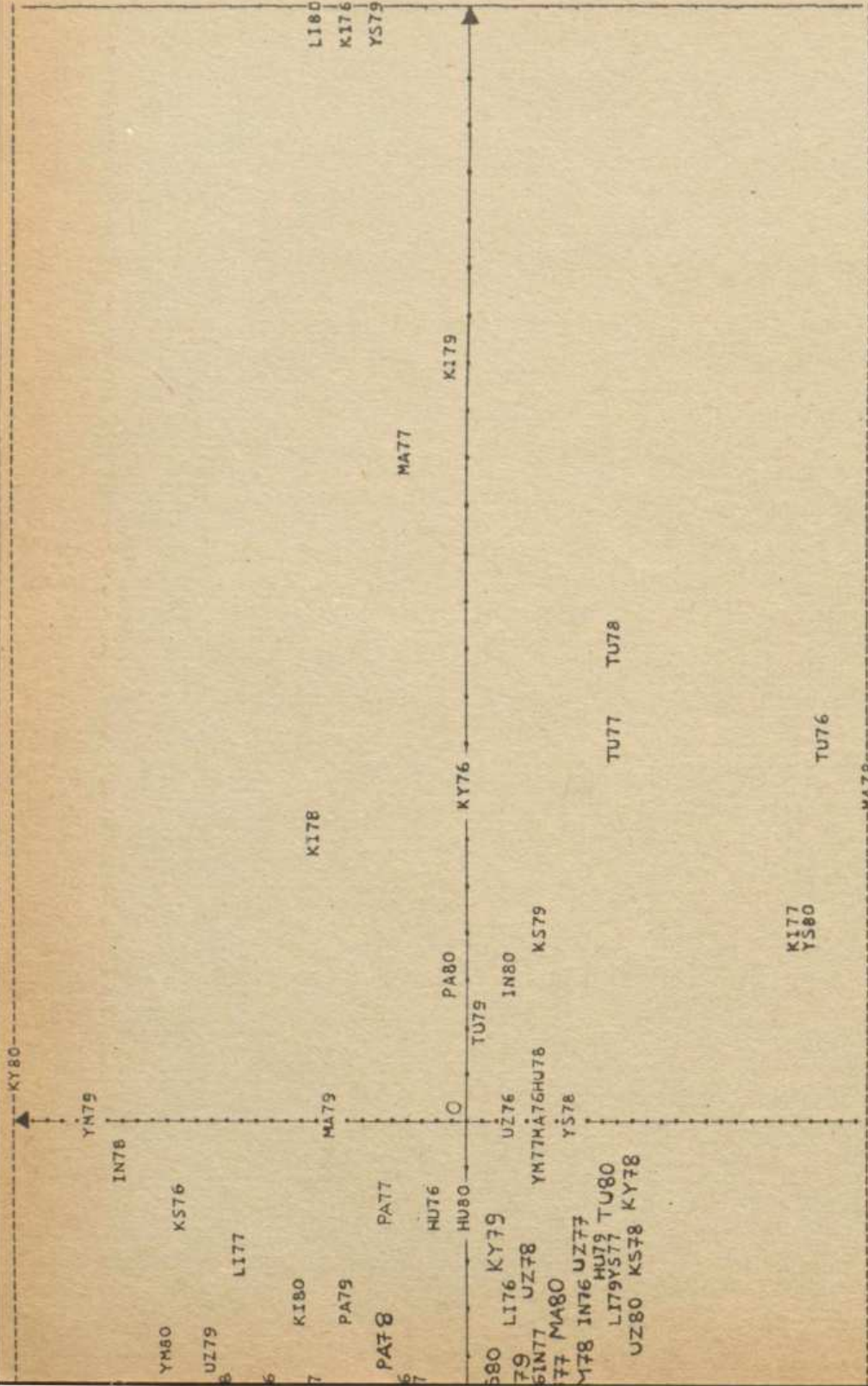
ŞEKİL-6.5: (60,25) BOYUTLU K(1,1,T) TABLOSU'NA AİT DEĞİŞKENLERİN 1. CI VE 2. CI BİLEŞENİN OLUŞAN (1,2) DÜZLEMİNDE GÖSTERİLMİŞİ.



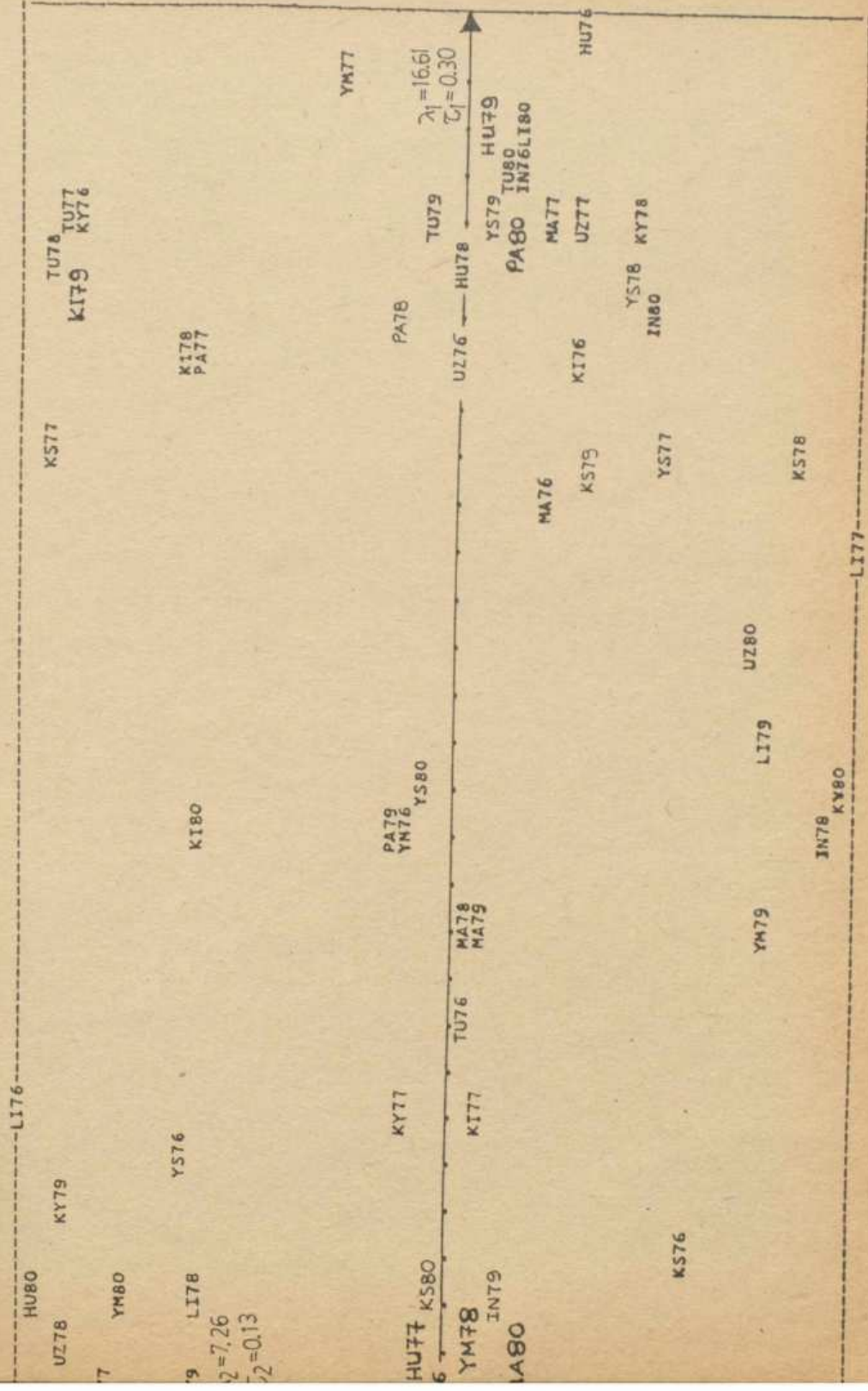
ŞEKL. 6.6: (60, 25) BOYUTLU KİT. J. T) TABLOSU'NA AIT DEĞİŞKENLERİN 2. Cİ VE 3. CÜ BİLEŞENİN OLUŞAN (2, 3) DÜZLEMİNDE GÖSTERİLMİ.



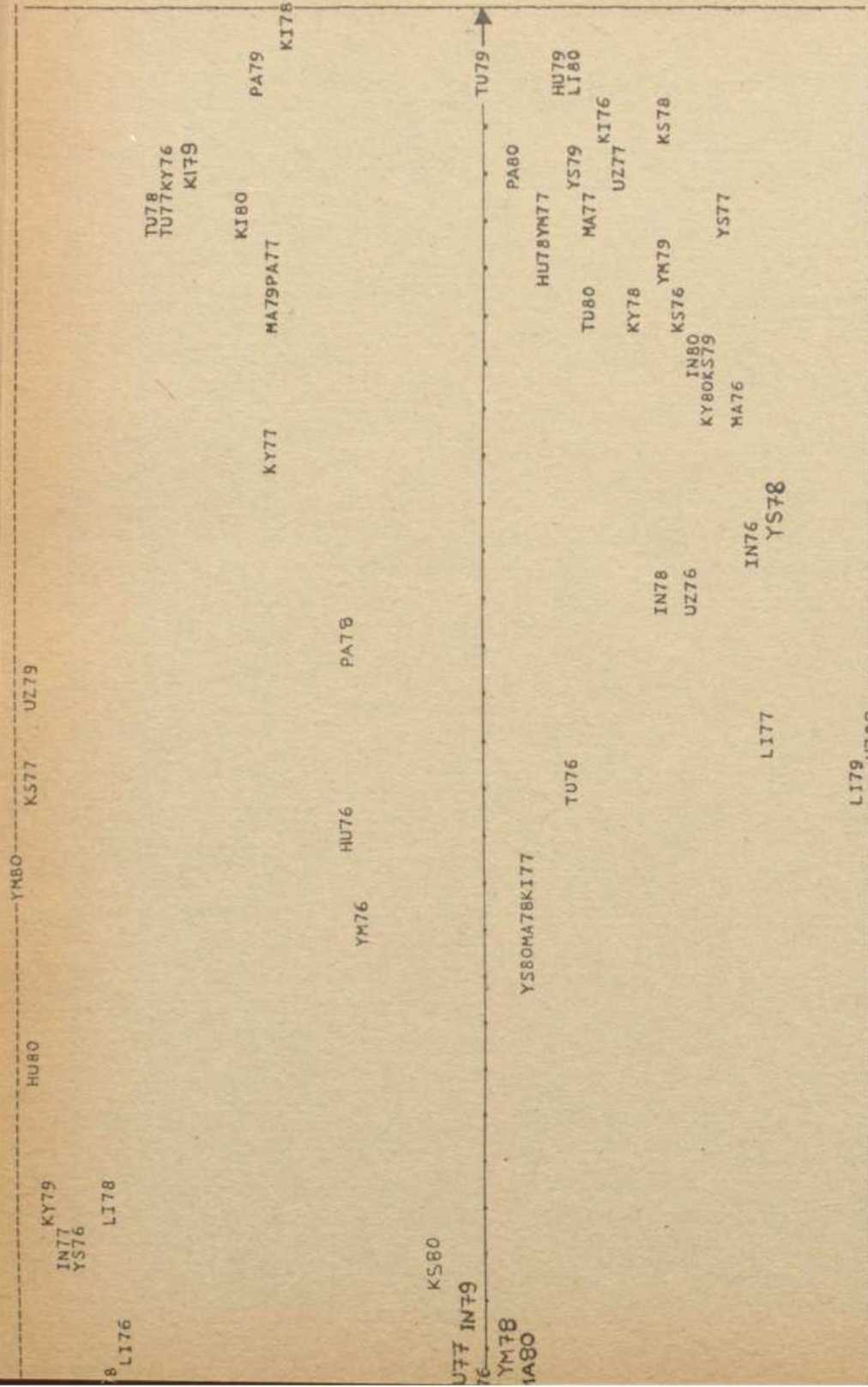
1.7: {60,25} BOYUTLU K(1,1,1) TABLOSU'NA AIT BIREYLERİN
 1.Cİ VE 2.Cİ BİLEŞENİN OLUSAN (1,2) DÜZLEMİNDE GÖS-
 TERİLMİŞ.



6.8: (60,25) BOYUTLU K(1,2,17) TABLOSU'NUN YENİDEN KURULUSU-
 NU VEREN FORMUL'UN ORTAYA KOYDUGU SONUÇLAR (BİREYLER
 İÇİN)



IL-6.9: (25, 60) BOYUTLU K(1,1,1,1) TABLOSUNA AIT BIREYLER-
IN-1, CI VE 2-CI BILEŞENDEN OLUŞAN (1,2) DÜZLEMİNDE
GÖSTERİLMİŞTİR.



KIL. 6.10: (25, 60) BOYUTLU K(I,J,T) TABLOSU'NUN YENIDEN KURULUSUNU YEREN FORMUL'UN ORTAYA KOYDUGU SONUÇLAR (B1-REYLER İÇİN)

REFERANSLAR

- (1) LAWLEY, D.N., MAXWEL, A.E.,: "Factor Analysis as a Statistical Method", Butterworths Group, London, 1963.
- (2) CATTEL, R.B.,: "Factor Analysis: An introduction to essentials I. The purpose and underlying models", Biometrics, 21, (1965), 190. II. The role of Factor Analysis in research, Biometrics, 21, (1965), 405.
- (3) HARMAN, H.H.,: "Modern Factor Analysis", University of Chicago Press, Chicago, 1960.
- (4) FRUCHTER, B.,: "Introduction to Factor Analysis", D. Van Nostrand Company, Inc. Princeton, New Jersey, 1954.
- (5) HORST, P.,: "Factor Analysis of Data Matrices", Holt Rinehart and Winston, New York, 1965.
- (6) ADCOK, C.J.,: "Simplified Factor Analysis", Occup. Psychol., Lond., 20, (1946), 188-198.
- (7) GUTTMAN, L.,: "The quantification of a glass of attributes: a theory and method of scale construction", In the Prediction of Personal Adjustment (P. Horst, ed.) New York, (1941), 251-364.
- (8) TORGERSON, W.S.,: "Theory and Methods of Scaling", New York: Wiley, 1958.
- (9) BENZECRI, J.P.,: "Statistical analysis as a tool to make patterns emerge from data", In Methodologies of Pattern Recognition (S. Watanabe, ed.) New York: Academic, (1969), 35-60.

- (10) BENZECRI, J.P.,: "Analyse des Données correspondances",
Dunod, 2, 1976.
- (11) WHITTAKER, R.H.,: "Gradient analysis of vegetation",
Biol. Rev., 42, (1967), 207-264.
- (12) KSHIRSAGAR, A.M.,: "Multivariate Analysis", Marcel
Dekker, Inc., New York, 1972.
- (13) KETTENRING, J.R.,: "Canonical analysis of several sets
of variables", Biometrika, 58, 3, (1971).
- (14) MORRISON, D.F.,: "Multivariate statistical Methods",
McGraw-Hill Book Co., New York, 1967.
- (15) SEAL, H.L.,: "Multivariate Statistical Analysis for
Biologists", Methven and Co. Ltd., London, 1964.
- (16) ANDERSON, T. W.,: "An Introduction to Multivariate
Statistical Analysis", New York: Wiley, 1958.
- (17) SPEARMAN, C.,: "The proof and measurement of association
between two things", Amer. Jour. Psychol., 15,
(1904), 201-293.
- (18) TUHRSTONE, L.L.,: "Multiple Factor Analysis", University
of Chicago Press, Chicago, 1947.
- (19) BURR, C.,: "The Factorial Analysis of qualitative data",
British Journal of Statistical Psychology, 3, 3,
(1969), 166-185.
- (20) HOTELLING, H.,: "The most predictable criterion", J. Educ.
Psychol., 26, (1935), 139-142.

- (21) HOTELLING, H.,: "Analysis of a Complex of statistical variables in to Principal Components", J. Educ. Psychol., 24, (1933), 417-441, 498-520.
- (22) BURT, C.,: "Alternative methods of Factor Analysis and their relations to pearson's methods of principal axes", Br. J. Psychol. Statist. 2, (1949), 98-121.
- (23) WILKS, S.S.,: "Mathematical Statistics", John Wiley and sons, Inc., New York, 1962.
- (24) GOWER, J.C.,: "Multivariate analysis and multidimensional geometry", The statistician, 17, (1967), 13-28.
- (25) KENDALL, M.G.,: "A Course in Multivariate Analysis", Griffin, London, 1975.
- (26) SPEARMAN, C.,: "General intelligence objectively determined and measured", Ame. J. Psychol., 15, (1904), 201-293.
- (27) WHITTLE, P.,: "On Principal Components and Least squares methods of Factor Analysis", Skandinavisk Aktuarietidskrift, 35, (1952), 223-239.
- (28) NASH, J.C., LEFKOVITCH, L.P.,: "Principal Components and regression by singular value decomposition on a small computer", Applied statistics, 25, 3, (1976), 210-215.
- (29) HIRSCHFELD, H.D.,: "A connection between correlation and contingency", Proc. Camb. Phil., Soc., 31, (1935), 520-524.
- (30) CAROLL, J.D.,: "Individual differences and multidimensional scaling", Polycoppy, Bell Telephon Laboratories, 1970.

- (31) CAROLL, J.D., CHANG, J.J.,: "Analyse of individual differences in multidimensional scaling via an N-Way generalisation of Eckart-Young decomposition", Psychometrika, 35, 3, (1970).
- (32) PEARSON, K.,: "On lines and planes of closest fit to systems of points in space", Phil. Mag., 2, 11, (1969), 559-572.
- (33) HOTELLING, H.,: "Relation between two sets of variables", Biometrika, 28, (1936), 321-377.
- (34) SHANNON, E.E., BREZONIK, P.L.,: "An index of eutrophication in American lakes", Journal of the Sanitary Engineering Division, Amer. Soc. of civil Engineers, 2, 37, (1977).
- (35) HAMMOND, W.H.,: "Factor Analysis as applied to social and economic data", Brit. J. Educ. Psychol., 16, 178 (Abstract of Ph. D. thesis; Univ. of Lond.), (1946).
- (36) STONE, R.,: "The analysis of market demand", Journal of the Royal Statistical Society, (1945), 108-297.
- (37) STONE, R.,: "The measurement of consumer's expenditure and behaviour in the United Kingdom", 1920-1938 (Cambridge University Press, Cambridge, Mass.), 1954.
- (38) JOLICOEUR, P., MOSIMAN, E.,: "Size and shape variation in the painted turtle: A principal component Analysis", Growth, 24, (1960), 339-354.
- (39) GNANADESIKAN, R.,: "Methods for Statistical Data Analysis of Multivariate Observations", John Wiley and Sons, New York, 1977.

- (40) HAWKINS, D.M.,: "On the investigation of alternative regressions by Principal Component Analysis", Applied Statistics, 22, (1973), 275-286.
- (41) HADDAD, C.L.,: "Testing income series an application of Principal Components", Journal of Econometrics, 4, (1976), 27-40.
- (42) HAMILTON, M.,: "Development of a rating scale for primary depressive illness", Brit. J. Soc. Clin. Psychol., 6, (1967), 278-296.
- (43) RAO, C.R., SLATER, P.,: "Multivariate analysis applied to differences between neurotic groups", Brit. J. Psychol. (Stat. sect.), 2, (1949), 17-29.
- (44) ANDERSON, J.A., BOYLE, J.A.,: "Computer diagnosis statistical aspects", Brit. Med. Bull., 24, (1968), 230-235.
- (45) ANDERSON, J.A.,: "Multivariate methods in clinical assessment", J.R. Statist. Soc. A., 139, 2, (1976), 161-165.
- (46) MARDIA, K.V., KENT, J.T., BIBBY, J.M.,: "Multivariate Analysis", United States Edition Published by Academic Press, London, 1979.
- (47) HOTELLING, H.,: "Simplified calculation of Principal Components", Psychometrika, 27, (1936), 1-35.
- (48) OBENCHAIN, R.L.,: "Regression optimality of Principal Components", The Annals of Mathematical statistics, 43, 4, (1972), 1317-1319.

- (31) CAROLL, J.D., CHANG, J.J.,: "Analyse of individual differences in multidimensional scaling via an N-Way generalisation of Eckart-Young decomposition", *Psychometrika*, 35, 3, (1970).
- (32) PEARSON, K.,: "On lines and planes of closest fit to systems of points in space", *Phil. Mag.*, 2, 11, (1969), 559-572.
- (33) HOTELLING, H.,: "Relation between two sets of variables", *Biometrika*, 28, (1936), 321-377.
- (34) SHANNON, E.E., BREZONIK, P.L.,: "An index of eutrophication in American lakes", *Journal of the Sanitary Engineering Division, Amer. Soc. of civil Engineers*, 2, 37, (1977).
- (35) HAMMOND, W.H.,: "Factor Analysis as applied to social and economic data", *Brit. J. Educ. Psychol.*, 16, 178 (Abstract of Ph. D. thesis; Univ. of Lond.), (1946).
- (36) STONE, R.,: "The analysis of market demand", *Journal of the Royal Statistical Society*, (1945), 108-297.
- (37) STONE, R.,: "The measurement of consumer's expenditure and behaviour in the United Kingdom", 1920-1938 (Cambridge University Press, Cambridge, Mass.), 1954.
- (38) JOLICOEUR, P., MOSIMAN, E.,: "Size and shape variation in the painted turtle: A principal component Analysis", *Growth*, 24, (1960), 339-354.
- (39) GNANADESIKAN, R.,: "Methods for Statistical Data Analysis of Multivariate Observations", John Wiley and Sons, New York, 1977.

- (40) HAWKINS, D.M.,: "On the investigation of alternative regressions by Principal Component Analysis", Applied Statistics, 22, (1973), 275-286.
- (41) HADDAD, C.L.,: "Testing income series an application of Principal Components", Journal of Econometrics, 4, (1976), 27-40.
- (42) HAMILTON, M.,: "Development of a rating scale for primary depressive illness", Brit. J. Soc. Clin. Psychol., 6, (1967), 278-296.
- (43) RAO, C.R., SLATER, P.,: "Multivariate analysis applied to differences between neurotic groups", Brit. J. Psychol. (Stat. sect.), 2, (1949), 17-29.
- (44) ANDERSON, J.A., BOYLE, J.A.,: "Computer diagnosis statistical aspects", Brit. Med. Bull., 24, (1968), 230-235.
- (45) ANDERSON, J.A.,: "Multivariate methods in clinical assessment", J.R. Statist. Soc. A., 139, 2, (1976), 161-165.
- (46) MARDIA, K.V., KENT, J.T., BIBBY, J.M.,: "Multivariate Analysis", United States Edition Published by Academic Press, London, 1979.
- (47) HOTELLING, H.,: "Simplified calculation of Principal Components", Psychometrika, 27, (1936), 1-35.
- (48) OBENCHAIN, R.L.,: "Regression optimality of Principal Components", The Annals of Mathematical statistics, 43, 4, (1972), 1317-1319.

- (49) NASH, J.C.,: "Small computer programs for ols and Principal components regression calculations", *Econometrika*, 44, 4, (1976), 190-195.
- (50) CHATFIELD, C., COLLINS, A.J.,: "Introduction to Multivariate Analysis", Chapman and Hall, London, 1980.
- (51) EASTMENT, H.T., KRZANOWSKI, W.J.,: "Cross-Validatory choice of the number of components from a Principal Component Analysis", *Technometrics*, 24, 1, (1982), 73-77.
- (52) COOLEY, W.W., LOHNES, P.R.,: "Multivariate Procedures for the Behavioral Sciences", John Wiley and sons, Inc., New York, 1967.
- (53) OKAMOTO, M.,: "Optimality of Principal Components", *Multivariate Analysis*, 2 (P.R. Krishnaiah, ed.) 673, New York: Academic Press, 1969.
- (54) CHEN, C.W.,: "An optimal property of principal components", *Communications in Statistics*, 3, 10, (1974), 979-983.
- (55) BENZECRI, J.P.,: "Sur l'Analyse des tableaux binaires associés à une correspondance multiple", Université Pierre et Marie Curie, 1976. Cahier de l'Analyse de Données, II, 1, 1976.
- (56) BENZECRI, J.P.,: "Approximation stochastique dans une Algèbre normée non commutative", *Bul. Soc. Math. France*, 97, (1976), 225-241.
- (57) FISHER, R.A.,: "The precision of discriminant function", *Ann. Eugen. London*, 10, (1940), 422-429.
- (58) NAOURI, J-Ch.,: "Analyse Factorielle des correspondances", Thèse de 3^o cycle, Paris, 1971.

- (59) CAZES, P.,: "Traitement des Problèmes Géologiques",
Thèse (Université Paris IV), 1970.
- (60) HATHEWAY, W.H.,: "Contingency-table analysis of rain
forest vegetation", In Statistical Ecology, 3,
Many Species Populations, Ecosystems and systems
Analysis, Pennsylvania State University Press,
(1971), 149-166.
- (61) MAUNG, K.,: "Measurement of association in a contingency
table with special reference to the pigmentation
of hair and eye colours of scottish school children",
Ann, Eugen. Lond., 11, (1941), 189-223.
- (62) BENZECRI, J.P.,: "Cours de Linguistique Mathématique",
Faculté des sciences de Rennes, 1964.
- (63) BENER, A.,: "Sürekli olmayan halde Tekabüller tablosunun
Faktörel Analizi metodunun temel bileşenli faktö-
rel analiz metodundan hareket edilerek incelenmesi
ve elde edilen modelin bir biyoloji problemine
uygulanması", Doktora Tezi, İ.T.Ü. Temel Bilimler
Fakültesi Yayınları, İstanbul, 1976.
- (64) ROMEDER, J.M.,: "Methode et Programme d' Analyse Dis-
criminant", Dunod, 1973.
- (65) DEMSTER, A.P.,: "Elements of Continuous Multivariate
Analysis", Addison-Wesley Publishing Company,
Massachusetts, 1969.
- (66) HILL, M.O.,: "Correspondance Analysis: A neglected mul-
tivariate method", Appl. Statist., 23, 3, (1974),
340-345.

- (67) KRZANOWSKI, W.J.,: "Between-Groups comparison of Principal Components", Journal of the American Statistical Association, 74, 367, (1979), 703-707.
- (68) HORST, P.,: "Relation among m sets of measures", Psychometrika, 26, (1965), 129-149.
- (69) MASSON, M.,: "Analyse non linéaire de données", C.R. Acad. Sc., 278, (1974).
- (70) ECKART, C., YOUNG, G.,: "The approximation of on matrix by another of lower rank", Psychometrika, 1, (1936), 211-218.
- (71) CAROLL, J.D.,: "Generalisation of canonical correlation to three or more sets of variables", Proc. Amer. Psych. Assoc., (1970), 227-228.
- (72) BENER, A.,: "Etude par l' analyse des Correspondances des interactions dan un tableau ternaire. Application à des données linguistiques", Thèse, Université Pierre et Marie Curie, 1981.
- (73) KENDALL, M.G., STUART, A.,: "The Advanced Theory of Statistics", Inference and Relationship, 2, London: Griffin, 1961.
- (74) WILKINSON, J.H.,: "The Algebraic Eigenvalue Problem", Clarendon Press, Oxford, 1965.
- (75) VELICER, W.F.,: "Determining the number of components from the matrix of partial correlations", Psychometrika, 41, (1976), 321-327.

- (76) ANDERSON, T.W.,: "Asymptotic theory for Principal Component Analysis", Annals of Mathematical Statistics, 34, (1958), 122-148.
- (77) D.İ.E., Dış Ticaret Yıllık İstatistikleri, B.D.İ.E., Ankara, 1976-1980.

TEŞEKKÜR

Bu çalışmayı yöneten, çalışma sırasında bilgi ve tecrübelerinden büyük ölçüde yararlandığım değerli hocam Sayın Prof. Dr. Cevdet KOÇAK'a en içten teşekkürlerimi sunarım.

Çalışmanın her aşamasında, kısıtlı zamanını büyük bir fedakârlıkla bana ayıran ve yardımlarını esirgemeyen kıymetli hocam Sayın Yrd. Doç. Dr. Aziz BENER'e sonsuz teşekkürlerimi sunarım.

ÖZGEÇMİŞ

Eyüp Sabri Türker, 1956 yılında Trabzon'un Çaykara ilçesinde doğmuş, lise öğrenimini Trabzon Lisesi'nde tamamlamıştır. 1973 yılında girdiği K.T.Ü. Temel Bilimler Fakültesi Matematik Mühendisliği Bölümü'nü 1977 yılında bitirmiştir. Aynı yıl S.D.M.M. Akademisi Matematik Kürsüsü'ne asistan olarak girmiştir. 1978 yılında Ege Üniversitesi'nde başladığı yüksek lisans öğrenimini 1980 yılında tamamlamıştır.

Halen İ.T.Ü. Sakarya Mühendislik Fakültesi Matematik Bölümü'nde araştırma görevlisi olarak görevine devam etmektedir. Evli ve bir çocuk babasıdır.

