

DISKRİMİNANT ANALİZİ VE DOĞRUSAL OLASILIK MODELLERİ ARASINDAKİ BENZERLİK VE FARKLAR

Dr. Talat FİRLAR

İ.Ü., Teknik Bilimler M.Y.O, Yardımcı Doçent

ABSTRACT: *This paper discusses these methods (Discriminant analysis and Linear probability models) and compares them. I believe that the results obtained from this stage will form a basis for the following stager (thematic).*

I-DOĞRUSAL OLASILIK MODELLERİ:

Kesikli bağımlı değişkenleri ölçmek amacıyla kurulan regresyon modelleridir. Bu Modellerin diğer regresyon modellerinden en önemli farkı , bağımlı değişkenin gözlenen birimin tercihine bağlı olarak değer almasıdır. Birimlerin tercihleri ikiden fazlaysa ,o zaman çoklu olasılık modelleri söz konusu olur.

Diskriminant Analizi ise, gözlem değerlerinin iki veya daha çok gruptan hangisine ait olduğunu araştırır ve her gözlemi taşıdığı özelliklere göre sınıflandırır, ilgili gruba ayırır. Bu Çalışmada Diskriminant Analizi ve Doğrusal olasılık Modelleri arasındaki ilişkiler ele alınacak, her iki yöntemin benzerlik ve farkları ortaya konulacaktır.

Diskriminant analizi; bir gözlem değerini, daha önce bazı özelliklere göre belirlenen gruplardan biri içinde sınıflandırmaya yarayan bir istatistiki teknik olarak ifade edilmektedir. Çok değişkenli diskriminant analizi, doğrusal ve kuadratik diskriminant analizi olmak üzere iki gruba ayrılır.

Bir doğrusal diskriminant fonksiyonu aşağıdaki gibi ifade edilebilir;

$$Z=B1*X1+B2*X2+.....+Vn* Xn$$

Burada,

Z= Diskriminant değerini,

B= Diskriminant katsayılarını,

X= Bağımsız değişken olan finansal oranları göstermektedir.

Görüldüğü gibi, diskriminant değeri (Z) bağımsız değişkenlerin doğrusal bir fonksiyonudur. Dolayısıyla,

herhangi bir bağımsız değişkenin etkisini yorumlamak kolaydır.

Kuadratik diskriminant fonksiyonu ise, daha karmaşık bir fonksiyon olup, bir değişkenin etkisini diğerlerinden arındırarak yorumlamak kolay değildir. 3 değişkenli bir kuadratik diskriminant fonksiyonu aşağıdaki gibidir.

$$Zi=B1*X1+B2*X2+B3*X2+B4*X1+B5*X1*X2+B6*X1*X3+B7*X2+B8*X2*X3+B9*X3$$

Yukarıdaki fonksiyondan görüldüğü gibi, "X"deki bir birim artışın Z üzerindeki etkisini yorumlamak için sadece, "B1" katsayısına bakmak yeterli olmayacaktır. Ayrıca, B4, B5 ve B6 gibi katsayıların da değerine bakmak gerekecektir. Şüphesiz ki; bu da kuadratik diskriminant fonksiyonunu yorumlamayı güçleştirmektedir. Bu nedenden dolayı, çoğu çalışmada doğrusal diskriminant analizi tercih edilmiştir.

Örneğin Diskriminant Analizi; işletme faaliyetini açıklayacak, işletme faaliyetindeki değişiklikleri özet olarak ortaya koyabilecek bir yöntemdir. Bir başka ifadeyle; diskriminant analizi yöntemi; finansal tablolarda önemli değişiklikler olup-olmadığını ve değişikliklerin en çok nereden kaynaklandığını açıklar.

Herhangi bir araştırmada kullanılan bireyler (veya birimler), gruplara ayrılırken birtakım önbilgilere göre subjektif ölçümlerle sınıflandırılmaktadır. Halbuki, sınıflandırma diskriminant analizi ile yapıldığında bireyleri gruplara ayırma objektif kriterlere göre yapılmaktadır. Bir başka ifadeyle, subjektif kriterlere göre yapılan ayırmada, ayırma konu olan birey ya da bireylerin sadece bir ya da birkaç özelliği dikkate alınarak sınıflandırılma yapılırken, diskriminant analizinde ise sözkonusu sınıflandırma, bireyin ait olduğu grup içindeki tüm özellikleri gözönüne alınarak gerçekleştirilmektedir. Böylece, diskriminant fonksiyonu; gözlem birimlerine ait birden çok değişkeni dikkate aldığından bunların yığılma göre sınıflandırılmasında, değişkenler arasındaki ilişkileri de gözönünde bulunduran bir özelliğe sahiptir.

Gözlemleri (işletmeleri), iki veya daha çok sayıdaki yığından birine ayırdetme, bir karar sonucudur. Bu nedenle, diskiriminant analizi bir istatistiksel karar alma, diskiriminant fonksiyonu ise karar almada kullanılan bir çeşit karar fonksiyonu olmaktadır.

Diskiriminant analizinin amacı; hatalı sınıflandırmayı en aza indirgeyecek bireyleri ait olduğu gruplara ayırmaktır. Diskiriminant fonksiyonu optimizasyona dayalı bir karar fonksiyonudur. Sözkonusu optimizasyon toplam yanlış sınıflandırma olasılığının en az yapılmasıdır. Gözlem birimlerinin yığınlara ayırıldırmada iki çeşit hata sözkonusu olmaktadır. Şöyle ki, G1 ve G2 gibi iki yığından birine ait bir gözlem gerçekte G1 yığına ait olduğu halde karar sonucu G2 yığına dahil edilse veya gerçekten G2 yığına ait olduğu halde karar sonucu G1 yığına dahil edilse, her iki durumda da gözlem yanlış sınıflandırılmış olmaktadır[1].

İşte bu gibi durumlarda, gözlemlerin hangi yığınlara dahil edileceği diskiriminant analiz tekniği ile belirlenir.

Çeşitli nedenlerle gözlemlerin iki veya daha çok yığından hangisine ait olduğunun bilinmesine gerek duyulur. İşte, birden çok özellik açısından gözlemlerin, yığınlara ayırıldırmaları sorunu, diskiriminant analizinin konusudur.

Diskiriminant fonksiyonunun elde edilmesi için çeşitli yaklaşımlar bulunmaktadır. Bu yaklaşımlar; genel yaklaşım, Bayes teoremi yaklaşımı, olabilirlik oran yaklaşımı ve Fisher'in varyans analizi yaklaşımıdır. Belirli varsayımlar altında tüm yaklaşımlar aynı sonucu verir. Yaklaşımlardaki farklılık, sadece olaya bakış açısından kaynaklanmaktadır. Temelde tüm yaklaşımlar toplam yanlış sınıflandırma olasılığını en aza düşürmek, ya da doğru tahmin yüzdesini en yükseğe çıkarmaktır.

II-REGRESYON TEKNİĞİ:

Değişkenler arasındaki ilişkiyi saptama yollarından biridir[2]. Regresyon analizi ise; iki değişken arasındaki ortalama ilişkinin matematiksel fonksiyonla ifade edilmesidir[3]. Regresyon analizinin temel amacı; iki ya da daha fazla sayıdaki bağımsız değişken ile bir bağımlı değişkenin tahmin edilmesi ya da tanımlanmasıdır. Bir başka deyişle; regresyon analizi tekniği ile iki ya da daha çok sayıdaki değişken arasında bir bağıntının olup-olmadığı ve bu bağıntının bir denklem (fonksiyonla) ile nasıl ifade edileceği araştırılabilir[4].

Teknik; verilerin ölçülebilir olduğu, bir bağımlı değişkenin iki veya daha çok fonksiyonun olduğu hallerde kullanılır. Tekniğin kullanılmasından amaç; bağımlı olarak kabul edilen değişkendeki değişimi (varyasyonu) bütün

bağımsız değişkenlerle olan birlikte değişime (kovaryans) dayanarak tahmin etmektir.

Çok değişkenli regresyon analizinde kullanılan fonksiyon aşağıdaki gibidir;

$$Y_i = B_0 + B_1X_1 + B_2X_2 + \dots + B_nX_n + U$$

Y_i = Regresyon derecesi,

B_0 = Regresyon denkleminin dikey eksenini kestigi nokta,

B = Regresyon katsayıları,

X_i = Bağımsız değişken,

U = Hata terimi

Katsayıların hesaplanmasında kullanılan temel yöntem 'en küçük kareler' yöntemidir.

III-LOJİSTİK REGRESYON ANALİZ TEKNİĞİ

Değişkenlerin bazılarının sürekli, bazılarının kesikli olması durumunda; diskiriminant analizine alternatif olarak lojistik regresyon önerilmektedir. Bir başka deyişle doğrusal olasılık modelinin dezavantajlarını ortadan kaldırmak için lojistik (logit) ve probit analiz tekniği kullanılmaktadır[5].

Sonuç değişkenin ikiden çok düzeyli kesikli değişken olması durumunda; bu değişkenin açıklayıcı değişkenler üzerindeki regresyon modeli, çoklu grup lojistik model adını alır.

IV-PROBİT ANALİZ TEKNİĞİ

Son zamanlarda, İşletmelerin başarısızlık tahmininde çok fazla tercih edilen modellerden biri, probit analiz tekniğidir. Bunun nedeni; çok değişkenli diskiriminant analizi ile çok değişkenli regresyon analiz tekniğine göre taşıdığı kuramsal üstünlüktür.

Zira; bir sorun olarak, doğrusal olasılık fonksiyonları olan çok değişkenli diskiriminant ve regresyon fonksiyonlarının parametreleri hesaplandıktan sonra, tahmini bağımlı değişken değerinin 0-1 aralığı dışına taşma ihtimali bulunmaktadır.

İşte; bu sorun, yani Z değerinin bağımsız değişken ya da değişkenlerin alacağı değer ne olursa olsun, 0-1 aralığında tutulabilmesi çözülür. Bu da; ancak birikimli bir olasılık fonksiyonunun kullanılmasıyla mümkündür. Probit,

birikimli olasılık fonksiyonu olduğundan, yukarıda ortaya çıkması mümkün sorunun çözümünde tercih edilir.

Bu yöntem, bir gözlem değerini daha önce bazı özelliklerine göre belirlenen gruplardan biri (başarılı ve başarısız) içinde sınıflandırmaya yarayan bir istatistik teknikidir.

Çok değişkenli diskriminant analizi, birbiriyle ilişkili, ancak farklı özellikteki sorunun çözümünde sağladığı sonuçlar açısından yararlı kabul edilmektedir. Bu sorunlar şunlardır[6];

1) Bir grup açıklayıcı değişkene (finansal oranlar gibi) bağlı olarak, iki ya da daha fazla grup (finansal açıdan başarılı ve başarısız işletme grupları gibi) arasında önemli bir farklılık olup-olmadığını belirlemek,

2) Başlangıçtaki açıklayıcı değişken sayısından daha az sayıda değişkenle gruplar arasındaki önemli farklılıkları açıklamak,

Elde edilen diskriminant fonksiyonu yardımıyla, grup üyeliği konusunda geleceğe yönelik öngörüde bulunmak.

Analize esas alınacak varsayımlar aşağıdaki gibidir;

i) Birbirinden farklı iki veya daha fazla grup söz konusu olmalıdır.

ii) Veriler yığından tesadüfi örnekleme yolu ile çekilmelidir.

iii) Bağımsız değişkenler çok boyutlu normal dağılım göstermelidir.

iv) Grupların sapma matrisleri birbirine eşittir.

Bu varsayımlardan başka, analizde zımnen kabul edilen diğer iki varsayım ise şunlardır;

v) Yığını oluşturan elamanların önsel olasılıkları bilinmektedir.

vi) Herhangi bir şirketin yanlış sınıflandırma maliyeti önceden bilinmektedir.

Bu varsayımlar, tartışmalı bir konudur. Analize ya da modele dahil edilen bağımsız değişkenler arasındaki korelasyon, yığın ve örnek sorunu, yaygın olarak tartışılmaktadır.

Bu arada, verilerin (finansal oranlar gibi) normal dağılımdan sapma gösterdiği durumda, çok boyutlu

modellerden doğrusal ve kuadratik diskriminant analizi ile çok değişkenli regresyon analiz modelinin uygulamasında sorun doğabilmektedir. Özellikle, modelin tahmin gücünün düşük çıktığında bu durumun payı olduğu düşünülür, tersine modelin tahmin gücü etkilenmemişse ya da yüksekse, verilerin normal dağıldığı ya da normale yakın dağılım gösterdiği düşünülür.

Çok değişkenli diskriminant analizinin; rastgele seçilmiş örnekleri sınıflandırmak amacıyla kullanılabilmesi için ön olasılıklarının ve eğer analiz sonucu elde edilen fonksiyonun karar verme amacıyla kullanılabilmesi için de yanlış sınıflandırma maliyetinin bilinmesi gerekmektedir.

V-SONUÇ

Modellerin genel başarısını ölçen F istatistiği her iki fonksiyon için aynıdır.

Her parametreyi tek tek test eden t testi ise, Diskriminant fonksiyonunda olduğu gibi doğrusal olasılık modelinde, bağımlı değişkenin ikili yapısına rağmen, ($n_1 + n_2 - k - 1$) serbestlik derecesiyle t dağılımına uygun olarak yapılır.

Her iki fonksiyonun başlangıç noktaları ve dağılım özellikleri tamamen farklıdır. Böyle olmasına rağmen her iki fonksiyonunda sonuçları aynıdır.

Regresyon analizinde genellikle bir sıfır hipotez ve bu hipoteze karşı bir alternatif hipotez tanımlanır. Eğer sıfır hipotez belli hata paylarıyla reddedilirse alternatif hipotez aynı hata payı ile kabul edilir. Diskriminant analizinde ise eşit koşullarda belirlenmiş iki grup vardır. Bir birim bir gruba ait olarak sınıflandırılmıyorsa, diğer gruba ait olarak sınıflandırılmalıdır.

Diskriminant analizi ile doğrusal olasılık modelinin en temel farkı değişkenlerinin dağılımlarıdır. Diskriminant analizinde belirlenen grupların normal dağıldığı varsayımı yapılmaktadır. Oysa doğrusal olasılık fonksiyonu daha önce de söylendiği gibi binom dağılımına uyar. Doğrusal olasılık modelinin yapısından kaynaklanan aksaklıkları gidermek için geliştirilmiş logistik fonksiyonda birikimli logistik dağılıma dayanır ve parametre tahmini genellikle en çok benzerlik yöntemi ile yapılır. Gerek doğrusal olasılık modeli gerekse logistik formun uygulandığı regresyon modellerinde tüm klasik regresyon modellerinde olduğu gibi bağımsız değişken x'in deterministik olduğu ve stokastik y bağımlı değişkenin hata terimi ile aynı dağılıma uyduğu varsayılır. Oysa diskriminant fonksiyonunda bağımlı değişken y kesindir ve bağımsız değişken x tesadüfidir. Doğrusal olasılık modelinde belli bir x değerine karşılık y'nin koşullu olasılığı belirlenirken diskriminant analizinde x ve y nin bileşik olasılıkları belirlenir. Doğrusal

olasılık modelinde x in belirlenmesi ,öncelikle y nin belirlenmesine bağlıdır. Diskriminant analizinde ise y nin belirlenmesi x in belirlenmesine bağlıdır[7].

Diskriminant analizi tahmincileri, logistik regresyondan ,en çok benzerlik yöntemi ile elde edilen tahmincilerle kıyaslanırsa , bağımsız değişkenin normal dağıldığı durumda, diskriminant analizi tahmincileri,en çok benzerlik tahmincileri ile aynı sonucu verirler. Hatta kimi durumlarda diskriminant analizi tahmincileri, asimptotik olarak en çok benzerlik tahmincilerinden daha etkindir.

KAYNAKÇA

- 1-Fırlar, Talat; Diskriminat Analizi ve Bir Uygulama, İ.Ü. S.B.E., Yayınlanmamış Doktora Tezi, İstanbul 1989
- 2-Kaptan, Saim; Bilimsel Araştırma ve İstatistik Teknikleri, Ankara 1995
- 3-Türkbal, Aydın; Bilimsel Araştırma Metodları ve Uygulamalı İstatistik, A.Ü. Yayınları, Erzurum 1981
- 4-Cinemre, Nalan; “Grupların Karşılaştırılmasında Birer İstatistiksel Yöntem Olarak Regresyon Analizi ve Ayrışım Analizi”, Dokuz Eylül Ü.İ.İ.B.F.Dergisi, İzmir 1987, Sayı:2
- 5-Özmucur, Süleyman; Geleceği Tahmin Yöntemleri, İstanbul Sanayi Odası Yayını, İstanbul 1990
- 6-F.M. Richardson-L.F.Davidson; “On Linear Discriminantion with Accounting Ratios”, Journal of Business Finance and Accounting, c.11, No:4, Kış 1984
- 7-Amemiya T. “Qualitative Response Models: A Survey”, Journal Of Economic Literature 19, December 1981.