



YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

öss sınav sonuçlarının okul bazınd..

Yüksek Lisans Tezi

ömer bilen

360
1

YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

ÖSS SINAV SONUÇLARININ OKUL BAZINDA
VERİ MADENCİLİĞİ İLE İNCELENMESİ

1100

İstatistikçi Ömer BİLEN

FBE İstatistik Anabilim Dalı İstatistik Programında
Hazırlanan

YÜKSEK LİSANS TEZİ

Tez Danışmanı: Doç. Dr. Ali Hakan BÜYÜKLÜ,

İSTANBUL, 2004

YILDIZ TEKNİK ÜNİVERSİTESİ
KÜTÜPHANE VE DOKÜMANTASYON
DAİRE BAŞKANLIĞI

Yer No (DDC) : R/360/1

Kayıt No : 2347.....

Geldiği Yer : Fen Bilimleri Ens.....

Tarih : 9/11/2004.....

Fiyat : 16.100.000.....

Fatura No :

Ayniyat No : 1/13.....

Ek :

ÖSS SINAV SONUÇLARININ OKUL BAZINDA VERİ MADENCİLİĞİ İLE İNCELENMESİ

İstatistikçi Ömer BİLEN

FBE İstatistik Anabilim Dalı İstatistik Programında

Hazırlanan

YÜKSEK LİSANS TEZİ

Prof. Dr. Sahamet Bulbul
Sahamet Bulbul

Tez Danışmanı: Doç. Dr. Ali Hakan BÜYÜKLÜ

Doç. Dr. Ali Hakan Büyüklü

Prof. Dr. Haldun Akın
Haldun Akın

İSTANBUL, 2004

İÇİNDEKİLER

Sayfa

SİMGE LİSTESİ	vi
KISALTMA LİSTESİ	vii
ŞEKİL LİSTESİ	viii
ÇİZELGE LİSTESİ	xii
ÖNSÖZ... ..	xix
ÖZET	xx
ABSTRACT	xxi
1. GİRİŞ	1
2. Veri Madenciliği	3
2.1 Veri Madenciliğinin Ortaya Çıkışı	7
2.2 Veri Madenciliği Niçin Tercih Ediliyor?	13
2.3 Birkaç Veri Madenciliği Uygulama Örneği	15
2.3.1 US WEST Firması'nda Bir Veri Madenciliği Uygulaması (Groth,1999)	15
2.3.2 Bass Export Firması'nda Bir Veri Madenciliği Uygulaması (Groth,1999)	16
3. Veri Madenciliği ve Diğer Disiplinler	17
3.1 Veri Tabanı Teknolojisi	18
3.1.1 İlişkisel Veritabanları	23
3.1.2 Veri Ambarları	25
3.1.2.1 Nesne Yönelimlilik	26
3.1.2.2 Entegre Olma	28
3.1.2.3 Zaman Değişimlilik	28
3.1.2.4 Geçmişe Dönük Değişimsizlik	28
3.1.3 İşlemsel (Transactional) Veritabanları	29
3.1.4 Uzman Veritabanı Sistemleri ve Uygulamaları	31
3.1.4.1 Nesne Yönelimli Veritabanları	31
3.1.4.2 Nesne İlişkisel Veritabanları	32
3.1.4.3 Uzamsal Veritabanları	32
3.1.4.4 Zamansal Veritabanları ve Zaman Serileri Veritabanları	32
3.1.4.5 Tekst Veritabanları ve Multimedya Veritabanları	33
3.1.4.6 Heterojen Veritabanları ve Daha Önceden Kullanılan (Legacy) Veritabanları	33
3.1.4.7 World Wide Web	34
3.2 Görselleştirme	34
3.2.1 Zaman Serileri	35
3.2.2 Konum Serileri	36
3.2.3 Nitel Seriler	36

3.2.4	Nicel Seriler	38
3.2.5	Serpilme Diyagramları	39
3.3	Karar Destek Sistemleri	40
3.3.1	OLTP Sistemleri	40
3.3.2	OLAP Sistemleri	44
3.3.2.1	Veri Ambarı Verisinin Analizi	45
3.3.2.2	Çok Boyutlu Veri Modeli	45
3.3.2.2.1	Bir Küpün Boyutları	46
3.3.2.2.2	Bir Küpün Ölçüleri	49
3.3.2.2.3	Bir Küpün Bölümleri (Partitions)	49
3.3.2.2.3.1	Küpün Bölmelere Ayrılmasının Faydaları	50
3.3.2.2.3.2	Küpün Depolanması Yöntemleri	50
3.3.2.2.4	Bir Küpte Kullanıcı Rollerini	51
3.3.2.2.5	Çok Boyutlu Veritabanlarında Kullanılan Star, Snowflakes, Fact Constellations Şemaları	51
3.3.2.2.5.1	Star Şeması	52
3.3.2.2.5.2	Snowflakes Şeması	52
3.3.2.2.5.3	Galaxy Şeması	53
3.3.2.2.6	OLAP Küpü Tablosu Örneği	55
3.3.2.2.7	OLAP Küplerinde Gerçekleştirilen İşlemler	56
3.3.2.2.8	Veri Ambarında Yer Alan Verilerin Analizi Süreci	60
3.3.3	OLTP Sistemleri İle OLAP Sistemlerinin Farkları (Han ve Kamber,2001)	62
3.4	Makine Öğrenmesi	63
3.4.1	Öğrenme Yaklaşımları	64
3.4.2	Öğrenme Yöntemleri	64
3.5	İstatistik	66
4.	Veri Madenciliği Metodolojisi	71
4.1	Karar Probleminin Belirlenmesi	73
4.2	Veri Hazırlama	73
4.2.1	Veri Temizleme	73
4.2.1.1	Kayıp Değerler	73
4.2.1.2	Aşırı Değerler	74
4.2.1.3	Uyumsuz (Inconsistent) Veri	75
4.2.2	Veri Entegrasyonu	76
4.2.3	Transformasyon	76
4.2.4	Veri İndirgeme	77
4.3	Veri Analizi	77
4.4	Sonuçların Yorumlanması	78
4.5	SEMMA Metodolojisi	78
5.	Veri Madenciliğinde Veri İndirgemesi İşlemlerinde Kullanılan İstatistik Tekniklerden Temel Bileşenler Analizi ile Kümeleme Analizinin İncelenmesi ...	80
5.1	Temel Bileşenler Analizi	81
5.1.1	Standardize Değişkenlerden Elde Edilen Temel Bileşenler	87
5.1.2	Varyans Kovaryans Matrisinde veya Korelasyon Matrisinde Özel Yapılar	88
5.1.3	Temel Bileşenler İle Değişkenler Arasındaki Korelasyon	89
5.1.4	Temel Bileşenlerin Rotasyonu	89
5.1.5	Temel Bileşenlerin Yorumlanması	90
5.1.6	Temel Bileşenlerin Grafik Gösterimi	90

5.1.7	Kaç Adet Temel Bileşenin Kullanılacağına Seçimi	91
5.2	Kümeleme Analizi	93
5.2.1	Kümeleme Analizi İçin Akış Diyagramı (Hair vd,1995)	96
5.2.2	Kümeleme Analizinde Değişken Türlerine Göre Benzerlik ve Uzaklık Ölçüleri ..	98
5.2.2.1	Aralık Ölçekli ve Oransal Ölçekli Değişkenler	98
5.2.2.1.1	Öklidyen Uzaklık Ölçüsü	98
5.2.2.1.2	Pearson Uzaklık Ölçüsü.....	99
5.2.2.1.3	Manhattan Uzaklık Ölçüsü	99
5.2.2.1.4	Minkowski Uzaklık Ölçüsü	99
5.2.2.1.5	Uzaklık Fonksiyonunun Özellikleri.....	100
5.2.2.1.6	Değişkenlerin Standardizasyonu ve Değişkenlerin Dönüştürülmesi.....	100
5.2.2.2	Nominal Ölçekli Değişkenler	102
5.2.2.2.1	İkili Değişkenler	102
5.2.2.2.2	İkili Olmayan Nominal Değişkenler.....	105
5.2.2.3	Ordinal Ölçekli Değişkenler	105
5.2.3	Kümeleme Metotlarının Sınıflandırılması.....	106
5.2.3.1	Hiyerarşik Kümeleme Metotları	107
5.2.3.1.1	Single Linkage.....	109
5.2.3.1.2	Complete Linkage.....	112
5.2.3.1.3	Average Linkage.....	114
5.2.3.2	Bölmeleme Metotları	117
5.2.3.2.1	K-Means Bölmeleme Metodu.....	118
5.2.3.2.2	K-Medoids Bölmeleme Metodu	120
5.2.3.3	Yoğunluk Bazlı Metotlar,Grid Bazlı Metotlar, Model Bazlı Metotlar.....	124
6.	UYGULAMA	125
6.1	Uygulamada Kullanılacak Olan Değişkenler	125
6.2	Uygulama Aşaması	127
6.2.1	Verilerin SAS Programında Analiz Edilecek Formata Getirilmesi.....	127
6.2.2	Enterprise Miner Modülünün Açılması	130
6.2.3	Analiz Aşaması	131
6.2.3.1	Veri Setinin Seçilmesi	131
6.2.3.2	Değişken Seçimi	140
6.2.3.3	Değişkenlerin Transformasyonu.....	145
6.2.3.4	Aşırı Değerlerin Filtrelenmesi	150
6.2.3.5	Temel Bileşenler Analizi	156
6.2.3.5.1	Temel Bileşenler Analizinin Sonuçları.....	162
6.2.3.6	Kümeleme Analizi	166
6.2.3.6.1	Kümeleme Analizinin Sonuçları	172
6.2.3.6.2	Küme Profilleri	181
7.	SONUÇLAR.....	198
	KAYNAKLAR.....	201

SİMGE LİSTESİ

λ	Özdeğer
μ	Anakütle Ortalaması
σ	Anakütle Standart Sapması
Σ	Varyans Kovaryans Matrisi
$\mathbf{l}$	Özdeğer Vektörü
E	Beklenen Değer Operatörü
e	Özvektör
Var	Varyans
Cov	Kovaryans

EKLER.....	203
EK 1 Veri Setinde Bulunan Nicel Değişkenlere İlişkin Histogramlar	204
EK 2 Veri Setinde Bulunan Nicel Değişkenlere İlişkin Kutu Diyagramları.....	216
EK 3 Kümelerin Nicel Değişkenler Bazında İstatistiksel Farklılığının Bonferroni Güven Bölgesi Yaklaşımı ile Testi.....	228
EK 4 Veri Setinde Bulunan Nicel Değişkenlerin Kümelere Göre İstatistikleri	264
EK 5 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Göre Dağılımları	277
EK 6 D40 Değişkenindeki Lise Türlerinin Kümelere Göre Dağılımları.....	293
EK 7 D41 Değişkenindeki Lise Türlerinin Kümelere Göre Dağılımları.....	294
EK 8 D42 Değişkenindeki Lise Türlerinin Kümelere Göre Dağılımları.....	295
EK 9 ÖSYM'nin Tanımladığı Lise Türlerinin Testlere Göre Kümelerdeki Net Ortalamaları.....	300
EK 10 D40 Değişkenindeki Lise Türlerinin Testlere Göre Kümelerdeki Net Ortalamaları ..	312
EK 11 D41 Değişkenindeki Lise Türlerinin Testlere Göre Kümelerdeki Net Ortalamaları ..	314
EK 12 D42 Değişkenindeki Lise Türlerinin Testlere Göre Kümelerdeki Net Ortalamaları ..	316
ÖZGEÇMİŞ.....	322

KISALTMA LİSTESİ

CRM	Customer Relation Management	9
DTS	Data Transformation Services	17
LAN	Local Area Network	19
MST	Minimum Spanning Tree	20
OLTP	On Line Transaction Processing	20
OLAP	On Line Analytic Processing	21
PAM	Partitioning Around Medoids	22
ROI	Risk of Investment	22
WWW	World Wide Web	23
Y.D.A.	Yabancı Dil Ağırlıklı	24
Sekil 3.30 Veri Ambianın Kuvvetli Sesi		25
Sekil 3.31 OLAP Server'da OLAP Kapıda FoodMart Veri Ambianın Analize Tahlil Edilen Nesneler		27
Sekil 3.32 OLAP Server'da Kullanılan FoodMart Veri Ambianın Kuvvetli Sesi		27
Sekil 3.33 DTS ile Veri Alınarak Bazı Veri Tabloları		28
Sekil 3.34 Veri Ambianında Verilerin Toplamı ve Ambianın Özellikleri		29
Sekil 3.35 Zaman Serisi Grafik		35
Sekil 3.36 Kuvvetli Sesi Gösterim		36
Sekil 3.37 Veri ve Çıktı Grafik Çizim		37
Sekil 3.38 Kuvvetli Sesi Veri Gösterim Çizim		38
Sekil 3.39 Kuvvetli Sesi Veri Gösterim Çizim ve Farklı Gösterim		39
Sekil 3.40 Kuvvetli Sesi Veri Gösterim Çizim ve Farklı Gösterim		39
Sekil 3.41 Kuvvetli Sesi Veri Gösterim Çizim ve Farklı Gösterim		40
Sekil 3.42 Kuvvetli Sesi Veri Gösterim Çizim ve Farklı Gösterim		41
Sekil 3.43 Kuvvetli Sesi Veri Gösterim Çizim ve Farklı Gösterim		42
Sekil 3.44 Kuvvetli Sesi Veri Gösterim Çizim ve Farklı Gösterim		43
Sekil 3.45 Kuvvetli Sesi Veri Gösterim Çizim ve Farklı Gösterim		44
Sekil 3.46 Kuvvetli Sesi Veri Gösterim Çizim ve Farklı Gösterim		45
Sekil 3.47 Kuvvetli Sesi Veri Gösterim Çizim ve Farklı Gösterim		46
Sekil 3.48 Kuvvetli Sesi Veri Gösterim Çizim ve Farklı Gösterim		47
Sekil 3.49 Kuvvetli Sesi Veri Gösterim Çizim ve Farklı Gösterim		48
Sekil 3.50 Kuvvetli Sesi Veri Gösterim Çizim ve Farklı Gösterim		49
Sekil 3.51 Kuvvetli Sesi Veri Gösterim Çizim ve Farklı Gösterim		50
Sekil 3.52 Kuvvetli Sesi Veri Gösterim Çizim ve Farklı Gösterim		51
Sekil 3.53 Kuvvetli Sesi Veri Gösterim Çizim ve Farklı Gösterim		52
Sekil 3.54 Kuvvetli Sesi Veri Gösterim Çizim ve Farklı Gösterim		53
Sekil 3.55 Kuvvetli Sesi Veri Gösterim Çizim ve Farklı Gösterim		54
Sekil 3.56 Kuvvetli Sesi Veri Gösterim Çizim ve Farklı Gösterim		55
Sekil 3.57 Kuvvetli Sesi Veri Gösterim Çizim ve Farklı Gösterim		56
Sekil 3.58 Kuvvetli Sesi Veri Gösterim Çizim ve Farklı Gösterim		57
Sekil 3.59 Kuvvetli Sesi Veri Gösterim Çizim ve Farklı Gösterim		58
Sekil 3.60 Kuvvetli Sesi Veri Gösterim Çizim ve Farklı Gösterim		59
Sekil 3.61 Kuvvetli Sesi Veri Gösterim Çizim ve Farklı Gösterim		60
Sekil 3.62 Kuvvetli Sesi Veri Gösterim Çizim ve Farklı Gösterim		61
Sekil 3.63 Kuvvetli Sesi Veri Gösterim Çizim ve Farklı Gösterim		62
Sekil 3.64 Kuvvetli Sesi Veri Gösterim Çizim ve Farklı Gösterim		63
Sekil 3.65 Kuvvetli Sesi Veri Gösterim Çizim ve Farklı Gösterim		64
Sekil 3.66 Kuvvetli Sesi Veri Gösterim Çizim ve Farklı Gösterim		65
Sekil 3.67 Kuvvetli Sesi Veri Gösterim Çizim ve Farklı Gösterim		66
Sekil 3.68 Kuvvetli Sesi Veri Gösterim Çizim ve Farklı Gösterim		67
Sekil 3.69 Kuvvetli Sesi Veri Gösterim Çizim ve Farklı Gösterim		68
Sekil 3.70 Kuvvetli Sesi Veri Gösterim Çizim ve Farklı Gösterim		69
Sekil 3.71 Kuvvetli Sesi Veri Gösterim Çizim ve Farklı Gösterim		70
Sekil 3.72 Kuvvetli Sesi Veri Gösterim Çizim ve Farklı Gösterim		71
Sekil 3.73 Kuvvetli Sesi Veri Gösterim Çizim ve Farklı Gösterim		72
Sekil 3.74 Kuvvetli Sesi Veri Gösterim Çizim ve Farklı Gösterim		73

ŞEKİL LİSTESİ

Şekil 2.1. Veritabanı Teknolojisinin Gelişimi ve Veri Madenciliği (Han vd,2001)	9
Şekil 3.1. Veri Madenciliğinin Yararlandığı Diğer Disiplinler	17
Şekil 3.2 Örnek veritabanı	19
Şekil 3.3 Örnek Tablo	20
Şekil 3.4 Örnek Form	20
Şekil 3.5 Örnek Sorgu	21
Şekil 3.6. Rapor Örneği	22
Şekil 3.7. Sayfa Örneği	22
Şekil 3.8. İlişkisel Veritabanında Bulunan Tablolar Arasındaki İlişkiler	23
Şekil 3.9. İki Farklı Tablonun Birleştirilmesi İle Oluşturulmuş Sorgunun Sonucu	24
Şekil 3.10 Veri Ambarının Kuruluş Süreci	25
Şekil 3.11 OLAP Server'da OLAP Kúpünde FoodMart Veri Ambarında Analize Dahil Edilen Nesneler	27
Şekil 3.12 OLAP Server'da Kullanılan FoodMart Veri Ambarına İlişkin Nesneler	27
Şekil 3.13 DTS İle Veri Alınabilecek Bazı Veritabanları	28
Şekil 3.14 Veri Ambarında Verilerin Toplanması ve Analizini Gösteren Diyagram	29
Şekil 3.15 Zaman Serisi Grafiği	35
Şekil 3.16 Konum Serilerinin Gösterimi	36
Şekil 3.17 Daire ve Çubuk Grafik Örneği	37
Şekil 3.18 Body Adlı SAS Veri Setindeki Cheek Histogram Çizimi	38
Şekil 3.19 Body Adlı SAS Veri Setindeki Cheek ve Face Değişkenlerinin Serpilme Diyagramı	39
Şekil 3.20 Body Adlı SAS Veri Setindeki Cheek Değişkeninin Kutu Diyagramı	39
Şekil 3.21 Sales Veri Kúpü Bileşenleri	46
Şekil 3.22 Sales Kúpüne İlişkin Boyutlar ve Boyutların Düzeyleri	47
Şekil 3.24 Bir Boyut ve Hiyerarşik Yapısı	48
Şekil 3.25 Bir Küboid Örneği	48
Şekil 3.26 Sales Kúpüne Ait Ölçüler	49
Şekil 3.27 Star Şeması	52
Şekil 3.28 Snowflakes Şeması	53
Şekil 3.29 Galaxy Şeması	54
Şekil 3.30 Bir Olap Kúpü Tablosu Örneği	55
Şekil 3.31 Boyutların Düzey Ayrıntılarına Göre Bir Olap Kúpü Tablosu Örneği	56
Şekil 3.32 Roll-up İşlemi Öncesi	57
Şekil 3.34 Drill-down İşlemi Öncesi	57
Şekil 3.35 Drill-down İşlemi Sonrası	58
Şekil 3.36 Slice İşlemi Öncesi	58
Şekil 3.37 Slice İşlemi Sonrası	59
Şekil 3.38 Dice İşlemi Öncesi	59
Şekil 3.39 Dice İşlemi Sonrası	59
Şekil 3.40 Pivot İşlemi Öncesi	60
Şekil 3.41 Pivot İşlemi Sonrası	60
Şekil 3.42 3 Katmanlı Veri Ambarı Mimarisi (Han ve Kamber,2001)	61
Şekil 3.43 Denetimli Öğrenme (Elmas,2003)	65
Şekil 3.44 Denetimsiz Öğrenme	65
Şekil 3.45 SAS Enterprise Miner Ürünün Veri Madenciliği Araçlarından Bazıları	68
Şekil 4.1 Veri Madenciliği Çalışmalarında Harcanan Çabanın Dağılımı (Cabena vd,1998)	71
Şekil 4.2 Veri Madenciliği Süreci	72
Şekil 4.3 Kayıp Değerler Bulunan Bir Veri Örneği	74

Şekil 4.4 Veri Madenciliği Teknikleri.....	78
Şekil 4.5 SEMMA Metodolojisi.....	79
Şekil 5.1 Sabit Yoğunluk Elipsi $x'\Sigma^{-1}x = c^2$ $\mu = 0$ ve $\rho = 0.75$	86
Şekil 5.2 Standardize Edilmiş Veri Matrisi	88
Şekil 5.3 İlk İki Temel Bileşenin Serpilme Diyagramı	90
Şekil 5.4 Yamaç Grafiği	92
Şekil 5.5 Veri Madenciliğinde Kullanılan Kümeleme Metotlarının Sınıflandırılması (Han ve Kamber 2001)	106
Şekil 5.6 Dendogram	107
Şekil 6.1 Farklı Veri Kaynağında Bulunan Veri Setinin SAS Programına Aktarılması İşlemi-1	128
Şekil 6.2 Farklı Veri Kaynağında Bulunan Veri Setinin SAS Programına Aktarılması İşlemi-2	128
Şekil 6.3 Farklı Veri Kaynağında Bulunan Veri Setinin SAS Programına Aktarılması İşlemi-3	129
Şekil 6.4 Farklı Veri Kaynağında Bulunan Veri Setinin SAS Programına Aktarılması İşlemi-4	129
Şekil 6.5 Enterprise Miner Modülünün Çalıştırılması.....	130
Şekil 6.6 Uygulama Akış Diyagramı.....	131
Şekil 6.7 Input Data Source Düğümünün Data Sekmesi.....	132
Şekil 6.8 Input Data Source Düğümünün Data Sekmesi.....	133
Şekil 6.9 Input Data Source Düğümünün Interval Variables Sekmesi.....	136
Şekil 6.10 Input Data Source Düğümünün Class Variables Sekmesi	138
Şekil 6.11 Input Data Source Düğümünün Notes Sekmesi	139
Şekil 6.12 Variable Selection Düğümünün Data Sekmesi	140
Şekil 6.13 Variable Selection Düğümünün Variables Sekmesi	141
Şekil 6.14 Variable Selection Düğümünün Manual Selection Sekmesi.....	143
Şekil 6.15 Variable Selection Düğümünün Manual Selection Sekmesi.....	144
Şekil 6.16 Variable Selection Düğümünün Output Sekmesi.....	144
Şekil 6.17 Transform Variables Düğümünün Data Sekmesi.....	145
Şekil 6.18 Transform Variables Düğümünün Variables Sekmesi.....	146
Şekil 6.19 Transform Variables Düğümünün Output Sekmesi	149
Şekil 6.20 Filter Outliers Düğümünün Data Sekmesi	150
Şekil 6.21 Filter Outliers Düğümünün Output Sekmesi.....	150
Şekil 6.22 Filter Outliers Düğümünün Settings Sekmesi	151
Şekil 6.23 Filter Outliers Düğümünün Interval Vars Sekmesi	152
Şekil 6.24 Filter Outliers Düğümünün Output Sekmesi.....	155
Şekil 6.25 Princomp/Dmneural Düğümünün Data Sekmesi	156
Şekil 6.26 Princomp/Dmneural Düğümünün Variables Sekmesi	157
Şekil 6.27 Princomp/Dmneural Düğümünün General Sekmesi	159
Şekil 6.28 Princomp/Dmneural Düğümünün Princomp Sekmesi	160
Şekil 6.29 Princomp/Dmneural Düğümünün Output Sekmesi.....	160
Şekil 6.30 Özdeğerlerin Yamaç (Scree) Grafiği.....	162
Şekil 6.31 Özdeğerlerin Varyans Açıklama Oranları Grafiği	162
Şekil 6.32 Clustering Düğümünün Data Sekmesi	166
Şekil 6.33 Clustering Düğümünün Variables Sekmesi	167
Şekil 6.34 Clustering Düğümünün Clusters Sekmesi.....	170
Şekil 6.35 Clustering Düğümünün Seeds Sekmesi	170
Şekil 6.36 Clustering Düğümünün Missing Values Sekmesi.....	171
Şekil 6.37 Clustering Düğümünün Output Sekmesi.....	171
Şekil 6.38 Kümelere İlişkin Daire Grafiği	172

Şekil 6.39 Kümelerin Birbirlerine Göre Konumları	174
Şekil 6.40 Kümelerde Bulunan Liselerin Ortalama Netlerin Toplam Değeri Bakımından Dağılımları	186
Şekil 6.41 Ortalama Netlerin Toplamına Göre Birinci Kümedeki Liselerin D42 Değişkeni Bazında Dağılımı	187
Şekil 6.42 Ortalama Netlerin Toplamına Göre İkinci Kümedeki Liselerin D42 Değişkeni Bazında Dağılımı	188
Şekil 6.43 Ortalama Netlerin Toplamına Göre Üçüncü Kümedeki Liselerin D42 Değişkeni Bazında Dağılımı	189
Şekil 6.44 Ortalama Netlerin Toplamına Göre Dördüncü Kümedeki Liselerin D42 Değişkeni Bazında Dağılımı	190
Şekil 6.45 Ortalama Netlerin Toplamına Göre Beşinci Kümedeki Liselerin D42 Değişkeni Bazında Dağılımı	191
Şekil 6.46 Matematik Netlerine Göre Kümelerdeki Liselerin Dağılımı	193
Şekil 6.47 Fizik Netlerine Göre Kümelerdeki Liselerin Dağılımı	194
Şekil 6.48 Kimya Netlerine Göre Kümelerdeki Liselerin Dağılımı	194
Şekil 6.49 Biyoloji Netlerine Göre Kümelerdeki Liselerin Dağılımı	195
Şekil 6.50 Türkçe Netlerine Göre Kümelerdeki Liselerin Dağılımı	195
Şekil 6.51 Tarih Netlerine Göre Kümelerdeki Liselerin Dağılımı	196
Şekil 6.52 Coğrafya Netlerine Göre Kümelerdeki Liselerin Dağılımı	196
Şekil 6.53 Felsefe Netlerine Göre Kümelerdeki Liselerin Dağılımı	197
Şekil Ek 1.1 D3 Değişkenine İlişkin Histogram	204
Şekil Ek 1.2 D4 Değişkenine İlişkin Histogram	204
Şekil Ek 1.3 D5 Değişkenine İlişkin Histogram	204
Şekil Ek 1.4 D6 Değişkenine İlişkin Histogram	205
Şekil Ek 1.5 D7 Değişkenine İlişkin Histogram	205
Şekil Ek 1.6 D8 Değişkenine İlişkin Histogram	205
Şekil Ek 1.7 D9 Değişkenine İlişkin Histogram	206
Şekil Ek 1.8 D10 Değişkenine İlişkin Histogram	206
Şekil Ek 1.9 D11 Değişkenine İlişkin Histogram	206
Şekil Ek 1.10 D12 Değişkenine İlişkin Histogram	207
Şekil Ek 1.11 D13 Değişkenine İlişkin Histogram	207
Şekil Ek 1.12 D14 Değişkenine İlişkin Histogram	207
Şekil Ek 1.13 D15 Değişkenine İlişkin Histogram	208
Şekil Ek 1.14 D16 Değişkenine İlişkin Histogram	208
Şekil Ek 1.15 D17 Değişkenine İlişkin Histogram	208
Şekil Ek 1.16 D18 Değişkenine İlişkin Histogram	209
Şekil Ek 1.17 D19 Değişkenine İlişkin Histogram	209
Şekil Ek 1.18 D20 Değişkenine İlişkin Histogram	209
Şekil Ek 1.19 D21 Değişkenine İlişkin Histogram	210
Şekil Ek 1.20 D22 Değişkenine İlişkin Histogram	210
Şekil Ek 1.21 D23 Değişkenine İlişkin Histogram	210
Şekil Ek 1.22 D24 Değişkenine İlişkin Histogram	211
Şekil Ek 1.23 D25 Değişkenine İlişkin Histogram	211
Şekil Ek 1.24 D26 Değişkenine İlişkin Histogram	211
Şekil Ek 1.25 D27 Değişkenine İlişkin Histogram	212
Şekil Ek 1.26 D28 Değişkenine İlişkin Histogram	212
Şekil Ek 1.27 D29 Değişkenine İlişkin Histogram	212
Şekil Ek 1.28 D30 Değişkenine İlişkin Histogram	213
Şekil Ek 1.29 D31 Değişkenine İlişkin Histogram	213
Şekil Ek 1.30 D32 Değişkenine İlişkin Histogram	213

Şekil 6.39 Kümelerin Birbirlerine Göre Konumları	174
Şekil 6.40 Kümelerde Bulunan Liselerin Ortalama Netlerin Toplam Değeri Bakımından Dağılımları	186
Şekil 6.41 Ortalama Netlerin Toplamına Göre Birinci Kümedeki Liselerin D42 Değişkeni Bazında Dağılımı	187
Şekil 6.42 Ortalama Netlerin Toplamına Göre İkinci Kümedeki Liselerin D42 Değişkeni Bazında Dağılımı	188
Şekil 6.43 Ortalama Netlerin Toplamına Göre Üçüncü Kümedeki Liselerin D42 Değişkeni Bazında Dağılımı	189
Şekil 6.44 Ortalama Netlerin Toplamına Göre Dördüncü Kümedeki Liselerin D42 Değişkeni Bazında Dağılımı	190
Şekil 6.45 Ortalama Netlerin Toplamına Göre Beşinci Kümedeki Liselerin D42 Değişkeni Bazında Dağılımı	191
Şekil 6.46 Matematik Netlerine Göre Kümelerdeki Liselerin Dağılımı	193
Şekil 6.47 Fizik Netlerine Göre Kümelerdeki Liselerin Dağılımı	194
Şekil 6.48 Kimya Netlerine Göre Kümelerdeki Liselerin Dağılımı	194
Şekil 6.49 Biyoloji Netlerine Göre Kümelerdeki Liselerin Dağılımı	195
Şekil 6.50 Türkçe Netlerine Göre Kümelerdeki Liselerin Dağılımı	195
Şekil 6.51 Tarih Netlerine Göre Kümelerdeki Liselerin Dağılımı	196
Şekil 6.52 Coğrafya Netlerine Göre Kümelerdeki Liselerin Dağılımı	196
Şekil 6.53 Felsefe Netlerine Göre Kümelerdeki Liselerin Dağılımı	197
Şekil Ek 1.1 D3 Değişkenine İlişkin Histogram	204
Şekil Ek 1.2 D4 Değişkenine İlişkin Histogram	204
Şekil Ek 1.3 D5 Değişkenine İlişkin Histogram	204
Şekil Ek 1.4 D6 Değişkenine İlişkin Histogram	205
Şekil Ek 1.5 D7 Değişkenine İlişkin Histogram	205
Şekil Ek 1.6 D8 Değişkenine İlişkin Histogram	205
Şekil Ek 1.7 D9 Değişkenine İlişkin Histogram	206
Şekil Ek 1.8 D10 Değişkenine İlişkin Histogram	206
Şekil Ek 1.9 D11 Değişkenine İlişkin Histogram	206
Şekil Ek 1.10 D12 Değişkenine İlişkin Histogram	207
Şekil Ek 1.11 D13 Değişkenine İlişkin Histogram	207
Şekil Ek 1.12 D14 Değişkenine İlişkin Histogram	207
Şekil Ek 1.13 D15 Değişkenine İlişkin Histogram	208
Şekil Ek 1.14 D16 Değişkenine İlişkin Histogram	208
Şekil Ek 1.15 D17 Değişkenine İlişkin Histogram	208
Şekil Ek 1.16 D18 Değişkenine İlişkin Histogram	209
Şekil Ek 1.17 D19 Değişkenine İlişkin Histogram	209
Şekil Ek 1.18 D20 Değişkenine İlişkin Histogram	209
Şekil Ek 1.19 D21 Değişkenine İlişkin Histogram	210
Şekil Ek 1.20 D22 Değişkenine İlişkin Histogram	210
Şekil Ek 1.21 D23 Değişkenine İlişkin Histogram	210
Şekil Ek 1.22 D24 Değişkenine İlişkin Histogram	211
Şekil Ek 1.23 D25 Değişkenine İlişkin Histogram	211
Şekil Ek 1.24 D26 Değişkenine İlişkin Histogram	211
Şekil Ek 1.25 D27 Değişkenine İlişkin Histogram	212
Şekil Ek 1.26 D28 Değişkenine İlişkin Histogram	212
Şekil Ek 1.27 D29 Değişkenine İlişkin Histogram	212
Şekil Ek 1.28 D30 Değişkenine İlişkin Histogram	213
Şekil Ek 1.29 D31 Değişkenine İlişkin Histogram	213
Şekil Ek 1.30 D32 Değişkenine İlişkin Histogram	213

Şekil Ek 1.31 D33 Değişkenine İlişkin Histogram.....	214
Şekil Ek 1.32 D34 Değişkenine İlişkin Histogram.....	214
Şekil Ek 1.33 D35 Değişkenine İlişkin Histogram.....	214
Şekil Ek 1.34 D36 Değişkenine İlişkin Histogram.....	215
Şekil Ek 1.35 D37 Değişkenine İlişkin Histogram.....	215
Şekil Ek 1.36 D38 Değişkenine İlişkin Histogram.....	215
Şekil Ek 2.1 D3 Değişkenine İlişkin Kutu Diyagramı	216
Şekil Ek 2.2 D4 Değişkenine İlişkin Kutu Diyagramı	216
Şekil Ek 2.3 D5 Değişkenine İlişkin Kutu Diyagramı	216
Şekil Ek 2.4 D6 Değişkenine İlişkin Kutu Diyagramı	217
Şekil Ek 2.5 D7 Değişkenine İlişkin Kutu Diyagramı	217
Şekil Ek 2.6 D8 Değişkenine İlişkin Kutu Diyagramı	217
Şekil Ek 2.7 D9 Değişkenine İlişkin Kutu Diyagramı	218
Şekil Ek 2.8 D10 Değişkenine İlişkin Kutu Diyagramı	218
Şekil Ek 2.9 D11 Değişkenine İlişkin Kutu Diyagramı	218
Şekil Ek 2.10 D12 Değişkenine İlişkin Kutu Diyagramı	219
Şekil Ek 2.11 D13 Değişkenine İlişkin Kutu Diyagramı	219
Şekil Ek 2.12 D14 Değişkenine İlişkin Kutu Diyagramı	219
Şekil Ek 2.13 D15 Değişkenine İlişkin Kutu Diyagramı	220
Şekil Ek 2.14 D16 Değişkenine İlişkin Kutu Diyagramı	220
Şekil Ek 2.15 D17 Değişkenine İlişkin Kutu Diyagramı	220
Şekil Ek 2.16 D18 Değişkenine İlişkin Kutu Diyagramı	221
Şekil Ek 2.17 D19 Değişkenine İlişkin Kutu Diyagramı	221
Şekil Ek 2.18 D20 Değişkenine İlişkin Kutu Diyagramı	221
Şekil Ek 2.19 D21 Değişkenine İlişkin Kutu Diyagramı	222
Şekil Ek 2.20 D22 Değişkenine İlişkin Kutu Diyagramı	222
Şekil Ek 2.21 D23 Değişkenine İlişkin Kutu Diyagramı	222
Şekil Ek 2.22 D24 Değişkenine İlişkin Kutu Diyagramı	223
Şekil Ek 2.23 D25 Değişkenine İlişkin Kutu Diyagramı	223
Şekil Ek 2.24 D26 Değişkenine İlişkin Kutu Diyagramı	223
Şekil Ek 2.25 D27 Değişkenine İlişkin Kutu Diyagramı	224
Şekil Ek 2.26 D28 Değişkenine İlişkin Kutu Diyagramı	224
Şekil Ek 2.27 D29 Değişkenine İlişkin Kutu Diyagramı	224
Şekil Ek 2.28 D30 Değişkenine İlişkin Kutu Diyagramı	225
Şekil Ek 2.29 D31 Değişkenine İlişkin Kutu Diyagramı	225
Şekil Ek 2.30 D32 Değişkenine İlişkin Kutu Diyagramı	225
Şekil Ek 2.31 D33 Değişkenine İlişkin Kutu Diyagramı	226
Şekil Ek 2.32 D34 Değişkenine İlişkin Kutu Diyagramı	226
Şekil Ek 2.33 D35 Değişkenine İlişkin Kutu Diyagramı	226
Şekil Ek 2.34 D36 Değişkenine İlişkin Kutu Diyagramı	227
Şekil Ek 2.35 D37 Değişkenine İlişkin Kutu Diyagramı	227
Şekil Ek 2.36 D38 Değişkenine İlişkin Kutu Diyagramı	227

ÇİZELGE LİSTESİ

Çizelge 2.1	Veri Madenciliğinin evrimi (Aldana,2000).....	11
Çizelge 2.2	Amerika Birleşik Devletleri'nde Kurumsal ve Kamusal Alana Ait Veri Madenciliği Araçlarının Sayısı (Aldana,2000).....	12
Çizelge 2.2	Devam Amerika Birleşik Devletleri'nde Kurumsal ve Kamusal Alana Ait Veri Madenciliği Araçlarının Sayısı (Aldana,2000).....	13
Çizelge 3.1	İşlemsel Veritabanı Tablo Örneği	30
Çizelge 3.2	Geleneksel Veritabanlarında Tablo Örneği	30
Çizelge 3.3	Aylık İmalat Sanayi Kapasite Kullanım Oranı (DİE,1999)	35
Çizelge 3.4	1991 Yılı Seçilmiş Bazı Dayanıklı Tüketim Mallarının Üretim Sayıları(DİE,1999)37	
Çizelge 3.5	Birinci Normal Biçim.....	41
Çizelge 3.6	İkinci Normal Biçim.....	42
Çizelge 3.7	Üçüncü Normal Biçim.....	42
Çizelge 3.9	İstatistik ve Veri Madenciliğinin Karşılaştırılması 2 (Ganesh,2002).....	70
Çizelge 4.1	Uyumsuz Veri Örneği	76
Çizelge 5.1	İkili Değişkenlerden Oluşan Örnek Kümeleme Veri Seti	102
Çizelge 5.2	İkili Değişkenler İçin Kontenjans Tablosu.....	103
Çizelge 5.3	Benzerlik Ölçüleri Formülleri (Johson ve Wichern,1982).....	104
Çizelge 5.4	Y.T.Ü. Fen Edebiyat Fakültesi Bölümlerine 1998-1999-2000 Yıllarında Yerleştirilmiş Olan Öğrencilerin Maksimum ve Minimum Puanları	109
Çizelge 5.5	Y.T.Ü. Fen Edebiyat Fakültesi Bölümlerinin Arasındaki Öklidyen Uzaklık Matrisi.....	109
Çizelge 6.1	Uygulamada Kullanılacak Olan Değişkenler ve Kodları	125
Çizelge 6.1	Devam Uygulamada Kullanılacak Olan Değişkenler ve Kodları	126
Çizelge 6.1	Devam Uygulamada Kullanılacak Olan Değişkenler ve Kodları	127
Çizelge 6.2	Veri Setinde Bulunan Değişkenler ve Değişkenlere İlişkin Bazı Bilgiler	133
Çizelge 6.2	Devam Veri Setinde Bulunan Değişkenler ve Değişkenlere İlişkin Bazı Bilgiler134	
Çizelge 6.2	Devam Veri Setinde Bulunan Değişkenler ve Değişkenlere İlişkin Bazı Bilgiler135	
Çizelge 6.3	Aralık ve Oransal Ölçekli Değişkenlere İlişkin Bazı İstatistikler	136
Çizelge 6.3	Devam Aralık ve Oransal Ölçekli Değişkenlere İlişkin Bazı İstatistikler.....	137
Çizelge 6.3	Devam Aralık ve Oransal Ölçekli Değişkenlere İlişkin Bazı İstatistikler.....	138
Çizelge 6.4	Temel Bileşenler Analizinde Kullanılmaya Uygun Değişkenler	141
Çizelge 6.4	Devam Temel Bileşenler Analizinde Kullanılmaya Uygun Değişkenler	142
Çizelge 6.4	Devam Temel Bileşenler Analizinde Kullanılmaya Uygun Değişkenler	143
Çizelge 6.5	Transformasyon Tablosu.....	146
Çizelge 6.5	Devam Transformasyon Tablosu	147
Çizelge 6.5	Devam Transformasyon Tablosu	148
Çizelge 6.5	Devam Transformasyon Tablosu	149
Çizelge 6.6	Analizde Kullanılacak Değişkenlere İlişkin Filtreleme Sınırları	152
Çizelge 6.6	Devam Analizde Kullanılacak Değişkenlere İlişkin Filtreleme Sınırları.....	153
Çizelge 6.6	Devam Analizde Kullanılacak Değişkenlere İlişkin Filtreleme Sınırları.....	154
Çizelge 6.7	Temel Bileşenler Analizinde Kullanılacak Değişkenler	157
Çizelge 6.7	Devam Temel Bileşenler Analizinde Kullanılacak Değişkenler.....	158
Çizelge 6.7	Devam Temel Bileşenler Analizinde Kullanılacak Değişkenler.....	159
Çizelge 6.8	Temel Bileşenlere İlişkin Özdeğerler Tablosu.....	163
Çizelge 6.9	Özdeğerlere İlişkin Özvektörler 1	164
Çizelge 6.10	Özdeğerlere İlişkin Özvektörler 2	164
Çizelge 6.10	Devam Özdeğerlere İlişkin Özvektörler 2.....	165
Çizelge 6.11	Kümeleme Analizinin Girdi Değişkenleri.....	167

Çizelge 6.11 Devam Kümeleme Analizinin Girdi Değişkenleri	168
Çizelge 6.11 Devam Kümeleme Analizinin Girdi Değişkenleri	169
Çizelge 6.12 Kümelere İlişkin Bazı İstatistikler.....	173
Çizelge 6.13 Küme Uzaklıkları Matrisi.....	174
Çizelge 6.14 Kümeleme Analizinde Kullanılan Değişkenlerin Önemlilik Katsayıları.....	176
Çizelge 6.15 İbrahim Turhan Lisesi Verileri İçin Temel Bileşen Skorlarının Hesaplanması	177
Çizelge 6.16 Kümeleme Kuralları	178
Çizelge 6.16 Devam Kümeleme Kuralları	179
Çizelge 6.16 Devam Kümeleme Kuralları	180
Çizelge 6.17 Kümelerin Testlerdeki Başarı Ortalamaları	182
Çizelge 6.18 Kümelerin Yüksek Öğretim Kurumlarına Öğrenci Yerleştirme Oranları.....	184
Çizelge Ek 3.1 D3 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı	228
Çizelge Ek 3.2 D4 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı	229
Çizelge Ek 3.3 D5 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı	230
Çizelge Ek 3.4 D6 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı	231
Çizelge Ek 3.5 D7 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı	232
Çizelge Ek 3.6 D8 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı	233
Çizelge Ek 3.7 D9 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı	234
Çizelge Ek 3.8 D10 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı	235
Çizelge Ek 3.9 D11 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı	236
Çizelge Ek 3.10 D12 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı	237
Çizelge Ek 3.11 D13 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı	238
Çizelge Ek 3.12 D14 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı	239
Çizelge Ek 3.13 D15 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı	240
Çizelge Ek 3.14 D16 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı	241
Çizelge Ek 3.15 D17 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı	242
Çizelge Ek 3.16 D18 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı	243
Çizelge Ek 3.17 D19 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı	244
Çizelge Ek 3.18 D20 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı	245
Çizelge Ek 3.19 D21 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı	246
Çizelge Ek 3.20 D22 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı	247
Çizelge Ek 3.21 D23 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı	248
Çizelge Ek 3.22 D24 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı	249
Çizelge Ek 3.23 D25 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı	250
Çizelge Ek 3.24 D26 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı	251
Çizelge Ek 3.25 D27 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı	252
Çizelge Ek 3.26 D28 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı	253
Çizelge Ek 3.27 D29 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı	254
Çizelge Ek 3.28 D30 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı	255
Çizelge Ek 3.29 D31 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı	256
Çizelge Ek 3.30 D32 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı	257
Çizelge Ek 3.31 D33 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı	258
Çizelge Ek 3.32 D34 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı	259
Çizelge Ek 3.33 D35 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı	260
Çizelge Ek 3.34 D36 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı	261
Çizelge Ek 3.35 D37 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı	262
Çizelge Ek 3.36 D38 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı	263
Çizelge Ek 4.1 D3 Değişkenine İlişkin İstatistikler	264
Çizelge Ek 4.2 D4 Değişkenine İlişkin İstatistikler	264
Çizelge Ek 4.3 D5 Değişkenine İlişkin İstatistikler	265
Çizelge Ek 4.4 D6 Değişkenine İlişkin İstatistikler	265

Çizelge Ek 4.5 D7 Değişkenine İlişkin İstatistikler	265
Çizelge Ek 4.6 D8 Değişkenine İlişkin İstatistikler	266
Çizelge Ek 4.7 D9 Değişkenine İlişkin İstatistikler	266
Çizelge Ek 4.8 D10 Değişkenine İlişkin İstatistikler	266
Çizelge Ek 4.9 D11 Değişkenine İlişkin İstatistikler	267
Çizelge Ek 4.10 D12 Değişkenine İlişkin İstatistikler	267
Çizelge Ek 4.11 D13 Değişkenine İlişkin İstatistikler	267
Çizelge Ek 4.12 D14 Değişkenine İlişkin İstatistikler	268
Çizelge Ek 4.13 D15 Değişkenine İlişkin İstatistikler	268
Çizelge Ek 4.14 D16 Değişkenine İlişkin İstatistikler	268
Çizelge Ek 4.15 D17 Değişkenine İlişkin İstatistikler	269
Çizelge Ek 4.16 D18 Değişkenine İlişkin İstatistikler	269
Çizelge Ek 4.17 D19 Değişkenine İlişkin İstatistikler	269
Çizelge Ek 4.18 D20 Değişkenine İlişkin İstatistikler	270
Çizelge Ek 4.19 D21 Değişkenine İlişkin İstatistikler	270
Çizelge Ek 4.20 D22 Değişkenine İlişkin İstatistikler	270
Çizelge Ek 4.21 D23 Değişkenine İlişkin İstatistikler	271
Çizelge Ek 4.22 D24 Değişkenine İlişkin İstatistikler	271
Çizelge Ek 4.23 D24 Değişkenine İlişkin İstatistikler	271
Çizelge Ek 4.24 D26 Değişkenine İlişkin İstatistikler	272
Çizelge Ek 4.25 D27 Değişkenine İlişkin İstatistikler	272
Çizelge Ek 4.26 D28 Değişkenine İlişkin İstatistikler	272
Çizelge Ek 4.27 D29 Değişkenine İlişkin İstatistikler	273
Çizelge Ek 4.28 D30 Değişkenine İlişkin İstatistikler	273
Çizelge Ek 4.29 D31 Değişkenine İlişkin İstatistikler	273
Çizelge Ek 4.30 D32 Değişkenine İlişkin İstatistikler	274
Çizelge Ek 4.31 D33 Değişkenine İlişkin İstatistikler	274
Çizelge Ek 4.32 D34 Değişkenine İlişkin İstatistikler	274
Çizelge Ek 4.33 D35 Değişkenine İlişkin İstatistikler	275
Çizelge Ek 4.34 D36 Değişkenine İlişkin İstatistikler	275
Çizelge Ek 4.35 D37 Değişkenine İlişkin İstatistikler	275
Çizelge Ek 4.36 D38 Değişkenine İlişkin İstatistikler	276
Çizelge Ek 5.1 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-A.Ö.Lisesi	277
Çizelge Ek 5.2 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Anadolu Denizcilik Meslek Lisesi.....	277
Çizelge Ek 5.3 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Anadolu Dış Ticaret Meslek Lisesi	277
Çizelge Ek 5.4 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Anadolu Güzel Sanatlar Lisesi	278
Çizelge Ek 5.5 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Anadolu Hazır Giyim ve Deri Hazır Giyim Lisesi.....	278
Çizelge Ek 5.6 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Anadolu İletişim Meslek Lisesi	278
Çizelge Ek 5.7 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Anadolu İmam Hatip Lisesi.....	279
Çizelge Ek 5.8 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Anadolu Kız Meslek Lisesi	279
Çizelge Ek 5.9 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Anadolu Lisesi	279
Çizelge Ek 5.10 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Anadolu Mahalli İdareler Meslek Lisesi	280

Çizelge Ek 5.11 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Anadolu Meslek Lisesi	280
Çizelge Ek 5.12 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Anadolu Moda Tasarımı Meslek Lisesi.....	280
Çizelge Ek 5.13 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Anadolu Otelcilik ve Turizm Meslek Lisesi.....	281
Çizelge Ek 5.14 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Anadolu Öğretmen Lisesi	281
Çizelge Ek 5.15 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Anadolu Sağlık Meslek Lisesi	281
Çizelge Ek 5.16 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Anadolu Sekreterlik Meslek Lisesi.....	282
Çizelge Ek 5.17 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Anadolu Teknik Lisesi.....	282
Çizelge Ek 5.18 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Anadolu Tekstil Meslek Lisesi	282
Çizelge Ek 5.19 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Anadolu Ticaret Meslek Lisesi.....	283
Çizelge Ek 5.20 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Anadolu Ticaret Meslek Lisesi.....	283
Çizelge Ek 5.21 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Çok Programlı Lise.....	283
Çizelge Ek 5.22 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Denizcilik ve Su Ürünleri Meslek Lisesi.....	284
Çizelge Ek 5.23 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Endüstri Meslek Lisesi	284
Çizelge Ek 5.24 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Fen Lisesi.....	284
Çizelge Ek 5.25 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Geleneksel Türk Sanatları Kız Meslek Lisesi	285
Çizelge Ek 5.26 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- İmam Hatip Lisesi.....	285
Çizelge Ek 5.27 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- İnşaat Teknik Lisesi	285
Çizelge Ek 5.28 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Kız Meslek Lisesi	286
Çizelge Ek 5.29 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Kuaförlük Meslek Lisesi.....	286
Çizelge Ek 5.30 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Matbaa Meslek Lisesi	286
Çizelge Ek 5.31 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Normal Lise	287
Çizelge Ek 5.32 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Ortopedik Cihaz ve Protez Teknisyen Okulu	287
Çizelge Ek 5.33 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Özel Akşam Lisesi	287
Çizelge Ek 5.34 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Özel Anadolu Otelcilik ve Turizm Meslek Lisesi	288

Çizelge Ek 5.35 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Özel Fen Lisesi	288
Çizelge Ek 5.36 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Özel Lise	288
Çizelge Ek 5.37 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Özel Ticaret Meslek Lisesi	289
Çizelge Ek 5.38 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Teknik Lise (Erkek)	289
Çizelge Ek 5.39 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Tekstil Meslek Lisesi	289
Çizelge Ek 5.40 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Tekstil Teknik Lisesi (Erkek)	290
Çizelge Ek 5.41 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Terzilik Meslek Lisesi	290
Çizelge Ek 5.42 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Terzilik Meslek Lisesi	290
Çizelge Ek 5.43 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Veteriner Sağlık Meslek Lisesi	291
Çizelge Ek 5.44 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Y.D.A. İmam Hatip Lisesi	291
Çizelge Ek 5.45 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Y.D.A. Lise	291
Çizelge Ek 5.46 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Yabancı Dille Öğretim Yapan Özel Lise	292
Çizelge Ek 5.47 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Yapı Meslek Lisesi	292
Çizelge Ek 5.48 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı- Ziraat Meslek Lisesi	292
Çizelge EK 6.1 Erkek Liselerinin Kümelere Göre Dağılımı	293
Çizelge EK 6.2 Karma Liselerin Kümelere Göre Dağılımı	293
Çizelge EK 6.3 Kız Liselerinin Kümelere Göre Dağılımı	293
Çizelge EK 7.1 Devlet Liselerinin Kümelere Göre Dağılımı	294
Çizelge EK 7.2 Özel Liselerin Kümelere Göre Dağılımı	294
Çizelge EK 8.1 Anadolu Liselerinin Kümelere Göre Dağılımı	295
Çizelge EK 8.2 Fen Liselerinin Kümelere Göre Dağılımı	295
Çizelge EK 8.3 Meslek Liselerinin Kümelere Göre Dağılımı	295
Çizelge EK 8.4 Meslek Liselerinin Kümelere Göre Dağılımı	296
Çizelge EK 8.5 Özel Liselerin Kümelere Göre Dağılımı	296
Çizelge EK 8.6 Özel Fen Liselerinin Kümelere Göre Dağılımı	296
Çizelge EK 8.7 Yabancı Dille Eğitim Yapan Özel Liselerin Kümelere Göre Dağılımı	297
Çizelge EK 8.8 Y.D.A. Liselerin Kümelere Göre Dağılımı	297
Çizelge EK 8.9 Y.D.A. Özel Liselerin Kümelere Göre Dağılımı	297
Çizelge EK 8.10 Anadolu Meslek Liselerinin Kümelere Göre Dağılımı	298
Çizelge EK 8.11 Akşam Liselerinin Kümelere Göre Dağılımı	298
Çizelge EK 8.12 Açık Öğretim Lisesinin Kümelere Göre Dağılımı	298
Çizelge EK 8.13 Özel Meslek Liselerinin Kümelere Göre Dağılımı	299
Çizelge EK 9.1 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Göre Matematik Net Ortalamaları	300
Çizelge EK 9.1 Devam ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Göre Matematik Net Ortalamaları	301

Çizelge EK 9.2 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Göre Fizik Net Ortalamaları	302
Çizelge EK 9.3 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Göre Kimya Net Ortalamaları	303
Çizelge EK 9.4 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Göre Biyoloji Net Ortalamaları	304
Çizelge EK 9.5 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Göre Türkçe Net Ortalamaları	305
Çizelge EK 9.6 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Göre Tarih Net Ortalamaları	306
Çizelge EK 9.7 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Göre Coğrafya Net Ortalamaları	307
Çizelge EK 9.8 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Göre Felsefe Net Ortalamaları	308
Çizelge EK 9.9 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Göre İngilizce Net Ortalamaları	309
Çizelge EK 9.10 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Göre Fransızca Net Ortalamaları	310
Çizelge EK 9.11 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Göre Almanca Net Ortalamaları	311
Çizelge EK 10.1 D40 Değişkenindeki Lise Türlerinin Kümelere Göre Matematik Net Ortalamaları	312
Çizelge EK 10.2 D40 Değişkenindeki Lise Türlerinin Kümelere Göre Fizik Net Ortalamaları	312
Çizelge EK 10.3 D40 Değişkenindeki Lise Türlerinin Kümelere Göre Kimya Net Ortalamaları	312
Çizelge EK 10.4 D40 Değişkenindeki Lise Türlerinin Kümelere Göre Biyoloji Net Ortalamaları	312
Çizelge EK 10.5 D40 Değişkenindeki Lise Türlerinin Kümelere Göre Türkçe Net Ortalamaları	312
Çizelge EK 10.6 D40 Değişkenindeki Lise Türlerinin Kümelere Göre Tarih Net Ortalamaları	313
Çizelge EK 10.7 D40 Değişkenindeki Lise Türlerinin Kümelere Göre Coğrafya Net Ortalamaları	313
Çizelge EK 10.8 D40 Değişkenindeki Lise Türlerinin Kümelere Göre Felsefe Net Ortalamaları	313
Çizelge EK 10.9 D40 Değişkenindeki Lise Türlerinin Kümelere Göre İngilizce Net Ortalamaları	313
Çizelge EK 10.10 D40 Değişkenindeki Lise Türlerinin Kümelere Göre Fransızca Net Ortalamaları	313
Çizelge EK 10.11 D40 Değişkenindeki Lise Türlerinin Kümelere Göre Almanca Net Ortalamaları	313
Çizelge EK 11.1 D40 Değişkenindeki Lise Türlerinin Kümelere Göre Matematik Net Ortalamaları	314
Çizelge EK 11.2 D40 Değişkenindeki Lise Türlerinin Kümelere Göre Fizik Net Ortalamaları	314
Çizelge EK 11.3 D40 Değişkenindeki Lise Türlerinin Kümelere Göre Kimya Net Ortalamaları	314

Çizelge EK 11.4 D40 Değişkenindeki Lise Türlerinin Kümelere Göre Biyoloji Net Ortalamaları	314
Çizelge EK 11.5 D40 Değişkenindeki Lise Türlerinin Kümelere Göre Türkçe Net Ortalamaları	314
Çizelge EK 11.6 D40 Değişkenindeki Lise Türlerinin Kümelere Göre Tarih Net Ortalamaları	315
Çizelge EK 11.7 D40 Değişkenindeki Lise Türlerinin Kümelere Göre Coğrafya Net Ortalamaları	315
Çizelge EK 11.8 D40 Değişkenindeki Lise Türlerinin Kümelere Göre Felsefe Net Ortalamaları	315
Çizelge EK 11.9 D40 Değişkenindeki Lise Türlerinin Kümelere Göre İngilizce Net Ortalamaları	315
Çizelge EK 11.10 D40 Değişkenindeki Lise Türlerinin Kümelere Göre Fransızca Net Ortalamaları	315
Çizelge EK 11.11 D40 Değişkenindeki Lise Türlerinin Kümelere Göre Almanca Net Ortalamaları	315
Çizelge EK 12.1 D42 Değişkenindeki Lise Türlerinin Kümelere Göre Matematik Net Ortalamaları	316
Çizelge EK 12.2 D42 Değişkenindeki Lise Türlerinin Kümelere Göre Fizik Net Ortalamaları	316
Çizelge EK 12.3 D42 Değişkenindeki Lise Türlerinin Kümelere Göre Kimya Net Ortalamaları	317
Çizelge EK 12.4 D42 Değişkenindeki Lise Türlerinin Kümelere Göre Biyoloji Net Ortalamaları	317
Çizelge EK 12.5 D42 Değişkenindeki Lise Türlerinin Kümelere Göre Türkçe Net Ortalamaları	318
Çizelge EK 12.6 D42 Değişkenindeki Lise Türlerinin Kümelere Göre Tarih Net Ortalamaları	318
Çizelge EK 12.7 D42 Değişkenindeki Lise Türlerinin Kümelere Göre Coğrafya Net Ortalamaları	319
Çizelge EK 12.8 D42 Değişkenindeki Lise Türlerinin Kümelere Göre Felsefe Net Ortalamaları	319
Çizelge EK 12.9 D42 Değişkenindeki Lise Türlerinin Kümelere Göre İngilizce Net Ortalamaları	320
Çizelge EK 12.10 D42 Değişkenindeki Lise Türlerinin Kümelere Göre Fransızca Net Ortalamaları	320
Çizelge EK 12.11 D42 Değişkenindeki Lise Türlerinin Kümelere Göre Almanca Net Ortalamaları	321

ÖNSÖZ

“ ÖSS Sınav Sonuçlarının Okul Bazında Veri Madenciliği İle İncelenmesi .” başlıklı yüksek lisans tezimin hazırlanmasında her zaman yanımda olan ve her türlü ruh halimde bana katlanan sevgili eşim Hatice’ye, anne ve babama, tezimin yazımı sırasında her türlü soruma sabırla cevap veren ve benden desteğini esirgemeyen hocam Sayın Doç. Dr. Ali Hakan Büyüklü’ye,

Tezimi satır satır okuyarak yapmış olduğum yazım hataları konusunda beni uyaran Sayın Öğretim Görevlisi Dr. Ertuğrul Efeoğlu’na,

Tezimi inceleyerek eleştirilerini bana aktaran Sayın Prof. Dr. Haldun Akpınar’a,

Çalışmamda uygulamada kullandığım SAS Enterprise Miner 8.0 ürününün kullanım izninin verilmesinde her türlü kolaylığı gösteren SAS Türkiye Ofis Müdürü Sayın Nuray Akmeriç’e ve programın kullanımını öğrenmem için yoğun çalışma ortamında bile bana zaman ayırarak yardımcı olan Sayın Fatih Köksal’a,

Tecrübelerinden faydalandığım Sayın Veysel Kamcı’ya ve yazıcısını kullanımına açarak bana destek olan dostum Dr. Serhat Çehreli’ye,

OLAP konusunda bana kaynak ve teknik destek veren istatistik bölümü öğrencisi Murat Can Dinçol’a,

Tezimin İngilizce özet kısmını yazan arkadaşım Arş.Görevlisi Ersel Topraktepe’ye,

Teşekkür etmeyi kendime bir borç bilirim.

Sağladığı yardımlardan dolayı değerli arkadaşım Pelin Biçen’e özel olarak teşekkürlerimi sunarım.

ÖZET

ÖSS SINAV SONUÇLARININ OKUL BAZINDA VERİ MADENCİLİĞİ İLE İNCELENMESİ

Ömer BİLEN

İstatistik, Yüksek Lisans Tezi

Veritabanı teknolojisindeki hızlı gelişmenin doğurduğu büyük ölçeklerdeki verilerin analiz edilerek yorumlanması ihtiyacına cevap veren veri madenciliği, verinin veritabanı ortamına aktarılmasından itibaren başlayarak analiz aşamasında elde edilen enformasyonun bilgiye dönüştürülmesi aşamasına kadar geçen uzun bir süreci kapsamaktadır.

Veri madenciliği veritabanı teknolojisi, karar destek sistemleri, görselleştirme, makine öğrenmesi ve istatistik gibi bir çok disiplinin bir arada kullanılmasına olanak sağlar. İstatistik disiplini veri madenciliği disiplinin ayrılmaz bir parçasıdır. Bu çalışmada kullanılan SAS Enterprise Miner 8.0 ürünü incelenirse SEMMA metodolojisinin her bir adımında bir istatistik tekniğe rastlamak mümkündür.

Bu çalışmada verilerin kaynağı olan veritabanlarına, veri saklamada ve raporlamada kullanılan OLAP teknolojisine ağırlıklı olarak yer verilmiş olup veri madenciliğinin diğer disiplinlerle olan ilişkisi incelenerek uygulamada kullanılan kümeleme ve temel bileşenler analizi teknikleri teorik olarak incelenmiştir.

Veri madenciliği metodolojisine yer verildikten sonra SAS Enterprise Miner 8.0 ürününün kullandığı SEMMA metodolojisi ele alınmıştır.

Uygulamada İstanbul ilindeki liseler 2001 ÖSS sınavındaki başarılarına göre kümelenmiş ve karar ağacı aracılığı ile oluşturulan kümeleme kuralları sayesinde, farklı dönemlerde liselerin 2001 yılındaki durumlarını koruyup korumadıkları incelenebilecektir.

Anahtar Kelimeler: Kümeleme analizi, temel bileşenler analizi, veri madenciliği, ÖSS, OLAP

Jüri:

1. Doç. Dr. Ali Hakan BÜYÜKLÜ (Danışman)
2. Prof. Dr. Haldun AKPINAR
3. Prof. Dr. Şehamet BÜLBÜL

Kabul tarihi:06.07.04

Sayfa Sayısı:322

ABSTRACT

THE EVALUATION of OSS SCHOOL BASED EXAMINATION RESULTS by DATA MINING

Ömer BİLEN

Statistics, M.S. Thesis

Data mining that answers the need of analysing and interpreting the data on a large scale coming out of the fast development in the database technology covers a long process including first of all the transfer of data to database and the translation of the gathered information into knowledge.

Data mining makes it possible to benefit from many disciplines like Database Technology, Decision Support Systems, Visualization, Machine Learning and Statistics. Statistics is integral part of data mining. If the product SAS Enterprise Miner 8.0 used in this study is examined, one can see easily a statistics technique at every step of SEMMA methodology.

In this study, databases that are source of data, have been considerably employed and OLAP Technology that is used in storage and reporting of data has been used as well. And we have examined the relation of data mining to other disciplines. And we have theoretically took a close look at techniques applied in analysis of clustering and principal components. After having presented data mining methodology, we have examined the methodology that SAS Enterprise Miner 8.0 uses.

In practice, colleges in İstanbul have been clustered according to their performances in the 2001 Entrance Examination for University (2001 ÖSS). And through the clustering rules that have been created by the decision tree, it has been possible to see other performances of those colleges in Entrance Examination for University at any year.

Keywords: Cluster, principal components, data mining, OSS, OLAP

Jury:

1. Assoc. Prof. Dr. Ali Hakan BÜYÜKLÜ (Supervisor)

Date:06.07.04

2. Prof. Dr. Haldun AKPINAR

Page:322

3. Prof. Dr. Şehamet BÜLBÜL

1. GİRİŞ

1960'lı yıllardan itibaren elektronik ortamda veri saklama işlemleri ile başlayan veritabanı teknolojisi gelişiminin 1970-1980'li yıllar arasında ivme kazanması ile birlikte ortaya çıkan büyük ölçeklerdeki veriden enformasyon elde etme ihtiyacının doğurduğu veri madenciliği hemen her alanda artan rekabet ortamında büyük önem kazandı.

Usama Fayyad, Gregory Piatetsky-Shapiro, Padhraic Smyth'e göre, veri madenciliği terimi sıklıkla istatistik, veri analist ve yönetim enformasyon sistem toplulukları tarafından kullanılmıştır(Fayyad vd.,1996). Adı geçen bilim adamları veri madenciliğini veritabanlarında bilgi keşfi sürecinin bir parçası olarak kabul etmektedir. Bu tez çalışması da, veri madenciliğinin veritabanı teknolojisi ile olan bağlantısından yola çıkılarak oluşturulmuştur. Fakat veri madenciliği terimi veri tabanlarında bilgi keşfi sürecine nazaran daha çok popüler olduğu için veritabanlarında bilgi keşfi süreci yerine veri madenciliği kavramı kullanılmıştır.

İkinci bölümde veri madenciliğine ilişkin değişik kaynaklarda yapılan tanımlardan yola çıkılarak bir veri madenciliği tanımı yapılmıştır. Veri madenciliğinin gelişimi gerek veritabanı teknolojisi gerekse veri madenciliğinin çözüm araçları ve çözümlerinin ilerlemesi göz önüne alınıp incelenerek veri madenciliğinin kullanım alanlarına değinilmiş ve başarılı birkaç veri madenciliği uygulamasına örnek verilmiştir.

Üçüncü bölümde veri madenciliğinin yararlandığı disiplinlerden veritabanı teknolojisi, karar destek sistemleri, görselleştirme, istatistik, makine öğrenmesi disiplinlerine değinilmiştir.

Dördüncü bölümde veri madenciliği metodolojisi konu alınmış ve SAS Enterprise Miner programının kullandığı SEMMA metodolojisine değinilmiştir.

Beşinci bölümde ise uygulamada kullanıcak olan istatistik tekniklerden temel bileşenler analizi ile birlikte kümeleme analizi teorik olarak anlatılmıştır. Kümeleme analizinde kullanılan algoritmalarından K-Means, K-Medoids ve hiyerarşik metotlardan single linkage, complete linkage ve average linkage algoritmalarına değinilerek her bir algoritma için örnek uygulama yapılmıştır.

Altıncı bölümde SAS Enterprise Miner 8.0 programında yapılan uygulama adım adım gösterilmiştir. Uygulama konusu Türkiye için büyük önem taşıyan ortaöğrenimden yüksek öğrenime geçişin bir adımı olan ÖSS sınavı olarak seçilmiştir. Uygulamada İstanbul ilinde bulunan liselerin yeni mezunlarının 2001 yılında göstermiş oldukları performansları

incelenmiştir. Bu liseler öğrencilerinin performanslarına göre kümelenmiş ve bir model oluşturulmuştur. Oluşturulan bu model ile liselerin farklı yıllarda bulundukları küme tahmin edilerek 2001 yılına göre ne durumda oldukları tahmin edilebilir. Ayrıca uygulama sonucunda ortaöğretim sistemindeki bazı aksaklıklarda ortaya çıkartılmıştır.

2. Veri Madenciliği

Bilgi teknolojisindeki gelişmeler sonucu verilerin toplanması, toplanan verilerin depolanması ve değişik birimlerde depolanan verilerin bir araya getirilmesi artık oldukça kolaylaşmıştır. Verilerin saklanması günümüzde oldukça kolay ve ucuz olması, büyük çaplı kuruluşları bu olanağı kullanmaya zorlamaktadır. Evlerinde bilgisayar kullanan kişilerin bile, gelir-gider durumlarını, ellerinde bulunan kitapların, cd'lerin vb. dökümlerini tutabilecekleri ve saklayabilecekleri veritabanı programları vardır. Başka bir deyişle, büyük çaplı kuruluşlardan evlerinde bilgisayar bulunan son kullanıcılara dek gerek tüzel gerekse özel tüm bilgisayar kullanıcıları birer veritabanı oluşturma gücüne sahiptir. Özellikle büyük kuruluşların çeşitli amaçlarla oluşturdukları büyük veritabanlarındaki veriler bilgiye dönüşmelidir ki yapılan veri toplama işi bu şirketler için kazanca dönüşebilsin. Unutulmamalıdır ki değerlendirilemeyen veri boşa giden emek ve sermayedir. Two Crows Corporation'un Data mining '99: Technology Report adlı yapıtında da söz edildiği gibi "bugünlerde veritabanlarının boyutları 1.000.000.000.000 byte'lara yani terabyte'lara ulaşmaktadır. (Two Crows Corporation,1999)". Bu büyüklükteki veritabanlarındaki verileri bilgiye dönüştürebilmek için istatistik, matematik, karar destek sistemleri, bilgisayar bilimi, yapay sinir ağları gibi çeşitli disiplinlerin çözüm amacıyla bir araya getirilerek veriyi enformasyona dönüştürme çalışmalarının yapılması işlemi **Veri Madenciliği**'dir.

Yukarıdaki tanıma ek olarak veri madenciliği için yapılmış farklı tanımlar aşağıda yer almaktadır.

"Veri Madenciliği, büyük veri tabanlarından **daha önceden bilinmeyen, doğru ve eyleme geçirilebilir** enformasyonu ayırtırma ve çok önemli kararların alınması aşamasında ayırtırılan bu enformasyonu kullanma sürecidir.(Cabena vd.,1998)". Tanımı incelediğimizde ayırtırılan enformasyona ilişkin önemli 3 özellik öne çıkmaktadır. Ayırtırılan enformasyon daha önceden bilinmemelidir. Örneğin; çocuk bezi satışları üzerine yapılan bir çalışmada çocuk bezlerinin 0-4 yaş grubu çocukları olan aileler tarafından alındığı sonucunun bulunması varolan bir bilginin veri tarafından doğrulanmasıdır. Bu çalışmada, çocuk bezlerinin yetişkinlerce de kullanıldığı enformasyonu elde edilirse, bu daha önceden bilinmeyen veya beklenilmeyen bir enformasyon olur. Ayırtırılan enformasyon doğru olmalıdır. Enformasyonun doğruluğu veri içerisinden alınacak değişik örneklerle de incelenmelidir. Elde edilen enformasyon doğru ise bu enformasyon değerlendirilip yeni fırsatlar yakalanmalıdır. Çocuk bezi örneğinde yetişkinlerin neden çocuk bezi kullandıkları enformasyonu araştırılarak, yeni satış politikaları oluşturulması yönünde kullanılmalıdır.

5- Büyük kuruluşlar için önemli enformasyon veya yapıları bulmak amacıyla bu büyük boyutlu veri tabanlarını analiz etmenin faydalı olması.

6- Yeni yapıları ve/veya yönelimleri keşfetmek için veri kaynaklarını analiz etme de kullanılabilecek birçok çeşitli veri madenciliği metodolojisinin bulunması.

“Veri madenciliği, otomatik veya yarı otomatik çözüm araçları (tool) ile büyük ölçeklerdeki verinin anlamlı yapılar ve kurallar keşfetmek üzere araştırılması (exploration) ve analiz edilmesidir.(Berry vd,1997)”. Tanım incelendiğinde, bilinmeyen yapıların ve kuralların bulunması için büyük veri kümelerinin araştırılması ve analiz edilmesin de veri madenciliğinin oynadığı rolün öne çıkarıldığı görülür. Bu araştırma ve analiz işlemlerini yapan veri madenciliği çözüm araçlarının ise otomatik veya yarı otomatik (insan yardımı olarak) olarak çalıştığı vurgulanmıştır.

“Veri Madenciliği, farklı endüstrilerde kullanılan birçok tekniği kapsayan bir terimdir.(Rud,2001)”. Bu tanımda veri madenciliğinin birçok endüstride kullanıldığı ve farklı veri madenciliği teknikleri kullanılarak sonuçlar üretildiği ortaya konulmuştur. Veri madenciliği teknikleri kullanarak ticari kuruluşlar, pazarlama sürecinin hemen hemen her aşamasında rakiplerine karşı üstünlük elde edebilir.

“Veri madenciliği , verilerin içerisindeki yapıların, ilişkilerin, düzensizliklerin, kuralların, ve istatistiksel olarak önemli olan yapıların yarı otomatik olarak keşfedilmesidir.(Vahaplar vd)”. Diğer tanımlarda da değinildiği gibi, bu tanımda da veri madenciliği teknikleri ile gerçekleşmesi umulmayan bilgilerin bulunabileceğine değinilmektedir. Buna ek olarak istatistiksel olarak önemli yapılardan söz edilerek veri madenciliğinin istatistik boyutu gözler önüne serilmiştir. Veri madenciliğinin istatistik ile ilişkisine daha sonraki bölümler içerisinde yer alan istatistik ve veri madenciliği alt başlığında değinilecektir.

“Veri madenciliği; istatistik, veritabanı teknolojisi, yapı tanıma (pattern recognition), makine öğrenmesi (machine learning) ve diğer alanları temel alan yeni bir disiplindir.(Hand,1998)” Yazar bu tanımda veri madenciliğinin farklı disiplinleri bir araya getirerek çözümler sunma özelliğini çıkarmıştır.

“Veri madenciliği, büyük veri setlerindeki ilginç, beklenilmeyen ya da değerli yapıların keşfedilmesidir.”* Bu tanımda ön plana çıkan, ortaya çıkarılan yapıların daha önceden bilinmeyen veya gerçekleşmesi öngörülmeyen (çocuk bezi örneği) özellikte olmasıdır.

“Veri madenciliği büyük hacimli verilerden bilginin ayrıştırılması sürecidir. Başka bir deyişle, veri madenciliği, öngörülmeyen, değerli yapıların ve ilginç ilişkilerin büyük gözlemsel veritabanlarında yer alan karmaşık veriler içerisinde keşfedilmesi bilimi veya sanatıdır.”**. Bu tanımda veri madenciliğine sanatsal bir yapı özelliği verilmesi söz konusudur.

“Pregibon (1996)’a göre veri madenciliğinin meydana çıkan alanı; istatistiğin, yapay zekanın ve veri tabanı araştırmasının karışımıdır.”*** Bu tanımda, veri madenciliğinin değişik disiplinlerden yararlanarak çözümler ürettiği ön plana çıkarılmaktadır.

Veri madenciliği tanımlarına baktığımızda genel olarak görünen, tanımların yazarların çalışma alanlarına göre farklılaşmasıdır. Bu tanımlarda öne çıkan noktalar şunlardır:

Veri Madenciliği;

- 1- Büyük ve karmaşık verilerle çalışır.
- 2- Her türlü veriyi kullanarak çözümler üretebilir.
- 3- İstatistik, yapay zeka, makine öğrenmesi, Veri tabanlarında bilgi keşfi, bilgisayar bilimi, yapı tanıma vb. gibi disiplinlerden faydalanır.
- 4- Daha önceden bilinmeyen, doğrulanabilir, etkinleştirilebilir enformasyon arar.
- 5- Otomatik veya yarı otomatik olarak çalışan çözüm araçları kullanır.
- 6- Birçok endüstride kullanılmaktadır.
- 7- Sorunlara göre değişen çözüm araçları vardır.
- 8- Hızla büyümekte olan bir sektördür.

* Hand,D.J.,Why Data Mining is More Than Statistics Writ Large, www.stat.fi/isi99/proceedings/arkisto/varasto/hand0490.pdf

** Ganesh,S., Data Mining Should It Be Included in the Statistics Curriculum?, www.stat.auckland.ac.nz/~iase/publications/1/314_gane.pdf

*** Data Mining, Statistical Methods Mining and History of Statistics,,stat.tamu.edu/ftp/pub/eparzen/future.pdf

9- Bir sanattır.

Yukarıda yer alan dokuz maddeden yararlanarak bir veri madenciliği tanımı yaparsak şöyle bir tanım yapabiliriz; “Veri madenciliği hızla büyümekte olan, büyük ve karmaşık veriler içerisinde herhangi bir veri ayrımı yapmaksızın , istatistik , yapay zeka, makine öğrenmesi, veri tabanlarında bilgi keşfi, yapı tanıma gibi disiplinlerden yararlanarak, birçok endüstride sorunların özelliklerine göre değişen otomatik veya yarı otomatik olarak çalışan çözüm araçları kullanarak, daha önceden bilinmeyen , doğrulanabilir , eyleme geçirilebilir enformasyon araştırma sanatıdır.”

2.1 Veri Madenciliğinin Ortaya Çıkışı

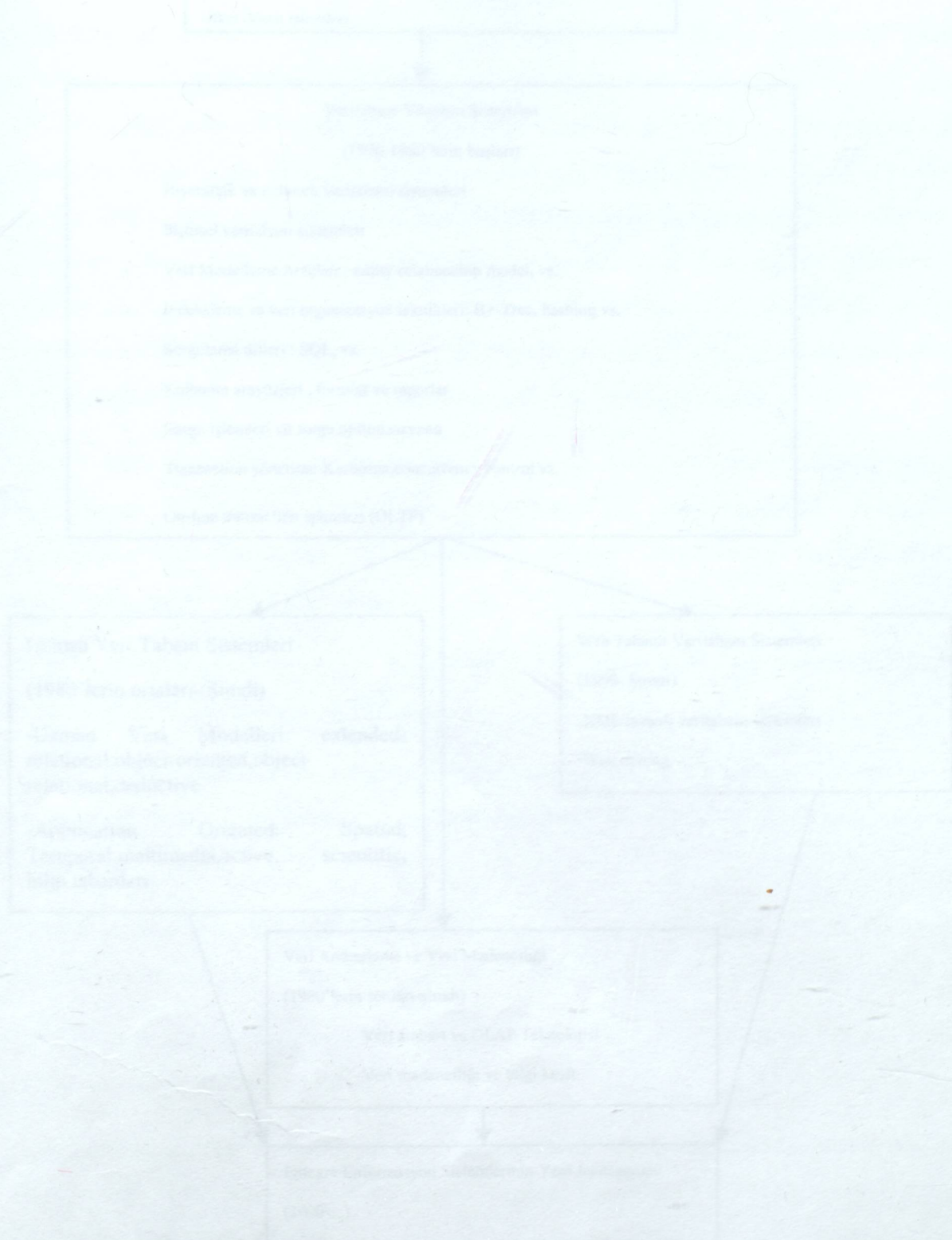
Buraya dek veri madenciliğinin tanımı yapıldı. Veri madenciliğinin uygulama alanlarına değinilerek veri madenciliğinin tercih edilme nedenlerine değinilmeye çalışıldı. Şimdiyse Jiawei Han ile Micheline Kamber’in yazmış oldukları Data Mining Concepts and Techniques adlı kitaplarının giriş bölümlerinde veri tabanı teknolojisinin evrimi olarak adlandırdıkları şekilden yararlanılarak veri madenciliğinin oluşumu incelenecek ve yazarların veri madenciliği için yapmış oldukları farklı bir tanıma yer verilecek.

“Gereksinmeler buluşların anasıdır.(Han vd, 2001)” sözleriyle kitabın Giriş bölümü başlıyor. Yeni bir şeyler aramamız için birtakım koşulların olgunlaşması gerekir ki bu olgunlaşan koşullar bizi yeni buluşlar yapmaya yönlendirsın. Veri madenciliği disiplinin de oluşumunda yaşanması ya da olgunlaşması gereken süreç veritabanı teknolojisinin evrim süreci olmuştur.

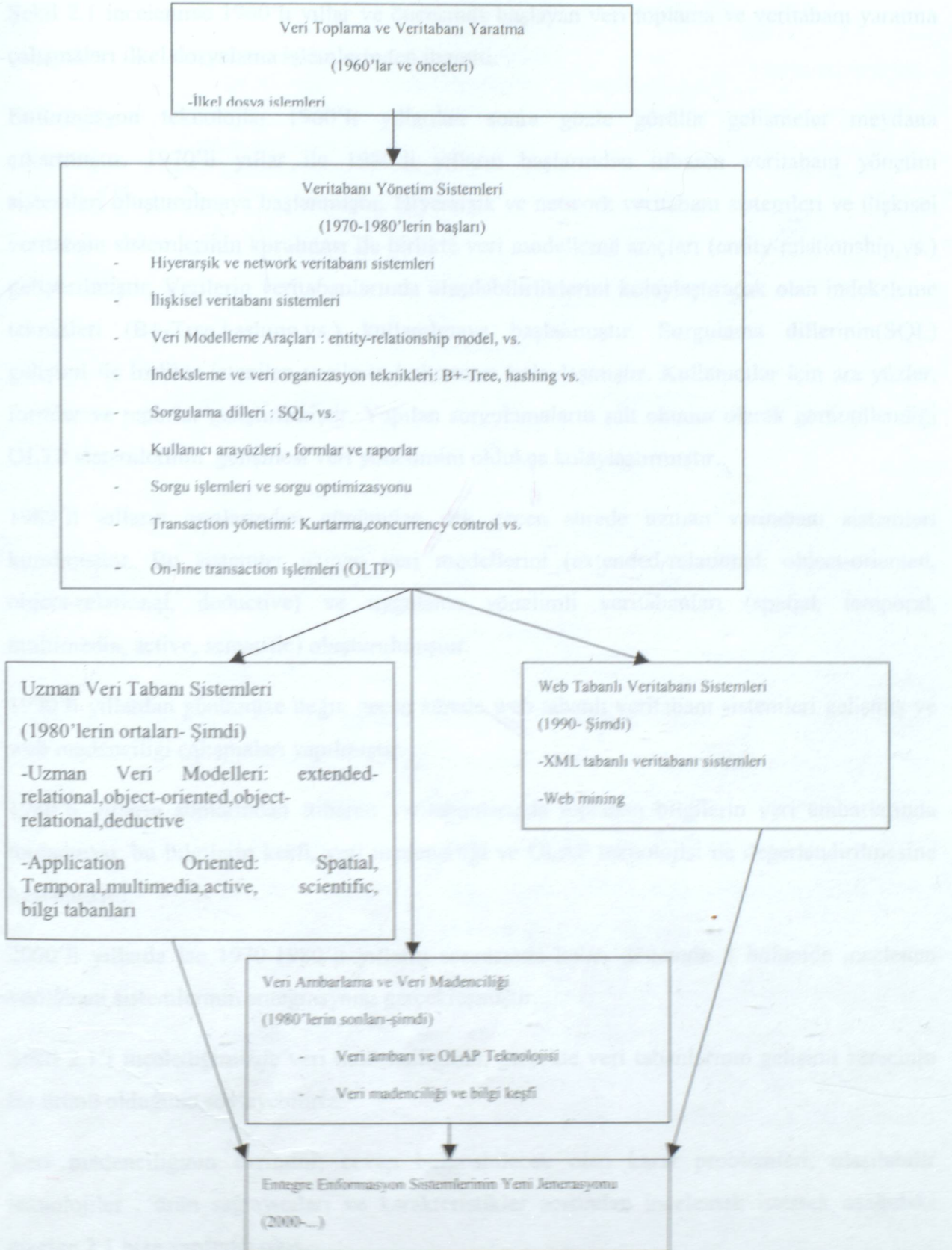
Veri madenciliğinin tanımı yapılırken “büyük ölçeklerdeki veri”den söz edilmişti. Veri madenciliğiyle ilgili yapılan tüm tanımlardaki ortak nokta “büyük ölçeklerdeki veri”dir. İşte bu ortak noktanın kaynağı ise veritabanlarıdır. Veritabanlarının oluşumu ve veritabanı sistemlerindeki gelişmelerin bir sonucu olarak ortaya çıkan büyük miktarlardaki veriyi değerlendirerek bu veri içerisinde gizli olarak bulunan yapıların, ilişkilerin ortaya çıkarılmasına duyulan gereksinme veri madenciliği disiplininin oluşumunda önemli bir rol oynamıştır. Veritabanı sistemlerinde yaşanan gelişme ve veri madenciliğinin bu gelişme içerisinde oynadığı rolü gösteren diyagram aşağıdaki şekilde yer almaktadır.

“Son yıllarda veri madenciliğinin enformasyon endüstrisinde oldukça fazla dikkat çekmesinin nedeni, büyük miktarlardaki veriye kolay ulaşılabilir olunması ve kısa zamanda doğmasından korkulan böyle büyük miktarlardaki verinin yararlı enformasyon ve bilgiye dönüştürülmesi ihtiyacıdır. Enformasyon ve bilgi kazanılması; iş yönetimi, üretim kontrolü ve Pazar

analizlerinden mühendislik dizaynı ve bilim araştırmalarına kadar birçok uygulamalar için kullanılabilir. (Han ve Kamber,2001)”



Şekil 2.1: Veritabanı Teknolojisinin Gelişimi ve Yapılandırılması (Han vd,2001)



Şekil 2.1. Veritabanı Teknolojisinin Gelişimi ve Veri Madenciliği (Han vd,2001)

Şekil 2.1 incelenirse 1960'lı yıllar ve öncesinde başlayan veri toplama ve veritabanı yaratma çalışmaları ilkel dosyalama işlemlerinden ibaretti.

Enformasyon teknolojisi 1960'lı yıllardan sonra gözle görülür gelişmeler meydana çıkartmıştır. 1970'li yıllar ile 1980'li yılların başlarından itibaren veritabanı yönetim sistemleri oluşturulmaya başlanmıştır. Hiyerarşik ve network veritabanı sistemleri ve ilişkisel veritabanı sistemlerinin kurulması ile birlikte veri modelleme araçları (entity-relationship, vs.) geliştirilmiştir. Verilerin veritabanlarında ulaşılabilirliklerini kolaylaştıracak olan indeksleme teknikleri (B+-Tree, hashing, vs.) kullanılmaya başlanmıştır. Sorgulama dillerinin (SQL) gelişimi ile birlikte istenilen verilerin bulunması kolaylaşmıştır. Kullanıcılar için ara yüzler, formlar ve raporlar geliştirilmiştir. Yapılan sorgulamaların salt okunur olarak görüntülediği OLTP sistemlerinin gelişmesi veri yönetimini oldukça kolaylaştırmıştır.

1980'li yılların ortalarından günümüze dek geçen sürede uzman veritabanı sistemleri kurulmuştur. Bu sistemler uzman veri modellerini (extended-relational, object-oriented, object-relational, deductive) ve uygulama yönelimli veritabanları (spatial, temporal, multimedia, active, scientific) oluşturulmuştur.

1990'lı yıllardan günümüze değin geçen sürede web tabanlı veritabanı sistemleri gelişmiş ve web madenciliği çalışmaları yapılmıştır.

1980'li yılların sonlarından itibaren veritabanlarında toplanan bilgilerin veri ambarlarında toplanması, bu bilgilerin keşfi, veri madenciliği ve OLAP teknolojisi ile değerlendirilmesine başlanmıştır.

2000'li yıllarda ise 1970-1980'li yılların sonrasında kalan dönemde 3 bölümde incelenen veritabanı sistemlerinin entegrasyonu gerçekleşmiştir.

Şekil 2.1'i incelediğimizde veri madenciliğinin, gerçekte veri tabanlarının gelişimi sürecinin bir ürünü olduğunu söyleyebiliriz.

Veri madenciliğinin evrimini; cevap bulunabilecek olan karar problemleri, ulaşılabilir teknolojiler, ürün sağlayıcıları ve karakteristikler açısından incelemek istersek aşağıdaki çizelge 2.1 bize yardımcı olur.

Çizelge 2.1 Veri Madenciliğinin evrimi (Aldana,2000)

Gelişim Adımları	Cevaplanan Karar Problemi	Kullanılabilen Teknolojiler	Ürün Sağlayıcıları	Karakteristikler
Veri Toplama (1960'lar)	"Benim toplam karım geçen 5 yılda ne kadardı?"	Bilgisayarlar, Teypler, Diskler	IBM,CDC	Geriye dönük , statik veri dağıtımı
Veri Erişimi (1980'ler)	"İngiltere'de geçen mart ayında birim satışları ne kadardı?"	İlişkisel Veritabanları, SQL, ODBC	Oracle,Sybase, Informix,IBM, Microsoft	Kayıt düzeyinde geriye dönük, dinamik veri dağıtımı
Veri Ambarlama ve Karar Destek Sistemleri (1990'lar)	"İngiltere'de geçen mart ayında birim satışları ne kadardı?"	OLAP, Çok Boyutlu Veritabanı Sistemleri, Veri ambarları	Pilot, Comshare, Arbor,Cognos, Microstrategy	Çoklu düzeylerde, geriye dönük dinamik veri dağıtımı
Veri Madenciliği (Bugün)	"Gelecek ay Boston'daki birim satışlar muhtemelen ne olabilir, niçin?"	İleri düzeyde algoritmalar, çok işlemcili bilgisayarlar, büyük veritabanları	Pilot, Lockheed, IBM,SGI, SPSS,SAS, Microsoft vs.	Geleceğe dönük ,proaktif enformasyon dağıtımı

1960'lı yıllar veri madenciliğinin gelişiminde veri toplama aşamasının gerçekleştiği yıllar olmuştur. Cevaplanabilen karar problemleri ayrıntıya girmeden belirli bir döneme ilişkin problemler iken, kullanılabilen teknolojiler bilgisayarlar, teypler ve diskler olmuştur. Bu teknolojilerden teypler ve diskler günümüzde de veritabanlarının yedeklenmesi için kullanılmaktadır. Bu yıllarda ürün sağlayıcı firmalar IBM ve CDC'dir ve yalnızca geriye dönük ve statik veri dağıtımı yapılmıştır. Bu yıllarda sadece geriye dönük aranılan veriye ulaşılırken bu verilerden enformasyon elde edilmediği görülmektedir.

1980'li yıllara geçildiğinde statik veri dağıtımında dinamik veri dağıtımına geçildiği görülmektedir. Elde edilen verilerden tüm veri içerisinden belirli dönemlere (ay) göre sorgulama yapmak mümkün olmuştur. İlişkisel veritabanlarının oluşumu , SQL ve ODBC ile veri kaynaklarına ulaşım bu yıllarda gerçekleşmiştir. Ürün sağlayıcılarındaki artışta dikkat çekicidir.

1990'lı yıllarda veri toplama ve saklamadaki gelişimin sonucu olarak veri ambarları oluşturulmaya başlanmış ve karar destek sistemleri devreye girmiştir. Farklı coğrafi bölgelerden gelen verilerin bir arada değerlendirilmeleri gerçekleşmiştir. 1980'li yıllarda tekli düzeylerde yapılan dinamik veri dağıtımı çoklu düzeylerde yapılmaya başlanmıştır. OLAP ve çok boyutlu veritabanları göze batan değişimlerdir. 1980'li yıllarda kullanılan OLTP

sistemleri yerini analitik işlemler yapma olanağı sağlayan OLAP sistemlerine bırakmıştır.

Şimdilerde ise veri madenciliği tam anlamı ile yapılmaya başlanmış olup geriye dönük yapılabilen veri dağıtımına ek olarak ileriye dönük tahminlere imkan veren proaktif enformasyon dağıtımı da yapılmaya başlanmıştır. Burada da ön plana çıkan değişim enformasyon dağıtımı olmuştur. Ürün sağlayan firma sayısı da 1960'lardan 1990'ların sonlarına dek bir elin parmaklarının sayısını geçmezken, şimdilerde oldukça çoktur. 1960'lı yıllarda yalnızca istenilen verinin elde edilmesiyle sonuçlanan işlemler artık şimdi geleceğe dönük tahminler ve bu tahminlerin nedenlerinin açıklanmasına dönüşen işlemlere dönüşmüştür. Çok işlemcili bilgisayarlar, büyük veritabanları ve gelişmiş algoritmalar bu değişimde büyük rol oynamışlardır.

Ayrıca veri madenciliğinin gelişimi ile birlikte kurumsal ve kamusal alana (Public Domain) ait veri madenciliği araçlarının sayısını gösteren çizelge aşağıda yer almaktadır.

Çizelge 2.2 Amerika Birleşik Devletleri'nde Kurumsal ve Kamusal Alana Ait Veri Madenciliği Araçlarının Sayısı (Aldana,2000)

<i>Araç Tipi</i>	<i>Şubat 1997</i>	<i>Şubat 1998</i>	<i>Kasım 1998</i>	<i>Ağustos 1999</i>
<i>Suites</i>	18	30	37	37
<i>Sınıflama</i>				
<i>Çoklu Yaklaşım</i>	4	8	8	10
<i>Karar Ağaçları</i>	14	18	18	19
<i>Karar Keşfi</i>	10	11	12	14
<i>Yapay Sinir Ağları</i>	50	60	76	100
<i>Bayesgil</i>	0	0	10	20
<i>Diğer</i>	6	7	12	13
<i>Sınıflama Toplam</i>	84	104	136	176

Çizelge 2.2 Devam Amerika Birleşik Devletleri'nde Kurumsal ve Kamusal Alana Ait Veri Madenciliği Araçlarının Sayısı (Aldana,2000)

Birliktelikler ve Diziler	0	0	6	13
Kümeleme	6	5	10	12
Görselleştirme	5	12	26	31
Text ve Web Madenciliği	0	5	7	15

Çizelge incelendiğinde veri madenciliği araçlarının üç yıl içerisindeki hızlı artışı dikkat çekmektedir. Bu da artık 1990'lı yıllarda veritabanı teknolojisinin gelişmesi sonucu ortaya çıkan depolanmış verilerden enformasyon elde etme isteğini de gözler önüne sermektedir.

2.2 Veri Madenciliği Niçin Tercih Ediliyor?

Veri madenciliğinin neden tercih edildiği konusuna girmeden önce Nobel ödülü sahibi Dr. Penzias'ın 1999 ocak ayında ComputerWorld'de vermiş olduğu bir röportajda veri madenciliğinin önemini vurguladığı şu sözlerine yer vermek istiyorum.

“Veri madenciliği oldukça önemli bir konuma gelecektir; ve firmalar müşterileriyle ilgili hiçbir bilgiyi atamayacaklardır, çünkü bu bilgiler oldukça değerli olacaktır. Eğer siz müşterilerinizin bilgilerini değerlendirmiyorsanız, firmanız sektörünüzün dışında kalacaktır.”(Groth,1999)

Veri madenciliği bölümünde de belirtildiği gibi büyük kuruluşların müşterileri ile ilgili veri saklamalarının ve bu bilgileri değerlendirmelerinin önemi Dr. Penzias'ın yukarıdaki sözleri ile de desteklenmektedir.

Veri madenciliği uygulamaları için bazı örnekler aşağıda yer almaktadır:

- Tıbbi araştırmalarda
 - DNA içerisindeki genlerin sıralarının belirlenmesi
 - Protein analizlerinin yapılması
 - Hastalık haritalarının hazırlanması
 - Hastalık tanıları
 - Sağlık politikalarına yön verilmesi
- Müşteri davranışlarının belirlenmesinde

- Doğrudan satış kampanyalarına cevap verme potansiyeli olan müşterilerin belirlenmesi
- Müşteri segmentasyonu
- Hile yoluyla suç işleyen müşterilerin saptanması
- Müşterileri elde tutma
- Kaybedilen müşterilerin geri kazanılması
- Kaybedilebilecek durumda olan müşterilerin tespiti
- Çapraz satışlar
- Müşteri sadakati
- Sepet analizleri yardımı ile marketlerde ürünlerin raflara dağılımının belirlenmesi
- Piyasaya sunulacak yeni ürünlerin müşteri potansiyelinin belirlenmesi
- Pazar segmentasyonunun yapılması
- Web hizmetlerinde
 - Elektronik ticaret yapan firmalar için müşteri davranışlarının belirlenmesi
 - Web Sitesini ziyaret eden kullanıcının daha önceki davranışlarına göre yönlendirilmesi
 - Web sitesi güvenliğinin sağlanması
 - Kullanıcı davranışlarına göre web sitesinin yenilenmesi
 - Kullanıcı profilinin belirlenmesi
- Bankacılık ve finans sektöründe
 - Stok tahmini (forecast)
 - Risk değerlendirme
 - Portföy yönetimi
 - Kar analizi
 - Kredi politikalarının belirlenmesi
- Sigorta Sektöründe
 - Risk analizi
 - Hileli işlemlerin belirlenmesi
 - Sigorta primlerinin belirlenmesi
- Telekomünikasyon hizmetlerinde
 - İletişim ağlarında sorunlu bölgelerin belirlenmesi
 - Kaçak hat kullanımlarının belirlenmesi
 - Kullanıcı davranışlarının belirlenmesi
 - Müşteri davranışlarına göre yeni hizmetlerin sunulması

- Fiyat politikalarının belirlenmesi

Yukarıda belirttiğimiz ve sayısını daha da arttırabileceğimiz farklı alanlardaki veri madenciliği uygulamaları ile elde edilebilecek avantajlar, günümüzde veri madenciliğine daha çok önem verilmesine neden olmaktadır. Veri madenciliği uygulamalarının genellikle müşteri ilişkilerinde etkin rol oynaması, işletmeciler için de veri madenciliğini çekici duruma getirmektedir. Walter Alberto Aldana ise (çizelge 2.1)'i şu şekilde yorumlamaktadır: “Veri madenciliği teknikleri araştırma ve ürün geliştirmenin gerçekleştiği uzun bir sürecin ürünüdür. Bu evrim verilerin ilk olarak bilgisayarlarda depolanmasıyla başladı ve kullanıcılara gerçek zaman içerisinde verileri arasında gezmelerine izin verecek olan birçok teknolojinin üretilmesi gibi veri erişimindeki ilerlemelerle sürdü. Veri madenciliği iş dünyasındaki uygulamalar için hazırdır, çünkü veri madenciliği olgunlaşmış olan üç teknoloji tarafından desteklenmektedir.

- Büyük veri koleksiyonu
- Güçlü çok işlemcili bilgisayarlar
- Veri madenciliği algoritmaları

Kurumsal veritabanları tahmin edilemeyecek oranlarda büyümektedirler. META Grub’un yapmış olduğu veri ambarı projelerine ilişkin ankette şu sonuçlar bulundu: “Ankete cevap verenlerin %59’u 1996’nın 2. çeyreğinde 50 gigabyte düzeyinden büyük veri ambarlarına sahip olmayı hedeflerken %19’unun bu miktara ulaştığı görülmüştür.”. Perakende gibi bazı endüstrilerde daha büyük rakamlara ulaşabilir.(Aldana,2000) ”

2.3 Birkaç Veri Madenciliği Uygulama Örneği

2.3.1 US WEST Firması’nda Bir Veri Madenciliği Uygulaması (Groth,1999)

US WEST, SAS Enterprise Miner™ ürününü kullanan 20 milyonun üzerinde müşterisi bulunan birleşik devletlerin en büyük telekomünikasyon şirketlerinden birisidir.

US WEST firması sektördeki birçok firma gibi düzensizlik (deregulation) nedeni ile pazar payında bir düşüş yaşamıştır. Pazarlama ilişkileri ikinci başkanı, yönetim kuruluna, şirketin CEO’suna CRM iş planı üzerinde bir soru yöneltti. “Müşteri karlılığınızı, pazarlama yatırımlarındaki müşterileri kazanma ve varolan müşterileri elde tutma çalışmalarını en iyi duruma çıkararak nasıl arttırabilirsiniz?”. Bu olay 1998 yılında US WEST firmasına, veri madenciliği platformu olarak SAS Enterprise Miner ürününü de kapsayan, bir girişimci pazarlama stratejisi geliştirmek için yol açtı.

SAS Enterprise Miner ürünü vasıtası ile , US WEST Firması pazarlama profesyonellerine hedeflenen pazarlama kampanyalarını planlamak, kurmak, yönetmek ve değerlendirmek için kampanya yönetimi software'inin kullanımına kolaylık sağlamıştır. Veri madenciliği ve kampanya yönetimi software'ini birleştirerek pazarlamacılar için çok zaman harcamaya neden olan bütün müşteri veritabanını skorlama ihtiyacını elemişlerdir. Manuel aracılık ile insan yanlışlarını indirgeyerek pazarlama dönemlerini kısaltmışlardır. Yüksek pazarlama ROI'si üretmek için küçük ve yüksekçe hedeflenmiş (highly-targeted) müşteri segmentasyonlarına olanak sağlamışlardır. Buradaki en büyük avantajlardan birisi ise, yalnızca özel kampanyaya ait kayıtların skorlanması nedeniyle kampanyanın hızlı ve daha iyi sonuçlara ulaşmasıdır.

2.3.2 Bass Export Firması'nda Bir Veri Madenciliği Uygulaması (Groth,1999)

Bass Export İngiltere'deki en büyük bira ihracatçısıdır. Şirket veri madenciliği yoluyla önemli kararlar almak için IBM şirketinin Intelligent MinerTM ürününü tercih etmiştir. Bu şirketin aynı zamanda 9 bira fabrikası vardır ve 80'den fazla denizaşırı ülkeye 60 çeşitten fazla bira dağıtmaktadır. IBM Quarterly'ye göre, Bass Export, büyük markaların ve iyi servisin kendilerinin İngiliz bira pazarındaki liderliklerini korumada yeterli olamayacağını anlamıştır. Bass Brewers'dan Mike Fisher şunları söylemiştir: "Her hafta , 23.000 müşterinin siparişlerine cevap vermekteyiz. Bu müşterilerin her birinin alışveriş alışkanlıklarını, bireysel hesabın her birinin karlılığını ve dağıttığımız her bir markanın karlılığını anlamak bizim için oldukça önemlidir."

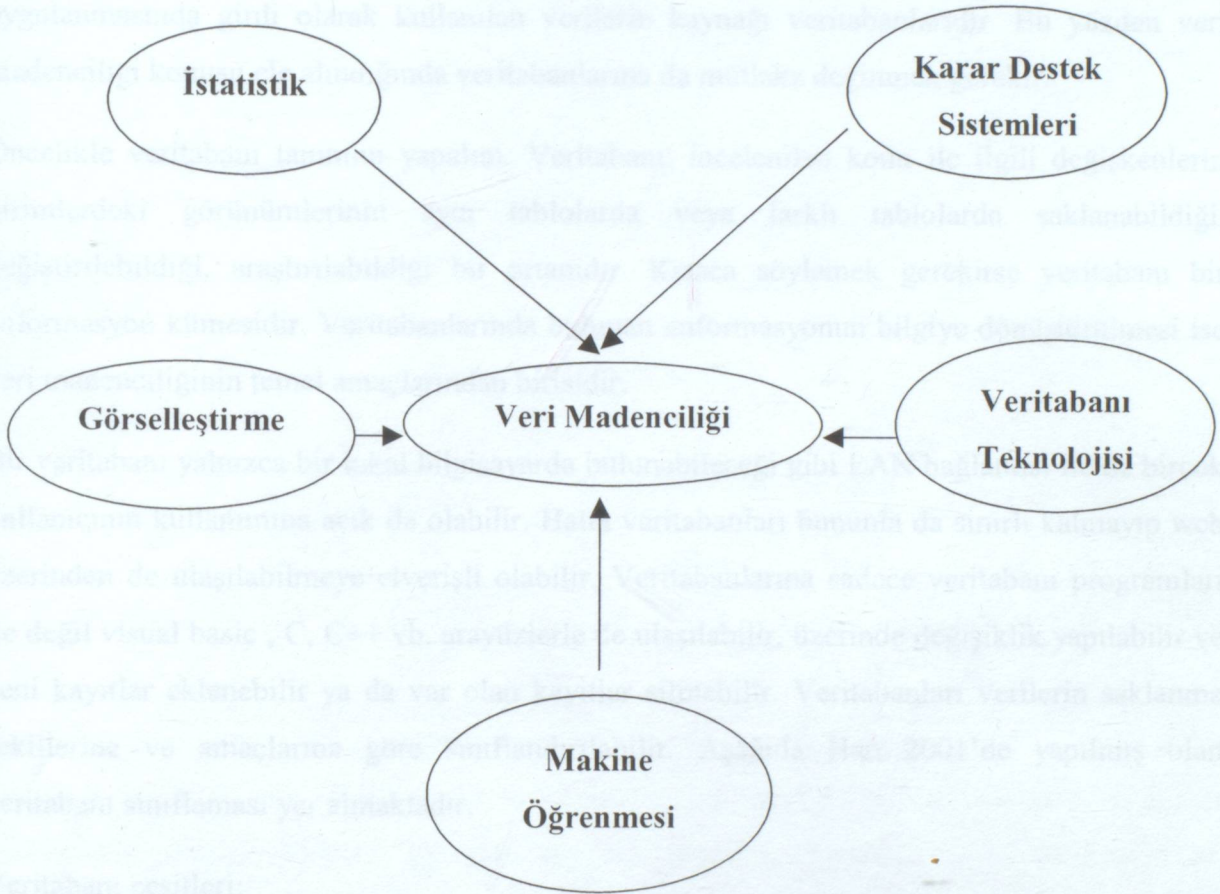
Yukarıda değinilen iki başarı öyküsü de, büyük müşteri potansiyeline ve bu müşterilere ait veritabanına sahip olan firmaların, veri madencilik ürünlerini kullanarak bu veriler içerisinde müşterilerine ait bilgilere ulaşarak kendi karlılık oranlarını artırmış olduklarını sergilemektedir.

Bu da, hemen hemen tüm pazarlarda artan rekabete karşılık firmaların veri madenciliği çözüm araçlarını kullanarak rekabete karşı direnebileceklerini göstermektedir.

Pazarda büyümekten çok, pazar payını korumak önemlidir. Pazar payını korumanın en önemli yolu da müşteriler ve bayilerle gerçekleştirilen tüm işlemlerin günü gününe izlenmesinden geçer. Böylece toplanan verilerden, veri madenciliği araçları ile enformasyon elde edilip bu enformasyonun gerçek hayatta kullanılabilir bilgiye dönüştürülmesi işin gereğidir.

3. Veri Madenciliği ve Diğer Disiplinler

Veri madenciliğinin tanımı yapılırken veri madenciliğinin farklı disiplinlerden yararlandığı da vurgulanmıştı. Bu bölümde veri madenciliğinin yararlandığı diğer disiplinlere yer verilecektir. Bu disiplinlere değinmeden önce, veri madenciliğinin merkezde yer aldığı ve diğer disiplinlerin veri madenciliğinin çevresinde kümелendiği Şekil 3.1’de yer verilecek.



Şekil 3.1. Veri Madenciliğinin Yararlandığı Diğer Disiplinler

Şekil 3.1’de veri madenciliğinin yararlandığı bazı disiplinler yer almaktadır. Seçilen disiplinler değişik kaynaklarda ortak olarak sözü edilen disiplinlerdir. Bu disiplinler veri madenciliği çalışmalarının hemen hemen her aşamasında karşımıza çıkabilir. Veri madenciliğini bu disiplinlerin bir bileşimi olarak düşünebiliriz. Veri madenciliğinin yararlandığı disiplinlere bulanık mantık, lojik programlama, paralel işlemler, çalışılan problemin türüne göre işletme, ekonomi gibi birçok disiplini ekleyebiliriz. Bu bölümün alt bölümlerinde Şekil 3.1’de gösterilen disiplinler incelenecek ve bu disiplinlerin veri madenciliği çalışmalarında oynadıkları rollere değinilecektir. İlerideki bölümlerde sözü edilecek olan veri madenciliği metodolojisindeki sıralama göz önünde tutularak veri

madenciliğinin yararlandığı disiplinlerle olan ilişkisi incelenecek. Kullanılan tekniklere ise, veri madenciliği metodolojisi içerisinde yer verilecektir.

3.1 Veri Tabanı Teknolojisi

Önceki bölümlerde veritabanı teknolojisinin gelişimi sürecinin bir sonucu olarak veri madenciliğinin gündeme geldiğine değinilmişti. Veri madenciliği tekniklerinin uygulanmasında girdi olarak kullanılan verilerin kaynağı veritabanlarıdır. Bu yüzden veri madenciliği konusu ele alındığında veritabanlarına da mutlaka değinmek gerekir.

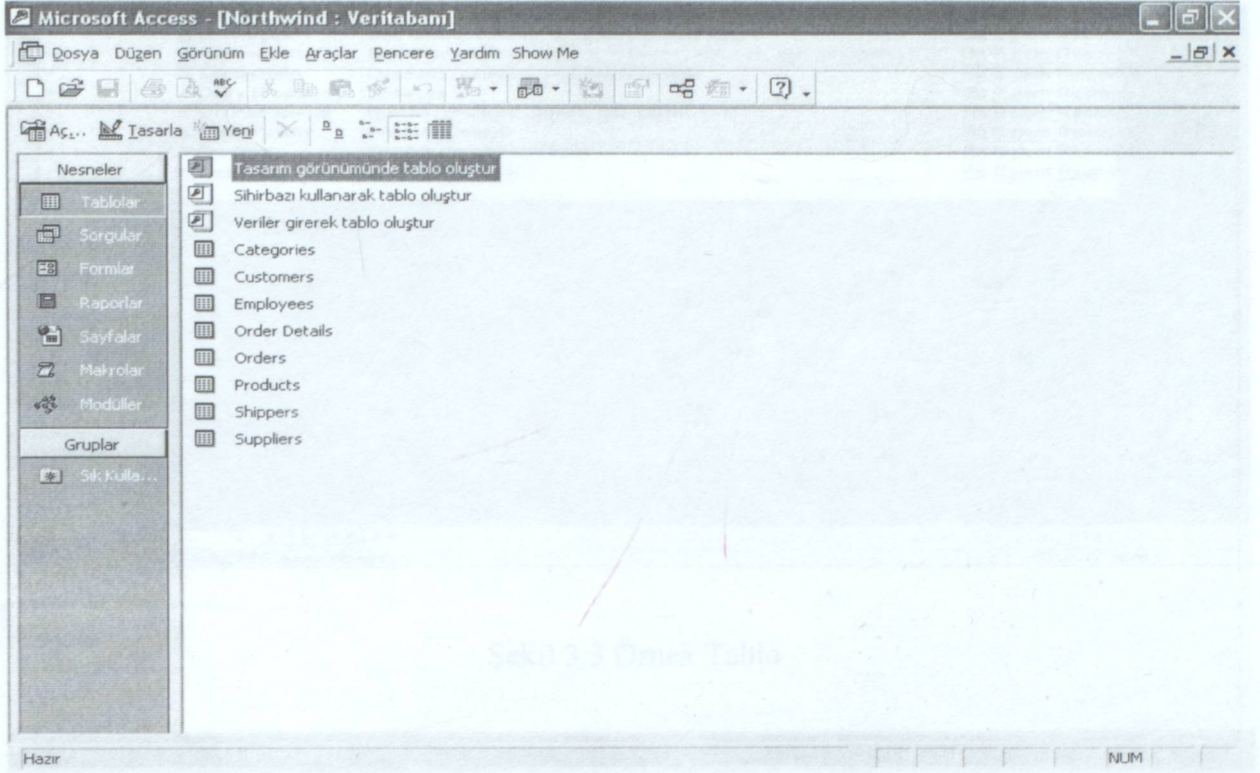
Öncelikle veritabanı tanımını yapalım. Veritabanı, incelenilen konu ile ilgili değişkenlerin birimlerdeki görünümünün aynı tablolarda veya farklı tablolarda saklanabildiği, değiştirilebildiği, araştırılabildiği bir ortamdır. Kısaca söylemek gerekirse veritabanı bir enformasyon kümesidir. Veritabanlarında bulunan enformasyonun bilgiye dönüştürülmesi ise veri madenciliğinin temel amaçlarından birisidir.

Bir veritabanı yalnızca bir lokal bilgisayarda bulunabileceği gibi LAN bağlantısı ile de birçok kullanıcının kullanımına açık da olabilir. Hatta veritabanları bununla da sınırlı kalmayıp web üzerinden de ulaşılabilmeyle elverişli olabilir. Veritabanlarına sadece veritabanı programları ile değil visual basic , C, C++ vb. arayüzlerle de ulaşılabilir, üzerinde değişiklik yapılabilir ve yeni kayıtlar eklenebilir ya da var olan kayıtlar silinebilir. Veritabanları verilerin saklanma şekillerine ve amaçlarına göre sınıflandırılabilir. Aşağıda Han 2001'de yapılmış olan veritabanı sınıflaması yer almaktadır.

Veritabanı çeşitleri:

- İlişkisel veritabanları
- Veri ambarları
- İşlemsel (transactional) veri tabanları
- Uzman veritabanı sistemleri ve uygulamaları
 - Nesne yönelimli veritabanları
 - Nesne ilişkisel veritabanları
 - Uzamsal (spatial) veritabanları
 - Zamansal (temporal) ve zaman serileri veritabanları
 - Text ve multimedya veritabanları
 - Heterojen ve daha önceden kullanılan (legacy) veritabanları
 - Web veritabanları

Aşağıdaki şekilde Microsoft Access programında yer alan örnek bir veritabanı yer almaktadır.



Şekil 3.2 Örnek veritabanı

Şekil 3.2'deki örnek veritabanının nesneler ve gruplar diye adlandırılan iki bölümden oluştuğu görülmektedir. Veritabanının en önemli kısmını nesneler bölümü oluşturur. Nesneler bölümünde tablolar, sorgular, formlar, raporlar, sayfalar, makrolar ve modüller yer alır. Verilerin saklandığı yer tablolardır. Formlar ise tabloların tek tek veya ortak değişkenler aracılığıyla birleştirilmeleri yoluyla oluşturulabilen görsel yapılardır. Formlar yoluyla tablolarda değişiklikler yapılabilir ve/veya tabloların satırları arasında dolaşılabilir. Aşağıdaki şekillerde örnek tablo ve formlar görülebilir.

Microsoft Access - [Categories : Tablo]

Dosya Düzen Görünüm Ekle Bçim Kayıtlar Araçlar Pencere Yardım Show Me

Category ID	Category Name	Description	Picture
1	Beverages	Soft drinks, coffees, teas, beers, and ales	Bit Eşlem Resmi
2	Condiments	Sweet and savory sauces, relishes, spreads, and seasonings	Bit Eşlem Resmi
3	Confections	Desserts, candies, and sweet breads	Bit Eşlem Resmi
4	Dairy Products	Cheeses	Bit Eşlem Resmi
5	Grains/Cereals	Breads, crackers, pasta, and cereal	Bit Eşlem Resmi
6	Meat/Poultry	Prepared meats	Bit Eşlem Resmi
7	Produce	Dried fruit and bean curd	Bit Eşlem Resmi
8	Seafood	Seaweed and fish	Bit Eşlem Resmi

Kayıt: 1 / 8

Number automatically assigned to a new category.

Şekil 3.3 Örnek Tablo

Microsoft Access - [Employees]

Dosya Düzen Görünüm Ekle Bçim Kayıtlar Araçlar Pencere Yardım Show Me

Nancy Davolio

Company Info | Personal Info

Employee ID: 1

First Name:

Last Name:

Title:

Reports To:

Hire Date:

Extension:



Kayıt: 1 / 1

Form/Görünüm

Şekil 3.4 Örnek Form

Sorgular ise tablolardan belirli bir amaca göre değişkenlerin seçimi ile oluşturulmuş yapılardır. Aşağıdaki şekilde bir sorgu örneği yer almaktadır. Şekilden de görüleceği üzere sorgular da gerçekte birer tablodur. Aşağıdaki tablo şu sorgunun çalıştırılması sonucu elde edilmiştir :

```
SELECT DISTINCTROW [Product Sales for 1997].CategoryName, Sum([Product Sales for 1997].ProductSales) AS CategorySales
```

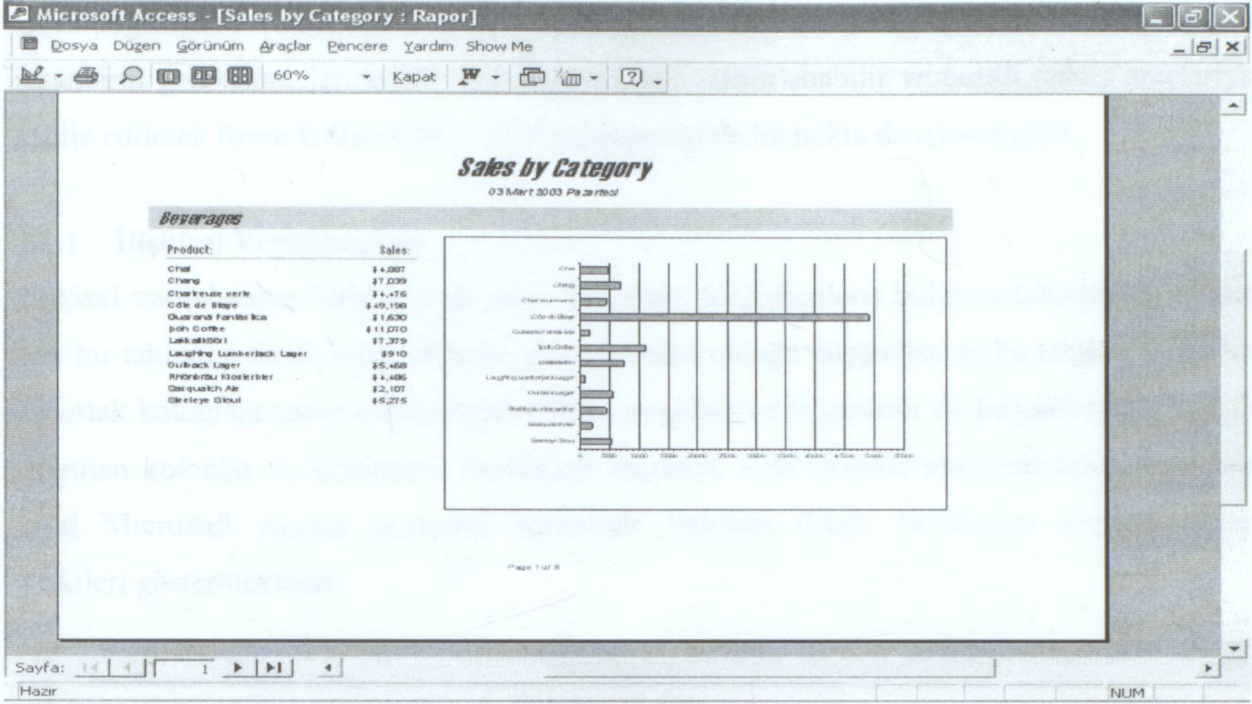
```
FROM [Product Sales for 1997]
```

```
GROUP BY [Product Sales for 1997].CategoryName;
```

Category Name	Category Sales
Beverages	\$102,074.29
Condiments	\$55,277.56
Confections	\$80,894.11
Dairy Products	\$114,749.75
Grains/Cereals	\$55,948.82
Meat/Poultry	\$81,338.06
Produce	\$53,019.98
Seafood	\$65,544.19

Şekil 3.5 Örnek Sorgu

Raporlar ise veritabanındaki tablolar yardımıyla oluşturulan yapılardır. Aşağıdaki şekilde bir rapor örneği görülmektedir. Bu raporlar visual basic programı desteği ile ve amaca uygun olarak oluşturulan SQL sorgulama dili kullanılarak hazırlanır. Şekil 3.6'da görülen rapor kategorilere göre satışları göstermektedir.



Şekil 3.6. Rapor Örneği

Sayfalar ise web sayfası dizaynındaki raporlardır.

View Orders by Company

Select a company from the list to display that customer's orders. Then click the arrow next to an order to see order details.

Select a company: **Comércio Mineiro**

Order ID	Order Date
10290	27-Ağu-1996
10466	06-Mar-1997
10494	02-Nis-1997
10969	23-Mar-1998
11042	22-Nis-1998

Orders 1-5 of 5

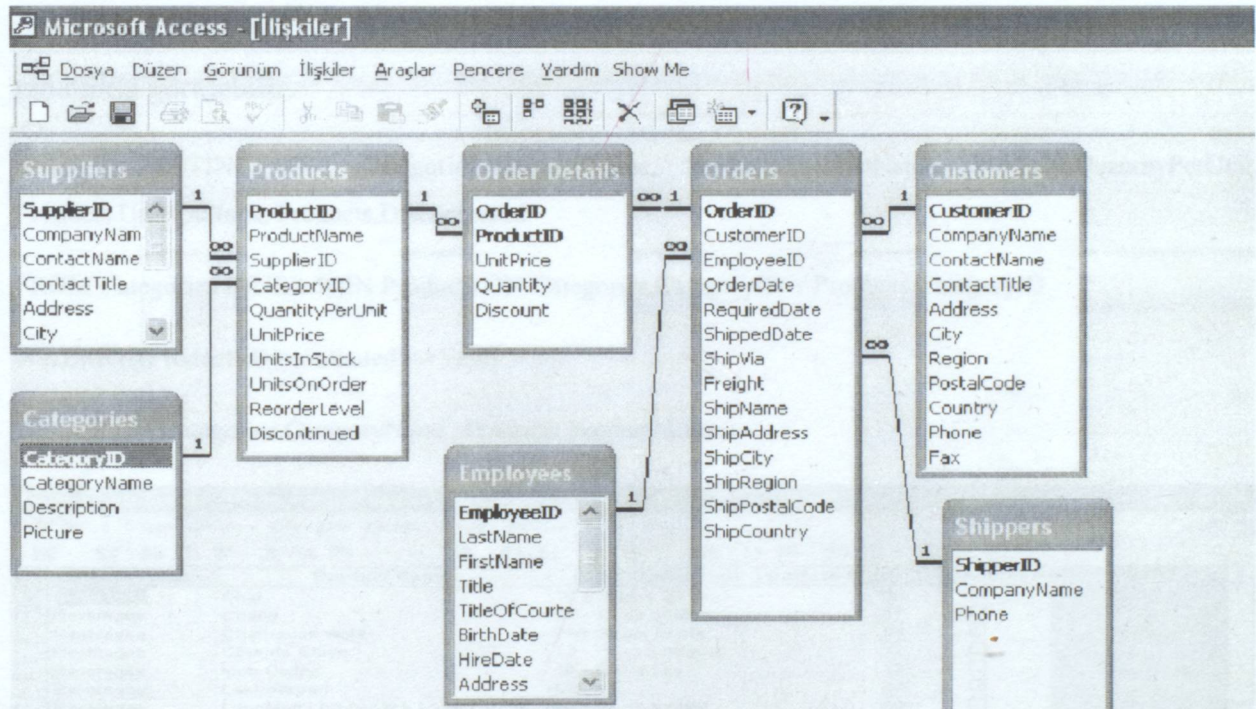
Şekil 3.7. Sayfa Örneği

Makrolar ve modüller ise raporların , sorguların , formların vb. hazırlanmasında kullanılan programlama araçlarıdır.

Görüldüğü üzere veritabanları içerisinde birçok enformasyon saklı olabilir, bu enformasyon insanların gereksinmelerine göre sınıflandırılabilir, raporlanabilir ve çeşitli analiz araçlarıyla analiz edilmek üzere kullanılabilir. Veri madenciliği de bu nokta da işleme girer.

3.1.1 İlişkisel Veritabanları

İlişkisel veritabanları birbirleri ile ortak kolonları (değişkenleri) bulunan tablolardan oluşur. Her bir tablonun farklı bilgi saklama gibi görevleri olduğu düşünülse de bu tablolar birbirleri ile ortak kolonlara sahip olduklarından SQL sorgulama dili yardımı ile birleştirilebilir, ya da istenilen kolonlar ve kolonların özellikleri seçilerek yeni tablolar oluşturulabilir. Aşağıdaki şekil Microsoft Access programı içerisinde bulunan örnek Northwind veritabanındaki ilişkileri göstermektedir.



Şekil 3.8. İlişkisel Veritabanında Bulunan Tablolar Arasındaki İlişkiler

Yukarıdaki şekilde yer alan tablolar arasındaki ilişkiler incelendiğinde tüm tabloların birbirlerine ortak bir kolon aracılığıyla bağlandıkları görülür. Örneğin; Suppliers tablosu Products tablosuna SupplierID kolonu aracılığıyla bağlanmıştır. İki tablo arasındaki ilişki Suppliers tablosundan Products tablosuna doğru 1'e sonsuzdur. Bu tanımlama şu anlama gelmektedir. Suppliers tablosundaki SupplierID kolonunda yer alan kodlar o kolonda tekrarlanmamaktadır. Fakat Products tablosunda yer alan SupplierID kolonundaki kodlar bir ürün sağlayıcıdan birden fazla ürün sağlanabileceğinden dolayı yinelenmektedir. Products

tablosu ise Order Details tablosuna ProductID kolonu aracılığıyla 1'e sonsuz bir ilişki ile bağlanmaktadır. Burada da bir önceki örneğe benzer biçimde Products tablosunda yer alan ProductID kolonundaki kodlar tekrarlanmamaktadır, fakat Order Details tablosundaki ProductID kolonundaki kodlar bir üründen birden çok sipariş edilebileceğinden tekrarlanmaktadır. Buradan da şu sonuca ulaşabiliriz: Eğer ilişki kurulan tablolardan herhangi birindeki ilişkiyi sağlayan değişken tek (unique) ise, ilişki işaretinin tablo yönündeki değeri 1, tek değilse sonsuz olarak gösterilir. Daha açıklayıcı bir örnek vermek gerekirse ; karı koca ilişkisi 1'e 1, baba çocuklar ilişkisi 1'e sonsuz, öğrenci öğretmen ilişkisi de sonsuza sonsuz ilişkiye örnek verilebilir.

Şekil 3.5'te görülen sorgu gibi bu tablolar arasındaki ilişkiler kullanılarak sorgular oluşturulabilir. Bu ilişkilerin kullanıldığı bir sorgu örneği aşağıda yer almaktadır. Sorguda Categories tablosunun Products tablosuna CategoryID kolonu aracılığıyla bağlandığı rahatlıkla görülebilir.

```
SELECT DISTINCTROW Categories.CategoryName, Products.ProductName, Products.QuantityPerUnit,
Products.UnitsInStock, Products.Discontinued
```

```
FROM Categories INNER JOIN Products ON Categories.CategoryID = Products.CategoryID
```

```
WHERE (((Products.Discontinued) <> Yes))
```

```
ORDER BY Categories.CategoryName, Products.ProductName;
```

Microsoft Access - [Products by Category: Seçme Sorgusu]

Dosya Düzen Görünüm Ekle Biçim Kayıtlar Araçlar Pencere Yardım Show Me

Category Name	Product Name	Quantity Per Unit	Units In Stock	Discontinued
Beverages	Chai	10 boxes x 20 bags	39	☐
Beverages	Chang	24 - 12 oz bottles	17	☐
Beverages	Chartreuse verte	750 cc per bottle	69	☐
Beverages	Côte de Blaye	12 - 75 cl bottles	17	☐
Beverages	Ipoh Coffee	16 - 500 g tins	17	☐
Beverages	Lakkalikööri	500 ml	57	☐
Beverages	Laughing Lumberjack Lager	24 - 12 oz bottles	52	☐
Beverages	Outback Lager	24 - 355 ml bottles	15	☐
Beverages	Röhnbräu Klosterbier	24 - 0.5 l bottles	125	☐
Beverages	Sasquatch Ale	24 - 12 oz bottles	111	☐
Beverages	Steeleye Stout	24 - 12 oz bottles	20	☐
Condiments	Aniseed Syrup	12 - 550 ml bottles	13	☐
Condiments	Chef Anton's Cajun Seasoning	48 - 6 oz jars	53	☐
Condiments	Genen Shouyu	24 - 250 ml bottles	39	☐
Condiments	Grandma's Boysenberry Spread	12 - 8 oz jars	120	☐
Condiments	Gula Malacca	20 - 2 kg bags	27	☐
Condiments	Louisiana Fiery Hot Pepper Sauce	32 - 8 oz bottles	76	☐
Condiments	Louisiana Hot Spiced Okra	24 - 8 oz jars	4	☐
Condiments	Northwoods Cranberry Sauce	12 - 12 oz jars	6	☐
Condiments	Original Frankfurt grüne Soße	12 boxes	32	☐
Condiments	Sirup d'érable	24 - 500 ml bottles	113	☐
Condiments	Veggie-spread	15 - 625 g jars	24	☐
Confections	Chocolade	10 pkgs.	15	☐
Confections	Gumbär Gummibärchen	100 - 250 g bags	15	☐
Confections	Maxilaku	24 - 50 g pkgs.	10	☐
Confections	NuNuCa Nuß-Nougat-Creme	20 - 450 g glasses	76	☐

Kayıt: 1 / 69

Name of food category.

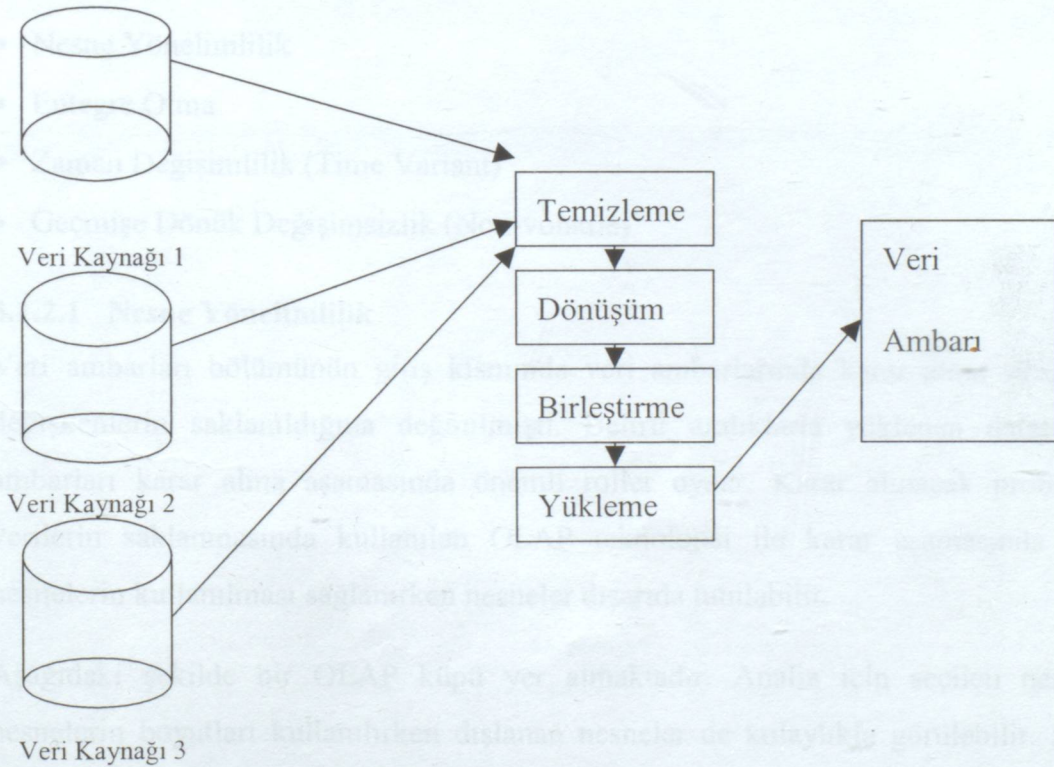
NUM

Şekil 3.9. İki Farklı Tablonun Birleştirilmesi İle Oluşturulmuş Sorgunun Sonucu

3.1.2 Veri Ambarları

Veritabanları konusu incelenirken verilen Northwind veritabanı bulunduran bir şirket düşünelim. Bu şirketin gerek yurtiçi gerekse yurtdışı olsun farklı coğrafi bölgelerde bayileri olsun. İncelenen veritabanı ise tüm şubelerde kullanılsın. Bu şirketin yöneticileri ürün bazında, dönem bazında , şube bazında vb. aylar, çeyrekler (Quarter) vb. için rapor almak istesin; ve ürün dağıtımını yapmak için bir modelleme yapma gereği duysun. Bu durumda farklı şubelere ait veritabanlarında bulunan enformasyonun ortak bir veritabanında toplanması gerekir. İşte birçok veritabanından alınarak birleştirilen verinin bulunduğu depoya “veri ambarı” denilir.

Veri ambarlarındaki veriler, geçmişe dönük bilgiler elde etmek ve bu verilerden ileriye dönük tahminler yapabilmek için saklanır. Önemli değişken grupları bazında yapılan depolama sırasında veri ambarlarına farklı dönemlerde yükleme yapıldığı için zaman boyutu da veri ambarlarına eklenmiş olur. Müşteriler, bayiler, ürünler, zaman vb. gibi konularda farklı kanallardan gelecek bilginin veri ambarında olması gereken tek bir formatta olması gerekir. Veri ambarlarında bulunan verilerin önemli özelliklerinden birisi de temiz ve dönüşümden geçmiş olmasıdır. Aşağıdaki şekil bir veri ambarının kullanım sürecini göstermektedir.



Şekil 3.10 Veri Ambarının Kuruluş Süreci

Şekil 3.10 incelenirse farklı veri tabanlarında bulunan veriler veri ambarına yüklenmeden önce belirli aşamalardan geçmektedir. Öncelikle veri ambarına yüklenecek olan data temizlenecek, gerekli dönüşümler yapılacak ve veriler birleştirildikten sonra veri ambarına yüklenecektir. Veri ambarına verilerin yüklenmesinden önce bu aşamaları geçirmesinin nedeni, veri ambarında bulunacak verinin temiz ve kullanıma hazır olması gerektiğindendir.

Veri ambarında toplanan veriler kullanılmaya hazır olduklarından kullanıcılar amaçlarına göre veri ambarlarından sorgulamalar yaparak analizleri için veri sağlarlar. Verinin elde edilmesinden sonraki aşama analiz aşamasıdır. Veri madenciliği araçları bu noktada işleme girer.

Veri ambarlarının kullanımı veri madenciliğinde önemli bir rol oynar. Karar destek sistemlerinin anlatıldığı bölümde veri ambarlarında yapılan veri madenciliği uygulamaları olan OLAP ve OLTP sistemlerine değinilecek ve SQL Server'da kümeleme ve karar ağacı örneği verilecektir. Şekil 3.11'de gösterilen diyagram daha ayrıntılı olarak yeniden verilecektir.

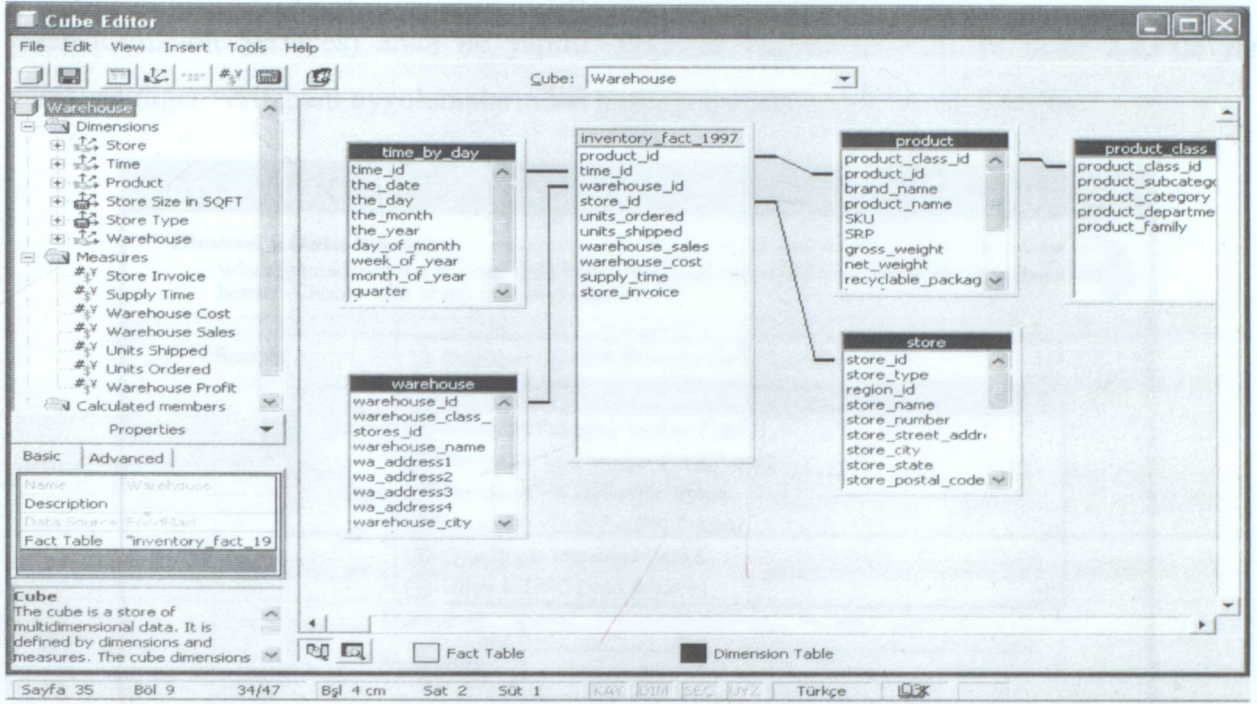
Veri ambarlarının önemli özellikleri aşağıda yer almaktadır.

- Nesne Yönelimlilik
- Entegre Olma
- Zaman Değişimlilik (Time Variant)
- Geçmişe Dönük Değişimsizlik (Non-volatile)

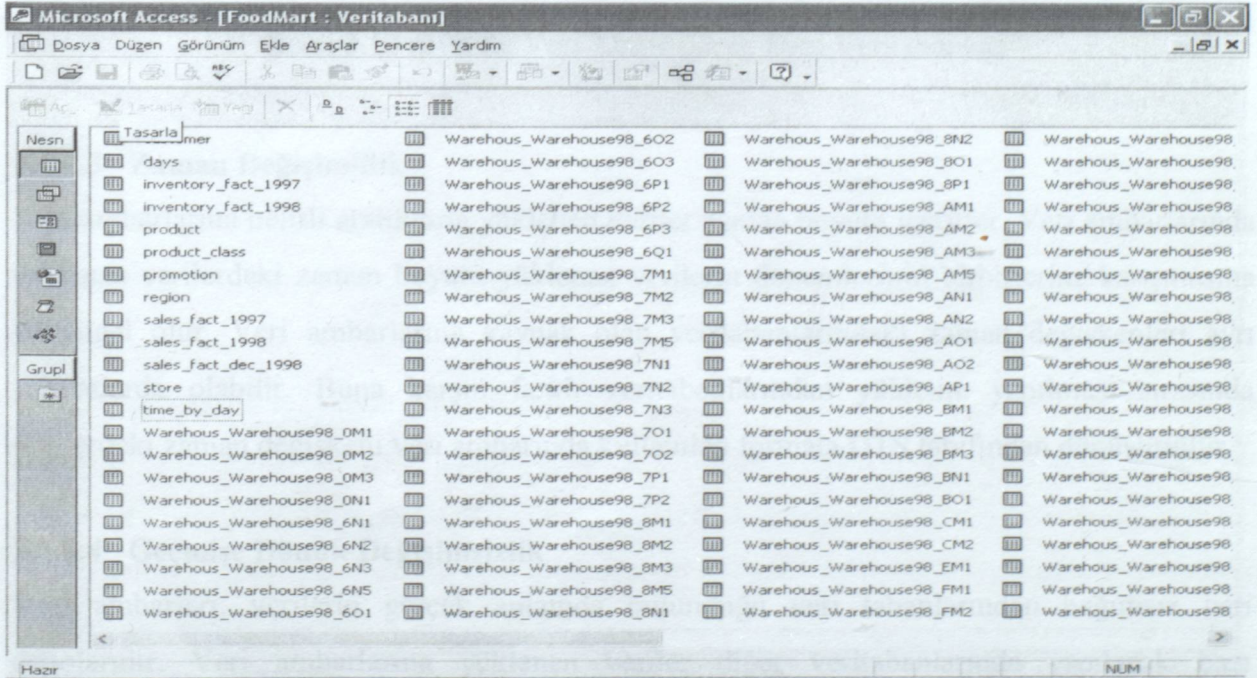
3.1.2.1 Nesne Yönelimlilik

Veri ambarları bölümünün giriş kısmında veri ambarlarında karar alma sürecinde önemli değişkenlerin saklandığına değinilmişti. Belirli aralıklarla yüklenen dataları olan veri ambarları karar alma aşamasında önemli roller oynar. Karar alınacak problem ile ilgili verilerin saklanmasında kullanılan OLAP teknolojisi ile karar aşamasında önemli olan nesnelerin kullanılması sağlanırken nesneler dışarıda tutulabilir.

Aşağıdaki şekilde bir OLAP küpü yer almaktadır. Analiz için seçilen nesneler ve bu nesnelerin boyutları kullanılırken dışlanan nesneler de kolaylıkla görülebilir. Şekil 3.11'de görülen nesnelerin Şekil 3.12'de görülen FoodMart veri ambarındaki tüm nesneler olmadığı açıktır.



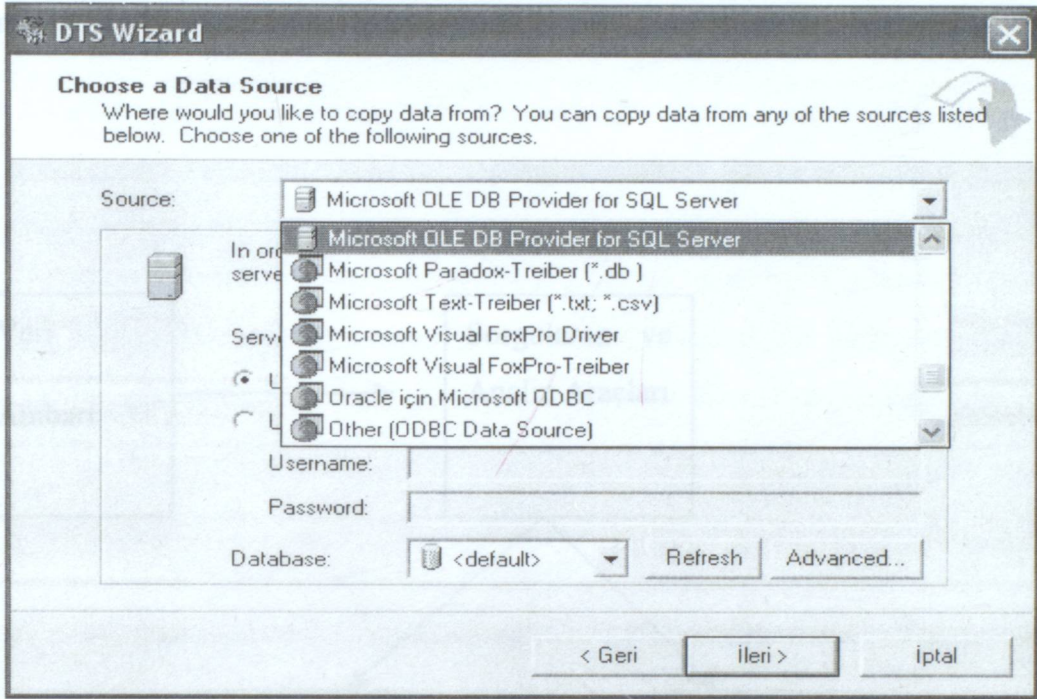
Şekil 3.11 OLAP Server’da OLAP Küpünde FoodMart Veri Ambarında Analize Dahil Edilen Nesneler



Şekil 3.12 OLAP Server’da Kullanılan FoodMart Veri Ambarına İlişkin Nesneler

3.1.2.2 Entegre Olma

Veri ambarları değişik coğrafi bölgelerde bulunan ve/veya değişik özellikleri olan veritabanlarının entegrasyonu sonucu oluşturulur. Bu entegrasyon SQL Server'da DTS (Data Transformation Services) aracı ile yapılır. DTS ile bağlanılabilecek ve Şekil 3.13'de yer almayan diğer veritabanı uygulamalarından bazıları ise Microsoft Excel, SAS'tır.



Şekil 3.13 DTS İle Veri Alınabilecek Bazı Veritabanları

3.1.2.3 Zaman Değişimlilik

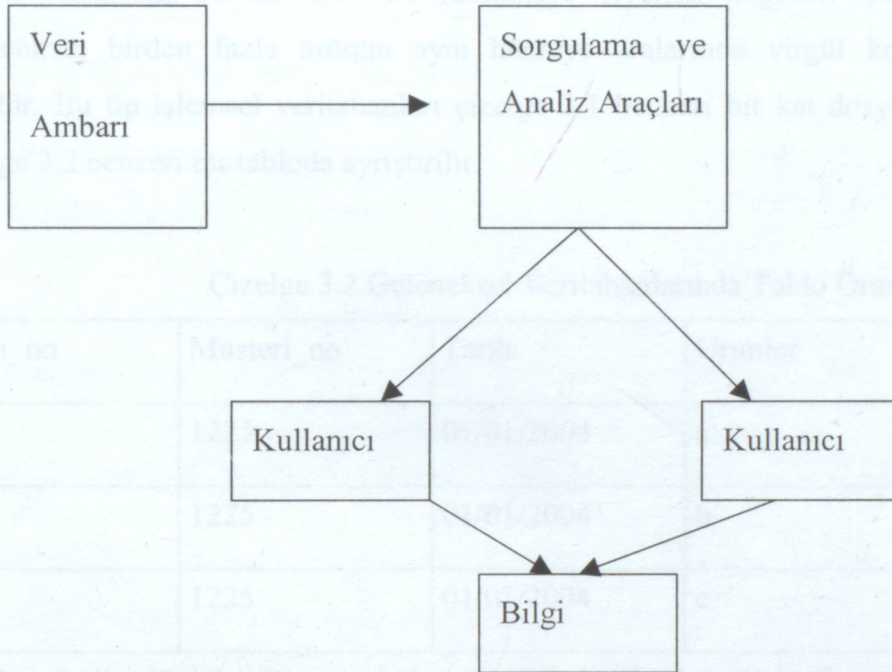
Veri ambarlarına belirli aralıklarla yüklenen veriler zaman boyutu içerirler. Veri ambarlarında saklanan verilerdeki zaman boyutu yüklenen verilerin dönemlerinin birbirlerine karışmasına da engel olur. Veri ambarlarına kaynak olan veritabanlarındaki zaman değişkenleri ayrı formatlarda olabilir. Buna karşın farklı veritabanlarından yükleme yapılması sırasında verilerdeki zaman değişkeni veri ambarında kullanılan formata DTS tarafından dönüştürülür.

3.1.2.4 Geçmişe Dönük Değişimsizlik

Veri ambarları, verilerin gerçek anlamda bulunduğu veri tabanlarından bağımsız veri depolarıdır. Veri ambarlarına yüklenen veriler diğer veritabanlarında yapılacak bazı yanlışlardan veya veri kayıplarından etkilenmezler. Fakat veri ambarları ile çalışırken çalışılacak konuyla ilgili metadata'lar oluşturulup bu datalar üzerinden çalışılmalıdır. Çünkü veri ambarına bağlanacak olan farklı kişiler tablolarda değişiklikler yapabilirler, bu yüzden

sorunlar yaşanabilir. Bununla birlikte veri ambarlarına veri yüklendikten sonra veride yeniden bir güncelleme yapılması genelde beklenmez. Bundan dolayı veri ambarına yüklenecek verinin doğruluğundan kesinlikle emin olunmalıdır.

Veri ambarlarına verilerin yüklenmesi ve analiz aşamaları aşağıdaki Şekil 3.14'de yer alan diyagramda görülebilir. Veri ambarlarının yapısına ayrıntılı olarak çok boyutlu veri modelleri konusunda yer verilecektir.



Şekil 3.14 Veri Ambarında Verilerin Toplanması ve Analizini Gösteren Diyagram

3.1.3 İşlemsel (Transactional) Veritabanları

İşlemsel veritabanlarındaki bir kayıt genellikle yalnızca bir işlemi temsil eder. Bu işlem bir müşterinin bir markette yapmış olduğu alışveriş sepeti içerisinde bulunan ürünler de olabilir, bir hastaneye giden hastaya yapılan işlemler de olabilir. İşlemsel veritabanları aşağıdaki çizelgedeki formattadır.

Çizelge 3.1 İşlemsel Veritabanı Tablo Örneği

Islem_no	Musteri_no	Tarih	Urunler	Tutar
12	1225	01/01/2004	a,b,c	12000000
13	1000	01/02/2004	a,d,e	25000000
14	1225	01/02/2004	d,e	18000000

Çizelge 3.1'deki Islem_no kolonu tek'tir (unique). Çünkü her işlem tek bir işlem numarası alabilir. Musteri_no kolonunda ise aynı kod tekrarlanabilir çünkü bir müşteri aynı ve/veya farklı tarihlerde bir veya birden fazla işlem gerçekleştirebilir. SQL ve MS ACCESS gibi birçok veritabanı bu tür bir veri saklamaya elverişli değildir. Çünkü Urunler kolonu incelenirse, birden fazla ürünün aynı hücreye aralarında virgül konularak kaydedildiği görülür. Bu tip işlemsel veritabanları çizelge 3.1 benzeri bir kat dosyasında (flat file) veya çizelge 3.2 benzeri bir tabloda ayrıştırılır.

Çizelge 3.2 Geleneksel Veritabanlarında Tablo Örneği

Islem_no	Musteri_no	Tarih	Urunler	Tutar
12	1225	01/01/2004	a	2000000
12	1225	01/01/2004	b	5000000
12	1225	01/01/2004	c	5000000

Çizelge 3.1'in 12 işlem numaralı kaydının geleneksel veritabanlarında saklanma biçimi çizelge 3.2'de görülmektedir. Çizelge 3.2'nin 12 işlem numaralı satırlarının toplamı çizelge 3.1'in 12 işlem numaralı satırının tutarına eşittir. Çizelge 3.2'de 12 numaralı işleme ilişkin Urunler kolonundaki hücrelerdeki ürünler ise Çizelge 3.1'in 12 numaralı işleminde alınan ürünler olduğu görülür. İşlemsel veritabanlarındaki verilere veri madenciliği tekniklerinden pazar sepet analizi (market basket analysis) uygulanabilir. Böylelikle tüm sepetler içerisinde gerçekleşen ürün birlikteliklerinden yola çıkılarak marketlerdeki raflarda ürünlerin yerleşim düzenlerinin belirlenmesi ya da müşteri profillerine göre ürün birliktelikleri araştırılarak müşterilere promosyonlar yapılması gibi uygulamalar yapılabilir.

3.1.4 Uzman Veritabanı Sistemleri ve Uygulamaları

Şekil 2.1’de veritabanı teknolojisinin gelişimi incelenirse uzman veritabanı sistemlerine geçişin ilişkisel veritabanı sistemlerinin iş dünyasında uzun süre kullanımından sonra olduğu görülür. Bu süre yaklaşık 10 yıl kadardır.

“Yeni veritabanı uygulamaları uzamsal veri (haritalar gibi), mühendislik dizayn verisi (binaların dizaynı, sistem bileşenleri veya entegre devreler), hipertekst ve multimedya verisi (tekst,imaj,video ve audio verisi) ve World Wide Web’i (internet tarafından ulaşılabilir kılınan büyük , geniş dağıtılmış enformasyon deposu (repository)) içerir . Bu uygulamalar kompleks nesne yapılarını kullanmak için etkin veri yapılarına ve ölçülebilir metotlara, değişken uzunluğunda kayıtlara, yarı yapılandırılmış veya yapılandırılmamış veriye, tekst ve multimedya verisine ve kompleks yapılarla ve dinamik değişimlerle veritabanı şemalarına gereksinim duyar (Han ve Kamber,2001) ”.

İşte inceleyeceğimiz uzman veritabanı sistemleri yukarıda sözü geçen gereksinimleri karşılamak üzere geliştirilmişlerdir.

3.1.4.1 Nesne Yönelimli Veritabanları

Bu veritabanlarında her bir birim (entity) bir nesne olarak kabul edilir. Northwind veritabanı örneğinde de Categories, Customers, Employees, Suppliers vs. bir nesne olabilir. Tüm nesneler birbirlerine aşağıdaki ilişkilerle bağlıdır.

- Değişkenlerin kümesi nesneyi oluşturur.
- Nesneler birbirleri ile mesajlar vasıtası ile ilişki kurarlar.
- Metotlar kümesi mesajları yerine getirmek için kodları tutarlar.

Her bir nesne sınıfının kendi içerisinde hiyerarşisi vardır. Categories sınıfı CategoryID,CategoryName,Description ve Picture alt sınıflarına ayrılır. Ve Employees tablosundaki title değişkeni ile ilgili bir sınıf oluşturulursa bu sınıf süper sınıfı olan Employee tablosundaki diğer değişkenlere ait bilgileri kullanabilir.

3.1.4.2 Nesne İlişkisel Veritabanları

Bu veritabanlarının kurulumu nesne ilişkisel modelleri baz alınarak kurulur. Bu model karmaşık nesneleri ve nesne oryantasyonunu kullanmak için zengin veri tipi sağlar. Bu veri tiplerini yönetmek için de ilişkisel sorgu dilleri kullanılır. Nesne ilişkisel veritabanları kompleks veri tiplerini , sınıf hiyerarşilerini ve nesne mirasını çok iyi kullanır.

Nesne ilişkisel ve nesne yönelimli veritabanlarında veri madenciliği benzer özellikler taşır. Her iki tip veritabanında kullanılan veri madenciliği teknikleri karmaşık nesne yapılarını , veri tiplerini , sınıf hiyerarşilerini değerlendirmek üzere geliştirilmiştir.

3.1.4.3 Uzamsal Veritabanları

Uzamsal ilişkili enformasyon içeren veritabanlarına tıp ve uydu imaj veritabanları örnek olarak gösterilebilir. Uzamsal veri raster ve vektör formatında gösterilebilir. Örneğin iki boyutlu bir uydu imajı bitmap veya pixel map'ler ile gösterilebilir. Haritalar ise vektör formatında gösterilebilir. Örneğin harita üzerinde yer alan binalar, barajlar temel geometrik şekiller olan çizgiler, noktalar, çokgenlerle gösterilebilir.

Coğrafi veritabanı uygulamaları deprem sonrası hasar belirlemesi, elektrik ve su gibi genel gereksinimlerin dağıtımı gibi durumlarda kullanılabilir.

Veri madenciliği teknikleri ise yerleşim yerlerinin nerelere kurulduğu, orman yangınlarının nerelerde meydana geldiği gibi desenleri keşfetmekte kullanılır.

3.1.4.4 Zamansal Veritabanları ve Zaman Serileri Veritabanları

Zamansal veritabanları ve zaman serileri veritabanlarının ortak özelliği zaman boyutu taşımalarıdır. Zamansal veritabanlarındaki değişkenler genellikle değişik anlamlara gelen zaman aralıklarını kullanırken zaman serileri veritabanlarındaki değişkenler ardışık ilişkileri gösteren zaman aralıklarını kullanır.

Bu veritabanlarında, veri madenciliği teknikleri kullanılarak zamana göre müşterilerin gelişleri arasında geçen süre tahmini, satışlar için geleceğe ilişkin tahminleri, dönemlere göre stokların belirlenmesi için tahminler yapılabilir. Kuruluşlar bu tahminler yoluyla stratejik kararlar alırlar ve geleceği daha iyi görebilmek için bu tahminlerden yararlanırlar.

3.1.4.5 Tekst Veritabanları ve Multimedya Veritabanları

Tekst veritabanları, incelenen özelliğe ilişkin kelimeler kullanılarak yapılan tanımlamaları içerir. Yalnız bu veritabanlarında yapılan tanımlamalar bir ya da iki kelimelik basit tanımlamalardan öte karmaşık cümleleri ve/veya paragrafları içerir. Karmaşık cümleler ve/veya paragraflardan oluşan tanımlamalar yapılandırılmamış veridir. WEB uygulamalarında ise bazı email mesajları gibi yarı yapılandırılmış, kütüphane uygulamalarında ise iyi yapılandırılmış tekst veritabanı uygulamalarına rastlanabilir.

Veri madenciliği kullanılarak cümleler ya da paragraflarda ortak olarak kullanılan kelimeler veya kelime grupları bulunabilir. Örneğin, bir doktor hastalarının şikayetlerini tuttuğu bir veritabanında bu hastalarına koymuş olduğu tanılara da yer verirse, belki de hasta şikayetleri ile tanı arasında önemli bir yapı bulunur ve hastaları dinleyerek hastaları muayene etmeden başarılı tanımlar koymak mümkün olabilir. Buna benzer uygulamalar hukuk, ekonomi, psikoloji gibi sosyal alanlarda da yapılabilir.

Multimedya veritabanlarında ise imaj, video, ses verileri saklanır. Bu tür veritabanları kullanılarak elektronik araçları ses komutlarıyla işaretler ile açıp kapatmak mümkün olabilir. Ancak bu tür verilerin saklanması için büyük ölçeklerde alana gerek vardır.

3.1.4.6 Heterojen Veritabanları ve Daha Önceden Kullanılan (Legacy) Veritabanları

Heterojen veritabanları farklı özellikte özerk veritabanı bileşenlerinden oluşur. Bu veritabanlarının bileşenleri olan veritabanları birbirlerinden oldukça farklı olabilir. Bu farklılıkları bir araya getirmek ya da bağdaştırmak oldukça güçtür.

Miras veritabanları da bir grup heterojen veritabanı sınıfına girer. Örneğin, teknolojik gelişme sonucu yeni bir veritabanı sistemine geçilirse eski veritabanı sistemindeki enformasyondan vazgeçilmez. Bu enformasyon yeni veritabanı sistemine uyarlanır. Bu durumdaki uyarlanan veritabanı daha önceden kullanılan veritabanı olarak adlandırılır.

Örneğin, iki üniversitenin not sistemi farklı olabilir. Bir üniversite harfli not sistemi kullanırken diğer bir üniversite rakamsal not sistemi kullanabilir. Aynı bölümlerin derslerinin kredi sayısı ve ders içerikleri farklı olabilir. Bu durumda iki üniversiteden gelen öğrencinin notlarının kıyaslanması oldukça güç olur. Veri madenciliği teknikleri kullanılarak bu üniversitelerin notlarının kıyaslanması mümkün olabilir. Bu durumlarda enformasyon değişimi söz konusudur.

3.1.4.7 World Wide Web

WWW dünya üzerindeki en büyük veritabanıdır diyebiliriz. Dünya üzerinde bulunan bütün web sitelerini, bu sitelerde bulunan dosyaları hatta bu dosyalar içerisinde yer alan bir kelimeye göre arama motorlarını kullanarak istenilen kaynağa ulaşmak olanaklıdır. En çok bilinen arama motorları arasında Yahoo!, MSN, Altavista, AOL, Google vardır. Bu arama motorları birbirleri ile de bağlantılı olarak çalışmaktadır ve örneğin bir arama motorunda ulaşılamayan bir bilgiye herhangi bir arama motorundan diğer arama motorlarına link kurarak ulaşmak söz konusu olabilir. Örneğin, Arabul.com arama motorunda aramanın sonucu bulunamazsa, arayıcıyı bir link ile Google arama motoruna ulaştırır.

Arama motorlarında yapılan aramalar sonucunda elde edilen sonuçlardan bir bölümü de aranılan konuyla dolaylı olarak ilgili kaynaklardır. Aranılan konuyla dolaylı olarak ilgili olan kaynakların neler olduğu, daha önceden yapılan aramalarda oluşan gizli yapılar bulunarak yapılabilir.

Elektronik ticaret de günümüzde önem kazanmıştır. Web üzerinde satış yapan amazon.com gibi web siteleri kullanıcılarına birer kullanıcı adı sağlayarak ve onlara ilişkin bazı bilgileri de alarak kullanıcı davranışlarını izlemekte ve ilginç yapıları yakalayıp kullanıcıların özelliklerine göre ürün sunmaktadırlar.

Yukarıdaki iki örnekten de görüleceği gibi veri madenciliği web uygulamalarında da önemli bir yer tutmaktadır.

3.2 Görselleştirme

Görselleştirme, veride bulunan bir değişkenin ya da değişken grubunun grafiksel olarak gösterilmesidir. Böylelikle değişkenlerin dağılımları hakkında bilgi sahibi olunur, değişken grupları arasında birtakım kümelenme, korelasyon gibi ilişkiler gözler önüne serilebilir. Değişkenlerde sapan değerlerin bulunup bulunmadığı da görselleştirme sayesinde kestirilebilir. Örneğin, bir öğrenci topluluğundaki öğrencilerin yaşları genelde 12 ile 16 arasında değişirken, 25 yaş gibi bir değer grafik gösterimde kolayca göze çarpabilir. Bir ülkenin gelir dağılımının grafiği çizilirse, dağılımın normal olup olmadığına ilişkin bilgi sahibi olunabilir. Ya da çalışanların aylık gelirleri ile yaşları arasında bir saçılım grafiği çizilirse yaş ile aylık gelir arasında bir korelasyon olup olmadığı yorumlanabilir. Coğrafi haritalar yoluyla da müşterilerin dağılımlarına göre alışveriş merkezlerinin konumlandırılması sağlanabilir. Prof. Dr. Neyran Orhunbilge, Tanımsal İstatistik Olasılık ve Olasılık Dağılımları adlı kitabında seri türlerine göre grafikleri beş grupta toplamının mümkün olduğunu

belirtmektedir.(Orhunbilge,2000)

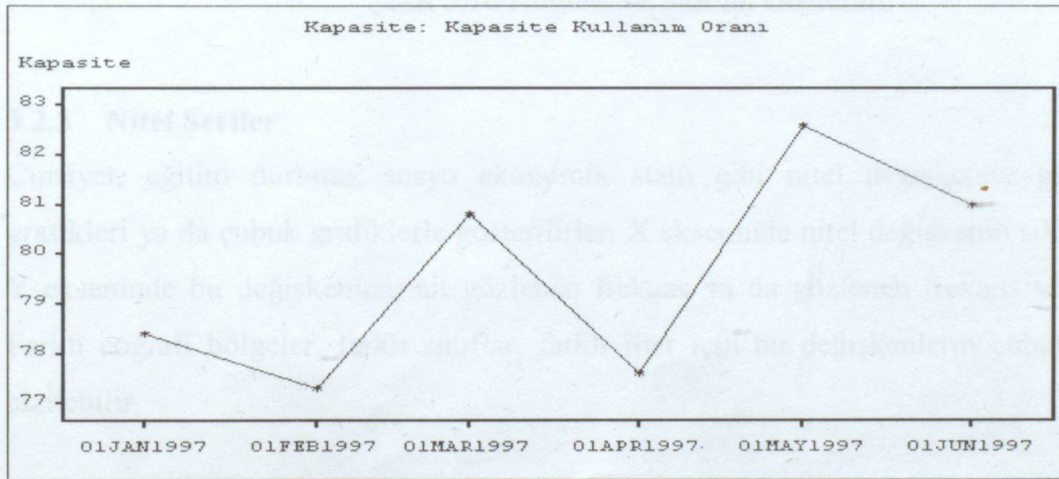
- Zaman Serileri
- Konum Serileri
- Nitel Seriler
- Nicel Seriler
- Serpilme Diyagramları

3.2.1 Zaman Serileri

Zaman serilerinin gösterimine en uygun olan grafikler çizgi grafikleridir. Zaman serileri genelde çizgi grafikleri ile gösterilse de, çubuk grafikler ile de gösterilebilir. Çizelge 3.3'te yer alan verilerin SAS programında çizilen zaman serisi grafik örneği yer almaktadır.

Çizelge 3.3 Aylık İmalat Sanayi Kapasite Kullanım Oranı (DİE,1999)

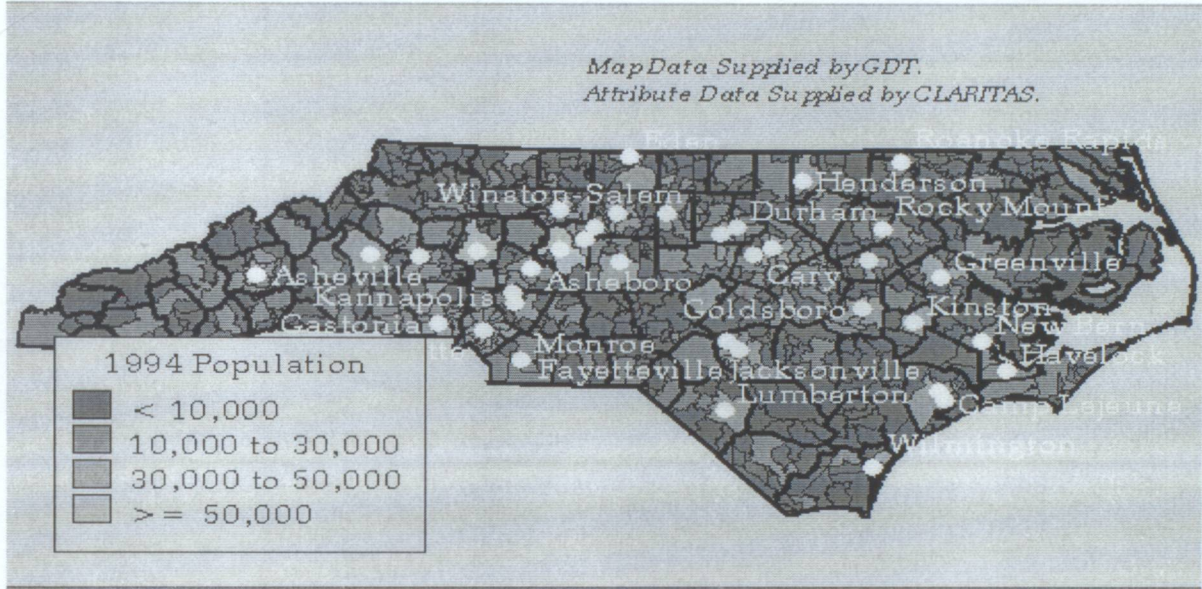
AYLAR	KULLANIM ORANI
1997 1	78,4
1997 2	77,3
1997 3	80,8
1997 4	77,6
1997 5	82,6
1997 6	81,0



Şekil 3.15 Zaman Serisi Grafiği

3.2.2 Konum Serileri

Konum serilerinde ise haritalar üzerinde incelenen olaylar resim, renk vb. kullanılarak gösterilirler. Bir ülkenin illerinin nüfus yoğunluklarının harita üzerinde gösterilmesi verilebilir. Şekil 3.16'da konum serilerinin haritalar üzerinde gösterildiği bir SAS programı çıktısı yer almaktadır..



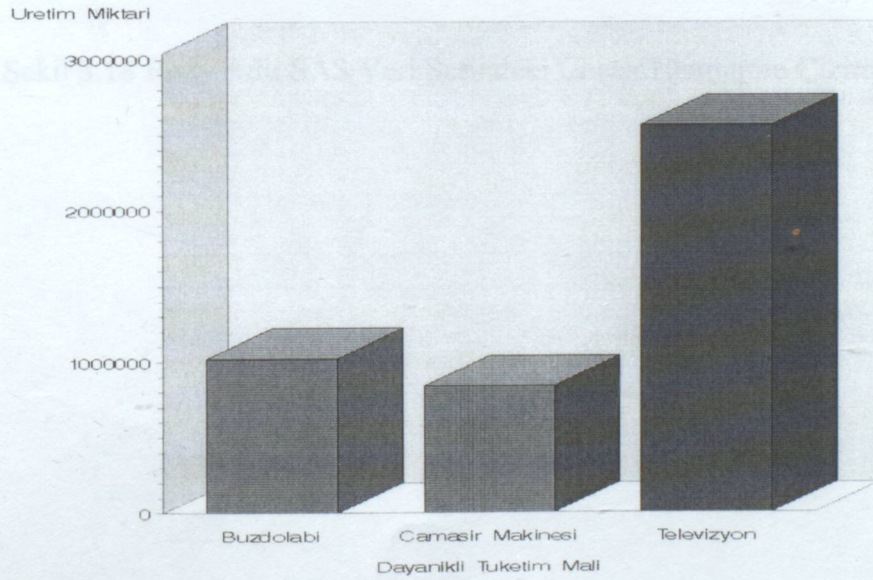
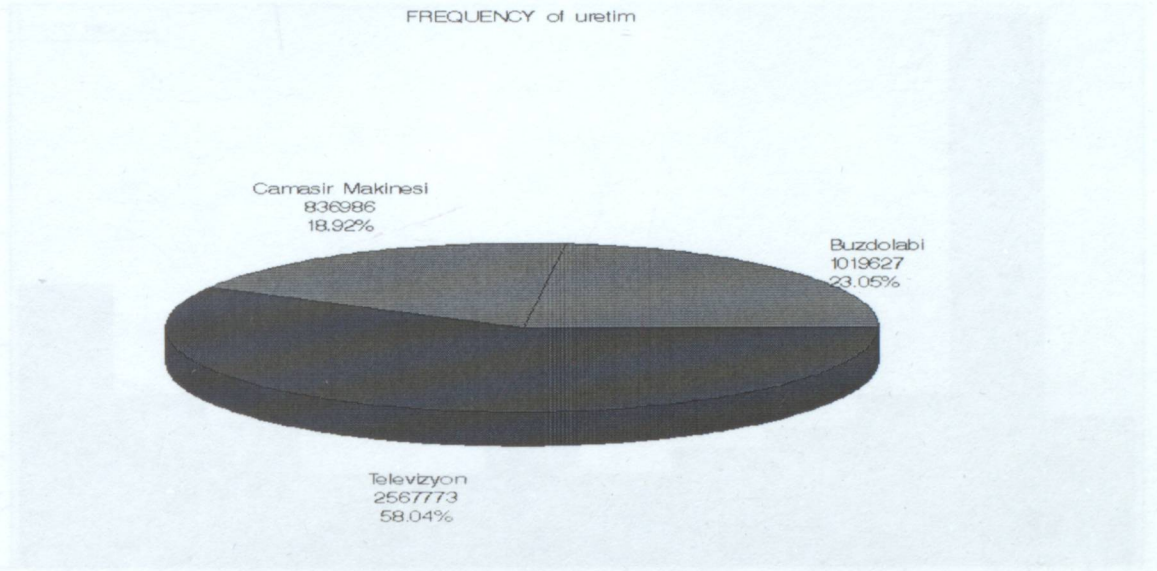
Şekil 3.16 Konum Serilerinin Gösterimi

3.2.3 Nitel Seriler

Cinsiyet, eğitim durumu, sosyo ekonomik statü gibi nitel değişkenler genellikle daire grafikleri ya da çubuk grafiklerle gösterilirler. X ekseninde nitel değişkenin sınıfları yer alırken Y ekseninde bu değişkenlere ait gözlenen frekans ya da gözlenen frekans yüzdesi yer alır. Farklı coğrafi bölgeler, farklı sınıflar, farklı iller için bu değişkenlerin çubuk grafikleri de çizilebilir.

Çizelge 3.4 1991 Yılı Seçilmiş Bazı Dayanıklı Tüketim Mallarının Üretim Sayıları(DİE,1999)

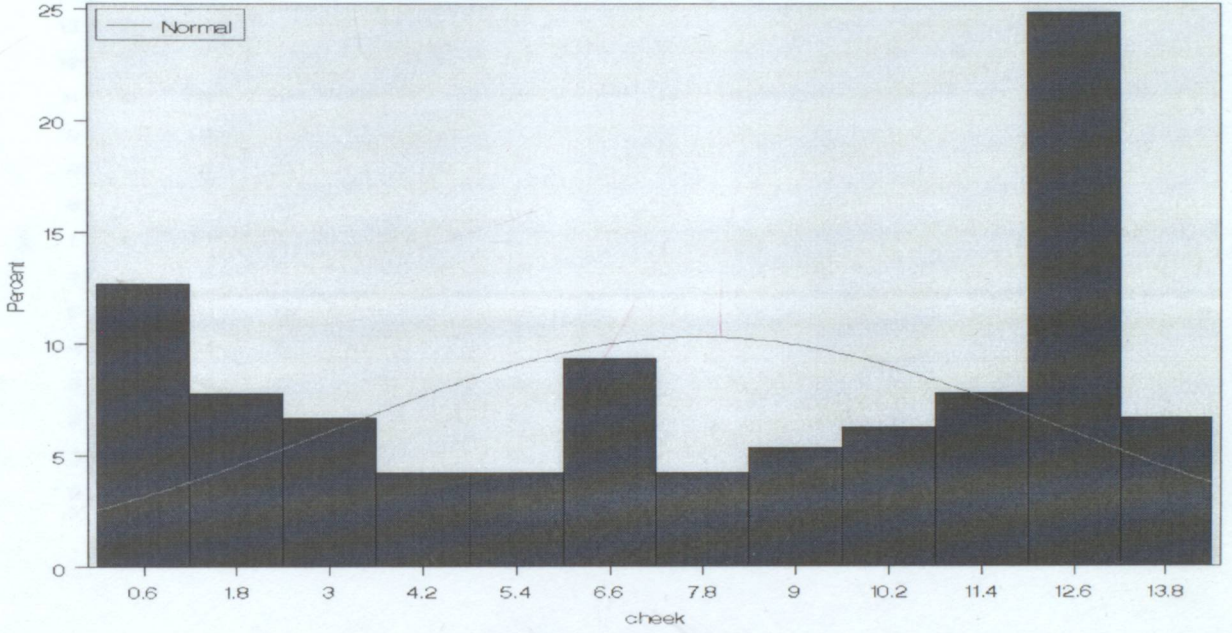
Dayanıklı Tüketim Mali	Üretim (Adet)
Çamaşır Makinesi	836986
Buzdolabı	1019627
Televizyon	2567773



Şekil 3.17 Daire ve Çubuk Grafik Örneği

3.2.4 Nicel Seriler

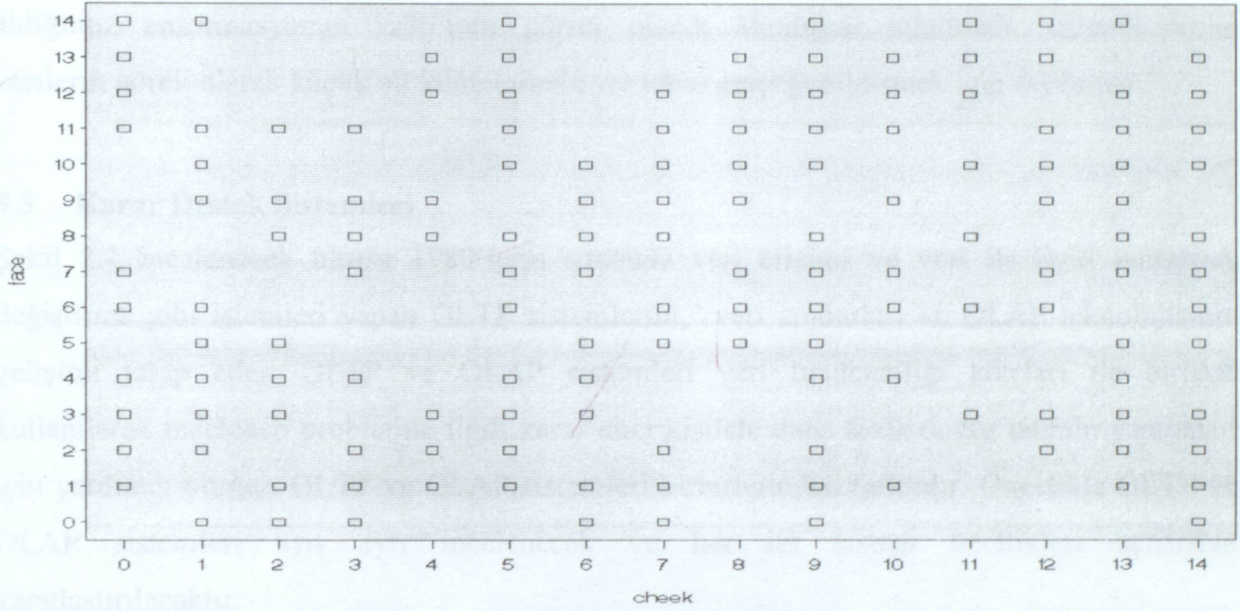
Yaş, ağırlık, kilo, nüfus gibi nicel seriler ölçeklendirilmemişler ise kartezyen grafiklerle ölçeklendirilmişler ise histogramlarla gösterilebilir. Histogramdaki sütunların orta noktaları birleştirilirse bir frekans poligonu oluşur. Histogramlar değişkenin dağılımı ile ilgili bilgi verir.



Şekil 3.18 Body Adli SAS Veri Setindeki Cheek Histogram Çizimi

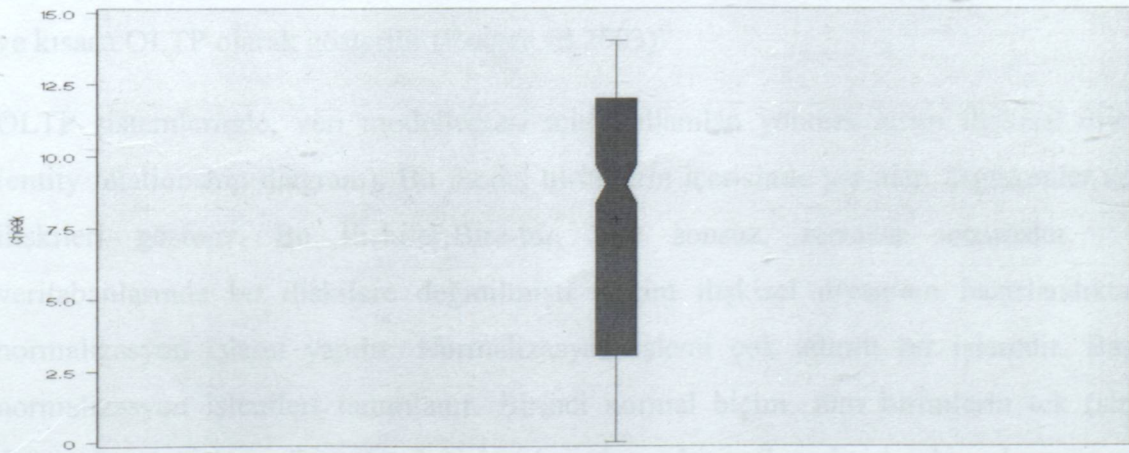
3.2.5 Serpilme Diyagramları

Serpilme diyagramları ile iki ayrı nicel özelliğin x ve y eksenlerindeki değerlerinin birbirlerine karşılık geldikleri noktalar birleştirilerek çizilen diyagramlardır. İki değişkenin arasındaki ilişki üzerinde bilgi verirler. Şekil 3.19 incelenirse face ve cheek değişkenlerinin değerleri arasında bir ilişki olmadığı görülür.



Şekil 3.19 Body Adlı SAS Veri Setindeki Cheek ve Face Değişkenlerinin Serpilme Diyagramı

Nitel ve nicel serilere ait çubuk grafikler ya da histogramlar normal dağılıma göre; sağa çarpık, sola çarpık, normal, j, ters j, u görünümünde olabilirler.



Şekil 3.20 Body Adlı SAS Veri Setindeki Cheek Değişkeninin Kutu Diyagramı

Veri madenciliğinde kullanılan görselleştirme araçlarından birisi de kutu (Box plot) diyagramlarıdır. Bu diyagramlarda verilerin ortalamaları etrafındaki dağılımları hakkında fikir sahibi olmamızı sağlar ve sapan değerleri gösterir.

Görselleştirmenin önemine Discovering Data Mining (Cabena vd,1997) adlı kitapta şöyle değinilmektedir. “Görselleştirme teknikleri veride gizli olarak bulunan yapıların tanımlanması için çok güçlü araçlardandır. Bu bir sürpriz değildir, Psikologlar uzun çalışmalar sonucu aldığımız enformasyonun %80’inin görsel olarak alındığını tanımladı. Görselleştirme verilerin görece olarak küçük alt kümelerinde yer tutan gerçeği bildirmek için faydalıdır.”

3.3 Karar Destek Sistemleri

Şekil 2.1 incelenecek olursa 1980’lerin sonunda veri erişimi ve veri ile ilgili kurtarma, değiştirme gibi işlemleri yapan OLTP sistemlerini, veri ambarları ve OLAP teknolojisinin gelişimi takip eder. OLTP ve OLAP sistemleri veri madenciliği araçları ile birlikte kullanılarak incelenen problemle ilgili karar alıcı kişilere daha fazla doğru tahmin yapmaları için yardımcı olurlar. OLTP ve OLAP sistemleri birbirlerinden farklıdır. Öncelikle OLTP ve OLAP sistemleri ayrı ayrı incelenecek ve her iki sistem özellikleri açısından karşılaştırılacaktır.

3.3.1 OLTP Sistemleri

Tarihsel gelişimde ilk olarak ortaya çıkan karar destek sistemleri OLTP sistemleridir. Veritabanlarının anlatıldığı bölümde işlemsel veri tabanlarına değinilmişti. “Operasyonel veya işlemsel veritabanları bireysel işlemleri hızlı ve etkin bir şekilde kaydetmek üzere tasarlanmışlardır. Bu tip işlem tabanlı etkileşim on-line transactional processing olarak bilinir ve kısaca OLTP olarak gösterilir.(Roiger vd,2003)”

OLTP sistemlerinde, veri modellemesi için kullanılan yöntem birim ilişkisel diyagramdır (entity relationship diagram). Bu model birimlerin içerisinde yer alan değişkenler arasındaki ilişkileri gösterir. Bu ilişkiler;Bire-bir, bire sonsuz, sonsuza sonsuzdur. İlişkisel veritabanlarında bu ilişkilere değinilmişti. Birim ilişkisel diyagram hazırlandıktan sonra normalizasyon işlemi yapılır. Normalizasyon işlemi çok adımlı bir işlemdir. Başlangıçta normalizasyon işlemleri tanımlanır. Birinci normal biçim, tüm birimlerin tek (single) bir değer alan değişkene (bir adet değişkenden oluşan birincil anahtar) sahip olmasını gerektirir. İkinci normal biçim ise birincil anahtarı birden fazla değişkenin oluşturduğu durumdur. Bu durumda tüm anahtar olmayan değişkenler, birincil anahtarı oluşturan değişkenlerin sadece

birisine bağıdır yani birincil anahtarı oluşturan değişkenlerden biri tektir (unique). Üçüncü normal biçim ise , birim ikinci normal biçimi sağlıyorsa ve diğer tüm değişkenler birincil anahtarı oluşturan değişkenlerin tamamına bağı ise gerçekleşir yani birincil anahtarı oluşturan iki veya daha fazla değişkenin her biri tek ise üçüncü normal biçim gerçekleşir. Aşağıdaki çizelgelerde sırası ile birinci normal biçim, ikinci normal biçim ve üçüncü normal biçim yer almaktadır.

Birinci normal biçimi gösteren çizelge 3.5’de Ürün Kodu değişkeni birincil anahtardır. Ürün tipi ve fiyat değişkenleri bu anahtara bağıdır.

İkinci normal biçimi gösteren çizelge 3.6’da ise birincil anahtarı ürün kodu ve müşteri kodu oluşturmaktadır. Bu değişkenlerden müşteri kodu tek değer almaktadır. Ürün kodu ise bir değeri birden fazla alabilmektedir. Fakat her iki değişkenin birimlerinin aynı kodu alması mümkün değildir.

Üçüncü normal biçimi çizelge 3.7’de ise birincil anahtarı oluşturan her iki değişkende tek değer almaktadır.

Eğer çizelge 3.7’de birincil anahtar tanımlanmazsa yani çizelge de birincil anahtar olarak gösterilen değişkenlerin aldıkları değerlerde aynı olursa bu durumda denormalizasyon gerçekleşmiş olur. Bu durumda ise ilişkisel birimlerin kombine edilmesi karar destek sistemleri için veri hazırlanması ihtiyacı karşılanmış olur.

Çizelge 3.5 Birinci Normal Biçim

Ürün Kodu (Birincil Anahtar)	Ürün Tipi	Fiyat
1	A	500000
2	B	700000
3	C	800000
4	A	500000

Çizelge 3.6 İkinci Normal Biçim

Ürün Kodu (Birincil Anahtar)	Ürün Tipi	Fiyat	Müşteri Kodu (Birincil Anahtar)	SES
1	A	500000	M1	AB
2	B	700000	M2	C1
3	C	800000	M3	C2
1	A	500000	M4	AB
2	B	700000	M5	D

Çizelge 3.7 Üçüncü Normal Biçim

Ürün Kodu (Birincil Anahtar)	Ürün Tipi	Fiyat	Müşteri Kodu (Birincil Anahtar)	SES
1	A	500000	M1	AB
2	B	700000	M2	C1
3	C	800000	M3	C2
4	A	500000	M4	AB
5	B	700000	M5	D

OLTP sistemlerinin bazı karakteristik özellikleri ise aşağıda yer almaktadır.

- Gerçekleşen işlemleri (transaction) eş anlı olarak izlemek üzere kurulmuşlardır.
- Veriler güncellenebilir (update).
- Yeni veriler ilave edilebilir.
- Bu sistemler veri girişine , değiştirilmesine, yüklenmesi esaslarına göre çalıştıklarından tüm iş ile ilgili süreçlerin gerçekleşmesi OLTP sistemlerinin performansına bağlıdır.
- Dalgalanmalı bir yapıya sahiptir. Her bir yeni satır eklenmesi ya da güncellenmesi sonucunda ortaya tutarsız raporlar çıkabilir.
- En son durumda gerçekleşen veri ile ilgili rapor sağlarlar.

OLTP sistemlerinin kullanıldığı bazı oluşumlar arasında

- Hanehalkı tüketim panelleri
- Perakende panelleri
- Bankalar
- Marketler
- Yayın evleri yer alır.

3.3.2 OLAP Sistemleri

“OLAP çok boyutlu çevrede (environment) veri analizini destekleyen sorgu bazlı bir metottur. OLAP insanlar tarafından üretilen hipotezleri doğrulamak veya yalanlamak ve manuel veri madenciliği gerçekleştirmek için önemli bir araçtır.(Roiger vd,2003)”

“OLAP ilişkisel ve iki boyutlu veritabanlarında depolanan enformasyondan çok boyutlu veri küpü kurmaya olanak sağlayan bir teknolojidir. Kullanıcılar veriyi gerçek dünyanın karmaşık karar problemlerine cevap vermek üzere kullanabilir. OLAP paylaşılmış çok boyutlu enformasyonun hızlı olarak analiz edilmesine de olanak sağlar.(Aldana,2000)”

Yukarıdaki iki tanım OLAP sistemlerinin temel amacını açıklamaktadır. Bu amaç, verilerin depolandığı veritabanı içerisinden gerçekte var olduğu düşünülen olayların ya da yapıların gerçekleşip gerçekleşmediğini kontrol etmeye aracı olur. Buna ek olarak bir şirket kendi ürünlerinin belirlediği periyotlarda (ay,çeyrek,hafta, vb dilimler) hangi bayilerinin hangi ürünlerini ne kadar sattığını, ne kadar sipariş verdiğini ve buna benzer iş ile ilgili sorulara OLAP sistemleri aracılığıyla cevap bulabilir. Bir önceki örnekteki şirketin en çok satış yapan bayisi İstanbul bayisi olsun, bu şirket en son verileri yüklenen dönemde bu bayinin satışlarının yine aynı oranda olup olmadığını kuracağı OLAP küpü aracılığıyla inceleyebilir. Görsel olarak bir farklılık (azalma ya da artma) söz konusu ise OLAP küpünde görselleştirdiği veriyi istatistik analizler kullanarak iki dönem arasındaki farklılığın anlamlı olup olmadığını test etmek amacı ile kullanabilir.

OLAP küpleri çok fazla ayrıntılı olarak OLTP sistemlerinde bulunan verilerin veri ambarlarına alındıktan sonraki aşamada alınacak karar için önemli boyutları kullanarak özetleme yapar. Örneğin, bir yıl boyunca günlük bazda yapılan satışlar aylık, 3 aylık , 6 aylık ya da bir yıllık periyotlarda ya da spesifik tanımlanmış periyotlarda incelenerek değerlendirme yapılabilir.

OLAP küpleri finansal raporlama, kalite izleme, karlılık analizleri, pazarlama analizleri gibi konularda faydalı olabilir. OLAP küpleri kurumsal düzeyde verileri depolamak amacıyla veri ambarlama tekniklerini kullanır. OLAP sistemleri web tabanlı uygulamaları da içerirler. Bu uygulamalar yoluyla yöneticiler istedikleri raporlara web bağlantısı ya da şirket içi ağ bağlantısı kurarak bağlanabilirler.

OLAP ile ilgili tanımlarda çok boyutlu verilerden söz edilmiştir. “Veri ambarları ve OLAP araçları çok boyutlu veri modeli üzerine kurulur. Bu model veriyi veri küpü biçiminde görüntüler.,(Han ve Kamber,2001)”

3.3.2.1 Veri Ambarı Verisinin Analizi

Veri ambarında bulunan verilerin analizi üç aşama da yapılır.

- Verilerin Raporlanması: Verilerden belirli değişkenler temel alınarak açıklayıcı raporlar hazırlanır.
- Verilerin Analizi: Veri analizi çok boyutlu veri analizi aracı ile yapılır. OLAP çok boyutlu bir veri analizi aracıdır.
- Bilgi Keşfi: OLAP küplerinin ya da tabloların kullanılarak veri madenciliği çalışması yapılmasıdır.

3.3.2.2 Çok Boyutlu Veri Modeli

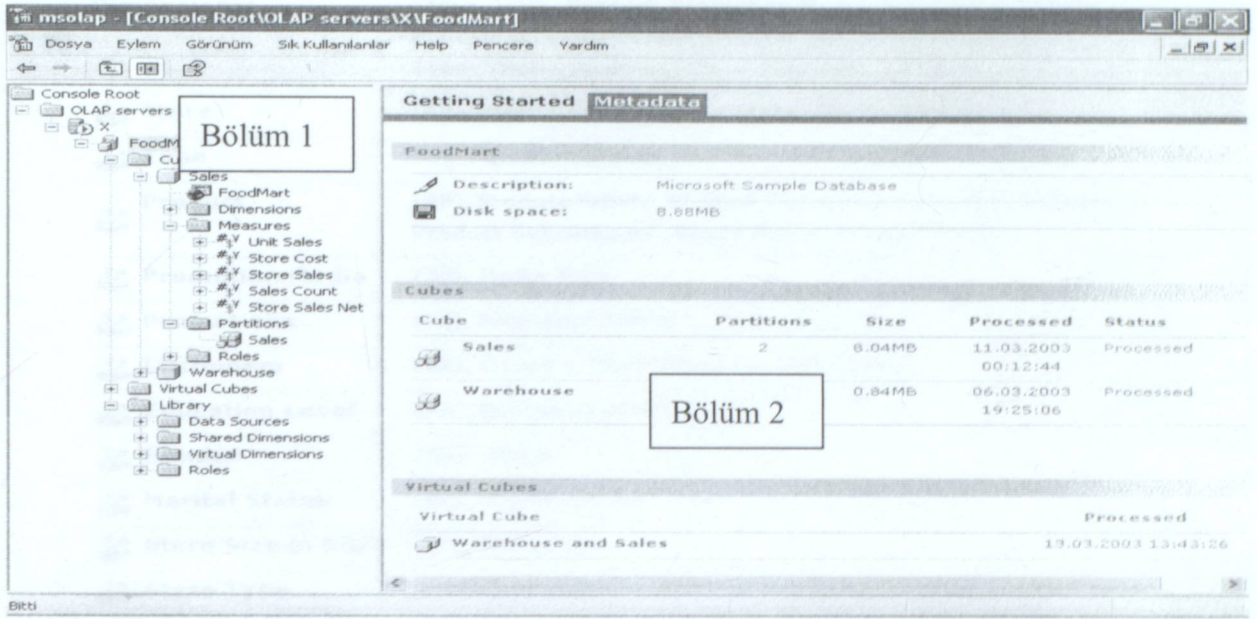
Veri ambarlarında farklı tablolar biçiminde bulunan veriler OLAP küpleri yoluyla çok boyutlu küpler biçiminde görüntülenebilir. Şekil 3.21’de SQL SERVER 97’de hazırlanmış olan Foodmart veri ambarı kullanılarak hazırlanan küplerin yer aldığı bir OLAP SERVER konsolu yer almaktadır. Bu konsolun birinci bölümünde, bir veri ambarı kullanılarak oluşturulan küpler ve bu küplerin bileşenleri, görsel küpler (var olan küplerin kombinasyonu ile oluşturulmuş küpler) ve içerisinde veri kaynaklarının, paylaşılan boyutların, görsel boyutlar ve roller yer alır. İkinci bölümde ise konsolun birinci bölümünde seçilen bileşen ait özellikler yer alır.

Konsoldaki ilk hiyerarşik bölünüm veri ambarlarıdır. Her yeni veri ambarı eklenişinde o ambara ilişkin alt bileşenler olan Küpler (Cubes), Görsel küpler (Virtual Cubes) ve Kütüphane (Library) bileşenleri o veri ambarının içerisine yerleşir.

Görsel küpler veri ambarındaki nesneler aracılığıyla hazırlanmış olan küplerin birleştirilmesi ile oluşturulan küplerdir.

Kütüphane ise, Veri Kaynaklarının (Data Sources), Paylaşılan Boyutların (Shared Dimensions), Görsel Boyutların (Virtual Dimensions) ve Rollerin (Roles) yer aldığı birimdir. Veri kaynakları server tarafından kullanılan veri ambarlarıdır. Paylaşılan boyutlar, oluşturulan küplerde ortak olarak kullanılan boyutlardır. Görsel boyutlar ise görsel küplerde kullanılan

boyutlardır.



Şekil 3.21 Sales Veri Küpü Bileşenleri

Oluşturulan küplere ait bileşenler ise konumuz içerisinde ayrıntılı olarak incelenecektir.

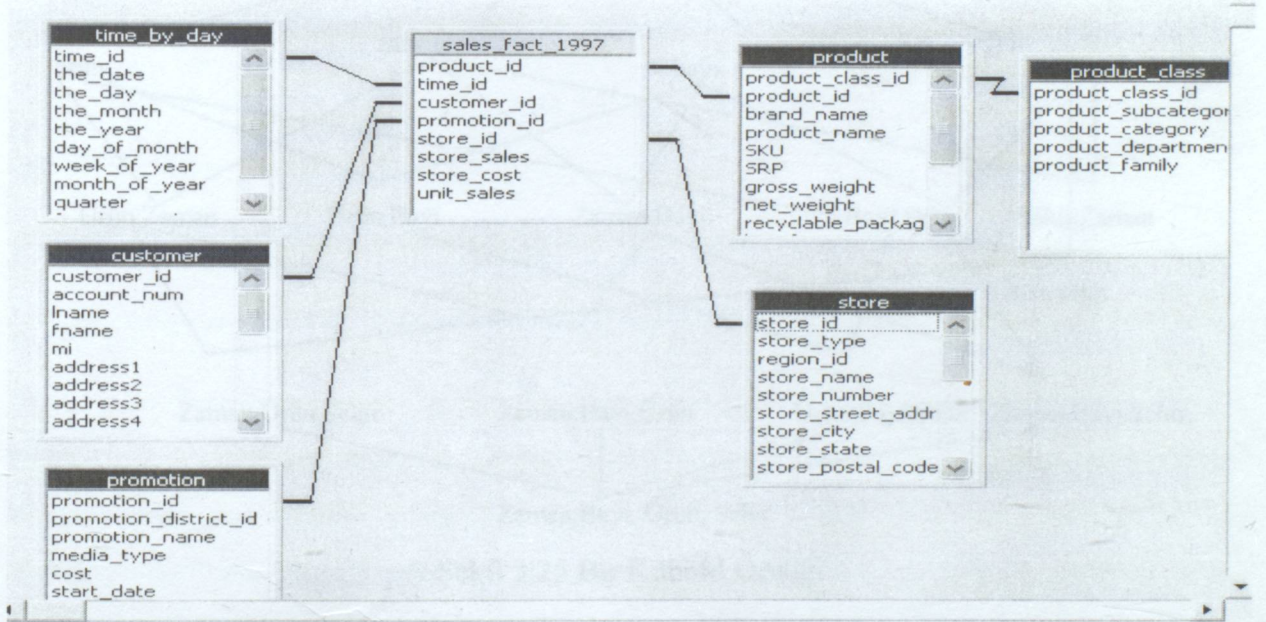
3.3.2.2.1 Bir Küpün Boyutları

Bir küp oluşturulurken bir ana veri tablosu kullanılır. Bu veri tablosu küpün boyutları aracılığı ile özetlenecek olan ölçü birimlerinin bulunduğu olay (fact) tablosudur. Olay tablosu boyut tabloları ile ilişkilendirilir. Boyut tabloları zaman, ürün, bayi, bölge, şehir gibi nesnelerden oluşur. Her bir boyut tablosu kendi içerisinde düzeylere ayrılır. Örneğin, ürün tablosunun düzeyleri Gazlı Meşrubatlar, Sıvı Yağlar, Bisküviler, Buzdolabı, Çamaşır Makinesi vb. düzeylere ayrılabilceği gibi Hızlı Tüketim Malları, Beyaz Eşya gibi düzeylere de ayrılabilir. Bu düzeyler ise boyut tablolarının bir kolonunu oluşturur. Boyutlar raporlamanın zenginleşmesini sağlarlar. Aşağıdaki şekillerden Şekil 3.22’de Sales küpüne ait boyutlar ve düzeyleri ve Şekil 3.23’de Sales küpünün olay tablosu ve olay tablosu ile ilişkili boyut tabloları görülmektedir.

Sales

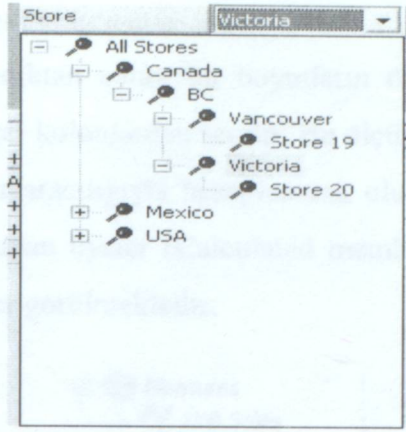
Dimensions:	Store, Time, Product, Promotion Media, Promotions, Customers, Education Level, Gender, Marital Status, Store Size in SQFT, Store Type, Yearly Income
 Store	(All), Store Country, Store State, Store City, Store Name
 Time	Year, Quarter, Month
 Product	(All), Product Family, Product Department, Product Category, Product Subcategory, Brand Name, Product Name
 Promotion Media	(All), Media Type
 Promotions	(All), Promotion Name
 Customers	(All), Country, State Province, City, Name
 Education Level	(All), Education Level
 Gender	(All), Gender
 Marital Status	(All), Marital Status
 Store Size in SQFT	(All), Store Sqft
 Store Type	(All), Store Type
 Yearly Income	(All), Yearly Income

Şekil 3.22 Sales Kúpüne İlişkin Boyutlar ve Boyutların Düzeyleri



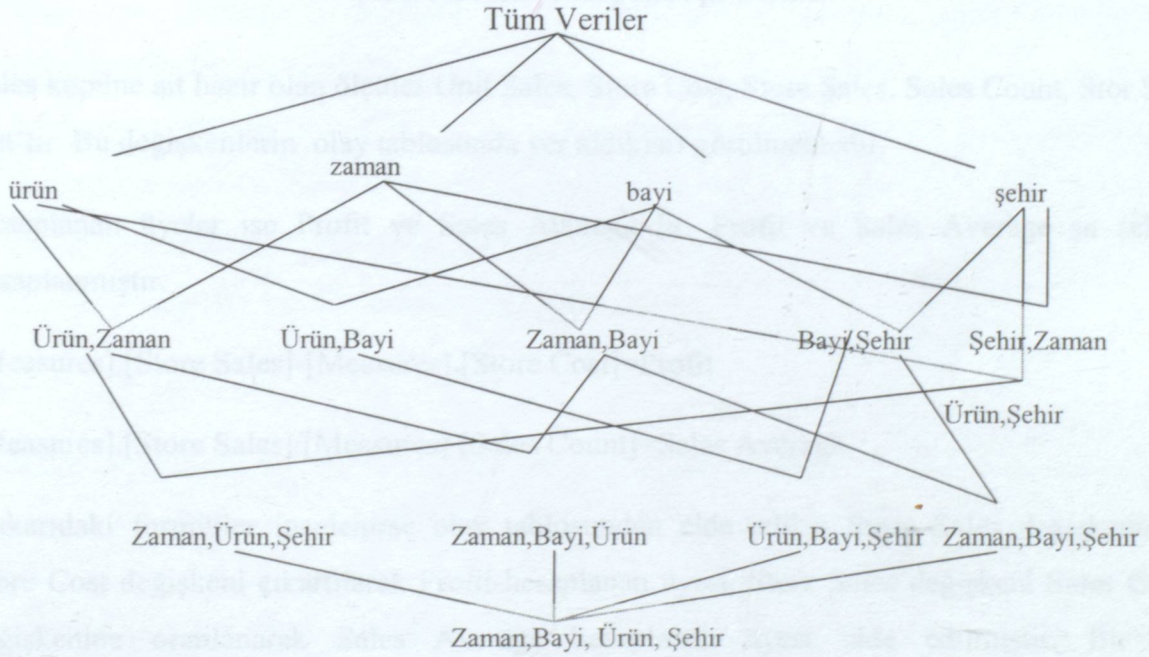
Şekil 3.23 Sales Kúpüne İlişkin Olay Tablosu ve Boyut Tabloları

OLAP küpü hazırlanırken kullanılan boyutlar, boyut tablolarının birer kolonları olduğuna değinilmişti. Şekil 3.22 incelenirse, Store'un bir boyut olduğu görülür. Store boyutu, Store boyut tablosundaki Store Country, Store State, Store City, Store Name düzeylerinden oluşturulmuştur. Şekil 3.24 ise bu düzeylerdeki hiyerarşik yapı görüntülenmektedir.



Şekil 3.24 Bir Boyut ve Hiyerarşik Yapısı

Olay tablosunda yer alan ölçüler 1,2,3,4 boyutlu küplerde özetlenebilirler. Şekil 3.25’de 1, 2, 3, 4 boyutlu olarak gösterilen küboidler yer almaktadır. Her bir küboid kendi boyutu kadar boyut tablosundan oluşur. Her bir küboid yapılan özetlemeye ilişkin bir derecedir.

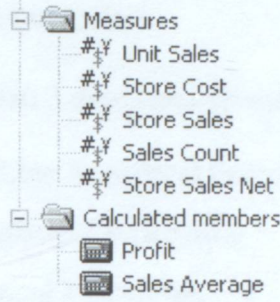


Şekil 3.25 Bir Küboid Örneği

Şekil 3.25’de yer alan küboidler bize 4 değişken üzerinde oluşturulabilecek her bir kübü göstermektedir. Oluşturulan OLAP küplerinin görünümleri de doğal olarak üç boyutlu değildir. Aşağıdaki şekillerde 1 boyutlu, 2 boyutlu ve 3 boyutlu küp örnekleri yer almaktadır. 4 boyut değişkeninin 1 boyutlu oluşturabilecekleri küp sayısı 4’ün 1’li kombinasyonları kadardır.

3.3.2.2.2 Bir K p n  l  leri

Bir k p n boyutları tanımlandıktan sonra bu boyutların  zet raporlarını olu turacak olan  l  ler olay tablosundaki integer kolonlardan se ilir. Bu  l  ler olay tablosundaki de i kenler olabilece i gibi bu de i kenler aracılı ıyla hesaplanarak olu turulabilen yeni de i kenler de olabilir. Bu  l  ler de hesaplanan  yeler (Calculated members) olarak adlandırılırlar.  ekil 3.26'da Sales k p ne ait  l  ler g r lmektedir.



 ekil 3.26 Sales K p ne Ait  l  ler

Sales k p ne ait hazır olan  l  ler Unit Sales, Store Cost, Store Sales, Sales Count, Stor Sales Net'tir. Bu de i kenlerin olay tablosunda yer aldıkları g r lmektedir.

Hesaplanan  yeler ise Profit ve Sales Average'dır. Profit ve Sales Average  u  ekilde hesaplanmı tır.

$$[\text{Measures}].[\text{Store Sales}] - [\text{Measures}].[\text{Store Cost}] = \text{Profit}$$

$$[\text{Measures}].[\text{Store Sales}] / [\text{Measures}].[\text{Sales Count}] = \text{Sales Average}$$

Yukarıdaki form ller incelenirse olay tablosundan elde edilen Store Sales de i keninden Store Cost de i keni  ıkarılarak Profit hesaplanan  yesi, Store Sales de i keni Sales Count de i kenine oranlanarak Sales Average hesaplanan  yesi elde edilmi tir. Bir k p  olu turulurken olay tablosundaki de i kenlerin elverdi i oranda yeni hesaplanabilir  yeler elde edilebilir.

3.3.2.2.3 Bir K p n B l mleri (Partitions)

Bir k p n b l mlere ayrılması o k p  i in analiz aracında bir depolama biriminin ayrılmasıdır. Olu turulan standart (default) bir k p  1 b l mden olu ur.  rne in, bir k p  zamana g re b l nebilir. 2001-2003, 1999-2000, 1998 ve  ncesi. Genellikle sorgular da yeni d nemler kullanılır. K p  olu turulurken yapılan b lmeleme ile sorgunun yapılaca ı b l mlerdeki veri sayısı azaldı ından, i lemler daha hızlı olacaktır.

3.3.2.2.3.1 K p n B lmelere Ayrılmasının Faydaları

- Tek bir b lmeleme i in sorgu yapılırken sorgu performansını deęerlendirmek kolaylařır.
- Eęer b lmeler farklı bilgisayarlardaysa b lmelerde yapılacak iřlemeler ve sorgular bilgisayarların iřlemci , bellek gibi birimlerinden daha iyi bir řekilde yararlanılabilir.
- Birbirlerinden farklı  zellikleri olan b lmelerde yapılacak sorgular iřlemleri daha da hızlandırabilir.

3.3.2.2.3.2 K p n Depolanması Y ntemleri

Bir k p n depolanması i in kullanılan 3 depolama y ntemi vardır.

- MOLAP : Mutidimensional On-Line Analytical Processing
- ROLAP: Relational On-Line Analytical Processing
- HOLAP: Hybrid On-Line Analytical Processing

Depolama y ntemlerinin ortak  zellikleri:

- 3 depolama y ntemi de gr plama dizaynı (aggregation), gr plama inřası, b lmeleme ve boyut inřası elementlerini i erir.
- 3 depolama y ntemi de boř gr plamaları kullanmaz.  rneęin, 6-12 yař grubu  ocuklar i ki alıřveriři yapmazlar. Bu grup i in hi bir disk depolaması yapılmaz.
- OLAP veritabanlarının hepsi veri patlamasından ka ınır.
- Her    depolama y ntemi de boyut yapılarını saklar.

3.3.2.2.3.2.1 MOLAP

MOLAP depolama y ntemi detaylı veri ve gr plamaların analiz server'da depolanması s z konusu olduęunda tercih edilir. Performans olarak en hızlı depolama teknięidir. Analiz server'ında olay tablosundan detaylı verinin bir kopyası vardır.MOLAP etkin veri sıkıřtırma y ntemleri kullanır. Bu y zden veri depolaması a ısından daha ekonomiktir. Yapılan sorgulama iliřkisel veritabanı sisteminde yapılmaz, analiz server'da yapılır. Bu y zden iliřkisel veritabanı kapansa da iřlemler yapılabilir.

3.3.2.2.3.2.2 ROLAP

MOLAP depolama y ntemi detaylı veri ve gr plamaların iliřkisel veritabanında depolanması s z konusu olduęunda tercih edilir. Analiz server'da yalnızca boyutsal hiyerarřiler kullanılır. Analiz server'da saklanan hiyerarřiler varsayılan ROLAP hiyerarřileri deęildir. Depolama iliřkisel veritabanında yapıldıęından, t m sorgular bu veritabanına ulařmak zorundadır. Bu

yüzden bu depolama yöntemi MOLAP depolama yöntemine oranla daha yavaştır. Analiz Server , ilişkisel veri tabanı ile kullanıcı arasında çok boyutlu ortalama bir bağlantı sağlar.

3.3.2.2.3.3 HOLAP

Detaylı veri ilişkisel veritabanlarında saklandığında ve gruplamalar çok boyutlu veri formatında saklandığında kullanılır. HOLAP disk depolamada yer tutma ve performans açısından ROLAP ve MOLAP arasındadır. Detaylı verinin kopyası oluşturulmadığından dolayı kullanılabilir disk oranı düşük olduğunda tercih edilebilir bir uygulamadır. Sorgu hızı diğer iki yöntem arasındadır. HOLAP ve MOLAP küplerinde işlem süresi yakındır.

3.3.2.2.4 Bir Küpte Kullanıcı Roller

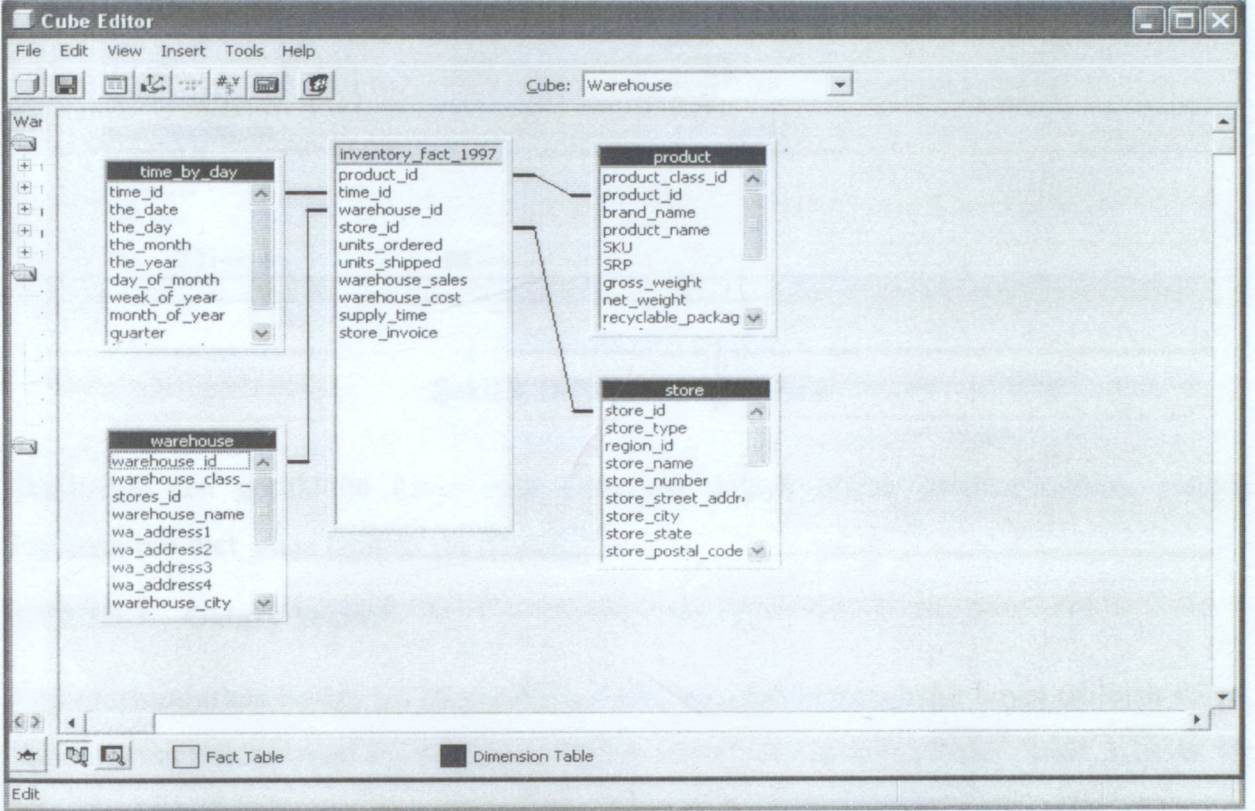
Kullanıcıların küp üzerinde işlem yapmaları ile ilgili rollerdir. Kimi kullanıcılar küp üzerinde değişiklik yapma, yenileme gibi işlemler yapma yetkisini ellerinde bulundururken kimi kullanıcılar yalnızca sorgu yapma yetkisini ellerinde bulundurlar. Veritabanı yetkilisi (Administrator) yetkileri kullanıcılar veren kişidir.

3.3.2.2.5 Çok Boyutlu Veritabanlarında Kullanılan Star, Snowflakes, Fact Constellements Şemaları

OLAP küpleri kurulurken kullanılan bu şemalar birim ilişkisel veri modelini kullanır ve nesne yönelimlidirler. Bu özellikleri ile on-line veri analizini gerçekleştirme de başarılıdırlar. Veri ambarı için en popüler veri modeli çok boyutlu veri modelidir. Bu veri modeli 3 şema ile gösterilir. Bu şemalar Stars, Snowflakes, Fact Constellements (Galaxy) şemasıdır.

3.3.2.2.5.1 Star Şeması

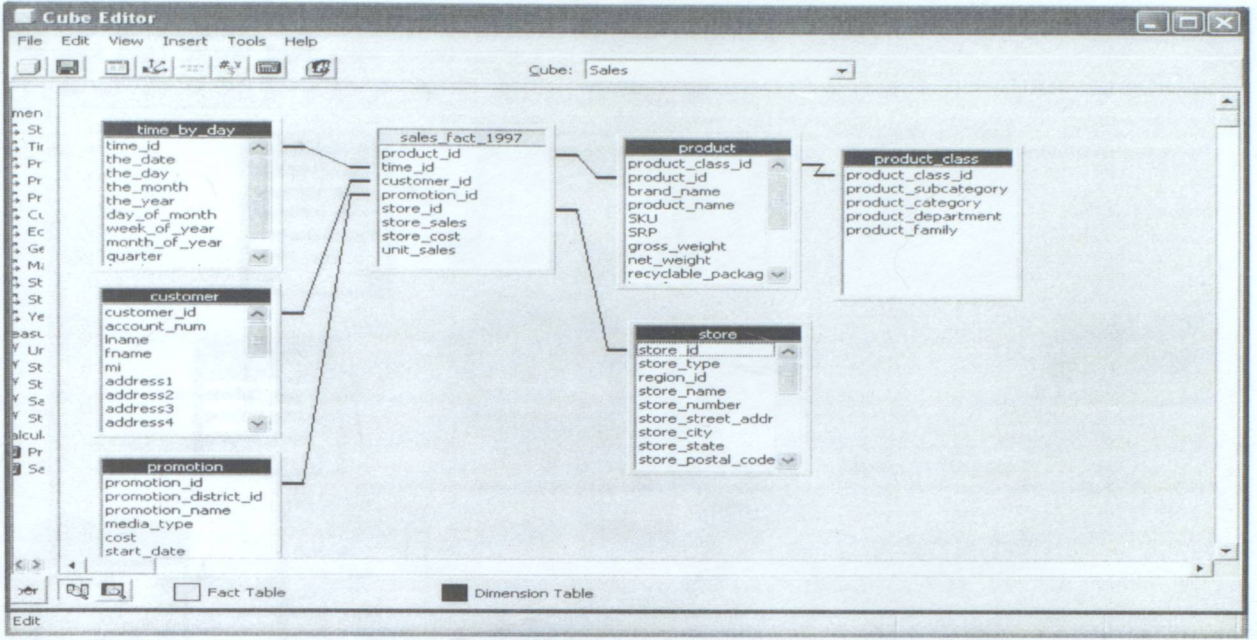
Bu şema, şeklinin yıldıza benzetilmesi nedeniyle bu adı almıştır. Merkezde olay tablosu yer alır ve çevresinde onunla ilişkili boyut tabloları bulunur. Boyut tabloları ile olay tablosu arasındaki ilişkiler çizgilerle gösterilir. Şekil 3.27’de SQL SERVER 97’de yer alan Warehouse kübünün bileşenleri kullanılarak oluşturulmuş bir star şeması görülmektedir.



Şekil 3.27 Star Şeması

3.3.2.2.5.2 Snowflakes Şeması

Snowflakes şeması da stars şemasına benzer. Yalnızca olay tablosuna bir ilişki ile bağlanan boyut tabloları dışında yine bu boyut tablolarıyla ilişkili olan boyut tabloları vardır. Şekil 3.28’de Sales küpüne ait bir snowflakes şeması yer almaktadır.

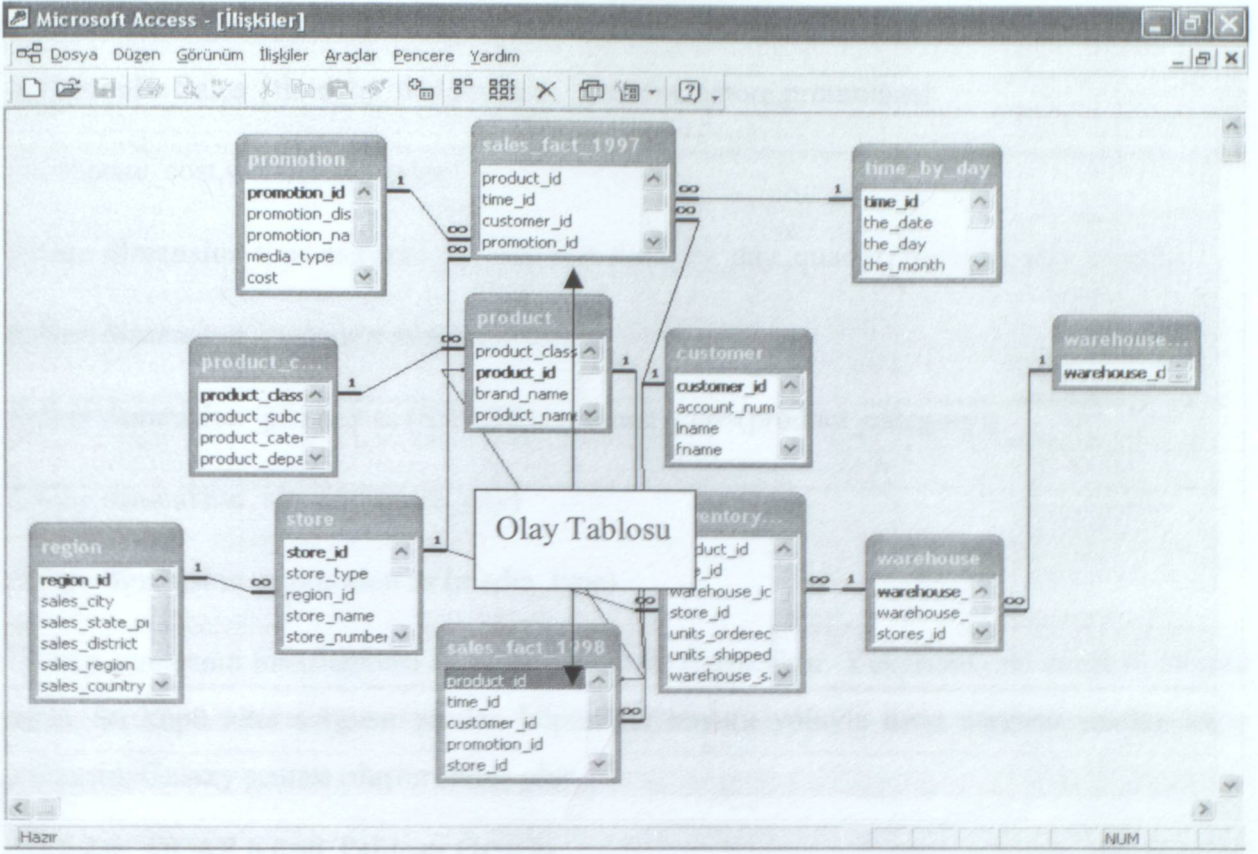


Şekil 3.28 Snowflakes Şeması

Şekil 3.28'den görüldüğü üzere olay tablosuna dolaylı olarak product tablosu yoluyla bağlanan product_class tablosu yer almaktadır.

3.3.2.2.5.3 Galaxy Şeması

Küp oluşturulurken birden çok olay tablosu ve bu olay tablolarına bağlı boyut tabloları varsa, oluşan şema “fact constellation” veya “galaksi şema” olarak adlandırılır. Şekil 3.29’da bir galaxy şema örneği yer almaktadır. Bu şekil SQL SERVER’da oluşturulamadığından, Microsoft Access XP’de oluşturulmuştur.



Şekil 3.29 Galaxy Şeması

Şekildeki oklarla gösterilen sales_fact_1997 ve sales_fact_1998 tabloları iki olay tablosudur ve diğer boyut tabloları ile ilişkilendirilmişlerdir. DMQL (Data Mining Query Language) yukarıda örnekleri verilen şemaların oluşturulması:

- Star Şemasının DMQL ile Oluşturulması:

```
define cube Warehouse [time_by_day,product, , warehouse,store]
```

```
warehouse_cost,warehouse_sales
```

```
define dimension time as (time_by_day.year time_by_day.,quarter, time_by_day .month)
```

```
define dimension warehouse as (country,city)
```

```
define dimension product as (SRP,SKU)
```

```
define dimension store as (store_city)
```


- Snowflakes Şemasının DMQL ile Oluşturulması:

define cube Sales [time_by_day,product, , customer,store,promotion]

warehouse_cost,warehouse_sales

define dimension time as (time_by_day.year time_by_day.,quarter, time_by_day .month)

define dimension customer as (country,city)

define dimension product as (SRP,SKU,product_class(product_category))

define dimension store as (store_city)

define dimension promotion as (media_type)

Galaxy şemasının hazırlanması da benzer biçimde oluşacaktır. Yukarıdaki iki sorguyu alt alta yazıp iki küpü altta sorgusu yazılan küpün bir boyutu yoluyla üstte sorgusu yazılan küpe bağlarsak Galaxy şeması oluşturulmuş olur.

3.3.2.2.6 OLAP Küpü Tablosu Örneği

Aşağıdaki Şekil 3.30'da bir Warehouse OLAP küpünde oluşturulmuş bir tablo yer almaktadır. Bu tabloda bütün mağazalarda, bütün mağaza türlerinde, bütün ambarlarda ve bütün mağaza büyüklüklerinde 1998 yılı itibari ile ölçü düzeylerinde gerçekleşen değerler görülmektedir. Bu değerler belirli bir mağaza, mağaza türü, ambar türü ve mağaza büyüklüğünde de elde edilebilir. Bu tabloda Şekil 3.31'de yer almaktadır.

MeasuresLevel	+ Product Family	+ Drink	+ Food	+ Non-Consumable
Store Invoice	181,818.84	15,801.34	131,606.76	34,410.74
Supply Time	18,942	1,833	13,495	3,614
Warehouse Cost	158,255.25	13,752.52	114,552.68	29,950.06
Warehouse Sales	348,484.72	30,352.70	251,794.94	66,337.09
Units Shipped	365,616	35,373	261,375	68,868
Units Ordered	399,628	38,646	285,839	75,143
Warehouse Profit	190,229.47	16,600.18	137,242.26	36,387.02

Şekil 3.30 Bir Olap Küpü Tablosu Örneği

The screenshot shows the 'Cube Editor' window. At the top, there is a menu bar (File, Edit, View, Insert, Tools, Help) and a toolbar. Below the toolbar, a dropdown menu shows 'Cube: Warehouse'. The main area contains a table with columns for 'MeasuresLevel' and four product categories: '+ Product Family', '+ Drink', '+ Food', and '+ Non-Consumable'. The table lists various measures like 'Store Invoice', 'Supply Time', 'Warehouse Cost', 'Warehouse Sales', 'Units Shipped', 'Units Ordered', and 'Warehouse Profit' with their corresponding values for each category. On the left side, there is a sidebar with a tree view showing the cube structure and a 'Cube' section with a 'The cube is' label.

MeasuresLevel	+ Product Family	+ Drink	+ Food	+ Non-Consumable
Store Invoice	15,725.38	1,147.86	11,465.83	3,111.69
Supply Time	1,582	141	1,080	361
Warehouse Cost	13,672.34	1,000.87	9,961.94	2,709.52
Warehouse Sales	30,336.79	2,241.48	22,054.90	6,040.41
Units Shipped	31,221	2,686	22,123	6,412
Units Ordered	33,958	2,950	23,831	7,177
Warehouse Profit	16,664.45	1,240.61	12,092.95	3,330.89

Şekil 3.31 Boyutların Düzey Ayrıntılarına Göre Bir Olap Küpü Tablosu Örneği

3.3.2.2.7 OLAP Küplerinde Gerçekleştirilen İşlemler

OLAP küplerinde Şekil 3.24'te yer alan bir boyutun hiyerarşik yapısı gösterilmiştir. Bir boyut tablosunda bir ya da birden çok düzey bulunabilir. Her düzey kendi şıkları tarafından gruplanır. Bu yapıdan yararlanarak OLAP küplerinde çeşitli işlemler yapılabilir. Bu işlemlerden bazıları şunlardır:

- Roll-up: Roll-up işlemi bir alt düzeyde alınan raporun bir üst düzeyde alınması işlemidir. Şekil 3.32 ve Şekil 3.33'de bir Roll-up işlemi görülmektedir. Şekil 3.32'de Mexico için mağaza detayında rapor alındığı görülmektedir. Bu detaylı rapordan Mexico için toplam rapora geçilmek istenirse Store Country düzeyinde yer alan Mexico alt hiyerarşisinin sol tarafında yer alan – işaretine mouse ile iki kez tıklattılır. Bu işlem Roll-up işlemidir.

		MeasuresLevel	
- Store Country	+ Store State	Store Invoice	Supply Time
All Stores	All Stores Total	7,619.05	7,619
+ Canada	Canada Total	634.92	635
- Mexico	Mexico Total	2,857.14	2,857
	+ DF	634.92	635
	+ Guerrero	317.46	317
	+ Jalisco	317.46	317
	+ Veracruz	317.46	317
	+ Yucatan	317.46	317
	+ Zacatecas	952.38	952
+ USA	USA Total	4,126.98	4,127

Şekil 3.32 Roll-up İşlemi Öncesi

		MeasuresLevel	
+ Store Country		Store Invoice	Supply Time
All Stores		7,619.05	7,619
+ Canada		634.92	635
+ Mexico		2,857.14	2,857
+ USA		4,126.98	4,127

Şekil 3.33 Roll-up İşlemi Sonrası

- Drill-down: Drill-down işlemi ise Roll-up işleminin tersidir. Bir düzeyin ya da düzeye ait alt hiyerarşinin kendi alt hiyerarşileri bazında tablolandırılmasıdır. Şekil 3.34 ve Şekil 3.35’de Drill-down işleminin adımları yer almaktadır. Şekil 3.34’te Store Country düzeyinin alt hiyerarşisinde yer alan Mexico alt hiyerarşisinin solundaki + işaretine mouse ile iki kez tıklanırsa Şekil 3-35 ’de yer alan şekildeki gibi Mexico’nun alt hiyerarşilerinde rapor alınır.

		MeasuresLevel	
+ Store Country		Store Invoice	Supply Time
All Stores		7,619.05	7,619
+ Canada		634.92	635
+ Mexico		2,857.14	2,857
+ USA		4,126.98	4,127

Şekil 3.34 Drill-down İşlemi Öncesi

		MeasuresLevel	
- Store Country	+ Store State	Store Invoice	Supply Time
All Stores	All Stores Total	7,619.05	7,619
+ Canada	Canada Total	634.92	635
	Mexico Total	2,857.14	2,857
	+ DF	634.92	635
	+ Guerrero	317.46	317
- Mexico	+ Jalisco	317.46	317
	+ Veracruz	317.46	317
	+ Yucatan	317.46	317
	+ Zacatecas	952.38	952
+ USA	USA Total	4,126.98	4,127

Şekil 3.35 Drill-down İşlemi Sonrası

- Slice ve Dice: Slice işlemi n boyutlu küpten bir boyutun bir alt hiyerarşisinin sabitlenip n-1 boyutlu bir küp elde edilmesidir. Şekil 3.36'da yer alan Country ve Product Family boyutları olan 2 boyutlu küp Şekil 3.37'de Product Family=Drink olan Country boyutu bazında rapor veren bir küp durumuna dönüştürülmüştür.

Dice işlemi ise, küpün boyutlarının düzeylerinden bazıları tarafından sınırlandırılmasıdır. Şekil 3.38 ve Şekil 3.39'da Dice işlemi öncesi ve sonrası durumlar görülmektedir.

Store Size in SQFT	All	Store Type	All
Time	1998	Warehouse	All Warehouses

+ Store Country	MeasuresLevel	+ Product Family		
		All Products	+ Drink	+ Food
	Store Invoice	800,000.00	74,358.97	574,358.97
	Supply Time	800,000	74,359	574,359
	Warehouse Cost	800,000.00	74,358.97	574,358.97
	Warehouse Sales	800,000.00	74,358.97	574,358.97
	Units Shipped	800,000	74,359	574,359
	Units Ordered	800,000	74,359	574,359
	Warehouse Profit	800,000.00	74,358.97	574,358.97
	Store Invoice	66,666.67	6,196.58	47,863.25
	Supply Time	66,667	6,197	47,863
	Warehouse Cost	66,666.67	6,196.58	47,863.25
	Warehouse Sales	66,666.67	6,196.58	47,863.25
	Units Shipped	66,667	6,197	47,863
	Units Ordered	66,667	6,197	47,863
	Warehouse Profit	66,666.67	6,196.58	47,863.25
	Store Invoice	300,000.00	27,884.62	215,384.62

Şekil 3.36 Slice İşlemi Öncesi

Store Size in SQFT	All	Store Type	All
Time	1998	Warehouse	All Warehouses
Product	Drink		

	+ Store Country			
MeasuresLevel	All Stores	+ Canada	+ Mexico	+ USA
Store Invoice	74,358.97	6,196.58	27,884.62	40,277.78
Supply Time	74,359	6,197	27,885	40,278
Warehouse Cost	74,358.97	6,196.58	27,884.62	40,277.78
Warehouse Sales	74,358.97	6,196.58	27,884.62	40,277.78
Units Shipped	74,359	6,197	27,885	40,278
Units Ordered	74,359	6,197	27,885	40,278
Warehouse Profit	74,358.97	6,196.58	27,884.62	40,277.78

Şekil 3.37 Slice İşlemi Sonrası

+ Product Family	+ Store Country	MeasuresLevel	
		Store Invoice	Supply Time
	All Stores	800,000.00	800,000
All Products	+ Canada	66,666.67	66,667
	+ Mexico	300,000.00	300,000
	+ USA	433,333.33	433,333
	All Stores	74,358.97	74,359
+ Drink	+ Canada	6,196.58	6,197
	+ Mexico	27,884.62	27,885
	+ USA	40,277.78	40,278
	All Stores	574,358.97	574,359
+ Food	+ Canada	47,863.25	47,863
	+ Mexico	215,384.62	215,385
	+ USA	311,111.11	311,111
	All Stores	151,282.05	151,282
+ Non-Consumable	+ Canada	12,606.84	12,607
	+ Mexico	56,730.77	56,731
	+ USA	81,944.44	81,944

Şekil 3.38 Dice İşlemi Öncesi

	MeasuresLevel	+ Product Family	
		Store Invoice	
+ Store Country	All Products	+ Drink	+ Food
All Stores	266,666.67	24,786.32	191,452.99
+ Canada	22,222.22	2,065.53	15,954.42
+ Mexico	100,000.00	9,294.87	71,794.87

Şekil 3.39 Dice İşlemi Sonrası

- Pivot: Oluşturulan OLAP küpünde boyutların yer değiştirmesidir. Örneğin, 2 boyutta X ekseninde yer alan değişkenin Y eksenine, Y ekseninde yer alan değişkenin X eksenine taşınması işlemidir.

	+ Store Country			
+ Product Family	All Stores	+ Canada	+ Mexico	+ USA
All Products	266,666.67	22,222.22	100,000.00	144,444.44
+ Drink	24,786.32	2,065.53	9,294.87	13,425.93
+ Food	191,452.99	15,954.42	71,794.87	103,703.70
+ Non-Consumable	50,427.35	4,202.28	18,910.26	27,314.81

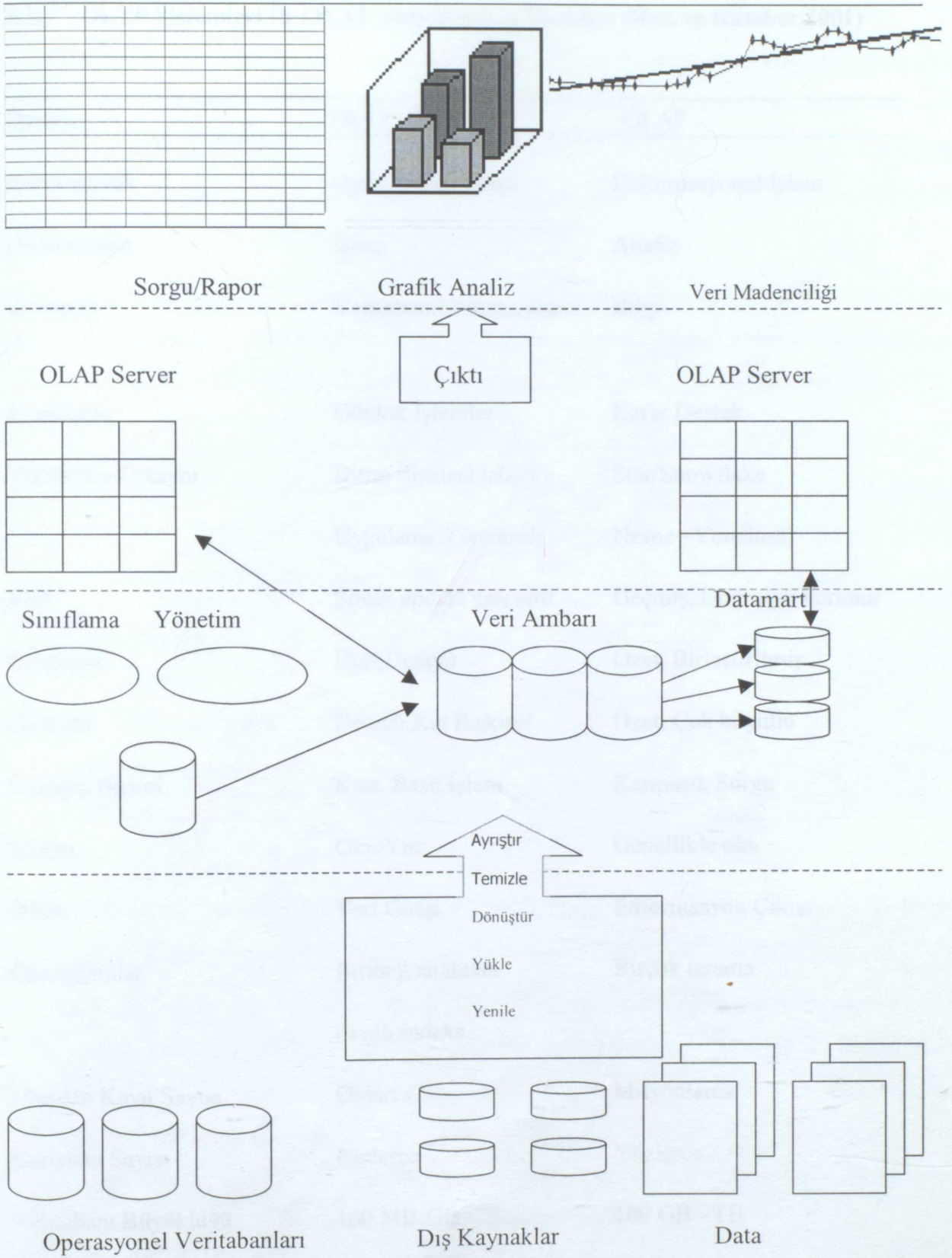
Şekil 3.40 Pivot İşlemi Öncesi

	+ Product Family			
+ Store Country	All Products	+ Drink	+ Food	+ Non-Consumable
All Stores	266,666.67	24,786.32	191,452.99	50,427.35
+ Canada	22,222.22	2,065.53	15,954.42	4,202.28
+ Mexico	100,000.00	9,294.87	71,794.87	18,910.26
+ USA	144,444.44	13,425.93	103,703.70	27,314.81

Şekil 3.41 Pivot İşlemi Sonrası

3.3.2.2.8 Veri Ambarında Yer Alan Verilerin Analizi Süreci

Şekil 3.42’de veri ambarına verilerin yüklenmesi, yüklenen verilerin yönetilmesi, OLAP Server’larına yüklenmesi ve analiz aşaması görülmektedir. Şekil görüldüğü üzere dört katmandan oluşmaktadır. En alttaki katman operasyonel veritabanları ve dış kaynaklardan oluşmaktadır. Veri yenilenme (refresh), transformasyon ve temizleme işlemlerinden sonra ayrıştırılarak veri ambarlarına yüklenir. Veri ambarlarına yüklenen verinin özellikleri (verinin türü, verinin boyutları, verinin boyutlarına ait düzeyler) metadata’da bulunur. Veri ambarında verinin yönetilmesi işlemi ve özel spesifik karar problemlerine ait oluşturulmuş veri depoları olan data martlara ya da veri OLAP Server’larına yüklenmesi işlemi gerçekleştirilir. OLAP Server’larında ise karar problemi ile ilgili değişkenler aracılığıyla küpler oluşturularak veri analizi için gerekli adıma uygun veriler oluşturulur. En son aşamada özet raporlar , grafiksel analiz ve veri madenciliği çalışmaları gerçekleştirilir.



Şekil 3.42 3 Katmanlı Veri Ambarı Mimarisi (Han ve Kamber,2001)

3.3.3 OLTP Sistemleri İle OLAP Sistemlerinin Farkları (Han ve Kamber,2001)

Özellik	OLTP	OLAP
Karakteristik	Operasyonel İşlem	Enformasyonel İşlem
Oryantasyon	İşlem	Analiz
Kullanıcı	VeritabanıUzmanı,yazıcı	Bilgi
Fonksiyon	Günlük İşlemler	Karar Destek
Veritabanı Dizaynı	Birim ilişkisel tabanlı	Star/Snowflake
	Uygulama Yönelimli	Nesne –Yönelimli
Veri	Şimdi update garantili	Geçmiş, Doğruluk korunur
Özetleme	İlkel,Detaylı	Özet, Birleştirilmiş
Görüntü	Detaylı,Kat İlişkisel	Özet, Çok boyutlu
Çalışma Birimi	Kısa, Basit işlem	Karmaşık Sorgu
Erişim	Oku/Yaz	Genellikle oku
Odak	Veri Girişi	Enformasyon Çıkışı
Operasyonlar	Birincil anahtara dayalı indeks	Birçok tarama
Ulaşılan Kayıt Sayısı	Onlarca	Milyonlarca
Kullanıcı Sayısı	Binlerce	Yüzlerce
Veritabanı Büyüklüğü	100 MB-GB	100 GB - TB
Öncelik	Performans,Yüksek ulaşım	Esneklik
Metrik	İşlem çıktısı (throughput)	Sorgu çıktısı, Cevap Süresi

3.4 Makine Öğrenmesi

Öğrenme, Simon tarafından, “zaman içerisinde yeni bilgilerin keşfedilmesi yolu ile davranışların iyileştirilmesi süreci” olarak tanımlanmaktadır. Öztemel bu tanımdan yola çıkarak makine öğrenmesini “bilgisayarın bir olay ile ilgili bilgileri ve tecrübeleri öğrenerek gelecekte oluşacak benzeri olaylar hakkında kararlar verebilmesi ve problemlere çözümler üretebilmesidir.(Öztemel,2003)” şeklinde tanımlamaktadır.

Makine öğrenmesi yapay zekanın bir alanıdır. Veri madenciliği tekniklerinden sınıflama, tahmin ve kümeleme tekniklerinde makine öğrenmesi kullanılmaktadır. Makine öğrenmesinde en önemli konu verinin içerisinde yer alan yapıların en iyi şekilde ortaya konulmasıdır.

Bilgisayarların öğrenmesi ise, incelenen olayla ilgili olarak alınan tüm örneklerin bilgisayar tarafından tek tek incelenmesi sonucu her adımda edindiği deneyimlerle gerçekleşir. Bilgisayar her adımda edindiği deneyimlerden olayla ilgili genellemeler yapar. Çeşitli öğrenme biçimleri şu şekilde sıralanabilir:

- Alışkanlıklardan öğrenme
- Görerek öğrenme
- Talimatlardan öğrenme
- Örneklerden öğrenme
- Deneylerle öğrenme
- Yapay sinir ağları ile öğrenme

Makine öğrenme sistemleri yukarıda belirtilen öğrenme biçimlerinden birini ya da birkaçını kullanabilir. Makine öğrenmesi çalışmaları genel olarak iki amaçla gerçekleştirilir.

- Farklı Çağrışım: Bu öğrenme amacı genellikle sınıflama ve tahmin problemlerinde kullanılır. Olayla ilgili örnekler ele alınarak, öğrenme biçimi ne olursa olsun, her olaydan elde edilen çağrışımların toplanması ve bu çağrışımlardan genelleme yapılmasıdır. Öğrenme yapacak sistemin girdileri ile çıktıları aynı değildir. Örneğin, enflasyonun geçmişe dönük verileri incelenirse, bu verilerin geçmişteki davranışları incelenerek gelecekteki davranışlarının tahmin edilmesi mümkün olabilir.
- Oto Çağrışım: Bu öğrenme amacı, bir olayın bilgisayara öğretilmesiyle ve karakterize edilen olayın başka zamanlarda tekrarlandığında tanınması durumlarında kullanılır. Örneğin, bir resim bir bilgisayara öğretilirse, resim yırtıldıktan sonra makineye bir örnek verildiği zaman, resimde bulunan eksik kısımlar makine tarafından tamamlanabilir.

3.4.1 Öğrenme Yaklaşımları

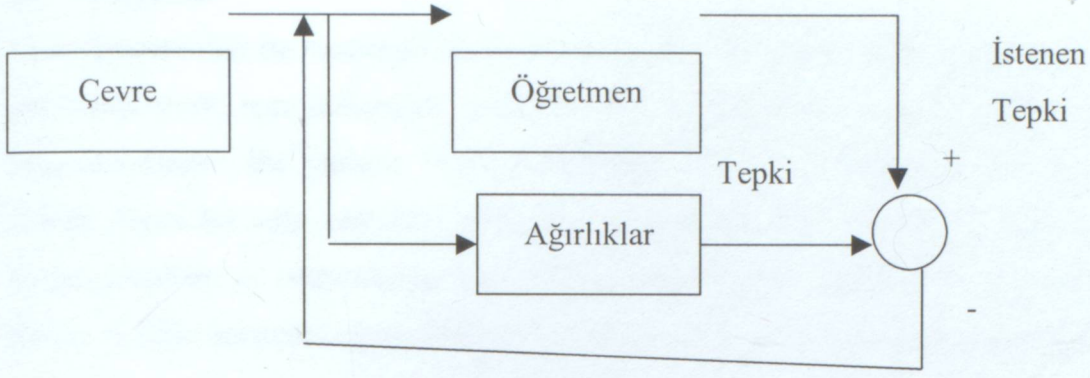
Bilgisayarların öğrenmesini sağlamak üzere kullanılan çeşitli yaklaşımlar söz konusudur. Bu yaklaşımlar şunlardır:

- Sembol İşleme Yöntemi: Geleneksel yapay zeka öğrenme teknolojisine dayanır. Bilgisayara belirli kurallar biçiminde sunulmuş olan bilgilerle karşılaştırma yoluyla öğrenme olayı gerçekleşir.
- Yapay Sinir Ağları: Öğrenme olayı, örneklerden öğrenilen bilgilerden genellemeler yoluyla gerçekleşir.
- İstatistiksel Yapı Tanıma: Bir veri setinin istatistiksel özellikleri ve dağılımı incelenerek veriye ilişkin genellemeler yapılır.
- Genetik Algoritmalar ve Evrimsel Programlama: Bu yöntemde başlangıç çözümleri atanır ve bu çözümlerden yeni çözümler türetilir. Bu işlemler, en iyi çözüm bulununcaya kadar bu işlemler tekrar edilir.
- Durum Tabanlı Öğrenme: Bu öğrenme yönteminde her durum birden çok örnekten oluşur ve bir olayı gösterir. Bir olay gerçekleşince nasıl karar verileceğine diğer olaylara bakılarak karar verilir.

3.4.2 Öğrenme Yöntemleri

Makine öğrenmesinde kullanılan iki tür öğrenme yöntemi vardır. Birinci tür öğrenme yöntemi “denetimli öğrenme” (supervised learning), ikinci tür öğrenme yöntemi ise “denetimsiz öğrenme” (unsupervised learning)’dir.

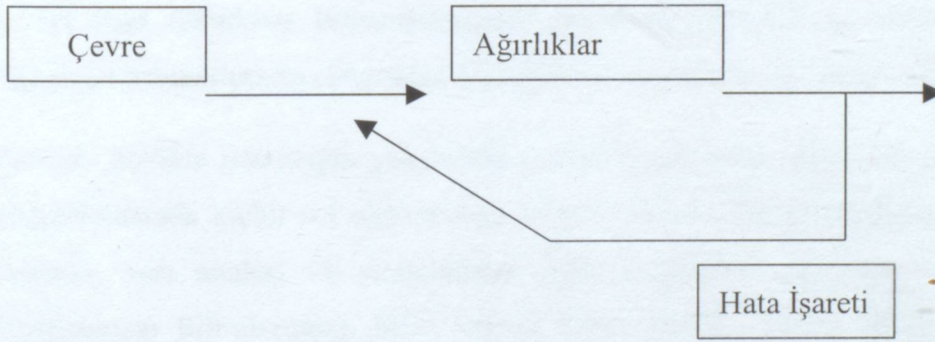
Denetimli öğrenme yönteminde bir bağımlı değişken vardır; bu bağımlı değişken, aranılan doğru cevaptır. Bağımlı değişkenin aldığı değerler, bağımsız değişkenler kullanılarak eğitim kümesi üzerinde tahmin edilmeye çalışılır. Tahmin sırasında, her bir bağımsız değişkene ait olan ağırlıklar bağımlı değişkeni tahmin etme oranı istenilen hata yüzdesinin altına düşünceye kadar değiştirilir. En son durumda elde edilen ağırlıklar kullanılarak değerlendirme kümesinde bağımlı değişkenin tahmini için kullanılır. Değerleme kümesinde, bu ağırlıklarla yapılan tahminlerin başarı oranı sınanır. Bundan sonraki aşama, test kümesinde bu ağırlıklarla yapılan tahminlerin başarı oranının bulunmasıdır. Bu başarı oranı da yeterli bulunursa model tahminler yapmak için kullanılmaya hazırdır. Danışmanlı öğrenme de ortaya çıkabilecek sorunlardan en önemlisi aşırı öğrenme (overfitting) sorunudur. Eğer eğitim kümesinde %100 oranında ya da bu orana çok yakın başarıyla tahminler elde edilmişse, kurulan model gerçek hayatta oluşacak olan yeni olayları tahmin etme de başarılı olamayacaktır. Aşağıdaki şekilde denetimli öğrenmenin akış diyagramı yer almaktadır.



Şekil 3.43 Denetimli Öğrenme (Elmas,2003)

Denetimli öğrenme kurallarına ; Delta, Genişletilmiş Delta ve Geri yayımlı öğrenme kuralları örnek olarak verilebilir.

Denetimsiz öğrenmedeyse denetimli öğrenmenin tersine ortada bir doğru cevap ya da bağımlı değişken yoktur. Denetimsiz öğrenmede var olan veride bulunan birimler birbirlerine olan uzaklıklarına göre ya da oluşturdıkları desenlere göre kümelenmeye çalışılır. Bu kümeleme işlemi, oluşturulan kümeler arasındaki uzaklık maksimum düzeye varıncaya kadar sürdürülür. Bu işlemin performans ölçümünün kontrol edilebileceği bir küme yoktur.



Şekil 3.44 Denetimsiz Öğrenme

Denetimsiz öğrenme kurallarına; Hebbian, Grossberg, Kohonen öz örgütlemeli öğrenme kuralları örnek olarak verilebilir.

3.5 İstatistik

“Veri madenciliği ile istatistiğin ilgilendikleri analizler, verinin modellenmesi ve bu görevleri gerçekleştirmek için kullanılan metotlar, bu iki disiplinin arasında büyük bir benzerlik oluşturmaktadır. Bir yazarın “Veri madenciliği nedir?” sorusuna vermiş olduğu cevabı : “Daha doğru bir adla istatistik” olmuştur. Veri madencileri her zaman istatistiksel araçları kullanmışlardır; ve istatistikçiler şimdi veri madenciliği problemlerine ilgi göstermektedirler. Bu iki disiplin arasındaki kesişimler her iki disiplin açısından da oldukça yararlı olacaktır.”*

“İstatistiksel açıdan bakınca, veri madenciliği büyük ölçeklerdeki karmaşık verinin bilgisayar yardımı ile otomatize edilmiş araştırılması ve analizi olarak görülebilir. Veri madenciliği ve istatistiğin genel özelliği, veriden öğrenmeleridir.

İstatistiğin geleneksel müfredatta küçük, temiz, statik, ve rasgele örneklenmiş ve sık olarak özel bir soruya cevap vermek için toplanmış verinin karakterize edilmesi gibi tanımlandığını hatırlayın. Bunlardan hiçbirisi veri madenciliği bağlamında (context) uygulanmaz. Klasik bir istatistikçiye göre binlerce gözlemden oluşan bir veri seti oldukça büyük görülebilirken bir veri madencisine göre oldukça küçüktür. Modern veritabanları sıklıkla milyonlarca kayıdı içerir ve bu büyüklükteki veriyi standart istatistiksel metotların fayda göstermediği bağlamına (context) açıkça koyar. Bu, işlenecek oldukça büyük verinin analizi sırasında, uyarılma ve dizisel bazı teknikleri birleştirilmesine gereksinim duyulacağı anlamına gelir. Bu konu, bilgisayar bilimcilerinin çalıştıkları yapı tanıma ve makine öğrenmesi alanlarına girer.

Bununla birlikte istatistiğin yukarıdaki paragraflarda sözü edilen büyük ölçeklerdeki verinin araştırılmasında hiçbir rol oynamadığı anlamı çıkmaz. Bütün bunlardan başka istatistik veri toplama, veri analizi ve modellemesi için mükemmel düzenlemeler öneren en önemli enformasyon bilimlerinden birisi olarak kabul edilir. Böyle olmakla birlikte, akademik istatistik, veri toplama, veri manipülasyonu ve veri yorumlama görünüşlerinin hepsini kapsamak için bakış açısını (purview) genişletmek zorundadır. Bu olay gerçekleşmedikçe, bilgisayar tabanlı disiplinler istatistikçilerden ve destekleme rolünü verdikleri istatistik disiplininden gereksinim duyduklarını alarak artan bir şekilde bu geniş perspektifi sunacaklardır.

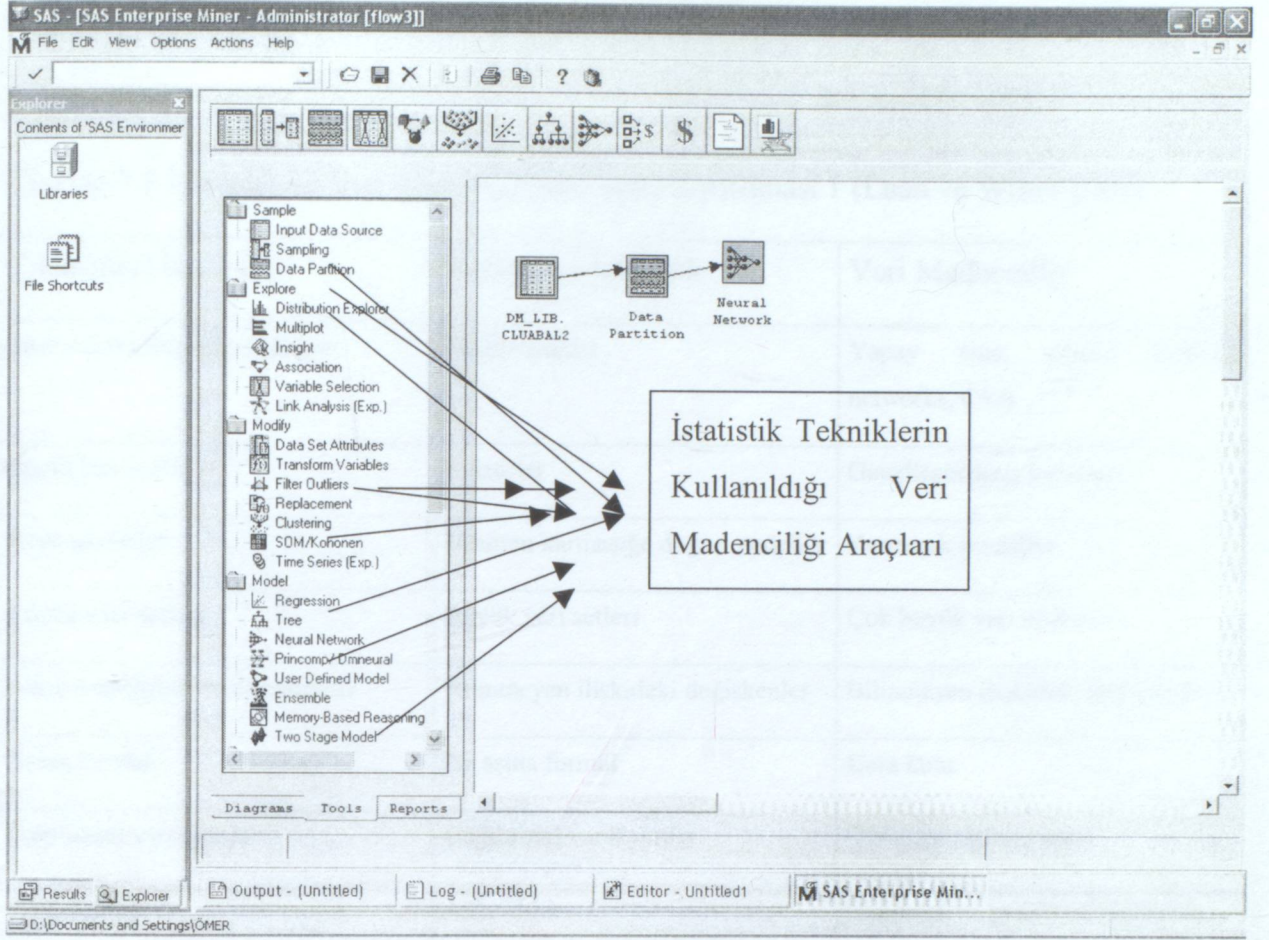
* What is data mining, <http://ugrad.stat.ubc.ca/~stat447j/resources/chapter1.pdf>

Veri toplama görünümlerini bir yana bırakırsak, en genel olarak bilinen veri madenciliği teknikleri; birliktelik kuralları, segmentasyon (kendini düzenleyen haritalar), karar ağaçları, yapay sinir ağları, görselleştirme teknikleridir. Bu teknikler istatistik disiplini dışında yer alan disiplinler tarafından geliştirilip sunulsa da, geleneksel istatistiksel teorilere ve fikirlere dayanırlar.(Ganesh,2002)”

Veri madenciliği ile istatistik disiplinleri yukarıdaki paragraflarda da belirtildiği gibi ortak noktalar barındırmalarına karşın, bazı özellikleri ile birbirlerinden ayrılırlar. Her iki disiplinin en önemli ortak noktası, bu disiplinlerin veri analizi ile ilgileniyor olmalarıdır. Veri madenciliğinin ortaya çıkışının 1980’li yılların sonlarına doğru olduğuna 2. bölümde değinilmişti. Veri madenciliğinin yararlanmış olduğu makine öğrenmesi, karar destek sistemleri, veritabanı teknolojisi, görselleştirme disiplinlerinin olduğu gibi, istatistik disiplininin de geçmişi veri madenciliği disiplininin daha geriye gitmektedir. Bu disiplinler bir araya getirilerek veri madenciliği disiplininin temelleri atılmıştır.

Bir veri madenciliği çalışması başlamadan önce yapılması gereken ilk çalışmalardan biri de ilgilenilen verinin değişkenlerinin dağılımının , ortalamasının , varyansının, eğer varsa değişkenler arasındaki korelasyonun öğrenilmesidir. Kütleyle ilişkin bu parametreler, ya da örneklem üzerinde çalışılıyorsa, kütleyle ilişkin parametrelerin tahmini uygulanılacak olan denetimli ya da denetimsiz veri madenciliği tekniğinin başarılı olabilmesi için gereklidir. Bir kümeleme analizinin başarısı veri setinde sapan değerlerin bulunmaması, değişkenler arasında korelasyonun olmaması durumunda daha başarılı sonuçlar verecektir. Yapılacak bir sınıflama çalışmasında kullanılacak olan regresyon , karar ağaçları, yapay sinir ağları çalışmalarının başarılı olmaları da değişkenlerin istatistiksel olarak analiz edilmeleri sonucu gereksinim duyulacak veri manipülasyonları sonucunda daha başarılı olacaklardır. Denetimli öğrenme sonucunda elde edilen modellerin başarılı olup olmadıkları da ki kare , f testi gibi istatistiksel testlerle sınanabilir. Kümeleme teknikleri kullanılarak elde edilen kümelerin birbirlerinden farklı olup olmadıklarının sınanmasında da “anova testi” ya da “t testi” kullanılabilir. Makine öğrenmesi teknikleri kullanılan çalışmalarda veri kümesi eğitim , test kümesi ve değerlendirme kümesi olarak ayrılmaktadır. Bu ayırma işleminde de örnekleme teknikleri kullanılmaktadır.

SAS Enterprise Miner ürününün veri madenciliği araçlarını incelersek de yine istatistiksel yöntemlerin bu araçlar içerisinde önemli bir yer kapladığını Şekil 3.45’de görürüz.



Şekil 3.45 SAS Enterprise Miner Ürünün Veri Madenciliği Araçlarından Bazıları

İstatistik ile veri madenciliğinin farklı olup olmadıkları tartışma konusu olsa da, veri madenciliği ile istatistik disiplinlerinin birbirlerinden ayrılmaları düşünülemez. Veri madenciliği teknikleri istatistik yöntemlere gereksinim duyar ve bu yöntemleri kullanır. Jing Luan ve Terence Willett AIR Conference’da sundukları bir makalede aşağıdaki çizelge 3.8’e yer vermişlerdir. Bu çizelgede çıkarımsal (inferential) istatistik , açıklayıcı (exploratory) istatistik ve veri madenciliği karşılaştırılmaktadır. Şekil 3.45 incelenecek olursa SAS Enterprise Miner ürününde çıkarımsal istatistik ve açıklayıcı istatistik tekniklerinin aşağıdaki çizelgede veri madenciliği teknikleri olarak ifade edilen yapay sinir ağları uygulamaları ile bütünleştiği görülmektedir. Bu durumda da görüldüğü gibi istatistik disiplini büyük veri setlerinde de çözümler üretebilmekte ve farklı disiplinlerle birlikte uyum içerisinde kullanılabilir. Aşağıdaki çizelgelerde belirtilen farklılıkları istatistik tekniklerinin genel uygulamaları için veri madenciliğinden ayrıştığı noktalar olarak dile getirebiliriz. Bununla birlikte, istatistiğin çizelge 3.9’da yer alan veri madenciliği uygulamalarında yetersiz kalacağını söyleyemeyiz.

Çizelge 3.8 İstatistik ve Veri Madenciliğinin Karşılaştırılması 1 (Luan ve Willet,2001)

Çıkarımsal İstatistik	Açıklayıcı İstatistik	Veri Madenciliği
t test, Anova, Regresyon, Kikare	Faktör analizi	Yapay sinir ağları, kohonen networks, C5.0
Öncül hipotezler	Faktörler	Genelleştirilmiş kurallar
Basit modeller	Basitten karmaşığa doğru modeller	Karmaşık modeller
Küçük veri setleri	Büyük veri setleri	Çok büyük veri setleri
Bilinen önemlilikde değişkenler	Bilinmeyen ilişkideki değişkenler	Bilinmeyen ilişkideki değişkenler
Aşına formül	Az aşına formül	Kara kutu
Dağılımsal varsayımlar	Dağılımsal varsayımlar	Öğrenme algoritmaları
Anlamlılık	Anlamlılık	Doğruluk ve genelleştirilebilirlik

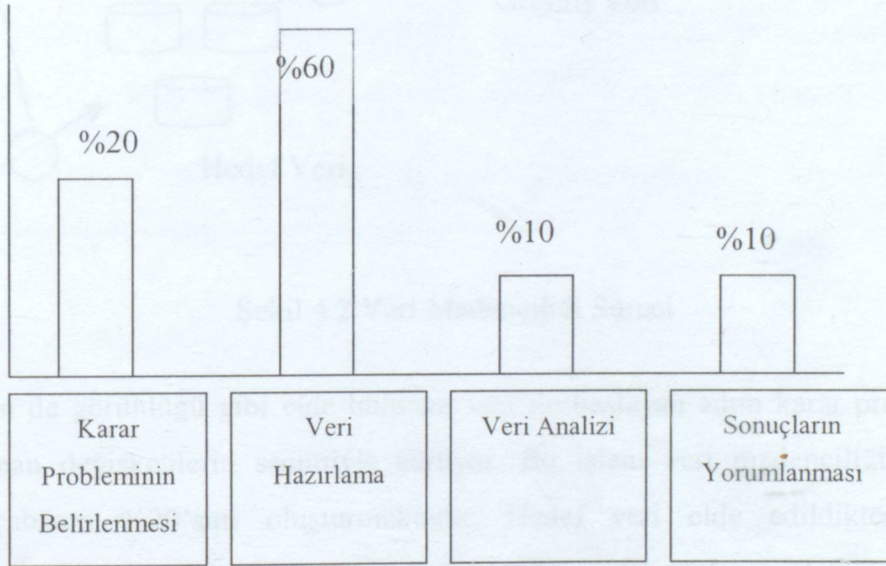
Çizelge 3.9’da ise veri madenciliği ile istatistik disiplinleri arasında bulunduğu iddia edilen ayrımlar yer almaktadır. Ama yine şekil 3.37’de görüldüğü gibi, istatistik tekniklerinin çizelge 3.9’da belirtilen farklılıklara uyum sağladığı görülmektedir.

Çizelge 3.9 İstatistik ve Veri Madenciliğinin Karşılaştırılması 2 (Ganesh,2002)

İstatistik	Veri Madenciliği
Yönetmek için Küçük n= Onlarca ya da yüzlerce p= bir avuç dolusu nadiren ondan fazla	Veri Büyüklüğü Çok Büyük n= yüz binlerce milyonlarca p= yüzlerce-binlerce
Veri bir modeli ya da özel bir sorunun cevabını test etmek için toplanır ya da bir araya konulur. (Genellikle Birinci el veri) Durum kontrol çalışmaları , Fikir araştırmaları anketleri, Deneysel tasarımlar, Gözlemsel çalışmalar	Veri Tipi Veri elektronik olarak toplanır ve mümkün bir gelecek kullanımı için depolanır. (Genellikle İkinci el veri) Süpermarket işlemleri verisi, telefon aramaları verisi, stok pazar verisi, internet trafik verisi
Veri insanlar tarafından bilgisayar yardımıyla işleme sokulur.	Veri İşlemleri Veri bilgisayar algoritmaları tarafından insan yardımı ile işleme tabi tutulur.
Model uydurma, Model test etme, Güven ve tahmin aralıkları,Daha iyi istatistiksel Prosedürler Gerçekleştirme, Metodların istatistiksel , ve matematiksel özelliklerini çalışmak	Tipik Görevler Yapılar araştırma ve tanıma, Sınıflama,Gruplama

4. Veri Madenciliği Metodolojisi

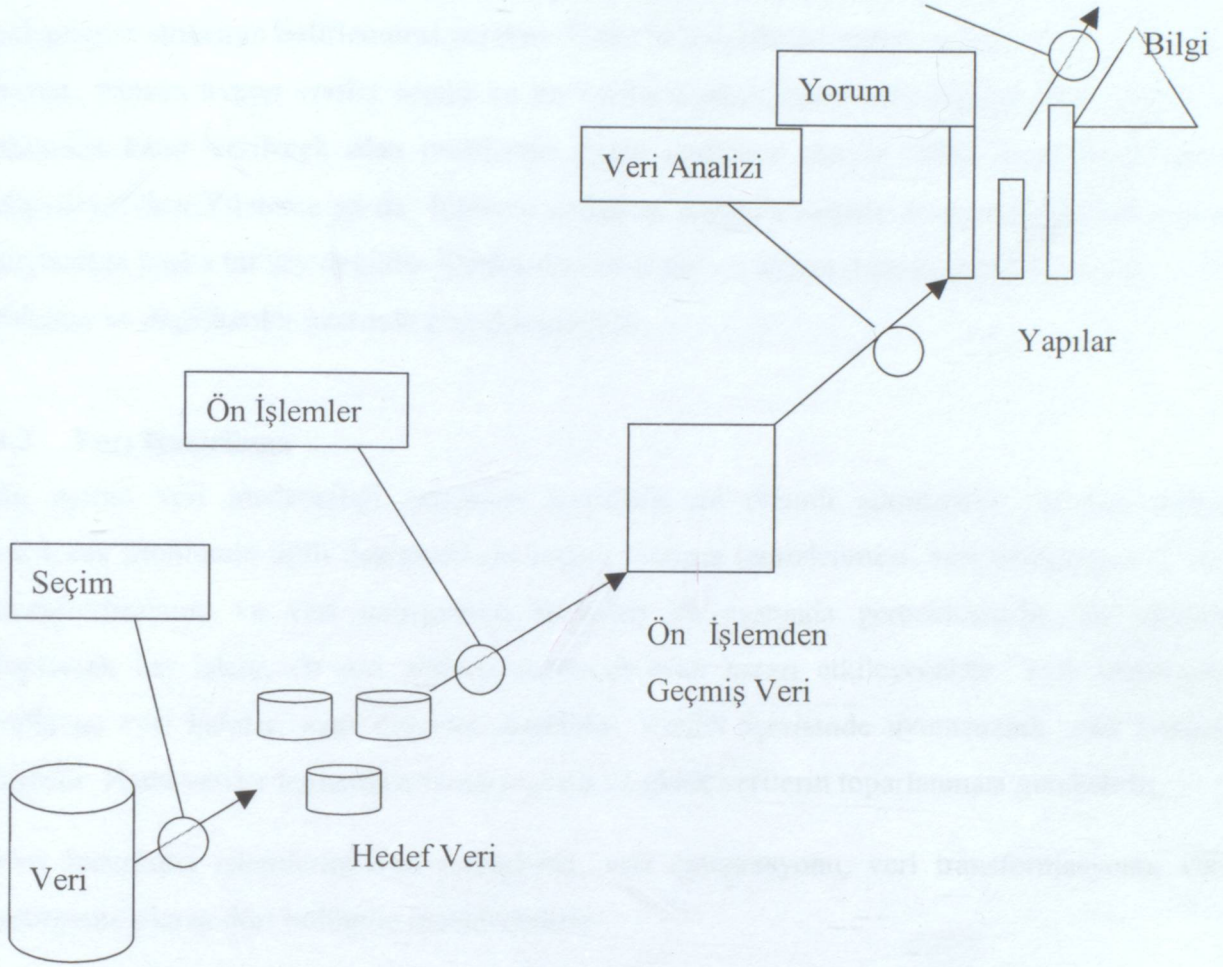
Veri madenciliği tanımı incelenirken veri madenciliğinin farklı disiplinlerden yararlandığı özellikle vurgulanmıştı. Veri madenciliğinin diğer disiplinlerle olan ilişkisi incelenirken, veri madenciliğinin veri analizinde yararlandığı yapay zeka, istatistik, görselleştirme disiplinlerinin yanı sıra verilerin saklanması ve analiz için veri elde edilmesi aşamalarında faydalandığı veritabanı teknolojisi ile karar destek sistemlerine de önemli ölçüde yer verildi. Çünkü veri madenciliği yalnızca hazır verinin analizinden ibaret değildir. Veri madenciliği, veri analizinin yanı sıra araştırılacak problemle ilgili veritabanının hazırlanmasından, verinin analize hazır hale getirilip analiz sonucunda elde edilen enformasyonun bilgiye dönüştürülmesi işlemlerini de içeren uzun bir süreçtir. Veri madenciliği sürecini kabaca dört bölümde düşünebiliriz. Bu bölümler; işle ilgili karar verilecek problemin tanımlanması ve problemle ilgili değişkenlerin seçilmesi, veri hazırlama, veri analizi ve sonuçların yorumlanmasıdır. Veri madenciliği çalışmaları sırasında harcanan çabanın bu dört bölüm içerisinde yüzdesel olarak dağılımı şekil 4.1’de görülmektedir.



Şekil 4.1 Veri Madenciliği Çalışmalarında Harcanan Çabanın Dağılımı (Cabena vd,1998)

Şekil 4.1’den de görüldüğü gibi veri madenciliği sürecinin en çok çaba gerektiren bölümü, veri hazırlama sürecidir.

Veri madenciliği sürecinin adımları aşağıda yer alan şekil 4.2’de ayrıntılı olarak görülebilir.



Şekil 4.2 Veri Madenciliği Süreci

Şekil 4.2’den de görüldüğü gibi elde bulunan veri ile başlayan adım karar problemine göre veride bulunan değişkenlerin seçimiyle sürüyor. Bu işlem veri madenciliği çalışmasında gösterilen çabanın %20’sini oluşturmaktadır. Hedef veri elde edildikten sonra veri madenciliği çalışmasında çabanın %60’ının gösterildiği veri hazırlama işlemine geçiliyor. Bu işlem sonucu analize hazır hale gelen veri analiz aşamasında kullanılıyor. Bu aşamada veri madenciliği çalışmasında gösterilen çabanın %10’u harcanmaktadır. En son aşama elde edilen yapıların yorumlanarak karar sürecinde kullanılmasıdır. Bu aşamada, veri madenciliği çalışmasında gösterilen çabanın %10’u harcanmaktadır.

Veri madenciliği çalışmasının aşamalarına aşağıda değinilmektedir.

4.1 Karar Probleminin Belirlenmesi

Veri madenciliği çalışması büyük veri tabanları içerisinde rasgele olarak bir enformasyon arama işlemi olarak görülmemelidir. Veri madenciliği çalışması yapmak için öncelikle çalışmanın amacının belirlenmesi gerekir. Amaç belirlendikten sonra, veri hazırlama işlemleri başlar. Amaca uygun veriler seçilir ve bu veriler analize hazır hale getirilir. Bu yüzden bu aşamada karar verilecek olan problemin doğru seçilmesi ileride tekrar başa dönülmesini engelleyecektir. Yüzlerce ya da binlerce değişken arasında rasgele araştırma yapmak zaman kaybından başka bir şey değildir. Çünkü öncelikle amaca uygun değişkenler seçilmelidir. Esas çalışma bu değişkenler üzerinde gerçekleşmelidir.

4.2 Veri Hazırlama

Bu aşama veri madenciliği çalışması içerisinde en önemli adımlardan birisidir. Karar verilecek problemle ilgili değişkenlerin seçimi, verinin temizlenmesi, veri entegrasyonu, veri transformasyonu ve veri indirgemesi işlemleri bu aşamada gerçekleştirilir. Bu adımda yapılacak her işlem, en son adımda verilecek olan kararı etkileyecektir. Veri ambarında toplanan veri hatalar, aşırı değerler içerebilir. Veriler içerisinde uyumsuzluk söz konusu olabilir. Hatta veriler tamamlanmamış olabilir ve eksik verilerin toparlanması gerekebilir.

Veri hazırlama işlemlerini veri temizleme, veri entegrasyonu, veri transformasyonu, veri indirgeme olarak dört bölümde inceleyebiliriz.

4.2.1 Veri Temizleme

Gerçek hayattan gelen veriler içerisinde mutlaka bazı sorunlar bulunabilir. Bu sorunlar içerisinde en çok göze batan sorunlarsa veri içerisinde kayıp değerlerin bulunması, verinin aşırı değerler (yüksek ve/veya düşük) içermesi, veri içerisinde uyumsuzluk bulunmasıdır.

4.2.1.1 Kayıp Değerler

Kayıp değerler üzerinde çalışılacak veride analiz aşamasında sorunlar oluştururlar. Örneğin, kümeleme analizinde kayıp değerlerin bulunduğu birimleri bir kümeye atanamaz, temel bileşenler analizinde ve faktör analizinde birimlere ait olan skorlar hesaplanamaz. Veri seti kayıp değerlerden uygun bir şekilde arındırılmalıdır.

Kayıp değerler; veri girişi hatalarından, eksik bilgi toplanmasından, bilginin alındığı birimin cevap vermekten kaçınmasından ya da birimin o bilgiye sahip olmamasından kaynaklanabilir.

Kayıp değerlerden veri setini arındırmak için yapılabilecek bazı işlemler şunlardır:

- Eğer maliyetli değilse ve uzun zaman almayacaksa eksik değerin ait olduğu birime başvurmak.
- Eksik değerin bulunduğu birimi veri setinden çıkarmak. Bu yöntem genellikle tercih edilmez.
- Eksik değer eğer tamamlanabiliyorsa manuel olarak eksik değerin tamamlanması. Örneğin, askerliğini yapmış bir kişinin cinsiyeti boş ise cinsiyetin erkek olarak tamamlanması.
- Değişkenin genel eğiliminin gösterdiği değeri kayıp değerin yerine atamak. Örneğin, değişken ortalaması, modu ya da medyanının kayıp değer yerine atanması.
- Eğer veri setindeki birimler arasında özellik bakımından farklılık söz konusuysa örneğin; cinsiyet, meslek gibi sınıflamalar söz konusu ise kayıp değerin yerine kayıp değerin ait olduğu sınıfın eğiliminin gösterdiği değerin atanması.
- Regresyon ya da karar ağacı modellemesi yaparak kayıp değerin olması gereken değerinin tahmin edilmesi.
- Kayıp değer yerine “Bilinmeyen” gibi global bir sabit atanması.

	BRANCH_ID	COMM_MKT_S	RET_MKT_SAI	BUSS_MKT_S	IND_PROFIT	I
►					-3408716	
	411	921	809	809	2545512	
	422	905	800	800	12703634	
					112289284	
					5998507	
	132	902	815	815	1154291930	
	183	902	815	815	348083171	
					-2964967	
					381689454	
					399819455	
					12142484	
	295	906	818	818	27401914	
					46346607	
	350	905	800	800	18038787	
	313	929	813	813	17997679	

Şekil 4.3 Kayıp Değerler Bulunan Bir Veri Örneği

4.2.1.2 Aşırı Değerler

Aşırı değerlerse birimlerin genel olarak bir değişkenin belirli bir aralıkta yer alan şıklarının dışında aşırı yüksek ya da aşırı düşük şıkları almalarıyla oluşur. Aşırı değerler, veri madenciliği sürecinin analiz aşamasında regresyon, kümeleme, temel bileşenler analizi, faktör analizi gibi uygulamalarda sorunlar doğururlar. Bu nedenle aşırı değerler veri setinde bulunması istenilmeyen bir durumdur.

Veri setinde bulunan aşırı değerleri bulmak için kullanılabilecek bazı yöntemler aşağıda yer almaktadır:

- Veriler küçükten büyüğe sıralanır. Sıralanmış veri bölmelere ayrılarak aşırı değerler bulunabilir.
- Veri seti kümeleme analizi ile kümeler ayrılır. Aşırı değerler içeren birimler bağımsız kümeler oluşturabilir.
- Değişkenlere ait kutu diyagramları çizilir. Kutu diyagramlarında aşırı değerler görülebilir.
- Değişkenlerin grafikleri aracılığıyla aşırı değerler bulunabilir.
- Temel bileşenler analizinde elde edilen ilk iki veya son iki temel bileşenin serpilme diyagramı incelenerek aşırı değerler bulunabilir.

Aşırı değerler bulunduktan sonra yapılacak işlem aşırı değerlerin arındırılmasıdır. Aşırı değerlerden arındırma işlemlerinden bazıları şunlardır:

- Aşırı değerlerin bulunduğu birim sayısı çok fazla değilse (buna çalışmayı yapan uzman karar verebilir) bu birimler analiz dışında tutulabilir.
- Aşırı değerler yerine değişkenin genel ortalaması kullanılabilir.
- Eğer veri setinde meslek ve cinsiyet gibi sınıflamalar varsa, aşırı değer bulunan değişkenin ait olunan sınıftaki ortalaması, modu ya da medyanı aşırı değerler yerine atanabilir.
- Aşırı değer bulunan değişkenler dışarıda tutularak bir regresyon, karar ağacı gibi modelleme yöntemi kullanılarak model kurulur. Kurulan modele göre aşırı değerlerin yerine hangi değer atanacağına karar verilebilir.
- Aşırı değer bulunan birimden bir önceki ve bir sonraki (n önceki; n sonraki) birimin o değerkende aldığı değerlerin ortalaması aşırı değer yerine konulabilir.

4.2.1.3 Uyumsuz (Inconsistent) Veri

Veri setinde bulunan değişkenlerin alabileceği bazı şıkların birden çok farklı biçimde gösterilmesi, aynı birimin iki değerkende almış olduğu değerlerin çelişkili olması ya da değerkene ait olmayan bir şıkkın veri setinde yer alması uyumsuz veriye örnek olarak gösterilebilir.

Çizelge 4.1 Uyumsuz Veri Örneği

Şehir	Bölge	Yaş
İstanbul	Marmara	-15
İSTANBUL	Marmara	15
Ankara	Doğu Anadolu	26

4.2.2 Veri Entegrasyonu

Veri ambarında yer alan veriler farklı tablolarda, farklı küplerde, veya kat (flat) dosyalarında yer alabilir. İncelenecek olan konu ile ilgili değişkenler bu farklı kaynaklardan toplanarak bir datamart ya da bir küp oluşturulabilir. Bu işleme veri entegrasyonu adı verilir.

4.2.3 Transformasyon

Analize dahil edilecek olan değişkenler yapılacak analizlerin varsayımlarına uydurulmak, ölçü birimlerinden arındırılmak, aşırı değerlerin etkisini azaltmak, nicel değişkenleri nitel değişkenlere dönüştürmek gibi nedenlerle transforme edilir. Transformasyon işlemlerinden bazıları şunlardır:

- Düzleştirme (Smoothing): Bölmeleme (binning), regresyon, kümeleme analizleri içerir. Genellikle aşırı değerleri arındırmak için kullanılır.
- Veri aşırı detaylıysa, veriyi daha özet bir hale getirmek için kullanılır. Örneğin, OLAP küplerindeki Roll-up işlemi.
- Aylık gelir düzeyi gibi düşük, orta, yüksek gibi sınıflanabilecek olan sürekli değişkenlerin genelleştirilerek nitel hale dönüştürülmesidir.
- Kümeleme analizinde daha ayrıntılı bir biçimde yer verilecek olan değişkenlerin standardizasyonu, 0-1 ya da -1-1 aralıklarına indirgenmesidir.

4.2.4 Veri İndirgeme

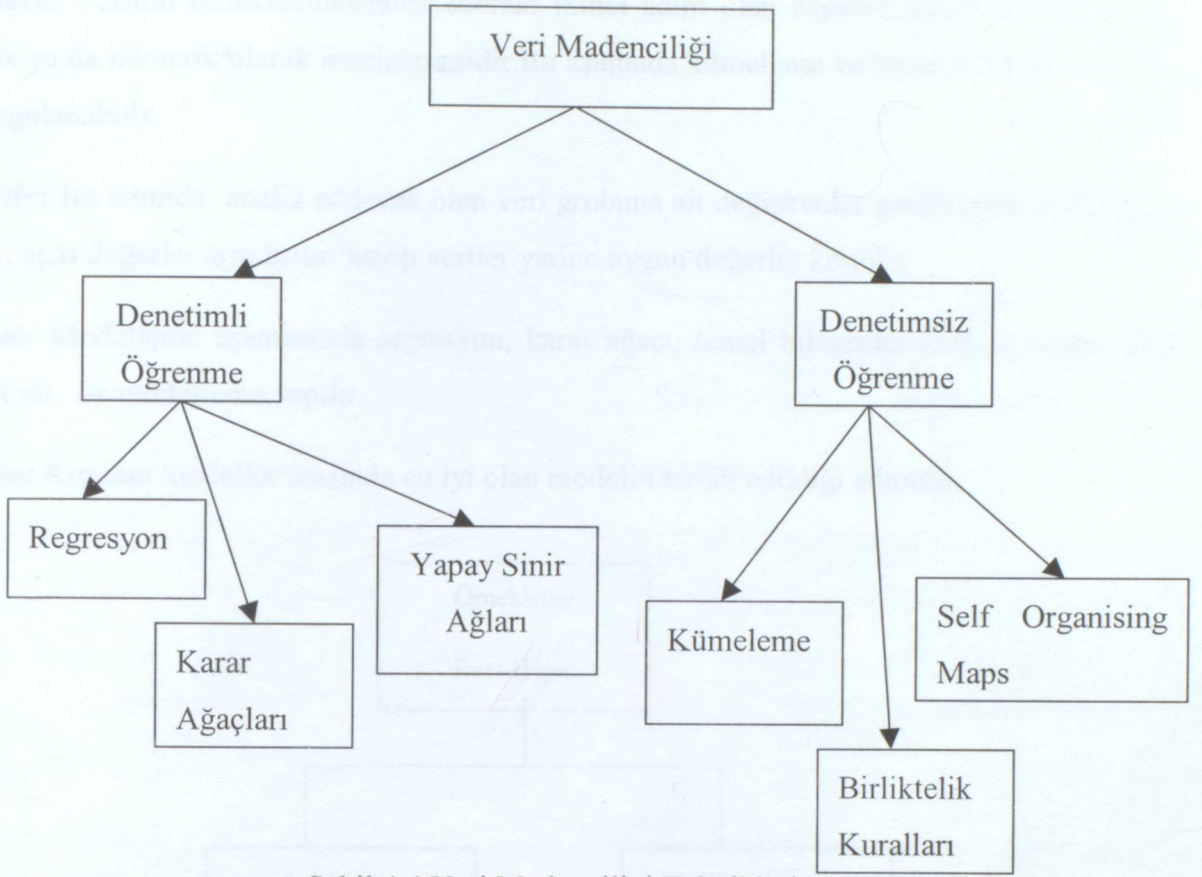
Bir veri ambarında bulunan oldukça çok değişken içerisinde seçilen karar problemi ile ilgili değişkenler de oldukça çok sayıda olabilir. Aynı zamanda değişken sayısının indirilmesi veri sayısının da azaltılması anlamına gelmemektedir. Birim sayısının azaltılması da veri indirgeme içerisinde incelenebilir. Bu yüzden veri indirgemeyi hem değişken sayısının azaltılması hem de birim sayısının azaltılması olarak görebiliriz.

Veri indirgeme tekniklerinden bazıları aşağıda yer almaktadır:

- Veri küplerinde çeyrekler bazında ya da aylar bazında satışların 6 aylık ve/veya yıllık satışlar olarak gösterilmesi.
- İlişkisiz, zayıf ilişkili veya gereksiz değişkenlerin kaldırılması. Forward selection, backward elimination, karar ağacı indirgeme yöntemleri bu yöntemlere örnek olarak verilebilir.
- Değişkenlerin bir araya getirilerek az sayıda değişkenle aynı açıklama oranının bulunması. 5. bölümde detaylı olarak incelemek olan temel bileşenler analizi ve faktör analizi bu yöntemlere örnek olarak verilebilir.
- Veri setindeki birimler heterojen bir yapıya sahipse bu verilerin homojen gruplara indirgenmesi. 5. bölümde detaylı olarak incelemek olan kümeleme analizi bu yöntemlere örnek olarak verilebilir.

4.3 Veri Analizi

Veriler analize uygun hale getirildikten sonra veri madenciliği teknikleri arasından karar probleminin çözümüne yardımcı olacak olan yöntem seçilir. Veri madenciliği teknikleri birbirlerinden bağımsız olarak kullanılabilecekleri gibi birbirleriyle etkileşim içerisinde de kullanılabilirler. Başka bir deyişle, veri madenciliği tekniklerinden birisinin çıktısı diğer bir veri madenciliği tekniğinin girdisi olabilir. Temel bileşenler analizi sonucunda elde edilen birimlere ilişkin skorlar kümeleme analizi, regresyon analizi için girdi olarak kullanılabilirler. Kümeleme analiziyle oluşturulan kendi içlerinde homojen fakat kendi aralarında heterojen gruplarasa diğer veri madenciliği teknikleri uygulanabilir. Genellikle bu tekniklerin bir arada kullanılması daha etkin sonuçlar doğurabilir. Veri madenciliği teknikleri aşağıdaki şekil 4.4'te görülebilir.



Şekil 4.4 Veri Madenciliği Teknikleri

4.4 Sonuçların Yorumlanması

Veri madenciliği metodolojisindeki en son aşama, veri analizi aşaması sonucunda elde edilen sonuçların yorumlanmasıdır. Bu aşamada elde edilen yapıların ve/veya modellerin görselleştirilmesi gerçekleştirilir. Bu yapılar ya da modeller gerçek hayata uygulanır.

Veri madenciliği araçları üreten firmalar da çeşitli metodolojiler geliştirmektedirler. Bu metodolojilerden birisi de SAS Enterprise Miner ürününde kullanılan SEMMA metodolojisidir.

4.5 SEMMA Metodolojisi

SAS Enterprise Miner ürününde kullanılan SEMMA metodolojisinin adımları Sample, Explore, (Modify), Model, Assess'dir.

Sample: Veri ambarından örnekleme yolu ile bir ya da birden fazla veri tablosu oluşturmaktır. Tüm veride veri madenciliği çalışması yapmak yerine örnekleme yolu ile veriden seçilen ve tüm veriyi temsil eden daha az birime sahip veriyle çalışmak daha etkin

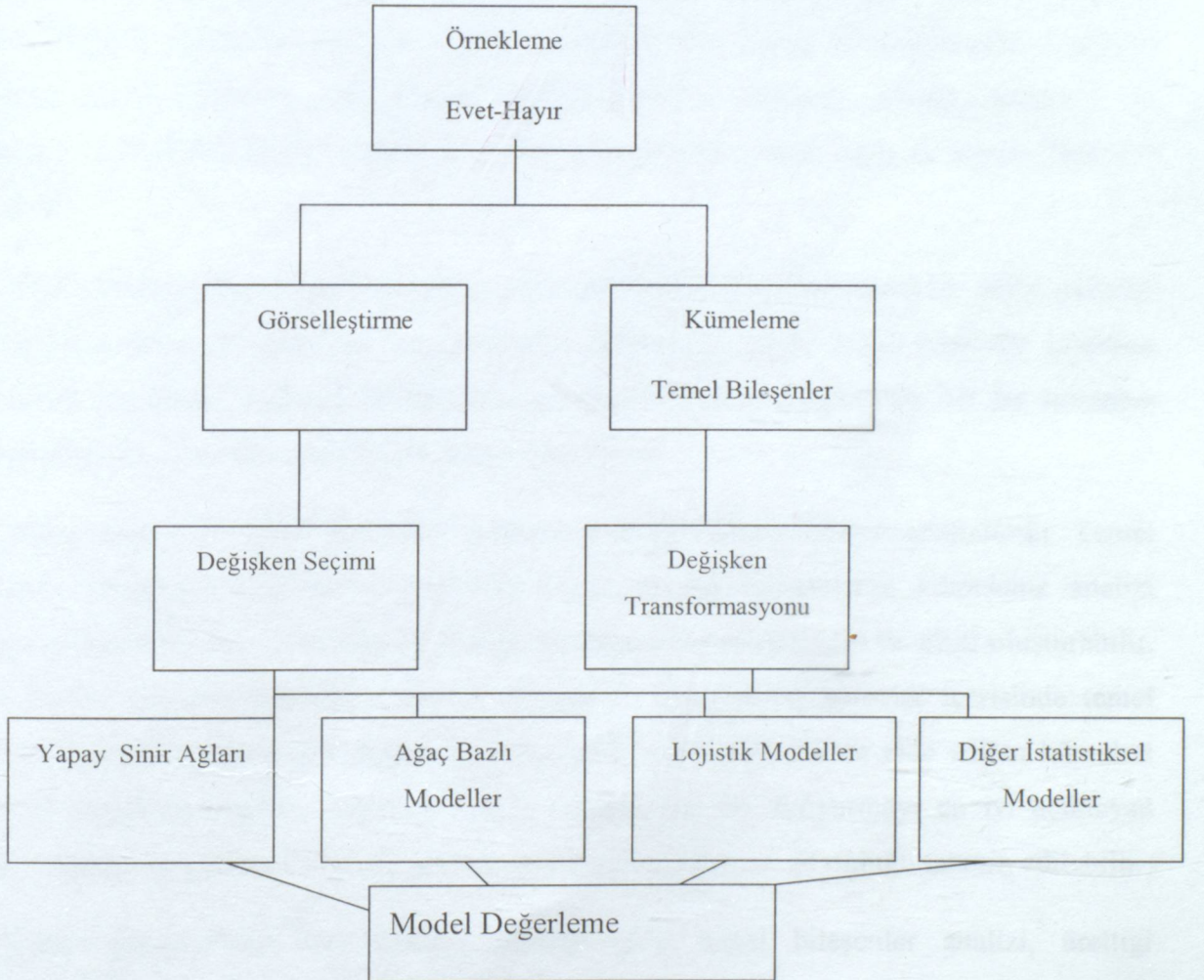
olur.

Explore: Verinin örneklenmesinden sonraki ikinci adım olan explore adımı verinin görsel olarak ya da nümerik olarak incelenmesidir. Bu aşamada kümeleme ve temel bileşenler analizi de uygulanabilir.

Modify: Bu adımda analiz edilecek olan veri grubuna ait değişkenler gerekiyorsa transforme edilir, aşırı değerler arındırılır, kayıp veriler yerine uygun değerler konulur.

Model: Modelleme aşamasında regresyon, karar ağacı, temel bileşenler analizi, yapay sinir ağları vb. ile modelleme yapılır.

Assess: Kurulan modeller arasında en iyi olan modelin tercih edildiği adımdır.



Şekil 4.5 SEMMA Metodolojisi

5. Veri Madenciliğinde Veri İndirgemesi İşlemlerinde Kullanılan İstatistik Tekniklerden Temel Bileşenler Analizi ile Kümeleme Analizinin İncelenmesi

Veri madenciliği tekniklerinin çok büyük veri setlerinde kullanıldığına önceki bölümlerde değinilmişti. Büyük veri setlerinde veri madenciliği teknikleri kullanılırken ortaya çıkabilecek sorunlardan bazıları; değişkenler arasında ortaya çıkan korelasyon yüzünden oluşturulacak regresyon modellerinde çoklu doğrusal bağlantı sorununun ortaya çıkması, kümeleme analizinde korelasyonlu değişkenlerin oluşturulan kümelerin tam ayrıştırılamamasına neden olması, büyük veri setleri içerisinde farklı özelliklere sahip kümelerin varlığının olması olasılığı ve bu kümelerin ayrıştırılarak her bir küme için özel modellemelere gerek duyulmasıdır.

Veri setinde bulunan değişkenlerin aralarında korelasyon olmayan daha az sayıda bileşenle temsil edilmesi, temel bileşenler analiziyle gerçekleştirilir. Temel bileşenler analizi sonucu değişken sayısı indirgenir, değişkenler arasındaki birim farklılığı ortadan kaldırılır ve aralarında korelasyon olmayan fakat tüm değişkenleri temsil eden daha az sayıda değişken elde edilir.

Kümeleme analizi ise, veri seti içerisinde farklı özellikleri olan birimlerin bir araya getirilip homojen grupların oluşturulması için kullanılır. Böylelikle veriler kendi içlerinde homojen ama farklı özellikleri bulunan alt kümelere ayrıştırılmış olur. Ayrıştırılan her bir homojen küme içerisinde, istenilen modelleme çalışmaları yapılır.

Kümeleme analizi ile temel bileşenler analizi bir arada kullanılabilecek analizlerdir. Temel bileşenler analiziyle elde edilen birimlere ilişkin skorlar kullanılarak kümeleme analizi gerçekleştirilebilir. Aynı zamanda bu skorlar bir regresyon modeli için de girdi oluşturabilir. Buna benzer biçimde kümeleme analizi sonucunda elde edilen kümeler içerisinde temel bileşenler analizi uygulanarak değişken indirgemesi yapılabilir. Ya da elde edilen kümeleri gösteren değişken, bağımlı değişken, olarak kullanılarak bu ayrıştırmayı en iyi açıklayan model kurulup veri setine katılacak yeni birimlerin hangi kümeye ait olduğu tahmin edilebilir.

Yukarıdaki paragraflarda söz konusu edildiği gibi; temel bileşenler analizi, ürettiği birbirlerinden bağımsız değişkenlerle diğer analizlere kaynak olabilmekte; ve kümeleme analizi ise, oluşturduğu homojen gruplarla daha başarılı modelleme çalışmalarına kaynak oluşturmaktadır. Her iki teknik de veri madenciliği çalışmalarında verimli olarak kullanılabildiklerinden , bu tez içinde incelenecektir.

5.1 Temel Bileşenler Analizi

Temel bileşenler analizi, veri setinde bulunan orijinal değişkenler arasındaki varyans kovaryans yapısını açıklamayla ilgilenir.

Temel bileşenler analizinin amaçları şunlardır (Özdamar,1999):

- 1- Veri İndirgemesi.
- 2- Tahmin yapmak.
- 3- Veri setini bazı yöntemlerin analiz edebileceği biçime sokmak.

Toplam sistemdeki değişkenliği göstermek için p bileşene gerek duyulmasına karşın, toplam sistemdeki değişkenlik daha az sayıda (k adet) temel bileşen ile de gösterilebilir. Oluşturulan k adet temel bileşen toplam sistemdeki değişkenliğin yeteri kadarını açıklıyorsa k adet temel bileşen p adet bileşen yerine kullanılabilir. Böylelikle p adet bileşene sahip n adet ölçüm, k adet temel bileşene sahip n adet ölçüme indirgenir. Analizler k adet temel bileşenle yapılsa da, yorumlar p adet bileşenle yorumlanabilir.

Temel bileşenler analizi, önceden bilinmeyen ilişkileri sıkça açığa çıkarır ve buradan genellikle sonuç olarak tahmin edilmeyen yorumlamalar yapılmasına olanak verir. Temel bileşenler analizi kanser tedavisinde hastaların radyoterapiye verdikleri çok sayıda tepkinin indirgenerek bu tepkinin daha basit bir ölçüyle gösterilmesinde; biyolojide, tohum üretme programlarında farklı ürünlerin ve protein yapılarının değişik ölçümlerinin seçim indeksi (selection index) adı verilen indekse dönüştürerek bu ürünlerin atalarına karar verilmesinde, çevreyle ilgili çalışmalarda, çok sayıda değişkenle ilgili veri toplandığı durumlarda daha kolay yorumlanabilir daha az sayıda değişken üretilmesinde; meteorolojide yeni meteorolojik değişkenler üretmek için kullanılabilir. Bu örneklerle ek olarak; spor , psikoloji, sosyoloji , ekonomi gibi bilimlerde de temel bileşenler analizinden yararlanılabilir.

Temel bileşenler analizi genel olarak tek başına kullanılmaz. “Temel bileşenler analizinin ürünü olarak ortaya çıkan skor değerleri; kümeleme, regresyon gibi analizler için girdi olarak kullanılır.(Johnson ve Wichern,1982)”

Temel bileşenleri cebirsel olarak göstermek gerekirse, p rasgele değişkenin $(X_1, X_2, \dots, X_p)$ özel lineer kombinasyonlarının elde edilmesi olarak dile getirebiliriz. Geometrik olaraksa $X_1, X_2, \dots, X_p$ orijinal eksenlerinin döndürülmesiyle yeni bir koordinat sistemi elde edilmesi olarak gösterilebilir. Elde edilen yeni eksenlerde toplam sistemdeki değişkenliğin ve kovaryans yapısının maksimum düzeyde açıklanabilirliği sağlanır. Temel bileşenler analizi varyans kovaryans matrisine veya korelasyon matrisine dayanır. “Bu matrislerin geliştirilmesi çok değişkenli normal dağılım varsayımına gereksinim duymaz. Öte yandan, temel bileşenler sabit yoğunluklu elipsoidler bakımından yararlı yorumları bulunan çok değişkenli normal popülasyonlar için türetilmiştir. (Johnson ve Wichern, 1982)”

$X' = [X_1, X_2, \dots, X_p]$ rasgele vektörü $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ öz değerleri bulunan Σ varyans kovaryans matrisine sahip olsun.

Aşağıdaki lineer kombinasyonları ele alalım:

$$\begin{aligned} Y_1 &= l'_1 X = l_{11}X_1 + l_{21}X_2 + \dots + l_{p1}X_p \\ Y_2 &= l'_2 X = l_{12}X_1 + l_{22}X_2 + \dots + l_{p2}X_p \\ &\vdots \\ Y_p &= l'_p X = l_{1p}X_1 + l_{2p}X_2 + \dots + l_{pp}X_p \end{aligned} \quad (5.1)$$

$$Y = l'X \quad (5.2)$$

$$Var(Y_i) = l'_i \Sigma_X l_i \quad i = 1, 2, \dots, p \quad (5.3)$$

$$Cov(Y_i, Y_k) = Cov(l'_i X, l'_k X) = l'_i \Sigma_X l_k \quad i, k = 1, 2, \dots, p \quad (5.4)$$

Temel bileşenler varyansları 5.3'e eşit olan korelasyonsuz $Y_1, Y_2, \dots, Y_p$ lineer kombinasyonlarıdır.

Birinci temel bileşen maksimum varyanslı, yani $Var(Y_1) = l'_1 \Sigma_X l_1$ 'i maksimum yapan lineer kombinasyondur. $Var(Y_1) = l'_1 \Sigma_X l_1$, herhangi bir l_1 'in herhangi bir sabit ile çarpılması ile artırılabilir. Bu belirsizliği (indeterminacy) ortadan kaldırmak için, katsayı vektörlerinin birim uzunlukta olmaları sağlanır. Bu işleme normalizasyon işlemi adı verilir. Vektörün elemanları vektörün uzunluğuna oranlanarak elde edilir. Normalize edilen vektörün uzunluğu ise 1 birim olur.

Buradan ;

birinci asal bileşen $=Var(Y_1)$ 'i maksimize eden $l_1'l_1 = 1$ olan $l_1'X$ lineer kombinasyonudur,
ikinci asal bileşen $=Var(Y_2)$ 'yi maksimize eden $l_2'l_2 = 1$ olan $l_2'X$ lineer kombinasyonudur,

$$Cov(l_1'X, l_2'X) = 0 \quad (5.5)$$

...

i. asal bileşen ise $=Var(Y_i)$ 'yi maksimize eden $l_i'l_i = 1$ olan $l_i'X$ lineer kombinasyonudur.

$$Cov(l_i'X, l_k'X) = 0 \quad i > k \quad (5.6)$$

Sonuç 5.1. $X' = [X_1, X_2, \dots, X_p]$ rasgele vektörü Σ varyans kovaryans matrisine sahip olsun. Σ 'nın özdeğer ve özvektör çiftleri ise $(\lambda_1, e_1), (\lambda_2, e_2), \dots, (\lambda_p, e_p)$ olsun. $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ olmak üzere i. temel bileşen

$$Y_i = e_i'X = e_{1i}X_1 + e_{2i}X_2 + \dots + e_{pi}X_p \quad i = 1, 2, \dots, p \quad (5.7)$$

Bu durumda aşağıdaki seçenekler gerçekleşir.

$$Var(Y_i) = e_i'\Sigma e_i = \lambda_i \quad i = 1, 2, \dots, p \quad (5.8)$$

$$Cov(Y_i, Y_k) = e_i'\Sigma e_k = 0 \quad i \neq k$$

Eğer bazı λ_i 'ler eşit ise, seçeneklere uyum gösteren katsayı vektörleri e_i 'ler ve bundan dolayı da Y_i 'ler tek değildir.

İspat (Johnson ve Wichern,1982) :

$$\max_{l \neq 0} \frac{l' \Sigma l}{l' l} = \lambda_1 \quad l_1 = e_1 \quad \text{ise} \quad (5.9)$$

Özvektörler normalize edildiğinden dolayı $e_1' e_1 = 1$ 'dir. Bundan dolayı

$$\max_{l \neq 0} \frac{l' \Sigma l}{l' l} = \lambda_1 = \frac{e_1' \Sigma e_1}{e_1' e_1} = e_1' \Sigma e_1 = \text{Var}(Y_1) \quad (5.10)$$

Benzer biçimde;

$$\max_{l \perp e_1, e_2, \dots, e_k} \frac{l' \Sigma l}{l' l} = \lambda_{k+1} \quad k = 1, 2, \dots, p-1 \quad (5.11)$$

$e_{k+1}', e_i = 0$ şartı ile $l = e_{k+1}$ alınır

$i=1, 2, \dots, k$ ve $k=1, 2, \dots, p-1$ için

$e_{k+1}' \Sigma e_{k+1} / e_{k+1}' e_{k+1} = e_{k+1}' \Sigma e_{k+1} = \text{Var}(Y_{k+1})$ elde edilir.

Yine $e_{k+1}', e_i = 0$ şartı ile $l = e_{k+1}$ alınır $\max_{l \perp e_1, e_2, \dots, e_k} \frac{l' \Sigma l}{l' l} = \lambda_{k+1} \quad k = 1, 2, \dots, p-1$

çözülduğünde $\lambda_{k+1} l' l = l' \Sigma l$ elde edilir. Buradan $\lambda_{k+1} e_{k+1}' e_{k+1} = e_{k+1}' \Sigma e_{k+1}$ bulunur.

$e_{k+1}' \Sigma e_{k+1} = \text{Var}(Y_{k+1})$ olduğundan ve $e_{k+1}' e_{k+1} = 1$ olduğundan $\text{Var}(Y_{k+1}) = \lambda_{k+1}$ elde edilir. Bu

durum $e_i \perp e_k$ 'yı göstermeyi gerektirir. Bunu da şu şekilde elde ederiz.

$\text{Cov}(Y_i, Y_k) = e_i' \Sigma e_k = e_i' \lambda_k e_k = \lambda_k e_i' e_k = 0 \quad i \neq k$ olduğunu bildiğimiz için bu durumun

gerçekleşmesi $e_i \perp e_k$ koşuluna bağlıdır. Buradan şu sonucu elde ederiz: Farklı özdeğerlere ait

özvektörlerin oluşturdukları matris diktir. Aynı özdeğerler söz konusu olursa özvektörler matrisi dik olacak şekilde seçilir.

Sonuç 5.1, temel bileşenlerin korelasyonsuz olduğunu ve varyanslarının varyans kovaryans matrisinin özdeğerlerine eşit olduğunu gösterir.

Sonuç 5.2. $X' = [X_1, X_2, \dots, X_p]$ Σ kovaryans matrisine ve $(\lambda_1, e_1), (\lambda_2, e_2), \dots, (\lambda_p, e_p)$

özdeğer özvektör çiftlerine sahip olsun. $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ olmak üzere bu özdeğerlere ait

özvektörlerle oluşturulan temel bileşenler $Y_1 = e_1' X, Y_2 = e_2' X, \dots, Y_p = e_p' X$ olsun. O zaman

$$\sigma_{11} + \sigma_{22} + \dots + \sigma_{pp} = \sum_{i=1}^p \text{Var}(X_i) = \lambda_1 + \lambda_2 + \dots + \lambda_p = \sum_{i=1}^p \text{Var}(Y_i) \quad (5.12)$$

İspat (Johnson ve Wichern,1982): $\sigma_{11} + \sigma_{22} + \dots + \sigma_{pp} = tr(\Sigma)$ P özvektörler matrisi ve Λ özdeğerlerden oluşan köşegen matris olmak üzere $\Sigma = P\Lambda P'$

$tr(\Sigma) = tr(P\Lambda P') = tr(\Lambda P P') = tr(\Lambda) = \lambda_1 + \lambda_2 + \dots + \lambda_p$ Böylece ispat tamamlanmış olur.

Buradan da k. bileşene ait toplam populasyon varyansı şöyle hesaplanır: $\lambda_k / \lambda_1 + \lambda_2 + \dots + \lambda_p$

Sonuç 5.3. Eğer $Y_1 = e'_1 X, Y_2 = e'_2 X, \dots, Y_p = e'_p X$ temel bileşenleri kovaryans matrisi Σ 'dan elde edilmiş ise,

$$\rho_{Y_i, X_k} = \frac{e_{ki} \sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}} \quad i, k = 1, 2, \dots, p \text{ korelasyon katsayısı formülü } Y_i \text{ ve } X_k \text{'lar arasındaki}$$

korelasyon katsayısını verir. Burada $(\lambda_1, e_1), (\lambda_2, e_2), \dots, (\lambda_p, e_p)$ çiftleri Σ 'nın özdeğer özvektör çiftleridir.

İspat (Johnson ve Wichern,1982): $l'_k = [0, \dots, 0, 1, 0, \dots, 0]$ olarak oluşturulsun. Bundan

dolayı $X_k = l'_k X$ olur ve $Cov(X_k, Y_i) = Cov(l'_k X, e'_i X) = l'_k \Sigma e_i$ elde edilir. $\Sigma e_i = \lambda_i e_i$ olduğundan $Cov(X_k, Y_i) = l'_k \lambda_i e_i$ olur. $l'_k e_i \lambda_i = \lambda_i e_{ki}$ 'ye eşit olur.

$Var(X_k) = \sigma_{kk}$ $Var(Y_i) = \lambda_i$ bilindiğinden bu değerler aşağıdaki formülde yerine yazılırsa 5.13 elde edilir.

$$\rho_{Y_i, X_k} = \frac{Cov(X_k, Y_i)}{\sqrt{Var(Y_i) Var(X_k)}} = \frac{\lambda_i e_{ki}}{\sqrt{\lambda_i \sigma_{kk}}} = \frac{\sqrt{\lambda_i} e_{ki}}{\sqrt{\sigma_{kk}}} \quad (5.13)$$

Yukarıdaki açıklamalarda temel bileşenler analizinin çok değişkenli normal dağılan X değişken kümesi için türetildiğine değinmiştik. X $N_p(\mu, \Sigma)$ dağılsın. X'in yoğunluk fonksiyonu μ merkezli elipsoidler üzerinde yer alır.

$$(X - \mu)' \Sigma^{-1} (X - \mu) = c^2 \quad (5.14)$$

5.14 formülü ile oluşan elipsoid $\pm c \sqrt{\lambda_i} e_i$ $i=1, 2, \dots, p$ olan noktalarda eksenleri keser.

(λ_i, e_i) 'ler Σ 'nın özdeğer ve özvektör çiftleridir. Elipsoidin i. eksen üzerinde bulunan A

noktaları koordinat sistemi üzerinde, $X_1, X_2, \dots, X_p$ eksenlerine paralel ve $e'_i = [e_{1i}, e_{2i}, \dots, e_{pi}]$

vektörünün elemanlarına göre orantısal koordinatlara sahip olacaktır. Bu noktaların varyans kovaryans matrislerinin tersi ise Σ^{-1} olsun. Bu noktalara ait elipsoidin denklemi aşağıdaki

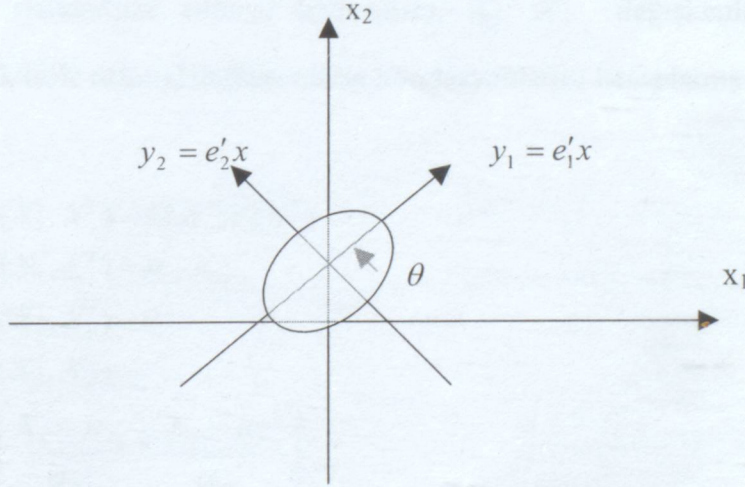
5.15 denklemi gibi olur. Denklem oluşturulurken işlem kolaylığı açısından $\mu = 0$ olarak alınmıştır.

$$c^2 = x' \Sigma^{-1} x = \frac{1}{\lambda_1} (e_1' x)^2 + \frac{1}{\lambda_2} (e_2' x)^2 + \dots + \frac{1}{\lambda_p} (e_p' x)^2 \quad (5.15)$$

$e_1' x, e_2' x, \dots, e_p' x$ x veri setinin temel bileşenleridir ve sırasıyla $y_1, y_2, \dots, y_p$ olarak gösterilirlerse (5.16) denklemi elde edilir.

$$c^2 = \frac{1}{\lambda_1} y_1^2 + \frac{1}{\lambda_2} y_2^2 + \dots + \frac{1}{\lambda_p} y_p^2 \quad (5.16)$$

(5-8) denklemi koordinat sisteminde $y_1, y_2, \dots, y_p$ eksenlerine sahip bir elipsoid tanımlar. En büyük özdeğer λ_1 ise, en büyük eksen bu özdeğere ait e_1 özvektörü yönünde uzanır. Kalan küçük eksenler ise $e_2, \dots, e_p$ yönlerinde uzanır. Şekil 5.1'de 2 boyutlu $\mu = 0$ ortalamalı x veri setine ait 2 temel bileşenin oluşturduğu yeni koordinat sistemi sabit yoğunluk elipsi görülmektedir.



Şekil 5.1 Sabit Yoğunluk Elipsi $x' \Sigma^{-1} x = c^2$ $\mu = 0$ ve $\rho = 0.75$

5.1.1 Standardize Değişkenlerden Elde Edilen Temel Bileşenler

“Ana bileşenler, veri setinde yer alan değişkenlerin ölçü birimlerinin farklı olduğu ve değişkenliklerinin farklı olduğu durumlarda standardize veri matrisinden ya da korelasyon matrisinden hesaplanmalıdır.(Özdamar,1999)”

Eğer orijinal veri setindeki değişkenlerin ölçü birimleri farklı ise, ya da değişkenlerin varyansları arasında farklılık söz konusu ise, değişkenler standardize edilir. Standardize değerler (5.17) formülü ile elde edilir:

$$z_{ij} = (X_{ij} - \mu_i) / \sigma_{ii} \quad (5.17)$$

Standardize edilmiş olan değişkenlerin korelasyon matrisleri ile varyans kovaryans matrisleri aynıdır. Benzer şekilde orijinal değişkenlerin korelasyon matrisi ile standardize edilmiş değişkenlerin korelasyon matrisleri aynıdır. Bundan dolayı temel bileşenleri hesaplarırken orijinal değişkenlerin korelasyon matrislerinin özdeğerlerine karşılık gelen özvektörlerini kullanmak ile standardize edilmiş olan değişkenlerin korelasyon matrislerinin özdeğerlerine karşılık gelen özvektörleri kullanmak arasında bir fark yoktur.

İspat: X'_1 değişkeni standardize edilmiş X_1 değişkeninin, X'_2 değişkeni standardize edilmiş X_2 değişkeninin standardize edilmiş hali olsun. X'_1 X'_2 değişkenlerinin kovaryansını hesapladığımız takdirde orijinal değişkenlerin korelasyonlarını hesaplamış oluruz.

$$Cov(X'_1, X'_2) = E(X'_1, X'_2) - E(X'_1)E(X'_2) \quad (5.18)$$

$$Cov(X'_1, X'_2) = E(X'_1, X'_2) - \mu_{X'_1} \mu_{X'_2}$$

$$Cov(X'_1, X'_2) = E(X'_1, X'_2) - 0$$

$$Cov(X'_1, X'_2) = E(X'_1, X'_2)$$

$$Cov(X'_1, X'_2) = E \left[\frac{X_1 - \mu_{X_1}}{\sigma_{X_1}} * \frac{X_2 - \mu_{X_2}}{\sigma_{X_2}} \right]$$

$$Cov(X'_1, X'_2) = \frac{1}{\sigma_{X_1} \sigma_{X_2}} Cov(X_1, X_2) = Cor(X_1, X_2)$$

Standardize edilmiş değişkenlerin oluşturduğu matris $Z = (S^{1/2})^{-1}(X - \mu)$ formülü ile hesaplanır ve yapısı aşağıdaki Şekil 5.2'deki gibi olur.

$$Z = \begin{matrix} \frac{X_{11} - \bar{X}_1}{\sqrt{\sigma_{11}}} & \frac{X_{12} - \bar{X}_1}{\sqrt{\sigma_{11}}} & \dots & \frac{X_{1p} - \bar{X}_1}{\sqrt{\sigma_{11}}} \\ \frac{X_{21} - \bar{X}_2}{\sqrt{\sigma_{22}}} & \frac{X_{22} - \bar{X}_2}{\sqrt{\sigma_{22}}} & \dots & \frac{X_{2p} - \bar{X}_2}{\sqrt{\sigma_{22}}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{X_{n1} - \bar{X}_p}{\sqrt{\sigma_{pp}}} & \frac{X_{n2} - \bar{X}_p}{\sqrt{\sigma_{pp}}} & \dots & \frac{X_{np} - \bar{X}_p}{\sqrt{\sigma_{pp}}} \end{matrix}$$

Şekil 5.2 Standardize Edilmiş Veri Matrisi

Korelasyon matrisi ile varyans kovaryans matrisinden elde edilen temel bileşenleri karşılaştırsak aşağıdaki sonuçlara ulaşırız.(Rencher,2002):

- 1- Korelasyon matrisinden hesaplanan varyans yüzdesi, varyans kovaryans matrisinden hesaplanan varyans yüzdesinden farklılaşacaktır.
- 2- Korelasyon matrisinden hesaplanan temel bileşen katsayıları, varyans kovaryans matrisinden elde edilen temel bileşen katsayılarından farklılaşacaktır.
- 3- Orijinal değişkenlerin korelasyon matrisi bakımından bileşenlerini ifade edersek, bu bileşenler varyans kovaryans matrisinden elde edilen bileşenler ile uyum sağlamayacaktır.
- 4- Korelasyon değerleri ölçekten bağımsız oldukları için korelasyon matrisinden elde edilen temel bileşenler de ölçekten bağımsız olacaklardır.
- 5- Korelasyon matrisinden elde edilen temel bileşenler tek değildir. Örneğin iki değişken söz konusu olduğunda ;

$$R = \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix} \quad \lambda_1 = 1 + r \quad \lambda_2 = 1 - r \text{ olur. Özvektörler ise; } e_1 = (0.707, 0.707) \text{ ve } e_2 = (0.707, -0.707) \text{ şeklinde gerçekleşir.}$$

5.1.2 Varyans Kovaryans Matrisinde veya Korelasyon Matrisinde Özel Yapılar

Varyans kovaryans matrislerindeki yapılar veya korelasyon matrislerindeki bazı özel yapılar yardımıyla temel bileşenlere ilişkin önceden tahminlerde bulunabiliriz. Örneğin, varyans kovaryans matrisindeki bir değişken, görece diğer değişkenlere oranla çok büyük varyansa sahip ise ilk temel bileşen bu değişkenden etkilenecektir. Başka bir durumsa, orijinal veri setindeki tüm değişkenlerin korelasyonsuz olması durumudur. Bu durumda varyans

kovaryans matrisi aşağıdaki biçimde oluşacaktır:

$$S = \begin{bmatrix} s_{11} & 0 & \cdots & 0 \\ 0 & s_{22} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & s_{pp} \end{bmatrix} \quad \text{Bu matrisin özdeğerleri ise } 0 = |S - \lambda I| = \prod_{i=1}^p (s_{ii} - \lambda) \text{ şeklinde}$$

oluşacak ve özdeğerler ise $\lambda_i = s_{ii}$ olacaktır. i.ci özdeğere ait özvektör ise $e_i = (0, \dots, 0, 1, 0, \dots, 0)$ olacaktır. Buradan $y_i = e_i' X = X_i$ olacaktır.

Perron Forbenius teoremine göre tüm korelasyonlar ya da kovaryanslar pozitif ise ilk özvektörün tüm elemanları pozitif olacaktır; ve izleyen diğer özvektörler bu ilk özvektöre dik olacağından, bu özvektörlerin bazı elemanları pozitif, bazı elemanları negatif olacaktır.

5.1.3 Temel Bileşenler İle Değişkenler Arasındaki Korelasyon

Temel bileşenler ile değişkenler arasındaki korelasyon temel bileşenlerin yorumlanmasında yardımcı olarak kullanılır. Orijinal veri setindeki bir değişkenin temel bileşenlerle korelasyonlarının karelerinin toplamı, o değişkenin temel bileşenler ile regresyonunun R^2 'sine eşittir.

$$r_{x_i, y_1}^2 + r_{x_i, y_2}^2 + \dots + r_{x_i, y_k}^2 = R_{x_i | y_1, \dots, y_k}^2 \quad (5.19)$$

Bu eşitliğin sağlanmasının nedeni, y temel bileşenlerinin birbirine dik olması nedeniyle korelasyonlarının sıfır olmasıdır.

5.1.4 Temel Bileşenlerin Rotasyonu

Temel bileşenler başlangıçta orijinal sistemin eksenlerinin rotasyonu sonucu elde edilirler. Bu bileşenler birbirleri arasında korelasyonsuzdur. Bu bileşenlerin yorumlanması tatmin edici bulunmazsa rotasyon işlemi yapılabilir.

Bununla birlikte rotasyon sonucu oluşacak olan yeni değişkenler korelasyonlu olabilirler ve maksimum varyansa ulaşamayabilirler.

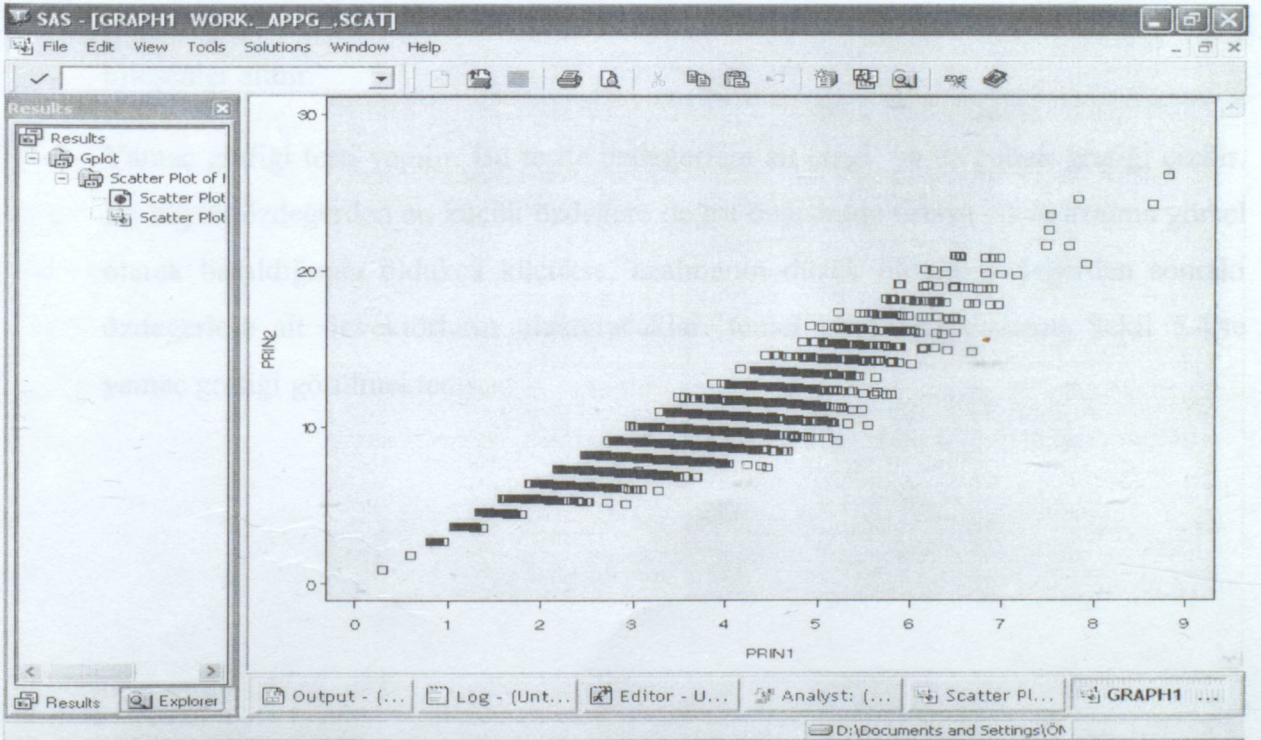
Genellikle, rotasyon işlemi faktör analizi sonucunda elde edilen eksenlere uygulanır. Kullanılan rotasyon işlemleri arasında varimax, quartimax, equamax, promax, direct oblimin işlemleri genellikle yer alır.

5.1.5 Temel Bileşenlerin Yorumlanması

Daha önce korelasyon ve varyans kovaryans matrislerinden elde edilen temel bileşenlerin uyumlu olmadığına değinilmişti. Çünkü varyans kovaryans matrisinden elde edilen temel bileşenler değişkenlerin ölçü birimlerinden etkilenirler. Bunun için varyans kovaryans veya korelasyon matrislerinin bileşenleri arasında seçim yaparken maksimum varyanslı bir değişken varsa, ya da değişkenler arasında ölçü birimleri farklıysa, korelasyon matrisi tercih edilir. Temel bileşenler ile orijinal değişkenler arasında gerçekleşecek olan korelasyon değerleri hangi temel bileşenin hangi değişkenlerin bileşiminden oluştuğu hakkında bilgi verir.

5.1.6 Temel Bileşenlerin Grafik Gösterimi

Şekil 5.1’de 2 boyutlu orijinal veri setindeki değişkenlerin rotasyonuyla elde edilen temel bileşenler görülmektedir. “Eğer boyut sayısı ikiden fazla ise, ilk iki temel bileşenin her bir gözlem için serpilme diyagramı çizilebilir. Düzlem üzerindeki en geniş veri yayılımını veren bu diyagram p boyutlu veri kümesinin projeksiyonuna eşdeğerdir.(Rencher,2002)” Şekil 5.3’de bir veri setine ilişkin ilk iki temel bileşenin SAS programında çizilmiş serpilme diyagramı yer almaktadır.



Şekil 5.3 İlk İki Temel Bileşenin Serpilme Diyagramı

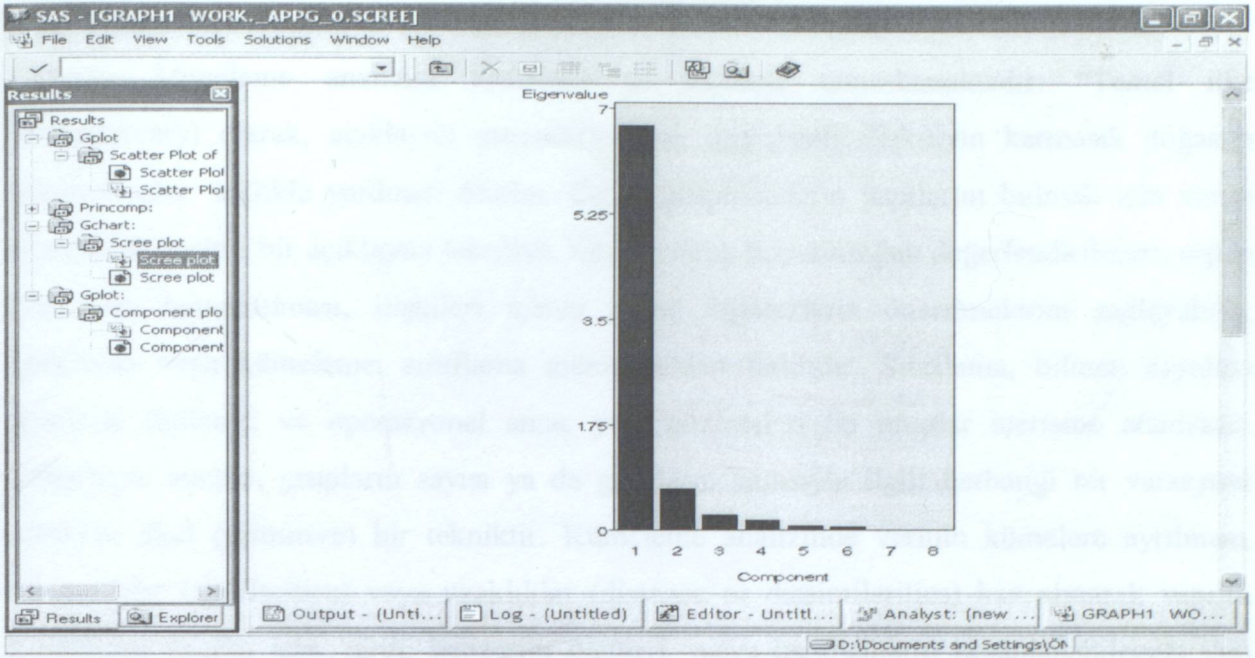
Gnanadesikan (1997) bu serpilme diyagramı sayesinde değişkenlerin arasındaki ilişkinin doğrusallığı, varyansı artıran ya da kovaryansı değiştiren sapan değerlerin varlığının anlaşılabilirliğine, ve bu işlem için yalnızca ilk iki ya da son iki temel bileşenin serpilme diyagramını incelemenin yeterli olacağına değinmiştir.

Bu serpilme diyagramlarından elde edilecek bir başka bilgi de birimlerin kümelenme eğilimidir. Şekil 5.3 incelenirse diyagramın uç noktalarında sapan değerlerin olduğu verinin genellikle orta noktalarda yoğunlaştığı söylenebilir.

5.1.7 Kaç Adet Temel Bileşenin Kullanılacağına Seçimi

Temel bileşenler analizi sonucunda elde edilen temel bileşenlerin sayısı orijinal veri setindeki değişkenlerin sayısına eşittir. Elde edilen tüm temel bileşenler genellikle kullanılmaz. Son olarak varyans açıklama oranı düşük olan birkaç temel bileşen dışlanır. Bu dışlama için değişik yöntemler söz konusudur. Bu yöntemler aşağıda yer almaktadır.

- 1- Toplam varyansın birikimli olarak %80'ninin açıklandığı temel bileşenden sonraki temel bileşenler göz ardı edilir.
- 2- Temel bileşenleri oluştururken özvektörlerin hesaplanmasında kullanılan özdeğerlerin ortalamasından yüksek olan özdeğerlere ait özvektörlerin oluşturdukları temel bileşenler alınır.
- 3- Yamaç grafiği testi yapılır. Bu testte özdeğerlere ait çizgi ya da çubuk grafiği çizilir. En büyük özdeğerden en küçük özdeğere doğru değişimde ortaya çıkan azalma görsel olarak bakıldığında oldukça küçükse, azalmanın düşük olduğu özdeğerden sonraki özdeğerlere ait özvektörlerin oluşturacakları temel bileşenler dışlanır. Şekil 5.4'te yamaç grafiği görülmektedir.



Şekil 5.4 Yamaç Grafiği

Şekil 5.4'teki yamaç grafiği incelenirse 3. sıradaki özdeğerden sonra değişim oranı azalmakta olduğundan, ilk 3 özdeğere ait özvektörün oluşturacağı temel bileşenler analizlerde kullanılır.

4- Büyük özdeğerlere ait büyük bileşenlerin anlamlılığının test edilmesi.

$\bar{\lambda} = \sum_{i=p-k+1}^p \lambda_i / k$ olmak üzere $u = \left(n - \frac{2p+11}{6} \right) \left(k \ln \bar{\lambda} - \sum_{i=p-k+1}^p \ln \lambda_i \right)$ test istatistiği hesaplanır. H_0

hipotezi ise $\lambda_{p-k+1} = \dots = \lambda_p$ 'dir. u test istatistiği $\chi^2_{\alpha, \frac{1}{2}(k-1)(k+1)}$ dağılır.

Yukarıdaki metotlar içerisinde uygulamada en çok tercih edilen metot 2. metottur.

5.2 Kümeleme Analizi

Johnson, kümeleme analizini kitabında şu biçimde tanımlamaktadır: “Temel ilke (Rudimentary) olarak, açıklayıcı prosedürler çok değişkenli ilişkilerin karmaşık doğasını anlamak için sıklıkla yardımcı olurlar. Doğal gruplamaların yapılarını bulmak için veriyi araştırmak önemli bir açıklayıcı tekniktir. Gruplamalar boyutluluğun değerlendirilmesi, sapan değerlerin tanımlanması, ilişkileri içeren ilginç hipotezlerin önerilmelerini sağlayabilir. Gruplama veya kümeleme, sınıflama metotlarından farklıdır. Sınıflama, bilinen sayıdaki gruplarla ilgilenir; ve operasyonel amaç yeni gözlemleri bu gruplar içerisine atamaktır. Kümeleme analizi, grupların sayısı ya da grupların yapısıyla ilgili herhangi bir varsayımı olmayan ilkel (primitive) bir tekniktir. Kümeleme analizinde verinin kümeler ayrılması, benzerlikler (similarities) veya uzaklıklar (distance or dissimilarities) baz alınarak yapılır. Kümeleme analizi için; girdi, benzerlik ölçüleri veya bu ölçülerin hesaplanabileceği veri setidir.(Johnson ve Wichern,1982)”

Kümeleme analizi için bir başka tanım da şu biçimde yapılmaktadır: “Kümeleme analizi, temel amacı nesneleri (birim) sahip oldukları karakteristik özellikleri baz alarak gruplamak olan çok değişkenli teknikler grubudur. Kümeleme analizi, nesneleri (cevaplayıcılar, ürünler veya diğer varlıklar (entity)) küme içerisinde çok benzer biçimde, kümeler arasında farklı olacak biçimde kümeler. Kümeleme işlemi başarılı olursa, bir geometrik çizim yapıldığında nesneler küme içerisinde birbirlerine çok yakın, kümeler ise birbirlerinden uzak olacaktır.(Hair vd,1995)”

Doğal gruplamaları gerçekleştirmenin zorluğunu gözler önüne sermek için 52’lik oyun kağıdı destesinde yer alan A,K,Q,J kartlarının gruplanmasını inceleyelim.

a) Birim olarak kartların gruplanması

	Maça	Sinek	Karo	Kupa
A	☐	☐	☐	☐
K	☐	☐	☐	☐
Q	☐	☐	☐	☐
J	☐	☐	☐	☐

b) Takım olarak kartların gruplanması

	Maça	Sinek	Karo	Kupa
A	☐	☐	☐	☐
K	☐	☐	☐	☐
Q	☐	☐	☐	☐
J	☐	☐	☐	☐

c) Renk takımlarına göre kartların gruplanması

	Maça	Sinek	Karo	Kupa
A	☐	☐	☐	☐
K	☐	☐	☐	☐
Q	☐	☐	☐	☐
J	☐	☐	☐	☐

d) Majör ve minör takımlara göre gruplama

	Maça	Sinek	Karo	Kupa
A	☐	☐	☐	☐
K	☐	☐	☐	☐
Q	☐	☐	☐	☐
J	☐	☐	☐	☐

Görüldüğü üzere 16 adet kart benzer özelliklerine göre değişik biçimlerde gruplanabilirler. Yukarıdaki dört adet gruplama daha da artırılabilir. Ama gerçek yaşamda verilerin olası bütün sınıflamalarını elde edip bunların içerisinde en iyisini seçmek olanaksız olabilir. Örneğin, bu 16 kart ikili gruplara ayrılmak istenirse, 32767 adet gruplama yapmak gerekir. Bunu yapmak zaman açısından imkansız olacağından, en iyi kümelemeyi yapmak için geliştirilen çeşitli algoritmalar kullanılır. Bu algoritmalarından birkaçı konunun ilerleyen bölümlerinde incelenecektir.

“Veri kümelemesi güçlü bir gelişme yaşamaktadır. Araştırmaya katkıda bulunan alanlar içerisinde veri madenciliği, istatistik, makine öğrenmesi, uzamsal (spatial) veritabanı teknolojisi biyoloji ve pazarlama yer alır. Veritabanlarında büyük ölçeklerde veri toplanması nedeni ile, veri madenciliği araştırmasında oldukça etkin bir konu haline gelmiştir.

İstatistiğin bir dalı olarak, kümeleme analizi yıllarca, uzaklık bazlı kümeleme baz alınarak çalışılmıştır. SAS, SPSS, S-PLUS gibi paket programlarda k-means, k-medoids, ve birçok farklı kümeleme metodu (algoritması) yer almıştır.

Veri madenciliğinde, büyük veritabanlarında etkin ve somut kümeleme analizi için metotlar

bulunması üzerine çalışılmıştır. Araştırmanın etkin konuları kümeleme metotlarının ölçülebilirliği, büyük veri tabanlarındaki nümerik ve kategorik datanın kümelenmesi için metotlar ve kompleks biçimde (shape) ve tipteki verinin kümelenmesi için metotların etkinliği olmuştur.(Han ve Kamber,2001”

Veri madenciliğinde kümeleme analizi için gerekli gereksinimler şunlardır:

Ölçeklendirilebilirlik (scalability): Birçok kümeleme algoritması 200’den az birimi bulunan veri setlerinde başarılıdır, ancak milyonlarca birimin yer aldığı veritabanlarında kullanılabilecek algoritmalara gereksinim vardır. Çünkü veritabanının belirli bir örneğinde kümeleme yapılması yanı sıra sonuçlara neden olabilir.

Değişik Tiplerdeki Değişkenlerle İşlem Yapabilme Yeteneği: Birçok veri madenciliği algoritması minimum aralık ölçekli veri setleri ile çalışmak üzere oluşturulmuştur. Fakat büyük veri tabanlarında nominal, ordinal, aralık ve oransal ölçekli değişkenler bir arada yer almaktadır.

Herhangi Bir Şekildeki Kümelerin Keşfi: Öklidyen ya da Manhattan uzaklık ölçülerine dayanan kümeleme algoritmaları küresel biçimde kümeler oluşturmaktadır. Bunların dışında farklı biçimlerde de kümeler söz konusu olabilir.

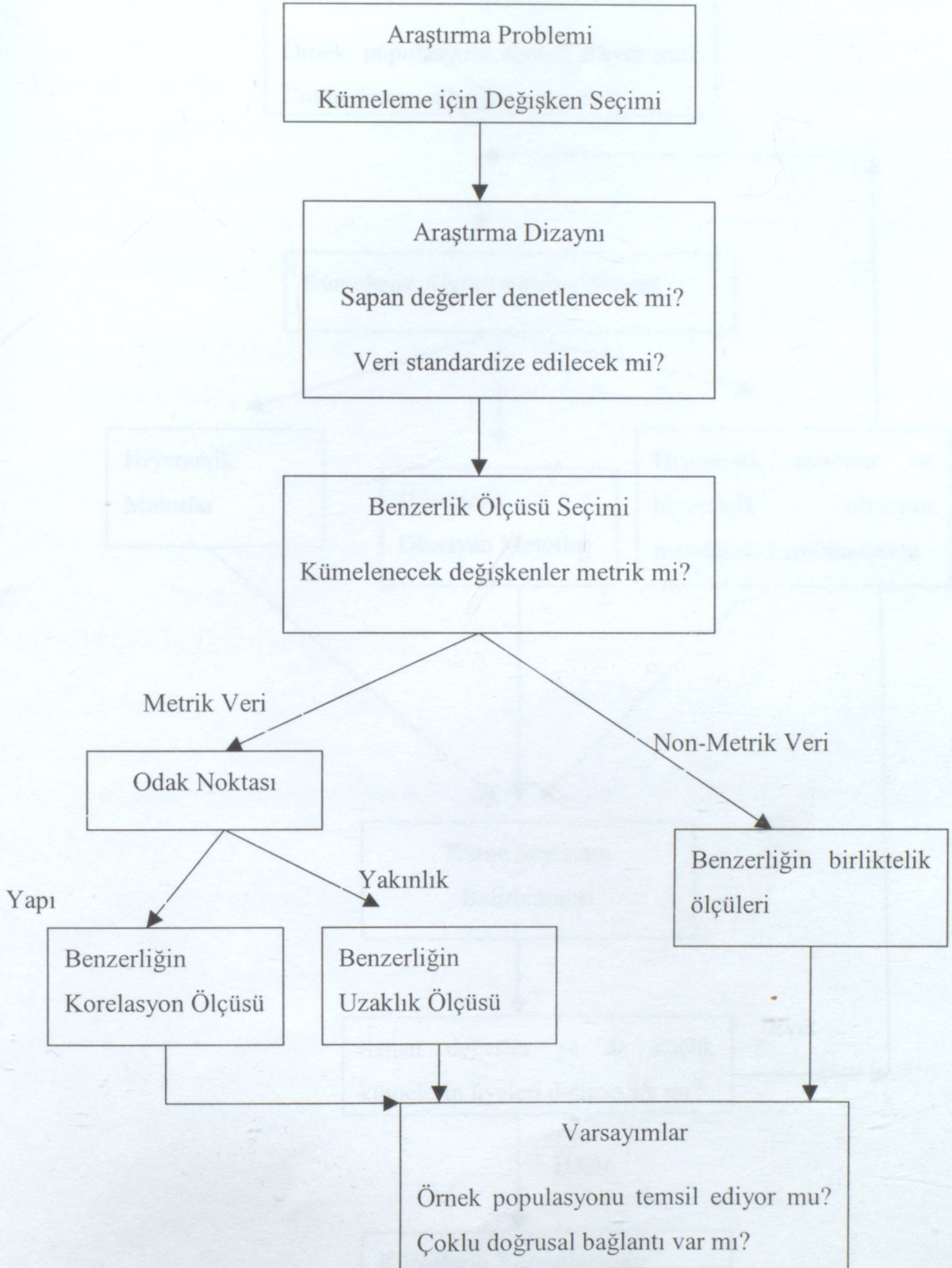
Girdi Parametrelerinin Seçiminde Minimum Alan Bilgisi Gereksinimi: Birçok kümeleme algoritması kümeleme analizi başlamadan önce küme sayısı, kullanılacak değişkenler gibi bazı parametrelere gereksinim duyar. Oldukça çok boyutu olan veritabanlarında karar vermek güç bir işlemdir.

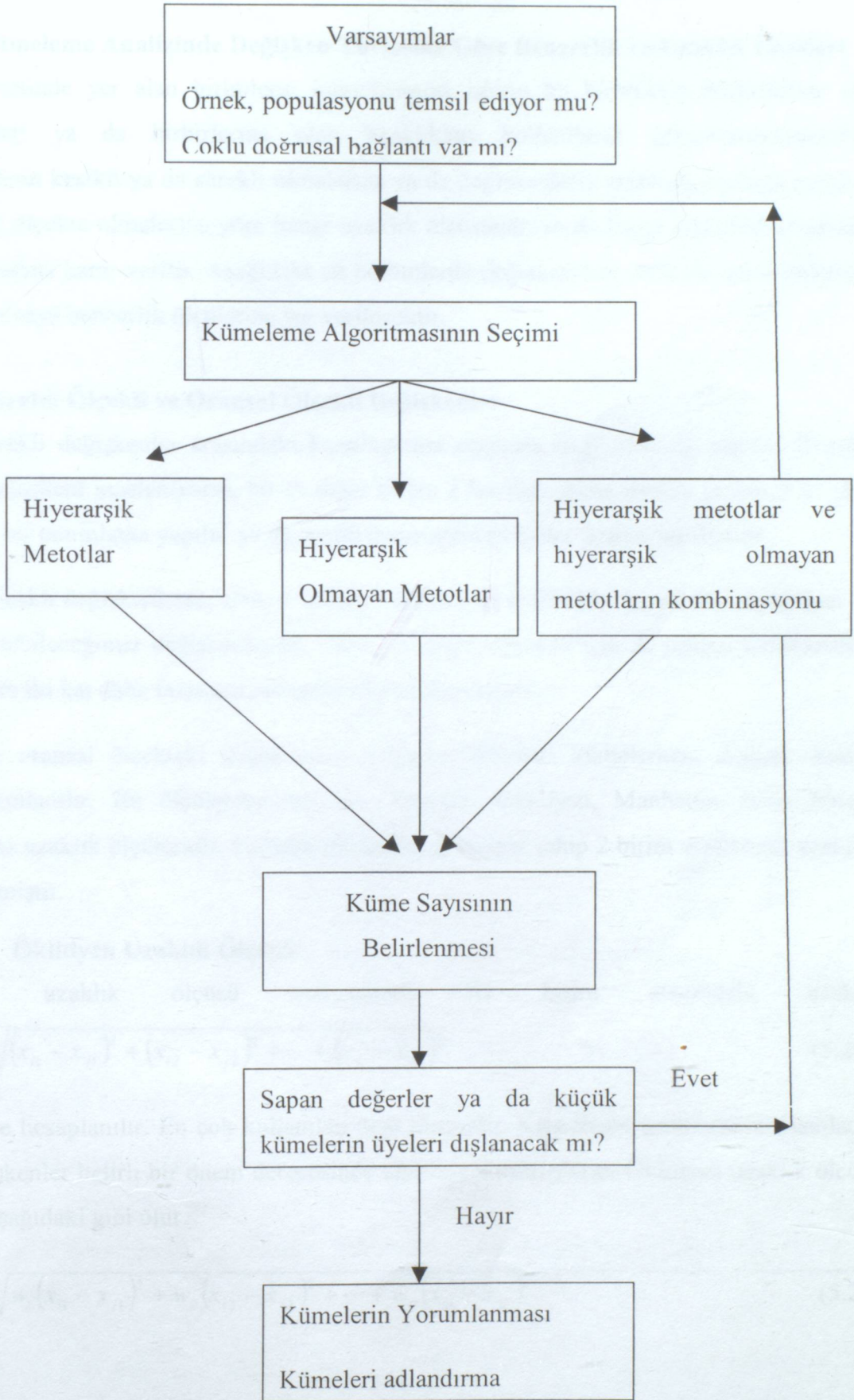
Gürültülü Veri İle Çalışma Yeteneği: Günümüzde birçok büyük veritabanında kayıp, eksik ya da yanlış verilere rastlanmaktadır. Bu tür veriler, kümeleme algoritmalarının oluşturacağı kümelerin iyi bir biçimde ayrışmasına neden olabilir.

Veri Setinin Sıralamasından Algoritmaların Etkilenmemesi: Bazı algoritmalar veri setinin sırasına göre küme merkezleri seçtikleri için, aynı veriyi ve aynı algoritmayı kullanarak farklı kümeler oluşturmak söz konusu olabilir.

Yorumlanabilirlik ve Kullanılabilirlik: Kümeleme analizi sonucunda oluşturulan kümeler özelliklerine göre ayrıştırılıp farklı analizlerde girdi olarak kullanılabilmelidir.

5.2.1 Kümeleme Analizi İçin Akış Diyagramı (Hair vd,1995)





5.2.2 Kümeleme Analizinde Değişken Türlerine Göre Benzerlik ve Uzaklık Ölçüleri

Bir veri setinde yer alan birimlerin kümelenmesi işlemi bu birimlerin birbirleriyle olan benzerlikleri ya da birbirlerine olan uzaklıkları kullanılarak gerçekleştirilmektedir. Değişkenlerin kesikli ya da sürekli olmalarına ya da değişkenlerin nominal, ordinal, aralık ya da oransal ölçekte olmalarına göre hangi uzaklık ölçüsünün ya da hangi benzerlik ölçüsünün kullanılacağına karar verilir. Aşağıdaki alt bölümlerde değişkenlerin türlerine göre kullanılan uzaklık ve/veya benzerlik ölçülerine yer verilecektir.

5.2.2.1 Aralık Ölçekli ve Oransal Ölçekli Değişkenler

Aralık ölçekli değişkenler arasındaki karşılaştırma oranıyla değil fark ile yapılır. Örneğin, sıcaklık değişkeni inceleniyorsa, bir ev diğer evden 2 kat daha sıcak demek yerine, 5 C° daha sıcak gibi bir tanımlama yapılır ya da suyun donmasına şu kadar derece var denilir.

Oransal ölçekli değişkenlerse; alan, uzunluk, uzaklık gibi aralarında kıyaslama yapılırken bir oran belirtebileceğimiz değişkenlerdir. Örneğin, oraya ulaşmak için A yolunu kullanırsak B yoluna göre iki kat daha fazla yol yürümüş oluruz, diyebiliriz.

Aralık ve oransal ölçekteki değişkenleri bulunan birimleri kümelerken, değişik uzaklık ölçüleri kullanılır. Bu ölçülerden bazıları; Pearson, Öklidyen, Manhattan (City block), Minkowski uzaklık ölçüleridir. Uzaklık formülleri p boyuta sahip 2 birim arasındaki uzaklığa göre verilmiştir.

5.2.2.1.1 Öklidyen Uzaklık Ölçüsü

Öklidyen uzaklık ölçüsü kullanılarak iki birim arasındaki uzaklık

$$d(i, j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{ip} - x_{jp})^2} \quad (5.20)$$

formülüyle hesaplanılır. En çok kullanılan ölçü birimidir. Kümeleme analizinde kullanılacak olan değişkenler belirli bir önem derecesinde ağırlıklandırılmışlarsa, Öklidyen uzaklık ölçüsü formülü aşağıdaki gibi olur.

$$d(i, j) = \sqrt{w_1(x_{i1} - x_{j1})^2 + w_2(x_{i2} - x_{j2})^2 + \dots + w_p(x_{ip} - x_{jp})^2} \quad (5.21)$$

5.2.2.1.2 Pearson Uzaklık Ölçüsü

Pearson uzaklık ölçüsü kullanılarak iki birim arasındaki uzaklık

$$d(i, j) = \sqrt{(x_{i1} - x_{j1})^2 / S_1^2 + (x_{i2} - x_{j2})^2 / S_2^2 + \dots + (x_{ip} - x_{jp})^2 / S_p^2} \quad (5.22)$$

formülü ile hesaplanır. Bu formülde kullanılan S_p , uzaklığın hesaplandığı değişkene ait varyanstır. Bununla birlikte farklı gruplar hakkında önceden bilgi sahibi olunmadığı için, uzaklık hesaplanmasında S değerinin kullanılması doğru olmaz. Bu nedenle Pearson uzaklık ölçüsü yerine genellikle Ökliden uzaklık ölçüsü tercih edilir. Kümeleme analizinde kullanılacak olan değişkenler belirli önem derecelerine göre ağırlıklandırılmışlarsa, Pearson uzaklık ölçüsü formülü aşağıdaki gibi olur.

$$d(i, j) = \sqrt{w_1(x_{i1} - x_{j1})^2 / S_1^2 + w_2(x_{i2} - x_{j2})^2 / S_2^2 + \dots + w_p(x_{ip} - x_{jp})^2 / S_p^2} \quad (5.23)$$

Pearson uzaklık ölçüsüne, “karesel Pearson uzaklığı” ya da “standardize öklid uzaklığı” adı da verilir.

5.2.2.1.3 Manhattan Uzaklık Ölçüsü

Manhattan uzaklık ölçüsü kullanılarak iki birim arasındaki uzaklık

$$d(i, j) = (|x_{i1} - x_{j1}| + |x_{i2} - x_{j2}| + \dots + |x_{ip} - x_{jp}|) \quad (5.24)$$

formülü ile hesaplanır. Bu ölçü de birimler arasındaki mutlak uzaklık kullanılır. Kümeleme analizinde kullanılacak olan değişkenler eğer belirli bir önem derecesinde ağırlıklandırılmışlarsa, Manhattan uzaklık ölçüsü formülü aşağıdaki gibi olur.

$$d(i, j) = (w_1|x_{i1} - x_{j1}| + w_2|x_{i2} - x_{j2}| + \dots + w_p|x_{ip} - x_{jp}|) \quad (5.25)$$

Manhattan uzaklık ölçüsüne, “city block uzaklık ölçüsü” adı da verilir.

5.2.2.1.4 Minkowski Uzaklık Ölçüsü

Minkowski uzaklık ölçüsü genel bir formüldür. Formülde yer alan m değerinin alacağı farklı değerlere göre yeni formüller türetir. Minkowski uzaklık ölçüsü kullanılarak iki birim arasındaki uzaklık

$$d(i, j) = \left[|x_{i1} - x_{j1}|^m + |x_{i2} - x_{j2}|^m + \dots + |x_{ip} - x_{jp}|^m \right]^{1/m} \quad (5.26)$$

formülü ile hesaplanır. Kümeleme analizinde kullanılacak olan değişkenler belirli önem derecelerine göre ağırlıklandırılmışlarsa, Minkowski uzaklık ölçüsü formülü aşağıdaki gibidir.

$$d(i, j) = \left[w_1 |x_{i1} - x_{j1}|^m + w_2 |x_{i2} - x_{j2}|^m + \dots + w_p |x_{ip} - x_{jp}|^m \right]^{1/m} \quad (5.27)$$

Minkowski uzaklık ölçüsündeki m değeri büyük ve küçük farklara verilen ağırlığı değiştirir. m=1 değerini alırsa, formül, Manhattan uzaklık ölçüsünün formülüne, m=2 değerini alırsa, formül Öklidyen uzaklık ölçüsü formülüne dönüşür.

5.2.2.1.5 Uzaklık Fonksiyonunun Özellikleri

1. $d(i, j) \geq 0$; Uzaklık negatif değildir
2. $d(i, i) = 0$; Her birimin kendisine olan uzaklığı sıfırdır.
3. $d(i, j) = d(j, i)$; Uzaklık fonksiyonu simetriktir.
4. $d(i, j) \leq d(i, h) + d(h, j)$; iki birimin arasındaki uzaklık bu iki birimin üçüncü bir birime olan uzaklıkları toplamından küçük olamaz.

5.2.2.1.6 Değişkenlerin Standardizasyonu ve Değişkenlerin Dönüştürülmesi

Veri matrisinde yer alan değişkenlerin ortalamaları ve varyansları birbirlerinden oldukça farklı olduğu durumlarda birimler arası uzaklıklar hesaplanırken, ortalaması daha büyük olan ve/veya varyansı daha büyük olan değişkenler, hesaplanılan uzaklık değerine daha büyük etki yapacaktır. Bu durumda kümeler oluşturulurken sistematik bir yanlış yapılmış olacaktır. Ayrıca değişkenlerde yer alan aşırı değerlerde uzaklık değerine etki eden başka bir faktördür. Aşırı değerler kümeleme analizi sonucunda ayrı kümeler olarak karşımıza çıkabilir. Bu gibi durumlar söz konusu olduğunda yapılacak işlem, değişkenlerin transforme edilmesidir. Verilerin transforme edilme işlemi, standardizasyon ya da belirli aralıklara indirgeme biçiminde gerçekleşir. Verilerin dönüştürülmesinde kullanılan bazı teknikler; z skorlarına dönüştürme, $-1 \leq x \leq +1$ aralığına indirgeme, $0 \leq x \leq +1$ aralığına indirgeme, ortalama 1 olacak biçimde indirgeme, standart sapma 1 olacak şekilde indirgeme, maksimum değer bir olacak biçimde indirgemedir.

- Z skorlarına dönüştürme

Bu yöntem, oransal ve aralık ölçekli veriler söz konusu olduğunda verilerin çok değişkenli normal dağılım gösterdikleri varsayımıyla verilere uygulanan bir yöntemdir. En çok tercih edilen yöntemdir. Veri madenciliği çalışmalarında veri sayımız çok büyük olduğundan değişkenlerin çok değişkenli normal dağılım gösterdikleri varsayılabilir.

$$z_{ik} = \frac{x_{ik} - \bar{x}_k}{s_k} \quad (5.28)$$

- $-1 \leq x \leq +1$ aralığına indirgeme

Eğer veri seti heterojen bir yapıda ise ve aşırı değerler söz konusu ise bu yöntem tercih edilir. Dönüştürülecek değişkenlerde + ya da - değerler var olduğunda ve $|x_{\min}| \leq x_{\max}$ olduğu durumlarda uygulanır. Dizideki en büyük değer $x_{\max}$ ise, formül indirgeme formülü şöyle oluşur:

$$x'_i = \frac{x_i}{x_{\max}} \quad (5.29)$$

- $0 \leq x \leq +1$ aralığına indirgeme

Veri seti heterojen yapıda ise ve aşırı değerler söz konusu ise değişkenlerin değerleri 0 ile 1 aralığına dönüştürülebilir. Dizideki en büyük değer $x_{\max}$, dizideki en küçük değer $x_{\min}$ ve rank $R = x_{\max} - x_{\min}$ olmak üzere indirgeme formülü şöyle oluşur:

$$x'_i = \frac{x_i - x_{\min}}{R} \quad (5.30)$$

- Ortalama 1 olacak biçimde indirgeme

Oluşturulacak olan indirgenmiş değişkenin ortalamasının pozitif ve 1 olması gerektiğinde uygulanan bir yöntemdir. Dönüştürme işlemi şu formülle yapılır:

$$x'_i = \frac{x_i}{\bar{x}} \quad (5.31)$$

Eğer ortalama sıfıra eşit ise formül şöyle olur:

$$x'_i = \frac{x_i + 1}{\bar{x} + 1} \quad (5.32)$$

- Standart sapma 1 olacak biçimde indirgeme

Eğer indirgenmiş değişkenin standart sapmasının 1 olması isteniyorsa, bu yöntem tercih edilir. İndirgeme formülü şöyledir:

$$x'_i = \frac{x_i}{s_x} \quad (5.33)$$

Orijinal değişkenin standart sapması 0 ise, bu dönüşüm uygulanmaz. Dönüşümün şart olduğu düşünülürse, diğer yöntemlerden uygun olan bir yöntem seçilir.

- Maksimum değer bir olacak biçimde indirgeme

İndirgenmiş değişkenin maksimum değerinin 1 olması isteniyorsa bu yöntem tercih edilir. İndirgeme formülü şöyledir:

$$x'_i = \frac{x_i}{x_{\max}} \quad (5.34)$$

Eğer dizide maksimum değer sıfır ise formül şöyle olur:

$$x'_i = \frac{x_i}{|x_{\min}|} + 1 \quad (5.35)$$

5.2.2.2 Nominal Ölçekli Değişkenler

Nominal ölçekli veriler aralarında büyüklük ya da küçüklük gibi herhangi bir karşılaştırmanın yapılamayacağı; sadece saç rengi, cinsiyet, medeni durum gibi nitel değişkenlerin sınıklarının 0,1,2 vb. kesikli nicel değerlere atanması ile oluşan değişkenlerdir. Nominal ölçekli değişkenleri kümeleme analizinde incelerken, ikili (binary), ikili olmayan nominal ölçekli değişkenler olarak ikiye ayırabiliriz.

5.2.2.2.1 İkili Değişkenler

Erkek-Kadın, Evet-Hayır, Olumlu-Olumsuz gibi salt iki değer alan nominal ölçekli değişkenlerdir. Bu değişkenlerde birimler arası uzaklık hesaplanırken aralık ve/veya oransal ölçekli verilerin hesaplanmasında kullanılan Pearson, Öklidyen, Manhattan (City block), Minkowski uzaklık ölçüleri tercih edilmez. Bu tür değişkenlerin kümelenmesinde benzerlik ölçüleri adı verilen ölçüler kullanılır. Örneğin, elimizde ikili değişkenlerin yer aldığı aşağıdaki veri seti olsun. Bu veri setinde öklidyen uzaklığı hesaplayalım.

Çizelge 5.1 İkili Değişkenlerden Oluşan Örnek Kümeleme Veri Seti

	D1	D2	D3	D4	D5
Birim 1	0	1	0	1	1
Birim 2	0	1	0	0	1
Birim 3	0	0	0	0	1

Çizelge 5.1’de yer alan örnek veri seti için kareli öklidyen uzaklık ölçüsünü hesaplayalım.

$$d(birim1, birim2) = (x_{birim1,p} - x_{birim2,p})^2 = (0-0)^2 + (1-1)^2 + (0-0)^2 + (1-0)^2 + (1-1)^2$$

$$d(birim1, birim2) = 1$$

İki birim arasındaki öklidyen uzaklığın karesi 1 çıkmıştır. Birim 3 ile birim 2 arasındaki öklidyen uzaklığın karesini de benzer biçimde hesaplayalım.

$$d(birim3, birim2) = (x_{birim1,p} - x_{birim2,p})^2 = (0-0)^2 + (0-1)^2 + (0-0)^2 + (0-0)^2 + (1-1)^2$$

$$d(birim2, birim2) = 1$$

Birim 3 ile birim 2 arasındaki öklidyen uzaklığın karesi de bir çıkmıştır. Birim 3 ile birim 1 verileri bakımından farklı değişkenler olmalarına karşın her iki birimin de birim 2’ye uzaklıkları aynı hesaplanmıştır. Hesaplanan değerler aynı zamanda eşleşmeyen değişken sayısına eşittir. Buna alternatif olarak kullanılan bir yöntem de, kontenjans tablosu yardımı ile hesaplanan benzerlik ölçüleridir. Kontenjans tablosundan elde edilen 1-1, 0-0, 0-1, 1-0 eşleşmelerin frekansları kullanılarak sözü edilen benzerlik ölçüleri hesaplanır. Çizelge 5.2’de kontenjans tablosu yer almaktadır.

Çizelge 5.2 İkili Değişkenler İçin Kontenjans Tablosu

		Birim j		
Birim i		1	0	Toplam
	1	a	b	a+b
	0	c	d	c+d
	Toplam	a+c	b+d	p=a+b+c+d

Çizelge 5.2’de yer alan kontenjans tablosundaki frekans değerleri kullanılarak aşağıdaki çizelge 5.3’de yer alan 8 adet benzerlik ölçüsü elde edilir.

Çizelge 5.3 Benzerlik Ölçüleri Formülleri (Johson ve Wichern,1982)

Katsayı	Rasyonel
1. $\frac{a+d}{p}$	1-1 ve 0-0 eşleşmeler için eşit ağırlık
2. $\frac{2(a+d)}{2(a+d)+b+c}$	1-1 ve 0-0 eşleşmeler için çift (double) ağırlık
3. $\frac{a+d}{a+d+2(b+c)}$	Eşleşmeyen çiftler için çift ağırlık
4. $\frac{a}{p}$	Payda 0-0 eşleşme yok
5. $\frac{a}{a+b+c}$	Payda ve payda da 0-0 eşleşme yok. 0-0 eşleşmeler ilişkisiz düşünülür.
6. $\frac{2a}{2a+b+c}$	Payda ve payda da 0-0 eşleşme yok. 1-1 eşleşmelere çift ağırlık
7. $\frac{a}{a+2(b+c)}$	Payda ve payda da 0-0 eşleşme yok Eşleşmeyen çiftlere çift ağırlık
8. $\frac{a}{b+c}$	0-0 eşleşmeler dışlandığında eşleşen çiftlerin eşleşmeyen çiftlere oranı.

Elimizde 2 adet farklı kontenjans tablosu olduğunu düşünelim. Çizelge 5.3'de yer alan 1 numaralı katsayıyı her iki tablo içinde hesaplamış olalım. 1. kontenjans tablosundan elde

edilen 1 numaralı katsayı $\frac{a_1+d_1}{p_1}$, 2. kontenjans tablosundan elde edilen 1 numaralı katsayı

ise $\frac{a_2+d_2}{p_2}$ olsun. Eğer $\frac{a_1+d_1}{p_1} \geq \frac{a_2+d_2}{p_2}$ ilişkisi sağlanmış ise hesaplanacak 2 numaralı

katsayılar için de $\frac{2(a_1+d_1)}{2(a_1+d_1)+b_1+c_1} \geq \frac{2(a_2+d_2)}{2(a_2+d_2)+b_2+c_2}$ eşitsizliği, aynı zamanda 3

numaralı katsayılar için de $\frac{a_1+d_1}{a_1+d_1+2(b_1+c_1)} \geq \frac{a_2+d_2}{a_2+d_2+2(b_2+c_2)}$ eşitsizliği sağlanır. İlk

3 katsayı da monotonik (Monotonicity) ilişkiden hoşlanan katsayılardır. 5,6,7 numaralı katsayılarda kendi aralarında benzer eşitsizlikleri sağlarlar. Monotoniklik çok önemli bir

durumdur. Single Linkage ve complete linkage hiyerarşik kümeleme metotları ilk 3 katsayı ile hesaplanan benzerlik ölçüleri kullanılarak hesaplandığında aynı kümeleri oluştururlar. Benzer biçimde 5,6,7 numaralı katsayılarda aynı kümeleri oluştururlar.

5.2.2.2 İkili Olmayan Nominal Değişkenler

Şıklarında iki durumdan daha fazla durum bulunduran değişkenlerin uzaklıklarının hesaplanması ise şu formülle sağlanır:

$$d(i, j) = \frac{p - m}{p} \quad (5.36)$$

p: toplam değişken sayısı, m: eşleşen değişken sayısı olmak üzere eşleşmeyen değişken sayısı (p-m) toplam değişken sayısına oranlanarak iki birim arasındaki uzaklık bulunur. Bu uzaklık 1'den çıkarılırsa, iki değişken arasındaki benzerlik katsayısı bulunur.

5.2.2.3 Ordinal Ölçekli Değişkenler

Ordinal değişkenler, aralarında bir sıra bulunan, ama sıralar arasındaki farkın ya da oranın yalnızca hangi birimin ön sırada yer aldığını gösteren kesikli değişkenlerdir. Örneğin, bir kurumdaki kişilerin unvanlarına göre sıralanması ya da bir sınav sonucunda öğrencilerin yalnızca sınavda almış oldukları dereceler ordinal ölçeklidir. Ordinal ölçekli değişkenlere sahip birimlerin kümelenmesi 3 aşamada gerçekleşir.

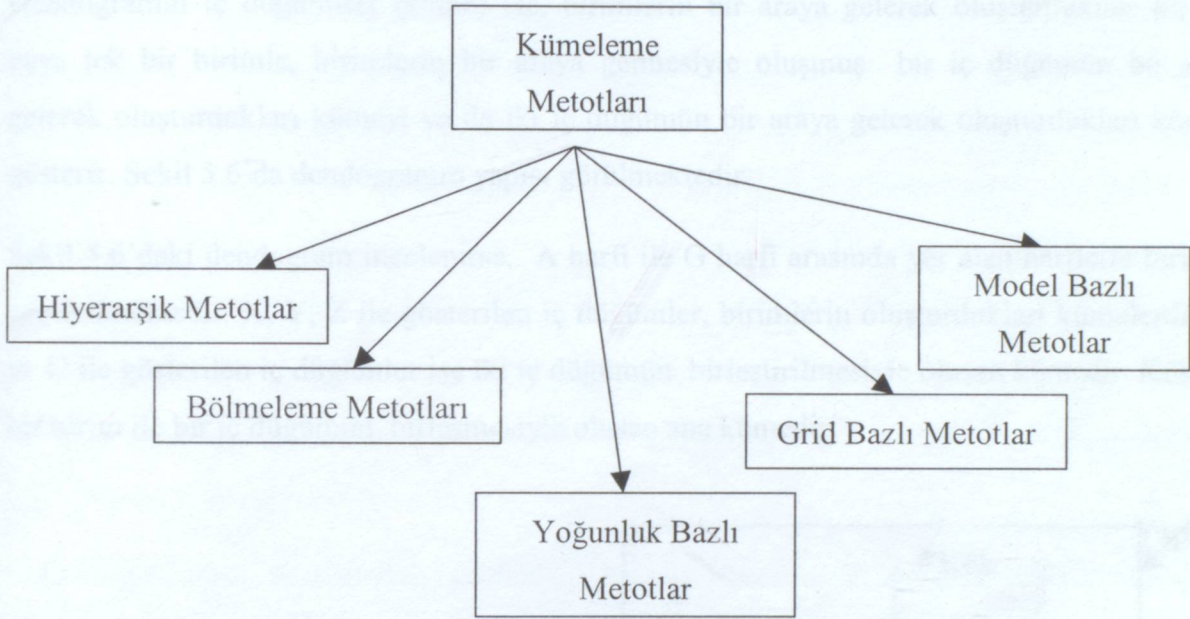
- x_f değişkenine ait i. birimin sıralaması M birim içerisinde r_{if} olur.
- Her bir ordinal değişken farklı sayıda durumlara sahip olacağından, her bir değişken aşağıdaki formül ile [0,1] aralığına indirgenir.

$$z_{if} = \frac{r_{if} - 1}{M_f - 1} \quad (5.37)$$

- [0,1] aralığına indirgenen değişkenlere ait birimler arasındaki uzaklıklar oransal ve aralık ölçekte kullanılan uzaklık ölçüleri ile bulunur.

5.2.3 Kümeleme Metotlarının Sınıflandırılması

Veri madenciliğinde, uygulamanın amacına, veri tipine, verinin büyüklüğüne göre kullanımda tercih edilebilecek farklı kümeleme metotları kullanılmaktadır. Değişik kaynaklarda kümeleme metotları farklı biçimlerde sınıflandırılmaktadır. Kümeleme metotlarının sınıflandırılmasında çok değişkenli veri analizi kitaplarında kullanılan en genel ayırım hiyerarşik ve hiyerarşik olmayan kümeleme metotları ayırımıdır. Veri madenciliği teknikleriyle ilgili kaynaklarda kullanılan kümeleme metotlarının sınıflandırılmasına ilişkin bir sınıflama örneği aşağıda yer almaktadır.



Şekil 5.5 Veri Madenciliğinde Kullanılan Kümeleme Metotlarının Sınıflandırılması
(Han ve Kamber 2001)

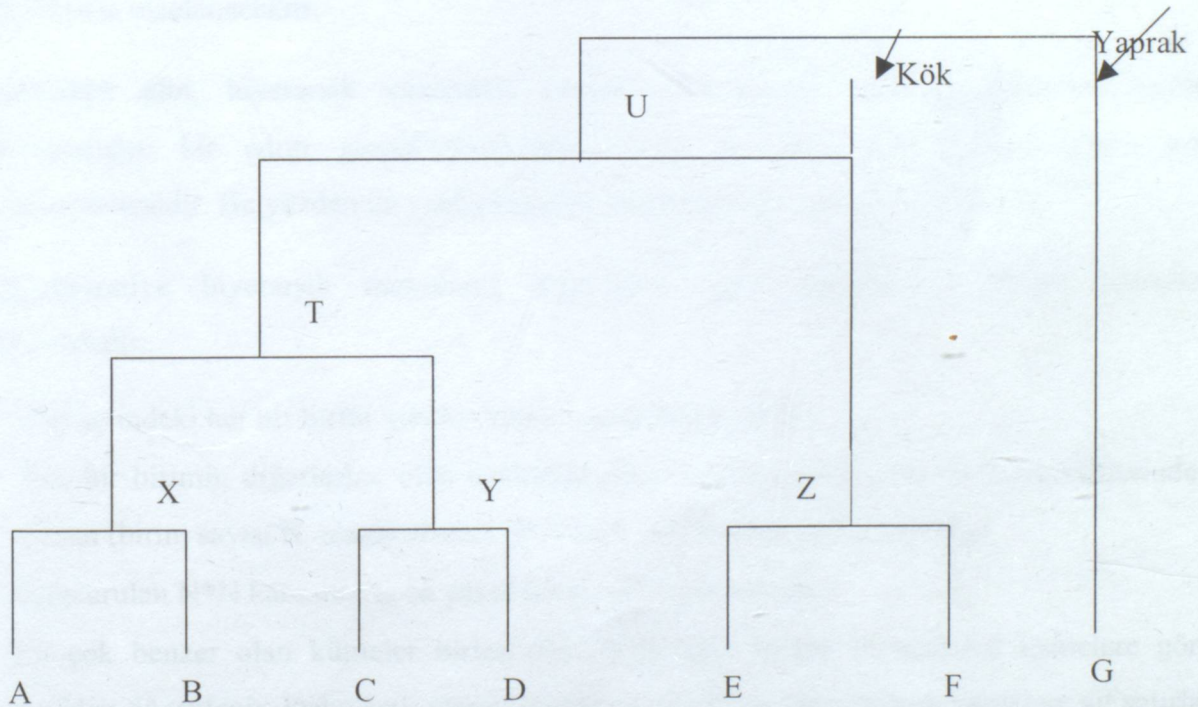
Şekil 5.5’de görülen kümeleme metotlarından hiyerarşik metotlar ve bölmeleme metotları çok değişkenli istatistik konusu içerisinde incelendiğinden, bu yöntemlere ağırlıklı olarak değinilecektir. Hiyerarşik metotlardan Single Link, Average Link, Complete Link algoritmaları ile bölmeleme (partitioning) metotlarından PAM, K- means, K-modes, algoritmaları incelenecektir. Yoğunluk bazlı metotlar, Grid bazlı metotlar ile Model bazlı metotlara ise genel olarak değinilecektir.

5.2.3.1 Hiyerarşik Kümeleme Metotları

Hiyerarşik kümeleme metotları, veri setinin birimlerinin birbirlerine olan uzaklık değerlerini kullanarak, veri setindeki birimlerin hiyerarşik ayrıştırmasını yapar. Hiyerarşik ayrıştırma sırasında, “ağaç veri yapısı” olarak da bilinen dendogram kullanılır. Dendogram, hiyerarşik kümeleme tekniğiyle elde edilen kümelerin görselleştirilmesini sağlar. Bir dendogramın yapısını kökler, iç düğümler ve yapraklar oluşturur.

Dendogramın kökü tüm birimlerin bir araya gelmesiyle oluşan ana kümeyi içerir. Dendogramın yapraklarıysa bir araya getirilmeyen tek bir birimden oluşan kümeleri içerir. Dendogramın iç düğümleri (küme) ise, birimlerin bir araya gelerek oluşturdukları kümeyi veya tek bir birimle, birimlerin bir araya gelmesiyle oluşmuş bir iç düğümün bir araya gelerek oluşturdukları kümeyi ya da iki iç düğümün bir araya gelerek oluşturdukları kümeyi gösterir. Şekil 5.6’da dendogramın yapısı görülmektedir.

Şekil 5.6’daki dendogram incelenirse, A harfi ile G harfi arasında yer alan harflerle birimler gösterilmektedir. X, Y, Z ile gösterilen iç düğümler, birimlerin oluşturdukları kümelerdir. T ve U ile gösterilen iç düğümler ise iki iç düğümün birleştirilmesiyle oluşan kümedir. Kök ise, bir birim ile bir iç düğümün birleşmesiyle oluşan ana kümedir.



Şekil 5.6 Dendogram

Hiyerarşik kümeleme metodunda, hiçbir birim küme dışında kalmaz. Şekil 5.6'dan da görüldüğü gibi, ilk aşamada A ile B birimleri bir araya gelerek X iç düğümünü (küme), C ile D birimleri bir araya gelerek Y iç düğümünü, E ile F birimleri bir araya gelerek Z iç düğümünü oluşturmuşlardır. G birimi tek başına bir küme olarak kalmıştır. İkinci aşamada, X ve Y iç düğümleri T iç düğümünü oluşturmuşlardır. Z iç düğümü ile G birimi yine bir önceki aşamadaki durumlarında kalmışlardır. Üçüncü aşamada ise, T ve Z iç düğümleri U iç düğümünü oluşturmuşlardır. En son aşamada ise U iç düğümü G birimi ile birleşerek kökü oluşturmuştur.

Dendogramın yapısının oluşturulmasına göre hiyerarşik kümeleme yöntemleri ikiye ayrılır. Bu yöntemler, agglomerative hiyerarşik kümeleme yöntemleri ile bölücü (divisive) kümeleme yöntemleridir. Hiyerarşik yapı oluşturulurken dendogramın kökünden birimlere doğru iniliyor ise bu yöntem bölücü hiyerarşik kümeleme yöntemi adı verilir. Bu durumun tersi söz konusu ise, yani kümeleme işlemi yapılırken her bir birim ayrı bir küme olarak düşünülüp ana küme elde ediliyorsa bu yöntem agglomerative hiyerarşik kümeleme yöntemi adı verilir.

Alt bölümlerde, agglomerative hiyerarşik kümeleme metodunun single linkage, average linkage ve complete linkage algoritmaları ve bölücü hiyerarşik kümeleme metodunun MST algoritması incelenecektir.

Görüldüğü gibi, hiyerarşik yöntemler iteratif yöntemlerdir. Bu işlemlerin en büyük olumsuzluğu, bir adım gerçekleştirildikten sonra bir daha tekrar aynı adıma geri dönülebilmesidir. Bu yüzden de yanlış kararları doğrulamaya izin vermemektedir.

Agglomerative hiyerarşik metodların algoritması genel olarak şu iteratif süreçleri içermektedir:

- Veri setindeki her bir birim ayrı bir küme olarak kabul edilir.
- Her bir birimin diğerlerine olan uzaklıklarından uzaklıklarından ya da benzerliklerinden oluşan (birim sayısı N olmak üzere) $N \times N$ kare matris oluşturulur. $D = \{d_{ik}\}$
- Oluşturulan $N \times N$ kare matris en yakın küme çiftleri araştırılır.
- En çok benzer olan kümeler birleştirilir. $N \times N$ kare matris birleştirilen kümelere göre yeniden düzenlenir. Daha açık olarak söylemek gerekirse, birleştirilen kümelere ait satırlar ve sütunlar silinir ve yeni oluşturulan küme veya kümeler ve diğer kümelerin oluşturduğu $(N-1) \times (N-1)$ kare matris oluşturulur.
- 3. ve 4. adımlar $N-1$ kez tekrar edilir ve dendogramın kök kısmı oluşunca kümeleme işlemi sona erer.

Single Linkage, Complete Linkage ve Average Linkage metotları uygulamaları için kullanılacak olan veri seti Yıldız Teknik Üniversitesi Fen Edebiyat Fakültesi bölümlerine 1998-1999-2000 yıllarında yerleştirilmiş olan öğrencilerin maksimum ve minimum puanları olacaktır. Çizelge 5.4'te veri tablosu yer almaktadır. Bölümler arası öklidyen uzaklıklarında Çizelge 5.5'de yer almaktadır.

Çizelge 5.4 Y.T.Ü. Fen Edebiyat Fakültesi Bölümlerine 1998-1999-2000 Yıllarında Yerleştirilmiş Olan Öğrencilerin Maksimum ve Minimum Puanları

	1998		1999		2000	
	Min	Max	Min	Max	Min	Max
- Fizik	188,381	197,205	184,242	196,360	181,664	194,122
- İstatistik	192,101	202,255	191,632	197,595	191,424	205,385
- Kimya	192,951	199,645	189,708	196,262	187,394	195,328
- Matematik	199,611	210,892	195,652	204,295	195,833	208,605
- Mütercim Tercümanlık (Frn.)	127,672	164,909	140,037	157,870	134,582	155,750

Çizelge 5.5 Y.T.Ü. Fen Edebiyat Fakültesi Bölümlerinin Arasındaki Öklidyen Uzaklık Matrisi

	Fizik	İstatistik	Kimya	Matematik	Mütercim Tercümanlık (Frn.)
Fizik	0	17,821	9,540	30,284	108,873
İstatistik	17,821	0	11,419	14,895	124,421
Kimya	9,540	11,419	0	22,764	117,323
Matematik	30,284	14,895	22,764	0,000	138,142
Mütercim Tercümanlık (Frn.)	108,873	124,421	117,323	138,142	0

5.2.3.1.1 Single Linkage

Single linkage yöntemi için girdi, birim çiftleri arasındaki uzaklık ya da benzerliklerin oluşturduğu $N \times N$ kare matristir. Kümeler, her biri ayrı bir küme olarak kabul edilen birimlerin birleştirilmesiyle oluşturulur. En yakın komşular, yani minimum uzaklığa ya da maksimum benzerliğe sahip olan küme çiftleri bir araya getirilir.

Eğer uzaklıklardan oluşan $N \times N$ kare matrisi kullanıyorsak, $D = \{d_{ik}\}$ en küçük uzaklık olmak üzere i . ve k . birimler bir araya getirilir. Yeni uzaklık matrisindeki uzaklıklar hesaplanırken i . ve k . birimlere ait uzaklıklar silinir. Bu uzaklıkların yerine yeni oluşan kümenin diğer birimlere olan uzaklıkları hesaplanır.

Yeni oluşturulan kümenin diğer kümelere uzaklığı ise U, V ve W kümeler olmak üzere aşağıdaki formül ile hesaplanır:

$$d_{(UV)W} = \min \{d_{UW}, d_{VW}\}$$



Single Linkage yönteminde oluşturulan kümeler dendogram aracılığıyla gösterilebilir.

Örnek:

	Fizik	İstatistik	Kimya	Matematik	Mütercim Tercümanlık (Frn.)
Fizik	0	17,821	9,540	30,284	108,873
İstatistik	17,821	0	11,419	14,895	124,421
Kimya	9,540	11,419	0	22,764	117,323
Matematik	30,284	14,895	22,764	0	138,142
Mütercim Tercümanlık (Frn.)	108,873	124,421	117,323	138,142	0

Minimum Uzaklık

Fizik ve Kimya kümeleri birleştirilir.

$$d_{(Fizik-Kimya)İstatistik} = \min \{d_{Fizik-İstatistik}, d_{Kimya-İstatistik}\}$$

$$d_{(Fizik-Kimya)İstatistik} = \min \{17.821, 11.419\} = 11.419$$

$$d_{(Fizik-Kimya)Matematik} = \min \{d_{Fizik-Matematik}, d_{Kimya-Matematik}\}$$

$$d_{(Fizik-Kimya)Matematik} = \min \{30.824, 22.764\} = 22.764$$

$$d_{(Fizik-Kimya)Müt-Ter} = \min \{d_{Fizik-Müt-Ter}, d_{Kimya-Müt-Ter}\}$$

$$d_{(Fizik-Kimya)Müt-Ter} = \min \{108.873, 117.323\} = 108.873$$

Minimum Uzaklık

	Fizik - Kimya	İstatistik	Matematik	Mütercim Tercümanlık (Frn.)
Fizik -Kimya	0	11,419	22,764	108,873
İstatistik	11,419	0	14,895	124,421
Matematik	22,764	14,895	0	138,142
Mütercim Tercümanlık (Frn.)	117,323	124,421	138,142	0

Fizik-Kimya bölümlerinin oluşturdukları küme ile İstatistik bölümü birleştirilir.

$$d_{(Fizik-Kimya-\text{İstatistik})Matematik} = \min\{d_{Fizik-Kimya,Matematik}, d_{\text{İstatistik},Matematik}\}$$

$$d_{(Fizik-Kimya-\text{İstatistik})Matematik} = \min\{22.764, 14.895\} = 14.895$$

$$d_{(Fizik-Kimya-\text{İstatistik})Müt-Ter} = \min\{d_{Fizik-Kimya,Müt-Ter}, d_{\text{İstatistik},Müt-Ter}\}$$

$$d_{(Fizik-Kimya-\text{İstatistik})Müt-Ter} = \min\{108.873, 124.421\} = 108.873$$

Minimum Uzaklık

	Fizik –Kimya – İstatistik	Matematik	Mütercim Tercümanlık (Frn.)
Fizik –Kimya- İstatistik	0	14.895	108,873
Matematik	14.895	0	138,142
Mütercim Tercümanlık (Frn.)	117,323	138,142	0

Fizik-Kimya-İstatistik bölümlerinin oluşturdukları küme ile Matematik bölümü birleştirilir.

$$d_{(Fizik-Kimya-\text{İstatistik}-Matematik)Müt-Ter} = \min\{d_{Fizik-Kimya-\text{İstatistik},Müt-Ter}, d_{Matematik-MütTer}\}$$

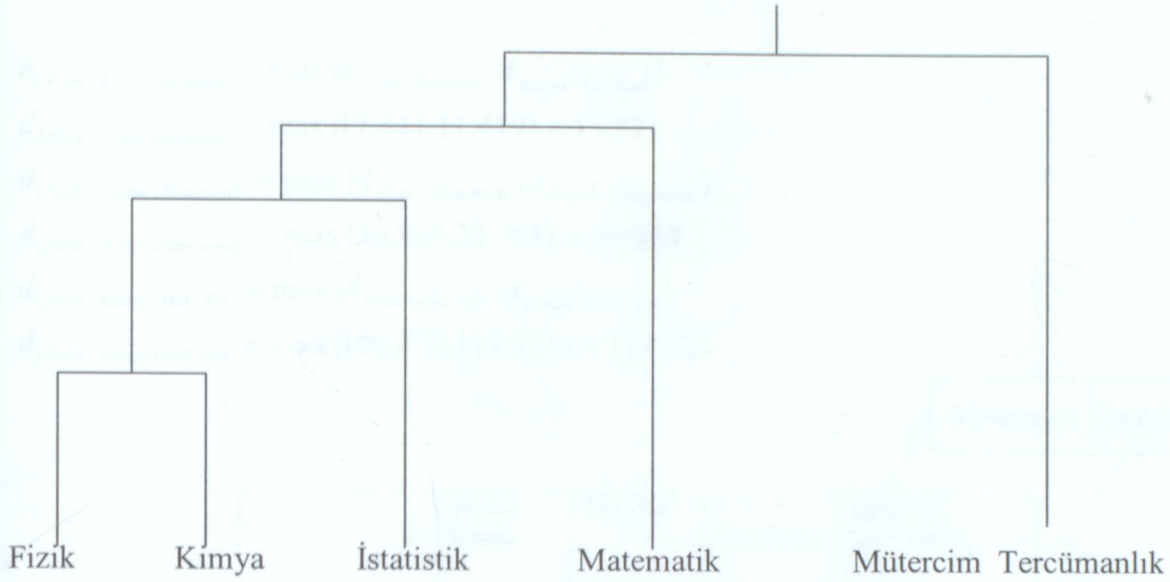
$$d_{(Fizik-Kimya-\text{İstatistik}-Matematik)Müt-Ter} = \min\{108.873, 138.142\} = 108.873$$

	Fizik –Kimya – İstatistik - Matematik	Mütercim Tercümanlık (Frn.)
Fizik –Kimya- İstatistik- Matematik	0	108,873
Mütercim Tercümanlık (Frn.)	108,873	0

En son adımda kalan iki küme de birleştirilerek işlem tamamlanır.

	Fizik –Kimya – İstatistik – Matematik-Mütercim Tercümanlık
Fizik –Kimya- İstatistik- Matematik- Mütercim Tercümanlık	0

Örneğe ilişkin dendogram aşağıda yer almaktadır.



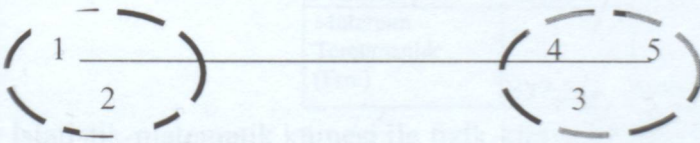
5.2.3.1.2 Complete Linkage

Complete linkage yöntemi single linkage yöntemiyle benzer biçimdedir. Ancak arada çok önemli bir fark vardır. Single linkage yönteminde başlangıçtaki $N \times N$ kare matris oluşturulduktan sonra birleştirilen iki kümenin diğer kümelerle uzaklıkları hesaplanmasında minimum uzaklıklar dikkate alınırken, complete linkage yönteminde maksimum uzaklıklar dikkate alınır. Ama iki küme birleştirilirken birbirlerine olan uzaklıkları minimum olarak hesaplanan kümeler yine bir araya getirilir.

Yeni oluşturulan kümenin diğer kümelerle olan uzaklığı ise U, V ve W kümeler olmak üzere aşağıdaki formül ile hesaplanır:

$$d_{(UV)W} = \max \{d_{UW}, d_{VW}\}$$

Complete Linkage yönteminde oluşturulan kümeler dendogram aracılığıyla gösterilebilir.



Örnek:

Minimum Uzaklık

	Fizik	İstatistik	Kimya	Matematik	Mütercim Tercümanlık (Frn.)
Fizik	0	17,821	9,540	30,284	108,873
İstatistik	17,821	0	11,419	14,895	124,421
Kimya	9,540	11,419	0	22,764	117,323
Matematik	30,284	14,895	22,764	0	138,142
Mütercim Tercümanlık (Frn.)	108,873	124,421	117,323	138,142	0

Fizik ve Kimya kümeleri birleştirilir.

$$d_{(Fizik-Kimya)İstatistik} = \max \{d_{Fizik-İstatistik}, d_{Kimya-İstatistik}\}$$

$$d_{(Fizik-Kimya)İstatistik} = \max \{17.821, 11.419\} = 17.821$$

$$d_{(Fizik-Kimya)Matematik} = \max \{d_{Fizik-Matematik}, d_{Kimya-Matematik}\}$$

$$d_{(Fizik-Kimya)Matematik} = \max \{30.824, 22.764\} = 30.824$$

$$d_{(Fizik-Kimya)Müt-Ter} = \max \{d_{Fizik-Müt-Ter}, d_{Kimya-Müt-Ter}\}$$

$$d_{(Fizik-Kimya)Müt-Ter} = \max \{108.873, 117.323\} = 117.323$$

Minimum Uzaklık

	Fizik - Kimya	İstatistik	Matematik	Mütercim Tercümanlık (Frn.)
Fizik -Kimya	0	17,821	30,824	117,323
İstatistik	17,821	0	14,895	124,421
Matematik	30,824	14,895	0	138,142
Mütercim Tercümanlık (Frn.)	117,323	124,421	138,142	0

Matematik ve İstatistik kümeleri birleştirilir.

$$d_{(İstatistik-Matematik)Fizik-Kimya} = \max \{d_{Fizik-Kimya,İstatistik}, d_{Fizik-Kimya,Matematik}\}$$

$$d_{(İstatistik-Matematik)Fizik-Kimya} = \max \{17.821, 30.824\} = 30.824$$

$$d_{(İstatistik-Matematik)MütTer} = \max \{d_{İstatistik,MütTer}, d_{Matematik,MütTer}\}$$

$$d_{(İstatistik-Matematik)MütTer} = \max \{124.421, 138.142\} = 138.142$$

Minimum Uzaklık

	Fizik - Kimya	İstatistik-Matematik	Mütercim Tercümanlık (Frn.)
Fizik -Kimya	0	30,824	117,323
İstatistik - Matematik	30,824	0	138,142
Mütercim Tercümanlık (Frn.)	117,323	138,142	0

İstatistik-matematik kümesi ile fizik-kimya kümesi birleştirilir.

$$d_{(İstatistik-Matematik-Fizik-Kimya)MütTer} = \max \{d_{Fizik-Kimya,MütTer}, d_{İstatistik-Matematik,MütTer}\}$$

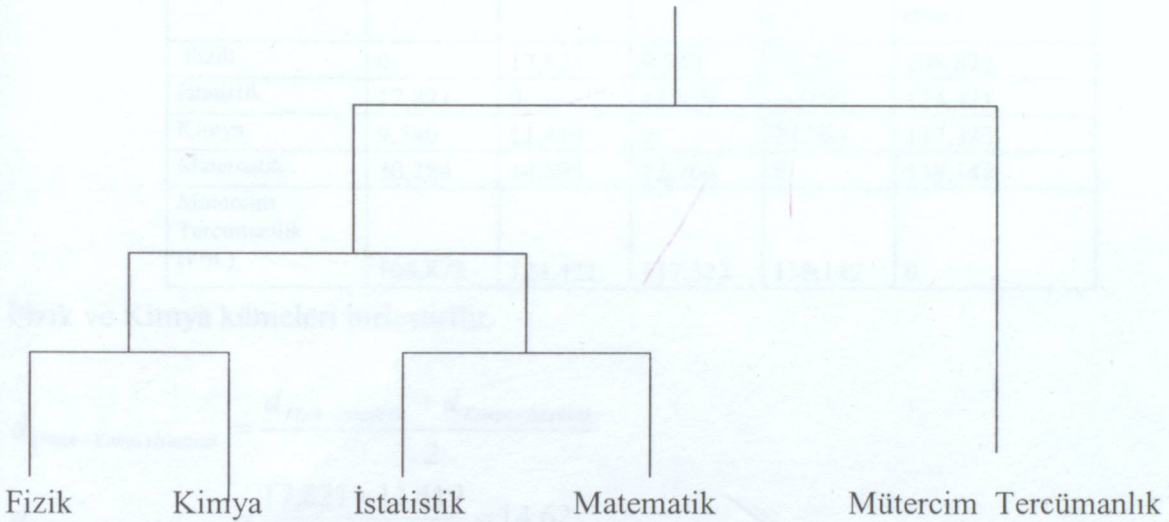
$$d_{(İstatistik-Matematik-Fizik-Kimya)MütTer} = \max \{117.323, 138.142\} = 138.142$$

	Fizik - Kimya-İstatistik-Matematik	Mütercim Tercümanlık (Frn.)
Fizik -Kimya-İstatistik-Matematik	0	138,142
Mütercim Tercümanlık (Frn.)	138,142	0

En son aşamada, kalan iki küme birleştirilerek kümeleme işlemi tamamlanır.

	Fizik –Kimya- İstatistik- Matematik- Mütercim Tercümanlık
Fizik –Kimya- İstatistik- Matematik- Mütercim Tercümanlık	0

Örneğe ilişkin dendrogram aşağıda yer almaktadır.



5.2.3.1.3 Average Linkage

Average Linkage yönteminde, uzaklıklardan ya da benzerliklerden oluşan $N \times N$ kare matriste minimum uzaklıkta olan ya da en çok benzeyen kümelerin birleştirilmesiyle oluşturulan yeni kümenin diğer birimlere olan uzaklıkları yeni oluşturulan kümenin diğer kümelerin birbirlerine olan uzaklıklarının ortalaması hesaplanılarak bulunur. Elde edilen yeni matriste ise, birbirine en az uzak olan ya da en çok benzer olan kümeler birleştirilir.

Yeni oluşturulan kümenin diğer kümelere olan uzaklığı ise U, V ve W kümeler olmak üzere aşağıdaki formül ile hesaplanır:

$$d_{(uv)w} = \frac{\sum_i \sum_k d_{ik}}{N_U N_V N_W}$$



Average Linkage yönteminde oluşturulan kümeler dendogram aracılığıyla gösterilebilir.

Örnek:

	Fizik	İstatistik	Kimya	Matematik	Mütercim Tercümanlık (Frn.)
Fizik	0	17,821	9,540	30,284	108,873
İstatistik	17,821	0	11,419	14,895	124,421
Kimya	9,540	11,419	0	22,764	117,323
Matematik	30,284	14,895	22,764	0	138,142
Mütercim Tercümanlık (Frn.)	108,873	124,421	117,323	138,142	0

Fizik ve Kimya kümeleri birleştirilir.

$$d_{(Fizik-Kimya)İstatistik} = \frac{d_{Fizik-İstatistik} + d_{Kimya-İstatistik}}{2}$$

$$d_{(Fizik-Kimya)İstatistik} = \frac{17,821 + 11,419}{2} = 14,62$$

$$d_{(Fizik-Kimya)Matematik} = \frac{d_{Fizik-Matematik} + d_{Kimya-Matematik}}{2}$$

$$d_{(Fizik-Kimya)Matematik} = \frac{30,284 + 22,764}{2} = 26,24$$

$$d_{(Fizik-Kimya)Müt-Ter} = \frac{d_{Fizik-MütTer} + d_{Kimya-MütTer}}{2}$$

$$d_{(Fizik-Kimya)Müt-Ter} = \frac{108,873 + 117,323}{2} = 113,098$$

Minimum Uzaklık

	Fizik - Kimya	İstatistik	Matematik	Mütercim Tercümanlık (Frn.)
Fizik -Kimya	0	14,62	26,524	113,098
İstatistik	14,62	0	14,895	124,421
Matematik	26,524	14,895	0	138,142
Mütercim Tercümanlık (Frn.)	113,098	124,421	138,142	0

Fizik-Kimya bölümlerinden oluşturulan küme ile istatistik kümesi birleştirilir.

$$d_{(Fizik-Kimya-\text{İstatistik})Matematik} = \frac{d_{Fizik,Matematik} + d_{Kimya,Matematik} + d_{İstatistik,Matematik}}{3}$$

$$d_{(Fizik-Kimya-\text{İstatistik})Matematik} = \frac{30,284 + 22,764 + 14,895}{3} = 22,647$$

$$d_{(Fizik-Kimya-\text{İstatistik})MütTer} = \frac{d_{Fizik,MütTer} + d_{Kimya,MütTer} + d_{İstatistik,MütTer}}{3}$$

$$d_{(Fizik-Kimya-\text{İstatistik})MütTer} = \frac{108,873 + 117,323 + 124,421}{3} = 116,872$$

Minimum Uzaklık

	Fizik – Kimya-İstatistik	Matematik	Mütercim Tercümanlık (Frn.)
Fizik – Kimya-İstatistik	0	22,647	116,872
Matematik	22,647	0	138,142
Mütercim Tercümanlık (Frn.)	116,872	138,142	0

Fizik-Kimya-İstatistik bölümlerinden oluşan küme ile Matematik kümesi birleştirilir.

$$d_{Fizik-Kimya-\text{İstatistik-Matematik},MütTer} = \frac{d_{Fizik-MütTer} + d_{Kimya-MütTer} + d_{İstatistik-MütTer} + d_{Matematik-MütTer}}{4}$$

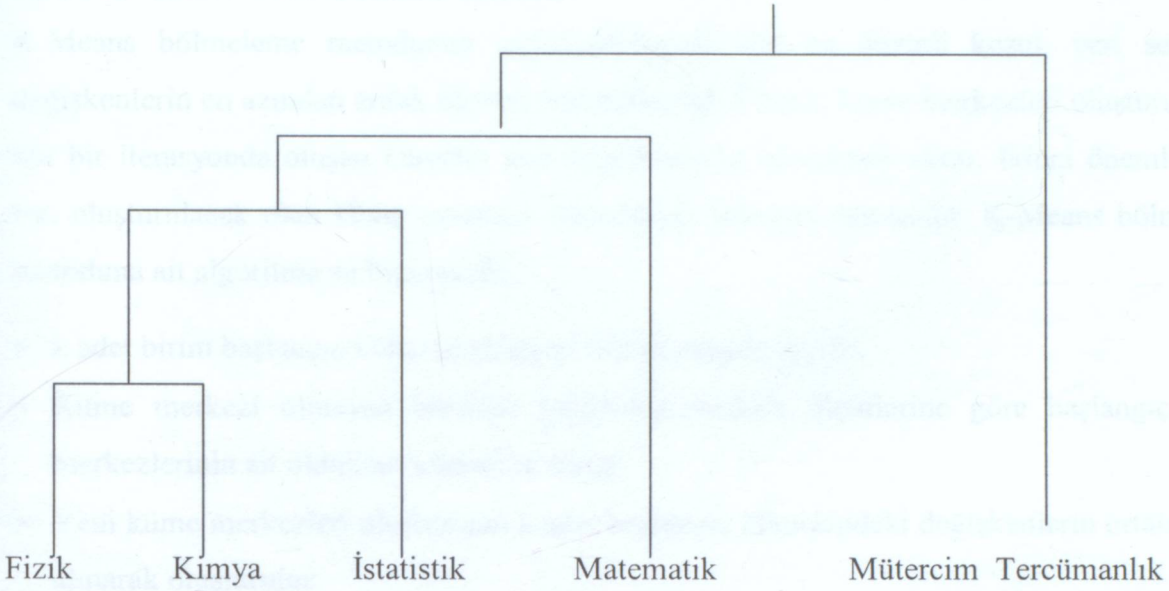
$$d_{Fizik-Kimya-\text{İstatistik-Matematik},MütTer} = \frac{108,873 + 117,323 + 124,421 + 138,142}{4} = 122,190$$

	Fizik – Kimya-İstatistik-Matematik	Mütercim Tercümanlık (Frn.)
Fizik – Kimya-İstatistik-Matematik	0	122,190
Mütercim Tercümanlık (Frn.)	122,190	0

En son adımda Fizik – Kimya-İstatistik-Matematik bölümlerinden oluşan küme ile Mütercim Tercümanlık kümesi birleştirilerek kümeleme işlemi tamamlanır.

	Fizik – Kimya-İstatistik-Matematik-Mütercim Tercümanlık
Fizik – Kimya-İstatistik-Matematik-Mütercim Tercümanlık	0

Örneğe ilişkin dendogram aşağıda yer almaktadır.



5.2.3.2 Bölmeleme Metotları

Bölmeleme metotları, hiyerarşik olmayan kümeleme metotlarındanıdır. Bu metotlar, n adet birimden oluşan veri setini başlangıçta belirlenen $k < n$ olmak üzere k adet kümeye ayırmak için kullanılırlar. Bölmeleme metotlarının hiyerarşik metotlardan en önemli farklarından birisi de budur. Bölmeleme metotları, hiyerarşik metotlara oranla daha büyük veri setlerine uygulanabilir. Bölmeleme metotlarında oluşturulacak k adet küme şu koşulları sağlar:

- Her bir küme en azından bir birim içerir.
- Her birim yalnızca bir gruba yerleştirilebilir.

Fakat fuzzy bölmeleme tekniklerinde 2. koşulun sağlanması beklenmez.

Bölmeleme metotlarında işlemler genellikle şu sıra ile yapılır: İlk olarak başlangıç kümeleri ya da başlangıç küme merkezleri gelişigüzel olarak belirlenir. Birimlerin, belirlenen kümelerin merkezlerine olan uzaklıklarına göre yeni küme merkezleri oluşturulur. Bu işlemler birbirlerinden farklı, kendi içlerinde homojen, birbirleri arasında benzerlik bulunmayan k adet küme oluşturuluncaya kadar sürdürülür. Görüldüğü gibi, bu metotlar iteratif süreçlerden oluşan metotlardır.

Bölmeleme metotları arasında en popüler olan metotlar K-Means ve K-Medoids bölmeleme metotlarıdır.

5.2.3.2.1 K-Means Bölmeleme Metodu

K-Means bölmeleme metodunun uygulanabilmesi için en önemli koşul, veri setindeki değişkenlerin en azından aralık ölçekte bulunmasıdır. Çünkü, küme merkezleri oluşturulurken her bir iterasyonda oluşan kümeler için değişkenlerin ortalaması alınır. İkinci önemli koşul ise, oluşturulacak olan küme sayısının başlangıçta biliniyor olmasıdır. K-Means bölmeleme metoduna ait algoritma şu biçimdedir:

- k adet birim başlangıç küme merkezleri olarak rasgele seçilir.
- Küme merkezi olmayan birimler belirlenen uzaklık ölçütlerine göre başlangıç küme merkezlerinin ait oldukları kümelere atanır.
- Yeni küme merkezleri oluşturulan k adet başlangıç kümesindeki değişkenlerin ortalamaları alınarak oluşturulur.
- Birimler en yakın oldukları oluşturulan yeni küme merkezlerine birimlerin uzaklıkları hesaplanarak kümeye atanır.
- Bir önceki küme merkezlerine olan uzaklıklar ile yeni oluşturulan küme merkezlerine olan uzaklıklar karşılaştırılır.
- Uzaklıklar makul görülebilir oranda azalmış ise, 4. adıma dönlür.
- Eğer çok büyük bir değişiklik söz konusu olmamış ise, iterasyon sona erdirilir.

İterasyonun durdurulması için kullanılan ölçütlerden birisi, kareli hata ölçütüdür. Bu ölçüt p veri uzayında bir nokta, m_i ise C_i kümesine ait ortalama ya da küme merkezi olmak üzere şu biçimdedir.

$$E = \sum_{i=1}^k \sum_{p \in C_i} |p - m_i|^2$$

K-Means metodunun başka bir versiyonu da K-Modes metodudur. Eğer veri setindeki kümeleme için kullanılacak olan değişkenler kategorik değişkenler ise, K-Modes metodu tercih edilir. Bu metotta, her bir iterasyonda belirlenen küme merkezleri, kümelerdeki değişkenlerin ortalaması yerine, modu hesaplanarak oluşturulur.

K-Means metodu kullanılan uzaklık ölçütlerinin farklılığına, iterasyonu durdurma kriterinin farklılığına, her bir iterasyonda küme merkezlerinin hesaplanmasında kullanılan yöntemin farklılığına göre değişik biçimlerde uygulanabilir.

K-Means bölmeleme metodunda algoritmanın durdurulması için kullanılabilecek olan diğer

yöntemlerden bazıları; varyans analizi, Hotelling T^2 istatistiğidir.

Örnek:

Çizelge 5.4'te yer alan veri setini K-Means algoritması yardımı ile 2 kümeye ayrılması şu biçimde gerçekleştirilir. Başlangıç öklidyen uzaklık matrisi aşağıdadır.

	Fizik	İstatistik	Kimya	Matematik	Mütercim Tercümanlık (Frn.)
Fizik	0	17,821	9,540	30,284	108,873
İstatistik	17,821	0	11,419	14,895	124,421
Kimya	9,540	11,419	0	22,764	117,323
Matematik	30,284	14,895	22,764	0	138,142
Mütercim Tercümanlık (Frn.)	108,873	124,421	117,323	138,142	0

Başlangıç Küme Merkezi olarak Fizik ve Mütercim Tercümanlık bölümleri seçilirse birimlerin küme merkezlerine olan uzaklıkları şöyle gerçekleşir.

$$\begin{aligned}
 d_{Fizik-Kimya} &= 9,540 & d_{MütTer-Kimya} &= 117,323 \\
 d_{Fizik-İstatistik} &= 17,821 & d_{MütTer-İstatistik} &= 124,421 \\
 d_{Fizik-Matematik} &= 30,284 & d_{MütTer-Matematik} &= 138,142
 \end{aligned}$$

Diğer bölümlerin, küme merkezleri Fizik ve Mütercim Tercümanlık olarak seçilen kümelere uzaklıkları karşılaştırılırsa Kimya, İstatistik, Matematik bölümleri, merkezi Fizik bölümü olan kümeye atanır.

Küme Numarası	Fizik	İstatistik	Kimya	Matematik	Mütercim Tercümanlık
1	x	x	x	x	
2					x

Mütercim Tercümanlık kümesi tek birimden oluştuğu için, küme içi uzaklık sıfır küme merkezi ise yine Mütercim Tercümanlık bölümüne ait olan değerler olacaktır. 2. küme merkezi ise Fizik, İstatistik, Kimya, Matematik birimlerinden oluşan kümenin değişkenler bazında ortalamaları hesaplanarak oluşturulur.

1.Küme Merkezi	193.261	202.4993	190.3085	198.628	189.0788	200.86
2.Küme Merkezi	127.672	164.909	140.037	157.87	134.582	155.75

Kümelemenin ilk adımı için kareli hata kriteri $E_1=57.645$

Birimler yeni küme merkezlerine olan uzaklıklarına göre yeniden kümelendir.

$$\begin{array}{ll}
 d_{1.küme-Kimya} = 6.902 & d_{2.küme-Kimya} = 117.323 \\
 d_{1.küme-İstatistik} = 5.495 & d_{2.küme-İstatistik} = 124.421 \\
 d_{1.küme-Matematik} = 16.644 & d_{2.küme-Matematik} = 138.142 \\
 d_{1.küme-Fizik} = 13.934 & d_{2.küme-Fizik} = 108.873 \\
 d_{1.küme-MütTer} = 122.099 & d_{2.küme-MütTer} = 0.000
 \end{array}$$

1. küme fizik, istatistik, kimya ve matematik bölümlerinden oluşurken 2. küme yine yalnızca Mütercim Tercümanlık bölümünden oluşur. Kümelemenin ikinci adımı için kareli hata ölçütü $E_2= 165.076$ olarak hesaplanır. $E_2 > E_1$ olduğundan iterasyon sonlandırılır.

Sonuçta ilk iterasyonda oluşan başlangıç küme merkezleri baz alınmak üzere kümeler oluşturulur.

5.2.3.2.2 K-Medoids Bölmeleme Metodu

K-Means bölmeleme metoduna benzer biçimde işlem yapan fakat her bir iterasyon sonucunda referans alınacak küme merkezlerini seçerken değişkenlerin ortalama değerini almayıp birimleri küme merkezi olarak alan bir bölmeleme metodudur. K-Means bölmeleme metodu küme merkezlerini oluştururken ortalama aldığı için aşırı değerlere duyarlı olmasına karşılık K-Medoids bölmeleme metodu, birimleri küme merkezi olarak seçtiğinden dolayı aşırı değerlere karşı duyarlı değildir.

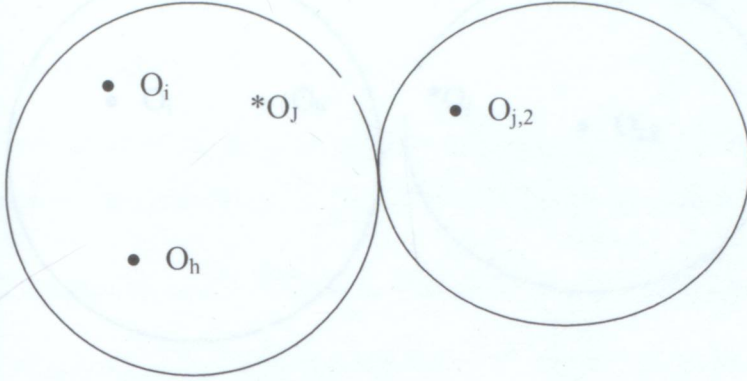
K-Medoids bölmeleme metodunun amacı her bir küme için temsili bir birim bulmaktır. Bu temsili birime medoid adı verilir. Medoid küme içerisine en merkezi birimdir. Medoidler aşırı değerlerin varlıklarından etkilenmedikleri gibi, medoidler incelenen birimlerin sırasına da bağlı değildir.

K-Medoids bölmeleme metoduna ait algoritma şu biçimdedir:

- k adet birim başlangıç medoidleri olarak rasgele seçilir.
- Her bir birim yakın olduğu medoide atanır.
- Medoid olmayan birimler sırayla seçilir.
- Sıra ile seçilen bir birim ile medoidin yer değiştirmesinin toplam maliyeti hesaplanır.
- Toplam maliyetler sıralanır.
- En küçük toplam maliyet sıfırdan küçükse, 3. adıma dönlür.
- Toplam maliyet sıfırdan büyük veya sıfıra eşit çıkarsa, iterasyon sona erer.

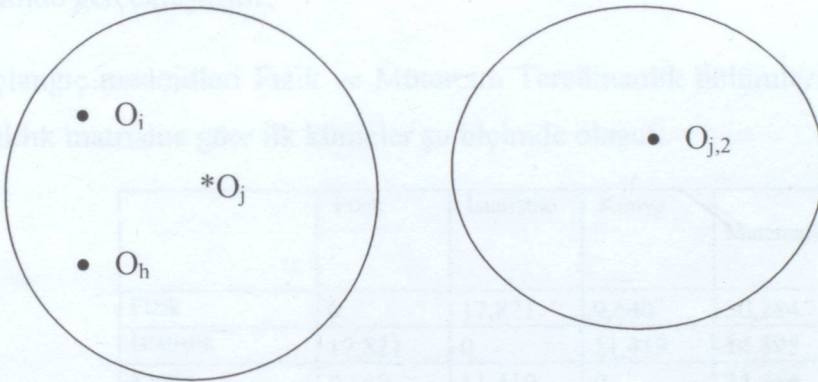
Medoidlerin değişmesinin maliyetinin hesaplanmasında 4 durum söz konusudur.

1.durum: O_i ve $O_{j,2}$ var olan medoidler, O_h rasgele seçilen medoid olsun.



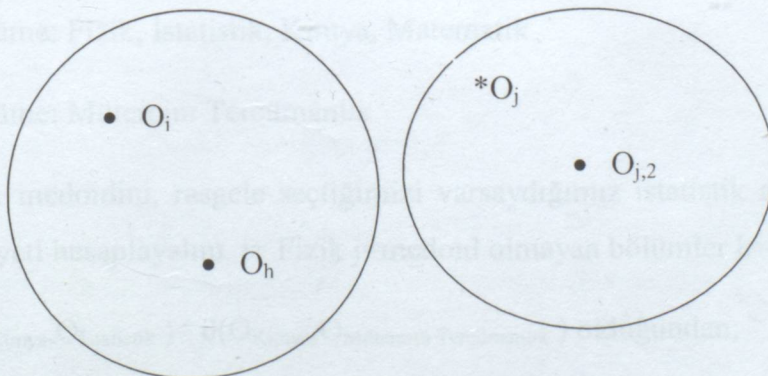
O_j birimi $O_{j,2}$ medoidine atanır. Bu maliyet ; $C_{jih} = d(O_j, O_{j,2}) - d(O_j, O_i)$ formülü ile hesaplanır.

2.durum:



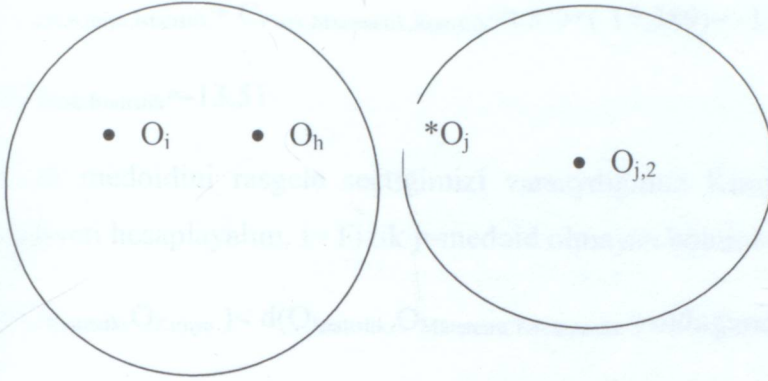
O_j yine aynı küme de kalır ve O_h medoidine atanır. Bu maliyet; $C_{jih} = d(O_j, O_h) - d(O_j, O_i)$ formülü ile hesaplanır.

3.durum:



O_j yine eski medoidinde kalır. $C_{jih} = 0$ olur.

4.durum:



O_j O_h medoidine atanır. Bu maliyet; $C_{jih} = d(O_j, O_h) - d(O_j, O_{j,2})$

Örnek:

Çizelge 5.4'te yer alan veri setini K-Medoids algoritması yardımı ile 2 kümeye ayrılması şu biçimde gerçekleştirilir.

Başlangıç medoidleri Fizik ve Mütercim Tercümanlık bölümleri seçilsin. Aşağıdaki Öklidyen uzaklık matrisine göre ilk kümeler şu biçimde oluşur.

	Fizik	İstatistik	Kimya	Matematik	Mütercim Tercümanlık (Frn.)
Fizik	0	17,821	9,540	30,284	108,873
İstatistik	17,821	0	11,419	14,895	124,421
Kimya	9,540	11,419	0	22,764	117,323
Matematik	30,284	14,895	22,764	0	138,142
Mütercim Tercümanlık (Frn.)	108,873	124,421	117,323	138,142	0

1. Küme: Fizik, İstatistik, Kimya, Matematik

2. Küme: Mütercim Tercümanlık

Fizik medoidini, rasgele seçtiğimizi varsaydığımız istatistik medoidi ile değiştirdiğimizdeki maliyeti hesaplayalım. $i = \text{Fizik}$ $j = \text{medoid olmayan bölümler}$ $h = \text{İstatistik}$ olsun.

$d(O_{\text{Kimya}}, O_{\text{İstatistik}}) < d(O_{\text{Kimya}}, O_{\text{Mütercim Tercümanlık}})$ olduğundan;

$C_{\text{Kimya, Fizik, İstatistik}} = d(\text{Kimya, İstatistik}) - d(\text{Kimya, Fizik}) = 11,419 - 9,540 = 1.879$

$d(O_{\text{Matematik}}, O_{\text{İstatistik}}) < d(O_{\text{Matematik}}, O_{\text{Mütercim Tercümanlık}})$ olduğundan;

$$C_{\text{Matematik, Fizik, İstatistik}} = d(\text{Matematik, İstatistik}) - d(\text{Matematik-Fizik}) = 14,895 - 30,284 = -15,389$$

$$C_{\text{Fizik, Kimya, İstatistik}} + C_{\text{Fizik, Matematik, İstatistik}} = 1,879 + (-15,389) = -13,51$$

$$TC_{\text{Fizik, İstatistik}} = -13,51$$

Fizik medoidini rasgele seçtiğimizi varsaydığımız Kimya medoidi ile değiştirdiğimizdeki maliyeti hesaplayalım. $i = \text{Fizik}$ $j = \text{medoid olmayan bölümler}$ $h = \text{Kimya}$ olsun.

$d(O_{\text{İstatistik}}, O_{\text{Kimya}}) < d(O_{\text{İstatistik}}, O_{\text{Mütercim Tercümanlık}})$ olduğundan;

$$C_{\text{İstatistik, Fizik, Kimya}} = d(\text{İstatistik, Kimya}) - d(\text{İstatistik, Fizik}) = 11,419 - 17,821 = -6,402$$

$d(O_{\text{Matematik}}, O_{\text{Kimya}}) < d(O_{\text{Matematik}}, O_{\text{Mütercim Tercümanlık}})$ olduğundan;

$$C_{\text{Matematik, Fizik, Kimya}} = d(\text{Matematik, Kimya}) - d(\text{Matematik-Fizik}) = 22,764 - 30,284 = -7,52$$

$$C_{\text{İstatistik, Fizik, Kimya}} + C_{\text{Matematik, Fizik, Kimya}} = -6,402 + (-7,52) = -13,902$$

$$TC_{\text{Fizik, Kimya}} = -13,902$$

Fizik medoidini rasgele seçtiğimizi varsaydığımız Matematik medoidi ile değiştirdiğimizdeki maliyeti hesaplayalım. $i = \text{Fizik}$ $j = \text{medoid olmayan bölümler}$ $h = \text{Kimya}$ olsun.

$d(O_{\text{İstatistik}}, O_{\text{Matematik}}) < d(O_{\text{İstatistik}}, O_{\text{Mütercim Tercümanlık}})$ olduğundan;

$$C_{\text{İstatistik, Fizik, Matematik}} = d(\text{İstatistik, Matematik}) - d(\text{İstatistik, Fizik}) = 14,895 - 17,821 = -2,926$$

$d(O_{\text{Kimya}}, O_{\text{Matematik}}) < d(O_{\text{Kimya}}, O_{\text{Mütercim Tercümanlık}})$ olduğundan;

$$C_{\text{Kimya, Fizik, Matematik}} = d(\text{Kimya, Matematik}) - d(\text{Kimya, Fizik}) = 22,764 - 9,540 = 13,224$$

$$C_{\text{İstatistik, Fizik, Matematik}} + C_{\text{Kimya, Fizik, Matematik}} = -2,926 + 13,224 = 10,298$$

$$TC_{\text{Fizik, Kimya}} = -10,298$$

Toplam maliyetler bulunduktan sonra;

$TC_{\text{Fizik, Kimya}} = -13,902 < TC_{\text{Fizik, İstatistik}} = -13,51 < TC_{\text{Fizik, Kimya}} = -10,298$ karşılaştırması sonucu negatif olan en küçük maliyeti veren medoid değişimi yapılır. Bu işlem pozitif toplam maliyet bulununcaya kadar tekrarlanır.

5.2.3.3 Yoğunluk Bazlı Metotlar, Grid Bazlı Metotlar, Model Bazlı Metotlar

Hiyerarşik kümeleme metotları ve bölmeleme metotları dışında kullanılan diğer kümeleme metotları arasında yoğunluk bazlı metotlar, grid bazlı metotlar ve model bazlı metotlar yer alır.

Rasgele biçimdeki kümelerin keşfedilmesi için yoğunluk bazlı metotlar kullanılmaktadır. Yoğunluk bazlı metotlar birimlerin veri uzayındaki dağılımlarını inceleyerek kümeleri birimlerin oluşturdukları bölgelerin yoğunluklarına göre oluşturmaktadırlar. Bu metotlara örnek olarak: DBSCAN (Density-Based Spatial Clustering of Applications with Noise), OPTICS (Ordering Points to Identify), DENCLUE (Density Based Clustering) metotları verilebilir.

Grid bazlı kümeleme metotları ise veri uzayını gridleri kullanarak sonlu sayıda hücrelere bölen metotlardır. Bu metotları kullanmanın kazancı, hızlı işlem süresidir. İşlem süresi birimlerin sayısından bağımsızdır. Fakat kümelenecek olan veri uzayında oluşturulacak her bir boyuttaki hücrelerin sayısına bağımlıdır. Bu metotlara örnek olarak: STING (Statistical Information Grid), Wave Cluster, CLIQUE (Clustering High Dimensional Space) metotları verilebilir.

Model bazlı kümeleme metotları, istatistiksel metotlar ve neural network metotlarından yararlanırlar. Bu metotlar verilerin kümeleneceği için matematiksel modellemeler oluştururlar. Bu metotlara örnek olarak, Sınıflama ağaçları ve Self Organising Maps metotları verilebilir.

6. UYGULAMA

Bu uygulama çalışmasında İstanbul ilinde bulunan ve 2001 yılı mezunları ÖSS sınavına katılan liseler kümeleme analizi ile gruplanacaklardır. Kümeleme analizi sonucunda oluşacak olan grupların ise profilleri incelenecektir.

Uygulamada kullanılacak olan veri seti ÖSYM Yayınları tarafından yayımlanan 2001-6 numaralı, Ortaöğretim Kurumlarına Göre 2001 Öğrenci Seçme Sınavı Sonuçları adlı kitaptan alınmıştır. Veri girişi işlemi, İstatistik Bölümü son sınıf öğrencilerinden Gürkan Mocan, ve Şükrü Türker Yalınkılıç tarafından yapılmıştır.

Uygulamada kullanılacak olan veri seti, İstanbul ilinde ÖSS sınavına öğrencileri aday olan tüm okullar olduğu için aynı zamanda ana küttedir. Toplam okul sayısı 640, toplam değişken sayısı ise 41'dir. Değişkenler içinden d5, d7, d9, d11, d13, d15 değişkenleri sırasıyla d4, d6, d8, d10, d12, d14 değişkenlerinin d3 değişkenine oranlanması ile elde edilmiştir. d16 değişkeni ise d3 değişkeninden d10, d12, d14 değişkenlerinin toplamı çıkarılıp, elde edilen sonuç d3 değişkenine oranlanarak oluşturulmuştur. D40 değişkeni eğitimin karma-kız-erkek olarak yapılmasına göre oluşturulmuştur. D41 değişkeni okulun devlet okulu ya da özel okul olma durumuna göre oluşturulmuştur. Uygulamada kullanılacak olan değişkenlerin tanımı ve kodları 6.1 bölümünde yer almaktadır.

6.1 Uygulamada Kullanılacak Olan Değişkenler

Çizelge 6.1 Uygulamada Kullanılacak Olan Değişkenler ve Kodları

Kod	Değişken
d1	Okul Kodu
d2	İl
d3	ÖSS Sınavına Giren Öğrenci Sayısı
d4	105 Puandan Az Puan Alan Öğrenci Sayısı
d5	105 Puandan Az Puan Alan Öğrenci Yüzdesi
d6	105-120 Puan Arasında Puan Alan Öğrenci Sayısı
d7	105-120 Puan Arasında Puan Alan Öğrenci Yüzdesi
d8	120 Puandan Fazla Puan Alan Öğrenci Sayısı

Çizelge 6.1 Devam Uygulamada Kullanılacak Olan Değişkenler ve Kodları

Kod	Değişken
d9	120 Puandan Fazla Puan Alan Öğrenci Yüzdesi
d10	Ön Lisans Programların Yerleştirilen Öğrenci Sayısı
d11	Ön Lisans Programların Yerleştirilen Öğrenci Yüzdesi
d12	Lisans Programlarına Yerleştirilen Öğrenci Sayısı
d13	Lisans Programlarına Yerleştirilen Öğrenci Yüzdesi
d14	Açık Öğretim Programlarına Yerleştirilen Öğrenci Sayısı
d15	Açık Öğretim Programlarına Yerleştirilen Öğrenci Yüzdesi
d16	Herhangi bir Programa Yerleştirilemeyen Öğrenci Yüzdesi
d17	Matematik Netleri Ortalaması
d18	Matematik Netleri Standart Sapması
d19	Fizik Netleri Ortalaması
d20	Fizik Netleri Standart Sapması
d21	Kimya Netleri Ortalaması
d22	Kimya Netleri Standart Sapması
d23	Biyoloji Netleri Ortalaması
d24	Biyoloji Netleri Standart Sapması
d25	Türkçe Netleri Ortalaması
d26	Türkçe Netleri Standart Sapması
d27	Tarih Netleri Ortalaması
d28	Tarih Netleri Standart Sapması
d29	Coğrafya Netleri Ortalaması
d30	Coğrafya Netleri Standart Sapması
d31	Felsefe Netleri Ortalaması
d32	Felsefe Netleri Standart Sapması

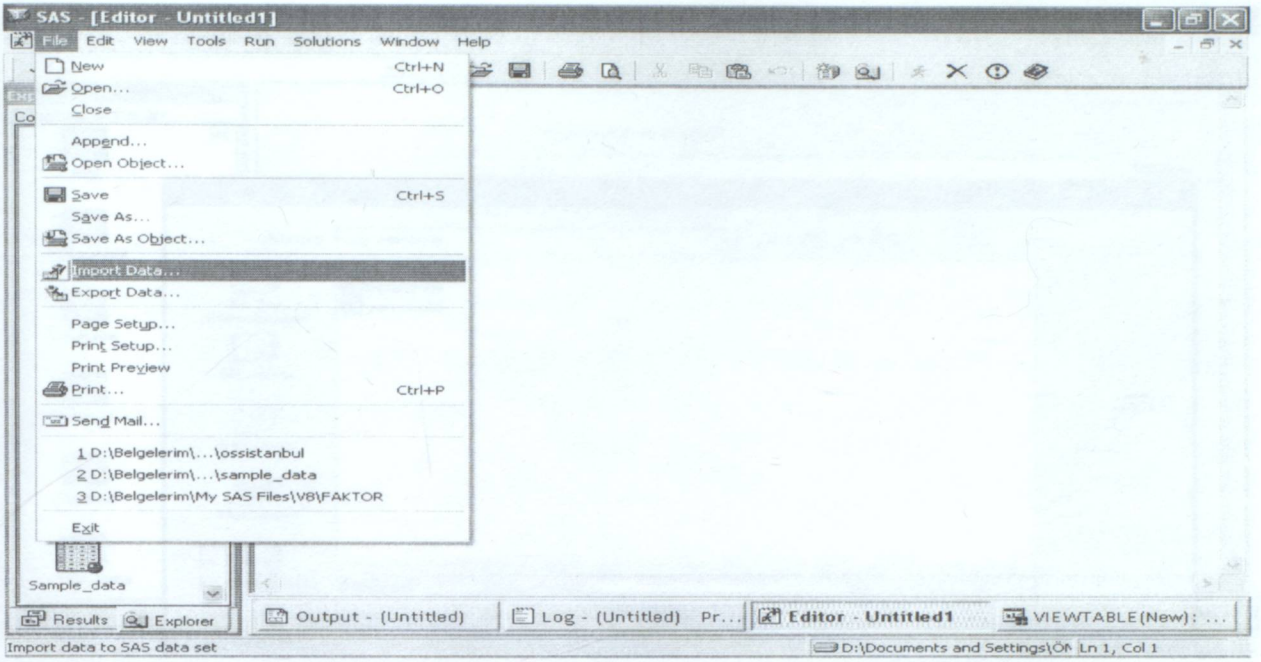
Çizelge 6.1 Devam Uygulamada Kullanılacak Olan Değişkenler ve Kodları

Kod	Değişken
d33	İngilizce Netleri Ortalaması
d34	İngilizce Netleri Standart Sapması
d35	Fransızca Netleri Ortalaması
d36	Fransızca Netleri Standart Sapması
d37	Almanca Netleri Ortalaması
d38	Almanca Netleri Standart Sapması
d39	Okul Türü
d40	Eğitim Türü (Karma-Erkek-Kız)
d41	Okul Türü (Devlet-Özel)
d42	Okul Türü (Anadolu-Fen-Meslek...)

6.2 Uygulama Aşaması

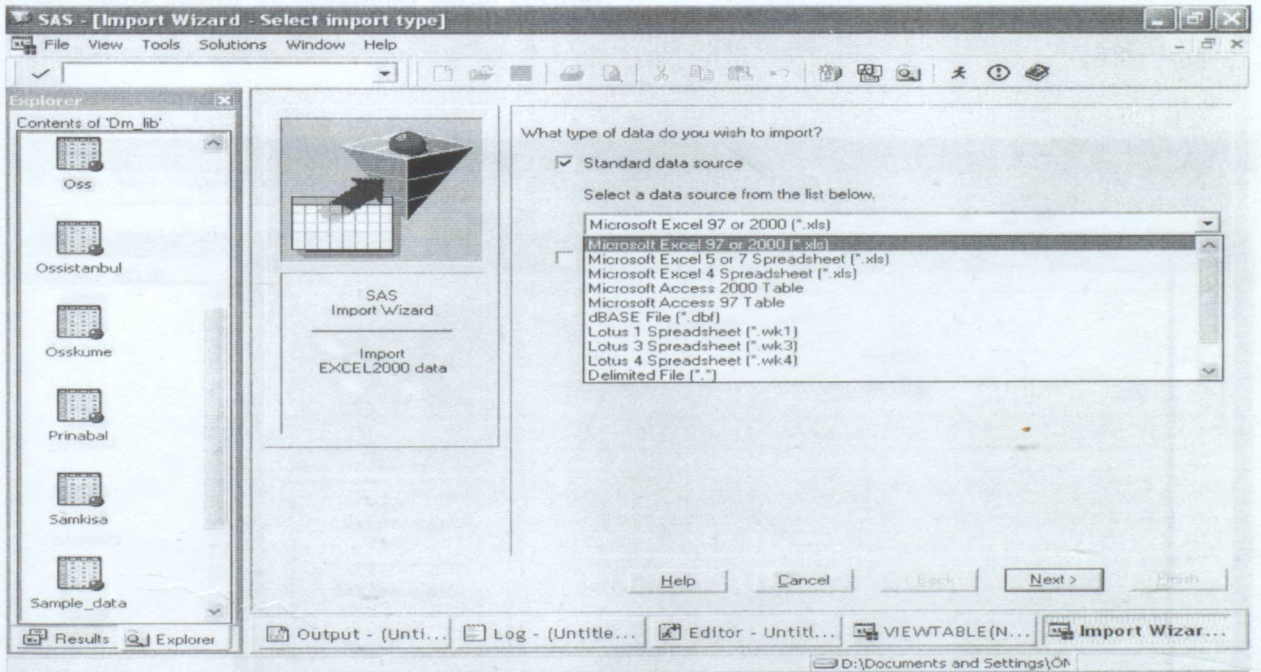
6.2.1 Verilerin SAS Programında Analiz Edilecek Formata Getirilmesi

Excel formatında hazırlanan verilerin SAS programının kullandığı dosya formatına dönüştürülmesi işlemleri aşağıda yer almaktadır. SAS program editoründen file sekmesinde yer alan Import Dataya tıklanır.



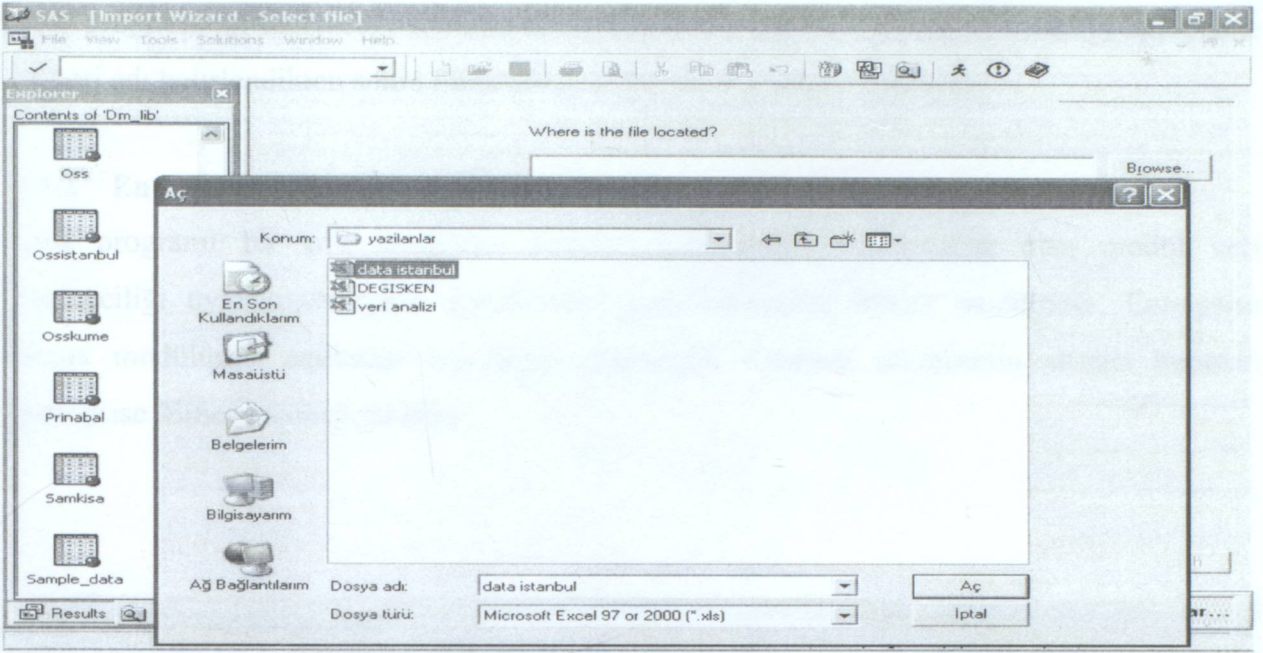
Şekil 6.1 Farklı Veri Kaynağında Bulunan Veri Setinin SAS Programına Aktarılması İşlemi-1

Veri kaynağının seçileceği pencere ekrana gelir.



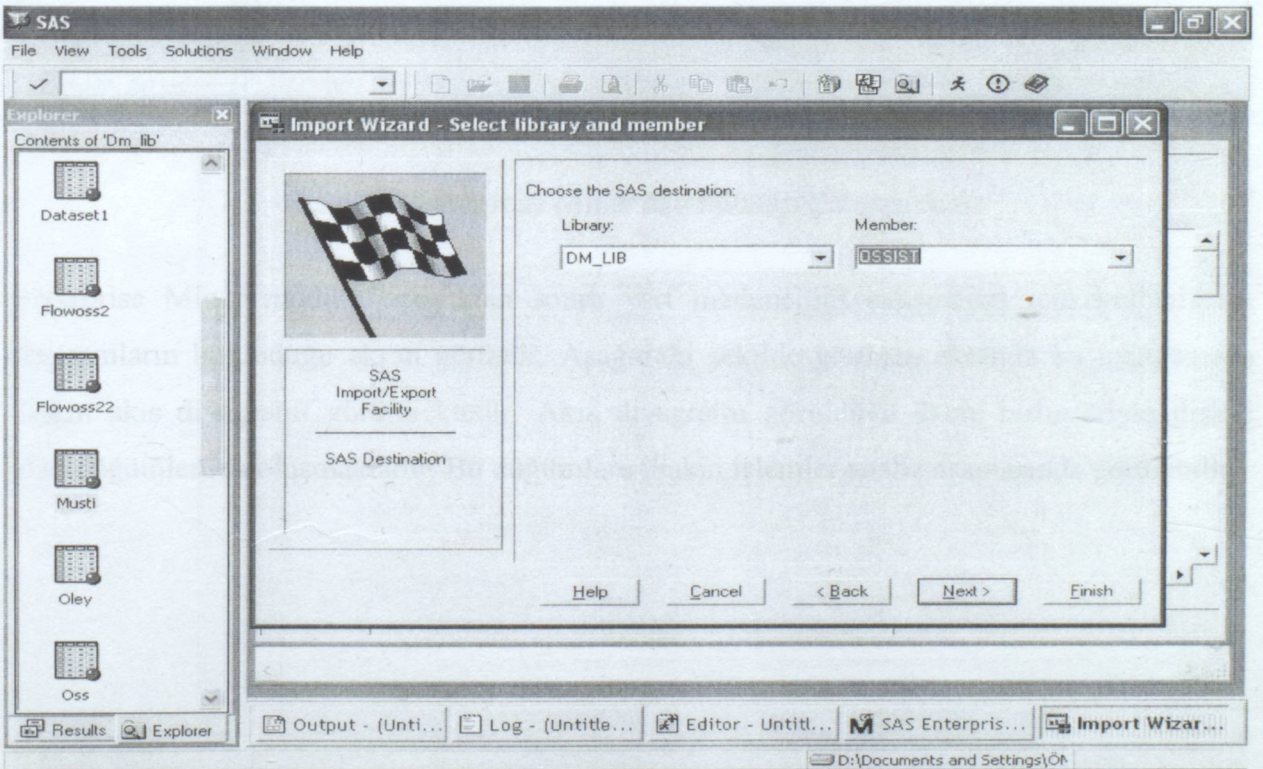
Şekil 6.2 Farklı Veri Kaynağında Bulunan Veri Setinin SAS Programına Aktarılması İşlemi-2

Veri kaynağı seçildikten sonra Next düğmesine tıklanır.



Şekil 6.3 Farklı Veri Kaynağında Bulunan Veri Setinin SAS Programına Aktarılması İşlemi-3

Verinin bulunduğu dosyanın adresi tanımlanır. Adres tanımlandıktan sonra tekrar Next düğmesine basılır ve aşağıdaki ekran görülür.

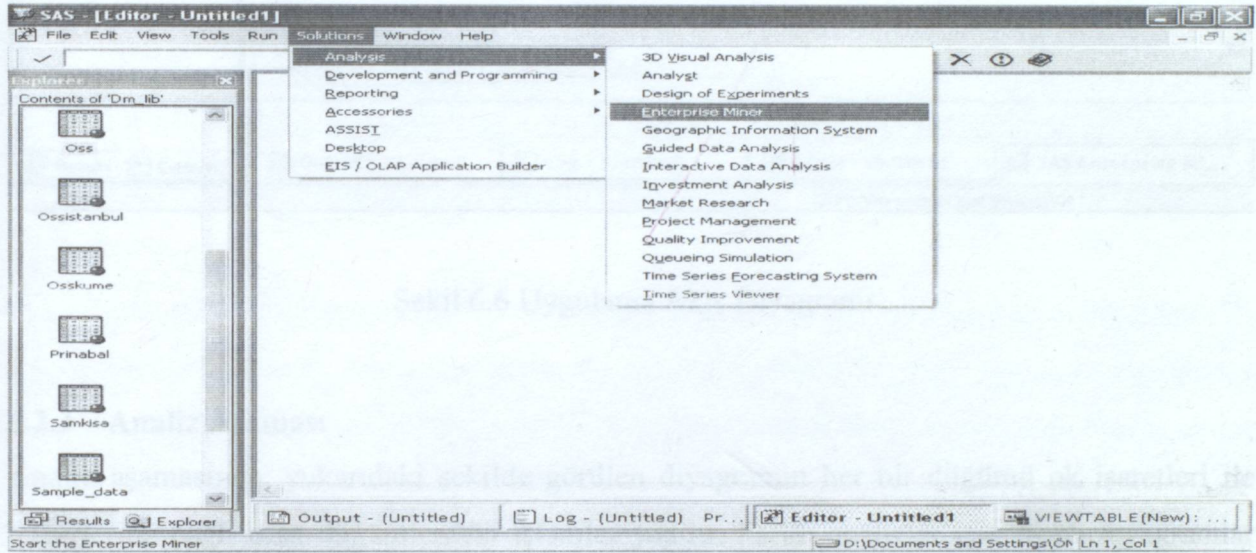


Şekil 6.4 Farklı Veri Kaynağında Bulunan Veri Setinin SAS Programına Aktarılması İşlemi-4

SAS programının tanıyacağı formata dönüştürülecek olan veri setinin bulunacağı kütüphane ve veri adı belirlendikten sonra Finish düğmesine basılır ve işlem bitirilir.

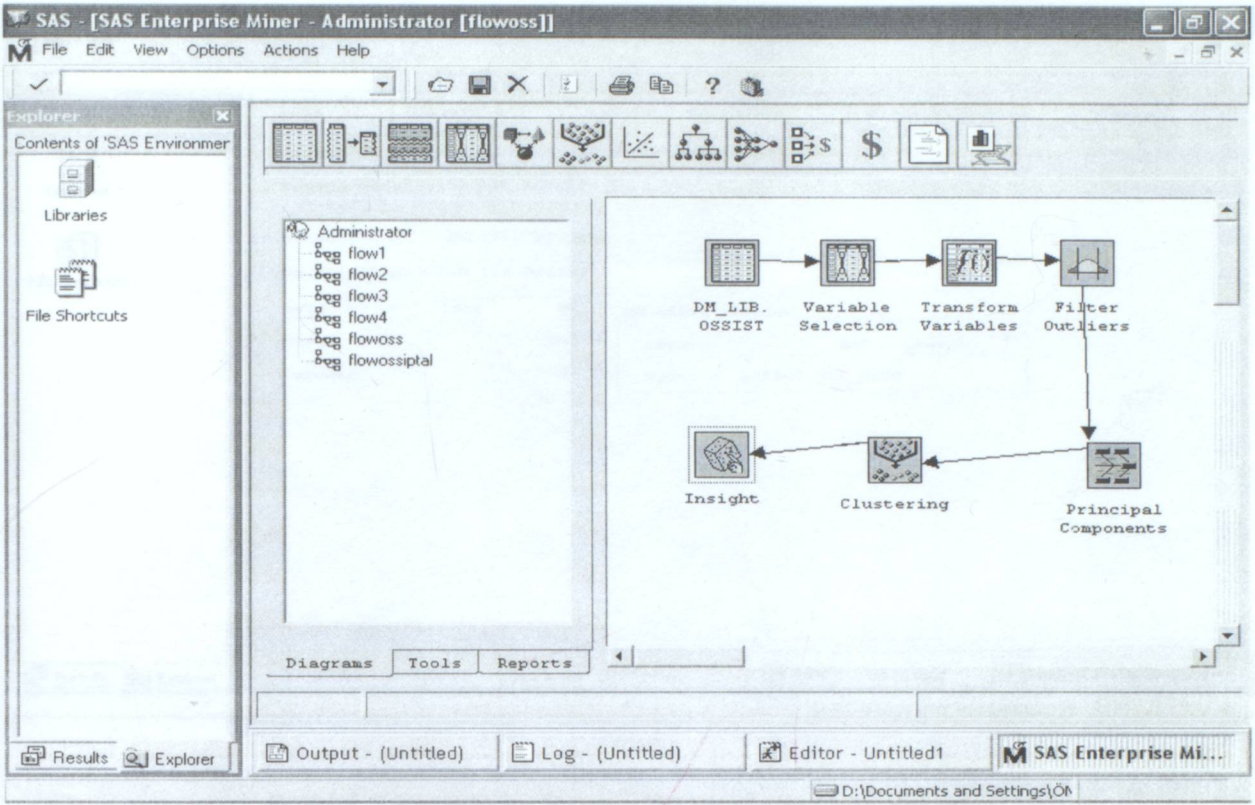
6.2.2 Enterprise Miner Modülünün Açılması

SAS programı bir çok modülden oluşur. Uygulamada kullanılacak olan modül veri madenciliği uygulamaları için geliştirilmiş olan Enterprise Miner modülüdür. Enterprise Miner modülünün açılması için SAS editöründe Analysis sekmesinin altında bulunan Enterprise Miner seçeneği tıklanır.



Şekil 6.5 Enterprise Miner Modülünün Çalıştırılması

Enterprise Miner modülü açıldıktan sonra veri madenciliği çalışmaları için kullanılacak diyagramların bulunduğu ekran görünür. Aşağıdaki şekilde görünen ekranda bu uygulamaya ilişkin akış diyagramı görülmektedir. Akış diyagramı görüldüğü üzere birbirleriyle ilişkili olan düğümlerden oluşmaktadır. Bu düğümlere ilişkin işlemler analiz aşamasında görülebilir.



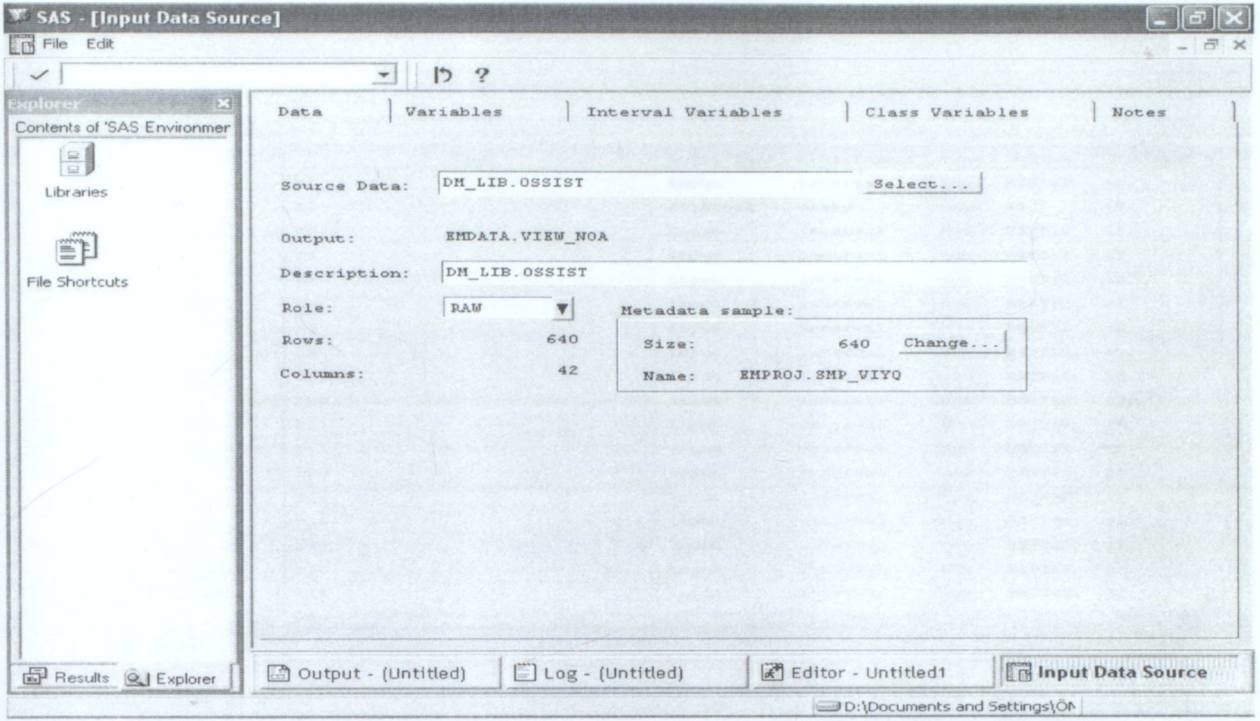
Şekil 6.6 Uygulama Akış Diyagramı

6.2.3 Analiz Aşaması

Analiz aşamasında, yukarıdaki şekilde görülen diyagramın her bir düğümü ok işaretleri ile birbirlerine bağlı olan düğümlerdeki işlemler yapılır. Farklı analizler için farklı diyagramlar oluşturulabilir. Bu uygulamada yukarıdaki şekilde yer alan adımlar tercih edilmiştir.

6.2.3.1 Veri Setinin Seçilmesi

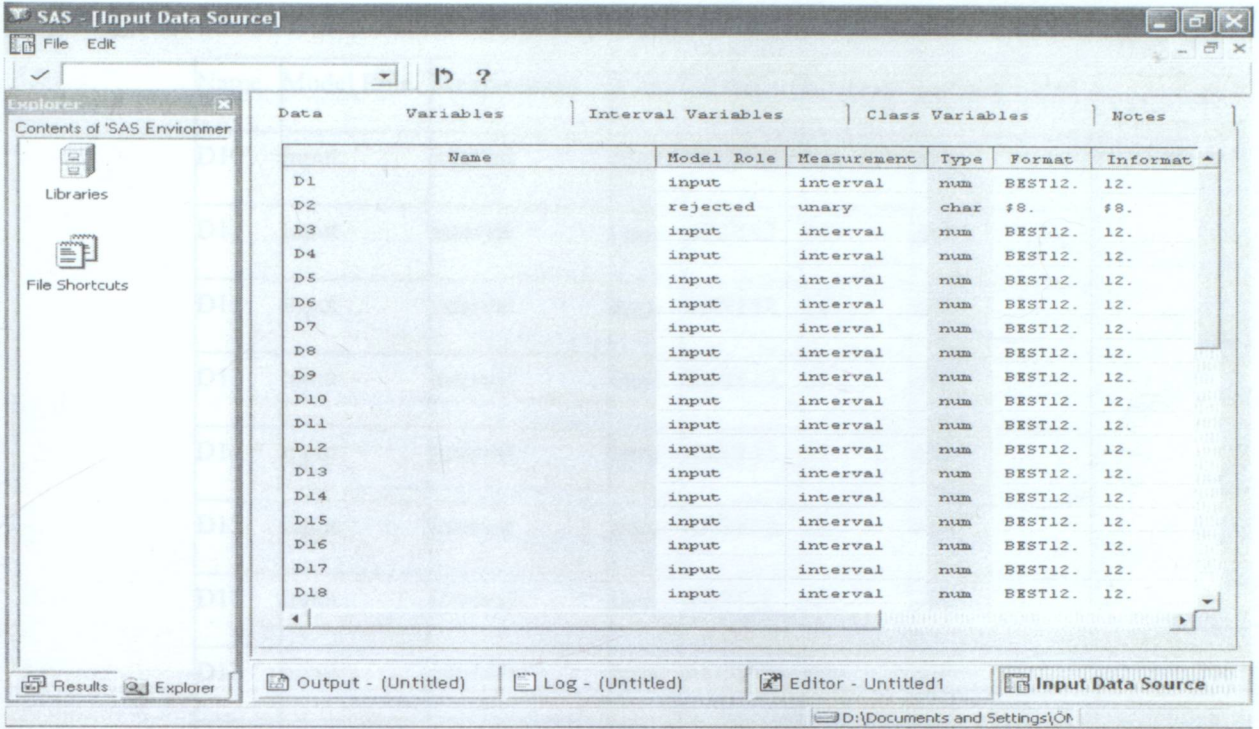
Analizde kullanılacak olan veri seti Input Data Source adı verilen düğümde seçilir. Veri Madenciliği çalışmalarının ilk başlangıcı bu adımla gerçekleşmektedir. Diyagramda bulunan Input Data Source düğümünün üzerinde mouse ile sağ klik yapıp open seçilirse şu ekran görünür.



Şekil 6.7 Input Data Source Düğümünün Data Sekmesi

Input Data Source düğümü açılınca dört sekmeden oluşan bir pencere karşımıza çıkar. İlk sekme verini seçileceği data sekmesidir. Bu sekmede verinin kaynağı ve tanımı yapıldıktan sonra metadata örnek büyüklüğü seçilir. Bu uygulamada tüm birimler metadata örneği olarak seçilmiştir.

Variables sekmesinde seçilen veri setinde bulunan değişkenler ve özellikleri görülür.



Şekil 6.8 Input Data Source Düğümünün Data Sekmesi

Variables sekmesinde veri setinde bulunan değişkenlerin kurulacak modeldeki rolleri, ölçekleri, türleri, formatları ve etiketleri yer alır. Aşağıdaki çizelgede uygulamada kullanılacak olan değişkenlerin özellikleri vardır.

Çizelge 6.2 Veri Setinde Bulunan Değişkenler ve Değişkenlere İlişkin Bazı Bilgiler

Name	Model Role	Measurement	Type	Format	Informat	Variable Label
D1	input	interval	num	BEST12.	12	d1
D2	rejected	unary	char	\$8	\$8	d2
D3	input	interval	num	BEST12.	12	d3
D4	input	interval	num	BEST12.	12	d4
D5	input	interval	num	BEST12.	12	d5
D6	input	interval	num	BEST12.	12	d6
D7	input	interval	num	BEST12.	12	d7
D8	input	interval	num	BEST12.	12	d8
D9	input	interval	num	BEST12.	12	d9

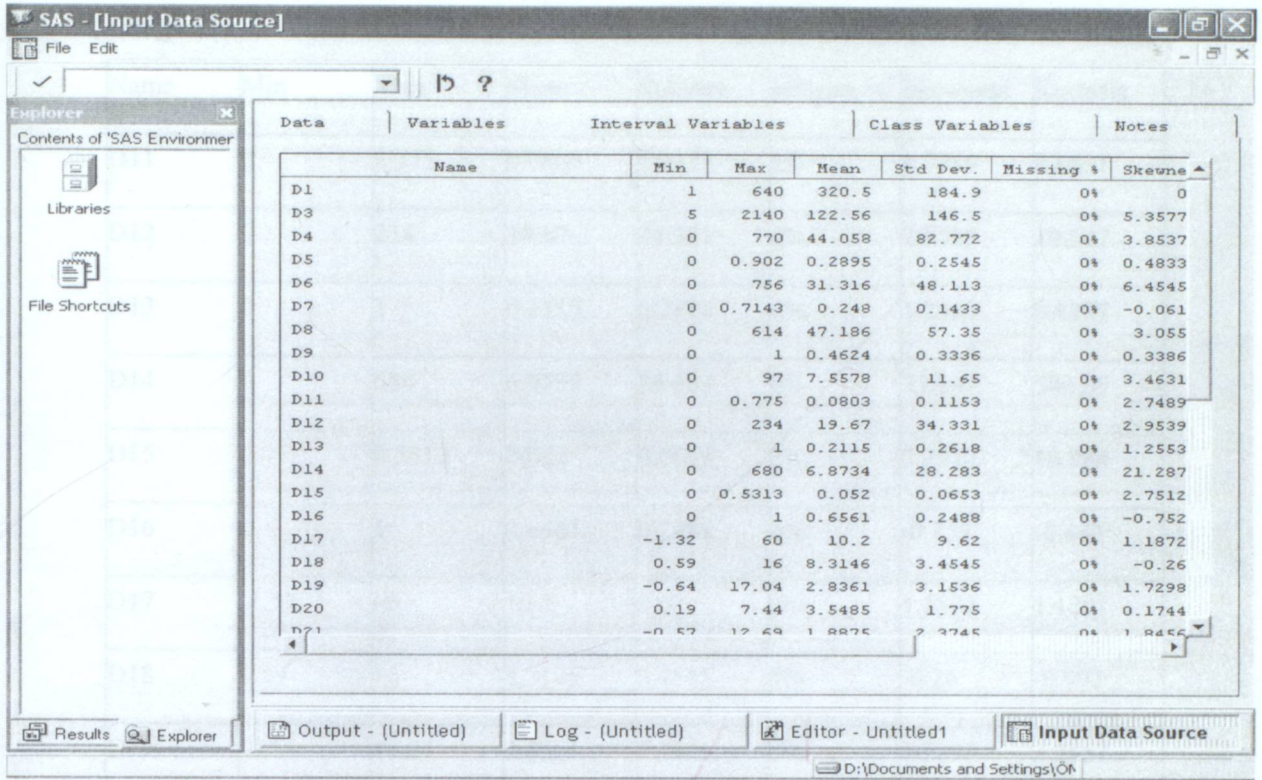
Çizelge 6.2 Devam Veri Setinde Bulunan Değişkenler ve Değişkenlere İlişkin Bazı Bilgiler

Name	Model Role	Measurement	Type	Format	Informat	Variable Label
D10	input	interval	num	BEST12.	12	d10
D11	input	interval	num	BEST12.	12	d11
D12	input	interval	num	BEST12.	12	d12
D13	input	interval	num	BEST12.	12	d13
D14	input	interval	num	BEST12.	12	d14
D15	input	interval	num	BEST12.	12	d15
D16	input	interval	num	BEST12.	12	d16
D17	input	interval	num	BEST12.	12	d17
D18	input	interval	num	BEST12.	12	d18
D19	input	interval	num	BEST12.	12	d19
D20	input	interval	num	BEST12.	12	d20
D21	input	interval	num	BEST12.	12	d21
D22	input	interval	num	BEST12.	12	d22
D23	input	interval	num	BEST12.	12	d23
D24	input	interval	num	BEST12.	12	d24
D25	input	interval	num	BEST12.	12	d25
D26	input	interval	num	BEST12.	12	d26
D27	input	interval	num	BEST12.	12	d27
D28	input	interval	num	BEST12.	12	d28
D29	input	interval	num	BEST12.	12	d29
D30	input	interval	num	BEST12.	12	d30
D31	input	interval	num	BEST12.	12	d31
D32	input	interval	num	BEST12.	12	d32
D33	rejected	interval	num	BEST12.	12	d33

Çizelge 6.2 Devam Veri Setinde Bulunan Değişkenler ve Değişkenlere İlişkin Bazı Bilgiler

Name	Model Role	Measurement	Type	Format	Informat	Variable Label
D34	rejected	interval	num	BEST12.	12	d34
D35	rejected	interval	num	BEST12.	12	d35
D36	rejected	interval	num	BEST12.	12	d36
D37	rejected	interval	num	BEST12.	12	d37
D38	rejected	interval	num	BEST12.	12	d38
D39	rejected	nominal	char	\$46	\$46	d39
D40	rejected	nominal	char	\$5	\$5	d40
D41	rejected	binary	char	\$6	\$6	d41
D42	rejected	nominal	char	\$11	\$11	d42

Interval Variables sekmesinde aralık ve oransal ölçekli değişkenlere ilişkin istatistikler yer alır.



Şekil 6.9 Input Data Source Düğümünün Interval Variables Sekmesi

Veri setinde bulunan aralık ve oransal ölçekli değişkenlere ait istatistikler aşağıdaki çizelgede görülmektedir.

Çizelge 6.3 Aralık ve Oransal Ölçekli Değişkenlere İlişkin Bazı İstatistikler

Name	Min	Max	Mean	Std Dev.	Missing %	Skewness	Kurtosis
D1	1	640	320.5	184.9	0%	0	-1.2
D3	5	2140	122.56	146.5	0%	5.3577	58.29
D4	0	770	44.058	82.772	0%	3.8537	19.808
D5	0	0.902	0.2895	0.2545	0%	0.4833	-0.931
D6	0	756	31.316	48.113	0%	6.4545	81.758
D7	0	0.7143	0.248	0.1433	0%	-0.061	-0.603
D8	0	614	47.186	57.35	0%	3.055	17.74
D9	0	1	0.4624	0.3336	0%	0.3386	-1.337
D10	0	97	7.5578	11.65	0%	3.4631	16.266

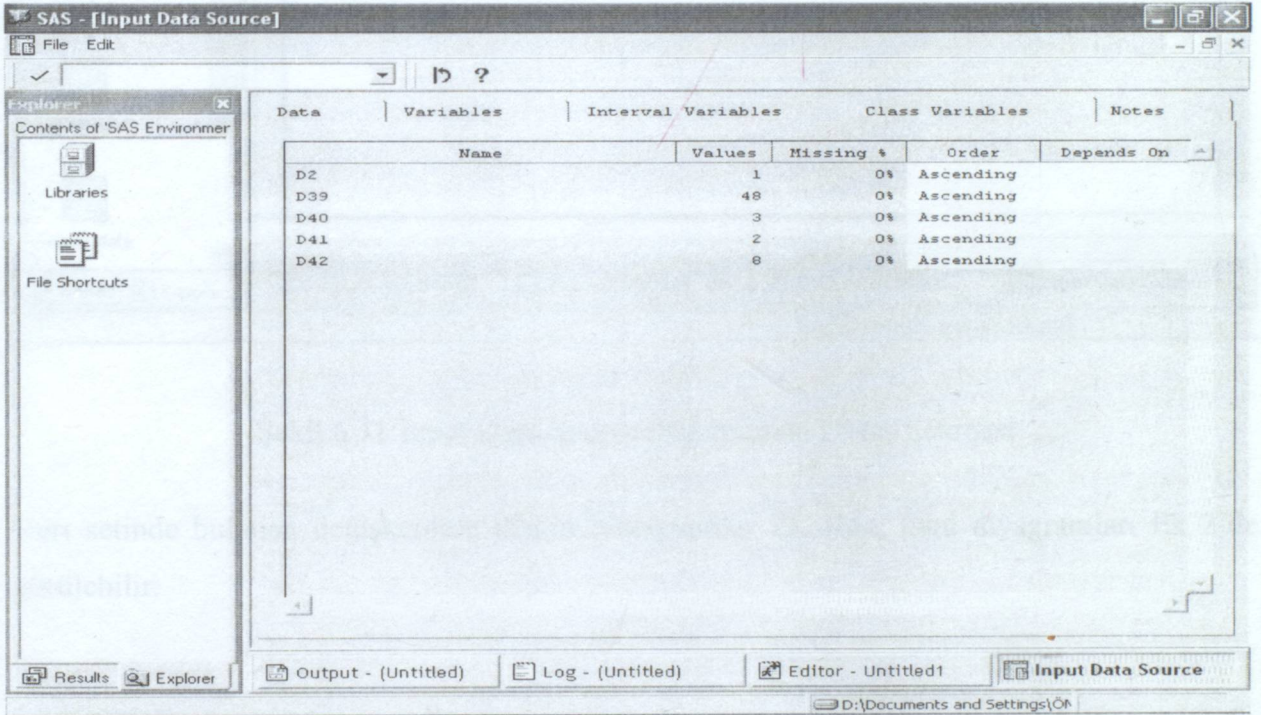
Çizelge 6.3 Devam Aralık ve Oransal Ölçekli Değişkenlere İlişkin Bazı İstatistikler

Name	Min	Max	Mean	Std Dev.	Missing %	Skewness	Kurtosis
D11	0	0.775	0.0803	0.1153	0%	2.7489	8.6631
D12	0	234	19.67	34.331	0%	2.9539	10.547
D13	0	1	0.2115	0.2618	0%	1.2553	0.4297
D14	0	680	6.8734	28.283	0%	21.287	504.09
D15	0	0.5313	0.052	0.0653	0%	2.7512	10.984
D16	0	1	0.6561	0.2488	0%	-0.752	-0.483
D17	-1.32	60	10.2	9.62	0%	1.1872	1.1201
D18	0.59	16	8.3146	3.4545	0%	-0.26	-0.902
D19	-0.64	17.04	2.8361	3.1536	0%	1.7298	3.2857
D20	0.19	7.44	3.5485	1.775	0%	0.1744	-1.123
D21	-0.57	12.69	1.8875	2.3745	0%	1.8456	3.5717
D22	0.08	5.72	2.5503	1.3813	0%	0.3427	-1.067
D23	-0.44	7.51	0.7743	1.0676	0%	2.6789	9.1132
D24	0.1	3.95	1.4894	0.7097	0%	0.914	0.2659
D25	0.4	41.46	23.128	7.9615	0%	0.1852	-0.934
D26	2.22	13.77	9.0683	1.5437	0%	-0.532	1.6338
D27	-0.48	15.27	7.1855	2.8026	0%	0.3916	-0.498
D28	1.65	7.26	4.7309	0.7578	0%	-0.246	0.7054
D29	0.71	12.14	4.4076	1.8254	0%	1.053	1.3096
D30	1.04	5.92	3.3096	0.7194	0%	0.2738	0.1256
D31	0.03	57	3.2551	2.629	0%	13.444	273.19
D32	0	4.59	2.9162	0.5976	0%	-0.615	1.2313
D33	0	92.52	10.365	24.549	0%	2.1545	3.0695
D34	0	33.68	3.2608	7.8606	0%	2.2846	3.8798

Çizelge 6.3 Devam Aralık ve Oransal Ölçekli Değişkenlere İlişkin Bazı İstatistikler

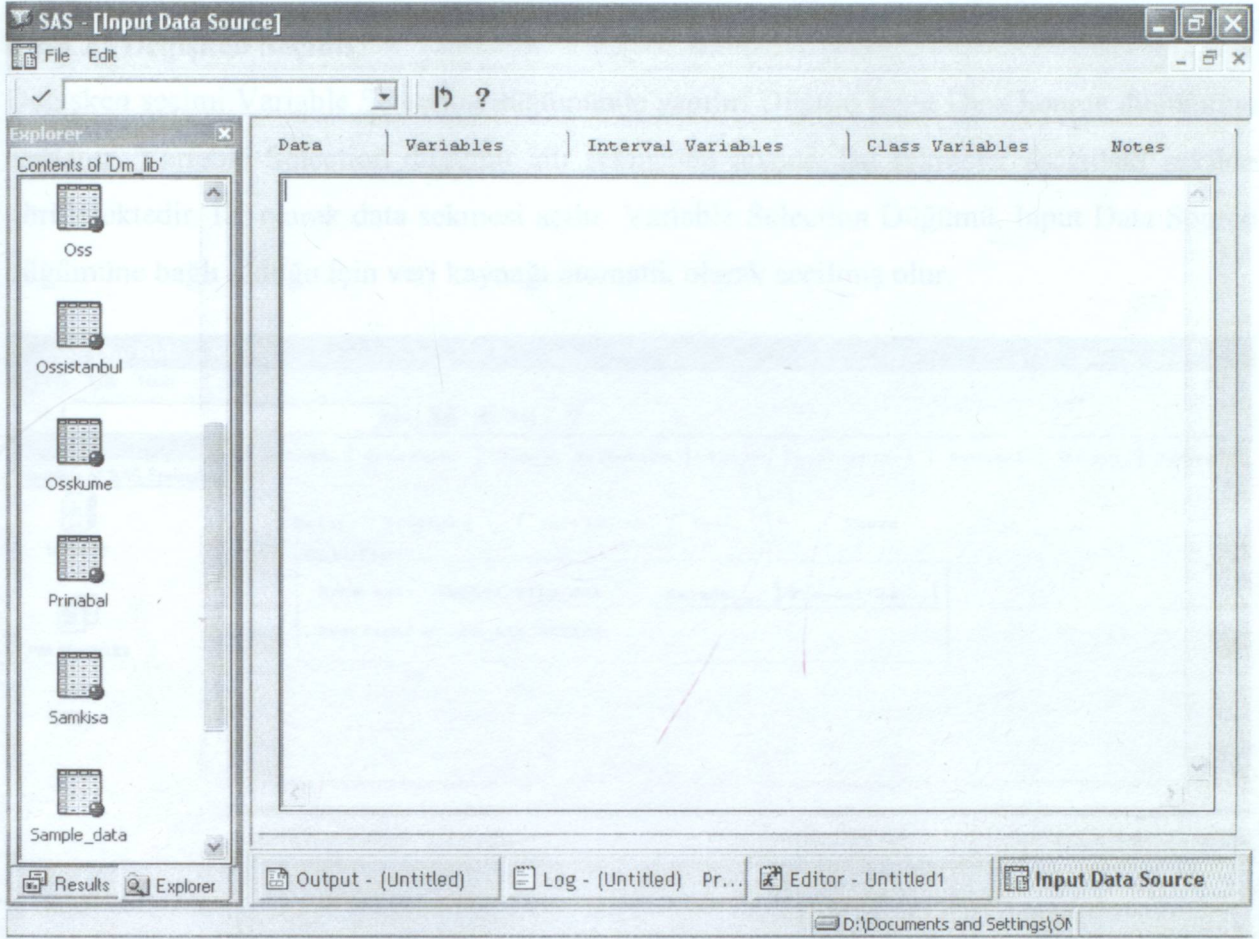
Name	Min	Max	Mean	Std Dev.	Missing %	Skewness	Kurtosis
D35	0	92.42	0.7616	7.8842	0%	10.393	107.55
D36	0	25.04	0.1107	1.3111	0%	14.479	237.01
D37	0	86.07	0.6385	6.6567	0%	10.604	113.38
D38	0	39.73	0.2029	2.2894	0%	12.839	181.54

Class Variables sekmesinde nominal ölçekli değişkenlere ilişkin kayıp veri, sınıf sayısı, sıralama ve bağımlı olduğu değişkenler görülmektedir.



Şekil 6.10 Input Data Source Düğümünün Class Variables Sekmesi

Notes sekmesinde kullanıcının veri setiyle ilgili not tutması için bir alan bulunur.



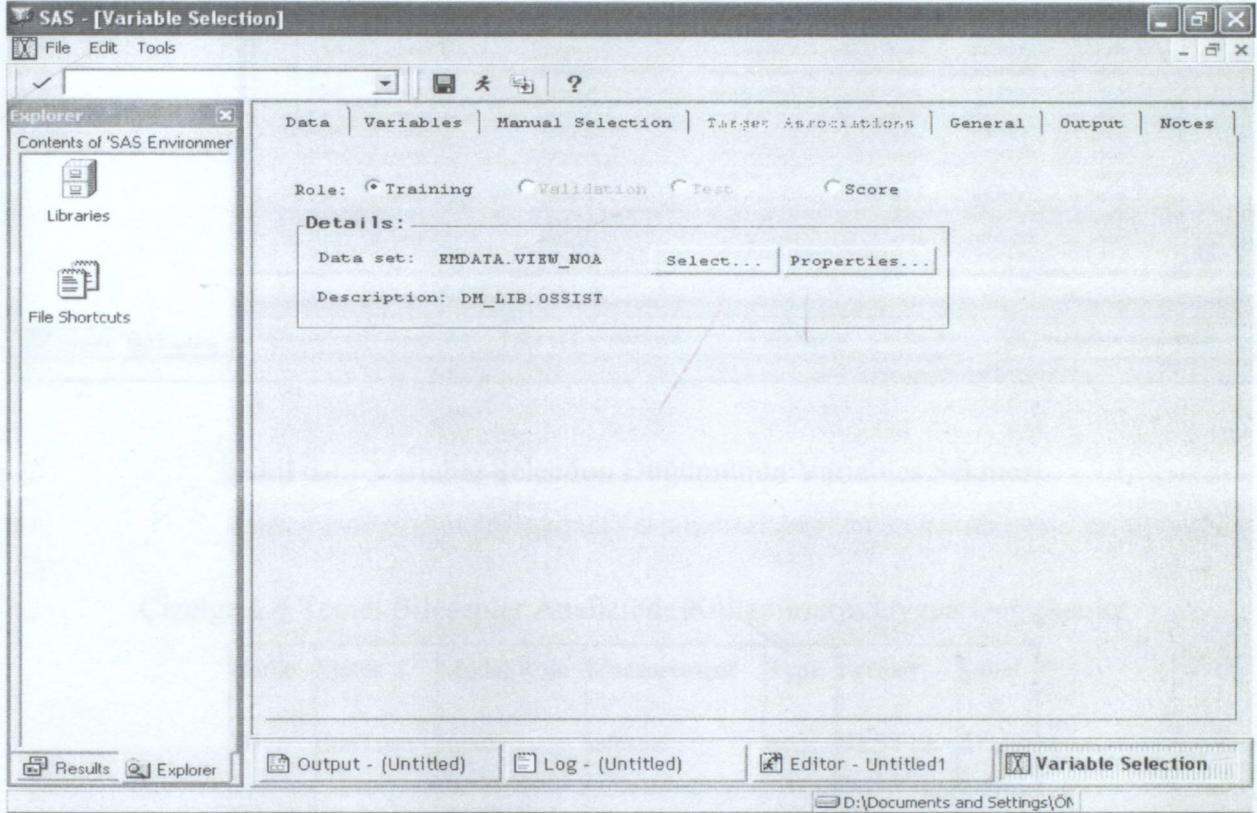
Şekil 6.11 Input Data Source Düğümünün Notes Sekmesi

Veri setinde bulunan değişkenlere ilişkin histogramlar Ek 1’da, kutu diyagramları Ek 2’de görülebilir.

Şekil 6.12 Variable Selection Düğümünde Data Sekmesi

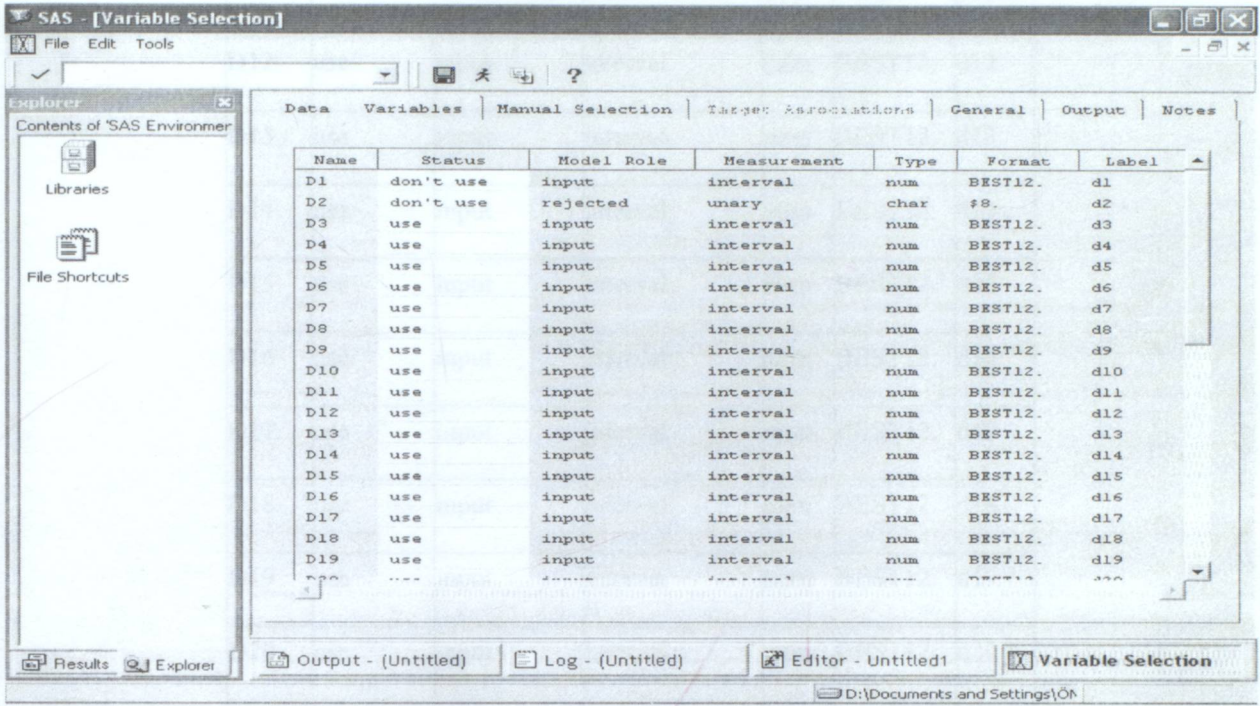
6.2.3.2 Değişken Seçimi

Değişken seçimi Variable Selection düğümünde yapılır. Düğüm Input Data Source düğümüne bağlanır. Variable Selection düğümü altı sekmeden oluşur. Bu sekmeler aşağıdaki şekilde görülmektedir. İlk olarak data sekmesi açılır. Variable Selection Düğümü, Input Data Source düğümüne bağlı olduğu için veri kaynağı otomatik olarak seçilmiş olur.



Şekil 6.12 Variable Selection Düğümünün Data Sekmesi

Variables sekmesinde değişkenlerin özellikleri görülür. Analizde kullanılmayacak olan değişkenler belirlenir. Bu uygulama için kullanılacak ve kullanılmayacak olan değişkenlerin listesi aşağıda yer almaktadır. Temel bileşenler analizini gerçekleştirebilmek için en azından aralık ölçekli değişkenler gerektiğinden nominal değişkenler dışlanmıştır. Yabancı dil sınavına giren öğrencilere ilişkin istatistiklerde temel bileşenler analizi dışında bırakılmıştır.



Şekil 6.13 Variable Selection Düğümünün Variables Sekmesi

Çizelge 6.4 Temel Bileşenler Analizinde Kullanılmaya Uygun Değişkenler

Name	Status	Model Role	Measurement	Type	Format	Label
D1	don't use	input	interval	num	BEST12.	d1
D2	don't use	rejected	unary	char	\$8	d2
D3	use	input	interval	num	BEST12.	d3
D4	use	input	interval	num	BEST12.	d4
D5	use	input	interval	num	BEST12.	d5
D6	use	input	interval	num	BEST12.	d6
D7	use	input	interval	num	BEST12.	d7
D8	use	input	interval	num	BEST12.	d8
D9	use	input	interval	num	BEST12.	d9
D10	use	input	interval	num	BEST12.	d10
D11	use	input	interval	num	BEST12.	d11

Çizelge 6.4 Devam Temel Bileşenler Analizinde Kullanılmaya Uygun Değişkenler

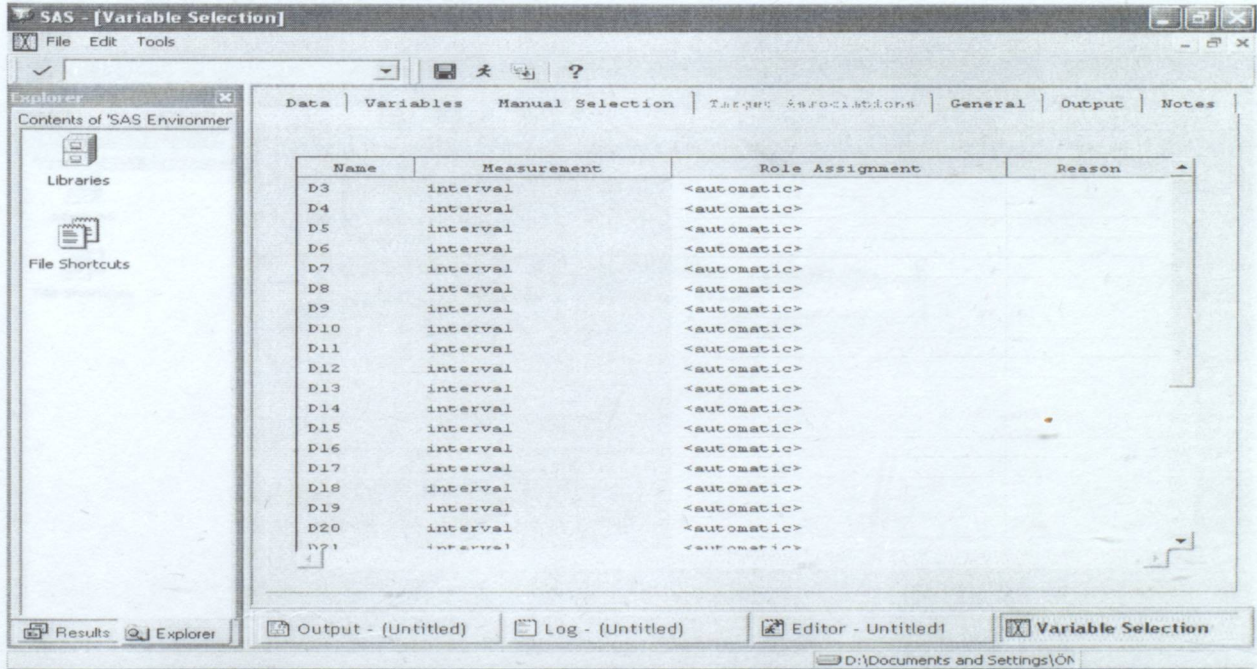
Name	Status	Model Role	Measurement	Type	Format	Label
D12	use	input	interval	num	BEST12.	d12
D13	use	input	interval	num	BEST12.	d13
D14	use	input	interval	num	BEST12.	d14
D15	use	input	interval	num	BEST12.	d15
D16	use	input	interval	num	BEST12.	d16
D17	use	input	interval	num	BEST12.	d17
D18	use	input	interval	num	BEST12.	d18
D19	use	input	interval	num	BEST12.	d19
D20	use	input	interval	num	BEST12.	d20
D21	use	input	interval	num	BEST12.	d21
D22	use	input	interval	num	BEST12.	d22
D23	use	input	interval	num	BEST12.	d23
D24	use	input	interval	num	BEST12.	d24
D25	use	input	interval	num	BEST12.	d25
D26	use	input	interval	num	BEST12.	d26
D27	use	input	interval	num	BEST12.	d27
D28	use	input	interval	num	BEST12.	d28
D29	use	input	interval	num	BEST12.	d29
D30	use	input	interval	num	BEST12.	d30
D31	use	input	interval	num	BEST12.	d31
D32	use	input	interval	num	BEST12.	d32
D33	don't use	rejected	interval	num	BEST12.	d33
D34	don't use	rejected	interval	num	BEST12.	d34
D35	don't use	rejected	interval	num	BEST12.	d35

Çizelge 6.4 Devam Temel Bileşenler Analizinde Kullanılmaya Uygun Değişkenler

Name	Status	Model Role	Measurement	Type	Format	Label
D36	don't use	rejected	interval	num	BEST12.	d36
D37	don't use	rejected	interval	num	BEST12.	d37
D38	don't use	rejected	interval	num	BEST12.	d38
D39	don't use	rejected	nominal	char	\$46	d39
D40	don't use	rejected	nominal	char	\$5	d40
D41	don't use	rejected	binary	char	\$6	d41
D42	don't use	rejected	nominal	char	\$11	d42

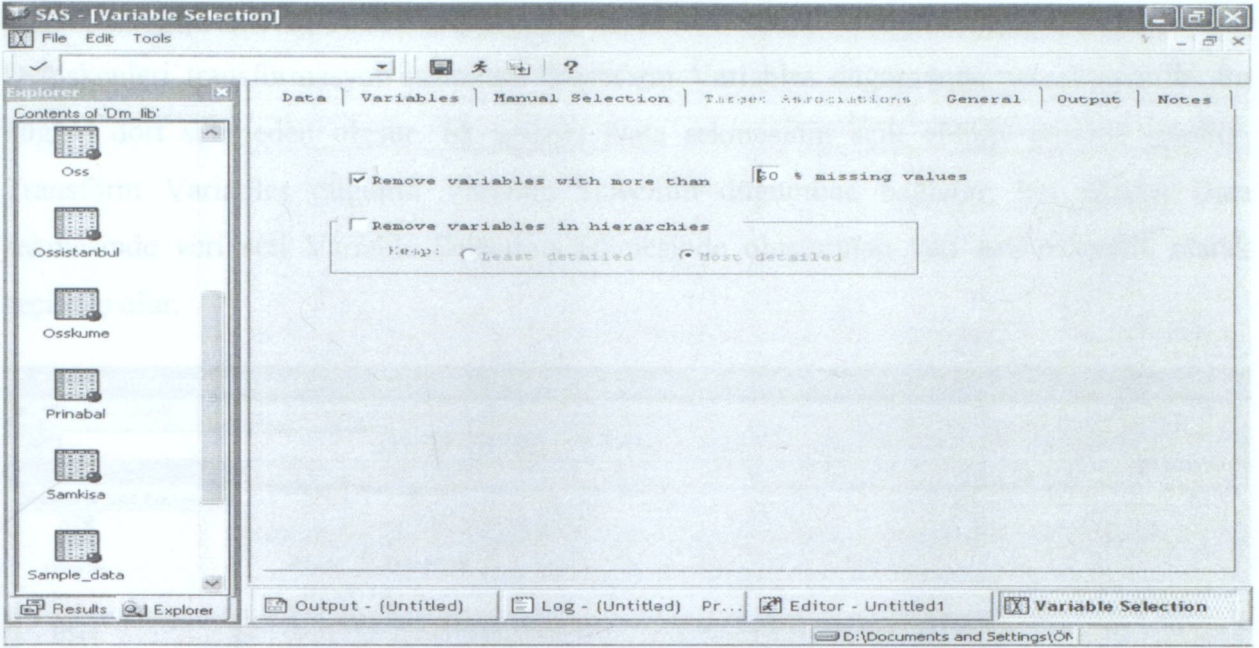
Şekil 6.15 Variable Selection Düğümünün Manual Selection Sekmesi

Manual Selection sekmesinde değişkenlerin rolleri manuel olarak belirlenir. Manuel olarak rol belirlenmezse rol atanması işlemi program tarafından otomatik olarak yapılır.



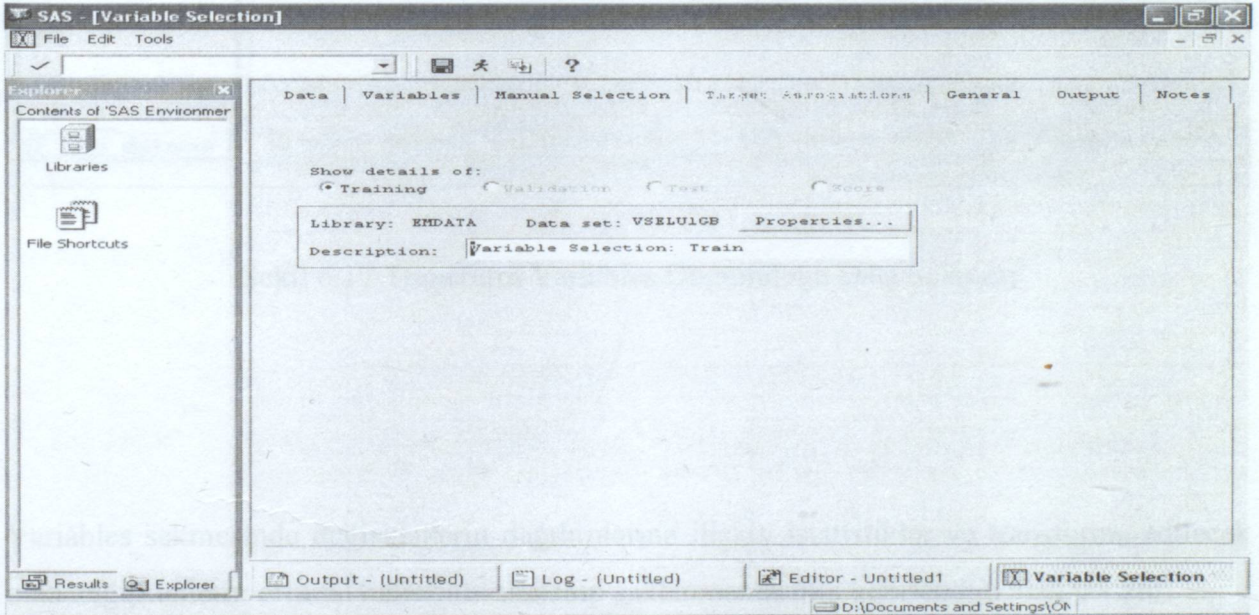
Şekil 6.14 Variable Selection Düğümünün Manual Selection Sekmesi

General sekmesinde ise kayıp değerler ve detaylı olan değişkenlerin dışlanma kriterleri belirlenir. Bu uygulama verisinde kayıp veri olmadığından ve detaylı değişkenlerin ayrıştırılmasına gerek olmadığından bu sekme kullanılmamıştır. Aşağıdaki şekilde General sekmesinin görüntüsü bulunmaktadır.



Şekil 6.15 Variable Selection Düğümünün Manual Selection Sekmesi

Output sekmesinde analizde kullanılacak seçilmiş değişken setinin oluşturduğu verinin adı ve bilgisayarda kaydedildiği adresi bulunur.

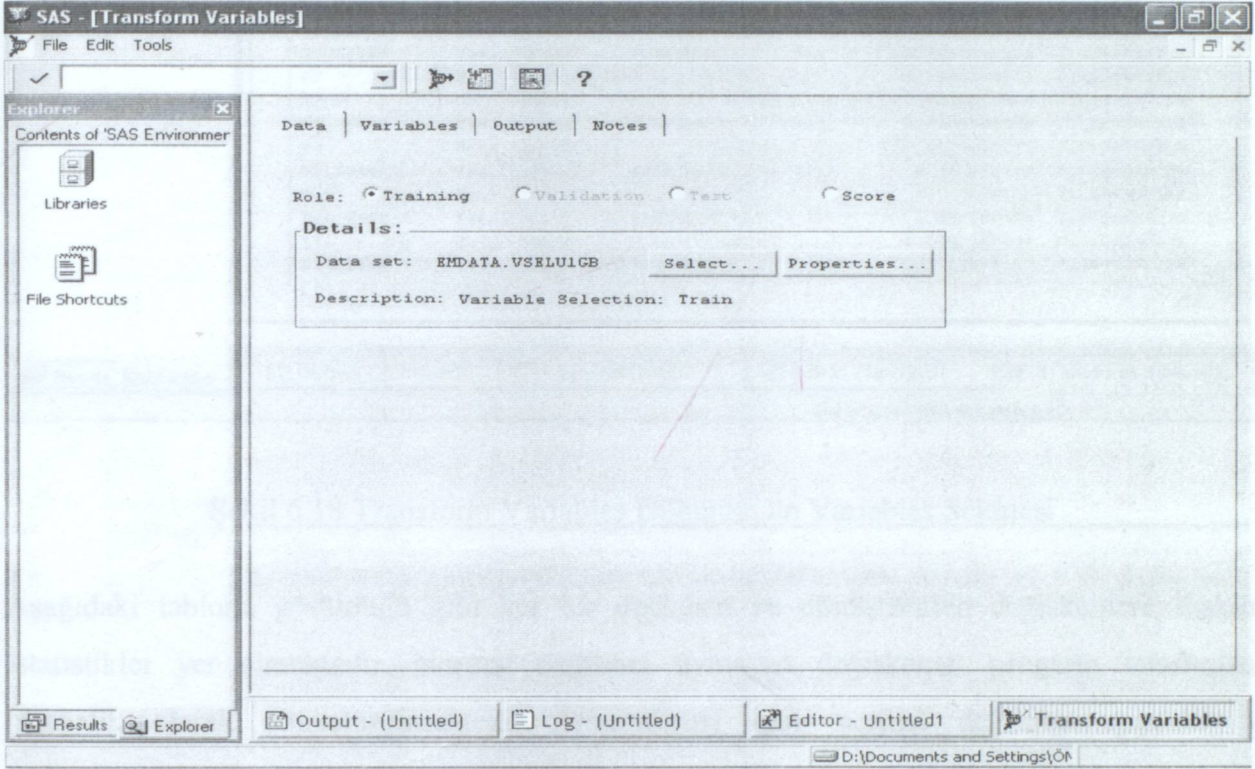


Şekil 6.16 Variable Selection Düğümünün Output Sekmesi

Notes sekmesi de yine Input Data Source düğümündeki gibi kullanıcı notları için ayrılmış bir alandır.

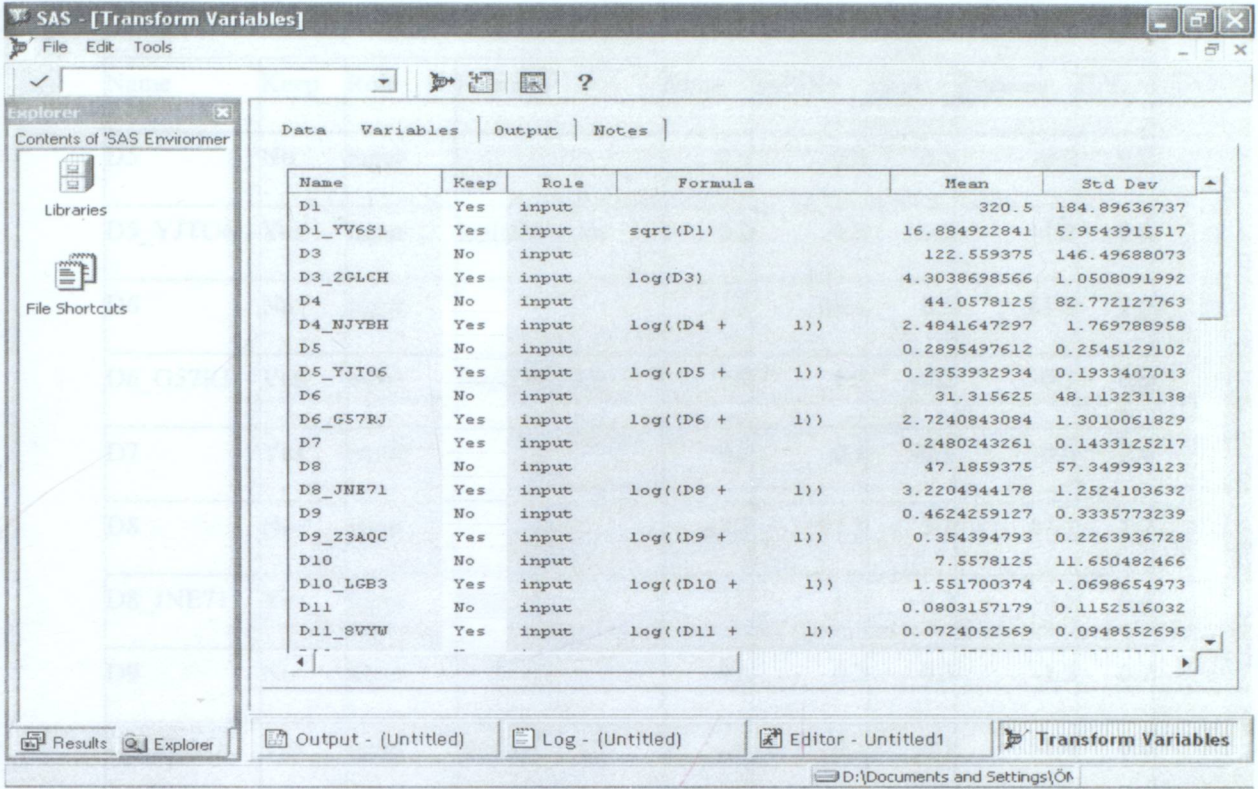
6.2.3.3 Değişkenlerin Transformasyonu

Değişkenleri transformasyon işlemleri Transform Variables düğümünde gerçekleştirilir. Bu düğüm dört sekmeden oluşur. İlk açılışta Data sekmesinin açık olduğu pencere görülür. Transform Variables düğümü Variable Selection düğümüne bağlanır. Bu yüzden Data sekmesinde veri seti Variable Selection sekmesinde oluşturulan veri seti otomatik olarak seçilmiş olur.



Şekil 6.17 Transform Variables Düğümünün Data Sekmesi

Variables sekmesinde değişkenlerin dağılımlarına ilişkin istatistikler ve transforme edilecek olan değişkenlerin orijinal hallerinin saklanıp saklanmamasının belirlendiği alan yer alır.



Şekil 6.18 Transform Variables Düğümünün Variables Sekmesi

Aşağıdaki tabloda görüldüğü gibi her bir değişken ve dönüştürülen değişkenlere ilişkin istatistikler yer almaktadır. Normal dağılıma uymayan değişkenler program tarafından otomatik olarak dönüştürülmektedir. Dönüştürme için, normal dağılıma uygunluğun maksimize edilmesi (maksimize normality) seçeneği tercih edildiğinden dönüştürmede kullanılacak olan formüller arasından en uygun formül program tarafından seçilmiştir.

Çizelge 6.5 Transformasyon Tablosu

Name	Keep	Role	Formula	Mean	Std Dev	Skew	Kurtosis	C.V.
D1	Yes	input		320.5	184.9	0.0	-1.2	0.6
D1_YV6S1	Yes	input	sqrt(D1)	16.9	6.0	-0.6	-0.6	0.4
D3	No	input		122.6	146.5	5.4	58.3	1.2
D3_2GLCH	Yes	input	log(D3)	4.3	1.1	-0.2	-0.4	0.2
D4	No	input		44.1	82.8	3.9	19.8	1.9
D4_NJYBH	Yes	input	log((D4 + 1))	2.5	1.8	0.1	-1.1	0.7

Çizelge 6.5 Devam Transformasyon Tablosu

Name	Keep	Role	Formula	Mean	Std Dev	Skew	Kurtosis	C.V.
D5	No	input		0.3	0.3	0.5	-0.9	0.9
D5_YJTO6	Yes	input	$\log((D5 + 1))$	0.2	0.2	0.3	-1.2	0.8
D6	No	input		31.3	48.1	6.5	81.8	1.5
D6_G57RJ	Yes	input	$\log((D6 + 1))$	2.7	1.3	-0.2	-0.5	0.5
D7	Yes	input		0.2	0.1	-0.1	-0.6	0.6
D8	No	input		47.2	57.3	3.1	17.7	1.2
D8_JNE71	Yes	input	$\log((D8 + 1))$	3.2	1.3	-0.3	-0.5	0.4
D9	No	input		0.5	0.3	0.3	-1.3	0.7
D9_Z3AQC	Yes	input	$\log((D9 + 1))$	0.4	0.2	0.1	-1.4	0.6
D10	No	input		7.6	11.7	3.5	16.3	1.5
D10_LGB3	Yes	input	$\log((D10+1))$	1.6	1.1	0.3	-0.6	0.7
D11	No	input		0.1	0.1	2.7	8.7	1.4
D11_8VYW	Yes	input	$\log((D11+1))$	0.1	0.1	2.4	6.1	1.3
D12	No	input		19.7	34.3	3.0	10.5	1.7
D12_KXSL	Yes	input	$\log((D12+1))$	2.0	1.5	0.4	-0.9	0.7
D13	No	input		0.2	0.3	1.3	0.4	1.2
D13_DG79	Yes	input	$\log((D13+1))$	0.2	0.2	1.0	-0.2	1.1
D14	No	input		6.9	28.3	21.3	504.1	4.1
D14_YQUQ	Yes	input	$\log((D14+1))$	1.3	1.1	0.5	0.2	0.8
D15	No	input		0.1	0.1	2.8	11.0	1.3
D15_ZKJ8	Yes	input	$\log((D15+1))$	0.0	0.1	2.3	7.8	1.2
D16	No	input		0.7	0.2	-0.8	-0.5	0.4
D16_F4IO	Yes	input	$\exp(D16)$	2.0	0.4	-0.4	-1.0	0.2
D17	No	input		10.2	9.6	1.2	1.1	0.9

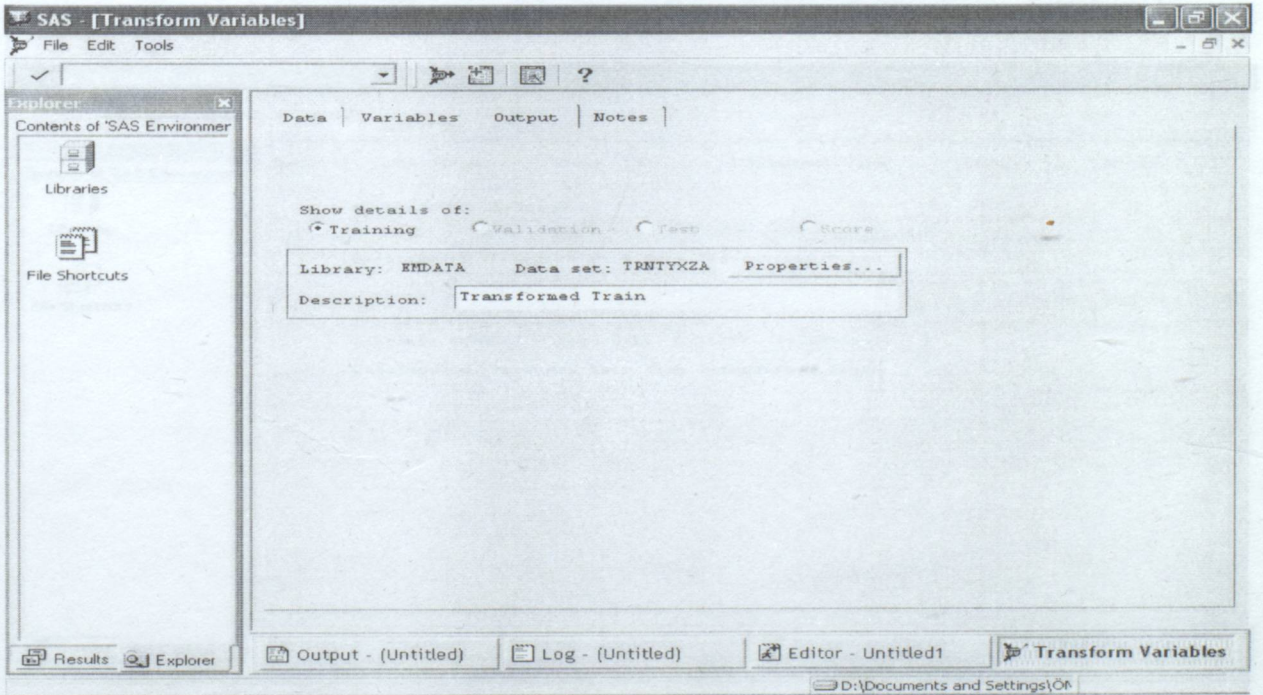
Çizelge 6.5 Devam Transformasyon Tablosu

Name	Keep	Role	Formula	Mean	Std Dev	Skew	Kurtosis	C.V.
D17_2HTL	Yes	input	$\log((D17+2.32))$	2.2	0.8	-0.1	-0.9	0.4
D18	Yes	input		8.3	3.5	-0.3	-0.9	0.4
D19	No	input		2.8	3.2	1.7	3.3	1.1
D19_606L	Yes	input	$\log((D19+1.64))$	1.3	0.6	0.5	-0.7	0.5
D20	No	input		3.5	1.8	0.2	-1.1	0.5
D20_W45K	Yes	input	$\sqrt{D20}$	1.8	0.5	-0.2	-0.9	0.3
D21	No	input		1.9	2.4	1.8	3.6	1.3
D21_ZI4D	Yes	input	$\log((D21+1.57))$	1.1	0.6	0.7	-0.4	0.5
D22	No	input		2.6	1.4	0.3	-1.1	0.5
D22_Y9JB	Yes	input	$\sqrt{D22}$	1.5	0.5	-0.1	-0.9	0.3
D23	No	input		0.8	1.1	2.7	9.1	1.4
D23_1FDO	Yes	input	$\log((D23+1.44))$	0.7	0.4	1.4	1.7	0.5
D24	No	input		1.5	0.7	0.9	0.3	0.5
D24_UN9H	Yes	input	$(D24)^{** 0.25}$	1.1	0.1	-0.1	0.5	0.1
D25	No	input		23.1	8.0	0.2	-0.9	0.3
D25_GCIE	Yes	input	$\sqrt{D25}$	4.7	0.9	-0.2	-0.2	0.2
D26	Yes	input		9.1	1.5	-0.5	1.6	0.2
D27	No	input		7.2	2.8	0.4	-0.5	0.4
D27_DQ06	Yes	input	$\sqrt{(D27+1.48)}$	2.9	0.5	0.0	-0.4	0.2
D28	Yes	input		4.7	0.8	-0.2	0.7	0.2
D29	No	input		4.4	1.8	1.1	1.3	0.4
D29_AFEO	Yes	input	$\log(D29)$	1.4	0.4	-0.2	0.5	0.3
D30	No	input		3.3	0.7	0.3	0.1	0.2
D30_VMZ5	Yes	input	$\sqrt{D30}$	1.8	0.2	-0.1	0.4	0.1

Çizelge 6.5 Devam Transformasyon Tablosu

Name	Keep	Role	Formula	Mean	Std Dev	Skew	Kurtosis	C.V.
D31	No	input		3.3	2.6	13.4	273.2	0.8
D31_ARE9	Yes	input	log(D31)	1.0	0.6	-1.2	5.1	0.6
D32	No	input		2.9	0.6	-0.6	1.2	0.2
D32_J25H	Yes	input	$((D32 + 1))^{** 2}$	15.7	4.5	0.0	0.1	0.3
D33	Yes	rejected		10.4	24.5	2.2	3.1	2.4
D34	Yes	rejected		3.3	7.9	2.3	3.9	2.4
D35	Yes	rejected		0.8	7.9	10.4	107.6	10.4
D36	Yes	rejected		0.1	1.3	14.5	237.0	11.8
D37	Yes	rejected		0.6	6.7	10.6	113.4	10.4
D38	Yes	rejected		0.2	2.3	12.8	181.5	11.3

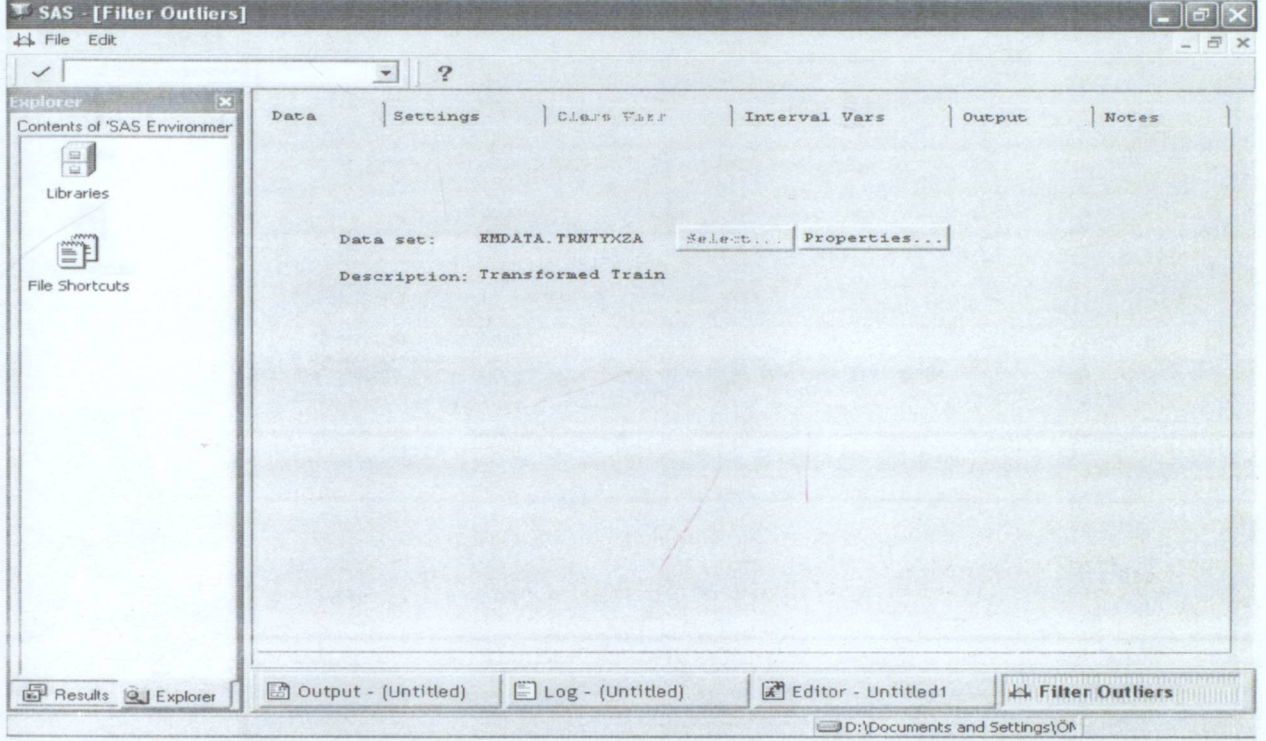
Output Sekmesinde transforme edilecek verinin adı ve bilgisayarda nereye kayıt edileceği görülür.



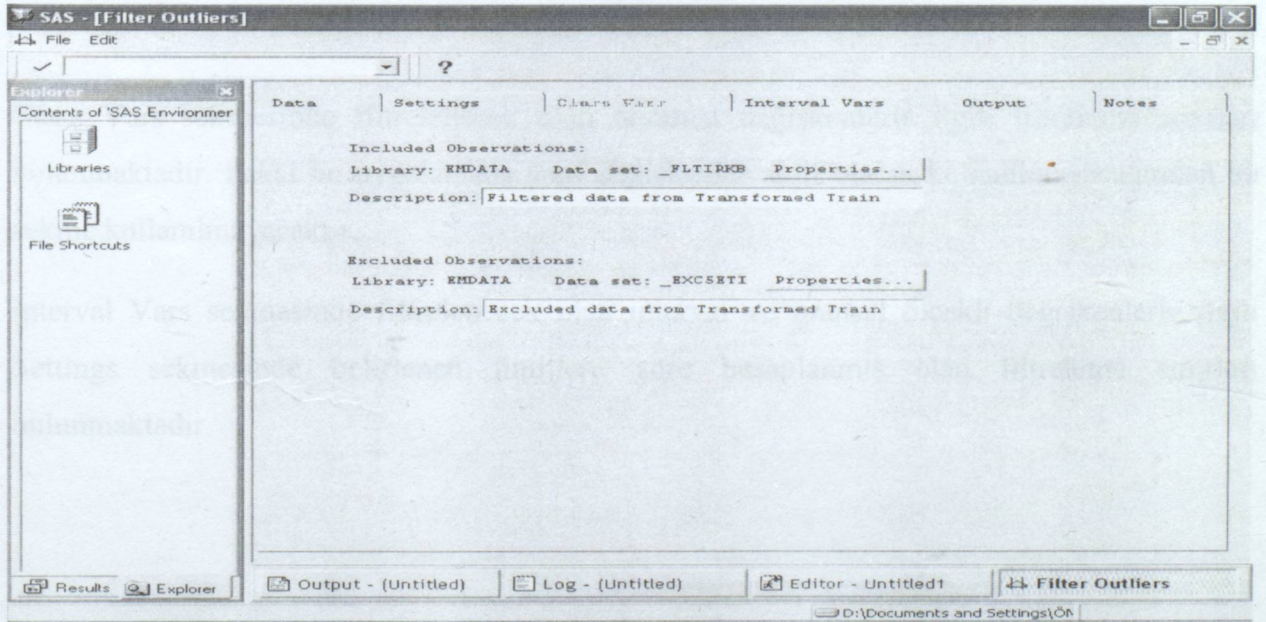
Şekil 6.19 Transform Variables Düğümünün Output Sekmesi

6.2.3.4 Aşırı Değerlerin Filtrelenmesi

Aşırı değerlerin filtrelenmesi işlemi Filter Outliers düğümünde gerçekleştirilir. Filter Outliers düğümü Transform Variables düğümüne bağlandığından, ilk açılan pencerede görülen data sekmesinde Transform Variables düğümünün çıktısı olan veri seti otomatik olarak görünür.

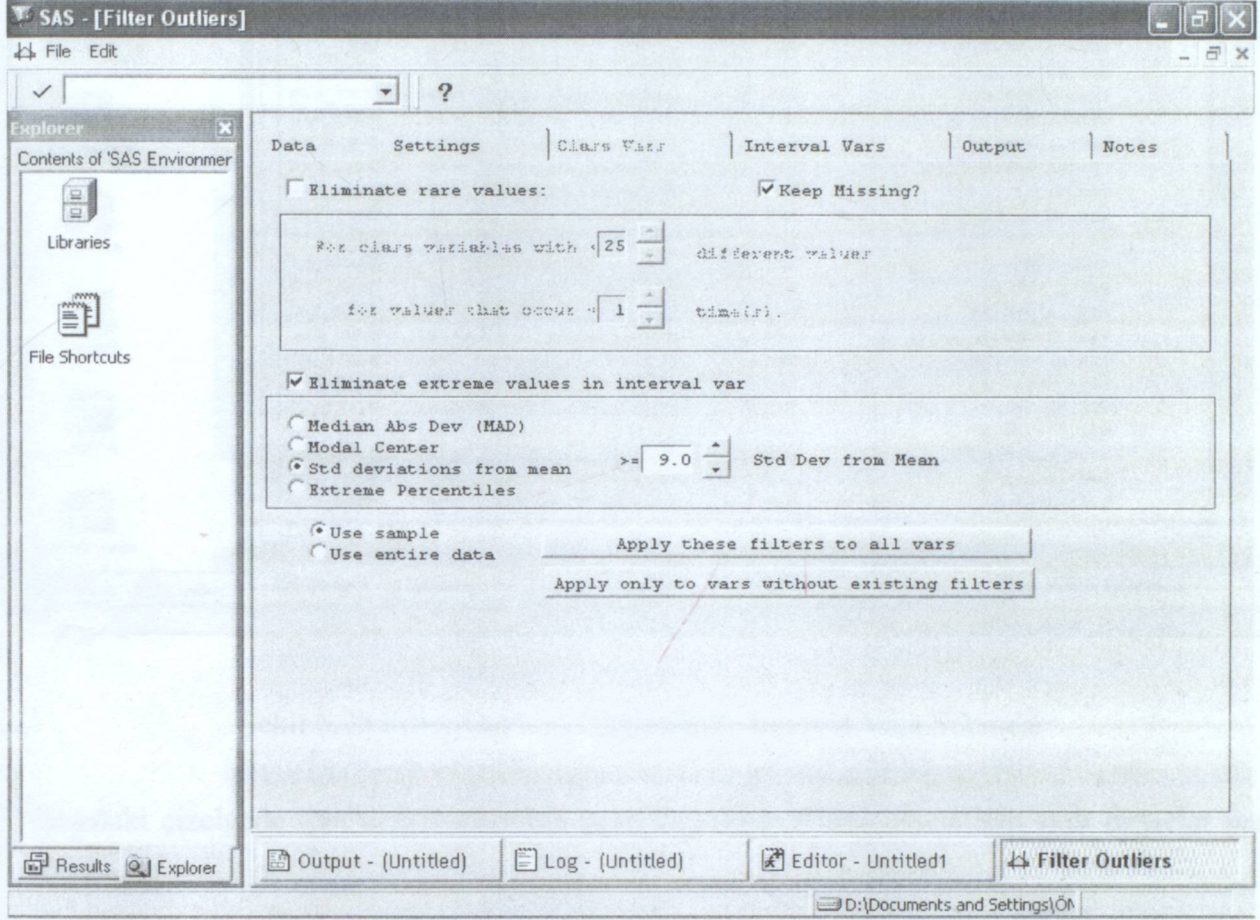


Şekil 6.20 Filter Outliers Düğümünün Data Sekmesi



Şekil 6.21 Filter Outliers Düğümünün Output Sekmesi

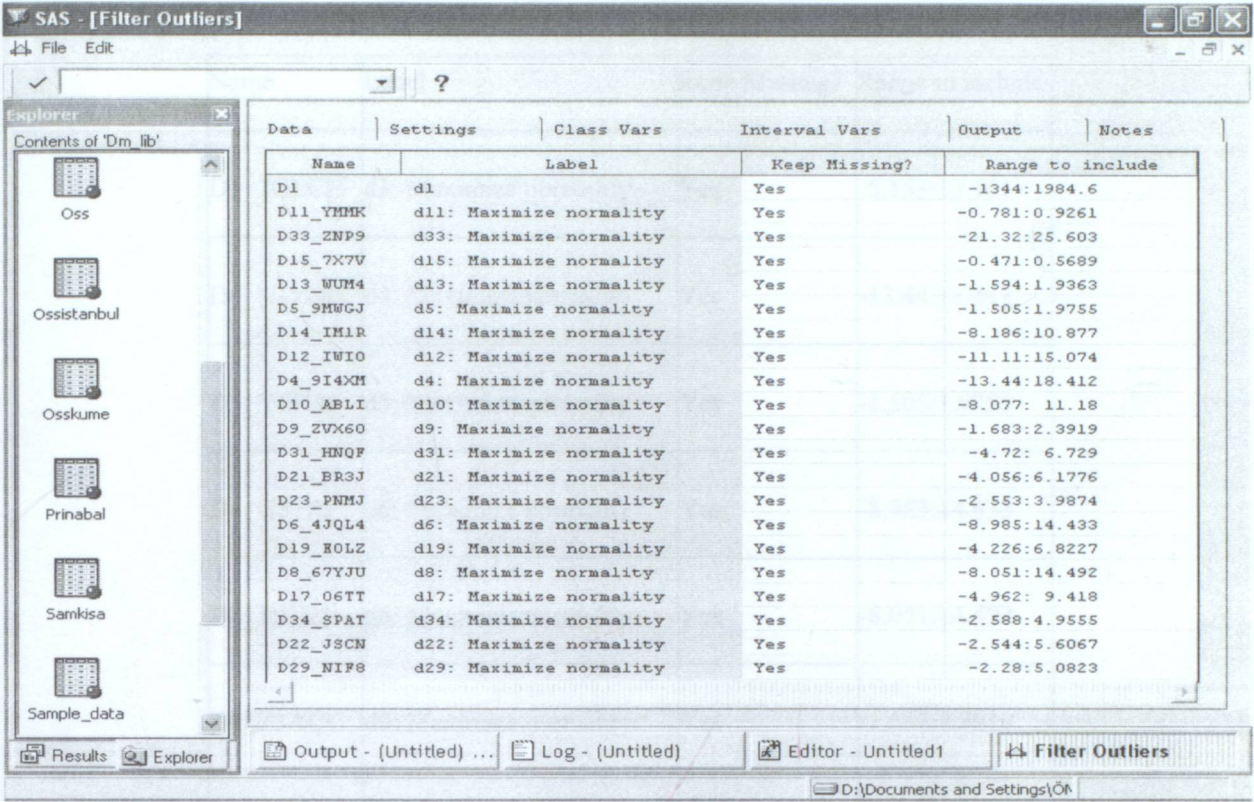
Settings sekmesinde elimine edilecek olan aşırı değerlerle ilgili kriterler belirlenir.



Şekil 6.22 Filter Outliers Düğümünün Settings Sekmesi

Class Vars sekmesinde filtrelenecek olan nominal değişkenlerle ilgili filtreleme sınırları bulunmaktadır. Fakat bu uygulamada sınıf değişkenler aktif olarak kullanılmayacağından bu sekme kullanılmayacaktır.

Interval Vars sekmesinde filtrelenecek olan interval ve oransal ölçekli değişkenlerle ilgili Settings sekmesinde belirlenen limitlere göre hesaplanmış olan filtreleme sınırları bulunmaktadır.



Şekil 6.23 Filter Outliers Düğümünün Interval Vars Sekmesi

Aşağıdaki çizelgede tüm değişkenlerdeki aşırı değerlerin filtrelenmesi için sınır değerler yer almaktadır.

Çizelge 6.6 Analizde Kullanılacak Değişkenlere İlişkin Filtreleme Sınırları

Name	Label	Keep Missing?	Range to include
D1	d1	Yes	-1344:1984.6
D7	d7	Yes	-1.042:1.5378
D18	d18	Yes	-22.78:39.405
D26	d26	Yes	-4.825:22.962
D28	d28	Yes	-2.089:11.551
D1_YV6S1	d1: Maximize normality	Yes	-36.7:70.474

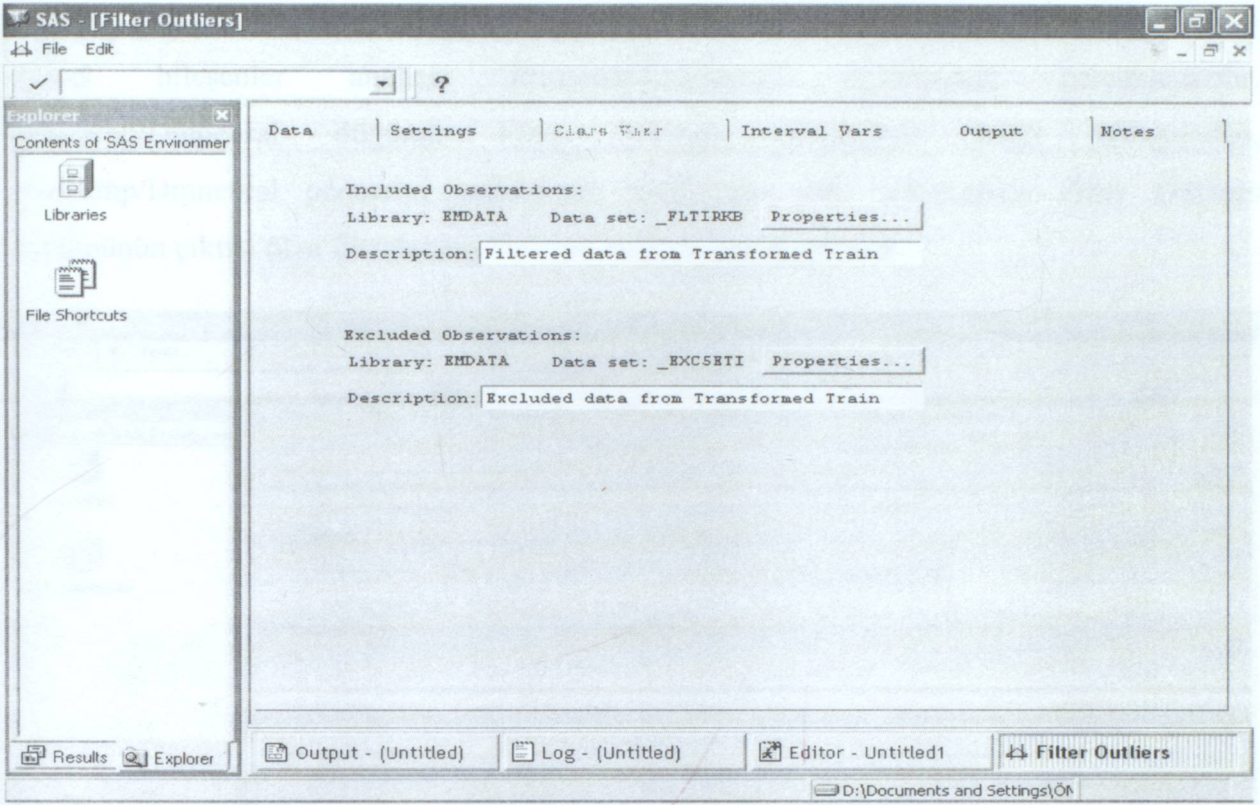
Çizelge 6.6 Devam Analizde Kullanılacak Değişkenlere İlişkin Filtreleme Sınırları

Name	Label	Keep Missing?	Range to include
D3_2GLCH	d3: Maximize normality	Yes	-5.153:13.761
D4_NJYBH	d4: Maximize normality	Yes	-13.44:18.412
D5_YJTO6	d5: Maximize normality	Yes	-1.505:1.9755
D6_G57RJ	d6: Maximize normality	Yes	-8.985:14.433
D8_JNE71	d8: Maximize normality	Yes	-8.051:14.492
D9_Z3AQC	d9: Maximize normality	Yes	-1.683:2.3919
D10_LGB3	d10: Maximize normality	Yes	-8.077: 11.18
D11_8VYW	d11: Maximize normality	Yes	-0.781:0.9261
D12_KXSL	d12: Maximize normality	Yes	-11.11:15.074
D13_DG79	d13: Maximize normality	Yes	-1.594:1.9363
D14_YQUQ	d14: Maximize normality	Yes	-8.186:10.877
D15_ZKJ8	d15: Maximize normality	Yes	-0.471:0.5689
D16_F4IO	d16: Maximize normality	Yes	-2.059:6.0265
D17_2HTL	d17: Maximize normality	Yes	-4.962: 9.418
D19_606L	d19: Maximize normality	Yes	-4.226:6.8227
D20_W45K	d20: Maximize normality	Yes	-2.687:6.3192

Çizelge 6.6 Devam Analizde Kullanılacak Değişkenlere İlişkin Filtreleme Sınırları

Name	Label	Keep Missing?	Range to include
D21_ZI4D	d21: Maximize normality	Yes	-4.055:6.1779
D22_Y9JB	d22: Maximize normality	Yes	-2.544:5.6067
D23_1FDO	d23: Maximize normality	Yes	-2.553:3.9874
D24_UN9H	d24: Maximize normality	Yes	-0.1:2.2635
D25_GCIE	d25: Maximize normality	Yes	-2.928:12.394
D27_DQ06	d27: Maximize normality	Yes	-1.405:7.2145
D29_AFEO	d29: Maximize normality	Yes	-2.28:5.0823
D30_VMZ5	d30: Maximize normality	Yes	0.0147:3.6019
D31_ARE9	d31: Maximize normality	Yes	-4.72: 6.729
D32_J25H	d32: Maximize normality	Yes	-24.82:56.207

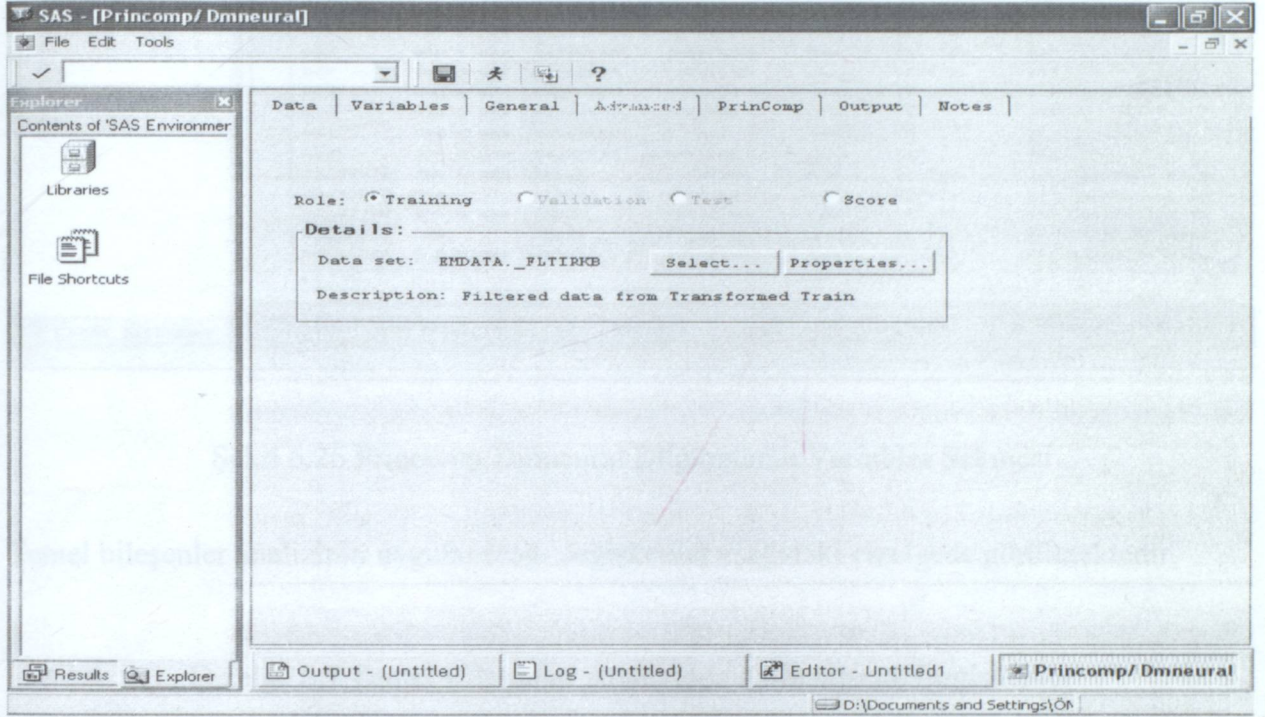
Output sekmesinde ise dışlanan gözlemler ve dışlanmayan gözlemlerin oluşturduğu veri setlerinin adları ve bilgisayarda kaydedilecekleri adresleri bulunur.



Şekil 6.24 Filter Outliers Düğümünün Output Sekmesi

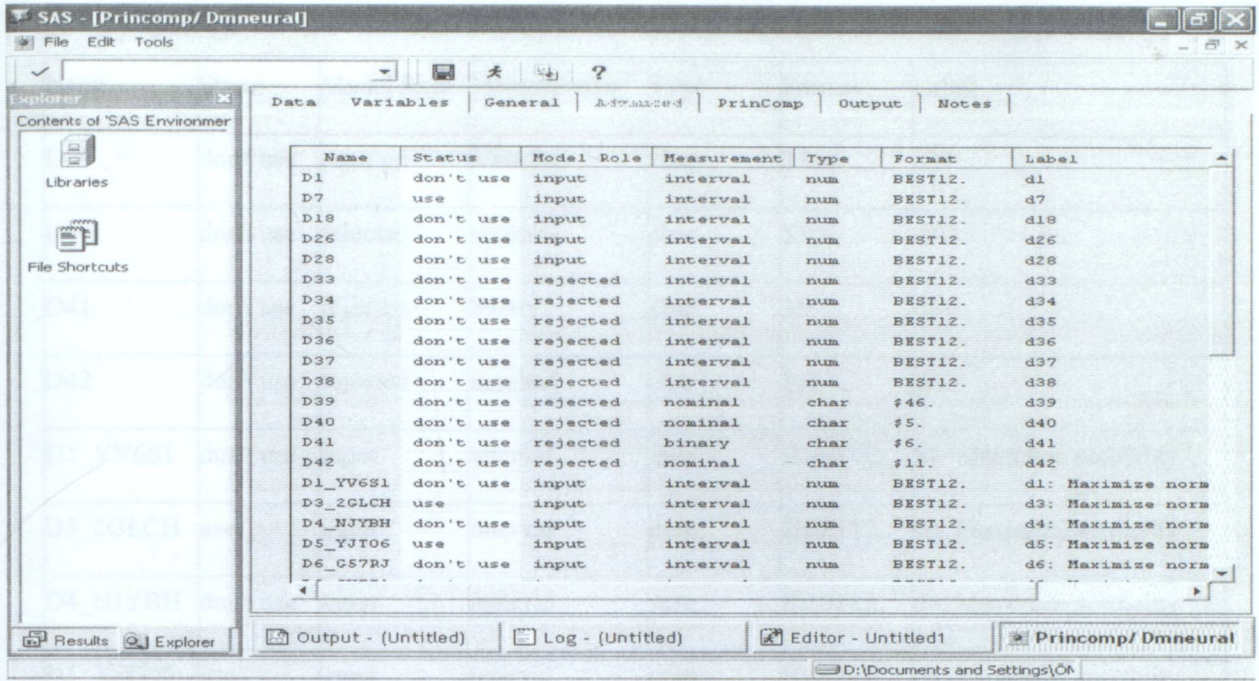
6.2.3.5 Temel Bileşenler Analizi

Temel bileşenler analizi, Princomp/Dmneural düğümünde gerçekleştirilir. Princomp/Dmneural düğümü Filter Outliers düğümüne bağlı olduğundan, Princomp/Dmneural penceresi açıldığında aktif olan data sekmesinde Filter Outliers düğümünün çıktısı olan filtrelenmiş veri seti kaynak olarak görülür.



Şekil 6.25 Princomp/Dmneural Düğümünün Data Sekmesi

Variables Sekmesinde temel bileşenler analizinde kullanılacak olan değişkenler seçilir.



Şekil 6.26 Princomp/Dmneural Düğümünün Variables Sekmesi

Temel bileşenler analizinin uygulanacağı değişkenler aşağıdaki çizelgede görülmektedir.

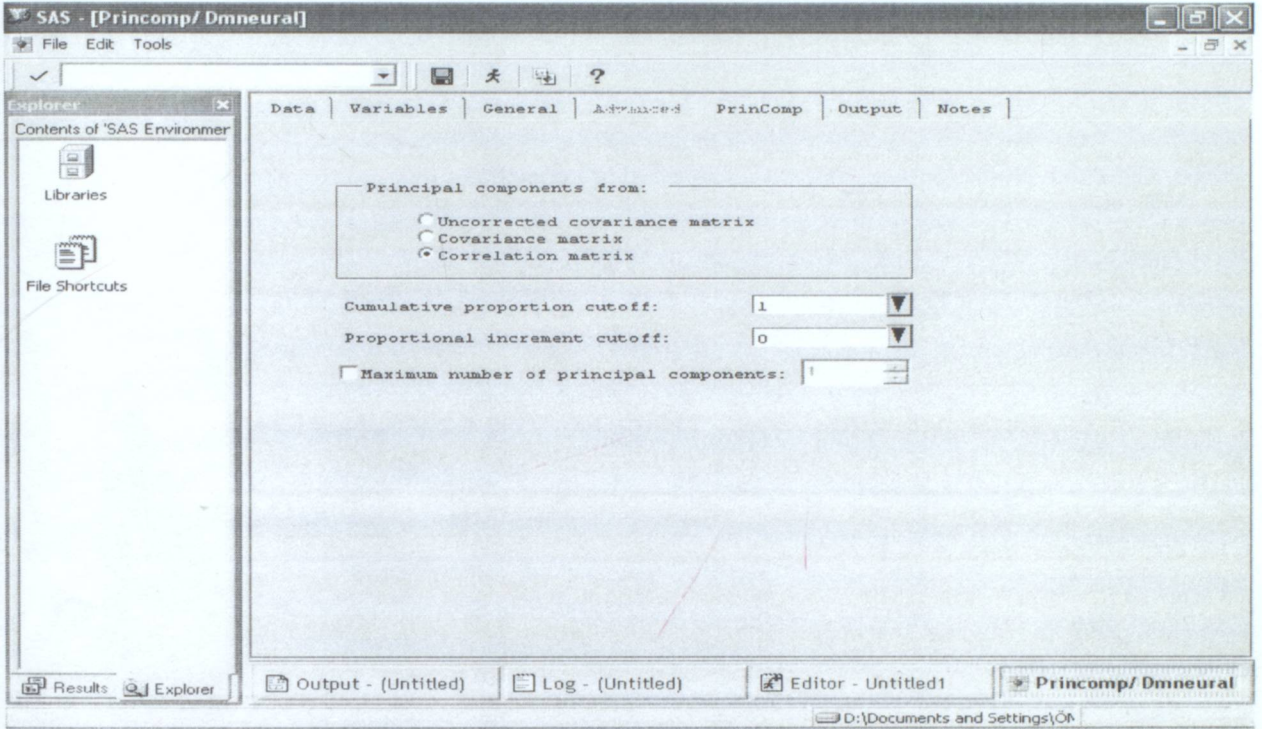
Çizelge 6.7 Temel Bileşenler Analizinde Kullanılacak Değişkenler

Name	Status	Model Role	Measurement	Type	Format	Label
D1	don't use	input	interval	num	BEST12.	d1
D7	use	input	interval	num	BEST12.	d7
D18	don't use	input	interval	num	BEST12.	d18
D26	don't use	input	interval	num	BEST12.	d26
D28	don't use	input	interval	num	BEST12.	d28
D33	don't use	rejected	interval	num	BEST12.	d33
D34	don't use	rejected	interval	num	BEST12.	d34
D35	don't use	rejected	interval	num	BEST12.	d35
D36	don't use	rejected	interval	num	BEST12.	d36
D37	don't use	rejected	interval	num	BEST12.	d37
D38	don't use	rejected	interval	num	BEST12.	d38

Çizelge 6.7 Devam Temel Bileşenler Analizinde Kullanılacak Değişkenler

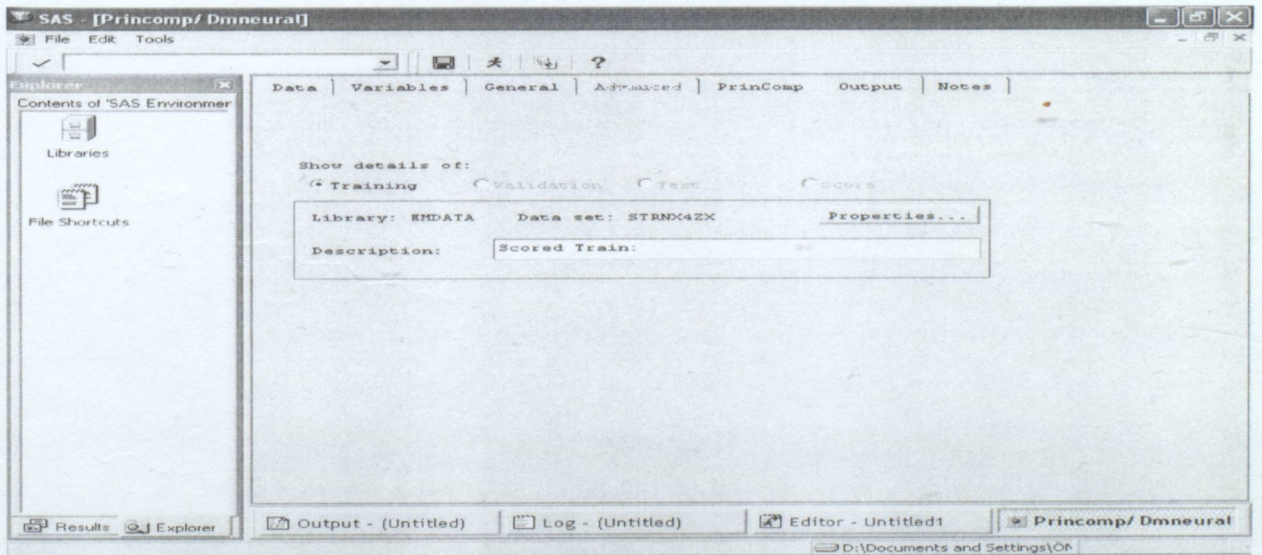
Name	Status	Model Role	Measurement	Type	Format	Label
D39	don't use	rejected	nominal	char	\$46	d39
D40	don't use	rejected	nominal	char	\$5	d40
D41	don't use	rejected	binary	char	\$6	d41
D42	don't use	rejected	nominal	char	\$11	d42
D1_YV6S1	don't use	input	interval	num	BEST12.	d1: Maximize normality
D3_2GLCH	use	input	interval	num	BEST12.	d3: Maximize normality
D4_NJYBH	don't use	input	interval	num	BEST12.	d4: Maximize normality
D5_YJTO6	use	input	interval	num	BEST12.	d5: Maximize normality
D6_G57RJ	don't use	input	interval	num	BEST12.	d6: Maximize normality
D8_JNE71	don't use	input	interval	num	BEST12.	d8: Maximize normality
D9_Z3AQC	use	input	interval	num	BEST12.	d9: Maximize normality
D10_LGB3	don't use	input	interval	num	BEST12.	d10: Maximize normality
D11_8VYW	use	input	interval	num	BEST12.	d11: Maximize normality
D12_KXSL	don't use	input	interval	num	BEST12.	d12: Maximize normality
D13_DG79	use	input	interval	num	BEST12.	d13: Maximize normality
D14_YQUQ	don't use	input	interval	num	BEST12.	d14: Maximize normality
D15_ZKJ8	use	input	interval	num	BEST12.	d15: Maximize normality
D16_F4IO	use	input	interval	num	BEST12.	d16: Maximize normality
D17_2HTL	use	input	interval	num	BEST12.	d17: Maximize normality
D19_606L	use	input	interval	num	BEST12.	d19: Maximize normality
D20_W45K	don't use	input	interval	num	BEST12.	d20: Maximize normality
D21_ZI4D	use	input	interval	num	BEST12.	d21: Maximize normality
D22_Y9JB	don't use	input	interval	num	BEST12.	d22: Maximize normality
D23_1FDO	use	input	interval	num	BEST12.	d23: Maximize normality

Princomp sekmesinde, temel bileşenlerin korelasyon ya da kovaryans matrislerinden mi oluşturulacağı seçimi yapılır. Bu uygulamada temel bileşenler korelasyon matrisinden elde edilecektir. Temel bileşenler açıklanan varyans yüzdesi %100 oluncaya kadar hesaplanacaktır.



Şekil 6.28 Princomp/Dmneural Düğümünün Princomp Sekmesi

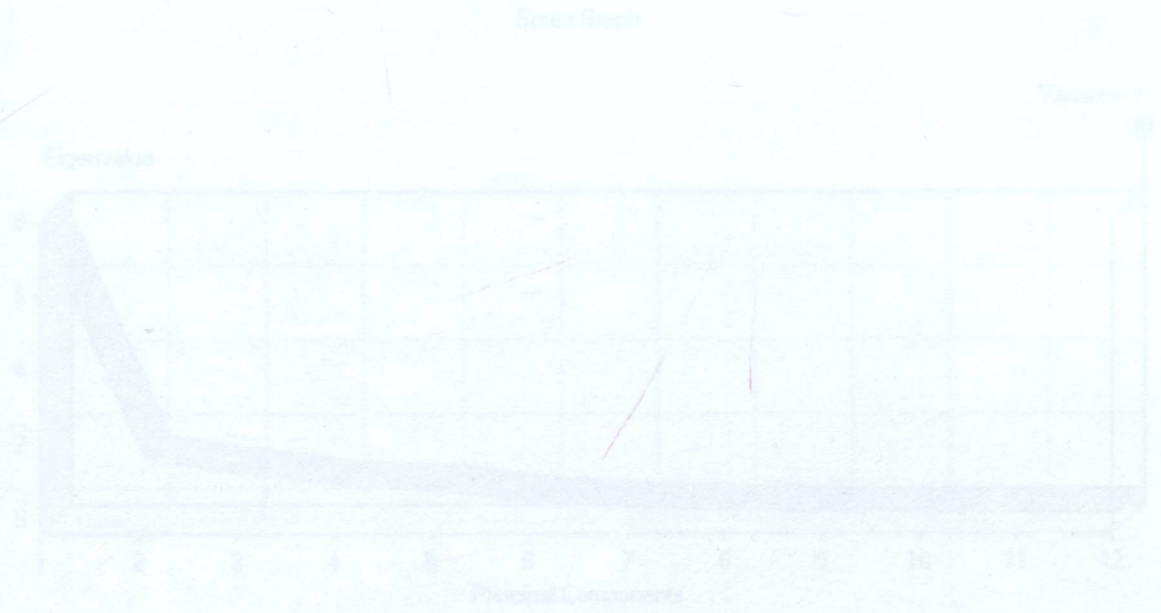
Output sekmesinde ise temel bileşenlerin eklendiği veri setinin adı ve bilgisayardaki adresi görülür.



Şekil 6.29 Princomp/Dmneural Düğümünün Output Sekmesi

Notes sekmesinde ise diğer düğümlere benzer olarak kullanıcı notları için ayrılmış bir alan bulunmaktadır.

Analiz sonuçları konularından bahsederken toplam 12 tane değişken analize dahil olan değişkenlerdir. Toplam 12 değişkeni açıklarız. Karşılıklı ilişkilerde elde edilen sonuçlara ilişkin grafik aşağıdaki şekilde görülmektedir. Çizelge 6.3'te ise sonuçlar, özdeğerlerin kümülatif ve kümülatif olmayan varyans açıklama oranları, özdeğerler arası fark görülmektedir.



Şekil 6.30 Özdeğerlerin Yanına (Scree) Grafiği

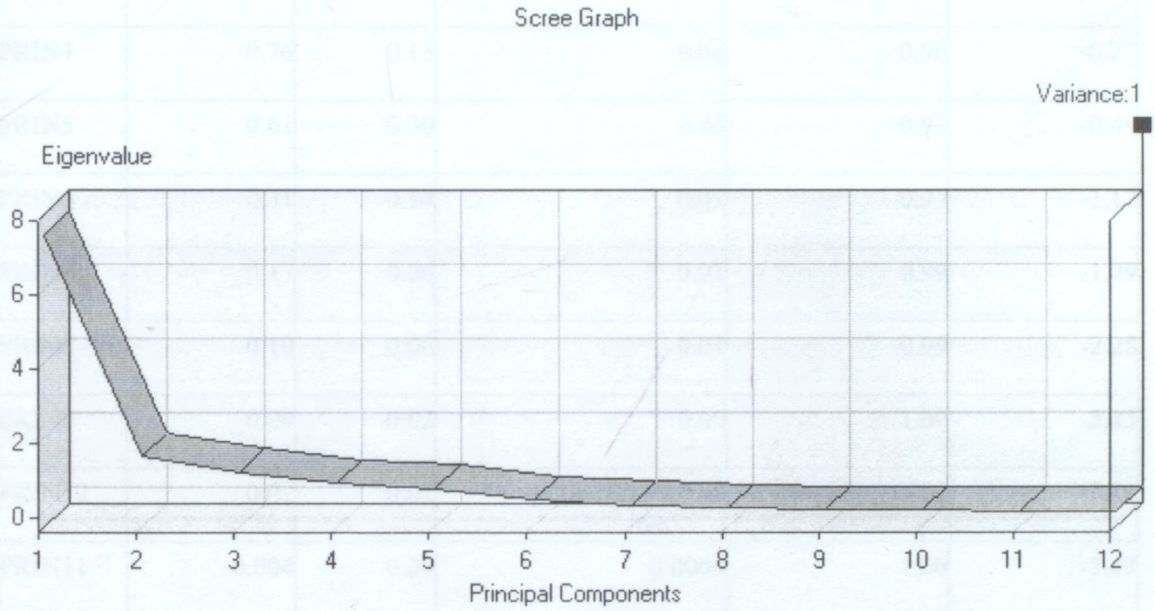
Bu özdeğerlerin varyans açıklama oranları ise aşağıdaki grafikte görülmektedir.



Şekil 6.31 Özdeğerlerin Varyans Açıklama Oranları Grafiği

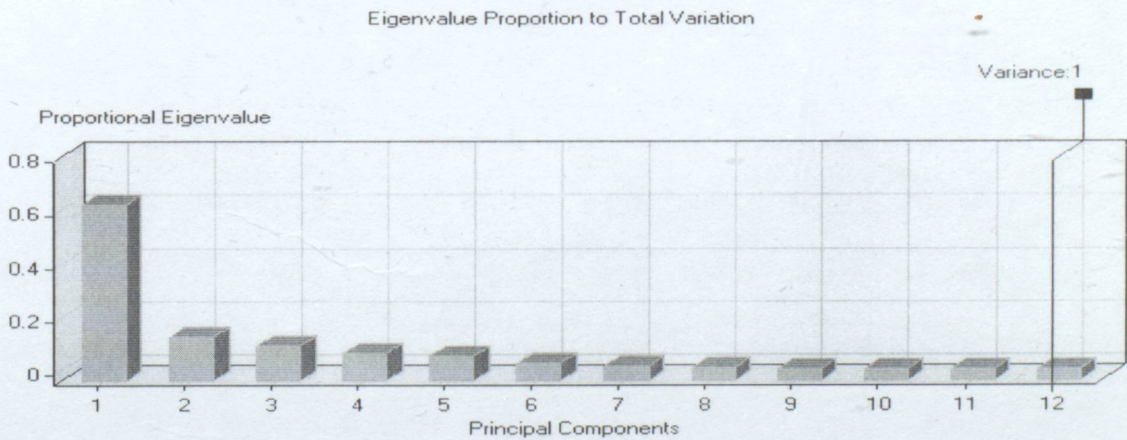
6.2.3.5.1 Temel Bileşenler Analizinin Sonuçları

Temel bileşenler analizi sonucun korelasyon matrisinden hesaplanan toplam 12 temel bileşen analize dahil olan değişkenlerdeki toplam değişimi açıklamıştır. Korelasyon matrisinden elde edilen özdeğerlere ilişkin grafik aşağıdaki şekilde görülmektedir. Çizelge 6.8'de ise özdeğerler, özdeğerlerin kümülatif ve kümülatif olmayan varyans açıklama oranları, özdeğerler arası fark görülmektedir.



Şekil 6.30 Özdeğerlerin Yamaç (Scree) Grafiği

Bu özdeğerlerin varyans açıklama oranları ise aşağıdaki grafikte görülmektedir.



Şekil 6.31 Özdeğerlerin Varyans Açıklama Oranları Grafiği

Çizelge 6.8 Temel Bileşenlere İlişkin Özdeğerler Tablosu

	NAME	EigenValue	Difference	ProportionalEigenvalue	Cum Proportional Eigenvalue	LogEigenvalue
1	PRIN1	7.44	5.97	0.62	0.62	2.01
2	PRIN2	1.47	0.41	0.12	0.74	0.39
3	PRIN3	1.07	0.31	0.09	0.83	0.07
4	PRIN4	0.76	0.15	0.06	0.90	-0.27
5	PRIN5	0.61	0.30	0.05	0.95	-0.49
6	PRIN6	0.31	0.14	0.03	0.97	-1.17
7	PRIN7	0.17	0.06	0.01	0.99	-1.79
8	PRIN8	0.10	0.06	0.01	0.99	-2.28
9	PRIN9	0.04	0.02	0.00	1.00	-3.23
10	PRIN10	0.02	0.02	0.00	1.00	-3.84
11	PRIN11	0.004	0.00	0.0004	1.00	-5.47
12	PRIN12	0.002	.	0.0001	1.00	-6.36

Prin* i. Temel bileşene ilişkin özdeğeri göstermektedir.

Çizelge 6.9 Özdeğerlere İlişkin Özvektörler 1

	NAME	PRIN1	PRIN2	PRIN3	PRIN4	PRIN5	PRIN6
1	d7	-0.254	-0.292	0.150	-0.009	0.741	0.040
2	D3_GLCH	-0.053	0.548	0.212	0.792	0.114	0.060
3	D5_YJTO6	-0.326	0.191	0.025	-0.171	-0.300	0.452
4	D9_Z3AQC	0.347	-0.067	-0.072	0.155	0.017	-0.423
5	D11_8VYW	-0.040	-0.647	0.431	0.371	-0.277	0.227
6	D13_DG79	0.346	0.083	-0.129	-0.030	-0.166	-0.137
7	D15_ZKJ8	-0.116	-0.196	-0.821	0.321	0.156	0.276
8	D16_F4IO	-0.312	0.287	0.154	-0.237	0.265	-0.067
9	D17_2HTL	0.346	-0.061	0.080	0.021	0.291	-0.032
10	D19_606L	0.346	0.015	0.122	-0.079	0.197	0.330
11	D21_ZI4D	0.351	0.064	0.069	-0.091	0.157	0.278
12	D23_1FDO	0.333	0.136	0.009	-0.086	0.042	0.523

Çizelge 6.10 Özdeğerlere İlişkin Özvektörler 2

	NAME	PRIN7	PRIN8	PRIN9	PRIN10	PRIN11	PRIN12
1	d7	-0.446	-0.075	-0.033	0.007	0.264	0.043
2	D3_GLCH	-0.081	0.037	-0.019	0.005	0.002	-0.001
3	D5_YJTO6	-0.076	0.296	0.206	0.100	0.616	0.091
4	D9_Z3AQC	0.283	-0.215	-0.045	-0.087	0.721	0.103
5	D11_8VYW	0.152	-0.042	0.002	0.037	-0.051	0.314
6	D13_DG79	-0.581	0.138	0.007	0.007	-0.066	0.670
7	D15_ZKJ8	0.183	0.056	-0.007	0.007	-0.034	0.180
8	D16_F4IO	0.487	-0.119	0.005	-0.021	-0.122	0.632
9	D17_2HTL	0.188	0.384	0.729	0.254	-0.074	-0.023

Çizelge 6.10 Devam Özdeğerlere İlişkin Özvektörler 2

	NAME	PRIN7	PRIN8	PRIN9	PRIN10	PRIN11	PRIN12
10	D19_606L	0.143	0.407	-0.251	-0.680	0.020	0.008
11	D21_ZI4D	0.130	0.099	-0.531	0.668	0.047	0.011
12	D23_1FDO	-0.062	-0.705	0.280	-0.089	-0.003	0.016

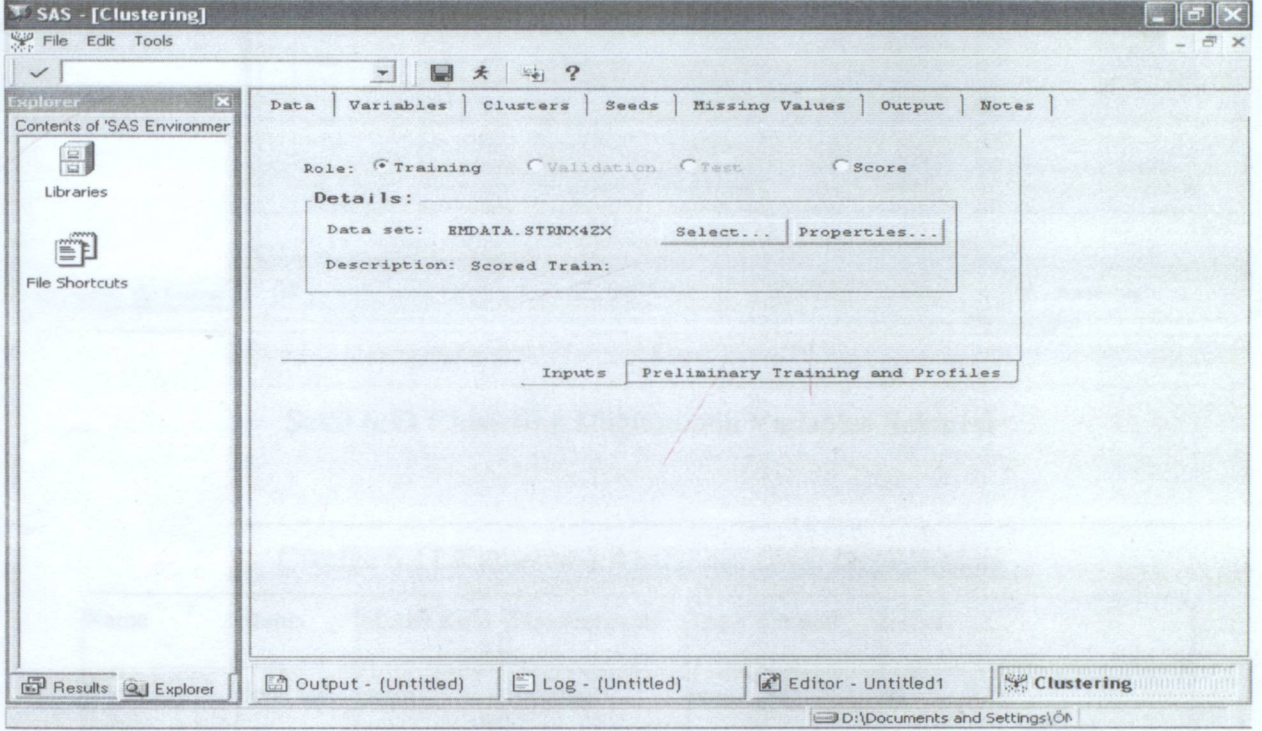


Şekil 6.32 Clustering Düğümünün Data Seçimi

Clustering düğümünün Variables sekmesinde ise analiz için kullanılacak değişkenler seçilir. Gerekli ise, standardizasyon yöntemi seçilir.

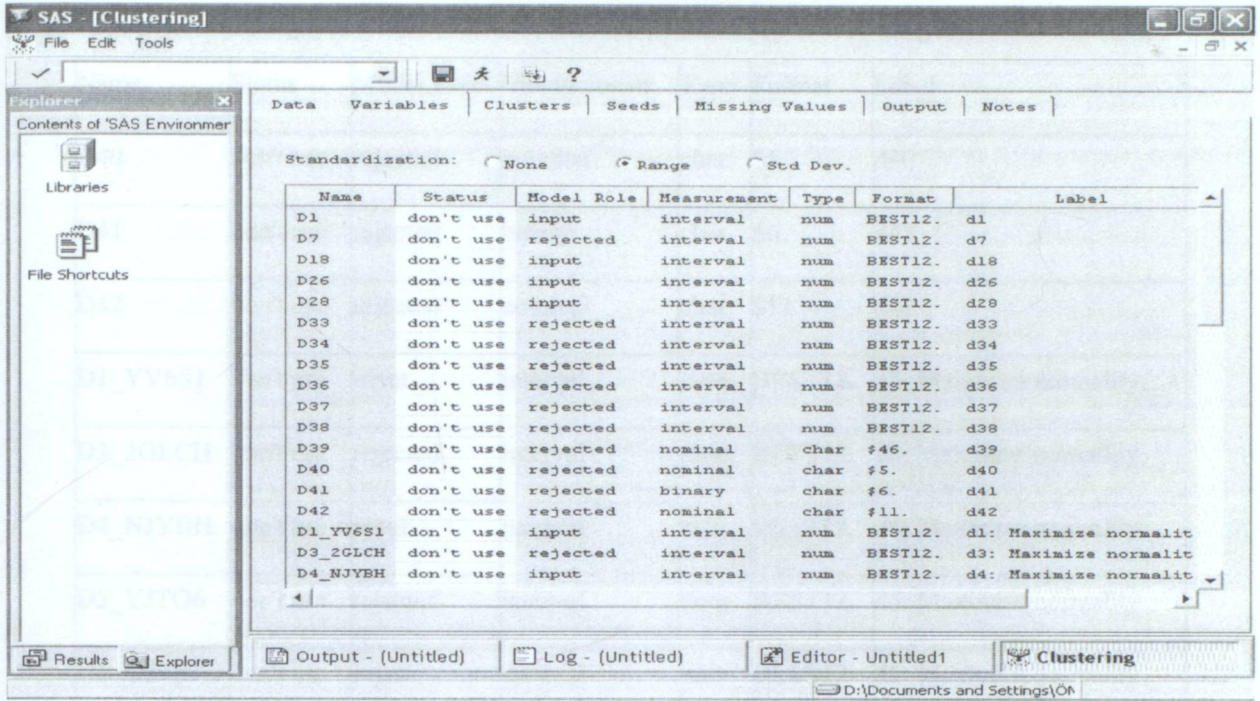
6.2.3.6 Kümeleme Analizi

SAS Enterprise Miner ürününde kümeleme analizi Clustering düğümünde gerçekleştirilir. Kümeleme analizi hiyerarşik ve hiyerarşik olmayan metotların bir arada kullanılması ile gerçekleştirilir. Clustering modülü açıldığında Data sekmesi aktif olarak görülür. Bu sekmede Clustering düğümünün bağlandığı temel bileşenler analizinin yapıldığı düğümün ürünü olan verinin bilgisayardaki adresi görülür.



Şekil 6.32 Clustering Düğümünün Data Sekmesi

Clustering düğümünün Variables sekmesinde ise analize dahil edilecek değişkenler seçilir. Gerekli ise, standardizasyon yöntemi seçilir.



Şekil 6.33 Clustering Düğümünün Variables Sekmesi

Çizelge 6.11 Kümeleme Analizinin Girdi Değişkenleri

Name	Status	Model Role	Measurement	Type	Format	Label
D1	don't use	input	interval	num	BEST12.	d1
D7	don't use	rejected	interval	num	BEST12.	d7
D18	don't use	input	interval	num	BEST12.	d18
D26	don't use	input	interval	num	BEST12.	d26
D28	don't use	input	interval	num	BEST12.	d28
D33	don't use	rejected	interval	num	BEST12.	d33
D34	don't use	rejected	interval	num	BEST12.	d34
D35	don't use	rejected	interval	num	BEST12.	d35
D36	don't use	rejected	interval	num	BEST12.	d36
D37	don't use	rejected	interval	num	BEST12.	d37
D38	don't use	rejected	interval	num	BEST12.	d38
D39	don't use	rejected	nominal	char	\$46	d39

Çizelge 6.11 Devam Kümeleme Analizinin Girdi Değişkenleri

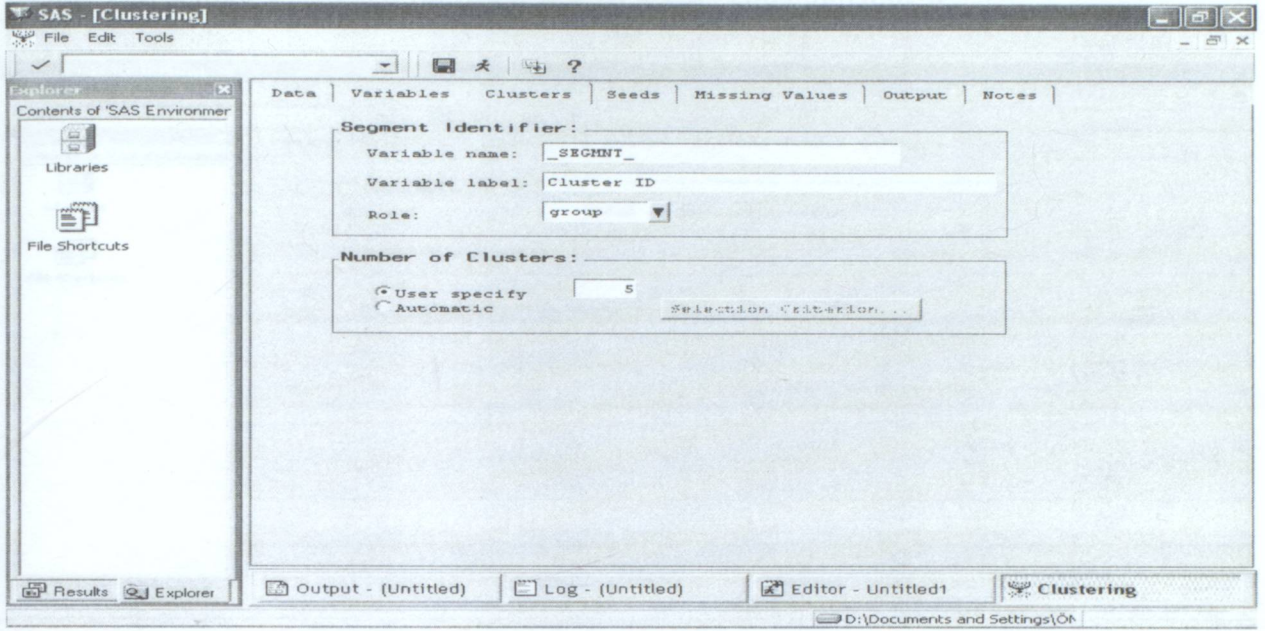
Name	Status	Model Role	Measurement	Type	Format	Label
D40	don't use	rejected	nominal	char	\$5	d40
D41	don't use	rejected	binary	char	\$6	d41
D42	don't use	rejected	nominal	char	\$11	d42
D1_YV6S1	don't use	input	interval	num	BEST12.	d1: Maximize normality
D3_2GLCH	don't use	rejected	interval	num	BEST12.	d3: Maximize normality
D4_NJYBH	don't use	input	interval	num	BEST12.	d4: Maximize normality
D5_YJTO6	don't use	rejected	interval	num	BEST12.	d5: Maximize normality
D6_G57RJ	don't use	input	interval	num	BEST12.	d6: Maximize normality
D8_JNE71	don't use	input	interval	num	BEST12.	d8: Maximize normality
D9_Z3AQC	don't use	rejected	interval	num	BEST12.	d9: Maximize normality
D10_LGB3	don't use	input	interval	num	BEST12.	d10: Maximize normality
D11_8VYW	don't use	rejected	interval	num	BEST12.	d11: Maximize normality
D12_KXSL	don't use	input	interval	num	BEST12.	d12: Maximize normality
D13_DG79	don't use	rejected	interval	num	BEST12.	d13: Maximize normality
D14_YQUQ	don't use	input	interval	num	BEST12.	d14: Maximize normality
D15_ZKJ8	don't use	rejected	interval	num	BEST12.	d15: Maximize normality
D16_F4IO	don't use	rejected	interval	num	BEST12.	d16: Maximize normality
D17_2HTL	don't use	rejected	interval	num	BEST12.	d17: Maximize normality
D19_606L	don't use	rejected	interval	num	BEST12.	d19: Maximize normality
D20_W45K	don't use	input	interval	num	BEST12.	d20: Maximize normality
D21_ZI4D	don't use	rejected	interval	num	BEST12.	d21: Maximize normality
D22_Y9JB	don't use	input	interval	num	BEST12.	d22: Maximize normality
D23_1FDO	don't use	rejected	interval	num	BEST12.	d23: Maximize normality
D24_UN9H	don't use	input	interval	num	BEST12.	d24: Maximize normality

Çizelge 6.11 Devam Kümeleme Analizinin Girdi Değişkenleri

Name	Status	Model Role	Measurement	Type	Format	Label
D25_GCIE	don't use	input	interval	num	BEST12.	d25: Maximize normality
D27_DQ06	don't use	input	interval	num	BEST12.	d27: Maximize normality
D29_AFE0	don't use	input	interval	num	BEST12.	d29: Maximize normality
D30_VMZ5	don't use	input	interval	num	BEST12.	d30: Maximize normality
D31_ARE9	don't use	input	interval	num	BEST12.	d31: Maximize normality
D32_J25H	don't use	input	interval	num	BEST12.	d32: Maximize normality
PRIN1	use	input	interval	num	BEST12.	
PRIN2	use	input	interval	num	BEST12.	
PRIN3	use	input	interval	num	BEST12.	
PRIN4	use	input	interval	num	BEST12.	
PRIN5	use	input	interval	num	BEST12.	
PRIN6	use	input	interval	num	BEST12.	
PRIN7	use	input	interval	num	BEST12.	
PRIN8	use	input	interval	num	BEST12.	
PRIN9	use	input	interval	num	BEST12.	
PRIN10	use	input	interval	num	BEST12.	
PRIN11	don't use	input	interval	num	BEST12.	
PRIN12	don't use	input	interval	num	BEST12.	

Çizelge 6.10'dan da görüldüğü gibi ilk 10 temel bileşen aracılığı ile kümeleme analizi gerçekleştirilmektedir. Bu temel bileşenler veri setindeki toplam değişimin %99,994'ünü açıklamaktadır.

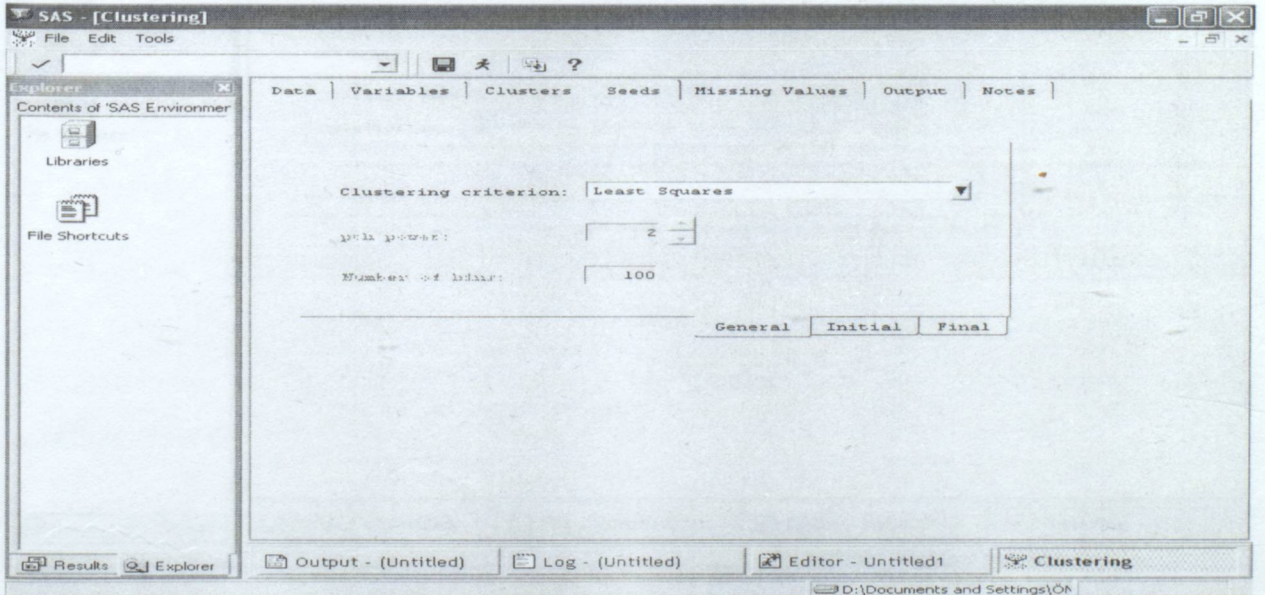
Clustering düğümünün Clusters sekmesinde kaç adet kümenin tercih edileceği belirlenir ya da programın küme sayısını belirlerken kullanacağı yöntem tercih edilir.



Şekil 6.34 Clustering Düğümünün Clusters Sekmesi

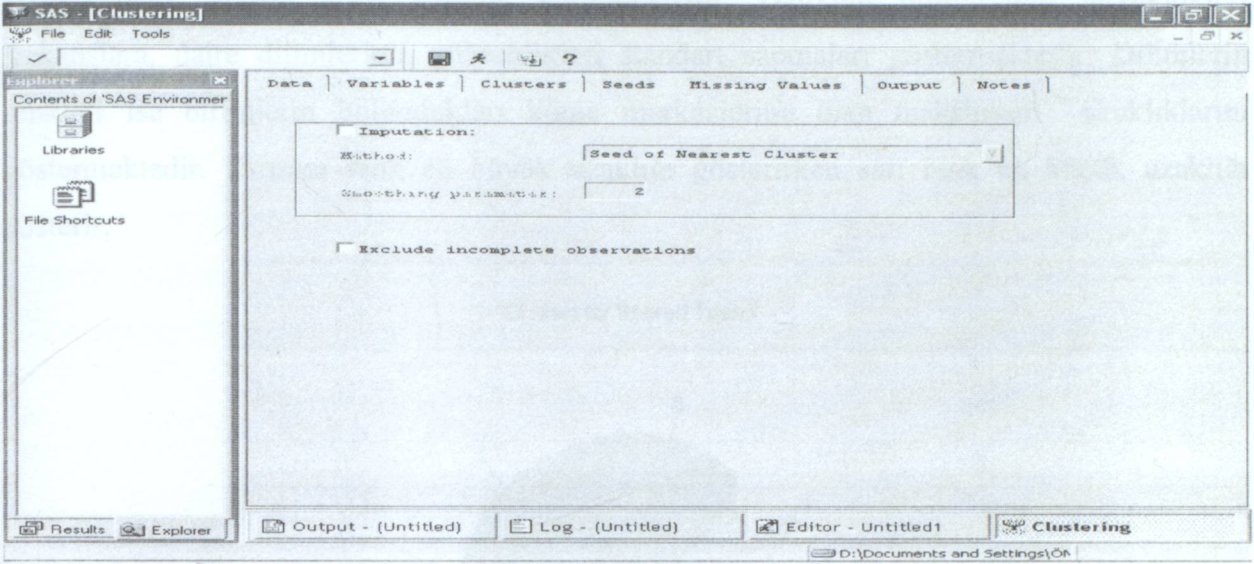
Şekil 6.25’den de görüldüğü gibi küme sayısı 5 olarak belirlenmiştir. Bu seçim farklı sayıdaki kümelerin sonuçlarının incelenmesi sonucunda gerçekleşmiştir.

Clustering düğümünün Seeds sekmesinde uzaklık kriteri (programda kullanılan adı ile clustering criterion) seçilir. Uzaklık kriteri olarak Least Squares Fast tercih edilmiştir.



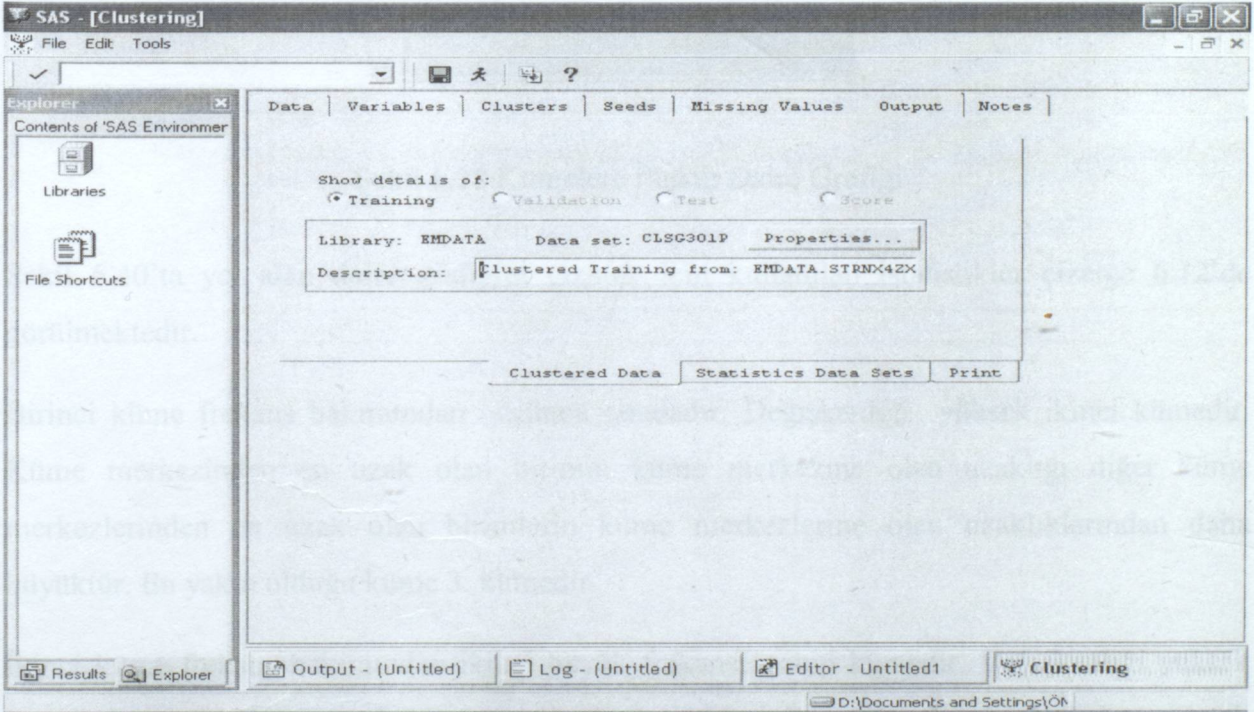
Şekil 6.35 Clustering Düğümünün Seeds Sekmesi

Clustering düğümünün Missing Values sekmesinde kayıp verilere hangi işlemin uygulanacağı seçilir. Veri setinde kayıp veri olmadığı için bu sekme kullanılmamıştır.



Şekil 6.36 Clustering Düğümünün Missing Values Sekmesi

Clusters düğümünün output sekmesinde kümeleme analizi sonucu istenen çıktılar seçilir ve oluşacak veri seti ile ilgili bilgiler görülür.

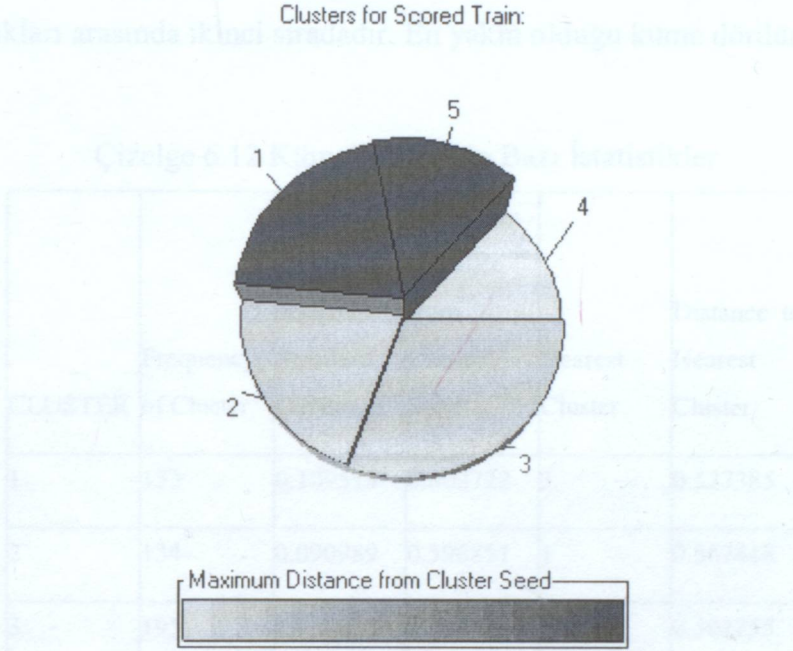


Şekil 6.37 Clustering Düğümünün Output Sekmesi

Clusters düğümünün notes sekmesinde ise kullanıcı notları için ayrılmış bir alan bulunur.

6.2.3.6.1 Kümeleme Analizinin Sonuçları

Kümeleme analizinde oluşturulan beş kümenin frekanslarına ve standart sapmalarına göre çizilen daire grafiği şekil 6.40'da görülmektedir. Dairenin dilimlerinin büyüklükleri frekanslara, daire dilimlerinin yükseklikleri standart sapmaları göstermektedir. Dilimlerin renkleri ise birimlerin bulundukları küme merkezlerine olan maksimum uzaklıklarını göstermektedir. Kırmızı renk en büyük uzaklığı gösterirken sarı renk en küçük uzaklığı gösterir.



Şekil 6.38 Kümelere İlişkin Daire Grafiği

Şekil 6.40'ta yer alan daire grafiğini çizmek için kullanılan istatistikler çizelge 6.12'de görülmektedir.

Birinci küme frekans bakımından üçüncü sıradadır. Değişkenliği yüksek ikinci kümedir. Küme merkezinden en uzak olan birimin küme merkezine olan uzaklığı diğer küme merkezlerinden en uzak olan birimlerin küme merkezlerine olan uzaklıklarından daha büyüktür. En yakın olduğu küme 3. kümedir.

İkinci küme frekans bakımından ikinci büyük frekansa sahip kümedir. Değişkenliği yüksek üçüncü kümedir. Küme merkezinden en uzak olan birimi diğer küme merkezlerinden en uzak olan birimlerin uzaklıkları arasında dördüncü sıradadır. En yakın olduğu küme birinci kümedir.

Üçüncü küme en büyük frekansa sahip kümedir. Değişkenliği diğer kümeler arasında ikinci

en düşük kümedir. Küme merkezinden en uzak olan birimi diğer küme merkezlerinden en uzak olan birimlerin uzaklıkları arasında üçüncü sıradadır. En yakın olduğu küme 4. kümedir.

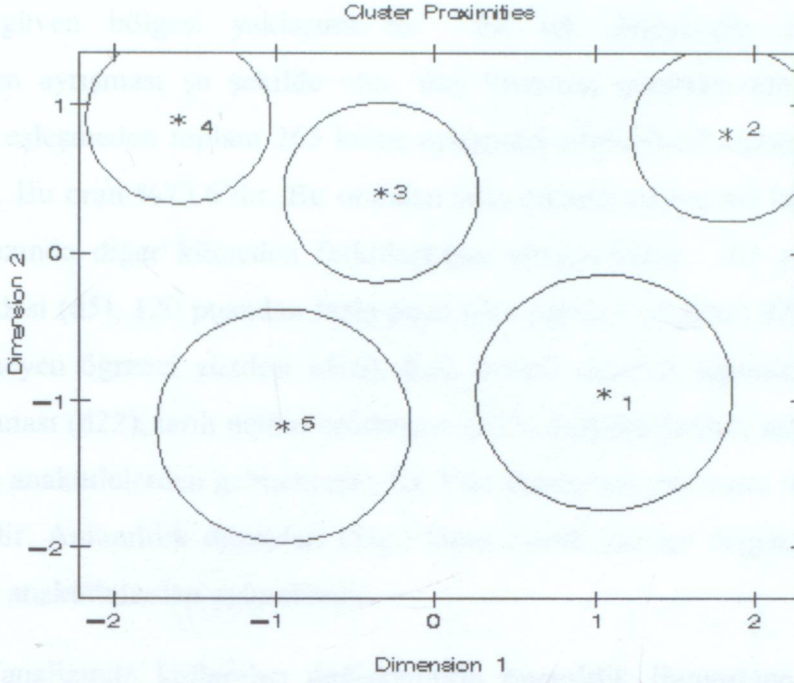
Dördüncü küme frekansı, değişkenliği ve küme merkezinden en uzak olan biriminin küme merkezine olan uzaklığı bakımından en düşük kümedir. En yakın olduğu küme üçüncü kümedir.

Beşinci küme frekans bakımından ikinci düşük frekanslı küme iken değişkenliği en yüksek kümedir. Küme merkezinden en uzak olan birimi diğer küme merkezlerinden en uzak olan birimlerin uzaklıkları arasında ikinci sıradadır. En yakın olduğu küme dördüncü kümedir.

Çizelge 6.12 Kümelere İlişkin Bazı İstatistikler

CLUSTER	Frequency of Cluster	Root-Mean-Square Standard Deviation	Maximum Distance from Cluster Seed	Nearest Cluster	Distance to Nearest Cluster
1	133	0.109519	0.802722	3	0.527385
2	134	0.090989	0.596851	1	0.567418
3	195	0.084834	0.611514	4	0.502255
4	86	0.077676	0.574372	3	0.502255
5	92	0.112881	0.795165	4	0.636573

Kümeleme analizinde oluşturulan kümelerin birbirlerine göre olan konumları şekil 6.41’de görülmektedir. Oluşturulan beş kümenin birbirlerinden tamamen ayrıldıkları şekilde açıkça görülmektedir. Grafik küme merkezlerinin uzaklık matrisi kullanılarak çok boyutlu ölçekleme analizi ile oluşturulur. Küme sınırlarını belirleyen çemberler ise küme merkezine en uzak olan birimlerin uzaklıkları kullanılarak çizilir.



Şekil 6.39 Kümelerin Birbirlerine Göre Konumları

Çok boyutlu ölçekleme analizinde kullanılan küme uzaklıkları matrisi aşağıdaki çizelgede görülmektedir.

Çizelge 6.13 Küme Uzaklıkları Matrisi

CLUSTER	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
1	0	1.937188	1.956742	3.249214	2.041455
2	1.937188	0	2.233136	3.431275	3.411159
3	1.956742	2.233136	0	1.398873	1.752621
4	3.249214	3.431275	1.398873	0	2.186306
5	2.041455	3.411159	1.752621	2.186306	0

Kümelerin tüm değişkenler bazında birbirlerinden ayrıldıkları şekil 6.41'de görüldü. Bonferoni güven bölgesi yaklaşımı ile tek tek değişkenler incelenirse kümelerin birbirlerinden ayrışması şu şekilde olur. Beş kümenin mümkün tüm değişkenler bazında toplam 360 eşleşmeden toplam 265 küme eşleşmesi istatistiksel olarak farklı anakütlelerden gelmektedir. Bu oran %73.6'dır. Bu orandan yola çıkarak da her bir kümenin en azından bir değişken bazında diğer kümeden farklılaştığını söyleyebiliriz. 105 puandan az puan alan öğrenci yüzdesi (d5), 120 puandan fazla puan alan öğrenci yüzdesi (d9), herhangi bir bölüme yerleştirilemeyen öğrenci yüzdesi (d16), fizik netleri standart sapması (d20), kimya netleri standart sapması (d22), tarih netleri ortalaması (d27) değişkenlerinde tüm kümeler istatistiksel olarak farklı anakütlelerden gelmektedir. Ek 3'de Bonferoni yaklaşımı ile elde edilen sonuçlar görülmektedir. Anlamlılık düzeyleri (Sig.) kalın olarak yazılan değişken çiftleri istatistiksel olarak farklı anakütlelerden gelmektedir.

Kümeleme analizinde kullanılan değişkenlerin önemlilik (importance) katsayıları çizelge 6.14'te görülmektedir. Önemlilik katsayısı 0 ile 1 arasında değerler alan kümeleme analizinde değişkenin kümeleri ne kadar iyi böldüklerini gösteren bir ölçüdür. nominal ve aralık-oransal ölçekli değişkenler için önemlilik katsayısının formülleri aşağıdaki gibidir.

Nominal ölçekli değişkenler için formül :

$$LOGWORTH = -\log((chi - square) * (p - value))$$

Chi-square değeri chi-square bağımsızlık testinden elde edilir.

Aralık-Oransal ölçekli değişkenler için formül

$$LOGWORTH = -\log((F - test) * (p - value))$$

F-test değeri ANOVA testinden elde edilir.

Altıncı ve onuncu temel bileşen, kümeleme analizinde hiçbir rol oynamazken birinci temel bileşen en önemli değişken olmuştur.

Çizelge 6.14 Kümeleme Analizinde Kullanılan Değişkenlerin Önemlilik Katsayıları

Name	Importance	Measurement	Type
PRIN1	1.000	interval	num
PRIN2	0.640	interval	num
PRIN3	0.408	interval	num
PRIN4	0.162	interval	num
PRIN5	0.129	interval	num
PRIN6	0.000	interval	num
PRIN7	0.190	interval	num
PRIN8	0.266	interval	num
PRIN9	0.131	interval	num
PRIN10	0.000	interval	num

Kümeleme analizi sonucunda oluşturulan kümeler aracılığı ile SAS programının oluşturduğu karar ağacı sonucunda kümelenme için kurallar elde edilmektedir. Bu kurallar çizelge 6.15’de görülmektedir. Çizelge 6.9 ve 6.10’da yer alan özdeğerler ile analize dahil olan değişkenlerin gerekli dönüşümler yapıldıktan sonra çarpılması ile elde edilen skor değerlerine göre yeni birimlerin hangi kümeye dahil olacaklarını tahmin edebiliriz. Esenler İbrahim Turhan Lisesi için bir örnek aşağıdaki çizelgede görülmektedir.

Çizelge 6.15 İbrahim Turhan Lisesi Verileri İçin Temel Bileşen Skorlarının Hesaplanması

	veri	prin1	prin2	prin7	prin9	veri*prin1	veri*prin2	veri*prin7	veri*prin9
d7	0.3643	-0.254	-0.292	-0.45	-0.03	-0.09254	-0.1064	-0.16248	-0.01202
D3_2GLCH	5.5947	-0.053	0.548	-0.08	-0.02	-0.29652	3.0659	-0.45317	-0.1063
D5_YJTO6	0.2718	-0.326	0.191	-0.08	0.21	-0.08859	0.05191	-0.02065	0.05598
D9_Z3AQC	0.2802	0.347	-0.067	0.283	-0.05	0.097236	-0.0188	0.0793	-0.01261
D11_8VYW	0.0184	-0.04	-0.647	0.152	0	-0.00074	-0.0119	0.0028	3.7E-05
D13_DG79	0.0786	0.346	0.083	-0.58	0.01	0.0272	0.00652	-0.04567	0.00055
D15_ZKJ8	0.0436	-0.116	-0.196	0.183	-0.01	-0.00506	-0.0086	0.00799	-0.00031
D16_F4IO	2.3514	-0.312	0.287	0.487	0.01	-0.73364	0.67486	1.14514	0.01176
D17_2HTL	1.8946	0.346	-0.061	0.188	0.73	0.655537	-0.1156	0.35619	1.38118
D19_606L	0.8879	0.346	0.015	0.143	-0.25	0.30721	0.01332	0.12697	-0.22286
D21_ZI4D	0.7608	0.351	0.064	0.13	-0.53	0.267043	0.04869	0.0989	-0.40399
D23_1FDO	0.5068	0.333	0.136	-0.06	0.28	0.16877	0.06893	-0.03142	0.14191
Skor= $\sum veri * prini$						0.305907	3.66893	1.10389	0.83332

Çizelge 6.15’de elde edilen skorlara göre çizelge 6.16’da yer alan kümeleme kuralları incelenirse Esenler İbrahim Turhan Lisesi’nin 36 numaralı düğümde yer alacaktır. Buna göre İbrahim Turhan Lisesi %97.7 olasılıkla üçüncü kümede yer alacaktır. Diğer değişkenlere ilişkin hesaplanan skorlar tabloya sığmayacağından gösterilmemiştir.

Çizelge 6.16 Kümeleme Kuralları

```

IF 2.647721 <= PRIN2
AND 1.869193 <= PRIN1
THEN
  NODE : 7
  N : 116
  1 : 0.0%
  2 : 100.0%
  3 : 0.0%
  4 : 0.0%
  5 : 0.0%

IF 1.869193 <= PRIN1 < 2.901241
AND PRIN2 < 2.647721
THEN
  NODE : 12
  N : 43
  1 : 88.4%
  2 : 11.6%
  3 : 0.0%
  4 : 0.0%
  5 : 0.0%

IF 2.901241 <= PRIN1
AND PRIN2 < 2.647721
THEN
  NODE : 13
  N : 10
  1 : 10.0%
  2 : 90.0%
  3 : 0.0%
  4 : 0.0%
  5 : 0.0%

IF PRIN2 < 3.338737
AND PRIN3 < 1.548415
AND PRIN1 < 0.166123
THEN
  NODE : 14
  N : 67
  1 : 0.0%
  2 : 0.0%
  3 : 0.0%
  4 : 0.0%
  5 : 100.0%

IF 3.338737 <= PRIN2
AND PRIN3 < 1.548415
AND PRIN1 < 0.166123
THEN
  NODE : 15
  N : 1
  1 : 0.0%
  2 : 0.0%
  3 : 0.0%
  4 : 100.0%
  5 : 0.0%

IF 0.023393 <= PRIN1 < 0.166123
AND 1.548415 <= PRIN3
THEN
  NODE : 17
  N : 9
  1 : 0.0%
  2 : 0.0%
  3 : 88.9%
  4 : 11.1%
  5 : 0.0%

```


Çizelge 6.16 Devam Kümeleme Kuralları

IF PRIN9 < 0.743718
 AND 2.521191 <= PRIN2
 AND 0.166123 <= PRIN1 < 1.869193
 THEN

NODE : 20
 N : 4
 1 : 0.0%
 2 : 0.0%
 3 : 25.0%
 4 : 0.0%
 5 : 75.0%

IF PRIN2 < 3.191231
 AND PRIN1 < 0.023393
 AND 1.548415 <= PRIN3

THEN
 NODE : 22
 N : 3
 1 : 0.0%
 2 : 0.0%
 3 : 100.0%
 4 : 0.0%
 5 : 0.0%

IF 3.191231 <= PRIN2
 AND PRIN1 < 0.023393
 AND 1.548415 <= PRIN3

THEN
 NODE : 23
 N : 81
 1 : 0.0%
 2 : 0.0%
 3 : 0.0%
 4 : 100.0%
 5 : 0.0%

IF PRIN5 < 1.94148
 AND PRIN8 < 0.739247
 AND PRIN2 < 2.521191
 AND 0.166123 <= PRIN1 < 1.869193

THEN
 NODE : 24
 N : 20
 1 : 0.0%
 2 : 0.0%
 3 : 0.0%
 4 : 0.0%
 5 : 100.0%

IF 1.94148 <= PRIN5
 AND PRIN8 < 0.739247
 AND PRIN2 < 2.521191
 AND 0.166123 <= PRIN1 < 1.869193

THEN
 NODE : 25
 N : 6
 1 : 83.3%
 2 : 0.0%
 3 : 0.0%
 4 : 0.0%
 5 : 16.7%

IF 0.166123 <= PRIN1 < 0.47459
 AND 0.739247 <= PRIN8
 AND PRIN2 < 2.521191

THEN
 NODE : 26
 N : 1
 1 : 0.0%
 2 : 0.0%
 3 : 0.0%
 4 : 0.0%
 5 : 100.0%

Çizelge 6.16 Devam Kümeleme Kuralları

IF 1.389379 <= PRIN7
 AND 0.743718 <= PRIN9
 AND 2.521191 <= PRIN2
 AND 0.166123 <= PRIN1 < 1.869193

THEN

NODE : 29

N : 13

1 : 38.5%

2 : 30.8%

3 : 30.8%

4 : 0.0%

5 : 0.0%

IF PRIN4 < 2.569825
 AND 0.47459 <= PRIN1 < 1.869193
 AND 0.739247 <= PRIN8
 AND PRIN2 < 2.521191

THEN

NODE : 30

N : 76

1 : 98.7%

2 : 0.0%

3 : 1.3%

4 : 0.0%

5 : 0.0%

IF 2.569825 <= PRIN4
 AND 0.47459 <= PRIN1 < 1.869193
 AND 0.739247 <= PRIN8
 AND PRIN2 < 2.521191

THEN

NODE : 31

N : 11

1 : 36.4%

2 : 0.0%

3 : 63.6%

4 : 0.0%

5 : 0.0%

IF 1.841677 <= PRIN1 < 1.869193
 AND PRIN7 < 1.389379
 AND 0.743718 <= PRIN9
 AND 2.521191 <= PRIN2

THEN

NODE : 33

N : 3

1 : 100.0%

2 : 0.0%

3 : 0.0%

4 : 0.0%

5 : 0.0%

IF 0.743718 <= PRIN9 < 1.24123
 AND 0.166123 <= PRIN1 < 1.841677
 AND PRIN7 < 1.389379
 AND 2.521191 <= PRIN2

THEN

NODE : 36

N : 175

1 : 0.6%

2 : 0.0%

3 : 97.7%

4 : 1.7%

5 : 0.0%

IF 1.24123 <= PRIN9
 AND 0.166123 <= PRIN1 < 1.841677
 AND PRIN7 < 1.389379
 AND 2.521191 <= PRIN2

THEN

NODE : 37

N : 1

1 : 100.0%

2 : 0.0%

3 : 0.0%

4 : 0.0%

5 : 0.0%

6.2.3.6.2 Küme Profilleri

Kümeleme analizi sonucunda oluşturulan beş kümenin analitik olarak birbirlerinden ayrıştıkları şekil 6.41’de ve Ek 3’de görülmektedir. Kümeleme analizini başarı sayabilmemiz için oluşturulan kümelerin matematik, fizik, kimya, biyoloji, türkçe, tarih, coğrafya, felsefe, ingilizce, almanca, fransızca sorularındaki netler ve/veya yüksek öğretim kurumlarına öğrenci yerleştirme oranları açısından birbirlerinden farklılaşmaları gerekir. Yabancı dil sınavına giren öğrenci sayıları farklılık göstermektedir. Bu yüzden yabancı dil testlerinde yapılan netlere ilişkin değişkenler kümeleme analizine girdi sağlayan temel bileşenler analizi dışında bırakılmışlardır.

Bu farklılaşmayı incelemek için her bir kümede bulunan liselerin 2001 ÖSS sınavında matematik, fizik, kimya, biyoloji, türkçe, tarih, coğrafya, felsefe, ingilizce, almanca, fransızca sorularındaki ortalama netlerini içeren çizelge 6.17 hazırlanmıştır. Çizelge, üstte incelenen değişken bazında küme ortalamalarının bulunduğu satır ve hemen altında bu ortalamalara göre kümelerin kendi içlerindeki sıralarını gösteren satırlardan oluşmaktadır. Kolonlar kümeleri göstermektedir. En sondaki kolonda ise değişkene ilişkin genel ortalama bulunmaktadır.

En altta bulunan skor satırı değişkenlerin her bir değişkende küme içinde almış oldukları sıralar toplanarak elde edilmiştir. Bu satırdaki en küçük toplam maksimum başarıyı, en büyük toplam minimum başarıyı göstermektedir. Genel sıra satırı da skor satırındaki kümelerin almış oldukları değerlere göre oluşturulmuştur.

Çizelge 6.18’de ise kümelerde bulunan liselerin sınava giren öğrenci sayıları, yüksek öğretim kurumlarında okuma hakkı kazanan öğrenci yüzdeleri ve 105 puandan az puan alan, 105-120 puan arası puan alan, 120 puandan fazla puan alan öğrenci yüzdelerinin ortalamaları yer almaktadır. Yine her bir kümenin bu değişkenler bazında sıra değerleri hesaplanmıştır.

Çizelge 6.17 ve 6.18’den yola çıkarak kümeler incelenecek. Ek 5’de yer alan ÖSYM’nin tanımladığı okul türlerine (d39 değişkeni), Ek 6’da yer alan d40 değişkenindeki okul türlerine, Ek 7’de yer alan d41 değişkenindeki okul türlerine, Ek 8’de yer alan d42 değişkenindeki okul türlerine göre küme profilleri yorumlanacaktır.

Çizelge 6.17 Kümelerin Testlerdeki Başarı Ortalamaları

	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
d17	12.7404	24.7393	6.2853	1.1285	1.482	10.1071
Sıra	2	1	3	5	4	
d19	3.5486	7.4947	1.5023	0.3316	0.1887	2.8361
Sıra	2	1	3	4	5	
d21	2.27	5.4691	0.8436	0.0598	0.0277	1.8875
Sıra	2	1	3	4	5	
d23	0.7824	2.2965	0.3423	0.0986	0.093	0.7743
Sıra	2	1	3	4	5	
d25	24.6071	33.3375	20.7244	14.7686	19.0257	23.1276
Sıra	2	1	3	5	4	
d27	7.3802	10.7803	6.4795	4.4794	5.694	7.1855
Sıra	2	1	3	5	4	
d29	4.3132	6.7137	3.9045	2.913	3.6484	4.4076
Sıra	2	1	3	5	4	
d31	3.2144	4.824	2.811	1.7173	2.7943	3.167
Sıra	2	1	3	5	4	
d33	9.2015	30.0437	6.0693	0.3876	1.8164	10.365
Sıra	2	1	3	5	4	
d35	0	3.6377	0	0	0	0.7616
Sıra	2	1	2	2	2	
d37	0	2.6807	0	0.5749	0	0.6385
Sıra	3	1	3	2	3	
Skor	23	11	32	46	44	
Genel Sıra	2	1	3	5	4	

Buradan elde edilen sonuçlar şunlardır:

- İkinci küme tüm test türlerinde en başarılı kümedir.
- Birinci küme almanca testi haricinde tüm test türlerinde ikinci en başarılı kümedir.
- Üçüncü küme Fransızca testi haricinde tüm test türlerinde üçüncü en başarılı kümedir.
- Dördüncü ve beşinci kümelerde bulunan liselerin öğrencileri tüm test türlerinde yakın netler çıkarmakla birlikte değişik test türlerinde dördüncülük ve beşinciliği paylaşmaktadırlar.
- Genel sıra dikkate alınırsa beşinci kümede bulunan liselerin öğrencileri başarı açısından dördüncü, dördüncü kümede bulunan liselerin öğrencileri ise başarı açısından beşinci sırada yer almaktadır.
- Kümeleme analizi sonucunda oluşturulan beş kümeyi öğrencileri il ortalamasının altında net yapan kümeler ve il ortalamasının üzerinde net yapan kümeler olarak iki gruba ayırmak istersek birinci ve ikinci kümeler öğrencileri il ortalamasının üzerinde net yapan diğer kümeler ise öğrencileri il ortalamasının altında net yapan kümeler olur.
- İkinci kümede bulunan liselerin öğrencileri türkçe testi haricinde kendisine en yakın net yapan birinci kümede bulunan liselerin öğrencilerinden genelde yaklaşık iki katı kadar fazla net yapmışlardır.
- İkinci kümede bulunan liselerin öğrencileri ile üçüncü kümede bulunan liselerin öğrencileri arasında da ikinci küme lehine benzer bir durum söz konusudur.
- Dördüncü ve beşinci kümelerde bulunan liselerin öğrencileri sayısal testlerde ortalama sıfır net civarında net yaparken sözel testlerde ise ortalama net sayısı türkçe testi haricinde 6 neti geçmemektedir.
- Dil netleri bakımından dördüncü kümede bulunan liselerin öğrencilerinin almanca testinde ikinci sırada oldukları görülmektedir. Buradaki başarı Almanya'dan gelen ve liseyi açıköğretim lisesinde bitirip dil sınavına giren öğrencilere bağlanabilir.

Kümeleri yalnızca testlerdeki başarı ortalamalarına göre değerlendirmek hatalı sonuçlara yol açabilir. Bu kümelerin öğrencilerini lisans, ön lisans ve açık öğretim programlarına yerleştirme yüzdelerini de incelemek gerekir. Çizelge 6.89’da her bir kümenin ÖSS sınavına giren öğrenci sayısı, 105 puandan az puan alan öğrenci yüzdesi, 105-120 puan arası puan alan öğrenci yüzdesi, 120 puandan fazla puan alan öğrenci yüzdesi, lisans programlarına yerleştirilen öğrenci yüzdesi, ön lisans programlarına yerleştirilen öğrenci yüzdesi ve açıköğretim programlarına yerleştirilen öğrenci yüzdesi değişkenleri için ortalama değerleri yer almaktadır.

Çizelge 6.18 Kümelerin Yüksek Öğretim Kurumlarına Öğrenci Yerleştirme Oranları

	1	2	3	4	5	Toplam
d3	25,6692	115,6194	142,2051	328,3721	38,7065	122,5594
	5	3	2	1	4	
d5	0,1571	1,21E-02	0,3149	0,6468	0,4976	0,2895
	4	5	3	1	2	
d7	0,2648	6,34E-02	0,34	0,266	0,281	0,248
	4	5	1	3	2	
d9	0,5781	0,9246	0,3451	8,72E-02	0,2215	0,4624
	2	1	3	5	4	
d11	0,1252	3,73E-02	6,72E-02	6,38E-02	0,1213	8,03E-02
	1	5	3	4	2	
d13	0,2668	0,5848	8,80E-02	5,45E-03	4,23E-02	0,2115
	2	1	3	5	4	
d15	5,06E-02	1,65E-02	4,89E-02	6,25E-02	0,1028	5,20E-02
	3	5	4	2	1	

Çizelge 6.89’dan elde edilen sonuçlar şunlardır:

- Dördüncü kümede bulunan liselerde ortalama 328 öğrenci ÖSS sınavına girmiştir. Bu ortalama değerle dördüncü küme ÖSS sınavına katılan öğrenci ortalaması bakımından birinci sırada yer almaktadır.
- Dördüncü kümede bulunan liselerde 105 puandan daha az puan alan öğrenci ortalaması %64,6’dır. Dördüncü küme alınan puan bakımından en başarısız kümedir.
- Dördüncü kümede bulunan liselerde sınava giren öğrencilerin %26,6’sı 105-120 puan arasında puan alırken bu kümede bulunan liselerde sınava giren öğrencilerin %8,7’si 120 puandan fazla puan almıştır.

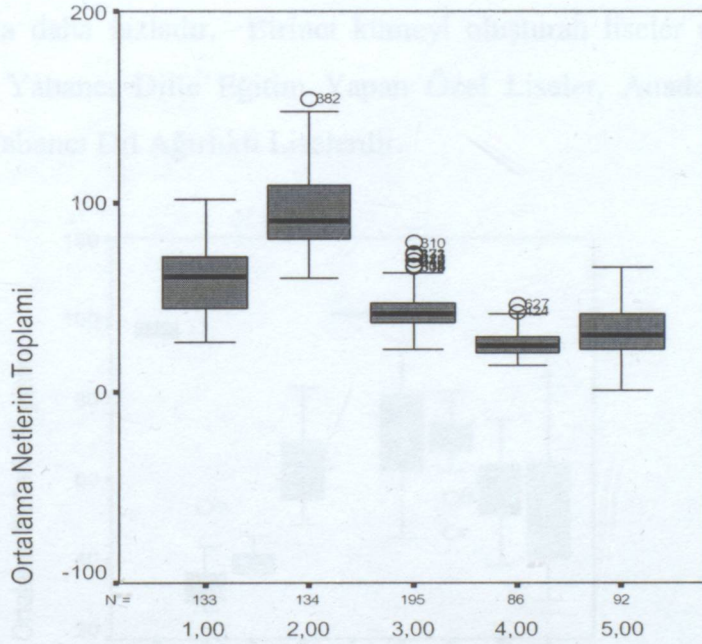
- Dördüncü kümede bulunan liselerde sınava giren öğrencilerin yaklaşık %6'sı ön lisans ve açık öğretim programlarına yerleşirken yaklaşık binde beşi lisans programlarına yerleştirilmiştir.
- Beşinci kümede bulunan liselerde ÖSS sınavına giren öğrenci sayısı ortalaması diğer kümeler arasında ikinci sırada yer alırken bu öğrencilerin yarısının 105 puan altında puan aldığı görülmektedir.
- Beşinci kümede bulunan liselerin öğrencilerinin ortalama %28'i 105-120 puan arasında puan alırken 120 puanı geçen öğrenci oranı %22'dir.
- Beşinci kümede bulunan liselerin öğrencilerinin ortalama %12'si ön lisans, ortalama %10'u açık öğretim programlarına yerleşirken lisans programlarına yerleştirilen öğrenci yüzdesi %4 dolaylarındadır.
- Üçüncü kümede bulunan liselerde ortalama 142 öğrenci ÖSS sınavına girerken bu öğrencilerden %31,5'i 105 puandan az, %34'ü 105-120 puan arası, %34,5'i 120 puandan fazla puan almıştır.
- Üçüncü kümede bulunan liselerin öğrencilerinin %6,7'si ön lisans, %8,8'i ön lisans, %4,9'u açıköğretim programlarına yerleştirilmiştir.
- İkinci kümede bulunan liselerde ortalama 116 öğrenci sınava girerken bu öğrencilerden %3,7'si ön lisans, %58,4'ü lisans, %1.65'i açıköğretim programlarına yerleştirilmiştir. İkinci küme yüksek öğretim programına öğrenci yerleştirme bakımından en başarılı kümedir.
- İkinci kümede bulunan liselerin öğrencilerinin %92'si 120 puan barajını aşmıştır.
- Birinci kümede bulunan liselerde ortalama 26 öğrenci ÖSS sınavına girerken bu öğrencilerden %57'si 120 puan barajını aşmıştır.
- Birinci kümede bulunan liselerin öğrencilerinin %12,5'i ön lisans, %26,7'si lisans, %5'i açıköğretim programlarına yerleştirilmiştir.
- Yerleştirilen öğrenci yüzdeleri açısından birinci ve ikinci küme ile diğer kümeler arasında gözle görülür bir fark söz konusudur.

Çizelge 6.17 ve 6.18'den elde ettiğimiz sonuçlara göre hem testlerdeki ortalama net sayısı hem de lisans programlarına yerleştirilen öğrenci yüzdesi bakımından kümeleri sıralar isek ikinci küme en başarılı küme, birinci küme ikinci başarılı küme ve üçüncü kümede orta derecede başarılı küme olarak nitelendirilebilir. Dördüncü ve beşinci kümelerde bulunan liselerin öğrencileri testlerde yaptıkları net bakımından başarısızdırlar. İki küme ortalama netler açısından benzer özellikler gösterse de beşinci kümenin ön lisans, lisans ve açık

öğretim programlarına öğrenci yerleştirme oranlarına ve puan dilimlerine göre öğrencilerin dağılmasına göre bariz üstünlüğünün olduğu görülmektedir. Bu farkın kaynağı ise orta öğretim başarı puanına bağlanabilir.

Yukarıda elde edilen enformasyondan yola çıkarak kümeleri isimlendirirsek, birinci kümeyi yüksek derecede performans gösteren liseler grubu, ikinci kümeyi çok yüksek performanslı liseler grubu, üçüncü kümeyi vasat liseler grubu, dördüncü kümeyi çok başarısız liseler grubu (ortalama netler ve yerleştirme oranları bakımından başarısız liseler), beşinci kümeyi başarısız liseler grubu (ortalama netler bakımından başarısız ortaöğretim başarı puanı avantajlı liseler) olarak adlandırabiliriz.

Her bir lisenin ÖSS sınavında yapmış oldukları testlerdeki netlerin ortalama değerlerinin toplamını alarak oluşturduğumuz değişkene göre liselerin dağılımı şekil 6.40'da görülmektedir*. Yabancı dil sınavında yapılan ortalama netler toplam dışarısında tutulmuştur. Bunun nedeni yabancı dil sınavına giren öğrenci sayısı ile esas sınava giren öğrenci sayılarının farklı olmasıdır.



Şekil 6.40 Kümelerde Bulunan Liselerin Ortalama Netlerin Toplam Değeri Bakımından Dağılımları

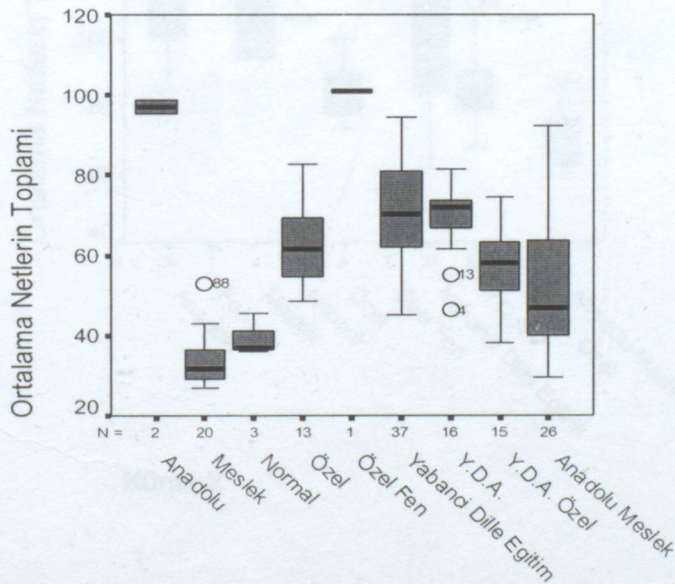
* D39, D40, D41, D42 değişkenlerinde tanımlanan lise türlerinin kümelerdeki testler bazındaki ortalama başarıları EK 9, 10, 11, 12'de, bu değişkenlerdeki lise türlerinin kümelerle göre dağılımıysa EK 5,6,7,8'de görülebilir.

Şekil 6.40'da aşırı değerler göz ardı edildiğinde birinci kümenin dağılımının sola çarpık, üçüncü ve dördüncü kümenin dağılımlarının normale yakın, ikinci ve beşinci kümelerin dağılımlarının ise sağa çarpık oldukları görülmektedir. Kümelerde ortalama netlerin toplamı bakımından aşırı değer olarak görülen liselerin hepsinin kümede bulunan liselerden daha başarılı olduklarını görürüz. İkinci kümede başarısı ile dikkati çeken lise İstanbul Atatürk Fen Lisesi'dir. Üçüncü kümede bulunan liselere göre başarısı ile dikkat çeken liseler ise Özel Marmara Lisesi, Y.D.A. Gaziosmanpaşa Plevne Lisesi, Özel Acıbadem Lisesi ve Özel Semiha Şakir Lisesi'dir. Dördüncü kümede bulunan liselere göre başarısı ile dikkat çeken liseler ise İstanbul Beşiktaş Anadolu Otelcilik ve Turizm Meslek Lisesi ve Bakırköy İmam Hatip Lisesi'dir.

Kümelerde Bulunan Liselerin D42 Değişkeni bazında Ortalama Netlerin Toplamı

Değişkenine Göre Dağılımları

Kümelerin ortalama netlerin toplamından elde edilen değişken bazında d42 değişkenine göre birinci kümedeki liselerin dağılımı şekil 6.41'de görülmektedir. Yüksek performans gösteren bu lise grubu öğrencilerini sınavla alan liselerden oluşmaktadır. Meslek liselerinin yoğunluğu ikinci kümeye oranla daha fazladır. Birinci kümeyi oluşturan liseler arasında frekansı en yüksek olan liseler Yabancı Dille Eğitim Yapan Özel Liseler, Anadolu Meslek Liseleri, Meslek Liseleri ve Yabancı Dil Ağırlıklı Liselerdir.

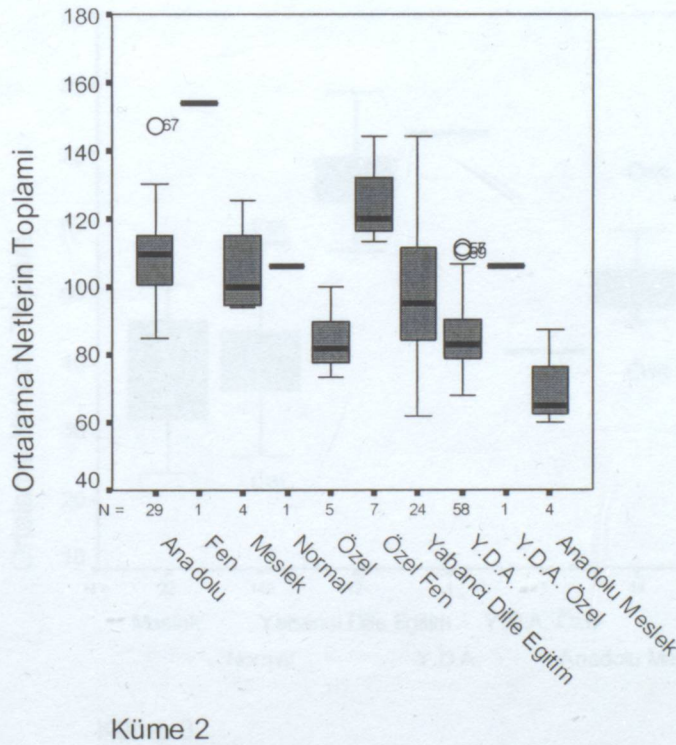


Küme 1

Şekil 6.41 Ortalama Netlerin Toplamına Göre Birinci Kümedeki Liselerin D42 Değişkeni Bazında Dağılımı

Yabancı dil ağırlıklı liseler ile meslek liseleri dışındaki lise türlerinde başarı ya da başarısızlık olarak dikkati çeken lise yoktur. Meslek liselerinde Askeri Lise başarısı ile Y.D.A. liselerde ise Y.D.A. Eyüp Alibeyköy Lisesi ile Y.D.A. Tuzla Tuğrul Bey Lisesi düşük ortalama net toplamı ile aşırı değer olarak görülmektedir. Anadolu liselerinin homojen olarak dağıldıkları diğer lise türlerinin ise dağılımlarının sola ya da sağa çarpık oldukları dikkati çekmektedir. Ortalama etrafındaki değişkenliği en fazla olan lise türleri Anadolu Meslek Lisesi, Yabancı Dille Eğitim Yapan Özel Lise ve Özel Lise türleridir. Bu lise türlerinin eğitimlerinin homojen olmadıkları dikkat çekmektedir. Normal liselerin ise oldukça fazla sola çarpık olduğu açıktır. Kümedeki en başarılı lise türü ise Özel Fen Lisesidir.

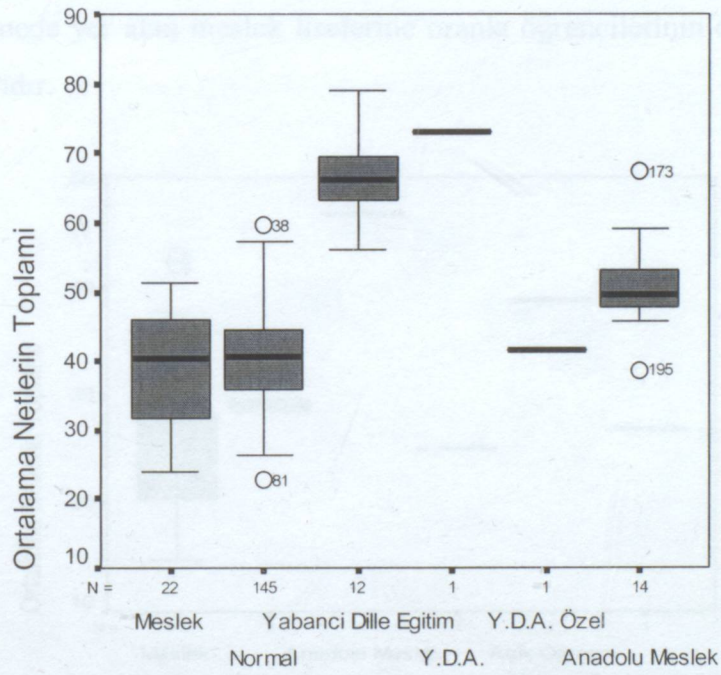
Kümelerin ortalama netlerin toplamından elde edilen değişken bazında d42 değişkenine göre ikinci kümedeki liselerin dağılımı şekil 6.42’de görülmektedir. Bu kümenin büyük çoğunluğunu Yabancı Dil Ağırlıklı Liseler, Anadolu Liseleri , Yabancı Dille Eğitim Yapan Özel Liseler, Fen Liseleri (Özel-Devlet) oluşturmaktadır. İkinci kümedeki liseler sınavla ya da not ortalaması başarısına göre öğrenci alan liselerdir.



Şekil 6.42 Ortalama Netlerin Toplamına Göre İkinci Kümedeki Liselerin D42 Değişkeni Bazında Dağılımı

En başarılı küme olan ikinci kümede bulunan Anadolu Liseleri ile Y.D.A. Liselerin dışında kalan türlerdeki liselerin aşırı değer olarak başarı veya başarısızlıkları ile göz önüne çıkmadıkları görülmektedir. Anadolu Liseleri arasında başarısı ile dikkat çeken lise İstanbul Lisesi, Y.D.A. Liseler arasında başarıları ile dikkat çeken liseler ise Y.D.A. İstanbul Kabataş Erkek Lisesi ve Y.D.A. Sakıp Sabancı Lisesi'dir. Bu kümedeki lise türlerinde de öğretimin homojen olmadığı görülmektedir. Lise türlerine göre dağılımların sağa ya da sol çarpık oldukları dikkat çekmektedir. Meslek Liseleri ile Yabancı Dille Eğitim Yapan Özel Liselerin ortalama etrafındaki değişimleri oldukça geniştir. En başarılı lise türleri Fen Lisesi ve Özel Fen Liseleridir.

Kümelerin ortalama netlerin toplamından elde edilen değişken bazında d42 değişkenine göre üçüncü kümedeki liselerin dağılımı şekil 6.43'de görülmektedir. Bu kümeye normal liseler grubu diyebiliriz. Normal liselerin yanı sıra Meslek Liseleri (Meslek-Anadolu Meslek) üçüncü kümedeki liseler genelde sınavsız öğrenci alan liselerdir. Başka bir deyişle bu liseler sınav kazanamayan öğrencilerin okuduğu liselerdir.

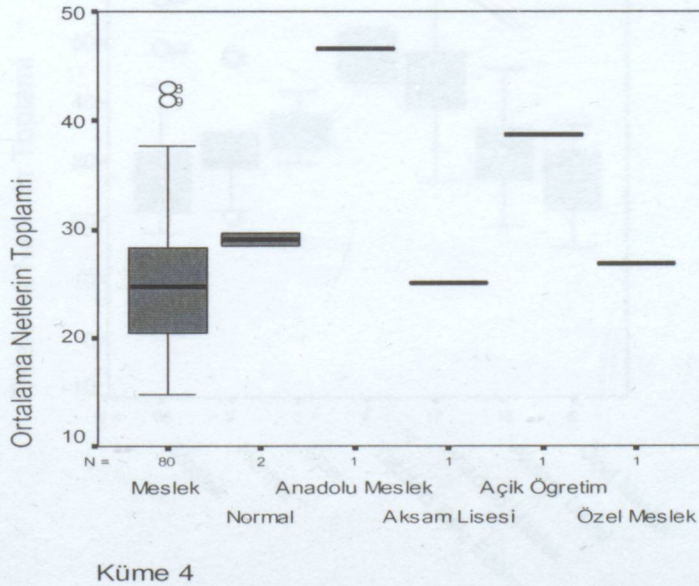


Küme 3

Şekil 6.43 Ortalama Netlerin Toplamına Göre Üçüncü Kümedeki Liselerin D42 Değişkeni Bazında Dağılımı

Üçüncü kümedeki lise türleri içinde başarı ya da başarısızlıkları ile aşırı değer olarak göze batan liselerin bulunduğu lise türleri Normal Liseler ve Anadolu Meslek Liseleridir. Normal liseler içinde yüksek ortalama net sayısı ile Ümraniye Nevzat Ayaz Lisesi, düşük ortalama net sayısı ile de Kartal Samandıra Lisesi dikkat çekmektedir. Anadolu Meslek Liseleri içinde yüksek ortalama net sayısı ile İstanbul Ticaret Odası Anadolu Ticaret Meslek Lisesi, düşük ortalama net sayısı ile de İstanbul Denizcilik ve Su Ürünleri Anadolu Meslek Lisesi dikkat çekmektedir. Normal Liselerin ve Yabancı Dille Eğitim Yapan Özel Liselerin dağılımlarının normale yakın olduğu görülmektedir. Bu küme içerisinde bulunan lise türlerinin de dağılımlarının homojen olmadığı görülmektedir. Fakat Meslek Liseleri ile Normal Liselerin kendi aralarında homojen dağıldıkları görülmektedir. Bu kümedeki en başarılı lise türü Yabancı Dil Ağırlıklı Lisedir.

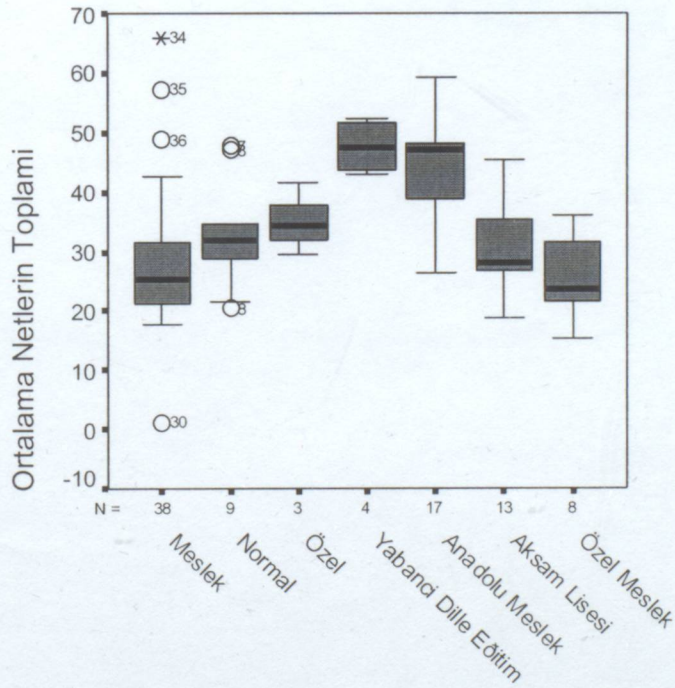
Kümelerin ortalama netlerin toplamından elde edilen değişken bazında d42 değişkenine göre dördüncü kümedeki liselerin dağılımı şekil 6.44'de görülmektedir. Bu kümeye sınavsız öğrenci alan meslek liseleri grubu diyebiliriz. Bu kümenin net bakımından aynı özelliklere sahip olduğu beşinci kümeden az puan alması ve/veya yerleştirme oranının düşük olması bu liselerin beşinci kümede yer alan meslek liselerine oranla öğrencilerinin daha az ortaöğretim başarı puanı almalarıdır.



Şekil 6.44 Ortalama Netlerin Toplamına Göre Dördüncü Kümedeki Liselerin D42 Değişkeni Bazında Dağılımı

Genel olarak Meslek Liselerinden oluşan dördüncü kümede Meslek Liseleri içinde başarısı ile dikkat çeken liseler İstanbul Beşiktaş Anadolu Otelcilik ve Turizm Meslek Lisesi ve Bakırköy İmam Hatip Lisesidir. Meslek Liselerinin dağılımının normal olduğu görülmektedir. Bu kümedeki en başarılı lise türü ise Anadolu Meslek Lisesidir.

Kümelerin ortalama netlerin toplamından elde edilen değişken bazında d42 değişkenine göre beşinci kümedeki liselerin dağılımı şekil 6.45’de görülmektedir. Bu küme dördüncü kümeye benzer şekilde genelde meslek liselerinden oluşsa da küme içinde Anadolu Meslek, Özel Meslek ve Özel Liselerin varlığı dikkat çekmektedir. Bu kümenin net bakımından aynı özelliklere sahip olduğu dördüncü kümeden fazla puan alması ve/veya yerleştirme oranının yüksek olması bu liselerin beşinci kümede yer alan meslek liselerine oranla öğrencilerinin daha az ortaöğretim başarı puanı almalarıdır.



Küme 5

Şekil 6.45 Ortalama Netlerin Toplamına Göre Beşinci Kümedeki Liselerin D42 Değişkeni Bazında Dağılımı

Testlerdeki Netlere Göre Kurumlarındaki Liselerin Dağılımlarına İncelenmesi

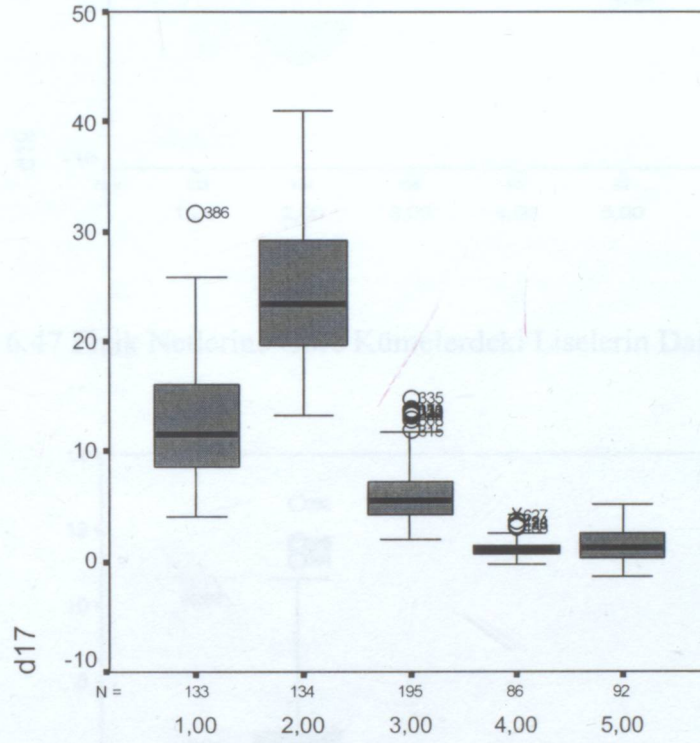
Beşinci kümeyi oluşturan lise türleri içinde Meslek Liseleri ve Normal Liseler içinde öğrencileri oldukça yüksek veya düşük netler yapan liseler aşırı değer olarak dikkat çekmektedir. Meslek Liselerinden yüksek net ortalaması ile dikkat çeken liseler İstanbul Avni Akyol Anadolu Güzel Sanatlar Lisesi, Özel Opera Güzel Sanatlar Lisesi ve Gaziosmanpaşa İmam Hatip Lisesi iken düşük net ortalaması ile Fatih İşitme Engelliler Çok Programlı Lisesi dikkat çekmektedir. Normal liselerden ise başarısı ile dikkat çeken liseler Silivri Gümüşyaka Lisesi ile Çatalca Durusu H.Ökten Lisesi iken başarısızlığı ile dikkat çeken lise Ümraniye Mustafa Kemal Lisesi'dir. Lise türlerinin bu kümede homojen olarak dağılmadığı görülmektedir. Normal Liseler, Özel Liseler ve Yabancı Dille Eğitim Yapan Özel Liselerin dağılımlarını normale yakın olduğu görülmektedir. Anadolu Meslek liselerinin ise aşırı derecede sağa çarpık olduğu dikkat çekmektedir.



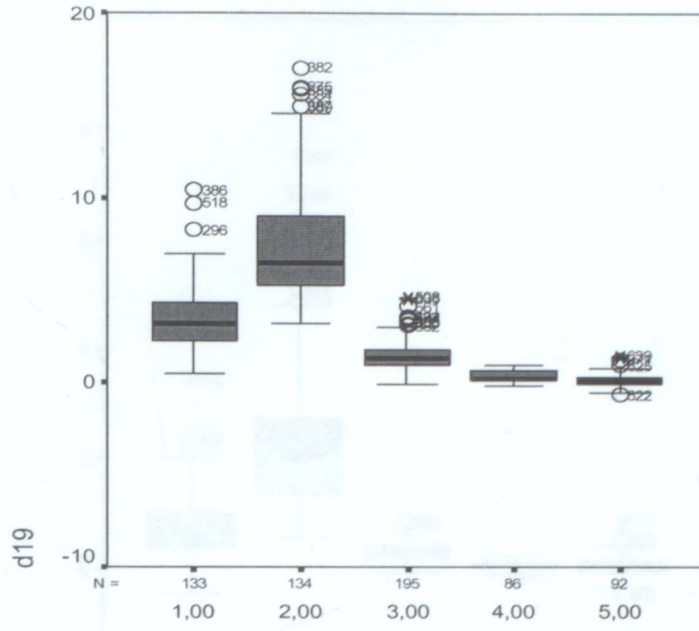
Şekil 6.46 Matematik Netlerine Göre Kurumlarındaki Liselerin Dağılımı

Testlerdeki Netlere Göre Kümelerdeki Liselerin Dağılımlarının İncelenmesi

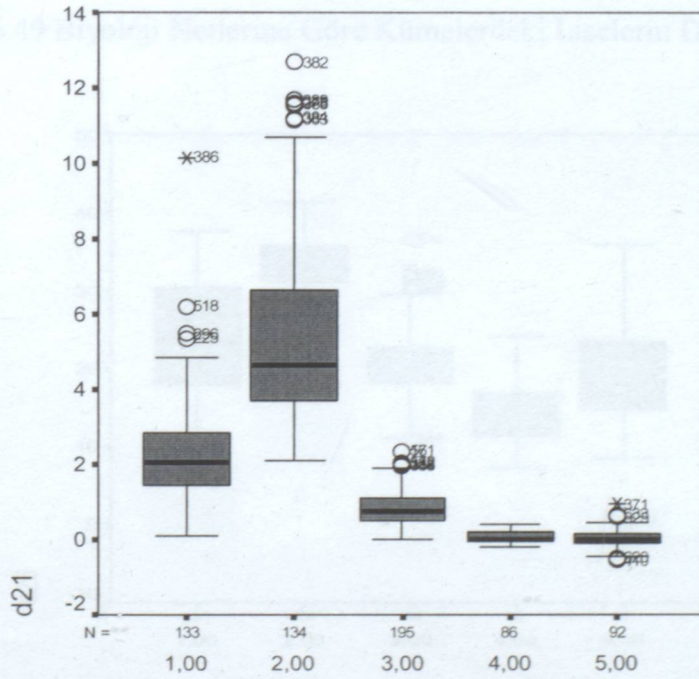
ÖSS sınavındaki matematik, fizik, kimya, biyoloji, türkçe, tarih, coğrafya, felsefe testleri için beş kümenin homojenliği şekil 6.44 ile 6.51 arasında bulunan şekillerde incelenirse, sayısal testlere oranla sözel testlerde homojenliğin daha fazla olduğu görülür. Ayrıca üçüncü, dördüncü ve beşinci kümelerin kendi aralarında tüm test türlerinde homojen olarak dağıldıklarını görmekteyiz.



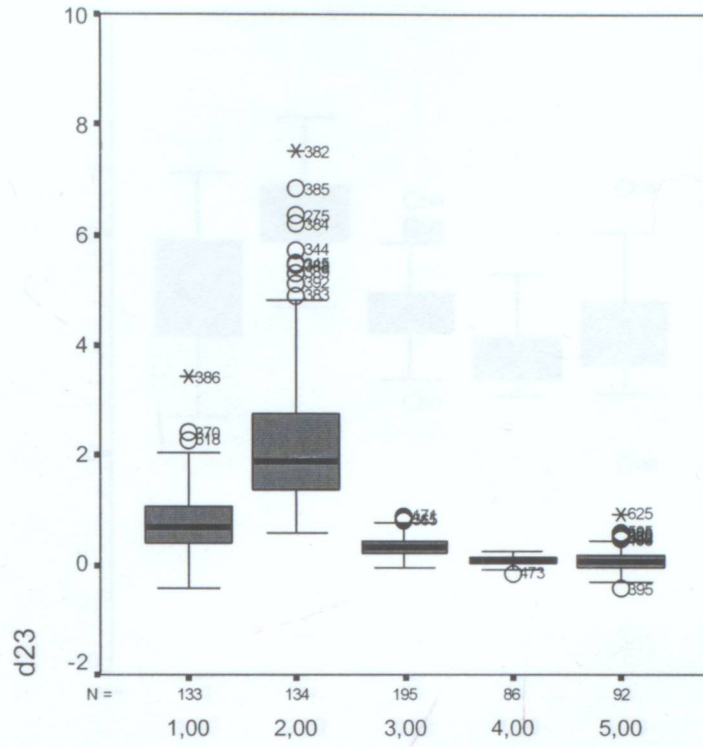
Şekil 6.46 Matematik Netlerine Göre Kümelerdeki Liselerin Dağılımı



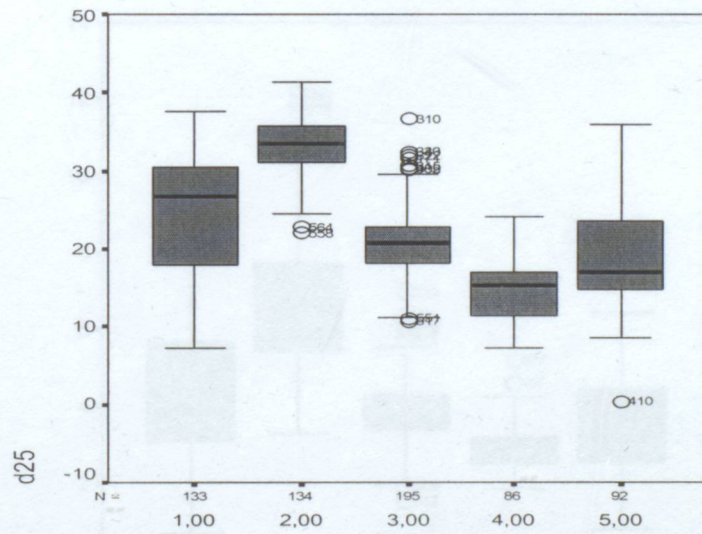
Şekil 6.47 Fizik Netlerine Göre Kümelerdeki Liselerin Dağılımı



Şekil 6.48 Kimya Netlerine Göre Kümelerdeki Liselerin Dağılımı

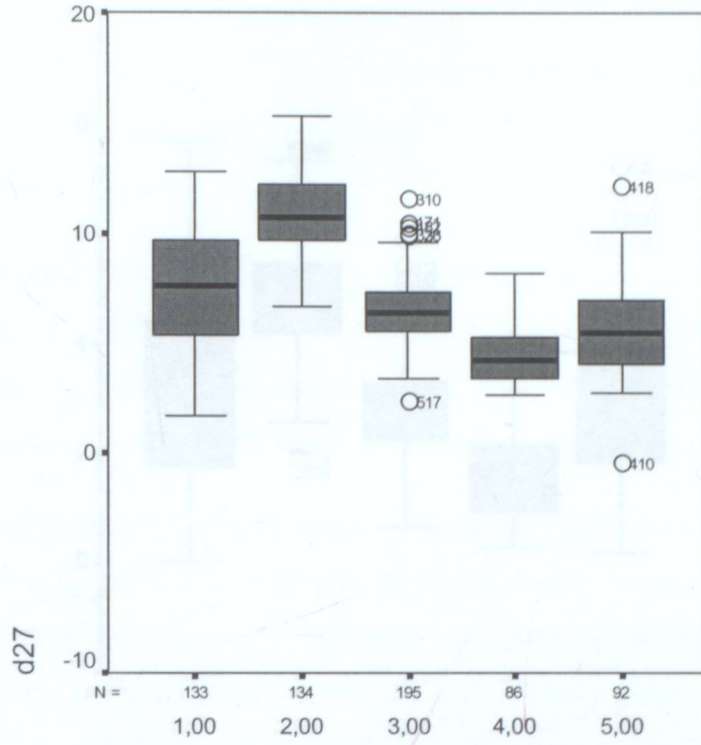


Şekil 6.49 Biyoloji Netlerine Göre Kümelerdeki Liselerin Dağılımı

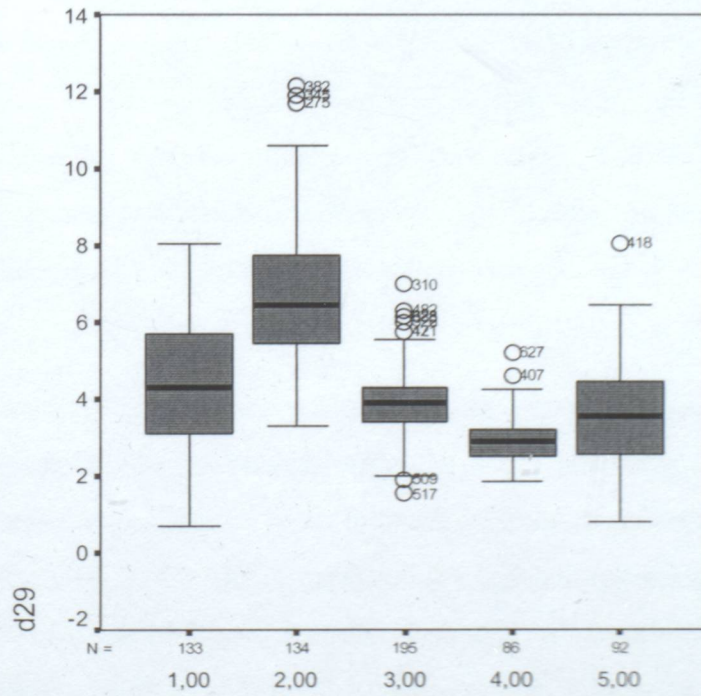


Şekil 6.50 Türkçe Netlerine Göre Kümelerdeki Liselerin Dağılımı

Şekil 6.52 Coğrafya Netlerine Göre Kümelerdeki Liselerin Dağılımı

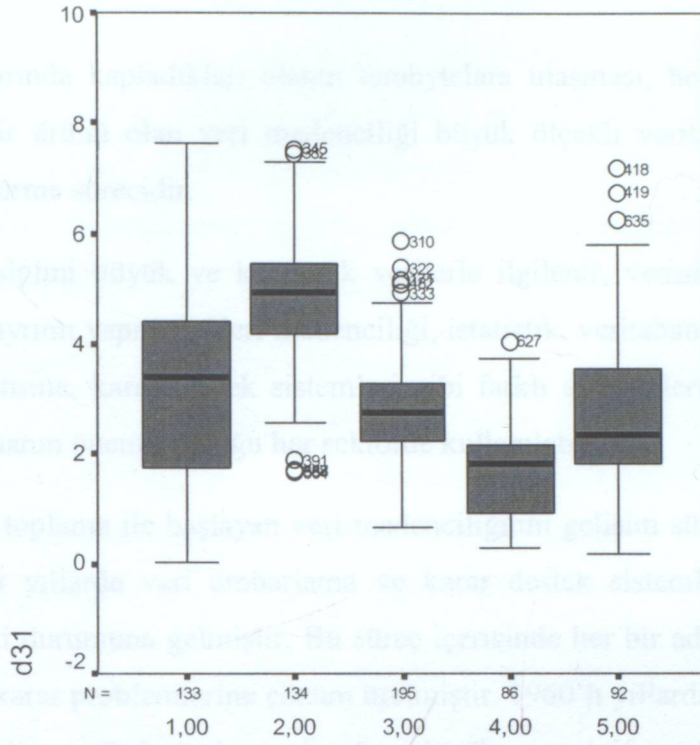


Şekil 6.51 Tarih Netlerine Göre Kümelerdeki Liselerin Dağılımı



Şekil 6.52 Coğrafya Netlerine Göre Kümelerdeki Liselerin Dağılımı

7. SONUÇLAR



Şekil 6.53 Felsefe Netlerine Göre Kümelerdeki Liselerin Dağılımı

7. SONUÇLAR

Verilerin veritabanlarında kapladıkları alanın terabytelara ulaşması, her alanda hızla artan rekabet ortamının bir ürünü olan veri madenciliği büyük ölçekli veritabanlarında bulunan veriyi bilgiye dönüştürme sürecidir.

Veri madenciliği disiplini büyük ve karmaşık verilerle ilgilenir, verinin özelliği ne olursa olsun herhangi veri ayrımı yapmaz. Veri madenciliği, istatistik, veritabanı teknolojisi, makine öğrenmesi, görselleştirme, karar destek sistemleri gibi farklı disiplinlerin entegrasyonundan oluşur ve veri toplamanın önemli olduğu her sektörde kullanılabilir.

1960'lı yıllarda veri toplama ile başlayan veri madenciliğinin gelişim süreci, 1980'li yıllarda veri erişimi, 1990'lı yıllarda veri ambarlama ve karar destek sistemlerinin kullanımı ile birlikte günümüzdeki durumuna gelmiştir. Bu süreç içerisinde her bir adım bir önceki adıma göre daha karmaşık karar problemlerine çözüm üretmiştir. 1960'lı yıllarda geriye dönük statik bilgiler elde edilirken günümüzde geleceğe dönük proaktif enformasyon dağıtımı yapılmaktadır. Bu gelişime paralel olarak da ürün sağlayıcılarındaki artış da dikkat çekicidir.

Veri madenciliği disiplininin karar problemini çözmede kullandığı çözüm araçları içerisinde istatistik teknikleri önemli bir yer tutmaktadır. Her ne kadar veri madenciliğinin başlı başına bir bilim dalı olarak ortaya çıktığı görülse de istatistik disiplininden de yoğun bir şekilde faydalandığını da inkar etmemek gerekir. Veri madenciliği disiplininin farklı disiplinlerin entegrasyonundan oluşan bir disiplin olduğu unutulmamalıdır. Tabi ki veri madenciliği ile ilgilenen kişiler uygulamalarında kendi uzmanlık alanlarına daha ağırlık vereceklerdir. Fakat öyle durumlar olacaktır ki çözümlerinde farklı disiplinlerden de faydalanma gereksinimi duyacaklardır.

Veri madenciliğine istatistik bağlamında yaklaştığımızda, bir çok istatistik tekniği yoğun bir şekilde kullandığı görülmektedir. Çalışmada kullanılan temel bileşenler analizi ile kümeleme analizi veri madenciliğinde kullanılan veri indirgeme tekniklerindendir. Temel bileşenler analizi ile kümeleme analizi teorik olarak incelenerek kümeleme metotlarından K-Means, K-Medoids, Tek Bağlantı(Single Linkage), Average Linkage ve Complete Linkage metotlarına ilişkin algoritmalar adım adım uygulama ile gösterilmiştir. Kümeleme analizi konusunda açıklanan bu algoritmalarından K-Medoids algoritması diğer metotlara göre sapan değerlere daha az duyarlı ve büyük veri setlerinde kullanılır.

Temel bileşenler analizi ve kümeleme analizi kullanılarak 2001 ÖSS sınavında İstanbul ilinde bulunan liseler mezun verdikleri öğrencilerinin performanslarına göre kümelenmişlerdir. Bu uygulama için SAS Enterprise Miner 8.0 ürünü kullanılmıştır. Temel bileşenler analizi kullanılarak korelasyonsuz ve daha az sayıda değişken elde edilerek kümeleme analizinde kullanılmaya uygun değişkenler üretilmiştir. Bu değişkenler kullanılarak kümeleme analizi gerçekleştirilmiştir.

Kümeleme analizi sonucunda liseler beş kümeye ayrılmışlardır. Bu kümelerin birbirlerinden net bir biçimde ayrıldığı görülmektedir. Birinci kümedeki liseler diğer kümeler arasında başarı bakımından ikinci sıradadır. Bu kümede ağırlıklı olarak Yabancı Dille Eğitim Yapan Özel Liseler, Anadolu Meslek Liseleri, Meslek Liseleri ve Yabancı Dil Ağırlıklı Liseler yer almaktadır. Bu liseler genellikle sınavla öğrenci alan liselerdir.

İkinci kümede yer alan liseler ise en başarılı liselerdir. Bu liselerin geneli sınavla öğrenci alan ya da not ortalamasına göre öğrenci alan liselerdir. Yabancı Dil Ağırlıklı Liseler, Anadolu Liseleri, Yabancı Dille Eğitim Yapan Özel Liseler, Fen Liseleri (Özel-Devlet) bu kümede bulunan lise türleri arasında yer almaktadır.

Üçüncü kümede yer alan liselerin büyük çoğunluğunu normal liseler oluşturmaktadır. Bu kümenin adını normal liseler kümesi olarak adlandırabiliriz. Bu kümede bazı Meslek, Anadolu Meslek ve Yabancı Dille Eğitim Yapan Özel Liseler bulunmaktadır. Bu kümede bulunan liseler genelde sınavsız öğrenci alan okullardır. Sınavla ya da not ortalaması ile öğrenci alan okulların bu kümede görülmesi dikkat çekicidir.

Dördüncü kümede bulunan liselerin geneli klasik Endüstri Meslek Liseleridir. Bu liseler başarı olarak en düşük liselerdir. Beşinci kümedeki liseler ile netleri yakın olsa da bu kümelerin yüksek öğretim kurumlarına öğrenci yerleştirme oranları düşüktür.

Beşinci kümede, dördüncü kümeye benzer şekilde genel meslek liseleri ağırlıklı olsa da küme içinde Anadolu Meslek, Özel Meslek ve Özel Liselerin varlığı dikkat çekmektedir. Bu kümede bulunan liselerin yüksek öğretim kurumlarına öğrenci yerleştirme oranları dördüncü kümeye göre daha yüksektir.

Kümeleme analizi sonrası uygulanan görselleştirme teknikleri ile de kümelerin okul bazında dağılımları gösterilmiştir. Buna göre, kümeler içinde okulların başarılarını da gözlemlemek mümkün olmuştur.

Sonuç olarak elde edilen kümeler sınavsız öğrenci alan liselerin başarı olarak (özellikle dördüncü ve beşinci kümeler) çok düşük performans gösterdiklerini ortaya koymaktadır. Buradan yola çıkılarak normal liseler de dahil olmak üzere meslek liselerinin yapılan ortalama net sayılarının azlığından dolayı rehabilitasyona gereksinim duyduğu açıktır. Kümeleme kuralları kullanılarak her bir lisenin farklı yıllardaki başarıları incelenebilir ve kümeleme kuralları bir başarı kriteri olarak kullanılabilir.

- Arslan, H. (1999), Türkiye Ekonomisi İstatistik ve Veri Analizi, Ankara: Ankara Üniversitesi İktisadi İstatistik Enstitüsü Yayınları.
- Banerjee, D., (2003), Yapay Sinir Ağları, Seçkin Yayıncılık, İstanbul.
- Banerjee, D., U. Gregory Paletsky-Shapiro, Paletsky-Shapiro, U. (2003), Towards a Unifying Framework for Data Mining, *Journal of Data Mining and Knowledge Discovery*, 7(1), 1-10.
- Banerjee, D., (2002), Data Mining: Statistics and Machine Learning, *Journal of Data Mining and Knowledge Discovery*, 6(1), 1-10.
- Banerjee, D., (1999), Data Mining: Statistics and Machine Learning, *Journal of Data Mining and Knowledge Discovery*, 3(1), 1-10.
- Banerjee, D., Anderson, R. K., Tishan, R. J., Bloch, W. F., (1999), *Data Mining: Statistics and Machine Learning*, Prentice Hall Publications, New Jersey.
- Banerjee, D., Kamber, M., (2001), *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers, USA.
- Banerjee, D. L., (1998), *Data Mining: Statistics and Machine Learning*, <http://www.stat.org/publications/98/98d.pdf>.
- Banerjee, D. L., Why Data Mining is More Than Statistics, *World Wide Web*, <http://www.w3.org/2000/09/proceedings/arkhiv/arkhiv0490.pdf>.
- Banerjee, R. A., Wichern, D. W., (1983), *Applied Multivariate Statistical Analysis*, 357, Prentice Hall Publications, USA.
- Banerjee, J., Willett, J., (2001), *Data Mining and Knowledge Management*, AIR Conference, Long Beach, CA.
- Banerjee, N. (2000), *Temel İstatistik: Olasılık ve Olasılık Dağılımları*, İÜ İktisadi İstatistik Fakültesi, İstanbul.
- Banerjee, K., (1999), *Peşer Programları ile İstatistiksel Veri Analizi*, Kasa Kitabevi, İstanbul.
- Banerjee, E., (2003), *Yapay Sinir Ağları*, Papaya Eğitim Yayıncılık, İstanbul.
- Banerjee, C. A., (2002), *Methods of Multivariate Analysis*, John Wiley & Sons Inc, Canada.
- Banerjee, J. R., Oetz, M. W., (2000), *Data Mining: A Tutorial Based Primer*, Pearson Education, USA.
- Banerjee, O., (2001), *Data Mining Cookbook*, John Wiley & Sons, USA.

KAYNAKLAR

- Aldana,W.,A.,(Mayıs 2000),Data Mining Industry: Emerging Trends and New Opportunities,B.S. Electrical Engineering and Computer Science MIT
- Berry,M., Linof, G., (1997), Data Mining Techniques, John Wiley & Sons,USA
- Cabena, P., Hadjinian, P.,Stadler, R.,(1998), Discovering Data Mining, Prentice Hall PTR, New Jersey
- Data Mining, Statistical Methods Mining and History of Statistics,stat.tamu.edu/ftp/pub/eparzen/ future .pdf
- DiE,(1999),Türkiye Ekonomisi: İstatistik ve Yorumlar,Ankara
- Elmas,Ç., (2003), Yapay Sinir Ağları,Seçkin Yayıncılık,Ankara
- Fayyad, U., Gregory Piatetsky-Shapiro, Padhraic Smyth, Knowledge Discovery and Data Mining: Towards a Unifying Framework, <http://www-aig.jpl.nasa.gov/kdd96>
- Ganesh,S.,(2002),Data Mining Should It Be Included in the Statistics Curriculum?, www.stat.auckland.ac.nz/~iase/publications/1/314_gane.pdf
- Groth, R., (1999), Data Mining: Building Competitive Advantage, Prentice Hall PTR, New Jersey
- Hair,J.F.,Anderson R.E.,Tatham,R.L.,Black, W.C.,(1995),Multivariate Data Analysis with Readings,Prentice Hall Publications,New Jersey
- Han, J.,Kamber, M.,(2001), Data Mining Concepts and Techniques,Morgan Kaufmann Publishers,USA
- Hand,D.J.,(1998), Data Mining: Statistics and More?, www.amstat.org/publications/tas/hand.pdf
- Hand,D.J.,Why Data Mining is More Than Statistics Writ Large, www.stat.fi/isi99/proceedings/arkisto/varasto/hand0490.pdf
- Johnson,R.A., Wichern,D.W.,(1982),Applied Multivariate Statistical Analysis,357,Prentice Hall Publications, USA
- Luan, J., Willett,T., (2001),Data Mining and Knowledge Management,AIR Conference, Long Beach, CA.
- Orhunbilge,N.,(2000), Tanımsal İstatistik Olasılık ve Olasılık Dağılımları,İ.Ü. İşletme Fak. Yay., İstanbul
- Özdamar,K.,(1999),Paket Programlar ile İstatistiksel Veri Analizi,Kaan Kitabevi,Eskişehir
- Öztemel E.,(2003),Yapay Sinir Ağları,Papatya Eğitim Yayıncılık,İstanbul
- Rencher, C. A.,(2002), Methods of Multivariate Analysis,John Wiley & Sons Inc.,Canada
- Roiger,J.R. ,Geatz, M. W., (2003),Data Mining A Tutorial Based Primer,Pearson Education,USA
- Rud, O.,(2001),Data Mining Cookbook, John Wiley & Sons,USA

Two Crows Corporation,(1999),Introduction to Data Mining and Knowledge Discovery III.Edition, www.twocrows.com/intro-dm.pdf

Vahaplar, A.,Dr. İnceoğlu, M., Veri Madenciliği ve Elektronik Ticaret, www.bayar.edu.tr/~baum/dokumanlar/inceoglu.doc

Veri Madencilik Uygulamaları ve Eylimi, courses.cs.deu.edu.tr/cse572/Veri%20MadencilikUygulamalar.doc

Westphal,C., Blaxton,T.,(1998),Data Mining Solutions,John Wiley & Sons,USA

What is data mining, <http://ugrad.stat.ubc.ca/~stat447j/resources/chapter1.pdf>

EKLER

EK 1 Veri Setinde Bulunan Nicel Değişkenlere İlişkin Histogramlar

EK 2 Veri Setinde Bulunan Nicel Değişkenlere İlişkin Kutu Diyagramları

EK 3 Kümelerin Nicel Değişkenler Bazında İstatistiksel Farklılığının Bonferroni Güven Bölgesi Yaklaşımı ile Testi

EK 4 Veri Setinde Bulunan Nicel Değişkenlerin Kümelere Göre İstatistikleri

EK 5 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Göre Dağılımları

EK 6 D40 Değişkenindeki Lise Türlerinin Kümelere Göre Dağılımları

EK 7 D41 Değişkenindeki Lise Türlerinin Kümelere Göre Dağılımları

EK 8 D42 Değişkenindeki Lise Türlerinin Kümelere Göre Dağılımları

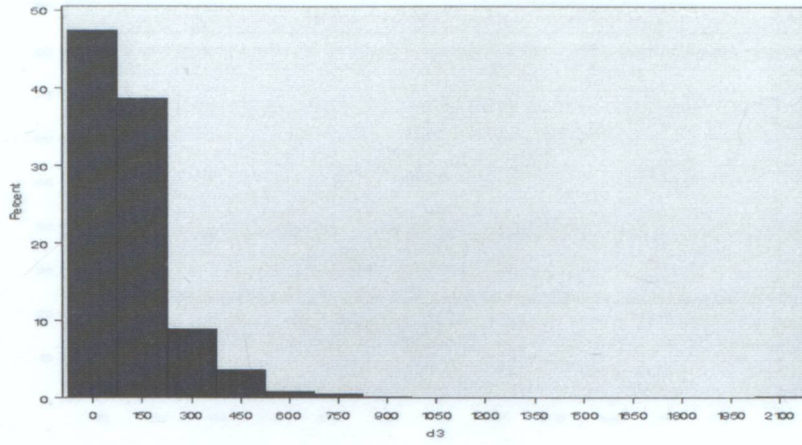
EK 9 ÖSYM'nin Tanımladığı Lise Türlerinin Testlere Göre Kümelerdeki Net Ortalamaları

EK 10 D40 Değişkenindeki Lise Türlerinin Testlere Göre Kümelerdeki Net Ortalamaları

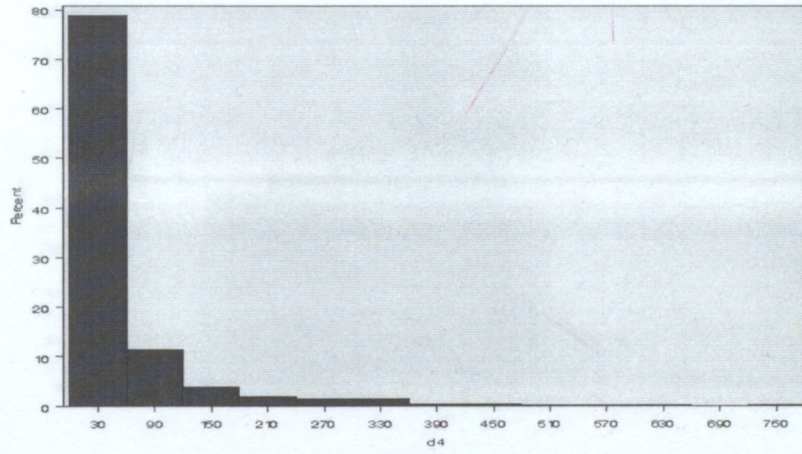
EK 11 D41 Değişkenindeki Lise Türlerinin Testlere Göre Kümelerdeki Net Ortalamaları

EK 12 D42 Değişkenindeki Lise Türlerinin Testlere Göre Kümelerdeki Net Ortalamaları

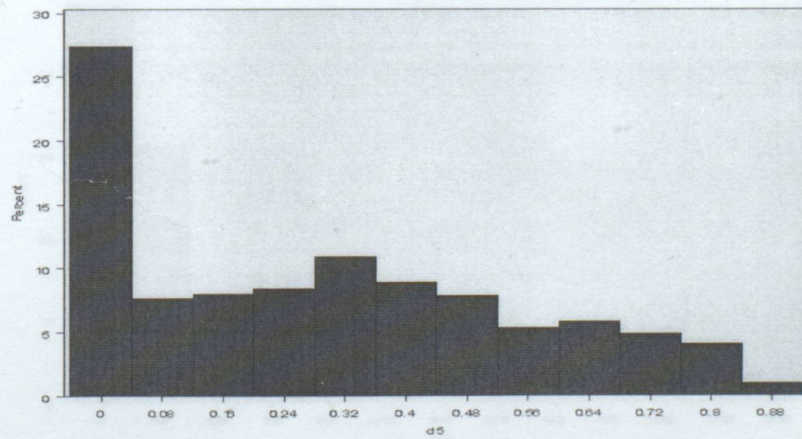
EK 1 Veri Setinde Bulunan Nicel Değişkenlere İlişkin Histogramlar



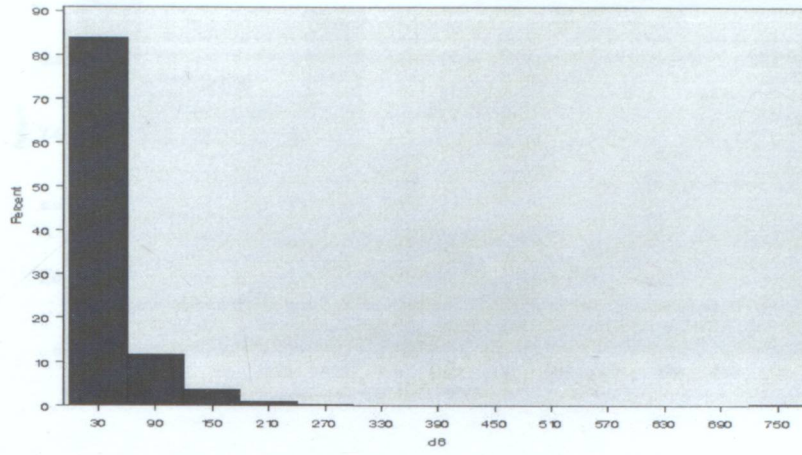
Şekil Ek 1.1 D3 Değişkenine İlişkin Histogram



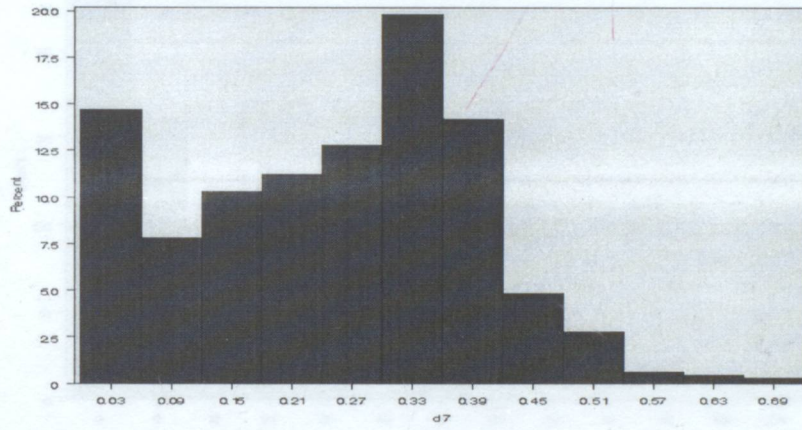
Şekil Ek 1.2 D4 Değişkenine İlişkin Histogram



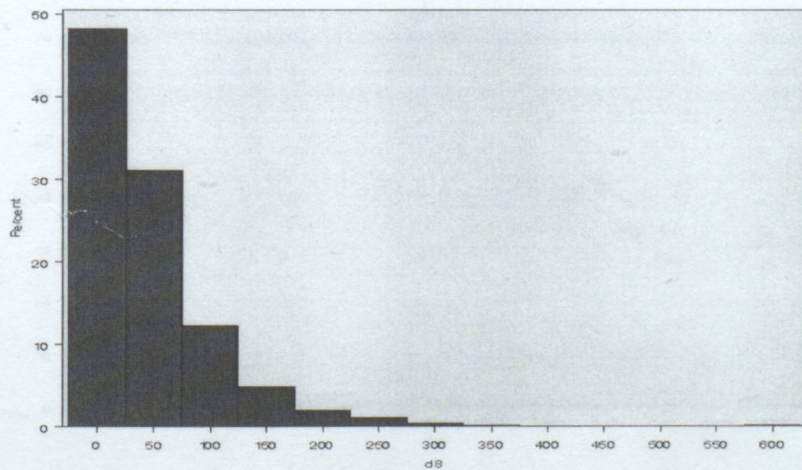
Şekil Ek 1.3 D5 Değişkenine İlişkin Histogram



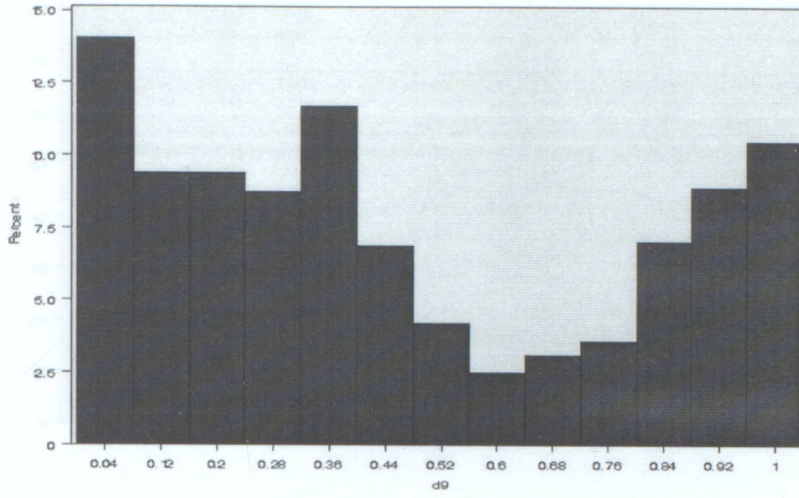
Şekil Ek 1.4 D6 Değişkenine İlişkin Histogram



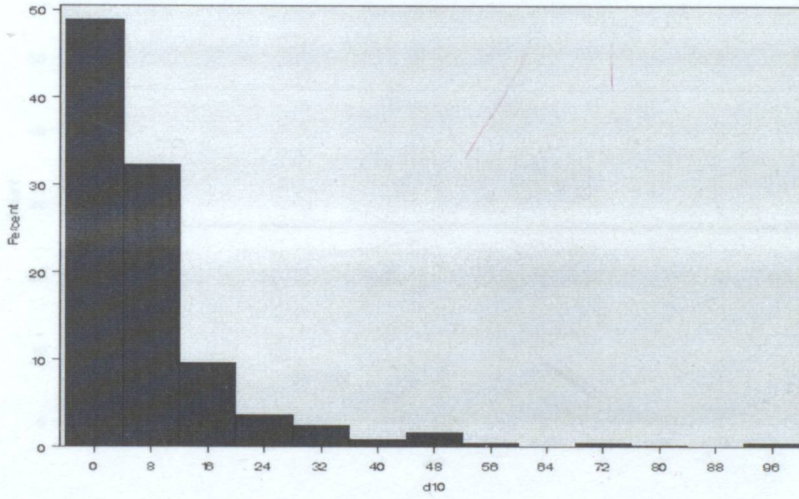
Şekil Ek 1.5 D7 Değişkenine İlişkin Histogram



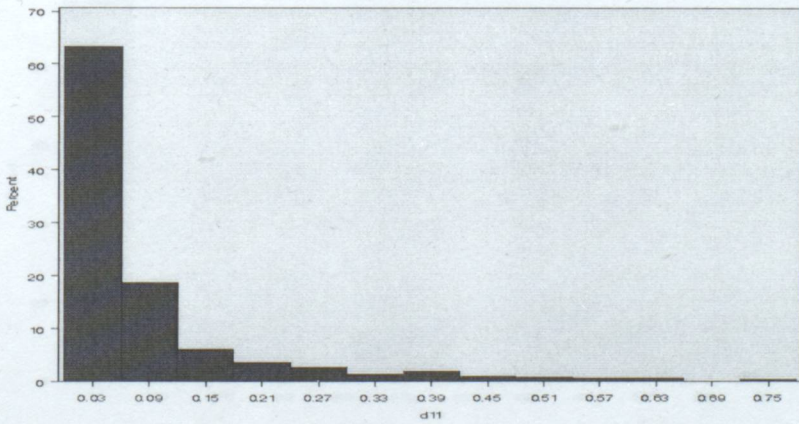
Şekil Ek 1.6 D8 Değişkenine İlişkin Histogram



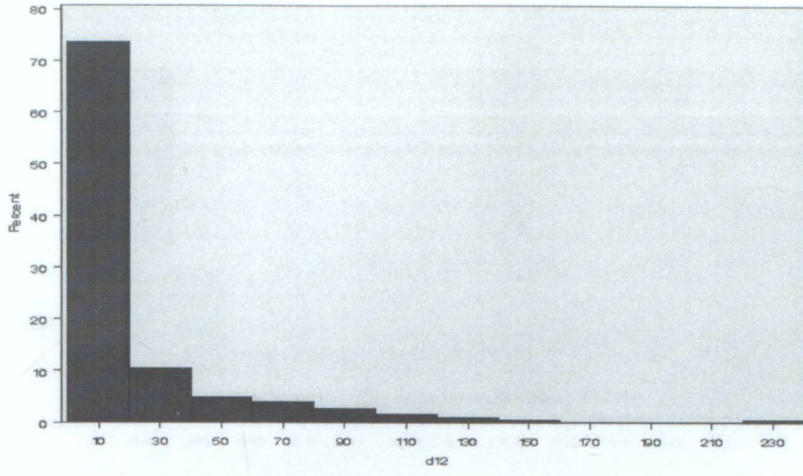
Şekil Ek 1.7 D9 Değişkenine İlişkin Histogram



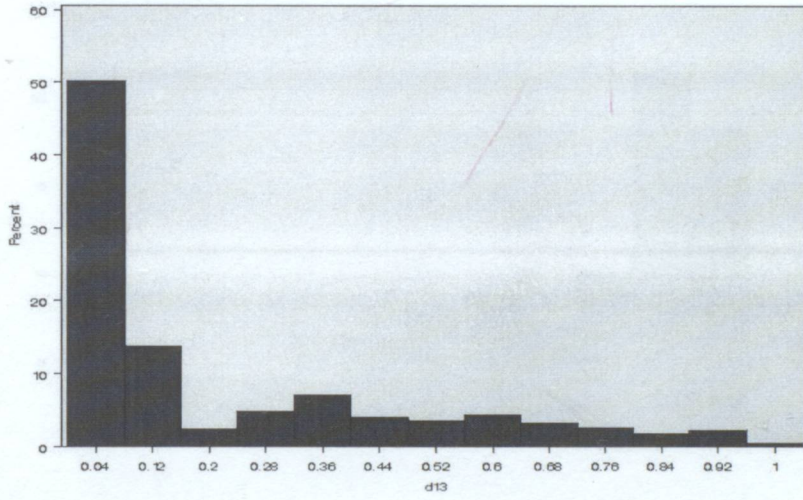
Şekil Ek 1.8 D10 Değişkenine İlişkin Histogram



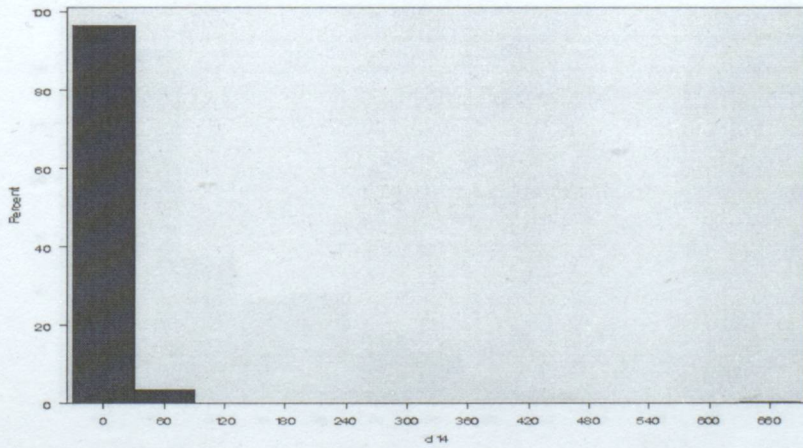
Şekil Ek 1.9 D11 Değişkenine İlişkin Histogram



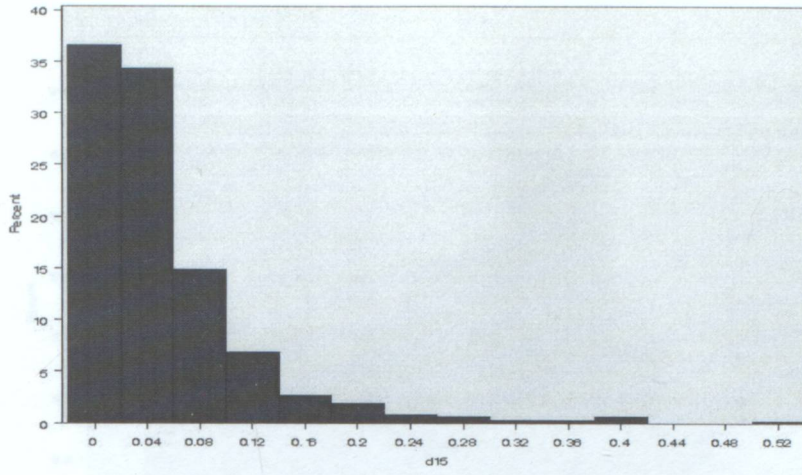
Şekil Ek 1.10 D12 Değişkenine İlişkin Histogram



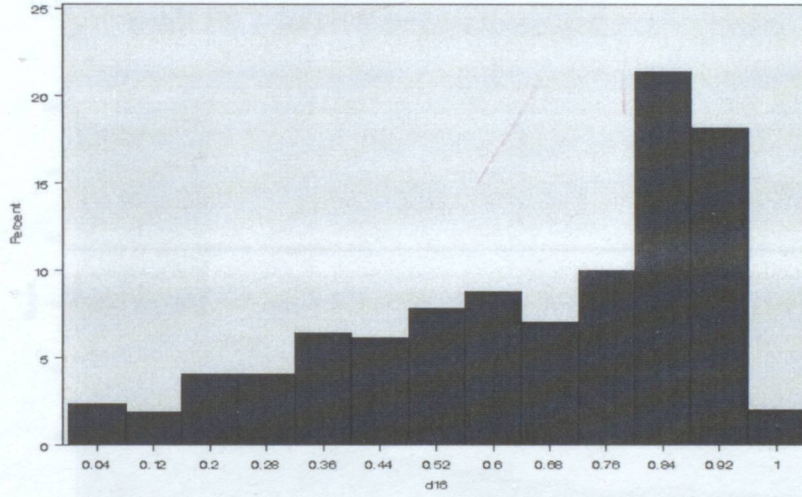
Şekil Ek 1.11 D13 Değişkenine İlişkin Histogram



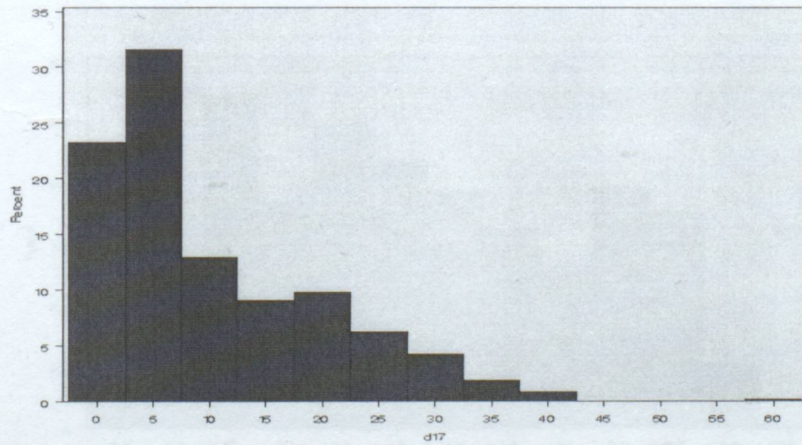
Şekil Ek 1.12 D14 Değişkenine İlişkin Histogram



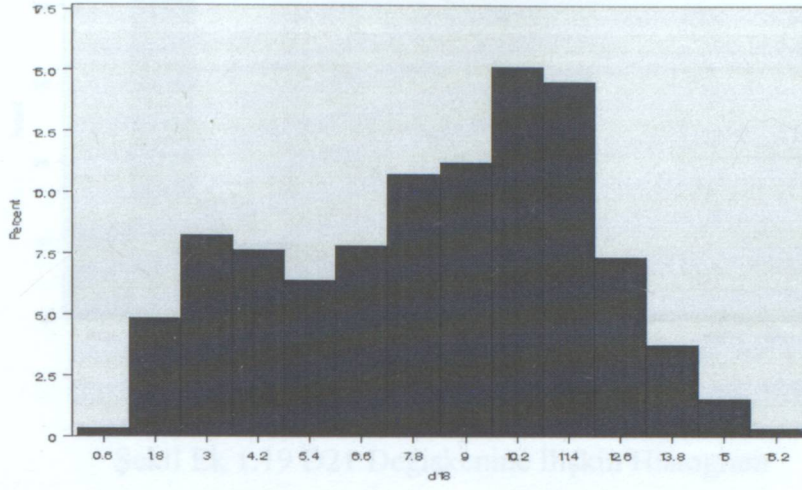
Şekil Ek 1.13 D15 Değişkenine İlişkin Histogram



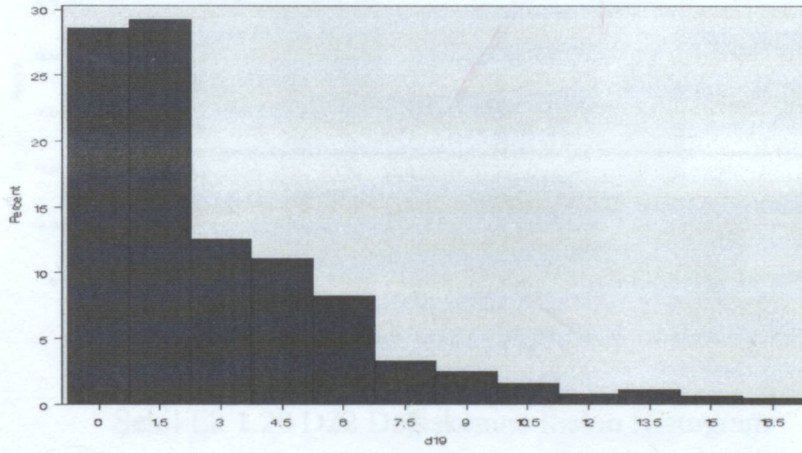
Şekil Ek 1.14 D16 Değişkenine İlişkin Histogram



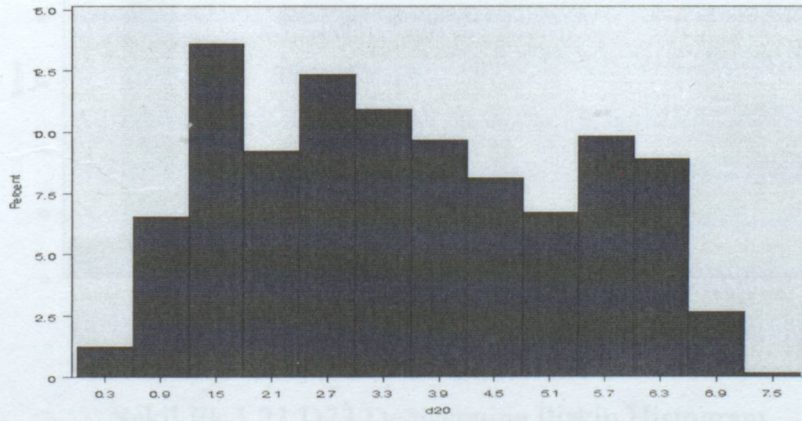
Şekil Ek 1.15 D17 Değişkenine İlişkin Histogram



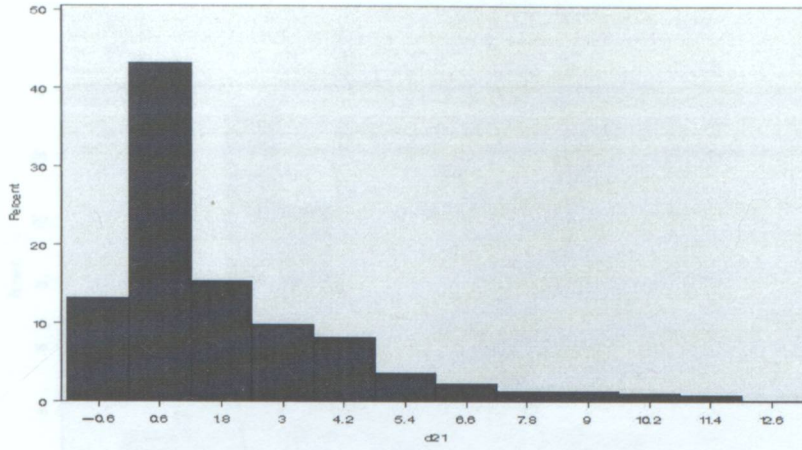
Şekil Ek 1.16 D18 Değişkenine İlişkin Histogram



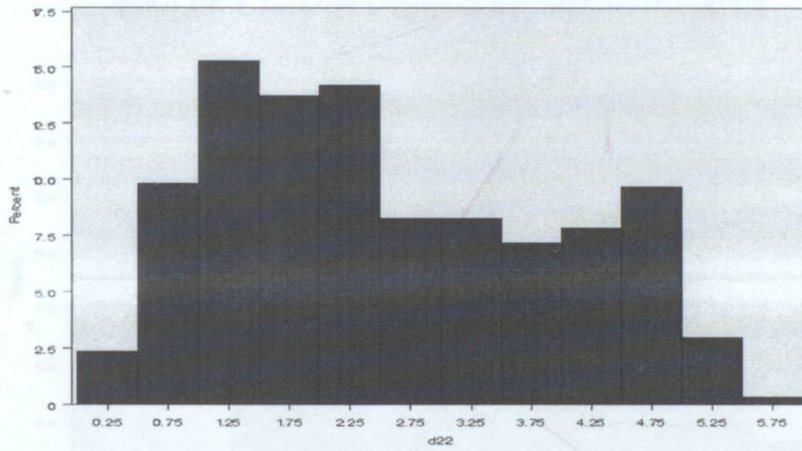
Şekil Ek 1.17 D19 Değişkenine İlişkin Histogram



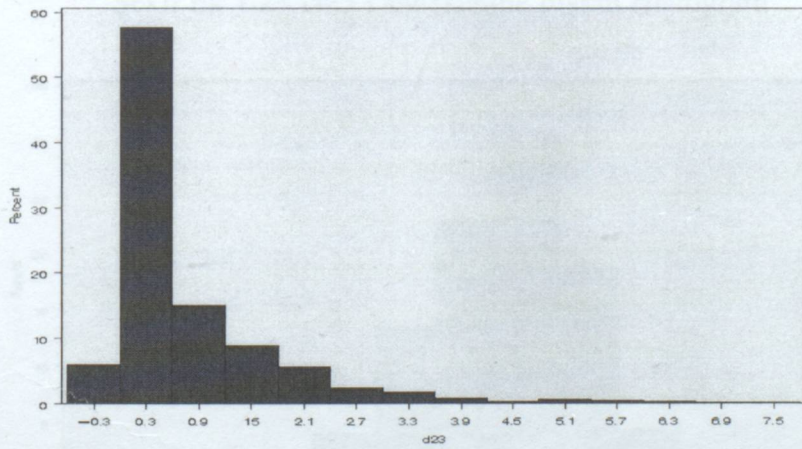
Şekil Ek 1.18 D20 Değişkenine İlişkin Histogram



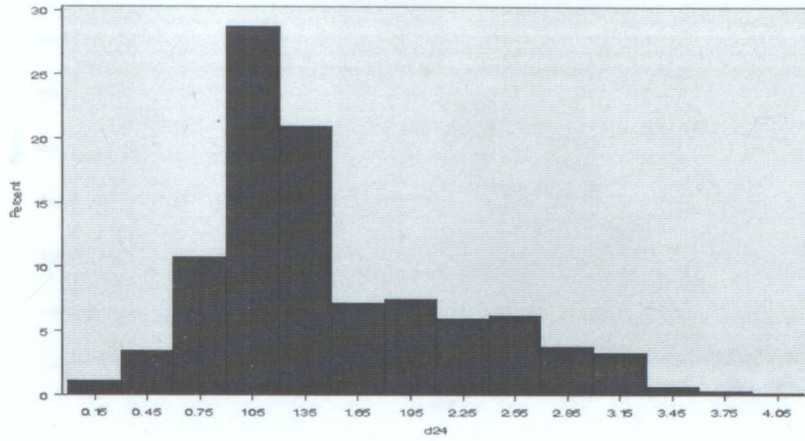
Şekil Ek 1.19 D21 Değişkenine İlişkin Histogram



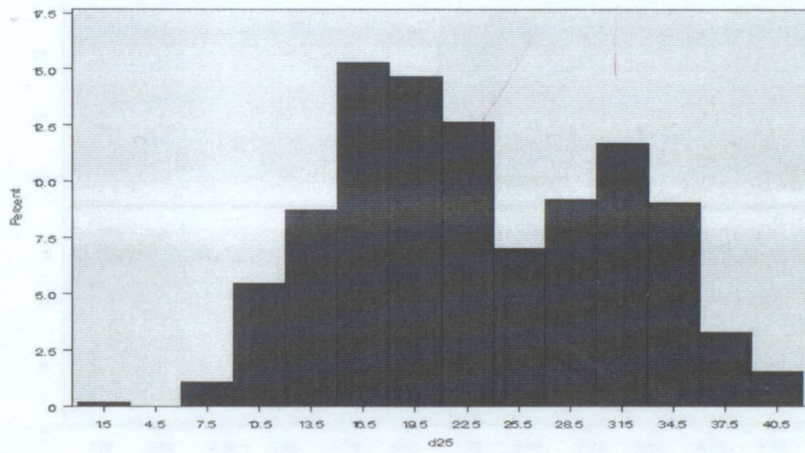
Şekil Ek 1.20 D22 Değişkenine İlişkin Histogram



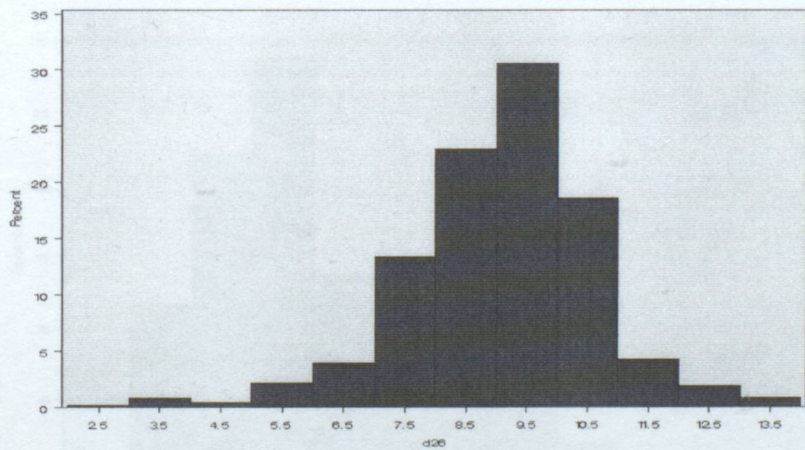
Şekil Ek 1.21 D23 Değişkenine İlişkin Histogram



Şekil Ek 1.22 D24 Değişkenine İlişkin Histogram

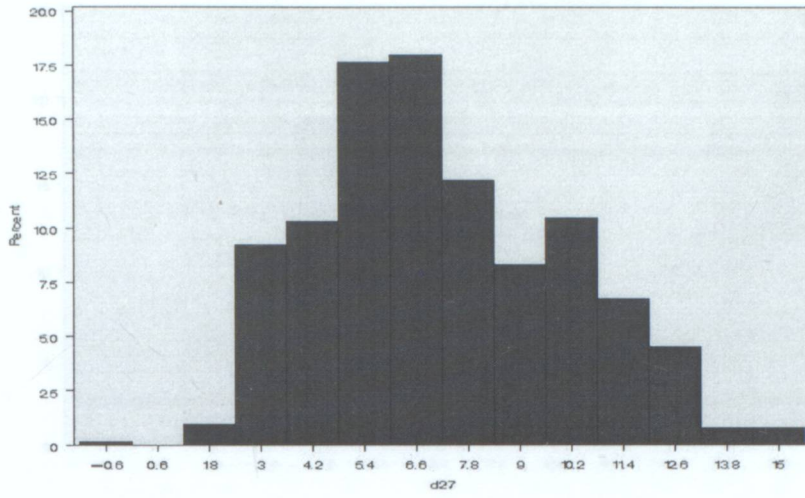


Şekil Ek 1.23 D25 Değişkenine İlişkin Histogram

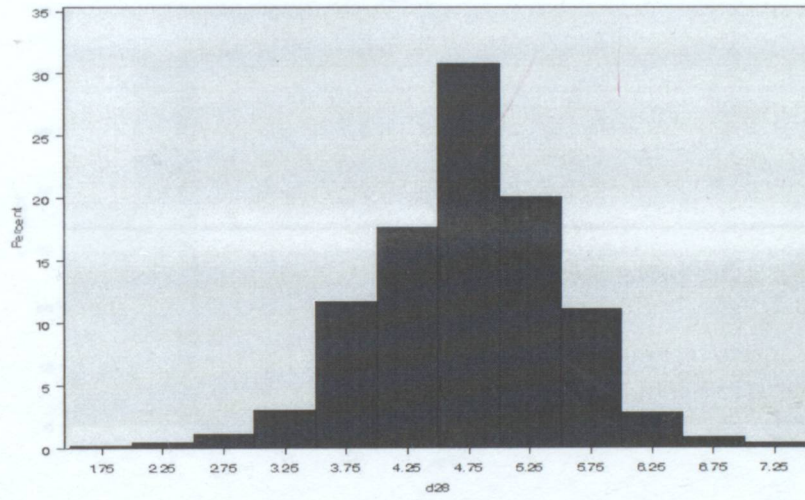


Şekil Ek 1.24 D26 Değişkenine İlişkin Histogram

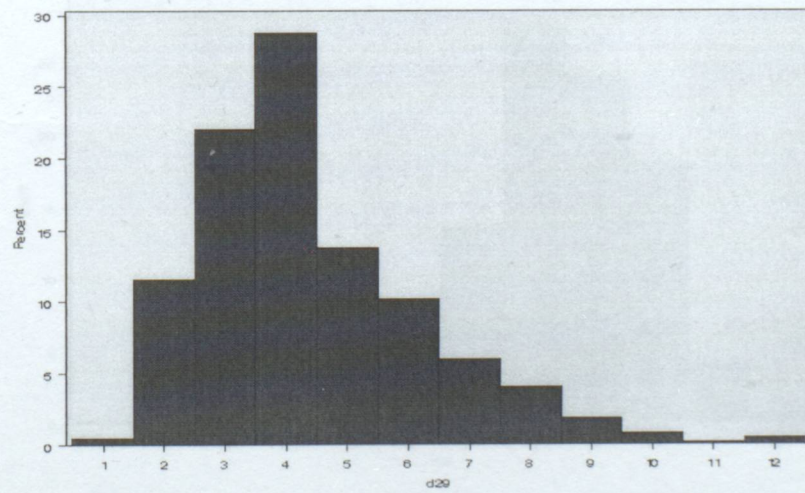
Şekil Ek 1.27 D29 Değişkenine İlişkin Histogram



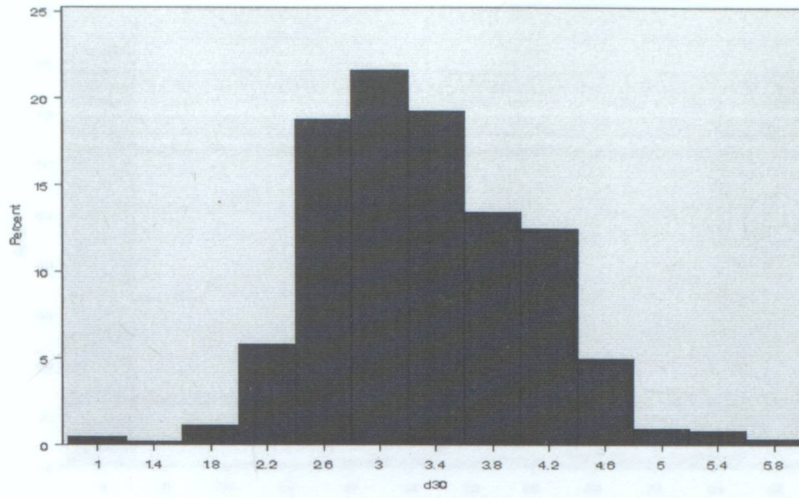
Şekil Ek 1.25 D27 Değişkenine İlişkin Histogram



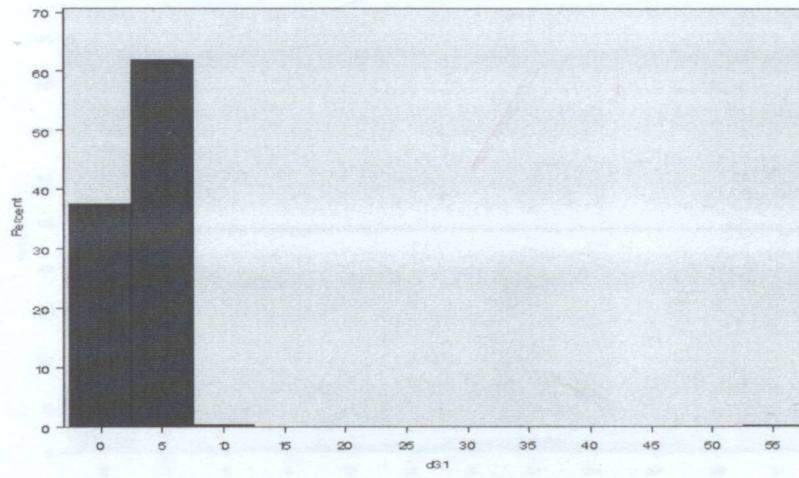
Şekil Ek 1.26 D28 Değişkenine İlişkin Histogram



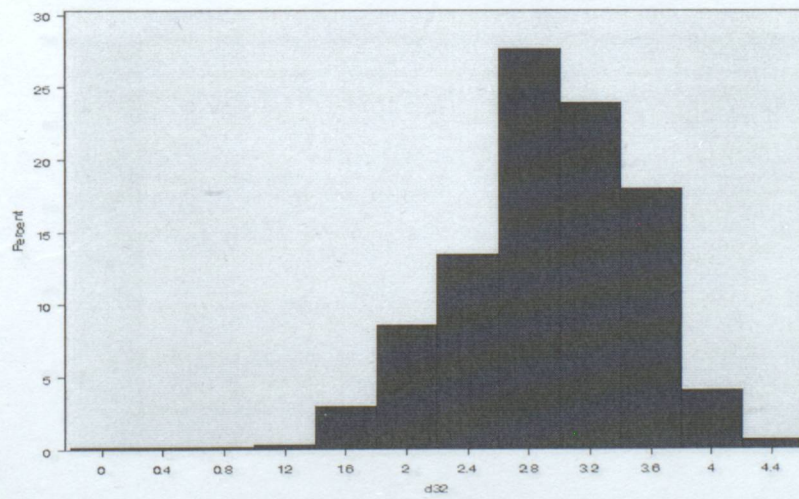
Şekil Ek 1.27 D29 Değişkenine İlişkin Histogram



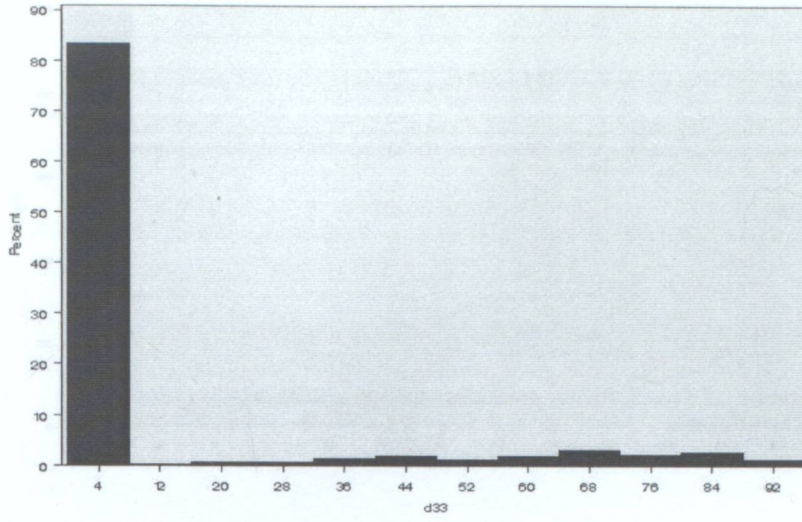
Şekil Ek 1.28 D30 Değişkenine İlişkin Histogram



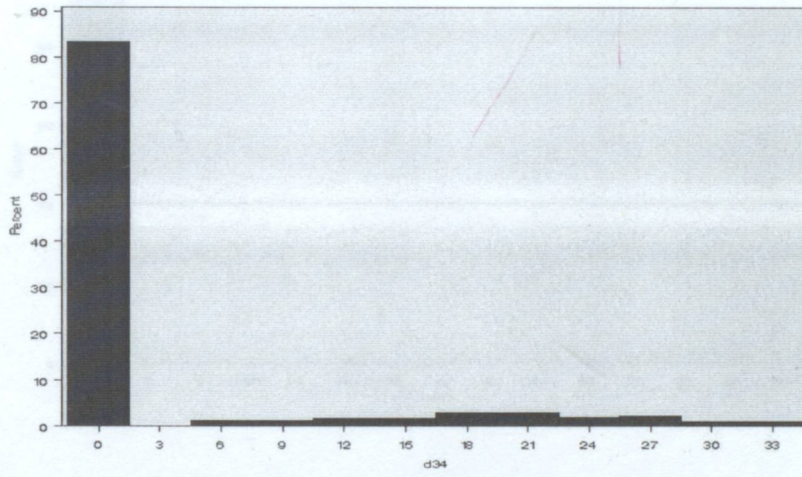
Şekil Ek 1.29 D31 Değişkenine İlişkin Histogram



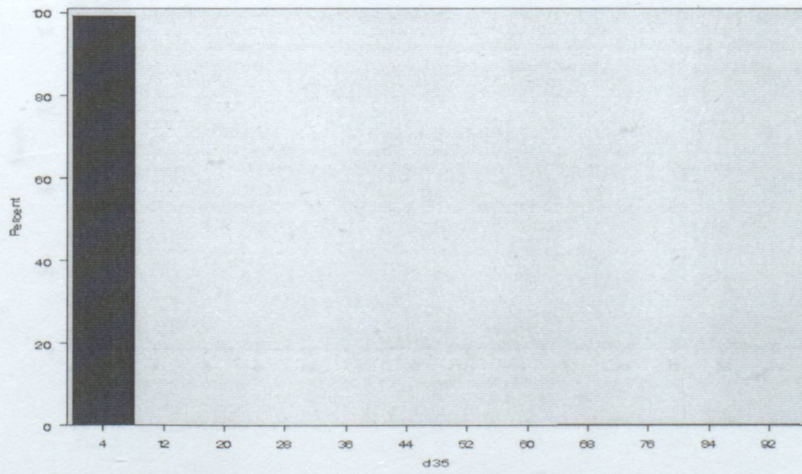
Şekil Ek 1.30 D32 Değişkenine İlişkin Histogram



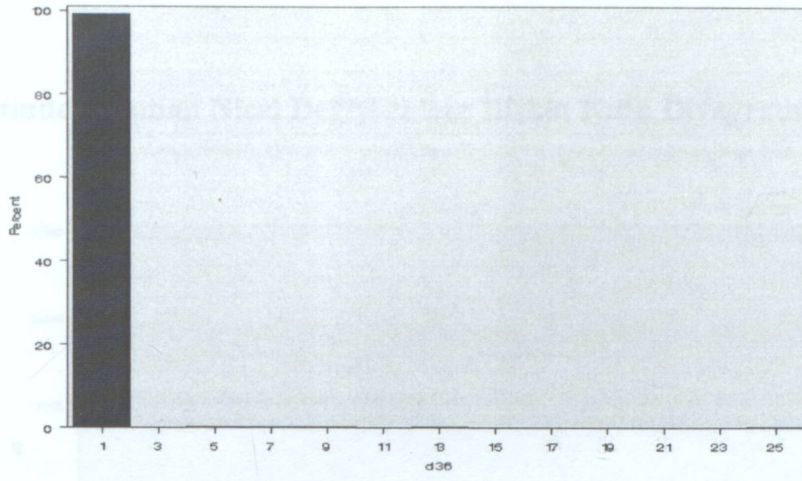
Şekil Ek 1.31 D33 Değişkenine İlişkin Histogram



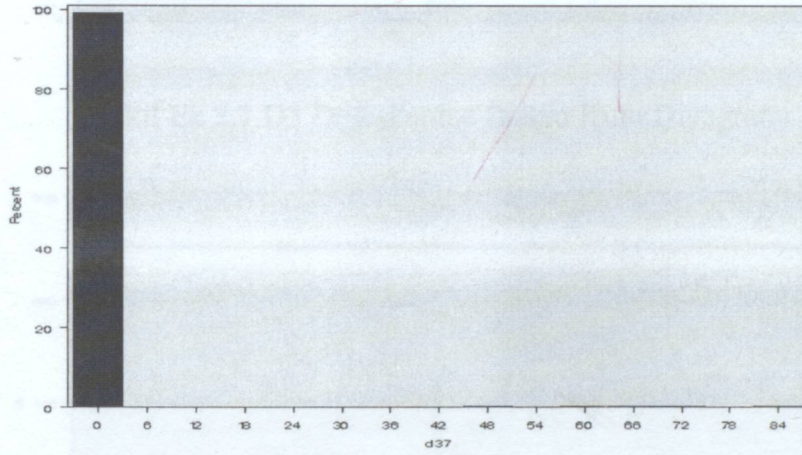
Şekil Ek 1.32 D34 Değişkenine İlişkin Histogram



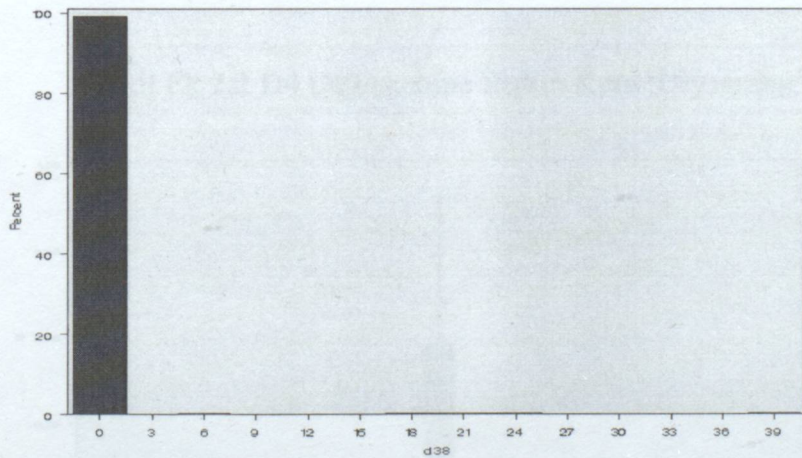
Şekil Ek 1.33 D35 Değişkenine İlişkin Histogram



Şekil Ek 1.34 D36 Değişkenine İlişkin Histogram

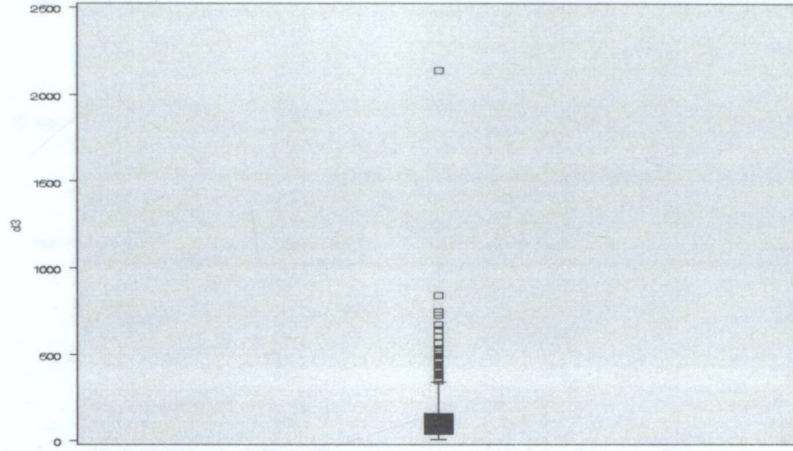


Şekil Ek 1.35 D37 Değişkenine İlişkin Histogram

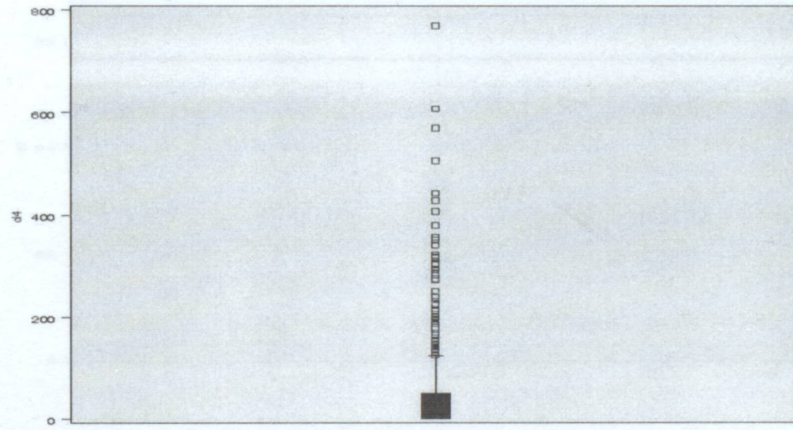


Şekil Ek 1.36 D38 Değişkenine İlişkin Histogram

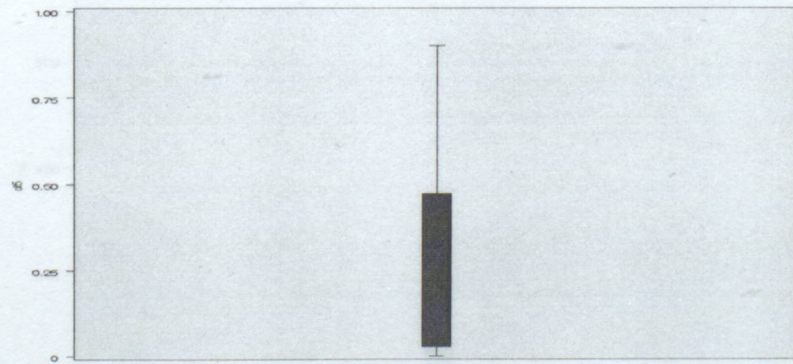
EK 2 Veri Setinde Bulunan Nicel Değişkenlere İlişkin Kutu Diyagramları



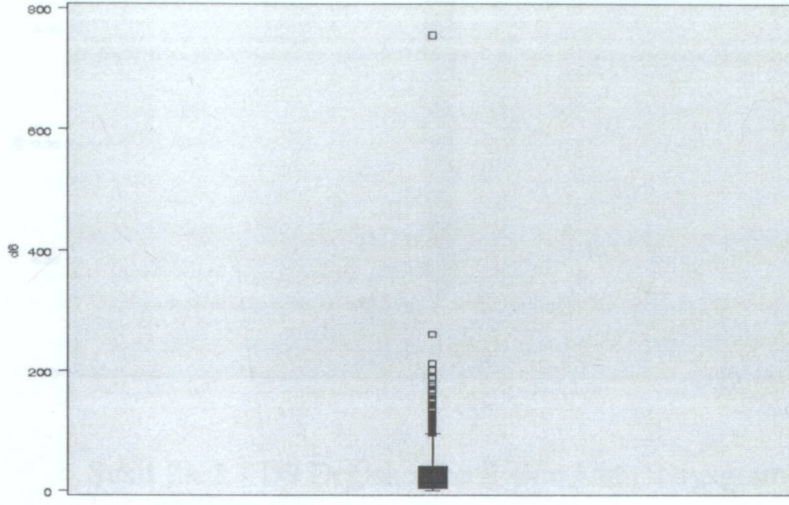
Şekil Ek 2.1 D3 Değişkenine İlişkin Kutu Diyagramı



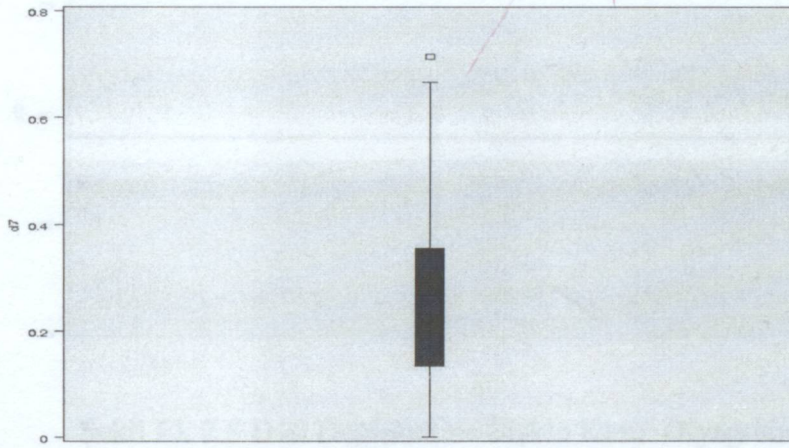
Şekil Ek 2.2 D4 Değişkenine İlişkin Kutu Diyagramı



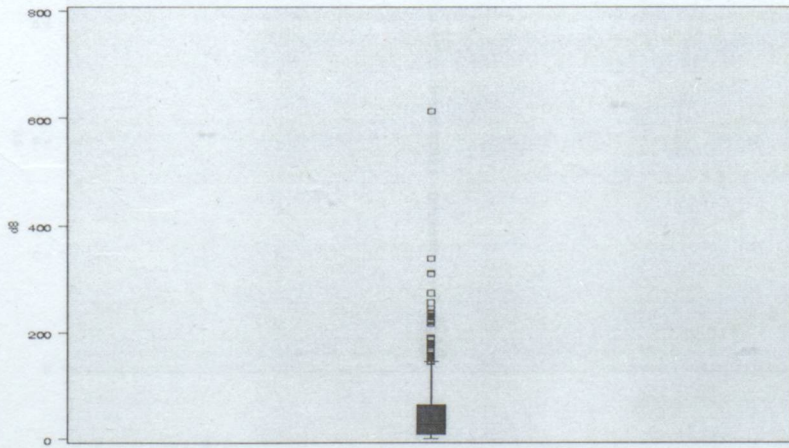
Şekil Ek 2.3 D5 Değişkenine İlişkin Kutu Diyagramı



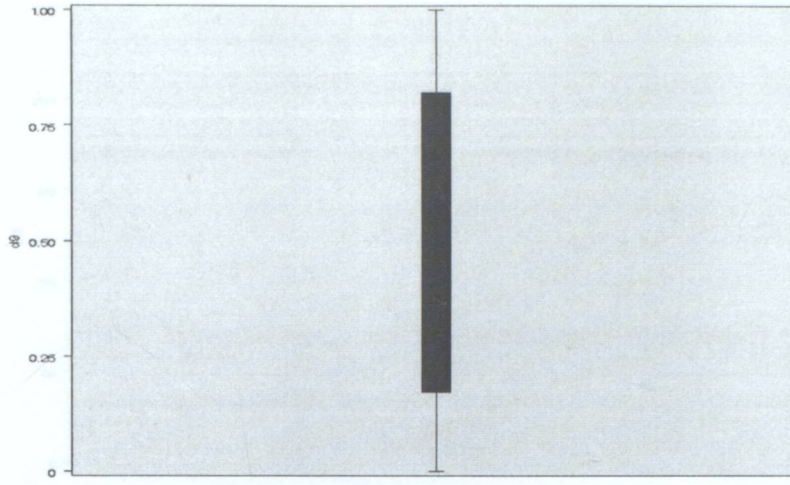
Şekil Ek 2.4 D6 Değişkenine İlişkin Kutu Diyagramı



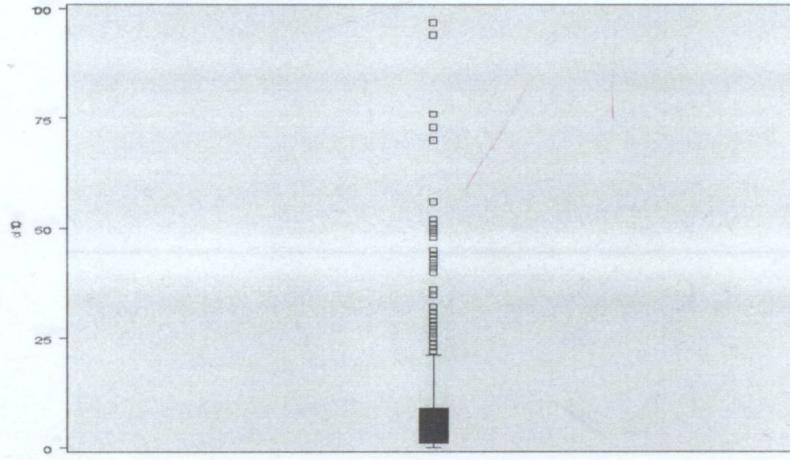
Şekil Ek 2.5 D7 Değişkenine İlişkin Kutu Diyagramı



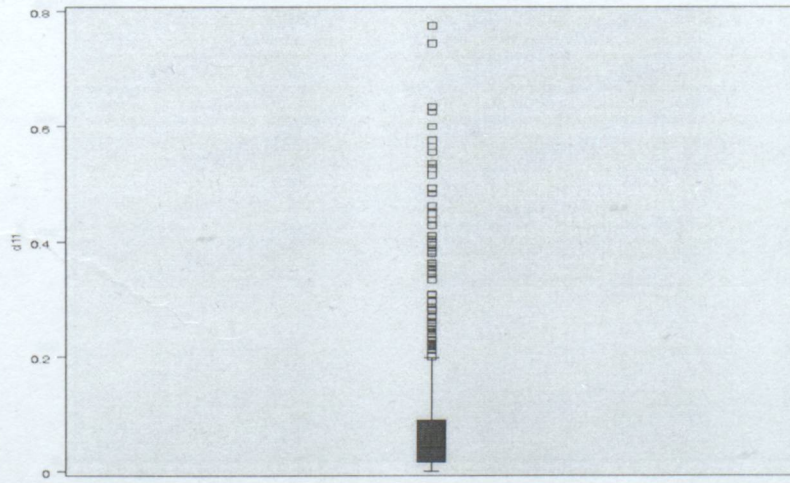
Şekil Ek 2.6 D8 Değişkenine İlişkin Kutu Diyagramı



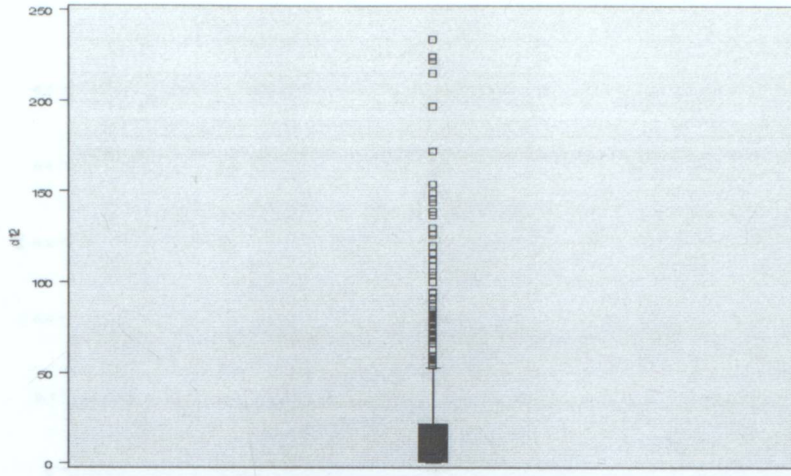
Şekil Ek 2.7 D9 Değişkenine İlişkin Kutu Diyagramı



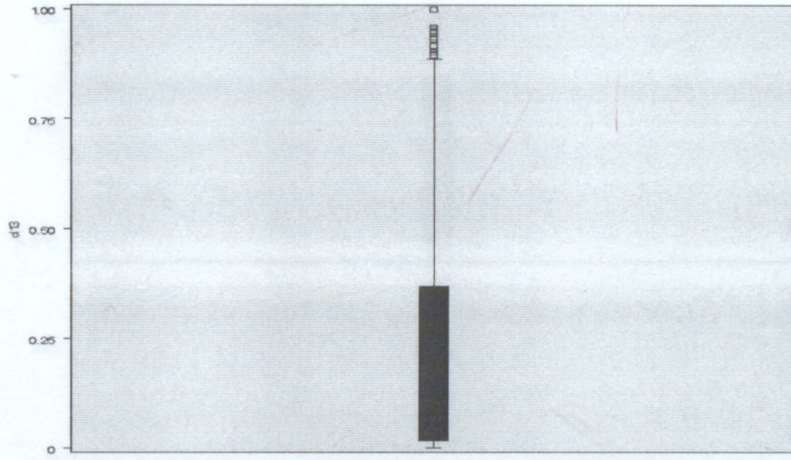
Şekil Ek 2.8 D10 Değişkenine İlişkin Kutu Diyagramı



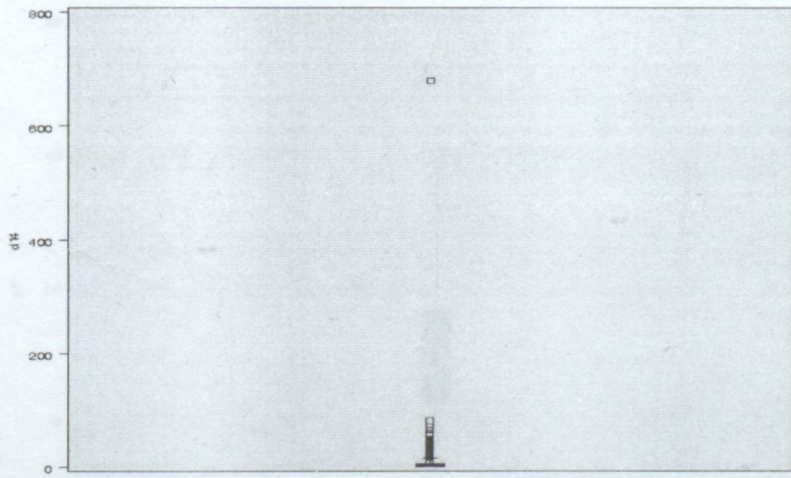
Şekil Ek 2.9 D11 Değişkenine İlişkin Kutu Diyagramı



Şekil Ek 2.10 D12 Değişkenine İlişkin Kutu Diyagramı

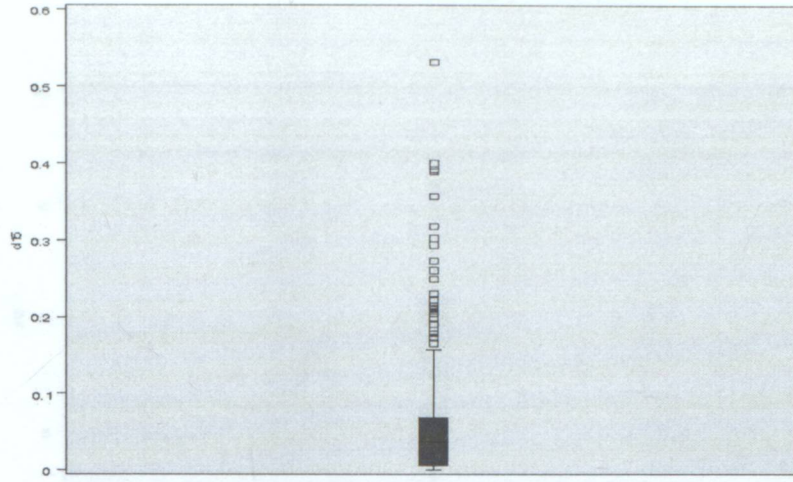


Şekil Ek 2.11 D13 Değişkenine İlişkin Kutu Diyagramı

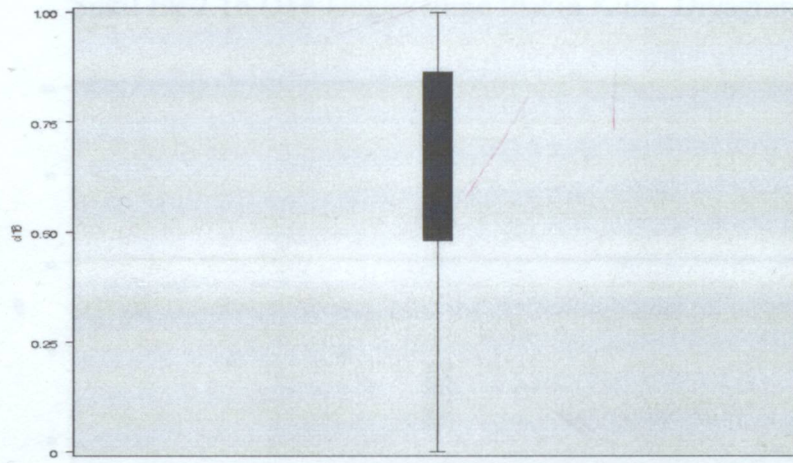


Şekil Ek 2.12 D14 Değişkenine İlişkin Kutu Diyagramı

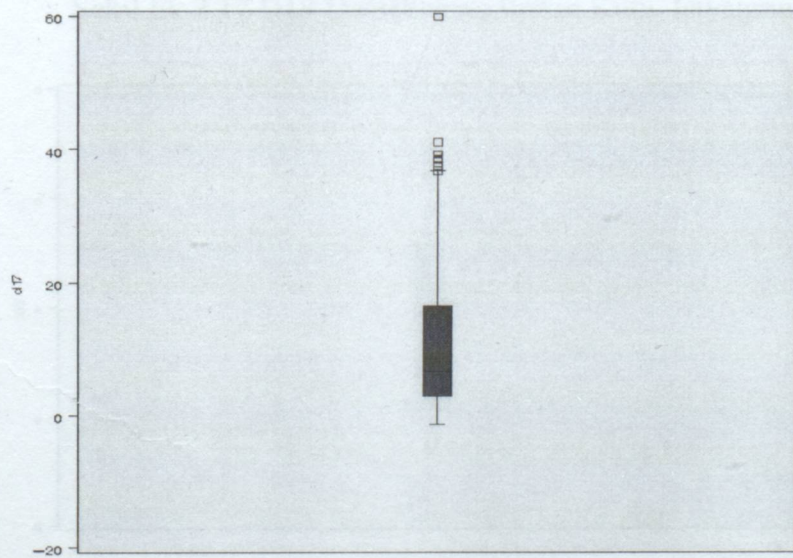
Şekil Ek 2.13 D17 Değişkenine İlişkin Kutu Diyagramı



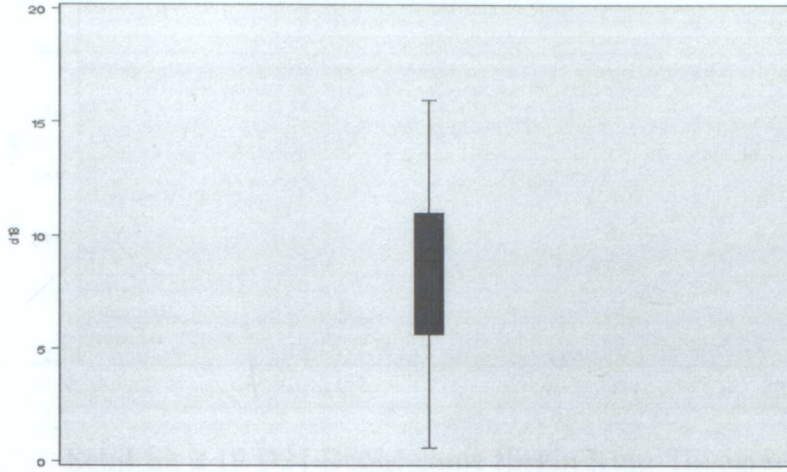
Şekil Ek 2.13 D15 Değişkenine İlişkin Kutu Diyagramı



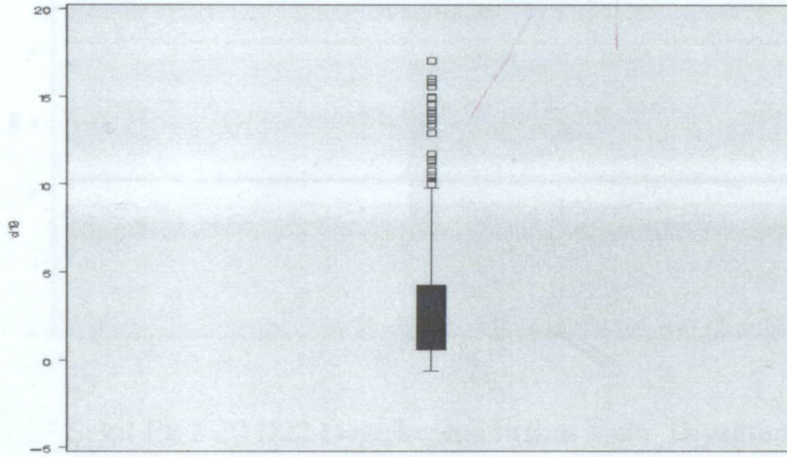
Şekil Ek 2.14 D16 Değişkenine İlişkin Kutu Diyagramı



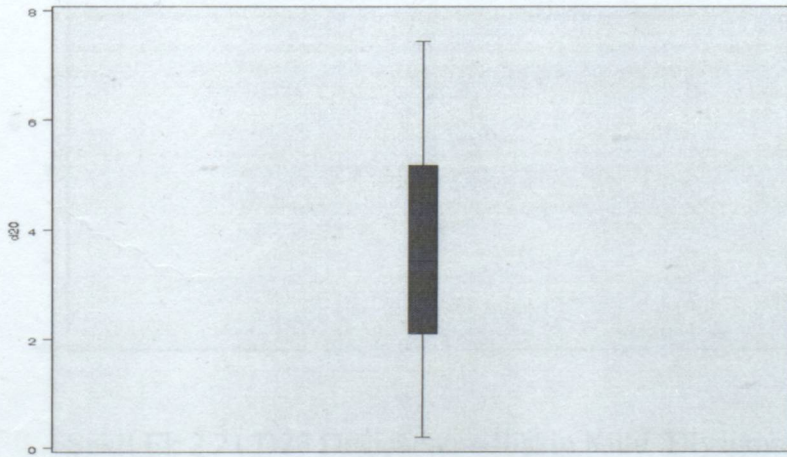
Şekil Ek 2.15 D17 Değişkenine İlişkin Kutu Diyagramı



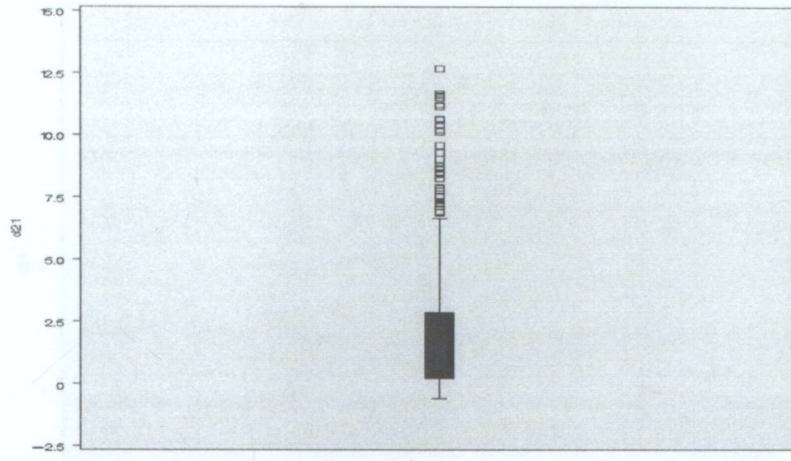
Şekil Ek 2.16 D18 Değişkenine İlişkin Kutu Diyagramı



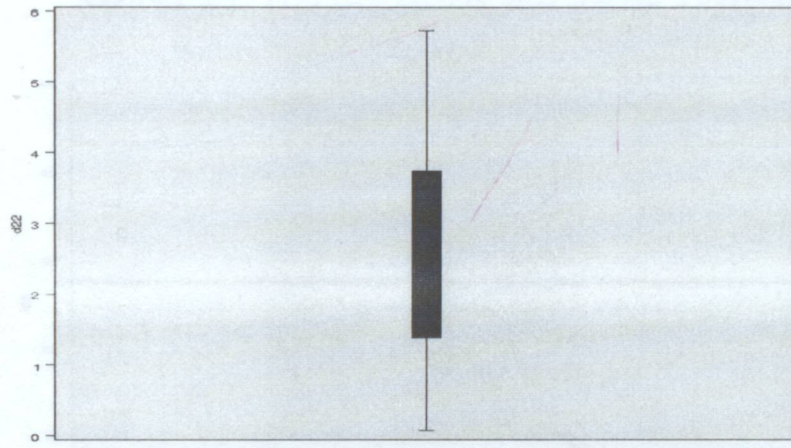
Şekil Ek 2.17 D19 Değişkenine İlişkin Kutu Diyagramı



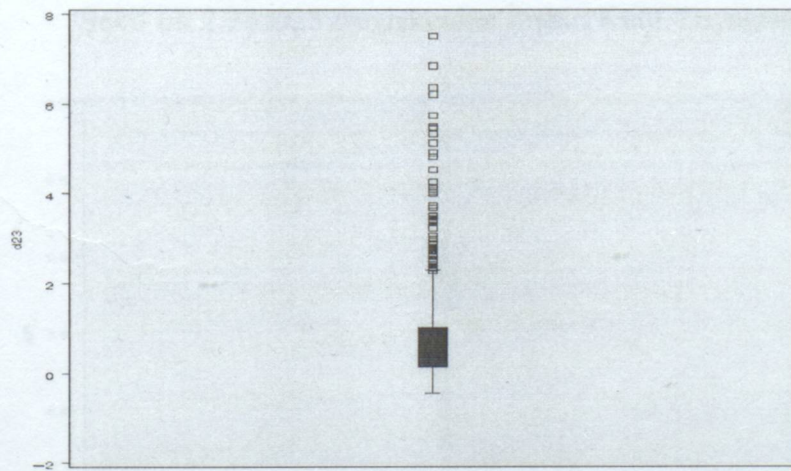
Şekil Ek 2.18 D20 Değişkenine İlişkin Kutu Diyagramı



Şekil Ek 2.19 D21 Değişkenine İlişkin Kutu Diyagramı

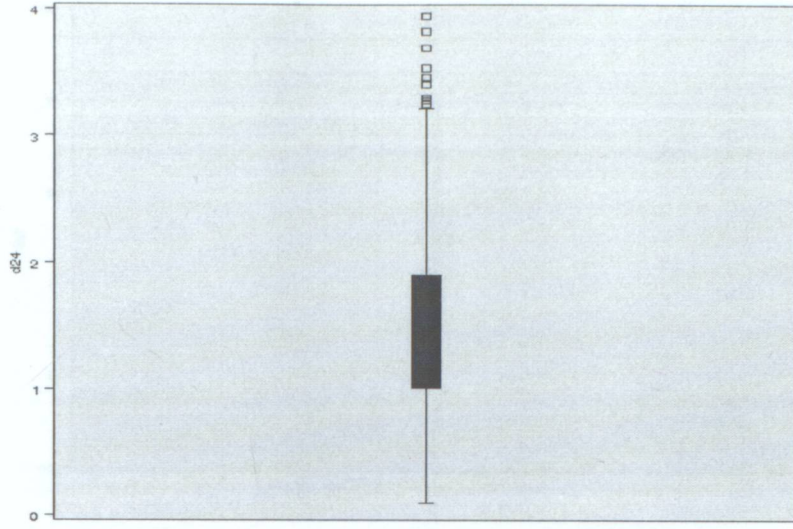


Şekil Ek 2.20 D22 Değişkenine İlişkin Kutu Diyagramı

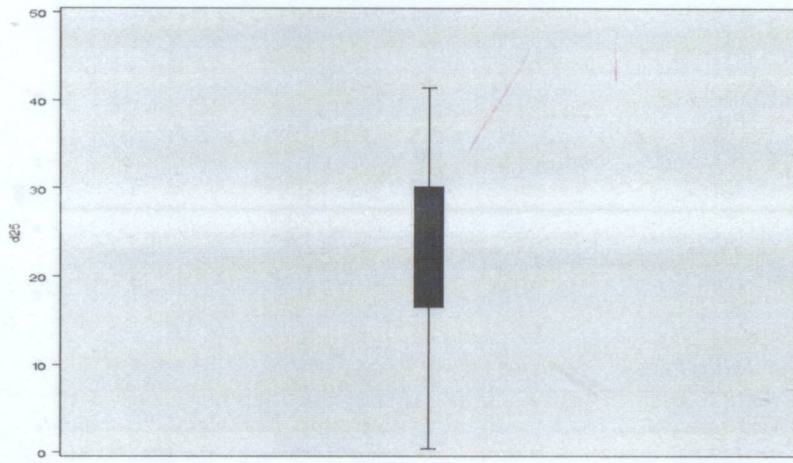


Şekil Ek 2.21 D23 Değişkenine İlişkin Kutu Diyagramı

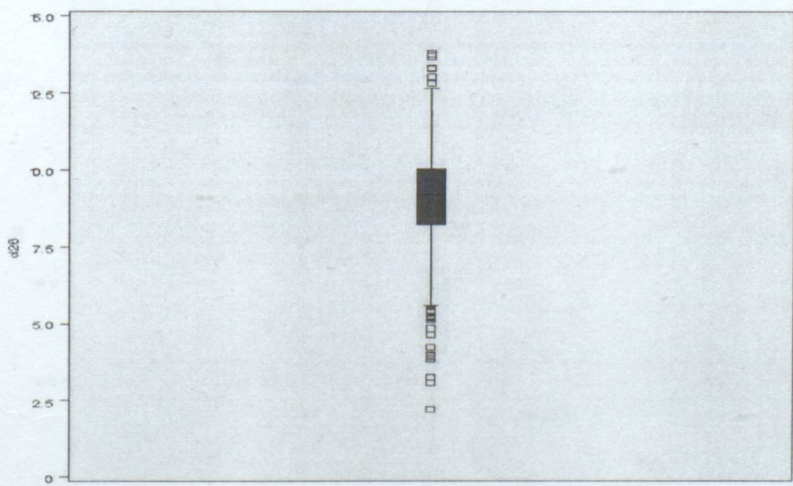
Şekil Ek 2.24 D26 Değişkenine İlişkin Kutu Diyagramı



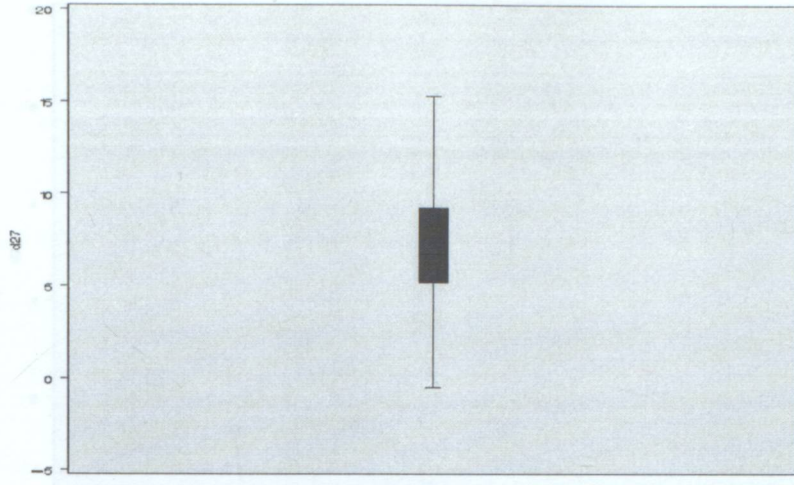
Şekil Ek 2.22 D24 Değişkenine İlişkin Kutu Diyagramı



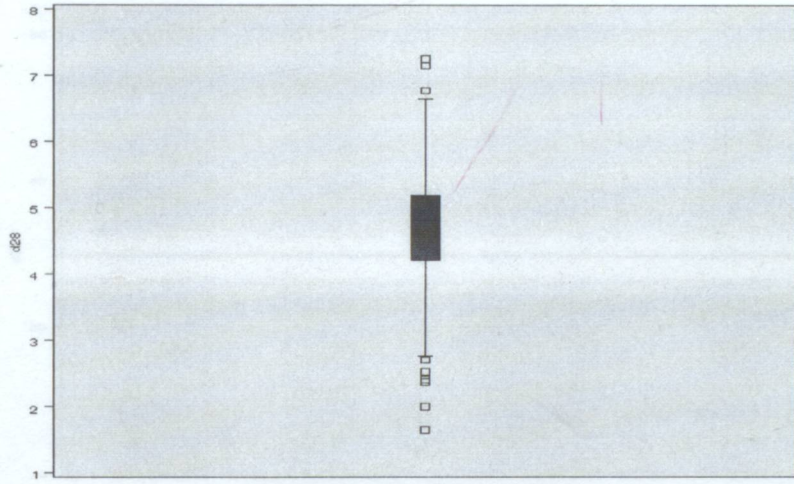
Şekil Ek 2.23 D25 Değişkenine İlişkin Kutu Diyagramı



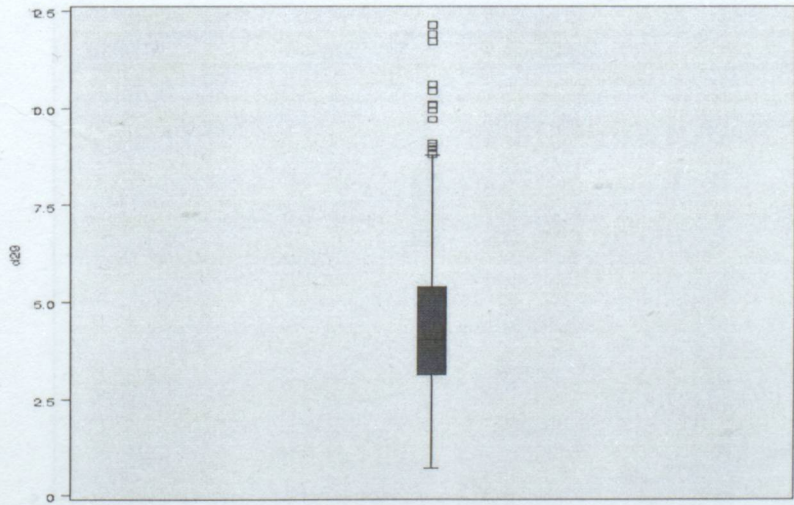
Şekil Ek 2.24 D26 Değişkenine İlişkin Kutu Diyagramı



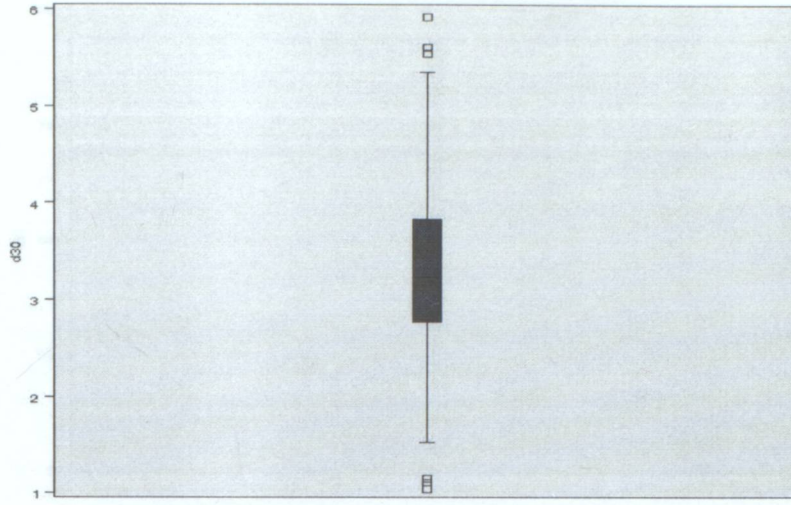
Şekil Ek 2.25 D27 Değişkenine İlişkin Kutu Diyagramı



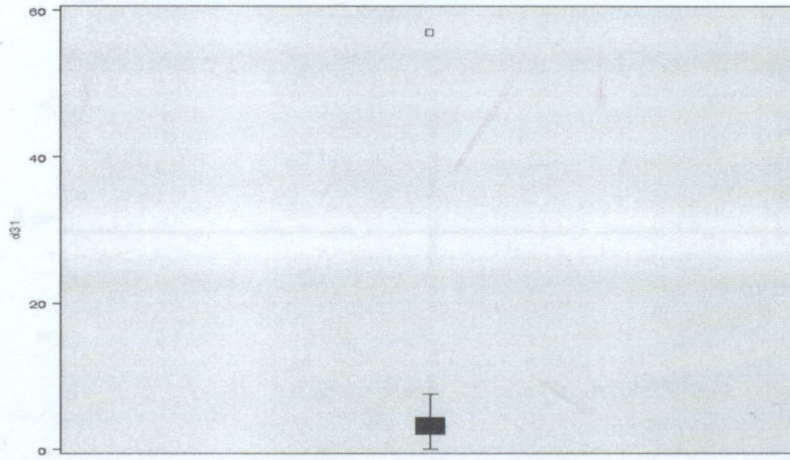
Şekil Ek 2.26 D28 Değişkenine İlişkin Kutu Diyagramı



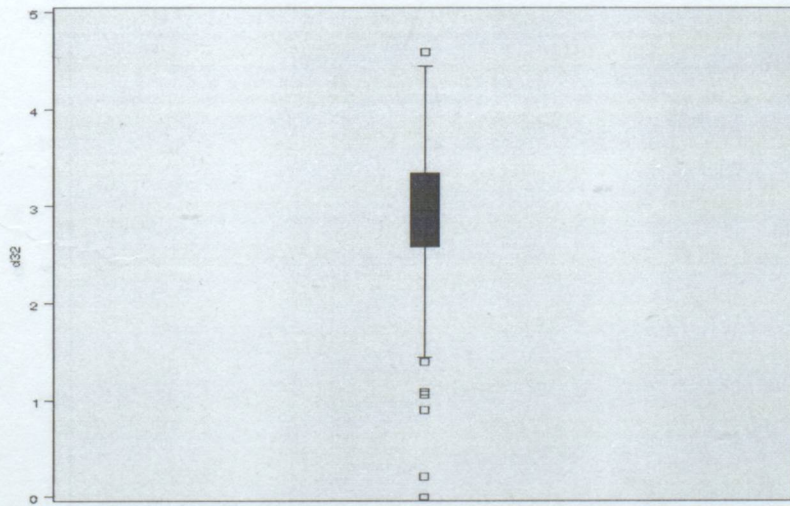
Şekil Ek 2.27 D29 Değişkenine İlişkin Kutu Diyagramı



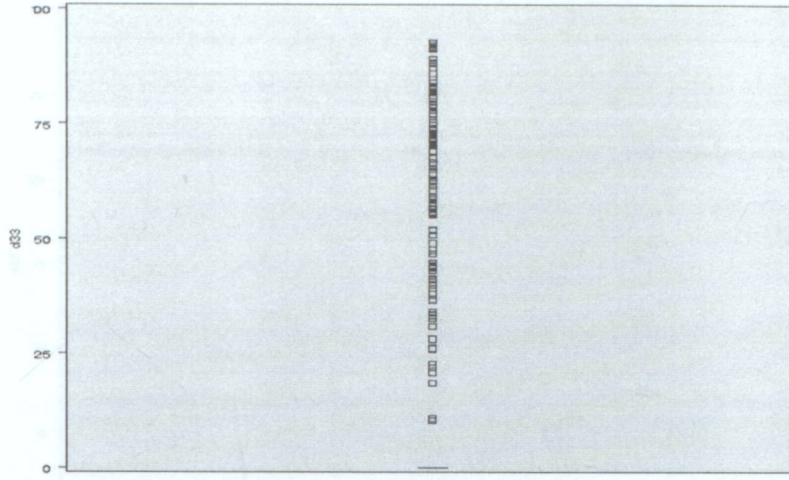
Şekil Ek 2.28 D30 Değişkenine İlişkin Kutu Diyagramı



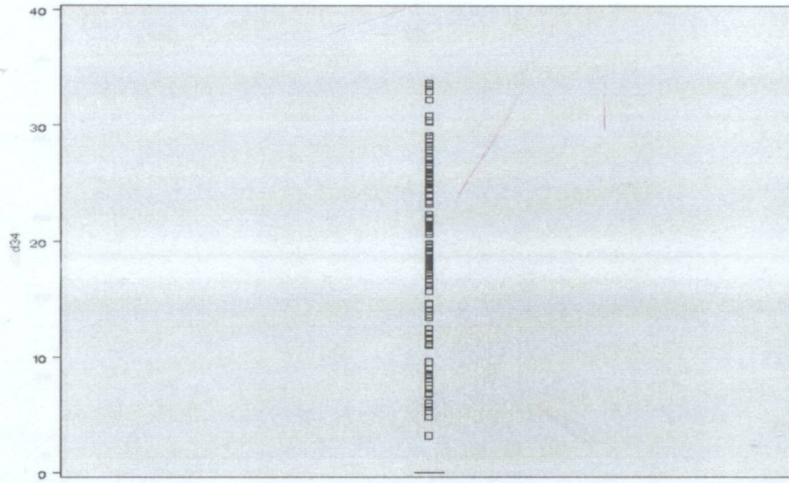
Şekil Ek 2.29 D31 Değişkenine İlişkin Kutu Diyagramı



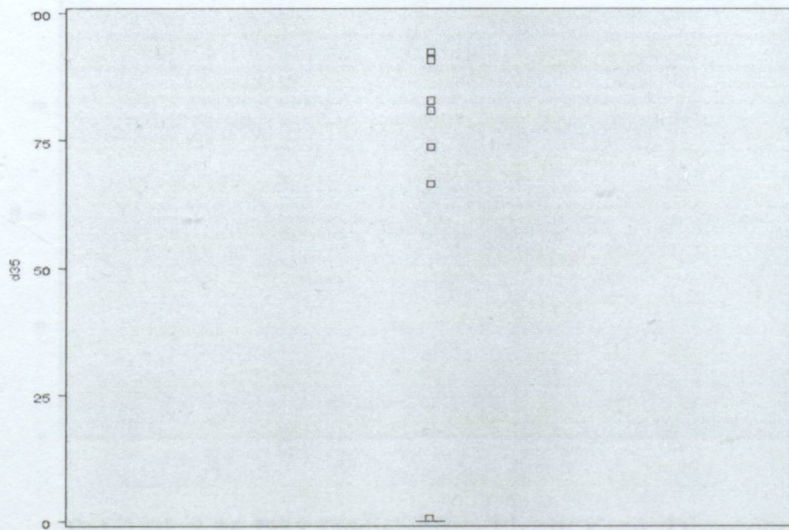
Şekil Ek 2.30 D32 Değişkenine İlişkin Kutu Diyagramı



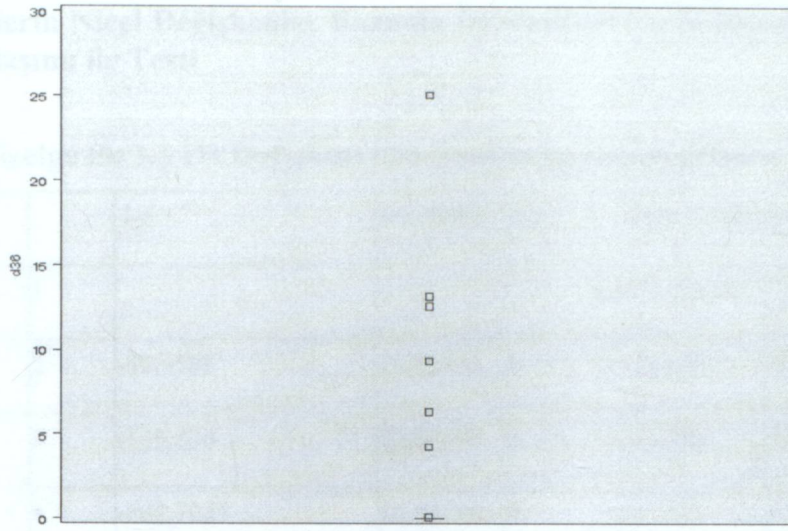
Şekil Ek 2.31 D33 Değişkenine İlişkin Kutu Diyagramı



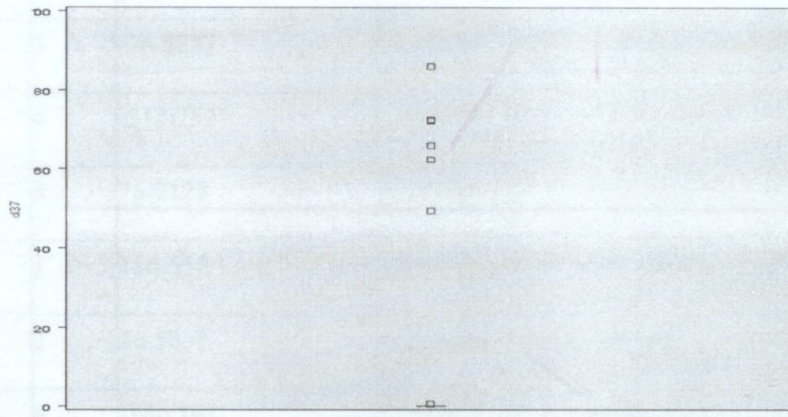
Şekil Ek 2.32 D34 Değişkenine İlişkin Kutu Diyagramı



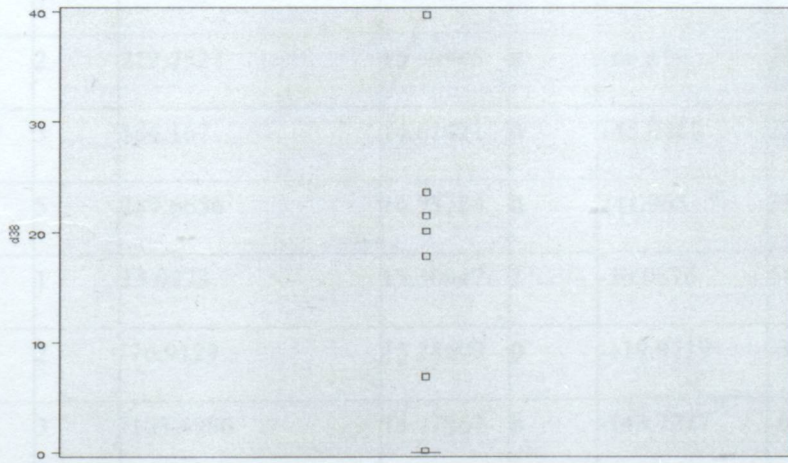
Şekil Ek 2.33 D35 Değişkenine İlişkin Kutu Diyagramı



Şekil Ek 2.34 D36 Değişkenine İlişkin Kutu Diyagramı



Şekil Ek 2.35 D37 Değişkenine İlişkin Kutu Diyagramı



Şekil Ek 2.36 D38 Değişkenine İlişkin Kutu Diyagramı

EK 3 Kümelerin Nicel Değişkenler Bazında İstatistiksel Farklılığının Bonferroni Güven Bölgesi Yaklaşımı ile Testi

Çizelge Ek 3.1 D3 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı

d3		I-J	Std. Error	Sig.	95% Confidence Interval	
I	J				Lower Bound	Upper Bound
1	2	-89.9502	13.81866	0	-128.8757	-51.0247
	3	-116.536	12.69645	0	-152.3003	-80.7716
	4	-302.7029	15.62196	0	-346.7081	-258.6977
	5	-13.0373	15.30947	1	-56.1623	30.0876
2	1	89.9502	13.81866	0	51.0247	128.8757
	3	-26.5857	12.66826	0.362	-62.2707	9.0992
	4	-212.7527	15.59905	0	-256.6933	-168.812
	5	76.9129	15.28609	0	33.8538	119.9719
3	1	116.536	12.69645	0	80.7716	152.3003
	2	26.5857	12.66826	0.362	-9.0992	62.2707
	4	-186.167	14.61421	0	-227.3334	-145.0005
	5	103.4986	14.27967	0	63.2745	143.7227
4	1	302.7029	15.62196	0	258.6977	346.7081
	2	212.7527	15.59905	0	168.812	256.6933
	3	186.167	14.61421	0	145.0005	227.3334
	5	289.6656	16.93384	0	241.965	337.3661
5	1	13.0373	15.30947	1	-30.0876	56.1623
	2	-76.9129	15.28609	0	-119.9719	-33.8538
	3	-103.4986	14.27967	0	-143.7227	-63.2745
	4	-289.6656	16.93384	0	-337.3661	-241.965

Çizelge Ek 3.2 D4 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı

d4		I-J	Std. Error	Sig.	95% Confidence Interval	
I	J				Lower Bound	Upper Bound
1	2	2.46	6.30824	1	-15.3096	20.2295
	3	-38.7147	5.79595	0	-55.0412	-22.3883
	4	-199.3958	7.13145	0	-219.4842	-179.3074
	5	-16.3066	6.9888	0.199	-35.9931	3.38
2	1	-2.46	6.30824	1	-20.2295	15.3096
	3	-41.1747	5.78308	0	-57.465	-24.8845
	4	-201.8558	7.12099	0	-221.9148	-181.7968
	5	-18.7665	6.97813	0.073	-38.4231	0.89
3	1	38.7147	5.79595	0	22.3883	55.0412
	2	41.1747	5.78308	0	24.8845	57.465
	4	-160.681	6.67141	0	-179.4736	-141.8885
	5	22.4082	6.51869	0.006	4.0458	40.7706
4	1	199.3958	7.13145	0	179.3074	219.4842
	2	201.8558	7.12099	0	181.7968	221.9148
	3	160.681	6.67141	0	141.8885	179.4736
	5	183.0892	7.73032	0	161.3138	204.8646
5	1	16.3066	6.9888	0.199	-3.38	35.9931
	2	18.7665	6.97813	0.073	-0.89	38.4231
	3	-22.4082	6.51869	0.006	-40.7706	-4.0458
	4	-183.0892	7.73032	0	-204.8646	-161.3138

Çizelge Ek 3.3 D5 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı

d5		I-J	Std. Error	Sig.	95% Confidence Interval	
I	J				Lower Bound	Upper Bound
1	2	0.145	0.01795	0	0.0945	0.1956
	3	-0.1577	0.0165	0	-0.2042	-0.1113
	4	-0.4897	0.0203	0	-0.5469	-0.4325
	5	-0.3404	0.01989	0	-0.3965	-0.2844
2	1	-0.145	0.01795	0	-0.1956	-0.0945
	3	-0.3028	0.01646	0	-0.3491	-0.2564
	4	-0.6347	0.02027	0	-0.6918	-0.5776
	5	-0.4855	0.01986	0	-0.5414	-0.4295
3	1	0.1577	0.0165	0	0.1113	0.2042
	2	0.3028	0.01646	0	0.2564	0.3491
	4	-0.3319	0.01899	0	-0.3854	-0.2785
	5	-0.1827	0.01855	0	-0.235	-0.1304
4	1	0.4897	0.0203	0	0.4325	0.5469
	2	0.6347	0.02027	0	0.5776	0.6918
	3	0.3319	0.01899	0	0.2785	0.3854
	5	0.1492	0.022	0	0.0873	0.2112
5	1	0.3404	0.01989	0	0.2844	0.3965
	2	0.4855	0.01986	0	0.4295	0.5414
	3	0.1827	0.01855	0	0.1304	0.235
	4	-0.1492	0.022	0	-0.2112	-0.0873

Çizelge Ek 3.4 D6 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı

d6		I-J	Std. Error	Sig.	95% Confidence Interval	
I	J				Lower Bound	Upper Bound
1	2	-0.6698	4.62786	1	-13.7059	12.3663
	3	-42.2903	4.25203	0	-54.2678	-30.3129
	4	-85.124	5.23178	0	-99.8613	-70.3867
	5	-4.1975	5.12713	1	-18.64	10.245
2	1	0.6698	4.62786	1	-12.3663	13.7059
	3	-41.6206	4.24259	0	-53.5714	-29.6697
	4	-84.4542	5.22411	0	-99.1699	-69.7385
	5	-3.5277	5.1193	1	-17.9482	10.8927
3	1	42.2903	4.25203	0	30.3129	54.2678
	2	41.6206	4.24259	0	29.6697	53.5714
	4	-42.8336	4.89428	0	-56.6202	-29.047
	5	38.0928	4.78225	0	24.6218	51.5638
4	1	85.124	5.23178	0	70.3867	99.8613
	2	84.4542	5.22411	0	69.7385	99.1699
	3	42.8336	4.89428	0	29.047	56.6202
	5	80.9264	5.67113	0	64.9516	96.9013
5	1	4.1975	5.12713	1	-10.245	18.64
	2	3.5277	5.1193	1	-10.8927	17.9482
	3	-38.0928	4.78225	0	-51.5638	-24.6218
	4	-80.9264	5.67113	0	-96.9013	-64.9516

Çizelge Ek 3.5 D7 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı

d7		I-J	Std. Error	Sig.	95% Confidence Interval	
I	J				Lower Bound	Upper Bound
1	2	0.2015	0.01261	0	0.1659	0.237
	3	-0.0752	0.01159	0	-0.1078	-0.0426
	4	-0.0011	0.01425	1	-0.0413	0.039
	5	-0.0162	0.01397	1	-0.0555	0.0232
2	1	-0.2015	0.01261	0	-0.237	-0.1659
	3	-0.2767	0.01156	0	-0.3092	-0.2441
	4	-0.2026	0.01423	0	-0.2427	-0.1625
	5	-0.2176	0.01395	0	-0.2569	-0.1783
3	1	0.0752	0.01159	0	0.0426	0.1078
	2	0.2767	0.01156	0	0.2441	0.3092
	4	0.0741	0.01334	0	0.0365	0.1116
	5	0.0591	0.01303	0	0.0224	0.0958
4	1	0.0011	0.01425	1	-0.039	0.0413
	2	0.2026	0.01423	0	0.1625	0.2427
	3	-0.0741	0.01334	0	-0.1116	-0.0365
	5	-0.015	0.01545	1	-0.0585	0.0285
5	1	0.0162	0.01397	1	-0.0232	0.0555
	2	0.2176	0.01395	0	0.1783	0.2569
	3	-0.0591	0.01303	0	-0.0958	-0.0224
	4	0.015	0.01545	1	-0.0285	0.0585

Çizelge Ek 3.6 D8 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı

d8		I-J	Std. Error	Sig.	95% Confidence Interval	
I	J				Lower Bound	Upper Bound
1	2	-91.7404	5.59182	0	-107.4919	-75.9889
	3	-35.5309	5.13771	0	-50.0032	-21.0586
	4	-18.1832	6.32154	0.042	-35.9902	-0.3761
	5	7.4667	6.19509	1	-9.9841	24.9175
2	1	91.7404	5.59182	0	75.9889	107.4919
	3	56.2096	5.1263	0	41.7694	70.6497
	4	73.5573	6.31227	0	55.7764	91.3382
	5	99.2072	6.18563	0	81.783	116.6313
3	1	35.5309	5.13771	0	21.0586	50.0032
	2	-56.2096	5.1263	0	-70.6497	-41.7694
	4	17.3477	5.91374	0.035	0.6894	34.006
	5	42.9976	5.77837	0	26.7206	59.2746
4	1	18.1832	6.32154	0.042	0.3761	35.9902
	2	-73.5573	6.31227	0	-91.3382	-55.7764
	3	-17.3477	5.91374	0.035	-34.006	-0.6894
	5	25.6499	6.8524	0.002	6.3475	44.9523
5	1	-7.4667	6.19509	1	-24.9175	9.9841
	2	-99.2072	6.18563	0	-116.6313	-81.783
	3	-42.9976	5.77837	0	-59.2746	-26.7206
	4	-25.6499	6.8524	0.002	-44.9523	-6.3475

Çizelge Ek 3.7 D9 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı

d9		I-J	Std. Error	Sig.	95% Confidence Interval	
I	J				Lower Bound	Upper Bound
1	2	-0.3465	0.02203	0	-0.4086	-0.2844
	3	0.2329	0.02024	0	0.1759	0.29
	4	0.4908	0.02491	0	0.4207	0.561
	5	0.3566	0.02441	0	0.2878	0.4253
2	1	0.3465	0.02203	0	0.2844	0.4086
	3	0.5794	0.0202	0	0.5225	0.6363
	4	0.8373	0.02487	0	0.7673	0.9074
	5	0.7031	0.02437	0	0.6344	0.7717
3	1	-0.2329	0.02024	0	-0.29	-0.1759
	2	-0.5794	0.0202	0	-0.6363	-0.5225
	4	0.2579	0.0233	0	0.1922	0.3235
	5	0.1236	0.02277	0	0.0595	0.1878
4	1	-0.4908	0.02491	0	-0.561	-0.4207
	2	-0.8373	0.02487	0	-0.9074	-0.7673
	3	-0.2579	0.0233	0	-0.3235	-0.1922
	5	-0.1342	0.027	0	-0.2103	-0.0582
5	1	-0.3566	0.02441	0	-0.4253	-0.2878
	2	-0.7031	0.02437	0	-0.7717	-0.6344
	3	-0.1236	0.02277	0	-0.1878	-0.0595
	4	0.1342	0.027	0	0.0582	0.2103

Çizelge Ek 3.8 D10 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı

d10		I-J	Std. Error	Sig.	95% Confidence Interval	
I	J				Lower Bound	Upper Bound
1	2	-0.8559	1.22651	1	-4.3108	2.599
	3	-4.4148	1.1269	0.001	-7.5892	-1.2405
	4	-18.8196	1.38656	0	-22.7254	-14.9139
	5	-1.3656	1.35883	1	-5.1933	2.462
2	1	0.8559	1.22651	1	-2.599	4.3108
	3	-3.5589	1.1244	0.016	-6.7262	-0.3916
	4	-17.9637	1.38453	0	-21.8638	-14.0637
	5	-0.5097	1.35675	1	-4.3315	3.3121
3	1	4.4148	1.1269	0.001	1.2405	7.5892
	2	3.5589	1.1244	0.016	0.3916	6.7262
	4	-14.4048	1.29712	0	-18.0587	-10.751
	5	3.0492	1.26743	0.164	-0.521	6.6193
4	1	18.8196	1.38656	0	14.9139	22.7254
	2	17.9637	1.38453	0	14.0637	21.8638
	3	14.4048	1.29712	0	10.751	18.0587
	5	17.454	1.503	0	13.2202	21.6878
5	1	1.3656	1.35883	1	-2.462	5.1933
	2	0.5097	1.35675	1	-3.3121	4.3315
	3	-3.0492	1.26743	0.164	-6.6193	0.521
	4	-17.454	1.503	0	-21.6878	-13.2202

Çizelge Ek 3.9 D11 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı

d11		I-J	Std. Error	Sig.	95% Confidence Interval	
I	J				Lower Bound	Upper Bound
1	2	0.0879	0.01353	0	0.0498	0.126
	3	0.058	0.01243	0	0.023	0.093
	4	0.0614	0.0153	0.001	0.0183	0.1045
	5	0.0039	0.01499	1	-0.0383	0.0461
2	1	-0.0879	0.01353	0	-0.126	-0.0498
	3	-0.0299	0.0124	0.162	-0.0648	0.005
	4	-0.0265	0.01527	0.837	-0.0695	0.0166
	5	-0.084	0.01497	0	-0.1262	-0.0418
3	1	-0.058	0.01243	0	-0.093	-0.023
	2	0.0299	0.0124	0.162	-0.005	0.0648
	4	0.0034	0.01431	1	-0.0369	0.0437
	5	-0.0541	0.01398	0.001	-0.0935	-0.0147
4	1	-0.0614	0.0153	0.001	-0.1045	-0.0183
	2	0.0265	0.01527	0.837	-0.0166	0.0695
	3	-0.0034	0.01431	1	-0.0437	0.0369
	5	-0.0575	0.01658	0.006	-0.1042	-0.0108
5	1	-0.0039	0.01499	1	-0.0461	0.0383
	2	0.084	0.01497	0	0.0418	0.1262
	3	0.0541	0.01398	0.001	0.0147	0.0935
	4	0.0575	0.01658	0.006	0.0108	0.1042

Çizelge Ek 3.10 D12 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı

d12		I-J	Std. Error	Sig.	95% Confidence Interval	
I	J				Lower Bound	Upper Bound
1	2	-60.4841	2.88972	0	-68.6241	-52.3441
	3	-4.6363	2.65504	0.813	-12.1152	2.8427
	4	5.0232	3.26682	1	-4.1791	14.2254
	5	5.9756	3.20147	0.624	-3.0425	14.9938
2	1	60.4841	2.88972	0	52.3441	68.6241
	3	55.8478	2.64915	0	48.3855	63.3102
	4	65.5073	3.26203	0	56.3186	74.696
	5	66.4598	3.19658	0	57.4554	75.4641
3	1	4.6363	2.65504	0.813	-2.8427	12.1152
	2	-55.8478	2.64915	0	-63.3102	-48.3855
	4	9.6595	3.05608	0.016	1.0509	18.2681
	5	10.6119	2.98612	0.004	2.2004	19.0235
4	1	-5.0232	3.26682	1	-14.2254	4.1791
	2	-65.5073	3.26203	0	-74.696	-56.3186
	3	-9.6595	3.05608	0.016	-18.2681	-1.0509
	5	0.9525	3.54115	1	-9.0225	10.9275
5	1	-5.9756	3.20147	0.624	-14.9938	3.0425
	2	-66.4598	3.19658	0	-75.4641	-57.4554
	3	-10.6119	2.98612	0.004	-19.0235	-2.2004
	4	-0.9525	3.54115	1	-10.9275	9.0225

Çizelge Ek 3.11 D13 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı

d13		I-J	Std. Error	Sig.	95% Confidence Interval	
I	J				Lower Bound	Upper Bound
1	2	-0.318	0.01909	0	-0.3718	-0.2642
	3	0.1788	0.01754	0	0.1294	0.2282
	4	0.2613	0.02158	0	0.2005	0.3221
	5	0.2244	0.02115	0	0.1649	0.284
2	1	0.318	0.01909	0	0.2642	0.3718
	3	0.4969	0.0175	0	0.4476	0.5462
	4	0.5794	0.02155	0	0.5187	0.6401
	5	0.5425	0.02112	0	0.483	0.602
3	1	-0.1788	0.01754	0	-0.2282	-0.1294
	2	-0.4969	0.0175	0	-0.5462	-0.4476
	4	0.0825	0.02019	0	0.0256	0.1394
	5	0.0456	0.01973	0.211	-0.01	0.1012
4	1	-0.2613	0.02158	0	-0.3221	-0.2005
	2	-0.5794	0.02155	0	-0.6401	-0.5187
	3	-0.0825	0.02019	0	-0.1394	-0.0256
	5	-0.0369	0.0234	1	-0.1028	0.029
5	1	-0.2244	0.02115	0	-0.284	-0.1649
	2	-0.5425	0.02112	0	-0.602	-0.483
	3	-0.0456	0.01973	0.211	-0.1012	0.01
	4	0.0369	0.0234	1	-0.029	0.1028

Çizelge Ek 3.12 D14 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı

d14		I-J	Std. Error	Sig.	95% Confidence Interval	
I	J				Lower Bound	Upper Bound
1	2	-0.4975	3.33054	1	-9.8792	8.8843
	3	-5.7437	3.06007	0.61	-14.3636	2.8761
	4	-24.9756	3.76517	0	-35.5816	-14.3696
	5	-2.2595	3.68985	1	-12.6533	8.1344
2	1	0.4975	3.33054	1	-8.8843	9.8792
	3	-5.2462	3.05327	0.862	-13.8469	3.3545
	4	-24.4781	3.75965	0	-35.0686	-13.8877
	5	-1.762	3.68422	1	-12.14	8.616
3	1	5.7437	3.06007	0.61	-2.8761	14.3636
	2	5.2462	3.05327	0.862	-3.3545	13.8469
	4	-19.2319	3.52228	0	-29.1537	-9.3101
	5	3.4842	3.44165	1	-6.2105	13.1789
4	1	24.9756	3.76517	0	14.3696	35.5816
	2	24.4781	3.75965	0	13.8877	35.0686
	3	19.2319	3.52228	0	9.3101	29.1537
	5	22.7161	4.08136	0	11.2194	34.2128
5	1	2.2595	3.68985	1	-8.1344	12.6533
	2	1.762	3.68422	1	-8.616	12.14
	3	-3.4842	3.44165	1	-13.1789	6.2105
	4	-22.7161	4.08136	0	-34.2128	-11.2194

Çizelge Ek 3.13 D15 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı

d15		I-J	Std. Error	Sig.	95% Confidence Interval	
I	J				Lower Bound	Upper Bound
1	2	0.0341	0.00737	0	0.0133	0.0548
	3	0.0016	0.00678	1	-0.0174	0.0207
	4	-0.0119	0.00834	1	-0.0354	0.0116
	5	-0.0523	0.00817	0	-0.0753	-0.0293
2	1	-0.0341	0.00737	0	-0.0548	-0.0133
	3	-0.0324	0.00676	0	-0.0515	-0.0134
	4	-0.046	0.00832	0	-0.0694	-0.0225
	5	-0.0863	0.00816	0	-0.1093	-0.0634
3	1	-0.0016	0.00678	1	-0.0207	0.0174
	2	0.0324	0.00676	0	0.0134	0.0515
	4	-0.0136	0.0078	0.827	-0.0355	0.0084
	5	-0.0539	0.00762	0	-0.0754	-0.0325
4	1	0.0119	0.00834	1	-0.0116	0.0354
	2	0.046	0.00832	0	0.0225	0.0694
	3	0.0136	0.0078	0.827	-0.0084	0.0355
	5	-0.0404	0.00904	0	-0.0658	-0.0149
5	1	0.0523	0.00817	0	0.0293	0.0753
	2	0.0863	0.00816	0	0.0634	0.1093
	3	0.0539	0.00762	0	0.0325	0.0754
	4	0.0404	0.00904	0	0.0149	0.0658

Çizelge Ek 3.14 D16 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı

d16		I-J	Std. Error	Sig.	95% Confidence Interval	
I	J				Lower Bound	Upper Bound
1	2	0.1961	0.02084	0	0.1374	0.2548
	3	-0.2385	0.01914	0	-0.2924	-0.1845
	4	-0.3109	0.02355	0	-0.3772	-0.2445
	5	-0.1761	0.02308	0	-0.2411	-0.1111
2	1	-0.1961	0.02084	0	-0.2548	-0.1374
	3	-0.4345	0.0191	0	-0.4883	-0.3807
	4	-0.5069	0.02352	0	-0.5732	-0.4407
	5	-0.3722	0.02305	0	-0.4371	-0.3072
3	1	0.2385	0.01914	0	0.1845	0.2924
	2	0.4345	0.0191	0	0.3807	0.4883
	4	-0.0724	0.02203	0.011	-0.1345	-0.0103
	5	0.0624	0.02153	0.039	0.0017	0.123
4	1	0.3109	0.02355	0	0.2445	0.3772
	2	0.5069	0.02352	0	0.4407	0.5732
	3	0.0724	0.02203	0.011	0.0103	0.1345
	5	0.1348	0.02553	0	0.0629	0.2067
5	1	0.1761	0.02308	0	0.1111	0.2411
	2	0.3722	0.02305	0	0.3072	0.4371
	3	-0.0624	0.02153	0.039	-0.123	-0.0017
	4	-0.1348	0.02553	0	-0.2067	-0.0629

Çizelge Ek 3.15 D17 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı

d17		I-J	Std. Error	Sig.	95% Confidence Interval	
I	J				Lower Bound	Upper Bound
1	2	-11.9989	0.49785	0	-13.4013	-10.5965
	3	6.4551	0.45742	0	5.1666	7.7436
	4	11.6119	0.56281	0	10.0265	13.1973
	5	11.2584	0.55156	0	9.7048	12.8121
2	1	11.9989	0.49785	0	10.5965	13.4013
	3	18.454	0.4564	0	17.1683	19.7396
	4	23.6108	0.56199	0	22.0277	25.1938
	5	23.2573	0.55071	0	21.706	24.8086
3	1	-6.4551	0.45742	0	-7.7436	-5.1666
	2	-18.454	0.4564	0	-19.7396	-17.1683
	4	5.1568	0.52651	0	3.6737	6.6399
	5	4.8033	0.51446	0	3.3542	6.2525
4	1	-11.6119	0.56281	0	-13.1973	-10.0265
	2	-23.6108	0.56199	0	-25.1938	-22.0277
	3	-5.1568	0.52651	0	-6.6399	-3.6737
	5	-0.3535	0.61008	1	-2.072	1.365
5	1	-11.2584	0.55156	0	-12.8121	-9.7048
	2	-23.2573	0.55071	0	-24.8086	-21.706
	3	-4.8033	0.51446	0	-6.2525	-3.3542
	4	0.3535	0.61008	1	-1.365	2.072

Çizelge Ek 3.16 D18 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı

d18		I-J	Std. Error	Sig.	95% Confidence Interval	
I	J				Lower Bound	Upper Bound
1	2	-0.0149	0.23237	1	-0.6695	0.6397
	3	2.0664	0.2135	0	1.465	2.6678
	4	6.9309	0.2627	0	6.1909	7.6709
	5	7.0883	0.25744	0	6.3631	7.8135
2	1	0.0149	0.23237	1	-0.6397	0.6695
	3	2.0813	0.21303	0	1.4813	2.6814
	4	6.9458	0.26231	0	6.2069	7.6847
	5	7.1032	0.25705	0	6.3791	7.8273
3	1	-2.0664	0.2135	0	-2.6678	-1.465
	2	-2.0813	0.21303	0	-2.6814	-1.4813
	4	4.8645	0.24575	0	4.1722	5.5567
	5	5.0219	0.24013	0	4.3455	5.6983
4	1	-6.9309	0.2627	0	-7.6709	-6.1909
	2	-6.9458	0.26231	0	-7.6847	-6.2069
	3	-4.8645	0.24575	0	-5.5567	-4.1722
	5	0.1574	0.28476	1	-0.6447	0.9595
5	1	-7.0883	0.25744	0	-7.8135	-6.3631
	2	-7.1032	0.25705	0	-7.8273	-6.3791
	3	-5.0219	0.24013	0	-5.6983	-4.3455
	4	-0.1574	0.28476	1	-0.9595	0.6447

Çizelge Ek 3.17 D19 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı

d19		I-J	Std. Error	Sig.	95% Confidence Interval	
I	J				Lower Bound	Upper Bound
1	2	-3.9461	0.20882	0	-4.5343	-3.3578
	3	2.0463	0.19186	0	1.5059	2.5868
	4	3.217	0.23607	0	2.5521	3.882
	5	3.36	0.23134	0	2.7083	4.0116
2	1	3.9461	0.20882	0	3.3578	4.5343
	3	5.9924	0.19143	0	5.4532	6.5316
	4	7.1631	0.23572	0	6.4991	7.8271
	5	7.306	0.23099	0	6.6553	7.9567
3	1	-2.0463	0.19186	0	-2.5868	-1.5059
	2	-5.9924	0.19143	0	-6.5316	-5.4532
	4	1.1707	0.22084	0	0.5486	1.7928
	5	1.3136	0.21578	0	0.7058	1.9214
4	1	-3.217	0.23607	0	-3.882	-2.5521
	2	-7.1631	0.23572	0	-7.8271	-6.4991
	3	-1.1707	0.22084	0	-1.7928	-0.5486
	5	0.1429	0.25589	1	-0.5779	0.8637
5	1	-3.36	0.23134	0	-4.0116	-2.7083
	2	-7.306	0.23099	0	-7.9567	-6.6553
	3	-1.3136	0.21578	0	-1.9214	-0.7058
	4	-0.1429	0.25589	1	-0.8637	0.5779

Çizelge Ek 3.18 D20 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı

d20		I-J	Std. Error	Sig.	95% Confidence Interval	
I	J				Lower Bound	Upper Bound
1	2	-1.2617	0.10109	0	-1.5465	-0.9769
	3	1.3676	0.09288	0	1.1059	1.6292
	4	2.8376	0.11428	0	2.5157	3.1596
	5	3.1994	0.11199	0	2.8839	3.5148
2	1	1.2617	0.10109	0	0.9769	1.5465
	3	2.6293	0.09267	0	2.3682	2.8903
	4	4.0993	0.11411	0	3.7779	4.4208
	5	4.4611	0.11182	0	4.1461	4.776
3	1	-1.3676	0.09288	0	-1.6292	-1.1059
	2	-2.6293	0.09267	0	-2.8903	-2.3682
	4	1.4701	0.10691	0	1.1689	1.7712
	5	1.8318	0.10446	0	1.5375	2.126
4	1	-2.8376	0.11428	0	-3.1596	-2.5157
	2	-4.0993	0.11411	0	-4.4208	-3.7779
	3	-1.4701	0.10691	0	-1.7712	-1.1689
	5	0.3617	0.12388	0.036	0.0128	0.7107
5	1	-3.1994	0.11199	0	-3.5148	-2.8839
	2	-4.4611	0.11182	0	-4.776	-4.1461
	3	-1.8318	0.10446	0	-2.126	-1.5375
	4	-0.3617	0.12388	0.036	-0.7107	-0.0128

Çizelge Ek 3.19 D21 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı

d21		I-J	Std. Error	Sig.	95% Confidence Interval	
I	J				Lower Bound	Upper Bound
1	2	-3.1991	0.15691	0	-3.6411	-2.7571
	3	1.4264	0.14432	0	1.0199	1.8329
	4	2.2102	0.17739	0	1.7105	2.7099
	5	2.2423	0.17384	0	1.7526	2.732
2	1	3.1991	0.15691	0	2.7571	3.6411
	3	4.6255	0.144	0	4.2199	5.0311
	4	5.4093	0.17713	0	4.9104	5.9083
	5	5.4414	0.17358	0	4.9525	5.9304
3	1	-1.4264	0.14432	0	-1.8329	-1.0199
	2	-4.6255	0.144	0	-5.0311	-4.2199
	4	0.7838	0.16608	0	0.316	1.2517
	5	0.8159	0.16228	0	0.3588	1.273
4	1	-2.2102	0.17739	0	-2.7099	-1.7105
	2	-5.4093	0.17713	0	-5.9083	-4.9104
	3	-0.7838	0.16608	0	-1.2517	-0.316
	5	0.0321	0.19229	1	-0.5096	0.5737
5	1	-2.2423	0.17384	0	-2.732	-1.7526
	2	-5.4414	0.17358	0	-5.9304	-4.9525
	3	-0.8159	0.16228	0	-1.273	-0.3588
	4	-0.0321	0.19229	1	-0.5737	0.5096

Çizelge Ek 3.20 D22 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı

d22		I-J	Std. Error	Sig.	95% Confidence Interval	
I	J				Lower Bound	Upper Bound
1	2	-1.1294	0.08164	0	-1.3594	-0.8994
	3	1.0833	0.07501	0	0.872	1.2946
	4	2.0135	0.0923	0	1.7535	2.2735
	5	2.3354	0.09045	0	2.0806	2.5902
2	1	1.1294	0.08164	0	0.8994	1.3594
	3	2.2127	0.07485	0	2.0018	2.4235
	4	3.1429	0.09216	0	2.8833	3.4025
	5	3.4648	0.09031	0	3.2104	3.7192
3	1	-1.0833	0.07501	0	-1.2946	-0.872
	2	-2.2127	0.07485	0	-2.4235	-2.0018
	4	0.9302	0.08634	0	0.687	1.1735
	5	1.2521	0.08437	0	1.0144	1.4897
4	1	-2.0135	0.0923	0	-2.2735	-1.7535
	2	-3.1429	0.09216	0	-3.4025	-2.8833
	3	-0.9302	0.08634	0	-1.1735	-0.687
	5	0.3218	0.10005	0.014	0.04	0.6037
5	1	-2.3354	0.09045	0	-2.5902	-2.0806
	2	-3.4648	0.09031	0	-3.7192	-3.2104
	3	-1.2521	0.08437	0	-1.4897	-1.0144
	4	-0.3218	0.10005	0.014	-0.6037	-0.04

Çizelge Ek 3.21 D23 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı

d23		I-J	Std. Error	Sig.	95% Confidence Interval	
I	J				Lower Bound	Upper Bound
1	2	-1.5141	0.08406	0	-1.7509	-1.2773
	3	0.4401	0.07723	0	0.2226	0.6577
	4	0.6838	0.09503	0	0.4161	0.9515
	5	0.6894	0.09312	0	0.4271	0.9517
2	1	1.5141	0.08406	0	1.2773	1.7509
	3	1.9542	0.07706	0	1.7372	2.1713
	4	2.1979	0.09489	0	1.9306	2.4652
	5	2.2035	0.09298	0	1.9415	2.4654
3	1	-0.4401	0.07723	0	-0.6577	-0.2226
	2	-1.9542	0.07706	0	-2.1713	-1.7372
	4	0.2437	0.0889	0.063	-0.0068	0.4941
	5	0.2492	0.08686	0.043	0.0045	0.4939
4	1	-0.6838	0.09503	0	-0.9515	-0.4161
	2	-2.1979	0.09489	0	-2.4652	-1.9306
	3	-0.2437	0.0889	0.063	-0.4941	0.0068
	5	0.0056	0.10301	1	-0.2846	0.2957
5	1	-0.6894	0.09312	0	-0.9517	-0.4271
	2	-2.2035	0.09298	0	-2.4654	-1.9415
	3	-0.2492	0.08686	0.043	-0.4939	-0.0045
	4	-0.0056	0.10301	1	-0.2957	0.2846

Çizelge Ek 3.22 D24 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı

d24		I-J	Std. Error	Sig.	95% Confidence Interval	
I	J				Lower Bound	Upper Bound
1	2	-0.9235	0.04905	0	-1.0616	-0.7853
	3	0.407	0.04506	0	0.2801	0.534
	4	0.5203	0.05545	0	0.3641	0.6765
	5	0.7974	0.05434	0	0.6443	0.9505
2	1	0.9235	0.04905	0	0.7853	1.0616
	3	1.3305	0.04496	0	1.2039	1.4572
	4	1.4438	0.05537	0	1.2878	1.5997
	5	1.7209	0.05426	0	1.5681	1.8737
3	1	-0.407	0.04506	0	-0.534	-0.2801
	2	-1.3305	0.04496	0	-1.4572	-1.2039
	4	0.1132	0.05187	0.294	-0.0329	0.2593
	5	0.3904	0.05068	0	0.2476	0.5331
4	1	-0.5203	0.05545	0	-0.6765	-0.3641
	2	-1.4438	0.05537	0	-1.5997	-1.2878
	3	-0.1132	0.05187	0.294	-0.2593	0.0329
	5	0.2771	0.0601	0	0.1078	0.4464
5	1	-0.7974	0.05434	0	-0.9505	-0.6443
	2	-1.7209	0.05426	0	-1.8737	-1.5681
	3	-0.3904	0.05068	0	-0.5331	-0.2476
	4	-0.2771	0.0601	0	-0.4464	-0.1078

Çizelge Ek 3.23 D25 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı

d25		I-J	Std. Error	Sig.	95% Confidence Interval	
I	J				Lower Bound	Upper Bound
1	2	-8.7304	0.64366	0	-10.5435	-6.9173
	3	3.8827	0.59139	0	2.2168	5.5485
	4	9.8385	0.72765	0	7.7888	11.8882
	5	5.5814	0.7131	0	3.5727	7.5901
2	1	8.7304	0.64366	0	6.9173	10.5435
	3	12.6131	0.59007	0	10.9509	14.2752
	4	18.5689	0.72659	0	16.5222	20.6156
	5	14.3118	0.71201	0	12.3062	16.3175
3	1	-3.8827	0.59139	0	-5.5485	-2.2168
	2	-12.6131	0.59007	0	-14.2752	-10.9509
	4	5.9558	0.68071	0	4.0383	7.8733
	5	1.6988	0.66513	0.109	-0.1748	3.5724
4	1	-9.8385	0.72765	0	-11.8882	-7.7888
	2	-18.5689	0.72659	0	-20.6156	-16.5222
	3	-5.9558	0.68071	0	-7.8733	-4.0383
	5	-4.257	0.78876	0	-6.4789	-2.0352
5	1	-5.5814	0.7131	0	-7.5901	-3.5727
	2	-14.3118	0.71201	0	-16.3175	-12.3062
	3	-1.6988	0.66513	0.109	-3.5724	0.1748
	4	4.257	0.78876	0	2.0352	6.4789

Çizelge Ek 3.24 D26 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı

d26		I-J	Std. Error	Sig.	95% Confidence Interval	
I	J				Lower Bound	Upper Bound
1	2	0.9734	0.17716	0	0.4743	1.4724
	3	-0.378	0.16278	0.205	-0.8366	0.0805
	4	0.8965	0.20028	0	0.3323	1.4607
	5	0.362	0.19628	0.656	-0.1909	0.9149
2	1	-0.9734	0.17716	0	-1.4724	-0.4743
	3	-1.3514	0.16242	0	-1.8089	-0.8939
	4	-0.0769	0.19999	1	-0.6402	0.4865
	5	-0.6113	0.19598	0.019	-1.1634	-0.0593
3	1	0.378	0.16278	0.205	-0.0805	0.8366
	2	1.3514	0.16242	0	0.8939	1.8089
	4	1.2745	0.18736	0	0.7467	1.8023
	5	0.74	0.18308	0.001	0.2243	1.2557
4	1	-0.8965	0.20028	0	-1.4607	-0.3323
	2	0.0769	0.19999	1	-0.4865	0.6402
	3	-1.2745	0.18736	0	-1.8023	-0.7467
	5	-0.5345	0.2171	0.141	-1.146	0.0771
5	1	-0.362	0.19628	0.656	-0.9149	0.1909
	2	0.6113	0.19598	0.019	0.0593	1.1634
	3	-0.74	0.18308	0.001	-1.2557	-0.2243
	4	0.5345	0.2171	0.141	-0.0771	1.146

Çizelge Ek 3.25 D27 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı

d27		I-J	Std. Error	Sig.	95% Confidence Interval	
I	J				Lower Bound	Upper Bound
1	2	-3.4001	0.2355	0	-4.0635	-2.7368
	3	0.9006	0.21638	0	0.2911	1.5101
	4	2.9007	0.26623	0	2.1508	3.6507
	5	1.6861	0.26091	0	0.9512	2.4211
2	1	3.4001	0.2355	0	2.7368	4.0635
	3	4.3008	0.2159	0	3.6926	4.9089
	4	6.3009	0.26584	0	5.552	7.0497
	5	5.0863	0.26051	0	4.3524	5.8201
3	1	-0.9006	0.21638	0	-1.5101	-0.2911
	2	-4.3008	0.2159	0	-4.9089	-3.6926
	4	2.0001	0.24906	0	1.2985	2.7017
	5	0.7855	0.24336	0.013	0.1	1.471
4	1	-2.9007	0.26623	0	-3.6507	-2.1508
	2	-6.3009	0.26584	0	-7.0497	-5.552
	3	-2.0001	0.24906	0	-2.7017	-1.2985
	5	-1.2146	0.28859	0	-2.0275	-0.4017
5	1	-1.6861	0.26091	0	-2.4211	-0.9512
	2	-5.0863	0.26051	0	-5.8201	-4.3524
	3	-0.7855	0.24336	0.013	-1.471	-0.1
	4	1.2146	0.28859	0	0.4017	2.0275

Çizelge Ek 3.26 D28 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı

d28		I-J	Std. Error	Sig.	95% Confidence Interval	
I	J				Lower Bound	Upper Bound
1	2	-0.1375	0.07921	0.83	-0.3607	0.0856
	3	0.2138	0.07278	0.034	0.0088	0.4188
	4	1.0348	0.08955	0	0.7825	1.287
	5	0.7143	0.08776	0	0.4671	0.9615
2	1	0.1375	0.07921	0.83	-0.0856	0.3607
	3	0.3514	0.07262	0	0.1468	0.5559
	4	1.1723	0.08942	0	0.9204	1.4242
	5	0.8518	0.08762	0	0.605	1.0986
3	1	-0.2138	0.07278	0.034	-0.4188	-0.0088
	2	-0.3514	0.07262	0	-0.5559	-0.1468
	4	0.8209	0.08377	0	0.585	1.0569
	5	0.5005	0.08185	0	0.2699	0.731
4	1	-1.0348	0.08955	0	-1.287	-0.7825
	2	-1.1723	0.08942	0	-1.4242	-0.9204
	3	-0.8209	0.08377	0	-1.0569	-0.585
	5	-0.3205	0.09707	0.01	-0.5939	-0.0471
5	1	-0.7143	0.08776	0	-0.9615	-0.4671
	2	-0.8518	0.08762	0	-1.0986	-0.605
	3	-0.5005	0.08185	0	-0.731	-0.2699
	4	0.3205	0.09707	0.01	0.0471	0.5939

Çizelge Ek 3.27 D29 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı

d29		I-J	Std. Error	Sig.	95% Confidence Interval	
I	J				Lower Bound	Upper Bound
1	2	-2.4005	0.16261	0	-2.8586	-1.9424
	3	0.4087	0.14941	0.064	-0.0121	0.8296
	4	1.4002	0.18383	0	0.8824	1.918
	5	0.6649	0.18016	0.002	0.1574	1.1723
2	1	2.4005	0.16261	0	1.9424	2.8586
	3	2.8092	0.14908	0	2.3893	3.2291
	4	3.8007	0.18356	0	3.2836	4.3178
	5	3.0654	0.17988	0	2.5587	3.5721
3	1	-0.4087	0.14941	0.064	-0.8296	0.0121
	2	-2.8092	0.14908	0	-3.2291	-2.3893
	4	0.9915	0.17197	0	0.5071	1.4759
	5	0.2561	0.16804	1	-0.2172	0.7295
4	1	-1.4002	0.18383	0	-1.918	-0.8824
	2	-3.8007	0.18356	0	-4.3178	-3.2836
	3	-0.9915	0.17197	0	-1.4759	-0.5071
	5	-0.7353	0.19927	0.002	-1.2967	-0.174
5	1	-0.6649	0.18016	0.002	-1.1723	-0.1574
	2	-3.0654	0.17988	0	-3.5721	-2.5587
	3	-0.2561	0.16804	1	-0.7295	0.2172
	4	0.7353	0.19927	0.002	0.174	1.2967

Çizelge Ek 3.28 D30 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı

d30		I-J	Std. Error	Sig.	95% Confidence Interval	
I	J				Lower Bound	Upper Bound
1	2	-0.6691	0.06428	0	-0.8502	-0.4881
	3	0.2864	0.05906	0	0.1201	0.4528
	4	0.811	0.07267	0	0.6063	1.0156
	5	0.6334	0.07121	0	0.4328	0.834
2	1	0.6691	0.06428	0	0.4881	0.8502
	3	0.9555	0.05893	0	0.7895	1.1215
	4	1.4801	0.07256	0	1.2757	1.6845
	5	1.3025	0.0711	0	1.1022	1.5028
3	1	-0.2864	0.05906	0	-0.4528	-0.1201
	2	-0.9555	0.05893	0	-1.1215	-0.7895
	4	0.5245	0.06798	0	0.3331	0.716
	5	0.347	0.06642	0	0.1599	0.5341
4	1	-0.811	0.07267	0	-1.0156	-0.6063
	2	-1.4801	0.07256	0	-1.6845	-1.2757
	3	-0.5245	0.06798	0	-0.716	-0.3331
	5	-0.1776	0.07877	0.245	-0.3994	0.0443
5	1	-0.6334	0.07121	0	-0.834	-0.4328
	2	-1.3025	0.0711	0	-1.5028	-1.1022
	3	-0.347	0.06642	0	-0.5341	-0.1599
	4	0.1776	0.07877	0.245	-0.0443	0.3994

Çizelge Ek 3.29 D31 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı

d31		I-J	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	-1.6096	0.14921	0	-2.0299	-1.1893
	3	0.4035	0.1371	0.034	0.0173	0.7896
	4	1.4971	0.16869	0	1.0219	1.9723
	5	0.4201	0.16531	0.113	-0.0456	0.8858
2	1	1.6096	0.14921	0	1.1893	2.0299
	3	2.0131	0.13679	0	1.6277	2.3984
	4	3.1067	0.16844	0	2.6322	3.5812
	5	2.0297	0.16506	0	1.5647	2.4946
3	1	-0.4035	0.1371	0.034	-0.7896	-0.0173
	2	-2.0131	0.13679	0	-2.3984	-1.6277
	4	1.0936	0.15781	0	0.6491	1.5382
	5	0.0166	0.15419	1	-0.4177	0.451
4	1	-1.4971	0.16869	0	-1.9723	-1.0219
	2	-3.1067	0.16844	0	-3.5812	-2.6322
	3	-1.0936	0.15781	0	-1.5382	-0.6491
	5	-1.077	0.18285	0	-1.5921	-0.5619
5	1	-0.4201	0.16531	0.113	-0.8858	0.0456
	2	-2.0297	0.16506	0	-2.4946	-1.5647
	3	-0.0166	0.15419	1	-0.451	0.4177
	4	1.077	0.18285	0	0.5619	1.5921

Çizelge Ek 3.30 D32 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı

d32		I-J	Std. Error	Sig.	95% Confidence Interval	
I	J				Lower Bound	Upper Bound
1	2	-0.6496	0.05642	0	-0.8085	-0.4907
	3	0.0561	0.05184	1	-0.0899	0.2022
	4	0.5767	0.06378	0	0.397	0.7564
	5	0.2854	0.06251	0	0.1093	0.4614
2	1	0.6496	0.05642	0	0.4907	0.8085
	3	0.7057	0.05172	0	0.56	0.8514
	4	1.2263	0.06369	0	1.0469	1.4057
	5	0.9349	0.06241	0	0.7591	1.1107
3	1	-0.0561	0.05184	1	-0.2022	0.0899
	2	-0.7057	0.05172	0	-0.8514	-0.56
	4	0.5206	0.05967	0	0.3525	0.6887
	5	0.2292	0.0583	0.001	0.065	0.3934
4	1	-0.5767	0.06378	0	-0.7564	-0.397
	2	-1.2263	0.06369	0	-1.4057	-1.0469
	3	-0.5206	0.05967	0	-0.6887	-0.3525
	5	-0.2914	0.06914	0	-0.4861	-0.0966
5	1	-0.2854	0.06251	0	-0.4614	-0.1093
	2	-0.9349	0.06241	0	-1.1107	-0.7591
	3	-0.2292	0.0583	0.001	-0.3934	-0.065
	4	0.2914	0.06914	0	0.0966	0.4861

Çizelge Ek 3.31 D33 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı

d33		I-J	Std. Error	Sig.	95% Confidence Interval	
I	J				Lower Bound	Upper Bound
1	2	-20.8422	2.72238	0	-28.5108	-13.1735
	3	3.1322	2.5013	1	-3.9137	10.178
	4	8.8139	3.07764	0.043	0.1446	17.4833
	5	7.3851	3.01608	0.146	-1.1108	15.881
2	1	20.8422	2.72238	0	13.1735	28.5108
	3	23.9743	2.49574	0	16.9441	31.0045
	4	29.6561	3.07313	0	20.9995	38.3127
	5	28.2272	3.01148	0	19.7443	36.7102
3	1	-3.1322	2.5013	1	-10.178	3.9137
	2	-23.9743	2.49574	0	-31.0045	-16.9441
	4	5.6818	2.87911	0.489	-2.4283	13.7919
	5	4.2529	2.8132	1	-3.6715	12.1774
4	1	-8.8139	3.07764	0.043	-17.4833	-0.1446
	2	-29.6561	3.07313	0	-38.3127	-20.9995
	3	-5.6818	2.87911	0.489	-13.7919	2.4283
	5	-1.4289	3.33609	1	-10.8262	7.9685
5	1	-7.3851	3.01608	0.146	-15.881	1.1108
	2	-28.2272	3.01148	0	-36.7102	-19.7443
	3	-4.2529	2.8132	1	-12.1774	3.6715
	4	1.4289	3.33609	1	-7.9685	10.8262

Çizelge Ek 3.32 D34 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı

d34		I-J	Std. Error	Sig.	95% Confidence Interval	
I	J				Lower Bound	Upper Bound
1	2	-4.6486	0.91789	0	-7.2342	-2.063
	3	-0.1494	0.84335	1	-2.525	2.2263
	4	2.5556	1.03768	0.14	-0.3674	5.4787
	5	2.0872	1.01692	0.405	-0.7774	4.9517
2	1	4.6486	0.91789	0	2.063	7.2342
	3	4.4993	0.84148	0	2.1289	6.8696
	4	7.2043	1.03615	0	4.2855	10.123
	5	6.7358	1.01537	0	3.8756	9.5959
3	1	0.1494	0.84335	1	-2.2263	2.525
	2	-4.4993	0.84148	0	-6.8696	-2.1289
	4	2.705	0.97074	0.055	-0.0294	5.4394
	5	2.2365	0.94852	0.187	-0.4353	4.9084
4	1	-2.5556	1.03768	0.14	-5.4787	0.3674
	2	-7.2043	1.03615	0	-10.123	-4.2855
	3	-2.705	0.97074	0.055	-5.4394	0.0294
	5	-0.4685	1.12482	1	-3.637	2.7
5	1	-2.0872	1.01692	0.405	-4.9517	0.7774
	2	-6.7358	1.01537	0	-9.5959	-3.8756
	3	-2.2365	0.94852	0.187	-4.9084	0.4353
	4	0.4685	1.12482	1	-2.7	3.637

Çizelge Ek 3.33 D35 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı *

d35		I-J	Std. Error	Sig.	95% Confidence Interval	
I	J				Lower Bound	Upper Bound
1	2	-3.6377	0.95081	0.001	-6.316	-0.9594
	3	0	0.8736	1	-2.4608	2.4608
	4	0	1.07489	1	-3.0278	3.0278
	5	0	1.05339	1	-2.9673	2.9673
2	1	3.6377	0.95081	0.001	0.9594	6.316
	3	3.6377	0.87166	0	1.1823	6.093
	4	3.6377	1.07331	0.007	0.6143	6.6611
	5	3.6377	1.05178	0.006	0.6749	6.6004
3	1	0	0.8736	1	-2.4608	2.4608
	2	-3.6377	0.87166	0	-6.093	-1.1823
	4	0	1.00555	1	-2.8325	2.8325
	5	0	0.98253	1	-2.7677	2.7677
4	1	0	1.07489	1	-3.0278	3.0278
	2	-3.6377	1.07331	0.007	-6.6611	-0.6143
	3	0	1.00555	1	-2.8325	2.8325
	5	0	1.16516	1	-3.2821	3.2821
5	1	0	1.05339	1	-2.9673	2.9673
	2	-3.6377	1.05178	0.006	-6.6004	-0.6749
	3	0	0.98253	1	-2.7677	2.7677
	4	0	1.16516	1	-3.2821	3.2821

Çizelge Ek 3.34 D36 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı

d36		I-J	Std. Error	Sig.	95% Confidence Interval	
I	J				Lower Bound	Upper Bound
1	2	-0.5287	0.1588	0.009	-0.976	-0.0813
	3	0	0.1459	1	-0.411	0.411
	4	0	0.17952	1	-0.5057	0.5057
	5	0	0.17593	1	-0.4956	0.4956
2	1	0.5287	0.1588	0.009	0.0813	0.976
	3	0.5287	0.14558	0.003	0.1186	0.9387
	4	0.5287	0.17926	0.033	0.0237	1.0336
	5	0.5287	0.17566	0.027	0.0338	1.0235
3	1	0	0.1459	1	-0.411	0.411
	2	-0.5287	0.14558	0.003	-0.9387	-0.1186
	4	0	0.16794	1	-0.4731	0.4731
	5	0	0.1641	1	-0.4622	0.4622
4	1	0	0.17952	1	-0.5057	0.5057
	2	-0.5287	0.17926	0.033	-1.0336	-0.0237
	3	0	0.16794	1	-0.4731	0.4731
	5	0	0.1946	1	-0.5482	0.5482
5	1	0	0.17593	1	-0.4956	0.4956
	2	-0.5287	0.17566	0.027	-1.0235	-0.0338
	3	0	0.1641	1	-0.4622	0.4622
	4	0	0.1946	1	-0.5482	0.5482

Çizelge Ek 3.35 D37 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı

d37		I-J	Std. Error	Sig.	95% Confidence Interval	
I	J				Lower Bound	Upper Bound
1	2	-2.6807	0.80673	0.009	-4.9531	-0.4082
	3	0	0.74121	1	-2.0879	2.0879
	4	-0.5749	0.91201	1	-3.1439	1.9941
	5	0	0.89376	1	-2.5176	2.5176
2	1	2.6807	0.80673	0.009	0.4082	4.9531
	3	2.6807	0.73957	0.003	0.5974	4.7639
	4	2.1058	0.91067	0.211	-0.4595	4.671
	5	2.6807	0.8924	0.028	0.1669	5.1944
3	1	0	0.74121	1	-2.0879	2.0879
	2	-2.6807	0.73957	0.003	-4.7639	-0.5974
	4	-0.5749	0.85317	1	-2.9782	1.8284
	5	0	0.83364	1	-2.3483	2.3483
4	1	0.5749	0.91201	1	-1.9941	3.1439
	2	-2.1058	0.91067	0.211	-4.671	0.4595
	3	0.5749	0.85317	1	-1.8284	2.9782
	5	0.5749	0.98859	1	-2.2099	3.3596
5	1	0	0.89376	1	-2.5176	2.5176
	2	-2.6807	0.8924	0.028	-5.1944	-0.1669
	3	0	0.83364	1	-2.3483	2.3483
	4	-0.5749	0.98859	1	-3.3596	2.2099

Çizelge Ek 3.36 D38 Değişkeni İçin Bonferroni Güven Bölgesi Yaklaşımı

d38		I-J	Std. Error	Sig.	95% Confidence Interval	
I	J				Lower Bound	Upper Bound
1	2	-0.6727	0.27888	0.161	-1.4583	0.1129
	3	0	0.25623	1	-0.7218	0.7218
	4	-0.462	0.31527	1	-1.3501	0.4261
	5	0	0.30897	1	-0.8703	0.8703
2	1	0.6727	0.27888	0.161	-0.1129	1.4583
	3	0.6727	0.25566	0.087	-0.0475	1.3929
	4	0.2107	0.31481	1	-0.6761	1.0975
	5	0.6727	0.3085	0.296	-0.1963	1.5417
3	1	0	0.25623	1	-0.7218	0.7218
	2	-0.6727	0.25566	0.087	-1.3929	0.0475
	4	-0.462	0.29494	1	-1.2928	0.3688
	5	0	0.28818	1	-0.8118	0.8118
4	1	0.462	0.31527	1	-0.4261	1.3501
	2	-0.2107	0.31481	1	-1.0975	0.6761
	3	0.462	0.29494	1	-0.3688	1.2928
	5	0.462	0.34175	1	-0.5007	1.4246
5	1	0	0.30897	1	-0.8703	0.8703
	2	-0.6727	0.3085	0.296	-1.5417	0.1963
	3	0	0.28818	1	-0.8118	0.8118
	4	-0.462	0.34175	1	-1.4246	0.5007

EK 4 Veri Setinde Bulunan Nicel Değişkenlerin Kümelere Göre İstatistikleri

Çizelge Ek 4.1 D3 Değişkenine İlişkin İstatistikler

d3	Mean	Std. Deviation	Minimum	Maximum	Skewness	Kurtosis	Range	N
1	25.67	13.39	5.00	66.00	0.78	0.29	61.00	133
2	115.62	65.93	15.00	342.00	1.08	1.28	327.00	134
3	142.21	87.73	29.00	508.00	1.34	1.91	479.00	195
4	328.37	264.53	87.00	2140.00	4.04	25.24	2053.00	86
5	38.71	23.59	5.00	98.00	0.48	-0.77	93.00	92
Toplam	122.56	146.50	5.00	2140.00	5.36	58.29	2135.00	640

Çizelge Ek 4.2 D4 Değişkenine İlişkin İstatistikler

d4	Mean	Std. Deviation	Minimum	Maximum	Skewness	Kurtosis	Range	N
1	3.64	4.12	0.00	18.00	1.33	1.23	18.00	133
2	1.18	2.34	0.00	14.00	3.29	12.11	14.00	134
3	42.35	26.49	0.00	121.00	0.62	-0.06	121.00	195
4	203.03	133.81	47.00	770.00	1.72	3.74	723.00	86
5	19.95	16.80	1.00	72.00	1.07	0.61	71.00	92
Toplam	44.06	82.77	0.00	770.00	3.85	19.81	770.00	640

Çizelge Ek 4.3 D5 Değişkenine İlişkin İstatistikler

d5	Mean	Std. Deviation	Minimum	Maximum	Skewness	Kurtosis	Range	N
1	0.16	0.16	0.00	0.64	0.95	-0.03	0.64	133
2	0.01	0.02	0.00	0.13	3.08	10.18	0.13	134
3	0.31	0.14	0.00	0.68	-0.03	-0.16	0.68	195
4	0.65	0.13	0.27	0.90	-0.63	0.05	0.63	86
5	0.50	0.23	0.02	0.89	-0.32	-0.99	0.87	92
Toplam	0.29	0.25	0.00	0.90	0.48	-0.93	0.90	640

Çizelge Ek 4.4 D6 Değişkenine İlişkin İstatistikler

d6	Mean	Std. Deviation	Minimum	Maximum	Skewness	Kurtosis	Range	N
1	6.25	4.16	0.00	21.00	0.76	0.55	21.00	133
2	6.92	6.93	0.00	29.00	1.05	0.51	29.00	134
3	48.54	31.70	4.00	169.00	1.18	1.17	165.00	195
4	91.37	90.72	9.00	756.00	4.85	33.66	747.00	86
5	10.45	7.09	1.00	31.00	0.82	-0.05	30.00	92
Toplam	31.32	48.11	0.00	756.00	6.45	81.76	756.00	640

Çizelge Ek 4.5 D7 Değişkenine İlişkin İstatistikler

d7	Mean	Std. Deviation	Minimum	Maximum	Skewness	Kurtosis	Range	N
1	0.26	0.16	0.00	0.71	0.39	-0.47	0.71	133
2	0.06	0.07	0.00	0.35	1.50	3.13	0.35	134
3	0.34	0.08	0.04	0.55	-0.62	1.51	0.51	195
4	0.27	0.07	0.09	0.41	-0.08	-0.59	0.32	86
5	0.28	0.11	0.05	0.62	0.52	0.18	0.56	92
Toplam	0.25	0.14	0.00	0.71	-0.06	-0.60	0.71	640

Çizelge Ek 4.6 D8 Değişkenine İlişkin İstatistikler

d8	Mean	Std. Deviation	Minimum	Maximum	Skewness	Kurtosis	Range	N
1	15.78	12.35	0.00	60.00	1.08	1.11	60.00	133
2	107.52	64.57	15.00	339.00	1.18	1.48	324.00	134
3	51.31	41.39	4.00	232.00	1.65	3.78	228.00	195
4	33.97	69.38	1.00	614.00	7.13	58.77	613.00	86
5	8.32	10.11	0.00	59.00	2.38	7.45	59.00	92
Toplam	47.19	57.35	0.00	614.00	3.06	17.74	614.00	640

Çizelge Ek 4.7 D9 Değişkenine İlişkin İstatistikler

d9	Mean	Std. Deviation	Minimum	Maximum	Skewness	Kurtosis	Range	N
1	0.58	0.28	0.00	1.00	-0.29	-1.13	1.00	133
2	0.92	0.09	0.54	1.00	-1.78	4.17	0.46	134
3	0.35	0.16	0.06	0.96	1.22	2.01	0.89	195
4	0.09	0.08	0.01	0.42	2.09	5.04	0.41	86
5	0.22	0.20	0.00	0.86	1.06	0.55	0.86	92
Toplam	0.46	0.33	0.00	1.00	0.34	-1.34	1.00	640

Çizelge Ek 4.8 D10 Değişkenine İlişkin İstatistikler

d10	Mean	Std. Deviation	Minimum	Maximum	Skewness	Kurtosis	Range	N
1	3.31	4.77	0.00	31.00	3.31	13.68	31.00	133
2	4.16	5.58	0.00	36.00	2.70	10.18	36.00	134
3	7.72	9.13	0.00	76.00	3.35	17.90	76.00	195
4	22.13	20.29	1.00	97.00	1.57	2.78	96.00	86
5	4.67	7.76	0.00	52.00	3.49	16.21	52.00	92
Toplam	7.56	11.65	0.00	97.00	3.46	16.27	97.00	640

Çizelge Ek 4.9 D11 Değişkenine İlişkin İstatistikler

d11	Mean	Std. Deviation	Minimum	Maximum	Skewness	Kurtosis	Range	N
1	0.13	0.14	0.00	0.78	1.73	3.45	0.78	133
2	0.04	0.06	0.00	0.41	3.96	19.44	0.41	134
3	0.07	0.10	0.00	0.53	2.96	9.33	0.53	195
4	0.06	0.04	0.01	0.26	1.74	6.07	0.25	86
5	0.12	0.17	0.00	0.74	1.89	3.04	0.74	92
Toplam	0.08	0.12	0.00	0.78	2.75	8.66	0.78	640

Çizelge Ek 4.10 D12 Değişkenine İlişkin İstatistikler

d12	Mean	Std. Deviation	Minimum	Maximum	Skewness	Kurtosis	Range	N
1	7.13	7.55	0.00	47.00	1.93	5.57	47.00	133
2	67.61	47.75	7.00	234.00	1.33	1.96	227.00	134
3	11.76	13.85	0.00	75.00	2.43	6.81	75.00	195
4	2.10	8.27	0.00	74.00	8.06	69.39	74.00	86
5	1.15	1.59	0.00	8.00	2.06	4.92	8.00	92
Toplam	19.67	34.33	0.00	234.00	2.95	10.55	234.00	640

Çizelge Ek 4.11 D13 Değişkenine İlişkin İstatistikler

d13	Mean	Std. Deviation	Minimum	Maximum	Skewness	Kurtosis	Range	N
1	0.27	0.22	0.00	1.00	0.81	0.35	1.00	133
2	0.58	0.21	0.08	1.00	0.02	-0.85	0.92	134
3	0.09	0.12	0.00	0.77	3.60	13.89	0.77	195
4	0.01	0.01	0.00	0.12	6.73	53.51	0.12	86
5	0.04	0.07	0.00	0.35	2.39	6.76	0.35	92
Toplam	0.21	0.26	0.00	1.00	1.26	0.43	1.00	640

Çizelge Ek 4.12 D14 Değişkenine İlişkin İstatistikler

d14	Mean	Std. Deviation	Minimum	Maximum	Skewness	Kurtosis	Range	N
1	1.34	2.25	0.00	17.00	3.41	17.84	17.00	133
2	1.84	2.38	0.00	13.00	2.08	5.50	13.00	134
3	7.08	6.23	0.00	32.00	1.45	2.21	32.00	195
4	26.31	73.56	0.00	680.00	8.45	75.65	680.00	86
5	3.60	3.81	0.00	20.00	2.05	5.38	20.00	92
Toplam	6.87	28.28	0.00	680.00	21.29	504.09	680.00	640

Çizelge Ek 4.13 D15 Değişkenine İlişkin İstatistikler

d15	Mean	Std. Deviation	Minimum	Maximum	Skewness	Kurtosis	Range	N
1	0.05	0.08	0.00	0.53	2.65	10.39	0.53	133
2	0.02	0.02	0.00	0.10	1.57	2.68	0.10	134
3	0.05	0.03	0.00	0.14	0.93	0.65	0.14	195
4	0.06	0.06	0.00	0.32	1.94	6.10	0.32	86
5	0.10	0.10	0.00	0.40	1.39	1.46	0.40	92
Toplam	0.05	0.07	0.00	0.53	2.75	10.98	0.53	640

Çizelge Ek 4.14 D16 Değişkenine İlişkin İstatistikler

d16	Mean	Std. Deviation	Minimum	Maximum	Skewness	Kurtosis	Range	N
1	0.56	0.20	0.00	0.92	-0.53	-0.13	0.92	133
2	0.36	0.18	0.00	0.71	-0.15	-0.97	0.71	134
3	0.80	0.16	0.06	0.97	-2.06	4.34	0.91	195
4	0.87	0.08	0.57	0.99	-1.21	2.34	0.42	86
5	0.73	0.19	0.13	1.00	-1.01	0.67	0.88	92
Toplam	0.66	0.25	0.00	1.00	-0.75	-0.48	1.00	640

Çizelge Ek 4.15 D17 Değişkenine İlişkin İstatistikler

d17	Mean	Std. Deviation	Minimum	Maximum	Skewness	Kurtosis	Range	N
1	12.74	5.18	4.08	31.71	0.71	0.51	27.63	133
2	24.74	6.34	13.34	41.04	0.52	-0.50	27.70	134
3	6.29	2.65	2.06	14.77	1.14	0.95	12.71	195
4	1.13	0.83	-0.22	4.17	1.19	2.14	4.39	86
5	1.48	1.48	-1.32	5.14	0.60	-0.42	6.46	92
Toplam	10.11	9.42	-1.32	41.04	1.06	0.29	42.36	640

Çizelge Ek 4.16 D18 Değişkenine İlişkin İstatistikler

d18	Mean	Std. Deviation	Minimum	Maximum	Skewness	Kurtosis	Range	N
1	10.89	2.25	4.71	16.00	-0.59	0.29	11.29	133
2	10.91	2.23	4.37	15.87	-0.58	0.42	11.50	134
3	8.82	1.65	5.06	13.09	0.10	-0.51	8.03	195
4	3.96	1.18	1.80	7.32	0.53	0.44	5.52	86
5	3.80	1.86	0.59	8.58	0.59	-0.59	7.99	92
Toplam	8.31	3.45	0.59	16.00	-0.26	-0.90	15.41	640

Çizelge Ek 4.17 D19 Değişkenine İlişkin İstatistikler

d19	Mean	Std. Deviation	Minimum	Maximum	Skewness	Kurtosis	Range	N
1	3.55	1.68	0.48	10.43	1.13	2.39	9.95	133
2	7.49	3.16	3.19	17.04	1.18	0.68	13.85	134
3	1.50	0.80	-0.02	4.64	1.33	2.07	4.66	195
4	0.33	0.31	-0.20	1.01	0.55	-1.04	1.21	86
5	0.19	0.35	-0.64	1.46	1.07	2.52	2.10	92
Toplam	2.84	3.15	-0.64	17.04	1.73	3.29	17.68	640

Çizelge Ek 4.18 D20 Değişkenine İlişkin İstatistikler

d20	Mean	Std. Deviation	Minimum	Maximum	Skewness	Kurtosis	Range	N
1	4.54	1.15	1.29	7.04	-0.16	-0.10	5.75	133
2	5.80	0.77	2.72	7.44	-1.43	3.25	4.72	134
3	3.17	0.81	0.98	5.44	0.06	0.05	4.46	195
4	1.70	0.50	0.90	2.63	0.29	-1.35	1.73	86
5	1.34	0.59	0.19	3.46	0.72	1.34	3.27	92
Toplam	3.55	1.78	0.19	7.44	0.17	-1.12	7.25	640

Çizelge Ek 4.19 D21 Değişkenine İlişkin İstatistikler

d21	Mean	Std. Deviation	Minimum	Maximum	Skewness	Kurtosis	Range	N
1	2.27	1.33	0.11	10.15	2.05	8.73	10.04	133
2	5.47	2.39	2.11	12.69	1.11	0.53	10.58	134
3	0.84	0.47	0.00	2.35	0.91	0.54	2.35	194
4	0.06	0.15	-0.22	0.40	0.50	-0.58	0.62	86
5	0.03	0.23	-0.57	0.93	0.83	2.42	1.50	92
Toplam	1.89	2.37	-0.57	12.69	1.85	3.57	13.26	639

Çizelge Ek 4.20 D22 Değişkenine İlişkin İstatistikler

d22	Mean	Std. Deviation	Minimum	Maximum	Skewness	Kurtosis	Range	N
1	3.25	1.01	0.86	5.59	-0.12	-0.35	4.73	133
2	4.38	0.62	1.79	5.72	-1.28	2.61	3.93	134
3	2.17	0.61	0.49	3.74	0.01	0.22	3.25	195
4	1.24	0.29	0.63	1.94	0.16	-0.98	1.31	86
5	0.91	0.44	0.08	2.35	0.75	1.77	2.27	92
Toplam	2.55	1.38	0.08	5.72	0.34	-1.07	5.64	640

Çizelge Ek 4.21 D23 Değişkenine İlişkin İstatistikler

d23	Mean	Std. Deviation	Minimum	Maximum	Skewness	Kurtosis	Range	N
1	0.78	0.56	-0.44	3.43	1.30	3.31	3.87	133
2	2.30	1.36	0.57	7.51	1.61	2.49	6.94	134
3	0.34	0.17	-0.04	0.90	0.62	0.22	0.94	195
4	0.10	0.09	-0.16	0.26	-0.05	-0.06	0.42	86
5	0.09	0.21	-0.44	0.93	0.94	2.38	1.37	92
Toplam	0.77	1.07	-0.44	7.51	2.68	9.11	7.95	640

Çizelge Ek 4.22 D24 Değişkenine İlişkin İstatistikler

d24	Mean	Std. Deviation	Minimum	Maximum	Skewness	Kurtosis	Range	N
1	1.60	0.58	0.45	3.19	0.20	-0.32	2.74	133
2	2.53	0.50	1.43	3.95	0.29	-0.33	2.52	134
3	1.20	0.26	0.17	1.89	-0.10	0.90	1.72	195
4	1.08	0.14	0.78	1.39	-0.16	-0.51	0.61	86
5	0.81	0.32	0.10	1.58	-0.13	-0.30	1.48	92
Toplam	1.49	0.71	0.10	3.95	0.91	0.27	3.85	640

Çizelge Ek 4.23 D24 Değişkenine İlişkin İstatistikler

d25	Mean	Std. Deviation	Minimum	Maximum	Skewness	Kurtosis	Range	N
1	24.61	7.61	7.30	37.55	-0.53	-0.79	30.25	133
2	33.34	3.57	21.99	41.46	-0.27	0.70	19.47	134
3	20.72	4.29	10.64	36.61	0.60	1.01	25.97	195
4	14.77	3.67	7.19	24.14	0.19	-0.43	16.95	86
5	19.03	6.19	0.40	35.97	0.44	0.32	35.57	92
Toplam	23.13	7.96	0.40	41.46	0.19	-0.93	41.06	640

Çizelge Ek 4.24 D26 Değişkenine İlişkin İstatistikler

d26	Mean	Std. Deviation	Minimum	Maximum	Skewness	Kurtosis	Range	N
1	9.33	1.82	5.19	13.77	0.02	-0.17	8.58	133
2	8.36	1.86	3.09	13.32	-0.53	0.38	10.23	134
3	9.71	0.83	5.37	12.66	-0.82	3.92	7.29	195
4	8.43	0.85	6.23	11.12	0.44	1.10	4.89	86
5	8.97	1.63	2.22	13.70	-0.41	2.72	11.48	92
Toplam	9.07	1.54	2.22	13.77	-0.53	1.63	11.55	640

Çizelge Ek 4.25 D27 Değişkenine İlişkin İstatistikler

d27	Mean	Std. Deviation	Minimum	Maximum	Skewness	Kurtosis	Range	N
1	7.38	2.72	1.72	12.82	-0.23	-0.87	11.10	133
2	10.78	1.77	6.73	15.27	0.14	-0.35	8.54	134
3	6.48	1.43	2.35	11.55	0.35	0.82	9.20	195
4	4.48	1.26	2.63	8.18	0.72	0.01	5.55	86
5	5.69	2.17	-0.48	12.12	0.38	0.24	12.60	92
Toplam	7.19	2.80	-0.48	15.27	0.39	-0.50	15.75	640

Çizelge Ek 4.26 D28 Değişkenine İlişkin İstatistikler

d28	Mean	Std. Deviation	Minimum	Maximum	Skewness	Kurtosis	Range	N
1	5.01	0.85	2.93	7.26	-0.13	-0.15	4.33	133
2	5.15	0.75	2.38	7.15	-1.28	3.13	4.77	134
3	4.80	0.38	3.46	6.03	-0.91	1.80	2.57	195
4	3.97	0.38	3.15	4.97	0.39	0.58	1.82	86
5	4.29	0.78	1.65	6.22	-0.57	1.37	4.57	92
Toplam	4.73	0.76	1.65	7.26	-0.25	0.71	5.61	640

Çizelge Ek 4.27 D29 Değişkenine İlişkin İstatistikler

d29	Mean	Std. Deviation	Minimum	Maximum	Skewness	Kurtosis	Range	N
1	4.31	1.69	0.71	8.06	0.24	-0.63	7.35	133
2	6.71	1.75	3.31	12.14	0.69	0.53	8.83	134
3	3.90	0.86	1.53	6.99	0.43	1.18	5.46	195
4	2.91	0.58	1.83	5.20	1.03	2.35	3.37	86
5	3.65	1.35	0.80	8.05	0.53	0.23	7.25	92
Toplam	4.41	1.83	0.71	12.14	1.05	1.31	11.43	640

Çizelge Ek 4.28 D30 Değişkenine İlişkin İstatistikler

d30	Mean	Std. Deviation	Minimum	Maximum	Skewness	Kurtosis	Range	N
1	3.46	0.77	1.52	5.54	-0.02	0.13	4.02	133
2	4.13	0.52	1.04	5.92	-1.38	10.74	4.88	134
3	3.17	0.34	2.19	4.21	-0.41	0.87	2.02	195
4	2.65	0.19	2.31	3.24	1.05	1.63	0.93	86
5	2.82	0.63	1.11	4.60	0.17	0.68	3.49	92
Toplam	3.31	0.72	1.04	5.92	0.27	0.13	4.88	640

Çizelge Ek 4.29 D31 Değişkenine İlişkin İstatistikler

d31	Mean	Std. Deviation	Minimum	Maximum	Skewness	Kurtosis	Range	N
1	3.21	1.73	0.03	7.64	-0.04	-0.89	7.61	133
2	4.82	1.07	1.67	7.54	-0.39	1.13	5.87	134
3	2.81	0.92	0.68	5.87	0.38	0.62	5.19	195
4	1.72	0.81	0.28	4.04	0.48	-0.01	3.76	86
5	2.79	1.39	0.17	7.17	0.98	0.61	7.00	92
Toplam	3.17	1.55	0.03	7.64	0.32	-0.58	7.61	640

Çizelge Ek 4.30 D32 Değişkenine İlişkin İstatistikler

d32	Mean	Std. Deviation	Minimum	Maximum	Skewness	Kurtosis	Range	N
1	2.92	0.74	0.21	4.59	-0.99	1.06	4.38	133
2	3.57	0.26	2.78	4.45	0.01	1.60	1.67	134
3	2.86	0.32	1.66	3.70	-0.92	1.77	2.04	195
4	2.34	0.37	1.52	3.12	-0.19	-0.62	1.60	86
5	2.63	0.49	0.00	3.46	-1.72	7.73	3.46	92
Toplam	2.92	0.60	0.00	4.59	-0.62	1.23	4.59	640

Çizelge Ek 4.31 D33 Değişkenine İlişkin İstatistikler

d33	Mean	Std. Deviation	Minimum	Maximum	Skewness	Kurtosis	Range	N
1	9.20	23.37	0.00	91.70	2.32	3.84	91.70	133
2	30.04	36.72	0.00	92.52	0.50	-1.62	92.52	134
3	6.07	16.71	0.00	84.60	2.87	7.66	84.60	195
4	0.39	3.59	0.00	33.33	9.27	86.00	33.33	86
5	1.82	9.06	0.00	63.71	5.38	30.23	63.71	92
Toplam	10.37	24.55	0.00	92.52	2.16	3.07	92.52	640

Çizelge Ek 4.32 D34 Değişkenine İlişkin İstatistikler

d34	Mean	Std. Deviation	Minimum	Maximum	Skewness	Kurtosis	Range	N
1	2.89	7.82	0.00	33.49	2.68	5.95	33.49	133
2	7.53	9.93	0.00	33.51	0.87	-0.69	33.51	134
3	3.03	7.98	0.00	33.68	2.46	4.54	33.68	195
4	0.33	3.06	0.00	28.36	9.27	86.00	28.36	86
5	0.80	3.90	0.00	24.40	5.02	24.98	24.40	92
Toplam	3.26	7.86	0.00	33.68	2.29	3.88	33.68	640

Çizelge Ek 4.33 D35 Değişkenine İlişkin İstatistikler

d35	Mean	Std. Deviation	Minimum	Maximum	Skewness	Kurtosis	Range	N
1	0.00	0.00	0.00	0.00	.	.	0.00	133
2	3.64	16.97	0.00	92.42	4.54	19.20	92.42	134
3	0.00	0.00	0.00	0.00	.	.	0.00	195
4	0.00	0.00	0.00	0.00	.	.	0.00	86
5	0.00	0.00	0.00	0.00	.	.	0.00	92
Toplam	0.76	7.88	0.00	92.42	10.39	107.55	92.42	640

Çizelge Ek 4.34 D36 Değişkenine İlişkin İstatistikler

d36	Mean	Std. Deviation	Minimum	Maximum	Skewness	Kurtosis	Range	N
1	0.00	0.00	0.00	0.00	.	.	0.00	133
2	0.53	2.83	0.00	25.04	6.51	47.52	25.04	134
3	0.00	0.00	0.00	0.00	.	.	0.00	195
4	0.00	0.00	0.00	0.00	.	.	0.00	86
5	0.00	0.00	0.00	0.00	.	.	0.00	92
Toplam	0.11	1.31	0.00	25.04	14.48	237.01	25.04	640

Çizelge Ek 4.35 D37 Değişkenine İlişkin İstatistikler

d37	Mean	Std. Deviation	Minimum	Maximum	Skewness	Kurtosis	Range	N
1	0.00	0.00	0.00	0.00	.	.	0.00	133
2	2.68	13.76	0.00	86.07	5.05	24.24	86.07	134
3	0.00	0.00	0.00	0.00	.	.	0.00	195
4	0.57	5.33	0.00	49.44	9.27	86.00	49.44	86
5	0.00	0.00	0.00	0.00	.	.	0.00	92
Toplam	0.64	6.66	0.00	86.07	10.60	113.38	86.07	640

Çizelge Ek 4.36 D38 Değişkenine İlişkin İstatistikler

d38	Mean	Std. Deviation	Minimum	Maximum	Skewness	Kurtosis	Range	N
1	0.00	0.00	0.00	0.00	.	.	0.00	133
2	0.67	3.61	0.00	23.68	5.49	29.34	23.68	134
3	0.00	0.00	0.00	0.00	.	.	0.00	195
4	0.46	4.28	0.00	39.73	9.27	86.00	39.73	86
5	0.00	0.00	0.00	0.00	.	.	0.00	92
Toplam	0.20	2.29	0.00	39.73	12.84	181.54	39.73	640

Çizelge Ek 5.2 ÖSYM'nin Tanımladığı İse Türevleri Etkilerine Dağılımı
Anadolu Denizcilik Meslek Lisesi

Anadolu Denizcilik Meslek Lisesi	Kısım 1	Kısım 2	Kısım 3	Kısım 4	Kısım 5	Toplam
Frekans						
Frekans İçi Frekans Yüzdesi			100.0%			100.0%
Kısım İçi Frekans Yüzdesi			0.5%			0.5%
Toplam Yüzdesi			0.5%			0.5%

Çizelge Ek 5.3 ÖSYM'nin Tanımladığı İse Türevleri Etkilerine Dağılımı
Anadolu Dış Ticaret Meslek Lisesi

Anadolu Dış Ticaret Meslek Lisesi	Kısım 1	Kısım 2	Kısım 3	Kısım 4	Kısım 5	Toplam
Frekans						
Frekans İçi Frekans Yüzdesi			100.0%			100.0%
Kısım İçi Frekans Yüzdesi			0.5%			0.5%
Toplam Yüzdesi			0.5%			0.5%

EK 5 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Göre Dağılımları

Çizelge Ek 5.1 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-A.Ö.Lisesi

A.Ö. Lisesi	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans				1		1
Değişken içi Frekans Yüzdesi				100.0%		100.0%
Küme İçi Frekans Yüzdesi				1.2%		0.2%
Toplam Yüzdesi				0.2%		0.2%

Çizelge Ek 5.2 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Anadolu Denizcilik Meslek Lisesi

Anadolu Denizcilik Meslek Lisesi	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans			3			3
Değişken içi Frekans Yüzdesi			100.0%			100.0%
Küme İçi Frekans Yüzdesi			1.5%			0.5%
Toplam Yüzdesi			0.5%			0.5%

Çizelge Ek 5.3 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Anadolu Dış Ticaret Meslek Lisesi

Anadolu Dış Ticaret Meslek Lisesi	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans	1					1
Değişken içi Frekans Yüzdesi	100.0%					100.0%
Küme İçi Frekans Yüzdesi	0.8%					0.2%
Toplam Yüzdesi	0.2%					0.2%

Çizelge Ek 5.4 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Anadolu Güzel Sanatlar Lisesi

Anadolu Güzel Sanatlar Lisesi	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans					2	2
Değişken içi Frekans Yüzdesi					100.0%	100.0%
Küme İçi Frekans Yüzdesi					2.2%	0.3%
Toplam Yüzdesi					0.3%	0.3%

Çizelge Ek 5.5 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Anadolu Hazır Giyim ve Deri Hazır Giyim Lisesi

Anadolu Hazır Giyim ve Deri Hazır Giyim Lisesi	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans					1	1
Değişken içi Frekans Yüzdesi					100.0%	100.0%
Küme İçi Frekans Yüzdesi					1.1%	0.2%
Toplam Yüzdesi					0.2%	0.2%

Çizelge Ek 5.6 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Anadolu İletişim Meslek Lisesi

Anadolu İletişim Meslek Lisesi	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans					1	1
Değişken içi Frekans Yüzdesi					100.0%	100.0%
Küme İçi Frekans Yüzdesi					1.1%	0.2%
Toplam Yüzdesi					0.2%	0.2%

Çizelge Ek 5.7 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Anadolu İmam Hatip Lisesi

Anadolu İmam Hatip Lisesi	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans	2	1				3
Değişken içi Frekans Yüzdesi	66.7%	33.3%				100.0%
Küme İçi Frekans Yüzdesi	1.5%	0.7%				0.5%
Toplam Yüzdesi	0.3%	0.2%				0.5%

Çizelge Ek 5.8 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Anadolu Kız Meslek Lisesi

Anadolu Kız Meslek Lisesi	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans	1		2		9	12
Değişken içi Frekans Yüzdesi	8.3%		16.7%		75.0%	100.0%
Küme İçi Frekans Yüzdesi	0.8%		1.0%		9.8%	1.9%
Toplam Yüzdesi	0.2%		0.3%		1.4%	1.9%

Çizelge Ek 5.9 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Anadolu Lisesi

Anadolu Lisesi	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans	2	29				31
Değişken içi Frekans Yüzdesi	6.5%	93.5%				100.0%
Küme İçi Frekans Yüzdesi	1.5%	21.6%				4.8%
Toplam Yüzdesi	0.3%	4.5%				4.8%

Çizelge Ek 5.10 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Anadolu Mahalli İdareler Meslek Lisesi

Anadolu Mahalli İdareler Meslek Lisesi	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans					1	1
Değişken içi Frekans Yüzdesi					100.0%	100.0%
Küme İçi Frekans Yüzdesi					1.1%	0.2%
Toplam Yüzdesi					0.2%	0.2%

Çizelge Ek 5.11 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Anadolu Meslek Lisesi

Anadolu Meslek Lisesi	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans	10	3	4		2	19
Değişken içi Frekans Yüzdesi	52.6%	15.8%	21.1%		10.5%	100.0%
Küme İçi Frekans Yüzdesi	7.5%	2.2%	2.1%		2.2%	3.0%
Toplam Yüzdesi	1.6%	0.5%	0.6%		0.3%	3.0%

Çizelge Ek 5.12 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Anadolu Moda Tasarımı Meslek Lisesi

Anadolu Moda Tasarımı Meslek Lisesi	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans					1	1
Değişken içi Frekans Yüzdesi					100.0%	100.0%
Küme İçi Frekans Yüzdesi					1.1%	0.2%
Toplam Yüzdesi					0.2%	0.2%

Çizelge Ek 5.13 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Anadolu Otelcilik ve Turizm Meslek Lisesi

Anadolu Otelcilik ve Turizm Meslek Lisesi	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans			1	1		2
Değişken içi Frekans Yüzdesi			50.0%	50.0%		100.0%
Küme İçi Frekans Yüzdesi			0.5%	1.2%		0.3%
Toplam Yüzdesi			0.2%	0.2%		0.3%

Çizelge Ek 5.14 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Anadolu Öğretmen Lisesi

Anadolu Öğretmen Lisesi	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans		1				1
Değişken içi Frekans Yüzdesi		100.0%				100.0%
Küme İçi Frekans Yüzdesi		0.7%				0.2%
Toplam Yüzdesi		0.2%				0.2%

Çizelge Ek 5.15 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Anadolu Sağlık Meslek Lisesi

Anadolu Sağlık Meslek Lisesi	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans	1					1
Değişken içi Frekans Yüzdesi	100.0%					100.0%
Küme İçi Frekans Yüzdesi	0.8%					0.2%
Toplam Yüzdesi	0.2%					0.2%

Çizelge Ek 5.16 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Anadolu Sekreterlik Meslek Lisesi

Anadolu Sekreterlik Meslek Lisesi	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans					1	1
Değişken içi Frekans Yüzdesi					100.0%	100.0%
Küme İçi Frekans Yüzdesi					1.1%	0.2%
Toplam Yüzdesi					0.2%	0.2%

Çizelge Ek 5.17 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Anadolu Teknik Lisesi

Anadolu Teknik Lisesi	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans	8					8
Değişken içi Frekans Yüzdesi	100.0%					100.0%
Küme İçi Frekans Yüzdesi	6.0%					1.3%
Toplam Yüzdesi	1.3%					1.3%

Çizelge Ek 5.18 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Anadolu Tekstil Meslek Lisesi

Anadolu Tekstil Meslek Lisesi	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans	1					1
Değişken içi Frekans Yüzdesi	100.0%					100.0%
Küme İçi Frekans Yüzdesi	0.8%					0.2%
Toplam Yüzdesi	0.2%					0.2%

Çizelge Ek 5.19 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Anadolu Ticaret Meslek Lisesi

Anadolu Ticaret Meslek Lisesi	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans	2		4		1	7
Değişken içi Frekans Yüzdesi	28.6%		57.1%		14.3%	100.0%
Küme İçi Frekans Yüzdesi	1.5%		2.1%		1.1%	1.1%
Toplam Yüzdesi	0.3%		0.6%		0.2%	1.1%

Çizelge Ek 5.20 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Anadolu Ticaret Meslek Lisesi

Askeri Lise	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans		2				2
Değişken içi Frekans Yüzdesi		100.0%				100.0%
Küme İçi Frekans Yüzdesi		1.5%				0.3%
Toplam Yüzdesi		0.3%				0.3%

Çizelge Ek 5.21 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Çok Programlı Lise

Çok Programlı Lise	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans	1		2	3	4	10
Değişken içi Frekans Yüzdesi	10.0%		20.0%	30.0%	40.0%	100.0%
Küme İçi Frekans Yüzdesi	0.8%		1.0%	3.5%	4.3%	1.6%
Toplam Yüzdesi	0.2%		0.3%	0.5%	0.6%	1.6%

Çizelge Ek 5.22 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Denizcilik ve Su Ürünleri Meslek Lisesi

Denizcilik ve Su Ürünleri Meslek Lisesi	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans				1		1
Değişken içi Frekans Yüzdesi				100.0%		100.0%
Küme İçi Frekans Yüzdesi				1.2%		0.2%
Toplam Yüzdesi				0.2%		0.2%

Çizelge Ek 5.23 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Endüstri Meslek Lisesi

Endüstri Meslek Lisesi	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans			2	27		29
Değişken içi Frekans Yüzdesi			6.9%	93.1%		100.0%
Küme İçi Frekans Yüzdesi			1.0%	31.4%		4.5%
Toplam Yüzdesi			0.3%	4.2%		4.5%

Çizelge Ek 5.24 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Fen Lisesi

Fen Lisesi	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans		1				1
Değişken içi Frekans Yüzdesi		100.0%				100.0%
Küme İçi Frekans Yüzdesi		0.7%				0.2%
Toplam Yüzdesi		0.2%				0.2%

Çizelge Ek 5.25 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Geleneksel Türk Sanatları Kız Meslek Lisesi

Geleneksel Türk Sanatları Kız Meslek Lisesi	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans					1	1
Değişken içi Frekans Yüzdesi					100.0%	100.0%
Küme İçi Frekans Yüzdesi					1.1%	0.2%
Toplam Yüzdesi					0.2%	0.2%

Çizelge Ek 5.26 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
İmam Hatip Lisesi

İmam Hatip Lisesi	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans			11	4	6	21
Değişken içi Frekans Yüzdesi			52.4%	19.0%	28.6%	100.0%
Küme İçi Frekans Yüzdesi			5.6%	4.7%	6.5%	3.3%
Toplam Yüzdesi			1.7%	0.6%	0.9%	3.3%

Çizelge Ek 5.27 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
İnşaat Teknik Lisesi

İnşaat Teknik Lisesi	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans	1					1
Değişken içi Frekans Yüzdesi	100.0%					100.0%
Küme İçi Frekans Yüzdesi	0.8%					0.2%
Toplam Yüzdesi	0.2%					0.2%

Çizelge Ek 5.28 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Kız Meslek Lisesi

Kız Meslek Lisesi	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans				9	17	26
Değişken içi Frekans Yüzdesi				34.6%	65.4%	100.0%
Küme İçi Frekans Yüzdesi				10.5%	18.5%	4.1%
Toplam Yüzdesi				1.4%	2.7%	4.1%

Çizelge Ek 5.29 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Kuaforluk Meslek Lisesi

Kuaforluk Meslek Lisesi	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans					1	1
Değişken içi Frekans Yüzdesi					100.0%	100.0%
Küme İçi Frekans Yüzdesi					1.1%	0.2%
Toplam Yüzdesi					0.2%	0.2%

Çizelge Ek 5.30 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Matbaa Meslek Lisesi

Matbaa Meslek Lisesi	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans					1	1
Değişken içi Frekans Yüzdesi					100.0%	100.0%
Küme İçi Frekans Yüzdesi					1.1%	0.2%
Toplam Yüzdesi					0.2%	0.2%

Çizelge Ek 5.31 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Normal Lise

Normal Lise	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans	3	1	145	2	9	160
Değişken içi Frekans Yüzdesi	1.9%	0.6%	90.6%	1.3%	5.6%	100.0%
Küme İçi Frekans Yüzdesi	2.3%	0.7%	74.4%	2.3%	9.8%	25.0%
Toplam Yüzdesi	0.5%	0.2%	22.7%	0.3%	1.4%	25.0%

Çizelge Ek 5.32 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Ortopedik Cihaz ve Protez Teknisyen Okulu

Ortopedik Cihaz ve Protez Teknisyen Okulu	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans					1	1
Değişken içi Frekans Yüzdesi					100.0%	100.0%
Küme İçi Frekans Yüzdesi					1.1%	0.2%
Toplam Yüzdesi					0.2%	0.2%

Çizelge Ek 5.33 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Özel Akşam Lisesi

Özel Akşam Lisesi	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans				1	13	14
Değişken içi Frekans Yüzdesi				7.1%	92.9%	100.0%
Küme İçi Frekans Yüzdesi				1.2%	14.1%	2.2%
Toplam Yüzdesi				0.2%	2.0%	2.2%

Çizelge Ek 5.34 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Özel Anadolu Otelcilik ve Turizm Meslek Lisesi

Özel Anadolu Otelcilik ve Turizm Meslek Lisesi	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans				1	4	5
Değişken içi Frekans Yüzdesi				20.0%	80.0%	100.0%
Küme İçi Frekans Yüzdesi				1.2%	4.3%	0.8%
Toplam Yüzdesi				0.2%	0.6%	0.8%

Çizelge Ek 5.35 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Özel Fen Lisesi

Özel Fen Lisesi	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans	1	7				8
Değişken içi Frekans Yüzdesi	12.5%	87.5%				100.0%
Küme İçi Frekans Yüzdesi	0.8%	5.2%				1.3%
Toplam Yüzdesi	0.2%	1.1%				1.3%

Çizelge Ek 5.36 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Özel Lise

Özel Lise	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans	13	5			3	21
Değişken içi Frekans Yüzdesi	61.9%	23.8%			14.3%	100.0%
Küme İçi Frekans Yüzdesi	9.8%	3.7%			3.3%	3.3%
Toplam Yüzdesi	2.0%	0.8%			0.5%	3.3%

Çizelge Ek 5.37 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Özel Ticaret Meslek Lisesi

Özel Ticaret Meslek Lisesi	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans					4	4
Değişken içi Frekans Yüzdesi					100.0%	100.0%
Küme İçi Frekans Yüzdesi					4.3%	0.6%
Toplam Yüzdesi					0.6%	0.6%

Çizelge Ek 5.38 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Teknik Lise (Erkek)

Teknik Lise (Erkek)	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans	17		7			24
Değişken içi Frekans Yüzdesi	70.8%		29.2%			100.0%
Küme İçi Frekans Yüzdesi	12.8%		3.6%			3.8%
Toplam Yüzdesi	2.7%		1.1%			3.8%

Çizelge Ek 5.39 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Tekstil Meslek Lisesi

Tekstil Meslek Lisesi	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans				1		1
Değişken içi Frekans Yüzdesi				100.0%		100.0%
Küme İçi Frekans Yüzdesi				1.2%		0.2%
Toplam Yüzdesi				0.2%		0.2%

Çizelge Ek 5.40 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Tekstil Teknik Lisesi (Erkek)

Tekstil Teknik Lisesi (Erkek)	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans	1					1
Değişken içi Frekans Yüzdesi	100.0%					100.0%
Küme İçi Frekans Yüzdesi	0.8%					0.2%
Toplam Yüzdesi	0.2%					0.2%

Çizelge Ek 5.41 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Terzilik Meslek Lisesi

Terzilik Meslek Lisesi	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans					1	1
Değişken içi Frekans Yüzdesi					100.0%	100.0%
Küme İçi Frekans Yüzdesi					1.1%	0.2%
Toplam Yüzdesi					0.2%	0.2%

Çizelge Ek 5.42 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Terzilik Meslek Lisesi

Ticaret Meslek Lisesi	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans				34	2	36
Değişken içi Frekans Yüzdesi				94.4%	5.6%	100.0%
Küme İçi Frekans Yüzdesi				39.5%	2.2%	5.6%
Toplam Yüzdesi				5.3%	0.3%	5.6%

Çizelge Ek 5.43 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Veteriner Sağlık Meslek Lisesi

Veteriner Sağlık Meslek Lisesi	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans					1	1
Değişken içi Frekans Yüzdesi					100.0%	100.0%
Küme İçi Frekans Yüzdesi					1.1%	0.2%
Toplam Yüzdesi					0.2%	0.2%

Çizelge Ek 5.44 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Y.D.A. İmam Hatip Lisesi

Y.D.A. İmam Hatip Lisesi	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans		1				1
Değişken içi Frekans Yüzdesi		100.0%				100.0%
Küme İçi Frekans Yüzdesi		0.7%				0.2%
Toplam Yüzdesi		0.2%				0.2%

Çizelge Ek 5.45 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Y.D.A. Lise

Y.D.A. Lise	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans	31	59	2			92
Değişken içi Frekans Yüzdesi	33.7%	64.1%	2.2%			100.0%
Küme İçi Frekans Yüzdesi	23.3%	44.0%	1.0%			14.4%
Toplam Yüzdesi	4.8%	9.2%	0.3%			14.4%

Çizelge Ek 5.46 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Yabancı Dille Öğretim Yapan Özel Lise

Yabancı Dille Öğretim Yapan Özel Lise	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans	37	24	12		4	77
Değişken içi Frekans Yüzdesi	48.1%	31.2%	15.6%		5.2%	100.0%
Küme İçi Frekans Yüzdesi	27.8%	17.9%	6.2%		4.3%	12.0%
Toplam Yüzdesi	5.8%	3.8%	1.9%		0.6%	12.0%

Çizelge Ek 5.47 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Yapı Meslek Lisesi

Yapı Meslek Lisesi	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans				1		1
Değişken içi Frekans Yüzdesi				100.0%		100.0%
Küme İçi Frekans Yüzdesi				1.2%		0.2%
Toplam Yüzdesi				0.2%		0.2%

Çizelge Ek 5.48 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Dağılımı-
Ziraat Meslek Lisesi

Ziraat Meslek Lisesi	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans					1	1
Değişken içi Frekans Yüzdesi					100.0%	100.0%
Küme İçi Frekans Yüzdesi					1.1%	0.2%
Toplam Yüzdesi					0.2%	0.2%

EK 6 D40 Değişkenindeki Lise Türlerinin Kümelere Göre Dağılımları

Çizelge EK 6.1 Erkek Liselerinin Kümelere Göre Dağılımı

Erkek	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans	18		7			25
Değişken içi Frekans Yüzdesi	72.0%		28.0%			100.0%
Küme İçi Frekans Yüzdesi	13.5%		3.6%			3.9%
Toplam Yüzdesi	2.8%		1.1%			3.9%

Çizelge EK 6.2 Karma Liselerin Kümelere Göre Dağılımı

Karma	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans	111	131	180	77	65	564
Değişken içi Frekans Yüzdesi	19.7%	23.2%	31.9%	13.7%	11.5%	100.0%
Küme İçi Frekans Yüzdesi	83.5%	97.8%	92.3%	89.5%	70.7%	88.1%
Toplam Yüzdesi	17.3%	20.5%	28.1%	12.0%	10.2%	88.1%

Çizelge EK 6.3 Kız Liselerinin Kümelere Göre Dağılımı

Kız	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans	4	3	8	9	27	51
Değişken içi Frekans Yüzdesi	7.8%	5.9%	15.7%	17.6%	52.9%	100.0%
Küme İçi Frekans Yüzdesi	3.0%	2.2%	4.1%	10.5%	29.3%	8.0%
Toplam Yüzdesi	0.6%	0.5%	1.3%	1.4%	4.2%	8.0%

EK 7 D41 Değişkenindeki Lise Türlerinin Kümelere Göre Dağılımları

Çizelge EK 7.1 Devlet Liselerinin Kümelere Göre Dağılımı

Devlet	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans	67	97	182	84	64	494
Değişken içi Frekans Yüzdesi	13.6%	19.6%	36.8%	17.0%	13.0%	100.0%
Küme İçi Frekans Yüzdesi	50.4%	72.4%	93.3%	97.7%	69.6%	77.2%
Toplam Yüzdesi	10.5%	15.2%	28.4%	13.1%	10.0%	77.2%

Çizelge EK 7.2 Özel Liselerin Kümelere Göre Dağılımı

Özel	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans	66	37	13	2	28	146
Değişken içi Frekans Yüzdesi	45.2%	25.3%	8.9%	1.4%	19.2%	100.0%
Küme İçi Frekans Yüzdesi	49.6%	27.6%	6.7%	2.3%	30.4%	22.8%
Toplam Yüzdesi	10.3%	5.8%	2.0%	0.3%	4.4%	22.8%

Çizelge EK 8.1 Meslek Liselerinin Kümelere Göre Dağılımı

Meslek	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans	20	4	23	30	28	105
Değişken içi Frekans Yüzdesi	19.0%	3.8%	21.9%	28.6%	26.7%	100.0%
Küme İçi Frekans Yüzdesi	15.0%	3.0%	11.3%	33.0%	21.3%	23.6%
Toplam Yüzdesi	3.1%	0.6%	3.4%	11.5%	7.9%	23.6%

EK 8 D42 Değişkenindeki Lise Türlerinin Kümelere Göre Dağılımları

Çizelge EK 8.1 Anadolu Liselerinin Kümelere Göre Dağılımı

Anadolu	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans	2	29				31
Değişken içi Frekans Yüzdesi	6.5%	93.5%				100.0%
Küme İçi Frekans Yüzdesi	1.5%	21.6%				4.8%
Toplam Yüzdesi	0.3%	4.5%				4.8%

Çizelge EK 8.2 Fen Liselerinin Kümelere Göre Dağılımı

Fen	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans		1				1
Değişken içi Frekans Yüzdesi		100.0%				100.0%
Küme İçi Frekans Yüzdesi		0.7%				0.2%
Toplam Yüzdesi		0.2%				0.2%

Çizelge EK 8.3 Meslek Liselerinin Kümelere Göre Dağılımı

Meslek	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans	20	4	22	80	38	164
Değişken içi Frekans Yüzdesi	12.2%	2.4%	13.4%	48.8%	23.2%	100.0%
Küme İçi Frekans Yüzdesi	15.0%	3.0%	11.3%	93.0%	41.3%	25.6%
Toplam Yüzdesi	3.1%	0.6%	3.4%	12.5%	5.9%	25.6%

Çizelge EK 8.4 Meslek Liselerinin Kümelere Göre Dağılımı

Normal	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans	3	1	145	2	9	160
Değişken içi Frekans Yüzdesi	1.9%	0.6%	90.6%	1.3%	5.6%	100.0%
Küme İçi Frekans Yüzdesi	2.3%	0.7%	74.4%	2.3%	9.8%	25.0%
Toplam Yüzdesi	0.5%	0.2%	22.7%	0.3%	1.4%	25.0%

Çizelge EK 8.5 Özel Liselerin Kümelere Göre Dağılımı

Özel	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans	13	5			3	21
Değişken içi Frekans Yüzdesi	61.9%	23.8%			14.3%	100.0%
Küme İçi Frekans Yüzdesi	9.8%	3.7%			3.3%	3.3%
Toplam Yüzdesi	2.0%	0.8%			0.5%	3.3%

Çizelge EK 8.6 Özel Fen Liselerinin Kümelere Göre Dağılımı

Özel Fen	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans	1	7				8
Değişken içi Frekans Yüzdesi	12.5%	87.5%				100.0%
Küme İçi Frekans Yüzdesi	0.8%	5.2%				1.3%
Toplam Yüzdesi	0.2%	1.1%				1.3%

Çizelge EK 8.7 Yabancı Dille Eğitim Yapan Özel Liselerin Kümelere Göre Dağılımı

Yabancı Dille Eğitim Yapan Özel Lise	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans	37	24	12		4	77
Değişken içi Frekans Yüzdesi	48.1%	31.2%	15.6%		5.2%	100.0%
Küme İçi Frekans Yüzdesi	27.8%	17.9%	6.2%		4.3%	12.0%
Toplam Yüzdesi	5.8%	3.8%	1.9%		0.6%	12.0%

Çizelge EK 8.8 Y.D.A. Liselerin Kümelere Göre Dağılımı

Y.D.A.	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans	16	58	1			75
Değişken içi Frekans Yüzdesi	21.3%	77.3%	1.3%			100.0%
Küme İçi Frekans Yüzdesi	12.0%	43.3%	0.5%			11.7%
Toplam Yüzdesi	2.5%	9.1%	0.2%			11.7%

Çizelge EK 8.9 Y.D.A. Özel Liselerin Kümelere Göre Dağılımı

Y.D.A. Özel	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans	15	1	1			17
Değişken içi Frekans Yüzdesi	88.2%	5.9%	5.9%			100.0%
Küme İçi Frekans Yüzdesi	11.3%	0.7%	0.5%			2.7%
Toplam Yüzdesi	2.3%	0.2%	0.2%			2.7%

Çizelge EK 8.10 Anadolu Meslek Liselerinin Kümelere Göre Dağılımı

Anadolu Meslek	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans	26	4	14	1	17	62
Değişken içi Frekans Yüzdesi	41.9%	6.5%	22.6%	1.6%	27.4%	100.0%
Küme İçi Frekans Yüzdesi	19.5%	3.0%	7.2%	1.2%	18.5%	9.7%
Toplam Yüzdesi	4.1%	0.6%	2.2%	0.2%	2.7%	9.7%

Çizelge EK 8.11 Akşam Liselerinin Kümelere Göre Dağılımı

Akşam Lisesi	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans				1	13	14
Değişken içi Frekans Yüzdesi				7.1%	92.9%	100.0%
Küme İçi Frekans Yüzdesi				1.2%	14.1%	2.2%
Toplam Yüzdesi				0.2%	2.0%	2.2%

Çizelge EK 8.12 Açık Öğretim Lisesinin Kümelere Göre Dağılımı

Açık Öğretim	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans				1		1
Değişken içi Frekans Yüzdesi				100.0%		100.0%
Küme İçi Frekans Yüzdesi				1.2%		0.2%
Toplam Yüzdesi				0.2%		0.2%

Çizelge EK 8.13 Özel Meslek Liselerinin Kümelere Göre Dağılımı

Özel Meslek	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Frekans				1	8	9
Değişken içi Frekans Yüzdesi				11.1%	88.9%	100.0%
Küme İçi Frekans Yüzdesi				1.2%	8.7%	1.4%
Toplam Yüzdesi				0.2%	1.3%	1.4%

EK 9 ÖSYM'nin Tanımladığı Lise Türlerinin Testlere Göre Kümelerdeki Net Ortalamaları

Cizelge EK 9.1 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Göre Matematik Net Ortalamaları

d17	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
A.Ö. Lisesi				1,95		2,0
Anadolu Denizcilik Meslek Lisesi			10,04			10,0
Anadolu Dış Ticaret Meslek Lisesi	11,3					11,3
Anadolu Güzel Sanatlar Lisesi				2,4		2,4
Anadolu Hazır Giyim ve Deri Hazır Giyim Lisesi				1,6		1,6
Anadolu İletişim Meslek Lisesi				3,9		3,9
Anadolu İmam Hatip Lisesi	19,7	18,2				19,2
Anadolu Kız Meslek Lisesi	6,6		6,8		2,3	3,4
Anadolu Lisesi	24,8	28,9				28,7
Anadolu Mahalli İdareler Meslek Lisesi				4,2		4,2
Anadolu Meslek Lisesi	8,9	15,2	9,4		2,0	9,3
Anadolu Moda Tasarımı Meslek Lisesi					1,6	1,6
Anadolu Otelcilik ve Turizm Meslek Lisesi			5,5	4,2		4,8
Anadolu Öğretmen Lisesi		26,4				26,4
Anadolu Sağlık Meslek Lisesi	12,4					12,4
Anadolu Sekreterlik Meslek Lisesi					1,7	1,7
Anadolu Teknik Lisesi	11,4					11,4
Anadolu Tekstil Meslek Lisesi	8,2					8,2
Anadolu Ticaret Meslek Lisesi	11,8		9,0		5,1	9,2
Askeri Lise		33,9				33,9
Çok Programlı Lise	8,4		3,8	1,3	0,5	2,2
Denizcilik ve Su Ürünleri Meslek Lisesi				1,9		1,9
Endüstri Meslek Lisesi			5,5	1,3		1,6
Fen Lisesi		41,0				41,0
Geleneksel Türk Sanatları Kız Meslek Lisesi					-1,3	-1,3
İmam Hatip Lisesi			4,6	3,1	2,3	3,6
İnşaat Teknik Lisesi	8,0					8,0
Kız Meslek Lisesi				0,2	0,1	0,1
Kuaförlük Meslek Lisesi					0,4	0,4
Matbaa Meslek Lisesi					0,0	0,0
Normal Lise	8,8	28,4	5,6	2,1	2,2	5,5
Ortopedik Cihaz ve Protez Teknisyen Okulu					2,8	2,8
Özel Akşam Lisesi				0,1	0,9	0,8
Özel Anadolu Otelcilik ve Turizm Meslek Lisesi				0,9	0,6	0,7
Özel Fen Lisesi	31,7	36,6				36,0
Özel Lise	11,1	21,1			3,5	12,4
Özel Ticaret Meslek Lisesi					0,7	0,7
Teknik Lise (Erkek)	9,7		8,6			9,4
Tekstil Meslek Lisesi				0,5		0,5
Tekstil Teknik Lisesi (Erkek)	5,8					5,8
Terzilik Meslek Lisesi					0,6	0,6
Ticaret Meslek Lisesi				0,9	0,9	0,9
Veteriner Sağlık Meslek Lisesi					1,8	1,8
Y.D.A. İmam Hatip Lisesi		22,7				22,7
Y.D.A. Lise	13,6	21,5	10,1			18,6
Yabancı Dille Öğretim Yapan Özel Lise	14,9	24,7	12,4		3,9	17,0

Çizelge EK 9.2 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Göre
Fizik Net Ortalamaları

d19	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
A.Ö. Lisesi				0,4		0,4
Anadolu Denizcilik Meslek Lisesi			3,6			3,6
Anadolu Dış Ticaret Meslek Lisesi	1,2					1,2
Anadolu Güzel Sanatlar Lisesi					0,0	0,0
Anadolu Hazır Giyim ve Deri Hazır Giyim Lisesi					0,1	0,1
Anadolu İletişim Meslek Lisesi					0,3	0,3
Anadolu İmam Hatip Lisesi	4,8	4,9				4,8
Anadolu Kız Meslek Lisesi	1,6		1,3		0,2	0,5
Anadolu Lisesi	7,2	9,3				9,2
Anadolu Mahalli İdareler Meslek Lisesi					0,5	0,5
Anadolu Meslek Lisesi	3,3	5,9	3,4		0,3	3,4
Anadolu Moda Tasarımı Meslek Lisesi					0,1	0,1
Anadolu Otelcilik ve Turizm Meslek Lisesi			0,5	0,5		0,5
Anadolu Öğretmen Lisesi		8,9				8,9
Anadolu Sağlık Meslek Lisesi	2,8					2,8
Anadolu Sekreterlik Meslek Lisesi					0,4	0,4
Anadolu Teknik Lisesi	4,8					4,8
Anadolu Tekstil Meslek Lisesi	1,7					1,7
Anadolu Ticaret Meslek Lisesi	1,1		1,4		0,6	1,2
Askeri Lise		12,1				12,1
Çok Programlı Lise	3,1		0,9	0,3	0,0	0,6
Denizcilik ve Su Ürünleri Meslek Lisesi				0,9		0,9
Endüstri Meslek Lisesi			1,9	0,7		0,8
Fen Lisesi		17,0				17,0
Geleneksel Türk Sanatları Kız Meslek Lisesi					-0,6	-0,6
İmam Hatip Lisesi			0,9	0,5	0,4	0,7
İnşaat Teknik Lisesi	2,3					2,3
Kız Meslek Lisesi				0,1	0,0	0,0
Kuaförlük Meslek Lisesi					-0,3	-0,3
Matbaa Meslek Lisesi					0,1	0,1
Normal Lise	2,6	9,3	1,3	0,6	0,2	1,3
Ortopedik Cihaz ve Protez Teknisyen Okulu					1,0	1,0
Özel Akşam Lisesi				0,0	0,0	0,0
Özel Anadolu Otelcilik ve Turizm Meslek Lisesi				0,1	0,1	0,1
Özel Fen Lisesi	10,4	14,7				14,2
Özel Lise	2,9	6,2			0,7	3,4
Özel Ticaret Meslek Lisesi					0,1	0,1
Teknik Lise (Erkek)	3,1		2,9			3,1
Tekstil Meslek Lisesi				0,1		0,1
Tekstil Teknik Lisesi (Erkek)	1,9					1,9
Terzilik Meslek Lisesi					0,2	0,2
Ticaret Meslek Lisesi				0,1	0,0	0,1
Veteriner Sağlık Meslek Lisesi					0,6	0,6
Y.D.A. İmam Hatip Lisesi		5,3				5,3
Y.D.A. Lise	3,5	5,7	2,8			4,9
Yabancı Dille Öğretim Yapan Özel Lise	3,9	7,3	2,7		0,8	4,6
Yapı Meslek Lisesi				0,3		0,3
Ziraat Meslek Lisesi					1,5	1,5
Toplam	3,5	7,5	1,5	0,3	0,2	2,8

Çizelge EK 9.3 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Göre
Kimya Net Ortalamaları

d21	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
A.Ö. Lisesi				0,2		0,2
Anadolu Denizcilik Meslek Lisesi			1,5			1,5
Anadolu Dış Ticaret Meslek Lisesi	0,4					0,4
Anadolu Güzel Sanatlar Lisesi					0,1	0,1
Anadolu Hazır Giyim ve Deri Hazır Giyim Lisesi					0,1	0,1
Anadolu İletişim Meslek Lisesi					0,2	0,2
Anadolu İmam Hatip Lisesi	3,3	3,0				3,2
Anadolu Kız Meslek Lisesi	1,0		0,7		0,1	0,3
Anadolu Lisesi	4,8	6,6				6,5
Anadolu Mahalli İdareler Meslek Lisesi					0,5	0,5
Anadolu Meslek Lisesi	1,9	3,4	1,5		0,0	1,9
Anadolu Moda Tasarımı Meslek Lisesi					-0,1	-0,1
Anadolu Otelcilik ve Turizm Meslek Lisesi			0,1	0,1		0,1
Anadolu Öğretmen Lisesi		6,4				6,4
Anadolu Sağlık Meslek Lisesi	2,2					2,2
Anadolu Sekreterlik Meslek Lisesi					0,3	0,3
Anadolu Teknik Lisesi	2,2					2,2
Anadolu Tekstil Meslek Lisesi	0,4					0,4
Anadolu Ticaret Meslek Lisesi	0,7		0,8		0,2	0,7
Askeri Lise		8,6				8,6
Çok Programlı Lise	0,6		0,6	0,0	-0,1	0,1
Denizcilik ve Su Ürünleri Meslek Lisesi				0,3		0,3
Endüstri Meslek Lisesi			0,4	0,2		0,2
Fen Lisesi		12,7				12,7
Geleneksel Türk Sanatları Kız Meslek Lisesi					0,0	0,0
İmam Hatip Lisesi			0,5	0,3	0,2	0,4
İnşaat Teknik Lisesi	1,1					1,1
Kız Meslek Lisesi				0,0	0,0	0,0
Kuaförlük Meslek Lisesi					-0,5	-0,5
Matbaa Meslek Lisesi					-0,2	-0,2
Normal Lise	2,0	6,7	0,8	0,4	0,0	0,8
Ortopedik Cihaz ve Protez Teknisyen Okulu					0,6	0,6
Özel Akşam Lisesi				-0,1	-0,1	-0,1
Özel Anadolu Otelcilik ve Turizm Meslek Lisesi				0,0	0,1	0,1
Özel Fen Lisesi	10,2	10,7				10,6
Özel Lise	1,9	4,1			0,2	2,2
Özel Ticaret Meslek Lisesi					-0,1	-0,1
Teknik Lise (Erkek)	1,6		1,2			1,5
Tekstil Meslek Lisesi				0,1		0,1
Tekstil Teknik Lisesi (Erkek)	0,8					0,8
Terzilik Meslek Lisesi					-0,1	-0,1
Ticaret Meslek Lisesi				0,0	-0,1	0,0
Veteriner Sağlık Meslek Lisesi					0,2	0,2
Y.D.A. İmam Hatip Lisesi		4,6				4,6
Y.D.A. Lise	2,6	4,3	1,9			3,7
Yabancı Dille Öğretim Yapan Özel Lise	2,5	5,4	1,6		0,2	3,1
Yapı Meslek Lisesi				0,1		0,1
Ziraat Meslek Lisesi					0,5	0,5
Toplam	2,3	5,5	0,8	0,1	0,0	1,9

Çizelge EK 9.4 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Göre
Biyoloji Net Ortalamaları

d23	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
A.Ö. Lisesi				0,2		0,2
Anadolu Denizcilik Meslek Lisesi			0,6			0,6
Anadolu Dış Ticaret Meslek Lisesi	0,0					0,0
Anadolu Güzel Sanatlar Lisesi					0,0	0,0
Anadolu Hazır Giyim ve Deri Hazır Giyim Lisesi					0,1	0,1
Anadolu İletişim Meslek Lisesi					0,0	0,0
Anadolu İmam Hatip Lisesi	1,4	1,0				1,3
Anadolu Kız Meslek Lisesi	0,4		0,3		0,1	0,2
Anadolu Lisesi	1,6	3,0				2,9
Anadolu Mahalli İdareler Meslek Lisesi					0,1	0,1
Anadolu Meslek Lisesi	0,6	1,3	0,7		0,1	0,7
Anadolu Moda Tasarımı Meslek Lisesi					0,2	0,2
Anadolu Otelcilik ve Turizm Meslek Lisesi			0,1	0,3		0,2
Anadolu Öğretmen Lisesi		2,9				2,9
Anadolu Sağlık Meslek Lisesi	2,0					2,0
Anadolu Sekreterlik Meslek Lisesi					0,2	0,2
Anadolu Teknik Lisesi	0,8					0,8
Anadolu Tekstil Meslek Lisesi	1,0					1,0
Anadolu Ticaret Meslek Lisesi	0,2		0,3		0,1	0,2
Askeri Lise		3,9				3,9
Çok Programlı Lise	0,7		0,3	0,2	0,1	0,2
Denizcilik ve Su Ürünleri Meslek Lisesi				0,3		0,3
Endüstri Meslek Lisesi			0,3	0,1		0,1
Fen Lisesi		7,5				7,5
Geleneksel Türk Sanatları Kız Meslek Lisesi					0,5	0,5
İmam Hatip Lisesi			0,2	0,2	0,1	0,2
İnşaat Teknik Lisesi	0,2					0,2
Kız Meslek Lisesi				0,0	0,1	0,1
Kuaförlük Meslek Lisesi					0,5	0,5
Matbaa Meslek Lisesi					0,0	0,0
Normal Lise	0,8	2,6	0,3	0,1	0,1	0,3
Ortopedik Cihaz ve Protez Teknisyen Okulu					0,9	0,9
Özel Akşam Lisesi				0,1	0,0	0,0
Özel Anadolu Otelcilik ve Turizm Meslek Lisesi				0,1	0,0	0,0
Özel Fen Lisesi	3,4	5,3				5,0
Özel Lise	0,7	1,6			0,1	0,8
Özel Ticaret Meslek Lisesi					0,1	0,1
Teknik Lise (Erkek)	0,4		0,4			0,4
Tekstil Meslek Lisesi				0,1		0,1
Tekstil Teknik Lisesi (Erkek)	0,1					0,1
Terzilik Meslek Lisesi					-0,3	-0,3
Ticaret Meslek Lisesi				0,1	0,0	0,1
Veteriner Sağlık Meslek Lisesi					0,6	0,6
Y.D.A. İmam Hatip Lisesi		1,9				1,9
Y.D.A. Lise	0,8	1,7	0,7			1,4
Yabancı Dille Öğretim Yapan Özel Lise	0,9	2,1	0,5		0,0	1,2
Yapı Meslek Lisesi				0,2		0,2
Ziraat Meslek Lisesi					0,5	0,5
Toplam	0,8	2,3	0,3	0,1	0,1	0,8

Çizelge EK 9.5 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Göre
Türkçe Net Ortalamaları

d25	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
A.Ö. Lisesi				20,7		20,7
Anadolu Denizcilik Meslek Lisesi			19,6			19,6
Anadolu Dış Ticaret Meslek Lisesi	33,8					33,8
Anadolu Güzel Sanatlar Lisesi					34,1	34,1
Anadolu Hazır Giyim ve Deri Hazır Giyim Lisesi					22,0	22,0
Anadolu İletişim Meslek Lisesi					32,3	32,3
Anadolu İmam Hatip Lisesi	34,4	34,7				34,5
Anadolu Kız Meslek Lisesi	25,0		26,9		24,8	25,1
Anadolu Lisesi	34,3	35,5				35,4
Anadolu Mahalli İdareler Meslek Lisesi					27,1	27,1
Anadolu Meslek Lisesi	16,5	24,0	20,4		23,8	19,2
Anadolu Moda Tasarımı Meslek Lisesi					27,2	27,2
Anadolu Otelcilik ve Turizm Meslek Lisesi			27,3	24,1		25,7
Anadolu Öğretmen Lisesi		35,4				35,4
Anadolu Sağlık Meslek Lisesi	28,1					28,1
Anadolu Sekreterlik Meslek Lisesi					27,4	27,4
Anadolu Teknik Lisesi	19,2					19,2
Anadolu Tekstil Meslek Lisesi	23,9					23,9
Anadolu Ticaret Meslek Lisesi	31,5		27,5		30,9	29,1
Askeri Lise		28,5				28,5
Çok Programlı Lise	23,9		18,9	15,3	11,7	15,5
Denizcilik ve Su Ürünleri Meslek Lisesi				14,4		14,4
Endüstri Meslek Lisesi			12,5	10,7		10,8
Fen Lisesi		41,0				41,0
Geleneksel Türk Sanatları Kız Meslek Lisesi					18,0	18,0
İmam Hatip Lisesi			22,8	20,9	21,5	22,1
İnşaat Teknik Lisesi	13,2					13,2
Kız Meslek Lisesi				15,8	14,6	15,0
Kuaförlük Meslek Lisesi					19,1	19,1
Matbaa Meslek Lisesi					13,2	13,2
Normal Lise	14,9	34,8	19,9	16,1	18,9	19,8
Ortopedik Cihaz ve Protez Teknisyen Okulu					17,0	17,0
Özel Akşam Lisesi				14,6	16,8	16,7
Özel Anadolu Otelcilik ve Turizm Meslek Lisesi				16,4	16,9	16,8
Özel Fen Lisesi	29,1	34,6				33,9
Özel Lise	28,3	31,5			20,2	27,9
Özel Ticaret Meslek Lisesi					12,3	12,3
Teknik Lise (Erkek)	11,9		13,6			12,4
Tekstil Meslek Lisesi				13,3		13,3
Tekstil Teknik Lisesi (Erkek)	15,1					15,1
Terzilik Meslek Lisesi					13,5	13,5
Ticaret Meslek Lisesi				16,5	17,2	16,6
Veteriner Sağlık Meslek Lisesi					13,9	13,9
Y.D.A. İmam Hatip Lisesi		36,0				36,0
Y.D.A. Lise	27,6	32,0	26,2			30,4
Yabancı Dille Öğretim Yapan Özel Lise	29,5	34,9	30,2		26,2	31,1
Yapı Meslek Lisesi				11,3		11,3
Ziraat Meslek Lisesi					19,2	19,2
Toplam	24,6	33,3	20,7	14,8	19,0	23,1

Çizelge EK 9.6 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Göre Tarih Net Ortalamaları

d27	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
A.Ö. Lisesi				7,3		7,3
Anadolu Denizcilik Meslek Lisesi			6,6			6,6
Anadolu Dış Ticaret Meslek Lisesi	11,5					11,5
Anadolu Güzel Sanatlar Lisesi					10,9	10,9
Anadolu Hazır Giyim ve Deri Hazır Giyim Lisesi				5,6		5,6
Anadolu İletişim Meslek Lisesi				9,9		9,9
Anadolu İmam Hatip Lisesi	11,4	12,5				11,8
Anadolu Kız Meslek Lisesi	5,1		7,0		7,1	6,9
Anadolu Lisesi	12,0	12,0				12,0
Anadolu Mahalli İdareler Meslek Lisesi				8,6		8,6
Anadolu Meslek Lisesi	5,2	7,5	6,5		7,3	6,1
Anadolu Moda Tasarımı Meslek Lisesi				6,7		6,7
Anadolu Otelcilik ve Turizm Meslek Lisesi			8,7	8,2		8,5
Anadolu Öğretmen Lisesi		12,4				12,4
Anadolu Sağlık Meslek Lisesi	5,8					5,8
Anadolu Sekreterlik Meslek Lisesi				7,9		7,9
Anadolu Teknik Lisesi	6,5					6,5
Anadolu Tekstil Meslek Lisesi	7,1					7,1
Anadolu Ticaret Meslek Lisesi	9,9		8,9		9,5	9,3
Askeri Lise		12,2				12,2
Çok Programlı Lise	8,3		6,0	4,5	3,5	4,8
Denizcilik ve Su Ürünleri Meslek Lisesi				4,2		4,2
Endüstri Meslek Lisesi			4,0	3,4		3,4
Fen Lisesi		15,3				15,3
Geleneksel Türk Sanatları Kız Meslek Lisesi					3,0	3,0
İmam Hatip Lisesi			7,9	6,9	7,2	7,5
İnşaat Teknik Lisesi	2,8					2,8
Kız Meslek Lisesi				3,6	3,7	3,7
Kuaförlük Meslek Lisesi					6,4	6,4
Matbaa Meslek Lisesi					3,3	3,3
Normal Lise	5,0	12,3	6,2	4,7	5,3	6,2
Ortopedik Cihaz ve Protez Teknisyen Okulu					5,9	5,9
Özel Akşam Lisesi				4,9	5,9	5,8
Özel Anadolu Otelcilik ve Turizm Meslek Lisesi				4,4	5,2	5,1
Özel Fen Lisesi	6,7	11,5				10,9
Özel Lise	8,3	9,7			5,5	8,2
Özel Ticaret Meslek Lisesi					4,5	4,5
Teknik Lise (Erkek)	3,3		3,6			3,4
Tekstil Meslek Lisesi				3,8		3,8
Tekstil Teknik Lisesi (Erkek)	3,0					3,0
Terzilik Meslek Lisesi					2,8	2,8
Ticaret Meslek Lisesi				5,1	5,4	5,2
Veteriner Sağlık Meslek Lisesi					4,0	4,0
Y.D.A. İmam Hatip Lisesi		11,3				11,3
Y.D.A. Lise	8,0	10,0	7,4			9,3
Yabancı Dille Öğretim Yapan Özel Lise	9,0	11,1	9,1		7,7	9,6
Yapı Meslek Lisesi				3,2		3,2
Ziraat Meslek Lisesi					8,3	8,3
Toplam	7,4	10,8	6,5	4,5	5,7	7,2

Çizelge EK 9.7 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Göre
Coğrafya Net Ortalamaları

d29	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
A.Ö. Lisesi				4,6		4,6
Anadolu Denizcilik Meslek Lisesi			3,6			3,6
Anadolu Dış Ticaret Meslek Lisesi	6,0					6,0
Anadolu Güzel Sanatlar Lisesi					7,1	7,1
Anadolu Hazır Giyim ve Deri Hazır Giyim Lisesi					4,4	4,4
Anadolu İletişim Meslek Lisesi					6,3	6,3
Anadolu İmam Hatip Lisesi	7,5	7,5				7,5
Anadolu Kız Meslek Lisesi	4,7		4,1		4,1	4,1
Anadolu Lisesi	7,6	8,0				8,0
Anadolu Mahalli İdareler Meslek Lisesi					4,6	4,6
Anadolu Meslek Lisesi	2,6	3,7	3,7		4,3	3,2
Anadolu Moda Tasarımı Meslek Lisesi					5,1	5,1
Anadolu Otelcilik ve Turizm Meslek Lisesi			6,2	5,2		5,7
Anadolu Öğretmen Lisesi		8,1				8,1
Anadolu Sağlık Meslek Lisesi	3,4					3,4
Anadolu Sekreterlik Meslek Lisesi					4,5	4,5
Anadolu Teknik Lisesi	3,7					3,7
Anadolu Tekstil Meslek Lisesi	4,3					4,3
Anadolu Ticaret Meslek Lisesi	5,6		4,5		5,3	5,0
Askeri Lise		7,8				7,8
Çok Programlı Lise	5,3		4,0	3,2	2,2	3,2
Denizcilik ve Su Ürünleri Meslek Lisesi				3,1		3,1
Endüstri Meslek Lisesi			2,5	2,4		2,5
Fen Lisesi		12,1				12,1
Geleneksel Türk Sanatları Kız Meslek Lisesi					0,9	0,9
İmam Hatip Lisesi			4,3	4,0	4,3	4,3
İnşaat Teknik Lisesi	1,6					1,6
Kız Meslek Lisesi				2,5	2,4	2,4
Kuaförlük Meslek Lisesi					3,7	3,7
Matbaa Meslek Lisesi					2,6	2,6
Normal Lise	3,6	7,6	3,8	3,3	3,7	3,8
Ortopedik Cihaz ve Protez Teknisyen Okulu					2,8	2,8
Özel Akşam Lisesi				3,5	4,4	4,4
Özel Anadolu Otelcilik ve Turizm Meslek Lisesi				3,0	3,5	3,4
Özel Fen Lisesi	6,8	7,5				7,4
Özel Lise	5,0	5,9			2,7	4,9
Özel Ticaret Meslek Lisesi					3,0	3,0
Teknik Lise (Erkek)	2,2		2,3			2,2
Tekstil Meslek Lisesi				2,4		2,4
Tekstil Teknik Lisesi (Erkek)	1,8					1,8
Terzilik Meslek Lisesi					1,9	1,9
Ticaret Meslek Lisesi				3,1	3,1	3,1
Veteriner Sağlık Meslek Lisesi					3,7	3,7
Y.D.A. İmam Hatip Lisesi		6,8				6,8
Y.D.A. Lise	4,2	5,7	4,3			5,2
Yabancı Dille Öğretim Yapan Özel Lise	5,4	7,4	5,4		4,5	6,0
Yapı Meslek Lisesi				2,1		2,1
Ziraat Meslek Lisesi					6,4	6,4
Toplam	4,3	6,7	3,9	2,9	3,6	4,4

Çizelge EK 9.8 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Göre
Felsefe Net Ortalamaları

d31	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
A.Ö. Lisesi				3,4		3,4
Anadolu Denizcilik Meslek Lisesi			1,4			1,4
Anadolu Dış Ticaret Meslek Lisesi	5,9					5,9
Anadolu Güzel Sanatlar Lisesi					7,0	7,0
Anadolu Hazır Giyim ve Deri Hazır Giyim Lisesi					2,9	2,9
Anadolu İletişim Meslek Lisesi					6,2	6,2
Anadolu İmam Hatip Lisesi	5,7	5,5				5,7
Anadolu Kız Meslek Lisesi	2,3		4,0		4,2	4,0
Anadolu Lisesi	4,7	5,3				5,3
Anadolu Mahalli İdareler Meslek Lisesi					4,5	4,5
Anadolu Meslek Lisesi	1,2	2,2	2,1		3,6	1,8
Anadolu Moda Tasarımı Meslek Lisesi					4,5	4,5
Anadolu Otelcilik ve Turizm Meslek Lisesi			4,4	4,0		4,2
Anadolu Öğretmen Lisesi		4,9				4,9
Anadolu Sağlık Meslek Lisesi	3,7					3,7
Anadolu Sekreterlik Meslek Lisesi					4,8	4,8
Anadolu Teknik Lisesi	2,0					2,0
Anadolu Tekstil Meslek Lisesi	3,0					3,0
Anadolu Ticaret Meslek Lisesi	5,2		4,6		4,9	4,8
Askeri Lise		3,2				3,2
Çok Programlı Lise	2,6		2,3	1,8	1,4	1,8
Denizcilik ve Su Ürünleri Meslek Lisesi				1,5		1,5
Endüstri Meslek Lisesi			0,9	0,8		0,8
Fen Lisesi		7,5				7,5
Geleneksel Türk Sanatları Kız Meslek Lisesi					0,2	0,2
İmam Hatip Lisesi			3,7	3,3	3,6	3,6
İnşaat Teknik Lisesi	0,6					0,6
Kız Meslek Lisesi				2,0	1,9	1,9
Kuaförlük Meslek Lisesi					2,2	2,2
Matbaa Meslek Lisesi					1,1	1,1
Normal Lise	1,8	4,7	2,7	1,8	2,4	2,7
Ortopedik Cihaz ve Protez Teknisyen Okulu					2,4	2,4
Özel Akşam Lisesi				2,0	2,6	2,5
Özel Anadolu Otelcilik ve Turizm Meslek Lisesi				2,0	2,5	2,4
Özel Fen Lisesi	2,9	4,2				4,1
Özel Lise	4,3	4,5			2,2	4,1
Özel Ticaret Meslek Lisesi					1,6	1,6
Teknik Lise (Erkek)	0,7		0,9			0,8
Tekstil Meslek Lisesi				1,0		1,0
Tekstil Teknik Lisesi (Erkek)	0,6					0,6
Terzilik Meslek Lisesi					1,6	1,6
Ticaret Meslek Lisesi				2,1	2,4	2,1
Veteriner Sağlık Meslek Lisesi					1,4	1,4
Y.D.A. İmam Hatip Lisesi		5,3				5,3
Y.D.A. Lise	3,6	4,6	3,8			4,3
Yabancı Dille Öğretim Yapan Özel Lise	4,3	5,3	4,5		4,2	4,7
Yapı Meslek Lisesi				0,6		0,6
Ziraat Meslek Lisesi					1,9	1,9
Toplam	3,2	4,8	2,8	1,7	2,8	3,2

Çizelge EK 9.9 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Göre
İngilizce Net Ortalamaları

d33	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
A.Ö. Lisesi				33,3		33,3
Anadolu Denizcilik Meslek Lisesi			0,0			0,0
Anadolu Dış Ticaret Meslek Lisesi	0,0					0,0
Anadolu Güzel Sanatlar Lisesi					0,0	0,0
Anadolu Hazır Giyim ve Deri Hazır Giyim Lisesi					0,0	0,0
Anadolu İletişim Meslek Lisesi					0,0	0,0
Anadolu İmam Hatip Lisesi	0,0	73,5				24,5
Anadolu Kız Meslek Lisesi	0,0		11,3		8,0	7,9
Anadolu Lisesi	0,0	23,3				21,8
Anadolu Mahalli İdareler Meslek Lisesi					0,0	0,0
Anadolu Meslek Lisesi	0,0	0,0	8,4		0,0	1,8
Anadolu Moda Tasarımı Meslek Lisesi					0,0	0,0
Anadolu Otelcilik ve Turizm Meslek Lisesi			0,0	0,0		0,0
Anadolu Öğretmen Lisesi		82,0				82,0
Anadolu Sağlık Meslek Lisesi	0,0					0,0
Anadolu Sekreterlik Meslek Lisesi					0,0	0,0
Anadolu Teknik Lisesi	0,0					0,0
Anadolu Tekstil Meslek Lisesi	0,0					0,0
Anadolu Ticaret Meslek Lisesi	0,0		18,0		0,0	10,3
Askeri Lise		0,0				0,0
Çok Programlı Lise	0,0		0,0	0,0	0,0	0,0
Denizcilik ve Su Ürünleri Meslek Lisesi				0,0		0,0
Endüstri Meslek Lisesi			0,0	0,0		0,0
Fen Lisesi		0,0				0,0
Geleneksel Türk Sanatları Kız Meslek Lisesi					0,0	0,0
İmam Hatip Lisesi			0,0	0,0	0,0	0,0
İnşaat Teknik Lisesi	0,0					0,0
Kız Meslek Lisesi				0,0	0,0	0,0
Kuaförlük Meslek Lisesi					0,0	0,0
Matbaa Meslek Lisesi					0,0	0,0
Normal Lise	0,0	82,8	2,0	0,0	0,0	2,4
Ortopedik Cihaz ve Protez Teknisyen Okulu					0,0	0,0
Özel Akşam Lisesi				0,0	0,0	0,0
Özel Anadolu Otelcilik ve Turizm Meslek Lisesi				0,0	0,0	0,0
Özel Fen Lisesi	0,0	0,0				0,0
Özel Lise	20,3	44,9			0,0	23,3
Özel Ticaret Meslek Lisesi					0,0	0,0
Teknik Lise (Erkek)	0,0		0,0			0,0
Tekstil Meslek Lisesi				0,0		0,0
Tekstil Teknik Lisesi (Erkek)	0,0					0,0
Terzilik Meslek Lisesi					0,0	0,0
Ticaret Meslek Lisesi				0,0	0,0	0,0
Veteriner Sağlık Meslek Lisesi					0,0	0,0
Y.D.A. İmam Hatip Lisesi		0,0				0,0
Y.D.A. Lise	6,0	33,1	51,8			24,4
Yabancı Dille Öğretim Yapan Özel Lise	20,9	39,0	54,9		23,7	32,0
Yapı Meslek Lisesi				0,0		0,0
Ziraat Meslek Lisesi					0,0	0,0
Toplam	9,2	30,0	6,1	0,4	1,8	10,4

Çizelge EK 9.10 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Göre
Fransızca Net Ortalamaları

d35	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
A.Ö. Lisesi				0,0		0,0
Anadolu Denizcilik Meslek Lisesi			0,0			0,0
Anadolu Dış Ticaret Meslek Lisesi	0,0					0,0
Anadolu Güzel Sanatlar Lisesi					0,0	0,0
Anadolu Hazır Giyim ve Deri Hazır Giyim Lisesi					0,0	0,0
Anadolu İletişim Meslek Lisesi					0,0	0,0
Anadolu İmam Hatip Lisesi	0,0	0,0				0,0
Anadolu Kız Meslek Lisesi	0,0		0,0		0,0	0,0
Anadolu Lisesi	0,0	5,4				5,1
Anadolu Mahalli İdareler Meslek Lisesi					0,0	0,0
Anadolu Meslek Lisesi	0,0	0,0	0,0		0,0	0,0
Anadolu Moda Tasarımı Meslek Lisesi					0,0	0,0
Anadolu Otelcilik ve Turizm Meslek Lisesi			0,0	0,0		0,0
Anadolu Öğretmen Lisesi		0,0				0,0
Anadolu Sağlık Meslek Lisesi	0,0					0,0
Anadolu Sekreterlik Meslek Lisesi					0,0	0,0
Anadolu Teknik Lisesi	0,0					0,0
Anadolu Tekstil Meslek Lisesi	0,0					0,0
Anadolu Ticaret Meslek Lisesi	0,0		0,0		0,0	0,0
Askeri Lise		0,0				0,0
Çok Programlı Lise	0,0		0,0	0,0	0,0	0,0
Denizcilik ve Su Ürünleri Meslek Lisesi				0,0		0,0
Endüstri Meslek Lisesi			0,0	0,0		0,0
Fen Lisesi		0,0				0,0
Geleneksel Türk Sanatları Kız Meslek Lisesi					0,0	0,0
İmam Hatip Lisesi			0,0	0,0	0,0	0,0
İnşaat Teknik Lisesi	0,0					0,0
Kız Meslek Lisesi				0,0	0,0	0,0
Kuaförlük Meslek Lisesi					0,0	0,0
Matbaa Meslek Lisesi					0,0	0,0
Normal Lise	0,0	0,0	0,0	0,0	0,0	0,0
Ortopedik Cihaz ve Protez Teknisyen Okulu					0,0	0,0
Özel Akşam Lisesi				0,0	0,0	0,0
Özel Anadolu Otelcilik ve Turizm Meslek Lisesi				0,0	0,0	0,0
Özel Fen Lisesi	0,0	0,0				0,0
Özel Lise	0,0	0,0			0,0	0,0
Özel Ticaret Meslek Lisesi					0,0	0,0
Teknik Lise (Erkek)	0,0		0,0			0,0
Tekstil Meslek Lisesi				0,0		0,0
Tekstil Teknik Lisesi (Erkek)	0,0					0,0
Terzilik Meslek Lisesi					0,0	0,0
Ticaret Meslek Lisesi				0,0	0,0	0,0
Veteriner Sağlık Meslek Lisesi					0,0	0,0
Y.D.A. İmam Hatip Lisesi		0,0				0,0
Y.D.A. Lise	0,0	0,0	0,0			0,0
Yabancı Dille Öğretim Yapan Özel Lise	0,0	13,8	0,0		0,0	4,3
Yapı Meslek Lisesi				0,0		0,0
Ziraat Meslek Lisesi					0,0	0,0
Toplam	0,0	3,6	0,0	0,0	0,0	0,8

Çizelge EK 9.11 ÖSYM'nin Tanımladığı Lise Türlerinin Kümelere Göre
Almanca Net Ortalamaları

d37	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
A.Ö. Lisesi				49,4		49,4
Anadolu Denizcilik Meslek Lisesi			0,0			0,0
Anadolu Dış Ticaret Meslek Lisesi	0,0					0,0
Anadolu Güzel Sanatlar Lisesi					0,0	0,0
Anadolu Hazır Giyim ve Deri Hazır Giyim Lisesi					0,0	0,0
Anadolu İletişim Meslek Lisesi					0,0	0,0
Anadolu İmam Hatip Lisesi	0,0	0,0				0,0
Anadolu Kız Meslek Lisesi	0,0		0,0		0,0	0,0
Anadolu Lisesi	0,0	9,4				8,8
Anadolu Mahalli İdareler Meslek Lisesi					0,0	0,0
Anadolu Meslek Lisesi	0,0	0,0	0,0		0,0	0,0
Anadolu Moda Tasarımı Meslek Lisesi					0,0	0,0
Anadolu Otelcilik ve Turizm Meslek Lisesi			0,0	0,0		0,0
Anadolu Öğretmen Lisesi		0,0				0,0
Anadolu Sağlık Meslek Lisesi	0,0					0,0
Anadolu Sekreterlik Meslek Lisesi					0,0	0,0
Anadolu Teknik Lisesi	0,0					0,0
Anadolu Tekstil Meslek Lisesi	0,0					0,0
Anadolu Ticaret Meslek Lisesi	0,0		0,0		0,0	0,0
Askeri Lise		0,0				0,0
Çok Programlı Lise	0,0		0,0	0,0	0,0	0,0
Denizcilik ve Su Ürünleri Meslek Lisesi				0,0		0,0
Endüstri Meslek Lisesi			0,0	0,0		0,0
Fen Lisesi		0,0				0,0
Geleneksel Türk Sanatları Kız Meslek Lisesi					0,0	0,0
İmam Hatip Lisesi			0,0	0,0	0,0	0,0
İnşaat Teknik Lisesi	0,0					0,0
Kız Meslek Lisesi				0,0	0,0	0,0
Kuaförlük Meslek Lisesi					0,0	0,0
Matbaa Meslek Lisesi					0,0	0,0
Normal Lise	0,0	0,0	0,0	0,0	0,0	0,0
Ortopedik Cihaz ve Protez Teknisyen Okulu					0,0	0,0
Özel Akşam Lisesi				0,0	0,0	0,0
Özel Anadolu Otelcilik ve Turizm Meslek Lisesi				0,0	0,0	0,0
Özel Fen Lisesi	0,0	0,0				0,0
Özel Lise	0,0	0,0			0,0	0,0
Özel Ticaret Meslek Lisesi					0,0	0,0
Teknik Lise (Erkek)	0,0		0,0			0,0
Tekstil Meslek Lisesi				0,0		0,0
Tekstil Teknik Lisesi (Erkek)	0,0					0,0
Terzilik Meslek Lisesi					0,0	0,0
Ticaret Meslek Lisesi				0,0	0,0	0,0
Veteriner Sağlık Meslek Lisesi					0,0	0,0
Y.D.A. İmam Hatip Lisesi		0,0				0,0
Y.D.A. Lise	0,0	0,0	0,0			0,0
Yabancı Dille Öğretim Yapan Özel Lise	0,0	3,6	0,0		0,0	1,1
Yapı Meslek Lisesi				0,0		0,0
Ziraat Meslek Lisesi					0,0	0,0
Toplam	0,0	2,7	0,0	0,6	0,0	0,6

EK 10 D40 Değişkenindeki Lise Türlerinin Testlere Göre Kümelerdeki Net Ortalamaları

Çizelge EK 10.1 D40 Değişkenindeki Lise Türlerinin Kümelere Göre Matematik Net Ortalamaları

d17	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Erkek	9,5		8,6			9,2
Karma	13,2	24,8	6,2	1,2	1,8	10,7
Kız	13,3	20,1	5,6	0,2	0,8	3,5
Toplam	12,7	24,7	6,3	1,1	1,5	10,1

Çizelge EK 10.2 D40 Değişkenindeki Lise Türlerinin Kümelere Göre Fizik Net Ortalamaları

d19	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Erkek	3,1		2,9			3,0
Karma	3,7	7,5	1,5	0,4	0,2	3,0
Kız	2,6	5,2	1,0	0,1	0,1	0,7
Toplam	3,5	7,5	1,5	0,3	0,2	2,8

Çizelge EK 10.3 D40 Değişkenindeki Lise Türlerinin Kümelere Göre Kimya Net Ortalamaları

d21	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Erkek	1,6		1,2			1,5
Karma	2,4	5,5	0,8	0,1	0,0	2,0
Kız	2,1	4,0	0,6	0,0	0,0	0,5
Toplam	2,3	5,5	0,8	0,1	0,0	1,9

Çizelge EK 10.4 D40 Değişkenindeki Lise Türlerinin Kümelere Göre Biyoloji Net Ortalamaları

d23	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Erkek	0,4		0,4			0,4
Karma	0,8	2,3	0,3	0,1	0,1	0,8
Kız	0,7	1,7	0,3	0,0	0,1	0,3
Toplam	0,8	2,3	0,3	0,1	0,1	0,8

Çizelge EK 10.5 D40 Değişkenindeki Lise Türlerinin Kümelere Göre Türkçe Net Ortalamaları

d25	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Erkek	12,1		13,6			12,5
Karma	26,4	33,4	20,8	14,6	19,4	23,8
Kız	31,2	32,3	24,5	15,8	18,1	20,6
Toplam	24,6	33,3	20,7	14,8	19,0	23,1

Çizelge EK 10.6 D40 Değişkenindeki Lise Türlerinin Kümelere Göre Tarih Net Ortalamaları

d27	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Erkek	3,3		3,6			3,3
Karma	8,0	10,8	6,6	4,6	6,1	7,5
Kız	8,8	8,9	6,5	3,6	4,8	5,4
Toplam	7,4	10,8	6,5	4,5	5,7	7,2

Çizelge EK 10.7 D40 Değişkenindeki Lise Türlerinin Kümelere Göre Coğrafya Net Ortalamaları

d29	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Erkek	2,2		2,3			2,2
Karma	4,6	6,8	4,0	3,0	4,0	4,6
Kız	5,7	5,1	4,2	2,5	2,9	3,4
Toplam	4,3	6,7	3,9	2,9	3,6	4,4

Çizelge EK 10.8 D40 Değişkenindeki Lise Türlerinin Kümelere Göre Felsefe Net Ortalamaları

d31	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Erkek	0,7		0,9			0,8
Karma	3,6	4,8	2,8	1,7	2,9	3,3
Kız	4,6	4,8	3,6	2,0	2,6	2,9
Toplam	3,2	4,8	2,8	1,7	2,8	3,2

Çizelge EK 10.9 D40 Değişkenindeki Lise Türlerinin Kümelere Göre İngilizce Net Ortalamaları

d33	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Erkek	0,0		0,0			0,0
Karma	10,5	29,7	6,3	0,4	1,5	11,2
Kız	15,4	46,4	6,1	0,0	2,7	6,3
Toplam	9,2	30,0	6,1	0,4	1,8	10,4

Çizelge EK 10.10 D40 Değişkenindeki Lise Türlerinin Kümelere Göre Fransızca Net Ortalamaları

d35	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Erkek	0,0		0,0			0,0
Karma	0,0	3,7	0,0	0,0	0,0	0,9
Kız	0,0	0,0	0,0	0,0	0,0	0,0
Toplam	0,0	3,6	0,0	0,0	0,0	0,8

Çizelge EK 10.11 D40 Değişkenindeki Lise Türlerinin Kümelere Göre Almanca Net Ortalamaları

d37	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Erkek	0,0		0,0			0,0
Karma	0,0	2,7	0,0	0,6	0,0	0,7
Kız	0,0	0,0	0,0	0,0	0,0	0,0
Toplam	0,0	2,7	0,0	0,6	0,0	0,6

EK 11 D41 Değişkenindeki Lise Türlerinin Testlere Göre Kümelerdeki Net Ortalamaları

Çizelge EK 11.1 D40 Değişkenindeki Lise Türlerinin Kümelere Göre Matematik Net Ortalamaları

d17	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Devlet	12,2	24,0	5,9	1,1	1,5	8,9
Özel	13,3	26,6	11,9	0,5	1,5	14,1
Toplam	12,7	24,7	6,3	1,1	1,5	10,1

Çizelge EK 11.2 D40 Değişkenindeki Lise Türlerinin Kümelere Göre Fizik Net Ortalamaları

d19	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Devlet	3,6	7,1	1,4	0,3	0,2	2,5
Özel	3,5	8,6	2,7	0,0	0,2	4,0
Toplam	3,5	7,5	1,5	0,3	0,2	2,8

Çizelge EK 11.3 D40 Değişkenindeki Lise Türlerinin Kümelere Göre Kimya Net Ortalamaları

d21	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Devlet	2,2	5,2	0,8	0,1	0,0	1,6
Özel	2,4	6,3	1,6	-0,1	0,0	2,8
Toplam	2,3	5,5	0,8	0,1	0,0	1,9

Çizelge EK 11.4 D40 Değişkenindeki Lise Türlerinin Kümelere Göre Biyoloji Net Ortalamaları

d23	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Devlet	0,7	2,1	0,3	0,1	0,1	0,7
Özel	0,8	2,7	0,5	0,1	0,0	1,1
Toplam	0,8	2,3	0,3	0,1	0,1	0,8

Çizelge EK 11.5 D40 Değişkenindeki Lise Türlerinin Kümelere Göre Türkçe Net Ortalamaları

d25	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Devlet	20,6	32,9	20,1	14,8	19,5	21,7
Özel	28,7	34,4	29,5	15,5	17,9	27,9
Toplam	24,6	33,3	20,7	14,8	19,0	23,1

Çizelge EK 11.6 D40 Değişkenindeki Lise Türlerinin Kümelere Göre Tarih Net Ortalamaları

d27	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Devlet	6,2	10,7	6,3	4,5	5,6	6,8
Özel	8,6	11,0	8,7	4,6	5,8	8,6
Toplam	7,4	10,8	6,5	4,5	5,7	7,2

Çizelge EK 11.7 D40 Değişkenindeki Lise Türlerinin Kümelere Göre Coğrafya Net Ortalamaları

d29	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Devlet	3,6	6,5	3,8	2,9	3,5	4,1
Özel	5,1	7,2	5,2	3,2	3,9	5,4
Toplam	4,3	6,7	3,9	2,9	3,6	4,4

Çizelge EK 11.8 D40 Değişkenindeki Lise Türlerinin Kümelere Göre Felsefe Net Ortalamaları

d31	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Devlet	2,3	4,8	2,7	1,7	2,9	2,9
Özel	4,2	5,0	4,3	2,0	2,6	4,1
Toplam	3,2	4,8	2,8	1,7	2,8	3,2

Çizelge EK 11.9 D40 Değişkenindeki Lise Türlerinin Kümelere Göre İngilizce Net Ortalamaları

d33	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Devlet	1,6	29,5	2,8	0,4	1,1	7,2
Özel	16,9	31,4	52,2	0,0	3,4	20,9
Toplam	9,2	30,0	6,1	0,4	1,8	10,4

Çizelge EK 11.10 D40 Değişkenindeki Lise Türlerinin Kümelere Göre Fransızca Net Ortalamaları

d35	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Devlet	0,0	1,6	0,0	0,0	0,0	0,3
Özel	0,0	8,9	0,0	0,0	0,0	2,3
Toplam	0,0	3,6	0,0	0,0	0,0	0,8

Çizelge EK 11.11 D40 Değişkenindeki Lise Türlerinin Kümelere Göre Almanca Net Ortalamaları

d37	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Devlet	0,0	2,8	0,0	0,6	0,0	0,7
Özel	0,0	2,3	0,0	0,0	0,0	0,6
Toplam	0,0	2,7	0,0	0,6	0,0	0,6

EK 12 D42 Değişkenindeki Lise Türlerinin Testlere Göre Kümelerdeki Net Ortalamaları

Çizelge EK 12.1 D42 Değişkenindeki Lise Türlerinin Kümelere Göre Matematik Net Ortalamaları

d17	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Anadolu	24,8	28,9				28,7
Fen		41,0				41,0
Meslek	9,3	29,2	5,9	1,1	0,8	3,4
Normal	8,8	28,4	5,6	2,1	2,2	5,5
Özel	11,1	21,1			3,5	12,4
Özel Fen	31,7	36,6				36,0
Yabancı Dille Eğitim Yapan Özel Lise	14,9	24,7	12,4		3,9	17,0
Y.D.A.	17,0	21,4	13,8			20,4
Y.D.A. Özel	10,1	27,0	6,5			10,9
Anadolu Meslek	10,8	16,0	8,7	4,2	2,5	8,3
Akşam Lisesi				0,1	0,9	0,8
Açık Öğretim				2,0		2,0
Özel Meslek				0,9	0,6	0,7
Toplam	12,7	24,7	6,3	1,1	1,5	10,1

Çizelge EK 12.2 D42 Değişkenindeki Lise Türlerinin Kümelere Göre Fizik Net Ortalamaları

d19	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Anadolu	7,2	9,3				9,2
Fen		17,0				17,0
Meslek	3,0	9,6	1,6	0,3	0,1	1,0
Normal	2,6	9,3	1,3	0,6	0,2	1,3
Özel	2,9	6,2			0,7	3,4
Özel Fen	10,4	14,7				14,2
Yabancı Dille Eğitim Yapan Özel Lise	3,9	7,3	2,7		0,8	4,6
Y.D.A.	4,2	5,6	3,0			5,3
Y.D.A. Özel	2,7	10,6	2,6			3,1
Anadolu Meslek	3,5	5,7	2,4	0,5	0,3	2,4
Akşam Lisesi				0,0	0,0	0,0
Açık Öğretim				0,4		0,4
Özel Meslek				0,1	0,1	0,1
Toplam	3,5	7,5	1,5	0,3	0,2	2,8

Çizelge EK 12.3 D42 Değişkenindeki Lise Türlerinin Kümelere Göre
Kimya Net Ortalamaları

d21	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Anadolu	4,8	6,6				6,5
Fen		12,7				12,7
Meslek	1,5	7,1	0,7	0,1	0,0	0,5
Normal	2,0	6,7	0,8	0,4	0,0	0,8
Özel	1,9	4,1			0,2	2,2
Özel Fen	10,2	10,7				10,6
Yabancı Dille Eğitim Yapan Özel Lise	2,5	5,4	1,6		0,2	3,1
Y.D.A.	3,1	4,3	2,4			4,0
Y.D.A. Özel	2,0	8,6	1,5			2,3
Anadolu Meslek	1,9	3,3	1,1	0,1	0,1	1,3
Akşam Lisesi				-0,1	-0,1	-0,1
Açık Öğretim				0,2		0,2
Özel Meslek				0,0	0,0	0,0
Toplam	2,3	5,5	0,8	0,1	0,0	1,9

Çizelge EK 12.4 D42 Değişkenindeki Lise Türlerinin Kümelere Göre
Biyoloji Net Ortalamaları

d23	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Anadolu	1,6	3,0				2,9
Fen		7,5				7,5
Meslek	0,4	3,2	0,3	0,1	0,1	0,2
Normal	0,8	2,6	0,3	0,1	0,1	0,3
Özel	0,7	1,6			0,1	0,8
Özel Fen	3,4	5,3				5,0
Yabancı Dille Eğitim Yapan Özel Lise	0,9	2,1	0,5		0,0	1,2
Y.D.A.	1,0	1,6	0,9			1,5
Y.D.A. Özel	0,6	3,4	0,5			0,8
Anadolu Meslek	0,7	1,2	0,4	0,3	0,1	0,5
Akşam Lisesi				0,1	0,0	0,0
Açık Öğretim				0,2		0,2
Özel Meslek				0,1	0,0	0,0
Toplam	0,8	2,3	0,3	0,1	0,1	0,8

Çizelge EK 12.5 D42 Değişkenindeki Lise Türlerinin Kümelere Göre
Türkçe Net Ortalamaları

d25	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Anadolu	34,3	35,5				35,4
Fen		41,0				41,0
Meslek	12,7	32,1	18,6	14,5	16,9	15,8
Normal	14,9	34,8	19,9	16,1	18,9	19,8
Özel	28,3	31,5			20,2	27,9
Özel Fen	29,1	34,6				33,9
Yabancı Dille Eğitim Yapan Özel Lise	29,5	34,9	30,2		26,2	31,1
Y.D.A.	28,2	32,0	31,6			31,2
Y.D.A. Özel	26,9	35,0	20,8			27,0
Anadolu Meslek	21,6	26,7	23,7	24,1	25,7	23,6
Akşam Lisesi				14,6	16,8	16,7
Açık Öğretim				20,7		20,7
Özel Meslek				16,4	14,6	14,8
Toplam	24,6	33,3	20,7	14,8	19,0	23,1

Çizelge EK 12.6 D42 Değişkenindeki Lise Türlerinin Kümelere Göre
Tarih Net Ortalamaları

d27	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Anadolu	12,0	12,0				12,0
Fen		15,3				15,3
Meslek	3,5	12,0	6,0	4,4	4,9	4,8
Normal	5,0	12,3	6,2	4,7	5,3	6,2
Özel	8,3	9,7			5,5	8,2
Özel Fen	6,7	11,5				10,9
Yabancı Dille Eğitim Yapan Özel Lise	9,0	11,1	9,1		7,7	9,6
Y.D.A.	8,2	10,0	10,5			9,6
Y.D.A. Özel	7,7	10,1	4,3			7,7
Anadolu Meslek	6,8	8,7	7,5	8,2	7,5	7,3
Akşam Lisesi				4,9	5,9	5,8
Açık Öğretim				7,3		7,3
Özel Meslek				4,4	4,9	4,8
Toplam	7,4	10,8	6,5	4,5	5,7	7,2

Çizelge EK 12.7 D42 Değişkenindeki Lise Türlerinin Kümelere Göre
Coğrafya Net Ortalamaları

d29	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Anadolu	7,6	8,0				8,0
Fen		12,1				12,1
Meslek	2,3	7,6	3,5	2,8	3,1	3,0
Normal	3,6	7,6	3,8	3,3	3,7	3,8
Özel	5,0	5,9			2,7	4,9
Özel Fen	6,8	7,5				7,4
Yabancı Dille Eğitim Yapan Özel Lise	5,4	7,4	5,4		4,5	6,0
Y.D.A.	4,2	5,7	5,8			5,4
Y.D.A. Özel	4,3	6,7	2,9			4,3
Anadolu Meslek	3,9	4,6	4,1	5,2	4,4	4,1
Akşam Lisesi				3,5	4,4	4,4
Açık Öğretim				4,6		4,6
Özel Meslek				3,0	3,3	3,2
Toplam	4,3	6,7	3,9	2,9	3,6	4,4

Çizelge EK 12.8 D42 Değişkenindeki Lise Türlerinin Kümelere Göre
Felsefe Net Ortalamaları

d31	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Anadolu	4,7	5,3				5,3
Fen		7,5				7,5
Meslek	0,8	4,1	2,4	1,7	2,3	1,9
Normal	1,8	4,7	2,7	1,8	2,4	2,7
Özel	4,3	4,5			2,2	4,1
Özel Fen	2,9	4,2				4,1
Yabancı Dille Eğitim Yapan Özel Lise	4,3	5,3	4,5		4,2	4,7
Y.D.A.	3,5	4,6	5,2			4,4
Y.D.A. Özel	3,8	4,5	2,5			3,7
Anadolu Meslek	2,5	3,0	3,1	4,0	4,3	3,2
Akşam Lisesi				2,0	2,6	2,5
Açık Öğretim				3,4		3,4
Özel Meslek				2,0	2,0	2,0
Toplam	3,2	4,8	2,8	1,7	2,8	3,2

Çizelge EK 12.9 D42 Değişkenindeki Lise Türlerinin Kümelere Göre İngilizce Net Ortalamaları

d33	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Anadolu	0,0	23,3				21,8
Fen		0,0				0,0
Meslek	0,0	20,5	0,0	0,0	0,0	0,5
Normal	0,0	82,8	2,0	0,0	0,0	2,4
Özel	20,3	44,9			0,0	23,3
Özel Fen	0,0	0,0				0,0
Yabancı Dille Eğitim Yapan Özel Lise	20,9	39,0	54,9		23,7	32,0
Y.D.A.	6,6	33,6	82,8			28,5
Y.D.A. Özel	5,3	0,0	20,8			5,9
Anadolu Meslek	0,0	18,4	9,2	0,0	4,3	4,4
Akşam Lisesi				0,0	0,0	0,0
Açık Öğretim				33,3		33,3
Özel Meslek				0,0	0,0	0,0
Toplam	9,2	30,0	6,1	0,4	1,8	10,4

Çizelge EK 12.10 D42 Değişkenindeki Lise Türlerinin Kümelere Göre Fransızca Net Ortalamaları

d35	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Anadolu	0,0	5,4				5,1
Fen		0,0				0,0
Meslek	0,0	0,0	0,0	0,0	0,0	0,0
Normal	0,0	0,0	0,0	0,0	0,0	0,0
Özel	0,0	0,0			0,0	0,0
Özel Fen	0,0	0,0				0,0
Yabancı Dille Eğitim Yapan Özel Lise	0,0	13,8	0,0		0,0	4,3
Y.D.A.	0,0	0,0	0,0			0,0
Y.D.A. Özel	0,0	0,0	0,0			0,0
Anadolu Meslek	0,0	0,0	0,0	0,0	0,0	0,0
Akşam Lisesi				0,0	0,0	0,0
Açık Öğretim				0,0		0,0
Özel Meslek				0,0	0,0	0,0
Toplam	0,0	3,6	0,0	0,0	0,0	0,8

Çizelge EK 12.11 D42 Değişkenindeki Lise Türlerinin Kümelere Göre
Almanca Net Ortalamaları

d37	Küme 1	Küme 2	Küme 3	Küme 4	Küme 5	Toplam
Anadolu	0,0	9,4				8,8
Fen		0,0				0,0
Meslek	0,0	0,0	0,0	0,0	0,0	0,0
Normal	0,0	0,0	0,0	0,0	0,0	0,0
Özel	0,0	0,0			0,0	0,0
Özel Fen	0,0	0,0				0,0
Yabancı Dille Eğitim Yapan Özel Lise	0,0	3,6	0,0		0,0	1,1
Y.D.A.	0,0	0,0	0,0			0,0
Y.D.A. Özel	0,0	0,0	0,0			0,0
Anadolu Meslek	0,0	0,0	0,0	0,0	0,0	0,0
Akşam Lisesi				0,0	0,0	0,0
Açık Öğretim				49,4		49,4
Özel Meslek				0,0	0,0	0,0
Toplam	0,0	2,7	0,0	0,6	0,0	0,6

ÖZGEÇMİŞ

Doğum tarihi 24.08.1979

Doğum yeri İstanbul

Lise 1993-1997

Lisans 1997-2001

Yabancı Dil Ağırlıklı Ümraniye Lisesi

Yıldız Üniversitesi Fen Edebiyat Fak.

İstatistik Bölümü

Yüksek Lisans 2001-

Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü

İstatistik Anabilim Dalı, İstatistik Programı

Çalıştığı kurum(lar)

1999-2001

HTP Araştırma ve Danışmanlık Hizmetleri A.Ş.

Üretim Uzmanı

2001-

Devam ediyor YTÜ Fen Edebiyat Fakültesi

İstatistik Bölümü Araştırma Görevlisi

