

ÇOK DEĞİŞKENLİ İSTATİSTİKLERİN ARAŞTIRMALARDA KULLANIMI

Araştırmalarda incelenen olaylar göstermektedir ki tek değişkenli istatistiklerin kullanılması problemi açıklamakta yetersiz ve eksik kalmaktadır. Tek değişkenli istatistiklerde çözümlenen olay tektir, tek değişken incelenmektedir. Bilimsel çalışmalar ise tek değişkenle açıklanamayacak kadar karmaşıktır. Araştırmaya konu olan bir problemin çözümünde kuşkusuz problemi etkileyen birçok faktör vardır ve çözülecek problem bu birçok faktörü dikkate alarak incelenmelidir. Bu nedenle tek değişkenli istatistiklerin sınırlılığı, çok değişkenli istatistiksel analizleri doğurmuştur. Böylece, tek değişkenli istatistiklerde varsayılan kısıtlamalar ortadan kalktığından araştırmalarda daha objektif ve tutarlı sonuçlar elde edilir. Tek değişkenli istatistiksel analizlerin en önemli kısıtlayıcı varsayımı ise birçok faktörün deneysel olarak kontrol altında tutulması ve her defasında tek bir faktörün etkisinin incelenmesidir. Bu şu demektir incelenen değişken üzerindeki faktörler türdeş ya da sabit tutulmasıdır. Oysaki çok değişkenli istatistiksel analizlerde bazı kontrollü denemeler dışında böyle bir kısıtlayıcıdan ya da özellikten söz edilmez. Çok değişkenli istatistiksel analizlerin en önemli varsayımı, verilerin çok değişkenli normal dağılımlı kitleden çekilmiş olduğudur. Çok değişkenli istatistiksel analizlerde, birden çok özelliğin analizi ile ilgilenildiğinden en az ikiden çok değişken söz konusudur.

	Tek Değişkenli Analizler	Çok Değişkenli Analizler
Değişken Sayısı (P)	P=1	P=2 (iki değişkenli) P>2 (çok değişkenli)
Varsayımı	Değişkeni etkileyen diğer faktörler(değişkenler) türdeş ya da sabittir.	N birimli p tane değişkenin normal dağılım gösterir.

Çok değişkenli istatistiksel analizler, incelenen olay ve çevresindeki çok sayıda içsel ve dışsal faktörleri dikkate alarak, problemi doğasındaki yapısına ilişkin bilgilere göre incelemek ve çözümlere ulaşmak için geliştirilmiş yöntemler bütünüdür (Özdamar,2002:1).

Çok değişkenli istatistiklerin uygulanma amaçları ise

- Veri İndirgeme
- Kümeleme ve Sınıflama
- Ölçekleme
- Hipotez Testleri (Hipotez Oluşturma)

A) Veri İndirgeme

Veri indirgemenin amacı, P sayıda değişken içeren veri setinin varyasyonunu açıklayan ve aralarında ilişki bulunmayan daha az sayıda değişkenle ($k < p$) veri yapısını açıklamaktır.

B) Kümeleme ve Sınıflama

Amaç, popülasyon özellikleri bilinmeyen yapılar hakkında prototip kümeler (grup, sınıf) belirleme çalışmalarına yardımcı olup daha önceden belirlenmiş gruplara yeni birimlerin atanmasını sağlamaktır.

C) Ölçekleme

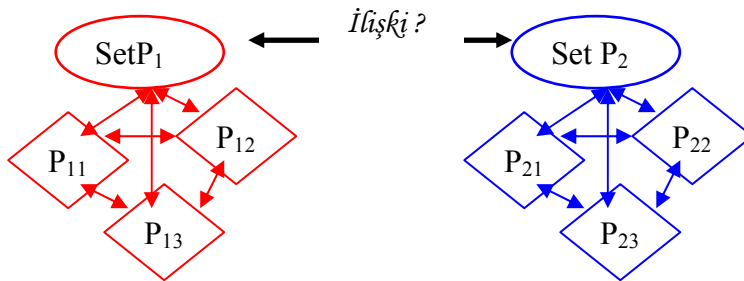
P sayıda değişken içeren p boyutlu ölçümlerden daha az sayıda değişken kullanarak birimlerin gösterilmesini, tanımlanmasını sağlamak ve birimlerin birbirleri ile $k < p$ boyutlu ölçekte benzerlik ve farklılıklarını incelemektir.

D) Çok değişkenli hipotezlerin test edilmesi

K toplumun çok değişkenli ortalamalar vektörünün eşitliği / farklılığı üzerine kurulacak hipotezleri test etmede ÇDİA yöntemleri kullanılır.

ÇOK DEĞİŞKENLİ İSTATİSTİKSEL ANALİZLER

Setlerarası Korelasyon Analizi: $P_1 > 2$ ve $P_2 > 2$ olmak üzere çok değişkenli iki ya da daha fazla değişkenler seti (grup) arasındaki korelasyonları açıklamak için kullanılan bu yöntem $P_1 > 2$ değişkenler setinin doğrusal bileşenleri ile diğer ($P_2 > 2$) değişkenler setinin doğrusal bileşenleri arasındaki korelasyonları inceleyen bir yöntemdir.



Örneğin, bir öğretmen öğrenme yeteneğini ölçen 3 ölçek (değişken) ile okuldaki başarıyı ölçen 4 ölçek (değişken) seti arasındaki ilişkiyi hesaplamada setler arası korelasyon analizini kullanabilir.

<i>Öğrenci No</i>	<i>Set1(Öğrenme Yeteneği)</i>			<i>Set2 (okuldaki başarı)</i>			
	Muhakeme	Anlama	Analitik	Türkçe	Mat	Fen Bil.	Sos. Bil.
<i>1</i>							
<i>2</i>	25	68	98	45	52	58	90
...	68	75	85	78	87	95	78
...	45	55	30	47	60	45	93
<i>N</i>	50	52	47	45	74	85	74

SPSS programında analiz için script yazılmalı ya da MATLAB, MİNİTAB, SAS ya da STATİSTİCA programlarından faydalanılabilir.

Setlerarası Korelasyon Analizi için,

- Değişken setlerinde yer alan değişkenlerin eşit sayıda olması zorunluluğu yoktur.
- Stevens'a göre setlerarası korelasyonun güvenilir ve tahminlerin tutarlı olması için birim sayısı (n), setlerdeki toplam değişken sayısının en az 20 katı olması önerilmektedir.
- Veri setinde aykırı değerlerin *bulunmamasına* dikkat edilmelidir.
- Veri setinde gereğinden fazla problemle ilgili olmayan değişkenlerin olmaması gerekir.
- Analiz edilecek değişkenler arasında tam korelasyon *bulunmamalıdır*.

Uyum Analizi: Çapraz tablo biçiminde gösterilen değişkenler arasındaki uyumluluğu ya da bir değişkenin kendi kategorileri arasındaki uyumluluğunda kullanılan *kategorik veri analizi* yöntemidir. Ayrıca bu istatistik, kategorik veri indirgeme yöntemi olarak da kullanılabilir.

Yöntem, çapraz tablo ya da iç içe çapraz tablo biçiminde gösterilen değişkenlerin alt sınıflarındaki yoğunlaşmalarını “öklid uzaklığı” ya da “pearson ki kare” uzaklığı cinsinden hesaplayarak ağırlıklı ana bileşenlere göre analiz eder.

Basit (r*c) ve Çoklu (r*c*m) Uyum analizi olarak iki farklı biçimde uygulanır.

Örnek: Cinsiyet ve yaşın günlük yaşam aktivitesi üzerindeki etkisi araştırılıyor olsun.

Cinsiyet: Kadın / Erkek (2 grup)

Yaş:<20 (1) ;20-39 (2) ; >40 (3)

Aktivite: Çokaktif (1) ; aktif (2); uyuşuk (3)

Cinsiyet	Yaş	Aktivite	Frekans
E	1	1	2
E	2	2	3
E	3	3	5
K	1	1	3
K	2	2	4
K	3	3	3

SPSS 11.0 programında analiz için: Spss<Analiz<Data Reduction<Correspondence Anaysis (P=2) / Optimal Scaling (p>2)

Hotelling T² Testi: İki ve daha çok değişkenli ($p \geq 2$) tek örnek ($g=1$) ve iki örnek ($g=2$) hipotezlerinin test edilir.

Örnek: Bir ilköğretim okulunda 8-A ($N_A=30$) ve 8-B ($N_B=33$) şubelerinde okuyan öğrencilerin, Türkçe, Matematik, Fen Bilgisi, Müzik, Edebiyat ders başarıları bakımından bu iki şubedeki öğrenciler farklılık göstermekte midir?

Öğrenci No	Türkçe	Mat.	Fen Bil.	Edebiyat	Müzik	Şube No
1	78	54	68	52	85	A
2	25	68	98	58	100	A
...	68	75	85	95	78	A
...	45	55	30	45	60	B
$N_A=30+N_B=33$ = 73	50	52	47	85	74	B

SPSS 11.0 programında analiz için: Spss<Analiz<General Linear Model(GML)<Multivariate

Çok Değişkenli Varyans Analizi (MANOVA): Grup sayısı ($g \geq 2$) ikiden fazla olduğunda Hotelling T² testinin alternatifidir.

Örnek2: 8-A($N_A=30$), 8-B($N_B=33$) ve 8-C ($N_C=27$) şubelerinde okuyan öğrencilerin Türkçe, Matematik, Fen ve Edebiyat ders başarıları şubelere göre farklılık göstermekte midir?

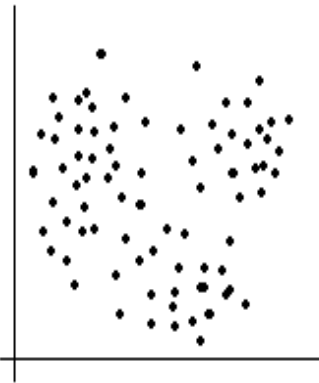
Öğrenci No	Türkçe	Mat.	Fen Bil.	Edebiyat	Şube No
1	78	54	68	52	A
2	25	68	98	58	A
2	68	75	85	95	A
4	45	55	30	45	B
...					B
...					C
99	50	52	47	85	C
100	50	52	47	85	C

SPSS 11.0 programında analiz için: Spss<Analiz<General Linear Model(GML)<Multivariate

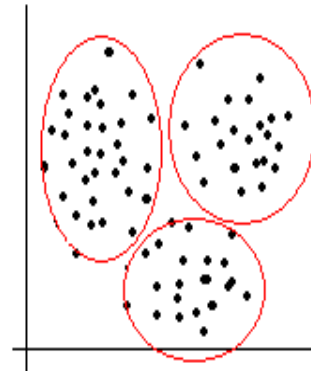
Çok Değişkenli Kovaryans Analizi (MANCOVA): Grup sayısı ($g \geq 2$) ikiden fazla normal dağılım gösteren $k \geq 2$ popülasyona dair kurulan hipotezler ortak değişkenlere göre test edilir.

Kümeleme Analizi: X veri matrisinde yer alan ve doğal gruplamaları kesin olarak bilinmeyen birimleri, değişkenleri ya da birim ve değişkenleri birbiri ile benzer olan alt kümelere (grup, sınıf) ayırmaya yardımcı olan yöntemler topluluğudur (Tatlıldil,1992). Yani kümeleme analizi, yapıları hakkında kesin bilgilerin bulunmadığı bir veri yığını içindeki birimleri, değişkenleri ya da birim ve değişkenleri birbiri ile benzer olan alt kümelere (grup, sınıf) ayırma yöntemidir. Kümeleme analizi (1) aşamalı ve aşamalı olmayan (küme sayısı belli) olmak üzere 2 şekilde yapılmaktadır.

Veri yığını



Kümeleme analizi sonucu veri yığınları



$$X(8 \times 3) = \begin{bmatrix} \text{Yaş} & \text{Boy} & \text{Ağırlık} \\ 25 & 165 & 55 \\ 32 & 183 & 65 \\ 28 & 175 & 74 \\ 26 & 167 & 60 \\ 24 & 183 & 65 \\ 32 & 174 & 72 \\ 26 & 155 & 80 \\ 29 & 164 & 70 \end{bmatrix}$$

N=8 ve p=3 Yaş, Boy ve Ağırlık değişkenlerinden elde edilmiş ölçümlere göre 8 bireyi 3 kümeye (aşamalı olmayan kümeleme yöntemi ile) ya da değişkenler dikkate alınarak doğal kümelere (aşamalı kümeleme yöntemi ile) ayırmak mümkündür.

Örnek: N sayıda öğrenciden oluşan bir grup olsun. Bu öğrencilerin de Matematik, Türkçe, Sos. Bil. ve Fen Bil. Ders başarı puanları ölçülmüş olsun. Buna göre öğrencileri kümeleme analizi yöntemiyle derslerdeki performansları bakımından farklı şubeler oluşturmada kullanabiliriz.

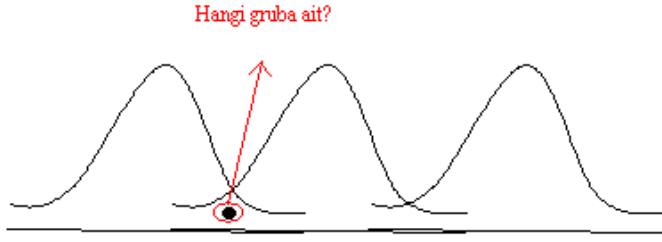
<i>Öğrenci No</i>	<i>Türkçe</i>	<i>Mat.</i>	<i>Fen Bil.</i>	<i>Edebiyat</i>	<i>Müzik</i>
1	78	54	68	52	85
2	25	68	98	58	100
...	68	75	85	95	78
...	45	55	30	45	60
N	50	52	47	85	74

SPSS 11.0 programında analiz için: Spss<Analiz<Classify<K-means cluster(aşamalı olmayan-küme sayısı belli) / Hierarchical Cluster (aşamalı-küme sayısı belli değil)

Kümeleme analizi ile kendi aralarında heterojen ve kendi içinde birbiri ile homojen kümeler oluşturulur.

Ayrırma Analizi: X veri setindeki değişkenlerin iki ve daha fazla gerçek gruplara ayrılmasını sağlayan, birimlerin p tane özelliğini ele alarak bu birimlerin doğal ortamdaki gerçek gruplarına, sınıflarına en uygun düzeyde atanmalarını sağlayacak fonksiyonlar türeten bir yöntemdir (Özdamar,2002). Ayrırma analizinin,

- 1- Grupları birbirinden ayırmayı sağlayan fonksiyonları bulmak
- 2- Hesaplanan fonksiyonlar aracılığıyla yeni gözlenen bir birimi sınıflama hatası minimum olacak biçimde g tane gruptan herhangi birine atamak amaçlıdır.



Örnek: 8-A, 8-B ve 8-C şubelerinde öğrencilerin Türkçe, Matematik ve Fen bilgisi ders başarılarına göre şubelere doğru yerleştirilip yerleştirilmediğini (sınıflamanın doğru yapılma olasılığını) ya da sonradan Türkçe, Matematik ve Fen bilgisi test başarı puanları sırasıyla 55, 65, 42 olan bir öğrenciyi istatistiksel olarak hangi şubeye dahil edileceği konusunda karar verirken ayırma analizinden yararlanmak mümkündür.

Öğrenci No	Türkçe	Mat.	Fen Bil.	Edebiyat	Müzik	Şube No
1	78	54	68	52	85	A
2	25	68	98	58	100	A
...	68	75	85	95	78	A
...	45	55	30	45	60	B
$N_A=30+N_B=33$ $= 73$	50	52	47	85	74	B

SPSS 11.0 programında analiz için: Spss<Analiz<Classify<Discrimination

Anabileşenler Analizi: Birbirleri ile ilişkili $p > 2$ değişken içeren veri matrislerinden, birbirleri ile bağımsız ve daha az sayıda yeni veri yapıları elde etmek amacıyla yararlanılan bir yöntemdir. Bu yöntemden aralarında yüksek düzeyde korelasyon bulunan verilerden daha az sayıda ve aralarında korelasyon bulunmayan yeni değişkenler türetmek ve veri indirgemesi yapmak amacıyla yararlanılır.

Kaç anabileşen olmalıdır? Sorusu içinse yaygın yönelimler şöyledir:

- *Birden büyük özdeğer sayısı kadar anabileşen seçilebilir.
- *Genel varyansın en az %67'sini açıklayan sayıda anabileşen seçilebilir
- *Grafik eğimi ile anabileşen seçimi yapılabilir.

Faktör Analizi: Bu yöntemin amacı $p \geq 2$ çok değişkenli veri yapılarını aralarında yüksek korelasyon bulunan değişkenleri bir araya getirerek yeni ve anlamlı faktör yapıları oluşturmaktır.

<i>Öğrenci No</i>	<i>Türkçe</i>	<i>Mat.</i>	<i>Fen Bil.</i>	<i>Edebiyat</i>	<i>Müzik</i>
1	78	54	68	52	85
2	25	68	98	58	100
...	68	75	85	95	78
...	45	55	30	45	60
100	50	52	47	85	74

SPSS 11.0 programında analiz için: Spss<Analiz<Scale<Multidimensional Scaling

Aynı örnek için eğer test başarılarına (değişkenlere) göre bir sınıf oluşturulacak ve her bir testi ayrı ayrı değerlendirip karar vermek yerine ÇBÖA'den yararlanarak hangi testi ya da testleri ölçüt olarak öğrenci seçimi yapacağımız hakkında karar vermemiz kolaylaşacaktır.

Not: ÇBÖA bir veri indirgeme yöntemi olarak kullanıldığında veriler özellikle nicel ise alternatif yöntem Faktör Analizidir. Benzer olguların oluşturduğu grupları belirlemek için kullanılacak ise alternatif yöntem aşamalı ya da aşamalı olmayan kümeleme analizidir.

ÖZDAMAR, Kazım., **Paket Programlar ile İstatistiksel Veri Analizi**, Kaan Kitapevi, Eskişehir, 1999

TATLIDİL, Hüseyin., **Uygulamalı Çok Değişkenli İstatistiksel Analiz**, Cem Web Ofset Ltd.Şti., Ankara, 1996