Kırklareli Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi
Kırklareli University Journal of the Faculty of Economics and Administrative Sciences
KLUJFEAS
ISSN: 2146-3417 | E-ISSN: 2587-2052
iibfdergi.klu.edu.tr | Yıl (Year): 2020 | Cilt (Volume): 9 | Sayı (Issue): 2 | dergipark.org.tr/tr/pub/klujfeas

DOĞRUSAL OLMAYAN KOVARYANS ANALİZİ MODELLERİ İLE SOSYAL GÜVENLİK VERİLERİ ÜZERİNE UYGULAMA

Cansu MERCAN

Yüksek Lisans Öğrencisi, Sivas Cumhuriyet Üniversitesi, SBE, Ekonometri Bölümü

cansuozkan08@hotmail.com | ORCID: 0000-0001-9986-616X

Necati Alp ERİLLİ

Doç. Dr., Ünvan, Sivas Cumhuriyet Üniversitesi, İİBF, Ekonometri Bölümü

aerilli@cumhuriyet.edu.tr | ORCID: 0000-0001-6948-0880

Özet

Kovaryans analizi deneme etkileri arasında anlamlı bir farklılık olup olmadığını araştıran, kontrol edilebilen faktörlerle deney boyunca kontrol edilemeyen ortak değişken veya değişkenleri de modelde birlikte değerlendirilen bir analiz yöntemidir. Kovaryans analizinin en temel varsayımlarından biri olan doğrusallık varsayımı, kovaryans analizi sonucunda oluşturulan modelin temsil gücünü etkileyen önemli bir etkendir. Doğrusal olmayan veriler üzerinde çok değişkenli kovaryans analizinin kullanılması, modelin temsil gücünü zayıflatacaktır. Bu çalışmada, doğrusal olmayan kovaryans analizi tanıtılmış, doğrusallık varsayımını sağlayan kovaryans analizi ile farklılıkları ortaya konmuş ve yapılan uygulama ile verinin doğrusal olup olmaması veya tam doğrusal durumlarının çok değişkenli kovaryans analizini nasıl etkilediği incelenmiştir.

Anahtar Kelimeler: Kovaryans Analizi, Doğrusallık, Kuadratik ANCOVA, Polinomial ANCOVA.

Bilgilendirme: Bu çalışma, Cansu Mercan tarafından yazılan "Doğrusal Olmayan Kovaryans Analizi Modelleri ile Sosyal Güvenlik Verileri Üzerine Uygulama" isimli yüksek lisans tezinden türetilmiştir.

Etik Beyanı: Bu çalışma "Araştırma ve Yayın Etiği" değerlerine uygun olarak hazırlanmıştır.

Araştırma Makalesi | Makale Geliş Tarihi: 26.06.2020 | Makale Kabul Tarihi: 25.09.2020
Research Article | Article Submission Date: 26.06.2020 | Article Acceptance Date: 25.09.2020

Kırklareli Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi
Kırklareli University Journal of the Faculty of Economics and Administrative Sciences
KLUJFEAS
ISSN: 2146-3417 | E-ISSN: 2587-2052
iibfdergi.klu.edu.tr | Yıl (Year): 2020 | Cilt (Volume): 9 | Sayı (Issue): 2 | dergipark.org.tr/tr/pub/klujfeas

NONLINEAR COVARIANCE ANALYSIS MODELS WITH APPLICATION ON SOCIAL SECURITY DATA

Abstract

Covariance analysis is an analysis method that investigates whether there is a significant difference between the effects of the experiment and the factors that cannot be controlled with the controllable factors and the variables that cannot be controlled throughout the experiment. Linearity assumption, which is one of the most basic assumptions of covariance analysis, is an important factor affecting the representation power of the model formed as a result of covariance analysis. Using multivariate covariance analysis on nonlinear data will weaken the representation power of the model. In this study, nonlinear covariance analysis is introduced. Differences were determined by the covariance analysis which provides the assumption of linearity, and whether the data is linear or exact linear conditions affects the multivariate covariance analysis. As a result of the application, it was found that the analysis to be made on nonlinear data is more appropriate to be done by non-linear covariance analysis.

Keywords: Covariance Analysis, Linearity, Quadratic ANCOVA, Polinomial ANCOVA.

Acknowledgement: This study is derived from the master thesis entitled “Application on Social Security Data with Nonlinear Covariance Analysis Models” written by Cansu Mercan.

Ethics Statement: This study has been prepared in accordance with the values of “Research and Publication Ethics.

Araştırma Makalesi | Makale Geliş Tarihi: 26.06.2020 | Makale Kabul Tarihi: 25.09.2020
Research Article | Article Submission Date: 26.06.2020 | Article Acceptance Date: 25.09.2020

Giriş

İstatistik, geçmiş ve şimdiki durum hakkında toplanmış sayısal verileri, geliştirilmiş olan bazı matematiksel tekniklerle analiz ederek gelecek hakkında karar vermemizi kolaylaştıran bir bilim dalıdır. Temel amaçlarının arasında durum tespiti yapmak, veri gruplarını karşılaştırmak, veri hakkında tahminlerde bulunmak ve karar vermek olan istatistik; hemen her alanda kullanılmaktadır.

Deney tasarımı, kontrol altındaki çeşitli durumların, deney birimlerinin bilinmeyen karakteristik özellikleri üzerindeki etkisini test etmek amacıyla uygulanan bir işlem veya süreç olarak tanımlanabilir. Deneyde birden fazla faktör olduğu durumlarda, faktör düzeylerinin kombinasyonları deneme olarak adlandırılır. Bu gibi birden fazla faktörün mevcut olduğu deneylerde temel amaç, faktörlerin ana etkileri ile beraber etkileşim etkilerini test etmektir. (Şenoğlu ve Acıtaş, 2011: 4).

Deneysel çalışmalarda deneysel hataya bağlı değişkenliği azaltmak ve deneme etkilerinin yansız tahminlerini elde etmek için deneysel kontrol yoluna gidilmektedir. Deneysel kontrol; birimlerin deneme düzeylerine rassal olarak atanması, deneklerin homojen gruplarda toplanması ve istatistiksel kontrol yapılması ile mümkün olmaktadır. İstatistiksel analizlerde önemli yeri olan Regresyon Analizi, Varyans Analizi ve Kovaryans Analizi bu amaçla kullanılan istatistiksel kontrol yöntemlerindendir.

Varyans analizi; ölçü ile belirtilen kitlelerde normal dağılıma uyan üç ya da daha fazla örneklem grupları arasındaki farklılığın önemli olup olmadığını araştıran ve bu farklılığı meydana getiren sebepleri kontrol etmede kullanılan istatistiksel bir tekniktir (Cantay, 2005). Başka bir ifade ile varyans analizi, ikiden fazla örneğin aynı ortalamaya sahip ana kitlelerden gelip gelmediği hakkında karar vermeye yarayan istatistiksel bir teknik olarak da açıklanabilir (Hair vd., 1995). Varyans analizinin genel işleyişi, farklı gözlem gruplarının varyanslarının karşılaştırılmasıyla bu grupların ortalamaları arasında bir fark olup olmadığını belirlenmesi şeklindedir (Lattin vd., 2003).

1. Kovaryans Analizi

Kovaryans analizi (ANCOVA) deneme etkileri arasında anlamlı bir farklılık olup olmadığını araştıran, kontrol edilebilen faktörlerle deney boyunca kontrol edilemeyen ortak değişken veya değişkenleri de modelde birlikte değerlendirilen bir analiz yöntemidir. Kovaryans analizinde ortak değişken veya değişkenlerin, bağımlı değişken üzerindeki etkisi arındırıldıktan sonra deneme etkileri hakkında bir sonuca varılır. Böylece deneyde hata oranı azaltılmış ve testin gücü artırılmış olur (Şahin, 2006; Şenoğlu ve Acıtaş 2011: 251).

Temel olarak kovaryans analizi grupların ortalamaları arasındaki farkın istatistiksel olarak anlamlı olup olmadığını incelemektedir. Kovaryans analizinde bağımlı değişken değerleri, ortak değişkene göre düzeltildikten sonra bağımsız değişkenin bağımlı değişken üzerindeki etkisi analiz edilir. Bu analizde gruplar arasındaki farklılık ölçülürken varyans analizinin yanı sıra regresyon analizi de kullanılmaktadır (Büyüköztürk, 1998).

Kovaryans analizinin iki temel amacı bulunmaktadır. Bunlardan birincisi; sonuçları etkileyebilecek araştırmacının kontrolü dışında kalan sistematik hatayı ortadan kaldırmaktadır. Böylelikle hata terimi küçülecek ve F testinin duyarlılığı artacaktır. İkincisi ise grup üyelerinin belirli karakteristikleri nedeniyle ortaya çıkan sonuçlar arasındaki farklılıklara açıklık getirmektir (Hair vd., 1995).

Bağımlı değişken ile ortak değişken(ler) arasında yüksek bir korelasyon varsa, kovaryans analizi kullanarak analiz yapmak, varyans analizi kullanarak analiz yapmaktan daha avantajlıdır. Kovaryans analizi tekniği kullanılarak yapılan analizler, varyans analizinde olduğu gibi bazı varsayımlara dayanır. Bu varsayımlar aşağıda verilmiştir (Şenoğlu ve Acıtaş, 2011:253):

1. Varsayım (Normallik): ε_{ij} hata terimleri 0 ortalama ve σ^2 varyans ile normal dağılıma sahiptir.

2. Varsayım (Eğimin Anlamlılığı): Ortak değişken ile bağımlı değişken arasındaki ilişkinin doğrusal olduğu varsayılır. Varsayımın doğruluğu, $H_0 : \beta = 0$ hipotezinin sınanmasıyla veya görsel bir yöntem olan ve uygulamada yaygın olarak kullanılan yayılım grafiğinin çizilmesiyle kontrol edilebilir. $H_0 : \beta = 0$ hipotezi ret edilmezse deneme etkileri arasında anlamlı bir fark olup olmadığını belirlemek için varyans analizi kullanmak yeterlidir.

3. Varsayım (Eğimlerin Homojenliği): Regresyon eğimlerinin homojen olduğu varsayılır. Varsayımın doğruluğu, $H_0 : \beta_1 = \beta_2 = ... = \beta_n$ veya H_0 : "Denemelere ait regresyon doğruları birbirine paraleldir" hipotezlerinin sınanmasıyla kontrol edilebilir. H_0 hipotezinin ret edilememesi, denemeler için bulunan regresyon doğrularının birbirine paralel olması demektir. Bu durumda, regresyon doğruları arasındaki mesafeler, ortak değişkenin tüm değerleri için aynı olacağından, önceden belirlenen herhangi bir ortak değişken değeri için, regresyon doğruları arasındaki mesafeleri karşılaştırmak, deneme etkilerini karşılaştırmaya denktir. H_0 hipotezinin ret edilmesi ise en az iki regresyon doğrusunun birbirine paralel olmadığını veya birbirini kestiğini gösterir. Bu durum, ortak değişken ile denemeler arasında etkileşim olduğunu ifade eder.

4. Varsayım (Ölçüm Hatası): Ortak değişkenlerde ölçüm hatasının bulunması, tahmin değerlerinde ve testlerin gücünde olumsuz etkilere yol açar.

5. Varsayım (Sabitlik): Ortak değişkenlerin sabit olduğu varsayılır. Bu varsayım pratikte çok geçerli değildir zira ortak değişkenlerin rasgele olduğu durumlar sabit olduğu

durumlardan daha yaygındır (Şenoğlu ve Acıtaş, 2011: 255; Hair vd., 1995; Weber ve Skillings, 2000).

1.2. Doğrusal Olmayan Kovaryans Analizi

Ortak değişken ve bağımlı değişken puanları arasındaki ilişki her zaman doğrusal olamamaktadır. ANCOVA modelinin altında yatan bir düşünceye göre, X ve Y arasındaki gruplar arası ilişki doğrusal olduğu için, araştırmacılar doğrusal olmama konusunun farkında olamayabilirler. ANCOVA analizinde; veriler doğrusal olmadığında kullanıldığı zaman, F-testinin gücü azalacak ve düzeltilmiş ortalamaların deneme etkileri zayıf gözükebilecektir (Huitema, 1980).

X ve Y arasındaki ilişkinin doğrusal olmaması iki sebepten dolayı olabilir: Değişkenlerin özellikleri ve değişkenlerdeki ölçekleme hataları. Değişkenlerin temel özelliklerinin ölçülebilmesinde doğrusal ilişki olmaması durumu, en önemli sorun olarak karşımıza çıkmaktadır. Doğrusal olmama durumunun sebebine bakılmaksızın, doğrusal ANCOVA modelinin doğrusal olmama durumu çok şiddetli ise doğrusal yöntemlerin uygulanması uygun olmamaktadır (Huitema, 2011:285).

2.2. Doğrusal Olmama Durumu

Verinin doğrusal olup olmadığını kontrol etmek için yapılması gereken ilk iş verinin grafiğini çizmektir. Bu ilk aşama, her gruptaki X gözlemlerine karşılık gelen Y gözlemlerinin birlikte grafiğini çizme ile gerçekleştirilir. Sorun yaratacak doğrusal olmama durumu genellikle serpme diyagramında gözlemlenen trend veya marjinal dağılımlar şeklinde açıkça görülebilmektedir. Doğrusal olmama durumunu belirlemek için daha hassas yaklaşımlar içerisinde ANCOVA modelinin görsel olarak incelenmesi ve verilere çeşitli alternatif modellerin uydurulması tavsiye edilmektedir (Huitema, 2011: 286).

Doğrusal olmama durumunun problem olduğuna karar verildiğinde, bir sonraki adımda şu iki yol izlenmelidir:

i. Değiştirilmiş verilere, doğrusal bir ilişki ile sonuçlanacak orijinal X ve/veya Y puanlarının değişmesi için bir yol aramak.

ii. Orijinal verilere uygun bir polinominal ANCOVA modeli uydurmak.

2.3. Veri Dönüştürme

Eğer X ve Y arasındaki ilişki, doğrusal olmayan bir ilişki olarak tanımlanmış ama monoton ise (yani, X arttıkça Y'nin de artması ama fonksiyonun doğrusal olmaması durumu var ise) X değişkenine dönüşüm uygulanmalıdır. İstenilen doğrusallığı sağladıkları için en sık kullanılan dönüştürme yöntemleri; logaritmik, karekök alma veya ters dönüştürmelerdir (Hunka, 1995).

Dönüştürme yöntemi seçildikten sonra, dönüştürülmüş veri üzerinde her zaman olduğu gibi ANCOVA uygulanır. Örneğin, eğer $\ln X$ ve Y arasındaki ilişkinin doğrusal olduğuna inanmak için sebep varsa, $\ln X$ ortak değişken olarak kullanılır ve ANCOVA uygulanır. Bununla birlikte, analizin yorumlanmasında X değişkeninden ziyade $\ln X$ 'in ortak değişken olduğunu belirtmek gerekir. Dönüştürülmüş ve dönüştürülmemiş veriler için hesaplanan ANCOVA sonuçlarının doğruluğunu tespit etmek için sonuçların birlikte grafikleri çizilerek karşılaştırılmalıdır (Huitema, 2011, s. 287).

2.4. Polinominal ANCOVA Modelleri

Eğer X ve Y arasındaki ilişki monoton değilse, basit bir dönüştürme bile doğrusallığı sağlamayabilir. Doğrusal olmayan ve monoton durumda, X değişkeninin değeri arttıkça Y değişkeninin değeri de artacaktır (Benson ve Hartz, 2000). Doğrusal olmayan ve monoton olmayan durumda ise X değişkeninin değeri arttıkça Y değişkeninin değeri sadece bir noktaya kadar artacak ve sonra X değişkeninin değeri arttıkça Y değişkeninin değeri azalacaktır. Monoton olmayan durumda X değişkenini $\ln X$'e dönüştürdüğümüzde, X değişkeninin değeri arttıkça $\ln X$ değerleri de artacaktır. $\ln X$ ve Y değişkenlerinin birlikte grafiği çizildiğinde doğrusal olmama durumu görülebilecektir. Bu durumda uygulanacak alternatif yöntem; ikinci derece polinominal (kuadratik) bir ANCOVA modelidir. Bu model Eşitlik.1'de verildiği gibi formüle edilir.

$$\overline{Y}_{ij} = \mu + \alpha_j + \beta_1(X_{ij} - \overline{X}_{..}) + \beta_2(X_{ij}^2 - \overline{X}_{..}^2) + \varepsilon_{ij} \quad (1)$$

Eşitlik.1'de kullanılan kısaltmalar, Y_{ij}; j. gruba ait i. bireyin bağımlı değişkene etkisini, μ; bağımlı değişkenin anakitle ortalamasını, ; j. deneme (işlem) etkisini, β_1; doğrusal regresyon katsayısını, β_2; eğim katsayısını, X_{ij}; j. gruba ait i. bireyin ortak değişkene etkisini, $\overline{X}_{..}$; ortak değişkenlerdeki tüm gözlemlerin ortalamasını, X_{ij}^2; j. gruba ait i. bireyin ortak değişkenin karesini, $\overline{X}_{..}^2$; ortak değişkenlerde gözlemlerin ortalamalarının karesini ve ε_{ij}; j. gruba ait i. bireyin hata terimi olarak tanımlanmıştır.

Bu model, eğim etkisi $\beta_2\left(X_{ij}^2 - \left(\overline{X}_{..}^2\right)\right)$ değerini içerdiği için doğrusal modelden farklılık göstermektedir. Eğer bağımlı değişken puanları, ortak değişkenlerin doğrusal fonksiyonu değil de kuadratik bir fonksiyonu ise bu modelin daha uygun bir model olduğu söylenebilir. Böylece uygulanacak metotları test etmek açısından daha başarılı olacaktır. Çoklu ortak değişken analizinde iki eş değişken varsa kuadratik ANCOVA, X ve X^2 kullanılarak hesaplanır.

Temel ANCOVA testi, regresyon testinin homojenliği, düzeltilmiş ortalamaların hesaplanması ve çoklu karşılaştırma testlerinin hepsi sıradan iki ortak değişkenli ANCOVA ile uygulanır. Eğer X ve Y değişkenleri arasındaki ilişki kuadratik fonksiyondan daha karmaşık yapıda ise yüksek derecede bir polinominal yapı daha kullanışlı olabilir. Üçüncü dereceden polinom (kübik) ANCOVA modeli Eşitlik.2'de verildiği gibi yazılır:

$$\overline{Y}_{ij} = \mu + \alpha_j + \beta_1(X_{ij} - \overline{X}_{..}) + \beta_2(X_{ij}^2 - \overline{X}_{..}^2) + \beta_3(X_{ij}^3 - \overline{X}_{..}^3) + \varepsilon_{ij} \quad (2)$$

Ortak değişken ve bağımlı değişken arasındaki ilişki kübik fonksiyon olduğunda bu model daha iyi bir uyum sağlayacaktır. Kübik ANCOVA'da, X, X^2, X^3 ortak değişken olarak kullanılır. Daha karmaşık fonksiyonlar için daha yüksek derecede polinomialler kullanılabilir, fakat bu çeşit durumlarla karşılaşmak son derece olağan dışıdır. Daha yüksek derecede polinomial modeller neredeyse her zaman örnek verilere daha basit polinomial modellerden daha iyi uyum sağlarlar ama bu durum daha karmaşık modellerin daha basitlere tercih edildiği anlamına gelmemektedir (Huitema, 1985). Gerektiğinden daha karmaşık bir model kullanmamaya dikkat edilmelidir. Modeli mümkün olduğu kadar basit tutmak için temelde iki neden vardır. İlk olarak, ANCOVA modele eklenen her yeni tanım için ANCOVA hata kare ortalamasından bir serbestlik derecesi kaybolur. Eğer örneklem sayısı fazla değil ise, serbestlik derecesi kaybı kolayca dengelenebilir. Gerekli olandan daha karmaşık bir model uygulamamanın ikinci sebebi ise tutumluluk ilkesidir.

Eğer doğrusal bir model bir veriye neredeyse kuadratik bir model kadar uyuyorsa, daha basit model seçilmelidir. Böylece sonuçların yorumlanması ve genellenmesi daha kolay olacaktır. Polinomial regresyon modellerin kullanımı ile ilişkili olarak bahsedilen ANCOVA ile ilişkili iki ek nokta daha vardır. İlk olarak, ortak değişkenin sabit bir değişken olması gerekmez. Bazen polinomial regresyonun sadece X sabitine uygun olduğuna dair yanlış bir inanış vardır. İkinci olarak, polinomial regresyon parametrelerini bazı çoklu regresyon yöntemleri ve bilgisayar programları ile hesaplamak bazen zordur. Çünkü bu programlar belirli veri setleri ile gerekli matrisin tersini veremezler (Huitema, 1980). X, X^2, X^3 ve bunun gibi değişkenler, birbirleriyle yüksek derecede ilişkili olduğundan bu problem genişletilebilir. Bu hesaplama güçlükleri genellikle regresyon analizi yapılmadan önce ham X puanlarını sapma puanlarına (yani ortalanmış puanlara) dönüştürerek azaltılabilir. Böylece, kuadratik ANCOVA'da X ve X^2 yerine $(X-\bar{X})$ ve $(X-\bar{X})^2$ ortak değişken olarak kullanılmalıdır.

Düzeltilmiş deneme etkisi, çoklu belirtme katsayıları R^2_{YX} ve $R^2_{YD,X}$'e bağlıdır. R^2_{YX} ; kuadratik regresyon ile açıklanan toplam değişkenlik oranını temsil etmektedir (Bağımlı değişken Y, bağımsız değişkenler X ve X^2 olmak üzere). $R^2_{YD,X}$ ise kuadratik regresyon ile açıklanan toplam değişkenlik oranı ve denemeleri temsil etmektedir. Bu nedenle iki katsayı arasındaki fark, bağımsız kuadratik regresyon ile hesaplanmış olan denemelerin değişkenlik oranını temsil eder. Bu yüzden iki katsayı temsilcisi arasındaki farkın oranı kuadratik regresyon tarafından hesaplanan çözümdür. Açıklanamayan değişkenlik oranı; $\left(1-R^2_{YD,X}\right)$ olarak ifade edilir. Kuadratik ANCOVA'nın genel hali Tablo.1'de verilmiştir (Huitema, 2011: 290):

Tablo 1: Kuadratik ANCOVA tablosu

Değişkenlik Kaynağı	Serbestlik Derecesi	Kareler Toplamı	Kareler Ortalaması	F Hesap Değeri
Düzeltilmiş Değerlendirme	$J-1$	$\left(R^2_{yD,X}-R^2_{yX}\right)KT$	$\frac{\left(R^2_{yD,X}-R^2_{yX}\right)}{(J-1)}$	$\frac{\left(R^2_{yD,X}-R^2_{yX}\right)/(J-1)}{\left(1-R^2_{yD,X}\right)/(N-J-2)}$
Kuadratik Hata$_w$	$N-J-2$	$\left(1-R^2_{yD,X}\right)KT$	$\frac{\left(1-R^2_{yD,X}\right)}{(N-J-2)}$	
Kuadratik Hata$_t$	$N-1-2$	$\left(1-R^2_{yX}\right)KT$		

Tablo.1'de KT; kareler toplamını, N; toplam gözlem sayısını, J ise değişken sayısını ifade etmektedir. Çoklu ANCOVA'da regresyonun homojenlik testinin önemli olması gibi, kuadratik ANCOVA'da da ayrı gruplar için kuadratik regresyonun homojenlik testi önemlidir. Bu test regresyon yüzeylerinin homojenlik testi ile aynı şekilde hesaplanır. Herhangi bir derecedeki kuadratik regresyon modeli için Homojenlik testinin genel formu Tablo.2'de verildği gibidir (Huitema, 2011: 291):

Tablo 2: Kuadratik ANCOVA modeli için homojenlik testinin genel formu

Değişkenlik Kaynağı	Serbestlik Derecesi	Kareler Toplamı	Kareler Ortalaması	F Hesap Değeri
Kuadratik Regresyonun Heterojenliği	$2(J-1)$	$\left(R^2_{yD,X,DX}-R^2_{yD,X}\right)KT$	$\frac{\left(R^2_{yD,X,DX}-R^2_{yD,X}\right)}{(2(J-1))}$	$\frac{\left(R^2_{yD,X,DX}-R^2_{yD,X}\right)}{\left(1-R^2_{yD,X,DX}\right)}\Big/\frac{(2(J-1))}{(N-J-3)}$
Kuadratik Hata$_i$	$N-(J-3)$	$\left(1-R^2_{yD,X,DX}\right)KT$	$\frac{\left(1-R^2_{yD,X,DX}\right)}{(N-J-3)}$	
Kuadratik Hata$_w$	$N-J-2$	$\left(1-R^2_{yD,X}\right)KT$		

Benzer şekilde herhangi bir derecedeki polinomial regresyon modeli için homojenlik testinin genel formu ise Tablo.3'de verildiği gibi olacaktır (Formüllerde belirtilen C değeri homojenlik derecesini ifade etmektedir) (Huitema, 2011: 291):

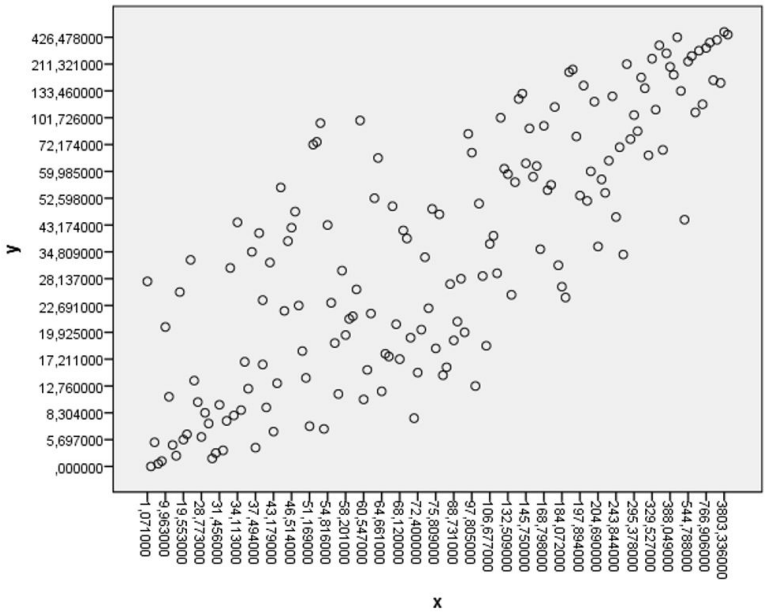
Tablo 3: Polinomial ANCOVA modeli için homojenlik testinin genel formu

Değişkenlik Kaynağı	Serbestlik Derecesi	Kareler Toplamı	Kareler Ortalaması	F Hesap Değeri
Polinomial Regresyonun Heterojenliği	$C\left(J-1\right)$	$\left(R^2_{yD,X,DX}-R^2_{yD,X}\right)K$	$\frac{\left(R^2_{yD,X,DX}-R^2_{yD,X}\right)}{C\left(J-1\right)}$	$\frac{\left(R^2_{yD,X,DX}-R^2_{yD,X}\right)}{\left(C\left(J-1\right)\right)}$ / $\frac{\left(1-R^2_{yD,X,DX}\right)}{\left(N-J-\left(C+1\right)\right)}$
Polinomial Hata$_i$	$N-\left(J-\left(C+1\right)\right)$	$\left(1-R^2_{yD,X,DX}\right)KT$	$\frac{1-R^2_{yD,X,DX}}{\left(N-J-\left(C+1\right)\right)}$	
Polinomial Hata$_w$	$N-J-C$	$\left(1-R^2_{yD,X}\right)KT$		

3. Uygulama

Uygulamada 2018 yılı Kasım ayına ait Sosyal Güvenlik verileri kullanılmıştır (www.sgk.gov.tr). "Sosyal Güvenlik Kapsamında Kişi Sayısı ve Türkiye Nüfusuna Oranı Tablosu" adı altında verilerin doğrusal olmadığı durumda Sosyal Güvenlik Kapsamında Aktif Çalışan Kişi Sayısı ile Bakmakla Yükümlü Tutulanların Sayısının 4/a ve 4/b li Sigortalı Sayısına Çalışan Sigortalı Sayısına Etkisi araştırılmıştır. Verilere, doğrusal ve doğrusal olmayan ANCOVA yöntemleri uygulanmış ve sonuçları tartışılmıştır. Yapılan uygulamalarda SPSS.23 ve Microsoft Excel paket programları kullanılmıştır. İstatistiksel anlamlılık düzeyi olan alfa değeri 0,05 olarak alınmıştır. Şekil.1'de verilen yayılım grafiğine göre verilerin tam doğrusal olmadığı söylenebilir.

Şekil 1: Aktif Çalışan Sayısı ile Sigortalı Çalışanların Yayılım Grafiği



Şekil.1'de verilen tam doğrusal olmayan yapıdaki veri için kuadratik ANCOVA çözümlemesi Tablo.4'de verilmiştir.

Tablo 4: Tam Doğrusal Olmayan Veri için Kuadratik ANCOVA Sonuçları

Değişkenlik Kaynağı	Serbestlik Derecesi	Kareler Toplamı	Kareler Ortalaması	F Hesap Değeri
Denemeler	2	229620,409	114810,204	25,443
Kuadratik Hata	156	703918,305	4512,296	
Genel	158	933538,714		

Çizelge.4'deki sonuçlara göre ANCOVA tablosundan elde edilen hesap değeri $F_{Hesap} = 25,443$ kritik değerinden büyük olduğundan $\left(F_{0,05;2;156} = 3,054\right)$ modelin anlamlı olduğu, deney ve kontrol grupları arasında istatistiksel farkın olduğunu söyleyebiliriz. Böylece kuadratik yani doğrusal olmayan ANCOVA modeli, tam doğrusal olmayan verileri temsil edecek şekilde kabul edilebilir.

Düzeltilmiş ortalama ve çoklu karşılaştırma işlemleri çoklu ANCOVA modeli gibi işlem görür. Düzeltilmiş ortalamalar, R^2_{Y123} üzerinden hesaplanan regresyon eşitliği ile elde edilir. Kesişim ve regresyon katsayıları $\beta_0 = 46,548$, $\beta_1 = -72,761$, $\beta_2 = 0,363$ ve $\beta_3 = -0,000$ olarak hesaplanmıştır.

1.gruba ait ortalama gölge değişken skoru 1, ortalama ortak değişken skoru 213,333 ve ortalama kareler ortak değişken skoru 280341,187 olarak hesaplanmıştır. Buna göre $\overline{Y}_1(düz)$ şu şekilde hesaplanır:

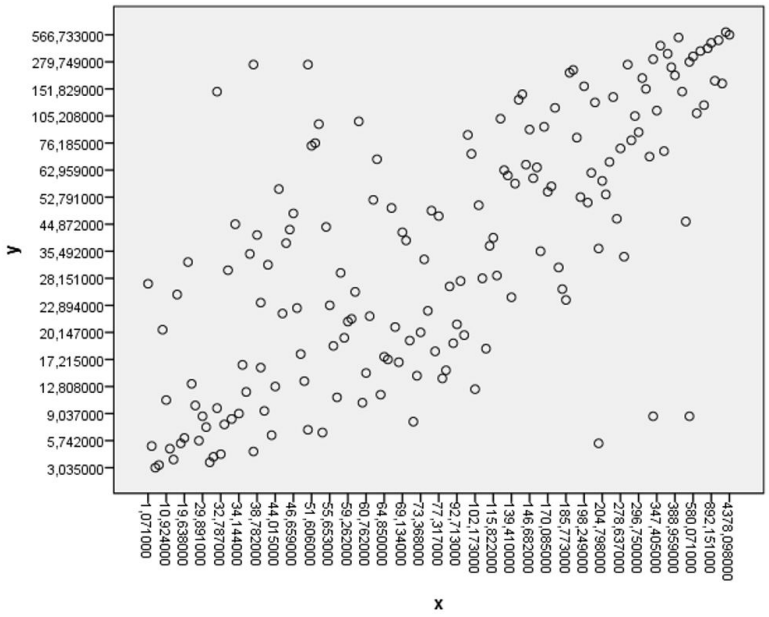
$$\overline{Y}_1(düz) = 46,548 - 72,761(1) + 0,363(213,333) - 0(280341,187) = 51,226$$

Benzer şekilde 2. gruba ait ortalama gölge değişken skoru 0, ortalama ortak değişken skoru 213,333 ve ortalama kareler ortak değişken skoru 280341,187 olarak hesaplanmıştır. Buna göre $\overline{Y}_2(düz)$ şu şekilde hesaplanır:

$$\overline{Y}_2(düz) = 46,548 - 72,761(0) + 0,363(213,333) - 0(280341,187) = 123,987$$

İkinci uygulama verisi ise ilk verilere doğrusallığı bozacak şekilde %3,7 oranında yeni gözlem eklenerek veriler doğrusal yapıdan uzaklaştırılmıştır. Şekil.2'de verilen yayılım grafiğinde uygulamada kullanılan verinin bir önceki uygulamadaki yayılım grafiğindekine göre doğrusallıktan uzaklaştığı görülmektedir.

Şekil 2: Orijinal Verinin %3,7 Arttırılmış Yayılım Grafiği



Şekil.2'de verilen doğrusal olmayan yapıdaki veri için kuadratik ANCOVA çözümlemesi Tablo.5'de verilmiştir.

Tablo 5: Doğrusal Olmayan Veri için Kuadratik ANCOVA Sonuçları

Değişkenlik Kaynağı	Serbestlik Derecesi	Kareler Toplamı	Kareler Ortalaması	F Hesap Değeri
Denemeler	2	255771	127885,72	27,409
Kuadratik Hata	162	755862	4665,814	
Genel	164	1011633		

Tablo.5'deki sonuçlara göre ANCOVA tablosundan elde edilen hesap değeri $F_{Hesap} = 27,409$ kritik değerinden büyük olduğundan $\left(F_{0,05;2;162} = 3,051\right)$ modelin anlamlı olduğu, deney ve kontrol grupları arasında istatistiksel farkın olduğunu söyleyebiliriz. Kuadratik ANCOVA modeli, doğrusal olmayan verileri temsil edecek şekilde kabul edilebilir.

Sonuç

Kovaryans analizi, varsayımların sağlanması durumunda güçlü ve yararlı bir analiz olmaktadır. Hata varyansını en aza indirgemek modelin gücünü artırır. Ayrıca kovaryans analizi, küçük örneklemlere ya da etki büyüklüğünün küçük olduğu durumlarda uygulandığında daha anlamlı sonuçlar vermektedir.

Geleneksel ANCOVA modeli varsayımlarındaki bağımlı değişken ve bağımsız değişkenler arasındaki ilişki her zaman doğrusal olmamaktadır. Yapılan analizlerde çok değişkenli kovaryans analizinin temel varsayımlarından olan doğrusallık varsayımının sağlanmaması problem teşkil etmektedir. Eğer doğrusal olmayan veriler üzerinde çok değişkenli kovaryans analizi kullanılıyorsa, kovaryans analizi modelinin temsil etme gücü zayıflamaktadır. Bu yüzden de çok değişkenli kovaryans analizi modellerinde doğrusallık şartının sağlanması önemlidir. Şiddetli doğrusal olmama durumu, grup içi XY yayılım grafiği ile kolaylıkla ortaya çıkarılabilecektir.

Eğer ilişki doğrusal değil fakat monoton ise basit dönüşümlere (genellikle X'e uygulanacak dönüşümler) Y ve dönüştürülmüş X arasında doğrusal ilişki bulunabilecektir. Değişkenlere dönüşüm uygulandıktan sonra verilere kovaryans analizi uygulanabilecektir.

Modelin doğrusal olmaması durumunda kovaryans analizi uygulanabilmesi için modelin doğrusal modele dönüştürülmesi gereklidir. Eğer bağımlı değişken ve bağımsız değişken arasındaki ilişki hem doğrusal olmayan hem de monoton değilse -yani bağımlı değişken artarken bağımsız değişken yalnızca bir noktaya kadar artıyor ve sonra bağımlı değişken artarken bağımsız değişken de artıyor ise- basit dönüşüm yaklaşımı da yeterli olmayacaktır. Bu durumda bağımsız değişkene polinomial yaklaşım uygulanacaktır. Genellikle bu durumda, kuadratik veya kübik Ancova modelleri kullanılmalıdır. Karmaşık polinom modellerin kullanımı sadece daha basit

modellerse eğer yeterli olabilir. Aksi halde polinom modeller de yetersiz olacaktır. Daha basit olan modeller tercih edilir çünkü karmaşık modellere dayalı sonuçlar daha zor yorumlanır ve genel olarak daha az kararlıdır. Polinom ANCOVA modelleri açıkça uygulanabildiğinde hesaplamalar çoklu ANCOVA'nın bir uzantısını içerir.

Uygulamada iki farklı veri kullanılmıştır. İki model için de hesaplanan F değerleri karşılaştırıldığında, ortak değişken ile bağımlı değişken arasındaki ilişki düşük olduğunda, kuadratik ANCOVA modelinin F değeri daha güçlü sonuçlar vermiştir. Yani doğrusal olmayan veriye uygulanacak en uygun modelin kuadratik ANCOVA modeli olduğu yapılan uygulamalarla gösterilmiştir.

Sonuç olarak, doğrusal olmayan veriler için uygulanacak en uygun modelin kuadratik ANCOVA modeli olduğu, doğrusal olduğu kabul edilen veriler için de doğrusal ANCOVA modeli olduğu değerlendirilmiştir. Ayrıca veri ne kadar doğrusal ise doğrusal ANCOVA modelinin F değerinin daha güçlü sonuçlar verdiği, verinin doğrusal olmadığı durumlarda da kuadratik ANCOVA modelinin F değerinin daha güçlü sonuçlar verdiği değerlendirilmiştir.

Her ne kadar doğrusal olmayan veriler için kuadratik ANCOVA modelinin daha uygun olsa da, analiz verileri için daha az karmaşık modelin tercih edilmesi daha sağlıklı sonuçlar verecektir. Eğer doğrusal olmayan bir veri için doğrusal model de uygunluk sağlıyorsa, daha az karmaşık model olan doğrusal modelin tercih edilmesi gerekmektedir. Yapılacak olan çalışmalarda doğrusal olmayan veriler için doğrusal olmayan ANCOVA analizinin doğruluğunun tespiti için daha büyük gözlem verileri kullanılarak simülasyon teknikleri kullanılabilir.

Kaynakça

Benson, K., Hartz, A.J., (2000). A comparison of observational studies and randomized controlled trials: Special articles. *New England Journal of Medicine*, 342, 1878–1886.

Büyüköztürk, Ş. (1998). *Kovaryans analizi (varyans analizi ile karşılaştırmalı bir inceleme).* Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi. Cilt:31, Sayı:1, s:1301.

Cantay, T. (2005). *Tesadüfi bloklar düzeninde kovaryans analizi ve uygulaması*. Selçuk Üniversitesi, Fen Bilimleri Enstitüsü, Yayınlanmamış Yüksek Lisans Tezi.

Hair J.F., Anderson R.E., Tatham R.L., Black W.C. (1995). *Multivariate data analysis with readings*. USA: Prentice Hall Pub.

Huitema, B.E., (1980). *The analysis of covariance and alternatives*. New York: Wiley and Sons Pub.

Huitema, B.E. (1985). Autocorrelation in applied behavior analysis: A myth. *Behavioral Assessment*, 7, 109–120.

Huitema, B.E. (2011). *The analysis of covariance and alternatives*. New Jersey: A. John Wiley And Sons, Inc., Publication.

Hunka, S. (1995). Identifying regions of significance in ANCOVA problems having nonhomogeneous regressions. *British Journal of Mathematical and Statistical Psychology*, 48, 161–188.

Lattin, J.M., Carroll D., Green, P. (2003). Analyzing Multivariate Data, Pacific Grove, CA : Thomson Brooks/Cole Canada.

Şahin, H. (2006). *Kovaryans analizi ve bir uygulama*. Gazi Üniversitesi, Fen Bilimleri Enstitüsü, Yayınlanmamış Yüksek Lisans Tezi.

Şenoğlu, B., Acıtaş, Ş. (2011). *İstatistiksel deney tasarımı*, Ankara: Nobel Yayınları (2. Basım).

Weber, D.C., Skillings, J.H. (2000). *A first course in the design of experiments: A Linear Models Approach*. New York: Crc Press Llc.

Kırklareli University Journal of the Faculty of Economics and Administrative Sciences (KLUJFEAS) ISSN: 2146-3417 | E-ISSN: 2587-2052
iibfdergi.klu.edu.tr | Year: 2020 | Volume: 9 | Issue: 2 | dergipark.org.tr/en/pub/klujfeas
Cansu MERCAN, Necati Alp ERİLLİ
Nonlinear Covariance Analysis Models With Application On Social Security Data, 315-331.

NONLINEAR COVARIANCE ANALYSIS MODELS WITH APPLICATION ON SOCIAL SECURITY DATA

Cansu MERCAN, Necati Alp ERİLLİ

Extended Abstract

Covariance analysis (ANCOVA) is an analysis method that investigates whether there is a significant difference between the effects of the experiment, and the common variables or variables, which cannot be controlled during the experiment, with controllable factors, are also evaluated together in the model. In covariance analysis, after the effect of the common variable or variables on the dependent variable is purified, a conclusion about the test effects is reached. Thus, the error rate in the experiment is reduced and the power of the test is increased.

Multivariate covariance analysis investigates whether the difference between group averages is statistically significant by examining multiple dependent variables at the same time. Linearity assumption, which is one of the most basic assumptions of covariance analysis, is an important factor affecting the representation power of the model created as a result of covariance analysis. Using multivariate covariance analysis on nonlinear data will weaken the representation power of the model.

The relationship between common variable and dependent variable scores is not always linear. According to an idea underlying the ANCOVA model, researchers may not be aware of the nonlinearity, since the relationship between groups between X and Y is linear. In ANCOVA analysis; When the data is used when it is not linear, the strength of the F-test will decrease and the trial effects of the corrected averages may appear weak.

The nonlinear relationship between X and Y can be for two reasons: properties of variables and scaling errors in variables. The fact that there is no linear relationship in measuring the basic properties of variables is the most important problem. Regardless of the cause of nonlinearity, if the nonlinearity of the linear ANCOVA model is very severe, it is not appropriate to apply linear methods.

If the relationship between X and Y is not monotonous, even a simple conversion may not provide linearity. In the nonlinear and monotonous state, the value of the Y variable will increase as the value of the X variable increases. In the nonlinear and non-monotonous state, as the value of the X variable increases, the value of the Y variable will only increase to a point, and then the value of the Y variable will decrease as the value of the X variable increases. When we convert the X variable to 'in the non-monotonous state, the values of the X variable will increase as the value increases.

When the variables of Y and Y are plotted together, nonlinearity can be seen. In this case, the alternative method to be applied; is a second degree polynominal (quadratic) ANCOVA model.

In this study, nonlinear covariance analysis is introduced. The covariance analysis, which provides the linearity assumption, has revealed its differences and how the data is linear or full linear has been examined with the application.

With the study, the effect of the number of people who are active in the scope of social security and the number of people who are obliged to take care of the number of 4 / a-4 / b employees, and the effect of the number of compulsory insured and the number of voluntary insured on gender were analyzed by statistical analysis. In statistical analysis; multivariate linear covariance analysis and multivariate nonlinear covariance analysis was used, and it was found that it would be more appropriate to perform analysis on nonlinear data by nonlinear covariance analysis.

Covariance analysis is a powerful and useful analysis if assumptions are provided. Minimizing the error variance increases the power of the model. In addition, covariance analysis gives more meaningful results when applied to small samples or when the effect size is small.

The relationship between dependent variable and independent variables in traditional ANCOVA model assumptions is not always linear. Failure to provide the linearity assumption, which is one of the basic assumptions of multivariate covariance analysis, poses a problem. If multivariate covariance analysis is used on nonlinear data, the representation power of the covariance analysis model is weakened. Therefore, it is important to provide the linearity condition in multivariate covariance analysis models. Severe non-linearity can be easily detected with the intra-group XY scatter plot.

If the relationship is not linear but monotonous, simple transformations (usually transformations to X) can have a linear relationship between Y and transformed X. After applying the transformation to variables, covariance analysis can be applied to the data.

If the model is not linear, the model must be converted to a linear model in order to apply covariance analysis. If the relationship between the dependent variable and the independent variable is not both linear and monotonous, that is, while the dependent variable increases, the independent variable only increases up to one point, and if the dependent variable increases while the independent variable increases, the simple transformation approach will not be sufficient. In this case, polynomial approach will be applied to the independent variable. Usually in this case, quadratic or cubic Ancova models should be used. The use of complex polynomial models can only be sufficient if they are simpler models. Otherwise, polynomial models will also be insufficient. Simpler models are preferred because results based on complex models are more difficult to interpret and generally less stable. When polynomial ANCOVA models are explicitly applicable, calculations include an extension of multiple ANCOVA.

Two different data were used in the application. When the F values calculated for both models are compared, when the relationship between the common variable and the dependent variable is low, the F value of the quadratic ANCOVA model gave stronger results. In other words, the most appropriate model to be applied to nonlinear data has been demonstrated with the applications made as the quadratic ANCOVA model.

As a result, it was evaluated that the most suitable model to be applied for nonlinear data is the quadratic ANCOVA model and for the data considered as linear, it is the linear ANCOVA model. In addition, the more linear the data, the F value of the linear ANCOVA model gives stronger results, and the quadratic ANCOVA model's F value gives stronger results when the data is not linear.

Although the quadratic ANCOVA model is more suitable for nonlinear data, choosing the less complex model for analysis data will yield healthier results. If the linear model is also suitable for a nonlinear data, the less complex model, the linear model, should be preferred. In the studies to be carried out, simulation techniques can be used by using larger observation data to determine the accuracy of nonlinear ANCOVA analysis for nonlinear data.