

Lojistik Regresyon Yöntemi ile Farklı Başarı Düzeyindeki Ülkelerin PISA Başarı Düzeylerini Etkileyen Önemli Değişkenlerin İncelenmesi

Investigation of Important Variables Affecting PISA Success Levels of Countries with Different Success Levels with Logistic Regression Method

Yusuf Kasap¹  Nuri Doğan²  Mustafa Köroğlu^{3*} 

¹ Dr. Hacettepe Üniversitesi, Eğitim Fakültesi, Ankara, Türkiye

² Prof. Dr., Hacettepe Üniversitesi, Eğitim Fakültesi, Ankara, Türkiye

³ Dr. Öğr. Üyesi, Erzincan Binali Yıldırım Üniversitesi, Eğitim Fakültesi, Erzincan, Türkiye

Makale Bilgileri

Geliş Tarihi (Received Date)

30.07.2022

Kabul Tarihi (Accepted Date)

25.10.2022

*Sorumlu Yazar

Tel.: +5535362142

Erzincan Binali

Yıldırım Üniversitesi,

Eğitim Fakültesi

mustafa.koroglu@erzincan.edu.tr

Öz: Bu çalışmada 2018 yılında PISA'ya katılan öğrencilere sınavla verilen öğrenci anketinden elde edilen 34 bağımsız değişken ile farklı başarı düzeyine sahip ülkelerin PISA okuduğunu anlama başarısını yordayabilen değişkenlerin belirlenmesi amaçlanmıştır. Bu amaç için PISA'ya katılan 79 ülke, başarı yüzdelik sırasına konuldu. Ardından eşit dilimlerle ülkeler alt, orta ve üst dilimler olarak ayrıldı. Yüzdelik dilimlerine alt, orta, üst başarı grubu ülkelerin verileri üzerinden oluşturulan üç ayrı lojistik model ile yapılan analizlerde okuduğunu anlama başarısını yordamada anlamlı değişkenler, 25-28 arasında değişmiştir. Daha sonra oluşturulan modellerin sınıflama performansında dikkate değer düşme olmayacak biçimde ayarlama yapılarak değişken sayısı azaltılıp her bir model için başarıyı en iyi yordayan 10 değişken seçilmiştir. Değişken sayısı azaltılarak oluşturulan modellerle yapılan analiz sonucunda başarıyı yordamada alt, orta, üst başarı grubu ülkeler için, önemli ortak yordayıcıların; evdeki eğitimsel eşyalar, okuma keyfi, sosyo-ekonomik durum indeksi, PISA testinin zorluk algısı, ailenin mal varlığı, yaşamın anlamı ve öğretmenin eğitimi yönlendirmesi değişkenleri olarak belirlenmiştir. Daha sonra, her bir grup için oluşturulan modellerin her birindeki 10 bağımsız değişken ile modellerin PISA okuduğunu anlama başarısını sınıflama ve kestirim performansları hesaplanmıştır.

Anahtar Kelimeler: Sınıflama, tahmin, PISA

Abstract: The objective of this study was to identify the variables that could predict the PISA reading comprehension success of countries with different success levels. This was done with 34 independent variables obtained from the student questionnaire given to the students who participated in PISA in 2018. For this purpose, 79 countries participating in PISA were put in order of success percentage. Then, with equal slices, countries were divided into lower, middle, and upper slices. Significant variables in predicting reading comprehension success in the analysis made with three different logistic models based on the data of low, middle and high achievement group countries in percentiles varied between 25-28. Afterwards, the number of variables was reduced by adjusting the classification performance of the models created in such a way that there would not be a notable decrease, and the 10 variables as the best-predicted success were selected for each model. As a result of the analysis made with the models created by reducing the number of variables, it was revealed that the important common predictors for the low, middle and high success group countries such as educational items at home, reading pleasure, socio-economic status index, PISA test's perception of difficulty, family wealth, meaning of life and teacher's orientation to education. Then, with 10 independent variables in each of the models created for each group, the models' performance in classifying and predicting PISA Reading Comprehension success was calculated.

Keywords: Classification, estimation, PISA

Kasap, Y., Doğan, N. ve Köroğlu, M. (2022). Lojistik regresyon yöntemi ile farklı başarı düzeyindeki ülkelerin PISA başarı düzeylerini etkileyen önemli değişkenlerin incelenmesi. *Erzincan Üniversitesi Eğitim Fakültesi Dergisi*, 24(3), 590-599. <https://doi.org/10.17556/erziefd.1151452>

Giriş

Teknolojinin gelişimiyle birlikte çok sayıda bireyden çok sayıda veri toplanmakta ve bu verileri analiz etmek için de istatistiksel teknikler geliştirilmektedir. Özellikle de bilgisayarların kullanılması ile bu teknikler daha etkili kullanılabilmektedir. Araştırmalardan elde edilen verilerin analizi sürecinde bu tekniklerin kullanılması ile araştırma amacına yönelik hangi değişkenlerin önemli, hangilerinin önemsiz olduğu belirlenmektedir. Uygun teknikleri kullanarak belirlenen önemli değişkenler sayesinde bireylerin başarı, alışveriş ve tüketim davranışları ile ilgili isabetli tahminler yapılmaktadır. Bu tahminlere dayanarak verilen kararlar doğrultusunda geleceğin planlanmasının önemi her geçen gün artmaktadır.

Eğitimde de son yıllarda hem uluslararası düzeyde hem de ulusal düzeyde yapılan çalışmalarda çok değişkenli veri toplama eğilimi yaygınlaşmıştır. Bu sınavlardan birisi de Uluslararası Öğrenci Değerlendirme Programıdır (PISA). PISA verileri kullanılarak, çok değişkenli istatistiksel modeller ile başarıyı yordayan veya başarı ile ilişkili bağımsız değişkenlerin hangileri olduğu saptanabilmektedir. PISA her 3 yılda bir yapılan, Matematik okuryazarlığı, Fen okur yazarlığı ve okuduğunu anlama testlerini içeren bir uygulamadır. Her bir PISA uygulamasında, bu üç testten birine ağırlık verilmektedir. PISA 2018'de ağırlık verilen başlıca alan okuduğunu anlamadır. Bu uygulamada temel amaçlar; metinleri anlama, kullanma, değerlendirme, metinler üzerinde düşünme ve metinlerle etkileşim kurma becerilerini ölçmek olarak belirlenmiştir (OECD, 2019). PISA 2018 uygulaması,

79 ülkenin çoğunluğunda bilgisayar üzerinden dijital olarak gerçekleştirilmiştir.

Alanyazında başarıyı yordayan faktörlerin belirlenmesine yönelik çalışmalar, uygulamalı istatistikteki, t testi, Mann-Whitney U Testi, Anova, Kruskal-Wallis H Testi, Faktör Analizi ve Çoklu Regresyon Analizi, gibi teknikleri kullanan başarıyı yordayan değişkenleri belirlemeye yönelik çalışmalardır (Akkurt ve Köse 2019; Berberoğlu ve diğerleri, 2017; Coşkuner, 2013; Erdas ve diğerleri, 2017; Güzle Kayır, 2012; Önder ve Gelbal, 2016; Özkan ve Güvendir, 2014; Uyulgan ve Akkuzu, 2017; Urfalı Dadandı ve diğerleri, 2018; Yıldırım, 2012; Yu ve diğerleri, 2012). İncelenen çalışmalarda PISA başarıları ile ilgili ve Lojistik Regresyon yöntemine dayanan Yu ve arkadaşlarının çalışmasıdır. Yu ve arkadaşları (2012) yaptıkları çalışmada, Kanada ve Amerika'daki öğrencilerin PISA Fen okur yazarlığı başarılarını yordamada 13 adet bağımsız değişken kullanarak Lojistik regresyon yöntemi ile elde edilen modeldeki ODDS oranlarına göre başarıyı yordayan önemli faktörleri belirleyip karşılaştırmışlar. Erdas ve arkadaşlarının yaptığı çalışmada, Türkiye'deki öğrencilerin PISA 2003, 2006 ve 2009 (Uluslararası Öğrenci Değerlendirme Programı) fen okuryazarlığı başarıları ile ilişkili olan değişkenleri araştırıldığı çalışmada anne-baba eğitim düzeyi değişkeni ile öğrencilerin fen okuryazarlığı başarıları arasında pozitif bir ilişki olduğunu tespit ettikleri gözlemlenmiştir. Ayrıca, öğrenci karakteristiklerinden olan Fen'e yönelik tutum ile öğrencilerin başarıları arasında pozitif ilişki olduğu ve kız öğrencilerin fen okuryazarlığı alanında erkeklerden daha başarılı oldukları rapor etmişlerdir (Erdas Kartal, Doğan ve Yıldırım, 2017).

Yapılan araştırmalardan anlaşıldığı gibi az sayıda değişken ile daha çok birkaç ülke ile sınırlı olan homojen gruplarla yapılan çalışmalar yoğunluktadır. Oysaki, farklı başarı gruplarında Lojistik Regresyon yönteminden elde edilen sonuçların değişip değişmediğini, yani yordayıcı değişkenlerin ve sayısının değişip değişmediğine yönelik çalışmaya rastlanmamıştır. Lojistik Regresyon yöntemi ile veri (örneklem) değiştikçe farklı sonuçlar verebileceğinin araştırılması önem teşkil etmektedir. Dolayısıyla bu araştırmada, PISA 2018 okuduğunu anlama başarılarına yönelik olarak her biri 3'er ülkeden oluşan alt, orta ve üst başarı grubuna sahip ülkeler dikkate alınmıştır.

Birçok ülkenin katıldığı PISA'da ülkelerin başarı düzeyleri birbirinden oldukça farklıdır. Bu nedenle; bu çalışmada, alt, orta ve üst başarı grubundaki ülkeler için başarıyı etkileyen değişkenlerin belirlenmesi ile ilgili gruptaki ülkelerin eğitim politikalarına yön verirken başarıyı olumlu etkileyen faktörlere daha fazla önem vermeleri sağlanabilir. Diğer yandan farklı başarı düzeylerinde ülkeler için başarıyı etkileyen aynı ve farklı faktörlerin neler olduğu saptanarak, düşük başarı düzeyindeki ülkelerin başarılarını artırmada daha doğru kararlar alınmaları sağlanabilir. Bu araştırmanın bu bakımdan alanyazına katkı sağlayacağı düşünülmektedir.

Bu araştırmanın temel amacı, PISA öğrenci anketinde bulunan 34 bağımsız değişkene ilişkin puanları kullanarak farklı başarı düzeyine sahip (alt, orta ve üst başarı grubundaki) ülkelerin iki sınıflı (başarısız-başarılı) PISA okuduğunu anlama başarılarını etkileyen önemli yordayıcıları belirlemektir. Çalışmanın diğer amacı ise belirlenen az sayıda daha önemli bağımsız değişkenlerin kullanılarak performansı yüksek Lojistik Regresyon sınıflama modellerinin elde edilip

edilmeyeceğini göstermektir. Böylece çalışmanın, Lojistik Regresyon sınıflama model performansını en yüksek yapan daha önemli bağımsız değişkenleri seçerek değişken sayısını azaltmaya yönelik az sayıda çalışmadan biri olması hedeflenmektedir.

Problem Cümlesi

Lojistik Regresyon yöntemini kullanarak uluslararası düzeyde PISA sınavına katılan farklı başarı düzeyindeki öğrencilerin okuduğunu anlama başarılarını tahmin etmede etkili olan öğrenci özellikleri nelerdir?

Alt problemler

Araştırmanın temel problem cümlesine dayalı olarak alt problemler aşağıda verilmiştir:

1. PISA öğrenci anketinde ölçülen öğrenci özelliklerinden yararlanarak okuduğunu anlama başarılarını tahmin etmede kullanılan lojistik regresyon yöntemi ile üst, orta, alt başarı düzeyindeki ülkelerin PISA Okuduğunu anlama başarılarını manidar yordayan önemli değişkenler ve bu değişkenlerden gruplarda ortak olanlar nelerdir?
2. PISA öğrenci anketinde ölçülen öğrenci özelliklerinden yararlanarak üst, orta, alt başarı düzeyindeki ülkelerin başarılarını belirlemede etkili değişkenlerin belirlenmesi sürecinde kullanılan lojistik regresyon yönteminin model performansı ne düzeydedir?

Yöntem

Araştırmanın Modeli

Bu araştırmada temel amaç, PISA sınavına giren öğrencilerin öğrenci anketinde bulunan sorulara verdikleri yanıtlardan yararlanarak alt, orta ve üst başarı düzeyine sahip ülkelerin okuduğunu anlama başarılarını tahmin etmede etkili olan öğrenci özelliklerini tespit etmenin yanı sıra çalışmada kullanılan lojistik regresyon yönteminin sınıflama performansının belirlenmesidir. Bu amaca yönelik öncelikle PISA başarılarını tahmin etmek için etkili olan değişkenler tespit edilecek, sonra da PISA başarılarını tahmin etmek için kullanılan lojistik regresyon yönteminin sınıflama doğruluk derecesi belirlenecektir.

Araştırmada öğrencilerin duyuşsal özelliklerini ölçen ölçekler ve sosyo-demografik özelliklerini ölçen anketlerle, farklı başarı düzeyine sahip ülkelerin PISA okuduğunu anlama puanlarına göre başarılı-başarısız olarak tahmin yapılması amaçlanmıştır. Tahminlerde etkili olan önemli değişkenlerin belirlenmesinde yordayıcı korelasyonel nicel araştırma tasarımı kullanılmıştır (Büyüköztürk ve diğerleri, 2018).

Çalışma Evreni ve Örneklemi

Çalışmanın amacı için ülkeler; PISA okuma puanı sıraları dikkate alınarak alt (zayıf), orta, üst (yüksek) başarı grubuna göre belirlendiği için amaçlı örnekleme yöntemi kullanılmıştır. Çünkü amaçlı örnekleme, derinlemesine araştırma yapabilmek amacıyla çalışmanın amacı bağlamında bilgi açısından zengin durumların seçilmesidir (Büyüköztürk ve diğerleri, 2018). Ayrıca, çalışma örnekleme ülkeler seçilirken kayıp veri oranının düşük olmasına ve farklı yüzdelik dilimlere dağılacak şekilde örnekleme yapılmasına

dikkat edilmiştir. Böylece çalışma örnekleminin sağlayacağı bilgilerin zenginleşmesi sağlanmıştır.

PISA 2018 uygulaması, 37'si OECD üyesi olmak üzere 79 ülkeden 612004 öğrencinin katılımıyla gerçekleştirilmiştir. PISA'ya katılacak okul ve öğrencilerin seçim işlemi, OECD tarafından iki aşamalı tabakalı tesadüfi yöntemle belirlenmektedir (MEB, 2019). Çalışma evrenini 2018 PISA'ya katılan ülkelerdeki 15 yaş grubu öğrencilerin tamamı oluşturmaktadır. Bu araştırmanın hedeflerine göre PISA 2018 sınavına katılan ülkelerin tamamından 9 ülke çalışma örneklemini seçilmiştir. Bu 9 ülke, 79 ülke içindeki başarı puanlarına göre yüzdelik sıralamaları dikkate alınarak belirlenmiştir. Son üçte birlik (%68-100) başarı sırasındaki dilimden seçilen ülkelerin yüzdelik sırası Endonezya için 91, Suudi Arabistan için 82 ve Brezilya için 72'dir. Böylece Endonezya, Suudi Arabistan ve Brezilya örneklemini, son %33'lük dilimdeki en başarısız ülkeleri temsil edecek şekilde alt başarı grubu ülkeler olarak adlandırılmıştır. %33 ile %67'lik başarı diliminden seçilen ülkelerin yüzdelik sırası ise Sırbistan için 57, Türkiye için 50 ve Macaristan için 41'dir. Böylece Sırbistan, Türkiye ve Macaristan örneklemini, %33 ile %67'lik dilim aralığındaki orta düzeyde başarılı ülkeleri temsil edecek şekilde orta başarı grubu ülkeler olarak adlandırılmıştır.

Son olarak, ilk üçte birlik başarı sırasındaki dilimden seçilen ülkelerin yüzdelik dilimleri sırası ile Slovenya için 27, ABD için 16 ve Finlandiya için 9'dur. Böylece Slovenya, ABD ve Finlandiya örneklemini, ilk %33'lük dilimdeki en başarılı ülkeleri temsil edecek şekilde üst başarı grubu ülkeler olarak adlandırılmıştır. Bu 9 ülkenin tamamı ise PISA'nın uygulandığı 79 ülkeyi temsil edecek şekilde PISA uygulama örneklemini olarak adlandırılmıştır. Bu durumda araştırmaya alt grup ülkeler için 28706, orta başarı grubu ülkeler için 18518, üst başarı grubu ülkeler için 16727 kişi dâhil edilmiştir. Alt, orta ve üst başarı grubu olmak üzere her bir grupta 3 ülkenin bulunduğu ve ülkelerin birleştirilmesi ile araştırmaya alınan toplam 9 ülkenin örneklem büyüklüğü 63951 kişiye ulaşmaktadır.

Tablo 1. Araştırma sürecinde kullanılan değişkenler ve kodları

Değişken Adı	Kodu	Değişken Adı	Kodu
Olumlu duygular	SWBP	BİT kaynakları	ICTRES
İleri düzey görevlere motivasyon	WORKMAST	Öğrenciden beklenen mesleki statü	BSMJ
Özyeterlik	RESILIENCE	Ebeveynlerin duygusal desteği	EMOSUPS
Yaşamın anlamı	EUDMO	Test dili öğrenme süresi	LMINS
Öğrenci rekabeti	PERCOMP	Toplam öğrenme süresi (Hafta)	TMINS
Öğretmenin okumaya teşviki	STIMREAD	Okula ait hissetme	BELONG
Öğrenci iş birliği	PERCOOP	Evdeki eğitim kaynakları	HEDRES
Okulun değeri	ATTLNACT	Öğretim adaptasyonu	ADAPTIVITY
Öğretmen Coşkusu	TEACHINT	Rekabete karşı tutumlar	COMPETE
Okuma keyfi	JOYREAD	Zorbalığa maruz kalma	BEINGBULLIED
Anne eğitim düzeyi	MISCED	Başarısızlık korkusu	GFOFAIL
Baba eğitim düzeyi	FISCED	Öğrenme amaçları	MASTGOAL
Disiplin İklimi	DISCLIMA	Ailenin mal varlığı	WEALTH
Öğretmen desteği	TEACHSUP	Okuma yetkinlik algısı	SCREADCOMP
Evdeki eğitimsel eşyalar	HOMEPOS	Okuma zorluk algısı	SCREADDIFF
Öğretmenin öğretimi yönlendirmesi	DIRINS	Pisa testinin zorluk algısı	PISADIFFF
Sosyo-ekonomik durum indeksi	ESCS	Okuduğunu anlama başarısı	PVREAD
Öğretmen geri bildirimi	PERFEED		

Veri Toplama Süreci

Araştırma sürecinde kullanılan veriler 2020 yılında paylaşımına açılan veri tabanından <http://www.oecd.org/pisa/data/2018database/> adresi kullanılarak elde edilmiştir. Araştırma kapsamına 2018 PISA verilerine ait toplamda 34 bağımsız ve okuduğunu anlama başarı düzeyine karşılık gelen on makul değerlerin ortalaması olan bir bağımlı değişken alınmıştır. Veriler incelendiğinde bağımlı değişkene ilişkin kayıp veriye rastlanmamıştır. Ancak, 34 bağımsız değişkene ilişkin yanıt vermeyen bireyler analizden çıkarıldıktan sonra saptanan az sayıda uç değerler SPSS programı kullanılarak verilerden çıkarılmıştır. Ayrıca tüm bağımsız değişkenlerin her birinde kayıp veriye rastlanmıştır. Bu bağımsız değişkenlere yönelik kayıp veriler tesadüfi dağıldığı için SPSS programı kullanılarak çoklu atama yöntemi ile atama yapılmıştır.

Veri Toplama Araçları

Araştırmada veriler 2018 yılında uygulanan PISA okuduğunu anlama sınavındaki test ve öğrenci anketinden elde edilmiştir. Kullanacağımız bağımlı değişkeni elde etmek için öncelikle okuduğunu anlama başarı puanı sürekli nicel değişken olarak elde edilmiş ve PVREAD olarak adlandırılmıştır. Burada PVREAD makul okuduğunu anlama başarı puanını ifade eder. PVREAD değişkeni PISA 2018 okuduğunu anlama başarı testinden her bir öğrencinin aldığı ve öğrencilerin sahip olabileceği yetenekler yelpazesini ifade eden (Wu ve Adams, 2002), 10 makul puan değerinin (PV1, PV2, ..., PV10) ortalamasıdır. Daha sonra, PVREAD puanının sınıflandırılması sürecinde PISA 2018 okuduğunu anlama OECD başarı ortalaması olan 487 kesme puanı olarak kullanılmıştır. PVREAD değeri 0-486,999 arası düşük ve 487-1000 arası yüksek olacak şekilde sınıflanarak elde edilmiştir (0=Başarısız, 1=Başarılı). Ayrıca, Lojistik Regresyon modelinde kullanılacak 34 bağımsız değişken ise alanyazındaki çalışmalarda en çok kullanılan değişkenler dikkate alınarak Tablo 2'de verildiği gibi belirlenmiştir.

Verilerin Analiz Yöntemi

Verilerin analizinde öncelikle PISA'ya katılan 79 ülke, başarı yüzdelik sırasına konuldu. Ardından eşit dilimlerle ülkeler alt, orta ve üst dilimler olarak ayrıldı. Yüzdelik dilimlerine ve kayıp veri azlığına göre alt, orta ve üst grubun her birinde üç ülke olmak üzere toplamda dokuz ülke seçildi. Verilerin analizinde SPSS 21 ve Excel programları kullanılmıştır. Araştırma kapsamında Tablo 2'de verilen 34 değişken bağımsız değişken ve iki düzeyli başarı değişkeni (0=başarısız, 1=başarılı) bağımlı değişken olarak alınmıştır. Verilerin analizi sürecinde, her birinde üç ülkenin yer aldığı alt, orta ve üst başarı gruplarındaki ülkelerin örneklem verileri kullanılarak PISA 2018 okuduğunu anlama başarısını yordamada etkili olan bağımsız değişkenlerin belirlenmesi ve bu değişkenlerin bağımlı değişkeni sınıflama performansını belirlemek için Lojistik Regresyon Modeli (Binary) kullanılmıştır.

Öncelikle alt, orta ve üst grup ülkeler örnekleminden gelen 34 adet bağımsız değişkene ait verilerin SPSS programı ile analiz edilmesi sonucunda 2 düzeyli PISA okuduğunu anlama başarı değişkenini yordamada anlamlı olan bağımsız değişkenlerin 25-28 adet aralığında değiştiği saptanmıştır. Ayrıca, 34 girdi değişken ile oluşturulan modellerin sınıflama performansları alt grup için %85,7, orta grup için %75,6 ve üst grup için %75,8 olarak hesaplanmıştır. Daha sonra, daha az sayıda değişken ile başarıyı yordamak için her bir modeldeki girdi değişken sayısında azalma olduğunda oluşturulacak modelin performansının ne düzeyde değiştiği incelenmiştir. İnceleme sonucunda, her bir modelin sınıflama performansında dikkate değer düşme olmayacak biçimde ayarlama yapılarak değişken sayısı azaltılıp başarıyı en iyi yordayan 10 değişken seçilmiştir. Son olarak, seçilen değişkenler üzerinden tekrardan Lojistik Regresyon yöntemi ile analiz yapılarak başarıyı yordamada anlamlı değişkenler ve modellerin sınıflama performansları belirlenmiştir. Elde edilen sonuçlara göre alt, orta ve üst grubun her birinde girdi olarak kullanılan 10 adet değişkenin tamamı anlamlı sonuçlar vermiştir. Ayrıca, seçilen 10 değişken ile modellerin sınıflama performansları alt grup için %85, orta grup için %73,3 ve üst grup için %72,5 olarak hesaplanmıştır. Bu hesaplanan performans kriterlerinin 34 bağımsız değişken kullanıldığında hesaplanan performans kriterlerinden önemli düzeyde bir farklılığa sahip olmadığı belirlenmiş ve böylece başarıyı tahmin etmede, daha az değişken ile yüksek performansa sahip modeller kurulabildiği gösterilmiştir.

Alt, orta ve üst grup ülkeler örnekleme üzerinden oluşturulan modellerin sınıflama performanslarını belirlemede Doğru Sınıflama Oranı (DSO) kullanılmıştır.

$$DSO = \frac{DP+DN}{DP+DN+YP+YN} \quad (1)$$

Formül (1)'de DP doğru pozitif, DN doğru negatif, YP yanlış pozitif, YN yanlış negatif durumların sayısını göstermektedir.

Lojistik Regresyon Modeli

Basit ve çoklu doğrusal regresyon modelinde bağımlı değişken eşit oranlı ya da eşit aralıklı ölçek düzeyindedir. Ancak bağımlı (yordanan) değişkenin iki veya daha fazla sınıflı olduğu durumda doğrusal regresyon yöntemlerinde parametre kestirimlerinde kullanılan en küçük kareler yöntemini

kullanmak, normal dağılıma uymayan hata terimi gibi bazı varsayımların sağlanamamasından dolayı uygun olmamaktadır. Bu durumda lojistik regresyon kullanmak uygun olmaktadır. İki sınıflı lojistik regresyonda, Y_i , Bernoulli olasılık dağılımına sahip rasgele değişkenin 1 değerini alma olasılığı, π_i ve 0 değerini alma olasılığı ise $(1 - \pi_i)$ olur.

Bağımsız (yordayıcı) değişkenler $X = (X_1, X_2, X_3, \dots, X_n)$ biçiminde ifade edildiği durumda, herhangi i 'inci gözlemin regresyon modeli $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ biçiminde gösterilen modelde bulunan bağımsız değişkenler için bir kısıtlama yoktur. Bu modelde Y_i bağımlı değişkeni $-\infty$ ile $+\infty$ arasında sayıda değer alabilmektedir. Ancak bağımlı değişkenin alacağı 0 ve 1 değerleri için model yapısal olarak değişmektedir.

$P(Y_i = 1)$ i 'inci gözlemin 1 değerini alma olasılığı olduğundan $E(\varepsilon_i) = 0$ olduğu için Y_i 'nin beklenen değeri; $E(Y_i) = 1 \cdot P(Y_i = 1) + 0 \cdot P(Y_i = 0) = P(Y_i = 1)$ olacaktır. Bağımlı değişken iki kategorili olduğu durumda $(0 \leq E(Y_i) = P(Y_i = 1) \leq 1)$ olduğundan en büyük sınırlama bağımlı değişkenin beklenen değeri üzerinde olacaktır. Bunun sonucunda;

$E(Y_i) = P(Y_i = 1) = \beta_0 + \beta_1 X_i$ olarak ifade edilebilir. Bu model bağımlı değişkeni 0 ile 1 arasında olasılık değeri alan "Doğrusal olasılık modeli" olarak adlandırılır (Tatlıdıl, 1996). Daha sonra bu sınırlamanın önüne geçmek için ilk olarak;

$E(Y_i) = P(Y_i = 1) = \beta_0 + \beta_1 X_i$ modelindeki olasılık değerlerinde $\frac{P}{1-P}$ (Odds) dönüşümü yapılarak bağımlı değişkenin sınırları 0 ile $+\infty$ aralığı olarak sonrada $\frac{P}{1-P}$ oranının doğal logaritması alınıp lojit dönüşümü yapılarak bağımlı değişkenin sınırları $-\infty$ ile $+\infty$ olarak genişletilir.

Dönüşümler sonucunda elde edilen model:

$$\ln \left[\frac{E(Y=1|X_i)}{E(Y=0|X_i)} \right] = \ln \left[\frac{P_i}{1-P_i} \right] = \beta_0 + \beta_1 X_i \text{ olarak ifade edilir. (2)}$$

Bu model lojistik model veya lojit model olarak adlandırılır (Hosmer ve Lemeshow, 2000). Modelden elde edilen olasılık değeri ise:

$$P_i = \frac{\exp(\beta_0 + \beta_1 X_i)}{1 + \exp(\beta_0 + \beta_1 X_i)} \text{ olarak ifade edilir (Collet, 2003). (3)}$$

Lojistik regresyon modelinde en çok olabilirlik yöntemi kullanılarak model katsayıları kestirilir. Bu yöntem ile kestirilen katsayıların anlamlılığı Wald testi veya G istatistiği ile elde edilir. Lojistik modelde bulunan "başarılı olmanın", "başarılı olmamaya oranı" $\frac{P}{1-P}$ ile gösterilip bu oran "Odds" olarak bilinir. Bu oran $0 \leq Odds \leq +\infty$ aralığında değer alıp lojistik regresyon modelinde değişkenlere ait katsayıların yorumunda kullanılır. 0 ile 1 arasında değer alırsa başarısızlık olasılığı başarılı olma olasılığından daha fazla, 1 olursa başarısız olma olasılığı ile başarılı olma olasılığı eşit, 1'den büyük olursa başarılı olma olasılığı başarısız olma olasılığından daha fazladır.

Bulgular ve Yorumlar

PISA okuduğunu anlama başarısını belirlemek için 34 bağımsız, 1 bağımlı değişkene alt grup ülkelerden 28706, orta grup ülkelerden 18518, üst grup ülkelerden 16727 öğrencinin yanıtlarına ilişkin 3 farklı veri kümesi ile çalışılmıştır. Her bir veri kümesi üzerinden lojistik regresyon analizi yapılmıştır.

Tablo 2.Alt grup ülkeler için lojistik regresyon modelindeki değişkenler

Değişken Kodu	B	S.H	Wald	sd	p	Beklenen(B)
HOMEPOS	0,490	0,046	114,328	1	,000	1,632
JOYREAD	0,475	0,024	404,153	1	,000	1,608
ESCS	0,468	0,026	331,893	1	,000	1,596
WORKMAST	0,401	0,021	362,333	1	,000	1,493
ICTRES	0,327	0,039	70,524	1	,000	1,386
GFOFAIL	0,244	0,019	172,343	1	,000	1,276
PISADIFF	-0,580	0,021	800,142	1	,000	0,560
WEALTH	-0,508	0,036	204,667	1	,000	0,602
EUDMO	-0,463	0,022	431,416	1	,000	0,629
DIRINS	-0,360	0,018	396,487	1	,000	0,697
Constant	-0,713	0,031	532,956	1	,000	0,490

Birinci Araştırma Sorusuna İlişkin Bulgular

PISA öğrenci anketinde ölçülen öğrenci özelliklerinden yararlanarak okuduğunu anlama başarısını tahmin etmede kullanılan lojistik regresyon yöntemi ile üst, orta, alt başarı düzeyindeki ülkelerin PISA okuduğunu anlama başarılarını manidar yordayan önemli değişkenler ve bu değişkenlerden gruplarda ortak olanların neler olduğuna ilişkin yapılan analizler sonucunda elde edilen bulgular verilmiştir. Analize alınan 34 bağımsız değişken üzerinden backward methodu ve Wald istatistiği kullanılarak yapılan analiz sonucunda alt grupta 10. adım sonunda 25, orta grupta 9. adım sonunda 26, üst grupta ise 7. adım sonunda 28 bağımsız değişken anlamlı olduğu için modele alınmıştır. Daha sonra her bir modelde okuduğunu anlama başarısını yordamada en etkili olan 10 tane değişken seçilerek backward (geriye doğru) methodu ve Wald istatistiği kullanılarak tekrardan analiz yapılmıştır. Analiz sonucunda alt, orta ve üst grubun her birinde 1. adım sonunda her bir gruptaki 10 adet girdi değişken tamamı anlamlı olduğu için modele alınıp analizlerin tamamı bu değişkenlerle gerçekleştirilmiştir.

Alt grup veri kümesine ilişkin analiz sonuçları Tablo 2’de verilmiştir.

Tablo 2 incelendiğinde, alt grup ülkeler örnekleminde PISA okuduğunu anlama başarısını olumlu etkileyen değişkenlerden en yüksek düzeyden en düşük düzeye doğru sıralaması HOMEPOS, JOYREAD, ESCS, WORKMAST, ICTRES, GFOFAIL biçimindedir. Bu değişkenlerden PISA Okuduğunu anlama başarısını olumlu yönde en iyi yordayan

HOMEPOS, JOYREAD, ESCS değişkenleridir. Bunlardan en etkili yordayıcı değişken ise HOMEPOS değişkenidir. HOMEPOS değişkeni bakımından alt grubun başarılı olma olasılığı başarısız olma olasılığının 1,632 katıdır. Diğer yandan alt grup ülkelerde PISA okuduğunu anlama başarısını olumsuz etkileyen değişkenlerden en yüksek düzeyden en düşük düzeye doğru sıralaması PISADIFF, WEALTH, EUDMO, DIRINS biçimindedir. Bu değişkenlerden yordayıcılığı en yüksek değişken ise PISADIFF değişkenidir. PISADIFF değişkeni bakımından alt grubun başarısız olma olasılığı başarılı olma olasılığının (1/0,560) 1,785 katıdır.

Tablo 3 incelendiğinde, orta grup ülkelerde PISA okuduğunu anlama başarısını olumlu etkileyen değişkenlerden en yüksek düzeyden en düşük düzeye doğru sıralaması HOMEPOS, JOYREAD, ESCS, ICTRES, EMOSUPS, STIMREAD biçimindedir. Bu değişkenlerden PISA okuduğunu anlama başarısını olumlu yönde en iyi yordayan HOMEPOS, JOYREAD değişkenleridir. Bunlardan en etkili yordayıcı değişken ise HOMEPOS değişkenidir. HOMEPOS değişkeni bakımından orta grubun başarılı olma olasılığı başarısız olma olasılığının 1,626 katıdır. Diğer yandan orta grup ülkelerde PISA okuduğunu anlama başarısını olumsuz etkileyen değişkenlerden en yüksek düzeyden en düşük düzeye doğru sıralaması WEALTH, PISADIFF, EUDMO, DIRINS biçimindedir. Bu değişkenlerden yordayıcılığı en yüksek değişken ise WEALTH değişkenidir. WEALTH değişkeni bakımından orta grubun başarısız olma olasılığı başarılı olma olasılığının (1/0,459) 2,17 katıdır.

Tablo 3.Orta grup ülkeler için lojistik regresyon modelindeki değişkenler

Değişken Kodu	B	S.H	Wald	sd	p	Beklenen(B)
HOMEPOS	0,486	0,050	92,895	1	0,000	1,626
JOYREAD	0,436	0,018	587,852	1	0,000	1,546
ESCS	0,433	0,026	283,157	1	0,000	1,542
ICTRES	0,325	0,040	65,346	1	0,000	1,384
EMOSUPS	0,309	0,019	259,775	1	0,000	1,362
STIMREAD	0,225	0,019	135,899	1	0,000	1,253
WEALTH	-0,778	0,049	256,821	1	0,000	0,459
PISADIFF	-0,579	0,020	846,390	1	0,000	0,560
EUDMO	-0,350	0,019	351,707	1	0,000	0,704
DIRINS	-0,318	0,020	259,412	1	0,000	0,728
Constant	-0,374	0,023	259,705	1	0,000	0,688

Tablo 4. Üst grup ülkeler için lojistik regresyon modelindeki değişkenler

Değişken Kodu	B	S.H	Wald	sd	p	Beklenen(B)
ESCS	0,596	0,031	379,593	1	0,000	1,814
JOYREAD	0,406	0,020	418,583	1	0,000	1,501
SCREADCOMP	0,381	0,023	286,229	1	0,000	1,464
TEACHSUP	0,297	0,022	176,648	1	0,000	1,346
GFOFAIL	0,266	0,020	181,140	1	0,000	1,305
HOMEPOS	0,153	0,039	15,462	1	0,000	1,165
PISADIFF	-0,608	0,021	799,557	1	0,000	0,545
DIRINS	-0,321	0,023	190,749	1	0,000	0,725
WEALTH	-0,275	0,033	70,123	1	0,000	0,760
EUDMO	-0,211	0,020	110,822	1	0,000	0,810
Constant	0,378	0,020	354,032	1	0,009	1,459

Tablo 5. Okuduğunu anlama başarısını yordayan önemli ortak değişkenler

Alt Grup Örnekleme	Orta Grup Örnekleme	Üst Grup Örnekleme
1.(-) PISADIFF	2.(-) PISADIFF	1.(-) PISADIFF
2.(-) WEALTH	1.(-) WEALTH	7.(-) WEALTH
3.(+) HOMEPOS	3.(+) HOMEPOS	10.(+) HOMEPOS
4.(+) JOYREAD	4.(+) JOYREAD	3.(+) JOYREAD
5.(+) ESCS	5.(+) ESCS	2.(+) ESCS
6.(-) EUDMO	6.(-) EUDMO	9. (-) EUDMO
8. (-) DIRINS	8.(-) DIRINS	5. (-) DIRINS

-, +: Lojistik regresyon modelindeki beta katsayısının işareti, 1,2, 3...: Önemli değişkenlerin önem sıralaması

Tablo 4 incelendiğinde, üst grup ülkelerde PISA okuduğunu anlama başarısını olumlu etkileyen değişkenlerden en yüksek düzeyden en düşük düzeye doğru sıralaması ESCS, JOYREAD, SCREADCOMP, TEACHSUP, GFOFAIL, HOMEPOS biçimindedir. Bu değişkenlerden PISA okuduğunu anlama başarısını olumlu yönde en iyi yordayan ESCS, JOYREAD ve SCREADCOMP değişkenleridir. Bunlardan en etkili yordayıcı değişken ise ESCS değişkenidir. ESCS değişkeni bakımından üst grubun başarılı olma olasılığı başarısız olma olasılığının 1,814 katıdır. Diğer yandan üst grup ülkelerde PISA okuduğunu anlama başarısını olumsuz etkileyen değişkenlerden en yüksek düzeyden en düşük düzeye doğru sıralaması PISADIFF, DIRINS, WEALTH, EUDMO biçimindedir. Bu değişkenlerden yordayıcılığı en yüksek değişken ise PISADIFF değişkenidir. PISADIFF değişkeni bakımından üst grubun başarısız olma olasılığı başarılı olma olasılığının (1/0,545) 1,834 katıdır.

Tablo 5 incelendiğinde alt, orta ve üst grupta PISA okuduğunu anlama başarısını yordamada en etkili olan 10 tane değişkenden yedi tanesinin ortak olduğu saptanmıştır. PISA okuduğunu anlama başarısını olumlu yönde yordayan önemli ortak değişkenler HOMEPOS, JOYREAD, ESCS değişkenleridir. Okuduğunu anlama başarısını olumlu yönde yordamada en etkili olan önemli ortak değişken ise HOMEPOS değişkenidir. Başarılı olumsuz yönde yordayan

önemli ortak değişkenler ise PISADIFF, WEALTH, EUDMO, DIRINS değişkenleridir. Başarılı olumsuz yönde yordamada en etkili olan önemli ortak değişken ise PISADIFF değişkenidir. Grupların tamamında ortak olmayan değişkenler ise; WORKMAST, ICTRES GFOFAIL, EMOSUPS, STIMREAD, SCREADCOMP, TEACHSUP biçimindedir. Bu değişkenlerden ICTRES alt ve orta grupta, GFOFAIL ise alt ve üst grupta başarıyı yordamada etkili olan önemli ortak değişkenler olarak saptanmıştır.

İkinci Araştırma Sorusuna İlişkin Bulgular

Tabloda 6'da verilen Omnibus testi sonuçları incelendiğinde, $p < 0,05$ olduğundan sabit terim dışındaki tüm lojit katsayılarının değerleri sıfırdan farklıdır. Yani, model katsayıları anlamlıdır. Dolayısıyla modellerin her birindeki bağımsız değişkenler bağımlı değişken üzerinde anlamlı bir fark oluşturmuştur. Böylece oluşturulan üç modelin de anlamlı olduğu sonucuna varılır.

Tabloda 7'de verilen Hosmer-Lemeshow testi sonuçlarına göre, alt grup, orta grup ve üst grup verileri üzerinden oluşturulan üç modelin tamamında $p > 0,05$ olduğu için oluşturulan modellerin tümünde model veri uyumunun olduğu söylenir.

Tablo 6. Model katsayılarının anlamlılık testi

Gruplar	Modeller	Adımlar	Chi-square	df	p
Alt grup	Model-1	Adım1	6181,458	10	0,000
Orta grup	Model-2	Adım 1	5132,391	10	0,000
Üst grup	Model-3	Adım 1	4911,303	10	0,000

Tablo 7. Uyum iyiliği testi

Grup	Modeller	Adım	Chi-square	df	p
Alt grup	Model-1	1	19,652	8	0,111
Orta grup	Model-2	1	7,424	8	0,492
Üst grup	Model-3	1	12,478	8	0,131

Tablo 8. Model performans kriterleri

Gruplar	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square	Doğru sınıflama Oranı
Alt grup	20231,414	0,194	0,322	85
Orta grup	19777,101	0,242	0,327	73,3
Üst grup	18008,083	0,254	0,341	72,5

Tablo 8 incelendiğinde alt grupta yapılan analizin Cox & Snell R Square değeri 0,194, Nagelkerke R Square değeri 0,322 ve doğru sınıflama oranı ise %85 olarak hesaplanmıştır. Bunlar için Cox ve Snell R Square ve Nagelkerke R Square değerlerinin 0.50 'nin üzerinde olması, McFadden R Square değerlerinin ise 0.20-0.40 arasında olması modelde ilişkinin güçlü olduğunu modelin uygun olduğunu göstermektedir (Cengiz, 2009; Alpar, 2011). Bu değerlere göre alt grup verileriyle oluşturulan çok değişkenli modelin bağımlı değişkeni makul düzeyde açıkladığı, modelin bireylerin başarı durumunu (başarısız-başarılı) doğru sınıflandırma oranının yüksek olduğu sonucuna varılmıştır. Orta grupta Cox & Snell R Square değeri 0,242, Nagelkerke R Square değeri 0,327 ve doğru sınıflama oranı ise %73,3 olarak hesaplanmıştır. Bu değerlere göre orta grup verileriyle oluşturulan çok değişkenli modelin bağımlı değişkeni makul düzeyde açıkladığı, modelin bireylerin başarı durumunu (başarısız-başarılı) doğru sınıflandırma oranını iyi düzeyde olduğu sonucuna varılmıştır. Üst grupta ise, Cox & Snell R Square değeri 0,254, Nagelkerke R Square değeri 0,341 ve doğru sınıflama oranı ise %72,5 olarak hesaplanmıştır. Bu değerlere göre üst grup verileriyle oluşturulan çok değişkenli modelin bağımlı değişkeni makul düzeyde açıkladığı, modelin bireylerin başarı durumunu (başarısız-başarılı) doğru sınıflandırma oranının iyi düzeyde olduğu sonucuna varılmıştır.

Sonuçlar ve Tartışma

Araştırmada alt, orta ve üst grup ülkeler olmak üzere her bir gruptaki öğrencilerin PISA okuduğunu anlama başarısını yordayan etkili değişkenler belirlenmiş ve bu gruplarda PISA okuduğunu anlama başarısını yordayan değişkenler karşılaştırılıp ortak olanları belirlenmiştir. Ayrıca araştırmada kullanılan lojistik regresyon yönteminin her bir gruptaki sınıflama performansı belirlenmiştir. Araştırma sonucuna göre alt, orta ve üst grubun her birinde okuduğunu anlama başarısını en iyi yordayan 10 değişken belirlenmiştir. Genel olarak gruplarda okuduğunu anlama başarısını olumlu yönde yordamada en etkili değişkenler HOMEPOS, JOYREAD ve ESCS değişkenleri iken olumsuz yönde yordamada en etkili değişkenler PISADIFF, WEALTH, EUDMO ve DIRINS değişkenleri olarak tespit edilmiştir.

Evdeki eğitimsel eşyalar değişkeni başarıyı yordamada alt ve orta grupta üçüncü sırada, üst grupta ise onuncu sırada öneme sahip ortak değişkendir. Elde edilen bu bulgu, alanyazında öğrencinin evdeki çalışma odası ve çalışma masası gibi ders çalışma ortamıyla ilgili evdeki eğitimsel eşyaların okuma başarısını artırdığına yönelik (Güzle Kayır, 2012) araştırma sonucu ile tutarlıdır. Elde edilen bu sonuç üst gruptaki ülkelerin sosyoekonomik düzeylerinin iyi olmasına bağlanabilir. Okuma keyfi değişkeni, başarıyı yordamada alt ve orta grupta dördüncü sırada, üst grupta ise üçüncü sırada öneme sahip ortak değişkendir. Bu bulgu, Tavşancıl ve arkadaşları (2019) tarafından yapılan araştırmada okuma keyfi değişkeninin PISA okuduğunu anlama başarısını önemli

ölçüde etkilediğine yönelik sonuç ile uyumludur. Sosyo-ekonomik durum değişkeni başarıyı yordamada alt ve orta grupta beşinci, üst grupta ise ikinci sırada öneme sahip ortak değişken olduğuna yönelik bulgu, 2018 PISA sınavına katılan tüm ülkelere yönelik PISA ön raporundan elde edilen araştırma sonuçlarıyla uyum göstermektedir (MEB, 2019). Bilgi iletişim teknolojileri (BİT) kaynakları değişkeni başarıyı yordamada alt grupta dokuzuncu, orta grupta ise yedinci sırada öneme sahip ortak değişken olarak saptanması, Urfalı Dadandı ve arkadaşları (2018) tarafından yapılan araştırmada BİT kaynaklarının okuduğunu anlama başarısını anlamlı düzeyde yordamasına yönelik elde edilen bulgu ile uyumludur.

Özetle, bulgular PISA testinin zorluk algısı, ailenin mal varlığı, evdeki eğitimsel eşyalar, okuma keyfi ve sosyo-ekonomik durum değişkenlerinin grupların tamamında PISA okuduğunu anlama başarısını yordamada önemli ölçüde etkiye sahip olduklarını göstermektedir. Dolayısıyla PISA testine yönelik zorluk algısı olan, ailesinin mal varlığı çok olan, evdeki eğitimsel eşyaları yetersiz olan, okumadan keyif almayan, sosyo-ekonomik durumu kötü olan öğrencilerin PISA okuduğunu anlama başarısı bakımından başarısız olacakları düşünülebilir. Bu sorunun önüne geçilmesi, uygun eğitim politikalarıyla sorunlara müdahale edilmesi ve bu değişkenlere yönelik farklılıkların yol açtığı olumsuzlukların giderilmesi ile mümkündür.

Yapılan araştırmada grupların tamamında kullanılan lojistik regresyon modelinin sınıflama performansının iyi düzeyde olduğu görülmüş ve ayrıntılı incelendiğinde modelin sınıflama performansının gruplara göre farklılık gösterdiği belirlenmiştir. Grupların sınıflama performansının en yüksek olandan en düşük olana doğru sırası; alt, orta ve üst başarı grubu ülkeler biçimindedir. En önemlisi ise model performansının değişken sayısının azalmasından önemli düzeyde etkilenmemesi bulgusudur. Bu sonuca dayalı olarak daha az sayıda değişken ile öğrencilerin PISA okuduğunu anlama başarısını yordamanın mümkün olduğu söylenebilir. Lojistik regresyon modeli ile alt, orta ve üst başarı grubu ülkeler örnekleme için yapılan araştırmada daha az sayıda değişken ile öğrencilerin PISA okuduğunu anlama başarısının iyi derecede yordanması, bu ülkelerdeki karar alıcıların eğitim politikalarına yön verirken PISA okuduğunu anlama başarısını olumlu etkileyen faktörlere daha fazla önem vermesiyle başarı artırılabilir.

Sonraki araştırmalar için ise şu öneriler sunulabilir:

1. Araştırmada kullanılan PISA okuduğunu anlama başarısı üç veya daha fazla kategoriye ayrılarak PISA okuduğunu anlama başarısını etkileyen değişkenlerin iki kategorili olduğu durumdaki etkili değişkenlere göre farklılık gösterip göstermediği incelenebilir.
2. Ayrıca kullanılan modelin PISA okuduğunu anlama başarısına yönelik sınıflama performansının PISA okuduğunu anlama başarısının üç veya daha fazla kategoriye ayrıldığında iki kategorili olduğu

durumdaki sınıflama performansına göre farklılık gösterip göstermeyeceği incelenebilir.

3. Araştırmanın bağımsız değişkenlerine ek olarak farklı değişkenler eklenerek başarıyı yordamada etkili olan önemli değişkenler belirlenebilir.

Yazar Katkı Oranı

Tüm yazarlar makalenin tüm süreçlerinde eşit oranda rol almışlardır. Tüm yazarlar çalışmanın son halini okumuş ve onaylamıştır.

Etik Kurul Beyanı

Bu çalışma Hacettepe Üniversitesi Akademik Etik Kurulundan alınan 30.06.2020 tarihli ve 51944218-300/00001133396 sayılı numaralı etik kurul izni ile yürütülmüştür.

Çatışma Beyanı

Yazarlar çalışma kapsamında herhangi bir kurum veya kişi ile çıkar çatışması bulunmadığını beyan etmektedir.

Kaynakça

- Akkurt, Z., ve Karabağ Köse, E. (2019). Öğrenci başarısının okul, öğretmen ve aileyle ilgili değişkenler açısından incelenmesi. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, 47, 1-16. <https://doi.org/10.9779/pauefd.451853>
- Alpar, R. (2011). *Çok Değişkenli İstatistiksel Yöntemler*. Ankara: Detay Yayıncılık.
- Berberoğlu, G., Çalışkan, M., ve Karşı, N. (2017). *PISA 2015 sonuçlarına göre Türk öğrencilerinin ulaştığı düşünme süreçleri ve bu süreçlerle ilişkili olan değişkenler*. Working Paper, August 2017
- Büyüköztürk, Ş., Kılıç Çakmak, E., Akgün, Ö. E., Karadeniz, Ş., ve Demirel, F. (2018). *Bilimsel Araştırma Yöntemleri*. Ankara: Pegem Yayıncılık.
- Cengiz, E. (2009). *ARCS motivasyon modelinin fen ve teknoloji dersinde öğrencilerin başarısına ve öğrenmenin kalıcılığına etkisi* (Yüksek Lisans Tezi, Fen Bilimleri Enstitüsü).
- Coşkun, T. (2013). *Uluslararası öğrenci başarı değerlendirme programı (PISA) 2009 uygulaması okuma becerileri okuryazarlığı'nı etkileyen faktörler*. (Yayınlanmamış Yüksek Lisans Tezi). Akdeniz Üniversitesi Eğitim Bilimleri Enstitüsü, Antalya.
- Erdas Kartal, E., Doğan, N., Yıldırım, S. (2017). Exploration of the Factors Influential on the Scientific Literacy Achievement of Turkish Students in PISA. *Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi (EFMED) Cilt 11, Sayı 1*, sayfa 320-339.
- Güzle Kayır, Ç. (2012). *PISA 2009 Türkiye verilerine dayanarak okuma becerileri alanında başarılı okullar ile başarısız okulları ayırt eden okul içi etmenler ve sosyo-ekonomik faktörler*. (Yayınlanmamış Yüksek lisans tezi). Akdeniz Üniversitesi Sosyal Bilimler Enstitüsü, Antalya.
- Ho Yu, C., Kaprolet, C., Jannasch Pennell, A., ve Digangi, S. (2012). 'A data mining approach to comparing american and canadian grade 10 students' PISA science test performance. *Journal of Data Science* 10, 441-464.
- Hosmer, D.W., ve Lemeshow, S. (2000). *Applied logistic regression*. John WileySons, Inc : Canada.
- MEB (2019). *PISA 2015 ulusal ön raporu*. Ankara: Meb.
- OECD (2019). *PISA 2018 Assessment and analytical framework*. Paris: OECD Publishing.
- Önder, R., Gelbal, S. (2016). PISA 2012 sonuçlarına göre matematik öz-yeterlik ve kaygı puanlarının bazı değişkenlere göre incelenmesi. *Eğitim ve Öğretim Araştırmaları Dergisi* 5(3),271-278.
- Özer Özkan, Y., ve Acar Güvendir, M. (2014). Socioeconomic factors of students' relation to mathematic achievement: comparison of PISA and OBBS *International Online Journal of Educational Sciences*, 6 (3), 776-789.
- Tatlıdil, H. (1992). *Uygulamalı çok değişkenli istatistiksel analiz*. Ankara: Engin Yayınları.
- Tavşancıl, E., Yıldırım, Ö., ve Bilican Demir, S. (2019). Direct and indirect effects of learning strategies and reading enjoyment on PISA 2009 reading performance. *Eurasian Journal of Educational Research*, 82, 169-190.
- Urfalı Dadandı, P., Dadandı, İ., ve Koca, F. (2018). PISA 2015 Türkiye sonuçlarına göre sosyoekonomik faktörler ile okuma becerileri arasındaki ilişkiler. *Uluslararası Türkçe Edebiyat Kültür Eğitim Dergisi*, 7(2), 1239-1252.
- Uyulgan, M. A., ve Akkuzu, N. (2017). Ortaöğretim öğrencilerinin akademik başarılarını yordayan kişisel faktörler üzerine bir araştırma. *Bartın Üniversitesi Eğitim Fakültesi Dergisi*, 6(3), 1211-1226. <https://doi.org/10.14686/buefad.341344>
- Wu, M., ve Adams, R.J. (2002). *Plausible Values – Why They Are Important*. International Objective Measurement Workshop, New Orleans.
- Yıldırım, Ö. (2012). *Okuduğunu anlama başarısıyla ilişkili faktörlerin aşamalı doğrusal modellemeyle belirlenmesi (PISA 2009 Hollanda, Kore ve Türkiye karşılaştırması)*. (Yayınlanmış Doktora Tezi). Ankara Üniversitesi Eğitim Bilimleri Enstitüsü Ankara.

Extended Summary

Introduction

The main purpose of this study is to determine the important predictors of the two-class (unsuccessful-successful) PISA reading comprehension achievement of countries with different achievement levels (low, medium, and high achievement groups) by using the scores of 34 independent variables in the PISA student questionnaire. The other aim of the study is to show whether high-performance logistic regression classification models can be obtained by using a small number of more important independent variables. The aim is that the study will be one of the few studies to reduce the number of variables by selecting the more important independent variables that make the Logistic Regression classification model performance the highest.

What are the student characteristics that are effective in predicting the reading comprehension achievement of students with different success levels who take the PISA exam at the international level using the Logistic Regression method?

Based on the main problem statement of the research, the sub-problems are given below.

1. What are the important variables that predict the PISA Reading Comprehension success of high, medium, and low-achievement countries, and what are the common ones among these variables with the logistic regression method used to estimate the reading comprehension achievement by making use of the student characteristics measured in the PISA student questionnaire?
2. What is the model performance of the logistic regression method used in determining the variables that are effective in determining the achievements of the high, medium, and low-achievement countries by making use of the student characteristics measured in the PISA student questionnaire?

Method

In this study, a relational quantitative research design was adopted since it aims at determining the variables that are effective in making successful or unsuccessful predictions according to the PISA Reading Comprehension scores of countries with different success levels, with scales measuring students' affective characteristics and questionnaires measuring socio-demographic characteristics (Büyüköztürk et al., 2018).

The PISA 2018 application was carried out with the participation of 61,042 students from 79 countries, 37 of which are OECD members. The selection process of schools and students to participate in PISA was determined by the OECD using a two-stage stratified random method (MEB, 2019). The population of the study consists of all 15-year-old students in the countries participating in the 2018 PISA. According to the objectives of this research, nine countries from all the countries participating in the PISA 2018 exam were selected as the study sample. These nine countries have been determined by taking into account the percentile rankings according to the achievement scores within 79 countries. The percentiles of countries selected from the last third (68-100%) in succession are 91 for Indonesia, 82 for Saudi Arabia, and 72 for Brazil. The samples of Indonesia, Saudi Arabia, and Brazil

were named as sub-achieving countries, representing the most unsuccessful countries in the last 33%. The percentage order of the countries selected from the 33% and 67% success rates is 57 for Serbia, 50 for Turkey, and 41 for Hungary. The samples of Serbia, Turkey, and Hungary were named as medium-success group countries, representing moderately successful countries in the range of 33% to 67%.

Finally, the percentiles of the countries selected from the first third of success, respectively, are 27 for Slovenia, 16 for the USA, and 9 for Finland. In the sample, Slovenia, the USA, and Finland are named as top achievement countries, representing the most prosperous countries in the top 33%. All of these nine countries are: It was called the PISA application sample to represent 79 countries where PISA was applied. In this case, 28706 people from lower-income countries, 18518 from medium-achievement countries, and 16727 individuals from high-achievement countries were included in the study. The sample size of the nine countries included in the study reaches 63951 people, in which there are three countries in each group as low, medium, and high achievement groups, and by combining the countries.

The data for the study came from the PISA reading comprehension test and a student questionnaire administered in 2018. To obtain the dependent variable that we will use, first of all, the reading comprehension success score was obtained as a continuous quantitative variable and named PVREAD. Here, PVREAD refers to the reasonable reading comprehension achievement score. The PVREAD variable is the average of ten reasonable point values (PV1, PV2,... PV10) obtained by each student from the PISA 2018 reading comprehension achievement test and expresses the range of abilities students can have (Wu & Adams, 2002). Then, in the process of classifying the PVREAD score, PISA 2018 Reading Comprehension was used as the OECD average of 487 cutoff score. The PVREAD value was calculated as low between 0–486,999 and high between 487–1000 (0=Failed, 1=Success).

In the analysis of the data, first of all, 79 countries participating in PISA were put in order of success percentage. Then, with equal slices, countries were divided into lower, middle, and upper segments. A total of nine countries, three in each of the lower, middle, and upper groups, were selected according to their percentiles and lack of missing data. The SPSS 21 and Excel programs were used in the analysis of the data. Within the scope of the study, 34 variables given in Table 2 were taken as the independent variables and the two-level success variable (0 = unsuccessful, 1 = successful) as the dependent variable. The logistic regression model was used to determine the independent variables that are effective in predicting the PISA 2018 reading comprehension success by using the sample data of the countries in the lower, middle, and upper achievement groups, each of which includes three countries, and to determine the performance of these variables in classifying the dependent variable.

Findings, Conclusions, and Discussion

In summary, the findings show that the variables of the PISA test's perception of difficulty, family assets, educational items at home, reading pleasure, and socio-economic status have a significant effect on predicting PISA reading comprehension achievement in all groups. Therefore, it can be thought that students who have a perception of difficulty in the PISA test,

have a large family, have insufficient educational equipment at home, do not enjoy reading, and have a poor socio-economic status will fail in terms of PISA reading comprehension success. To prevent this problem, it is possible to eliminate the negativities caused by the differences in these variables by intervening with the appropriate educational policies.

In the study, it was seen that the classification performance of the logistic regression model used in all the groups was at a good level, and when examined in detail, it was determined that the classification performance of the model differed according to the groups. The order of the classification performance of the groups from the highest to the lowest; low-achievement group countries, medium and high-achievement countries. The most important finding is that the model performance is not significantly affected by the decrease in the number of variables. Based on this result, it can be said that it is possible to predict students' PISA reading comprehension success with fewer variables. In the study conducted with the logistic regression model for the sample of low, medium, and high-achievement countries, the success of the students in PISA reading comprehension was well predicted with fewer variables, and the decision-makers in these countries gave more importance to the factors that positively affect PISA reading comprehension success while guiding their education policies. It can be increased.

The following recommendations can be made for future research:

The PISA reading comprehension achievement used in the research can be divided into three or more categories, and it can be examined whether the variables affecting the PISA reading comprehension achievement differ from the effective variables in the case where there are two categories.

In addition, it can be examined whether the classification performance of the model used for PISA reading comprehension achievement will differ according to the classification performance in the case where PISA reading comprehension achievement is divided into three or more categories and has two types.

By adding different variables in addition to the independent variables of the research, important variables that are effective in predicting success can be determined.

Author Contribution Rates

All authors equally took part in all processes of the article.

Ethical Declaration

The purposes and procedure of the current study were granted approval from the ethical committee of the Hacettepe University (Ethics Committee's Decision Date: 30.06.2020), Ethics Committee Approval Issue Numbers: 51944218-300/00001133396).

Conflict Statement

The authors declare that there is no conflict of interest with any institution or person within the scope of the study.