T.C.
MİMAR SİNAN ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
İSTATİSTİK ANABİLİM DALI
DOKTORA PROGRAMI

PARAMETRİK OLMAYAN ÇOK DEĞİŞKENLİ ANALİZ TEKNİĞİ : HOMOJENLEŞTİRME ANALİZİ

DOKTORA TEZİ

Hazırlayan : Meltem G. (ÇATANA) TUNA

Tez Yöneticisi : Doç. Dr. Gülay (BAŞARIR) KIROĞLU

İstanbul, 1999

ÖZET

PARAMETRİK OLMAYAN ÇOK DEĞİŞKENLİ ANALİZ TEKNİĞİ: HOMOJENLEŞTİRME ANALİZİ

TUNA, Meltem Gülsün
Doktora Tezi, İstatistik Anabilim Dalı
Tez Yöneticisi : Doç. Dr. Gülay (Başarır) Kıroğlu
Mayıs, 1999, 100 sayfa

Çok değişkenli parametrik olmayan teknikler özellikle son yıllarda dünya literatüründe önemli bir yer tutmaya başlamıştır. Gelişen bilgisayar teknolojisi de parametrik yöntemlere kıyasla daha gerçekçi olan bu tür tekniklerin uygulanmasını kolaylaştırmaktadır. Özellikle grafiksel gösterim ağırlıklı teknikler büyük ilgi görmektedir.

Bu çalışmanın amacı kategorik verilerin analizi için geliştirilmiş olan grafiksel gösterime dayalı homojenleştirme analizini (Homogeneity Analysis) incelemektir. Homojenleştirme analizinde nesne ve kategoriler için kayıp fonksiyonu olarak tanımlanan belirli kriterleri en iyileyen puanlar belirlenir daha sonra bu puanlar yardımıyla grafikler oluşturulur. Çalışmada kayıp fonksiyonları ve bunların çözümlerinde kullanılan algoritmalar incelenmiş ve elde edilen grafiklerin yorumları sunulmuştur. Uygunluk analizi (Correspondence Analysis) de benzer bir teknik olmakla beraber sadece iki değişken için uygulanmaktadır. Özetle uygunluk analizi iki değişkenin oluşturduğu iki boyutlu çapraz tablolar için kullanılmaktadır. Bu nedenle homojenleştirme analizi çoklu uygunluk analizi olarak da adlandırılabilmektedir. Çalışmada uygunluk analizi incelenmiş, grafiklerin oluşturulabilmesi için gerekli olan koordinat hesaplarına değinilmiştir. Her iki analiz sonucunda elde edilen grafiklerin kategoriler arasındaki ilişkiyi belirlemek için kullanımının yanı sıra kategorileri sınıflandırmak için kullanımı üzerinde de durulmuştur. Grafiklerin incelenmesini kolaylaştırmak amacıyla lojistik regresyon analizinin kullanımı önerilmiştir. Analizlerden elde edilen koordinatlara uygulanan lojistik analizi yardımıyla ulaşılan fonksiyonların ayrımsama için kullanılabileceği gösterilmiştir.

Homojenleştirme analizi ile ilişkili olduğu düşünülen çok değişkenli analiz tekniklerinden çok boyutlu ölçekleme, doğrusal olmayan temel bileşenler analizi ve doğrusal olmayan kanonik korelasyona da değinilmiştir. Bu tekniklerle elde edilebilecek grafiklere örnekler verilmiş ve teknikler arası farklılıklar açıklanmıştır.

Uygulamada bir otomotiv şirketinin bayileri üzerinde gerçekleştirdiği araştırma sonuçları kullanılmış ve SPSS paket programının 8.0 sunumundan faydalanılmıştır. Elde edilen grafiklerin çeşitli açılardan yorumları sunulmuştur. Yararlı olabileceği düşünülen komut dizinlerine ve araştırmada kullanılan soru formuna yer verilmiştir.

Anahtar Kelimeler : Kategorik Veri, Homojenleştirme Analizi, Uygunluk Analizi, SPSS CATEGORİES, HOMALS, PRINCALS, OVERALS.

Bilim Dalı Sayısal Kodu: 406. 02. 01

SUMMARY

A NON-PARAMETRIC MULTIVARIATE ANALYSIS TECHNIQUE: HOMOGENEITY ANALYSIS

TUNA, Meltem Gülsün

Ph.D. in Statistics

Supervisor: Assoc. Prof. Dr. Gülay (Başarır) KIROĞLU

May, 1999, 100 pages

Non-parametric multivariate techniques have an important place in literature in recent years. These techniques are more realistic than parametric ones and are easier to apply with the developing computer technology. Especially graphical approaches of data analysis have been calling attention.

The main objective of this study is to analyse the homogeneity analysis which is a popular graphical method for the categorical data. Homogeneity analysis minimize the loss function and determines the objects scores and categories quantification's. Graphs are obtained with the help of these scores and quantification's. Since correspondence analysis is also analysed in this study is a very similar technique which is used for two variables summarised by a contingency table, homogeneity analysis is also known as multiple correspondence analysis. Both methods produce plots in which both objects and categories are presented as points in a low – dimensional space. Plots obtained form theses two techniques can be used to explain the similarities between categories as well as to classify the similar categories. In order to simplify the analysis of the graphs obtained from both of the techniques the logistic regression is proposed. It is shown that classification will be easier by using the line obtained from the logistic regression analysis applied to the coordinates of the categories. A Other multivariate techniques related to homogeneity analysis are also explained and comparisons of the methods are given.

The results of a research made for the services of an automotive and SPSS statistical software is used for applications. Graphs obtained from the analysis is also interpreted from different aspects. Command syntax that will be necessary is given at the end.

Key words: Categorical Data, Homogeneity Analysis, Correspondence Analysis, SPSS,CATEGORIES, HOMALS, PRINCALS, OVERALS.

Science Code: 406.02.01

İÇİNDEKİLER

TABLO VE ŞEKİLLER DİZİNİ

SAYFA

GÖSTERİM

Birden fazla değişkenin ve çok sayıda birimin incelendiği çok değişkenli analiz tekniklerinde elde edilen verilerin gösterimlerinde matris ve vektörlerden faydalanılmaktadır. Bu çalışmada orjinal veri matrisi **H** ile gösterilmiş ve aksi belirtilmediği sürece koyu büyük harfler matris gösteriminde kullanılmıştır. **H** matrisi değişken sayısı (m) kadar sütuna ve nesne sayısı (n) kadar satıra sahiptir. Analiz yöntemlerinde kullanılacak olan matrislerde yapılması gereken düzenlemeler ve matrisin uyması gereken koşullar yeri geldiğinde açıklanmıştır.

Değişkenler h_j ile gösterilmiştir, yani, j. değişken h_j'dir ve **H** matrisindeki j. değişkene ait sütununu belirtmek için de $\mathbf{h}_j$ ifadesi kullanılmıştır. Değişkenler vektörü ise **h** ile gösterilmiştir.

Doğrusal olmayan çok değişkenli analiz tekniklerinden bir çoğunda gösterge matrisine ihtiyaç duyulmaktadır. Kategorik değişkenlerin oluşturduğu bu matris **G** ile gösterilmiştir. **G** gösterge matrisi değişken sayısı kadar alt matristen oluşmaktadır. Adı geçen bu matrisler $\mathbf{G}_j$ ile ifade edilirler. Her $\mathbf{G}_j$ alt matrisi h_j değişkeninin kategori sayısı (k_j) kadar sütuna ve nesne sayısı kadar satıra sahiptir. Her sütunun sadece bir elemanı bire eşitken diğer elemanlarının değeri sıfır ise **G** matrisi tam gösterge matrisi olarak adlandırılır ve çalışmamızda kullanılan matrisler hep bu tür matrislerdir. Başka bir ifade ile her birim, her değişkende sadece bir kategoriye düşmektedir ve bir birimin h_j değişkeninde düştüğü kategori dışındaki kategorilerdeki değeri sıfırdır.

Teknikleri açıklamakta kullanılan diğer matrisler ise **C** ve **D** matrisleridir. İkili marjinaller matrisi olarak adlandırılan **C** matrisi **C = G'G** matris çarpımından elde edilir. Tekli marjinaller matrisi olarak adlandırılan **D** matrisi ise **C** matrisinin köşegen elemanlarını köşegeninde barındıran ve diğer bütün elemanları sıfır olan bir matristir.

Kayıp fonksiyonlarında geçen SSQ (X) ifadesi ise parantez içerisindeki matris veya vektörün elemanlarının kareleri toplamıdır.

BÖLÜM I

GİRİŞ

Tarihi 1700'lü yıllara dayanan, verinin toplanma, organizasyon, yorum ve rapor halinde sunum bilimi olarak da bilinen istatistik, sadece psikoloji ve sosyoloji gibi sosyal bilimlerde değil tıbbi çalışmalarda da geniş uygulama alanı bulmaktadır. Uygulama alanlarındaki gelişmelerin yanı sıra ilerleyen bilgisayar teknolojisi sayesinde çok daha ayrıntılı veri derlenmesi mümkün hale gelmiştir. Derlenen bu çok büyük ve ayrıntılı verinin oluşturduğu matrislerinin incelenmesi ve yorumlanması hızla gelişmekte olan çokdeğişkenli analiz yöntemleri ile mümkündür.

Uygulamada ve akademik alanda istatistikçiler tarafından çok sıkça kullanılan çoklu regresyon analizi, diskriminant analizi, temel bileşenler analizi, faktör analizi, çoklu varyans analizi gibi tekniklerin yanı sıra kanonik korelasyon analizi, kümeleme analizi, çok boyutlu ölçekleme, konjoint analizi, homojenlik ve uygunluk analizi gibi nispeten daha az bilinen, daha çok araştırmacılar tarafından kullanılan fakat etkisi, kullanılabilirliği ve popülaritesi gün geçtikçe artan bu tekniklerin hepsi kısaca çok değişkenli analiz teknikleri adı altında toplanmaktadır.

Birçok bilim dalında, ortaya çıkan sonucun hangi etkenlerden ne ölçüde etkilendiğinin belirlenmesi büyük önem taşımaktadır. Bu gibi durumlarda öncelikle olay tanımlanır ve olayın gerçekleşmesine yol açtığı düşünülen etkenler belirlenmeye çalışılır. Sonuç bağımlı değişken, etkenler ise bağımsız (açıklayıcı) değişkenler olarak adlandırılır. Böylece problem bir bağımlı ve bir veya birden fazla bağımsız değişken arasındaki neden sonuç ilişkisini ortaya çıkarma problemi haline gelir. Regresyon analizi bu tür problemlerin çözümünde kullanılan yöntemlerin başında gelir.

Çoklu regresyon analizi, ilgilenilen bir bağımlı değişken ile birden fazla bağımsız değişken arasındaki ilişkinin incelenmesini sağlayan analiz tekniğidir. Bu teknikte amaç, değerleri bilinen bağımsız değişkenleri kullanarak öğrenilmek istenen bağımlı değişkene ulaşılmasıdır. Bağımlı ve bağımsız değişkenlerin metrik yani ölçülebilir olması gereken bu teknikte hangi değişkenin bağımlı hangi değişkenin bağımsız olarak alınacağına dikkat edilmelidir. Çoklu regresyon analizi ayrıca bağımlı ve bağımsız değişkenler arasındaki ilişkinin derecesinin belirlenmesinde de kullanılmaktadır.

Bağımlı değişkenin kesikli olması durumunda genel doğrusal regresyon modeli için yapılan "bağımlı değişkenin sürekliliği varsayımı" bozulmaktadır. Bağımlı değişkenin 0,1 gibi iki düzeyli olduğu durumlarda değişkenler arası ilişkiyi incelemek için regresyon analizinin kullanılması genellikle anlamsız olmaktadır. Böyle durumlarda lojistik regresyon analizi önerilmektedir. Lojistik regresyon analizinde bağımlı değişken logit dönüşümü $\{\ln(p/1-p)\}$ kullanılmasıyla $(-\infty , +\infty)$ aralığına yayılarak kesikli veya sürekli bağımsız değişkenler ile iki düzeyli bağımlı değişken arasındaki ilişkiler tahmin edilebilir. Kestirim için iteratif yöntemler kullanan bu tekniğin sonuca ulaşamaması durumunda başlangıç verisi olarak kullanılan diskriminant analizi sonuçları esas alınabilmektedir. Anılan yöntem ikiden fazla düzeyi bulunan bağımlı değişkenler için de kullanılabilir. Bu durumda genel olarak bağımlı değişkenin kategori sayısının bir eksiği kadar lojistik model oluşturulur. Hosmer ve Lemeshov 1989'da lojistik regresyon analizinin kullanım amaçlarının modelde önemli olan değişkenleri teşhis etmekle beraber veriye yeni alınacak bir gözlemin değerlerine bakılarak yeni gözlemin var olan gruplardan hangisine atanacağını da belirlemek olduğunu belirtmişlerdir (Başarır, 1990, s. 28 - 30).

Metrik değişkenlerin varlığı durumunda veri yapısındaki gruplanma bilindiğinde iki veya daha çok sınıfta tanımlanabilecek hipotetik bağımlı değişken oluşturularak çoklu diskriminant analizi kullanılır. Söz konusu yöntem bütün örneğin bağımlı değişkenin bilinen sınıflarına göre sınıflandırılmasında kullanılabilecek niteliktedir. Analizin temel amacı sınıflar arasındaki farkı belirlemek ve birkaç metrik bağımsız değişkene dayanarak bir nesnenin hangi sınıfa ait olduğuna karar verebilmeyi sağlamaktır. Değişken sayısının grup sayısından küçük olduğu durumlarda diskriminant analizi yerine yüksek dereceden diskriminant fonksiyonu ya da lojistik regresyon analizinin kullanılması önerilmektedir.

Temel bileşenler analizinde bağımlı bağımsız değişken ayırımı yapılmaz. Bu analiz tekniğinde amaç, p tane değişkenin birbirinden bağımsız öyle kombinasyonlarını bulmaktır ki, bütün örnek başlangıçtaki değişken sayısından çok daha az sayıda olan bu kombinasyonlarla açıklanabilsin. Birbirinden bağımsız bu kombinasyonlardan p tane, yani değişken sayısı kadar elde edilebilirken, kombinasyonların bir kısmının varyansının göz ardı edilebilecek kadar küçük olduğu gözlenecek ve bu da boyut indirgemeyi sağlayacaktır.

Temel bileşenler analizinde olduğu gibi faktör analizinde de bağımlı bağımsız değişken ayrımı yapılmaz Bu analiz yönteminin başlıca amacı boyut indirgemek, bağımsızlık sağlamak, kavramsal anlamlılığa ulaşmak ve çok sayıdaki değişken arasındaki ilişkiyi değişkenlerin ortak boyutunu, yani faktörünü göz önüne alarak incelemektedir.

Çokdeğişkenli varyans analizi birkaç bağımsız değişken ile iki veya daha fazla metrik bağımlı değişken arasındaki ilişkiyi belirlemek için kullanılır. Söz konusu yöntem esas olarak değişken ortalamalarını karşılaştırmaya yöneliktir.

Kanonik korelasyon analizi, bir metrik bağımlı değişken ile birkaç metrik veya metrik olmayan bağımsız değişken arasındaki ilişkiyi ölçen çoklu regresyon analizinin daha genişletilmiş bir hali olarak düşünülebilir. Bu analiz yönteminde bir yerine birden fazla metrik bağımlı değişken mevcuttur. Amaç, her bir değişken kümesinden aralarındaki korelasyon en yüksek olacak şekilde doğrusal kombinasyonlar türetmek ve önem testlerini kullanarak en önemli kombinasyonlara ulaşmak dolayısıyla boyut indirgemeyi sağlamaktır.

Kümeleme analizi, kişi veya nesnelerin anlamlı alt gruplarını oluşturmak için kullanılabilecek bir yöntemdir. Bu analiz tekniğinde amaç kişi veya nesneleri benzerliklerine dayanarak mümkün olduğunca az sayıda ve birbirinden tamamen ayrık kümelere atamaktır. Diskriminant analizinin aksine kümeleme analizinde küme (grup) sayısı önceden bilinmemekte, bu sayı örnekteki elemanlara göre analiz sonunda ortaya çıkmaktadır. Bu nedenle kümeleme analizi örneğe girecek veya çıkacak herhangi bir elemandan etkileneceğinden sadece uygulandığı örnek için geçerli olup sonraki örnek seçimlerinde kullanılamaz.

Çok boyutlu ölçekleme tekniğinin amacı, benzerlikler veya tercih konusunda bireylerin kararlarını çok boyutlu uzayda uzaklıklar cinsinden belirlemektir. Örneğin A ve B ürünleri tüketiciler tarafından diğer ürünlere nazaran birbirlerine benzer kabul edilmişlerse, çok boyutlu ölçekleme bu iki ürünü grafikte birbirine çok yakın olarak yerleştirecektir. Daha çok bir haritalama tekniği olarak düşünülebilir.

Pazar araştırmaları ve tercih verilerinde kullanılabilen konjoint analizinin amacı her bir değişkenin ve değişkenlerin her bir düzeyinin kişilerin ürün seçimi üzerindeki etkisini araştırmaktır. Kişilerin seçiminde en yararlı olan bileşeni belirleme ve simülasyon yapma olanakları tanıyan, son zamanlarda çok gündemde olan bir tekniktir.

Uygunluk analizi, temel bileşenler analizi ile yakından ilişkili, araştırıcı bir tekniktir. Bu teknik, iki yönlü çapraz tabloların satır ve sütunları arasındaki ilişkilerin çok boyutlu uzayda gösterimini sağlar.

Çokdeğişkenli uygunluk analizi olarak da tanımlanabilen homojenlik analizi değişkenler arasındaki homojenliği en büyüklemek için kullanılan bir yöntemdir. Birden fazla kategorisi bulunan birden fazla değişkenle ilgilenen homojenlik analizi, kategorilerin, değişkenlerin homojenliği en büyüklenecek şekilde puanlanmasını sağlar. Elde edilen puanlar kategorilerin grafik üzerinde gösterimi için kullanılırlar.

Yapılan araştırmaların yorumları sırasında homojenlik analizinden elde edilen tablolarda ayrımsamayı daha belirginleştirmek için bir veya daha fazla doğruya ihtiyaç duyulduğu gözlenmiştir. Bu doğrular araştırmacılar tarafından el yordamıyla çizilmekte dolayısıyla her araştırmacı aynı veri kümesi için farklı sonuçlara varabilmektedir. Çalışmamızda bu karışıklığa yol açmayacağını umduğumuz ve lojistik regresyondan faydalanarak elde edilen ayrımsama doğrularının hesaplanmasına da yer verilmiştir.

Bu çalışmada son zamanların en gözde analiz tekniklerinden biri olan homojenlik analizi üzerinde durulmuş, homojenlik analizinin ilişkili olduğu diğer analiz teknikleri ile olan benzerlik ve ayrılıklarına değinilmiştir. Uygulamaya son derece açık olan bu tekniğin daha iyi anlaşılabilmesine yardımcı olmak amaçlanmış ve bu amaç doğrultusunda bir otomotiv firmasının bayileri üzerinde gerçekleştirilen bir araştırmadan alınan gerçek verilerle uygulamalar yapılmıştır.

Bu amaca yönelik olarak çalışmanın izleyen ikinci bölümünde homojenlik analizi detaylı olarak incelenmiştir. Homojenlik analizinin temelini oluşturduğu düşünülen ve pazar araştırmalarında sıkça kullanılan uygunluk analizi bu bölümün ilk alt kısmıdır. Homojenlik analizini oluşturan kayıp fonksiyonu ve en küçükleme algoritmaları gibi konular da ikinci alt kısımda incelenmiştir. Üçüncü bölüm homojenlik analizi ile ilişkili olduğu düşünülen diğer çok değişkenli analiz yöntemlerine ayrılmıştır. İlk alt kısımda grafiksel gösterim sağlayan çok değişkenli analiz yöntemleri arasında geçmişi çok eskilere dayanan çok boyutlu ölçekleme analizi incelenmiştir. Doğrusal ve doğrusal olmayan temel bileşenler analizi ile ilgili temel bilgiler 3.2. Bölümde kanonik korelasyon analizi ise 3.3. Bölümde sunulmuştur. Yukarıda değinilen bütün analiz teknikleri ile yapılan uygulamalar kullanılan verilerin açıklamaları ile beraber dördüncü bölümde sunulmuştur. Çalışmanın son bölümü olan beşinci bölümde ise analiz tekniklerinin uygulamalarından elde edilen sonuçların kısa bir özeti verilmiş ve bu alanda ileride yapılabileceği düşünülen çalışmalara değinilmiştir.

Meltem G. TUNA
İstanbul, Mayıs 1999

BÖLÜM II

HOMOJENLEŞTİRME ANALİZİ

Homojenleştirme analizi (homogeneity analysis) sayısal olmayan veri yapısını resimlemeyi amaçlar. Bu amaca yönelik olarak nesne ve kategoriler için belirli kriterleri en iyileyen puanlar belirlenir daha sonra bu puanlar yardımıyla grafikler oluşturulur. Çoklu uygunluk analizi olarak da adlandırılan yönteme geçmeden önce yöntemin temelini oluşturan uygunluk analizinin incelenmesi yerinde olacaktır.

2.1. Uygunluk Analizi

Satır ve sütunları kategorik değişkenlerce tanımlanan ve hücreleri frekansları gösteren tablolar ilk kez 1904 yılında Karl Pearson tarafından çapraz tablo olarak adlandırılmıştır. Temellerinin o yıllarda atıldığı düşünülen bu tür tabloların kullanımı günümüze kadar önemini yitirmeden ulaşmıştır. Yule (1900, 1912), Pearson (1904, 1913) ve Goodman ve Kruskal (1959) gibi bir çok teorisyenin ilgisini çekmiştir. 1922-1962 yılları arasında değişik zamanlarda Fisher tarafından incelenmiştir. Fleiss (1981) ve Hennekes ve During (1987) tarafından klinik denemelerdeki kullanımının avantaj ve dezavantajları üzerinde durulmuştur. Sıralı kategorilerin söz konusu olduğu çalışmalar Fisher tarafından 1940 yılında başlatılmış ve Maung (1941), Williams (1952), Lancaster (1957,1969), Kendall ve Stuart (1961) tarafından geliştirilmiştir (Agresti, 1990, s. 13, 26-28, 296). Bu çalışmalarda daha çok değişkenlerin bağımsızlığının testi veya belirli bir dağılıma uygunluğu araştırılmıştır. Pearson tarafından öne sürülen, bağımsızlık veya uygunluk testinde kullanılan χ^2 istatistiği yoğun ilgi görmüş, her temel istatistik kitabında yerini almış ve bu alanda en çok kullanılan istatistik olmuştur. Bununla beraber bu istatistiğin hücre frekansı beşten az olan çapraz tablolarda sorun çıkartması, değişkenler bazında kalması ve kategoriler arası ilişkileri tek tek inceleme imkanı vermemesi nedeniyle daha detaylı yöntemlere ihtiyaç duyulmuştur. Çok sayıda değişkenin kullanıldığı detaylı araştırmalar için geliştirilmiş bir çok yöntem vardır. Bu yöntemlerden bir tanesi de uygunluk analizidir.

Çok değişkenli kategorik verilerin ve çapraz tabloların grafiksel gösteriminde kullanılan açıklayıcı bir veri analiz tekniği olarak da tanımlanabilen uygunluk analizinin tarihi 60 yıl öncesine kadar dayanmaktadır (Weller, 1990, s.15). Bununla birlikte yöntemin gerçek atılımını 1960'lı yılların başlarında J. P. Benzecri ve arkadaşlarının Rennes Üniversitesi'nde yaptığı çalışmalarla kaydettiğini söylemek yanlış olmaz. Benzer gelişmeler J de Leeuw ve C. Hayashi'nin çalışmaları ile Hollanda ve Japonya'da da gözlenmiştir. Benzecri'nin 'L'Analyse de Donnes' isimli Fransızca kitabı uygunluk analizi konusundaki bir çok çalışmaya temel oluşturmuştur. Bununla beraber kullanıldığı bir çok ülkede uygunluk analizinin anlaşılmasında M. J. Greenacre tarafından 1984 yılında yayınlanan 'Theory and Applications of Correspondence Analysis' kitabın büyük etkisi olduğu bilinmektedir. Carlier ve Kroonenberg (1996) uygunluk analizini üç boyutlu çapraz tablolara genelleştirmeye çalışmışlar, bu yolda önemli mesafeler almışlardır. Boik (1996) ise kategorik değişkenlerin kategorileri arasındaki ilişki yerine sadece değişkenler arasındaki ilişkiyi belirlemek için uygunluk analizinin kullanımı üzerinde durmuştur. Kategorik ve sürekli değişkenlerin bir karışımını içeren veri kümelerinin analizi Kiers (1991) tarafından ayrıntılı biçimde incelenmiştir. Bu yayınlarda uygunluk analizinin çok değişkenli durumda kullanılabileceği belirtilmekle birlikte uygulamada daha çok iki değişkenli çapraz tabloların kullanımı ağırlık kazanmaktadır.

Uygunluk analizi Devrik Ortalamalar Yöntemi (Method of Reciprocal Averaging), Toplamsal Ölçekleme (Additive Scoring), Uygun Ölçekleme (Appropriate Scoring), Kalitatif Veri İçin Temel Bileşenler Analizi (Principle Components Analysis of Qualitative Data) gibi birçok farklı isimle anılıp küçük faklılıklarla uygulanmıştır (Weller, 1990, s. 14). Adı geçen bütün bu uygulamalarda çalışmamızda açıklanan hesaplamalar esas alınmıştır.

Uygunluk analizi, negatif olmayan veri matrisini satır ve sütunları nokta olarak belirten özel bir tür grafiksel gösterime dönüştüren açıklayıcı çok değişkenli analiz yöntemlerinden birisidir. Birçok veri matrisine uygulanabilen bu yöntem daha çok iki kategorik değişkenin yer aldığı iki yönlü çapraz tabloların analizinde kullanılmaktadır. Genel olarak çapraz tabloyu oluşturan ilk değişkenin i kategorisi ile ikinci değişkenin j kategorisi kapsamındaki elemanların sayısı n_{ij} iken; i. satırın toplamı n_{i+}, j. sütunun toplamı n_{+j} ile gösterilmekte, genel toplam n_{++} ile ifade edilmektedir. Bahsedilen bu ifadeler aşağıda verilen tabloda gösterilmiştir.

Tablo 2.1: Çapraz Tabloların Genel Tanımı (N)

I \ J	1	2	...	j	...	Toplam
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+}
:	:	:		:		:
Toplam	n_{+1}	n_{+2}	...	n_{+j}	...	n_{++}

Uygunluk analizinde en çok kullanılan kavramlardan biri profildir. Bir frekanslar kümesinin profili frekansların toplam frekansa bölümü olarak tanımlanır. Örneğin i satırının profiline, i. satırdaki bütün frekansların teker teker n_{i+} ile gösterilen satır toplamına bölünmesi ile ulaşılabilir. Satır profillerinin genel biçimi Tablo 2.2'de verilmiştir. Ortalama olarak adlandırılan son satır ortalama satır profilidir ve sütun toplamlarının genel toplama bölümü ile elde edilmektedir. Ortalama satır profilinin elemanları sütun ağırlıklarını belirtir. Bir sütunun ağırlığı ise o sütunu yaratan özelliğin önemini gösterir. Ortalama satır profili hariç olmak üzere diğer profillerin oluşturduğu matris R ile, ortalama satır profili ise c sütun vektörü ile gösterilebilir. Ortalama satır profilinin r yerine c ile gösterilmesinin nedeni satır profil elemanlarının aynı zamanda sütun ağırlıklarını vermesidir.

Tablo 2.2: Satır Profilleri (R,c)

I \ J	1	2	...	J	...
1	n_{11}/n_{1+}	n_{12}/n_{1+}	...	n_{1j}/n_{1+}	...
2	n_{21}/n_{2+}	n_{22}/n_{2+}	...	n_{2j}/n_{2+}	...
:	:	:		:	
I	n_{i1}/n_{i+}	n_{i2}/n_{i+}	...	n_{ij}/n_{i+}	...
:	:	:		:	
Ortalama	n_{+1}/n_{++}	n_{+2}/n_{++}	...	n_{+j}/n_{++}	...

Sütun profilleri satır profilleri ile aynı özelliklere sahiptirler. Ortalama sütun profili hariç olmak üzere diğer profillerin oluşturduğu matris C ile , ortalama sütun profili ise r sütun vektörü ile Tablo 2.3'de olduğu gibi gösterilebilir.

Tablo 2.3: Sütun Profilleri (C,r)

I \ J	1	2	. . .	J	. . .	Ortalama
1	n_{11}/n_{+1}	n_{12}/n_{+2}	. . .	n_{1j}/n_{+j}	. . .	n_{1+}/n_{++}
2	n_{21}/n_{+1}	n_{22}/n_{+2}	. . .	n_{2j}/n_{+j}	. . .	n_{2+}/n_{++}
:	:	:		:		:
I	n_{i1}/n_{+1}	n_{i2}/n_{+2}	. . .	n_{ij}/n_{+j}	. . .	n_{i+}/n_{++}
:	:	:		:		:

Ortalama profiller satır ve sütunlar arasındaki benzerlik ve farklılıkların gözlenmesinde kullanılabilecek önemli araçlardır. Satır veya sütunlar arasında hiç fark olmaması durumunda bütün profillerin kendi ortalama profillerindeki değerlere çok yakın değerler içereceği ve küçük farklılıkların sadece örneklemden kaynaklanabileceği açıktır. Bilindiği gibi bu farklılıkların gerçekten örneklemden kaynaklanıp kaynaklanmadığının testinde χ^2 istatistiği kullanılmaktadır. Profillerin ortalama profillerden ne kadar saptığını grafiksel bir yöntemle gösteren uygunluk analizinde ise χ^2 uzaklığı ile yakın ilişkisi bulunan toplam hareketsizlik (inertia) veya kısaca hareketsizlik olarak adlandırılan bir istatistik kullanılmaktadır. Hareketsizliğin genel formu ve χ^2 ile olan ilişkisi aşağıdaki eşitlikle açıklanmaktadır.

$$\text{Hareketsizlik} = \frac{\chi^2}{n_{++}} = \sum_i \frac{n_{i+}}{n_{++}} \left[\sum_j \frac{\left(\frac{n_{ij}}{n_{i+}} - \frac{n_{+j}}{n_{++}} \right)^2}{\frac{n_{+j}}{n_{++}}} \right] \tag{2.1}$$

Hareketsizliğin alabileceği en küçük değer sıfır iken en büyük değer problemin boyutuna eşittir. Problemin boyutu doğru olarak [min (satır sayısı, sütun sayısı) - 1] bağıntısıyla belirlenmekle birlikte uygulamada yüksek boyutlu grafiklerin gösterim zorluğu nedeniyle genellikle 2 olarak alınmaktadır. Bu en yüksek değere bütün noktalar en uç durumlara yerleştiğinde yani, sapma çok büyük olduğunda ulaşılabilmektedir. Sıfır değerine ulaşmak ise ancak bütün noktalar ortalama profillerinin oluşturduğu merkezde toplanmışsa yani, sapma yoksa mümkündür.

Temel eksenlere göre satır ve sütun koordinatları aşağıda verilmiş olan standardize edilmiş matrisin tekil değer ayrıştırması (singular value decomposition) ile elde edilebilir. (Greenacre, 1993, s.181)

$$D_r^{-1/2}\left[\frac{1}{n_{++}}N - rc^T\right]D_c^{-1/2} = XD_\alpha Y^T \tag{2.2}$$

Burada D_r ve D_c sırasıyla ortalama sütun ve satır profillerinin köşegen matrisleri iken r ve c daha önce de belirtildiği gibi ortalama sütun ve satır profillerini göstermektedir. Bu ayrıştırma ile elde edilen tekil değerler (D_α) temel hareketsizliklerin kare köküdür:

$$D_\alpha = D_\lambda^{1/2} \tag{2.3}$$

Yukarıda bahsedilen tekil değer ayrıştırması kullanılarak hareketsizliği en iyi açıklayan eksenler belirlenmiş olur. Profillerin temel ve standart koordinatları da yine tekil değer ayrıştırması yardımıyla hesaplanabilmektedir. Aşağıdaki tabloda temel ve standart koordinatlar verilmiştir.

Tablo 2.4 : Temel ve Standart Koordinatlar

	Satır	Sütun
Temel Koordinatlar	$D_r^{-1/2}XD_\alpha$	$D_c^{-1/2}YD_\alpha$
Standart Koordinatlar	$D_r^{-1/2}X$	$D_c^{-1/2}Y$

Temel ve standart koordinatlar kullanılarak elde edilebilecek asimetrik ve simetrik olarak adlandırılan iki tür harita mevcuttur. Asimetrik haritalarda kullanılan nokta kümelerinden biri için (örneğin satırlar) temel koordinatlar diğer nokta kümesi için ise standart koordinatlar kullanılmaktadır. Satır profilleri için standart koordinatların sütun profilleri için temel koordinatların kullanıldığı asimetrik haritaların incelenmesiyle yapılan analiz sütun analizi olarak adlandırılır ve sadece sütunlar (sütundaki değişkenin kategorileri) arasındaki benzerlik veya farklılıkların önemli olduğu araştırmalarda kullanılır. Benzer şekilde sütun profilleri için standart koordinatların satır profilleri için temel koordinatların kullanıldığı asimetrik haritaların incelenmesiyle yapılan analiz ise satır analizi olarak adlandırılır ve sadece satırlar (satırdaki değişkenin kategorileri) arasındaki benzerlik veya farklılıkların önemli olduğu araştırmalarda kullanılır. Simetrik haritalarda ise hem satır hem sütun profilleri için temel koordinatlar kullanılmaktadır. Satır ve sütunların kendi aralarındaki ilişkileri yerine birbirleri ile olan ilişkilerinin incelenmesi gerektiğinde bu tür haritalar kullanılır (Greenacre, 1993, s. 63-74). Bu grafiklerin Greenacre (1993, s.181-183) tarafından ayrıntılı biçimde açıklanmış olan hesaplamalarına burada değinilmemiştir. Uygulama bölümünde ise daha çok kullanılan paket programın özelliklerine yer verilmiştir.

Uygunluk analizi yardımıyla elde edilen haritalar tekniğin kuramsal alt yapısı hakkında fikir sahibi olmayan bir kişinin bile nesneler ve kategoriler arasındaki ilişkiler konusunda bilgilenmesini sağlar. Gösterim kolaylığı nedeniyle iki boyut kullanılmakta fakat hareketsizlik hesapları daha fazla boyut için de yapılabilmektedir. İki boyuta indirgenen gösterimde ayrımsamayı kolaylaştırmak amacıyla bir veya daha fazla sayıda doğruya ihtiyaç duyulabilmektedir. Uygulamalarda bu doğruların araştırmacılar tarafından el yordamı ile çizilmeye çalışıldığı gözlenmiş ve analiz sonucunda elde edilen bilgiler kullanılarak doğrular için kuramsal bir altyapı oluşturulup oluşturulamayacağı araştırılmıştır.

Amaç farklı kategorileri birbirinden daha kolay ayırmada yardımcı olacak bir doğruya ulaşmaktır. Bu amaç doğrultusunda akla ilk gelen satır veya sütun noktaları için bir regresyon doğrusu elde etmek olabilir. Boyutlardan birinin bağımlı diğerinin bağımsız değişkene karşılık geldiği varsayılarak regresyon analizine başvurmanın akılcı olacağı düşünülebilirse de, boyutlar birbirinden bağımsız olduğundan bu düşünce sağlıklı olmayacak ve sadece x ve y eksenlerine çok yakın doğrular elde edilebilecektir. Halbuki bu eksenler uygunluk analizi ile zaten en iyi şekilde değerlendirilmektedir.

Ayrımsamayı kolaylaştırmayı sağlayacak bir doğru arandığı düşüncesi akla Diskriminant Analizini getirmektedir. Hipotetik olarak oluşturulan bir Y değişkeni ile diğer değişkenler için diskriminant analizi uygulaması genellikle değişken sayısı grup sayısından büyük olduğunda anlamlı sonuçlar vermemekte ve yorum güçleşmektedir. Diskriminant analizinde normallik varsayımı da söz konusu olduğundan başka analiz yöntemlerinin kullanılması gerekliliği kendiliğinden ortaya çıkmaktadır.

Bağımlı değişkenin kesikli olması durumunda değişkenler arası ilişkiyi incelemek için lojistik regresyon analizi önerilmektedir. Lojistik regresyon analizinde bağımlı değişken logit {ln(p/1-p)} dönüşümü kullanılarak (-∞ , +∞) aralığına yayılır ve böylece kesikli veya sürekli bağımsız değişkenler ile iki düzeyli bağımlı değişken arasındaki denge sağlanarak ilişkiler kolayca tahmin edilebilir. Bu yöntem ikiden fazla düzeyi bulunan bağımlı değişkenler için de kullanılabilir. Ayrıca, ayırmada önemli olan değişkenleri teşhis etmekle beraber veriye yeni alınacak bir gözlemin değerlerine bakılarak söz konusu gözlemin var olan gruplardan hangisine atanacağı da belirlenebilmektedir (Başarır, 1990, s. 111-115). Ardışık işlemler kullanan bu tekniğin sonuca ulaşamaması durumunda başlangıç verisi olarak kullanılan diskriminant analizi sonuçları esas alınabilmektedir.

Yapılan araştırmanın amacına uygun bir değişkenin orijinal değerlerine göre veya varsa daha önce elde edilmiş bilgiler kullanılarak değişken kategorileri gruplanır. Grup numaralarını veren değişkenin bağımlı değişken olduğu kabul edilir. Değişken kategorilerini iki gruba ayırmayı sağlayacak uygun bir değişken bulunamadığı takdirde kümeleme analizine baş vurulabilir. Uygunluk analizinden her kategori için elde edilen birinci ve ikinci boyut değerleri bağımsız değişkenler olarak alınır. Yukarıda açıklandığı biçimde elde edilen ve grup numaralarını içeren değişken bağımlı değişken ile bağımsız değişkenlere lojistik regresyon analizi uygulanır. Analiz sonucunda ulaşılan ayrımsama fonksiyonu lojistik regresyon fonksiyonu için ayrımsama noktası olarak alınan kritik değere eşitlenerek bir doğru denklemi oluşturulur. Farklı kritik değerlerin seçilmesi mümkündür. Bununla beraber en yüksek doğru ayrımsama düzeyine genellikle 0.5 kritik değerinde ulaşılabilmektedir. Böylece uygulamaya alınan bütün değişken kategorilerini iki gruba ayırmayı sağlayan bir doğruya ulaşılmış olur.

Yukarıda anlatılan işlemler birçok veri kümesine uygulanabilir. Bu işlemlerle ilgili daha geniş açıklamalar son bölümdeki uygulamalar kısmında verilmiştir.

Uygunluk analizi temel bileşenler (principle components), log-lineer ve çok boyutlu ölçekleme (multidimensional scaling) gibi çok değişkenli analiz yöntemleri ile yakından ilgilidir. Söz konusu analiz, sürekli yerine kategorik verilerle ilgilenmesi nedeniyle temel bileşenler analizinden, dağılım hakkında bir varsayıma ve model hakkında bir hipoteze gerek duymadığı için log-lineer analizden, değişkenler arasındaki ilişki ile kategoriler arasındaki ilişkiyi aynı uzayda gösterebilmesi özelliği nedeniyle çok boyutlu ölçeklemeden farklıdır.

Uygunluk analizi, SPSS ve LISP gibi istatistik paket programlar yardımıyla kolaylıkla uygulanabilen ve çapraz tabloların analizinde büyük kolaylık sağlayan bir yöntemdir. Pazar araştırmaları için araştırma şirketlerinin kullanımına çok uygun olan bu yöntem aynı zamanda daha ileri safhalardaki analiz yöntemleri için kategoriler ve değişkenler arasındaki ilişkilerin belirlenmesi için de kullanılabilmektedir. Bu nedenle çalışmanın uygulama bölümünde SPSS paket programının CORRESPONDENCE kategorisi kullanılarak uygunluk analizinin bir uygulamasına yer verilmiştir. Açıklamaları ayrıntılı biçimde sunulan uygulama sonuçlarından da anlaşılabileceği gibi çapraz tabloların incelenmesinde çok büyük kolaylık sağlayan uygunluk analizi ile bilgisayar programları yardımıyla anında elde edilebilen bir kaç haritayı yorumlamak ve anlamlı sonuçlara ulaşmak mümkündür.

2.2. Homojenleştirme Analizi

Kategorik değişkenler arasındaki ilişkiyi inceleyen log lineer modeller tekniği değişkenler arasındaki ilişki yapısını kullanarak çapraz tablo hücrelerini modellemeye çalışır. Goodman 1979 ve 1981 yıllarında, Agresti ise 1983'te sıralı kategorik değişkenlerin söz konusu olduğu çapraz tabloların bu yöntemle incelenmesi üzerinde durmuştur. İki boyut için bu modellere kolaylıkla ulaşılabilmekte fakat daha fazla boyut için aynı kolaylık söz konusu olmaktadır. Log lineer modeller analizi değişkenler arası etkileşimleri ortaya çıkartırken uygunluk ve homojenleştirme analizi gibi grafiksel gösterim ağırlıklı analiz teknikleri bu etkileşimleri görsel olarak ortaya koymaktadır (Van der Heijden, Worsley, 1988). Ayrıca ikiden fazla değişkenin bulunmasının bile araştırmayı zora koşmaması homojenleştirme analizinin önemli bir avantajıdır.

Optimal ölçekleme fikrinin çokdeğişkenli analize uygulanması 1978'lerde Young, De Leeuw ve Takane ile başlamıştır. Homojenleştirme analizinde olduğu gibi çok boyutlu ölçeklemenin de temelini oluşturan optimal ölçekleme, kategorik verilerin kodlanmasında odaklanmış olup daha eski çokdeğişkenli analiz tekniklerinin gelişimi ile çok boyutlu ölçeklemedeki gelişmeler arasında bağ kurulmasına yardımcı olmuştur (Van der Burg, de Leeuw 1988).

Çokdeğişkenli analiz tekniklerinde dönüştürmeler sadece ön bilgilere ve elde edilen verilere dayanılarak yapılmaktadır. Homojenlik analizinin temelini oluşturan optimal ölçekleme tekniğinde ise hedef kısaca, bir kayıp fonksiyonunun belirli kısıtlar altında en küçüklenmesi ile uygun dönüşümlere ulaşmaktır.

Çoklu Uygunluk analizi (Multiple Correspondence Analysis) (Koster, 1989), En İyi ve Uygun Ölçekleme Yöntemleri (Optimal Scoring and Appropriate Scoring Methods), Homojenleştirme analizi (Homogeneity Analysis), Niteleme Yöntemi (Quantification Method) (Tenenhaus, 1985) gibi birçok farklı isim verilmiş olan Homojenleştirme Analizi değişkenler arası homojenliği en büyüklemek için kullanılan bir yöntemdir. Anılan yöntem çok kategorili birden fazla değişken ile ilgilenir.

Değişkenlerin homojenliğinin enbüyüklenmesi amacıyla çokdeğişkenli veri analizinde kullanılan homojenleştirme analizi, bir kayıp fonksiyonu ve bu fonksiyonun en iyi çözümünü bulmak için kullanılacak bir yöntem yardımıyla kategorik verilerin analizini sağlayan teknik olarak da tanımlanabilir. Kategori sayısı birden çok olan birden fazla değişkenle ilgilenen homojenleştirme analizi, kategorilerin değişkenlerin homojenliğini en büyükleyecek şekilde puanlanmasını sağlar.

Kategorik verilerin incelenmesinde kullanıldığından ve temel bileşenler analizi ile olan benzerliğinden dolayı, kategorik temel bileşenler analizi, uygunluk analizi ile olan ilişkisinden dolayı da çokdeğişkenli uygunluk analizi olarak adlandırıldığına rastlanmaktadır.

Daha genel anlamda ise homojenleştirme analizi çeşitli çok değişkenli analiz tekniklerinin uygulanmasında kullanılabilecek, değişkenler arası homojenliği arttırmayı sağlayacak işlemler bütünü olarak da tanımlanabilir. Kısaca, homojenleştirme analizinin sadece bir analiz tekniği olmaktan çok bir düşünce tarzı olduğu söylenebilir. Bu durum Albet Gifi tarafından ortaya atılan ve Gifi - Sistemi olarak adlandırılan sistemde de homojenleştirme analizinin doğrusal olmayan çok değişkenli analiz için bir başlangıç noktası olarak kullanılmasıyla açıklanabilir (Van der Burg, de Leeuw 1988).

Homojenleştirme analizinin başta gelen amacı her bir nesneye bir skor ve her bir kategoriye bir puan atamaktır. Bir nesnenin skoru o nesnenin ait olduğu kategorilerin puanlarının ortalaması ile bir kategorinin puanı, içerdiği nesnelerin skorlarının ortalaması ile orantılıdır.

Nesne skorları ve kategori puanlarının belirlenmesi için kullanılan kayıp fonksiyonuna geçmeden önce kayıp fonksiyonunda yararlanılacak olan birtakım ifadelerin açıklanması gerekmektedir. **H** ile gösterilen veri matrisinin satır sayısı nesne sayısı n'ye, sütun sayısı değişken sayısı m'ye eşittir. h_j ise **H**'nin j. sütununu, yani j. değişkene ait sütununu belirtmek için kullanılmaktadır. Araştırma kapsamındaki j değişkeni, k_j değişik değer alabiliyorsa değişkenin k_j kategoriye sahip olduğu söylenir.

Söz konusu değişkene ait olan ve hazırlanışı gösterim bölümünde ayrıntılı biçimde açıklanan $\mathbf{G}_j$ gösterge matrisinin boyutu n x k_j dir. Daha önce de değinildiği gibi, gösterge matrisi hangi kategorilerin hangi nesneleri içerdiğini belirtmek için kullanılır.

Homojenleştirme analizi, homojenliği en büyükleyecek kategori puanlarını belirlemeyi sağlar. Bu kategori puanları k_j elemanlı y_j vektörü ile gösterilmektedir. Dolayısıyla $\mathbf{G}_j y_j$, n tane nesnenin j değişkeni yardımıyla puanlamasını göstermektedir. Gösterimin ortaya koyduğu gibi, aynı kategoriye düşen nesneler aynı puanı alacaklardır.

Gekçekleştirilen tanımlarla kayıp fonksiyonu aşağıdaki gibi düzenlenebilir:

$$\sigma\,(x;y) = m^{-1}\,\Sigma_j\,SSQ(\,x - \mathbf{G}_j\,y_j) \qquad (2.4)$$

Burada kullanılan SSQ(.) ifadesi parantez içindeki vektör ya da matrisin elemanlarının kareleri toplamını, x ise n elemanlı (nx1 boyutlu) nesne skorları vektörünü belirtmektedir.

Kategori puanları sayesinde elde edilen nesne puanları ($\mathbf{G}_j\,y_j$) ile nesne skorları (x) arasındaki fark homojenlikten sapmaya karşılık geldiğinden, σ (x;y) fonksiyonu kayıp fonksiyonu olarak değerlendirilir. Tam bir homojenlikten söz edebilmek için her bir değişken yardımıyla bulunan puanlamanın nesne skorlarına eşit olması, yani $x = \mathbf{G}_1\,y_1 = \mathbf{G}_2\,y_2 = \ldots = \mathbf{G}_m\,y_m$ eşitliğinin sağlanması zorunludur.. Homojenliğin en büyüklenmesi, bir anlamda homojenlikten sapmanın en küçüklenmesi olduğundan homojenleştirme analizinde amaç σ (x;y) kayıp fonksiyonunu en küçüklemektir. Başka bir deyişle, amaç σ (x;y) kayıp fonksiyonunu en küçükleyecek nesne skorlarını ve kategori puanlarını hesaplamaktır.

Her bir değişken için oluşturulan $\mathbf{G}_j$ gösterge matrislerinin sırayla yan yana eklenmesinden oluşan matris $\mathbf{G} = [\, \mathbf{G}_1 , \mathbf{G}_2 , \ldots , \mathbf{G}_j , \ldots , \mathbf{G}_m \,]$ ile gösterilirken, kategori puanları vektörlerinin sırayla alt alta birbirine eklenmesi ile elde edilen vektör $y' = [y_1', y_2', \ldots , y_j', \ldots , y_m']$ ile ifade edilmektedir. Bu durumda $\mathbf{G}y/m$ şeklindeki bir ifade, nesnelere değişkenlerin kategorileri tarafından verilen puanların ortalaması anlamına gelecektir. Hesaplanan bu nesne puanlarının nesne skorları ile orantılı olması gerekmektedir.

Benzer yaklaşımla nesne skorları kullanılarak $\mathbf{D}_j^{-1}\mathbf{G}_j'x$ ile aynı kategorideki nesne skorlarının ortalaması olarak hesaplanan kategori skorlarının da kategori puanları ile orantılı olması gerekmektedir. Uygun çözüme (feasible solution) ulaşılabilmesi bu orantıların sağlanması durumunda mümkün olur.

Yukarıdaki açıklamalar doğrultusunda sağlanması gereken ifadeler

$$x \propto \mathbf{G}y/m \quad (2.5)$$

$$y \propto \mathbf{D}^{-1}\mathbf{G}'x \quad (2.6)$$

şeklinde açıklanabilir.

Nesne skorları (x) ile kategori puanlarının (y) en iyi (optimum) çözümlerine dalgalı en küçük kareler (alternating least squares) yöntemine dayalı algoritmalar yardımıyla ulaşılabilir. Dalgalı en küçük kareler yönteminin kullanıldığı bu algoritmalarda kayıp fonksiyonu ilk adımda puan (skor) sabit tutulup skora (puana) göre en küçüklenirken daha sonraki adımda skor (puan) sabit tutulup puana (skora) göre en küçüklenmektedir. Her bir adımda diğer değere göre en küçükleme yapıldığından 'dalgalı' nitelemesi yönteme son derece uygun düşmektedir Bu çerçevede dalgalı en küçük kareler yöntemine bağlı kalınarak tanımlanan iki farklı algoritma kullanılabilir. Bunlardan birincisi skorlar üzerine kısıt getiren 'Normalleştirilmiş Skorlar Algoritması', diğeri ise puanlar üzerine kısıt getiren 'Normalleştirilmiş Ağırlıklar Algoritması'dır. Bu çalışmanın sonunda yer alan uygulama SPSS bilgisayar paket programında bulunan ve ayrıntılı açıklaması Uygulama Bölümü'nde sunulan HOMALS programı yardımıyla yapılmıştır. HOMALS programı normalleştirilmiş skorlar algoritmasını kullandığından (Gifi, 1990, s. 106) burada da sadece bu algoritmanın açıklanması ile yetinilmiştir.

Kategoriler için tek bir puanlamanın hesaplanabildiği ve (2.4) no'lu eşitlikte açıklanan kayıp fonksiyonu için Normalleştirilmiş Skorlar Algoritmasının adımları aşağıda verilmiştir.

(1) Nesne skorları düzeltilir : $\tilde{x} \leftarrow G\tilde{y} / m$

(2) Nesne skorları normalleştirilir: $x^{+} \leftarrow \sqrt{n}\,\tilde{x}(\tilde{x}'\tilde{x})^{-1/2}$

(3) Kategori puanları düzeltilir: $y^{+} \leftarrow D^{-1}G'x$

(4) Yakınsaklık testi yapılır: x^+ ve y^+ yeterince dengelenmedikçe, başka bir ifade ile kayıp fonksiyonu arzu edilen seviyeye inmedikçe $y^+ \rightarrow \tilde{y}$ ile birinci adıma dönülür. (Gifi, 1990, s. 88)

Normalleştirilmiş skorlar algoritmasında nesne skorları $x'x = 1$ kısıtını sağlamak zorundadır. Algoritmaya, kendisi sıfırdan farklıyken ($x \neq 0$) ortalamanın sıfır, varyansın bir olması için kareleri toplamı n ile normalleştirilmiş, tekdüze rasgele bir x seçeneği ile başlanır. Daha önce sözü edilen olan $y \propto \mathbf{D}^{-1}\mathbf{G}'x$ orantısı eşitlik olarak alınarak ilk $\tilde{y}$ kategori puan kümesi oluşturulduktan sonra algoritmanın uygulamasına geçilir (Gifi, 1990).

Yukarıda bahsedilen $\sigma(x;y)$ kayıp fonksiyonunda kategoriler için sadece tek bir puanlama söz konusudur. Oysa homojenleştirme analizinin en büyük özelliklerinden biri sadece tek bir puanlama ile sınırlı kalınmamasıdır. Her bir değişkenin kategorileri için eşanlı olarak birden fazla ve hatta boyut sayısı olarak tanımlanan p farklı puanlama bulunabilir. Elde edilmesi mümkün bu p tane puanlama, p-boyutlu puanlama olarak da adlandırılmaktadır. Bir j değişkeni için p tane puanlama k_j x p boyutlu $\mathbf{Y}_j$ matrisinde toplanır ve çoklu puanlama olarak adlandırılır. Dolayısıyla, $\mathbf{G}_j\mathbf{Y}_j$ şeklindeki bir ifade, n tane nesnenin j değişkeni sayesindeki p tane puanlamasını gösterecektir. Bu durumda X, nxp boyutlu nesne skorları matrisi, yani her bir nesne için p tane skor içeren matris iken kayıp fonksiyonunun aşağıdaki gibi düzenlenmesi gerekir:

$$\sigma(\mathbf{X},\mathbf{Y}) = m^{-1} \Sigma_j \, SSQ(\mathbf{X} - \mathbf{G}_j\mathbf{Y}_j) \quad (2.7)$$

Yeni kayıp fonksiyonu $\mathbf{X'X} = \mathbf{I}$ ve u tüm elemanları 1'e eşit olan n boyutlu sütun vektör iken $u'\mathbf{X} = \mathbf{0}$ kısıtları altında en küçüklenir. $u'\mathbf{X} = \mathbf{0}$ kısıtı **X**'in sütun ortalamalarından sapmaları içeren matris olmasını sağlarken, $\mathbf{X'X} = \mathbf{I}$ kısıtı **X**'in sütunlarının birim varyanslı ve birbirlerinden bağımsız olmalarını sağlar (van der Burg, de Leeuw 1988).

Tek puanlamada olduğu gibi tam bir homojenlikten söz edebilmek için her bir değişken yardımıyla bulunan çoklu nesne puanlamalarının nesne skorlarına eşit olması, başka bir deyişle $\mathbf{X} = \mathbf{G}_1\mathbf{Y}_1 = \mathbf{G}_2\mathbf{Y}_2 = \ldots = \mathbf{G}_m\mathbf{Y}_m$ eşitliğinin sağlanması gerekir. Eşitliklerin sağlanmaması homojenlikten sapma olduğunu gösterir. Bu sapma, birden fazla puanlamaya ulaşılmasını sağlayan HOMALS programında $\sigma(\mathbf{X},\mathbf{Y})$ kayıp fonksiyonu ile ifade edilmekte ve tek puanlamada bahsedilen normalleştirilmiş skorlar algoritması ile en küçüklenmektedir. Algoritmanın adımları aynı olmakla birlikte tek fark x ve y vektörleri yerine her bir sütunda farklı bir skor ve puanlama bulunduran **X** ve **Y** matrislerinin kullanılmasıdır.

Sütunları standardize edilmiş **G** matrisinin $(\mathbf{GD}^{-1/2})$ tekil değer ayrıştırması (singular value decomposition)

$$\mathbf{GD}^{-1/2} = \mathbf{V\Psi W'} \tag{2.8}$$

eşitliği ile ifade edildiğinde, dalgalı en küçük kareler yöntemi ile ulaşılacak olan durağan vektör çifti $x^* = v_1$ ve $y^* = \psi_1\mathbf{D}^{-1/2}w_1$ olarak açıklanacaktır Gifi 1990).

Tekil değer ayrıştırmasının ortaya koyduğu gibi ψ_1, $\mathbf{GD}^{-1/2}$'in ilk tekil değeri iken v_1 ve w_1 sırasıyla bu tekil değere karşılık gelen sağ ve sol vektörlerdir.

Daha önce bahsedilen $y \propto \mathbf{D}^{-1}\mathbf{G}'x$ orantısının eşitlik olarak kabul edilmesi ile ulaşılan $\mathbf{D}_j y_j = \mathbf{G}_j'x$ durağanlık eşitliğinin kullanılması ile

$$\sigma(x,y) = m^{-1} \Sigma_j \, SSQ(x - \mathbf{G}_j y_j) \qquad (2.9)$$

kayıp fonksiyonu sadece kategori puanlamaları cinsinden yazılabilir.

$$\sigma(x,y) = x'x + m^{-1} \Sigma_j \, y_j' \mathbf{D}_j \, y_j - 2\, m^{-1} \Sigma_j \, y_j' \mathbf{G}_j' x$$

$$= 1 - m^{-1} \Sigma_j \, y_j' \mathbf{D}_j \, y_j = 1 - m^{-1} y' \mathbf{D}\, y \qquad (2.10)$$

Kayıp fonksiyonunun dalgalı en küçük kareler yöntemi ile en küçüklenmesi durumunda ulaşılacak olan $y^* = \Psi_1 \mathbf{D}^{-1/2} w_1$ ifadesi kullanıldığında ile kayıp fonksiyonunun en küçük değerinin $1 - \Psi_1^2 / m$ olduğu görülecektir. $\sigma(*,*)$ kayıp fonksiyonunun değerini belirtmesi durumunda bu ifade aşağıdaki eşitlikle gösterilebilir:

$$\sigma(*,*) = 1 - m^{-1} \Psi_1^2 \, w_1' \mathbf{D}^{-1/2} \mathbf{D}\, \mathbf{D}^{-1/2} w_1 = 1 - \Psi_1^2 / m \qquad (2.11)$$

$x \propto \mathbf{G}y/m$ ve $y \propto \mathbf{D}^{-1}\mathbf{G}'x$ orantılarından $x \propto (\mathbf{GD}^{-1}\mathbf{G}'/m)x$ yazılabileceği görülmektedir. Bilindiği gibi $\mathbf{GD}^{-1/2} = \mathbf{V\Psi W'}$ tekil değer ayrıştırmasının varlığı durumunda, $\mathbf{GD}^{-1/2}\,\mathbf{D}^{-1/2}\,\mathbf{G}'/m$ matrisinin özdeğer ayrıştırması aşağıdaki eşitlikle gösterilebilir:

$$\mathbf{GD}^{-1/2}\,\mathbf{D}^{-1/2}\,\mathbf{G}'/m = \mathbf{V\Psi W'W\Psi V'}/m = \mathbf{V}\,(\Psi^2/m)\,\mathbf{V'} \qquad (2.12)$$

Bu durumda Ψ^2/m matrisinin $\mathbf{GD}^{-1}\mathbf{G}'/m$ matrisinin özdeğerlerini içerdiğini ve nesne skorlarının da (x) bu matrisin özvektörleri olduğu söylenebilir. $\mathbf{GD}^{-1}\mathbf{G}'/m$

matrisinin en büyük özdeğeri de Ψ_1^2/m olacaktır. Başka bir deyişle, kayıp fonksiyonu en küçük değerine, $\mathbf{GD}^{-1}\mathbf{G}'/m$ matrisinin en büyük özdeğerinin bir ile farkından ulaşır.

Aynı düşünce tarzı çoklu çözüme uygulandığında elde edilen bağıntı aşağıda verilmiştir.

$$\sigma(\mathbf{X},\mathbf{Y}) = \mathrm{tr}(\mathbf{X}'\mathbf{X}) - m^{-1} \Sigma_j \, \mathrm{tr} \, \psi_j'\mathbf{D}_j\mathbf{Y}_j = p - m^{-1} \Sigma_j \, y_s'\mathbf{D} \, y_s \qquad (2.13)$$

Dalgalı en küçük kareler yöntemi ile ulaşılacak olan durağanlık çifti ise $\mathbf{X}^* = \mathbf{V}_p$ ve $\mathbf{Y}^* = \mathbf{D}^{-1/2}\mathbf{W}_p \, \psi_p$ eşitlik çifti ile açıklanır. Burada kullanılan p alt indisi, seçilen ilk p tekil değerine karşılık gelmektedir. Bu durumda en küçük kayıp aşağıdaki eşitlikle belirtilebilir:

$$\sigma(*,*) = p - \Sigma_s \, \psi_s^2 / m \qquad (2.14)$$

Kısaca ifade etmek gerekirse, çoklu çözüm için oluşturulan kayıp fonksiyonunun en küçük değeri, $\mathbf{GD}^{-1}\mathbf{G}'/m$ matrisinin en büyük s özdeğeri toplamı ile p boyut sayısı arasındaki farka eşittir. ψ_s^2 / m s. boyuttaki ayırım ölçülerinin ortalamasıdır (Gifi 1990, s. 109). Ayırım ölçüsü değişkenin boyuta katkısı olarak tanımlanmaktadır. Değişkenin boyuta katkısı olmaması halinde ayırım ölçüsü 0 olacak yani değişkenin kategorileri orijinle çakışacaktır.

Belirtilen kayıp fonksiyonlarının en küçüklenmesi ile elde edilen nesne skorları ve kategori puanlarının çok boyutlu uzayda gösterimi homojenleştirme analizinin en büyük özelliklerindendir. Bu gösterimler, hesaplamalar hakkında fikir sahibi olmayan bir kişinin bile nesneler ve kategoriler arasındaki ilişkiler hakkında bilgilenmesini sağlar. Gösterim kolaylığı nedeniyle iki boyut kullanılır ve her boyut için ayrı ayrı hesaplanan nesne skorları ve kategori puanları ile grafiksel gösterime ulaşmak oldukça kolaydır. Böylece elde edilen grafikte ayrımsamayı kolaylaştırmak amacıyla bir veya daha fazla sayıda doğruya ihtiyaç duyulabilmektedir. Uygunluk analizinde olduğu gibi homojenleştirme analizi için de lojistik regresyondan faydalanılabilir. Uygunluk analizinin son bölümündeki açıklamalar burada da geçerliliğini korumaktadır. Tek fark ikiden fazla değişkenin söz konusu olmasıdır. Lojistik regresyon yardımıyla çizilen doğrular artık sadece iki değişkenin değil analize katılan bütün değişkenlerin sınıflandırılmasına yardımcı olmaktadır.

BÖLÜM III

HOMOJENLEŞTİRME ANALİZİ İLE İLİŞKİLİ DİĞER ÇOK DEĞİŞKENLİ ANALİZ TEKNİKLERİ

3.1. Çok Boyutlu Ölçekleme

Çok değişkenli analiz yöntemlerinin büyük çoğunluğu varyans, kovaryans ve korelasyon gibi istatistikler ile dağılımla ilgili varsayımlar gerektiren yöntemlerdir. Çok değişkenli analiz yöntemlerinden biri olan çok boyutlu ölçekleme yöntemi ise birimler arası benzerlik veya farklılıkların kullanıldığı gösterime dayalı bir yöntemdir. Bu nedenle tarihsel gelişimi diğer çok değişkenli yöntemlerin tarihsel gelişiminden biraz farklı olmuştur.

Çok değişkenli analiz tekniklerinin pek çoğunun temeli 1901'lerde atılıp o zamandan beri ilerleme kaydedilmesine rağmen verilerin grafiksel gösterimine dayalı teknikler üzerindeki ilgi 1940 ve 1960 yılları arasında kesintiye uğramıştır. Young ve Householder 1938 yılında aralarındaki öklit uzaklıkları verilmiş olan noktaların yerleşim biçimini oluşturmak için faktör analizi ile yakından ilişkili bir teorem ortaya atmışlardır. Yine aynı yıl Richardson, çok boyutlu ölçekleme için bu teoreme dayanan ilk metrik yöntemi geliştirmiştir. Aradan 20 yıl gibi uzun bir zaman geçmiş konu ile ilgili bu çalışmalar ancak 1958 yılında Torgerson tarafından tekrar gözden geçirilmiştir (Kruskal 1986, s. 22).

Shephard 1962 yılında tüm ilgisini benzerlik ve uzaklıklar arasındaki ilişki üzerine yoğunlaştırmış ve sonuçta metrik olmayan çok boyutlu ölçeklemeye dayanan ilk bilgisayar programını ortaya çıkarmıştır. Shephard'dan iki yıl sonra Kruskal iki bilgisayar programı daha geliştirilmiştir (Kruskal, 1986, s. 23). Bilgisayar kullanımının daha da yaygınlaşması ile çok boyutlu ölçekleme gibi gösterime dayalı tekniklere olan ilgi artarak sürmüştür.

Çok boyutlu ölçeklemede biri metrik diğeri metrik olmayan iki yöntem üzerinde durulmaktadır. Metrik çok boyutlu ölçekleme yönteminde bireyler arasındaki uzaklıklar veya farklılıklar matrisi bilinmekte ve bu matris yardımıyla nesnelerin daha az boyutlu bir grafikte birer nokta halinde gösterimi sağlanmaya çalışılmaktadır. Farklılıklar matrisini oluşturan verilerin elde edilmesi diğer veri toplama yöntemlerinden oldukça farklıdır. Bu tür matrisleri oluşturmada kullanılabilecek verileri toplama yöntemleri 'MultivariateData Analysis: With Readings' adlı kitapta ayrıntılı biçimde incelenmiştir. Metrik olmayan çok boyutlu ölçekleme yönteminde ise uzaklıklar yerine sadece uzaklık sıraları kullanılabilmektedir. Uzaklık sıraları 'mümkün olduğunca' korunarak nesnelerin grafiksel gösterimine ulaşılmaya çalışılır.

Boyut ne olursa olsun grafik üzerine yerleştirilmiş nesnelere ait uzaklıklar matrisine, sadece matematiksel işlemlerle kolayca ulaşılabilmektedir. Fakat aynı ifade uzaklıklar matrisi verilen nesnelerin grafik üzerinde gösterilmesi probleminde kullanılamaz. Bu durumda bu tür problemlerin çözümü için geliştirilmiş olan metrik çok boyutlu ölçekleme yöntemi kullanılır. Uzaklıklar matrisi verildiğinde nesnelerin bir grafik üzerinde gösterimini sağlayan bu yöntem sayesinde sadece aralarındaki uzaklıklar bilinen nesnelerin birbirleri ile olan ilişkilerinin bir bakışta görülebileceği grafikler oluşturmak mümkündür. Bu gösterimlerde, bütün nesnelerin tek bir çizgi üzerinde yerleştirildiği tek boyut, nesnelerin bir düzlemde yerleştirildiği iki boyut ve nesnelerin bir uzayda yerleştirildiği üç boyut kullanılması mümkündür. Söz konusu yöntem grafiksel gösterime dayalı olduğundan daha fazla boyuttan söz etmek mümkün değildir. Gerçek verilerde kaç boyuta ihtiyaç duyulacağı kesin olarak bilinemeyeceğinden birkaç boyutun denenmesi gerekebilmektedir (Tatlıdil, 1996, s. 353-467).

Nesneler arası farklılıklar matrisi δ ile gösterilsin. Bu durumda δ_{ij}, i nesnesi ile j nesnesi arasındaki farkı gösteren i. satırın j. sütun elemanı olacaktır. Çok boyutlu ölçekleme nesneler arasındaki uzaklıklar $\{d_{rs}\}$, farklılıklar veya farklılıkların bir fonksiyonuna $\{f(\delta_{rs})\}$ yaklaşık olarak eşit $\{d_{rs} \approx f(\delta_{rs})\}$ olacak şekilde grafiksel bir gösterim sağlar. Bunun için de temelde iki yöntem kullanılır. Bunlar; Klasik ölçekleme ve En küçük kareler ölçeklemesidir (Cox ve Cox 1994, s. 22).

Temeli 1930'lu yıllara dayanan klasik ölçeklemede nesnelerin koordinatları spektral ayrıştırmadan faydalanılarak bulunur. p boyutlu öklit uzayındaki nesnelerin koordinat vektörleri

$$r = 1, 2, \ldots, n \quad \text{için} \quad x_r = (x_{r1}, \ldots, x_{rp})$$

ile gösterildiğinde r ve s nesneleri arasındaki uzaklık aşağıdaki gibi yazılabilir.

$$d_{rs}^2 = (x_r - x_s)^T (x_r - x_s) = x_r^T x_r + x_s^T x_s - 2\, x_r^T x_s$$

Koordinat vektörlerinin çarpımlarının oluşturduğu B ($= x_r^T x_s$) matrisi, $X = [x_1,\ldots,x_n]^T$, nxp boyutlu koordinatlar matrisi i ken , $B = X X^T$ olarak da yazılabilir. Aynı matrisin spektral ayrışımının da $B = V\Lambda V^T$ olduğu düşünülürse, p boyutlu öklit uzayı için koordinatlar matrisi $X = V\Lambda^{1/2}$ şeklinde yazılabilir. Tanımından dolayı B matrisi simetrik bir matristir. Ayrıca pozitif yarı tanımlıdır ve rankı p'ye eşit olduğundan (rank B = rank X = p) p tane negatif ve sıfır olmayan özdeğere sahiptir. Dolayısıyla Λ sadece sıfırdan farklı özdeğerlerin oluşturduğu köşegen matris (Λ= köş$(\lambda_1,\lambda_2, \ldots,\lambda_p)$) iken V bu özdeğerlere karşılık gelen özvektörlerin matrisidir.

Farklılıkların pozitif yarı tanımlı olmayan bir B matrisi oluşturmaları durumunda farklılıklar matrisinin köşegen elemanları dışındaki elemanlarına uygun bir sabit eklenebilir. Ekleme sonucunda uzaklıklar $d_{rs} = \delta_{rs} + c(1- \delta_{rsKR})$ olarak alınarak B matrisi pozitif yarı tanımlı hale dönüştürülür. Burada δ_{rsKR} Kronecker delta'yı, c ise uygun sabiti belirtmek için kullanılmıştır (Cox ve Cox, 1994, s.26).

Gerekli olabilecek en yüksek boyut; B matrisinin pozitif yarı tanımlı olması halinde sıfırdan farklı özdeğer sayısına, diğer durumlarda ise pozitif özdeğer sayısına eşittir. Bu yaklaşımla belirlenen boyutlar en yüksek boyutlardır. Uygulamada daha küçük boyutların seçilmesi gerekir. Bunun için seçilen boyutla değişimin ne oranda açıklanabildiği büyük önem taşır. p boyutla açıklanan değişim oranının ölçüsü pozitif yarı tanımlı B matrisi için

$$\sum_{i=1}^{p} \lambda_i \Big/ \sum_{i=1}^{n-1} \lambda_i \tag{3.1}$$

pozitif yarı tanımlı olmayan B matrisi için

$$\sum_{i=1}^{p} \lambda_i \Big/ \sum_{i=1}^{n-1} |\lambda_i| \quad \text{veya} \quad \sum_{i=1}^{p} \lambda_i \Big/ \sum (\text{pozitif özdeg erler}) \tag{3.2}$$

şeklinde verilebilmektedir. Boyutun seçimi bu ölçülere dayanılarak yapılabilmektedir (Cox ve Cox, 1994, s. 29).

En küçük kareler yöntemi, çok boyutlu ölçeklemede, 1970'lerin ortalarına kadar fazla ilgi gören bir yöntem olmamasına karşın popülaritesi, gelişen bilgisayar teknolojisi ile birlikte hızla artmıştır. ALSCAL ve SMACOF gibi bir çok paket program bu yöntemi kullanmaktadır (Cox ve Cox, 1994, s. 39).

En küçük kareler yönteminde farklılıkların dönüştürülmesi sürekli monoton bir fonksiyon kullanımıyla gerçekleşir. Bu fonksiyonun seçimi daha çok kullanılan veriye ve araştırmanın amacına bağlıdır. En temel ve çok kullanılan dönüşüm fonksiyonu α ve β parametreleri en küçük kareler yöntemi ile belirlenen

$$f(\delta_{rs}) = \alpha + \beta\, \delta_{rs} \tag{3.3}$$

fonksiyonudur. Bunun yanında (3.4) numaralı eşitlikte verilen bir logaritmik fonksiyon ve (3.5) numaralı eşitlikte verilen bir üstel fonksiyon'da dönüşüm fonksiyonlarına örnek olarak verilebilirler.

$$f(\delta_{rs}) = \beta.\log(\delta_{rs}) \tag{3.4}$$

$$f(\delta_{rs}) = \alpha + \beta.\exp(\delta_{rs}) \qquad (3.5)$$

Borg ve Groenen (1997, s. 161 - 162) aşağıda verilen ikinci dereceden polinom gibi monoton olmayan fonksiyonların da aynı amaçla kullanılabileceğini belirtmişlerdir.

$$f(\delta_{rs}) = \alpha + \beta.\ \delta_{rs} + \theta\ \delta_{rs}^{2} \qquad (3.6)$$

Koordinatlar, W_{rs} araştırmanın özellikleri doğrultusunda belirlenen ağırlıkları gösterirken,

$$S = \frac{\sum_{r \neq s} W_{rs}\left(d_{rs} - f(\delta_{rs})\right)^2}{\sum_{r \neq s} d_{rs}^2} \qquad (3.7)$$

değeri en küçüklenecek şekilde belirlenir. $\{d_{rs}\}$ uzaklıkları öklit uzaklığı olmak zorunda değildir. Araştırmanın özelliklerine göre Chebychev, Minkowski gibi bir çok farklı uzaklık arasından seçim yapılabilir. Uygulama bölümünden de görülebileceği gibi kullanılmak istenen uzaklık cinsi SPSS paket programının çok boyutlu ölçekleme bölümünde sunulan bir çok alternatif arasından seçilebilmektedir.

Konuyla ilgili kitap ve yayınlarda bir çok çeşidine rastlanan ve stres olarak adlandırılan S değerinin birkaç farklı tanımı aşağıda verilmiştir (Tatlıdil, 1996, s. 356; Borg, Groenen, 1997, s. 160):

$$\text{STRES } 1 = \sum_{i}^{n}\sum_{j}^{n}\left(d_{ij} - \hat{d}_{ij}\right)^2 \qquad (3.8)$$

$$\text{STRES } 2 = \sum_{i}^{n}\sum_{j}^{n} w_{ij}\left(d_{ij} - \hat{d}_{ij}\right)^2 \qquad (3.9)$$

$$\text{STRES } 3 = \sum_{i}^{n}\sum_{j}^{n}\left(d_{ij}^{2} - \hat{d}_{ij}^{2}\right) \qquad (3.10)$$

Bilgisayar paket programlarında da klasik ölçekleme yöntemi yerine, aralarında küçük farklılıklar bulunan bu stres değerleri ile, en küçük kareler yöntemi kullanılmaktadır.

Metrik çok boyutlu ölçekleme yönteminde araştırıcıya büyük kolaylıklar sağlayan uzaklıklar veya farklılıklar matrisinin varlığı söz konusudur. Halbuki, çok boyutlu ölçeklemenin yaygın biçimde kullanıldığı pazar araştırmalarında karşılaşılan bazı durumlarda böyle bir kolaylıktan söz etmek mümkün değildir. Bilindiği gibi, pazar araştırmalarında müşterilerin, servislerin veya ürünlerin en temel özelliklerini belirlemekte çok boyutlu ölçekleme yönteminden faydalanılmaktadır. Bu tür araştırmaların bazılarında ise uzaklıklar veya farklılıklar matrisi değil sadece uzaklık sıra sayılarına (Tatlıdil , 1996, s. 361) ulaşmak mümkün olmaktadır. Ne uzaklıklar matrisine ne de farklılıkların sıra sayılarına ulaşmak mümkün olduğunda eldeki verilerin korelasyonlarının kullanılması gerekebilmektedir (Kruskal, 1986). Bu gibi durumlarda metrik olmayan çok boyutlu ölçekleme yöntemi kullanılmaktadır.

Önceden olduğu gibi nesneler arası farklılıklar matrisi δ ile gösterilsin. Metrik olmayan çok boyutlu ölçekleme nesneler arasındaki uzaklıkların $\{d_{rs}\}$ farklılıklara $\{\delta_{rs}\}$ 'olabildiğince' eşit olacak şekilde veya farklılıkların sıra sayıları 'olabildiğince' korunacak şekilde grafiksel gösterimini sağlar. Sıra sayıları söz konusu olduğundan, çarpma bölme gibi matematiksel işlemlere dayalı fonksiyonların {(3.3)-(3.6)} kullanımı fazla anlamlı olmamaktadır. Bu sorunu ortadan kaldırmak için oransızlık veya eşitsizlik (disparity) olarak adlandırılan yeni değerler ($\hat{d}_{rs}$) tanımlanır. Eşitsizlikler $\{\hat{d}_{rs}\}$ nesneler arası fark sıralarının korunmasını sağlarken bazı gruplaşmalara izin verir.

Metrik olmayan çok boyutlu ölçekleme yönteminde de çok boyutlu uzaydaki gerçek şekil ile indirgenmiş uzayda kestirilen şekil arasındaki farklılığın ifadesi olan farklı stress değerleri (kayıp fonksiyonu) kullanılabilmektedir. Stress değerlerin en bilinen üçü aşağıda verilmiştir (Cox ve Cox, 1994, s. 43, 44, 55)

$$S_1 = \min_{\phi}\left\{\sum_{r,s}\left|d_{rs}-\hat{d}_{rs}\right|\right\} \tag{3.11}$$

$$S_2 = \left\{\frac{\sum_{rs}\left(d_{rs}-\hat{d}_{rs}\right)^2}{\sum_{rs} d_{rs}^2}\right\}^{1/2} \tag{3.12}$$

$$S_3 = \left\{\frac{\sum_{r,s}\left(d_{rs}-\hat{d}_{rs}\right)^2}{\sum_{r,s}\left(d_{rs}-d_{..}\right)^2}\right\}^{1/2} \quad , d_{..} = d_{rs}\text{'lerin ortalaması} \tag{3.13}$$

Bu çalışma bilgisayar uygulamalarına ağırlık vermek üzere planlandığından veri toplama yöntemlerine değinilmemiş, sadece toplanan veriler üzerinde gerekli düzenlemeler yapıldıktan sonra ulaşılan sonuçların açıklamaları üzerinde durulmuştur.

Çok boyutlu ölçekleme ile uygunluk analizinin pek çok ortak noktası vardır. Örneğin her iki teknik de nesneleri daha az boyutlu uzayda grafiklerle açıklar. Basit anlamda çok boyutlu ölçekleme bir grup nesne ile ilgilenirken uygunluk analizinde satır ve sütun olmak üzere iki grup nesne veya değişken söz konusudur. Uygunluk analizinde verilerin negatif olmaması gerekirken çok boyutlu ölçekleme ile sıra sayıları bile incelenebilmektedir.

3.2. Temel Bileşenler Analizi

3.2.1. Doğrusal Temel Bileşenler Analizi

Temelinin 1901 yılında Karl Pearson tarafından dik en küçük kareler yöntemi ile düzlemler oluşturulması ve korelasyon elipsoidinin temel eksenlerinin kullanılması şeklinde atıldığı belirtilmekle birlikte, doğrusal temel bileşenler analizi'ne en büyük katkının 1933 ve 1936 yıllarında Hotelling tarafından yapıldığı da kabul edilmektedir (Morrison, D. F., 1967, s. 222 , Mardia K. V. et. al, 1979, s. 213). İlk olarak 1933'de Hotelling tarafından veri analiz tekniği olarak ele alınan bu yöntemde Hotelling ve Girshick (1936) homojenlik fikrini başlama noktası olarak kullanmışlar ve en yüksek varyanslı doğrusal bileşimi aramışlardır. Bu doğrusal bileşim değişkenlerle korelasyonlarının kareleri toplamı en büyük olan doğrusal bileşimle aynıdır. Diğer bileşimlere diklik kısıtı altında bir sonraki en iyi ikinci çözümden ulaşılır. Eckart ve Young (1936), genel bir p (boyut) için temel bileşenler analizini, ilk p eşanlı bileşenin uyacağı en iyilik (optimality) özelliklerini formüle ederek tanıtmışlardır. Ayrıca 1906'da Schmidt tarafından daha genel olarak tanımlanmış olan tekil değer ayrıştırmasının en küçük kareler özelliğini de kullanmışlardır. Araştırmacılar ve teorisyenler tarafından çokça kullanılan bu yönteme pek çokdeğişkenli analiz kitabında rastlamak mümkündür. Fransa'da da büyük ilgi görmüş olan bu yönteme Cailliez ve Pagès tarafından hazırlanan Introduction à l'Analyse des Données (1976) kitabında da geniş yer verilmiştir (Gifi, 1990, s. 151 - 152).

Esas olarak boyut indirgeme ve değişkenler arasındaki bağımlılığı yok etme amacıyla için kullanılan temel bileşenler analizi birçok çalışmada tek başına kullanıldığı gibi başka analiz tekniklerine veri hazırlamak için de kullanılmaktadır. Temel bileşenler analizi ile ulaşılan doğrusal bileşimlerin en önemli özelliği orijinal değişkenler ile aralarındaki korelasyon kareleri toplamının enbüyük olmasıdır. Korelasyon kareleri toplamını en büyükleyen ilk bileşene birinci temel bileşen denir. Bu bileşen varyansın en büyük kısmını açıklar. İkinci temel bileşen ise birinci temel bileşene dik olan ve korelasyon kareleri toplamı birinci temel bileşenden sonra en büyük olan bileşendir. İkinci temel bileşen birinci temel bileşenden sonra kalan varyansın en büyük kısmını açıklar. Her temel bileşen, kendisinden önceki bileşenlere diktir ve korelasyon kareler toplamı kendisinden önceki bileşenlerden sonraki en büyük korelasyon kareler toplamıdır. Orijinal bileşen sayısı kadar temel bileşen elde edilebilir. İlk bir kaç bileşenin varyansı yeterince açıklaması durumunda araştırmaya sadece bu bileşenlerle devam edilmesi temel bileşenler analizinin boyut indirgeme özelliğidir. Yeni bulunan temel bileşenlerin sayısı ile başlangıçtaki değişken sayısı arasındaki fark kadar temel bileşenin sadece kalabalık yarattığı ve göz ardı edilebileceği söylenir. Boyut indirgeme, bileşenlerin toplam varyansın yaklaşık 2/3'ünü açıklaması durumunda anlamlı ve kullanılabilirdir. Temel bileşenler analizi bütün orijinal değişkenler birbirlerinden bağımsızken daha az sayıda değişkenle çalışma kolaylığını yani boyut indirgemeyi sağlayamaz.

Temel bileşenler analizindeki **H** veri matrisi ölçülebilir, yani sayısal değişkenlerden oluşmaktadır. Bu analiz tekniği sayısal değişkenler için geliştirilmiş olmakla birlikte ilerleyen bölümlerde açıklandığı gibi sırasal ve kategorik değişkenler için de düzenlenebilir.

3.2.1.1 Temel Bileşenlerin Elde Edilmesi

Çoğunlukla olduğu gibi çalışma kapsamındaki değişkenlerin ölçü birimlerinin birbirinden çok farklı olması durumunda, orijinal veri matrisi yerine standartlaştırılmış veri matrisinin kullanılması gerekir. Veri matrisi **H**'nin sütunları eşit olarak normalize edilir, yani her satır kendi kareler toplamına bölünmesiyle standartlaştırılmış veri matrisine ulaşılır. Daha önceki bölümde yapılan açıklamalardan da anlaşılacağı gibi temel bileşenler orijinal verilerin doğrusal bileşenidir, dolayısıyla **H** veri matrisi ile aynı boyutta olan ve temel bileşenleri içeren X matrisi için aşağıdaki eşitlik yazılabilir:

$$\mathbf{X}_{nxm} = \mathbf{A}_{pxp}\,\mathbf{H}_{pxn} \tag{3.14}$$

Bu eşitlikte kullanılan **A** matrisi değişkenlerin katsayılarını içeren matristir ve dönüşüm matrisi olarak da adlandırılır. Bilindiği gibi elde edilen yeni değişkenlerin yani temel bileşenlerin birbirine dik olmaları gerekir. Temel bileşenlerin dik olabilmesi için **X**'in varyans matrisinin köşegen matris olması gerekmektedir. (4.1) eşitliği ile varyans matrisinin köşegen matris olmasını sağlayacak birden fazla **A** matrisi bulunabilir. Bu durumun önüne geçmek için bazı kısıtlara ihtiyaç vardır. Bunlardan ilki $a_1'a_1=1$ ile açıklanan kısıtdır. Bu kısıt yardımıyla A matrisinin elemanları olan a_j vektörlerinin $|\mathbf{R} - \lambda\mathbf{I}| = 0$ eşitliğinden bulunacak olan λ_j özdeğerlerine karşılık gelen özvektörler olacağı bilinmektedir. (Tatlıdil, **H.**, 1996, s. 141). Burada kullanılan **R** değişkenler arası korelasyon matrisidir. **H** standartlaştırılmış veri matrisi ve **D**, **H'H** matrisinin köşegen elemanlarının oluşturduğu köşegen matris iken korelasyon matrisi $\mathbf{R}=\mathbf{D}^{-1/2}\mathbf{H}'\mathbf{H}\mathbf{D}^{-1/2}$ eşitliğinden elde edilir.

Yukarıdaki açıklamalardan da anlaşılabileceği gibi, temel bileşenler analizi basit olarak örnek kovaryans veya korelasyon matrisinin özdeğerlerinin bulunmasından ibarettir. Çünkü bir temel bileşenin varyansı kovaryans veya korelasyon matrisinin özdeğeridir (Manly B. F. J., 1989, s. 59). Bu özdeğerlerin bir kısmı sıfırdır ve değişken sayısı kadar özdeğer vardır. En büyük özdeğer birinci temel bileşenin varyansına eşittir.

Özdeğer ve özvektör cinsinden açıklanan temel bileşenler analizinin kayıp fonksiyonu cinsinden, yani homojenlik fikri ile ifadesi aşağıdaki gibidir:

$$\sigma(\mathbf{X},\mathbf{A}) \equiv m^{-1} \Sigma_j \, SSQ(\mathbf{X} - h_j a_j') \qquad (3.15)$$

Burada **X** nxp boyutlu, **A** ise mxp boyutlu matrislerdir. a_j', **A** matrisinin j. satırıdır. İlgilenilen ağırlıklandırma planının boyutu, bir başka deyişle kullanılacak temel bileşen sayısı ise p'dir.

Temel bileşenlere (4.2) eşitliği ile verilen kayıp fonksiyonunun **X** ve **A** üzerinden enküçüklenmesi ile ulaşılabilir. Kayıp fonksiyonunun enküçüklenmesi homojenlikten sapmanın enküçüklenmesi olarak da açıklanabilir. İkinci bölümde incelenen homojenleştirme analizinde, kayıp fonksiyonlarının en küçüklenmesinde kullanılan algoritma buradaki kayıp fonksiyonu için de kullanılabilir. Algoritma Gifi (1990, s. 152) tarafından ayrıntıları ile açıklanmıştır.

3.2.2. Doğrusal Olmayan Temel Bileşenler Analizi

Değişkenler arasındaki bağımlılık yapısının ortadan kaldırılması (homojenliğin enbüyüklenmesi) ve boyutun indirgenebilmesi için değişkenler yerine sadece değişkenlerin doğrusal bileşenlerinin, yani doğrusal temel bileşenlerinin kullanılması değil değişkenlerin doğrusal olmayan bileşenlerinin, yani doğrusal olmayan temel bileşenlerinin de kullanılması mümkündür. Doğrusal olmayan kombinasyonların kullanıldığı analiz tekniği Doğrusal Olmayan Temel Bileşenler Analizi olarak adlandırılmaktadır. elişen bilgisayar teknolojisi sayesinde uygulanabilirliği oldukça aartan yöntem özellikle son yıllarda büyük ilgi görmektedir.

Doğrusal olmayan temel bileşenler analizinde, bileşenlerin orjinal değişkenlerin doğrusal bileşeni olması gibi bir kısıt söz konusu değildir. Orjinal değişkenlerin herhangi bir fonksiyonu bu amaç için kullanılabilir. Bu fonksiyon veri matrisinin sütunlarının, yani h_j'lerin bir fonksiyonudur ve Φ_j (h_j) ile gösterilebilir. Temel bileşenler analizinin amaçlarını gerçekleştirebilecek fonksiyon aşağıda verilmiş olan kayıp fonksiyonunun en küçüklenmesi ile elde edilebilir.

$$\sigma(x;a;\Phi) = m^{-1} \Sigma_j \, SSQ(x - a_j\Phi_j(h_j)) \tag{3.16}$$

şeklinde tanımlanan kayıp fonksiyonu, x skorları, a_j ağırlıkları ve doğrusal olmayan Φ_j transformasyonları üzerinden enküçüklenmelidir. a_j ağırlıkları ve Φ_j (h_j) fonksiyonu ile birleştirilebilir. Dolayısıyla yukarıdaki kayıp fonksiyonu

$$\sigma(x;\Phi) = m^{-1} \Sigma_j \, SSQ(x - \Phi_j(h_j)) \tag{3.17}$$

şeklinde de ifade edilebilir. Kayıp fonksiyonunun $x'x=1$ veya $SSQ(\Phi_j (h_j)) = 1$ kısıtları altında enküçüklenmesi ile doğrusal olmayan temel bileşenlere ulaşılır.

Doğrusal olmayan temel bileşenler analizinin, doğrusal olmayan bileşenlerin kullanımına olanak vermesi dışında nominal değişkenlerle de işlem yapabilmesi gibi bir özelliği vardır. Bu analiz tekniğinde nominal, ordinal ve sayısal değişkenlerin bir arada kullanılması mümkündür. Böyle durumlarda h_j değişkeni için bir $\mathbf{G}_j$ gösterge matrisinin kullanılması söz konusu olacaktır. $\mathbf{G}_j$ gösterge matrisinin j değişkeninin kategori sayısı olan k_j ile belirlenen sütun sayısı, uygulamada kullanılan birim sayısı olan n'den çok küçük olmalıdır. Bu, ya değişkenlerin başlangıçtan beri birçok ilişki içeren kategorik formda olduğu veya sınırlı sayıda kategoriye ayrıldığı durumlarda mümkündür. G_j'deki k_j tane vektör, bütün olası doğrusal olmayan transformasyonların oluşturduğu n boyutlu uzayın, k_j boyutlu alt uzayını kapsamaktadır. Optimal puanlamadan sonra verideki ilişkiler geçerliliğini sürdürür.

Böylece bazı y_j katsayı vektörleri için $\Phi_j(h_j) = \mathbf{G}_j\, y_j$ yazmak mümkün olacaktır. Bu durumda kayıp fonksiyonu aşağıdaki formu alacaktır.

$$\sigma(x;y) = m^{-1}\, \Sigma_j\, SSQ(\, x - \mathbf{G}_j\, y_j) \tag{3.18}$$

Yukarıdaki ifade çok boyutlu forma da dönüştürülebilir. Çok çözümlü doğrusal olmayan temel bileşenler analizi için kayıp fonksiyonu

$$\sigma(\mathbf{X};\mathbf{Y}) = m^{-1}\, \Sigma_j\, SSQ(\, \mathbf{X} - \mathbf{G}_j\, \mathbf{Y}_j) \tag{3.19}$$

şeklinde yazılır. Bu kayıp fonksiyonu $\mathbf{Y}_j$'lerin kısıtsız, $\mathbf{X}$'in ise $\mathbf{X}'\mathbf{X}= \mathbf{I}$ eşitliğini sağlaması kısıtı ile enküçüklendiğinden homojenleştirme analizindeki kayıp fonksiyonu ile aynıdır. Bilindiği gibi homojenleştirme analizinde $\mathbf{X}$ matrisinin sütun ortalamalarından sapmaları içeren matris olmasını sağlayan $u'\mathbf{X}= 0$ koşulu da söz konusudur. Doğrusal olmayan temel bileşenler analizinde ise bu koşullara ek olarak

$\mathbf{Y}_j = y_j a_j'$

$u'\mathbf{D}_j y_j = 0$

$y_j{}'\mathbf{D}_j y_j = 1$

koşulları da bulunmaktadır (Gifi,1990, s.162).

Doğrusal olmayan temel bileşenler analizi için çeşitli kayıp fonksiyonları ve bu fonksiyonları kullanan değişik bilgisayar programları vardır (Gifi, 1990, s. 166). Çalışmanın uygulama bölümünde doğrusal olmayan temel bileşenler analizinin uygulamasında kullanılan SPSS paket programının PRINCALS kategori seçeneğinden faydalanıldığından bu bölümde sadece adı geçen programda kullanılan kayıp fonksiyonuna yer verilmiştir. Diğer kayıp fonksiyonları ve bilgisayar programları ile ilgili bilgiler Gifi (1990, s. 157) tarafından ayrıntılı olarak verilmiştir.

Çalışmanın uygulama bölümünde, SPSS paket programının PRINCALS kategorisi kullanılarak verilere doğrusal olmayan temel bileşenler analizi de uygulanmıştır. Ulaşılan sonuçlar HOMALS kategorisi yardımıyla uygulanan Homojenleştirme Analizi'nden elde edilen sonuçlar ile olan benzerlikler ve farklılıklar açısından ayrıntılı biçimde incelenmiş ve yorumlanmıştır.

3.3. Kanonik Korelasyon Analizi

Çokdeğişkenli analiz yönteminin pek çoğunda olduğu gibi kanonik korelasyon analizi de bir takım varsayımlar altında anlamlı olmaktadır. Kanonik korelasyon analizinde verinin, bir rastlantı örneği olması yanında çokdeğişkenli normal dağılıma da uyması gerekmektedir. Bilindiği gibi bunun için her bir değişkenin tek başına normal dağılıma sahip olması yetmez aynı zamanda değişkenlerin bileşik dağılımının da normal olması gerekir. İstatistiksel paket programları kullanılarak bu tür varsayımların geçerliliği test edilebilir.

Adından da anlaşılabileceği gibi doğrusal kanonik korelasyon analizi değişkenler arasındaki doğrusal bileşenlerin korelasyonlarını en büyükleme analizidir ve sayısal değişkenler için geliştirilmiştir. Kategorik değişkenlerin söz konusu olduğu ve doğrusal kombinasyonların yanı sıra doğrusal olmayan bileşenlerin de göz önünde bulundurulduğu araştırmalar için ise doğrusal olmayan kanonik korelasyon analizi geliştirilmiştir. Bu bölümde iki tür kanonik korelasyon analizi kısaca açıklanmış ve uygulamalar 4. Bölüm'e bırakılmıştır.

3.3.1. Doğrusal Kanonik Korelasyon Analizi

Çoklu regresyon analizinin genişletilmiş hali olarak da tanımlanabilen kanonik korelasyon analizi, aynı kişilere uygulanmış iki psikolojik test serisi arasındaki ilişkinin araştırılması için Hotelling (1935) tarafından geliştirilmiştir. Ekonomik uygulamaları ise 1942 yılında Waugh tarafından gerçekleştirilmiş ve bu uygulamalar 1946 - 1952 yıllarında Tintner tarafından, 1961 yılında da Kendall tarafından incelenmiştir (Morrison, D. F., 1967, s. 213 - 218). Bilindiği gibi çoklu regresyon analizi bir bağımlı değişken ile birden fazla bağımsız değişken arasındaki ilişkinin incelenmesini sağlayan analiz tekniğidir. Kanonik korelasyon analizi ise iki ya da daha çok değişken kümesi arasındaki ilişkiyi inceler.

Değişkenler arasındaki bağımlılık yapısı ortadan kaldırılmış her bir değişken kümesinden aralarındaki korelasyon en büyük olacak şekilde bulunan doğrusal değişken bileşenlerine kanonik değişkenler adı verilir. Her bir kanonik değişken kendisinden önceki kanonik değişkenden bağımsızdır ve ondan sonra en büyük korelasyona, yani açıklama oranına sahiptir. Bir değişkenin analize olan katkısı kendi kümesinde bulunan diğer değişkenlerden bağımsız olarak verdiği bilgiye ve diğer kümelerde bulunan değişkenlerle olan ilişkisine bağlıdır. En küçük değişken kümesindeki değişken sayısı kadar kanonik değişken elde edilebilir. Fakat boyut indirgeme amacı da olan bu teknikte sadece anlamlı bulunan kanonik değişkenlerin kullanılıp yorumlanması ile değişken kümeleri arasındaki ilişki açıklanabilir. Kanonik değişkenlerin ve dolayısıyla kanonik korelasyon katsayılarının önemlerinin kontrol edilmesinde kullanılabilecek iki tür test vardır: Bartlett testi ve Roy'un en büyük özdeğer yaklaşımı. Bu testler Tatlıdil (1992, s. 225 - 228)'de ayrıntılı biçimde açıklanmıştır.

Kanonik korelasyon analizinde kullanılan formülasyonlar genellikle iki değişken kümesi için verilmiş olmakla birlikte, bu formülasyonlar ikiden fazla değişken kümesi için de geliştirilebilir.

3.3.1.1. Kanonik Değişkenlerin Elde Edilmesi

Kanonik korelasyon analizindeki **h** değişken vektörü iki ya da daha fazla değişken grubundan oluşan parçalı bir vektördür. Başlangıç olarak sadece iki değişken grubu göz önüne alındığında aşağıdaki gibi ifade edilebilir.

$$\mathbf{h} = \begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \end{bmatrix}$$

Birinci değişken grubunda p, ikinci değişken grubunda q değişken bulunması durumunda $\mathbf{h}_1$ ve $\mathbf{h}_2$ alt vektörlerinin boyutları sırasıyla px1 ve qx1 olacaktır. Bu durumda **H** veri matrisi de $\mathbf{H}_1$ ve $\mathbf{H}_2$ alt matrislerinden oluşacak ve aşağıdaki gibi gösterilebilecektir

$\mathbf{H} = [\, \mathbf{H}_1\ \mathbf{H}_2 \,]$

Ulaşılmak istenen doğrusal bileşenler yani kanonik değişkenler $Z_k = \mathbf{h}_k\, a_k$ ile gösterildiğinde, amaç birim varyanslı Z_k değişkenleri arasındaki korelasyonu en büyüklemek olarak açıklanabilir. Tahmin edilebileceği gibi her bir veri kümesi için sonsuz sayıda doğrusal bileşen oluşturulabilir. Oluşturulabilecek bu bileşenlerden, iki küme arasındaki ilişkiyi yorumlamayı sağlayacak olanların seçilebilmesi için bir takım koşulların konulması gerekmektedir. Bu koşullardan ilki, yukarıdaki açıklamalarda da belirtilen, kanonik değişkenlerin birim varyanslı olması koşuludur. İkincisi ise küme içindeki ve kümeler arasındaki kanonik değişkenlerin birbirinden bağımsız olması gerekliliği şeklinde açıklanabilir. Bu koşullar göz önünde bulundurularak yapılacak olan enbüyükleme için Lagranj çarpanları kullanılmaktadır. Kanonik değişkenlerin hesaplanması ile ilgili ayrıntılı açıklamalar herhangi bir çokdeğişkenli analiz kitabından elde edilebilir.

Bulunan kanonik değişkenlerin kendi kümelerindeki ve diğer kümelerdeki orijinal değişkenlerle olan ilişkilerinin açıklanması da önemlidir. Kanonik değişkenlerin orijinal değişkenlerle olan korelasyonu, o orijinal değişkenin kanonik değişkene olan katkısını gösterir.

İkiden fazla değişken kümesinin bulunduğu durumlarda değişken vektörü

$$\mathbf{h} = \begin{bmatrix} \mathbf{h}_1 \\ . \\ . \\ . \\ \mathbf{h}_k \end{bmatrix}$$

ile gösterilir ve iki değişken kümesi için verilen formülasyonlar bu durum için genelleştirilebilir (Tatlıdil, 1996, s. 217).

Homojenlik fikrinin temel olarak alınması durumunda kanonik korelasyon analizi için uygun olan kayıp fonksiyonu aşağıda ki gibi yazılabilir:

$$\sigma(\mathbf{X}, \mathbf{A}) = K^{-1} \sum_k SSQ(\mathbf{X} - H_k A_k) \tag{3.20}$$

Burada K mevcut değişken küme sayısını verirken A_k vektörleri kanonik değişkenlerin katsayılarını içermektedir. Daha önce de değindiğimiz koşullar bu kayıp fonksiyonu için de geçerlidir ve aşağıda olduğu gibi verilebilir.

$u'\mathbf{X} = 0$

$\mathbf{X}'\mathbf{X} = 1$

Kayıp fonksiyonundaki $\mathbf{H}_k$ matrisini oluşturan sütunların ortalamadan sapmalar şeklinde ifade edilmesi varsayımı mevcuttur (Gifi, 1990, s. 193).

3.3.2. Doğrusal Olmayan Kanonik Korelasyon Analizi

Önceki bölümde de belirtildiği gibi klasik yani doğrusal kanonik korelasyon analizi, sayısal değişkenler arasındaki ilişki doğrusal bileşenlerle açıklamaya çalışmaktadır. Birçok bilim dalında ise sayısal yerine kategorik veya sırasal değişkenlerin kullanımı söz konusu olup bu değişkenlerin doğrusal olmayan bileşenlerinin de göz önüne alınması gerekebilmektedir. Doğrusal olmayan kanonik korelasyon analizi bu tür durumlar için geliştirilmiş olup birçok bilgisayar paket programına da dahil edilmiş durumdadır. Çalışmanın uygulama bölümünde kullanılan SPSS paket programının OVERALS kategorisi de bu programlardan biridir.

Homojenlik fikri yardımıyla açıklanan doğrusal olmayan kanonik korelasyon analizinde daha önce sözü edilen gösterge matrislerinden faydalanılmaktadır. Bu analiz yönteminde kullanılan kayıp fonksiyonu aşağıdaki gibi verilebilmektedir.

$$\sigma\,(\mathbf{X},\mathbf{Y}) = \mathbf{K}^{-1} \sum_{\mathbf{k}} \mathbf{SSQ}\left[\mathbf{X} - \sum_{j,J_k} \mathbf{G}_j \mathbf{Y}_j\right] \tag{3.21}$$

Bu kayıp fonksiyonunda m değişkenin $\mathbf{h_k}$ ile belirtilen K kümeye bölündüğü düşünülmektedir. $\mathbf{Y_j}$'ler puanlama matrislerini belirtmekte kullanılmaktadır. Kayıp fonksiyonunda kullanılan **X** matrislerinin aşağıdaki eşitliklerle özetlenen koşulları sağlaması gerekmektedir:

$\mathbf{u'X = 0}$

$\mathbf{X'X = I}$

Doğrusal olmayan kanonik korelasyon analizi bilgisayar programları yardımı olmaksızın uygulanamayacak kadar karmaşık olmasından dolayı son zamanlarda ilgi çekmeye başlamış yöntemlerden biridir. Birbiri ile ilişkili değişkenlerin söz konusu olduğu araştırmalarda değişken grupları oluşturularak incelemenin kolaylaşması sağlanır. Değişkenlerin sırasal, sınıfsal veya sayısal olarak kullanılabilmesi tekniğin uygulanabilirliğini arttırmaktadır. Kanonik korelasyon analizinin homojenleştirme ve temel bileşenler analizlerinden farkı değişkenlerin gruplanmasıdır. Değişkenler arasındaki ilişki yerine grup içindeki değişkenlerin bileşenleri arasındaki ilişkiye önem verilir.

BÖLÜM IV

UYGULAMA

Bu bölümde önceki bölümlerde açıklanan çokdeğişkenli analiz tekniklerinin uygulanmasına aynı sıra ile yer verilmiştir. Her alt kısmın sonunda o kısımda elde edilen sonuçlar daha önceki kısımlarda elde edilen sonuçlarla karşılaştırılmış, benzerlik ve farklılıklar açıklanmıştır.

Uygulamalarda Renault Mais Motorlu Araçlar A.Ş.'nin bayileri üzerinde yaptırdığı araştırma sonuçları kullanılmıştır. Bayiler sadece numaralarla belirtilmiş isimler saklı tutulmuştur. Yaptırılan araştırmanın amacı bayi performansını belirlemektir. Bu amaçla her bayiden müşteri listesi istenmiş, bu listelerden rasgele seçilen müşterilere telefonla ulaşılmış ve 18 soru içeren anket formunun cevaplanması sağlanmıştır. Bununla beraber uygulamamızda sadece şirketi en çok ilgilendiren memnuniyet sorularına yer verilmiştir.

Sorular araç tipi, servise geliş nedeni, genel memnuniyet ve bakım onarım kalitesi memnuniyeti olarak sıralanabilir. Araç tipleri aracın motor tipine göre 13 grupta, servise geliş nedeni ise mekanik (1), kaporta (2) ve diğer (3) olmak üzere 3 grupta toplanmıştır. Genel memnuniyet için hiç memnun değilim (1), memnun değilim (2), memnunum (3) ve çok memnunum (4) olmak üzere 4 seçenek sunulmuştur. Bakım onarım kalitesi için ise cevaplar ise memnunum (1), memnun değilim (2) ve diğer (3) olmak üzere 3 seçenekle sınırlandırılmıştır. Grafiklerde servise geliş nedeni mekanik-kaporta olarak, genel memnuniyet genel soru olarak ve bakım onarım kalitesi memnuniyeti ise soru 8 olarak kısaltılmıştır. Anket formunun bir örneği Ek 1'de sunulmuştur.

Aylık olarak toplanmış olan verilerin 1998 yılı Ocak ayına ait olanları alınmış diğer aylara ait veriler ise sadece kontrol amaçlı olarak kullanılmıştır.

Uygulamalar SPSS paket programının 8.0 versiyonu ile yapılmış, veri düzenlemesi için Excel'den faydalanılmıştır.

4.1. Uygunluk Analizi

2.Bölüm'de de açıklandığı gibi uygunluk analizi iki değişken veya özelliğin kategorilerinin, aralarındaki uzaklıklar benzerliklerini verecek şekilde, grafik üzerinde gösterimini sağlayan bir yöntemdir. Uygunluk analizi uygulamaları için SPSS paket programının CATEGORİES bölümünde yer alan CORRESPONDENCE seçeneği kullanılmıştır. Uygulamanın yapılabilmesi için bir nesne veya bireye karşılık gelen değişken veya özellik değeri kategori numarası olacak şekilde sütunların değişken, satırların ise bireyleri göstermesi yeterlidir. CORRESPONDENCE seçeneği çapraz tabloyu oluşturur ve kategori koordinatlarını hesapladıktan sonra hareketsizlik ve tekil değerlerle birlikte grafikleri de sunar.

Birkaç çeşit grafik elde etmek mümkündür. En çok kullanılanlar sadece satır skorlarının saçılım grafiğini veren RPOINTS, sadece sütun skorlarının saçılım grafiğini veren RPOINTS ve satır ve sütun skorlarının birlikte saçılımını veren BIPLOT seçeneğidir.

Araştırmanın niteliğine ve amacına göre normalizasyonun da belirlenmesi gerekir. Simetrik normalizasyon (SYMMETRICAL) satır ve sürunlar arası benzerlik ve farklılıkların önemli olduğu araştırmalarda kullanılan seçenektir. Normalizasyon seçimi kullanıcı tarafından belirlenmediğinde program simetrik normalizasyonu (default) kullanır.

Diğer normalizasyon çeşitleri ise PRINCIPAL, RPRINCIPAL ve CPRINCIPAL olarak adlandırılmaktadır. Hem satırların kendi aralarındaki benzerlikleri hem de sütunların kendi aralarındaki benzerlikleri gerekli olduğunda PRINCIPAL seçeneği kullanılır. Bu seçenekte satır ve sütunlar arası benzerliklerden söz edilemez. Satırların kendi aralarındaki ilişkinin gösterimi için RPRINCIPAL, sütunların kendi aralarındaki ilişkinin gösterimi için CPRINCIPAL seçeneği uygundur.

Grafiklerde ki-kare ölçümünün yanı sıra öklit ölçümünün de kullanılması mümkündür. Bu durumda ulaşılan grafik çok boyutlu ölçekleme tekniği ile elde edilen grafikle çok benzer olmaktadır.

Uygulamalarımızda simetrik normalizasyon, biplot ve ki-kare ölçüm seçenekleri yanında burada üzerinde durulmamış olan diğer komutlar için de programın uygun gördüğü seçenekler (default) kullanılmıştır.

Bu kısımdaki uygulamalarla iki değişken olduğunda kategoriler arası ilişkiyi belirlemeyi sağlayan ve gruplamayı kolaylaştıran uygunluk analizi grafikleri sunulmuş ve yorumları yapılmıştır. Daha önceki bölümlerde de belirtildiği gibi çok sayıda ve benzer kategorinin söz konusu olması halinde, kategorileri gruplamayı kolaylaştırmak için lojistik regresyon analizindeki ayrım fonksiyonundan elde edilen doğrudan faydalanılmıştır. Bunun için kategorilerin birinci ve ikinci boyuttaki değerleri bağımsız değişkenler olarak alınmaktadır. Yapılan araştırmaya uygun olacak şekilde kategorilerin iki gruba toplanmasında uygun özelliklerden veya daha önceden edinilmiş olan izlenimlerden faydalanılmaktadır. Kategorilerin grup numaralarını içeren değişken ise bağımlı değişken olarak kullanılabilmektedir. Birinci ve ikinci boyut değerlerinin bir dosyada toplanmasını sağlayan komut dizimi (syntax) Ek 2'de sunulmuştur.

Çalışmamızda her bayii için genel memnuniyet sorusuna verilen cevapların ortalaması alınmıştır. Daha önceki deneyimler ve şirket değerlendirmeleri de bunlara eklenerek bayiiler 4 gruba ayrılmıştır. Kategorilerin grup numaralarını içeren değişkenden 'bayii sınıflandırma değişkeni' olarak söz edilecektir. Benzer şekilde gerek duyulan diğer değişkenler için de sınıflandırma değişkeni oluşturulmuştur.

İlk uygulama genel memnuniyet sorusu (Genel Soru) ile bakım onarım kalitesi memnuniyeti (Soru 8) arasındaki uygunluk analizi sonucunu içermektedir. Sadece bu uygulama için CORRESPONDENCE çıktılarının büyük kısmı ayrıntılı biçimde Tablo 4.1 ve Grafik 4.1'de sunulmuştur. Diğer uygulamalarda sadece gerekli görülen grafiklere ve yorumlarına yer verilmiştir.

Tablo 4.1. Genel Memnuniyet (Genel Soru) Ve Bakım Onarım Kalitesi Memnuniyeti (Soru 8) İçin Uygunluk Analizi Sonuçları

Tablo 4.1.1. Çapraz Tablo (Uygunluk Analizi Tablosu)

	SORU 8			
GENEL SORU	**1**	**2**	**3**	**Satır Toplamı**
1	26	41	3	70
2	31	67	10	108
3	350	91	41	482
4	378	22	11	411
Sütun Toplamı	785	221	65	1071

Genel memnuniyet (genel soru) ve bakım onarım kalitesi memnuniyeti (soru 8) için 1071 kişiden alınan cevaplar Tablo 4.1.1'de çapraz tablo biçiminde sunulmuştur.

Tablo 4.1.2. Boyutlar Hareketsizlik ve Ki –Kare Değerleri

Boyut	Tekil Değer (singular value)	Hareket-Sizlik (Inertia)	Ki-Kare	Sig.	Hareketsizlik Yüzdesi (Proportion of Inertia)	
					Boyutlarda	**Kümülatif**
1	0,484	0,234			0,957	0,957
2	0,103	0,011			0,043	1,000
Toplam		0,244	261,831	0,000(a)	1,000	1,000
(a) Serbestlik derecesi : 6						

Tablodan da görülebileceği gibi toplam hareketsizlik ve ki-kare değerleri sırasıyla 0,244 ve 261,831 olarak hesaplanmıştır. Bu değerler (2.1) no'lu eşitlik kullanılarak da elde edilebilmektedir. Hesaplanan ki-kare değerinin 6 serbestlik dereceli ki-kare tablo değeri (α=0.05 için 12.592) ile karşılaştırılması, satır ve sütun değişkenlerinin bağımsız olduğunun söylenebilmesi için yeterli neden olmadığını ortaya koymaktadır. Dolayısıyla kategorilerin ilişkilerinden bahsetmek anlamsız olmayacaktır. Tekil değerlere ise (2.2) no'lu eşitlikte verilmiş olan matris çarpımının tekil değer ayrıştırması ile ulaşılabilmektedir. Boyuttaki hareketsizlik ise o boyutun tekil değerinin karesine eşittir. Hareketsizliğin % 95,7'sinin birinci boyutla açıklandığı görülmektedir. Dolayısıyla birinci boyutun ayrımsamada ikinci boyuta göre çok daha başarılı olduğu söylenebilir.

Tablo 4.1.3. Satır Noktalarının Özellikleri (Overview Row Points)

GENEL SORU	**Ağırlık** (Mass)	**Koordinatlar** (Score in Dimension)		**Hareketsizlik** (Inertia)
		1	**2**	
1	0,065	-1,318	-0,626	0,058
2	0,101	-1,521	-0,030	0,113
3	0,450	0,023	0,329	0,005
4	0,384	0,597	-0,271	0,069
Toplam	1,000			0,244

Ağırlık sütunu, satır toplamlarının genel toplama bölünmesiyle elde edilen satır ağırlıklarını vermektedir. Üçüncü ve dördüncü sütunlarda verilen koordinatlar yardımıyla kategoriler grafiğe yerleştirilmektedir. Son sütundaki hareketsizlik değerleri ise toplam hareketsizliğin satır kategorilerine göre dağılımını vermektedir. Bu değerler (2.1) no'lu eşitlikte her satır için hesaplanan toplam terimleridir. Genel hareketsizliğin büyük bölümünün 2 numaralı seçenek tarafından yaratıldığı görülmektedir.

Genel soru kategorilerinin boyut hareketsizliğine katkısı Tablo 4.1.4'de verilmiştir. Bir noktanın boyut hareketsizliğine katkısı, adı geçen noktanın ağırlıklandırılmış koordinat karesinin ağırlıklandırılmış koordinat kareleri toplamına bölümü ile elde edilir. Burada bahsedilen ağırlıklar Tablo 4.1.3'ün ikinci sütununda yer alan satır ağırlıklarıdır. Tablo 4.1.4 incelendiğinde birinci boyutun hareketsizliğine en büyük katkının genel memnuniyet sorusunun 2 ile gösterilen 'memnun değilim' seçeneği tarafından yapıldığı görülmektedir. İkinci boyutta ise aynı ifade 3 numaralı 'memnunum' seçeneği için kullanılabilmektedir.

Tablo 4.1.4. Genel Soru Kategorilerinin Boyut Hareketsizliğine Katkısı

GENEL SORU	**Noktalarin Boyut Hareketsizliğine Katkısı** (Contributions of Point to Inertia of Dimension)	
	1	2
1	0,235	0,250
2	0,482	0,001
3	0,001	0,474
4	0,282	0,275
TOPLAM	1,000	1,000

Tablo 4.1.5. Boyutların Genel Soru Kategorilerinin Hareketsizliğine Katkısı

GENEL SORU	**Boyutların Nokta Hareketsizliğine Katkısı** (Contributions of Dimension to Inertia of Point)		
	1	2	**TOPLAM**
1	0,954	0,046	1,000
2	1,000	0,000	1,000
3	0,023	0,977	1,000
4	0,958	0,042	1,000

Tablo 4.1.5 ile kategorilerdeki hareketsizliğin boyutlara göre dağılımı (boyutların nokta hareketsizliğine katkıları) verilmiştir. Boyutların nokta

hareketsizliğine katkıları hesaplanırken de ağırlıklandırılmış koordinat karelerinin boyuttaki yüzdelerinden faydalanılır. Yorumlama kolaylığı sağlamak açısından, paket program bu yüzdeleri toplamları 1 olacak şekilde hesaplayıp tablo halinde sunmaktadır.

Sütun noktalarının özellikleri Tablo 4.1.6'da verilmiştir. Ağırlık sütunu sütun toplamlarının genel toplama bölünmesiyle elde edilen sütun ağırlıklarını vermektedir. Sütun kategorilerinin koordinatları üçüncü ve dördüncü sütunlarda verilmektedir. Son sütundaki hareketsizlik değerleri ise toplam hareketsizliğin sütun kategorilerine göre dağılımını vermektedir. Bu değerlere 2.1 no'lu eşitlikte satır toplamları yerine sütun, sütun toplamları yerine de satır toplamları konularak ulaşılabilmektedir. Adı geçen tabloya göre genel hareketsizliğe en büyük katkının 2 numaralı sütun tarafından yapıldığı söylenebilmektedir.

Tablo 4.1.6. Sütun Noktalarının Özellikleri (Overview Column Points)

SORU 8	**Ağırlık** (Mass)	**Koordinatlar** (Score in Dimension)		**Hareketsizlik** (Inertia)
		1	**2**	
1	0,733	0,401	-0,057	0,057
2	0,206	-1,316	-0,165	0,173
3	0,061	-0,370	1,248	0,014
Toplam	1,000			0,244

Bakım onarım kalitesi memnuniyetini belirten Soru 8 kategorilerinin boyutlardaki ve boyutların bu kategorilerdeki hareketsizliklere katkısı 4.1.7 ve 4.1.8 no'lu tablolarda verilmiştir. Tablolardaki değerler Tablo 4.1.4 ve Tablo 4.1.5'deki değerlerle aynı şekilde hesaplanmaktadır. Tablo 4.1.7 incelendiğinde birinci boyuta en büyük katkının ikinci kategori tarafından, ikinci boyuta en büyük katkının ise üçüncü kategori tarafından yapıldığı görülmektedir.

Tablo 4.1.7. Soru 8 Kategorilerinin Boyut Hareketsizliğine Katkısı

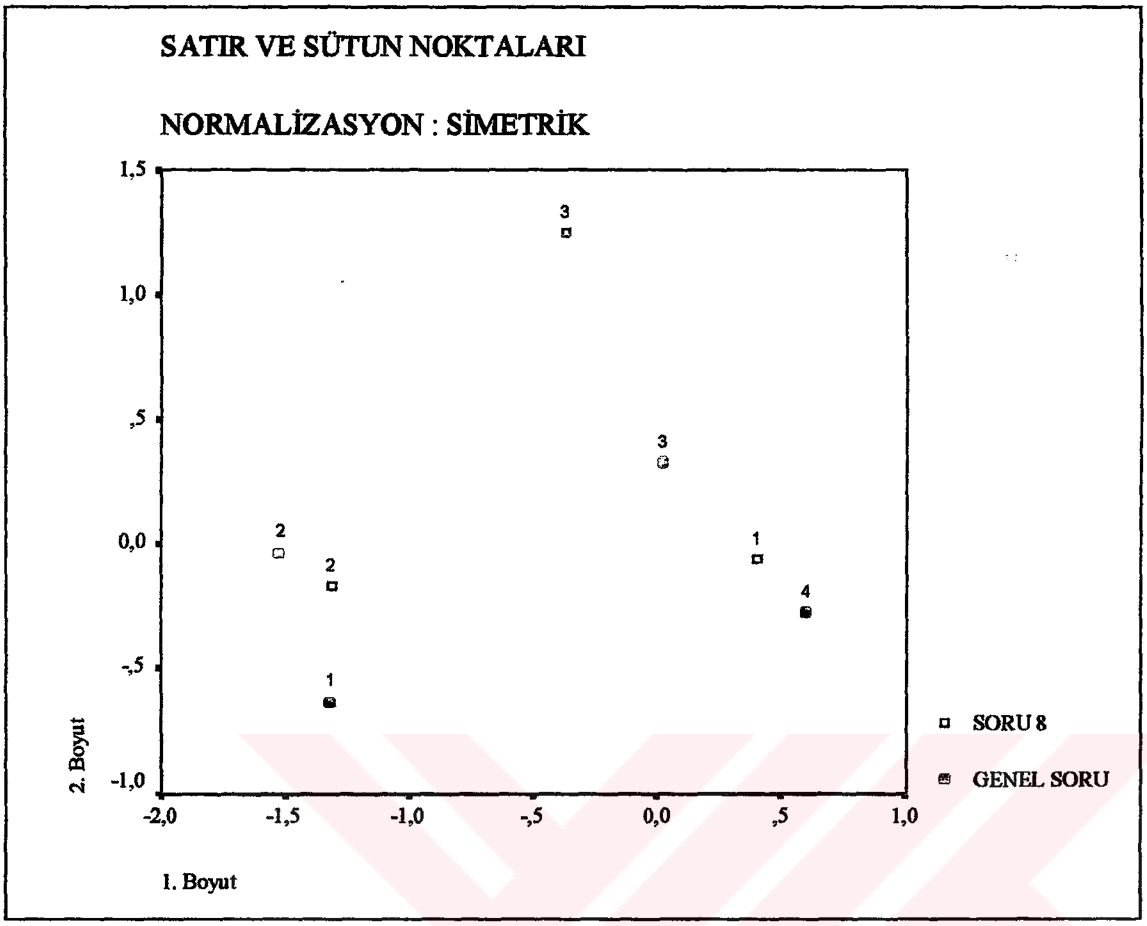
SORU 8	Noktalarin Boyut Hareketsizliğine Katkısı (Contributions of Point to Inertia of Dimension)	
	1	2
1	0,244	0,023
2	0,739	0,055
3	0,017	0,922
TOPLAM	1,000	1,000

Boyutların sütun kategorilerinin yani bakım onarım kalitesi memnuniyet düzeylerinin hareketsizliğine katkısı Tablo 4.1.8'de verilmiştir. Birinci boyutun 1 ve 2 numaralı sütunlardaki, ikinci boyutun ise üçüncü kategorideki hareketsizliği daha iyi açıkladığı görülmektedir.

Tablo 4.1.8. Boyutların Soru 8 Kategorilerinin Hareketsizliğine Katkısı

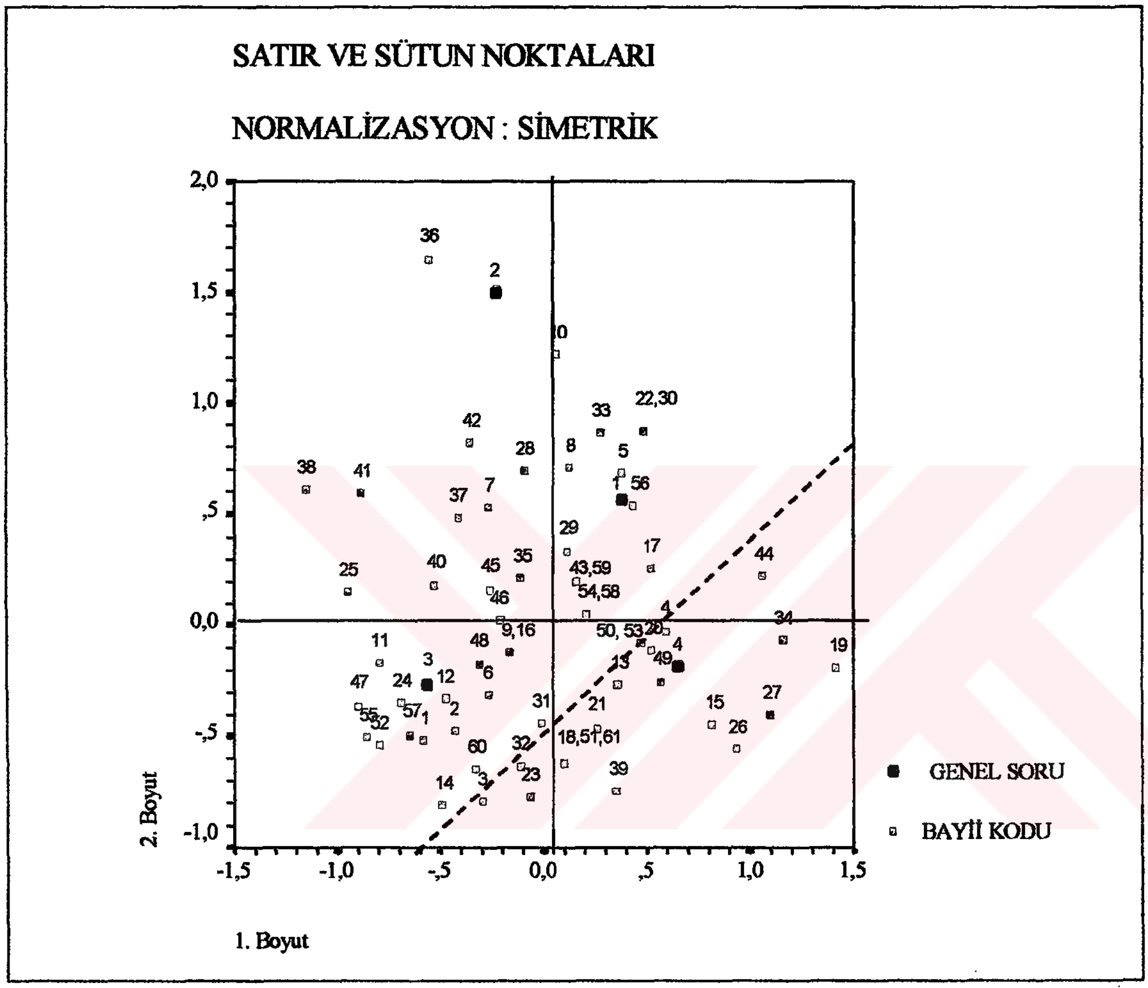
SORU 8	Boyutların Nokta Hareketsizliğine Katkısı (Contributions of Dimension to Inertia of Point)		
	1	2	TOPLAM
1	0,996	0,004	1,000
2	0,997	0,003	1,000
3	0,293	0,707	1,000

Çok fazla kategorisi bulunan değişkenler söz konusu olduğunda yukarıdaki tabloların incelenmesi bu kadar da kolay olmamaktadır. Bunun yanında aynı bilgilerin bir çoğuna daha kolay incelenebilen uygunluk analizi grafikleriyle de ulaşmak mümkündür. Bir grafiğin incelenmesiyle bu bilgilere ulaşılabilmesi, çok sayıda değişken arasındaki ikili ilişkilerin incelenmesi gereken uygulamalarda çok yararlı olmaktadır.



Şekil 4.1: Genel Memnuniyet ve Bakım Onarım Kalitesi Memnuniyeti İçin Uygunluk Analizi Grafiği

İlk uygulamaya ait olan genel memnuniyet ve bakım onarım kalitesi memnuniyeti için uygunluk analizi grafiği Şekil 4.1'de verilmiştir. Grafik incelendiğinde bakım onarım kalitesi memnuniyetinin (soru8) 2 numaralı 'memnun değilim' seçeneği ile genel memnuniyetin 1 ve 2 numaralı 'hiç memnun değilim' ve 'memnun değilim' seçeneklerinin birbirine çok yakın yerleştiği görülebilmektedir. Aynı ifadeler Soru 8'in 'memnunum' (1) ve genel memnuniyetin 'memnunum' (3) ile 'çok memnunum' (4) seçenekleri için de kullanılabilir. Her iki soru da memnuniyet sorusu olduğundan bu beklenen bir sonuçtur. Sorular arasında bir çelişki olmadığını ve genel memnuniyetin bakım onarım kalitesi memnuniyeti ile yakından ilişkili olduğu gerçeğini göstermektedir. Soru 8'in 'diğer' (3) seçeneğinin diğer kategorilerden uzağa yerleştirilmiş olması da gayet mantıklı ve anlamlıdır. Genel memnuniyet sorusunda kararsızlığa yer verilmemişken Soru 8'in 3 numaralı seçeneği bakım onarım kalitesi hakkındaki kararsızlığı belirtmektedir.



Şekil 4.2: Genel Memnuniyet Ve Bayii Kodu İçin Uygunluk Analizi Grafiği

Genel soru ile bayii kodu için uygunluk analizinden elde edilen grafik Şekil 4.2 verilmiştir. Bu grafik yardımıyla bayilerin bir bütün olarak değerlendirilmesi durumunda nasıl sınıflandırılabileceği görülmektedir. Genel sorunun 2 ve 3 numaralı (memnun değilim, memnunum) seçeneklerinin grafiğin solunda ve 1 ve 4 numaralı (hiç memnun değilim, çok memnunum) seçeneklerinin sağda yer alması nedeniyle birinci boyutun en iyi ve en kötülerle, yani uç noktalarla orta karar bayileri ayırdığı söylenebilir. Benzer şekilde genel sorunun memnuniyetsizliği belirten 1 ve 2 numaralı seçeneklerinin üst kısımda, memnuniyeti belirten 3 ve 4 numaralı seçeneklerinin ise grafiğin alt kısmında bulunması dolayısıyla ikinci boyutun iyilerle kötüleri ayırmakta daha başarılı olduğu söylenebilir.

Daha özele inerek örnek vermek gerekirse 'çok memnunum' seçeneğine yakın konuşlanmış olan 4, 49, 20 numaralı bayilerin müşteriler tarafından bütün olarak çok başarılı bulundukları, 'hiç memnun değilim' seçeneğine yakın yerleştirilmiş olan 56, 5, 8, 33, 20, 30 numaralı bayilerin ise bütün olarak başarısız bulundukları söylenebilmektedir. Genel sorunun 3 numaralı 'memnunum' seçeneği çevresindeki (12, 24, 2, 6 gibi) bayilerin de başarılı bayiler arasında sayılması gerekmektedir. Memnuniyetsizliğe yakın yerleştirilmiş olmalarından dolayı 36 ve 10 numaralı bayilerin bazı yönlerden başarılı bulunmadığı ve bu yönlerin düzeltilmesi halinde başarılı bayiler arasına geçebilecekleri görülmektedir. Merkezden ve diğer kategorilerden hayli uzakta bulunan 38, 41, 25 numaralı bayii müşterilerinin bayiinin başarısı hakkında kararsız olduğu fakat başarısızlığa daha yakın gördüğü de söylenebilmektedir.

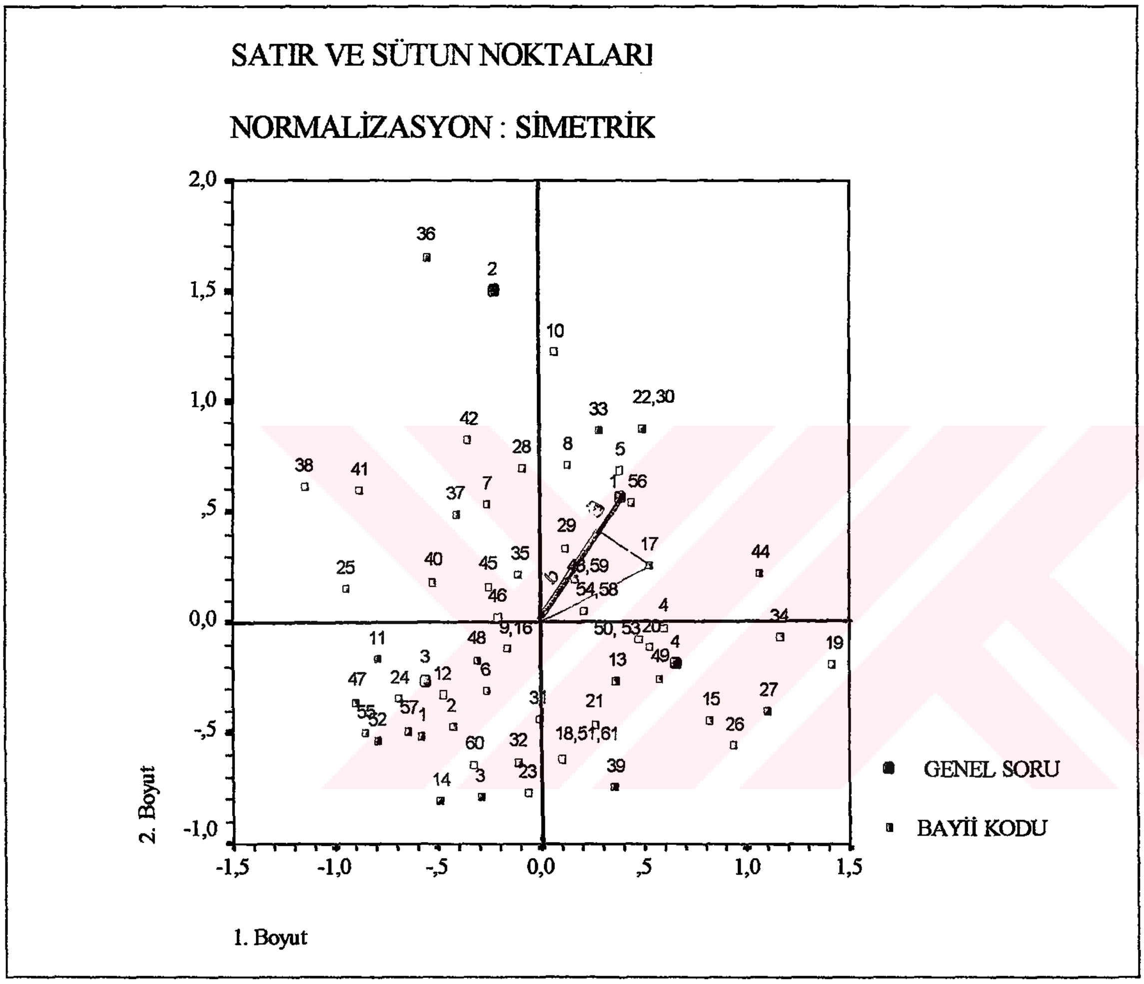
Bayilerin çoğu gayet rahatlıkla sınıflandırılabilecekken, özellikle 1 ve 4 numaralı genel soru kategorileri arasında yer alan bayilerin (50, 4, 17, 54, 58, gibi) hangi gruba dahil edileceğine karar vermek zordur. Bu kategorilerin birbirlerinden çok farklı olması zorluğu bir kat daha arttırmaktadır. Bu gibi durumlarda lojistik regresyon analizi sınıflandırmayı kolaylaştırır. Önemli olan birbirine çok yakın yerleştirilmiş olan başarılı ve başarısız bayileri birbirinden ayırmaktır. Grafiğin sağ alt kısmında genel memnuniyetin 4 numaralı kategorisi çevresinde yerleşik olan ve başarılı bulunan bayiler teşhis edilmeye daha uygun gözükmektedir. Dolayısıyla zaten var olan bayii sınıflandırma değişkeninde 4. tip olarak belirlenen bayiler bir gruba diğerleri bir grupta yer alacak şekilde iki grup yapılır. Koordinatlar bağımsız değişkenler, iki gruplu bayii sınıflandırma değişkeni de bağımlı değişken olarak alınır ve lojistik regresyon analizi uygulanır. Ulaşılan lojistik fonksiyon aşağıda verilmiştir.

$$f = -2.8661 + 5.3825\,(\text{Boyut1}) - 6.7513\,(\text{Boyut2}) \quad (4.1)$$

Bu fonksiyonun kritik atama noktası (cut off point) olan 0.5'e eşitlenmesi ile elde edilen doğru Şekil 4.2'deki grafikte kesikli çizgi ile gösterilmiştir. Programın kritik atama noktası için otomatik olarak seçtiği değer 0.5'tir. Bu değerin isteğe ve araştırmanın amacına göre belirlenmesi mümkündür. Kritik atama noktasının belirlenmesine Ek 2'de verilen komut dizinlerinde değinilmiştir.

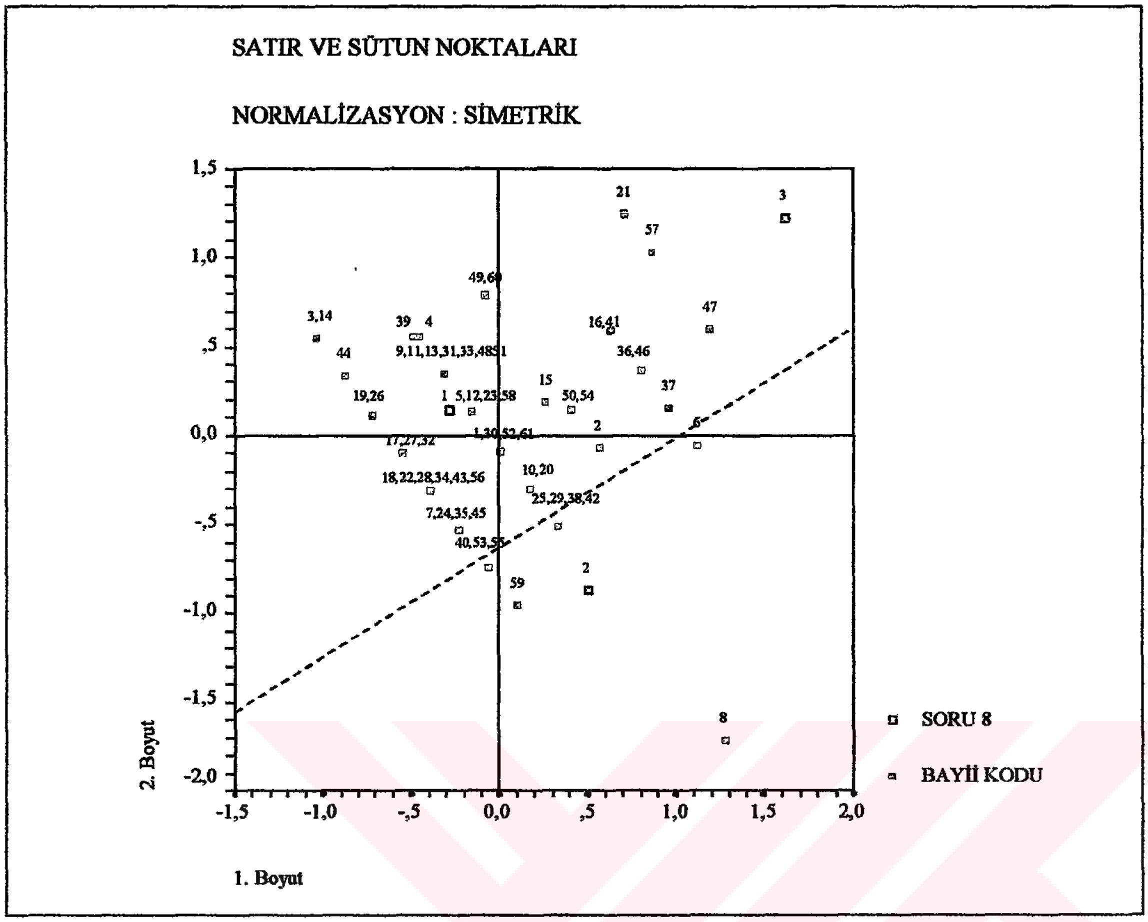
Lojistik fonksiyonundan elde edilen doğru, iyi bayilerle diğerlerini birbirinden ayırmaktadır. Böylece hangi gruba atanacağına karar verilemeyen 4 ve 50 numaralı bayiler başarılı 17, 54 ve 58 numaralı bayiler ise başarısız olarak nitelendirilebilir.

Bir profilin ortalama profilden farkı grafikten yaklaşık olarak elde edilebilir. Bunun için ilgilenilen değişkenin i kategorisine orijinden bir doğru çizilir. Diğer değişkenin kategorilerinden yine ilgilenilen bir tanesi (j) seçilerek bu doğru üzerine izdüşümü alınır. Bu izdüşümün orijine uzaklığı ile i kategorisinin orijine olan uzaklığı çarpımından (scalar product) elde edilen değer, j kategorisinin i kategorisindeki profili ile aynı kategorideki ortalama profil arasındaki farkı verecektir. Aşağıdaki grafikte (Şekil 4.3) bu yöntem uygulanarak 1 numaralı genel soru kategorisinde 17 numaralı bayiinin ortalama bayii profilinden farkı araştırılmıştır.



Şekil 4.3: Profil Farklarının İzdüşüm Yardımıyla İncelenmesi

Grafikte (a) ile gösterilen 1 numaralı genel soru kategorisinin orijine uzaklığı 0.74 birimdir. 17 numaralı bayiinin bu doğruya izdüşümü alınmıştır. İki doğrunun kesişim noktasının orijine olan uzaklığı (b) ise 0.48 birimdir. Bu iki değerin çarpımı olan 0.3626 'da 17 numaralı bayiinin 1 numaralı genel soru kategorisinde bulunan bayilerin ortalamasından farkını vermektedir.

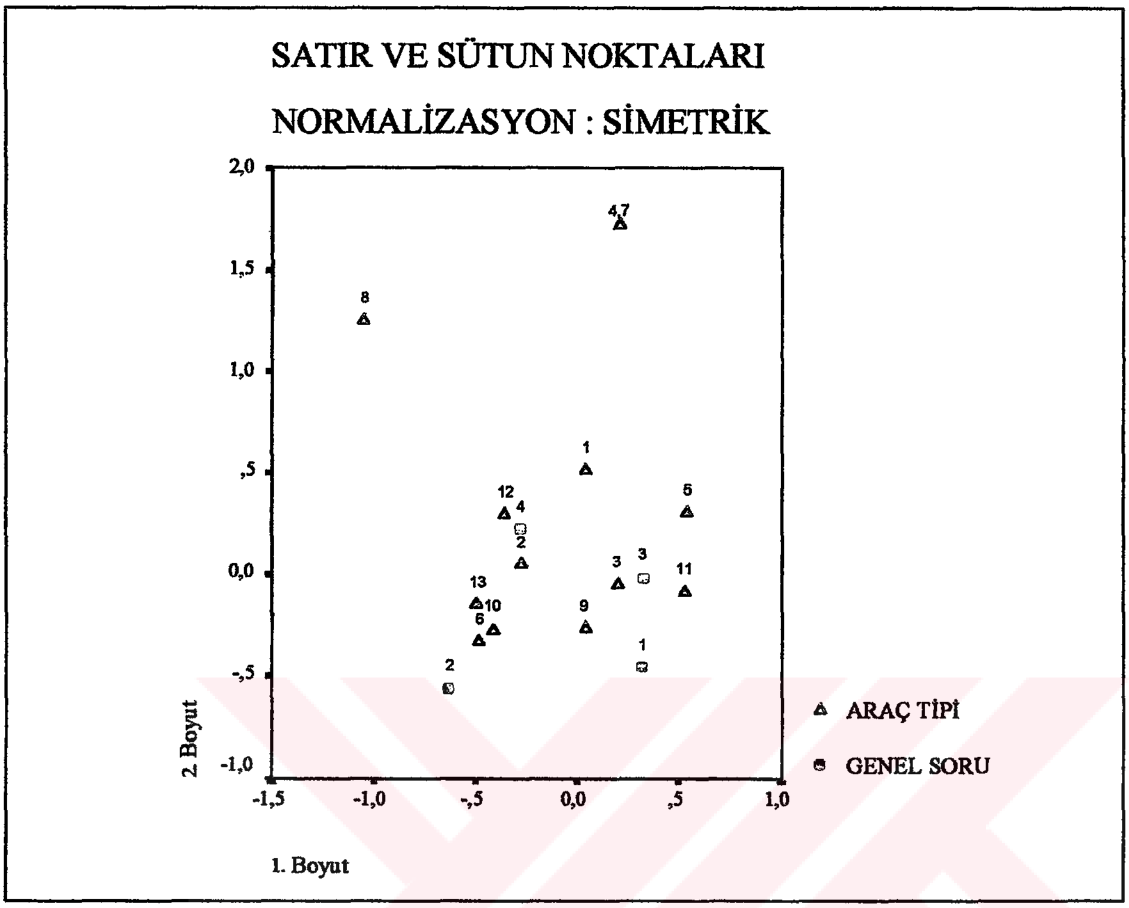


Şekil 4.4: Bakım Onarım Kalitesi Memnuniyeti (Soru 8) ve Bayii Kodu İçin Uygunluk Analizi Grafiği

Şekil 4.4'deki grafik bayilerin sadece bakım onarım kalitesi açısından değerlendirilmesi durumunda nasıl sınıflandırılacaklarını gözler önüne sermektedir. Memnuniyeti belirten 1 seçeneğinin grafiğin sol üst kısmında, memnuniyetsizliği belirten 2 numaralı seçeneğin sağ alt kısmında yer almasından dolayı hiçbir boyutun ayrımsamada diğerine göre daha başarılı ya da başarısız olarak nitelenmesi mümkün değildir. Kararsızlığı belirten 3 numaralı seçeneğin her iki boyuttan neredeyse aynı miktarda uzağa yerleşmiş olması nedeniyle de boyutların üstünlüğünden söz edilemez. Başka bir deyişle boyut hareketsizlikleri birbirine çok yakın değerlerdir. Bakım onarım kalitesi açısından ne başarılı ne de başarısız bulunan bayiler 3 seçeneğine en yakın olan 21, 57 ve 47 numaralı bayilerdir. Başarılı bulunan bayilerden birkaçı (5, 12, 23, 58) ve (9, 11, 13, 31, 33, 48, 51) olarak sıralanabilir. Başarısız bayiler arasında ise 8, 59, (40, 53, 55) ve (25, 29, 38, 42) sayılabilir. Özellikle (7, 24, 35, 45) ve (10, 20) gibi 1 ve 2 numaralı seçenekler arasında bulunan bayilerin sınıflandırılması zordur. Soru 8'in 1 numaralı 'memnunum' seçeneği, 2 ve 3 numaralı 'memnun değilim' ve 'diğer' seçeneğine kıyasla merkeze daha yakın yerleşmiştir. Bu durumda saçılımı daha çok memnuniyetsizliğin oluşturduğu söylenebilir. Dolayısıyla memnuniyetsiz olunan bayilerin diğerlerinden ayrılması gerektiği düşünülmüştür. Bunun için yine lojistik regresyon analizi kullanılmalıdır. Bağımlı değişken olarak, çok memnun (4) ve memnun (3) şeklinde sınıflanan bayilerin bir gruba diğerlerinin ise ikinci gruba ayrılmasını sağlayacak şekilde düzenlenen bayii sınıflama değişkeni kullanılmıştır. Böylece başarısız bulunan, yani memnun olunmayan bayilerin diğerlerinden ayrılması amaçlanmıştır. Bağımsız değişkenler daha önce de açıklandığı gibi bayilerin birinci ve ikinci boyuttaki koordinat değerleridir. Lojistik regresyon yardımıyla ulaşılan fonksiyon aşağıda verilmiştir. Bu fonksiyonun belirttiği doğru ise grafik üzerinde kesikli çizgi ile gösterilmiştir. Kritik değer önceden olduğu gibi 0.5 olarak alınmıştır.

$$f = -0.4571 + 0.9457\,(\text{Boyut } 1) - 1.5129\,(\text{Boyut } 2) \qquad (4.2)$$

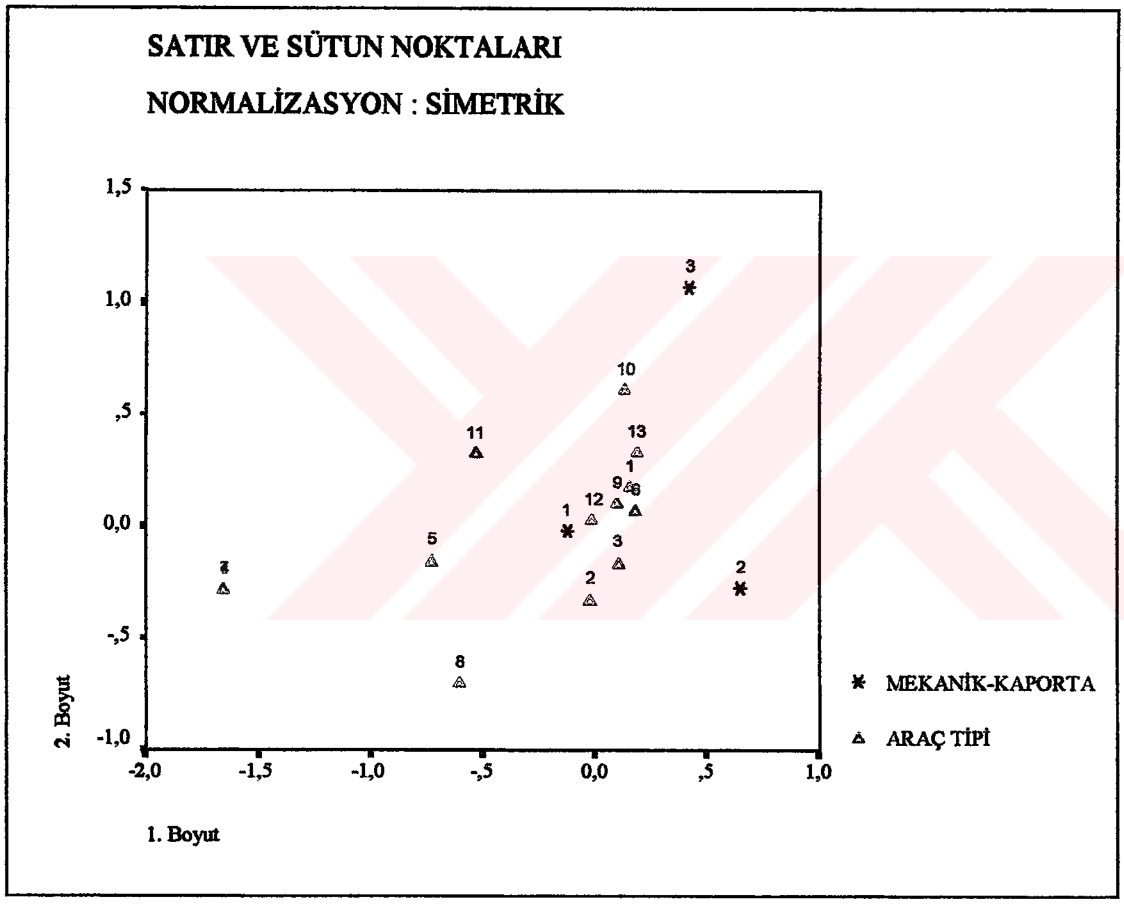
Çizilen bu doğru sayesinde (7, 24, 35, 45) ve (10, 20) bayilerinin başarılı 6 numaralı bayiinin ise başarısız bulunduğu söylenebilmektedir.



Şekil 4.5: Genel Memnuniyet ve Araç Tipi İçin Uygunluk Analizi Grafiği

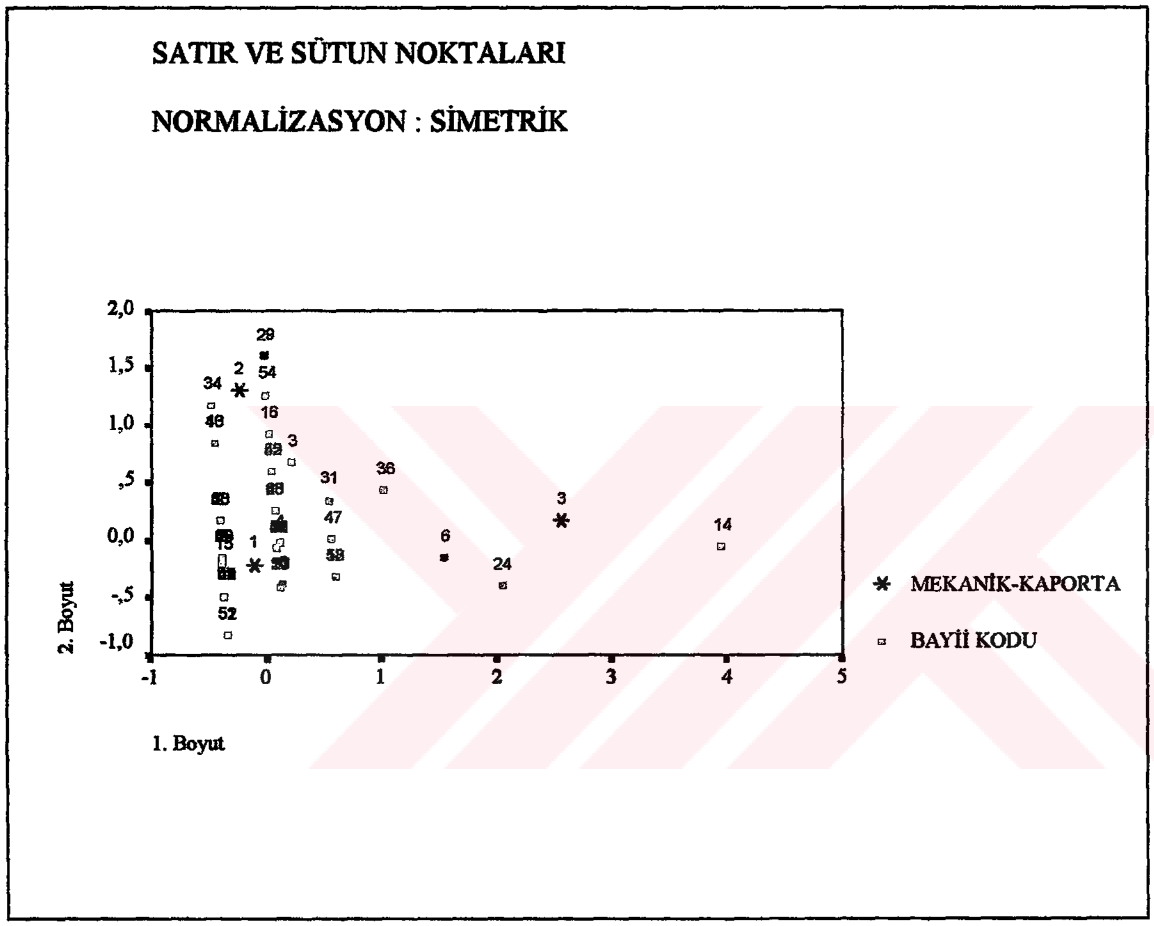
Araç tipi ve genel memnuniyet sorusu için yapılan uygunluk analizinden elde edilen grafik Şekil 4.5'te verilmiştir. Genel memnuniyet sorusu araçlar hakkındaki memnuniyeti değil bayiler hakkındaki memnuniyeti yansıtmaktadır. Buna rağmen araç tipinin bayii memnuniyetini etkileyip etkilemediğinin anlaşılabilmesi için araç tipi ve genel soru için uygunluk analizi yapılması uygun bulunmuştur. Analiz sonucunda oluşan grafikte bütün araç tiplerinin 4 ve 3 numaralı (çok memnunum, memnunum) seçenekleri çevresinde bulunmasından dolayı, araç tipinin bayii memnuniyetini etkilemediği söylenebilmektedir. Bayileri başarısız bulan araç tipi sahiplerinin belirlenmesi için yapılacak bir lojistik regresyon analizi de belirgin bir ayrımsama yapılamayacağı için bir çözüme ulaşamayacaktır. Dolayısıyla araç tipleri açısından memnuniyetin belirlenebilmesi bu amaca yönelik farklı bir araştırmanın yapılmasını gerektirmektedir.

Yukarıdaki grafikten çıkarılabilecek bir sonuç 4, 7 ve 8 numaralı araç tiplerinin diğer araç tiplerinden biraz daha farklı olduğudur. Bunun nedeni az rastlanır özel yapım araçlar olmaları ve farklı ilgiye maruz kalmalarıdır.



Şekil 4.6: Bayiye Geliş Nedeni (Mekanik - Kaporta) ve Araç Tipi İçin Uygunluk Analizi Grafiği

Araçların bayiye geliş nedenlerinin araştırılması amaçlandığında Şekil 4.6'ya başvurulabilir. Şeklin incelenmesi sonucunda araçların mekanik arızalar yüzünden bayiye geldikleri görülecektir. Kaporta problemi ve diğer tür isteklerle bayilere yapılan başvurular mekanik problemlere nazaran çok azdır. 4, 7 ve 8 numaralı araç tipleri yine farklılıkları ile göze çarpmaktadırlar. Bunlar nadir rastlanan araçlar olmalarından dolayı diğerlerinden farklı yerleştirilmişlerdir.



Şekil 4.7: Bayiye Geliş Nedeni ve Bayii Kodu İçin Uygunluk Analizi Grafiği

Bayilerin daha çok hangi nedenle gelen araçlarla ilgilendiği Şekil 4.7'deki grafikten anlaşılabilmektedir. Söz konusu grafiğin ortaya koyduğu gibi 6, 14 ve 24 numaralı bayiler daha çok mekanik ve kaporta dışında kalan isteklerle gelen müşterilerle; 16, 29, 34, 40, 54 numaralı bayiler ise daha çok kaporta problemleriyle ilgilenmektedirler. Diğer bayilerin mekanik problemi yoğunluklu oldukları açıktır.

4.2. Homojenleştirme Analizi

Çok sayıda kategorik değişkenin grafiksel gösteriminden hareketle incelenmesini sağlayan homojenleştirme analizi son yıllarda geliştirilmiş olan bir çok paket program yardımıyla kolayca uygulanabilmektedir. Uygulamamızda SPSS paket programının 8.0 sunumunun CATEGORİES bölümünde bulunan HOMALS seçeneği kullanılmıştır. Uygulamanın yapılabilmesi için bir nesne veya bireye karşılık gelen değişken veya özellik değerleri kategori numarası olacak şekilde sütunların değişkenleri, satırların ise nesne veya bireyleri göstermesi yeterlidir. HOMALS seçeneği marjinal frekansları, kategori koordinatlarını ve özdeğerleri hesaplar ve grafiklerle birlikte sunar.

HOMALS seçeneği ile marjinal frekanslar, özdeğerler, ardışık işlemlerin sonuçları, kategori puanları ve nesne skorları arasından istenenler seçilerek tablolar halinde alınabilir. Ayrıca kategori puanlarının, nesne skorlarının ve ayırım ölçülerinin garafiklerine de ulaşılabilmektedir. Nesne skorları saklanabileceği gibi, boyut sayısı da belirlenebilmektedir.

Homojenleştirme analizinin incelendiği 2.2. bölümde verilen normalleştirilmiş skorlar algoritmasının (4) numaralı son adımında kullanılan yakınsama değeri (convergence) ve en yüksek ardışık işlem sayısı da belirlenebilmektedir. Bu değerler belirlenmediği sürece program otomatik olarak yakınsama değeri için 0.00001 ve en yüksek ardışık işlem sayısı için de 100 değerini kullanmaktadır.

Tablo 4.2 Genel memnuniyet (genel soru), bakım onarım kalitesi memnuniyeti (soru 8) ve bayii kodu için homojenleştirme analizi sonuçları

Tablo 4.2.1 Hareketsizlik ve ayırım ölçüleri

		Her Değişken ve Her Boyut İçin Ayırım Ölçüleri (Discrimination Measures per Variable per Dimension)		
Boyut	**Özdeğer**	**Bayi Kodu**	**Genel Soru**	**Soru 8**
1	0,5542	0,356	0,668	0,639
2	0,4618	0,589	0,486	0,310

Tablo 4.2.1'in ikinci sütununda boyutların özdeğerleri verilmiştir. Bu değerlere göre iki boyut için yapılan araştırmada birinci boyutun ayrımsama açısından ikinciye nazaran daha başarılı olduğu görülmektedir. Ayırım ölçüleri incelendiğinde ise birinci boyuttaki saçılımın büyük çoğunluğunun soru 8 tarafından, ikinci boyuttaki saçılımın büyük çoğunluğunun ise bayii kodu tarafından yaratıldığı görülmektedir. Genel sorunun iki boyuta etkisi birbirine çok yakındır. Bu sonuçlar Şekil 4.8'de verilen ayırım ölçüleri grafiğinde daha belirgin olarak görülmektedir.

Tablo 4.2.2'de bayii kodu için kategori puanları, Tablo 4.2.3'de genel memnuniyet için kategori puanları ve Tablo 4.2.4'de bakım onarım kalitesi memnuniyeti için kategori puanları listelenmiştir.

Tablo 4.2.2 Bayii Kodu İçin Kategori Puanları

Boyut	Bayii Kodu Kategori Puanları (Category Quantifications)								
	1	2	3	4	5	6	7	8	9
1	-0,06	-0,09	1,11	0,77	-0,26	-0,45	-0,36	-2,01	0,41
2	0,65	0,86	0,45	-0,37	-0,70	0,88	-0,34	-0,94	0,29

Boyut	Bayii Kodu Kategori Puanları (Category Quantifications)								
	10	11	12	13	14	15	16	17	18
1	-0,67	0,21	0,14	0,57	1,05	0,41	0,08	0,37	0,40
2	-0,59	0,87	0,57	-0,16	0,63	-0,34	0,89	-0,84	-0,19

Boyut	Bayii Kodu Kategori Puanları (Category Quantifications)								
	19	20	21	22	23	24	25	26	27
1	0,97	-0,05	0,44	-0,30	0,60	-0,17	-0,66	0,99	0,71
2	-1,45	-0,55	0,97	-1,19	0,43	0,35	0,70	-0,88	-1,17

Boyut	Bayii Kodu Kategori Puanları (Category Quantifications)								
	28	29	30	31	32	33	34	35	36
1	-0,32	-0,67	-0,39	0,62	0,50	-0,12	0,23	0,01	-1,20
2	-0,55	-0,45	-0,87	0,28	0,05	-0,59	-1,45	-0,30	0,43

Boyut	Bayii Kodu Kategori Puanları (Category Quantifications)								
	37	38	39	40	41	42	43	44	45
1	-0,83	-1,05	1,06	-0,44	-0,81	-0,98	-0,07	0,53	-0,37
2	0,69	0,65	0,13	-0,01	1,11	-0,18	-0,62	-1,39	-0,28

Boyut	Bayii Kodu Kategori Puanları (Category Quantifications)								
	46	47	48	49	50	51	52	53	54
1	-0,27	-0,40	0,19	0,71	0,19	0,65	-0,12	0,03	-0,06
2	0,78	1,87	0,36	0,01	-0,06	0,22	0,85	-0,79	0,08

Boyut	Bayii Kodu Kategori Puanları (Category Quantifications)						
	55	56	57	58	59	60	61
1	-0,22	-0,16	0,17	0,19	-0,57	0,60	0,32
2	0,56	-1,00	1,82	-0,20	-0,76	0,97	0,13

Tablo 4.2.3. Genel memnuniyet (genel soru) için kategori puanları

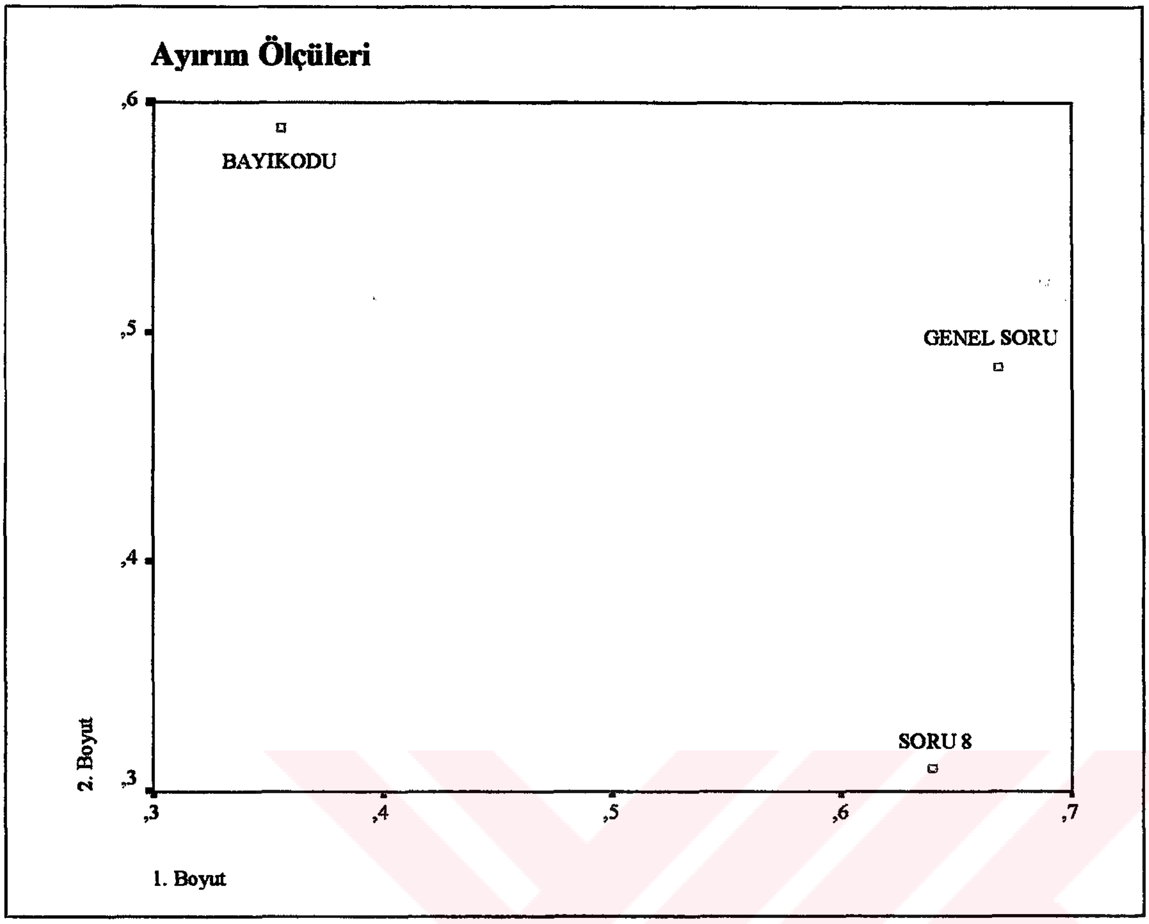
Boyut	Genel Memnuniyet için Kategori Puanları (Category Quantifications)			
	1	2	3	4
1	-1,43	-1,78	-0,03	0,75
2	-1,03	-0,51	0,76	-0,58

Tablo 4.2.4. Bakım onarım kalitesi memnuniyeti (Soru 8) için kategori puanları

Boyut	**Bakım Onarım Kalitesi Memnuniyeti (Soru 8) için Kategori Puanları** (Category Quantifications)		
	1	2	3
1	0,47	-1,50	-0,53
2	-0,07	-0,38	2,14

Tablo 4.2.2, 4.2.3 ve 4.2.4 numaralı tablolar birlikte incelendiğinde 8, 42 numaralı bayii koordinatlarının ((-2,01; -0,94) ve (-0,98; - 0,18)), genel sorunun 1 ve 2 ((-1,43; -1,03) ve (-1,78; -0,51)) soru 8'in 2 (-1,50; -0,51) numaralı seçeneklerinin koordinatları ile çok yakın olduğu görülecektir. Buna göre 8 ve 42 numaralı bayilerin her iki memnuniyet açısından başarısız bulunduğu söylenebilir. Aynı şekilde genel sorunun 4 (0,75; -0,58) ve soru 8'in 1 (0,47; -0,07) numaralı seçenekleri ile yakın koordinatlara sahip olmalarından dolayı 4 (0,77; -0,37), 15 (0,41; -0,34) numaralı bayilerin her iki memnuniyet açısından başarılı bulundukları görülebilmektedir. Bu sonuçlara ulaşabilmek için tabloların bütün hücrelerinin tek tek incelenmesi ve birbiri ile karşılaştırılması yeterli olmakla birlikte Şekil 4.9'da verilen kategori puanları grafiğinin incelenmesi çok daha kolaydır.

Homojenleştirme analizi sonucunda elde edilen ayırım ölçülerine ait grafik Şekil 4.8'de verilmiştir. Tablo 4.2.1 için yapılan açıklamaların aynısı burada da yapılabilir. Başka bir ifade ile birinci boyutun ayırım ölçüsünün büyüklüğü nedeniyle (0,639) soru 8'in daha çok birinci boyutla, ikinci boyutun ayırım ölçüsünün büyüklüğü nedeniyle de (0,589) bayii kodunun ikinci boyutla açıklandığını söylemek yanlış olmaz. Sonuç olarak genel soru her iki boyutla ortak olarak açıklanmaktadır.

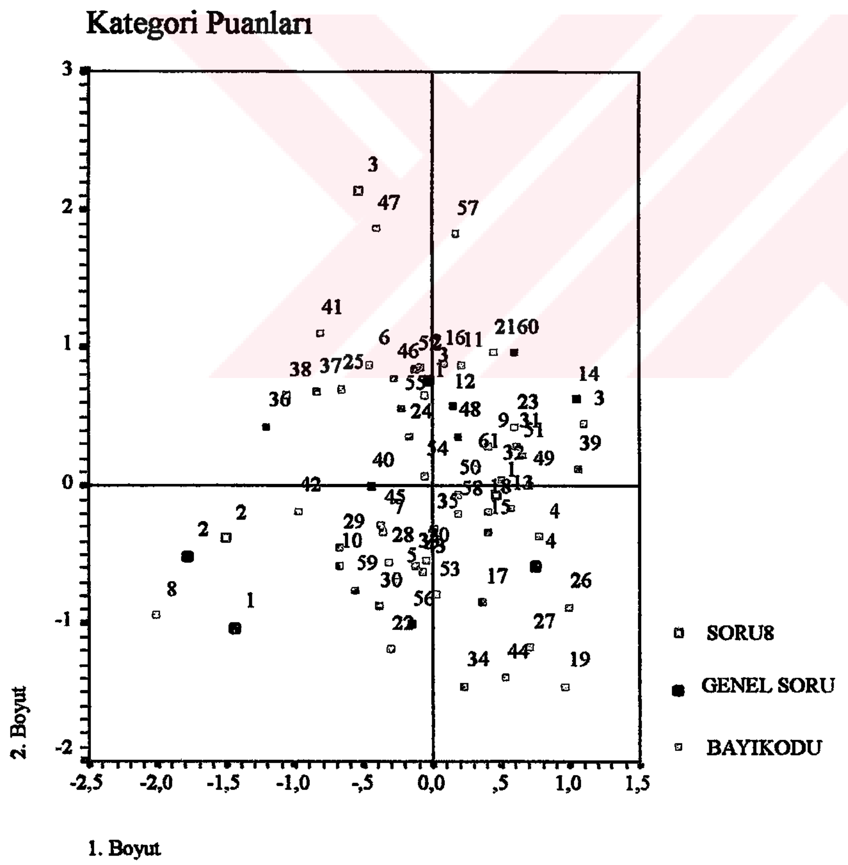


Şekil 4.8. Genel Memnuniyet, Bakım Onarım Kalitesi Memnuniyeti ve Bayii Kodu İçin Ayırım Ölçüleri Grafiği

Şekil 4.9'da verilen kategori puanları grafiğinde, soru 8'in 2 numaralı seçeneği ile genel sorunun 1ve 2 numaralı seçenekleri birbirlerine çok yakın yerleşmişlerdir. Bu beklenen bir durumdur. Aynı ifade soru 8'in 1 ve genel sorunun 3, 4 seçenekleri için de geçerlidir. Birinci boyut iyilerle kötüleri ayıran en iyi boyut iken, ikinci boyut bu ayırım konusunda daha zayıf kalmıştır. Bu sonuç, Şekil 4.1'deki grafikten ulaşılan sonuçla aynıdır.

42 ve 8 numaralı bayiler her iki memnuniyet sorusuna göre başarısız, 4, 13, 15, 18 numaralı bayiler ise her ikisine göre başarılı bulunmaktadır. 47 ve 57 numaralı

bayiler hakkında bir kararsızlık söz konusudur. 19, 26, 27, 34 ve 44 numaralı bayiiler genel memnuniyet açısından çok başarılı bulunmuşlardır. Bununla beraber bakım onarım kalitesinin bu memnuniyete katkısının diğer etkenlere nazaran daha az olduğu da göz önünde bulundurulmalıdır. Bu ifadenin kullanılmış olmasının nedeni adı geçen bayilerin soru 8'in memnuniyeti belirten1 numaralı seçeneğinden uzağa fakat genel sorunun 4 numaralı 'çok memnunum' seçeneğine yakın yerleştirilmiş olmasıdır. 1, 46 ve 52 numaralı bayiler gibi genel sorunun 3 numaralı seçeneğine yakın yerleşmiş olan bayilerin ise genel olarak başarılı bulunduğu söylenebilir.



Şekil 4.9. Genel Memnuniyet, Bakım Onarım Kalitesi Memnuniyeti ve Bayii Kodu İçin Homojenleştirme Analizi Grafiği

Merkeze yakın yerleşmiş olan ve haklarında belirgin bir karara varılamayan diğer bayilerin sınıflandırılması için yine lojistik regresyon analizinden yararlanılmıştır. Birden fazla lojistik fonksiyonuna ulaşılabilmektedir. Şekil 4.10'daki grafikte kesikli çizgi ile gösterilen doğru, 4. gruba dahil olan bayilerin belirlenmesi için kullanılmıştır. Bunun için bayi sınıflama değişkeninde 4 değerini alan bayiler bir gruba diğer bayiler ise ikinci gruba dahil kabul edilmiştir. Doğrunun elde edildiği lojistik fonksiyon aşağıda verilmiştir.

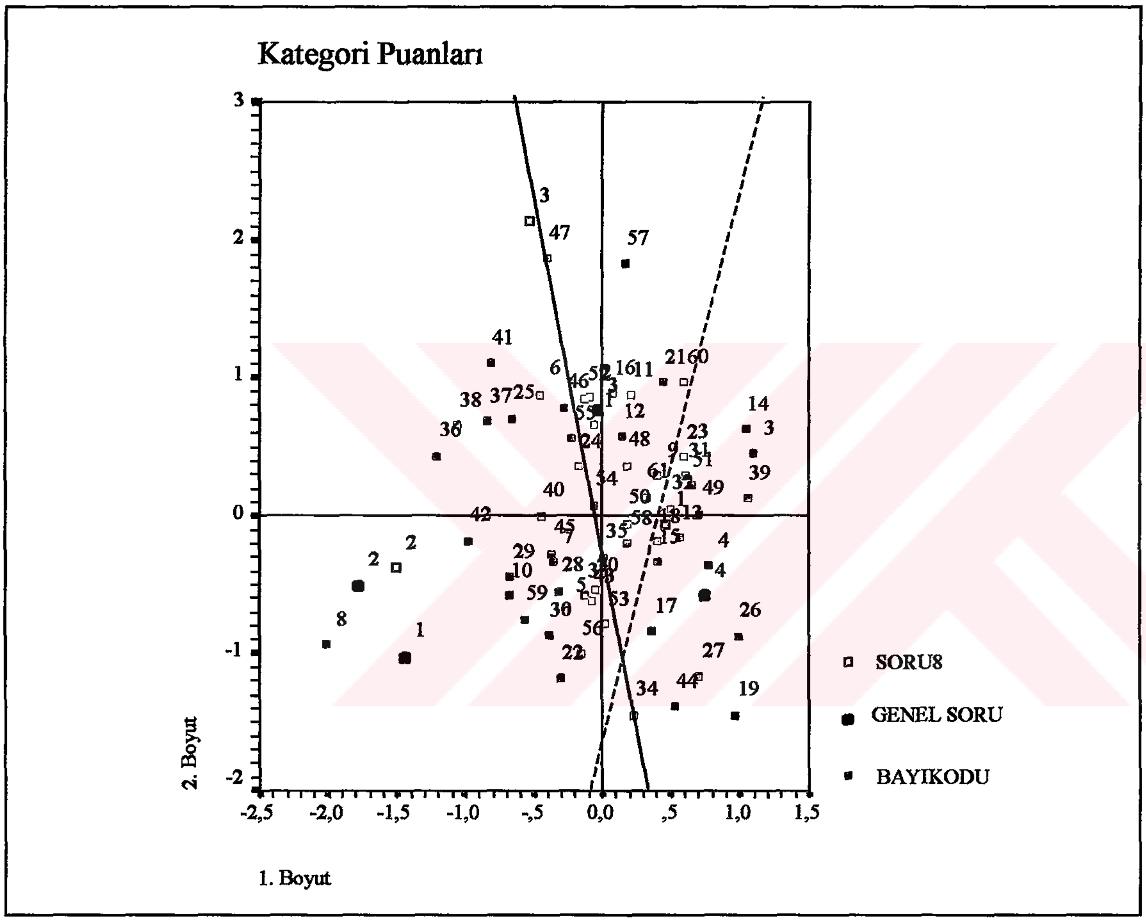
$$f = -2.649 + 7.1173\ (\text{Boyut } 1) - 1.8331\ (\text{Boyut } 2)$$

Kritik atama noktası yine 0.5 olarak alınmıştır. Doğrunun sağında kalan bayilerin çok başarılı bayiler olarak belirlenmesi mümkündür. Ayrıca 3. gruba dahil bireylerin de diğerlerinden ayrılması gerekli olabilir. Bunun için bir başka lojistik fonksiyonuna ihtiyaç duyulacaktır. Bu durumda bayi sınıflama değişkeninde 3 ve 4 değerlerini alan bayiler bir gruba diğer bayiler ise ikinci gruba dahil kabul edilerek lojistik regresyon analizi uygulanabilir. Böylece ulaşılan lojistik fonksiyon aşağıda verilmiştir.

$$f = 0.713 + 4.7286\ (\text{Boyut } 1) + 0.8216\ (\text{Boyut } 2)$$

Kritik atama noktasının 0.5 olarak alınması ile ulaşılan doğru Şekil 4.10'daki grafikte düz çizgi gösterilmiştir. Bu doğrunun solunda kalan bayilerin başarılı ve çok başarılı bulunan bayiler olduğu söylenebilir. Bununla grafik üzerinde gösterilen doğrular arasında kalan bayilerin 3. kategoriye dahil olduğu başka bir deyişle başarılı bulundukları gözlenmektedir. Benzer şekilde kalan iki grubu (1 ve 2) birbirinden ayırmak için üçüncü bir lojistik fonksiyonu daha elde edilebilir.

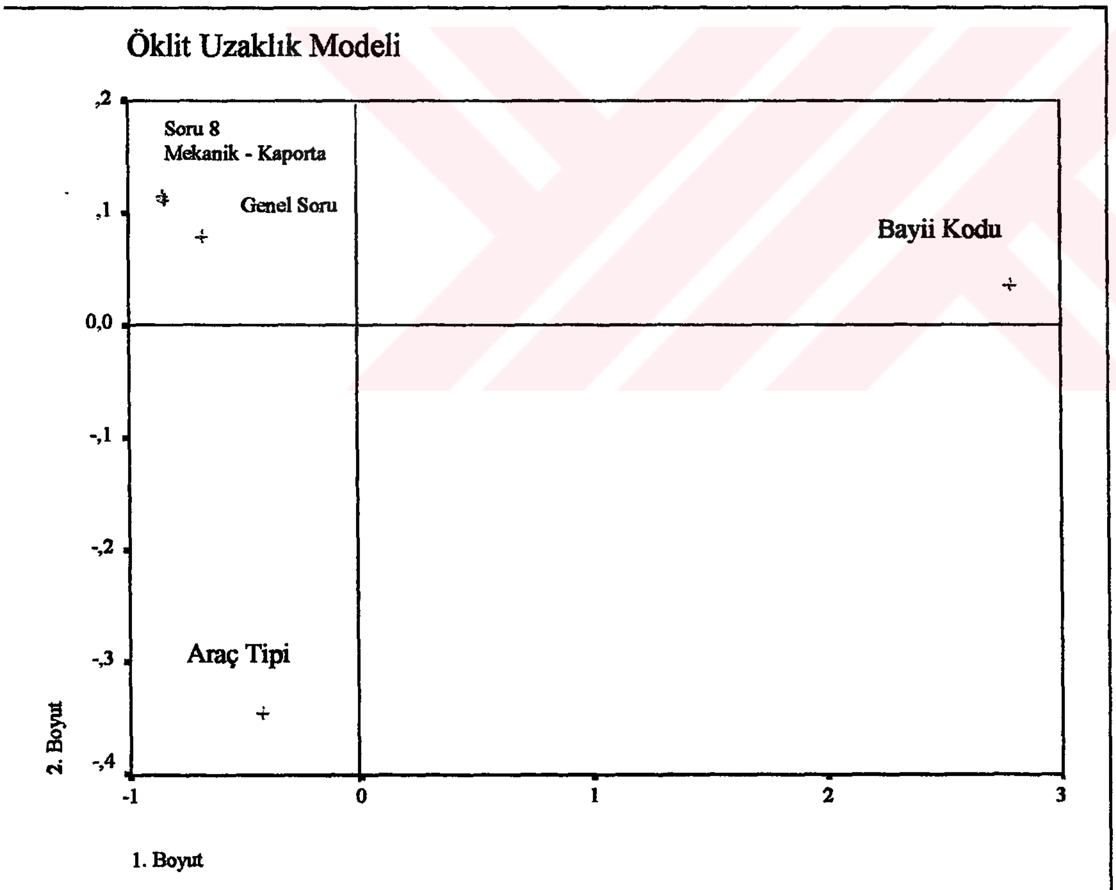
Yukarıda yapılan açıklamalardan da anlaşılabileceği gibi sınıflamayı kolaylaştırmak amacıyla birden fazla lojistik fonksiyonuna başvurulabilmektedir.



Şekil 4.10: Homojenleştirme Analizinde Lojistik Regresyonun Kullanımı

4.3. Çok Boyutlu Ölçekleme

Geçmişi uygunluk ve homojenleştirme analizinden daha eskilere dayanan, farklılıklar matrisinin grafiksel gösterimini sağlayan çok boyutlu ölçekleme uygulamalarında da SPSS paket programı kullanılmıştır. Amaç uygunluk ve homojenleştirme analizin ayrıntılı biçimde incelenmesi olduğundan bu ve izleyen alt bölümlerde sadece yöntemler arası farklılığı belirtmek için birer grafiğin verilmesiyle yetinilmiştir.



Şekil 4.11: Çok Boyutlu Ölçekleme Grafiği

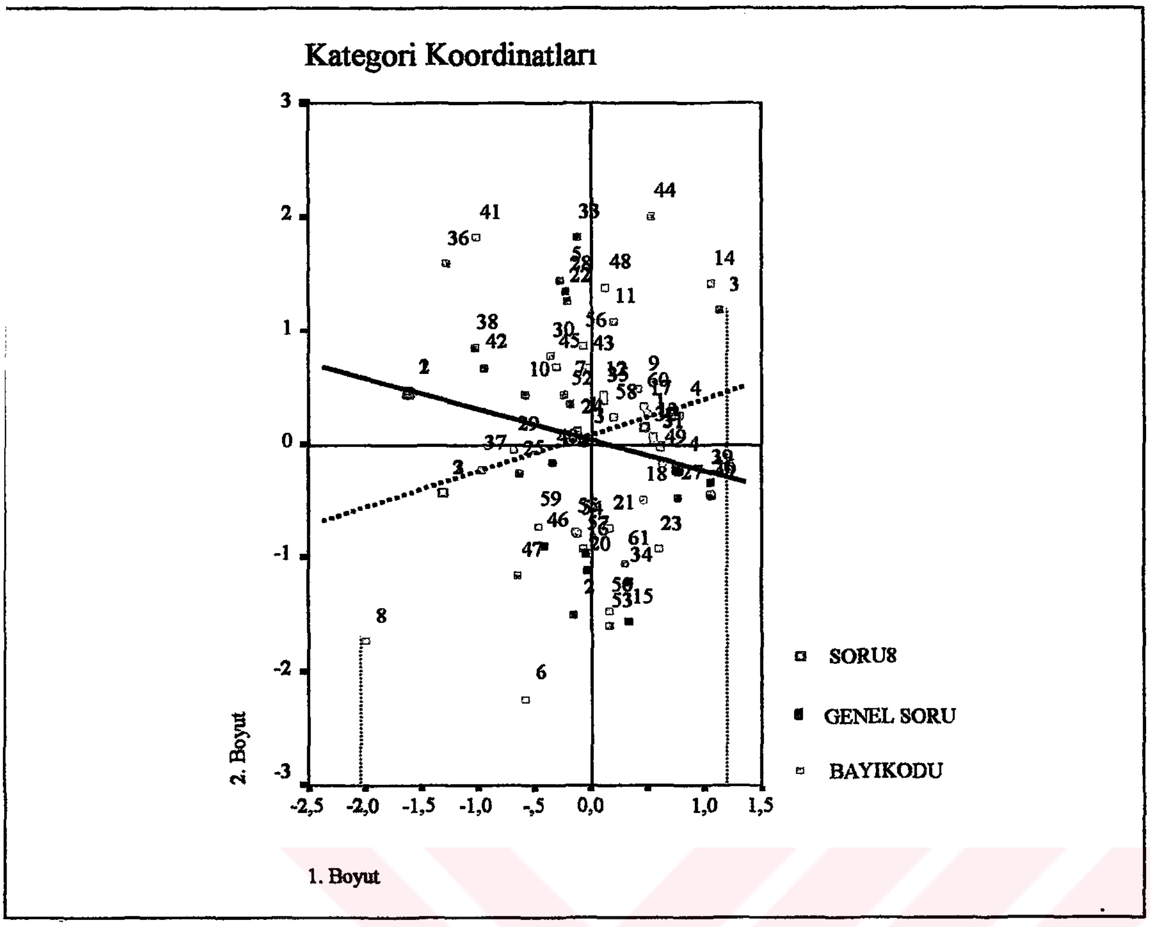
Bu uygulamanın sonucunda elde edilen grafik. Şekil 4.11'de sunulmuştur. Anılan grafiğin ortaya koyduğu gibi çok boyutlu ölçekleme tekniği ile sadece değişkenler arası uzaklıklara ulaşılabilmekte, kategori bazına inilememektedir. Ayrıca değişkenler arasındaki uzaklıkların, değişkenler arasındaki benzerlik veya ilişkiler cinsinden yorumu da gerçekleştirilememektedir. Sözü edilen uzaklıklar, grafiğin başlığından da anlaşılabileceği gibi öklit uzaklıkları olup sadece fiziksel değerler olarak algılanmaları gerekir. Dolayısıyla kategoriler arası ilişkiler arandığında homojenleştirme analizi, farklılıklar matrisi yardımıyla nesneler arasındaki uzaklıkları gösteren haritaya ihtiyaç duyulduğunda çok boyutlu ölçekleme analizi kullanılmalıdır.

4.4 Doğrusal Olmayan Temel Bileşenler Analizi

Değişkenler arasındaki bağımlılığı ortadan kaldırılmak ve boyut indirgemek için kullanılan temel bileşenler analizinin uygulaması için SPSS paket programının PRINCALS seçeneğinden faydalanılmıştır. Bu seçenek doğrusal olmayan temel bileşenler analizi uygulamasına imkan vermektedir. HOMALS seçeneğinde bütün değişkenlerin çoklu sınıfsal (multinominal) olması gerekirken, PRINCALS ile çoklu sınıfsal değişkenlerin yanı sıra sırasal (ordinal) ve/veya kesikli sayısal (discrete numeric) değişkenlerin kullanılması da mümkündür.

Uygulamalarımızda kullanılan genel soru ve soru 8'in sırasal değişken olarak alınması da mümkündür. Bu değişkenlerin sırasal alınmasıyla gerçekleştirilen uygulama sonucunda elde edilen grafik Şekil 4.12'de sunulmuştur. Grafik incelendiğinde soru 8 ve genel sorunun memnuniyetsizlik belirten seçeneklerinin solda, memnuniyet belirten seçeneklerinin ise sağda bulunduğu görülmektedir. Böylece bayilerin memnuniyete sıralanabileceği görülmektedir. Homejenleştirme analizinden elde edilen sonuçlarla benzer olmasından dolayı bu sonuç olağandır. Ayrıca her iki sorunun kategorilerinin birer doğru üzerinde memnuniyetsizlikten memnuniyete doğru sıra ile yerleştiği de görülebilecektir. Kalın doğru genel soru için, kesikli kalın doğru ise soru 8 için çizilmiştir. Bu durum soru değişkenlerinin sırasal olarak tanımlanmasından ve temel bileşenler analizinin özelliklerinden ileri gelmektedir. Böylece, bütün kategorilerin birinci eksene izdüşümlerinin alınmasıyla bayilerin soru 8 ve genel soruya göre memnuniyetsizlikten memnuniyete doğru sıralanması mümkün olacaktır.

En soldaki ve en sağdaki bayilerin birinci boyuta izdüşümlerinin alınmasıyla iki memnuniyeti sorusunun aynı anda göz önünde bulundurulması durumunda 8 numaralı bayiinin en kötü, 3 numaralı bayiinin ise en iyi olduğu söylenebilir. Bütün bayiler için benzer sıralamanın yapılması mümkündür.

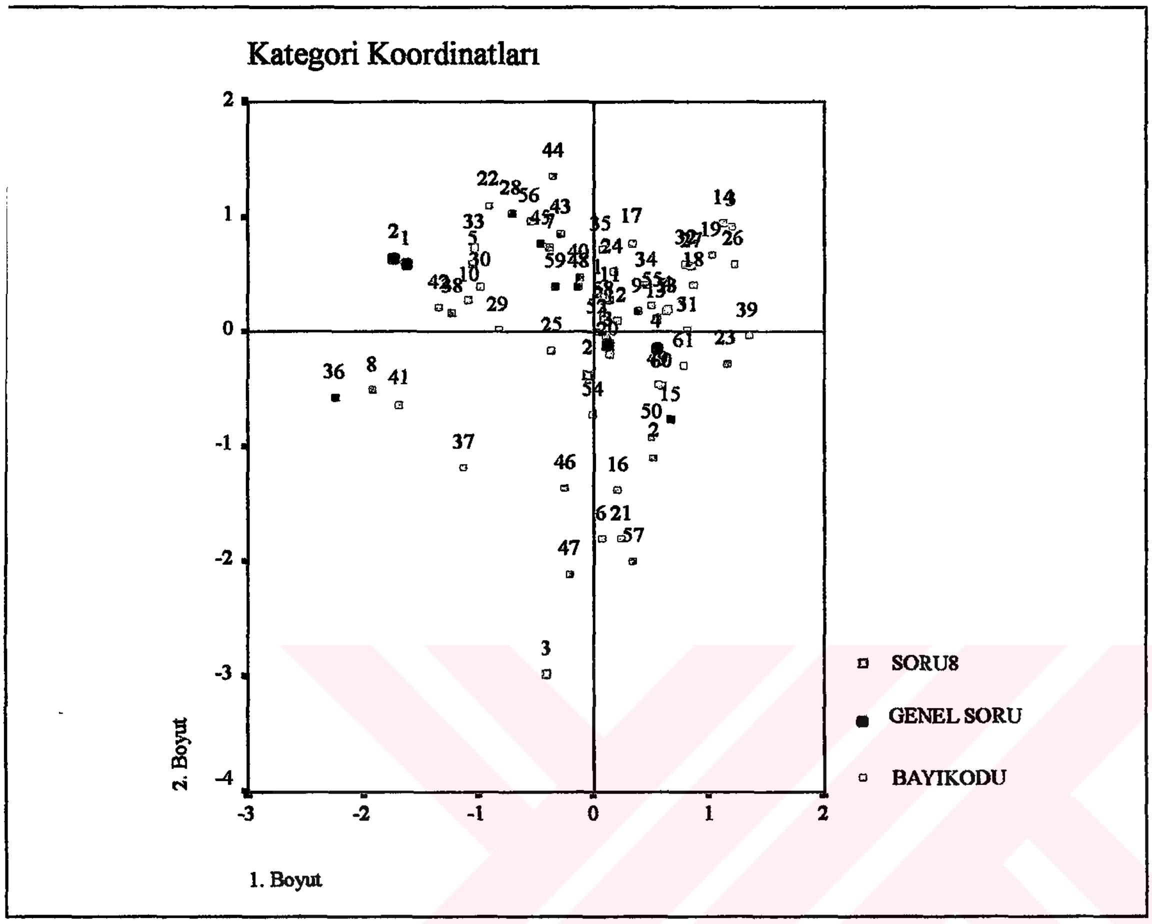


Şekil 4.12: Bakım Onarım Kalitesi Memnuniyeti (Soru 8) , Genel memnuniyeti ve Bayii Kodu İçin Temel Bileşenler Analizi Grafiği

4.5. Doğrusal Olmayan Kanonik Korelasyon Analizi

SPSS paket programı değişken setleri arasındaki ilişkiyi inceleyen kanonik korelasyon analizi uygulanmasını da mümkün kılmaktadır. Söz konusu analiz için CATEGORIES grubunda bulunan OVERALS seçeneğinin kullanılması gerekir.

Kanonik korelasyon analizinin homojenleştirme ve temel bileşenler analizinden farkı değişkenler arasındaki bağımlılığı teker teker değişken bazında değil, değişkenlerin oluşturdukları kümeler arasında incelemesidir. Böyle incelenmesine rağmen grafikte bütün değişkenlerin kategorilerinin yerleştiriliyor olması incelemeyi kolaylaştırmaktadır.



Şekil 4.13: Bakım Onarım Kalitesi Memnuniyeti (Soru 8) , Genel Memnuniyet ve Bayii Kodu İçin Kanonik Korelasyon Analizi Grafiği

İlgilenilen değişkenlerle yapılan kanonik korelasyon analizi sonucu elde edilen grafik Şekil 4.13'te verilmiştir. Bu analizde soru 8 ve genel soru birbirlerine olan yakınlıkları nedeniyle bir küme olarak ve bayii kodu ayrı bir küme olarak değerlendirilmiştir. Kanonik korelasyon analizi küme içindeki değişkenler arasındaki bağımlılığı ortadan kaldırdıktan sonra değişken kümeleri arasındaki ilişkiyi araştırdığı için soru 8 ve genel soru farklı boyutlarda etkin görünmektedir.

BÖLÜM V

SONUÇ

Çok değişkenli parametik olmayan teknikler özellikle son yıllarda dünya literatüründe önemli bir yer tutmaya başlamıştır. Gelişen bilgisayar teknolojisi ve paket programlar da parametrik yöntemlere nazaran daha gerçekçi görünen bu tür tekniklerin uygulanmasını kolaylaştırmaktadır.

Bilindiği gibi homojenleştirme analizi parametrik olmayan kategorik verilerin incelenmesinde kullanılan önemli yöntemlerden biri olup ikiden fazla değişkenin bulunduğu durumlarda grafiksel gösterim sağlaması nedeniyle incelemeyi kolaylaştıran bir analiz tekniğidir. Çalışmanın amacı söz konusu tekniğin daha uygulanabilir olmasını sağlamak ve diğer çok değişkenli analiz teknikleri ile olan ilişkisini açıklamaya çalışmaktır.

Bu amaç doğrultusunda konuyla ilgili olduğu düşünülen teknikler incelenmiş ve karşılaştırılmıştır. Homojenleştirme analizinden elde edilen grafikler gerekli görüldüğünde oldukça ayrıntılı biçimde yorumlanmıştır. Yorumlar sırasında karşılaşılan ayrımsama problemlerini ortadan kaldırmak için lojistik regresyon analizine başvurulmuştur. Bu analiz yardımıyla grafik üzerinde bir veya daha fazla doğruya ulaşılabileceği ve bu doğrular yardımıyla ayrımsamanın daha kolay yapılabileceği gösterilmiştir. Böylece araştırmacılar tarafından el yordamıyla çizilen doğrulara kuramsal bir alt yapı oluşturulmuş olmaktadır. Analiz tekniklerinden elde edilen grafiklerin değişik açılardan yorumlarına yer verilmiştir.

İki kategorik değişkenin bulunması durumunda kullanılabilecek bir yöntem olan uygunluk analizi ile ulaşılan sonuçlar, iki değişkenle uygulanan homojenleştirme analizi ile ulaşılan sonuçlarla karşılaştırılmış ve aralarındaki benzerliğe dikkat çekilmiştir. Homojenleştirme analizinin çoklu uygunluk analizi olarak da adlandırılmasının nedeni budur.

Sadece değişkenler bazında kalan bir grafiğe ulaşmayı sağlayan çok boyutlu ölçekleme tekniği, kategorilerin ilişkisi araştırıldığında yetersiz kalmaktadır. Doğrusal olmayan temel bileşenler analizi, kategorilerin sıralaması için kullanılabilecek bir yöntem olduğundan sınıflandırmada kullanılan homojenleştirme analizinden farklıdır. Doğrusal olmayan kanonik korelasyon analizi, değişken kümesi içinde bağımsızlığı sağladıktan sonra değişken kümeleri arasındaki ilişkiyi ortaya koymaktadır. Bu nedenle, anılan teknik bütün kategorilerin birbiri ile olan ilişkisi yerine farklı kümelerde bulunan değişkenlerin kategorileri arasındaki ilişkinin önemli olduğu araştırmalarda kullanılmalıdır.

Çalışmamızda bahsedilen yöntemler sosyal bilimler alanındaki araştırmalarda da güvenle kullanılabilir. Gelecek çalışmalar için lojistik regresyonun ikiden fazla grup için kullanımının geliştirilmesi uygun gözükmektedir. Bununla beraber eksik (missing value) ve aykırı değerlerin (outlier) bulunduğu veri analizi de üzerinde çalışılması gereken konulardır.

KAYNAKLAR

AGRESTI, A.,(1983). A Survey of Strategies for Modeling Cross-Classification Having Ordinal Variables. JASA, Vol. 78, No: 381, 184-198.

AGRESTI, A.(1990). Categorical Data Analysis, John Wiley and Sons Ltd.

BARNETT, V., (1981). Interpreting Multivariate Data, John Wiley and Sons Ltd.

BARTHOLOMEW, D. J., (1980). Factor Analysis for Categorical Data. J. R. Statist. Soc. B., Vol.42, No. 3, 293 - 321.

BASILEVSKY, A., (1994). Statistical Factor Analysis and Related Methods Theory and Applications, John Wiley&Son's Inc

BAŞARIR, G.,(1990). Çok Değişkenli Verilerde Ayrımsama Sorunu ve Lojistik Regresyon Analizi, Doktora Tezi, Hacettepe Üniversitesi.

BOIK, R.,(1996). An Efficient Algorithm for Joint Correspondence Analysis. Psychometrika, Vol.61, No.2, 255 - 269.

BORG, I. and GROENEN, P., (1997). Modern Multidimensional Scaling. Theory and Applications. Springer.

BÖCKENHOLT, U., BÖCKENHOLT, I., (1990). Canonical Analysis of Contingency Tables With Linear Constraints. Psychometrika, Vol.55, No.4, 633 - 639.

CHOULAKIAN, V., (1988). Exploratory Analysis of Contingency Tables by Loglinear Formulations and Generalizations of Correspondence Analysis. Psychometrika, Vol.53, No.2, 235 - 250

CARLIER, A., KROONENBERG, P. M., (1996). Decompositions and Biplots in Three-Way Correspondence Analysis. Psychometrika, Vol.61, No.2, 355 - 373.

CARROLL, J. D., GREEN, P. E. and SCHAFFER, M.C. (1986). Interpoint Distance Comparisons in Correspondense Analysis. Journal of Marketing Research, Vol. 23, 271-280.

CARROLL, J. D., GREEN, P. E. and SCHAFFER, M.C. (1987). Comparing Interpoint Distances in Correspondense Analysis: A Clarification. Journal of Marketing Research, Vol. 24, 445-450.

CARROLL, J. D., GREEN, P. E. and SCHAFFER, M.C. (1989). Reply to Greenacre's Commentary on the Carroll-Green-Schaffer Scaling of Two-Way Correspondense Analysis Solutions. Journal of Marketing Research, Vol. 26, 366-368.

COX, T. F. and COX, M. A. A., (1994). Multidimensional Scaling. Chapman & Hall.

DAWKINS, B. P., ANDREAE, P. M. and O'CONNOR, P. M., (1994). Analysis of Olympic Heptathlon Data. JASA, Vol. 89, No: 427, 1100-1106.

DE LEEUW, J., and VAN DER HEIJDEN, P. G. M.,(1988). Correspondense Analysis of Incomplete Contingency Tables. Psychometrika, Vol.53, No:2, 223-233.

FIENBERG, S. E., (1992). A Brief History of Statistics Three and One-Half Chapters: A Review Essay. Statistical Science, Vol.2, 208 - 225.

FRIENDLY, M. (1994). Mosaic Displays for Multiway Contingency Tables. JASA, Vol. 89, No: 425, 190-200.

GIFI, A., (1990). Nonlinear Multivariate Analysis. John Wiley & Sons.

GILULA, Z. and RITOV, Y., (1990). Inferential Ordinal Correspondense Analysis: Motivation, Derivation and Limitations. International Statistical Review, Vol. 58, 2, 99-108

GOODMAN, L. A., (1979). Simple Models for the Analysis of Association in Cross-Classifications Having Ordered Categories. JASA, Vol.74, No:367, 537-552.

GOODMAN, L. A., (1981). Association Models and Canonical Correlation in the Analysis of Cross-Classifications Having Ordered Categories. JASA, Vol.76, No:374, 320-334.

GOODMAN, L.A., (1986). Some Useful Extensions of the Usual Correspondence Analysis Approach in Analysis of Contingency Tables. Intenational Statistical Review, Vol. 54, 3, 243-309

GOODMAN, L.A., (1991). Measures, Models and Graphical Displays in the Analysis of Cross-Classified Data. JASA, Vol.86, No. 416, 1085-1038.

GOWER, J. C., (1990). Fisher's Optimal Scores and Multivariate Correspondence Analysis. Biometrics, Vol. 46, 947-961.

GREENACRE, M. J. and HASTİE, T., (1987). The Geometric Interpretation of Correspondense Analysis. JASA, Vol. 82, No: 398, 437-447.

GREENACRE, M. J., (1988). Correspondence Analysis of Multivariate Categorical Data by Weighted Least - Squares. Biometrica, Vol. 75, 3, 457-467.

GREENACRE, M. J.,(1989). The Carroll-Green-Schaffer Scaling in Correspondense Analysis: A Theoretical and Empirical Appraisal. Journal of Marketing Research, Vol. 26, 358-365.

GREENACRE, M. J.,(1993). Correspondence Analysis in Practice. Academic Press.

HAWKINS, D.M., (1982). Topics in Applied Multivariate Analysis. Cambridge University Press.

HENNEBERT, M., LEES, A.,(1991). Environmental Gradients in Carbonate Sediments and Rocks Detected by Correspondence Analysis: Example From the Recent of Norway and the Dinantian of Southwest England. Sedimentation, Vol. 38, 623 - 642.

HOFFMAN, D. L. and FRANKE, G. R. (1986). Correspondense Analysis: Graphical Representation of Categorical Data in Marketing Research. Journal of Marketing Research, Vol. 23, 213-227.

HUBERT, L., ARABIE, P., (1992). Correspondence Analysis and Optimal Structural Representations. Psychometrika, Vol.56, No.1, 119 - 140.

JACKSON, D. A., SOMERS K. M., (1991). Putting Things in Order: The Ups and Downs of Detrended Correspondence Analysis. The American Naturalist, Vol. 137, No.5, 704 - 712.

JACKSON, J. E., (1991). A User's Guide to Principle Components, John Wiley&Son's Inc

JAMBU, M. and LEBEAUX, M. D., (1983). Cluster Analysis and Data Analysis. North Holland Publishing Company.

KIERS, H. A. L., (1991). Simple Structure in Component Analysis Techniques for Mixtures of Qualitative and Quantitative Variables. Psychometrika, Vol.56, No.2, 197 - 212.

KOSTER, J. T. A.,(1989). Mathematical Aspects of Multiple Correspondence Analysis for Ordinal Variables, DSWO Press.

KRZONOWSKI, W. J., (1988). Principles of Multivariate Analysis A User's Perspective. Oxford Science Publications.

MANLY, B. F. J., (1989). Multivariate Statistical Methods. Chapman & Hall.

MARDIA, K. V.,KENT, J. T., BIBBY,J. M., (1979). Multivariate Analysis. Academic Press.

MORRISON, D. F., (1967). Multivariate Statistical Methods. McGraw - Hill inc.

NISHISATO, S.,(1986). Classification with a Variety of Categorical Data. Classification as Tool of Research. Elsevier Science Publishers B.V., 353 - 359.

NISHISATO, S., (1993). On Quantifying Different Types of Categorical Data. Psyhometrika, Vol. 58, No:4, 617-629.

REYMENT, R., JÖRESKOG, K. G., (1993). Applied Factor Analysis in the Natural Sciences. Cambridge Uni. Press.

ROMANAZZI, M., (1992). Influence in Canonical Correlation Analysis. Psychometrika, Vol.57, No.2, 237 - 259.

STURT, E., (1981). Computerized Construction in Fortran of a Discriminant Function for Categorical Data. Appl. Statist., Vol. 30, No. 3, 213 - 222.

TENENHAUS, M. and YOUNG, F. W.,(1985). An Analysis and Synthesis of Multiple Correspondense Analysis, Optimal Scaling, Dual Scaling, Homogeneity Analysis and Other Methods for Quantifying Categorical Multivariate Data. Psychometrika, Vol. 50, No:1, 91-119.

TER BRAAK, C. J., (1986). Canonical Correspondence Analysis: A New Eigenvector Technnique for Multivariate Direct Gradient Analysis. Ecology, Vol. 67(5), 1167-1179.

TUNA, M. ve KIROĞLU, G., (1996). Uygunluk Analizi Üzerine Bir Uygulama. Marmara Üniversitesi İstatistik ve Ekonometri Araştırma ve Uygulama Merkezi Dergisi Sayı 1, 125 -134.

TUNA, M. (1998). Homojenlik ve Uygunluk Analizleri Üzerine Bir Uygulama. İstatistik Konferansı Bildiriler Kitabı , D.İ.E. ve Gazi Üniversitesi, 345 – 351.

ULUÇHAN, Y. E., (1992). Correspondence Analysis Approach in the Analysis of Contingency Tables. Eastern Mediterrenean University, Master Thesis.

VAN DER BURG, E., DE LEEUW, J. AND VERDEGAAL, R., (1988). Homogeneity Analysis With k Sets of Variables: An Alternating Least Squares Method With Optimal Scaling Features. Psychometrika, Vol.53, No.2, 177 - 197.

VAN DER HEIJDEN, P. G. M. and DE LEEUV, J.,(1985). Correspondense Analysis Used Complementary to Loglinear Analysis. Psychometrika, Vol. 50, No: 4, 287-291.

VAN DER HEIJDEN, P. G. M., and Worsley, K. J. (1988). Comment on "Correspondense Analysis Used Complementary to Loglinear Analysis". Psychometrika, Vol. 53, No: 2, 287-291.

WELLER, S. C., (1990). Metric Scaling : Correspondence Analysis, Sage Publ.

EKLER

EK 1

SORU FORMU

0 – Genel olarak, otomobilinizi götürdüğünüz bu servisten ne kadar memnunsunuz? Memnuniyetinizi 1 ila 4 arasında puanlar mısınız? 1 = Hiç memnun değilim, 4= Çok memnunum anlamındadır.

- Hiç memnun değilim 1
- Memnun Değilim 2
- Memnunum 3
- Çok memnunum 4

1 – Servise geldiğinizde, iyi karşılanmanızdan % 100 memnun kaldınız mı?

- % 100 Memnunum 1
- % 100 Memnun Değilim 2
- Ekstra... 3

2 – Servise gitmek için randevu aldınız mı?

- Evet 1
- Hayır 2
- Ekstra... 3

3 – Servise gittiğinizde aracınızı beklemeden teslim etmenizden % 100 memnun kaldınız mı?

- % 100 Memnunum 1
- % 100 Memnun Değilim 2
- Ekstra... 3

4 – Otomobilinize yapılacak bakım ve onarım ile ilgili, size verilen bilgi ve önerilerden % 100 memnun kaldınız mı?

- % 100 Memnunum 1
- % 100 Memnun Değilim 2
- Ekstra.. 3

5 – Özellikle dikkat edilmesini istediğiniz noktalara özen gösterilmesinden % 100 memnun kaldınız mı?

- % 100 Memnunum 1
- % 100 Memnun Değilim 2
- Ekstra.. 3

6 – Otomobilinizin onarımının belirtilen zamanda tamamlanmasından % 100 memnun kaldınız mı?

- % 100 Memnunum 1
- % 100 Memnun Değilim 2
- Ekstra.. 3

7 – İstediğiniz işlerin bakım ve onarımlarının gerçekleşmesinden % 100 memnun kaldınız mı?

- % 100 Memnunum 1
- % 100 Memnun Değilim 2
- Ekstra.. 3

8 – Yapılan bu bakım ve onarımların kalitesinden % 100 memnun kaldınız mı?

- % 100 Memnunum 1
- % 100 Memnun Değilim 2
- Ekstra.. 3

9 – Aracınız size yıkanmış ve temiz olarak tesliminden % 100 memnun kaldınız mı?

- % 100 Memnunum 1
- % 100 Memnun Değilim 2
- Ekstra.. 3

10 – İlgili kişilere kolay ulaşabilmenizden % 100 memnun kaldınız mı?

- % 100 Memnunum 1
- % 100 Memnun Değilim 2
- Ekstra.. 3

11 – Serviste çalışanların size gösterdikleri ilgiden % 100 memnun kaldınız mı?

- % 100 Memnunum 1
- % 100 Memnun Değilim 2
- Ekstra.. 3

12 – Faturada belirtilen işlerden dolayı sorduğunuz sorular için aldığınız cevaplardan % 100 memnun kaldınız mı?

- % 100 Memnunum 1
- % 100 Memnun Değilim 2
- Ekstra.. 3

13 – Faturada yapıldığı belirtilen işler sizin talep etmiş olduğunuz işlerin karşılığımıydı? Bu konudan % 100 memnun kaldınız mı?

- % 100 Memnunum 1
- % 100 Memnun Değilim 2
- Ekstra.. 3

14 - Ödediğiniz ücret yapılan hizmetlerin karşılığı mıydı? Bu konudan % 100 memnun kaldınız mı?

- % 100 Memnunum 1
- % 100 Memnun Değilim 2
- Ekstra.. 3

15 – Talep ederek yaptırmış olduğunuz bu işlemlerden dolayı tekrar servise gitme ihtiyacı hissettiniz mi?

- Evet 1
- Hayır... 2
- Ekstra.. 3

16 – Otomobilinizi servise niye götürmüştünüz?

- Mekanik.. 1
- Kaporta/Boya... 2
- Diğer... 3

17 – Servise götürdüğünüzde otomobiliniz kaç kilometrede idi?

(.............. Km.)

18 – Aracınız serviste yedek parça olmadığı için bekledi mi?

- Evet 1
- Hayır... 2

EK 2

KOMUT DİZİNLERİ

Uygunluk Analizi

Uygunluk analizi komut dizini uygulama kısmında uygunluk analizi için verilen ilk örnekte kullanılan komut dizinidir.

CORRESPONDENCE TABLE = GENSOR(1,4) BY SORU8(1,3)
/OUTFİLE = SCORE(UYGSKOR).

Bu komut dizini sayesinde kategorilerin boyutlara göre skorları UYGSKOR dosyasında toplanmış ve diğer seçenekler için programın geçerli (default) seçenekleri kullanılmış olmaktadır.

Homojenleştirme Analizi

Homojenleştirme analizi komut dizini uygulama kısmında homojenlik analizi için verilen ilk örnekte kullanılan komut dizinidir.

HOMALS VARİABLES = SORU8(3) GENSORU(4) BAYİKODU(61)
/MATRIX = OUT (HOMSKOR).

Bu komut dizini sayesinde kategorilerin boyutlara göre skorları HOMSKOR dosyasında toplanmış ve diğer seçenekler için programın geçerli (default) seçenekleri kullanılmış olmaktadır.

Lojistik Regresyon Analizi

Uygulamalar sırasında kullanılan lojistik regresyon analizi komut dizini aşağıda verilmiştir.

LOGISTIC REGRESSION VARIABLES = BAGDEG WITH BOYUT1, BOYUT2
/ CRITERIA = CUT 0.5.

BAGDEG iki sınıflı bağımlı değişken iken BOYUT1 ve BOYUT2 uygunluk veya homojenleştirme analizi sonucunda ulaşılan birinci ve ikinci boyut için kategori koordinatlarıdır. Kritik atama noktası CUT değeri ile belirlenebilmektedir. Yukarıda verilen dizinde adı geçen değer 0.5 olarak alınmıştır.