Dumlupınar Üniversitesi Sosyal Bilimler Dergisi
Sayı: 18 Ağustos 2007

Uyum Analizinin Teorik Esasları ve Regresyon Analizi İle Benzerliğinin Grafiksel Boyutta Karşılaştırılması

Nevin UZGÖREN*

Özet: İstatistiğin temel amaçlarından birisi değişkenler arasındaki ilişkiyi incelemektir. Bu amaçla kullanılan yöntemlerden olan Uyum Analizi; ele alınan değişkenlerin kategorik formda olduğu ve özellikle gözlem sayısının yetersizliği nedeniyle ki-kare analizinin uygulanamadığı durumlarda kullanılan başlıca yöntemdir. Uyum analizi, her bir değişkenin kategorileri arasındaki ilişkileri ve aynı zamanda değişkenler arasındaki kategorik bazda çapraz ilişkileri harita adı verilen grafikler yardımıyla incelemeyi sağlayan bir yöntemdir.Bu doğrultuda çalışmanın ilk aşamasında amaç, sadece iki değişken arasındaki ilişkiyi inceleyen basit uyum analizinin temel esasları ile birlikte haritaların nasıl oluşturulduğunu ve nasıl yorumlanacağını teorik olarak sunmaktır. Çalışmanın ikinci aşamasındaki amaç ise; uyum analizinin, genelde nicel değişkenler arasındaki ilişkiyi inceleyen çoklu regresyon analizi ile olan benzerliğini grafiksel boyutta karşılaştırmaktır.

Anahtar kelimeler: Uyum analizi, Regresyon analizi, Toplam değişkenlik, Artık, İstatistiksel grafikler

Theoretical Basis of Correspondance Analysis and Comparison of Its Similarities With Regression Analysis on The Graphic

Abstract: One of the main aims of statistics is to examine the relations between variables. Correspondance Analysis is an essential method used when variables is in the categorical form and chi-square analysis can not be applied because of the inadequate numbers of observation. Correspondance analysis is a method providing a tool to examine both the relationships between categories of each variable and categorical across relationships between variables by the help of graphics, named map. In this direction, the first step of study is to present the principal bases of simple correspondance analysis which examine the relationship between two variables, and how to construct and interprof the maps. The aim of the second step of study is to compare similarity of correspondance analysis with multiple regression analysis grafically, generally examining the relationship between quantitive variables.

Keywords: Correspondence analysis, Regression analysis, Total inertia, Residual, Statistical graphics

GİRİŞ

Uyum analizinin temel amacı iki veya daha fazla kategorik değişken arasındaki birlikteliği, değişkenlerin kategorilerini az-boyutlu bir uzayda noktalar şeklinde göstererek analiz etmektir (Clausen, 1998). Son

* Yrd.Doç.Dr., Dumlupınar Üniversitesi İ.İ.B.F., İşletme Bölümü Sayısal Yöntemler ABD.

dönemlerde kategorik verilerin analizinde sıklıkla kullanılan yöntemlerden biri olan uyum analizinin esası, değişkenlerin kategorileri arasındaki benzerlikleri ya da farklılıkları uzaklıklar cinsinden ifade ederek, elde edilen sonuçları harita adı verilen grafikler yardımıyla görsel olarak sunmaktır (Özdamar, 2002).

Bu yöntem ilk olarak Fransa'da 1960'lı yılların sonlarında Jean-Paul Benzecri tarafından geliştirilmiştir, ancak son zamanlarda İngilizce konuşan ülkelerde artan bir popülariteye sahip olduğu görülmektedir. Ayrıca birçok ülkede birbirlerinden bağımsız olarak, optimal scaling, reciprocal averaging, optimal scoring, quantification method ya da homogeneity analysis adı altında benzer tekniklerin geliştirildiği görülmektedir (http://www.statsoft.com/textbook /stcoran.html, 2004).

Uyum analizinin ilk aşaması kategorik değişkenlere ait yanıtları çapraz tablolar (kontenjans tabloları) yardımıyla frekanslar halinde özetleyerek, grafiksel formda sunmaktır. Analizin ikinci aşaması ise, araştırmadaki farklı sorular arasındaki ilişkileri açıklamaktır. Eğer bir araştırmada sorular nicel yanıtlar içeriyorsa, regresyon analizi, faktör analizi, ana bileşenler analizi gibi korelasyona dayalı yöntemler kullanılabilmektedir (Greenacre, 2002). Anket çalışmalarında ortaya çıkan kategorik yanıt durumunda ise, değişkenler arasındaki ilişki genellikle çapraz tabloları içeren ki-kare analizi ile incelenmektedir. Ancak çapraz tabloların analizinde, ki-kare analizi özellikle satır ve sütun değişkenlerine ait kategori sayısının çok olması ve buna bağlı olarak göze frekanslarının yetersiz kalması durumunda kullanılamaz hale gelmektedir. Bu gibi durumlarda uyum analizi, kategorik verilerin anlaşılmasını ve yorumlanmasını kolaylaştıran ve veri analizine grafiksel bir yaklaşım sunan çok değişkenli bir analiz yöntemi olarak karşımıza çıkmaktadır (Seyfullahoğulları, 2003). Dolayısıyla küçük tablolarda grafiğin nasıl bir şekil alabileceği tasavvur edilebileceği için, uyum analizi küçük tablolardan ziyade çok daha büyük tabloların analizinde ve yorumlanmasında daha büyük bir önem arz etmektedir (http://www.statsoft.com/textbook/stcoran.html, 2004). Bu bağlamda uyum analizi, aynı zamanda bir frekans tablosunun ki-kare (ya da phi-kare $\varphi^2 = \frac{\chi^2}{n} = \Lambda^2 =$ toplam inertia) değerinin ayrıştırılmasına yönelik bir teknik olarak da tanımlanabilmektedir (Clausen, 1998).

Bu doğrultuda çalışmanın amacı ilk olarak iki kategorik değişkeni içeren basit uyum analizinin temel esasları ile birlikte haritaların oluşturulması ve yorumlanmasına ilişkin teorik bilgileri vermek; ikinci olarak değişkenler arasındaki ilişkileri analiz etmek için kullanılan ve gerçekte analiz prosedürü olarak çok farklı görünen regresyon ve uyum analizlerinin benzerliklerini grafiksel boyutta göstermektir.

UYUM ANALİZİNİN TEORİK ESASI

Daha önce belirtildiği gibi uyum analizi, iki-boyutlu[1] bir kontenjans tablosunun kategorileri arasındaki birlikteliği resmetmek için geliştirilen istatistiksel bir tekniktir. Satır ve sütunlar sırası ile i (i=1,2,…I) ve j (j=1,2,…,J) olarak tanımlandığında, iki-boyulu bir kontenjans tablosu aşağıdaki gibi gösterilebilir:

Tablo 1: İki-boyutlu kontenjans tablosunun genel gösterimi

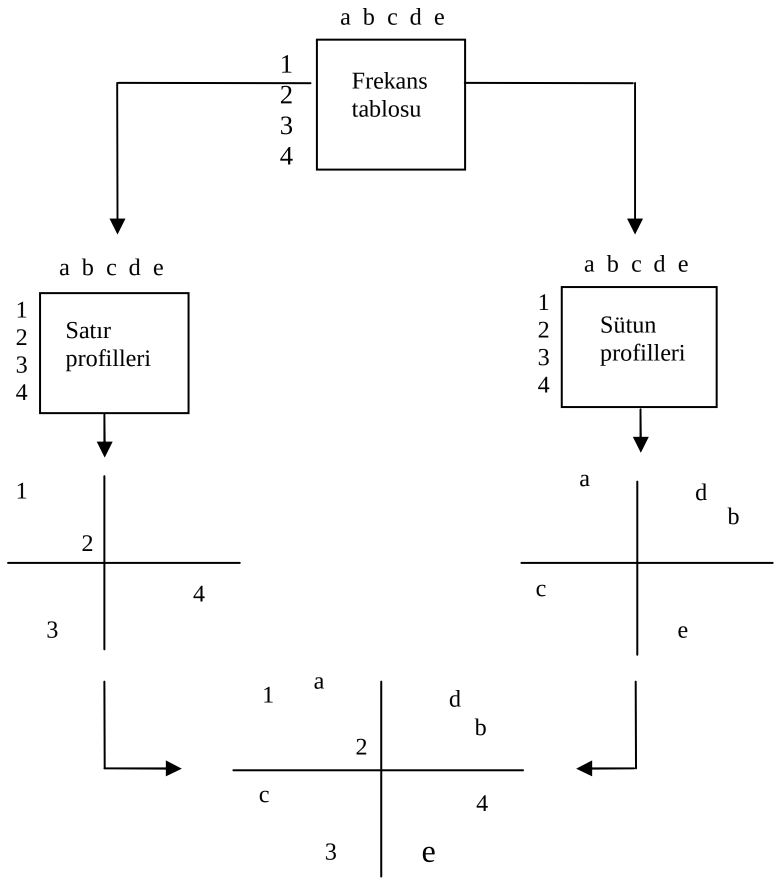
	1	2	…	J	Satır toplamı
1	n_{11}	n_{12}	…	n_{1J}	n_{1+}
2	n_{21}	n_{22}	…	n_{2J}	n_{2+}
⋮	⋮	⋮		⋮	⋮
I	n_{I1}	n_{I2}	…	n_{IJ}	n_{I+}
Sütun toplamı	n_{+1}	n_{+2}	…	n_{+J}	n

Burada $n_{ij,}$ i. satır ve j. sütundaki gözlemlenen frekans değerini ve n toplam gözlem sayısını göstermektedir.

Uyum analizi, göze frekansları yardımıyla iki değişkenin gözlemlenen birlikteliğini sunan ve genel olarak yukarıdaki gibi tanımlanan iki-boyutlu bir kontenjans tablosundan yararlanarak, bir değişkenin belirli düzeylerinin diğer bir değişkenin bazı düzeyleriyle birlikteliğinin olup olmadığının belirlenmesidir. Bu doğrultuda uyum analizi iki-boyutlu bir uzayda, iki-boyutlu bir kontenjans tablosunun satır ve sütunlarını tablodaki birliktelikleriyle tutarlı olacak şekilde noktalar halinde göstermek için geliştirilen geometrik bir tekniktir (Lee, 2006). Bu durumda problem noktalara en iyi uyan iki-boyutlu uzayı bulabilmektir. Uyum analizinin analitik süreci aşağıda gösterileceği gibi üç aşamadan oluşmaktadır (Clausen, 1998):

---

[1] Uyum analizi ilişki araştırılacak değişken sayısının üç veya daha fazla olması durumunda çok-boyutlu tablolara da uygulanabilmekte ve çoklu uyum analizi olarak isimlendirilmektedir. Ancak bu çalışmada sadece iki değişkeni içeren iki-boyutlu tablolar dikkate alınmıştır.



Şekil 1: Uyum Analizinin Üç Aşamasının Şematik Gösterimi

Bu sürece göre ilk olarak satır ve sütun profilleri hesaplanmakta, ikinci aşamada satır ve sütun profilleri iki-boyutlu uzayda ayrı ayrı resmedilmekte ve son aşamada ise satır ve sütun profilleri iki boyutlu ortak bir harita üzerinde gösterilmektedir. Uyum analizinde harita adı verilen bu grafiklerin gözlemlenen frekanslara göre değil, frekansların satır içindeki nispi önemini gösteren satır profillerine ve benzer şekilde frekansların sütun içindeki nispi önemini gösteren sütun profillerine göre çizildiği görülmektedir. Bu nedenle uyum analizinin anlaşılabilmesi için profil, mass, ki-kare uzaklığı ve toplam inertia (toplam değişkenlik) gibi dört temel kavramın açıklanması gerekmektedir.

Uyum Analizinin Temel Kavramları

<u>Profiller:</u> Bir kontenjans tablosunun yorumlanmasında her bir gözedeki gözlemlenen frekansları yorumlamak uygun değildir. Çünkü her bir satır ve sütun farklı sayıda yanıt içerir. Her bir satır ve sütunun kıyaslanması için aynı esasın benimsenmesi gerekir. Bu esas da satır ve sütun toplamlarını baz almak suretiyle, her bir satır ve sütun toplamını 1' e eşit kılacak şekilde nispi frekansları hesaplamaktır. Elde edilen bu satır ve sütun nispi frekans değerleri de satır ve sütun profilleri olarak kabul edilir (Greenacre, 1994). Dolayısıyla satır (sütun) profilleri; satır (sütun) değişkeninin kategorilerinin sütun (satır) değişkeninin her bir kategorisi için hesaplandığı nispi frekans değerlerini gösterir (Clausen, 1998). Uyum analizinde genellikle ilgilenilen satır profilleridir, çünkü genel olarak açıklanan değişkenin kategorileri için sütunlar, bir ya da daha fazla açıklayıcı değişkenin kategorileri için satırlar kullanılır (Greenacre, 1994). Bu nedenle satır ve sütunlar için yapılacak işlemler benzer prosedürler gerektirdiğinden tekrardan kaçınmak üzere, yapılacak açıklamalarda genelde satır değişkeni dikkate alınacaktır. Satır profillerinin genel gösterimi aşağıdaki tabloda özetlenmiştir:

Tablo 2: Satır Profillerinin genel gösterimi

	1	2	…	J	Satır toplamı	**Satır mass**
1	$a_{11} = \frac{n_{11}}{n_{1+}}$	$a_{12} = \frac{n_{12}}{n_{1+}}$	…	$a_{1J} = \frac{n_{1J}}{n_{1+}}$	1	$\frac{n_{1+}}{n}$
2	$a_{21} = \frac{n_{21}}{n_{2+}}$	$a_{21} = \frac{n_{21}}{n_{2+}}$	…	$a_{2J} = \frac{n_{2J}}{n_{2+}}$	1	$\frac{n_{2+}}{n}$
⋮	⋮	⋮		⋮	⋮	
I	$a_{I1} = \frac{n_{I1}}{n_{I+}}$	$a_{I2} = \frac{n_{I2}}{n_{I+}}$	…	$a_{IJ} = \frac{n_{IJ}}{n_{I+}}$	1	$\frac{n_{I+}}{n}$
Ortalama Satır profili	$a_{+1} = \frac{n_{+1}}{n}$	$a_{+2} = \frac{n_{+2}}{n}$	**…**	$a_{+J} = \frac{n_{+J}}{n}$	**1**	

Herhangi bir satır profili (ya da sütun profili) bir matematiksel vektör olarak tanımlanabilir ve her vektör uzayda bir nokta olarak gösterilebilir. Bu durumda herhangi bir profil elemanı uzayda bir koordinat meydana getirir. Örneğin sütun sayısı üç olduğunda her satır profili üç-boyutlu bir uzayda noktalar olarak gösterilebilir. Eğer iki satır kategorisine ait profiller birbirlerine benzer ise, bu kategoriler birbirlerine daha yakın noktalar, aksi durumda uzak noktalar olarak yer alır (Clausen, 1998). Ancak birçok uygulamada ele alınan uzay üçten daha büyük boyuta sahiptir, dolayısıyla her bir profilin pozisyonunu belirlemek zordur. Ancak profillerin boyutları indirgenerek, -örneğin iki-boyutlu bir uzay gibi-, daha kolay bir formatta profilleri göstermek mümkün olabilmektedir (Greenacre, 1994)

Ayrıca ortalama satır ve ortalama sütun profili olmak üzere iki ayrı kavram daha vardır. Ortalama satır profili farklı sütunlardaki toplam gözlem sayısının genel toplama bölünmesi (n_{+j} / n) ve ortalama sütun profili de farklı satırlardaki toplam gözlem sayısının genel toplama bölünmesi (n_{i+} / n) ile elde edilen sonuçtur. Bu noktalar merkez olarak isimlendirilir ve noktaların ana eksenlerin orijine göre yerini belirler. Eğer bir profil ortalama profilden çok farklıysa nokta orijinden uzak, buna karşılık ortalama profile yakınsa profiller merkeze yakın yer alacaklardır. Eğer kategoriler eşit profillere sahipse, tüm noktalar merkezde toplanacaktır (Clausen, 1998).

Mass: Uyum analizinde ikinci önemli kavram satır ve sütun masslarıdır. Satır massları (marjinal satır profilleri), satırlara ilişkin marjinal frekansların genel toplama bölünmesiyle (n_{i+}/n) elde edilir ve analizde her bir satır profilini ağırlıklandırmak için kullanılır. Bu ağırlıklandırma sisteminin amacı, her bir yanıtın her bir profil noktasına eşit katkıda bulunmasını sağlamaktır (Greenacre, 1994, 10). Diğer bir ifade ile massların analizde belirli bir profilin öneminin ölçüsü olduğu söylenebilir. Ayrıca satır massları ortalama sütun profiline, sütun massları da ortalama satır profiline karşılık gelir (Clausen, 1998).

Ki-kare uzaklıkları: Çok boyutlu bir uzayda noktalar arasındaki uzaklıklar öklid uzaklıkları olarak bilinir ve bu uzaklıklar Pythagorean formülü kullanılarak hesaplanabilir.

i ve i' noktaları arasındaki öklid uzaklığı = s(i, i') = $\sqrt{\sum_j \left(a_{ij} - a_{i'j}\right)^2}$ (1)

Burada a_{ij} i. satır ve j. sütun profilini, $a_{i'j}$ i'. satır ve j. sütun profilini göstermektedir. Bu formül satır ve sütun profillerine uygulandığında, farklı satırlar ve sütunlar arasındaki öklid uzaklıkları hesaplanabilir. Ancak uygulamalarda genellikle öklid uzaklığı yerine ki-kare uzaklığı kullanılmaktadır. Ki-kare uzaklığı bir tartılı öklid uzaklığı olup, tartılar ortalama profil elemanlarının tersidir.

i ve i' noktaları arasındaki ki-kare uzaklığı = d(i, i') = $\sqrt{\sum_j \frac{\left(a_{ij} - a_{i'j}\right)^2}{a_{+j}}}$ (2)

Burada tartı, profillerin kendi arasındaki ağırlıklandırmayı değil, uzaydaki boyutların farklı ağırlıklandırmasını gösterir. Böylece daha az sıklıkla ortaya çıkan yanıt seçeneklerinin profillerarası uzaklığa katkısının daha yüksek, daha fazla sıklıkla ortaya çıkan yanıt seçeneklerinin ise, daha az katkı sağladığı anlamı ortaya çıkar (Greenacre, 1994).

Burada dikkat edilmesi gereken husus, bu uzaklıkların farklı değişkenlerin kategorileri arasında değil, sadece aynı değişkenin farklı kategorileri arasında hesaplanabilmesidir. Uyum analizi bu ki-kare metriğine dayalı olup, ki-kare istatistiğinin ayrıştırılmasına yönelik bir teknik olarak da tanımlanabilir (Clausen, 1998).

Toplam inertia (Λ^2): Uyum analizinde varyans kavramı ki-kare uzaklıkları ile ilgilidir. Bunun için genellikle inertia terimi benimsenir ve inertia ile varyans terimleri eşanlamlı terimler olarak kullanılır. Toplam inertia, profil noktalarının merkez etrafındaki dağılımlarına ilişkin bir mesafe ölçümü olup, aşağıdaki formül yardımıyla hesaplanır (Clausen, 1998):

$$\Lambda^2 = \sum_i r_i d_i^2 \qquad (3)$$

Yukarıdaki eşitlikte d_i i. satır noktasının merkeze olan uzaklığı ve r_i i. satır noktanın mass (tartı) değerini gösterir.

Arzu edilen Λ^2'nin büyük bir değer almasıdır. Bu ise, satır noktalarının merkeze olan ki-kare uzaklıklarının (d_i) artması ile mümkündür. Ortalama profiller merkez olarak kabul edildiğinden, d_i' nin maximum olabilmesi ise, satır profillerinin ortalama satır profillerinden uzaklaşması ile gerçekleşebilir. Profiller ortalama profillere yakın değerler aldıkça, profil noktaları da o ölçüde merkeze yakın bir yerde konumlanır. Bu durum ise, ele alınan satır kategorisinin sütun kategorilerinden bağımsız (ilişkisiz) olduğunu gösterir. Gözelerdeki değişimin bir ölçüsü olan toplam inertia değeri sıfıra yaklaştıkça, satır profilleri de merkez etrafında toplanacak, toplam inertia değeri sıfırdan uzaklaştıkça satır profilleri de o ölçüde merkezden uzaklaşacaktır. Böylece satır profil noktalarının merkezden uzaklaşması ise satır kategorilerinin sütun kategorileri ile bağımlılığının arttığı anlamına gelecektir (Greenacre, 1994). Satır profilleri kutuplaştığında, yani profiller tamamen bir kategoride yoğunlaştığında profillerarası uzaklık da maximuma ulaşır. Bu da ilgili satır kategorisinin sütun kategorisi ile yüksek bir birliktelik gösterdiği anlamına gelir.

Toplam inertia kavramı aşağıda gösterileceği gibi, Pearson ki-kare istatistiği ile direk olarak ilişkilidir:

$$\chi^2 = \Lambda^2 n \qquad (4)$$

Ayrıca toplam inertia phi-kare cinsinden aşağıdaki gibi de yazılabilir (Clausen, 1998):

$$\varphi^2 = \frac{\chi^2}{n} = \Lambda^2 \tag{5}$$

Kategorik değişkenler arasındaki ilişkiyi belirlemede kullanılan en yaygın yöntemlerden birisi ki-kare analizidir. Ki-kare istatistiğinin bilinen formülüne göre her bir gözeye ait beklenen frekanslar, satır ve sütun bağımsızlığı altında, "(satır toplamı x sütun toplamı) / genel toplam"dır. Beklenen değerden her sapma toplam ki-kare değerine belli ölçüde katkı sağlayacaktır. Uyum analizine bu açıdan bakıldığında, beklenen değerlerden sapmaların az-boyutlu bir grafik ile gösterilebildiği ve toplam ki-karenin (ya da toplam inertia=ki-kare/n) ayrıştırılmasına yönelik bir yöntem olduğu görülebilir (http://www.statsoft.com/textbook/stcoran.html, 2004).

İstatistiksel analizlerde çok kullanılmasına rağmen, ki-kare değerini yorumlamak kolay değildir. Çünkü her bir gözedeki gözlemlenen frekans değerleri iki katına çıkartıldığında, nispi uzaklıkların değişmemesine rağmen ki-kare değeri artacaktır. Bu nedenle uyum analizinde gözelerdeki toplam değişkenliğin belirlenmesinde ki-kare istatistiği yerine, ki-kare istatistiğinin toplam gözlem sayısına bölümünü ifade eden toplam inertia ölçüsü kullanılır. Biraz önce belirtildiği gibi büyük bir inertia, satır ve sütunlar arasındaki birlikteliğin yüksek olduğunu, sıfıra yakın bir inertia değeri ise birlikteliğin olmadığı anlamına gelir (Benzerci, 2004).

Her ne kadar amaç toplam inertianın ayrıştırılması olsa da, toplam inertia ile ki-kare arasındaki ilişkiden dolayı uyum analizi, bir frekans tablosunun toplam ki-kare (ya da phi-kare) değerinin ayrıştırılması tekniği olarak da tanımlanır. Toplam inertia (ya da ki-kare) bir özdeğerler kümesi yardımıyla ayrıştırılır (Clausen, 1998). Satırlara karşı gelen frekanslar toplamı sütun toplamlarına, sütunlara karşı gelen frekanslar toplamı da satır toplamlarına eşit olması gerektiğinden her bir satırda (J-1) ve her bir sütunda da (I-1) bağımsız giriş vardır. Yani satır ve sütun toplamları bilindiğinde geri kalan girişlerin ne olduğu bulunabilir. Bundan dolayı, iki yönlü bir tablo için maximum özdeğerlerin sayısı ve aynı zamanda boyutların sayısı (J-1) ve (I-1)' in minimumuna eşittir (http://www.statsoft.com/textbook/stcoran.html, 2004). Bu özdeğerler boyutların nispi önemini ifade eder ya da her bir boyutun toplam inertianın ne kadarlık bir kısmını açıkladığını ifade eder. Veri matrisinin özdeğerleri hesaplandığında toplam inertianın en fazla ilk boyutla, daha sonra ikinci boyutla ve azalan miktarlarla diğer boyutlarla açıklandığı görülür (Clausen, 1998).

Her bir noktanın koordinatları ve aynı zamanda masslar kullanılarak her bir boyutun özdeğerleri hesaplanabilir (Clausen, 1998):

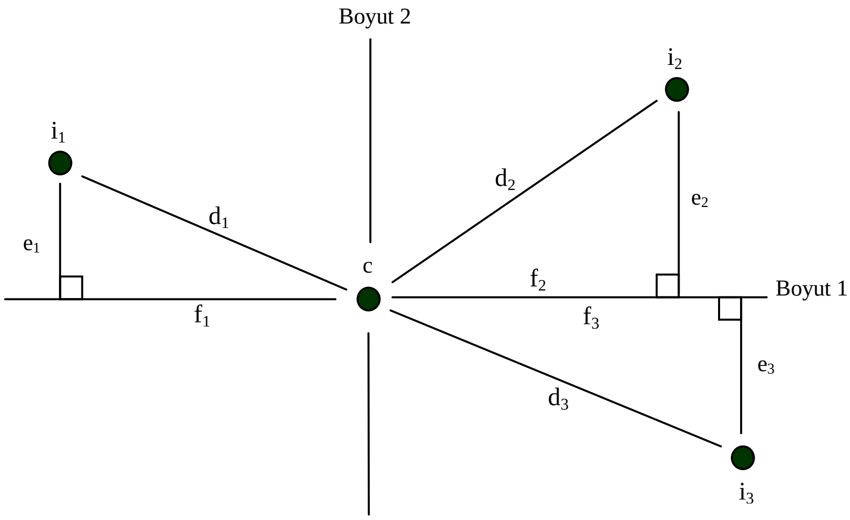
$$\lambda_k^2 = \sum_i r_i f_{ik}^2 \qquad (6)$$

Burada r_i i. satır noktasının mass değeri ve f_{ik} k. boyut üzerinde i. noktanın koordinat karesidir (bkz. Şekil 2). Özdeğerler her iki küme içinde aynıdır. Noktaların herhangi bir boyut için koordinat değeri ne kadar artarsa (ya da noktaların eksenlere uzaklığı ne kadar azalırsa), o boyutun toplam inertiayı açıklama miktarı da o ölçüde artacaktır.

Uyum Analizinde Geometrik Gösterimi Ve Haritaların Yorumu

Uyum analizinin temel esası, boyut azaltılarak noktaların bir alt uzayda (genellikle iki-boyutlu uzayda)gösterilmesidir. Bu alt uzay her bir noktanın mass değeri ile ağırlıklandırıldığı tartılı en küçük kareler yardımıyla uygun hale getirilir ve noktalar arasındaki uzaklıklar (d) ve böylece alt uzay ki-kare uzaklığına göre belirlenir (Greenacre, 2002).

Uyum analizinde haritaların oluşturulabilmesi için, ki-kare uzaklıkları (d) kullanılarak satır (yada sütun) noktaları arasındaki uzaklıklar ve farklı noktaların merkeze ya da eksenlerin orijinine uzaklıklarının hesaplanması gerekir. Bu durumda geriye kalan problem çok- boyutlu uzayda tüm noktalara en yakın olan eksenleri bularak noktaları iki-boyutlu uzayda gösterebilmektir. Yakınlık ölçüsü olarak, noktaların eksenlerden uzaklığının tartılı kareler toplamı kullanılabilir ($\sum re^2$), burada r tartı olarak tanımlanan satır masslarıdır. Böylece amaç $\sum re^2$ 'yi minimize etmektir (Clausen, 1998). Ayrıca boyutların satır veya sütun noktaları arasındaki uzaklık maximize kılınacak şekilde oluşturulduğuna da dikkat edilmelidir (http://www.statsoft.com/ textbook/stcoran.html, 2004). Daha önceden belirtildiği gibi, uyum analizi üç aşamadan oluşmaktadır, ilk aşama satır ve sütun profillerinin hesaplanması, ikinci aşama satır ve sütun profillerinin ayrı haritalarda gösterilmesidir. Aşağıda yer alan şekil satır noktalarının iki boyutlu uzaydaki geometrik gösterimine ilişkindir (Clausen, 1998):



Şekil 2: İki-Boyutlu Bir Uzayda Satır Noktalarının Geometrik Gösterimi

Burada i satır profil noktalarını, c merkezi, d i ve c (merkez) arasındaki uzaklığı, e i. noktanın eksene olan uzaklığını, f ise koordinatları göstermektedir.

Uyum analizinin üçüncü ve son aşaması satır ve sütun profillerini iki-boyutlu uzayda aynı harita üzerinde göstermektir. Her satır ve sütun profili bir matematiksel vektör olduğundan her profil noktası çok-boyutlu bir uzayda noktalar olarak gösterilebilecektir. Uygulamalarda boyut sayısı genelde üçden fazladır ve her profilin harita üzerindeki pozisyonunu belirlemek oldukça zordur. Ancak bazı profillerin ihmal edilerek noktaların iki-boyutlu bir uzayda gösterilmesi mümkündür Dolayısıyla uyum analizlerinde kullanılan grafikler, profil noktalarının gerçek pozisyonlarının bir yaklaşığıdır, çünkü onlar profil uzayında optimal görünen düzlemden elde edilmiştir (Greenacre, 1994).

Uyum analizinde genellikle ilk iki eksen (bazen ilk üç eksen) yoluyla yaratılan düzlemle çalışılır. Bazen de nadiren eksen 1 ve 3 ya da 2 ve 3 gibi eksenlerle de çalışılması sözkonusudur. Analitik sonuçların gösteriminde birkaç grafiksel yöntem vardır. Bunlar içerisinde en çok kullanılan ikisi asimetrik ve simetrik grafiklerdir. Asimetrik grafik, yanlış yorumlamalara engel olan en iyi grafiktir. Satır ve sütun noktaları iki farklı scala ile şekillenir, ancak uygulama sonuçları içinde nadiren görünür. Simetrik grafik en yaygın grafiktir, ancak aynı zamanda en tartışmalısıdır. Burada satır ve sütun noktaları aynı scalaya sahip grafik üzerinde gösterilir ve satır-sütun noktaları tüm grafiğin üzerinde yayılır. Bu grafiklerde dikkat edilmesi gereken husus, eğer satır ve sütun noktaları içindeki uzaklıklar ki-kare uzaklığından elde edilmişse, satır ve sütun noktaları arasındaki uzaklıktan söz edilememesidir (Benzecri, 2004).

Noktaların görünümüne ilişkin yorum, noktalar arasındaki ki-kare uzaklıklarına dayanır ve bu uzaklıklar her bir nokta kümesi için tanımlanır. Eğer satır (ya da sütun) noktaları birbirlerine yakın uzanıyorlarsa, bu iki noktanın profillerinin benzer olduğunu gösterir (Clausen, 1998). Yani iki-boyutlu çözümün yeterli bir uygunluk sağladığı varsayılırsa (ilk iki boyutun açıkladığı yüzde yeterli ise), satır noktaları sütunlara karşı benzer profiller gösterdiğinde birbirine yakın olacaktır. Aynı durum sütun noktaları içinde geçerlidir (Everitt and Dunn, 2001). Ayrıca bir noktanın profil değerleri ortalama profillere yakın ise, orijine yakın bir noktada konumlanacaktır.

Satır ve sütun noktaları arasındaki uzaklıklarla ilgilenildiğinde ilişki daha karmaşıktır, çünkü bu uzaklıklar ki-kare uzaklığı ile tanımlanamaz. Dolayısıyla yorum uzaklıklara göre değil, aşağıdaki şekle göre yapılabilir:

n_{ij} : i. satır ve j. sütuna ait gözenin gözlemlenen değeri

n'_{ij} : i. satır ve j. sütuna ait gözenin bağımsızlık altında hesaplanan beklenen değeri olmak üzere eğer $n_{ij} > n'_{ij}$ ise, i. satır ve j. sütun noktaları birbirine yakın, eğer $n_{ij} < n'_{ij}$ ise, i. satır ve j. sütun noktaları birbirine uzak olacaktır (Clausen, 1998). Ayrıca $n_{ij} > n'_{ij}$ ise, i. satır ve j. sütun arasında pozitif bir birliktelik, $n_{ij} < n'_{ij}$ ise, i. satır ve j. sütun arasında negatif bir birlikteliğin olduğu ve n_{ij} değeri n'_{ij} değerine yakın olduğunda bir birlikteliğin olmadığı anlamına gelecektir (Everitt and Dunn, 2001). Dolayısıyla herhangi bir satır noktasının herhangi bir sütun noktasına yakın pozisyon alması durumunda aralarındaki uzaklık yorumlanamayacağı gibi, iki noktanın benzer olduğu da söylenemez. Ancak ilgili satır noktasının ilgili sütun noktasına yönelik oransal frekans değerinin diğer sütun noktalarına göre daha yüksek olduğu, yani gözlenme sıklığının daha yüksek olduğu anlamına gelir (http://www.statsoft.com/textbook /stcoran.html, 2004).

Uyum analizinde elde edilen sonuçların grafikle gösterimi yorumların daha kolay yapılmasını ve her bir değişkene ait kategoriler arasındaki ilişkilerin daha anlaşılır olmasını sağlamaktadır. Her bir noktanın orijinden uzaklığı o noktanın yani kategorinin önemini ifade etmektedir. Grafik üzerinde orijinden ilgilenilen noktaya bir doğru çizildiğinde, diğer noktalara da orijinden başka noktalar çizildiğinde ilgilenilen noktaya ait doğru ile diğer doğrular arasındaki açı, ilgilenilen nokta ile diğer noktalar arasındaki ilişkiyi göstermektedir. Açının küçüklüğü ilişkinin büyüklüğünü, açının büyüklüğü ise ilişkinin küçüklüğünü ifade etmektedir (Palmer, 1993).

Uygulamalarda genellikle simetrik grafikler kullanılmakta ve bu grafiklerle beraber toplam inertia değeri ile birlikte ele alınan boyutların (genelde ilk iki boyut) toplam inertiayı açıklama yüzdeleri verilmektedir. Örneğin %96 gibi bir açıklama yüzdesi toplam inertianın %96'sının ele alınan iki boyut ile açıklandığını geri kalan %4'ünün ise diğer boyutlarla açıklandığını gösterir. Bu iki yüzde sırasıyla toplam inertianın düzlem ve artık ile açıklanan yüzdesini verir (Greenacre, 1994).

REGRESYON VE UYUM ANALİZİNİN BENZERLİKLERİNİN GRAFİKSEL BOYUTTA İRDELENMESİ

Regresyon ve uyum analizlerinin ortak özelliği, değişkenler arasındaki ilişkileri inceleyen yöntemler olmasıdır. Regresyon analizi daha çok nicel verilerin analizinde kullanılmakla birlikte uyum analizi daha çok kategorik verilerin analizinde kullanılmaktadır. Regresyon analizi değişkenler arasındaki ilişkiyi matematiksel bir denklem ile ifade ederken, uyum analizi iki veya daha fazla değişkenin kategorileri arasındaki ilişkileri ya da bir değişkenin kendi kategorileri arasındaki ilişkileri grafiksel bir formda ifade etmektedir.

Her iki analizinde esası toplam değişkenliği bileşenlerine ayırmaktır. Regresyon analizinde bağımlı değişken Y'deki toplam değişimin, yani Y'nin kendi ortalamasından olan *uzaklıklarının* kareli toplamının $[\sum y^2 = \sum (Y-\bar{Y})^2]$ bileşenlerine ayrıştırılması, uyum analizinde ise, gözelerdeki toplam değişimin ya da başka bir ifadeyle profil noktalarının merkezden olan *uzaklıklarının* ölçümünü ifade eden toplam inertianın (Λ^2) bileşenlerine ayrılması sözkonusudur. Toplam değişkenliğin bileşenlerine ayrıştırılmasına yönelik olarak her iki analizde de temel esas, artık kareler toplamını minimize ederek, düzlem ile açıklanan değişkenliği maksimize etmektir. Regresyon analizinde artık kareler toplamı ($\sum e^2$) regresyon ile açıklanamayan, hatalara bağlı değişkenliği gösterirken, uyum analizinde artık kareler toplamı ($\sum r_i e_i^2$) haritada gösterilmeyen boyutların açıkladığı değişkenliği gösterir. Her iki yöntemde de bu değer ne kadar küçük olursa tahminlere ilişkin yorumlar da o ölçüde güvenilir olur.

Bu doğrultuda çalışmanın amacı toplam değişkenliğin birer göstergeleri olan $\sum y^2$ (Y'deki toplam değişkenlik) ve Λ^2 (toplam inertia) ifadelerinin bileşenlerine ayrıştırılması işlemini grafiksel boyutta göstermektir.

Çoklu Regresyon Analizinde Toplam Değişkenliğin Bileşenlerine Ayrıştırılması

Regresyon analizinin temel amacı bağımlı değişken ile bağımsız değişken(ler) arasındaki ilişkiyi matematiksel bir denklem ile açıklamaktır. Genel olarak nicel değişkenlerin analizinde kullanılan regresyon analizi, basit ve çoklu olmak üzere ikiye ayrılmaktadır. Bir bağımlı değişken ile birden fazla bağımsız değişken arasındaki ilişkiyi inceleyen çoklu regresyon analizi bir bağımsız değişken içeren basit regresyon analizinin doğal bir uzantısıdır. Ancak çoklu regresyon analizi yapmak basit regresyon analizi yapmaktan birçok açıdan daha zordur. Özellikle ikiden fazla bağımsız değişken varsa direk olarak ya verilere ya da modele uyan üçten daha fazla boyutlu bir grafik çizmek mümkün değildir (Kleinbaum vd, 1988). Bu nedenle bağımlı değişkendeki toplam değişimi açıklamaya yönelik grafiksel açıklamalar sadece iki bağımsız değişkeni içeren regresyon modeline göre yapılacaktır.

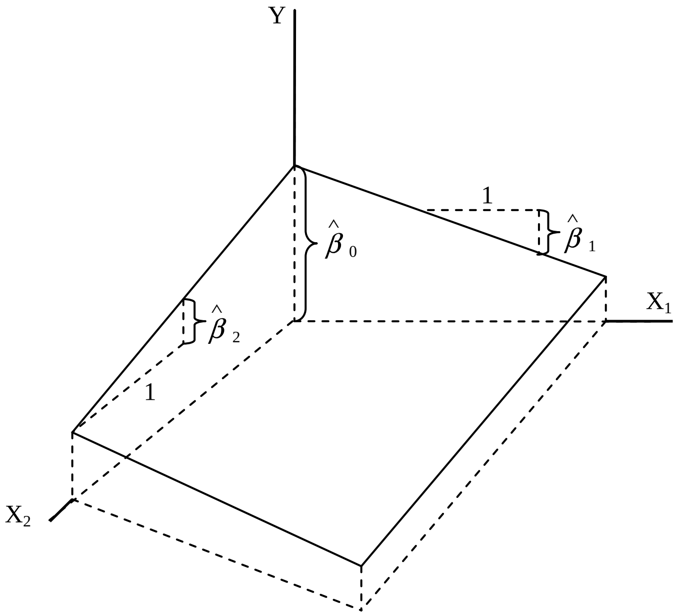
Çoklu regresyon modelinin genel formu;

$$Y = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots\dots\dots + \beta_k X_{ik} + \varepsilon_i$$

şeklinde tanımlanır ve Y bağımlı değişkeni, X'ler bağımsız değişkenleri ve ε istatistiksel hata terimini gösterir. Ayrıca β_0 sabit (ya da regresyon sabiti) ve $\beta_1, \beta_2, \dots, \beta_k$ regresyon eğimleri (ya da kısmi regresyon katsayıları) olarak ifade edilir. Buradaki β_k katsayıları, $X_{i1}, X_{i2}, \dots, X_{i,k-1}$ sabit tutulduğunda $E(Y_i)$ üzerinde X_{ik}'nın etkisini ölçer (Kmenta, 1971).

Bir bağımsız değişkenin (Y, X) varlığında bir doğru oluşturan regresyon denklemi, iki bağımsız değişkenin varlığında üç boyutlu (X_1, X_2, Y) uzayda iki boyutlu bir düzlemin denklemini verir (Weisberg, 1980) (Bkz. Şekil 3).

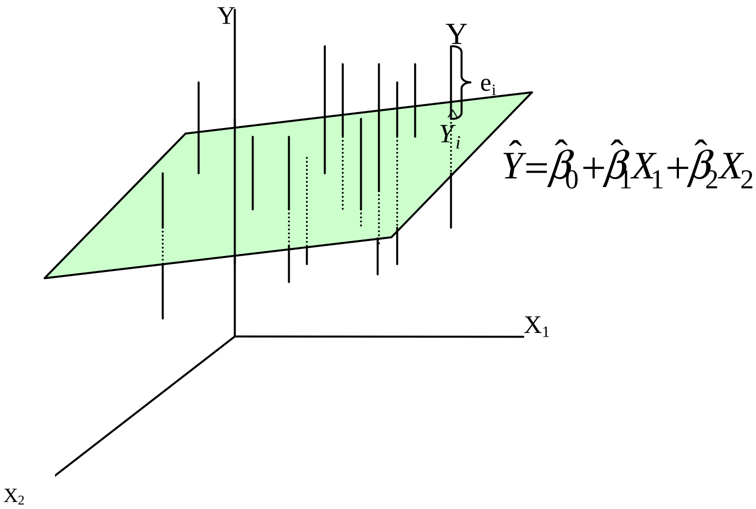
$\hat{\beta}_0$ bu düzlemin Y eksenini kestiği nokta olup $\hat{\beta}_1$ ve $\hat{\beta}_2$ eğim katsayılarını gösterir. Problem, (X_{11} , X_{21}, Y_1), (X_{12} , X_{22}, Y_2),………, (X_{1n} , X_{2n}, Y_n) noktalarının saçılımına en iyi uyan üç boyutlu uzayda bir yüzey bulmaktır, burada (X_{1i} , X_{2i}, Y_i) örneklemindeki i. birime ait X_1, X_2 ve Y değerlerini gösterir. Bu durumda regresyon denklemi X_1 ve X_2'nin çeşitli kombinasyon değerlerine karşılık gelen Y'nin ortalama değerleri tarafından tanımlanan yüzeyidir. Diğer bir ifadeyle regresyon denklemi, X_1 ve X_2'nin her bir farklı değer çiftine karşılık gelen $\mu_{y|X_1, X_2}$ ortalamalı ve $\sigma^2{}_{Y|X_1,X_2}$ varyanslı Y değerlerinin bir dağılımıdır.



Şekil 3: İki Açıklayıcı Değişkenli Bir Doğrusal Regresyon Düzlemi

İki boyutlu uzayda en basit eğri bir düz doğru iken üç boyutlu uzayda istatistiksel model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$ formuna sahipken en basit yüzey bir düzlemdir. Bu nedenle basit regresyonda en iyi doğrunun bulunması bir bağımsız değişken içeren modellerin ilk aşaması iken, en iyi uyan düzlemin bulunması, üç-boyutlu uzayda en iyi uyan yüzeye karar vermenin (sık sık) ilk aşamasıdır (Kleinbaum vd, 1988). Üç boyutlu bir uzayda verilere en iyi uyan grafiksel gösterim Şekil 4'de verilmiştir.

Şekil 4'de görüldüğü gibi, Y_i değerleri bu düzlem çerçevesinde kümelenmiştir ve Y_i değerlerinin düzlem üzerindeki izdüşümleri ise kestirim değerlerini $(\hat{Y}_i)$ ve gözlem değerlerinin kestirim değerlerinden sapması da örneklem hata terimlerini (ya da artıkları, e) vermektedir (Erar, 1983).



Şekil 4: Üç-Boyutlu Uzayda Eniyi-Uyan Düzlem

Bu durumda amaç gözlemlere en iyi uyan düzlemi bulmaktır, bu amaca yönelik olarak EKY'de dikkate alınan ölçüt basit regresyon analizinde olduğu gibi, gözlemlenen Y_i değerleri ile bunlara karşılık gelen kestirim değerleri ($\hat{Y}$) arasındaki uzaklığın (e_i) kareler toplamını minimize etmektir. Diğer bir ifadeyle, β_0, β_1 ve β_2'nin en küçük kareler kestirimlerinin bulunması için

$$\sum e_i^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i})^2 \qquad (7)$$

miktarının minimize edilmesidir (Kleinbaum vd, 1988).

Basit regresyonda olduğu gibi Y'deki toplam değişkenlik iki bileşenin toplamından oluşmaktadır (Kmenta, 1971):

$$\sum_i (Y_i - \bar{Y})^2 = \sum_i \left[\left(\hat{Y}_i + e_i \right) - \bar{Y} \right]^2$$

$$= \sum_i \left(\hat{Y}_i - \bar{Y} \right)^2 + \sum_i e_i^2 + 2 \sum_i (\hat{Y}_i - \bar{Y}) e_i$$

($2 \sum_i \left(\hat{Y}_i - \bar{Y} \right) e_i = 0$ olduğundan)

$$\sum_i \left(Y_i - \bar{Y} \right)^2 = \sum_i \left(\hat{Y}_i - \bar{Y} \right)^2 + \sum_i e_i^2 \qquad (8)$$

Y'deki toplam değişkenlik (TD) = Regresyon Kareler Toplamı (RKT)+ Artık kareler Toplamı (AKT)

Amaç AKT'nı minimize edecek ya da RKT'nı maksimize edecek regresyon denkleminin katsayılarını bulmaktır. Dolayısıyla RKT oluşturulan düzlemin noktalara olan uyumunun bir göstergesi olarak değerlendirilir. Bu amaç doğrultusunda hesaplanacak belirlilik katsayısı uyum iyiliğinin bir göstergesi olarak değerlendirilir. Sıfır ile bir arasında değişen değerler alan belirlilik katsayısı Y'deki toplam değişimin X'ler ile ya da düzlem ile açıklanan yüzdesini gösterir ve aşağıdaki gibi tanımlanır:

$$R^2 = \frac{RKT}{TD} = 1 - \frac{AKT}{TD} \tag{9}$$

R^2=1 olması durumunda mükemmel bir uyumun, R^2'nin sıfıra yakın değerler alması durumunda da uyumun çok iyi olmadığı belirtilir (Mirer, 1983).

Uyum Analizinde Toplam Değişkenliğin Bileşenlerine Ayrıştırılması

Uyum analizinde verilerdeki toplam değişkenliğin (varyansın) ölçüsü toplam inertia olarak tanımlanır. 1.1'de açıklandığı gibi toplam inertia (Λ^2) aşağıdaki formül yardımıyla hesaplanmakta olup, amaç bu değeri maksimum kılacak düzlemi bulmaktır:

$$\Lambda^2 = \sum_i r_i d_i^2$$

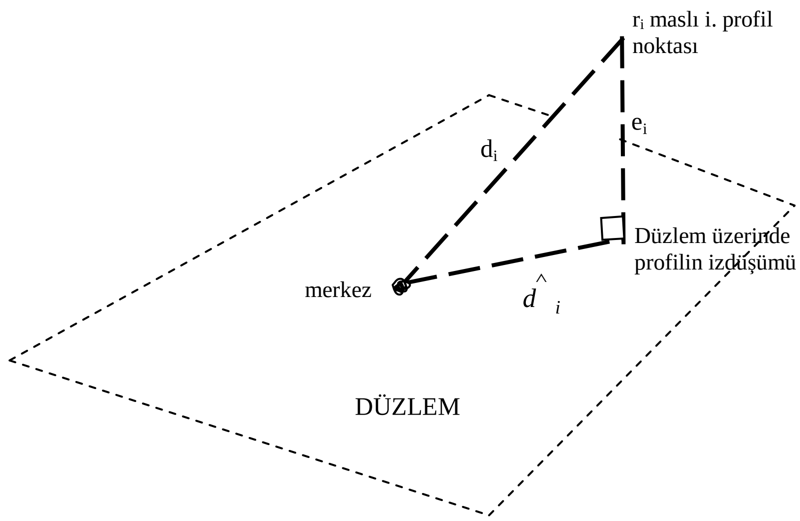
Burada r_i satır massı ve d_i i. profilin ortalama profile (merkez) olan ki-kare uzaklığının karesini göstermektir.

Verilen herhangi bir profil noktasının düzlemden uzaklığı, profil ve düzlem arasındaki en küçük ki-kare uzaklığı ile hesaplanabilir. Profile en yakın düzlem noktası profilin izdüşümü olarak ifade edilir. Profilin izdüşümünden olan uzaklığı e_i ve merkezden izdüşüme olan düzlem uzaklığı $\hat{d}_i$ ile tanımlandığında Şekil 5 elde edilir.

Merkez, izdüşüm ve profil dik açılı bir üçgen formundadır ve bu üç unsura Pythagoras teoremi uygulanırsa, profil noktasının merkeze olan uzaklığı

$$d_i^2 = \hat{d}_i^2 + e_i^2 \tag{10}$$

formunda yazılabilir.



Şekil 5: Düzlem Üzerinde Bir Profil Noktasının Genel Görünümü

Böylece regresyon analizinde olduğu gibi toplam inertia ($\Lambda^2 = \sum_i r_i d_i^2$) iki bileşene ayrıştırılabilir:

$$\sum_i r_i d_i^2 = \sum_i r_i \hat{d}_i^2 + \sum_i r_i e_i^2 \tag{11}$$

Toplam inertia = Düzlem inertiası + Artık inertiası

Profil noktalarının düzleme olan yakınlığı, noktaların düzleme olan uzaklıklarının (yani artık inertiası)tartılı kareli uzaklığıyla ölçülür ve analiz amacı bu miktarı minimize edecek düzlemi bulmaktır. Bu eşitlik aynı zamanda artık inertiasının minimize edilmesinin, düzlem inertiasının maximize edilmesine eşit olduğunu göstermektedir. Bu amaç için genellikle düşük boyutlu bir altuzay seçilir. Artık inertiası iki-boyutlu bir formatta profillerin ihmal edilmesiyle kaybedilir ve analiz bu kaybın minimum olduğu bir düzlem bulur. Ya da eşdeğer olarak bulunacak düzlem, iki-boyutlu bir gösterimde mümkün olan maximum inertiayı elinde bulunduran düzlemdir (Grenacre, 1994).

SONUÇ VE ÖNERİLER

Basit uyum analizinde amaç iki boyutlu bir kontenjans tablosundan yararlanarak, her satır ve sütunu bir nokta olarak göstermek suretiyle tablonun "haritasını" çizmektir. Bu yaklaşım ana bileşenler analizine oldukça benzerdir. Çünkü ilk olarak tablonun toplam varyansının bir ölçüsü (toplam inertia) elde edilir ve daha sonra bu toplam varyans optimal bir şekilde “ana eksenlere” ayrıştırılır (Grenacre, 2002).

Daha önce uyum analizinin sıra ve sütun sayısının çok fazla olduğu ve gözlem sayısının yetersiz kaldığı durumlarda uygulanan önemli bir yöntem olduğu belirtilmiştir. Dolayısıyla sıra veya sütun sayısının artması, satır ve sütun noktalarının çok boyutlu bir uzayda gösterilmesi anlamına gelmektedir. Ancak uyum analizi bazı boyutları ihmal ederek iki-boyutlu bir kontenjans tablosunun satır ve sütunlarını, tablodaki birliktelikleriyle tutarlı pozisyonlarını az-boyutlu bir uzayda (genellikle iki-boyutlu) göstermeyi sağlamaktadır. Böylelikle uyum analizi çok karmaşık tabloların haritalar yardımıyla kolay bir şekilde yorumlanması kolaylığını sağlamaktadır. Harita çiziminde amaç, genellikle toplam varyansın (toplam inertianın) iki ana eksen tarafından büyük bir ölçüde açıklanmasıdır. Uygulamada arzu edilen toplam inertianın yüksek olması (bu değerin yüksek olması ele alınan iki değişken arasındaki ilişkinin yüksek olması anlamına gelir) ve çizilen iki boyutlu haritanın toplam inertianın büyük bir bölümünü açıklamasıdır.

Verilerdeki toplam değişimin bir göstergesi olan toplam inertianın iki bileşenden oluştuğu birinci bileşenin düzlem inertiası ikinci bileşenin ise artık inertiası olduğu ve bu özelliğinin çoklu regresyon analizine oldukça benzer olduğu teorik ve grafiksel olarak gösterilmiştir. Genellikle nicel değişkenlerin analizinde kullanılan regresyon analizi, yöntem olarak uyum analizinden çok farklı olsa da her ikisinde de amaç değişkenler arasındaki ilişkiyi açıklamak ve toplam değişkenliği bileşenlerine ayırarak çözüm geliştirmektir. Aralarındaki fark, uyum analizinde toplam değişkenlik (toplam inertia) kavramı ile bir kontenjans tablosunun gözlemlenen frekansı ile bağımsızlık altında hesaplanan beklenen frekansları arasındaki farklılık ifade edilirken, regresyon analizinde toplam değişkenlik kavramı ile bağımlı değişken Y’nin gözlem değerlerinin ortalamadan ($\bar{Y}$) olan farklılıkları ifade edilmektedir.

Uyum analizinde olduğu gibi regresyon analizinde de toplam değişkenliğin iki bileşene ayrıldığı ve amacın açıklanamayan değişimin bir ölçüsü olan artık kareler toplamını ($\sum e^2$) minimize etmek olduğu teorik ve grafiksel olarak gösterilmiştir.

Ayrıca çalışmada uygulamalarda sıklıkla kullanılan iki boyutlu simetrik grafiklerde noktaların pozisyonlarına göre nasıl yorumlanacağı açıklanmış olup, bundan sonra uyum analizi kullanılarak yapılacak olan uygulamaya yönelik çalışmalarda yorum kolaylığının sağlanması hedeflenmiştir.

KAYNAKLAR

Benzecri J.P., (2004). “Correspondence Analysis”, http://www.micheloud.com/FXM/COR/e/into.htm, (06.10.2004).

Clausen S.E., (1998). *Applied Correspondence Analysis-An Introduction*, Sage Publication, ISBN:0-7619-1115-4, USA.

______, (2004). “Correspondence Analysis”, http://www.statsoft.com/textbook/ stcoran html, (30.09.2004).

Erar A., (1983). *Regresyon Çözümlemesi*, Hacettepe Üniversitesi İstatistik Bölümü Ders Notları, Ankara.

Everitt B.S.and Dunn G., (2001). *Applied Multivariate Data Analysis*, Oxford University Press Inc., New York.

Greenacre M. and Blasıus J., (1994). *Correspondence Analysis in the Social Sciences*, Academic Press Limited, ISBN: 0-12-104570-6, USA.

Greenacre M., (2002). “Correspondence Analysis of the Spanish National Health Survey”, *Gac Sanit*, Vol:16, No:2, Mar.-Apr., Barcelona.

Kleınbaum D.C., Kupper L.C. and Muller K.E., (1988). *Applied Regression Analysis and Other Multivariable Methods*, PWS-KENT Publishing Co. ISBN:0-87150-123-6, USA.

Kmenta J., (1971) *Elements of Econometrics*, Macmillan Publishing Company, ISBN 0-02-365070-2, New York.

Lee B.L., (2006). “Correspondence Analysis”, http://www.uv.es/prodat/vista, (10.11.2006).

Mirer T.W., (1988). *Economic Statistics and Econometrics*, Macmillan Publishing Company, ISBN:0-02-381821-2, New York.

Özdamar. K. , (2002). *Paket Programlar İle İstatistiksel Veri Analizi-2,* Kaan Kitabevi, Eskişehir.

Palmer M.W., (1993). “Putting Things ın Even Better Order the Advantages of Conanical Correspondence Analysis”, *Ecology*, 74 (Cangür Ş., Sığırlı D., Ediz B., Ercan İ. ve Kan İ., (2005). “Türkiye’deki Özürlü Grupların Yapısının Çoklu Uyum Analizi İle İncelenmesi”, *Uludağ Üniversitesi Tıp Fakültesi Dergisi*, 31 (3) 153-157’ den alınmıştır)

Seyfullahoğulları A., (2003). “Çapraz Tabloların Analizi ve Ticari Malların değerlendirilmesiyle İlgili Bir Uygulama”, *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, Sayı:4, Aralık.

Weısberg S., (1980). *Applied Linear Regression*, John Wiley & Sons, ISBN:0-471-04419-9, Canada.