



Lojistik Regresyonda Aşırı Yayılım Durumunda Williams Yönteminin Etkinliği: BİST 100 Firmaları için Mali Başarısızlığın Belirlenmesi

Funda H. Sezgin

Yrd.Doç. Dr., İstanbul Üniversitesi, Mühendislik Fak.,Endüstri Mühendisliği Böl.
hfundasezgin@yahoo.com

Öz

Lojistik regresyonda, gözlenen varyansın beklenen varyansdan büyük olması aşırı yayılım olarak tanımlanmaktadır. Problemin belirlenmesinde, Pearson ki-kare ile sapma uyum istatistiklerinden yararlanılır. Aşırı yayılım durumunda model uyum istatistiklerine ve parametre tahminlerine güvenilmeyeceği için, bu durumu dikkate alan farklı yöntemler kullanılmalıdır. Bu yöntemlerin en bilineni Williams metodudur. Bu çalışmada, BİST 100 endeksinde işlem gören firmalara yönelik mali başarısızlığı belirlemek amaçlı bir sınıflama lojistik regresyon analizi ile gerçekleştirilmiştir. Analiz sonucunda, aşırı yayılım tespit edilerek, Williams metodu ile parametreler yeniden tahmin edilmiş ve doğru sınıflama oranının arttığı gözlenerek yöntemin etkinliği belirlenmeye çalışılmıştır.

Anahtar kelimeler: Aşırı Yayılım, Lojistik Regresyon, Mali Başarısızlık, Faktör Analizi, Williams Yöntemi

Effectiveness of Williams Method in the Case of Overdispersion in Logistic Regression: Application of Financial Distress for BIST 100 Index

Abstract

In logistic regression, having higher observed variation than expected variation is called as overdispersion. Pearson chi-square and deviance goodness of statistics are used to determine overdispersion in logistic regression. Different methods can be used in the case of overdispersion where goodness of fit of the model and parameter estimations are not confidential. Williams method is widely used for overdispersion situations. In this study, logistic regression analysis was built in order to classify firms of BIST 100 index for financial distress. At the end of this analysis overdispersion was determined and parameters were re-estimated by Williams method in order to detect the efficiency of the method.

Keywords: Overdispersion, Logistic Regression, Financial Distress, Factor Analysis, Williams Method

Giriş

Lojistik regresyon, bağımlı değişkeninin kategorik olarak, ikili veya çoklu gözlemlendiği durumlarda bağımsız değişkenlerle sebep-sonuç ilişkisini belirlemede yararlanılan bir yöntemdir. Bağımsız değişkenlerin değerlerine göre bağımlı değişkenin beklenen değerinin olasılığını belirleme yöntemi olan lojistik regresyon, aynı zamanda bağımsız değişkenlerin etkilerine dayanarak verilerin sınıflandırılmasında da kullanılabilir.

Bağımlı değişkenin sürekli olduğu ve özellikle artıklar üzerinde normalliğin yanı sıra çeşitli varsayımların yapıldığı doğrusal regresyon modelinde, bağımlı değişkenin iki ya da daha çok düzey içeren kesikli bir değişken olması durumunda doğrusal regresyon modeli için öngörülen varsayımlar bozulmakta ve hata teriminin binom dağılım göstermesi sebebiyle gözlem varyansları eşit olmamaktadır. Bu durumda, verilerin analizi için lojistik regresyon analizinin kullanılması önerilmektedir (İyit ve Genç, 2005). Böylece, lojistik regresyon modeli, genelleştirilmiş doğrusal modellerin binom dağılımlı bağımlı değişkenler için elde edilmiş olan özel bir biçimi olarak tanımlanabilir. Ayrıca, sürekli ve kategorik değişkenlerin birlikte analiz edilmesine olanak tanıyarak büyük kolaylık sağlamaktadır.

Lojistik regresyonda, gözlenen varyansın, beklenen varyanstan büyük olması aşırı yayılım olarak tanımlanmaktadır. Sapma (deviance) ve Pearson ki-kare uyum istatistikleri lojistik regresyonda yayılım olup olmadığını belirlemede kullanılan istatistiklerdir. Eğer, uyum istatistikleri sonucunda aşırı yayılım söz konusu ise, aşırı yayılımı dikkate alarak çözümleme yapan modellerin kullanılması gerekmektedir. Lojistik regresyonda, aşırı yayılımı açıklayan yöntemlerden biri de Williams analizidir. Bu yöntemde, aşırı yayılım modele dahil edilerek giderilmeye çalışılır (Czado, 2003). Aşırı yayılımın göz ardı edilmesi durumunda, uyum istatistikleri beklenen düzeyde gerçekleşmeyeceği gibi sınıflama olasılıklarına güven kalmayacaktır. Aşırı yayılımın nedenleri; başarı olasılıkları arasındaki değişim, ikili cevaplar arasındaki korelasyon, yanlış bağlantı fonksiyonunun kullanılması, modelde olması gereken bağımsız değişkenlerin dışarıda kalması ve örnek büyüklüğünün yetersiz kalması olarak sıralanabilir.

Bu çalışmada, BİST 100 endeksinde işlem gören hisse senetlerinin Eylül 2016 dönemine ilişkin verileri için mali başarısızlığının belirlenmesine yönelik bir sınıflama analizi yapılmış, aynı zamanda mali başarısızlık üzerinde hangi finansal oranların belirleyici olduğu araştırılmıştır. Analiz sonucunda, aşırı yayılım belirlendiği için Williams yöntemi kullanılmış ve problem giderilerek uygun bir sınıflama analizi gerçekleştirilmiştir.

Lojistik Regresyon Modeli

Lojistik regresyonda, genelleştirilmiş doğrusal modeller kullanılarak en çok olabilirlik yöntemi ile parametre tahmini elde edilmektedir. Bu yöntem ile



binom dağılımının olasılık fonksiyonu üssel olarak tanımlanarak verilerin doğrudan beklenen değeri yerine, beklenen değer doğrusal kombinasyonu kullanılmaktadır (Akbulut ve Rençber, 2015). Genelleştirilmiş doğrusal modeller, bağımsız değişkenlerin doğrusal yapısını bağımlı değişkeninin beklenen değerine bağlayan bir bağlantı fonksiyonunu (link function) kullanmaktadır. Lojistik regresyonda, söz konusu bağlantı fonksiyonu logit dönüşüm ile verilmektedir (Stokes vd., 2000).

Lojistik regresyon modelinde, y iki kategorili bağımlı değişken $1 \times p$ boyutlu bağımsız değişken vektörü x olmak üzere x 'e göre y 'nin koşullu ortalamasını gösteren $\pi(x) = E(y/x)$ niceliği kullanılır. Lojistik regresyon modeli,

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p}} \quad (1)$$

biçiminde yazılır. Burada β_0 sabit, β_p regresyon katsayısı ve $\pi(x)$, bağımlı değişkenin x 'e bağlı olarak 1'e eşit olması koşullu olasılığıdır. Bu modelde, $-\infty$ ile $+\infty$ arasında değer alabilen bağımsız değişkenlere dayanarak iki kategorili bağımlı değişkenin değerlerini tahmin etmek güçtür. Bu durumu ortadan kaldırmak için en iyi çözüm, bağımlı değişkenin olasılık değerinin $-\infty$ ile $+\infty$ arasında tanımlı hale getirilmesidir. Bu amaçla, $\pi(x)$ 'e dayanarak eşitlik 1'e lojit dönüşüm uygulanmasıyla eşitlik (2) elde edilir.

$$g(x) = \ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad (2)$$

modeline ulaşılır. Lojit $g(x)$, x 'in aralığına bağlı kalarak $-\infty$ ile $+\infty$ arasında değer alabilen doğrusal bir fonksiyondur (Agresti, 1996).

Lojistik regresyonda, $\beta_0, \beta_1, \dots, \beta_p$ katsayılarını tahmin etmede en çok olabilirlik yöntemi yaygın olarak kullanılmaktadır. İki kategorili bağımlı değişkenin i .nci birimin değeri y_i olarak ele alındığında ve x_i i .nci birim için bağımsız değişken vektöründeki değerlerini göstermek üzere $(x_{1i}, x_{2i}, \dots, x_{pi}, y_i)$ n tane bağımsız gözlem çiftinden oluşan bir örneklem olarak tanımlanabilir ($i = 1, 2, \dots, n$). Eğer bağımlı değişken y , 0 ya da 1 olarak kodlanmışsa $\pi(x)$, verilen x için y 'nin 1'e eşit olmasının koşullu olasılığını $(P(y=1/x))$ verir. $1 - \pi(x)$ ise, verilen x için y 'nin 0'a eşit olmasının koşullu olasılığını $(P(y=0/x))$ verir. Böylece (x_i, y_i) çiftleri için, $y_i = 1$ 'in olabilirlik fonksiyonuna katkısı $\pi(x_i)$ ve $y_i = 0$ 'ın olabilirlik fonksiyonuna katkısı $1 - \pi(x_i)$ ifadesine eşittir. Burada $\pi(x_i)$, x_i 'de hesaplanan



$\pi(x)$ olasılığını tanımlar. (x_i, y_i) çiftleri için olasılık ya da olabilirlik fonksiyonuna katkı,

$$\zeta(x_i) = \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i} \quad (3)$$

eşitliği ile elde edilir. Gözlemlerin bağımsız olduğu varsayıldığı için olabilirlik fonksiyonu,

$$I(\beta) = \prod_{i=1}^n \zeta(x_i) = \prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i} \quad (4)$$

ile gösterilir ve en çok olabilirlik yöntemi bu eşitliği maksimum yapan β vektörünün tahmin edilmesini sağlar (Hosmer ve Lemeshow, 2000).

Lojistik regresyon modelinin varsayımları şöyle sıralanabilir:

Doğrusallık: Logit değeri ile sürekli yapıda bağımsız değişkenler arasında doğrusal bir ilişki olmalıdır. Bu sınamak için grafiksel yöntemlerden veya Box-Tidwell yönteminden yararlanılmaktadır. Box-Tidwell yönteminde, sürekli bağımlı değişkenler (X_i) için modele etkileşim terimleri ($X_i \ln X_i$) eklenmekte ve etkileşim terimi ile logit arasında istatistiksel olarak anlamlı bir ilişki varsa doğrusallığın olmadığına kanaat getirilmektedir (Alpar, 2011).

Hata Terimlerinin Bağımsızlığı: Modele ait veri setindeki gözlem değerleri birbirinden bağımsız olmalı yani, aynı istatistiki karar birimine ait birden fazla eşzamanlı gözlem bulunmamalıdır. Bu varsayımdan sapma aşırı yayılıma da sebep olmaktadır (Field, 2009).

Çoklu Doğrusal Bağlantı: Bağımsız değişkenler arasındaki yüksek dereceli (0,70 ve üzeri) korelasyonların sebep olacağı çoklu doğrusal bağlantı, parametre tahminlerinin güvenilirliğini azaltmaktadır. Çünkü, bu durumda parametre tahmincilerine ait standart hatalar büyüyeceğinden Wald testi hesap değeri büyüyecek ve böylece önemli bir değişken önemsiz gibi kabul edilebilecektir. Bu sebeple, varyans şişirme çarpanının (VIF) 5'ten büyük olmamasına dikkat edilmeli, aksi halde ilgili bağımsız değişken modelden çıkarılmalıdır (Aydın ve Arı, 2016).

Beklenen Frekans Sayısı: Modelde yer alan değişkenlere ait beklenen frekans değerleri 0 olmamalı, 1'den büyük olmalı ve 5'ten küçük olanlar da toplam gözlem sayısının %20'sini aşmamalıdır. Uyum iyiliği testlerinin başarısı buna bağlıdır (Wilson ve Lorenz, 2015).

Tam Ayrışım (Complete Separation): Bağımlı değişken, bağımsız değişken ya da değişken seti tarafından mükemmel bir şekilde tahmin ediliyorsa tam ayrışım durumu söz konusudur. Bu durumda, bağımsız değişken ya da



değişkenlerin belirli bir değer aralığı için bağımlı değişken 0, diğer aralık için ise 1 değerini almaktadır. Tam ayrışım durumunda, parametre standart hataları çok büyük değerler almakta ve böylece 1.tip hata olasılığı azalmaktadır. Bu durum daha çok özellikle gözlem sayısının yetersiz ve değişken sayısının fazla olduğu durumda ortaya çıkmaktadır (Aydın ve Arı, 2016).

Aşırı Yayılım (Overdispersion): Aşırı yayılım, modelde gözlenen varyansın beklenen varyanstan büyük olması durumudur ve parametre tahmincilerine ait standart hataların çok küçük çıkarak tahminciler için güven aralıklarının daralmasına ve 1.tip hata olasılığının artmasına sebep olmaktadır. Bu durumun tespit edilebilmesi için aşırı yayılım parametresinin (ki-kare veya sapma uyum iyiliği istatistiği / serbestlik derecesi) 1'den büyük ama özellikle de 2'den büyük olmamasına dikkat edilmelidir (Slaton vd., 2000).

Lojistik Regresyonda Aşırı Yayılım

Aşırı yayılım, kesikli verilerin analizinde önemli bir kavramdır. Lojistik regresyonda bağımlı değişkenin gözlenen varyansının beklenen varyanstan büyük çıkması aşırı yayılım olarak tanımlanır (Cameron ve Trivedi, 1998). Bu sorun, standart hata tahminlerini ve modelin uyum iyiliği ölçütlerini etkileyeceği için, aşırı yayılımı dikkate alarak çözümlene yapan farklı yöntemlere ihtiyaç duyulur. Bu yöntemlerin en bilineni ve yaygın kullanılanı Williams analizidir. Lojistik regresyonda aşırı yayılımın farklı nedenleri arasında en sık rastlanılan, başarı olasılıkları arasındaki değişim ve ikili cevap değişkenleri arasındaki korelasyondur. Y bağımlı değişkeninin beklenen değeri ve varyansı,

$$E(Y_i) = n_i \pi_i \text{ ve } \text{Var}(Y_i) = n_i \pi_i (1 - \pi_i) \quad (5)$$

biçiminde verilebilir. Sözelimi ikili cevap değişkenleri arasında korelasyon var ise, $Y_i = \sum_{j=1}^{n_i} R_{ij}$ için, $R_{ij}=1$ başarı ve $R_{ij}=0$ diğer durumlarını göstermek üzere, $P(R_{ij} = 1) = \pi_i$ olacaktır. Bu durumda $k \neq j$ olmak üzere;

$$\text{Var}(Y_i) = \sum_{j=1}^{n_i} \text{Var}(R_{ij}) + \sum_{j=1}^{n_i} \sum_{k=j}^{n_i} \text{cov}(R_{ij}, R_{ik}) \quad (6)$$

$$\text{Var}(Y_i) = n_i \pi_i (1 - \pi_i) + \sum_{j=1}^{n_i} \sum_{k=j}^{n_i} \text{cov}(R_{ij}, R_{ik}) \quad (7)$$

$$\text{Var}(Y_i) = n_i \pi_i (1 - \pi_i) [1 + (n_i - 1)\phi] \quad (8)$$

elde edilir (Yeşilova ve Kasap, 2008).



Böylece, beklenen varyans ile eşitlik (8)'i kullanarak elde edilen gözlenen varyans birbirine eşit değildir. Ayrıca, $\phi = 0$ olduğunda gözlenen varyans beklenen varyansa eşit olacağından aşırı yayılım olmayacaktır. Ters durum $\phi > 0$ olduğunda ise, veri kümesinde aşırı yayılım söz konusudur. Uygulamalarda genellikle aşırı yayılım durumuyla karşılaşılır (Cox, 1983).

Sapma ve Pearson ki-kare istatistikleri lojistik regresyonda yaygın olarak kullanılan uyum istatistikleridir. Uyum istatistiklerinin yayılım parametre değerleri, hesap değerlerinin kendi serbestlik derecesine bölünmesi ile elde edilmektedir. Sapma ve Pearson ki-kare uyum istatistiği değerlerinin bire eşit ya da çok yakın çıkması aşırı yayılımın olmadığı, birden büyük çıkmasının ise aşırı yayılımın olduğu anlamına gelir. Aşağıda χ_p^2 olarak verilen değer Pearson ki-kare ve χ_D^2 olarak belirtilen değer ise sapma ölçütleridir (Allison,1999).

$$\chi_p^2 = \sum_{i=1}^m \sum_{j=1}^{k+1} \frac{(r_{ij} - n_i \hat{p}_{ij})^2}{n_i \hat{p}_{ij}} \quad \chi_D^2 = 2 \sum_{i=1}^m \sum_{j=1}^{k+1} r_{ij} \log \left(\frac{r_{ij}}{n_i \hat{p}_{ij}} \right) \quad (9)$$

Sapma uyum istatistiklerinin serbestlik derecelerine bölünmesi ile,

$$\hat{\sigma}^2 = \left\{ \begin{array}{l} \chi_p^2 / (mk - q) \\ \chi_D^2 / (mk - q) \\ (sabit)^2 \end{array} \right\} \quad (10)$$

ifadesi yazılabilir.

Burada, m alt yığın (subpopulation) sayısı, k + 1 kategori sayısı (cevap düzeyi), r_{ij} ise, i.nci gözlemin j.inci kategorisine bağlı toplam ağırlıktır. Eşitlik 9'da yer alan \hat{p}_{ij} , i.nci gözlem için j.nci cevap düzeyinin tahmin edilen olasılık değeridir. Aşırı yayılımın belirlenmesi için kullanılan bu ki-kare istatistikleri mk - q serbestlik derecedir. Burada q değeri tahmin edilen parametre sayısını göstermektedir. Bu tahminlerin tutarlı olabilmesi için, alt yığınların yeterli tekrarı sağlayabilecek sayıda olması gerekir. Verilerin bu yeterliliği sağlayamadığı durumda aşırı yayılım istatistikleri güvenli sonuçlar üretmez (Saefuddin vd., 2011).

Williams Yöntemi

Veri kümesinin büyüklüğü n olmak üzere, i.nci gözlem için r_i / n_i gözlenen oran değeri olsun. Ayrıca, i.nci gözlem için cevap olasılık değeri P_i rasgele değişkeni olsun. P_i 'nin ortalama ve varyansı,



$$E(P_i) = p_i \quad V(P_i) = \Phi p_i (1 - p_i) \quad (11)$$

biçiminde verilebilir. Burada Φ parametresi negatif olmayan ve bilinmeyen skala parametresi (scale parameter) olarak tanımlanmıştır. r_i 'nin ortalama değeri ve varyansı ise,

$$E(r_i) = n_i p_i \quad V(r_i) = n_i p_i (1 - p_i) [1 + (n_i - 1)\Phi] \quad (12)$$

olarak ifade edilebilir (Hilbe, 2009). Williams yönteminde bilinmeyen Φ parametresinin tahmini, Pearson ki-kare istatistiğinin yaklaşık beklenen değerine eşitlenmesiyle elde edilmektedir. Burada, ω_i^* 'nin i .nci gözleme ilişkin ağırlık değeri olduğu varsayıldığında, Pearson ki-kare istatistiği,

$$\chi^2 = \sum_{i=1}^n \frac{\omega_i^* (r_i - n_i \hat{p}_{ij})^2}{n_i \hat{p}_{ij} (1 - \hat{p}_i)} \quad (13)$$

biçiminde verilebilir. $g'(\cdot)$ değeri, $g(\cdot)$ bağlantı fonksiyonunun birinci türevi ise, χ^2 'nin yaklaşık beklenen değeri,

$$E_{\chi^2} = \sum_{i=1}^n \omega_i^* (1 - \omega_i^* v_i d_i) [1 + \Phi (n_i - 1)] \quad (14)$$

olacaktır. Eşitlik (14)'de $v_i = n_i / (p_i (1 - p_i) [g'(p_i)]^2)$ ve $d_i = \hat{\alpha}_i + \chi_i' \hat{\beta}$ doğrusal tahmincisinin varyansını göstermektedir. Bilinmeyen Φ parametresi iteratif işlem kullanılarak tahmin edilmektedir (Williams, 2006).

İlk aşamada $\omega_i^* = 1$ ve p_i yaklaşık olarak r_i / n_i değerini alarak başlanır. Φ 'nın başlangıç tahmini,

$$\hat{\phi}_0 = \frac{\chi^2 - (n - m)}{\sum (n_i - 1)(1 - v_i d_i)} \quad (15)$$

olarak ifade edilir.

Burada m toplam parametre sayısını göstermektedir. Ağırlıkların başlangıç tahminleri,

$$\hat{\omega}_{i0}^* = [1 + (n_i - 1)\hat{\phi}_0]^{-1} \quad (16)$$

biçiminde olur. Modelin ağırlıklandırılmış uyumu yapıldıktan sonra $\hat{\beta}$ ve χ^2 yeniden hesaplanır. Daha sonraki aşamada Φ 'nin tahmini,



$$\hat{\Phi}_1 = \frac{\chi^2 - \sum \omega_{i0}^* (1 - \omega_{i0}^* v_i d_i)}{\omega_{i0}^* (n-1) (1 - \omega_{i0}^* v_i d_i)} \quad (17)$$

elde edilir. Bu iteratif işlemler χ^2 uyum istatistiğinin kendi serbestlik derecesine yaklaşıncaya kadar devam eder (SAS, 2007).

Uygulama

Çalışmada, BİST 100 endeksinde işlem gören firmalar için, Eylül 2016 dönemine ilişkin mali başarısızlığın tahminine yönelik bir sınıflama analizi uygulanmıştır. Aynı bilanço yapısına sahip 80 firma için analize dahil edilerek, bankalar, sigorta şirketleri, faktoring kuruluşları kapsam dışı bırakılmıştır. İlk aşamada, faktör analizi yardımıyla söz konusu firmalar için başarılı-başarısız ayrımı gerçekleştirilmiştir. Böylece, bağımlı değişken mali başarılı firmalar için "0" ve mali başarısızlar için "1" değeri atanarak oluşturulmuştur. Bağımsız değişkenler, firmaların mali yapısını yansıtacak likidite oranları, mali yapı oranları, faaliyet oranları, karlılık oranları başlıklarını içeren oran değerleridir. Çalışmada kullanılan finansal oranlar Altman (2005) ile Suarez ve Susman (2004) çalışmaları doğrultusunda seçilerek Tablo 1'de verilmiştir.

Tablo 1: Açıklayıcı Faktör Analizinde Kullanılan Finansal Oranlar

Likidite Oranı	Net İşletme Sermayesi Devir Hızı
Cari Oran	Dönen Varlıklar Devir Hızı
Nakit Oran	Aktif Toplamı Devir Hızı
Özsermaye Çarpanı	Net Kar Marjı
Kısa Vadeli Borçlar / Toplam Aktifler	Esas Faaliyet Kar Marjı
Kısa Vadeli Borçlar / Toplam Borçlar	Faaliyet Kar Marjı
Uzun Vadeli Borçlar / Devamlı Sermaye Oranı	Özsermaye Kar Marjı
Duran Varlıklar / Özsermaye Oranı	Aktif Kar Marjı
Duran Varlıklar / Devamlı Sermaye Oranı	Ebitta / Net Satışlar

Her bir finansal oran kullanılarak açıklayıcı faktör analizinin uygulanabilirliğine yönelik ön testler SPSS 22.0 sürümü kullanılarak gerçekleştirilmiştir. Buna göre, veri setinin Kaiser-Meyer-Olkin (KMO)



örneklem yeterliği iyi düzey olan 0.70 değerinin üzerinde 0.916 bulunmuştur. Analize yer alan değişkenlerin tutarlılığını ölçen Bartlett küresellik testi istatistiksel olarak anlamlı ($\chi^2= 897.453$ ve $p= .001$) bulunmuş, korelasyon matrisinin birim matrise eşit olduğunu ileri süren sıfır hipotezi reddedilmiştir. Böylece, değişkenler arasında korelasyon bulunduğu, bağımsızlaştırma ve boyut indirgemenin yani; faktör analizinin uygunluğu belirlenmiştir. Testler sonucunda açıklayıcı faktör analizi için kullanılacak örneklemin yeterli olduğu ve ölçekte yer alan ifadelerin iç tutarlılığının sağlandığı görülmüştür.

Veri setinin uygunluğunun yapılan testlerle onaylanmasının ardından faktör yapısının ortaya konulması amacıyla faktör tutma yöntemi olarak Varimax döndürme metodu ile temel bileşenler analizi yöntemi uygulanmıştır. Faktör yapısında, 4 faktör öz değerlerinin 1'den yüksek olacak şekilde elde edilmiştir. Böylece, toplam varyansın %82.278'ini açıklayan 4 faktörlük bir yapı belirlenmiştir. Veri kümesini oluşturan firmaların faktör analizi yardımıyla mali başarılı-başarısız ayırımı, genel faktörün negatif olan ve olmayan değerlerine göre belirlenerek, bağımlı değişken ikili kategorik düzeyde oluşturulmuştur. Genel faktör sütununda negatif değer alan firmalar mali başarısız olarak "0" kod ve pozitif olanlar başarılı olarak "1" kod almıştır.

Faktörleşme sonucunda, kavramsal anlamlılık da sağlandığı için, bağımsız değişkenler olarak bu faktör serilerinin kullanılması uygun görülmüştür. Lojistik regresyon analizi, çoklu bağlantı sorununa hassas bir yöntem olduğundan, indirgenmiş ve bağımsızlığı sağlanmış bağımsız değişken tercihi sonuçların güvenilirliği açısından önem taşımaktadır.

Adımsal seçim metodlarından, olabilirlik oran test istatistiğine dayalı ileriye doğru eleme yöntemi (Forward LR) kullanılmış ve lojistik modele ilişkin elde edilen sonuçlar Tablo 3'te verilmiştir. Burada, Faktör_1: likidite, Faktör_2: mali yapı, Faktör_3: faaliyet oranları ve Faktör_4: karlılık oranlarını belirtmektedir.

Tablo 2: Lojistik Regresyon Parametre Tahminleri

Bağımsız değişkenler	$\hat{\beta}$	Se($\hat{\beta}$)	Wald	p	Exp($\hat{\beta}$)
Faktör_1	1.277	.264	23.456	.000*	3.586
Faktör_2	.970	.271	12.840	.000*	2.639
Faktör_3	.907	.202	20.067	.000*	2.476



Faktör_4	.137	.237	.333	.564	.872
----------	------	------	------	------	------

*0.05 düzeyinde istatistik anlamlı

Yukarıda verilen modelin uyumunun belirlenmesinde, Hosmer-Lemeshow testi ($\chi^2 = 3.64$ ve $p = 0.35$) kullanılarak mali başarılı-başarısız ayırımında model yeterli bulunmuştur. Ayrıca, Cox and Snell R^2 değeri %78.3 ve Nagelkerke R^2 değeri % 81.7 olarak yüksek bir değerde elde edilmiştir. Model sonuçlarında, likidite, mali yapı ve karlılık oranı faktörleri %5 düzeyinde anlamlı ve önemli çıkmıştır. Buna karşılık, faaliyet oranı değişkeni istatistik anlamlı ve önemli bulunamamıştır. Mali başarılı olup olmama üzerinde risk düzeyini gösteren $\text{Exp}(\hat{\beta})$ değerleri büyüklük sırasıyla likidite, mali yapı, karlılık oranları olarak elde edilmiştir. Bu model için doğru sınıflama yüzdesi %74.4 olarak elde edilmiştir.

Elde edilen sonuç değerlerine göre model olumlu görünmektedir. Lojistik regresyonda aşırı yayılım uyum istatistiklerine bakılmadıkça, göze çarpan bir problem olmadığı için, araştırmacı herşeyin yolunda gittiğini düşünebilir. Bu nedenle kontrol mekanizmasının önemi büyüktür.

Tablo 3: Lojistik Regresyon Uyum İstatistik Değerleri

Uyum İstatistikleri	Serbestlik Derecesi	Hesap Değeri	Sonuç Değeri
Sapma	79	497.032	6.235
Pearson Ki-kare	79	426.775	5.168

Tablo 3'den görüleceği üzere, uyum istatistik değerleri 1'in üzerinde çıktığı için aşırı yayılım söz konusudur. Aşırı yayılımı modellemek için Williams yöntemi uygulanmıştır. Bu çözümlene sonucunda elde edilen parametre katsayıları Tablo 5'de verilmiştir.

Tablo 4: Williams Yöntemi İle Elde Edilen Lojistik Regresyon Parametre Tahminleri

Bağımsız değişkenler	$\hat{\beta}$	$\text{Se}(\hat{\beta})$	Wald	p	$\text{Exp}(\hat{\beta})$
Faktör_1	1.261	.266	22.449	.000*	3.530
Faktör_2	.940	.279	11.379	.001*	2.560



Faktör_3	.555	.216	6.571	.010*	1.741
Faktör_4	.880	.264	11.140	.002*	2.411

*0.05 düzeyinde istatistik anlamlı

Hosmer-Lemeshow testi ($\chi^2 = 2.98$ ve $p = 0.12$) kullanılarak mali başarılı-başarısız ayırımında model yeterli bulunmuştur. Cox and Snell R^2 değeri %88.24 ve Nagelkerke R^2 değeri %91.05 olarak ilk modele göre artış göstermiştir. Bu yöntemde elde edilen parametre tahmin sonuçlarında ele alınan bağımsız değişkenlerin tümü istatistik anlamlı ve önemli çıkmıştır. Mali başarılı olup olmama üzerinde risk düzeyini gösteren $\text{Exp}(\hat{\beta})$ değerleri büyüklük sırasıyla likidite, mali yapı, karlılık ve faaliyet oranları olarak elde edilmiştir. Ayrıca, doğru sınıflama yüzdesi %89.1 olarak ilk modele göre daha yüksek bir değerdir. Williams çözümlü model hem uyum iyiliği göstergeleri açısından hem de sınıflama yüzdesi olarak daha başarılı sonuçlar vermiştir. Bu modele ilişkin kontrol amaçlı aşırı yayılım sonuçları verilmiştir.

Tablo 5: Williams Yöntemi İle Elde Edilen Lojistik Regresyon Uyum İstatistik Değerleri

Uyum İstatistikleri	Serbestlik Derecesi	Hesap Değeri	Sonuç Değeri
Sapma	79	68.361	0.872
Pearson Ki-kare	79	59.527	0.704

Uyum istatistiği değerleri 1'in altında çıktığı için aşırı yayılımın modellenmesinde Williams yönteminin başarılı sonuçlar verebildiği anlaşılmıştır. Günümüzde ortaya çıkan rekabet sonucunda birçok firma mali sıkıntılarla karşı karşıya kalmaktadır. İleride karşılaşılabilecek sorunları önceden görerek önlem almak, krizlerden ve mali sıkıntılardan en az zararlı çıkmak açısından önemli bir stratejidir. Dolayısıyla, mali başarısızlığın doğru bir yöntemle güvenilir tahmin edilmesi oldukça önem kazanmaktadır. Lojistik regresyon, sınıflama açısından etkin bir yöntem olmakla birlikte, olabilecek bazı problemler bilinip giderildiğinde daha güvenilir sonuçlara ulaşılabilecektir. Mali başarısızlığın belirlenmesi, özellikle kriz dönemlerinde daha önem taşıdığı için hata yapılmaması gereken konulardan biridir.



Sonuç

Son yıllarda oldukça geniş kullanıma sahip lojistik regresyon analizi, özellikle gözlemlerin gruplara atanmasında tercih edilen yöntemlerden biridir. Hesaplamalarda sağladıkları kolaylık ve analiz sonuçlarının daha rahat yorumlanabilmesi nedeniyle kullanım alanları gün geçtikçe yaygınlaşan ikili veriler, bir çok bilim dalında kullanılmaktadır. Özellikle sağlık, finans ve sosyal bilimlerde kullanılan veriler bu yapıda olabilmektedir. Lojistik regresyon analizi; sürekli, kesikli, ya da bunların herhangi bir karışımı olan veri kümelerinden kategorik bir sonucu tahmin etmeye olanak sağlamaktadır.

Kategorik veri kümelerine uygulanan Poisson ve lojistik gibi regresyon yöntemlerinin uygulanabilir olup olmadığını gösteren uyum istatistikleri (sapma ve Pearson ki-kare) büyük önem taşımaktadır. Uygulanan regresyon yönteminin doğruluğu bakımından, her iki uyum istatistiği değerlerinin 1'e eşit ya da çok yakın olması istenmektedir. Poisson ve lojistik gibi regresyon analizlerinde, söz konusu uyum istatistikleri ortalama ile varyans arasındaki ilişkiye bağlı olarak değişmektedir.

Lojistik regresyon aşırı yayılım, gözlenen varyansın beklenen varyanstan büyük çıkması durumunda gerçekleşmektedir. Aşırı yayılıma birçok neden gösterilebilir. Bu nedenlerden en önemli olanları, başarı olasılıkları arasındaki değişim, ikili cevaplar arasındaki korelasyon, yanlış bağlantı fonksiyonun kullanılması, modelde olması gereken önemli terimlerin olmaması ve örnek büyüklüğünün yetersiz olması şeklinde verilebilir. Aşırı yayılım sonucunda model uyum istatistikleri ve tahmin edilen parametreler güvenilir sonuçlar üretmeyeceği için, aşırı yayılımı dikkate alan farklı yöntemlere ihtiyaç duyulmaktadır. Bunlar arasından yaygın kullanılanlardan biri de Williams yöntemi olmaktadır.

Bu çalışmada, BİST 100 endeksinde işlem gören ve bilanço yapıları aynı olan 80 sanayi firmasına yönelik mali başarısızlığın belirlenmesi hedeflenmiştir. Faktör analizi yardımıyla ele alınan finansal oranlar için firmalar başarılı-başarısız biçiminde sınıflanarak bağımlı değişken oluşturulmuştur. Bağımsız değişken kümesi faktör analizi ile belirlenen likidite, mali yapı, faaliyet oranları ve karlılık oranları olarak dört ana başlıkta ele alınmıştır ve veri kümesine lojistik regresyon uygulanmıştır. Daha sonra, elde edilen uyum istatistik değerleri 1'den çok büyük çıktığı için veri kümesinde aşırı yayılım olduğu saptanmıştır. Bu durumda, lojistik regresyon analizi sonuçlarının kullanılmasında sapmalı parametre tahminlerine neden olacağı için aşırı yayılımı modele dahil eden yöntemlerden, Williams analizi kullanılarak aşırı yayılım kontrol altına alınmıştır.

Williams çözümlemesi sonrasında, hem sapma hem de Pearson ki-kare uyum istatistiklerinin aşırı yayılım değerleri 1'e oldukça yakın bulunarak



problem giderilmiştir. Ayrıca, modelin uyum iyiliği ölçütlerinin değeri ve doğru sınıflama yüzdesi değeri artış göstererek, uygun çözümleme sonrasında modelin daha ideal sonuçlar ürettiği görülmüştür. Böylece, aşırı yayılım problemi için Williams yönteminin etkin çözümleme alternatifi olduğu ortaya konulmuştur. BİST 100 endeksinde işlem gören firmaların, Eylül 2016 verileri için mali başarısız duruma düşme olasılığını arttırıcı değişkenlerin, likidite ve mali yapı oranları altındaki finansal değerlerden kaynaklı risk faktörü olabileceği belirlenmiştir.

Kaynakça

- Agresti, A. (1996). *Categorical Data Analysis*, John Wiley & Sons Inc., New York.
- Akbulut, R. ve Rençber, Ö. F. (2015). Veri Zarflama ve Lojistik Regresyon Analizi ile Çimento İşletmelerinde Finansal Performansa Dayalı Etkinliklerin Değerlendirilmesi, *Uluslararası Alanya İşletme Fakültesi Dergisi*, 7(3), 123-135.
- Allison, P. D. (1999). *Logistic Regression Using the SAS System: Theory and Application*, SAS Institute Inc., USA.
- Alpar, R. (2011). *Çok Değişkenli İstatistiksel Yöntemler*, Detay Yayıncılık, 3. Baskı, Ankara.
- Altman, E. (2005). *Corporate Financial Distress and Bankruptcy*, Third Edition, John and Wiley & Sons Inc., New York.
- Aydın, N. ve Arı, E. (2016). Hanehalkı Otomobil Talebini Belirleyen Etkenlerin İkili Lojistik Regresyon Yöntemiyle Analizi: Türkiye Örneği, *Kastomonu Üniversitesi, İ.İ.B.F. Dergisi*, 12 (1), 1-22.
- Cameron, A. C. ve Trivedi, P. K. (1998). *Regression Analysis of Count Data*, Cambridge University Press, United Kingdom.
- Cox, R. (1983). Some Remarks on Overdispersion, *Biometrika*, 70 (1), 269-274.
- Czado, C. (2003). *Overdispersion in Logistic Regression*, Erişim:[<http://www-m4.ma.tum.de/courses/GLM/lec5.pdf>], Erişim Tarihi: 28.09.2016.
- Field, A. (2009). *Discovering Statistics Using SPSS*, Sage Publications, 3 th Edition, New York.
- Hilbe, J. M. (2009). *Logistic Regression Models*, Chapman & Hall/ CRC Press, New York.



- Hosmer, D. W. ve Lemeshow, S. (2000). *Applied Logistic Regression*, Second Edition, John Wiley & Sons Inc., New York.
- İyit, N. ve Genç A. (2005). Lojistik Regresyon Analizi Yardımıyla Denekte Menopoz Evresine Geçişe İlişkin Bir Sınıflandırma Modelinin Elde Edilmesi, S.Ü. Fen Edebiyat Fak. Fen Dergisi, 25 (1), 19-27.
- Saefuddin, A. Setiabudi N. A. ve Achsani, N. A. (2011). The Effect of Overdispersion on Regression Based Decision with Application to Churn Analysis on Indonesian Mobile Phone Industry, *European Journal of Scientific Research*, 60(4), pp. 602-610
- SAS (2007). *SAS / STAT Software: Changes and Enhancements to SAS / SDTAT Software in Versions 7 and 8*, SAS Institute Inc., North Carolina.
- Slaton, T. L. Piegorsch, W. W., ve Durham, S. D. (2000). Estimation and Testing with Overdispersed Proportions Using the Beta- Logistic Regression Model of Heckman and Willis, *Biometrics*, 56(1), 125-133.
- Suarez, J. ve Susman O. (2004). Financial Distress, Bankruptcy Law and The Business Cycle, Erişim:[[http://www.finance.ox.ac.uk/file links/finecon_papers/2004fe07.pdf](http://www.finance.ox.ac.uk/file_links/finecon_papers/2004fe07.pdf)], Erişim Tarihi: 19.08.2016.
- Stokes, M. E. Davis, C. S. ve Koch G. G. (2000). *Categorical Data Analysis Using The SAS System*, John Wiley & Sons Inc., New York.
- Williams, R. (2006). Generalized Ordered Logit / Partial Proportional Odds Models for Ordinal Dependent Variables, *The Stata Journal*, 6(1), 58-82.
- Wilson J. R. ve Lorenz K. A. (2015). *Modeling Binary Correlated Responses Using SAS, SPSS and R, ICOSA Book Series İn Statistics 9*, Springer Publishing., Switzerland.
- Yeşilova, A. ve Kasap, İ. (2008). Lojistik Regresyonda Meydana Gelen Aşırı Yayılımın İncelenmesi, Yüzüncü Yıl Üniversitesi, Ziraat Fakültesi, *Tarım Bilimleri Dergisi*, 18 (1), 21-25.

