



Betimsel İstatistiğin Yeni Boyutları: Çok Boyutlu Çözümleme ve İstatistik Mühendisliği

*New Dimensions of Description Statistics: Multi-Dimensional Analysis
and Statistical Engineering*

Adnan MAZMANOĞLU¹

Öz: Bu çalışma, Türkiye'nin yetiştirdiği istatistik bilim alanında değerli çalışmaları olan Prof. Dr. Merih İpek'in İstanbul Üniversitesi İktisat Fakültesi C. T. Tütengil'e' Armağan olarak yayınlanmış bir makalesinden esinlenerek yazılmıştır. Bugünün koşullarına bir adaptasyon bir uyarlama makalesidir. Bilhassa çok boyutlu çözümleme ve çok boyutlu çözümleme tekniklerinin yıllar önce betimsel istatistik yöntemleriyle nasıl bağdaştığını gösteren M.İPEK, istatistikçilere, adı geçen çalışması kariyer yapmaları durumunda nasıl davranmaları gerektiğine ışık tutacağına inancım tamdır. Çünkü istatistikteki yeni eğilimler ve yöntemler incelendiği zaman gerek betimsel istatistiğin çok boyutlu çözümleme teknikleri konusunda gerekse veri bilimi, veri çözümleme ve istatistik mühendisliği disiplinlerine bir katkı sağlamak M. İPEK'in yüksek lisans ve doktora tez danışmanım olmasından dolayı benim üstlenmemin doğru olduğuna inandım. Henüz veri analistliği, veri bilimcisi, büyük veri (big data) kullanma hatta istatistik mühendisliği kavramları ortaya çıkmamıştı. Çalışmamda bunlara da yer vererek bir sentez oluşturdum. Bununla da bir tür istatistikteki günceli yakalamış oldum.

Anahtar Kelimeler: Betimsel İstatistik ~ Tümevarımcı İstatistik, Çok Boyutlu Çözümleme, Çok Boyutlu Veri Çözümleme, Büyük Veri (Big Data), Asal Bileşenler Analizi, Veri Analisti, Veri Bilimcisi, istatistik Mühendisliği

¹ Prof. Dr. Toros Üniversitesi Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü,
33140, Mersin, Türkiye. adnan.mazmanoglu@toros.edu.tr

Abstract: *This study is valuable work in the field of statistics science, Turkey has trained Prof. Dr. Merih İPEK, an article published as present to Istanbul University Faculty of Economics, C. T. Tütengil. An adaptation to today's conditions is an adaptation article. M.İPEK, which shows how multidimensional analysis and multidimensional analysis techniques are compatible with descriptive statistical methods years ago, is completely confident that it will shed light on how the statisticians should behave if they do the career in question. Because when new trends and methods in statistics were examined, I believed that it was correct to make a contribution both to the multidimensional analysis techniques of descriptive statistics and to the discipline of data science, data analysis and statistical engineering. The concepts of data analytics, data scientist, using big data or even statistical engineering had not yet emerged. I created a synthesis by including these in my work. With this, I caught the diary in some kind of statistics.*

Keywords: *Descriptive Statistics ~ Interference Statistics, Multidimensional Analysis, Multidimensional Data Analysis, Big Data, Principal Components Analysis, Data Analyst, Data Scientist, Statistical Engineering*

1. GİRİŞ

İstatistik belirsizliğin bilimidir. İstatistik ne olduğu sorusuyla değil de, ne olabileceği, neyin olanaklı neyin olası olduğu sorusuyla ilgilenir. Dış etkiler, koşullar, karışıklıklar, olayların sonsuz çeşitliliği sürekli evrensel bir etkileşim içinde olmaları, rastlantı denen nesnel gerçeği ortaya çıkarır. Rastlantı ise, bir süreçte kendine özgü çizgileriyle zorunluyu tamamlar. Öyleyse herhangi bir olay hem rastlantıyı hem de zorunluyu içerir. Önemli olan, görünümün arkasında özün aranması gibi, rastlantının ardından zorunluyu ortaya çıkarmak, başka bir deyişle, temel yasaya, ana nedene ya da nedenlere ulaşabilmek için, genel eğilimi, genel gelişme doğrultusunu bulmaktır.

1.1. Tümevarımcı İstatistikte Çok Boyutlu Çözümleme

Betimsel istatistik veya buna paralel olarak özünde tümevarımcılık adıyla anılan: Örneğe (ya da örneklerle) dayanarak ana kütle özelliklerine ilişkin tahminlerde, yorumlarda, genellemelerde bulunmak. Bunları yaparken de rastlantısal dağılımlardan yararlanmak². Zamanla istatistiğin sıkıcı bir bölümü olarak görüldü ve ders kitapları dahil istatistiğe ayrılan yerler

² Merih İPEK, “C. O. Tütengil’e Armağan”, İ.Ü. İktisat Fakültesi Yayınları, İstanbul, 1982

daraldı. Bilindiği gibi bilgisayarların işlem yapabilme hızlarının gelişimiyle *betimsel istatistik* yeni boyutlar kazandı. Böylece istatistik sorunlar yeni bir yaklaşımla ele alınarak satırlarca simetrik matrisler köşegenleştirilebildi; sınıflamalar kolayca yapılabilir duruma geldi. Hatta kuramsal temeli çok eski tarihlere uzanan, ancak uygulama olanaksızlığından gelişemeyen «*betimsel istatistik yöntemleri*», değişik türdeki teknikleriyle hızla gelişerek tümevarımcı yöntemlerin karşısına çıkmış oldu.

Bilgisayar yazılım olanaklarının artması ve gelişmiş bilgisayarlara paralel olarak gelişmesi gençlerin daha çok ilgisini çekmeye başladı. Bu genç kuşak «*olasılık temeline dayanan*» tümevarımcı model kurucu eski yöntemleri; türlü varsayımlar içinde kısıtlı kalmakla suçluyorlardı. Hatta denilebilir ki bu varsayımların sağlam bir temele oturduğundan kuşku duyanlar, yeni yöntemlerde kendilerini doğrulayacak kanıtlar da buldular bu kanıtlar elden ele dolaşmaya başlıyordu. Amaç *çok boyutlu verilerin çözülmesi* konusunda özlü, genel bilgi vermek, ayrıntılara, matematiksel teorik açıklamalara girmeden, yeni ya da yenilenmiş *betimsel istatistiklerinin* en çok kullanılanlarını, aralarındaki ayırım ve ilişkileri, uygulama alanlarını açık bir şekilde tanımlamak istiyoruz³.

1.2. Çok Boyutlu Çözümleme (Veri Çözümlemesi) Nedir? Ne Değildir?

1.2.1. İstatistik Veriler

İstatistik veri bilimidir. Genellikle iki tür kümeden oluşur. Bireyler, bireylere ilişkin özellikler diyebiliriz. Gözlenen, birtakım özellikleri ölçülen insanlar, aileler, insan dışı canlılar, işletmeler, ülkeler, olaylar, bireylerdir. Bunlara ilişkin özellikler ise nicel ya da niteldir. Yani her gözlenebilen veya deneysel birim bir nesnedir. Bu nitel özelliklerin türlü görünüşleri ya da ayrışmaları sıralanabilir. Nitel özellikler doğal sayısal ölçekte ölçülmeyen birimlerdir; bunlar yalnızca bir grup kategoriden birine sınıflandırılabilir. Bazı nitel özelliklerin türlü görünüşleri ya da ayrışmaları sıralanabilir, başarı derecesi, unvan, tercih sırası gibi,

³ Merih İPEK, “C. O. Tütengil’e Armağan”, İ.Ü. İktisat Fakültesi Yayınları, s.466, İstanbul, 1982

çoğununki sıralanamaz – cinsiyet, göz rengi gibi – Birkaç yıl öncesine kadar «klasik» diye adlandırılan tümevarımcı model kurucu istatistik, genellikle az sayıda birim ve bir ya da birkaç özellik üstünde çalışır. Oysa bireyler ya da karmaşık olaylar, çoğunlukla sayısız özelliklerle karşımıza çıkar⁴.

Öte yandan günümüzde, birçok alanda yapılabilen anketlerden çok sayıda birimi, sayısız özellikleriyle ve de bunların ayrışmalarıyla gözlemek olanaklıdır. Böylece derlenip toplanan, sınıflandırılan, özetlenen, düzenlenen, ayrıştırılan veriler(bilgiler), kabaca üç tip tablo ile sunulabilir⁵.

- a. Birey x Özellikler Tabloları
- b. Kontenjans Tabloları
- c. İlişki(yakınlık) Tabloları

Bu tablolara ilişkin açıklamalar:

a. *Birey x Özellikler tabloları*, genellikle n satırı bireylerden, p sütunu özelliklerden oluşan bir $n \times p$ boyutlu X değişken matrisi ile gösterilebilir. Çünkü istatistikte çeşitli grupların içerdiği birimler (yani bireyler) üç veya daha çok şıklı (çoğu zaman nitel \sim kalitatif) bir vafsa (özelige) bölünmelere sahip olabilir. Tablo, özellikler nicelse, sayısal değerleri içerir; nitelse 0 ve 1 gösterge değişkenlerden oluşabilir⁶. Tabloların izlenebilmesini sağlamak için başvuru gösterge değişkenler (ya da kodlar), bireyde (veya birimde) bir özelliğin belli bir ayrışımının olup olmadığını (1, 0) gösterir. Nicel – Nitel özelliklerin birlikte yer aldığı bir tabloda, değişkenin aldığı değerler sınıflanarak nicel özellik için de gösterge değişken kullanılabilir⁷.

b. Kontenjans aslında gerçekleşeceği belirsiz ve şüpheli olan konular (maddeler) demektir. Bir diğer anlamı ise kontenjans ya da bağımlılık tabloları olup, iki nitel özelliğin ayrışımalarının kesiştiği yerde ortak sıklıklar (frekanslar) ya da sıklık oranları yer alır. Çapraz adı da verilen bu

⁴ Merih İPEK, “C. O. Tütengil’e Armağan”, İ.Ü. İktisat Fakültesi Yayınları, İstanbul, 1982

⁵ James T. McClave, T. Sincich, Statistics, International Edition, USA, 1987

⁶ Adnan Mazmanoğlu, Herkes için Temel İstatistik Yöntemleri ve Uygulamaları, 2. Baskı, Ocak 2018, Ankara

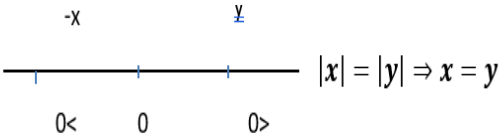
⁷ Merih İPEK, “C. O. Tütengil’e Armağan”, İ.Ü. İktisat Fakültesi Yayınları, age s.497, İstanbul, 1982

tablolarda gruplandırılan bireyler ayırdedilemezler. Bu tabloların da işlenebilmesi için, her özelliğe, bir gösterge değişken tablosu bağlanabilir (ya da bir tablo ilişkilendirilebilir). İki özelliğin ayrışmalarına ilişkin tablolar X_1 ve X_2 ile gösterilirse ($X_1, X_2 \in X$), $X_1 X_2$ skaler matrisiyel çarpımının sonucu bir kontenjans tablosudur.

c. İlişki (yakınlık) tabloları ise benzerlik ya da benzersizliklere ilişkin yakınlık (ilişki) tablosu çoğunlukla simetrik olur.

Kontenjans Tablosuna Örnek						
X_2e ait iki özelliğe ilişkin ayrışım Tablo: 4 erli aralıklardan oluşan yaş grupları içinde emekli olanların sayısı X_1: Eğitim sektörü; X_2: Sağlık sektörü						
$(X_1) \setminus (X_2)$	35-39	39-43	43-47	47-51	51-55	Toplam
35-39	2	1	1			4
39-43	3	1	2	1		7
43-47	1	1	2	1		5
47-51		1	6	5	2	14
51-55				6	4	10
Toplam	6	4	11	13	6	40

Uzaklıkların ya da bunların terslerinin karşılığı olarak pozitif sayılar içerir. Genellikle matematik anlamda bir uzaklığın aşağıdaki üç önermeyi sağlaması gerekir:

i. $d(x, y) = 0 \iff x = y$ ise 

ii. $d(x, y) = d(y, x)$ (simetri)

iii. her (x, y, Z) üçlüsü için $d(x, y) \leq d(x, z) + d(y, z)$
(üçgensel eşitsizlik)⁸

Bu önermelerden (iii). Önerme \sim koşul gerçekleşmiyorsa, uzaklığın bir “benzersizlik” olduğu söylenir. Diğer yandan, uzaklıkla eşanlamlı olarak “metrik” sözcüğü kullanılırken (iv). Bir koşulu gerçekleştiren uzaklığa da $-d(x, y) \leq \max \{ d(x, z) d(y, z) \}$ –“ültrametrik” denir⁹.

⁸ d harfi İngilizce uzaklık sözcüğü “distance” in birinci harfidir.

⁹ Merih İPEK, “C. O. Tütengil’e Armağan”, İ.Ü. İktisat Fakültesi Yayınları, age s.468’de yapılan atf; F. Cailliez, J.P. pages : 121.

Derleme sonunda elde edilen çok boyutlu, çok sayıda verili tabloları, daha kavranabilir bir duruma sokmak, ilk bakışta görülemeyeni görebilmek için ciddi bir tasarımdır bir özümsemedir. İşte bu evrede betimsel istatistik yeni boyutlarıyla çok boyutlu bir veri çözümlemesi olarak karşımıza çıkar.

1.2.2. Çok boyutlu çözümleme (Veri çözümlemesi) ve Büyük Veri (Big Data)

“Verilerin Çözümlemesine” değinmeden önce “Çok Boyutlu Çözümlemenin” ulaştığı boyutlar hakkında yani Büyük Veri (Big Data) hakkında bir açıklamayı uygun görüyoruz.

Unutmayalım ki veri devrimi daha yeni başlıyor. Artan bir hızla veri, eğitimden sağlığa, pazarlamadan iletişime hemen her alanda hayatımızda merkezi bir konuma oturuyor. Veri arttıkça *veri, bilim, yazılım* gibi sözcüklerin anlamları değişiyor. İnsanların enformasyona (bilgiye) ulaşmalarının ve karar vermelerinin biçimleri değişiyor. Yukarıda da söz ettiğimiz gibi, şimdi inanılmaz bir "*veri seli*" ile karşı karşıyayız. Bugünlerde enformasyon teknolojisi ile ilgili birçok kişi ve şirket, sürekli **Büyük Veri**'yi konuşuyor. Terabyte'lar ve petabyte'larla ölçülebilen, hep kullandığımız bilgisayar ve yöntemlerle analiz edemeyeceğimiz veriler, "**Büyük Veri**" olarak adlandırılıyor. Hatta *Büyük Verinin* "yeni petrol" olduğunu söyleyenlerin çokluğunu unutmayalım. Onun içindir ki istatistik, herkes için vazgeçilmez bilimler üstü bir bilim dalıdır¹⁰. Böyle devasa verileri çözümlemek için veri artık karakter değiştiriyor. İçerisinde bulunduğumuz Samanyolu galaksisinde 2 milyar yıldız olduğu söyleniyor. Eğer yıldızların her biri bir bayt olsaydı, 1 PB'lık veriye ulaşabilmek için 5.000 Samanyolu galaksisine ihtiyacımız olacaktı. 1 petabayt aynı zamanda 10.000 saatlik TV programına da denk geliyor. Bu programları hiç uyumadan seyretmeniz, ancak iki senede mümkün olurdu

Büyük Verinin tanımına gelince '**Büyük veri**' artan hacim, hız ve çeşitlilik gösteren veri kümeleri olarak ifade edilmektedir. Ayrıca veriyi

-
- ¹⁰ 1Gigabyte: 1024 MB;
 - 1Terabyte: 1024 GB;
 - 1 Petabyte: 1024 TB;

Tabii ki 1MB'ın 1024 KB olduğunu unutmayalım.

iyice anlamak ve içselleştirmek için maliyet etkin ve yenilikçi işleme yöntemleri talep eden büyük hacimli, hızlı ve çeşitli veriler olarak da tanımlanabilir.

Genellikle Fransız ekolünde kısaca "*verilerin çözümlenmesi*" diye adlandırılan, "*çok boyutlu çözümleme*", "*çok değişkenli çözümleme*" deyimleri de kullanılan *betimsel istatistik yöntemleri*, ham verilerin ya da bunlardan oluşturulan tabloların, ayıklanıp indirgenilmesini, sınıflanmasını sağlar. Nicel ve nitel, çok sayıda değişkenin topluca incelenmesine olanak verir. Bunu da çoğunlukla, aradaki ilintileri, benzerlikleri ya da ayrımları ortaya çıkararak, anlamlı sonuçları, önemli noktaları belirginleştirerek gerçekleştirir. Bunlar olurken, bireylerle değişkenler geometrik uzaylara sokulur. Bir düzeyde görüntülemek ya da türdeş sınıflarda toplamak üzere dönüşümlere uğratılır. Öte yandan, tüm bu süreç boyunca, varsayımlar olabildiğince aza indirgenir; model kurmamaya çalışılır. Yitirilen bilgi de olabildiğince azdır. Bu duruma göre, *gerek işe karıştığı evre açısından, çok boyutlu çözümleme betimsel istatistiğe yaklaşır; onun yeni bir aşaması olarak ortaya çıkar. Gerekse, temel amaç açısından betimseldir: Bilimlerin tümü için geçerli yasalar çıkarmaya çalışmaz. Genellemeye, model kurmayla uğraşmayıp, bireylerin "yapısıyla", "olayların özülle" ilgilenir*¹¹.

1.2.3. Çok boyutlu çözümlemenin temeli ve uygulama alanı

Veri çözümlemesi ya da çok boyutlu çözümleme, 20. yüzyıl başında Spearman ile Pearson'un çalışmalarına dayanır. 1930 larda, bu çalışmalardan yararlanan Hotelling, yeni ya da yenilenen çözümleme yöntemlerinin kaynaklandığı "*başlıca bileşenler çözümlemesi*" ve "*kümelere arası çözümleme*" yöntemlerinin temelini kurar. 1960'a dek yavaş gelişen, ancak uygulanamayan yöntemler, iki olgunun etkisi ve bunların etkileşimiyle son 15-20 yıl içinde hızla gelişir:

a) Bilgisayarların veri işleme hızının gelişmesi

¹¹ M. İPEK'in kaynak referansları: J.M. Bouroche, G. Saporta: a.g.y., s:6; F. Cailliez, J.P. Pagès: a.g.y., Morlat'nın önsözü; L. Lebart, A. Morineau, N. Tabard: Techniques de la description statistique, s:1; L. Lebart, J.P. Fénelon: Statistique et informatiques appliquées, s:7.

b) Değişik türdeki alanlarda anketlerin düzenlenmesi.

İstatistiğin önemli bir bölümünü oluşturma yolundaki bu yöntemlerin uygulama alanı çok genişdir: Sosyoloji, psikoloji, meteoroloji, iktisat, tıp, biyoloji, uluslararası ilişkiler, şehirleşme, çevre kirliliği, jeoloji, dilbilimi, coğrafya, tarih gibi hemen hemen tüm bilim dallarında kullanılmışlardır ya da kullanılabilirler. Son iki alanda yapılan sayısız derlemeler, söz konusu yöntemlerle kolaylıkla işlenebileceği gibi, sosyoloji-iktisat alanında yapılmış ya da yapılacak sosyoekonomik anketlerden çok yararlı sonuçlar çıkarılmıştır ya da çıkarılabilir. Günümüzde, kamuoyu davranışları konusunda duyulan bilgi edinme gereksinimi büyük önem kazanmıştır. Özellikle bu gereksinimi karşılayabilen "*çok boyutlu çözümleme teknikleri*" nin, özet olarak, karmaşık olayların söz konusu olduğu tüm alanlara uygulanabileceği söylenebilir. Kullanılmakta olan *çok değişkenli istatistik tekniklerinden* en popüler olanlardan çoklu doğrusal regresyon modeli, lojistik regresyon analizi, probit regresyon modeli (probit regression models), faktör analizi (factor analysis), ayırma (discriminant) analizi, kümeleme analizi (cluster analysis) ile çok boyutlu ölçekleme (multidimensional scaling, MDS) sayılabilir. Bunları biraz daha açıklamaya çalışalım.

2. BELLİ BAŞLI YÖNTEMLERİN SINIFLANMASI VE AÇIKLANMASI

2.1. Belli başlı yöntemlerin sınıflanması

Veri çözümlemesi ya da çok boyutlu çözümleme yöntemlerinin bir bölümü yenilenir, yenileri de sürekli ortaya çıkarılırken, bunları sınıflamak oldukça güçtür. Bu nedenle bazı istatistikçilerin, "hazır reçeteler", "yemek tarifleri", "araç koleksiyonu" gibi deyimler kullandıkları oluyor¹². Kimi istatistikçiler, çok boyutlu çözümlemenin, Birey x özellik tablosuna ilişkin X matrisinin çokluk (frekans) sütunları üzerinde yapılan işlemlerden kaynaklandığını, ancak bu işlemlerin matrisin niteliğine değil, değişkenlerle aralarındaki ilintilerin (ilişkinin) yarattığı sorunların niteliğine bağlı olduğunu belirtiyorlar. Bu arada türlü yöntemlere örnek olarak "*faktör*", "*regresyon*", "*korelasyon*", "*varyans*", "*path*" analizlerini,

¹² M. Vollé: a.g.y., s:5.

"kümeler arası çözümleme" ve "ayırıcı" çözümlemeyi, aralarındaki ilişkilerle gösteriyorlar¹³. Kimi istatistikçiler de, veri çözümlemesini "faktör" analizi olarak tanımlayıp, bu analiz içinde gösterdikleri "Başlıca bileşenler" ve "Bağlantı" çözümlemelerinin "Modelsiz ekonometri" ye yol açtığını söylüyorlar¹⁴. Kuşkusuz, verileri ayıklayan bu karmaşık görünümlü yöntemlerin arasında da bir ayıklama olacaktır. Şimdilik, daha çok Caillez ve Pagès'nin yapıtındaki sınıflamadan esinlenerek, şöyle bir sınıflama yapmayı uygun bulduk:

- A) Başlıca amacı bireyleri sınıflama olan teknikler: Otomatik sınıflama
B) Başlıca amacı, alt uzayda görüntüleme yoluyla bireyleri betimlemek olan teknikler: Başlıca bileşenler çözümlemesi ve uzaklıklar tablosunun faktör analizi.
C) Başlıca amacı, iki ya da daha çok özellik arasındaki ilintiyi betimleme olan teknikler.

Son gruptaki teknikler çok sayıdadır. Değişkenlerin sayı ve niteliklerine göre aralarındaki ayırım ya da bağlantıları daha iyi belirlemek üzere, aşağıdaki tabloyu veriyoruz (Caillez ve Pagès, s:218).

İlintileri betimleyen yöntemler	Özelliklerin sayı ve niteliği	
	1. grup	2. grup
	Açıklanacak değişkenler	Açıklayıcı değişkenler
Çoklu regresyon analizi	1 nicel	p nicel
Varyans analizi	1 nicel	P nitel
Kovaryans analizi	1 nicel	p_1 nicel + p_2 nitel
Kümeler arası çözümleme	q nicel	p nicel
Ayırıcı çözümleme	1 nitel	p nicel
Bağlantı çözümlemesi	1 nitel	1 nitel
Çok boyutlu varyans analizi	q nicel	p nitel
Çok boyutlu kovaryans analizi	q nicel	p_1 nicel + p_2 nitel

¹³ Bkz: M. İPEK'in kaynak referansları: Van de Geer: Introduction to Multivariate Analysis for the Social Sciences, s:85-92.

¹⁴ Ch. Labrousse: Introduction à l'économétrie, s:94.

Bu tabloda tüm yöntemlere yer verilmediği gibi, ikiden çok gruba ilişkin genelleştirilmiş yöntemler de gösterilmemiştir. Bunlar tek tek ele alındığında, aralarındaki ilintiler daha açıklığa kavuşacaktır. Ancak burada, regresyon, varyans, kovaryans analizleri gibi, tümevarımcı istatistik çerçevesinde işlenmiş yöntemleri göz önünde tutmayacağız.

Tabloyla ilgili son bir noktaya değinelim: Pagès ve Caillez'nin yapıtında "*doğrusal çözümleme teknikleri*" adı altında toplanan ve "*başlıca bileşenler*" çözümlemesini de içeren bu tekniklerden "*Bağlantı çözümlemesi*", bazı yazarlarca, doğrusal olmayan bir teknik diye tanımlanıyor¹⁵.

2.2. Belli başlı çok boyutlu çözümleme yöntemlerinin açıklanması

2.2.1. Otomatik sınıflama (Taximomie)

"*Eski boyutlarıyla*" betimsel istatistikte, veriler derlendikten sonraki aşama, bunların tablo ve grafiklerle derli toplu sunulmasıdır. Bunun için de çokluk verilerin sınıflanması gerekir. Gözlenen birimlerin ve de birimlere ilişkin özelliklerin sayısı çoksa, sınıflama oldukça güçleşir.

Güçlü bilgisayarlarla bu sorunun üstesinden gelinerek geniş veri yığınlarına erişilmiş, sayısız sınıflama yolları bulunmuştur¹⁶.

"*Taximomie*" adı verilen sınıflamanın amacı, bireyleri, az sayıda türdeş sınıflarda toplamak, böylece birey sayısını azaltarak onları betimlemektir. Çözümlenen tablo, ya n birey arasındaki uzaklık tablosu, ya da p eksen üzerinde bireylerin koordinatları tablosudur¹⁷.

Sayısız tekniklerden oluşan otomatik sınıflama, bir kümenin, bir ya da birçok bölmelemesini tanımlamaktan başka bir şey değildir. Ancak burada önemli olan "en iyi" bölmelemeyi yapmaktır. Başka deyişle,

i) Ele alınan p özellik karşısında sınıflar olabildiğince birbirinden ayrık,

¹⁵ J.M. Bouroche, G. Saporta, a.g.y., s:108.

¹⁶ Biyoloji, iktisat, muhasebe gibi alanlarda değişen sınıflama kavramı ve sınıflamanın tarihçesi, felsefesi için bkz: M. Vollé, a.g.y., s:190-205.

¹⁷ Merih İPEK'in adı geçen çalışmasında konuyla ilgili kaynaklar: J.M. Bouroche, G. Saporta, a.g.y., s:49.

ii) Her sınıf olabildiğince türdeş olmalıdır.

Bununla birlikte, böyle bir bölmeleme, sınıflar belli sayıda olursa gerçekleşebilir. Oysa sınıfları belli sayıda tutmanın türlü sakıncaları oluyor. Henüz hiçbir yöntem k gibi belli sayıda sınıfa optimal bölmelemeye hızla ulaşmayı sağlayamıyor. Bu nedenle de, hızla çalışan yöntemlerle yetiniliyor.

Otomatik sınıflama yöntemleri başlıca iki grupta toplanıyor:

a) Hiyerarşik sınıflama

b) Doğrudan sınıflama

a) *Hiyerarşik* sınıflamada, gitgide genişleyen sınıflara doğru bir bölmeleme dizisi oluşur. Bölmeler birbirine karşıt değil, bir sınıflama ağacı biçiminde iç içedir. Botanik, zooloji gibi doğa bilimlerinde yapılmış sınıflamalar bu türdendir. Günlük yaşamda doküman sınıflaması, yönetsel bölmeler, vb. hiyerarşik (iç içe) sınıflamaya birer örnek olarak gösterilebilir.

Günümüzde hiyerarşik sınıflama, geniş çapta pazarlama, bir ülkenin iktisadi coğrafyası, çocuk edebiyatında önerilen modeller, hastane yapımında aynı kata konacak servis grupları, devlet harcamalarına göre sınıflama vb. gibi yerlerde uygulanmıştır.

Hiyerarşik sınıflamanın pratikte iki sakıncası görülür: Birey sayısı fazla olunca ($n > 400$ gibi) uzaklıkların hesabı güçleşir. Bilgisayarlarda büyük bellek, uzun hesap zamanı gerekir. Bu nedenlerle daha hızlı, daha ucuz yöntemlere gerek duyulur. Bu da aşıldı.

b) Hiyerarşik olmayan sınıflamada bölmeleme doğrudan gerçekleşir. k sınıfa doğrudan bölmelemede k' yı ya yöntem belirler ya da bir maksimumla bir minimum arasında değişir. Yukarıda da değinildiği gibi, belli sınıf sayısı için "en iyi" bölmelemenin büyük güçlüğünden, yaklaşık çözümler veren hızlı algoritmalara başvurulur. Bunlar arasında "oynak odaklar çevresinde toplama", "dinamik bulutlar", "varyansa göre toplama" sayılabilir¹⁸.

¹⁸ Merih İPEK'in adı geçen çalışmasında konuyla ilgili kaynaklar: Değişik teknikler için bkz. J.M. Bouroche, G. Saporta, a.g.y., s:50-53, A. Henry-Labordère, a.g.y., s:60-77, M. Vollé, a.g.y., s:230-251. Özellikle de: Benzecri'nin (I. cilt: Taximonomie) ve Pagès-Caillez'nin adı geçen yapıtları.

Doğrudan sınıflama, mimarlık, pazarlama, sosyoloji, iktisat vb. gibi türlü alanlarda kullanılır. Örneğin, kilise, sendika, dernek gibi kurumlara karşı tutumlar; sosyoekonomik niteliklerine ve siyasal eğilimlerine göre kişilerin gruplandırılması; İtalyan demiryollarının toplu taşımacılık açısından kara ve demiryollarını modernleştirip yeniden düzenlemek üzerine yaptığı bir araştırma, doğrudan sınıflamanın uygulandığı alanlar arasında gösterilebilir¹⁹.

Son olarak, otomatik sınıflamaya ilişkin iki noktaya işaret edelim:

Daha sonra görülecek "*başlıca bileşenler*", "*bağlantı*" çözümlenmeleri gibi yöntemlerin sağladığı grafikler üzerinde de bölmelemeler yapılabilir. Ancak, "*gözle*" yapılabilen bu tür bölmelemelere güvenilmez. Grafikler tüm bilgileri vermediği gibi, gözlem sayısı çok olunca sınıflama da can sıkır. Bu nedenle, çokluk, yöntemleri tümleyen bir yaklaşımla sorunlar ele alınır. Ya görüntüleme yöntemlerinin ortaya çıkardığı varsayımları doğrulamak ya da bozmak için, ikinci adımda, otomatik sınıflama yöntemlerine başvurulur. Ya da otomatik sınıflamayla kabaca bir bölmeleme yapıldıktan sonra, "*başlıca bileşenler*", "*bağlantı*" çözümlenmeleri kullanılarak, sonuçlar daha yakından incelenir²⁰.

Öte yandan otomatik sınıflamanın özelliği, çok değişik ölçütlerin kullanılmasıdır. Matematiksel istatistiğe hemen hiç yer verilmez. Tümevarımcılık yerine, hesap olanaklarını hep en iyi biçimde kullanıp betimlemeyi geliştirmek, öteki yöntemlerden daha çok sınıflamada görülür.

Ancak, kesin kurallar olmadığından, sayısız ölçüt ve yollar teknik seçimini güçleştirir. Bu nedenle de bir yandan sınıflama programlarını düzeltme (veriler tablosunun boyutu, hızlı hesap açısından), bir yandan da elde edilen sonuçların kalitesini yükseltme yolunda arayışlar sürmektedir.

¹⁹ Uygulamalar için bkz. J.P. Benzecri'nin adı geçen yapıtı: P. Bertier, J.M. Bouroche: a.g.y., s:87-103; K. Kurtuluş: Pazarlama Araştırmaları, s:477.

²⁰ Bkz: J.M. Bouroche, G. Saporta, a.g.y., s:59.

2.3. Başlıca bileşenler çözümlemesi: *Bileşenler Sistemi ve SPSS*²¹

Önce tüm sosyal bilimler için çığır açan faktör analizinin tarihsel gelişimine bir göz atalım. Konumuz “çok boyutlu çözümleme yöntemleri” olduğuna göre, Faktör Analizi çok değişkenli istatistik tekniklerin biri olup en çok kullanılanı da odur. Buna ek olarak “istatistik biliminin yüksek düzeyde” çok önemli uğraşı alanlarından da en önemlisidir. Hatta ortaya çıkışı ve yoğun kullanımı psikoloji alanında gözlemlendiğinden, bu yöntem istemeyerek de psikolojik bir öğretim tekniği görülmüştür²². Ama en önemlisi Faktör ANALİZİ’nin temel niteliği bilgiyi kısaltma ve özetleme olmakla birlikte, değişkenlerin tümü analiz kapsamına girer. Değişkenler kümesi içinde yer alan inceleme konularının birbirleriyle ilişkisini ve bu ilişkinin gücünü saptamak üzere değişkenler, gerçek durumlarından daha küçük bir doğrusal birleşimler grubu ile tanımlar²³. Bu açıklamadan bu yöntemle insan yetenekleri, alışkanlıkları, becerileri ve davranışları gibi psikolojik olguların açıklanması için etkin bir matematik modelleme olanağı sağlamıştır denebilir. Kuramsal olarak konuya ilişkin çalışmalarıyla önderlik yapan Spearman, Burt, Kelley, Thurstone, Holzinger ve Thomson tarafından yapılmıştır²⁴. Bilhassa 1901 yılında Faktör Analizi konusunda istatistiksel açıdan önemli çalışmalar yapan Karl Pearson²⁵ ilk kez “Asal Eksen Yöntemi” adıyla yayınlanarak—Asal Bileşenler Analizi” nin bir tür temelini atmıştır. Aşağıda da daha geniş açıklamalarda bulunacağız ve birçok bilim adamının uğraş alanı durumuna gelen istatistiksel gelişmeler konu edilecektir.

²¹ H.NIE NORMAN, Hull C.Hadlai, Jenkins Jean G., Stein BRENNER KARIN, BENTDALE H., SPSS (Statistical Package for The Social Sciences), Mc GRAW-HILL BOOK COMPANY, USA, 1975

²² H. Durucasu, “Asal Bileşen Analizi ve Bir Uygulama Denemesi, “Y.Lisans Tezi, Anadolu Üni., 1991

²³ Adnan MAZMANOĞLU, “Faktör Analizi ve Bilgisayarlarda Modern Faktör Analizi Yöntemlerinin Kullanımı, Y. Lisans Tezi, İstanbul Üniversitesi, 1979

²⁴ H. Durucasu, “Asal Bileşen Analizi ve Bir Uygulama Denemesi, “Y.Lisans Tezi, s. 18-23, Anadolu Üni., 1991

²⁵ Karl Pearson (1895-1980)-The father of Statistics, Pearson after earning a law degree at Cambridge Uni. And a Ph.D. in political science at the Uni. of Heidelberg (Germany), Pearson became a professor of applied mathematics at University College in London. The term *standard deviation* and its associated symbol (σ); developing the distributions of the correlation Coefficient.

2.3.1. Asal Bileşenler Analizi²⁶

Kısaca bu teknik gözlenen bir değişken kümesini daha önceden ifade ettiğimiz gibi yeni bir değişkenler kümesine dönüştüren doğrusal bir teknik olduğunu söylemiştik. Söz konusu doğrusal dönüşüm sonrasında elde edilecek yeni değişkenler birbirlerine dik (orthogonal) olan asal bileşenler olacaktır. Bu tekniğin uygulanacağı değişkenlerin yapısına ilişkin herhangi özel varsayıma gerek yoktur. Bu teknikte asal bileşenlerin, özgün verideki varyansın nedenini tüm diğer olası doğrusal bileşimlerden daha iyi açıklayan bir doğrusal bileşim oluşturması amaçlanır. Bundan dolayı ilk bileşen (ya da faktör) veride yer alan doğrusal ilişkilerin en iyi ifade edildiği boyut olmaktadır. İkinci bileşen, özgün veriden ilk faktör tarafından yutulan (absorb) varyans sonrasında kalan varyans bölümünü mümkün en yüksek oranda üstlenecek yalnız kendine özgü bir nitelik taşıyan yani orijinal değişkenlerin doğrusal bileşimi olarak elde edilir. İzleyen bileşenler benzer mantığı kullanarak, verideki varyans tükeninceye kadar hesaplamaya devam edecektir. SPSS'e gelince (STATISTICAL PACKAGE THE SOCIAL SCIENCES) bu istatistik yazılım paketinde beş adet faktör analizi alt programları vardır. Asal bileşenleri içerdiğinden şöyle bir sınıflandırma yapmak mümkündür. PA1 ve PA2 tekrarsız ve tekrarlı asal faktörleme (Principal Factoring) programlarıdır. RAO, ALPHA ve IMAGE ise sırasıyla RAO, kanonik faktörlemesinin uygulanmasını sağlayan yazılım programlarıdır. Her beş programın sahip olduğu üç temel özellik vardır. Birincisi tüm faktörler ortogonal olarak hesaplanmaktadır. İkinci nitelik faktörlerin, yukarıda da ifade ettiğimiz gibi önem sırasına göre sıralanmasıdır. Yani çıktıda (output) belirtilen ilk faktör en önemli bileşendir. İkinci belirtilen ikinci, üçüncü belirtilen üçüncü önemli bileşendir. Üçüncü özellik ise ilk faktörün genel bir faktör olmasında diğer bir deyişle her değişkenle anlamlı şekilde yüklenmiş bir faktördür (Factor Loading). Bunu izleyen faktörler ise çift kutuplu olup hem pozitif hem de negatif faktör yüklemeye sahiptirler²⁷.

²⁶ H. HARMAN HARRY, "Modern Factor ANALYSIS", The University of CHICAGO PRESS, 1970

²⁷ Daha geniş bilgi için Bkz: Adnan MAZMANOĞLU, "Faktör Analizi ve Bilgisayarlarda Modern Faktör Analizi Yöntemlerinin Kullanımı, Y. Lisans Tezi, İstanbul Üniversitesi, 1979

Çok boyutlu çözümlemenin ana kaynaklarından biri, görüldü ki "Başlıca bileşenler çözümlemesi (Analyse en composantes principales)" dir. Ancak, 1933 lerde Hotelling tarafından öne sürülen ve Anderson (1958) ile Kshirsagar'ın (1972) geliştirdikleri biçimiyle yöntem, "klasik" istatistik çerçevesine sokulur. Çünkü bu çerçevede sorun, çok boyutlu normal dağılımın göstergesi elipsin başlıca eksenlerinin aranmasıdır ve bu da rastlantısal örneğe dayanarak kestirim yoluyla yapılır²⁸.

Öte yandan, psikologlarca kullanılan ve özgül varyansların sıfır ya da eşit kabul edildiği özel durumları işleyen Horst(1965), Harman(1966) gibi istatistikçiler de, "klasik faktör analizci" olarak kabul edilirler²⁹. Burada amaç, bireylerin betimlenmesi değil, özellikler arasındaki yakınlıkların betimlenmesidir. Bu tür çözümlemenin artık pek kullanılmadığı, psikologların aşırı yorumları nedeniyle de gözden düştüğü öne sürülür³⁰.

Çok boyutlu çözümlemeciler açısından "başlıca bileşenler çözümlemesi", özelliklere ayrıcalıklı bir rol tanımadan bunları işler. Özel bir modele, istatistik varsayımlara başvurulmaz. Veriler, bazı cebirsel ve geometrik ölçütlere göre temsil edilir. Daha açıkçası, temeli Pearson'a uzanan, ancak bilgisayarların gelişimine koşut olarak değişimler gösterip büyük önem kazanan bu yöntemin amacı, Birey x özellik tablosundaki bilginin özünü çekmek ve yorumu kolaylaştırmak üzere görüntülemektir. Bunu yaparken, "yakın" ve "uzak" bireyleri ortaya çıkararak onları betimler³¹.

Matematiksel olarak amaç şöyle dile getirilebilir: Söz konusu tabloya ilişkin x matrisi p noktalı bir bulut oluşturur. Amaç, bu nokta bulutuna, onu en iyi yansıtan bir alt uzay vektörü uydurmaktır. Böylece, noktaların, iyi belirlenen bir eksen, bir düzlem ya da hiper düzlemdeki izdüşümleri incelenebilir³².

²⁸Merih İPEK'in çalışmasında baş vurulan kaynaklar listesi: Lebart - Morineau - Tabard, a.g.y., s:7.

²⁹ Lebart - Morineau - Tabard, a.g.y., s:7; Lebart - Fénélon, a.g.y., s:195.

³⁰ Lebart - Fénélon, a.g.y., s:195, Caillez - Pagès, a.g.y., s:272. J. Torrens Ibern de, "Modèles et méthodes de l'analyse factorielle" adlı yapıtında, bilgisayarların kullanımıyla gelişen yöntemin, hesap teknikleriyle, klasik yöntemden uzaklaştığını belirtiyor (s: 90).

³¹ Caillez - Pagès, a.g.y., s:221.

³² Ch. Labrousse, a.g.y., s:96.

"Başlıca bileşenler", başlangıçtaki özelliklerin lineer bileşimleri olan, "yapma" ya da "sentetik" özelliklerdir. Demek ki, özellik sayısında indirgeme, bunlar arasından seçimle değil, yeni özellik oluşturarak sağlanır.

Yöntemin uygulama alanlarına örnek olarak, biyometride bazı hayvan ya da organlar üzerinde ölçümler, iktisatta ailelerin çeşitli kalemlere harcamaları, EU (OECD) ülkelerinin iktisadi durumlarının silueti, yıllara göre devlet harcamaları, öğrencilerin kaynaşma açısından birbirlerine verdikleri notlar vb. gösterilebilir³³.

"Başlıca bileşenler yöntemi" nin uzantısı, "uzaklıklar" ya da "yakınlıklar" tablosunun faktör analizidir. Bu sonuncu yönteme, değişkenlerin aldıkları değerler bilinmeyip, yalnız bireyler arasındaki uzaklıklar bilindiğinde başvurulur.

Özellikle, psikoloji ya da piyasa araştırmalarında kullanılır: Yeni bir ürünün piyasa sürümü, tüketici gereksinmelerinin irdelenmesi, reklam kampanyaları, bölgelere ilişkin "gelişme senaryoları", siyasal görünümlü bazı kişiler arasında yakınlık gibi³⁴.

2.3.2. İlişkileri (ilişkileri) Betimleyen Yöntemler

2.3.2.1. Kümeler arası korelasyon çözümlemesi (Kanonik Korelasyon analizi)

Kümeler arası korelasyon çözümlemesi, hazır programları olmakla birlikte, verilerin çözümlenmesinde en az kullanılan bir yöntemdir. Bu durum, sonuçların kullanımı ve yorumunda doğan büyük güçlüklerden ileri geliyor.

Ancak, yöntemin kuramsal önemi çok büyüktür. Öteki yöntemler arasında bir köprü görüyor. Bunların çoğu, kümeler arası çözümlemenin özgül (serbest) uygulamaları niteliğindedir.

³³ Merih İPEK'in çalışmasında baş vurulan kaynaklar listesi : Değişik uygulamalar için bkz: Bertier- Bouroche, a.g.y., s:115-123; Lebart - Fénélon, a.g.y., s:219; Lebart - Morineau - Tabard, a.g.y., s:27; Bouroche - Saporta, a.g.y., s:37; Labordère, a.g.y., s:6; 11; Caillez - Pagès, a.g.y., s:277-294.

³⁴ Uygulamalar için bkz: Bertier - Bouroche, a.g.y., s:205; Labordère, a.g.y., s:14.

Kümeler arası çözümlemenin (kanonik analiz) amacı, aynı birey kümesinde gözlenen iki nicel özellik grubu arasındaki ilintileri incelemek ve birleşimleştirmektir (sentetize etmektir)³⁵. Başka deyişle, iki nicel özellik grubunun denk sayılıp sayılmayacağını araştırmak söz konusudur.

Doğrusal çözümlemede denklikten şu anlaşılıyor: X_1, X_2, \dots, X_p gibi 2 özellik kümesi olsun. X_i ve X_j özelliklerinin tüm doğrusal bileşimleriyle dile gelen "öngörü güçleri" çakışıyor, iki kümenin "denk" olduğu söylenir³⁶.

Bu duruma göre, kümeler arası çözümlemede yapılan iş, iki gruba ilişkin doğrusal bileşimler arasında korelasyonu en fazla olan bileşimleri aramak oluyor.

Yukarıda verilen tablodan anlaşılacağı gibi, gruplardan biri tek değişkene indirgenirse, çoklu regresyon analizi söz konusu oluyor. Ayrıca, varyans analizinin, ayırıcı çözümlemenin ve bağlantı çözümlemesinin de, kümeler arası çözümlemenin özel durumları olduğu görülüyor. Buna karşılık, "başlıca bileşenler" çözümlemesinin özgül bir yanı var: İki özellik grubu arasındaki çözümlemenin değil de "genelleştirilmiş kümeler arası çözümlemenin" özel bir durumudur.

Bu sonuncu çözümlemede ikiden çok özellik grubu ele alınıyor. Her grupta tek özellik bulunursa, "*başlıca bileşenler çözümlemesi*" den başka bir şey olmuyor.

Demek ki, "*başlıca bileşenler çözümlemesi*, bir yandan öteki çok boyutlu yöntemlere dolaylı bağlanırken, bir yandan da aralarında ayırım yapılmayan sayısal özellikleri ele alarak özgürlüğünü koruyor³⁷.

Kümeler arası korelasyon çözümlemesinin uygulama alanı ve çözmeye çalıştığı sorunlarla ilgili şu örnekler verilebilir:

İki laboratuvarın nehir kirliliğine ilişkin ayrı analizleri: Bunlar özdeş kabul edilebilir mi? Kirlenmeyle savaşımda bu analizlerden yararlanılabilir mi? Psikoteknik testlerle okul başarıları: Notların tümü öğrenciyi betimleyebiliyor mu? Yeteneklerini ortaya çıkarabiliyor mu? İki kümeden

³⁵ Parçaların ve öğelerin bir araya getirilip bir bütün olarak birleştirilmesi.

³⁶ Cailliez - Pagès, a.g.y., s:351.

³⁷ Bouroche - Saporta, a.g.y., s:21, 82; Vollé, a.g.y., s:190.

yararlanarak yeteneklerine göre öğrenci yönlendirilebilir mi³⁸? 30 kobay farenin bir labirentten çıkış kapısı bulmak için harcadığı süre ve taşıdıkları stresin aynı olmadığı gibi.

Kümelere arası çözümleme genelleştirildiğinde, birçok değişken kümesini birlikte inceleme olanağını verdiği göre, önemi ve türlü yöntemler arasındaki yeri büyüktür. Özellikle insana değin bilim dallarında, davranış ya da tutuma ilişkin değişkenler arasındaki bağlantılar ortaya çıkarmaya büyük gereksinim duyuluyor. Bu nedenle yöntemin kullanma olanaklarının da artması bekleniyor.

2.3.2.2. Ayırıcı çözümleme (Diskriminant analiz)³⁹

Ayırıcı çözümleme, açıklanacak 1 nitel değişkenle, açıklayıcı p nicel değişken kümesi arasındaki ilişkileri ortaya çıkarmayı sağlıyor. Daha ayrıntılı olarak çözümlemenin amacı ve ele alınan sorunlar şöyle açıklanabilir:

Bireyler q sayıda sınıfa ayrılmıştır ya da ayrılabilir. Bir sınıfta yer alma, sınıf sayısınca ayrışımı olan nitel bir değişkene (Y) yol açar. Demek ki, birey kümesinin, ayrık q sınıflı bir bölmelemesi tanımlanmıştır ya da tanımlanabilir. Öte yandan, ayrı bireylerde p nicel özellik ölçülür. Bu duruma göre,

- i) Nicel değişkenlerin tümünde ya da alt kümede q sınıf ayırt edilebiliyor mu?
- ii) Yalnızca nicel değişkenlerin bilinmesi, bireyleri, kendi öz sınıflarına yeniden ayırmayı sağlıyor mu?
- iii) Yalnızca nicel değişkenler bilindiğinde, hangi sınıfa ait olduğu bilinmeyen bir bireyi en az riskle bir sınıfa vermek olası mı?
- iv) Nicel değişken sayısı, iyi ayrımı sağlamaya yetecek ölçüde azaltılabilir mi?

³⁸ Örnekler için bkz: Bertier - Boroche, a.g.y., s:149-161; K. Kurtuluş, a.g.y., s:431; Van de Geer, a.g.y., s: 164, 169.

³⁹ Editör Şeref KALAYCI: SPSS Uygulamalı Çok Değişkenli İstatistik Teknikleri, Asil Yayın Dağıtım Ltd. Şti., Ankara, 2010

İkiye ayrılabilen çözümleme yöntemi, bu soruların yanıtlanmasını sağlıyor:

α) Ayırıcı etken çözümlemesi (diskriminant faktör analizi), açıklanacak değişkenle, açıklayıcı değişken kümesi arasındaki ilintiyi betimliyor. Görüntüleme yoluyla yaptığı bu betimlemede yukarıda ifade ettiğimiz (i) sorunuyla uğraşılıyor.

β) Karar amaçlı çözümleme, açıklayıcı değişkenlerin aldıkları değerlerden, açıklanacak değişkenin ayrışmalarını görmeyi sağlıyor. Dolayısıyla i) ve iii) sorunlarını içeriyor.

Öte yandan, yine karar ve / ya da betimsel amaçlı "adım adım (Stepwise)" denen çözümleme, ii) ve iv) sorunlarıyla ilgileniyor⁴⁰.

Ayırıcı çözümleme, nitel değişkene sayısal kodlama arayışı açısından, "Kümeler arası çözümleme" nin özel bir yöntemi olarak kabul ediliyor. Ayrıca, eski özelliklerin bileşimi yeni bir özelliği belirleme açısından da, "Başlıca bileşenler çözümlemesi" nin özel durumu oluyor⁴¹.

Ancak burada maksimum varyanslı yeni bir özellik söz konusu değildir.

Ayırıcı çözümlemede,

- Aynı sınıfa ait bireyler için olabildiğince yakın,
- Ayırı sınıflara ait bireyler için olabildiğince ayırık,

bir özellik ya da özellikler aranıyor. Başka deyişle, değişkenlerin değerlerinin sınıflar içinde olabildiğince az, sınıflar arasında olabildiğince fazla dağıldığı lineer bileşimler (diskriminant fonksiyonlar) bulunmaya çalışılıyor⁴².

Bununla birlikte, otomatik sınıflamada olduğu gibi türdeş sınıflara ayırma da söz konusu değildir burada. Birey kütesine ilişkin hem nicel değişkenler, hem de nitel değişken biliniyor. Y tarafından gerçekleştirilen bir bölmeleme zaten var. Amaç, bu bölmelemeyi daha açıklığa

⁴⁰Merih İPEK'in çalışmasında baş vurulan kaynaklar listesi: Bertier - Bourroche, a.g.y., s:165.

⁴¹ Matematiksel ilişkiler için bkz: Cailliez - Pagès, a.g.y., s:412; Bourroche - Saporta, a.g.y., s:115.

⁴² Bourroche - Saporta, a.g.y., s:110; Nie, Hull, Jenkins..., SPSS, s:435.

kavuşturmak, gerektiğinde sınıfı bilinmeyen bireyleri de belirli sınıflara dağıtmaktır⁴³.

Ayırıcı çözümlemenin, öteki yöntemler arasında en geniş uygulama alanı bulan bir yöntem olduğu söylenebilir. Tıp, maden araştırmaları, antropoloji, tarih, hukuk, iktisat gibi.

Birkaç örnek verelim:

Tıpta: i) Şizofren, sinir hastası, normal diye ayrılan üç gruba anksiyeteye işaret edebilecek 30 dolayında belirtiyeye ilişkin-kâbus, kramp, çarpıntı vb. gibi-sorular sorularak, elde edilen verilerin normal kişiyi ötekilerden ayırt etmeyi sağlayıp sağlamadığı, belirtiler yardımıyla değişik anksiyeteye biçimlerinin belirlenip belirlenemeyeceği gibi sorunlarla uğraşılması.

ii) Sarılığa yakalanmış hastalar grubunun ilaçla tedavi görenler ve ameliyat edilenler diye ikiye ayrıldıktan sonra hastaneye girdiklerinde geçirdikleri radyolojik, biyolojik, klinik incelemelerin dökümünün yapılması. Bu bilgilerin iki grubu iyi ayırıp ayırmadığının, daha az veri ile ayırımın sağlanıp sağlanmayacağına, yeni bir hastanın ameliyat edilip edilmeyeceğine kararda bu bilgilerden nasıl yararlanılacağına aranması.

Antropolojide: i) Burnun yüksekliği, derinliği gibi verilerden yararlanarak Hindistan'daki çeşitli kastların belirlenmesi.

ii) Kadın ve erkek alanına ilişkin veriler göz önünde tutularak, bir kazı sırasında bulunan kafatasının ait olduğu kişinin cinsiyetinin saptanması.

İktisatta: i) Bir bankanın, çeşitli verilere göre, kredi alan müşterilerini iyi, orta, kötü diye ayırması ve bu verilere göre, kredi isteyen yeni müşterilere nasıl davranacağına karar vermesi.

ii) Çeşitli kentsel ulaşım biçimleri arasında seçim için bir karar mekanizmasının oluşturulması.

Tarihte: Sağ ve sol milletvekillerinin kullanmış oldukları sözcüklere göre iki sınıfa ayrılmaları. Sözcükler daha aza indirilerek bu ayırımın sağlanabilmesi⁴⁴.

⁴³ Bir bireyin hangi sınıfa ait olduğunun bilinmeme nedenleri için bkz: J.M. Romeder, Méthodes et programmes d'analyse discriminante, s: 11-14.

⁴⁴ Merih İPEK'in çalışmasında baş vurulan kaynaklar listesi Örnekler için bkz: Bertier - Bouroche, a.g.y., s:113; Bouroche - Saporta, a.g.y., s:117 - 122;

2.3.2.3. Bağlantı çözümlemesi (connected ~ connectivity analysis)

2.3.2.4.

Bugünkü boyutlarıyla bağlantı çözümlemesi, "kontenjans (ya da bağımlılık ~ ~ contingency~ dependency) tabloları" çözümlemesinin ikiden çok nitel değişkene uzatılmasıdır. Yöntemin kuramsal ilkeleri *Fisher'e* dek uzanıyor. Bununla birlikte, söz konusu başlangıç, tümevarımcı istatistik çerçevesinde yer alıyor. Oysa son yıllarda, yöntemin, cebirsel ve geometrik özelliklere betimleyici yanı ağırlık kazanmıştır: Veriler tablonun satırlarıyla sütunlarını oluşturan kütleler arasındaki bağlantının en iyi temsiliyle, bağımlılığın yapısı incelenerek belirgin noktaların ortaya çıkarılmasıyla uğraşılıyor.

Çözümleme, ana çizgileriyle, "*Başlıca bileşenler çözümlemesi*" ne benziyor: Veriler tablosunun çözülmesi, az bilgi kaybıyla alt uzaya iz düşüm vb. gibi. Ancak, ele alınan değişkenler niteldir, dolayısıyla veriler de pozitif tam sayılardır. Giderek bu verilerin sıklık ya da olasılıkların birlikte dağılımı olması gerekmiyor. Bundan başka, gösterge değişkenlere dayanarak yapılan bu çözümleme, başlıca bileşenlerinkinden daha iyi sonuç veriyor. Bu nedenle, sayısal veriler (ya da nicel değişkenler) gösterge değişkenlere dönüştürülerek söz konusu çözümlemenin yapıldığı da oluyor. Bu son durumda bilgi kaybedileceği düşünülürse de, doğrusal çözümlemeden kurtulma gibi bir yarar sağladığı ileri sürülüyor⁴⁵.

Grafik yorumunda da, "*Başlıca bileşenler çözümlemesi*" deki gibi, bir yandan noktalar ve başlıca düzeyler arasındaki yakınlık, öte yandan, bir eksenin belirlenmesinde her noktanın oynadığı rol göz önünde tutuluyor. Yalnız, değişkenler nitel olduğundan, korelasyona değil, katkılara bakılıyor.

Bağımlılık çözümlemesi ayrıca, ayırıcı çözümlemenin, dolayısıyla kümeler arası çözümlemenin özgül(serbest) uygulamalarından sayılıyor⁴⁶.

K. Kurtuluş, a.g.y., s:446; Romeder, a.g.y., s: 9 - 14.

⁴⁵ Bouroche - Saporta, a.g.y., s:108.

⁴⁶ Merih İPEK'in çalışmasında baş vurulan kaynaklar listesi : Yöntemler arasındaki matematiksel ilişkiler için bkz: Lebart - Fénélon, a.g.y., s:254; Lebart - Morineau - Tabard, a.g.y., s: 102-106.

Nitel değişkenleri ele aldığından çok gözde bir yöntem olan bu çözümlemede, Fransa'da Benzécri'nin 1960 sıralarında başlattığı çalışmalar oldukça yönlendirici olmuştur⁴⁷.

Özellikle anket sonuçlarının incelenmesinde çok güçlü bir araç olarak kabul edilen bu tekniğin, uygulama alanlarına ilişkin şu örnekler verilebilir:

- Lise diploma türlerinin bölgelere göre dağılımı
- Televizyonda seyredilen filmlerle seyircilerin yaş, öğrenim durumu, oturulan yer v.b. gibi türlü nitelikleri arasındaki bağlantı
- Fransız başkanlık seçimlerine ilişkin olarak, adaylarla mahallelerin karşılaştırılması
- Avrupa ülkelerinin, bütçelerine göre dağılımlarına bakılarak, politikalarında özdeşliğin aranması
- Ölüm nedenleriyle, meslek kategorileri ve bölgeler arasındaki bağlantıların incelenmesi
- Ailelerin tüketim kalemleriyle, sosyal yapılarının karşılaştırılması
- Yine ailelerin, tüketim harcamalarıyla, sosyoekonomik durumlarının, gösterge değişkenlere dönüştürme yoluyla çözümlenmesi
- Pazarlama konusunda çeşitli çalışmalar⁴⁸

Yine *nitel* değişkenlerin ele alındığı başka bir anket 1978-1979 ders yılında İstanbul Üniversitesine yeni kaydolan 4518 öğrenci arasından Mediko-Sosyal Merkezi görevlilerince hazırlanmıştır. Genellikle ankette tamamıyla öğrencilerin üniversite öncesi ve üniversite sırasında karşılaştığı veya karşılaşacağı sosyo-ekonomik problemlere ilişkin sorular sorulmuştur. Uygulamadaki amaç bu problemlerde öğrencinin sosyolojik durumuna ve yaşamına en etken faktörlerin bulunması ve gözlem hakkında genel bir sonuç çıkartma olan bu uygulama *A. Mazmanoğlu* tarafından, 1979 yılında değerlendirilmesi İstanbul Üniversitesi Haydar Furgaç Elektronik Hesap ve Araştırma(istatistiksel) merkezinde SPSS yazılım paketi kullanılarak yapılmıştır. Toplam 58 değişkenli (çoklu değişken

⁴⁷ Bkz: Benzécri, a.g.y, 2. cilt (Analyse des Correspondences).

⁴⁸ Başta Benzécri, a.g.y, 2. cilt (Analyse des Correspondences) olmak üzere, öteki a.g.y. larda değişik uygulama örnekleri bulunabilir.

çözümlemesi) bir Faktör Analizi uygulamasıdır. PA1 (PRINCIPAL FACTORING WITHOUT ITERATION)faktör alt programı uygulanarak birbirleriyle ilişkili 11 değişken arasında yapılan uygulama sonucunda elde edilen nitel değişkenler I. Faktör listesi altında “Doğum Yeri”, “Yaşam Bölgesi” ve “Ailesinin Yaşam Bölgesi” değişkenleri en önemli faktörler olarak karşımıza çıkmıştır⁴⁹.

3. VERİ ANALİZİ VE VERİ BİLİMİ

3.1. Veri Bilimi: 13 Aralık 2013 TÜİK tarafından İstatistik Araştırma Dergisinde Fikri AKDENİZ’in “İstatistikte Yeni Eğilimler ve Yöntemler” adlı makalesinde, yine başka bir kaynaktan aktardığını ifade ettiği bir yazı şöyle bir başlık atmış: YENİ BİR MESLEK: VERİ BİLİMCİSİ WORLD ECONOMIC FORUM (20 Nisan 2018) iş & yaşam dengesi puanlarına göre belirlenen dünyadaki en popüler 25 meslek grubunda en iyi işler sıralamasında 1. Sırada “Veri Bilimcisi”, 20. Sırada da “Veri Analisti” yer almaktadır. Neden? Yukarıda Fikri AKDENİZ’in naklettiği yazıda internetteki bilgi ve veri bombardımanına akıllı cihazlardan gelenler de eklendikçe iş daha da karmaşık bir hal alıyor. Tahminlere göre 2020 yılında, otomobilden ev aletlerine ve telefonlara kadar 50 milyar cihaz veri üretecek ve birbirleriyle iletişim olacaktır. İleriye yönelik öngörülerde bulunmak ve karar almak isteyen şirketler için bu “veri tsunamisi ~Big Data” yı doğru analiz edebilmek kritik önem taşıyacak. Bu durum, adına “veri bilimci” denilebilecek yeni bir mesleği doğuruyor. Bunları çalıştıran şirketler de önemli avantajlar sağlayacak. Örneğin tüm dünyada büyük şirketlere veri ambarı, kampanya yönetimi ve büyük veri çözümleri sunan bir şirket olan TERADATA⁵⁰’ya göre örneğin, bir perakendeci bu sayede kar payını yüzde 60 artırebilecektir. Bugün bile ABD sağlık sektörü, veri bilimcileri kullanarak yılda 300 milyar dolarlık tasarruf edebilecek

⁴⁹ Adnan MAZMANOĞLU, “Faktör Analizi ve Bilgisayarlarda Modern Faktör Analizi Yöntemlerinin Kullanımı, Y. Lisans Tezi, İstanbul Üniversitesi, 1979

⁵⁰ Türkiye’nin ilk Veri Bilimcisi Teradata, <http://www.itnetwork.com.tr/turkiyenin-ilk-veri-bilimcisi-teradata-turkiyede-goreve-basladi/>

durumdadır. Bu yüzden birkaç yıl içinde veri bilimcilere talep patlayacak ve 2020 de bu alanda yaklaşık 200 bin kişilik istihdam açığı doğabilecektir.

3.2. Veri İşleme: Veri Analizcileri (Data Analysts) ve Veri Bilimciler (Data Scientists) kime denir?

► Mademki **Büyük veri**’ artan hacim, hız ve çeşitlilik gösteren veri kümeleri olarak ifade edilmektedir. Ayrıca ’veriyi iyice anlamak ve içselleştirmek için maliyet etkin ve yenilikçi işleme yöntemleri talep eden büyük hacimli, hızlı ve çeşitli veriler olarak da tanımlanabilir. Büyük Veri kullanan bazı Avrupa ülkeleri için bir örnek verelim.

► Luxemburg Sosyal Güvenlik Kurumundan gelen büyük boyutlu veriyi işleyerek yıllık brüt kazanç verisini açıklamaktadır.

► Fransa gelir vergisi kayıt istemi ve devlet memurları kayıt sistemlerini kullanarak, kazanç ve işgücü maliyeti istatistiklerini yayınlamaktadır.

► Hollanda 2005 yılından bu yana kazanç istatistiklerini çalışanların verdikleri kazanç beyanlarından oluşan vergi veri tabanından oluşturmaktadır.

Uygulama örneklerine göre istatistiklerin iki farklı süreç içerdiğine dikkat etmek gerekir: 1. veri kümelerinin tanımlanması (descriptive statistics) ve 2. örneklemeye dayalı veri kümeleri hakkında sonuçların alınması (tahminler, kararlar, öngörü vb. ~making estimates, decisions, predictions, etc.). Böylece, istatistik uygulamaları iki alana ayrılabilir: Tanımlayıcı (descriptive statistics) ve çıkarımsal istatistikler (inferential statistics). İşte tanımlayıcı veri analizcilerin veri düzenlenmesi içindeki rolleri; varolan raporlama, durumu açıklamakla görevli kişi olarak tanımlandı⁵¹. Devasa büyüklükteki bu çok hızlı veri aktarımının altından kalkmak, veri aktarımı, kurumların durumları anında algılayarak hızlı karşılık vermelerine olanak sağlamaktadır. Büyük veri kullanımı için teknolojik kapasitenin ve altyapının oluşturulması yanında büyük veriye ilişkin araçları kullanabilecek, modelleme yapabilecek ve yorumlayabilecek **nitelikli veri bilimcisi personelinin yetiştirilmesi için**, veriyi matematiksel

⁵¹ F. AKDENİZ, “İstatistikte Yeni Eğilimler ve Yöntemler”, TÜİK, İstatistik Araştırma Dergisi S.39

işlemler ve istatistiksel yöntemlerle (problemi tanımlama, değişken(ler)i tanımlama, belirli bir kütle tanımlama, içinden alınacak örneği saptama vb.) sunuma hazırlayabilmek doğrusal (lineer) cebir ~ matris cebri, sayısal analiz, makine dili kodlama gibi alanlarda çalışmış olma gerekliliğini ifade etmeliyiz (Akdeniz, F, 2013). Bunun yanında hazır istatistik yazılım paketleri hakkında bilgi, beceri ve deneyim kazanmış olmaları gerekir. ***Farklı analiz yöntemlerini geliştirme (anlık verinin kontrolü, analizi vs.) büyük veriyi yönetmek için geliştirilen teknolojilerinden haberdar olmak. Bunun yanında bunları kullanabilecek beceri ve bellek kıvraklığına sahip insanların en az ileri teknolojiler kadar da önemli olduklarının ayırımının farkına varan kurumların veri bilimcileri gereksinimleri günden güne artmaktadır.*** Inferential (Predictive) Analysis~ çıkarımsal istatistik analizi yapan veri bilimcileri hakkında Veri bilimcilerin her ne kadar veri organizasyonlar içinde henüz rolleri tanımlanmamış iken, yukarıda açıkladığım büyük veri yönetimi için, bir istatistik bilim adamı gibi hipotezler kurup, bu hipotezlerin doğruluğunu ya da yanlışlığını test etmek için araştırmalar yapar. Hatta *veri odaklı uygulamalar* geliştirir demektedir makalesinde. Bu uygulamaların birkaç önemli karakteristiği olmasına rağmen, bizce veriden yeni “*veri elde etmek*” diye ifade edilebilir. F. Akdeniz devamla;

- Adı geçen uygulamalar veriden yararlanarak ortaya çıkar.
- Bu uygulamaların kullanımı sonucunda yeni veri ortaya çıkar.
- Yeni ortaya çıkan bu veri, uygulamaların daha verimli bir iyileştirme için kullanılır.

Veri Odaklı Uygulamalarda en can alıcı nokta; farklı sonuçlar üretebilmek. Veri bilimcilerinin yarattığı katma değer, veriyi görsel olarak sunabilmek, verileri ayrıştırabilmek ve düzenlemek. Aynı zamanda, bu verileri yorumlayabilmek için gelişmiş algoritmalar hazırlamak ve bu şekilde işlenen veriler ile iş kararları verilmesini sağlamaktır. Çağımızın en cazip ve aranan mesleklerinden biri olan veri bilimciliği için nasıl bir eğitim ve alt yapı gerekiyor bunu düşünmek gerekir. Çünkü henüz veri bilimci yetiştiren bir akademik kurum yok. Ağırlıklı olarak Bilgisayar Bilimleri, Matematik, İstatistik, Yönetim ve Bilişim sistemleri eğitime sahip olan

kişiler, ekonomi ve işletme alanında da eğitim görmüş olabiliyor. Donanım, yazılım ve bilişim teknolojisindeki hızlı gelişmenin çok önem kazandığı günümüzde, veri bilimcilerin ne kadar değerli bir insan kaynağı olduğu konusunda her sektör uyum içindeler.

3.2.1. Veri Bilimcisinin Kurum Bazındaki İşlevleri Nelerdir?

Veri bilimcisi özetle, mesleğin kapsamlı kavrayışını teknik beceri ve istatistikî bilgi ile birleştirerek veri toplama ve yorumlama amacındaki kuruluşlar için yöntemler ortaya koyar. Veri bilimci bir kuruluşa cevap bekleyen soruları belirlemesinde, uygun verinin toplanması için yöntem ve teknolojik araçlar geliştirmesinde ve toplanmış veriden cevaplar çıkarmak için istatistikî modeller kurmasında yardım eder. Kısaca bu profesyoneller, veri çözümlemesi kullanarak, bir kuruluşa karar alma (making decision) ve müşteriler için yararlı ürün ve hizmet geliştirme konularında yardım eder ve bununla ilgili teknik beceri ve mesleki kavrayışa sahiptir.

Veri bilimcisinin işinin dakik ayrıntıları, çalıştığı sektöre bağlıdır. Biri belirli bir verinin toplanması için ihtiyaç duyulan programlamaya odaklanmış olabilirken, diğeri verinin etkinliğini ya da doğruluğunu arttıracak eşsiz yöntemler için var olan araçları kullanmaya odaklanmış olabilir. Bir başkası ise, veri toplamak ve çözümlemek için var olan araçlarla yenilerini birleştirip, kuruluşun yeni bir ürün ya da hizmet sunmasına yardım edebilir.

Örneğin, veri bilimciler sosyal ağ sitesindeki uygunluk uygulamalarının birçoğunun geliştirilmesine yardımcı olurlar. Her kullanıcının iş durumu ve eğitim bilgileri verisi ve sonra, bu bilgiler şimdiki bağlantıları ile karşılaştırılır. Uygulama, her bireyin geçmişi ve şimdiki bağlantılarına dayanarak, yeni bağlantılar, iş yönlendirmeleri veya ilgi duyulabilecek ürün ve hizmetleri tavsiye eder. Sonuçlar, genel olarak, kullanıcının ana sayfasına ya da profilinde görünür. Bu uygulamaları geliştiren veri bilimciler teknik beceri ve yaratıcılık kullanarak, web sitelerinin daha yararlı kullanıcı deneyimi yaratmasına yardım eder.

Benzer uygulamalar veri bilimcilerin site ziyaretçileri, var olan müşteriler hakkında bilgi toplanmasına, çözümlenmesine, yorumlanmasına ve raporlamasına izin verir. Kuruluşun hedefleri doğrultusunda bu tip bilgiler, alışveriş davranışları yaratmada ya da farklı

pazarlama stratejilerini denemede kullanılabilir. Örneğin, birçok site müşteri davranışlarına dayanarak, buna uygun reklamlar yayınlamaktadır. Bu uygulamaları kullanıma açmadan önce, uzman bir veri bilimci müşteri bilgilerini toplamının, çözümlemenin ve uygun sonucun gösterilmesinin araçlarını programlamalıdır.

Farklı sektörler, ya da aslında aynı sektördeki farklı firmalar, veri bilimcinin günlük işlerine gelindiğinde, farklı ihtiyaçlara sahiptir. Görevler değişirken, ihtiyaç duyulan beceriler aynı kalır. Bu alanda çalışan uzmanlar, veri toplamak ve veriyi yönetmek için uygun araçları geliştirecek teknik beceri ve programlamaya gereksinim duyarlar. Bu uzmanlar, ayrıca hangi veriyi toplayacakları konusunda ve eldeki bilgiyi türlü yorumlama yolları için, yaratıcılık ve eleştirel düşünme ve sektörün gereksinimlerini kavrama becerisi yetisine sahip olmalıdır⁵².

4. İSTATİSTİKSEL DÜŞÜNME VE İSTATİSTİK MÜHENDİSLİĞİ

Bu başlık altında yazılacak çok şeyin olduğunu söylemek gerekir. Fakat bir fikir vermek ve sonunda veri analistlerine ve veri bilimcilerine de yararlı açıklamalarda bulunmak istiyoruz. “Statistical Thinking” adlı kitabının 2. Baskısında Roger Hoerl ve Ron Snee, (2005) istatistiksel düşünmenin; iş performansını iyileştirme, iş süreçlerini iyileştirmek için istatistiksel düşünceyi uygulama kapasitesinin geliştirilerek iş liderliğine kadar uzanan bir sürece hazırlanabileceğini ifade ediyorlar. Bunu daha çok etki yaratmak için birden çok istatistiksel aracı mantıklı bir şekilde kullanmamıza yardımcı olan bir disiplin olan “*istatistik mühendisliğine*” bağlamaktadırlar. Bunun sayesinde yeni istatistiksel düşünmenin “*iş performansını artırma*” olarak bir anlam kazandığını iddia etmektedirler. İstatistiksel düşünce, yöneticilere değişkenlikle nasıl mücadele edilebileceği, sistem ve süreçler üzerinde etkili kararlar alabilmek için nasıl veri toplanacağı ve bu verilerin nasıl kullanılacağını gösterir.

⁵² N’den N’ye Gezinti istatistik dergisi, “Görsel Veri MADENCİLİĞİ Nedir?”, s.50-51, Kasım- Aralık 2012, sayı 9, Ankara:<http://www.wisegeek.com/what-is-visual-data-mining-htm>

Tüm iş faaliyetlerinde ve endüstride, doğa bilimlerinde karşılaştığımız problemlerin birçoğunda problemi yaratan olaylardaki değişkenlerin değeri önceden bilinmediğinden belirsizlikler vardır. Belirsizliğin olduğu yerlerde mutlaka değişkenlikte olacaktır. Yöneticiler bu belirsizlik ortamında karar vermek için iş süreçlerinde var olan değişkenliğin yapısını, nasıl dağıldığını anlamak durumundadırlar. Çünkü verilen hatalı kararlar maliyetlerin yükselmesine ve çabaların boşa gitmesine neden olacaktır (Makrymichalos ve Antony 2005)⁵³. İşte, İstatistik Mühendislik (Statistical Engineering) , süreç değişkenliğini azaltma amaçlı bir algoritmadır. Bu algoritma, özellikle üretim ve montaj süreçlerindeki kronik problemleri çözmek için tasarlanmıştır. İstatistik Mühendislik algoritmasının temeli, süreçte maliyet azaltmada etkili olacak değişkenlikleri fark etmenin, süreç davranışları hakkındaki bilgi düzeyinin artması ile doğru orantılı olduğu söylenebilir. Hatta süreç bilgisini deneysel olarak arttırmanın yolu olarak değişik işlevler olarak **Sor, Planla, Veri Topla, Analiz Et, Karar Ver** (Question, Plan, Data, Analysis, Conclusion = QPDAC) döngüsü önerilmektedir (M. TANIK, C. ÖZLER, 2014). Son olarak Whitaker ve Johnson (2001)'a göre kaliteyi ve iş performansını arttırmak isteyen tüm yöneticiler "*istatistiksel*" düşünebilmeli ve temel istatistik teknikleri uygulayabilmelidir. Çünkü Juran (1989) işletmelerde kalitenin güvence altına alınması için uygulanması gereken üç süreci aşağıdaki gibi tanımlamıştır:

1. Kalite Planlaması
2. Kalite Kontrol
3. Kalite İyileştirme⁵⁴.

Görüldüğü gibi, yeni boyutlarıyla betimsel istatistik yöntemleri, yan yana sıralanmış, birbirinden ayrık bir yöntem dizisi oluşturmuyor. Aralarında yakın ilişki vardır. Bilgisayarların yenilenmelerine ya da yenilerinin ortaya çıkmasına yol açtığı bu tekniklerin, ister "**güçlü araçlarla kaynaklara**

⁵³ M. TANIK, C. ÖZLER, "İstatistiksel Mühendislik Algoritmasının Altı Sigma Projelerinde Kullanılması Üzerine Bir Araştırma, Sıtkı Koçman Üni., S.B. Enstitüsü Dergisi, Sayı 32, 2014 Bahar, 1-15, Muğla

⁵⁴ M. TANIK, C. ÖZLER, "İstatistiksel Mühendislik Algoritmasının Altı Sigma Projelerinde Kullanılması Üzerine Bir Araştırma, s.2, Sıtkı Koçman Üni., S.B. Enstitüsü Dergisi, Sayı 32, 2014 Bahar, 1-15, Muğla

dönüş" diye kabul edilsin, ister "**modelsiz istatistik**" olarak öne sürülsün, bugün için, "**tümevarımcı**" istatistiğin karşısına güçlenmiş olarak çıktığı bir gerçektir: Çok boyutlu, çok sayıda verilere yanaşmayı, hızlı bilgi edinmeyi sağlamakla kalmayıp, bunları bir düzene koymaya, değişkenleri indirgerken özellikler arasındaki bağlantıları ortaya çıkarmaya, ilk bakışta görülmeyen yapıyı röntgen gibi görüntülemeye olanak veriyor. Bu arada, olağandışı durumları, ölçüm yanlışlarını, soruşturma bozukluklarını da ortaya çıkarıyor. Ayrıca, sosyokültürel gösterge gibi, sentetik indekslerin yapılmasına da yol açıyor⁵⁵.

Ancak, bu durumun, "**klasik yaklaşım**" denen, tümevarımcı istatistiği destekleyeceğini kabul etmek olanaksızdır. Karmaşık veriler nasıl bir gerçekse, olasılık, rastlantı da bir gerçektir. Varsayımlar kısıtlayıcı gibi görülsede, bunlara dayanan modeller, bugüne dek yapılan genellemeler küçümsenemez. Kuşkusuz, her alanda olduğu gibi, istatistik dalları arasında bir etkileşim olacaktır. Bir yandan "**veri çözümleme**" teknikleri yerlerine otururken yenileri ortaya çıkacak; öte yandan, tümevarımcı istatistiğe yeni ufuklar açılacak, model kurma kolaylaşacak, yeni gelişmeler söz konusu tekniklerden kaynaklanacaktır.

Bugün için önemli olan, "**çok boyutlu çözümleme**" tekniklerinin iyi anlaşılması, matematiksel ilkelerin iyi bilinmesidir. Hepsinin hazır programları olsa da⁵⁶, başta lineer cebir, matris hesabı olmak üzere, yöntemlerin özü bilinmeden, bunlar kullanılamaz. Ayrıca sonuçların yorumu da beceri, ustalık ve deneyim gerektirir.

KAYNAKÇA

Editör Şeref KALAYCI: SPSS Uygulamalı Çok Değişkenli İstatistik Teknikleri, Asil Yayın Dağıtım Ltd. Şti., Ankara, 2010

Adnan MAZMANOĞLU, "Faktör Analizi ve Bilgisayarlarda Modern Faktör Analizi Yöntemlerinin Kullanımı, Y. Lisans Tezi, İstanbul Üniversitesi, 1979

⁵⁵ Sentetik indekslere ilişkin olarak bkz: Lebart - Morineau - Tabard, a.g.y., s:221, 268, 280; SPSS (Statistical Package for the Social Science), s:529.

⁵⁶ Başta SPSS olmak üzere, kaynakçada verilen yapıtların çoğunda, yöntemlere ilişkin programlar da bulunmaktadır.

- Benjamin FRUCTER, “Introduction to Factor Analysis”, D. Von Nostrand Company, 1968
- H. Durucasu, “Asal Bileşen Analizi ve Bir Uygulama Denemesi”, “Y.Lisans Tezi, s. 18-23, Anadolu Üni., 1991
- H.NIE NORMAN, Hull C.Hadlai, Jenkins Jean G. , Stein BRENNER KARIN, BENTDALE H., SPSS (Statistical Package for The Social Sciences), Mc GRAW-HILL BOOK COMPANY, USA, 1975
- BENNETT Spencer and BOWERS David, “An Introduction to Mutivariate Techniques for Social and Behavioural Sciences”, Mc Millan Press, 1976
- Adnan MAZMANOĞLU, “Herkes İçin TEMEL İSTATİSTİK YÖNTEMLERİ ve Uygulamaları I, 2. Basım, Nobel Akademik Basım Yayın Dağ. , Ocak 2018 Ankara
- James T.Mc Clave, P. G. Benson, T. Sincich, “Statistics”, Pearson, U.S.A., 2011
- McClave, Sincich, “Statistics”, International Edition, USA, 2011
- John Neter, W. Wasserman, M.H.Kutner, “Applied Linear Statistical Models”, IRWIN, USA, 1985
- M. TANIK, C. ÖZLER, “İstatistiksel Mühendislik Algoritmasının Altı Sigma Projelerinde Kullanılması Üzerine Bir Araştırma, Sıtkı Koçman Üni., S.B. Enstitüsü Dergisi, Sayı 32, 2014 Bahar, 1-15, Muğla
- STEFAN H. STEINER, R. JOCK MACKAY, “Statistical Engeneering An Algorlthm for Reducing Variation in Manufacturing Processes”, ASQ, 2005
- Roger Hoerl, Ren Snee, “Statistical Thinking”, John Wiley & Sons, 2012