
GAZİ ÜNİVERSİTESİ
Gazi
1926

MADDE TEPKİ KURAMI KAPSAMINDA SINIFLAMA DOĞRULUĞU İNDEKSLERİNİN KARŞILAŞTIRILMASI

Nurşah Yakut

YÜKSEK LİSANS TEZİ

EĞİTİMDE ÖLÇME VE DEĞERLENDİRME ANABİLİM DALI

GAZİ ÜNİVERSİTESİ

EĞİTİM BİLİMLERİ ENSTİTÜSÜ

AĞUSTOS, 2022

TELİF HAKKI VE TEZ FOTOKOPİ İZİN FORMU

Bu tezin tüm hakları saklıdır. Kaynak göstermek koşuluyla tezin teslim tarihinden itibaren 24 (yirmi dört) ay sonra tezden fotokopi çekilebilir.

YAZARIN

Adı : Nurşah

Soyadı : YAKUT

Bölümü : Eğitim Bilimleri

İmza :

Teslim Tarihi : …/…/…

TEZİN

Türkçe Adı : Madde Tepki Kuramı Kapsamında Sınıflama Doğruluğu İndekslerinin Karşılaştırılması

İngilizce Adı : Comparison of Classification Accuracy Indices Under the Item Response Theory

ETİK İLKELERE UYGUNLUK BEYANI

Tez yazma sürecinde bilimsel ve etik ilkelere uyduğumu, yararlandığım tüm kaynakları kaynak gösterme ilkelerine uygun olarak kaynakçada belirttiğimi ve bu bölümler dışındaki tüm ifadelerin şahsıma ait olduğunu beyan ederim.

Yazar Adı Soyadı: Nurşah YAKUT

İmza:

JÜRİ ONAY SAYFASI

Nurşah YAKUT tarafından hazırlanan "Madde Tepki Kuramı Kapsamında Sınıflama Doğruluğu İndekslerinin Karşılaştırılması" adlı tez çalışması aşağıdaki jüri tarafından oy birliği ile Gazi Üniversitesi Eğitim Bilimleri Anabilim Dalı Eğitimde Ölçme ve Değerlendirme Bilim Dalı'nda Yüksek Lisans tezi olarak kabul edilmiştir.

Danışman: Doç. Dr. Emine ÖNEN

Eğitimde Ölçme ve Değerlendirme, Gazi Üniversitesi ……

Başkan: Doç. Dr. Emine ÖNEN

Eğitimde Ölçme ve Değerlendirme, Gazi Üniversitesi ……

Üye: Doç. Dr. Hamide Deniz GÜLLEROĞLU

Eğitimde Ölçme ve Değerlendirme, Ankara Üniversitesi ……

Üye: Prof. Dr. Dilara BAKAN KALAYCIOĞLU

Eğitimde Ölçme ve Değerlendirme, Gazi Üniversitesi ……

Tez Savunma Tarihi: 29/07/2022

Bu tezin Eğitimde Ölçme ve Değerlendirme Anabilim Dalı'nda Yüksek Lisans tezi olması için şartları yerine getirdiğini onaylıyorum.

Prof. Dr. Şaban Çetin

Eğitim Bilimleri Ensitüsü Müdürü ……

Canım Anneme

TEŞEKKÜR

Yüksek lisans eğitimim boyunca çalışmamın her aşamasında bana yardımcı olan, değerli vaktini ve bilimsel desteğini sunan, bilgi ve tecrübesi ile bana yol gösteren çok değerli danışman hocam Doç. Dr. Emine ÖNEN'e sonsuz saygı ve şükranlarımı sunarım.

Tez savunma jürimde yer alan ve süreç boyunca akademik gelişimimi destekleyen çok kıymetli hocam Prof. Dr. Dilara BAKAN KALAYCIOĞLU'na teşekkürlerimi sunarım. Tez savunma jürimde yer alarak görüş ve önerileri ile önemli katkılar sunan değerli hocam Doç. Dr. Hamide Deniz GÜLLEROĞLU'na çok teşekkür ederim.

Tüm yaşamım boyunca bana güvenen ve destek olan, varlıklarını hissederek kendimi güvende hissettiğim, sahip olduğum için şanslı olduğumu düşündüğüm sonsuz güç kaynağım değerli annem Emine YAKUT'a, babam Adem YAKUT'a ve kardeşim Sümeyye YAKUT'a,

Eğitim hayatım boyunca bana destek veren adını sayamadığım bütün arkadaşlarıma ve kıymetli hocalarıma sonsuz teşekkürlerimi sunarım.

Nurşah YAKUT

MADDE TEPKİ KURAMI KAPSAMINDA SINIFLAMA DOĞRULUĞU İNDEKSLERİNİN KARŞILAŞTIRILMASI

(Yüksek Lisans Tezi)

Nurşah Yakut

GAZİ ÜNİVERSİTESİ

EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Ağustos, 2022

ÖZ

Eğitim ortamlarında uygulanan ölçme araçlarından elde edilen ölçümlere dayalı olarak bireyler hakkında tanılama ve yerleştirmeye yönelik kararlar verilmekte ve sınıflama doğruluğu, bu kararların doğruluğuna işaret etmektedir. Eğitsel ölçümlere dayalı bu kararlar özellikle öğrencilerin yaşamlarına yön verici nitelikte olduğundan, bu ölçümler için sınıflama doğruluğunun incelenmesi önem taşımaktadır. Bu araştırmada Madde Tepki Kuramına dayalı yöntemlerden Rudner'ın, Guo'nun ve Lee'nin yöntemleri kullanılarak hesaplanan sınıflama doğruluğu ve tutarlılığı indekslerinin doğruluk derecelerinin- 1PLM, 2PLM ve 3PLM altında- farklı test uzunluğu ve örneklem büyüklükleri açısından karşılaştırmalı olarak incelenmesi amaçlandığından bu çalışma simülasyon çalışması olarak tasarlanmıştır. Simülasyon çalışması yoluyla test uzunluğu ve örneklem büyüklüğü değişimlenerek, hangi koşul(lar) altında hangi yönteme dayalı olarak daha az yanlılıkla bu indekslerin kestirildiğini değerlendirmek mümkün olmaktadır. Bu çalışmada -1PLM, 2PLM ve 3PLM altında- iki farklı test uzunluğu (k=15, k=30) ve örneklem büyüklüğü (n=500, n=1000) koşulları altında R Studio programında 'irtoys' paketi kullanılarak veri setleri üretilmiştir. Sınıflama doğruluğu ve sınıflama tutarlılığı indeksleri ile bunlara ilişkin yanlılık değerleri 'cacIRT' paketi kullanılarak hesaplanmıştır. Sınıflama doğruluğu ve tutarlılığı indekslerine ilişkin değerlerin, madde sayısının artışına bağlı olarak dikkate değer bir farklılık gösterdiği gözlenmiştir. Lee'nin yöntemine dayalı hesaplanan kappa değerleri ile Guo'nun yöntemine dayalı hesaplanan FP ve FN değerlerinin daha yüksek yanlılık

değerlerine sahip olduğu gözlenmiştir. Rudner'ın yöntemine dayalı hesaplanan sınıflama doğruluğu ve tutarlılığı indekslerinin ise diğer yöntemlere kıyasla, daha düşük mutlak yanlılık değerlerine sahip olduğu görülmektedir. Rudner'ın yönteminin büyük örneklem gruplarında uygulanmasının sınıflama kararları açısından daha uygun olacağı düşünülmektedir. Araştırma sonucunda, Türkiye'de uygulanan merkezi sınavlarda sınıflama doğruluğu ve sınıflama tutarlılığının incelenmesinde büyük örneklem gruplarında uygulandığında daha iyi performans göstermesi ve diğer yöntemlere kıyasla daha düşük mutlak yanlılık değerlerine sahip olması açısından Rudner'ın yönteminin kullanılması önerilmektedir.

Anahtar Kelimeler : Madde Tepki Kuramı, Sınıflama Doğruluğu, Sınıflama İndeksleri
Sayfa Adedi : 132
Danışman : Doç. Dr. Emine ÖNEN

COMPARISON OF CLASSIFICATION ACCURACY INDICES UNDER THE ITEM RESPONSE THEORY

(M. S. Thesis)

Nurşah Yakut

GAZI UNIVERSITY

GRADUATE SCHOOL OF EDUCATIONAL SCIENCES

August, 2022

ABSTRACT

Individual diagnostic and placement decisions are made using data from educational measurement instruments, and classification accuracy indicates how accurate these decisions are. It is critical to investigate the classification accuracy for educational measures since decisions based on these measurements have a significant impact on students' lives. This is a simulation study since the aim is to compare the classification accuracy and consistency indices calculated using Rudner's, Guo's, and Lee's item response theory-based approaches under 1PLM, 2PLM, and 3PLM in terms of different test lengths and sample sizes. It is possible to examine under which condition(s) and which method these indices are calculated with less bias by changing the test length and sample size through the simulation study. Under the conditions of two different test lengths (k=15, k=30) and sample sizes (n=500, n=1000) under 1PLM, 2PLM, and 3PLM, data sets were generated using the 'irtoys' package in R Studio. Classification accuracy and consistency indices and related bias values were calculated using the 'cacIRT' package. As the number of items increased, the classification accuracy and consistency indices showed a significant difference. For Kappa values calculated using Lee's approach and FP and FN values calculated using Guo's method, higher bias values were observed. It can be seen that Rudner's method-based classification accuracy and consistency indices have lower absolute bias values than other methods. In terms of classification decisions, it is considered that Rudner's method would work better when applied to large sample sizes. Since Rudner's method performs better when applied to large sample groups and has lower absolute bias values than other methods, it is recommended to use it for examining the classification accuracy and consistency in the national exams used in Turkey.

Key Words : Item Response Theory, Classification Accuracy, Classification Indices
Page Number : 132
Supervisor : Associate Professor Emine ÖNEN

İÇİNDEKİLER

TELİF HAKKI VE TEZ FOTOKOPİ İZİN FORMU…….i

ETİK İLKELERE UYGUNLUK BEYANI……..ii

JÜRİ ONAY SAYFASI…….iii

ÖZ……vi

ABSTRACT……..viii

İÇİNDEKİLER……x

TABLOLAR LİSTESİ…….xiii

ŞEKİLLER LİSTESİ……...xiv

SİMGELER VE KISALTMALAR LİSTESİ……xv

BÖLÜM I…….1

GİRİŞ……1

Problem Durumu…….....1

Sınıflama Doğruluğu ve Sınıflama Tutarlılığı Kavramları…….....3

Klasik Test Kuramına Dayalı Yöntemler…….....6

Madde Tepki Kuramına Dayalı Yöntemler……...8

Genel Sınıflama Doğruluğu ve Sınıflama Tutarlılığı İndeksleri……14

Sınıflama Analizleri……17

İlgili Araştırmalar……...20

Araştırmanın Amacı……...26

Araştırmanın Önemi……..28

Sınırlılıklar…….....28

BÖLÜM II……..29

YÖNTEM……...29

Araştırmanın Modeli…….....29

Verilerin Simülasyonu……29

Ölçüm Güvenirliği……..31

Ölçüm Yorumlarının ve Kullanımlarının Geçerliği……32

Verilerin Analizi…….33

Varsayımların Test Edilmesi…….33

***Tek Boyutluluk*…….33**

***Yerel Bağımsızlık*…….35**

***MKE'nin Monotonik Artış Göstermesi*……...35**

Model Veri Uyumu…….36

BÖLÜM III……40

BULGULAR, SONUÇ VE TARTIŞMA……40

Bulgular ve Yorumlar……40

Sonuç ve Tartışma…….....61

Öneriler……63

KAYNAKLAR……...65

EKLER……...75

EK 1. Bilinen Madde Parametreleri…….76

EK 2. Üretilen Veri Setlerinin 1PLM, 2PLM ve 3PLM İçin Yerel Bağımsızlık Testi Sonuçları……77

EK 3. Madde Karakteristik Eğrileri……...103

EK 4. Üretilen Veri Setlerinin 1PLM, 2PLM ve 3PLM Altında Test Edilmesine İlişkin Madde Uyum İndeksleri……..109

EK 5. Farklı Simülasyon Koşulunda Üretilen Her Bir Veri Setinin Farklı Kesme Puanları ve Yöntemler İçin Üretildiği Model Altında Test Edilmesi Durumunda Sınıflama İndekslerinin Değerleri…….121

EK 6. Farklı Simülasyon Koşulunda Üretilen Her Bir Veri Setinin Farklı Kesme Puanları ve Yöntemler İçin Yanlış Model Altında Test Edilmesi Durumunda Sınıflama İndekslerinin Değerleri…….123

EK 7. Farklı Simülasyon Koşulunda Üretilen Her Bir Veri Setinin Farklı Kesme Puanları ve Yöntemler İçin Üretildiği Model Altında Test Edilmesi Durumunda Sınıflama İndekslerinin Mutlak Yanlılık Değerleri……127

EK 8. Farklı Simülasyon Koşulunda Üretilen Her Bir Veri Setinin Farklı Kesme Puanları ve Yöntemler İçin Yanlış Model Altında Test Edilmesi Durumunda Sınıflama İndekslerinin Mutlak Yanlılık Değerleri……129

TABLOLAR LİSTESİ

Tablo 1. *Sınıflama Tutarlılığı Karışıklık Matrisi*……...4
Tablo 2. *Sınıflama Doğruluğu Karışıklık Matrisi*……..5
Tablo 3. *Farklı Örneklem Büyüklüğü ve Madde Sayısı İçin Farklı Modeller Altında Üretilen Veri Setleri*…… 29
Tablo 4. *Marjinal Güvenirlik Katsayıları*……31
Tablo 5. *Üretilen Veri Setlerine İlişkin AFA Sonuçları*……...33
Tablo 6. *Üretilen Veri Setlerinin Farklı Modeller Altında Test Edilmesine İlişkin Elde Edilen Uyum İstatistikleri ve Hata Değerleri*……...36

ŞEKİLLER LİSTESİ

Şekil 1. Belirli bir theta kestiriminde koşullu sınıflama doğruluğu, FP ve FN oranlarını temsil eden normal dağılım eğrisinin altında kalan alan……9

SİMGELER VE KISALTMALAR LİSTESİ

CA	Classification Accuracy
CC	Classification Consistency
KTK	Klasik Test Kuramı
MTK	Madde Tepki Kuramı
FP	False Positive
FN	False Negative
TKE	Test Karakteristik Eğrisi
ML	Maximum Likelihood
1PLM	1 Parametreli Lojistik Model
2PLM	2 Parametreli Lojistik Model
3PLM	3 Parametreli Lojistik Model
AFA	Açımlayıcı Faktör Analizi
PISA	Programme for International Student Assessment
YGS	Yükseköğretime Geçiş Sınavı
ÖSYM	Ölçme, Seçme ve Yerleştirme Merkezi
TIMSS	Trends in International Mathematics and Science Study
RMSEA	Root Mean Square Error of Approximation
SRMR	Standardized Root Mean Square Residual
CFI	Comparative Fit Index
TLI	Tucker Lewis Index

BÖLÜM I

GİRİŞ

Bu bölümde problem durumu, problem cümlesi, araştırmanın amacı, önemi ve sınırlılıklar yer almaktadır.

Problem Durumu

Eğitim; "bireyin davranışlarında kendi yaşantıları yoluyla kasıtlı ve istendik davranış değişikliği meydana getirme süreci" olarak tanımlanmaktadır (Demirel, 1999). Tanımına bakıldığında, eğitimin, girdi, süreç, çıktı ve kontrol öğelerine sahip bir süreç olduğu görülmektedir. Eğitim sürecinde, her öğrencinin istenilen davranış değişikliğini ne ölçüde gerçekleştirdiği ve öğrencinin istenilen beceri ve yetenekleri ne ölçüde kazandığı belirlenmeye çalışılır (Atılgan, Kan ve Doğan, 2017). Burada ölçülmeye veya gözlenmeye çalışılan değişkenler genellikle; başarı, ilgi, motivasyon, yetenek gibi psikolojik değişkenlerdir. Söz konusu değişkenlerin birçoğunun fiziksel özellikleri bilinmez ve bu sebeple fiziksel boyutları tanımlanamaz. Bu süreç sonucunda istenilen davranışların ortaya çıkıp çıkmadığını veya ne ölçüde gerçekleştiğini ortaya çıkarma, öğrencileri yönlendirme ve benzeri amaçlarla yapılan tüm değerlendirmeler geçerli ve güvenilir ölçüm sonuçlarına dayandırılmalıdır (Atılgan, Kan ve Doğan, 2017).

Eğitimde uygulanan birçok testin nihai amacı sınava giren bireyleri performanslarına ve önceden belirlenmiş kriter ve standartlara göre uygun gruba yerleştirmektedir. Eğitim sürecinde bireyler hakkında alınan kararların çoğu geçti/kaldı, başarılı/başarısız şeklinde değerlendirilir. Aslında bu değerlendirmeler, ölçme sonuçlarının sınıflama düzeyinde ifade edilmesidir. Ölçme sonuçları hangi ölçekte olursa olsun, bir ölçüte dayanarak elde edilen

ölçümler sınıflama düzeyinde ifade edilebilir. Bu aşama, değerlendirme aşamasını oluşturur ve asıl amaç da bu nihai karara ulaşabilmektir.

Öğrenciler hakkında önemli kararların alındığı merkezi sınavlar, sınava girenler üzerinde önemli etkilere sahiptir. Ölçütün üstünde yer alan bir birey için kaldı/başarısız kararı vermek, o bireyin hak ettiği fırsatlardan mahrum kalması anlamına gelirken, aslında kalacak bir birey için ise geçti/başarılı kararı vermek de potansiyel olarak hem bireyler hem de toplum açısından istenmeyen sonuçlara yol açabilir. Örneğin, akademik bir program için yeterli donanıma sahip olmayan bir öğrenci, hatalı olarak sınıflandırıldığı durumda, diğer öğrencilerin gerisinde kalabilir ve bundan olumsuz etkilenebilir. Benzer şekilde, alanında uzman ve yetkin olmayan bir hekime, yanlış bir kararla uzmanlık yetkisi verilirse, hastaların güvenliği ve sağlık hizmetlerinin kalitesi riske atılmış olur. Bu nedenle, bireylerin potansiyellerinin en uygun şekilde kullanılabilmesi ve karar verme sürecinin etkili bir şekilde yönetilmesi hem bireyler hem de toplum için hayati önem taşımaktadır (Cizek ve Bunch, 2007).

Eğitim sisteminde bireyler hakkında önemli kararların alındığı sınavlar, yüksek risk içeren sınavlar olarak tanımlanmaktadır (Kumandaş ve Kutlu, 2010; Kumandaş, 2013). Yüksek risk içeren sınavların sonuçlarına göre öğrencilerin bir üst eğitim kademesine yerleştirilmesi ve sonrasında gidebileceği eğitim kademelerini belirlemesi açısından bu sınavların sonuçlarına göre alınan kararlar bireyler üzerinde önemli etkilere sahiptir. Türkiye'de özellikle ortaöğretim kurumları arasında gözlemlenen fiziksel, sosyal ve kültürel açılardan niteliksel farkların ve kurumlar arası başarı farkının fazla olması, liseye geçişte uygulanan sınavların önemini arttırmaktadır. Benzer şekilde ortaöğretimde gidilen eğitim kurumunun yükseköğretimde iyi bir eğitim almanın belirleyicisi olarak görülmesi de ortaöğretime seçme amacıyla uygulanan sınavların önemini ortaya çıkarmaktadır (Özkan ve Turan, 2021). Bu sınavların sonuçlarına dayalı olarak bireyler hakkında alınan kararlar, bireylerin yaşamını doğrudan etkilemekte olup bu kararların doğruluğunun incelenmesi bireyler ve toplum açısından büyük bir önem taşımaktadır.

Bu doğrultuda, Hambleton ve Novick (1973), sınıflama kararlarını değerlendirmede kullanılabilecek iki indeks önermiştir. Bunlardan ilki olan sınıflama tutarlılığı (classification consistency; CC), bireylerin tekrarlı ya da paralel ölçümlerde de aynı kategoride sınıflandırılma derecesini ifade etmektedir. Sınıflama tutarlılığı kavramı, tutarlılık tanımı tekrarlı ölçümler gerektirdiği için sınıflama güvenirliği olarak ele alınmaktadır. Sınıflama

doğruluğu (classification accuracy; CA) ise tek bir uygulama üzerinde, gözlenen sınıflama ile gerçek sınıflamanın uyum derecesini ifade eder (Lee vd., 2000). Sınıflama doğruluğu bireyin gözlenen puanına dayalı olarak yapılan sınıflamanın, gerçek puanına dayalı yapılan sınıflamaya göre doğruluğunun derecesini ifade ettiği için sınıflama geçerliği ile ilgili bir kavramdır.

Tekrarlı ölçümlerden elde edilen veriler mevcutsa, sınıflama tutarlılığının hesaplanması basittir. Ancak, gerçek testlerde genellikle testin yalnızca tek bir formu uygulanmaktadır. Ayrıca, sınıflama doğruluğunun hesaplanması sürecinde bireylerin gerçek puanları bilinmemekte ve bunların kestirilmesi gerekmektedir. Bu nedenlerle, tek uygulamaya dayalı sınıflama doğruluğu ve tutarlılığı hesaplama yöntemleri geliştirilmiş ve önerilmiştir.

Literatürde, Klasik Test Kuramı (KTK) kapsamında geliştirilen çok sayıda yöntem önerilmiştir (Cohen, 1960; Hanson ve Brennon, 1990; Huynh, 1976; Lee, Brennon ve Wan, 2009; Livingston ve Lewis, 1995; Subkoviak, 1976). Bu yöntemler, gözlenen puanların rapor edildiği ve kesme puanlarının ham puan ölçeğinde olduğunu varsaymaktadır.

Son yıllarda, Madde Tepki Kuramı (MTK) kapsamında da sınıflama doğruluğu ve tutarlılığının hesaplanmasına yönelik yöntemler geliştirilmiştir (Guo, 2006; Lee, 2010; Rudner, 2001, 2005). Bu yöntemlerin tümü MTK modellerine dayalı olsa da temelinde yer alan dağılım varsayımları ve sınıflama kararlarının ham puan ölçeğine mi yoksa gizil yetenek ölçeğine mi dayalı olduğu açısından farklılık göstermektedir.

Sınıflama Doğruluğu ve Sınıflama Tutarlılığı Kavramları

Başarı testlerinden elde edilen test puanları, yorumlama amacıyla iki farklı türde bilgi sağlar (Glaser, 1963). Test puanlarının yorumu, kullanım amacına bağlı olarak, norm referanslı ya da kriter referanslı olarak tanımlanabilir. Norm referanslı testlerden elde edilen bilgiler, bireyin önceden belirlenmiş bir popülasyondaki konumunu belirlemeye yardımcı olur. Norm referanslı testlere örnek olarak Eğitim Yetenek Testi (Scholastic Aptitude Test; SAT) verilebilir. Kriter referanslı testlerden elde edilen bilgiler ise bireyin performansının önceden belirlenmiş bazı standartları karşılayıp karşılamadığını gösterir. Hem norm referanslı hem de kriter referanslı testlerde, sınava giren bireylerin tekrar ve tekrar test edilmesi durumunda, bireylerin gerçek yeteneklerinde bir değişiklik olmadığı varsayılarak, tutarlılık ve tekrarlanabilirlik test puanlarının istenilen bir özelliğidir. Paralel (eşdeğer) test güvenirliği, eşdeğer yarılar güvenirliği ve Cronbach alfa iç tutarlılık gibi geleneksel güvenirlik ölçütleri,

bireylerin test puanlarının istikrarı, tutarlılığı ve birbirleriyle göreli sıralaması hakkında bilgi verir. Ayrıca, kriter referanslı testlerde, gerçek test puanlarının kendilerine değil sınıflama kararlarının yorumlarına odaklanılır. Bu geleneksel ölçütler, bireyin paralel bir test formunda tutarlı bir şekilde sınıflandırılıp sınıflandırılmayacağı ya da bireyin kendi gerçek kategorisinde doğru bir şekilde sınıflandırılıp sınıflandırılmayacağı hakkında bilgi sağlamada yetersiz olmaktadır.

Sınıflama doğruluğu ve sınıflama tutarlılığı kavramları ilk olarak Hambleton ve Novick (1973) tarafından, bireylerin kesme puanları gibi önceden tanımlanmış bazı standartlarla karşılaştırıldığı, kriter referanslı testlerin güvenirliği ve geçerliği için alternatif ölçütler olarak önerilmiştir. Aynı veya paralel formu ile tekrarlanan testlerde, bireylerin tutarlı bir şekilde aynı kategoride sınıflandırılma derecesi olarak tanımlanan sınıflama tutarlılığı ile güvenirliği değerlendirmenin daha uygun olacağını öne sürmüşlerdir. Geçerlik ile ilgili ise sınıflama doğruluğunun, sınıflama kararlarının doğruluk derecesine ilişkin bilgi sağladığını ifade etmişlerdir. Bir başka deyişle, sınıflama doğruluğu, bireylerin gerçek sınıflamalarının bilindiği varsayılarak, bireylerin gözlenen test verilerine dayalı olarak yapılan sınıflamalarıyla, gerçek sınıflamalarının ne derece uyuştuğunu ölçer. Sınıflama tutarlılığı kavramı, K'nın belirli bir testteki sınıflamaların sayısını temsil ettiği durumda, KxK karışıklık matrisi olarak kavramsallaştırılabilir (Tablo 1). Bu durumda, testin paralel bir formunun uygulandığı ya da testin tekrar uygulandığı varsayılmaktadır. Sınıflama tutarlılığını hesaplamada, en yaygın olarak kullanılan indeks, genel sınıflama tutarlılığı oranıdır (*p*).

Tablo 1.

Sınıflama Tutarlılığı Karışıklık Matrisi

		2. uygulamaya dayalı sınıflama kararları		
		Geçti	Kaldı	Marjinal Oran
1. uygulamaya dayalı sınıflama kararları	Geçti	Tutarlı Sınıflama	Tutarsız Sınıflama	$p_{.1}$
	Kaldı	Tutarsız Sınıflama	Tutarlı Sınıflama	$p_{.2}$
	Marjinal Oran	$P_{1.}$	$P_{2.}$	

Matriste yer alan her hücre, testin her iki uygulamasında da tutarlı bir şekilde aynı kategoride sınıflandırılan bireylerin oranını temsil eder ve *p*, her iki uygulamada da tüm kategorilerde tutarlı bir şekilde sınıflandırılan bireylerin genel oranını temsil etmektedir. Sınıflama

tutarlılığını ölçmek için yaygın olarak kullanılan bir başka indeks, Cohen'in kappa katsayısıdır (Cohen, 1960). κ ile gösterilen katsayı şu şekilde ifade edilir:

$$\kappa = \frac{p - p_c}{1 - p_c} \quad \text{(Eşitlik 1)}$$

Burada p_c iki test formu arasındaki şans eseri elde edilen tutarlılık derecesidir. İki test formunun birbirinden bağımsız olduğunun varsayıldığı durumda, p_c şu şekilde ifade edilir:

$$p_c = \sum_{k=1}^{K} (p_{.k})(p_{k.}) \quad \text{(Eşitlik 2)}$$

Burada $p_{\cdot k}$ ve $p_{k\cdot}$, iki test formunda da k. kategoride sınıflandırılan bireylerin marjinal oranlarıdır. Sınıflama doğruluğu kavramı da benzer şekilde KxK karışıklık matrisi olarak sunulabilir (Tablo 2).

Tablo 2.

Sınıflama Doğruluğu Karışıklık Matrisi

		Sınıflama Kararları	
		Geçti	Kaldı
Gerçek Sınıflamalar	Geçti	Doğru Sınıflama	FN
	Kaldı	FP	Doğru Sınıflama

Burada, aynı testin ya da testin paralel formlarının iki kez uygulanması yerine, test yalnızca bir kez uygulanır ve bireylerin gerçek sınıflamalarının bilindiği varsayılır. Genel sınıflama doğruluğu oranı (γ) doğru sınıflamaların toplamına eşittir. Genel yanlış pozitif (False Positive; FP) oranı, γ^+ olarak gösterilir ve bireylerin gerçek sınıflamasının üzerinde bir kategoride hatalı olarak sınıflandırılma oranıdır. Genel yanlış negatif (False Negative; FN) oranı ise γ^- olarak gösterilir ve bireylerin hatalı bir şekilde daha düşük kategoride sınıflandırılma oranını ifade etmektedir.

Swaminathan, Hambleton ve Algina (1974), ilk kez iki uygulamadan elde edilen örnek test verileriyle sınıflama tutarlılığı indekslerinin hesaplanmasını göstermiştir. Ancak, aynı sınava giren bireylerin iki kez test edilmesi ve iki uygulamadan veri toplanması gerçek hayatta kullanışlı ve pratik değildir. Ayrıca, bireylerin gerçek puanları ya da gerçek sınıflamaları bilinmemekte ve bunların kestirilmesi gerekmektedir. Tek bir uygulamadan elde edilen verileri kullanarak sınıflama indekslerini hesaplamak için bazı yöntemlere sahip olmak pratik açıdan önemli ve gereklidir.

Klasik Test Kuramına Dayalı Yöntemler

Huynh'un Yöntemi

Huynh (1976), yalnızca ikili puanlanan maddelerden oluşan testlerin tekli uygulamalarından elde edilen test puanlarına dayalı sınıflama tutarlılığı indekslerinin hesaplanmasına yönelik bir yöntem önermiştir. Bu yaklaşımda, bireylerin gerçek puanlarının iki parametreli bir beta dağılımını, gözlenen puanların da bir beta binom dağılımını takip ettiği varsayılır. Binom modelinin kullanımı, testlerdeki tüm maddelerin, bilinen maddeler evreninde mükemmel şekilde yer değiştirebilir olduğunu varsayar. Bir başka deyişle, maddelerin zorluk açısından benzer olduğunu ve bir dereceye kadar içerik homojenliğini ifade eder. Ayrıca, belirtilen aynı evrenden alınan aynı sayıda maddeden oluşan paralel bir test formunun da aynı bireylere uygulandığını varsayar.

Farklı örneklerden alınan test puanlarının yerel olarak bağımsız olduğu varsayımı altında, iki paralel test formundan (x ve y) alınan test puanları, ortak yoğunluk fonksiyonu ile iki değişkenli iki parametreli bir beta binom dağılımını takip eder:

$$f(x, y) = \frac{\binom{J}{x}\binom{J}{y}}{\mathrm{B}(\alpha,\beta)} \mathrm{B}(\alpha + x + y, 2j + \beta - x - y) \qquad \text{(Eşitlik 3)}$$

Burada α ve β, beta dağılımının parametreleridir. B beta fonksiyonudur, j ikili puanlanan maddelerin toplam sayısıdır. X ve y, iki paralel test formunun uygulamalarından elde edilen toplam puanlardır.

Genel sınıflama tutarlılığı oranı, Eşitlik 3'te verilen aynı kategoriye ait hem x hem de y puanlarına sahip bireylerin olasılığına eşittir.

Hanson ve Brennan (1990), bireylerin gerçek puanları için, gerçek puan dağılımında bir alt ve bir üst sınır içeren dört parametreli beta dağılımını uygulayarak Huynh'un yöntemini genişletmişlerdir. Ayrıca, sınıflama doğruluğu, FP ve FN indekslerinin hesaplanmasındaki yöntemleri de göstermişlerdir. Dört parametreli beta binom modelinin, gözlenen puan dağılımlarına daha iyi model uyumu sağladığı ve sınıflama indekslerinde tercih edilen model olduğu sonucuna ulaşmışlardır. Ek olarak, dört parametreli beta modelini dört parametreli beta bileşik binom modeliyle karşılaştırmışlar, ancak ayırt edilemez sonuçlara ulaşmışlardır.

Subkoviak'ın Yöntemi

Subkoviak (1976) tarafından önerilen yöntem, paralel test formlarından alınan puanların bağımsız olarak dağıldığı ve Huynh'un yönteminde varsayılanlara benzer bir binom dağılımı izlediği varsayımlarına dayanmaktadır. İki yöntem arasındaki en önemli fark, bireylerin gerçek puanlarına ilişkin dağılımın Subkoviak'ın yönteminde varsayılmamasıdır. Bireyin gerçek puanı, maksimum olabilirlik (maximum likelihood; ML) kestirimi veya test puanlarının ortalaması, varyansı ve güvenirliği gibi tamamlayıcı bilgileri içeren bir doğrusal regresyon yöntemi kullanılarak kestirilir. Bireylerin gerçek puanlarına ilişkin hiçbir dağılım varsayılmadığından, genel sınıflama tutarlılığı oranı, her bir bireye dayalı hesaplanan koşullu sınıflama tutarlılığı oranının ortalaması alınarak elde edilir. Binom modeli, maddeler arasındaki doğru yanıt olasılığının sabit kaldığını varsaydığından, basitliği nedeniyle tercih edilmektedir. Ancak Subkoviak (1976), binom modelinin, maddeler arasında doğru yanıt olasılığının değişkenlik göstermesine izin vererek gerçek test durumlarına daha iyi bir yaklaşım sağlayacağına dikkat çekmiştir.

Livingston ve Lewis'in Yöntemi

Huynh ve Subkoviak'ın yaklaşımının sınırlılıklarından biri, her iki yöntemin de yalnızca ikili puanlanan maddelerden oluşan test puanlarına uygulanabilir olmasıdır. Livingston ve Lewis (1995), Huynh'un yöntemini, karma testler için sınıflama indekslerinin hesaplanmasına uygun olacak şekilde genişletti. Ayrıca, etkili test uzunluğu kavramını tanıttılar. Bu kavram, karma testlere ilişkin gözlenen puanlarla aynı kesinlikte toplam puanlar elde etmek amacıyla gerekli olan kategorik, ikili puanlanan, yerel bağımsız ve eşit derecede zor test maddelerinin sayısı olarak tanımlanmaktadır. Örneğin, çoklu puanlanan maddelerden oluşan bir testte her maddenin 0 ile 3 puan arasında olası bir puan aralığı olsun. Bu test, ikili puanlanan 20 maddelik başka bir test ile aynı güvenirliğe sahipse, test için etkili test uzunluğu 20 olacaktır. Etkili test uzunluğu şu şekilde kestirilebilir:

$$j_{LL} = \frac{(\mu_x - X_{min})(X_{max} - \mu_x) - r\sigma_x^2}{\sigma_x^2(1-r)} \qquad \text{(Eşitlik 4)}$$

Burada X_{min} ve X_{max} orijinal puan ölçeğindeki en düşük ve en yüksek puan, μ_x ve σ_x^2 gözlenen puanların ortalaması ve varyansı, r ise güvenirlik katsayısıdır. Etkili test uzunluğu belirlendikten sonra, orijinal gözlenen puanlar 0 ile j_{LL} arasında değişen yeni bir ölçeğe

dönüştürülür ve sınıflama indekslerinin hesaplanmasındaki sürecin geri kalanı Hanson ve Brennon (1990) tarafından açıklanan süreç ile aynıdır.

Madde Tepki Kuramına Dayalı Yöntemler

Bu bölümde, Rudner'ın, Guo'nun ve Lee'nin yöntemleri incelenmiştir. Öncelikle, her bir birey için koşullu sınıflama doğruluğu ve sınıflama tutarlılığı indekslerinin hesaplanmasındaki yöntemler tanıtılmıştır. Genel sınıflama doğruluğu ve sınıflama tutarlılığı indekslerini elde etme yöntemleri sonraki bölümde sunulmuştur.

Rudner'ın Yöntemi

Rudner (2001), ikili puanlanan maddelerden oluşan testler için sınıflama doğruluğu indekslerinin hesaplanmasına yönelik bir yöntem önermiş ve daha sonra bu yöntemi çoklu puanlanan maddelerden oluşan testler için genişletmiştir (Rudner, 2005). Bu yaklaşımda, sınıflama kararları verilirken, bireylerin test puanları ve kullanılan kesme puanlarının, örtük özellik için sürekli theta (θ) ölçeğinde eşleştirildiği ve kestirilen theta puanlarının doğrudan theta ölçeğinde belirlenen kesme puanlarıyla (theta kesme puanı) karşılaştırıldığı varsayılır. Herhangi bir bireyin gerçek theta puanı bilinmediğinden, kestirilen theta puanı ($\hat{\theta}$), onun gerçek puanı (θ) için en iyi puan kestirimi olarak ele alınır. Kestirilen theta puanı ($\hat{\theta}$), yeterince fazla sayıda madde ve örneklem büyüklüğü ile maksimum olabilirlik kestirimi ile elde edildiğinde, θ'nın ortalaması ve aşağıda formülü verilen θ kestiriminin standart hatası (se(θ)) ile asimptotik olarak normal dağılır:

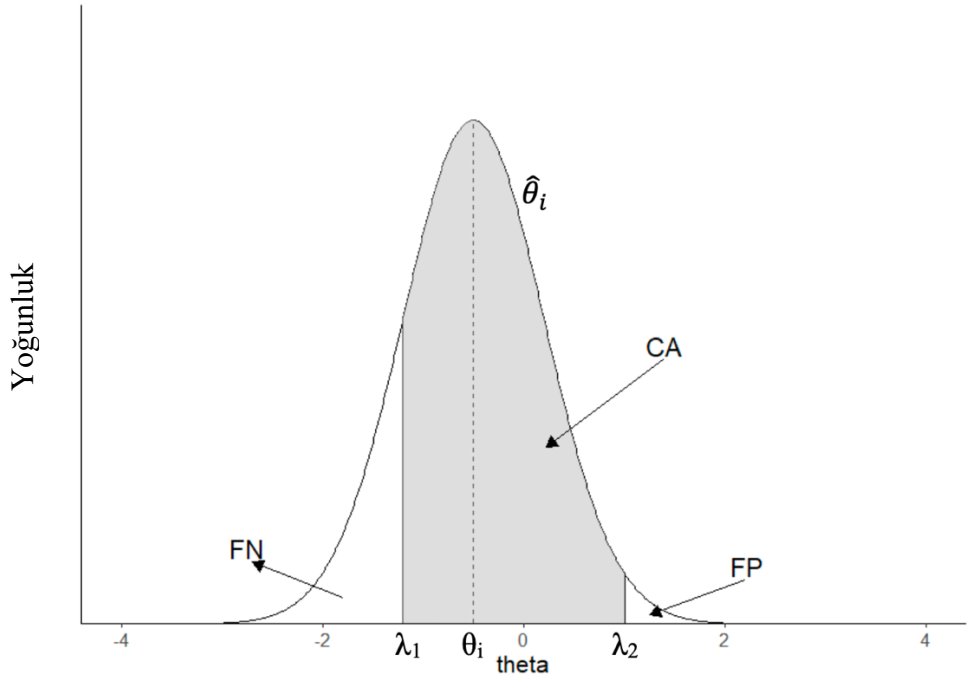
$$se(\theta_i) = \frac{1}{\sqrt{I(\theta_i)}} \qquad \text{(Eşitlik 5)}$$

Burada i, sınava giren i. bireyi gösterir. $I(\theta_i)$, sınava giren i.birey için θi de değerlendirilen test bilgi fonksiyonudur. Test bilgi fonksiyonu, verilen MTK modeline ve madde parametrelerine göre değerlendirilen madde bilgi fonksiyonlarının toplamıdır. Bireyleri K kategoriden herhangi birine sınıflandırırken, K-1 tane theta kesme puanı ($\lambda_1 < \lambda_2 < ... < \lambda_{K-1}$) sürekli theta ölçeğine yerleştirilir. İki ek theta puanı ($\lambda_0 = -\infty$ and $\lambda_K = \infty$) hesaplama amaçları için, olası theta puanlarının teorik olarak alt ve üst sınırlarını temsil etmek üzere dahil edilir.

Eğer $\hat{\theta_i}$, λ_{k-1} 'den büyük ya da eşit ve λ_k 'den küçük ise, birey k.kategoride sınıflandırılır. Bireyin (i) k.kategoride sınıflandırılma olasılığı şu şekilde verilir:

$$\Pr(K = k|\theta_i) = \Phi\left(\frac{\lambda_k - \theta_i}{se(\theta_i)}\right) - \Phi\left(\frac{\lambda_{k-1} - \theta_i}{se(\theta_i)}\right) \quad \text{(Eşitlik 6)}$$

Burada ϕ, ortalaması θ_i ve kestiriminin standart hatası se(θ_i) olan birikimli normal dağılım fonksiyonudur. Eşitlik 6, iki theta kesme puanı arasında ve normal dağılımın, N(θ_i, se(θ_i)), altında kalan alanı temsil eder. Bireyin gerçek sınıflaması bilinmemektedir, ancak gözlenen verilere dayalı yapılan sınıflama kararları, bireyin gerçek sınıflamasının en iyi kestirimi olarak alınabilir. Bireyin kendi gerçek sınıfında sınıflandırılma olasılığı olan koşullu sınıflama doğruluğu (γ_{θ_i}), θ_i ve gerçek kategoriyi tanımlayan theta kesme puanları kullanılarak Eşitlik 6 ile hesaplanabilir. Koşullu sınıflama doğruluğu indekslerinin hesaplanması Şekil 1'de grafiksel olarak gösterilmiştir. Dikey kesikli çizgi, bireylerin θ_i 'sinin konumunu temsil eder ve $\hat{\theta_i}$ 'nin dağılımı, se(θ_i) 'nin kestiriminin standart hatasıyla θ_i civarında asimptotik olarak normaldir. İki dikey düz çizgi, kullanılan iki theta kesme puanını (λ_1 ve λ_2) temsil eder. θ_i, iki theta kesme puanı arasında kaldığından, bireyin gerçek sınıflaması ortada yer alır. Birey, $\hat{\theta_i}$ değeri kullanılan iki theta kesme puanı arasındaysa, orta kategoride doğru bir şekilde sınıflandırılır. Birey için doğru sınıflandırılma olasılığı ya da bireyin koşullu sınıflama doğruluğu, iki theta kesme puanı ile sınırlanan normal dağılım eğrisinin altındaki alana eşittir. Sonrasında, ilgili bölgelerde normal dağılım eğrisinin altındaki alan hesaplanarak koşullu FP ve FN hesaplanabilir.



Şekil 1. Belirli bir theta kestiriminde koşullu CA, FP ve FN oranlarını temsil eden normal dağılım eğrisinin altında kalan alan. Chau, L. H. (2018). Evaluating the correctness of IRT-based methods in computing classification consistency and accuracy indices in the presence of model misspecification (Doctoral dissertation, University of British Columbia) kaynağından alınmıştır.

Orijinal Rudner yöntemi, yalnızca sınıflama doğruluğu indekslerini hesaplamak amacıyla geliştirilmiştir. Sınıflama tutarlılığı indekslerini hesaplama yöntemi, orijinal yöntemin bir uzantısı olarak önerilmiştir (Li, 2006; Wyse & Hao, 2012). K kategorili bir testte, her bir kategoriye θ_i ile sınıflandırılan bireye ilişkin olasılıklar, ilgili theta kesme puanları arasındaki normal dağılım eğrisinin altındaki alana eşit olan Eşitlik 6 ile hesaplanabilir. Bireylere, aynı theta kesme puanlarına ve MTK parametrelerine sahip maddelerden oluşan varsayımsal bir paralel test formu uygulandığında, her bir kategoriye aynı θ_i ile sınıflandırılan bireye ilişkin olasılıklar, Eşitlik 6 kullanılarak aynı şekilde hesaplanabilir.

Bu koşullar altında, varsayımsal paralel test formu için elde edilen olasılıklar, uygulanan gerçek test formundan elde edilen olasılıklar ile aynı olacaktır. İki test formunun bağımsız olduğu varsayıldığında, bireyin iki test formunda da tutarlı bir şekilde sınıflandırılma olasılığı, uygulanan gerçek test formunda her kategoride sınıflandırılma olasılığının karesinin toplamına eşittir.

Koşullu sınıflama tutarlılığı (p_{θ_i}) şu şekilde hesaplanabilir:

$$p\theta_i = \sum_{k=1}^{K}[\Pr(K = k|\theta_i)]^2 \quad \text{(Eşitlik 7)}$$

Burada K, sınıflandırılma kategorilerinin toplam sayısıdır. $Pr(K = k|\theta_i)$ ise k.kategoride verilen θ_i de sınıflandırılma olasılığıdır. Koşullu sınıflama tutarlılığı, tüm kategorilerde sınıflandırılma olasılıklarının karelerinin toplamıdır.

Guo'nun Yöntemi

Rudner'ın yönteminin bir dezavantajı, bireylerin kestirilen thetalarının normallik varsayımına bağlı olması ve theta kesme puanları arasındaki normal dağılım eğrisinin altında kalan alanı hesapladıktan sonra her bir kategorideki sınıflandırılma olasılıklarının hesaplanmasıdır. Pratikte normallik varsayımı yalnızca asimptotik olarak geçerlidir ve hiçbir zaman tam olarak karşılanmaz (Wyse ve Hao, 2012). Guo (2006), kestirilen thetalar için normallik varsayımı gerektirmeyen, ancak verilerin olasılığına dayanan Rudner'ın yönteminin bir uzantısı olarak bir yöntem geliştirmiştir. Guo'nun yöntemi, Rudner'ın yöntemine benzer şekilde hem gözlenen puanların hem de kullanılan kesme puanlarının theta ölçeğinde eşlendiğini varsayar. K kategorili bir test için, K-1 theta kesme puanı ($\lambda_1 < \lambda_2 < ... < \lambda_{K-1}$) ile birlikte olası theta puanlarının alt ve üst sınırlarını temsil eden iki theta puanı ($\lambda_0 = -\infty$ ve $\lambda_K = \infty$), theta ölçeğine yerleştirilir. Eğer $\hat{\theta i}$, $\lambda_k{-}1$ 'den büyük ya da eşit ve λ_k 'den küçükse, birey k.kategoride sınıflandırılır. MTK'nin tek boyutluluk ve yerel bağımsızlık varsayımları altında, maddelere verilen gözlenen yanıtlar ve bilinen madde parametreleri göz önüne alındığında, bireyin herhangi bir θ'ya sahip olma olasılığı şu şekilde verilebilir:

$$L(u_{i1}, u_{i2}, \dots, u_{ij}|\theta) = \prod_{j=1}^{J} P_{ij}^{u_{ij}}(1 - P_{ij})^{u_{ij}} \quad \text{(Eşitlik 8)}$$

j toplam madde sayısını, uij bireyin (i) j maddesine verdiği yanıtı temsil eder. Pij ise bireyin j maddesi için doğru bir yanıt alma olasılığıdır. Bireyin theta kesme puanları $\lambda_k{-}1$ ve λ_k ile sınırlanan k.kategoride sınıflandırılma olasılığı şu şekilde verilir:

$$L_{ki} = \int_{\theta=\lambda_{k-1}}^{\lambda_k} L(u_{i1}, u_{i2}, \dots, u_{ij}|\theta)d\theta \quad \text{(Eşitlik 9)}$$

Eşitlik 9'daki integrasyonu hesaplama açısından mümkün kılmak için, minimum ve maksimum theta kesme puanları bazı keyfi büyük değerler ile değiştirilir. Örneğin, Guo'nun (2006) orijinal makalesinde gösterilen yöntemlerde, minimum ve maksimum theta kesme

puanları -3 ve +3'e eşit olarak ayarlanmıştır. Ek olarak, theta kesme puanları arasında bir dizi eşit aralıklı theta noktası (örn: 100 nokta) oluşturularak integrasyon işlemi toplama işlemi ile değiştirilir. Eşitlik 9, toplam formülü kullanılarak yeniden yazılabilir:

$$L^*_{ki} = \sum_{\Theta=\lambda_{k-1}}^{\lambda_k} L\left(u_{i1}, u_{i2}, \dots, u_{iJ} \middle| \theta\right) \qquad \text{(Eşitlik 10)}$$

Eşitlik 10'da olabilirlik fonksiyonu tarafından verilen değerler, doğrudan olasılık olarak alınamaz. Bir birey için tüm kategorilerdeki olasılıkların toplamı 1'e eşit olacak şekilde bir normalleştirme adımı gereklidir. K. kategoride sınıflandırmanın beklenen olasılığı şu şekilde verilir:

$$Pr(K = k|u_{i1}, u_{i2}, \dots, u_{iJ}) = \frac{L^*_{ki}}{\sum_{k=1}^{K} L^*_{ki}} \qquad \text{(Eşitlik 11)}$$

Burada payda yer alan L^*_{ki}, kesme puanları arasındaki eşit aralıklı theta noktaları ile Eşitlik 10'dan elde edilir. Paydada yer alan ifade ise her bir kategoride sınıflandırma olasılıklarının toplamı olan normalizasyon sabitidir. Bireyin gerçek sınıflamasının en iyi kestirimi olarak kabul edilen bireyin gözlenen puanlarına göre sınıflaması, bireyin $\hat{\theta_i}$ 'si ile kullanılan theta kesme puanları karşılaştırılarak belirlenebilir. Daha sonra, bireyin koşullu sınıflama doğruluğu olan, gerçek sınıflamada yer alma olasılığı Eşitlik 11'den elde edilebilir. Guo'da Rudner'a benzer şekilde, çalışmasında sadece sınıflama doğruluğunu hesaplamaya yönelik bir yöntem önermiştir. Wyse ve Hao (2012), Guo'nun yöntemini genişletti ve sınıflama tutarlılığını hesaplamak için bir yöntem önerdiler. Geliştirdikleri yöntem, Rudner'ın yönteminin uzantısına benzer bir mantıktadır. Aynı theta kesme puanları ve aynı maddelere sahip paralel test formunun uygulandığı varsayılarak, iki test formunda da maddelere aynı yanıtları veren bireyin, paralel ve orijinal test formunda her bir kategoride sınıflandırılma olasılığı aynı olacaktır. Sonrasında, iki test formunun bağımsız olduğu varsayılarak, koşullu sınıflama tutarlılığı (p_i) şu şekilde hesaplanabilir:

$$p_i = \sum_{k=1}^{K} \left[Pr(K = k|u_{i1}, u_{i2}, \dots, u_{iJ})\right]^2 \qquad \text{(Eşitlik 12)}$$

Burada, $Pr(K = k|u_{i1}, u_{i2}, \dots, u_{iJ})$, Eşitlik 10'dan elde edilen k kategorisine sınıflandırılmanın normalleştirilmiş olasılığıdır.

Lee'nin Yöntemi

Lee'nin (2010) yöntemi ya da bazen MTK-özyinelemeli yöntem olarak da adlandırılan yöntem, sınıflama kararlarının ham puan ölçeğinde verildiği Rudner'ın ve Guo'nun yönteminden farklıdır. Başka bir deyişle, kesme puanları ham puan metriğinde kullanılır ve bireylerin ham puanları ile ham kesme puanları karşılaştırılarak sınıflama kararları alınır. Lee'nin yöntemi, ikili puanlanan MTK modellerinin kullanımını inceleyen Huynh (1990) ile Schulz, Kolen ve Nicewander (1999), çoklu puanlanan MTK modellerini dikkate alarak bu yöntemi genişleten Wang, Kolen ve Harris (2000) tarafından daha önce geliştirilen ve önerilen yöntemlerin bir genellemesidir. Bireyleri K kategoriye sınıflandıran bir test olduğu varsayılırsa, ham puan ölçeğinde K-1 kesme puanı ($c_1 < c_2 < \ldots c_{K-1}$), olası minimum ve maksimum ham puanları ($c_0 = 0$ ve $c_K = X_{max}$) temsil eden iki ham puan ile birlikte ham puan ölçeğinde kullanılır. Tüm maddelerin ikili puanlandığı durumda, *Xmax* madde sayısına eşit olacaktır. Bireyin ham puanı $c_k{-}1$ 'den büyük ya da eşit ve c_k 'den küçük ise birey k.kategoride sınıflandırılır veya ham puanı maksimum puana eşit ise otomatik olarak en yüksek kategoride sınıflandırılır. Lee'nin yöntemindeki ilk adım, bireyin $\hat{\theta i}$ ve madde parametrelerine bağlı olası ham puanların her birinin gözlenme olasılığını hesaplamaktır. Herhangi bir olası ham puan (x), $\boldsymbol{u_i}$ = [$u_{i}1$, $u_{i}2$, ... , u_{ij}] vektörü olarak gösterilebilen maddelere verilen yanıt örüntüsünün bir fonksiyonudur. Burada *i* bireyi, *j* maddeyi, *uij* ise birey (i) tarafından *j* maddesine verilen yanıtı temsil eder. Bir test yalnızca ikili puanlanan maddelerden oluşuyorsa, *uij* yalnızca 0 veya 1 değerlerini alır. Her bir x ham puanının gerçekleşmesi $\boldsymbol{u_i}$'deki maddelerin toplanmasına eşittir. Bir x ham puanının gözlenme olasılığı şu şekilde ifade edilebilir:

$$\Pr\left(X = x \middle| \widehat{\theta}_i\right) = \sum_{\forall ui\ için\ 1 r ui = x} \prod_{J=1}^{J} P_{ij}^{u_{ij}} (1 - P_{ij})^{u_{ij}} \qquad \text{(Eşitlik 13)}$$

Burada $\prod_{J=1}^{J} P_{ij}^{u_{ij}} (1 - P_{ij})^{u_{ij}}$, Eşitlik 8 ile aynı tanıma sahiptir ve Eşitlik 13, x ham puanı verecek olan tüm olası madde yanıt/tepki örüntülerinin olasılıklarının toplanmasını gerektirir. Bu formülü kullanmanın, olası tepki modellerinin sayısı, maddelerin sayısı ile katlanarak arttığından dolayı, kullanışsız bir hale gelebileceği görülebilir. Bununla birlikte, Lord ve Wingersky (1984) tarafından ikili puanlanan maddeler için özyinelemeli bir formül geliştirildi ve sonrasında çoklu puanlanan maddelerde de kullanılması için formül genişletildi (Thissen, Pommerich, Billeaud & Williams, 1995). Sınıflama doğruluğu indekslerinin hesaplanması, Rudner'ın ve Guo'nun yöntemindeki kadar kolay değildir,

çünkü birey için gerçek yetenek theta ölçeğinde kestirilirken, sınıflama kararları ham puan ölçeğinde alınır. Bireyin gerçek sınıflamasını belirlemek için, $\hat{\theta_i}$, test karakteristik eğrisi (TKE) ile beklenen ya da gerçek ham puana ($\tau\hat{\theta_i}$) dönüştürülür. Daha sonra, $\tau\hat{\theta_i}$, ham puan ölçeğinde ölçüldüğü ve gerçek sınıflamayı belirlemek için ham kesme puanları ile karşılaştırıldığı şekliyle, bireyin gerçek yeteneğinin en iyi kestirimi olarak ele alınabilir. Eğer kesme puanları orijinal olarak theta ölçeğinde kullanılıyorsa, bunların TKE aracılığıyla ham puan ölçeğine dönüştürülmesi ve ardından mümkün olan en yakın ham puana yuvarlanması gerekmektedir. Sonrasında, gerçek kategori k içinde sınıflandırılma olasılığı şu şekilde hesaplanabilir:

$$\Pr\,(K = k|\widehat{\theta_\iota} = \sum_{x=c_{k-1}}^{c_{k-1}} \Pr\,(X = x|\widehat{\theta_\iota}) \qquad \text{(Eşitlik 14)}$$

Bu, bireyin $\hat{\theta_i}$ 'si göz önüne alındığında, tüm olası yanıt örüntülerinden k kategorisindeki her bir olası ham puanın gözlenme olasılıklarının toplanmasını gerektirir. Sınıflama tutarlılığı indekslerinin hesaplanması, aynı kesme puanlarına ve aynı MTK parametrelerine sahip maddelere ilişkin bir paralel test formunun uygulandığı ve her bir kategoride yer alma olasılıklarının her iki uygulamada da aynı olacağı yönündeki benzer varsayımları kabul eder. Koşullu sınıflama tutarlılığı şu şekilde verilir:

$$p_{\theta_i} = \sum_{k=1}^{K}\left[\Pr\,(K = k|\widehat{\theta_\iota})\right]^2 \qquad \text{(Eşitlik 15)}$$

Burada Pr($K = k|\hat{\theta_i}$) Eşitlik 14'ten elde edilen bir kategoride sınıflandırılma olasılığıdır.

Genel Sınıflama Doğruluğu ve Sınıflama Tutarlılığı İndeksleri

Yukarıda bahsedilen koşullu indeksler yalnızca belirli bir birey ya da theta seviyesi için bilgi sağlamaktadır. Bir grup için verilen kararların genel sınıflama doğruluğu ve sınıflama tutarlılığı indekslerini yorumlamak için genellikle koşullu indeksler tercih edilmez. Koşullu sınıflama doğruluğu ve sınıflama tutarlılığı indekslerini marjinalleştirmek için iki yöntem vardır. Bunlar P-yöntemi ve D-yöntemidir (Lee, 2010).

P-yöntemi ya da bireye dayalı yöntemde, her birey için koşullu indeksler ilk olarak yukarıda açıklanan yöntemler kullanılarak hesaplanır. Genel indeksler, tüm koşullu indekslerin toplanması ve toplam birey sayısı üzerinden ortalamasının alınmasıyla hesaplanır:

$$\gamma = \frac{1}{N}\sum_{i=1}^{N}\gamma_{\theta_i} \qquad \text{(Eşitlik 16)}$$

$$p = \frac{1}{N}\sum_{i=1}^{N} p_{\theta_i} \quad \text{(Eşitlik 17)}$$

Burada N, bireylerin toplam sayısını temsil etmektedir.

D-yöntemi ya da dağılıma dayalı yöntem, bir theta dağılımının gözlenen verilerden kestirilmesini veya standart normal dağılım gibi belirli bir dağılımı takip ettiğini varsayar. Genel indeksler, koşullu indekslerin uygulanabilir theta alanı üzerinde integrasyonu ile hesaplanır:

$$\gamma = \int_{\theta=-\infty}^{\infty} \gamma_\theta g(\theta) d\theta \quad \text{(Eşitlik 18)}$$

$$p = \int_{\theta=-\infty}^{\infty} p_\theta g(\theta) d\theta \quad \text{(Eşitlik 19)}$$

Burada $g(\theta)$, kestirilen ya da varsayılan dağılımdan elde edilen θ'nın yoğunluğudur. D-yöntemi, theta dağılımının uygun bir şekilde belirli bir dağılımı takip ettiği varsayıldığında, genel sınıflama doğruluğu ve sınıflama tutarlılığı indekslerini hesaplamak için uygun bir yol sağlamaktadır. Bu durumda, kestirimler yalnızca MTK parametrelerini ve kesme puanı konumlarını gerektirir ve her bireyin θ'sının kestirilmesine gerek kalmamaktadır. Ancak, sonuçların yalnızca varsayılan yetenek dağılımı hedef yetenek dağılımına uygun ölçüde yakın olduğunda, geçerli olması beklenir. Rudner'ın ve Lee'nin yöntemlerinde hem P-yöntemi hem de D-yöntemi uygulanabilir, ancak Guo'nun yöntemi dayalı sınıflama doğruluğu ve tutarlılığı indekslerinin hesaplanmasında sadece P-yöntemi kullanılabilir. Bunun sebebi, Guo'nun yöntemine dayalı sınıflama doğruluğu ve sınıflama tutarlılığı indekslerinin hesaplanması sürecinin, MTK parametreleri, kesme puanı konumları ve bireylerin gözlenen madde tepkilerini gerektiren olabilirlik fonksiyonu tarafından yönlendirilmesidir. Kestirilen thetalar, bireylerin gözlenen sınıflandırılmış kategorilerini belirlemek dışında, her bir kategoriye sınıflandırılma olasılıklarının hesaplanmasında rol oynamaz. Hem gerçek hem de simüle edilmiş verilere dayanan çalışmalar, hem P-yöntemi hem de D-yöntemlerinin, gözlenen thetalar normal dağıldığında çok benzer sonuçlar verdiğini göstermiştir (Lathrop & Chen, 2013; Lee, 2010). Ayrıca, Lee (2010) bilgisayar uyarlamalı testler gibi, testin odak noktası her bir birey üzerinde olduğunda, P-yönteminin daha uygun olduğunu, grup düzeyinde istatistiklere yoğunlaşıldığında ise D-yönteminin tercih edildiği belirtmiştir. Wyse ve Hao (2012), Guo'nun yönteminin genellikle hesaplanan en yüksek genel sınıflama doğruluğu ve sınıflama tutarlılığı oranlarını ürettiğini, ardından Rudner'ın ve ardından Lee'nin yönteminin geldiğini göstermiştir.

Çoğu durumda, üç MTK tabanlı yöntem tarafından hesaplanan genel sınıflama doğruluğu ve sınıflama tutarlılığı oranları arasındaki farklar önemsizdir; ancak, daha kısa bir test uzunluğunda (örneğin, 30 madde), özyinelemeli tabanlı Lee'nin yöntemi ile maksimum olabilirlik (ML) kestirimi tabanlı Rudner'ın veya Guo'nun yöntemi arasındaki farklar .04 ila .05'e yaklaşabilir ve bu da hesaplama yönteminin bazı durumlarda söz konusu indeksler üzerinde bazı etkileri olabileceğini gösterebilir.

Sınıflama doğruluğu ve tutarlılığı indeksleri arasındaki bazı temel benzerlikleri ve farklılıkları vurgulamak önemlidir. İlk olarak, her yöntemde farklı dağılım varsayımlarının olduğu görülmektedir. Guo tabanlı yöntemler için, tek bir dağılım varsayılmaz ve indekslerin temelindeki beklenen olasılıklar, olabilirlik fonksiyonları tarafından yönlendirilir. Bu olabilirlik işlevleri bireyin yanıt örüntüsüne bağlı olarak değişebilir. Rudner tabanlı yöntemler için, indeksler hesaplanırken, bireylerin yetenek kestiriminin normal dağılım gösterdiği varsayılır. Guo ise kestirilen thetalar için normallik varsayımı gerektirmeyen, verilerin olasılığına dayanan bir yöntem geliştirmiştir. Her üç yöntem de bireye dayalı yöntemler olarak sınıflandırılabilir. Bununla birlikte, Lee tabanlı yöntemlerde kesme puanları ham puan ölçeğinde kullanılır ve sınıflama kararları bireylerin ham puanları ile ham kesme puanları karşılaştırılarak yapılır. Rudner ve Guo tabanlı yöntemlerde ise gözlenen puanların ve kullanılan kesme puanlarının theta ölçeğine yerleştirildiği varsayılır. Bu yönüyle Rudner ve Guo tabanlı yöntemlerden farklıdır. Lee tabanlı yöntemlerde bireyin gerçek yeteneği theta ölçeğinde kestirilirken, sınıflama kararları ham puan ölçeğinde yapıldığı için diğer yöntemlere göre hesaplamasının zor olduğu düşünülebilir. Farklı varsayımlar ve yöntemler, farklı sınıflama doğruluğu ve tutarlılığı kestirimlerine yol açabilir. Bununla birlikte, tüm yöntemler, MTK modellerinin özelliklerine ve varsayımlarına dayanmaktadır. Bu, bireylerin yetenek kestirimlerinin ve kullanılan madde parametrelerinin, temel alınan parametrelerin iyi kestirimleri olduğu varsayımını içermektedir. Bahsedilen yöntemler dışında, sınıflamaların doğruluğunu incelemek amacıyla sınıflama analizleri de kullanılmaktadır.

Sınıflama Analizleri

Sınıflama analizlerinde birden fazla bağımsız değişkene ilişkin değerler kullanılarak iki ya da daha çok kategorili bağımlı değişkenler (geçti/kaldı ya da başarılı/başarısız gibi) belirlenmeye çalışılır. Sınıflama analizlerinin amacı, bağımsız değişkenlerin kombinasyonu ile bağımlı değişken açısından grup üyeliğini belirlemektir. Eğitim bilimlerinde sınıflama analizleri araştırmacılar tarafından çeşitli amaçlar ile kullanılabilir. Bunlar, farklı okul türlerine kayıt yaptıran bireyler arasındaki farklılıkları belirlemek, yapılan sınıflamaların doğruluğunu incelemek ya da öğrencilerin sınıflamalarını etkileyen cinsiyet, sosyoekonomik düzey gibi değişkenleri belirlemek olabilir. Bu kapsamda, belirtilen amaçlar için kullanılan sınıflama analizlerine ilişkin başlıca teknikler (ayırma analizi, lojistik regresyon analizi, kümeleme analizi) aşağıda açıklanmıştır.

Ayırma (diskriminant) analizi, bireylerin sınıflandırılmasında bağımsız değişkenlerin kullanılması esasına dayanan istatistiksel bir analiz tekniğidir. Burada amaç, bağımsız değişkenleri doğrusal olarak bir araya getirerek hatalı sınıflama oranını en düşük seviyede tutmak ve bireyleri yalnız bir gruba yerleştirmektir (Dillon & Goldstein, 1984). Ayırma analizi, grupların önceden bilinmesi durumunda kullanılır ve bireylere ilişkin faktörler ya da değişkenler dikkate alınarak, bu faktörlere göre bireylerin gerçek sınıflarına en doğru sınıflamayla yerleştirilmeleri sağlanır. Ayırma analizi sonucu elde edilen doğru sınıflama oranı ve kanonik korelasyon değerleri sınıflamaların doğruluğuna ilişkin bilgi sağlamaktadır. Örneğin, Güzeller ve Kelecioğlu (2006), Ortaöğretim Kurumları Öğrenci Seçme Sınavı'nda alt test puanlarına dayalı yapılan sınıflamaların geçerliğini incelemek üzere ayırma analizini kullanmışlardır. Divjak ve Oreški (2009) ise yaptıkları çalışmada, ayırma analizi ile öğrencilerin akademik performansını etkileyen faktörleri ve bunların öğrencilerin sınıflandırılmasında ne derece etkili olduğunu incelemiştir.

Lojistik regresyon analizinin temel amacı, bireyleri farklı gruplara ayırarak sınıflandırmak ve bireylerin üyesi olduğu grubu kestirmede bir regresyon denklemi oluşturmaktır (Çokluk, 2010). Aynı zamanda, bu analiz tekniğinde en az değişken kullanılarak en iyi uyum düzeyine sahip olacak şekilde bağımlı değişken ve bağımsız değişkenler kümesine ilişkin genel olarak kabul edilebilir bir model kurulması amaçlanır (Aktaş & Erkuş, 2009). Lojistik regresyon analizi, doğrusal regresyon analizine oldukça benzemektedir fakat aralarındaki fark, lojistik regresyon analizindeki bağımlı değişkenin kategorik yapıda olmasıdır (Atar, 2012). Eğitim alanında yapılan çalışmaların çoğunda problem kategorik sonuçların kestirimi ile ilgilidir

(Çokluk, 2010). Örneğin, bir öğrenci okulda akademik olarak başarılı/başarısız olarak sınıflandırılabilir ya da bir öğrencide öğrenme güçlüğü vardır/yoktur şeklinde karar verilebilir (Peng, Lee & Ingersoll, 2002). Lojistik regresyon analizi sonucunda elde edilen doğru sınıflama oranı, Cox ve Snell R kare ve Nagelkerke R kare değerleri ile sınıflama kararlarının doğruluğu incelenmektedir. Kutlu, Yıldırım, Bilican ve Kumandaş (2011) yaptıkları çalışmada, öğrencilerin okuduğunu anlamada ne kadar başarılı olduğunu lojistik regresyon analizi ile incelemişlerdir. Lojistik regresyon analizinde, ayırma analizindeki normallik, doğrusallık, varyans-kovaryans matrislerinin eşitliği gibi sayıltıların karşılanması gerekmediğinden kullanım kolaylığı söz konusudur. Lojistik regresyon tekniğinin ayırma analizinden farkı, gruba ilişkin olasılık fonksiyonunun doğrusal olan bir yöntemle hesaplanmamasıdır. Ayrıca, ayırma analizinin yalnızca sürekli bağımsız değişkenlere uygulanabilmesine rağmen, lojistik regresyon analizinde bağımsız değişkenler yalnız sürekli ya da süreksiz ve hem sürekli hem de süreksiz değişkenlerden oluşabilir. Lojistik regresyon analizinin ayırma analizine göre daha doğru sınıflama yaptığı bazı çalışmalar bulunmaktadır (Barön, 1991; Rausch & Kelley, 2009). Bu analiz, regresyon denklemiyle oluşturulduğu ve diğer analiz yöntemlerine kıyasla daha kullanışlı olması sebebiyle, kategorik veri analizinde önemli görülmektedir (Kılıç, 2000).

Kümeleme analizi, verileri ya da bireyleri sahip oldukları özelliklere göre gruplara ayırmaya ya da sınıflara yerleştirmeye yardımcı olan bir analiz tekniğidir. Bu analizde, gruplanmamış veriler ya da bireyler benzerliklerine göre sınıflandırılmaktadır. Bu yönüyle ayırma analizine benzemektedir. Kümeleme analizi, lojistik regresyon analizine yakınlığı bulunan bir tekniktir fakat lojistik regresyon analizinden farklı olarak bu teknikte grup sayısı ve grup üyelikleri bilinmemektedir (Çokluk, 2010). Bu teknikte gözlemler, benzerlik ya da uzaklık ölçütlerine dayalı olarak kümelendirilir, gruplandırılır (Çokluk, 2010). Kümeleme analizi sonucunda elde edilen aglomeratif çizelge ve ağaç grafiği, kümelerin benzerliklerine dolayısıyla sınıflamaların doğruluğuna ilişkin bilgi sağlamaktadır. Kümeleme geçerliğinin incelenmesinde ve uygun küme sayısının seçilmesinde silhouette indeksi kullanılır. Oluşacak her bir küme silhouette çizimi ile gösterilir ve bu çizim, yer aldıkları kümelerle uyumlu elemanları ve yanlış kümede yer alan elemanları gösterir. Örneğin, kümeleme analizinin kullanıldığı çalışmalardan birinde öğrencilerin güdülenme durumlarına göre sınıflandırılması kümeleme analizinin dahil olduğu dört teknik ile incelenmiştir (Ertürk, 2016). Ayrıca, kümeleme analizinde toplam puanların normalliği haricinde herhangi bir sayıltı bulunmamaktadır.

Bu araştırmada KTK'ye dayalı yöntemler ya da sınıflama analizlerinin değil MTK'ye dayalı yöntemlerin incelenmesi amaçlanmıştır. Son yıllarda MTK'nin ölçme ve değerlendirme süreçlerinde önemli bir yer tuttuğu ve çeşitli alanlarda etkin bir şekilde kullanıldığı görülmektedir. Çünkü MTK'de KTK'den farklı olarak bireylerin madde yanıt örüntüleri dikkate alınmakta ve daha hassas kestirimler yapılabilmektedir. MTK altında elde edilen birey puanları bireylerin yanıt örüntülerine dayalı olarak elde edildiği için doğası gereği bu puanlar KTK altında elde edilen puanlardan farklılaşmaktadır. Söz konusu farklılaşmadan dolayı KTK'ye dayalı geliştirilen indekslerin MTK altında elde edilen birey puanlarına dayalı sınıflama çalışmaları için kullanılması sonuçların doğruluğunu düşürebilmektedir. Bu durum MTK'ye dayalı sınıflama indekslerinin geliştirilmesinin önünü açmıştır. Bu doğrultuda başta Rudner'ın, Guo'nun ve Lee'nin indeksi olmak üzere sınıflama doğruluğu ve sınıflama tutarlılığını değerlendirmek amacıyla çeşitli indeksler geliştirilmiştir (Bourque, Goodman, Hambleton ve Han, 2004; Guo, 2006; Lee, 2010; Rudner, 2001, 2005). Bu çalışmada da MTK'ye dayalı yöntemlerin incelenmesi kararlaştırılmıştır.

Hem KTK hem de MTK tabanlı yöntemlerde, sınıflama indeksleri, gözlenen veriler ve bazı matematiksel modeller kullanılarak hesaplanır ve gözlenen verinin altında yatan gerçek model ile veriler için örtük model arasında her zaman farklılıklar olduğundan, bir dereceye kadar modelin yanlış belirlenme durumu ortaya çıkacaktır. Sınıflama indekslerinin hesaplanmasında modelin yanlış tanımlanmasının, dolayısıyla model uyumsuzluğunun etkisinin önemi birkaç araştırmacı tarafından belirtilmiştir (Chau, 2018; Hanson & Brennon, 1990; Lathrop & Chen, 2013; Lee, 2010). Sınıflama indekslerinin hesaplanmasında MTK tabanlı yöntemler arasındaki temel varsayım ve yöntemlerdeki farklılıklara rağmen, hesaplanan indekslerin doğruluğu, kalibrasyon yöntemlerinde yanlış tanımlanmış bir MTK modeli kullanılırsa olumsuz etkilenebilir.

MTK'ye dayalı yöntemlerden Rudner'ın, Guo'nun ve Lee'nin yönteminin karşılaştırıldığı alan yazındaki benzer çalışmalarda (Chau, 2018) sınıflama doğruluğu ve sınıflama tutarlılığı indekslerini etkilediği belirtilen test uzunluğu ve örneklem büyüklüğü faktörlerinin etkisinin sadece belli bir sabit değer alınarak incelendiği görülmektedir. Bu çalışmada, söz konusu faktörlerin değişimlenerek etkilerinin daha detaylı çalışılması amaçlanmıştır. Bu bağlamda, Türkiye'de eğitim ortamlarında kullanılan alt testler incelendiğinde, madde sayılarının genellikle kısa ve orta uzunlukta olduğu görülmektedir. Örneğin Liselere Geçiş Sistemi (LGS) 10 ve 20 maddelik testlerden oluşurken, Temel Yeterlilik Testi (TYT) ve Alan Yeterlilik Testi (AYT) ise en az 5 maddelik, en fazla 40 maddelik alt testlerden oluşmaktadır.

Dolayısıyla, bu yöntemlerin daha kısa testler ve daha az sayıda bireyden oluşan gruplara uygulandığında sınıflama doğruluğu ve sınıflama tutarlılığı indekslerinin ne derece doğru kestirildiğinin incelenmesinin önemli olduğu düşünülmüştür. Bu sebeple, bu çalışmada test uzunluğu ve örneklem büyüklüğü faktörleri değişimlenerek, sınıflama doğruluğu ve sınıflama tutarlılığı indekslerinin bu faktörlere bağlı değişimlerinin incelenmesi, aynı zamanda söz konusu indekslerin model yanlış tanımlaması durumuna duyarlılığının araştırılması amaçlanmıştır.

İlgili Araştırmalar

Türkiye'de ve yurtdışında sınıflama doğruluğunun incelenmesi ile ilgili çok sayıda çalışma (Atar, Sayın & Atar, 2013; Büyükatak, 2016; Chau, 2018; Çelikten & Çakan, 2019; Guo, 2006; Lathrop & Cheng, 2014; Lee, 2010; Lee, Hanson & Brennan, 2002; Özkan & Doğan, 2013; Rudner, 2001, 2005; Sen & Cohen, 2020; Wyse & Hao, 2012) bulunmaktadır. Türkiye'deki çalışmalara bakıldığında sınıflama doğruluğunun incelenmesinde sınıflama analizlerinin sıklıkla kullanıldığı görülmektedir. Fakat MTK'ye dayalı yöntemlerden Rudner, Guo ve Lee'nin yöntemlerinin karşılaştırmalı olarak incelendiği çalışmaların (Chau, 2018; Diao & Sireci, 2018; Wyse & Hao, 2012) sınırlı sayıda olduğu görülmektedir.

Martineau (2007) çalışmasında, Rudner'ın yöntemini incelemiş ve bu yöntemin kullanılabileceği en uygun koşulları araştırmıştır. Rudner'ın indeksinin, pozitif yönlü yanlılığa sahip olduğu ve örneklem büyüklüğünün çok fazla sayıda olduğu durumlarda kullanılmasının uygun olduğu sonucuna ulaşılmıştır.

Wyse ve Hao (2012), MTK modellerinde kullanılabilecek iki yeni sınıflama tutarlılığı indeksi önermiştir. Bu indekslerin Rudner'ın ve Guo'nun sınıflama doğruluğu indeksleri ile ilişkili olduğu belirtilmiştir. Aynı zamanda bu çalışmada Rudner'ın ve Guo'nun yöntemine dayalı hesaplanan sınıflama doğruluğu ve tutarlılığı indeksleri ile MTK-özyinelemeli yöntemler karşılaştırılmıştır. Guo'nun yöntemine dayalı hesaplanan indekslerin en yüksek sınıflama doğruluğu ve sınıflama tutarlılığı değerlerine sahip olduğu sonucuna ulaşılmıştır.

Chau (2018) ise çalışmasında model yanlış tanımlaması durumunun, Rudner'ın, Guo'nun ve Lee'nin yöntemine dayalı hesaplanan sınıflama doğruluğu ve tutarlılığı indeksleri üzerindeki etkisini araştırmıştır. Söz konusu indekslerin model yanlış tanımlaması durumundan etkilenme dereceleri incelenmiştir. Ayrıca, theta ölçeğinde yer alan kesme puanının konumu da indekslerin üzerindeki etkisi açısından incelenmiştir. Genel olarak, kesme puanı testin en

fazla bilgiye sahip olduğu noktaya yerleştirildiğinde, MTK'ye dayalı bu üç yöntem kullanılarak hesaplanan sınıflama doğruluğu ve tutarlılığı indekslerinin gerçek değerlerine daha yakın sonuçlar verdiği gözlenmiştir. Bununla birlikte, model yanlış tanımlaması durumlarında elde edilen sınıflamaların, doğru model altında yapılan sınıflamalardan farklı olduğuna ulaşılmıştır Diao ve Sireci (2018) yaptıkları çalışmada, sınıflama doğruluğu ve sınıflama tutarlılığı indekslerini hesaplamak amacıyla kullanılabilecek mevcut yöntemleri ve bu amaçla kullanılabilecek ücretsiz yazılımların açıklamalarını sunmuştur.

Lee, Hanson ve Brennan (2002) yaptıkları çalışmada, tek bir uygulamaya dayalı test verilerini kullanarak çoklu sınıflandırmalar için farklı sınıflama doğruluğu ve tutarlılığı indekslerini hesaplamaya yönelik yöntemlerden bahsetmiştir. Ayrıca, sınıflama doğruluğu ve tutarlılığı indekslerini üç farklı model (iki parametreli beta binom, dört parametreli beta binom ve üç parametreli lojistik model) altında incelemişlerdir. Elde edilen bulgular, sırasıyla MTK modelinin, dört parametreli beta binom modelin ve iki parametreli beta binom modelinin verilere daha iyi uyum sağladığına işaret etmiştir.

Sen ve Cohen (2020), çok düzeyli karma MTK modellerinde test ve örneklem özelliklerinin model seçimi ve sınıflama doğruluğu üzerindeki etkisini incelemiştir. Elde edilen sonuçlar, örneklem büyüklüğü küçük olduğunda çok düzeyli karma MTK modellerinin daha az doğru kestirim sonuçları verdiğini göstermiştir.

Lathrop ve Cheng (2014), sınıflama doğruluğu ve tutarlılığı indekslerini hesaplamak için parametrik olmayan bir yöntem önermişlerdir. Parametrik olmayan sınıflama doğruluğu ve sınıflama tutarlılığı indekslerinin etkililiği, farklı MTK modelleri, test uzunlukları ve yetenek dağılımları ile çeşitli koşullarda üretilen simülasyon veri setleri ile test edilmiştir. Sınıflama doğruluğu için parametrik olmayan bu yaklaşım ile, Lee'nin ve Livingston ve Lewis'in yönteminden daha iyi sonuçlar elde edilmiştir. Parametrik olmayan sınıflama tutarlılığı indeksi ise Lee'nin yöntemine benzer şekilde performans göstermiş ve yetenek dağılımları normal olmadığında Livingston ve Lewis'in yönteminden daha iyi performans göstermiştir.

Çelikten ve Çakan (2019) yaptıkları çalışmada, MTK kapsamında farklı kestirim yöntemlerinden elde edilen yetenek kestirimlerine göre bireylerin sınıflandırılmasına ilişkin sınıflama doğruluğu ve sınıflama tutarlılığı indekslerini farklı örneklem büyüklükleri açısından karşılaştırmıştır. Bu çalışmada Rudner'ın yöntemine dayalı hesaplanan sınıflama doğruluğu ve tutarlılığı indeksleri kullanılmıştır. Nonbayesian yetenek kestirimlerine dayalı

yapılan sınıflamaların, Bayesian yöntemlerinden daha daha doğru ve tutarlı sonuçlar verdiği görülmüştür.

Bu araştırmalarda sınıflama doğruluğunun incelenmesi için mevcut yöntemlerin etkililiğinin araştırıldığı ya da yeni yöntemlerin önerildiği görülmektedir. Bunun yanı sıra, sınıflama doğruluğu ve tutarlılığı indekslerini etkileyen farklı değişkenlerin etkisinin incelendiği de görülmektedir. Sınıflama doğruluğunu etkileyebilecek model, test uzunluğu, yetenek kestirimi, örneklem büyüklüğü, kesme puanı ve benzeri birçok değişken olabileceği için bu tarz çalışmaların yapılması ve sınıflama doğruluğunu etkileyen faktörlerin araştırılması büyük öneme sahiptir. Sınıflama doğruluğunu incelemek amacıyla sınıflama analizlerinin de kullanıldığı çok sayıda çalışma (Artino & Stephens, 2009; Atar, 2012; Atar, Sayın & Atar, 2013; Büyükatak, 2016; Ceylan, 2009; Ceylan & Akerson, 2014; Divjak & Oreški, 2009; Ertürk, 2016; Güzeller & Kelecioğlu, 2006; Kalender, 2015; Kutlu vd., 2011; McCoach & Siegle, 2003; Onu & John, 2016; Özkan, 2016; Özkan & Anıl, 2014; Özkan & Doğan, 2013; Taşdemir & Şahin, 2016) bulunmaktadır.

McCoach ve Siegle (2003) yaptıkları çalışmada, üstün yetenekli yüksek başarılı öğrenciler ile üstün yetenekli düşük başarılı öğrencileri ayıran değişkenleri belirlemeyi ve bu değişkenlerin öğrencileri ne derecede doğru sınıflandırdığını lojistik regresyon analizi ile incelemişlerdir. Çalışmanın örneklemi 178 öğrenciden oluşmaktadır. Analiz sonucunda, etkisi incelenen 5 değişkenden sadece 2 değişkenin grup üyeliğini yordamada istatistiksel anlamda manidar olduğu görülmüştür. Modelin öğrencileri doğru sınıflama oranı ise %81,8 bulunmuştur.

Güzeller ve Kelecioğlu (2006) yaptıkları çalışmada, Ortaöğretim Kurumları Öğrenci Seçme Sınavı'nda alt test puanlarına bağlı yapılan yerleştirmelerin geçerliğini ayırma analizi ile incelemiştir. Çalışmanın örneklemini 967 öğrenci oluşturmaktadır. Ayırma fonksiyonunda en ayırıcı değişkenin fen alt testi, ikinci ayırıcı değişkenin de matematik alt testi olduğu görülmüştür. Doğru sınıflama oranının ise özel fen liselerinde %36,7, Anadolu liselerinde %52,7 ve resmi fen liselerinde %96 olduğu görülmüştür.

Divjak ve Oreški (2009), organizasyon ve bilişim fakültesi öğrencilerinin akademik performanslarını etkileyen değişkenleri ve bu değişkenlerin öğrencileri başarı durumları bakımından ne derecede doğru sınıflandırdığını incelemek amacıyla ayırma analizini kullanmıştır. Araştırmanın örneklemi, toplam 223 öğrenciden oluşmaktadır. Analiz

sonucunda, etkisi incelenen 30 değişkenden 8'inin anlamlı bulunduğu ve sınıflama doğruluğunun %75,22 olduğu sonucuna ulaşılmıştır.

Artino ve Stephens (2009), çevrimiçi öğrenme ortamında lisans ve lisansüstü öğrencileri arasındaki potansiyel farkları akademik motivasyon ve öz düzenleme açısından incelemek için lojistik regresyon analizini kullanmıştır. Çalışmanın örneklemi 194 öğrenciden oluşmaktadır. Analiz sonucunda, 6 değişkenin öğrencileri sınıflandırmada etkili olduğu ve modelin sınıflama doğruluğunun %79,9 olduğu görülmüştür.

Ceylan (2009) Türkiye'deki yüksek ve düşük performans gösteren okullar arasındaki farklılıkları, Uluslararası Öğrenci Değerlendirme Programı (Programme for International Student Assessment; PISA) 2006 öğrenci anketinden sağlanan beş örtük değişken bağlamında incelemiştir. Araştırmanın örneklemini toplamda 938 öğrenci oluşturmaktadır. Örtük değişkenleri tespit etmek amacıyla öğrenci anketindeki yanıtlara faktör çözümlemesi yapılmıştır ve sonrasında fen okuryazarlığında yüksek ve düşük performans gösteren okullar arasındaki farkları belirlemek için ayırma analizi yapılmıştır. Analizler sonucunda, okulların beş örtük değişkenden dört örtük değişken açısından farklılaştığı gözlenmiştir.

Kutlu vd. (2011) ilköğretim 5.sınıf öğrencilerinin okuduğunu anlamada başarılı olup olmadığını ve bu durumu etkileyen faktörleri lojistik regresyon analizi ile incelemişlerdir. Araştırmanın örneklemi 279 öğrenciden oluşmaktadır. Analiz sonucunda, etkisi incelenen 7 değişkenden 5'inin öğrencilere başarılarına göre ayırmada etkisinin olduğu görülmüştür.

Atar (2012) yaptığı çalışmada, özel yetenek sınavı, Yükseköğretime Geçiş Sınavı (YGS) ve ağırlıklı ortaöğretim başarı puanlarının öğrencilerin resim iş öğretmenliği programına yerleşme durumlarına etkisini ayırma analizi ile, alt puanların bireyleri ayırt edebilme derecesini (sınıflama doğruluğunu) lojistik regresyon analizi ile incelemiştir. Çalışmanın örneklemini, resim iş öğretmenliği özel yetenek sınavlarına katılan 305 aday oluşturmaktadır. Ayrıca bu çalışmada, Ölçme, Seçme ve Yerleştirme Merkezi (ÖSYM)'nin kullandığı puan türü ağırlıkları için alternatif bir hesaplama yöntemi önerilmiştir. Alt puanların yerleştirme puanlarını etkileme derecesi ayırma analizi ile incelenmiştir. Lojistik regresyon analizi sonucunda ise önerilen yöntem ile elde edilen toplam doğru sınıflama yüzdesinin ÖSYM'nin kullandığı yöntem ile elde edilen toplam doğru sınıflama yüzdesinden daha yüksek olduğu görülmüştür. Ayrıca ayırma analizinden elde edilen bulgular, lojistik regresyon analizinden elde edilen bulguları destekler niteliktedir.

Atar, Sayın ve Atar (2013), müzik eğitimi programına başvuran bireylerin programa yerleşmelerinde kullanılan alt puan türlerinin yerleştirme puanını yordama derecelerini belirleyip bireylerin programa yerleşme durumlarına göre sınıflama doğruluğunu ayırma analizi ile incelemişlerdir. Çalışmanın örneklemi, 100 öğrenciden oluşmaktadır. Elde edilen bulgular sonucunda, programa yerleşen ve yerleşemeyen bireylerin sınıflama doğruluğu %94 bulunmuştur.

Özkan ve Doğan (2013) yaptıkları çalışmada, ilköğretim sekizinci sınıf öğrencilerinin okuma becerilerindeki başarıyı etkileyen değişkenleri ve bu değişkenlerin öğrencileri sınıflandırma derecelerini lojistik regresyon analizi ile incelemişlerdir. Çalışmanın örneklemi, 3004 öğrenciden oluşmaktadır. Okuma becerilerindeki başarıyı etkilemesi beklenen 5 değişkenden 4'ünün (cinsiyet, evdeki kitap sayısı, sınıfta okumaya ayrılan zaman ve okul dışı okumaya ayrılan zaman) bireyleri gruplara ayırmada etkili olduğu görülmüştür.

Özkan ve Anıl (2014), PISA'da son zamanlarda üst sıralarda olan Kore ve orta sıralarda olan Türkiye ve Danimarka'nın mesleki gelişim ihtiyacı, mesleki gelişim etkinliklerine katılım ve bu etkinliklerin etki düzeyi, mesleki gelişime ayrılan zaman ve öğretmenlerin zorunlu olarak katıldığı mesleki gelişim gün sayısı değişkenlerine dayalı olarak doğru sınıflandırılma derecelerini incelemiştir. Çalışma kapsamında 7916 öğretmenden veri toplanmıştır. Ayırma analizi sonucunda, değişkenlerin hepsinin grupları ayırmada etkili olduğu görülmüştür.

Ceylan ve Akerson (2014), Uluslararası Matematik ve Fen Eğilimleri Araştırması (Trends in International Mathematics and Science Study; TIMSS) 2007 verilerini kullanarak Amerika'daki yüksek ve düşük performans gösteren okullar arasındaki farkları belirlemek için ayırma analizini kullanmıştır. Araştırmanın örneklemi, 1465 öğrenciden oluşmaktadır. Yapılan analizler sonucunda, yüksek ve düşük performans gösteren okullara doğru bir şekilde sınıflandırılmış öğrenci oranının %80,7 olduğu görülmüştür.

Kalender (2015) çalışmasında, PISA 2012'de düşük sosyoekonomik düzeyde, yüksek ve düşük başarılı öğrenciler arasındaki okuma becerileri farklılıklarını etkileyen faktörleri ve bu faktörlerin öğrencileri sınıflama düzeyini lojistik regresyon analizi ile incelemiştir. Çalışmanın örneklemi, 1200 öğrenciden oluşmaktadır. Elde edilen bulgulara göre, lojistik regresyon analizi ile kurulan model ile bireylerin doğru sınıflama oranı %79,4 bulunmuştur. Akademik anlamda üstün başarılı öğrencilerin düşük başarılı öğrencilere göre öğretmen ve okula karşı olumlu tutum sergiledikleri de gözlenmiştir.

Onu ve John (2016) Abia Eyaleti'nde yer alan bir okuldaki öğrencileri, ilk yıllarındaki akademik performansları açısından iki gruba (istatistik ile bilgisayar bilimleri) ayırmak için ayırma analizini kullanmışlardır. Toplamda 120 öğrenci aldıkları dört derse göre gruplara ayrılmışlardır. Analiz sonucunda, sınıflama doğruluğunun %92,5 olduğu görülmüştür.

Ertürk (2016) araştırmasında, üniversite öğrencilerinin akademik güdülenme durumlarını etkileyen değişkenleri belirleyip öğrencilerin güdülenme durumlarına göre sınıflandırılmasını faktör analizi, kümeleme analizi, lojistik regresyon analizi ve ayırma analizi ile incelemiştir. Çalışmanın örneklemi 671 öğrenciden oluşmaktadır. Elde edilen sonuçlar incelendiğinde, ayırma analizi ile elde edilen toplam doğru sınıflama yüzdesinin, lojistik regresyon analizi sonucunda elde edilen toplam doğru sınıflama yüzdesinden yüksek olduğu sonucuna ulaşılmıştır.

Taşdemir ve Şahin (2016), üniversite öğrencilerinin ölçme ve değerlendirme ders başarılarını incelemek amacıyla lojistik regresyon ve ayırma analizini kullanmışlardır. Bu amaçla, 410 öğrenci dersten geçme kalma durumlarına göre iki gruba ayrılmıştır. Öğrencilerin geçme kalma durumunu etkileyen 3 değişken ile yapılan analizler sonucunda, ayırma analizi ile elde edilen toplam doğru sınıflama yüzdesinin, lojistik regresyon analizi sonucunda elde edilen toplam doğru sınıflama yüzdesinden daha yüksek olduğu gözlenmiştir.

Özkan (2016) yaptığı çalışmada, PISA 2012'ye katılan 149 öğrenciden elde edilen verileri kullanarak okulları başarılarına göre sınıflamada etkili olan değişkenleri belirleyip bu değişkenlerin okulları sınıflama doğruluğunu lojistik regresyon analizi ile incelemiştir. Bu amaçla, okullar okul başarı puanına göre başarılı ve başarısız olmak üzere iki gruba ayrılmıştır. Analiz sonucunda, okul başarılarını etkileyeceği düşünülen 10 değişkenden 6'sının (okulda öğrencilere sunulan ekstra etkinlikler, eğitimsel kaynakların kalitesi, öğretmen başına düşen öğrenci sayısı, öğretim faaliyetlerini aksatan öğrenci ile ilişkili değişkenler ve öğretmen katılımı) anlamlı etkisi görülmüştür.

Büyükatak (2016) çalışmasında, PISA 2012'de yer alan matematiğe yönelik duyuşsal özelliklerin bölge, cinsiyet ve okul türü değişkenlerine göre sınıflama doğruluğunu karesel ayırma analizi ile incelemiştir. Toplamda 4848 öğrenciden toplanan veriler ile yapılan analiz sonucunda, cinsiyet ve okul türü değişkenlerinin sınıflama yapmak için kullanılabileceği, bölge değişkeninin ise kullanılamayacağı belirtilmiştir.

Yukarıda özetlenen araştırmalar incelendiğinde, genel olarak sınıflama doğruluğunu incelemek amacıyla sınıflama analizlerinden en sık kullanılan yöntemlerin lojistik regresyon ve ayırma analizleri olduğu görülmektedir. Özellikle lojistik regresyon analizinde normallik, doğrusallık gibi sayıltıların karşılanması gerekmediğinden, kullanım kolaylığı sebebiyle birçok alanda tercih edilmektedir. Ayrıca, bireylere ilişkin sınıflamaların doğruluğunun yanında sınıflamaları etkileyen faktörlerin etkisinin de birçok çalışmada incelendiği görülmüştür. Yukarıda özetlenen araştırmalarda, sınıflama doğruluğunu incelemek amacıyla birçok analiz yönteminin geliştirildiği ve alan yazında bunlarla ilgili çok sayıda çalışmanın yapıldığı görülmüştür. MTK'ye dayalı yöntemlerden Rudner'ın, Guo'nun ve Lee'nin yöntemine dayalı çalışmalara da (Chau, 2018; Wyse & Hao, 2012) alan yazında rastlanmış olsa da farklı değişkenlerin indekslerin doğruluğu üzerindeki etkilerinin araştırılmasının önemli olduğu düşünülmektedir.

Araştırmanın Amacı

Bu araştırmanın amacı sınıflama doğruluğu ve sınıflama tutarlılığı indekslerini kestirmede MTK'ye dayalı yöntemlerden Rudner'ın, Guo'nun ve Lee'nin yönteminin etkililiğini, 1 Parametreli Lojistik Model (1PLM), 2 Parametreli Lojistik Model (2PLM) ve 3 Parametreli Lojistik Model (3PLM) altında farklı test uzunluğu ve örneklem büyüklükleri açısından karşılaştırmalı olarak incelemektir.

Yanlılık, sınıflama kararlarının doğruluğunu sistematik bir şekilde etkilemektedir (Wang & Wang, 2001). Bu sebeple araştırmada hesaplanan sınıflama doğruluğu ve tutarlılığı indekslerinin doğruluğu yanlılık değerleri hesaplanarak incelenmiştir. Ayrıca bu indekslerin model yanlış tanımlamasına duyarlılığı da araştırılacaktır. Bu genel amaç çerçevesinde, aşağıdaki sorulara yanıt aranacaktır:

1. 1PLM altında üretilen veri 1PLM altında test edildiğinde;
    a. Her bir yöntem altında kestirilen sınıflama doğruluğu ve tutarlılığı indeksleri test uzunluğuna bağlı olarak farklılaşmakta mıdır?
    b. Her bir yöntem altında kestirilen sınıflama doğruluğu ve tutarlılığı indeksleri örneklem büyüklüğüne bağlı olarak farklılaşmakta mıdır?
    c. Test uzunluğuna bağlı olarak, farklı yöntemlerle kestirilen sınıflama doğruluğu ve tutarlılığı indekslerinin doğruluk derecesi farklılaşmakta mıdır? Hangi yöntem daha düşük bir yanlılık değeri ile sonuçlanmaktadır?

d. Örneklem büyüklüğüne bağlı olarak, farklı yöntemlerle kestirilen sınıflama doğruluğu ve tutarlılığı indekslerinin doğruluk derecesi farklılaşmakta mıdır? Hangi yöntem daha düşük bir yanlılık değeri ile sonuçlanmaktadır?

2. 2PLM altında üretilen veri 2PLM altında test edildiğinde;
    a. Her bir yöntem altında kestirilen sınıflama doğruluğu ve tutarlılığı indeksleri test uzunluğuna bağlı olarak farklılaşmakta mıdır?
    b. Her bir yöntem altında kestirilen sınıflama doğruluğu ve tutarlılığı indeksleri örneklem büyüklüğüne bağlı olarak farklılaşmakta mıdır?
    c. Test uzunluğuna bağlı olarak, farklı yöntemlerle kestirilen sınıflama doğruluğu ve tutarlılığı indekslerinin doğruluk derecesi farklılaşmakta mıdır? Hangi yöntem daha düşük bir yanlılık değeri ile sonuçlanmaktadır?
    d. Örneklem büyüklüğüne bağlı olarak, farklı yöntemlerle kestirilen sınıflama doğruluğu ve tutarlılığı indekslerinin doğruluk derecesi farklılaşmakta mıdır? Hangi yöntem daha düşük bir yanlılık değeri ile sonuçlanmaktadır?
3. 3PLM altında üretilen veri 3PLM altında test edildiğinde;
    a. Her bir yöntem altında kestirilen sınıflama doğruluğu ve tutarlılığı indeksleri test uzunluğuna bağlı olarak farklılaşmakta mıdır?
    b. Her bir yöntem altında kestirilen sınıflama doğruluğu ve tutarlılığı indeksleri örneklem büyüklüğüne bağlı olarak farklılaşmakta mıdır?
    c. Test uzunluğuna bağlı olarak, farklı yöntemlerle kestirilen sınıflama doğruluğu ve tutarlılığı indekslerinin doğruluk derecesi farklılaşmakta mıdır? Hangi yöntem daha düşük bir yanlılık değeri ile sonuçlanmaktadır?
    d. Örneklem büyüklüğüne bağlı olarak, farklı yöntemlerle kestirilen sınıflama doğruluğu ve tutarlılığı indekslerinin doğruluk derecesi farklılaşmakta mıdır? Hangi yöntem daha düşük bir yanlılık değeri ile sonuçlanmaktadır?
4. Belirli bir MTK modeli altında üretilen veri, farklı bir MTK modeli altında test edildiğinde yani model yanlış tanımlaması durumunda;

a. Sınıflama doğruluğu ve tutarlılığı indekslerinin doğruluk derecesi farklılaşmakta mıdır? Hangi sınıflama doğruluğu ve tutarlılığı indeks(ler)i model yanlış tanımlamasından daha az etkilenmektedir?

Araştırmanın Önemi

Eğitim alanında kullanılan testler ile bireylerin yaşamlarını etkileyecek seçme ve sınıflamaya dayalı kararlar alınmaktadır. Bu kapsamda sınıflama doğruluğu ve sınıflama tutarlılığı indekslerinin incelenmesi bireyler ve toplum açısından oldukça önemlidir. Bu araştırmada sınıflama doğruluğu ve tutarlılığı indeksleri 1PLM, 2PLM ve 3PLM altında, MTK'ya dayalı yöntemlerden Rudner'ın, Guo'nun ve Lee'nin yöntemlerine dayalı olarak farklı test uzunluğu ve örneklem büyüklüğüne göre karşılaştırılmıştır. Alan yazında söz konusu yöntemlerin etkililiğinin araştırıldığı çalışmalar (Chau, 2018; Wyse & Hao, 2012) bulunmasına rağmen, test uzunluğu ve örneklem büyüklüğünün etkisinin araştırıldığı bu tarz bir çalışmaya alan yazında rastlanmamıştır.

Sınıflama doğruluğu ve tutarlılığı indekslerinin hesaplanmasında MTK tabanlı yöntemlerin performansının değerlendirilmesi ve karşılaştırılmasına yönelik birkaç çalışma olmasına rağmen, bilindiği kadarıyla, yanlış model tanımlamasının MTK tabanlı sınıflama indekslerini hesaplama üzerindeki etkisini inceleyen yeterince çalışma mevcut değildir. Bu araştırma, Türkiye'de sınıflama doğruluğu ve tutarlılığı indekslerinin incelenmesinde görülen eksiklik ve Türkçe alan yazında bu tarz bir çalışmaya rastlanmamış olması nedenleriyle önemli görülmektedir.

Sınırlılıklar

Bu çalışmadan elde edilen sonuçlar yalnızca MTK'ye dayalı sınıflama doğruluğu ve sınıflama tutarlılığı indekslerinden elde edilen sonuçlarla sınırlıdır.

Ayrıca, bu çalışmada üretilen veri setleri kesme puanlarına göre üretilmemiştir. Sınıflama doğruluğu ve sınıflama tutarlılığı indekslerinin hesaplanması sürecinde farklı kesme puanına ilişkin indeks değerleri hesaplanmıştır.

BÖLÜM II

YÖNTEM

Araştırmanın Modeli

Bu araştırmada sınıflama doğruluğu ve tutarlılığı indekslerinin farklı koşullar altında karşılaştırılması amaçlandığından araştırma koşullarının oluşturulması için ilk olarak simülasyon veri setleri üretilmiştir. Bu sebeple araştırma bir simülasyon çalışmasıdır. Simülasyon çalışmaları korelasyonel, deneysel ve betimsel olabilir (Erkuş, 2013). Bu araştırma, farklı sınıflama doğruluğu ve tutarlılığı indekslerinin çeşitli koşullar altında incelenip birbirleriyle karşılaştırılması açısından betimsel bir araştırmadır. Betimsel araştırmalar, verilen bir durumu olabildiğince tüm yönleriyle inceleyen ve raporlayan araştırmalardır (Fraenkel, Wallen, & Hyun, 2012).

Verilerin Simülasyonu

Araştırma amacı doğrultusunda veriler R (R Core Team, 2017) yazılım ortamında 1PL, 2PL ve 3PL model altında iki farklı test uzunluğu (15 ve 30) ve iki farklı örneklem büyüklüğüne (500 ve 1000) dayalı olarak üretilmiştir. Alan yazındaki çalışmalara (Chau, 2018; Lathrop & Cheng, 2014; Wyse & Hao, 2012) bakıldığında, sınıflama doğruluğu ve tutarlılığı indeksleri için örneklem büyüklüğü ve test uzunluğunun önemli bir değişken olduğu görülmüştür. Bu çalışma kapsamında örneklem büyüklükleri, sınıflama doğruluğu ve sınıflama tutarlığı indeks değerlerinin örneklem büyüklüğüne göre nasıl değişeceğini gözlemlemek üzere n=500 ve n=1000 olarak belirlenmiştir. Alan yazındaki çalışmalar incelendiğinde (Chen, de la Torre ve Zhang, 2013; Minchen, de la Torre ve Liu, 2017; Minchen ve de la Torre, 2018; Terzi ve de la Torre, 2018), test uzunluklarının kısa testler için 15 ve uzun testler için 30 maddeden oluştuğu görülmektedir. Dolayısıyla, bu çalışmada madde sayısı 15 ve 30 olarak belirlenmiştir. Bu doğrultuda bu çalışmada, Tablo 1'de sunulan

12 simülasyon koşulu için ayrı ayrı veri setini üretmek amacıyla R Studio programı *irtoys* (Partchev, 2017) paketi, *sim* komutu kullanılmıştır.

Tablo 3.

Farklı Örneklem Büyüklüğü ve Madde Sayısı İçin Farklı Modeller Altında Üretilen Veri Setleri

Simülasyon Koşulu	Model	Örneklem Büyüklüğü	Madde Sayısı
1	1PLM	500	15
2	1PLM	500	30
3	1PLM	1000	15
4	1PLM	1000	30
5	2PLM	500	15
6	2PLM	500	30
7	2PLM	1000	15
8	2PLM	1000	30
9	3PLM	500	15
10	3PLM	500	30
11	3PLM	1000	15
12	3PLM	1000	30

Tablo 1'de sunulan simülasyon koşullarına yönelik olarak veri setlerini üretmek için öncelikle bireylerin yetenek parametreleri için (-4, +4) aralığında, ortalaması 0 ve standart sapması 1 olan normal dağılımdan (N(0,1)) toplam 500 ve 1000 tane değer, R yazılımında seçkisiz olarak seçilmiştir. Araştırmada test edilen her bir model altında 1000 bireyin yer aldığı veri setlerini üretmek için yine aynı dağılımdan 1000 tane seçkisiz değer seçilmiştir. Bu değerler bilinen theta değerleri olarak ele alınmıştır.

Veri setleri, MTK modellerinden 1PLM, 2PLM ve 3PLM temel alınarak, bilinen theta değerleri ve aşağıda açıklanan yol izlenerek elde edilen madde parametre değerlerine dayalı olarak üretilmiştir. Her bir simülasyon koşulu için veri üretmek üzere madde parametrelerinin belirlenmesinde aşağıdaki yol izlemiştir:

- a parametresi için değerler, orta ve yüksek ayırıcılık düzeylerini temsil edebilmesi açısından tek biçimli dağılımdan U[0,5, 2,0] olacak şekilde (Kingsbury ve Weiss, 1980);
- b parametresi için değerler, gerçek uygulamadaki değerlere yakın olabilmesi açısından, normal dağılımdan N(-0,5, 1,5) olacak şekilde (Thompson, 2009; Warm, 1989);

- c parametresi için değerler yine gerçek bir uygulama düşünülerek, normal dağılımdan N(0,20, 0,05) olacak şekilde (Thompson, 2009) çekilmiştir.

Bu doğrultuda 15 maddelik veri setlerini üretmek üzere 15 madde için; a parametresi için tek biçimli dağılımdan U[0,5, 2,0] olacak şekilde 15 tane değer, b parametresi için normal dağılımdan N(-0,5, 1,5) olacak şekilde 15 tane değer ve c parametresi için normal dağılımdan N(0,20, 0,05) olacak şekilde 15 tane değer seçkisiz olarak çekilmiştir. Bu değerler bilinen madde parametreleri olarak ele alınmış ve ekte sunulmuştur (Bakınız EK 1). Bu bağlamda 1. ve 3. simülasyon koşulları için, b parametresi için çekilen 15 tane seçkisiz değer kullanılarak 1PLM altında; 5. ve 7. simülasyon koşulları için, b ve a parametreleri için çekilen seçkisiz değerler kullanılarak 2PLM altında ve 9. ve 11. simülasyon koşulları için ise b, a, c parametreleri için çekilen seçkisiz değerler kullanılarak 3PLM altında veri setleri üretilmiştir. Ardından 2. ve 4. simülasyon koşulları için, b parametresi için çekilen 30 tane seçkisiz değer kullanılarak 1PLM altında; 6. ve 8. simülasyon koşulları için, b ve a parametreleri için çekilen seçkisiz değerler kullanılarak 2PLM altında ve 10. ve 12. simülasyon koşulları için ise b, a, c parametreleri için çekilen seçkisiz değerler kullanılarak 3PLM altında veri setleri üretilmiştir.

Ölçüm Güvenirliği

Bu çalışmada, 12 simülasyon koşulu için üretilen veri setlerinin her birine yönelik olarak, MTK'ye dayalı olarak hesaplanan marjinal güvenirlik katsayıları hesaplanarak güvenirlik kanıtı olarak değerlendirilmiştir. Yüksek marjinal güvenirlik değeri, test edilen modellerden elde edilen puanların güvenirliğini yüksek olduğunu göstermektedir (Md Desa, 2012).

Tablo 4'te ilgili simülasyon koşullarında üretilen veri setleri için hesaplanan marjinal güvenirlik katsayıları verilmiştir.

Tablo 4.

Marjinal Güvenirlik Katsayıları

Simülasyon Koşulu	Marjinal Güvenirlik Katsayısı
1	0,715
2	0,826
3	0,722
4	0,829
5	0,700
6	0,842
7	0,702
8	0,839
9	0,636
10	0,798
11	0,647
12	0,785

Bu çalışmada her bir simülasyon koşulunda üretilen veri seti, bir ölçme aracından elde edilen ölçümler olarak ele alındığından, Tablo 4'te sunulan güvenirlik katsayılarının söz konusu ölçümlerin kabul edilebilir düzeyde güvenilir olduklarına işaret ettikleri düşünülmektedir.

Ölçüm Yorumlarının ve Kullanımlarının Geçerliği

Bu araştırmada tek boyutluluğun test edilmesinde kullanılan açımlayıcı faktör analizi sonuçları, aynı zamanda geçerlik kanıtı olarak ele alınmıştır. Alanyazında (Fleck, Poirier-Littre, Guelfi, Bourdel, ve Loo, 1995; Houben, Leeuw, Vlaeyen, Goubert, ve Picavet, 2005; Romera, Delgado-Cohen, Perez, Caballero, ve Gilaberte, 2008) faktör yük değerleri için sınır değerin 0,25 olarak alındığı çalışmalar mevcuttur. Bu kapsamda her bir veri setine yönelik açımlayıcı faktör analizi (AFA) çözümünde faktör yüklerinin 0,25'den büyük olduğu görülmektedir (Bakınız Tablo 5). AFA'dan elde edilen bulgular incelendiğinde, her bir simülasyon koşulunda üretilen veri setlerinin tek boyutlu bir yapıya sahip olduğu gözlenmiş ve bu bulgular yapı geçerliğine ilişkin kanıt olarak değerlendirilmiştir. Bunun yanı sıra; Reise (1990), MTK kapsamında düşük madde uyum indekslerinin zayıf geçerlik göstergesi olduğunu belirtmektedir. Veri setleri kendi üretildikleri model altında test edildiğinde elde edilen madde uyum düzeyleri, geçerliğe ilişkin kanıt olarak ele alınmıştır (Bakınız EK 4).

Verilerin Analizi

Varsayımların Test Edilmesi

Çalışmada yer alan veri setleri için öncelikle MTK varsayımlarının karşılanıp karşılanmadığı incelenmiştir.

Tek Boyutluluk

MTK'nin varsayımlarından biri olan tek boyutluluk, MTK modelleri için tek bir örtük özellik olduğunu varsaymaktadır. Tek boyutluluk varsayımını test etmek üzere açımlayıcı faktör analizi (AFA) yapılabilmektedir (Hambleton ve Swaminathan, 1985). Analiz sonucuna göre maddeler tek bir faktöre yükleniyorsa, tek boyutluluk varsayımının karşılandığı kabul edilir. Bu çalışmada tek boyutluluk varsayımını test etmek için her bir veri seti üzerinde tetrakorik korelasyon matrisine dayalı açımlayıcı faktör analizi yapılmıştır. Faktör analizi için R Studio programında *psych* paketi kullanılmıştır (Revelle, 2015). Üretilen her bir veri seti için uygulanan açımlayıcı faktör analizine (AFA) ilişkin bulgular Tablo 5'de verilmiştir.

Tablo 5.

Üretilen Veri Setlerine İlişkin AFA Sonuçları

Maddeler	Faktör Yükleri											
	Simülasyon Koşulları											
	1	2	3	4	5	6	7	8	9	10	11	12
m1	0,57	0,49	0,46	0,55	0,75	0,73	0,78	0,69	0,34	0,48	0,36	0,59
m2	0,43	0,50	0,49	0,55	0,34	0,37	0,36	0,35	0,34	0,56	0,45	0,54
m3	0,55	0,45	0,53	0,49	0,59	0,49	0,44	0,43	0,45	0,62	0,52	0,60
m4	0,51	0,63	0,50	0,50	0,53	0,40	0,56	0,42	0,39	0,54	0,48	0,56
m5	0,54	0,49	0,51	0,51	0,70	0,67	0,83	0,74	0,45	0,39	0,40	0,38
m6	0,49	0,52	0,52	0,50	0,75	0,76	0,69	0,77	0,66	0,69	0,57	0,63
m7	0,60	0,43	0,53	0,56	0,60	0,51	0,63	0,61	0,43	0,60	0,48	0,59
m8	0,63	0,42	0,50	0,49	0,49	0,52	0,50	0,49	0,55	0,67	0,48	0,62
m9	0,49	0,50	0,55	0,49	0,66	0,64	0,58	0,71	0,30	0,28	0,44	0,33
m10	0,47	0,43	0,51	0,48	0,65	0,60	0,65	0,66	0,36	0,62	0,40	0,62
m11	0,45	0,48	0,47	0,54	0,39	0,34	0,30	0,32	0,30	0,31	0,39	0,37
m12	0,48	0,51	0,54	0,58	0,79	0,47	0,67	0,51	0,32	0,40	0,30	0,27
m13	0,43	0,54	0,56	0,51	0,53	0,57	0,51	0,59	0,52	0,45	0,52	0,50
m14	0,49	0,56	0,51	0,59	0,33	0,42	0,34	0,39	0,57	0,45	0,37	0,37
m15	0,57	0,51	0,52	0,50	0,56	0,65	0,62	0,60	0,53	0,61	0,54	0,60
m16	-	0,44	-	0,50	-	0,61	-	0,54	-	0,65	-	0,57
m17	-	0,58	-	0,52	-	0,42	-	0,49	-	0,48	-	0,37
m18	-	0,37	-	0,39	-	0,76	-	0,65	-	0,55	-	0,56
m19	-	0,53	-	0,56	-	0,69	-	0,70	-	0,58	-	0,48
m20	-	0,51	-	0,46	-	0,67	-	0,64	-	0,31	-	0,29
m21	-	0,43	-	0,43	-	0,55	-	0,55	-	0,28	-	0,37
m22	-	0,51	-	0,59	-	0,72	-	0,73	-	0,64	-	0,56
m23	-	0,59	-	0,48	-	0,55	-	0,68	-	0,59	-	0,66
m24	-	0,54	-	0,39	-	0,46	-	0,53	-	0,45	-	0,41
m25	-	0,50	-	0,46	-	0,70	-	0,51	-	0,55	-	0,46
m26	-	0,51	-	0,48	-	0,42	-	0,38	-	0,46	-	0,40
m27	-	0,53	-	0,59	-	0,67	-	0,71	-	0,34	-	0,53
m28	-	0,50	-	0,50	-	0,37	-	0,40	-	0,31	-	0,34
m29	-	0,54	-	0,48	-	0,45	-	0,39	-	0,54	-	0,49
m30	-	0,59	-	0,55	-	0,32	-	0,31	-	0,62	-	0,54
öz değer	4,01	7,73	3,96	7,83	5,29	9,63	5,10	9,57	3,0	7,97	3,06	7,48
açıklanan varyans(%)	27	26	26	26	35	32	34	32	20	27	20	25

Faktör yük değerleri için sınır değer 0,30 olarak alınmıştır (Brown, 2015). Elde edilen sonuçlar incelendiğinde, 10. simülasyon koşulunda üretilen veri setinde yer alan iki madde (m9 ve m21) ile 12. simülasyon koşulunda üretilen veri setinde yer alan iki madde (m12 ve m20) hariç hepsinin 0,30 ya da 0,30'dan büyük değerlere sahip olduğu görülmektedir. Bu maddelere ait faktör yükleri ise 0,25 ve üzerindedir ve alanyazında faktör yük değerleri için sınır değerin 0,25 olarak alındığı görülmektedir (Fleck, Poirier-Littre, Guelfi, Bourdel, ve Loo, 1995; Houben, Leeuw, Vlaeyen, Goubert, ve Picavet, 2005; Romera, Delgado-Cohen,

Perez, Caballero, ve Gilaberte, 2008). Bu doğrultuda tüm maddelerin ilgili faktörü yeterince temsil edebildiği sonucuna ulaşılmıştır. Veri setlerine ait maddelerin çoğunluğuna ait faktör yüklerinin 0,30'dan büyük olması veri setlerinin tek boyutluluk varsayımını karşıladığını göstermektedir.

Yerel Bağımsızlık

Alan yazında tek boyutluluk varsayımının karşılanmış olmasının, yerel bağımsızlık için kanıt olarak sunulabildiği belirtilmektedir (Embretson ve Reise, 2000; Hambleton vd., 1991). Ek olarak, Yen (1984) tarafından geliştirilen Q_3 istatistiği kullanılarak da yerel bağımsızlık varsayımı test edilebilmektedir. Yerel madde bağımsızlığı indeksi olarak ortaya konulan Q_3 istatistiği, her bir madde çifti arasında hesaplanan artık değerlerin korelasyonudur. Bu indeksi hesaplamak için öncelikle her bir yanıtlayıcı için yetenek kestirimi yapılır ve sonuçlar yanıtlayıcının her bir maddedeki beklenen performansını kestirmek için kullanılır. Artık değeri hesaplamak için ise yanıtlayıcının bir maddedeki beklenen ve gözlenen performansı arasındaki sapma hesaplanmalıdır. Q_3 istatistiği testte yer alan her bir madde çifti için elde edilmektedir. Sonuç olarak seçilen madde çiftinin Q_3 istatistiği, sapma puanlarının tüm yanıtlayıcılardaki korelasyonu olarak belirtilir. Bu indeksin değerinin 0,20'den küçük olması, ilgili madde çifti için yerel bağımsızlık varsayımının sağlandığına işaret etmektedir (Chen ve Thissen, 1997; DeAyala, 2009). Bu doğrultuda, üretilen veri setleri için yerel bağımsızlık varsayımının karşılanıp karşılanmadığının test edilmesi amacıyla 1PLM, 2PLM ve 3PLM altında bütün madde çiftlerine yönelik Q_3 istatistiği hesaplanmıştır. Q_3 istatistiğinin hesaplanması amacıyla R Studio programı *sirt* paketinden (Robitzsch, 2020) faydalanılmıştır ve her bir madde çifti için bu varsayımın karşılandığı görülmüştür. Hesaplanan Q_3 istatistikleri, ekte tablo halinde sunulmuştur (Bakınız EK 2).

MKE'nin Monotonik Artış Göstermesi

MTK'nin varsayımlarından bir diğeri de madde karakteristik eğrisinin monotonik olarak artış göstermesidir. Bu varsayım, yetenek düzeyi arttıkça bireyin maddeyi doğru yanıtlama olasılığının artmasını ifade etmektedir (Hambleton vd., 1991; Hambleton ve Swaminathan, 1985). Bu varsayım için, her bir simülasyon koşuluna dayalı olarak üretilen veri setlerinde yer alan her madde için madde karakteristik eğrisi elde edilerek incelenmiştir. Her bir

simülasyon koşuluna dayalı olarak elde edilen madde karakteristik eğrileri ekte sunulmuştur (Bakınız EK 3). Bu incelemelere dayalı olarak her bir madde için örtük değişken ile madde yanıtları arasında doğrusal bir ilişki olduğu, yani yetenek düzeyi arttıkça maddeyi doğru yanıtlama olasılığının da arttığı gözlenmiştir.

Model Veri Uyumu

Model uyumunun yeterli düzeyde olması, MTK'nin etkili ve geçerli sonuçlar vermesi için gerekli olan bir özelliktir. MTK'de model uyumu, her bir model için genel model uyumu ve madde düzeyinde uyum olmak üzere iki şekilde incelenebilmektedir. Bu çalışmada model-veri uyumu, her bir model için genel model-veri uyumu olarak ve madde düzeyinde olacak şekilde ele alınmıştır. Model-veri uyumuna ilişkin istatistikler, R studio programında *mirt* paketi (Chalmers, 2020) kullanılarak elde edilmiştir. Her bir simülasyon koşulunda üretilen veri setleri öncelikle veri seti hangi MTK modeline dayalı olarak üretilmiş ise o MTK modeli altında test edilmiş ve uyum istatistikleri (Comparative Fit Index - CFI, Tucker Lewis Index - TLI) ve hata değerleri (Root Mean Square Error of Approximation - RMSEA, Standardized Root Mean Square Residual - SRMR) elde edilmiştir. Bu araştırmada ayrıca, hesaplanan sınıflama doğruluğu ve tutarlılığı indekslerinin model yanlış tanımlamasına duyarlılığının da incelenmesi amaçlandığından her bir veri seti, veri üretiminde temel alınan MTK modelinden farklı diğer iki MTK modeli altında da test edilmiştir. Araştırma kapsamında her bir simülasyon koşulunda üretilen veri setleri 1PLM, 2PLM ve 3PLM altında test edildiğinde elde edilen uyum istatistikleri (CFI, TLI) ve hata değerleri (RMSEA, SRMR) Tablo 6'da verilmiştir.

Tablo 6.

Üretilen Veri Setlerinin Farklı Modeller Altında Test Edilmesine İlişkin Elde Edilen Uyum İstatistikleri ve Hata Değerleri

Simülasyon Koşulu	Test Edilen Model	RMSEA	SRMR	TLI	CFI
1	1PLM	0,008	0,042	0,996	0,996
1	2PLM	0,101	0,118	0,506	0,577
1	3PLM	0,081	0,115	0,683	0,773
2	1PLM	0,008	0,042	0,995	0,995
2	2PLM	0,109	0,119	0,358	0,402
2	3PLM	*0,039*	0,107	*0,915*	*0,927*
3	1PLM	0	0,028	1,00	1,00
3	2PLM	0,112	0,113	0,414	0,498
3	3PLM	*0,052*	0,106	0,874	*0,910*
4	1PLM	0,004	0,031	0,999	0,999
4	2PLM	0,123	0,101	0,217	0,271
4	3PLM	*0,023*	0,099	*0,972*	*0,976*
5	1PLM	0,190	0,123	-0,494	0
5	2PLM	0,008	0,041	0,997	0,997
5	3PLM	*0,027*	0,097	*0,969*	*0,978*
6	1PLM	0,267	0,132	-1,899	0
6	2PLM	0,011	0,037	0,994	0,995
6	3PLM	0,082	0,139	0,726	0,763
7	1PLM	0,197	0,123	-0,539	0
7	2PLM	0,006	0,029	0,998	0,999
7	3PLM	*0,008*	0,095	*0,997*	*0,998*
8	1PLM	0,265	0,128	-1,871	0
8	2PLM	0	0,024	1,00	1,00
8	3PLM	*0.046*	0,127	*0,910*	*0,923*
9	1PLM	0,209	*0,074*	-2,844	0
9	2PLM	*0.060*	0,085	0,675	0,721
9	3PLM	0	0,030	1,00	1,00
10	1PLM	0,120	0,099	0,203	0,203
10	2PLM	*0,055*	0,109	0,828	0,840
10	3PLM	0,011	0,041	0,993	0,993
11	1PLM	0,221	*0,069*	-3,112	0
11	2PLM	*0,079*	0,127	0,475	0,550
11	3PLM	0	0,023	1,00	1,00
12	1PLM	0,108	0,083	0,232	0,232
12	2PLM	*0,030*	0,095	*0,940*	*0,945*
12	3PLM	0,005	0,029	0,998	0,998

RMSEA ve SRMR değerlerinin sıfıra (0) yakın olması, daha iyi model-veri uyumuna işaret etmektedir (De Ayala, 2009; Maydeu-Olivares, Cai, ve Hernández, 2011; Maydeu-Olivares ve Joe, 2005, 2006, 2014). RMSEA ve SRMR değerlerinin 0,05'in altında olması iyi uyuma,

0,08'in altında olması ise kabul edilebilir düzeyde model uyumuna işaret etmektedir (Kline, 2015). CFI ve TLI değerlerinin ise 0,90'dan büyük olmasının kabul edilebilir düzeyde model uyumuna, 0,95'den büyük olmasının da iyi uyuma işaret ettiği kabul edilmektedir (Hu & Bentler, 1999; Kline, 2015). Genel model uyumu bazında incelendiğinde, tüm veri setlerinin kendi üretildiği model altında test edildiğinde modelin veriye daha iyi uyum gösterdiği gözlenmiştir. Ayrıca Tablo 6'da italik yazılan değerler, ilgili veri setinin, kendi üretildiği model dışında farklı bir model altında test edilmesi durumunda modele yeterli veya kabul edilebilir düzeyde uyum sergilediğine işaret etmektedir. Bu doğrultuda tablodaki değerler incelenecek olursa; 1, 2, 3 ve 4. simülasyon koşullarında üretilen veri setleri 1PLM'e dayalı olarak üretildikleri için bu veri setlerinin 2PLM ve 3PLM altında test edilmeleri model yanlış tanımlaması durumuna işaret etmektedir. Söz konusu veri setleri 2PLM altında test edildiğinde incelenen tüm uyum istatistikleri ve hata değerlerinin kötü/yetersiz uyuma işaret ettiği görülmektedir. Bu veri setleri 3PLM altında test edildiğinde ise 2, 3 ve 4. simülasyon koşullarında üretilen veri setlerine ait RMSEA ve CFI değerleri ile 2. ve 4. simülasyon koşullarında üretilen veri setlerine ait TLI değerlerinin kabul edilebilir düzeyde, yanlış model uyumuna işaret ettiği gözlenmiştir. Tablodaki değerler dikkate alındığında; 5, 6, 7 ve 8. simülasyon koşullarında üretilen veri setleri 2PLM altında üretildikleri için bu veri setlerinin 1PLM ve 3PLM altında test edilmeleri model yanlış tanımlaması olarak ele alınmaktadır. Bu veri setleri 1PLM altında test edildiğinde incelenen tüm uyum istatistikleri ve hata değerlerinin kötü/yetersiz uyuma işaret ettiği görülmektedir. Aynı veri setleri, 3PLM altında test edildiğinde ise 5, 7 ve 8. simülasyon koşullarında üretilen veri setlerine ilişkin RMSEA, TLI ve CFI değerleri, veri setlerinin kabul edilebilir düzeyde, yanlış modele uyum sergilediğine işaret etmektedir. Aynı şekilde, 9, 10, 11 ve 12. simülasyon koşullarında üretilen veri setlerine baktığımızda 3PLM altında üretildikleri için 1PLM ve 2PLM altında test edilmeleri bu veri setleri için model yanlış tanımlaması durumuna işaret etmektedir. Bu veri setleri 1PLM altında test edildiğinde sadece 9. ve 11. simülasyon koşullarında üretilen veri setlerine ait SRMR değerlerinin; 2PLM altında test edildiğinde ise 9, 10, 11 ve 12. simülasyon koşullarında üretilen veri setlerine ait RMSEA değerlerinin ve 12. simülasyon koşulunda üretilen veri setine ait TLI ve CFI değerlerinin kabul edilebilir düzeyde, yanlış modele uyum sergilediğine işaret ettikleri gözlenmiştir.

Elde edilen bulgular, genel model uyumu kapsamında 3PLM'in yanlış model olarak tanımlandığı durumlarda, bazı uyum istatistikleri ve hata değerleri için kabul edilebilir düzeyde veriye uyum sergilediğine işaret etmektedir.

Genel model veri uyumunun dışında, madde düzeyindeki uyum istatistikleri Darrel Bock'un (1972) χ^2 indeksi kullanılarak hesaplanmıştır. Madde düzeyindeki uyumu değerlendirmek için R studio programında *ltm* paketi (Rizopoulos, 2018) kullanılmıştır. 1PLM, 2PLM ve 3PLM için ilgili uyum değerleri ekte tablo halinde sunulmuştur (Bakınız EK 4). Madde uyumu kapsamında hesaplanan χ^2 değerinin manidar olması ($p<.001$), maddelerin söz konusu modele iyi uyum sergilemediğini göstermektedir. Öncelikle, üretilen tüm veri setlerinin kendi üretildiği model altında test edilmesi durumunda tüm maddelerin uyum sergilediği gözlenmiştir. Elde edilen değerlere bakıldığında; 1, 2 ve 4. simülasyon koşullarında üretilen veri setlerinin 2PLM ve 3PLM altında test edildiğinde de tüm maddelerin söz konusu modellere uyum sergilediği görülmüştür. 1PLM altında ve 3. simülasyon koşulunda üretilen veri setinde yer alan madde 9 ve madde 12'nin 2PLM ve 3PLM altında, madde 13 ve madde 14'ün ise 3PLM altında uyum sergilemediği görülmektedir. 2PLM altında ve 5. simülasyon koşulunda üretilen veri setinde yer alan 6 maddenin (1, 3, 4, 5, 10 ve 15); 6. simülasyon koşulunda üretilen veri setinde yer alan 4 maddenin (1, 6, 20 ve 22); 7. simülasyon koşulunda üretilen veri setinde yer alan 5 maddenin (1, 4, 5, 10 ve 15) ve 8. simülasyon koşulunda üretilen veri setinde yer alan 10 maddenin (1, 5, 6, 9, 10, 19, 20, 22, 23 ve 27) 1PLM altında uyum sergilemediği görülürken 7. simülasyon koşulunda üretilen veri setinde yer alan 5 maddenin (1, 3, 4, 5 ve 10) 3 PLM altında uyum sergilemediği görülmektedir. 3PLM altında ve 9. simülasyon koşulunda üretilen veri setinde yer alan madde 6'nın 1PLM ve 2PLM altında uyum sergilemediği; 10. simülasyon koşulunda üretilen veri setinde yer alan madde 8 ve madde 30'un 1PLM altında uyum sergilemediği; 11. simülasyon koşulunda üretilen veri setinde yer alan 6 maddenin (2, 3, 4, 6, 13 ve 15) 1PLM altında, 7 maddenin (3, 4, 6, 8, 9, 13 ve 15) 2PLM altında uyum sergilemediği; 12. simülasyon koşulunda üretilen veri setinde yer alan 7 maddenin (3, 6, 8, 10, 18, 23 ve 30) 1PLM altında uyum sergilemediği görülmektedir. Elde edilen bulgular, 1PLM, 2PLM ve 3PLM altında üretilen veri setlerinde yer alan maddelerin büyük bir kısmının, kendisinden daha fazla parametrenin kestirildiği model altında da iyi uyum sergilediğine işaret etmektedir.

BÖLÜM III

BULGULAR, SONUÇ VE TARTIŞMA

Bulgular ve Yorumlar

Bu araştırma kapsamında öncelikli olarak araştırma amacı doğrultusunda her bir simülasyon koşulunda üretilen veri setleri hangi model altında üretildiyse o model altında test edilmiştir. Buna dayalı olarak, farklı simülasyon koşullarında ve Rudner'ın, Guo'nun ve Lee'nin yöntemlerine dayalı olarak CA ve CC indeksleri ile bunlara ilişkin mutlak yanlılık değerleri hesaplanmıştır. Ardından söz konusu veri setleri model yanlış tanımlaması durumunda test edilmiş ve buna dayalı olarak CA ve CC indeksleri ile bunlara ilişkin mutlak yanlılık değerleri hesaplanmıştır. Hesaplanan CA ve CC indekslerine ilişkin değerler ve bu değerlere ilişkin mutlak yanlılık değerleri ekte sunulmuştur (Bakınız EK 5, EK 6, EK 7 ve EK 8). Aşağıda bulgular, araştırma soruları bağlamında sunularak yorumlanmıştır.

1. 1PLM altında üretilen veri 1PLM altında test edildiğinde;
    a. Her bir yöntem altında kestirilen sınıflama doğruluğu ve tutarlılığı indeksleri test uzunluğuna bağlı olarak farklılaşmakta mıdır?

1PLM altında üretilen N=500 veri setinin 1PLM altında test edilmesine ilişkin bulgular incelendiğinde; her bir kesme puanı değerinde, madde sayısının artışına bağlı olarak, her bir yöntem altında hesaplanan genel CA ve CC oranları ile kappa değerlerinin arttığı, Rudner'ın yöntemine dayalı hesaplanan FP ve FN oranlarının ise her bir kesme puanı değeri için azaldığı gözlenmiştir. Lee'nin ve Guo'nun yöntemlerine dayalı hesaplanan FP ve FN oranlarının, kesme puanına bağlı olarak farklılaştığı ancak sistematik bir farklılaşmanın bulunmadığı anlaşılmaktadır. Lee'nin yöntemine dayalı hesaplanan FP oranları, madde sayısının artışına bağlı olarak, her bir kesme puanı değeri için düşüş gösterirken FN oranları -0,75 ve 0 kesme puanlarında düşüş, 0 kesme puanında artış göstermiştir. Guo'nun

yöntemine dayalı hesaplanan FP oranları madde sayısının artışına bağlı olarak -0,75 kesme puanında artış, 0 ve 0,75 kesme puanlarında azalış gösterirken; FN oranları -0,75 kesme puanında azalış, 0 ve 0,75 kesme puanında artış göstermiştir.

1PLM altında üretilen N=1000 veri setinin 1PLM altında test edilmesine ilişkin bulgular incelendiğinde; her bir kesme puanı değeri için, madde sayısının artışına bağlı olarak, her bir yöntem altında hesaplanan genel CA ve CC oranları ile kappa değerlerinin artış gösterdiği dikkat çekmektedir. Her bir yönteme dayalı hesaplanan FN oranlarının, madde sayısının artışına bağlı olarak, -0,75 ve 0 kesme puanlarında azalış gösterirken 0,75 kesme puanında artış gösterdiği gözlenmiştir. Kesme puanının -0,75 olduğu durumda Lee'nin ve Rudner'ın yöntemine dayalı hesaplanan FP oranlarının, madde sayısının artışına bağlı olarak azalış gösterdiği, Guo'nun yöntemine dayalı hesaplanan FP oranlarının ise artış gösterdiği görülmektedir. Kesme puanının 0 olduğu durumda ise Lee'nin yöntemine dayalı hesaplanan FP oranı, madde sayısının artışına bağlı olarak artış gösterirken; Rudner'ın ve Guo'nun yöntemine dayalı hesaplanan FP oranları azalış göstermiştir. Son olarak, kesme puanı değerinin 0,75 olduğu durumda her bir yöntem altında hesaplanan FP oranlarının madde sayısının artışına bağlı olarak azalış gösterdiği gözlenmiştir.

Elde edilen bulgular kesme puanları dikkate alınmadan incelendiğinde; Rudner'ın yöntemine dayalı hesaplanan genel CA ve CC oranları, kappa değerleri ve FP oranları, Guo'nun yöntemine dayalı hesaplanan genel CA ve CC oranları ve kappa değerleri ile Lee'nin yöntemine dayalı hesaplanan kappa değerleri için madde sayısının artışına bağlı olarak dikkate değer farklılıkların olduğu dikkat çekmektedir. Bu farklılıklar, genel CA ve CC oranları ile kappa değerlerinde artış, FP oranlarında azalış olarak gözlenmektedir.

b. Her bir yöntem altında kestirilen sınıflama doğruluğu ve tutarlılığı indeksleri örneklem büyüklüğüne bağlı olarak farklılaşmakta mıdır?

1PLM altında üretilen K=15 veri setinin 1PLM altında test edilmesine ilişkin bulgular incelendiğinde; örneklem büyüklüğünün artışına bağlı olarak, her bir yöntem altında hesaplanan genel CA ve CC oranlarının -0,75 kesme puanında azalış, 0 ve 0,75 kesme puanlarında artış gösterdiği gözlenmiştir. Rudner'ın, Guo'nun ve Lee'nin yöntemine dayalı hesaplanan FP, FN oranları ve kappa değerlerinin ise kesme puanına bağlı olarak farklılaştığı ancak sistematik bir farklılaşmanın bulunmadığı anlaşılmaktadır. Lee'nin ve Rudner'ın yöntemine dayalı hesaplanan kappa değerleri, örneklem büyüklüğünün artışına bağlı olarak; -0,75 kesme puanı için azalış, 0 ve 0,75 kesme puanları için artış gösterirken, Guo'nun

yöntemine dayalı hesaplanan kappa değerleri her bir kesme puanında artış göstermiştir. Lee'nin yöntemine dayalı hesaplanan FP oranları örneklem büyüklüğünün artışına bağlı olarak; her bir kesme puanı değerinde azalış gösterirken, Rudner'ın ve Guo'nun yöntemine dayalı hesaplanan FP oranları 0 kesme puanında artış, 0,75 kesme puanında azalış göstermiştir. Kesme puanının -0,75 olduğu durumda ise örneklem büyüklüğünün artışına bağlı olarak; Rudner'ın yöntemine dayalı hesaplanan FP oranında düşüş gözlenirken, Guo'nun yöntemine dayalı hesaplanan FP oranında ise değişim gözlenmemiştir. Rudner'ın ve Guo'nun yöntemine dayalı hesaplanan FN oranlarında, örneklem büyüklüğündeki artışa bağlı olarak; -0,75 kesme puanı değerinde artış, 0 ve 0,75 kesme puanı değerlerinde ise azalış görülmektedir. Lee'nin yöntemine dayalı hesaplanan FN oranlarında örneklem büyüklüğünün artışına bağlı olarak; -0,75 ve 0,75 kesme puanı değerlerinde artış, 0 kesme puanı değerinde azalış gözlenmiştir.

1PLM altında üretilen K=30 veri setinin 1PLM altında test edilmesine ilişkin bulgular incelendiğinde; örneklem büyüklüğünün artışına bağlı olarak, her bir yöntem altında hesaplanan genel CC oranları ve kappa değerlerinin -0,75 kesme puanında azalış gösterirken, 0 kesme puanında artış gösterdiği göze çarpmaktadır. Kesme puanının 0,75 olduğu durumda, örneklem büyüklüğündeki artışa bağlı olarak, Lee'nin ve Rudner'ın yöntemine dayalı hesaplanan genel CC oranlarının azalış gösterirken; Guo'nun yöntemine dayalı hesaplanan genel CC oranının ise artış gösterdiği gözlenmiştir. Örneklem büyüklüğünün artışına bağlı olarak, her bir yöntem altında hesaplanan kappa değerlerinin 0,75 kesme puanı değeri için arttığı bulunmuştur. Her bir yönteme dayalı hesaplanan genel CA oranlarının, örneklem büyüklüğünün artışına bağlı olarak -0,75 kesme puanında azalış gösterirken, 0 kesme puanında artış gösterdiği gözlenmiştir. Kesme puanının 0,75 olduğu durumda ise, örneklem büyüklüğündeki artışa bağlı olarak, Lee'nin yöntemine dayalı hesaplanan genel CA oranının azalırken, Rudner'ın ve Guo'nun yöntemine dayalı hesaplanan genel CA oranlarının artış gösterdiği gözlenmiştir. Rudner'ın ve Guo'nun yöntemine dayalı hesaplanan FP oranları, örneklem büyüklüğündeki artışa bağlı olarak; -0,75 ve 0,75 kesme puanlarında azalırken, 0 kesme puanında artış göstermektedir. Lee'nin yöntemine dayalı hesaplanan FP oranının ise örneklem büyüklüğündeki artışa bağlı olarak -0,75 kesme puanı değeri için değişim göstermediği, 0 ve 0,75 kesme puanı değerleri için ise artış gösterdiği gözlenmiştir. Örneklem büyüklüğündeki artışa bağlı olarak, her bir yöntem altında hesaplanan FN oranlarının -0,75 kesme puanı değeri için artış gösterirken, 0 kesme puanı değeri için azalış gösterdiği görülmektedir. Kesme puanının 0,75 olduğu durumda ise,

örneklem büyüklüğündeki artışa bağlı olarak, Lee'nin yöntemine dayalı hesaplanan FN oranının azaldığı, Rudner'ın ve Guo'nun yöntemine dayalı hesaplanan FN oranlarının arttığı gözlenmiştir.

Elde edilen bulgular kesme puanları dikkate alınmadan incelendiğinde; Rudner'ın, Guo'nun ve Lee'nin yöntemine dayalı hesaplanan genel CA ve CC oranları, kappa değerleri, FP ve FN oranları için örneklem büyüklüğünün artışına bağlı olarak dikkate değer bir farklılaşmanın olmadığı sonucuna ulaşılmıştır.

c. Test uzunluğuna bağlı olarak, farklı yöntemlerle kestirilen sınıflama doğruluğu ve tutarlılığı indekslerinin doğruluk derecesi farklılaşmakta mıdır? Hangi yöntem daha düşük bir yanlılık değeri ile sonuçlanmaktadır?

1PLM altında üretilen N=500 veri setinin 1PLM altında test edilmesine ilişkin bulgular incelendiğinde; her bir kesme puanı için, madde sayısının artışına bağlı olarak, her bir yöntem altında hesaplanan genel CC oranları ve kappa değerlerine ilişkin mutlak yanlılık değerlerinin azaldığı gözlenmiştir. Her bir kesme puanı için, madde sayısındaki artışa bağlı olarak, Lee'nin ve Guo'nun yöntemine dayalı hesaplanan genel CA oranlarına ilişkin mutlak yanlılık değerlerinin azaldığı görülmüştür. Rudner'ın yöntemine dayalı hesaplanan genel CA oranına ilişkin mutlak yanlılık değerlerinin, madde sayısındaki artışa bağlı olarak, -0,75 kesme puanında artış gösterdiği, 0 kesme puanında değişim göstermediği, 0,75 kesme puanında ise azaldığı görülmektedir. Rudner'ın, Guo'nun ve Lee'nin yöntemine dayalı hesaplanan FP ve FN oranlarına ilişkin mutlak yanlılık değerlerinin ise kesme puanına bağlı olarak farklılaştığı ancak sistematik bir farklılaşmanın bulunmadığı görülmüştür. Rudner'ın ve Guo'nun yöntemine dayalı hesaplanan FP ve FN oranlarına ilişkin mutlak yanlık değerlerinin, madde sayısındaki artışa bağlı olarak; -0,75 ve 0,75 kesme puanlarında azalış gösterirken, 0 kesme puanında artış gösterdiği sonucuna ulaşılmıştır. Lee'nin yöntemine dayalı hesaplanan FN oranlarına ilişkin mutlak yanlılık değerleri her bir kesme puanı için, madde sayısındaki artışa bağlı olarak, azalış gösterirken FP oranlarına ilişkin mutlak yanlılık değerleri ise -0,75 kesme puanı için artış, 0 ve 0,75 kesme puanı değerleri için azalış göstermiştir.

1PLM altında üretilen N=1000 veri setinin 1PLM altında test edilmesine ilişkin bulgular incelendiğinde; her bir kesme puanı değeri için, madde sayısındaki artışa bağlı olarak, her bir yöntem altında hesaplanan genel CC oranları, kappa değerleri ve FN oranlarına ilişkin mutlak yanlılık değerlerinin azaldığı dikkat çekmektedir. Her bir yöntem altında hesaplanan

genel CA ve CC oranlarına ilişkin mutlak yanlılık değerlerinin, kesme puanına bağlı olarak farklılaştığı ancak sistematik bir farklılaşmanın bulunmadığı gözlenmiştir. Lee'nin ve Rudner'ın yöntemine dayalı hesaplanan genel CA oranlarına ilişkin mutlak yanlılık değerlerinin, madde sayısındaki artışa bağlı olarak; -0,75 ve 0 kesme puanlarında azalırken; 0,75 kesme puanında arttığı gözlenmiştir. Guo'nun yöntemine dayalı hesaplanan genel CA oranına ilişkin mutlak yanlılık değerlerinin ise her bir kesme puanı değeri için, madde sayısının artışına bağlı olarak, azaldığı görülmektedir. Rudner'ın ve Guo'nun yöntemine dayalı hesaplanan FP değerlerine ilişkin mutlak yanlılık değerlerinin, madde sayısının artışına bağlı olarak; -0,75 ve 0,75 kesme puanlarında azalış, 0 kesme puanında artış gösterdiği görülmektedir. Lee'nin yöntemine dayalı hesaplanan FP oranlarına ilişkin mutlak yanlılık değerleri ise her bir kesme puanı değeri için, madde sayısının artışına bağlı olarak, artış göstermiştir.

Elde edilen bulgular kesme puanları dikkate alınmadan incelendiğinde; Rudner'ın ve Guo'nun yöntemine dayalı hesaplanan genel CC oranları ve kappa değerlerine ilişkin mutlak yanlılık değerleri ile Lee'nin yöntemine dayalı hesaplanan kappa değerleri ve FN oranlarına ilişkin mutlak yanlılık değerleri için, madde sayısının artışına bağlı olarak, dikkate değer farklılıkların olduğu göze çarpmaktadır. Bu farklılıklar, söz konusu değerlere ilişkin mutlak yanlılık değerlerinde azalış olarak gözlenmiştir.

1PLM altında üretilen K=15 veri setinin 1PLM altında test edilmesine ilişkin bulgular kesme puanı dikkate alınmaksızın incelendiğinde; Rudner'ın yöntemine dayalı hesaplanan genel CA oranı, kappa değeri ve FP oranı, Lee'nin yöntemine dayalı hesaplanan genel CC oranı ve FP oranı ile Guo'nun yöntemine dayalı hesaplanan FN oranının daha düşük mutlak yanlılık değerine sahip olduğu bulunmuştur.

1PLM altında üretilen K=30 veri setinin 1PLM altında test edilmesine ilişkin bulgular kesme puanı dikkate alınmaksızın incelendiğinde; Lee'nin yöntemine dayalı hesaplanan genel CA, FP ve FN oranları ile Rudner'ın yöntemine dayalı hesaplanan genel CC oranı ve kappa değerinin daha düşük mutlak yanlılık değerine sahip olduğu görülmüştür.

d. Örneklem büyüklüğüne bağlı olarak, farklı yöntemlerle kestirilen sınıflama doğruluğu ve tutarlılığı indekslerinin doğruluk derecesi farklılaşmakta mıdır? Hangi yöntem daha düşük bir yanlılık değeri ile sonuçlanmaktadır?

1PLM altında üretilen K=15 veri setinin 1PLM altında test edilmesine ilişkin bulgular incelendiğinde; örneklem büyüklüğünün artışına bağlı olarak, her bir yöntem altında

hesaplanan genel CC oranları ve kappa değerlerine ilişkin mutlak yanlılık değerleri -0,75 kesme puanında artış gösterirken, 0 ve 0,75 kesme puanlarında azalış göstermiştir. Her bir yöntem altında hesaplanan genel CA, FP ve FN oranlarına ilişkin mutlak yanlılık değerlerinin, kesme puanına bağlı olarak farklılaştığı ancak sistematik bir farklılaşmanın bulunmadığı anlaşılmaktadır. Lee'nin yöntemine dayalı hesaplanan genel CA ve FP oranlarına ilişkin mutlak yanlılık değerlerinin her bir kesme puanında, örneklem büyüklüğünün artışına bağlı olarak, azaldığı gözlenirken aynı yönteme dayalı hesaplanan FN oranlarına ilişkin mutlak yanlılık değerlerinin -0,75 ve 0 kesme puanlarında azalış; 0,75 kesme puanında artış gösterdiği sonucuna ulaşılmıştır. Rudner'ın ve Guo'nun yöntemine dayalı hesaplanan FN oranlarına ilişkin mutlak yanlılık değerlerinin, örneklem büyüklüğünün artışına bağlı olarak; -0,75 kesme puanında azalırken, 0 ve 0,75 kesme puanlarında arttığı görülmektedir. Rudner'ın yöntemine dayalı hesaplanan genel CA oranına ilişkin mutlak yanlılık değerlerinin, örneklem büyüklüğünün artışına bağlı olarak, -0,75 ve 0 kesme puanlarında arttığı, 0,75 kesme puanında değişim göstermediği; FP oranlarına ilişkin mutlak yanlılık değerlerinin ise -0,75 ve 0 kesme puanlarında azaldığı, 0,75 kesme puanında arttığı gözlenmiştir. Guo'nun yöntemine dayalı hesaplanan genel CA oranına ilişkin mutlak yanlılık değerlerinin ise örneklem büyüklüğünün artışına bağlı olarak -0,75 ve 0,75 kesme puanlarında azalış, 0 kesme puanında artış gösterirken; FP oranlarına ilişkin mutlak yanlılık değerlerinin her bir kesme puanında artış gösterdiği görülmüştür.

1PLM altında üretilen K=30 veri setinin 1PLM altında test edilmesine ilişkin bulgular incelendiğinde, örneklem büyüklüğündeki artışa bağlı olarak, her bir yöntem altında hesaplanan kappa değerlerine ilişkin mutlak yanlılık değerlerinin -0,75 kesme puanında artış, 0 ve 0,75 kesme puanlarında azalış gösterdiği gözlenmiştir. Her bir yöntem altında hesaplanan genel CA, CC, FP ve FN oranlarına ilişkin mutlak yanlılık değerlerinin, kesme puanına bağlı olarak farklılaştığı ancak sistematik bir farklılaşmanın bulunmadığı görülmektedir. Lee'nin ve Rudner'ın yöntemine dayalı hesaplanan genel CC oranlarına ilişkin mutlak yanlılık değerlerinin, örneklem büyüklüğünün artışına bağlı olarak, -0,75 ve 0,75 kesme puanlarında arttığı, 0 kesme puanında azaldığı görülmektedir. Rudner'ın yöntemine dayalı hesaplanan FP ve Guo'nun yöntemine dayalı hesaplanan genel CC, FP ve FN oranlarına ilişkin mutlak yanlılık değerlerinin, örneklem büyüklüğündeki artışa bağlı olarak, -0,75 kesme puanında artış, 0 ve 0,75 kesme puanlarında azalış gösterdiği gözlenmiştir. Lee'nin yöntemine dayalı hesaplanan genel CA oranına ilişkin mutlak yanlılık değerlerinin, örneklem büyüklüğündeki artışa bağlı olarak, -0,75 ve 0 kesme puanlarında

azaldığı, 0,75 kesme puanında arttığı; FP oranlarına ilişkin mutlak yanlılık değerlerinin -0,75 kesme puanında azaldığı, 0 ve 0,75 kesme puanlarında arttığı; FN oranlarına ilişkin mutlak yanlılık değerlerinin ise -0,75 kesme puanında arttığı, 0 ve 0,75 kesme puanlarında azaldığı gözlenmiştir. Rudner'ın yöntemine dayalı hesaplanan genel CA ve FN oranlarına ilişkin mutlak yanlılık değerlerinin, örneklem büyüklüğünün artışına bağlı olarak, -0,75 kesme puanında azalış, 0 kesme puanında artış; 0,75 kesme puanında ise genel CA oranına ilişkin mutlak yanlılık değerlerinin artış, FN oranına ilişkin mutlak yanlılık değerlerinin azalış gösterdiği görülmüştür. Guo'nun yöntemine dayalı hesaplanan genel CA oranına ilişkin mutlak yanlılık değerlerinin ise, örneklem büyüklüğündeki artışa bağlı olarak, -0,75 ve 0 kesme puanlarında artış, 0,75 kesme puanında azalış gösterdiği gözlenmiştir.

Elde edilen bulgular kesme puanları dikkate alınmadan incelendiğinde; örneklem büyüklüğünün artışına bağlı olarak, yalnızca Rudner'ın yöntemine dayalı hesaplanan genel CA oranına ilişkin mutlak yanlılık değeri için dikkate değer bir farklılık görülmüştür. Bu farklılık, söz konusu değere ilişkin mutlak yanlılık değerinde artış olarak göze çarpmaktadır.

1PLM altında üretilen N=500 veri setinin 1PLM altında test edilmesine ilişkin bulgular kesme puanı dikkate alınmaksızın incelendiğinde; Rudner'ın yöntemine dayalı hesaplanan genel CA, CC oranları ve kappa değerleri ile Lee'nin yöntemine dayalı hesaplanan FP ve FN oranlarının daha düşük mutlak yanlılık değerlerine sahip olduğu bulunmuştur.

1PLM altında üretilen N=1000 veri setinin 1PLM altında test edilmesine ilişkin bulgular kesme puanı dikkate alınmaksızın incelendiğinde; Rudner'ın yöntemine dayalı hesaplanan genel CC, FP oranları ve kappa değeri ile Lee'nin yöntemine dayalı hesaplanan genel CA, FP ve FN oranlarının daha düşük mutlak yanlılık değerlerine sahip olduğu gözlenmiştir.

2. 2PLM altında üretilen veri 2PLM altında test edildiğinde;
    a. Her bir yöntem altında kestirilen sınıflama doğruluğu ve tutarlılığı indeksleri test uzunluğuna bağlı olarak farklılaşmakta mıdır?

2PLM altında üretilen N=500 veri setinin 2PLM altında test edilmesine ilişkin bulgular incelendiğinde; her bir kesme puanı değeri için, madde sayısının artışına bağlı olarak, her bir yöntem altında hesaplanan genel CA ve CC oranları ile kappa değerlerinin arttığı, Lee'nin yöntemine dayalı hesaplanan FP ve FN oranlarının azaldığı gözlenmiştir. Rudner'ın yöntemine dayalı hesaplanan FN oranlarının, madde sayısının artışına bağlı olarak, her bir kesme puanında azaldığı; FP oranlarının ise -0,75 kesme puanı değerinde arttığı, 0 ve 0,75 kesme puanı değerlerinde azaldığı görülmektedir. Guo'nun yöntemine dayalı hesaplanan FP

oranlarının, madde sayısının artışına bağlı olarak, her bir kesme puanında azalış gösterdiği; FN oranlarının ise -0,75 ve 0 kesme puanlarında azalış, 0,75 kesme puanında artış gösterdiği sonucuna ulaşılmıştır.

2PLM altında üretilen N=1000 veri setinin 2PLM altında test edilmesine ilişkin bulgular incelendiğinde; her bir kesme puanı değeri için, her bir yöntem altında hesaplanan genel CA, CC oranları ve kappa değerlerindeki değişim ile Lee'nin ve Rudner'ın yöntemine dayalı hesaplanan FP ve FN oranlarındaki değişimin, madde sayısının artışına bağlı olarak, N=500 veri setinde görülen farklılaşma ile aynı olduğu dikkat çekmektedir. Guo'nun yöntemine dayalı hesaplanan FN oranlarının, madde sayısının artışına bağlı olarak, her bir kesme puanında azalırken; FP oranlarının ise -0,75 kesme puanında arttığı, 0 ve 0,75 kesme puanlarında azaldığı görülmektedir.

Elde edilen bulgular kesme puanları dikkate alınmadan incelendiğinde; Rudner'ın yöntemine dayalı hesaplanan genel CA, CC, FN oranları ve kappa değeri, Guo'nun yöntemine dayalı hesaplanan genel CA, CC oranları ve kappa değeri ile Lee'nin yöntemine dayalı hesaplanan genel CC oranı ve kappa değeri için madde sayısının artışına bağlı olarak, dikkate değer farklılıkların olduğu görülmektedir. Bu farklılıklar, genel CA, CC oranları ve kappa değerinde artış, FN oranında azalış olarak gözlenmektedir.

b. Her bir yöntem altında kestirilen sınıflama doğruluğu ve tutarlılığı indeksleri örneklem büyüklüğüne bağlı olarak farklılaşmakta mıdır?

2PLM altında üretilen K=15 veri setinin 2PLM altında test edilmesine ilişkin bulgular incelendiğinde; örneklem büyüklüğünün artışına bağlı olarak, her bir yönteme dayalı hesaplanan genel CA, CC, FP ve FN oranları ile kappa değerlerinin kesme puanına bağlı olarak farklılaştığı ancak sistematik bir farklılaşmanın bulunmadığı görülmektedir. Örneklem büyüklüğünün artışına bağlı olarak, Lee'nin ve Rudner'ın yöntemine dayalı hesaplanan kappa değerlerinin -0,75 kesme puanında arttığı, 0 ve 0,75 kesme puanlarında azaldığı gözlenirken; Guo'nun yöntemine dayalı hesaplanan kappa değerlerinin -0,75 ve 0 kesme puanlarında artış, 0,75 kesme puanında azalış gösterdiği gözlenmiştir. Rudner'ın ve Guo'nun yöntemine dayalı hesaplanan FP oranlarının, örneklem büyüklüğünün artışına bağlı olarak, -0,75 ve 0 kesme puanlarında azaldığı, 0,75 kesme puanında değişim göstermediği gözlenirken; Lee'nin yöntemine dayalı hesaplanan FP oranlarının -0,75 kesme puanında arttığı, 0 ve 0,75 kesme puanında azaldığı görülmektedir. Rudner'ın yöntemine dayalı hesaplanan FN oranları her bir kesme puanında, örneklem büyüklüğünün artışına bağlı

olarak, azalırken; Lee'nin yöntemine dayalı FN oranları -0,75 ve 0 kesme puanı değerlerinde artmış, 0,75 kesme puanı değerinde azalmıştır. Guo'nun yöntemine dayalı hesaplanan FN oranlarının ise -0,75 ve 0,75 kesme puanlarında artış, 0 kesme puanında azalış gösterdiği gözlenmiştir. Lee'nin yöntemine dayalı hesaplanan genel CA oranının, örneklem büyüklüğünün artışına bağlı olarak, -0,75 ve 0 kesme puanı değerlerinde azaldığı, 0,75 kesme puanı değerinde arttığı görülmektedir. Rudner'ın yöntemine dayalı hesaplanan genel CA oranının -0,75 kesme puanında arttığı, 0 ve 0,75 kesme puanlarında azaldığı; Guo'nun yöntemine dayalı hesaplanan genel CA oranının ise -0,75 ve 0 kesme puanlarında arttığı, 0,75 kesme puanında azaldığı görülmektedir. Lee'nin yöntemine dayalı hesaplanan genel CC oranının, örneklem büyüklüğünün artışına bağlı olarak, -0,75 ve 0 kesme puanlarında değişim göstermediği, 0,75 kesme puanında arttığı; Rudner'ın yöntemine dayalı hesaplanan genel CC oranının -0,75 kesme puanında arttığı, 0 ve 0,75 kesme puanlarında azaldığı gözlenmiştir. Guo'nun yöntemine dayalı hesaplanan genel CC oranının ise -0,75 ve 0 kesme puanlarında artış, 0,75 kesme puanında azalış gösterdiği görülmüştür.

2PLM altında üretilen K=30 veri setinin 2PLM altında test edilmesine ilişkin bulgular incelendiğinde; her bir yöntem altında hesaplanan kappa değerlerinin, örneklem büyüklüğünün artışına bağlı olarak, -0,75 kesme puanında azalış gösterirken, 0 ve 0,75 kesme puanlarında artış gösterdiği gözlenmiştir. Her bir yöntem altında hesaplanan genel CA, CC, FP ve FN oranlarının, kesme puanına bağlı olarak farklılaştığı ancak sistematik bir farklılaşmanın bulunmadığı anlaşılmaktadır. Lee'nin yöntemine dayalı hesaplanan genel CA ve CC oranlarının, örneklem büyüklüğünün artışına bağlı olarak, -0,75 ve 0,75 kesme puanlarında azaldığı, 0 kesme puanında arttığı görülmektedir. Rudner'ın yöntemine dayalı hesaplanan genel CA ve CC oranlarının, örneklem büyüklüğünün artışına bağlı olarak, -0,75 kesme puanında azaldığı, 0 ve 0,75 kesme puanlarında arttığı; Guo'nun yöntemine dayalı hesaplanan genel CA ve CC oranlarının, -0,75 kesme puanında azaldığı, 0 ve 0,75 kesme puanlarında arttığı gözlenmiştir. Lee'nin ve Rudner'ın yöntemine dayalı hesaplanan FN oranlarının, örneklem büyüklüğünün artışına bağlı olarak, -0,75 kesme puanında artış, 0 kesme puanında azalış ve 0,75 kesme puanında değişim göstermediği gözlenmiştir. Guo'nun yöntemine dayalı hesaplanan FN oranlarının ise, örneklem büyüklüğünün artışına bağlı olarak, -0,75 kesme puanında artış, 0 ve 0,75 kesme puanlarında azalış gösterdiği görülmüştür. Lee'nin yöntemine dayalı hesaplanan FP oranlarının, örneklem büyüklüğünün artışına bağlı olarak, -0,75 ve 0,75 kesme puanlarında arttığı, 0 kesme puanında azaldığı; Rudner'ın yöntemine dayalı hesaplanan FP oranlarının -0,75 kesme puanında arttığı, 0 ve

0,75 kesme puanlarında azaldığı; Guo'nun yöntemine dayalı hesaplanan FP oranlarının ise -0,75 kesme puanında değişim göstermediği, 0 ve 0,75 kesme puanlarında azaldığı gözlenmiştir.

Elde edilen bulgular kesme puanları dikkate alınmadan incelendiğinde; Rudner'ın, Guo'nun ve Lee'nin yöntemine dayalı hesaplanan genel CA, CC, FP, FN oranları ve kappa değerleri için örneklem büyüklüğünün artışına bağlı olarak dikkate değer bir farklılaşmanın olmadığı görülmektedir.

c. Test uzunluğuna bağlı olarak, farklı yöntemlerle kestirilen sınıflama doğruluğu ve tutarlılığı indekslerinin doğruluk derecesi farklılaşmakta mıdır? Hangi yöntem daha düşük bir yanlılık değeri ile sonuçlanmaktadır?

2PLM altında üretilen N=500 veri setinin 2PLM altında test edilmesine ilişkin bulgular incelendiğinde; her bir kesme puanı değeri için, madde sayısının artışına bağlı olarak, her bir yöntem altında hesaplanan genel CC oranları ve kappa değerlerine ilişkin mutlak yanlılık değerlerinde azalış gözlenmiştir. Her bir yöntem altında hesaplanan genel CA, FP ve FN oranlarına ilişkin mutlak yanlılık değerlerinin, kesme puanına bağlı olarak farklılaştığı ancak sistematik bir farklılaşmanın bulunmadığı anlaşılmaktadır. Lee'nin yöntemine dayalı hesaplanan genel CA, FP ve FN oranlarına ilişkin mutlak yanlılık değerlerinin, madde sayısının artışına bağlı olarak, -0,75 kesme puanında artış, 0 ve 0,75 kesme puanlarında azalış gösterdiği görülmektedir. Rudner'ın yöntemine dayalı hesaplanan genel CA oranına ilişkin mutlak yanlılık değerlerinin, madde sayısının artışına bağlı olarak, -0,75 ve 0,75 kesme puanlarında arttığı, 0 kesme puanında azaldığı; FP oranına ilişkin mutlak yanlılık değerlerinin her bir kesme puanında azaldığı; FN oranına ilişkin mutlak yanlılık değerlerinin ise -0,75 ve 0,75 kesme puanlarında azaldığı, 0 kesme puanında arttığı gözlenmiştir. Guo'nun yöntemine dayalı hesaplanan genel CA ve FN oranlarına ilişkin mutlak yanlılık değerlerinin, madde sayısının artışına bağlı olarak, her bir kesme puanı değerinde azalış gösterdiği; FP oranlarına ilişkin mutlak yanlılık değerlerinin -0,75 kesme puanında değişim göstermediği, 0 kesme puanında artış, 0,75 kesme puanında azalış gösterdiği görülmektedir.

2PLM altında üretilen N=1000 veri setinin 2PLM altında test edilmesine ilişkin bulgular incelendiğinde; her bir kesme puanı değeri için, madde sayısının artışına bağlı olarak, her bir yöntem altında hesaplanan genel CC oranları ve kappa değerlerine ilişkin mutlak yanlılık değerlerinin azaldığı gözlenmiştir. Her bir yöntem altında hesaplanan genel CA, FP ve FN oranlarına ilişkin mutlak yanlılık değerlerinin, kesme puanına bağlı olarak farklılaştığı ancak

sistematik bir farklılaşmanın bulunmadığı görülmüştür. Lee'nin yöntemine dayalı hesaplanan genel CA ve FP oranlarına ilişkin mutlak yanlılık değerlerinin, madde sayısının artışına bağlı olarak, -0,75 kesme puanında arttığı, 0 ve 0,75 kesme puanlarında azaldığı; FN oranına ilişkin mutlak yanlılık değerlerinin ise -0,75 ve 0,75 kesme puanlarında arttığı, 0 kesme puanında azaldığı gözlenmiştir. Rudner'ın yöntemine dayalı hesaplanan genel CA oranına ilişkin mutlak yanlılık değerlerinin, madde sayısının artışına bağlı olarak, her bir kesme puanı değerinde artış gösterdiği; FP oranına ilişkin mutlak yanlılık değerlerinin, her bir kesme puanı değerinde azalış gösterdiği; FN oranına ilişkin mutlak yanlılık değerlerinin ise -0,75 ve 0,75 kesme puanlarında azalış, 0 kesme puanında artış gösterdiği görülmektedir. Guo'nun yöntemine dayalı hesaplanan genel CA oranına ilişkin mutlak yanlılık değerlerinin, madde sayısının artışına bağlı olarak, her bir kesme puanı değerinde azaldığı; FP oranına ilişkin mutlak yanlılık değerlerinin -0,75 ve 0,75 kesme puanlarında azaldığı, 0 kesme puanında arttığı; FN oranına ilişkin mutlak yanlılık değerlerinin ise -0,75 ve 0 kesme puanlarında azaldığı, 0,75 kesme puanında arttığı gözlenmiştir.

Elde edilen bulgular kesme puanları dikkate alınmadan incelendiğinde; her bir yönteme dayalı hesaplanan genel CC oranları ve kappa değerleri, Rudner'ın yöntemine dayalı hesaplanan FP oranları ile Guo'nun yöntemine dayalı hesaplanan genel CA oranlarına ilişkin mutlak yanlılık değerleri için, madde sayısının artışına bağlı olarak, dikkate değer farklılıkların olduğu gözlenmiştir. Bu farklılıklar, söz konusu değerlere ilişkin mutlak yanlılık değerlerinde azalış olarak gözlenmiştir.

2PLM altında üretilen K=15 veri setinin 2PLM altında test edilmesine ilişkin bulgular kesme puanı dikkate alınmaksızın incelendiğinde; Rudner'ın yöntemine dayalı hesaplanan genel CA, CC, FN oranları ve kappa değerleri, Guo'nun yöntemine dayalı hesaplanan FP oranı ile Lee'nin yöntemine dayalı hesaplanan FN oranının daha düşük mutlak yanlılık değerlerine sahip olduğu görülmektedir.

2PLM altında üretilen K=30 veri setinin 2PLM altında test edilmesine ilişkin bulgular kesme puanı dikkate alınmaksızın incelendiğinde; Rudner'ın yöntemine dayalı hesaplanan genel CC, FN oranları ve kappa değerleri, Lee'nin yöntemine dayalı hesaplanan genel CA ve FP oranları ile Guo'nun yöntemine dayalı hesaplanan genel CA oranının daha düşük mutlak yanlılık değerlerine sahip olduğu görülmektedir.

d. Örneklem büyüklüğüne bağlı olarak, farklı yöntemlerle kestirilen sınıflama doğruluğu ve tutarlılığı indekslerinin doğruluk derecesi farklılaşmakta mıdır? Hangi yöntem daha düşük bir yanlılık değeri ile sonuçlanmaktadır?

2PLM altında üretilen K=15 veri setinin 2PLM altında test edilmesine ilişkin bulgular incelendiğinde; örneklem büyüklüğünün artışına bağlı olarak, her bir yöntem altında hesaplanan genel CA, CC, FP ve FN oranları ve kappa değerlerine ilişkin mutlak yanlılık değerlerinin, kesme puanına bağlı olarak farklılaştığı ancak sistematik bir farklılaşmanın bulunmadığı anlaşılmaktadır. Rudner'ın yöntemine dayalı hesaplanan genel CC oranı ve kappa değerlerine ilişkin mutlak yanlılık değerlerinin, örneklem büyüklüğünün artışına bağlı olarak, -0,75 kesme puanında azaldığı, 0 ve 0,75 kesme puanlarında arttığı; genel CA oranına ilişkin mutlak yanlılık değerlerinin ise -0,75 kesme puanında arttığı, 0 ve 0,75 kesme puanlarında azaldığı görülmektedir. Rudner'ın yöntemine dayalı hesaplanan FP oranına ilişkin mutlak yanlılık değerlerinin, örneklem büyüklüğündeki artışa bağlı olarak, -0,75 ve 0,75 kesme puanlarında arttığı, 0 kesme puanında azaldığı; FN oranına ilişkin mutlak yanlılık değerlerinin ise -0,75 ve 0 kesme puanlarında arttığı, 0,75 kesme puanında azaldığı gözlenmiştir. Guo'nun yöntemine dayalı hesaplanan genel CA ve CC oranları ile kappa değerlerine ilişkin mutlak yanlılık değerlerinin, örneklem büyüklüğünün artışına bağlı olarak, -0,75 ve 0 kesme puanlarında azalış, 0,75 kesme puanında artış gösterdiği görülmektedir. Guo'nun yöntemine dayalı hesaplanan FP oranına ilişkin mutlak yanlılık değerlerinin, örneklem büyüklüğündeki artışa bağlı olarak, -0,75 ve 0,75 kesme puanlarında arttığı, 0 kesme puanında azaldığı; FN oranına ilişkin mutlak yanlılık değerlerinin ise -0,75 kesme puanında arttığı, 0 ve 0,75 kesme puanlarında azaldığı gözlenmiştir. Lee'nin yöntemine dayalı hesaplanan genel CA oranına ilişkin mutlak yanlılık değerlerinin, örneklem büyüklüğünün artışına bağlı olarak, -0,75 kesme puanında azalış, 0 ve 0,75 kesme puanlarında artış gösterdiği görülmüştür. Lee'nin yöntemine dayalı hesaplanan FP oranına ilişkin mutlak yanlılık değerlerinin her bir kesme puanı değerinde azaldığı; FN oranına ilişkin mutlak yanlılık değerlerinin ise -0,75 ve 0,75 kesme puanlarında azaldığı, 0 kesme puanında arttığı gözlenmiştir. Lee'nin yöntemine dayalı hesaplanan genel CC oranına ilişkin mutlak yanlılık değerlerinin, örneklem büyüklüğünün artışına bağlı olarak, -0,75 ve 0 kesme puanlarında değişim göstermediği, 0,75 kesme puanında azaldığı; kappa değerlerine ilişkin mutlak yanlılık değerlerinin ise -0,75 kesme puanında azaldığı, 0 ve 0,75 kesme puanlarında arttığı görülmektedir.

2PLM altında üretilen K=30 veri setinin 2PLM altında test edilmesine ilişkin bulgular incelendiğinde; örneklem büyüklüğünün artışına bağlı olarak, Rudner'ın ve Guo'nun yöntemine dayalı hesaplanan genel CC oranları ve kappa değerlerine ilişkin mutlak yanlılık değerlerinin -0,75 kesme puanında artış, 0 ve 0,75 kesme puanlarında azalış gösterdiği gözlenmiştir. Lee'nin yöntemine dayalı hesaplanan genel CA, CC ve FN oranlarına ilişkin mutlak yanlılık değerlerinin, örneklem büyüklüğünün artışına bağlı olarak, -0,75 ve 0,75 kesme puanlarında arttığı, 0 kesme puanında azaldığı görülmüştür. Lee'nin yöntemine dayalı hesaplanan FP oranına ilişkin mutlak yanlılık değerlerinin -0,75 kesme puanında azaldığı, 0 ve 0,75 kesme puanlarında arttığı; kappa değerine ilişkin mutlak yanlılık değerlerinin ise -0,75 kesme puanında arttığı, 0 ve 0,75 kesme puanında azaldığı gözlenmiştir. Rudner'ın yöntemine dayalı hesaplanan FP ve FN oranlarına ilişkin mutlak yanlılık değerlerinin, örneklem büyüklüğünün artışına bağlı olarak, -0,75 ve 0 kesme puanlarında artış, 0,75 kesme puanında değişim göstermediği; genel CA oranına ilişkin mutlak yanlılık değerlerinin ise -0,75 ve 0,75 kesme puanlarında azalış, 0 kesme puanında artış gösterdiği gözlenmiştir. Guo'nun yöntemine dayalı hesaplanan genel CA ve FP oranlarına ilişkin mutlak yanlılık değerlerinin, örneklem büyüklüğünün artışına bağlı olarak, her bir kesme puanı değerinde artış gösterdiği; FN oranına ilişkin mutlak yanlılık değerlerinin ise -0,75 ve 0,75 kesme puanlarında artış, 0 kesme puanında azalış gösterdiği gözlenmiştir.

Elde edilen bulgular kesme puanları dikkate alınmadan incelendiğinde; Rudner'ın, Guo'nun ve Lee'nin yöntemine dayalı hesaplanan genel CA, CC, FP, FN oranları ve kappa değerlerine ilişkin mutlak yanlılık değerleri için örneklem büyüklüğünün artışına bağlı olarak dikkate değer bir farklılaşmanın olmadığı görülmüştür.

2PLM altında üretilen N=500 veri setinin 2PLM altında test edilmesine ilişkin bulgular kesme puanı dikkate alınmaksızın incelendiğinde; Rudner'ın yöntemine dayalı hesaplanan genel CA, CC, FN oranları ve kappa değerleri ile Guo'nun yöntemine dayalı hesaplanan FP oranının daha düşük mutlak yanlılık değerlerine sahip olduğu gözlenmiştir.

2PLM altında üretilen N=1000 veri setinin 2PLM altında test edilmesine ilişkin bulgular kesme puanı dikkate alınmaksızın incelendiğinde; Rudner'ın yöntemine dayalı hesaplanan genel CC oranı ve kappa değeri, Lee'nin yöntemine dayalı hesaplanan genel CA ve FN oranları ile Guo'nun yöntemine dayalı hesaplanan FP oranının daha düşük mutlak yanlılık değerlerine sahip olduğu gözlenmiştir.

3. 3PLM altında üretilen veri 3PLM altında test edildiğinde;
    a. Her bir yöntem altında kestirilen sınıflama doğruluğu ve tutarlılığı indeksleri test uzunluğuna bağlı olarak farklılaşmakta mıdır?

3PLM altında üretilen N=500 veri setinin 3PLM altında test edilmesine ilişkin bulgular incelendiğinde; her bir kesme puanı değerinde, her bir yöntem altında hesaplanan genel CA ve CC oranlarına ilişkin değerlerin, madde sayısının artışına bağlı olarak arttığı gözlenmiştir. Rudner'ın ve Guo'nun yöntemine dayalı hesaplanan kappa değerlerinin, madde sayısının artışına bağlı olarak, her bir kesme puanında, artış gösterdiği; Lee'nin yöntemine dayalı hesaplanan kappa değerlerinin ise -0,75 ve 0 kesme puanlarında artış, 0,75 kesme puanında azalış gösterdiği görülmüştür. Lee'nin ve Guo'nun yöntemine dayalı hesaplanan FN oranlarının her bir kesme puanında, madde sayısının artışına bağlı olarak, azaldığı; Rudner'ın yöntemine dayalı hesaplanan FN oranının ise -0,75 ve 0 kesme puanlarında azaldığı, 0,75 kesme puanında arttığı gözlenmiştir. Rudner'ın yöntemine dayalı hesaplanan FP oranının her bir kesme puanı değerinde, azaldığı; Lee'nin yöntemine dayalı hesaplanan FP oranının -0,75 kesme puanında arttığı; 0 ve 0,75 kesme puanlarında azaldığı; Guo'nun yöntemine dayalı hesaplanan FP oranının ise -0,75 ve 0 kesme puanlarında arttığı, 0,75 kesme puanında azaldığı görülmektedir.

3PLM altında üretilen N=1000 veri setinin 3PLM altında test edilmesine ilişkin bulgular incelendiğinde; her bir kesme puanı değeri için, Lee'nin ve Rudner'ın yöntemine dayalı hesaplanan genel CA, CC, FP, FN oranları ve kappa değerlerindeki değişimin, madde sayısının artışına bağlı olarak, N=500 veri setinde görülen farklılaşma ile aynı olduğu gözlenmiştir. Guo'nun yöntemine dayalı hesaplanan genel CC, FP oranları ve kappa değerinin, madde sayısının artışına bağlı olarak, her bir kesme puanı değerinde artış gösterirken, FN değerinin azalış; genel CC oranının ise -0,75 ve 0 kesme puanlarında artış, 0,75 kesme puanında azalış gösterdiği gözlenmiştir.

Elde edilen bulgular kesme puanları dikkate alınmadan incelendiğinde; Rudner'ın yöntemine dayalı hesaplanan genel CA, CC, FP oranları ve kappa değeri, Guo'nun yöntemine dayalı hesaplanan genel CA, CC oranları ve kappa değeri ile Lee'nin yöntemine dayalı hesaplanan genel CA ve CC oranlarına ilişkin değerler için madde sayısının artışına bağlı olarak, dikkate değer farklılıkların olduğu gözlenmiştir. Bu farklılıklar, genel CA, CC oranları ve kappa değerlerinde artış, FP oranında azalış olarak gözlenmiştir.

b. Her bir yöntem altında kestirilen sınıflama doğruluğu ve tutarlılığı indeksleri örneklem büyüklüğüne bağlı olarak farklılaşmakta mıdır?

3PLM altında üretilen K=15 veri setinin 3PLM altında test edilmesine ilişkin bulgular incelendiğinde; örneklem büyüklüğünün artışına bağlı olarak, Lee'nin yöntemine dayalı hesaplanan genel CA, CC oranları ve kappa değerinin -0,75 ve 0 kesme puanlarında azalış, 0,75 kesme puanında artış gösterdiği görülmektedir. Lee'nin yöntemine dayalı hesaplanan FP oranının, örneklem büyüklüğünün artışına bağlı olarak, -0,75 ve 0,75 kesme puanlarında azalış, 0 kesme puanında artış gösterdiği; FN oranının ise -0,75 kesme puanında artış, 0 ve 0,75 kesme puanında azalış gösterdiği gözlenmiştir. Rudner'ın yöntemine dayalı hesaplanan genel CA ve CC oranlarının her bir kesme puanında, örneklem büyüklüğünün artışına bağlı olarak azaldığı görülmüştür. Rudner'ın yöntemine dayalı hesaplanan kappa değerinin -0,75 ve 0 kesme puanlarında azaldığı, 0,75 kesme puanında arttığı; FP oranının -0,75 kesme puanında azaldığı, 0 ve 0,75 kesme puanlarında arttığı; FN oranının ise -0,75 kesme puanında arttığı, 0 ve 0,75 kesme puanlarında azaldığı gözlenmiştir. Guo'nun yöntemine dayalı hesaplanan genel CC oranı ve kappa değerinin, her bir kesme puanında, örneklem büyüklüğünün artışına bağlı olarak, azaldığı; genel CA oranının -0,75 kesme puanında azaldığı, 0 kesme puanında arttığı, 0,75 kesme puanında değişim göstermediği görülmektedir. Guo'nun yöntemine dayalı hesaplanan FP oranının -0,75 kesme puanında değişim göstermediği, 0 kesme puanında artış, 0,75 kesme puanında azalış gösterdiği; FN oranının ise -0,75 ve 0,75 kesme puanlarında artış, 0 kesme puanında azalış gösterdiği gözlenmiştir.

3PLM altında üretilen K=30 veri setinin 3PLM altında test edilmesine ilişkin bulgular incelendiğinde; örneklem büyüklüğünün artışına bağlı olarak, her bir yöntem altında hesaplanan kappa değerleri ile Rudner'ın ve Guo'nun yöntemine dayalı hesaplanan genel CC oranlarının, her bir kesme puanında azaldığı görülmektedir. Lee'nin yöntemine dayalı hesaplanan genel CC oranının ise örneklem büyüklüğünün artışına bağlı olarak -0,75 ve 0 kesme puanlarında azalış, 0,75 kesme puanında artış gösterdiği gözlenmiştir. Lee'nin ve Rudner'ın yöntemine dayalı hesaplanan FN oranlarının, örneklem büyüklüğünün artışına bağlı olarak, -0,75 ve 0 kesme puanlarında artış, 0,75 kesme puanında azalış gösterdiği; Guo'nun yöntemine dayalı hesaplanan FN oranının ise -0,75 kesme puanında artış, 0 ve 0,75 kesme puanlarında azalış gösterdiği görülmüştür. Lee'nin yöntemine dayalı hesaplanan genel CA oranının, örneklem büyüklünün artışına bağlı olarak, -0,75 ve 0,75 kesme puanlarında artış, 0 kesme puanında azalış gösterdiği; FP oranının ise -0,75 ve 0 kesme

puanlarında artış, 0,75 kesme puanında değişim göstermediği gözlenmiştir. Rudner'ın yöntemine dayalı hesaplanan genel CA oranının, örneklem büyüklüğünün artışına bağlı olarak, -0,75 ve 0 kesme puanlarında azalış, 0,75 kesme puanında artış gösterdiği; FP oranının ise her bir kesme puanında artış gösterdiği görülmektedir. Guo'nun yöntemine dayalı hesaplanan genel CA oranının, örneklem büyüklüğünün artışına bağlı olarak, her bir kesme puanında azalış gösterdiği; FP oranının ise -0,75 kesme puanında azalış, 0 ve 0,75 kesme puanında artış gösterdiği gözlenmiştir.

Elde edilen bulgular kesme puanları dikkate alınmadan incelendiğinde; Rudner'ın, Guo'nun ve Lee'nin yöntemine dayalı hesaplanan genel CA, CC, FP, FN oranları ve kappa değerleri için örneklem büyüklüğünün artışına bağlı olarak dikkate değer bir farklılaşmanın olmadığı görülmüştür.

c. Test uzunluğuna bağlı olarak, farklı yöntemlerle kestirilen sınıflama doğruluğu ve tutarlılığı indekslerinin doğruluk derecesi farklılaşmakta mıdır? Hangi yöntem daha düşük bir yanlılık değeri ile sonuçlanmaktadır?

3PLM altında üretilen N=500 veri setinin 3PLM altında test edilmesine ilişkin bulgular incelendiğinde; her bir kesme puanı değeri için, madde sayısının artışına bağlı olarak, Rudner'ın ve Guo'nun yöntemine dayalı hesaplanan genel CA, CC, FP oranları ve kappa değerlerine ilişkin mutlak yanlılık değerlerinde azalış görülmektedir. Rudner'ın yöntemine dayalı hesaplanan FN oranınına ilişkin mutlak yanlılık değerlerinde, madde sayısının artışına bağlı olarak, -0,75 kesme puanında azalış, 0 ve 0,75 kesme puanında artış görülmüştür. Guo'nun yöntemine dayalı hesaplanan FN oranına ilişkin mutlak yanlılık değerlerinde ise -0,75 ve 0 kesme puanlarında azalış, 0,75 kesme puanında artış görüldüğü gözlenmiştir. Lee'nin yöntemine dayalı hesaplanan genel CA oranına ilişkin mutlak yanlılık değerlerinin, madde sayısının artışına bağlı olarak, -0,75 ve 0,75 kesme puanlarında arttığı, 0 kesme puanında azaldığı gözlenmiştir. Lee'nin yöntemine dayalı hesaplanan FP oranına ilişkin mutlak yanlılık değerlerinin -0,75 ve 0,75 kesme puanlarında azaldığı, 0 kesme puanında arttığı; FN oranına ilişkin mutlak yanlılık değerlerinin ise -0,75 kesme puanında azaldığı, 0,75 kesme puanında arttığı, 0 kesme puanında değişim göstermediği gözlenmiştir. Lee'nin yöntemine dayalı hesaplanan genel CC oranına ilişkin mutlak yanlılık değerlerinin ise her bir kesme puanında azaldığı; kappa değerine ilişkin mutlak yanlılık değerlerinin -0,75 ve 0 kesme puanlarında azaldığı, 0,75 kesme puanında arttığı görülmüştür.

3PLM altında üretilen N=1000 veri setinin 3PLM altında test edilmesine ilişkin bulgular incelendiğinde; madde sayısının artışına bağlı olarak, Rudner'ın yöntemine dayalı hesaplanan genel CA, CC, FP oranları ve kappa değeri ile Guo'nun yöntemine dayalı hesaplanan genel CC oranı ve kappa değerine ilişkin mutlak yanlılık değerlerinin, her bir kesme puanı değeri için, azaldığı gözlenmiştir. Guo'nun yöntemine dayalı hesaplanan genel CA, FP ve FN oranları ile Lee'nin yöntemine dayalı hesaplanan genel CA oranı ve kappa değerine ilişkin mutlak yanlılık değerlerinde, madde sayısının artışına bağlı olarak, -0,75 ve 0 kesme puanlarında azalış, 0,75 kesme puanında artış görüldüğü gözlenmiştir. Lee'nin yöntemine dayalı hesaplanan FP oranına ilişkin mutlak yanlılık değerlerinin, madde sayısının artışına bağlı olarak, -0,75 ve 0,75 kesme puanlarında azaldığı, 0 kesme puanında arttığı görülmüştür. Lee'nin yöntemine dayalı hesaplanan FN oranına ilişkin mutlak yanlılık değerlerinin -0,75 kesme puanında azaldığı, 0 ve 0,75 kesme puanlarında arttığı; genel CC oranına ilişkin mutlak yanlılık değerlerinin ise her bir kesme puanında azaldığı görülmektedir. Rudner'ın yöntemine dayalı hesaplanan FN oranına ilişkin mutlak yanlılık değerlerinde ise madde sayısının artışına bağlı olarak, -0,75 ve 0,75 kesme puanlarında azalış, 0 kesme puanında artış görüldüğü gözlenmiştir.

Elde edilen bulgular kesme puanları dikkate alınmadan incelendiğinde; Rudner'ın yöntemine dayalı hesaplanan genel CA, CC, FP oranları ve kappa değeri, Guo'nun yöntemine dayalı hesaplanan genel CC oranı ve kappa değeri ile Lee'nin yöntemine dayalı hesaplanan genel CC oranına ilişkin mutlak yanlılık değerleri için, madde sayısının artışına bağlı olarak, dikkate değer farklılıkların olduğu gözlenmiştir. Bu farklılıklar, söz konusu değerlere ilişkin mutlak yanlılık değerlerinde azalış olarak gözlenmiştir.

3PLM altında üretilen K=15 veri setinin 3PLM altında test edilmesine ilişkin bulgular kesme puanı dikkate alınmaksızın incelendiğinde; Guo'nun yöntemine dayalı hesaplanan genel CA, CC oranları ve kappa değeri ile Lee'nin yöntemine dayalı hesaplanan FP ve FN oranlarının daha düşük mutlak yanlılık değerlerine sahip olduğu gözlenmiştir.

3PLM altında üretilen K=30 veri setinin 3PLM altında test edilmesine ilişkin bulgular kesme puanı dikkate alınmaksızın incelendiğinde; Guo'nun yöntemine dayalı hesaplanan genel CA, FP, FN oranları ve kappa değeri ile Lee'nin yöntemine dayalı hesaplanan genel CC oranının daha düşük mutlak yanlılık değerlerine sahip olduğu gözlenmiştir.

d. Örneklem büyüklüğüne bağlı olarak, farklı yöntemlerle kestirilen sınıflama doğruluğu ve tutarlılığı indekslerinin doğruluk derecesi farklılaşmakta mıdır? Hangi yöntem daha düşük bir yanlılık değeri ile sonuçlanmaktadır?

3PLM altında üretilen K=15 veri setinin 3PLM altında test edilmesine ilişkin bulgular incelendiğinde; örneklem büyüklüğünün artışına bağlı olarak, Lee'nin yöntemine dayalı hesaplanan genel CC oranı ve kappa değeri ile Rudner'ın yöntemine dayalı hesaplanan kappa değerine ilişkin mutlak yanlılık değerlerinin -0,75 ve 0 kesme puanlarında arttığı, 0,75 kesme puanında azaldığı görülmektedir. Guo'nun yöntemine dayalı hesaplanan genel CC oranı ve kappa değeri ile Rudner'ın yöntemine dayalı hesaplanan genel CC oranına ilişkin mutlak yanlılık değerlerinin, örneklem büyüklüğünün artışına bağlı olarak, her bir kesme puanı değerinde arttığı gözlenmiştir. Guo'nun yöntemine dayalı hesaplanan genel CA, FP ve FN oranlarına ilişkin mutlak yanlılık değerlerinde, örneklem büyüklüğünün artışına bağlı olarak, -0,75 kesme puanında artış, 0 ve 0,75 kesme puanlarında azalış görüldüğü gözlenmiştir. Rudner'ın yöntemine dayalı hesaplanan genel CA ve FP oranlarına ilişkin mutlak yanlılık değerlerinin, örneklem büyüklüğünün artışına bağlı olarak, her bir kesme puanında azaldığı; FN oranına ilişkin mutlak yanlılık değerlerinin ise -0,75 ve 0,75 kesme puanlarında arttığı, 0 kesme puanında değişim göstermediği görülmüştür. Lee'nin yöntemine dayalı hesaplanan genel CA oranına ilişkin mutlak yanlılık değerlerinin, örneklem büyüklüğünün artışına bağlı olarak, -0,75 kesme puanında değişim göstermediği, 0 kesme puanında artış, 0,75 kesme puanında ise azalış gösterdiği gözlenmiştir. Lee'nin yöntemine dayalı hesaplanan FP oranına ilişkin mutlak yanlılık değerlerinde, her bir kesme puanı değerinde azalış görülürken; FN oranına ilişkin mutlak yanlılık değerlerinde -0,75 ve 0,75 kesme puanlarında artış, 0 kesme puanında azalış görülmüştür.

3PLM altında üretilen K=30 veri setinin 3PLM altında test edilmesine ilişkin bulgular incelendiğinde; örneklem büyüklüğünün artışına bağlı olarak, Rudner'ın yöntemine dayalı hesaplanan genel CC, FP oranları ve kappa değeri, Guo'nun yöntemine dayalı hesaplanan genel CC oranı ve kappa değeri ile Lee'nin yöntemine dayalı hesaplanan kappa değerine ilişkin mutlak yanlılık değerlerinde, her bir kesme puanı değeri için artış görüldüğü gözlenmiştir. Lee'nin yöntemine dayalı hesaplanan genel CC, FP ve FN oranları ile Rudner'ın ve Guo'nun yöntemine dayalı hesaplanan FN oranlarına ilişkin mutlak yanlılık değerlerinin, örneklem büyüklüğünün artışına bağlı olarak, -0,75 ve 0 kesme puanlarında arttığı, 0,75 kesme puanında azaldığı görülmektedir. Lee'nin yöntemine dayalı hesaplanan genel CA oranına ilişkin mutlak yanlılık değerlerinde, örneklem büyüklüğünün artışına bağlı

olarak, -0,75 ve 0,75 kesme puanlarında azalış, 0 kesme puanında artış görülmüştür. Rudner'ın yöntemine dayalı hesaplanan genel CA oranına ilişkin mutlak yanlılık değerlerinde, -0,75 kesme puanında artış, 0 ve 0,75 kesme puanlarında azalış görüldüğü gözlenmiştir. Guo'nun yöntemine dayalı hesaplanan genel CA oranına ilişkin mutlak yanlılık değerlerinde, örneklem büyüklüğünün artışına bağlı olarak, -0,75 kesme puanında azalış, 0 ve 0,75 kesme puanlarında artış görülürken; FP oranına ilişkin mutlak yanlılık değerlerinde -0,75 ve 0,75 kesme puanlarında artış, 0 kesme puanında azalış görülmektedir. Elde edilen bulgular kesme puanları dikkate alınmadan incelendiğinde; Rudner'ın, Guo'nun ve Lee'nin yöntemine dayalı hesaplanan genel CA, CC, FP ve FN oranları ile kappa değerlerine ilişkin mutlak yanlılık değerleri için, örneklem büyüklüğünün artışına bağlı olarak, dikkate değer bir farklılaşmanın olmadığı görülmüştür.

3PLM altında üretilen N=500 veri setinin 3PLM altında test edilmesine ilişkin bulgular kesme puanı dikkate alınmaksızın incelendiğinde; Guo'nun yöntemine dayalı hesaplanan genel CA, FP, FN oranları ve kappa değeri, Lee'nin yöntemine dayalı hesaplanan genel CC ve FN oranları ile Rudner'ın yöntemine dayalı hesaplanan FP oranının daha düşük mutlak yanlılık değerlerine sahip olduğu gözlenmiştir.

3PLM altında üretilen N=1000 veri setinin 3PLM altında test edilmesine ilişkin bulgular kesme puanı dikkate alınmaksızın incelendiğinde; Guo'nun yöntemine dayalı hesaplanan genel CA, FP oranları ve kappa değeri, Lee'nin yöntemine dayalı hesaplanan genel CC oranı ile Rudner'ın yöntemine dayalı hesaplanan FN oranının daha düşük mutlak yanlılık değerlerine sahip olduğu gözlenmiştir.

4. Belirli bir MTK modeli altında üretilen veri, farklı bir MTK modeli altında test edildiğinde yani model yanlış tanımlaması durumunda;
    a. Sınıflama doğruluğu ve tutarlılığı indekslerinin doğruluk derecesi farklılaşmakta mıdır? Hangi sınıflama doğruluğu ve tutarlılığı indeks(ler)i model yanlış tanımlamasından daha az etkilenmektedir?

1PLM altında üretilen veri setinin 2PLM altında test edilmesine ilişkin bulgular kesme puanı dikkate alınmaksızın incelendiğinde; sınıflama doğruluğu ve tutarlılığı indekslerine ilişkin mutlak yanlılık değerlerinde farklılaşmakta olduğu görülmektedir. Genel CA oranları ve kappa değerlerine ilişkin mutlak yanlılık değerlerinde artış gözlenmiştir. Ayrıca, FP, FN oranları ve kappa değerlerine ilişkin mutlak yanlılık değerlerinde, kullanılan yönteme bağlı olarak, dikkate değer bir farklılaşma söz konusudur. Lee'nin yöntemine dayalı hesaplanan

FP ve FN oranları ile Rudner'ın yöntemine dayalı hesaplanan kappa değerinin daha düşük mutlak yanlılık değerlerine sahip olduğu dikkat çekmiştir.

1PLM altında üretilen veri setinin 3PLM altında test edilmesine ilişkin bulgular kesme puanı dikkate alınmaksızın incelendiğinde; sınıflama doğruluğu ve tutarlılığı indekslerine ilişkin mutlak yanlılık değerlerinde farklılaşmakta olduğu görülmüştür. Genel CC oranları ve kappa değerlerine ilişkin mutlak yanlılık değerlerinde artış görülmektedir. Kullanılan yönteme bağlı olarak, FP, FN oranları ve kappa değerlerine ilişkin mutlak yanlılık değerlerinde dikkate değer bir farklılaşma gözlenmiştir. Lee'nin yöntemine dayalı hesaplanan FP ve FN oranları ile Guo'nun yöntemine dayalı hesaplanan kappa değerinin daha düşük mutlak yanlılık değerlerine sahip olduğu görülmüştür.

1PLM altında üretilen veri setlerinin model yanlış tanımlaması durumunda test edilmesine ilişkin bulgular kesme puanı dikkate alınmaksızın incelendiğinde; sınıflama doğruluğu ve tutarlılığı indekslerine ilişkin mutlak yanlılık değerlerinde, madde sayısının artışına bağlı olarak, farklılaşma olduğu dikkat çekmektedir. Bu farklılaşma söz konusu indekslere ilişkin mutlak yanlılık değerlerinin, madde sayısının artışına bağlı olarak, azalması yönündedir.

1PLM altında üretilen veri setlerinin model yanlış tanımlaması durumunda test edilmesine ilişkin bulgular kesme puanı dikkate alınmaksızın incelendiğinde; sınıflama doğruluğu ve tutarlılığı indekslerine ilişkin mutlak yanlılık değerlerinde, örneklem büyüklüğünün artışına bağlı olarak, dikkate değer bir farklılaşmanın olmadığı sonucuna ulaşılmıştır.

2PLM altında üretilen veri setinin 1PLM altında test edilmesine ilişkin bulgular kesme puanı dikkate alınmaksızın incelendiğinde; sınıflama doğruluğu ve tutarlılığı indekslerine ilişkin mutlak yanlılık değerlerinde farklılaşmakta olduğu görülmektedir. Genel CA, CC, FP, FN oranları ve kappa değerlerine ilişkin mutlak yanlılık değerlerinde artış gözlenmiştir. FP oranları ve kappa değerlerine ilişkin mutlak yanlılık değerlerinde, kullanılan yönteme bağlı olarak, dikkate değer bir farklılaşma gözlenmiştir. Lee'nin yöntemine dayalı hesaplanan FP oranı ile Guo'nun yöntemine dayalı hesaplanan kappa değerinin daha düşük mutlak yanlılık değerlerine sahip olduğu söylenebilir.

2PLM altında üretilen veri setinin 3PLM altında test edilmesine ilişkin bulgular kesme puanı dikkate alınmaksızın incelendiğinde; sınıflama doğruluğu ve tutarlılığı indekslerine ilişkin mutlak yanlılık değerlerinde farklılaşmakta görülmektedir. Genel CA, FP oranları ve kappa değerlerine ilişkin mutlak yanlılık değerlerinde artış görülmüştür. Kullanılan yönteme bağlı olarak ise sadece kappa değerlerine ilişkin mutlak yanlılık değerlerinde dikkate değer bir

farklılaşma söz konusudur. Guo'nun yöntemine dayalı hesaplanan kappa değerinin daha düşük mutlak yanlılık değerlerine sahip olduğu gözlenmiştir.

2PLM altında üretilen veri setlerinin model yanlış tanımlaması durumunda test edilmesine ilişkin bulgular kesme puanı dikkate alınmaksızın incelendiğinde; sınıflama doğruluğu ve tutarlılığı indekslerine ilişkin mutlak yanlılık değerlerinin, madde sayısının artışına bağlı olarak, farklılaştığı göze çarpmaktadır. Bu farklılaşma söz konusu indekslere ilişkin mutlak yanlılık değerlerinin, madde sayısının artışına bağlı olarak, azalması şeklindedir.

2PLM altında üretilen veri setlerinin model yanlış tanımlaması durumunda test edilmesine ilişkin bulgular kesme puanı dikkate alınmaksızın incelendiğinde; sınıflama doğruluğu ve tutarlılığı indekslerine ilişkin mutlak yanlılık değerlerinde, örneklem büyüklüğünün artışına bağlı olarak, dikkate değer bir farklılaşmanın olmadığı görülmektedir.

3PLM altında üretilen veri setinin 1PLM altında test edilmesine ilişkin bulgular kesme puanı dikkate alınmaksızın incelendiğinde; sınıflama doğruluğu ve tutarlılığı indekslerine ilişkin mutlak yanlılık değerlerinde farklılaşmakta olduğu gözlenmiştir. Genel CA, FP ve FN oranlarına ilişkin mutlak yanlılık değerlerinde artış görülmektedir. FN oranları ve kappa değerlerine ilişkin mutlak yanlılık değerlerinde, kullanılan yönteme bağlı olarak, dikkate değer bir farklılaşma olduğu görülmüştür. Rudner'ın yöntemine dayalı hesaplanan FN oranı ile Guo'nun yöntemine dayalı hesaplanan kappa değerinin daha düşük mutlak yanlılık değerlerine sahip olduğu görülmüştür.

3PLM altında üretilen veri setinin 2PLM altında test edilmesine ilişkin bulgular kesme puanı dikkate alınmaksızın incelendiğinde; sınıflama doğruluğu ve tutarlılığı indekslerine ilişkin mutlak yanlılık değerlerinde farklılaşma gözlenmiştir. Genel CA ve FN oranlarına ilişkin mutlak yanlılık değerlerinde artış görülmektedir. FN oranları ve kappa değerlerine ilişkin mutlak yanlılık değerlerinde, kullanılan yönteme bağlı olarak, dikkate değer bir farklılaşma olduğu gözlenmiştir. Rudner'ın yöntemine dayalı hesaplanan FN oranı ile Guo'nun yöntemine dayalı hesaplanan kappa değerinin daha düşük mutlak yanlılık değerlerine sahip olduğu görülmektedir.

3PLM altında üretilen veri setlerinin model yanlış tanımlaması durumunda test edilmesine ilişkin bulgular kesme puanı dikkate alınmaksızın incelendiğinde; sınıflama doğruluğu ve tutarlılığı indekslerine ilişkin mutlak yanlılık değerlerinin, madde sayısının artışına bağlı olarak, farklılaştığı dikkat çekmiştir. Söz konusu indekslere ilişkin mutlak yanlılık değerlerinin, madde sayısının artışına bağlı olarak, azaldığı sonucuna ulaşılmaktadır.

3PLM altında üretilen veri setlerinin model yanlış tanımlaması durumunda test edilmesine ilişkin bulgular kesme puanı dikkate alınmaksızın incelendiğinde; sınıflama doğruluğu ve tutarlılığı indekslerine ilişkin mutlak yanlılık değerlerinde, örneklem büyüklüğünün artışına bağlı olarak, dikkate değer bir farklılaşmanın olmadığı bulunmuştur.

1PLM, 2PLM ve 3PLM'e dayalı üretilen veri setlerinin, üretildiği model dışındaki modeller altında test edildiği durumlar incelendiğinde; sınıflama doğruluğu ve tutarlılığı indekslerine ilişkin mutlak yanlılık değerlerinin test edildiği modele bağlı olarak farklılaştığı anlaşılmıştır. Model yanlış tanımlaması durumlarından, 3PLM altında üretilen veri setinin 1PLM altında test edilmesi durumunda indeks değerlerinin en yüksek mutlak yanlılık değerlerine sahip olduğu görülmüştür. 1PLM altında üretilen veri setinin 2PLM altında test edilmesi ve 2PLM altında üretilen veri setinin 3PLM altında test edilmesi durumunda da indekslerin en düşük mutlak yanlılık değerlerine sahip olduğu görülmektedir. Madde sayısının artışına bağlı olarak, model yanlış tanımlaması durumunda, sınıflama doğruluğu ve sınıflama tutarlılığı indekslerine ilişkin mutlak yanlılık değerlerinde azalma olduğu göze çarpmaktadır. Örneklem büyüklüğünün artışına bağlı olarak ise model yanlış tanımlaması durumlarında söz konusu indekslere ilişkin mutlak yanlılık değerlerinde dikkate değer bir farklılaşmanın olmadığı görülmüştür.

Sonuç ve Tartışma

Bu araştırmada, sınıflama doğruluğu ve sınıflama tutarlılığı indekslerini kestirmede MTK tabanlı yöntemlerden Rudner'ın, Guo'nun ve Lee'nin yönteminin etkililiği, 1PLM, 2PLM ve 3PLM altında farklı test uzunluğu ve örneklem büyüklükleri açısından karşılaştırmalı olarak incelenmiştir. Ayrıca, bu indekslerin doğruluğu model yanlış tanımlaması durumlarında da test edilmiştir. Bu amaç doğrultusunda ilgili veri setleri için madde ve yetenek parametreleri kestirilmiş, genel uyum ve madde uyum indeksleri hesaplanmıştır. Analizlere geçilmeden önce MTK'nin varsayımlarından tek boyutluluk ve yerel bağımsızlık varsayımlarının karşılanıp karşılanmadığı incelenmiş ve varsayımların tüm veri setleri ve modeller için karşılandığı görülmüştür. Ayrıca her bir simülasyon koşulunda üretilen veri setleri için ayrı ayrı hesaplanan marjinal güvenirlik katsayıları da veri setlerinin güvenilir ölçümler sağladığını göstermiştir. Yapılan AFA sonucunda testin tek boyutlu olduğu görülmüş ve geçerlik kanıtı olarak sunulmuştur. Sınıflama doğruluğu ve sınıflama tutarlılığı kavramlarının incelenmesi açısından, simülasyon veri setlerinin ölçümlerinin güvenilir ve geçerli olması önemli görülmektedir.

Test uzunluğunun, sınıflama doğruluğu ve sınıflama tutarlılığı indeksleri ile bunlara ilişkin mutlak yanlılık değerleri üzerinde dikkate değer bir etkisinin olduğu görülmektedir. Madde sayısı arttıkça, indekslere ilişkin değerler artmış, yanlılık değerleri azalmıştır. Bununla birlikte, örneklem büyüklüğünün sınıflama doğruluğu ve sınıflama tutarlılığı indeksleri ile bu indekslere ilişkin mutlak yanlılık değerleri üzerinde herhangi bir etkisinin olmadığı görülmüştür. Benzer şekilde, model yanlış tanımlaması durumunda da test uzunluğunun, sınıflama doğruluğu ve sınıflama tutarlılığı indekslerine ilişkin mutlak yanlılık değerleri üzerinde dikkate değer bir etkisinin olduğu görülmektedir. Madde sayısının artışına bağlı olarak mutlak yanlılık değerlerinde azalış gözlenmiştir. Örneklem büyüklüğünün ise söz konusu indekslere ilişkin mutlak yanlılık değerleri üzerinde dikkate değer bir etkisinin olmadığı görülmüştür. Lathrop ve Cheng (2013)'de çalışmalarında benzer sonuçlara ulaşmışlardır. Madde sayısının artışına bağlı olarak, mutlak yanlılık değerlerinde azalma görülmüştür fakat söz konusu değerler örneklem büyüklüğünden etkilenmemiştir. Ancak, standart hataların örneklem büyüklüğüne bağlı olarak azaldığı gözlenmiştir. Bu durum, daha büyük örneklemler için sınıflama doğruluğu kestiriminin değişkenliğinin azaldığı anlamına gelmektedir.

Genel olarak, Guo'nun yöntemine dayalı hesaplanan FP ve FN oranları ile Lee'nin yöntemine dayalı hesaplanan genel CC oranları ve kappa değerlerinin en yüksek mutlak yanlılık değerlerine sahip olduğu gözlenmektedir. Rudner'ın yöntemine dayalı hesaplanan sınıflama doğruluğu ve sınıflama tutarlılığı indekslerinin ise diğer yöntemlere kıyasla, daha düşük mutlak yanlılık değerlerine sahip olduğu görülmektedir. Bu bulgu, Chau (2018) ve Lathrop ve Chen (2013)'in çalışmalarından elde edilen bulgular ile de desteklenmektedir. Söz konusu çalışmalarda da Rudner'ın yönteminin daha iyi performans gösterdiği görülmektedir. Wyse ve Hao (2012)'nun çalışmasında ise Rudner'ın yönteminin uç değerler için daha düşük performans gösterdiği sonucuna ulaşılmıştır. Bu bulgular, Martineau (2007)'nun çalışmasındaki bulgular ile birlikte değerlendirildiğinde, Rudner'ın yönteminin büyük örneklem gruplarında ve uç değerlerin fazla olmadığı durumlarda uygulanmasının sınıflama kararları açısından daha uygun olacağı düşünülmektedir. Bu nedenle, Türkiye'de uygulanan merkezi sınavlarda sınıflama doğruluğu ve sınıflama tutarlılığının incelenmesinde büyük örneklem gruplarında uygulandığında daha iyi performans göstermesi ve diğer yöntemlere göre daha düşük mutlak yanlılık değerlerine sahip olması açısından Rudner'ın yönteminin kullanılması tercih edilebilir.

Model yanlış tanımlaması durumunda, indekslere ilişkin mutlak yanlılık değerlerinin farklılaştığı görülmektedir fakat sistematik olarak dikkate değer bir farklılık söz konusu değildir. 3PLM altında üretilen veri setinin 1PLM altında test edilmesi durumunda sınıflama doğruluğu ve tutarlılığı indekslerinin en yüksek mutlak yanlılık değerine sahip olduğu görülmüştür. Ayrıca, tüm model yanlış tanımlaması durumlarında Lee'nin indeksine dayalı hesaplanan kappa değerleri en yüksek mutlak yanlılık değerlerine sahiptir. Bu bulgular, Chau (2018)'nun çalışmasından elde edilen bulgular ile desteklenmektedir. Söz konusu çalışmada, model yanlış tanımlaması durumlarında elde edilen değerler, doğru model altında hesaplanan değerler ile benzerlik göstermektedir. Buna rağmen model yanlış tanımlaması durumunda, sınıflama kararları, theta ölçeğinin bir kalibrasyon diğerine kaymasından dolayı, her kategoride sınıflandırılan bireylerin farklı oranlarının belirtildiği durum ile aynı olmayacaktır. Örneğin, doğru MTK modeli altında theta ölçeğinde ve bir kesme puanı kullanarak uygulanan bir lisans sınavında bireylerin yaklaşık %20'si hakkında geçti kararı verilir. Ancak, sınıflama doğruluğu ve sınıflama tutarlılığı indekslerinin hesaplanmasında MTK modeli yanlış tanımlanırsa, hesaplanan indeksler minimum düzeyde etkilenmiş görünse de aslında değerlendirilmekte olan sınıflama kararları uygulamada verilenlerle aynı olmayacaktır (Chau, 2018). Ek olarak, Chau (2018)'nun çalışmasında kesme puanlarının konumu yüksek test bilgisine sahip olduğu durumda sınıflama doğruluğu ve tutarlılığı indekslerinin yüksek değerlere sahip olduğu görülmüştür. Bu çalışmada da farklı kesme puanlarında, sınıflama doğruluğu ve sınıflama tutarlılığı indekslerinin ve bunlara ilişkin mutlak yanlılık değerlerinin farklılaştığı görülmüştür. Ancak sistematik bir farklılaşma söz konusu değildir. Ayrıca bu çalışmada, veri setleri simüle edilirken standart normal dağılımdan elde edilen yetenek parametreleri kullanılmıştır. Bu sebeple bulgular, yalnızca belirtilen simülasyon koşullarında kullanılan dağılımlara sahip madde ve yetenek parametrelerine sahip testlere genellenebilir.

Öneriler

Ulusal ve uluslararası uygulanan testler ile bireylerin, belirli bir alandaki yeterlilik düzeyleri belirlenmeye çalışılmaktadır. Bu doğrultuda bu tür testlerde sınıflama doğruluğu ve sınıflama tutarlılığı indekslerinin doğruluğu önem kazanmaktadır. Bu çalışmada, simülasyon koşullarında yalnızca ikili puanlanan maddelere dayalı veri setleri dikkate alınırken, çoklu

puanlanan maddelere dayalı veri setlerinden elde edilen ölçümler için de sınıflama doğruluğu ve tutarlılığı indekslerinin doğruluğunun araştırılması pratik öneme sahiptir.

Bu çalışmada örneklem büyüklüğünün, sınıflama doğruluğu ve tutarlılığı indeksleri ile bu indekslere ilişkin mutlak yanlılık değerleri üzerinde dikkate değer bir etkisinin olmadığı görülmüştür. Bu bulgu Lathrop ve Cheng (2013)'in çalışmasından elde edilen bulgular ile desteklenmektedir fakat söz konusu çalışmada standart hataların örneklem büyüklüğüne bağlı olarak azaldığı görülmektedir. Bu durum, daha büyük örneklem grupları için sınıflama doğruluğu kestiriminin değişkenliğinin azaldığı anlamına gelmektedir. Bu anlamda daha büyük örneklem grupları ile çalışarak örneklem büyüklüğünün sınıflama doğruluğu ve tutarlılığı indeksleri üzerindeki etkisini incelemek faydalı olacaktır.

Bu çalışmada, veri setleri, standart normal dağılımdan elde edilen yetenek parametreleri ile simüle edilmiştir. Ayrıca madde parametreleri de belirli dağılımlara dayalı olarak seçilmiştir ve buna dayalı olarak veri üretimi gerçekleştirilmiştir. Yetenek parametresinin normal olmayan bir dağılımdan seçildiği ve madde parametrelerinin de farklı dağılımlardan seçildiği durumlarda veri üretilip, bu dağılımların indekslerin doğruluğu üzerindeki etkileri de araştırılmaya değerdir.

Bu çalışmada, model yanlış tanımlaması durumu, bir model altında üretilen simülasyon verisinin başka bir model altında test edilme durumu olarak ele alınmıştır. Ancak model yanlış tanımlaması durumu gelecek araştırmalarda farklı bir şekilde ele alınıp incelenebilir. Örneğin, verinin çok boyutluluğu gibi model uyumsuzluğuna yol açabilecek başka durumlar açısından da ileriki simülasyon çalışmalarında incelemeler yapılabilir, model yanlış tanımlaması farklı durumlar açısından incelenebilir.

KAYNAKLAR

Aktaş, C., & Erkuş, O. (2009). Lojistik Regresyon Analizi ile Eskişehir'in Sis Kestiriminin İncelenmesi Investigation of Fog Forecastıng of Eskişehir Using Logistic Regression Analysis. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, *8*(16), 47-61.

Artino Jr, A. R., & Stephens, J. M. (2009). Academic motivation and self-regulation: A comparative analysis of undergraduate and graduate students learning online. *The Internet and Higher Education*, *12*(3-4), 146-151.

Atar, H. Y. (2012). Resim-iş öğretmenliği özel yetenek sınavlarının sınıflama doğruluğu üzerine bir çalışma. *Eğitim ve Bilim*, *37*(163).

Atar, H. Y., Sayın, A., & Atar, B. (2013). Müzik öğretmenliği özel yetenek sınavlarının sınıflama doğruluğunun incelenmesi. *Kuram ve Uygulamada Eğitim Bilimleri*, *13*(3), 1567-1577.

Atılgan, H., Kan, A., & Doğan, N. (2017). *Eğitimde ölçme ve değerlendirme*. Ankara: Anı.

Barön, A. E. (1991). Misclassification among methods used for multiple group discrimination the effects of distributional properties. *Statistics in medicine*, *10*(5), 757-766.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. New York: Guilford.

Büyükatak, E. (2016). *PISA 2012'deki matematiğe yönelik duyuşsal özelliklerin bölge, okul türü ve cinsiyete göre sınıflama doğruluğunun incelenmesi.* (Yüksek lisans tezi). https://tez.yok.gov.tr adresinden erişilmiştir.

Ceylan, A. G. D. E. (2009). PISA 2006 sonuçlarına göre Türkiye'de fen okuryazarlığında düşük ve yüksek performans gösteren okullar arasındaki farklar. *Van Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi*, *6*(2), 55-75.

Ceylan, E., & Akerson, V. (2014). ABD'de TIMSS'e göre düşük ve yüksek performans göstermiş okulların karşılaştırılması. *Eğitim ve Bilim*, *39*(173).

Chen, J., de la Torre, J., & Zhang, Z. (2013). Relative and absolute fit evaluation in cognitive diagnosis modeling. *Journal of Educational Measurement*, *50*(2), 123-140.

Cizek, G. J. & Bunch, M. B. (2007). *Standard Setting: A Guide to Establishing and Evaluating Performance Standards on Tests*. London: Sage.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, *20*(1), 37-46.

Çelikten, S., & Çakan, M. (2019). Bayesian ve nonbayesian kestirim yöntemlerine dayalı olarak sınıflama indekslerinin TIMSS-2015 matematik testi üzerinde incelenmesi. *Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi*, *13*(1), 105-124.

Çokluk, Ö. (2010). Lojistik regresyon analizi: Kavram ve uygulama. *Kuram ve uygulamada eğitim bilimleri*, *10*(3), 1357-1407.

Chalmers, P. (2020). *Package 'mirt'*. Retrieved from https://cran.r-project.org/web/packages/mirt/mirt.pdf

Chau, L. H. (2018). *Evaluating the correctness of IRT-based methods in computing classification consistency and accuracy indices in the presence of model misspecification*. Doctoral dissertation, University of British Columbia, Vancouver.

Chen, W. H., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*, *22*(3), 265-289.

Darrell Bock, R. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, *37*(1), 29-51.

Demirel, Ö. (1999). *Kuramdan uygulamaya eğitimde program geliştirme*. Ankara: Pegem Akademi.

De Ayala, R. J. (2009). *The Theory and Practice of Item Response Theory. Methodology in the Social Sciences*. New York: Guilford.

Diao, H., & Sireci, S. G. (2018). Item response theory-based methods for estimating classification accuracy and consistency. *Journal of Applied Testing Technology*, *19*(1), 20-25.

Dillon, W. R., & Goldstein, M. (1984). *Multivariate analysis: Methods and applications*. New York: Wiley.

Divjak, B., & Oreski, D. (2009). Prediction of academic performance using discriminant analysis. *Proceedings of the ITI 2009 31st International Conference on Information Technology Interfaces*, 225-230.

Embretson, S. E., & Reise, S. P. (2013). *Item response theory*. New York: Psychology Press.

Erkuş, A. (2013). *Davranış bilimleri için bilimsel araştırma süreci*. Ankara: Seçkin

Ertürk, Z. (2016). *Ölçeklerin faktör yapısını belirlemede kullanılan açımlayıcı faktör analizi ve kümeleme analizi ile verilerin sınıflandırılmasında kullanılan diskriminant ve lojistik regresyon analizi tekniklerinin karşılaştırılması.* (Yüksek lisans tezi). https://tez.yok.gov.tr adresinden erişilmiştir.

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to design and evaluate research in education*. New York: McGraw-hill.

Fleck, M. P. D. A., Poirier-Littre, M. F., Guelfi, J. D., Bourdel, M. C., & Loo, H. (1995). Factorial structure of the 17-item Hamilton depression rating scale. *Acta Psychiatrica Scandinavica*, *92*(3), 168-172.

Guo, F. (2006). Expected classification accuracy using the latent distribution. *Practical Assessment, Research, and Evaluation*, *11*(1), 6.

Güzeller, C. O. & Kelecioğlu, H. (2006). Ortaöğretim kurumları öğrenci seçme sınavının sınıflama geçerliği üzerine bir çalışma. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 30*(30), 140-148.

Hambleton, R. K., & Novick, M. R. (1973). Toward an integration of theory and method for criterion referenced tests 1, 2. *Journal of Educational Measurement, 10*(3), 159-170.

Hambleton, R. K., & Swaminathan, H. (1985). I*tem Response Theory: Principles and Applications.* Boston: Kluwer Nijhoff.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). F*undamentals of item response theory*. London: Sage.

Hanson, B. A., & Brennan, R. L. (1990). An investigation of classification consistency indexes estimated under alternative strong true score models. *Journal of Educational Measurement, 27*(4), 345-359.

Houben, R. M., Leeuw, M., Vlaeyen, J. W., Goubert, L., & Picavet, H. S. J. (2005). Fear of movement/injury in the general population: factor structure and psychometric properties of an adapted version of the Tampa Scale for Kinesiophobia. *Journal of behavioral medicine, 28*(5), 415-424.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal, 6*(1), 1-55.

Huynh, H. (1976). On the reliability of decisions in domain-referenced testing. *Journal of educational measurement,* 253-264.

Huynh, H. (1990). Computation and statistical inference for decision consistency indexes based on the Rasch model. *Journal of Educational Statistics, 15*(4), 353-368.

Kalender, İ. (2015). Üstün başarılı öğrenci profilinin PISA 2012'ye dayalı olarak analizi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 6*(1), 158- 172.

Kılıç, S. (2000). *Lojistik regresyon analizi ve pazarlama araştırmalarında bir uygulama.* Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.

Kingsbury, G. G., & Weiss, D. J. (1980). *A Comparison of Adaptive Sequential, and Conventional Testing Strategies for Mastery Decisions*. Retrieved from https://apps.dtic.mil/sti/pdfs/ADA094478.pdf

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. New York: Guilford.

Kumandaş, H., & Kutlu, O. (2010). High stakes testing: does secondary education examination involve any risks? *Procedia-Social and Behavioral Sciences, 9*, 758-764.

Kumandaş, H. (2013). *Yükseköğretime öğrenci seçmede ve yerleştirmede kullanılan sınavların ortaöğretime devam eden öğrencilerin okul başarıları üzerinde oluşturduğu risk faktörleri.* Doktora Tezi, Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.

Kutlu, Ö., Yıldırım, Ö., Bilican, S. ve Kumandaş, H. (2011). İlköğretim 5. sınıf öğrencilerinin okuduğunu anlamada başarılı olup-olmama durumlarının kestirilmesinde etkili olan değişkenlerin incelenmesi. *Eğitimde ve Psikolojide Ölçme Değerlendirme Dergisi, 2*(1), 132-139.

Lathrop, Q. N., & Cheng, Y. (2013). Two approaches to estimation of classification accuracy rate under item response theory. *Applied Psychological Measurement, 37*(3), 226-241.

Lathrop, Q. N., & Cheng, Y. (2014). A nonparametric approach to estimate classification accuracy and consistency. *Journal of Educational Measurement, 51*(3), 318-334.

Lee, W. C. (2010). Classification consistency and accuracy for complex assessments using item response theory. *Journal of Educational Measurement, 47*(1), 1-17.

Lee, W. C., Brennan, R. L., & Wan, L. (2009). Classification consistency and accuracy for complex assessments under the compound multinomial model. *Applied Psychological Measurement, 33*(5), 374-390.

Lee, W. C., Hanson, B. A., & Brennan, R. L. (2000). *Procedures for Computing Classification Consistency and Accuracy Indices with Multiple Categories*. Retrieved from https://www.act.org/content/dam/act/unsecured/documents/ACT_RR2000-10.pdf

Lee, W. C., Hanson, B. A., & Brennan, R. L. (2002). Estimating consistency and accuracy indices for multiple classifications. *Applied Psychological Measurement, 26*(4), 412-432.

Li, S. (2006). *Evaluating the consistency and accuracy of proficiency classifications using item response theory*. University of Massachusetts Amherst.

Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of educational measurement, 32*(2), 179-197.

Lord, F. M., & Wingersky, M. S. (1984). Comparison of IRT true-score and equipercentile observed-score" equatings". *Applied Psychological Measurement, 8*(4), 453-461.

Martineau, J. A. (2007). An expansion and practical evaluation of expected classification accuracy. *Applied Psychological Measurement, 31*(3), 181-194.

Maydeu-Olivares, A., Cai, L., & Hernández, A. (2011). Comparing the fit of item response theory and factor analysis models. *Structural Equation Modeling: A Multidisciplinary Journal, 18*(3), 333-356.

Maydeu-Olivares, A., & Joe, H. (2005). Limited-and full-information estimation and goodness-of-fit testing in 2 n contingency tables: A unified framework. *Journal of the American Statistical Association, 100*(471), 1009-1020.

Maydeu-Olivares, A., & Joe, H. (2006). Limited information goodness-of-fit testing in multidimensional contingency tables. *Psychometrika, 71*(4), 713-732.

Maydeu-Olivares, A., & Joe, H. (2014). Assessing approximate fit in categorical data analysis. *Multivariate Behavioral Research, 49*(4), 305-328.

McCoach, D. B., & Siegle, D. (2003). Factors that differentiate under achieving gifted students from high-achieving gifted students. *Gifted Child Quartetly*, *47*(2), 144-154

Md Desa, Z. N. D. (2012). *Bi-factor multidimensional item response theory modeling for subscores estimation, reliability, and classification* (Doctoral dissertation). Retrieved from https://kuscholarworks.ku.edu/handle/1808/10126

Minchen, N., & de la Torre, J. (2018). A general cognitive diagnosis model for continuous-response data. *Measurement: Interdisciplinary Research and Perspectives, 16*(1), 30-44.

Minchen, N. D., de la Torre, J., & Liu, Y. (2017). A cognitive diagnosis model for continuous response. *Journal of Educational and Behavioral Statistics, 42*(6), 651-677.

Onu, J. O., & John, O. O. (2016). An Application of Discriminant Analysis to the Classification of Students on the Basis of their Academic Performances. *Statistics*, *56*, 4.

Özer Özkan, Y. (2016). Examining the Effective Variables on Classification of School's Success through PISA 2012 Turkey Data. *International Online Journal of Educational Sciences*.

Özkan, Y. Ö. ve Doğan, B. (2013). İlköğretim 8. sınıf öğrencilerinin okuma becerilerinin kestirilmesinde etkili olan değişkenlerin belirlenmesi. *The Journal of Academic Social Science Studies, 6*(4), 667-680.

Özkan, Y. Ö. ve Anıl, D. (2014). Öğretmen mesleki gelişim değişkenlerinin ayırt edicilik düzeyi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 29*(4), 205- 216.

Özkan, Y. Ö. & Turan, S. (2021). Düşündürdükleri ve ötesiyle yüksek riskli sınav gerçeği. *Alanyazın, 2*(1), 59-63.

Partchev, I. (2017). *Package 'irtoys'*. Retrieved from https://cran.r-project.org/web/packages/irtoys/irtoys.pdf

Peng, C. Y. J., Lee, K. L., & Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *The journal of educational research*, *96*(1), 3-14.

Rausch, J. R., & Kelley, K. (2009). A comparison of linear and mixture models for discriminant analysis under nonnormality. *Behavior Research Methods, 41*(1), 85-98.

Reise, S. P. (1990). A comparison of item-and person-fit methods of assessing model-data fit in IRT. *Applied Psychological Measurement, 14*(2), 127-137.

Revelle, W. (2015). *Package 'psych'*. Retrieved from https://cran.r-project.org/web/packages/psych/psych.pdf

Rizopoulos, D. (2018). *Package 'ltm'*. Retrieved from https://cran.r-project.org/web/packages/ltm/ltm.pdf

Robitzsch, A. (2020). *Package 'sirt'*. Retrieved from https://cran.r-project.org/web/packages/sirt/sirt.pdf

Romera, I., Delgado-Cohen, H., Perez, T., Caballero, L., & Gilaberte, I. (2008). Factor analysis of the Zung self-rating depression scale in a large sample of patients with major depressive disorder in primary care. *BMC psychiatry*, *8*(1), 1-8.

Rudner, L. M. (2001). Computing the expected proportions of misclassified examinees. *Practical Assessment, 7*(14), 1-5.

Rudner, L. M. (2005). Expected classification accuracy. *Practical Assessment, 10*(13), 1-4.

Sen, S., & Cohen, A. S. (2020). The impact of test and sample characteristics on model selection and classification accuracy in the multilevel mixture IRT model. *Frontiers in psychology, 11*, 197.

Schulz, E. M., Kolen, M. J., & Nicewander, W. A. (1999). A rationale for defining achievement levels using IRT-estimated domain scores. *Applied Psychological Measurement, 23*(4), 347-362.

Subkoviak, M. J. (1976). Estimating reliability from a single administration of a criterion-referenced test. *Journal of Educational Measurement,* 265-276.

Swaminathan, H., Hambleton, R. K., & Algina, J. (1974). Reliability of criterion-referenced tests: A decision-theoretic formulation. *Journal of Educational Measurement, 11*(4), 263-267.

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2007). *Using multivariate statistics*. Boston: Pearson.

Taşdemir, F., & Şahin, D. B. (2016). The examination of the assessment and evaluation course achievements of university students based on discriminant and logistic regression analysis. *SHS Web of Conferences*, 26.

Terzi, R., & De la Torre, J. (2018). An iterative method for empirically-based Q-matrix validation. *International Journal of Assessment Tools in Education, 5*(2), 248-262.

Thissen, D., Pommerich, M., Billeaud, K., & Williams, V. S. (1995). Item response theory for scores on tests including polytomous items with ordered responses. *Applied Psychological Measurement, 19*(1), 39-49.

Thompson, N. A. (2009). Item selection in computerized classification testing. *Educational and Psychological Measurement, 69*(5), 778-793.

Wang, T., Kolen, M. J., & Harris, D. J. (2000). Psychometric properties of scale scores and performance levels for performance assessments using polytomous IRT. *Journal of Educational measurement, 37*(2), 141-162.

Wang, S., & Wang, T. (2001). Precision of Warm's weighted likelihood estimates for a polytomous model in computerized adaptive testing. *Applied Psychological Measurement, 25*(4), 317-331.

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika, 54*(3), 427-450.

Wyse, A. E., & Hao, S. (2012). An evaluation of item response theory classification accuracy and consistency indices. *Applied Psychological Measurement, 36*(7), 602-624.

Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement, 8*(2), 125-145.

EKLER

EK 1. Bilinen Madde Parametreleri

K=15 İçin Bilinen Madde Parametreleri

	1PLM	2PLM		3PLM		
	b	b	a	b	a	c
1	-1,34	-0,79	1,74	1,56	1,46	0,21
2	-0,85	1,75	0,68	-0,18	1,19	0,17
3	1,84	0,55	0,95	-1,96	1,48	0,23
4	-0,39	0,41	1,03	0,77	1,53	0,26
5	-0,31	-1,31	1,95	-0,61	0,85	0,23
6	2,07	-2,78	1,95	0,32	1,93	0,23
7	0,19	-2,24	1,30	-2,50	1,74	0,18
8	-2,40	-4,23	1,00	1,00	1,84	0,19
9	-1,53	-2,89	1,48	0,45	0,95	0,14
10	-1,17	-0,64	1,42	1,72	1,87	0,12
11	1,34	-0,11	0,52	1,03	1,21	0,19
12	0,04	-2,54	1,65	1,20	0,73	0,23
13	0,10	-0,77	1,04	-0,44	1,22	0,21
14	-0,33	-1,34	0,65	0,45	1,02	0,24
15	-1,33	-1,41	1,34	-0,38	1,67	0,29

K=30 İçin Bilinen Madde Parametreleri

	1PLM	2PLM		3PLM		
	b	b	a	b	a	c
1	-1,34	0,51	1,74	-2,05	1,46	0,22
2	-0,85	1,98	0,68	-0,68	1,19	0,11
3	1,84	1,36	0,95	-1,23	1,48	0,26
4	-0,39	-2,84	1,03	-0,17	1,53	0,18
5	-0,31	1,27	1,95	-0,43	0,85	0,24
6	2,07	0,36	1,95	-0,86	1,93	0,21
7	0,19	1,21	1,30	-1,55	1,74	0,17
8	-2,40	-2,36	1,00	0,05	1,84	0,11
9	-1,53	0,53	1,48	0,57	0,95	0,19
10	-1,17	-1,28	1,42	-0,32	1,87	0,18
11	1,34	0,04	0,52	0,88	1,21	0,26
12	0,04	-3,26	1,65	-1,20	0,73	0,25
13	0,10	0,34	1,04	-0,16	1,22	0,25
14	-0,33	0,84	0,65	-0,97	1,02	0,25
15	-1,33	0,09	1,34	-2,67	1,67	0,26
16	2,18	-0,39	1,14	-2,63	1,87	0,20
17	0,25	3,10	1,13	-3,08	0,95	0,23
18	-3,45	-1,84	1,90	-0,35	1,38	0,22
19	0,55	0,65	1,51	-2,32	1,19	0,16
20	-1,21	1,84	1,64	-2,89	0,75	0,22
21	-2,10	1,20	1,06	-3,69	0,78	0,17
22	-0,83	-0,88	1,59	0,06	1,70	0,22
23	-2,04	-2,16	1,53	-0,54	1,78	0,19
24	-1,59	1,28	0,94	0,25	1,21	0,21
25	-1,44	-3,09	1,37	-1,01	1,15	0,34
26	-3,03	-0,26	0,60	0,86	1,56	0,22
27	0,76	-1,50	1,64	-3,27	1,26	0,21
28	-0,27	0,96	0,68	-0,71	0,64	0,25
29	-2,21	-2,65	0,75	-0,97	1,12	0,14
30	1,38	0,60	0,51	0,06	1,76	0,17

EK 2. Üretilen Veri Setlerinin 1PLM, 2PLM ve 3PLM İçin Yerel Bağımsızlık Testi Sonuçları

15 Madde ve 500 Örneklem Büyüklüğüne Sahip Veri Setlerinin 1PLM, 2PLM ve 3PLM İçin Yerel Bağımsızlık Testi Sonuçları

Madde çiftleri		1PLM	2PLM	3PLM
Madde 1	Madde 2	Q_3	Q_3	Q_3
m1	m2	-0,146	-0,167	-0,118
m1	m3	0,017	-0,104	0,005
m1	m4	-0,083	-0,088	-0,126
m1	m5	-0,036	-0,034	-0,090
m1	m6	-0,033	0,035	-0,107
m1	m7	-0,045	-0,005	0,000
m1	m8	-0,028	-0,065	-0,113
m1	m9	-0,096	0,033	-0,107
m1	m10	-0,078	-0,134	-0,027
m1	m11	0,009	-0,107	-0,053
m1	m12	-0,177	0,058	-0,087
m1	m13	-0,074	-0,087	-0,023
m1	m14	-0,059	-0,135	-0,126
m1	m15	-0,088	-0,070	-0,054
m2	m3	-0,060	-0,154	-0,022
m2	m4	-0,077	-0,075	-0,044
m2	m5	-0,071	-0,073	-0,099
m2	m6	-0,138	-0,012	-0,063
m2	m7	-0,062	-0,012	-0,009
m2	m8	-0,022	0,022	-0,115
m2	m9	-0,088	-0,057	-0,040
m2	m10	-0,101	-0,101	-0,087
m2	m11	-0,043	-0,056	-0,095
m2	m12	-0,115	-0,022	-0,074
m2	m13	-0,057	-0,169	-0,087
m2	m14	-0,070	-0,113	-0,112
m2	m15	-0,036	-0,084	-0,101
m3	m4	-0,078	-0,112	-0,034
m3	m5	0,026	-0,053	-0,046
m3	m6	-0,117	-0,045	-0,105
m3	m7	-0,137	-0,021	0,030
m3	m8	-0,023	-0,013	-0,054
m3	m9	-0,016	-0,011	-0,039
m3	m10	0,009	-0,102	-0,023
m3	m11	-0,182	-0,163	-0,050
m3	m12	-0,123	-0,015	-0,033
m3	m13	-0,095	-0,026	-0,028
m3	m14	0,003	-0,179	-0,069
m3	m15	-0,081	-0,045	-0,079

m4	m5	-0,152	-0,060	-0,080
m4	m6	-0,044	-0,007	-0,082
m4	m7	-0,072	-0,092	-0,020
m4	m8	0,020	0,029	-0,097
m4	m9	-0,069	-0,081	-0,061
m4	m10	-0,055	-0,119	-0,141
m4	m11	-0,125	-0,147	-0,057
m4	m12	-0,049	-0,011	-0,036
m4	m13	-0,097	-0,058	-0,117
m4	m14	-0,120	-0,186	-0,074
m4	m15	-0,059	-0,113	-0,136
m5	m6	-0,062	-0,104	-0,079
m5	m7	-0,112	-0,039	0,024
m5	m8	-0,138	-0,069	0,028
m5	m9	-0,016	-0,093	-0,154
m5	m10	-0,080	-0,059	-0,032
m5	m11	-0,141	-0,085	-0,060
m5	m12	-0,033	-0,022	-0,074
m5	m13	-0,092	-0,092	-0,075
m5	m14	-0,079	-0,059	-0,037
m5	m15	-0,086	-0,082	-0,134
m6	m7	-0,024	0,148	-0,060
m6	m8	0,019	-0,037	-0,029
m6	m9	-0,117	-0,005	-0,032
m6	m10	-0,013	-0,005	-0,047
m6	m11	-0,030	0,012	-0,106
m6	m12	-0,051	-0,009	-0,130
m6	m13	-0,033	-0,018	-0,057
m6	m14	-0,085	-0,136	-0,030
m6	m15	-0,035	-0,156	-0,038
m7	m8	0,004	-0,016	-0,062
m7	m 9	-0,005	-0,146	-0,041
m7	m10	-0,117	0,022	-0,057
m7	m11	-0,132	-0,095	-0,094
m7	m12	-0,076	-0,171	-0,008
m7	m13	-0,156	-0,045	-0,025
m7	m14	-0,017	-0,083	0,020
m7	m15	-0,063	-0,126	-0,042
m8	m9	-0,060	-0,090	-0,061
m8	m10	-0,089	-0,064	-0,070
m8	m11	0,010	-0,009	-0,099
m8	m12	-0,080	-0,003	-0,164
m8	m13	-0,075	-0,087	-0,092
m8	m14	-0,152	0,038	-0,044
m8	m15	0,016	-0,042	-0,005
m9	m10	-0,184	0,104	-0,150
m9	m11	-0,065	-0,113	-0,149
m9	m12	-0,045	-0,081	-0,061

m9	m13	-0,095	-0,076	-0,062
m9	m14	-0,040	-0,116	-0,138
m9	m15	-0,044	0,116	-0,075
m10	m11	-0,018	-0,097	-0,022
m10	m12	-0,045	-0,068	-0,076
m10	m13	-0,072	-0,154	-0,117
m10	m14	-0,055	-0,072	-0,114
m10	m15	-0,106	-0,149	0,034
m11	m12	-0,089	-0,043	-0,103
m11	m13	0,014	-0,195	-0,108
m11	m14	-0,124	-0,064	-0,088
m11	m15	-0,053	-0,117	-0,061
m12	m13	-0,135	-0,036	-0,140
m12	m14	-0,046	-0,067	-0,033
m12	m15	-0,058	-0,037	-0,113
m13	m14	-0,105	-0,072	-0,013
m13	m15	-0,063	-0,051	-0,015
m14	m15	-0,178	-0,043	-0,119

30 Madde ve 500 Örneklem Büyüklüğüne Sahip Veri Setlerinin 1PLM, 2PLM ve 3PLM İçin Yerel Bağımsızlık Testi Sonuçları

Madde çiftleri		1PLM	2PLM	3PLM
Madde 1	Madde 2	Q_3	Q_3	Q_3
m1	m2	-0,115	-0,163	0,028
m1	m3	-0,032	-0,031	-0,062
m1	m4	-0,059	-0,005	-0,043
m1	m5	0,007	-0,023	0,080
m1	m6	-0,050	0,023	-0,018
m1	m7	-0,025	-0,123	-0,102
m1	m8	-0,031	0,019	-0,052
m1	m9	-0,082	0,047	0,043
m1	m10	-0,044	-0,034	-0,004
m1	m11	0,060	-0,085	-0,030
m1	m12	0,020	-0,069	0,046
m1	m13	-0,088	-0,022	0,018
m1	m14	0,063	-0,044	-0,031
m1	m15	-0,057	0,055	-0,025
m1	m16	-0,087	0,051	0,051
m1	m17	-0,102	-0,047	-0,069
m1	m18	0,038	-0,011	0,071
m1	m19	-0,053	-0,087	-0,102
m1	m20	-0,112	-0,051	-0,168
m1	m21	-0,090	-0,030	-0,010
m1	m22	-0,046	0,030	-0,030
m1	m23	-0,076	-0,029	-0,031
m1	m24	0,046	-0,051	-0,033
m1	m25	0,005	0,013	-0,169
m1	m26	-0,038	-0,122	-0,018
m1	m27	-0,045	0,034	0,003
m1	m28	-0,050	-0,081	-0,022
m1	m29	0,038	-0,049	-0,005
m1	m30	0,003	-0,067	-0,106
m2	m3	-0,042	-0,047	0,042
m2	m4	-0,027	-0,071	-0,027
m2	m5	0,046	-0,105	-0,113
m2	m6	-0,053	-0,081	-0,093
m2	m7	-0,031	-0,015	-0,031
m2	m8	0,076	0,029	-0,020
m2	m9	-0,054	-0,115	-0,035
m2	m10	0,030	0,043	-0,108
m2	m11	0,005	-0,064	-0,095
m2	m12	-0,060	-0,015	0,007
m2	m13	-0,034	-0,065	-0,160
m2	m14	-0,026	-0,086	-0,038
m2	m15	0,005	0,052	0,116
m2	m16	-0,062	-0,006	-0,063

m2	m17	-0,032	-0,078	0,046
m2	m18	-0,087	-0,012	0,010
m2	m19	-0,042	-0,049	-0,012
m2	m20	-0,036	-0,016	-0,021
m2	m21	-0,044	-0,094	0,010
m2	m22	-0,081	-0,063	-0,064
m2	m23	-0,092	0,039	-0,045
m2	m24	-0,075	-0,136	-0,037
m2	m25	-0,006	-0,031	0,046
m2	m26	0,038	-0,044	-0,098
m2	m27	-0,098	0,036	0,032
m2	m28	-0,058	-0,006	-0,001
m2	m29	-0,095	0,007	-0,092
m2	m30	-0,080	0,016	-0,087
m3	m4	-0,014	-0,015	-0,059
m3	m5	0,004	0,027	-0,079
m3	m6	-0,038	-0,071	0,115
m3	m7	-0,042	0,031	-0,001
m3	m8	0,006	-0,115	-0,031
m3	m9	0,023	-0,129	0,007
m3	m10	-0,129	-0,031	-0,111
m3	m11	-0,038	-0,075	-0,152
m3	m12	-0,048	-0,024	-0,092
m3	m13	-0,119	-0,091	-0,029
m3	m14	-0,080	-0,073	0,002
m3	m15	-0,036	0,010	0,043
m3	m16	0,010	-0,041	0,055
m3	m17	-0,054	-0,049	0,075
m3	m18	-0,008	-0,039	-0,058
m3	m19	0,041	-0,060	-0,077
m3	m20	-0,041	-0,077	-0,105
m3	m21	-0,010	-0,070	-0,029
m3	m22	-0,110	-0,003	0,026
m3	m23	0,024	-0,042	-0,002
m3	m24	0,019	0,029	0,034
m3	m25	-0,058	0,017	-0,058
m3	m26	0,019	0,004	-0,090
m3	m27	-0,118	-0,058	0,033
m3	m28	0,006	-0,034	-0,075
m3	m29	0,006	-0,081	0,060
m3	m30	-0,043	0,007	-0,068
m4	m5	-0,081	0,024	-0,059
m4	m6	-0,002	0,005	-0,063
m4	m7	-0,041	-0,067	-0,004
m4	m8	-0,023	0,024	-0,041
m4	m9	-0,061	0,057	-0,039
m4	m10	-0,057	-0,066	-0,082
m4	m11	0,073	0,001	-0,001

m4	m12	-0,047	-0,069	-0,041
m4	m13	0,003	-0,090	-0,164
m4	m14	-0,111	-0,003	-0,133
m4	m15	-0,075	-0,105	0,017
m4	m16	-0,091	-0,008	-0,008
m4	m17	-0,089	-0,008	0,033
m4	m18	-0,049	0,069	0,038
m4	m19	-0,072	0,051	0,062
m4	m20	0,074	0,030	-0,019
m4	m21	0,030	0,052	-0,002
m4	m22	-0,003	-0,063	-0,042
m4	m23	0,039	-0,051	-0,014
m4	m24	0,052	0,013	-0,092
m4	m25	-0,035	-0,113	-0,040
m4	m26	0,037	-0,074	-0,011
m4	m27	-0,103	-0,126	0,027
m4	m28	-0,082	0,031	-0,022
m4	m29	-0,161	-0,128	-0,069
m4	m30	-0,042	-0,109	-0,101
m5	m6	-0,059	-0,005	-0,127
m5	m7	-0,046	0,086	-0,044
m5	m8	0,028	-0,005	-0,073
m5	m9	0,017	-0,014	0,000
m5	m10	-0,057	-0,002	-0,022
m5	m11	-0,071	-0,070	0,016
m5	m12	-0,118	0,006	0,004
m5	m13	-0,086	-0,037	0,045
m5	m14	-0,036	-0,090	-0,002
m5	m15	-0,086	0,010	-0,060
m5	m16	0,027	-0,030	-0,002
m5	m17	0,033	-0,003	-0,015
m5	m18	-0,034	-0,016	0,003
m5	m19	-0,065	0,064	-0,062
m5	m20	-0,061	0,043	0,013
m5	m21	-0,126	-0,023	-0,047
m5	m22	0,027	-0,063	-0,116
m5	m23	-0,028	-0,060	-0,050
m5	m24	-0,023	-0,035	-0,008
m5	m25	-0,145	-0,004	-0,168
m5	m26	-0,024	-0,056	-0,059
m5	m27	-0,003	-0,006	-0,063
m5	m28	-0,090	-0,077	-0,053
m5	m29	-0,013	-0,042	-0,080
m5	m30	-0,095	-0,085	-0,140
m6	m7	-0,068	-0,070	0,067
m6	m8	-0,033	-0,007	-0,022
m6	m9	-0,068	0,143	-0,151
m6	m10	-0,011	-0,069	0,029

m6	m11	-0,088	-0,016	-0,087
m6	m12	-0,113	0,024	0,013
m6	m13	0,019	0,031	-0,004
m6	m14	0,013	-0,085	-0,029
m6	m15	0,057	-0,177	0,001
m6	m16	0,002	-0,042	0,011
m6	m17	-0,084	-0,057	0,020
m6	m18	-0,004	-0,004	-0,102
m6	m19	0,015	-0,010	0,015
m6	m20	-0,019	-0,028	-0,104
m6	m21	-0,058	-0,079	-0,093
m6	m22	-0,011	0,016	0,007
m6	m23	0,028	0,044	-0,025
m6	m24	-0,007	-0,051	-0,061
m6	m25	0,002	0,012	0,037
m6	m26	-0,015	-0,074	-0,039
m6	m27	-0,003	-0,003	-0,020
m6	m28	-0,066	-0,096	-0,082
m6	m29	0,049	-0,050	0,019
m6	m30	-0,114	-0,107	0,014
m7	m8	-0,111	0,035	-0,003
m7	m9	-0,083	-0,025	-0,068
m7	m10	-0,079	-0,040	-0,084
m7	m11	0,005	-0,158	-0,020
m7	m12	-0,016	0,020	-0,085
m7	m13	0,008	-0,093	-0,001
m7	m14	-0,076	-0,102	-0,019
m7	m15	0,010	-0,059	-0,036
m7	m16	-0,076	-0,055	0,049
m7	m17	-0,050	-0,048	-0,050
m7	m18	-0,106	-0,026	0,014
m7	m19	-0,012	0,064	0,035
m7	m20	-0,033	-0,037	-0,014
m7	m21	-0,063	-0,044	-0,060
m7	m22	-0,049	-0,015	-0,024
m7	m23	0,016	-0,006	0,006
m7	m24	-0,063	0,074	-0,116
m7	m25	-0,060	0,017	0,020
m7	m26	-0,020	-0,095	0,001
m7	m27	-0,063	0,013	-0,007
m7	m28	0,003	-0,052	-0,029
m7	m29	-0,096	-0,067	-0,076
m7	m30	-0,016	0,006	0,076
m8	m 9	0,005	0,010	-0,053
m8	m10	-0,047	-0,096	0,023
m8	m11	-0,048	-0,048	-0,065
m8	m12	-0,061	-0,017	-0,050
m8	m13	0,020	-0,054	-0,013

m8	m14	0,014	-0,019	-0,026
m8	m15	-0,059	-0,048	-0,040
m8	m16	0,045	0,039	0,023
m8	m17	-0,014	-0,009	-0,142
m8	m18	-0,039	-0,065	-0,094
m8	m19	-0,050	-0,021	0,063
m8	m20	-0,025	0,023	-0,073
m8	m21	-0,152	0,035	-0,003
m8	m22	-0,107	-0,152	0,002
m8	m23	-0,051	0,058	-0,056
m8	m24	0,011	-0,120	-0,104
m8	m25	-0,058	0,059	-0,009
m8	m26	-0,026	-0,042	-0,100
m8	m27	-0,052	-0,154	-0,016
m8	m28	0,007	-0,050	-0,015
m8	m29	-0,015	-0,066	0,010
m8	m30	-0,042	0,025	-0,028
m9	m10	-0,124	-0,164	-0,137
m9	m11	-0,036	-0,116	-0,102
m9	m12	-0,062	-0,014	-0,077
m9	m13	0,017	0,021	-0,101
m9	m14	-0,029	-0,029	-0,045
m9	m15	0,038	-0,105	-0,103
m9	m16	-0,033	-0,041	-0,104
m9	m17	-0,044	0,036	0,020
m9	m18	0,063	0,008	0,005
m9	m19	-0,046	-0,121	-0,034
m9	m20	-0,100	-0,026	-0,027
m9	m21	-0,037	-0,073	0,078
m9	m22	-0,054	-0,003	-0,092
m9	m23	0,001	-0,016	-0,027
m9	m24	-0,061	-0,083	-0,050
m9	m25	-0,064	-0,005	-0,009
m9	m26	-0,074	-0,052	-0,053
m9	m27	0,015	0,031	-0,012
m9	m28	-0,027	-0,044	-0,038
m9	m29	0,031	0,028	-0,047
m9	m30	-0,024	-0,109	-0,079
m10	m11	-0,004	0,015	-0,125
m10	m12	-0,059	0,005	-0,082
m10	m13	-0,057	-0,098	-0,051
m10	m14	0,025	-0,013	-0,014
m10	m15	0,002	-0,050	-0,057
m10	m16	-0,002	0,042	0,041
m10	m17	-0,069	-0,038	-0,003
m10	m18	-0,084	-0,024	-0,090
m10	m19	-0,015	-0,009	-0,010
m10	m20	0,026	0,017	0,070

m10	m21	-0,005	0,006	0,069
m10	m22	-0,149	-0,151	-0,019
m10	m23	-0,045	-0,023	-0,054
m10	m24	-0,037	-0,014	-0,055
m10	m25	-0,028	-0,067	-0,101
m10	m26	-0,032	-0,023	0,001
m10	m27	0,036	-0,109	0,018
m10	m28	-0,009	-0,086	0,012
m10	m29	-0,075	-0,018	0,002
m10	m30	-0,034	0,056	0,028
m11	m12	-0,010	-0,017	0,008
m11	m13	-0,022	-0,054	0,080
m11	m14	-0,033	-0,015	-0,066
m11	m15	-0,039	-0,066	-0,010
m11	m16	-0,086	-0,062	-0,013
m11	m17	-0,092	-0,039	-0,057
m11	m18	0,021	0,010	-0,075
m11	m19	-0,083	-0,055	0,008
m11	m20	0,024	-0,076	0,056
m11	m21	-0,028	-0,041	-0,034
m11	m22	-0,139	-0,103	-0,048
m11	m23	-0,073	-0,003	-0,095
m11	m24	-0,043	-0,059	-0,094
m11	m25	-0,076	-0,011	-0,074
m11	m26	0,018	-0,011	0,053
m11	m27	-0,090	-0,047	-0,057
m11	m28	-0,007	-0,005	-0,102
m11	m29	-0,013	-0,072	-0,102
m11	m30	-0,032	-0,020	-0,056
m12	m13	-0,092	-0,003	-0,081
m12	m14	0,026	0,040	-0,021
m12	m15	-0,061	0,003	-0,020
m12	m16	-0,004	-0,021	0,034
m12	m17	-0,034	0,005	0,096
m12	m18	0,002	-0,027	-0,086
m12	m19	-0,105	-0,027	-0,131
m12	m20	-0,099	0,002	0,058
m12	m21	-0,065	0,019	-0,030
m12	m22	0,017	0,062	-0,056
m12	m23	-0,063	-0,070	-0,096
m12	m24	-0,088	0,024	-0,094
m12	m25	0,010	0,010	-0,078
m12	m26	0,014	-0,013	-0,036
m12	m27	0,029	0,024	0,020
m12	m28	-0,023	-0,040	-0,007
m12	m29	-0,004	-0,037	-0,081
m12	m30	-0,007	-0,088	-0,038
m13	m14	-0,077	-0,069	0,032

m13	m15	-0,042	-0,011	-0,047
m13	m16	-0,078	-0,082	-0,050
m13	m17	-0,053	0,007	-0,034
m13	m18	-0,052	-0,024	0,006
m13	m19	-0,017	-0,097	-0,053
m13	m20	-0,062	-0,047	-0,060
m13	m21	0,003	-0,076	-0,047
m13	m22	-0,064	0,071	-0,044
m13	m23	-0,070	-0,098	0,002
m13	m24	-0,038	-0,050	-0,106
m13	m25	-0,017	0,007	-0,054
m13	m26	0,018	-0,075	-0,008
m13	m27	-0,004	-0,014	-0,021
m13	m28	-0,047	-0,027	-0,148
m13	m29	0,033	0,102	-0,033
m13	m30	-0,097	-0,098	-0,055
m14	m15	-0,080	-0,016	-0,100
m14	m16	-0,044	-0,065	0,027
m14	m17	-0,081	0,081	-0,009
m14	m18	-0,031	-0,025	-0,121
m14	m19	-0,001	-0,115	-0,034
m14	m20	-0,028	-0,078	-0,067
m14	m21	-0,034	-0,083	-0,068
m14	m22	-0,025	-0,050	-0,047
m14	m23	0,033	-0,042	-0,069
m14	m24	-0,082	-0,076	0,054
m14	m25	-0,124	0,016	-0,080
m14	m26	-0,002	-0,006	-0,076
m14	m27	-0,153	-0,029	0,008
m14	m28	-0,048	-0,024	-0,090
m14	m29	-0,015	-0,008	-0,055
m14	m30	-0,031	-0,037	0,047
m15	m16	0,028	-0,056	-0,103
m15	m17	-0,032	0,017	0,011
m15	m18	-0,046	0,013	0,059
m15	m19	-0,067	-0,129	-0,062
m15	m20	0,016	0,058	-0,053
m15	m21	0,063	-0,036	-0,053
m15	m22	-0,018	0,024	-0,016
m15	m23	-0,066	-0,023	0,016
m15	m24	-0,070	-0,057	0,015
m15	m25	0,035	-0,016	0,010
m15	m26	-0,045	-0,072	0,065
m15	m27	-0,036	-0,050	0,003
m15	m28	-0,100	-0,002	-0,034
m15	m29	-0,103	-0,053	0,011
m15	m30	-0,170	-0,143	-0,045
m16	m17	-0,027	0,016	0,084

m16	m18	-0,090	-0,070	0,052
m16	m19	-0,084	-0,082	0,026
m16	m20	0,013	-0,049	-0,029
m16	m21	0,020	0,009	-0,090
m16	m22	-0,080	-0,035	-0,023
m16	m23	-0,054	-0,052	-0,059
m16	m24	-0,013	-0,109	-0,017
m16	m25	-0,061	0,011	0,034
m16	m26	-0,104	-0,143	-0,086
m16	m27	0,041	-0,069	0,094
m16	m28	-0,003	-0,145	-0,029
m16	m29	-0,031	0,026	0,016
m16	m30	0,018	0,010	-0,022
m17	m18	0,017	0,011	-0,036
m17	m19	0,005	-0,095	-0,112
m17	m20	-0,111	-0,013	-0,041
m17	m21	-0,057	-0,018	-0,011
m17	m22	0,024	-0,013	-0,083
m17	m23	0,005	0,016	-0,103
m17	m24	-0,042	0,013	-0,081
m17	m25	0,041	0,002	-0,020
m17	m26	-0,046	-0,029	0,015
m17	m27	0,019	-0,007	-0,037
m17	m28	-0,162	-0,083	-0,015
m17	m29	0,043	-0,056	0,007
m17	m30	0,010	-0,039	-0,061
m18	m19	0,012	-0,005	-0,044
m18	m20	0,056	0,000	-0,051
m18	m21	-0,019	-0,008	-0,131
m18	m22	-0,064	-0,068	-0,103
m18	m23	-0,022	0,087	-0,052
m18	m24	0,058	-0,026	0,048
m18	m25	-0,046	0,010	0,038
m18	m26	0,009	-0,057	-0,129
m18	m27	0,003	-0,034	-0,036
m18	m28	0,050	-0,041	0,004
m18	m29	-0,037	-0,112	-0,087
m18	m30	0,015	-0,109	-0,091
m19	m20	-0,077	0,068	-0,018
m19	m21	-0,061	-0,013	-0,115
m19	m22	-0,006	-0,040	-0,006
m19	m23	-0,013	-0,065	0,014
m19	m24	-0,073	0,012	-0,045
m19	m25	-0,093	-0,024	-0,017
m19	m26	-0,006	-0,005	-0,004
m19	m27	-0,041	-0,007	0,031
m19	m28	-0,035	-0,025	-0,064
m19	m29	0,023	0,044	-0,073

m19	m30	-0,054	-0,080	0,068
m20	m21	0,055	-0,026	0,014
m20	m22	0,035	0,008	-0,048
m20	m23	-0,037	-0,022	-0,075
m20	m24	-0,111	-0,057	-0,068
m20	m25	-0,077	-0,006	-0,077
m20	m26	-0,124	-0,082	0,013
m20	m27	-0,080	-0,009	0,020
m20	m28	-0,005	0,004	-0,006
m20	m29	-0,165	0,014	-0,047
m20	m30	0,093	-0,077	0,038
m21	m22	-0,017	-0,029	-0,031
m21	m23	-0,048	-0,118	0,022
m21	m24	-0,025	-0,050	-0,093
m21	m25	0,015	0,014	-0,012
m21	m26	-0,049	0,006	0,037
m21	m27	-0,007	0,029	-0,060
m21	m28	-0,041	-0,103	0,012
m21	m29	-0,034	-0,047	-0,007
m21	m30	0,011	-0,054	-0,031
m22	m23	-0,039	-0,036	0,033
m22	m24	-0,050	0,001	-0,028
m22	m25	0,006	-0,063	0,000
m22	m26	-0,024	-0,053	-0,124
m22	m27	-0,083	0,012	0,038
m22	m28	-0,078	-0,019	-0,066
m22	m29	0,039	-0,065	0,019
m22	m30	0,021	-0,100	-0,017
m23	m24	0,003	0,010	0,015
m23	m25	-0,097	-0,035	0,025
m23	m26	-0,040	-0,043	-0,112
m23	m27	-0,047	-0,031	-0,066
m23	m28	-0,030	-0,042	-0,104
m23	m29	0,009	-0,067	-0,008
m23	m30	-0,042	0,051	-0,013
m24	m25	0,034	-0,040	0,000
m24	m26	-0,003	-0,068	-0,044
m24	m27	-0,056	0,006	-0,028
m24	m28	-0,040	-0,028	-0,030
m24	m29	-0,042	-0,019	-0,132
m24	m30	-0,013	-0,092	-0,024
m25	m26	0,015	-0,033	-0,054
m25	m27	-0,033	-0,069	0,014
m25	m28	0,008	-0,012	-0,089
m25	m29	-0,081	-0,106	0,019
m25	m30	-0,006	-0,008	-0,032
m26	m27	-0,017	-0,022	-0,064
m26	m28	-0,038	-0,062	-0,006

m26	m29	-0,016	0,050	-0,032
m26	m30	-0,037	-0,017	-0,091
m27	m28	-0,059	-0,042	-0,091
m27	m29	-0,009	-0,043	0,031
m27	m30	0,006	-0,121	-0,066
m28	m29	-0,021	-0,095	-0,047
m28	m30	-0,054	-0,017	-0,018
m29	m30	-0,026	-0,035	-0,038

15 Madde ve 1000 Örneklem Büyüklüğüne Sahip Veri Setlerinin 1PLM, 2PLM ve 3PLM İçin Yerel Bağımsızlık Testi Sonuçları

Madde çiftleri		1PLM	2PLM	3PLM
Madde 1	Madde 2	Q_3	Q_3	Q_3
m1	m2	-0,039	-0,128	-0,080
m1	m3	-0,035	-0,063	-0,038
m1	m4	-0,046	-0,060	-0,131
m1	m5	-0,037	-0,023	-0,085
m1	m6	-0,055	-0,032	-0,080
m1	m7	-0,082	-0,050	-0,030
m1	m8	-0,062	0,021	-0,071
m1	m9	-0,076	0,039	-0,078
m1	m10	-0,098	-0,079	-0,062
m1	m11	-0,085	-0,120	-0,076
m1	m12	-0,119	-0,039	-0,085
m1	m13	-0,119	-0,141	-0,092
m1	m14	-0,093	-0,127	-0,096
m1	m15	-0,058	-0,047	-0,046
m2	m3	-0,046	-0,153	-0,043
m2	m4	-0,091	-0,146	-0,058
m2	m5	-0,039	-0,087	-0,059
m2	m6	0,007	-0,003	-0,013
m2	m7	-0,077	-0,031	-0,012
m2	m8	-0,108	-0,031	-0,089
m2	m9	-0,102	-0,007	-0,119
m2	m10	-0,073	-0,101	-0,052
m2	m11	-0,085	-0,074	-0,129
m2	m12	-0,115	-0,021	-0,082
m2	m13	-0,085	-0,122	-0,132
m2	m14	-0,099	-0,076	-0,028
m2	m15	-0,099	-0,062	-0,130
m3	m4	-0,084	-0,157	-0,037
m3	m5	-0,079	-0,110	-0,038
m3	m6	-0,089	-0,028	-0,052
m3	m7	-0,089	-0,036	0,056
m3	m8	0,012	-0,043	-0,067
m3	m9	-0,077	-0,045	-0,042
m3	m10	0,001	-0,075	-0,029
m3	m11	-0,103	-0,118	-0,058
m3	m12	-0,136	-0,028	-0,043
m3	m13	-0,040	-0,110	-0,055
m3	m14	-0,064	-0,094	-0,053
m3	m15	-0,008	-0,100	0,026
m4	m5	-0,157	-0,058	-0,039
m4	m6	-0,044	-0,021	-0,097
m4	m7	-0,120	-0,040	-0,013
m4	m8	-0,072	-0,035	-0,049

m4	m9	-0,057	-0,083	-0,067
m4	m10	-0,128	-0,059	-0,065
m4	m11	-0,060	-0,166	-0,130
m4	m12	-0,093	-0,027	-0,122
m4	m13	-0,009	-0,063	-0,059
m4	m14	-0,067	-0,158	-0,053
m4	m15	-0,059	-0,032	-0,127
m5	m6	-0,113	0,090	-0,098
m5	m7	-0,091	0,013	-0,054
m5	m8	0,017	-0,057	-0,141
m5	m9	-0,087	-0,011	-0,085
m5	m10	-0,063	-0,034	-0,083
m5	m11	-0,014	-0,114	-0,079
m5	m12	-0,080	-0,033	-0,063
m5	m13	-0,118	-0,106	-0,114
m5	m14	-0,123	-0,143	-0,035
m5	m15	-0,096	-0,023	-0,065
m6	m7	-0,050	-0,110	-0,039
m6	m8	-0,010	-0,083	-0,038
m6	m9	-0,087	0,010	-0,076
m6	m10	-0,028	-0,015	-0,070
m6	m11	-0,080	-0,097	-0,106
m6	m12	-0,036	0,021	-0,115
m6	m13	-0,087	0,000	-0,056
m6	m14	-0,055	-0,071	-0,115
m6	m15	-0,009	-0,056	-0,043
m7	m8	-0,107	0,065	-0,015
m7	m 9	-0,039	-0,023	-0,027
m7	m10	-0,052	-0,114	-0,026
m7	m11	-0,080	-0,115	-0,001
m7	m12	-0,069	-0,087	-0,067
m7	m13	-0,058	-0,059	-0,024
m7	m14	-0,086	-0,083	-0,070
m7	m15	-0,090	-0,051	-0,034
m8	m9	-0,038	-0,033	-0,115
m8	m10	-0,078	-0,019	-0,088
m8	m11	-0,021	-0,053	-0,037
m8	m12	-0,070	-0,141	-0,081
m8	m13	-0,061	-0,071	-0,077
m8	m14	-0,102	0,059	-0,105
m8	m15	-0,022	-0,099	-0,055
m9	m10	-0,040	-0,100	-0,096
m9	m11	-0,097	-0,026	-0,092
m9	m12	-0,012	-0,005	-0,094
m9	m13	-0,069	-0,058	-0,063
m9	m14	-0,049	-0,046	-0,063
m9	m15	-0,081	-0,045	-0,053
m10	m11	-0,074	-0,179	-0,032

m10	m12	-0,048	-0,025	-0,083
m10	m13	-0,135	-0,070	-0,033
m10	m14	-0,042	-0,125	-0,096
m10	m15	-0,123	-0,065	-0,068
m11	m12	-0,033	-0,051	-0,068
m11	m13	-0,118	-0,045	-0,092
m11	m14	-0,079	-0,072	-0,104
m11	m15	-0,057	-0,160	-0,059
m12	m13	-0,111	-0,085	-0,048
m12	m14	-0,069	-0,073	-0,102
m12	m15	-0,079	-0,027	-0,114
m13	m14	-0,044	-0,150	-0,105
m13	m15	-0,021	-0,112	-0,014
m14	m15	-0,130	-0,089	-0,114

30 Madde ve 1000 Örneklem Büyüklüğüne Sahip Veri Setlerinin 1PLM, 2PLM ve 3PLM İçin Yerel Bağımsızlık Testi Sonuçları

Madde çiftleri		1PLM	2PLM	3PLM
Madde 1	Madde 2	Q_3	Q_3	Q_3
m1	m2	-0,045	-0,107	-0,039
m1	m3	-0,015	-0,036	0,021
m1	m4	-0,036	-0,011	-0,038
m1	m5	-0,065	0,017	-0,004
m1	m6	0,031	0,003	0,021
m1	m7	0,013	-0,047	0,003
m1	m8	-0,059	0,003	0,026
m1	m9	-0,047	-0,032	-0,059
m1	m10	-0,055	-0,033	-0,030
m1	m11	0,017	-0,105	0,006
m1	m12	-0,070	-0,004	-0,046
m1	m13	-0,035	-0,050	0,011
m1	m14	-0,041	-0,056	-0,112
m1	m15	-0,030	0,049	0,066
m1	m16	-0,019	-0,036	0,038
m1	m17	-0,023	-0,068	0,026
m1	m18	-0,016	-0,037	-0,030
m1	m19	-0,025	-0,005	0,017
m1	m20	-0,024	0,000	-0,055
m1	m21	-0,018	-0,043	-0,020
m1	m22	-0,055	0,010	-0,018
m1	m23	-0,046	0,011	0,034
m1	m24	-0,084	-0,132	-0,002
m1	m25	-0,054	-0,012	-0,043
m1	m26	-0,068	-0,054	-0,059
m1	m27	-0,034	-0,017	-0,028
m1	m28	-0,040	-0,081	-0,035
m1	m29	0,023	-0,048	-0,044
m1	m30	-0,089	-0,048	-0,063
m2	m3	-0,023	-0,006	-0,029
m2	m4	-0,045	0,011	0,003
m2	m5	-0,065	-0,047	-0,090
m2	m6	-0,023	-0,078	-0,037
m2	m7	-0,037	-0,019	-0,017
m2	m8	-0,021	-0,029	-0,029
m2	m9	-0,060	-0,069	-0,090
m2	m10	0,019	-0,076	-0,032
m2	m11	-0,008	0,016	0,014
m2	m12	-0,054	-0,018	-0,044
m2	m13	-0,043	-0,008	-0,068
m2	m14	-0,054	-0,042	-0,057
m2	m15	-0,055	-0,069	0,003
m2	m16	-0,009	-0,008	0,023

m2	m17	-0,009	-0,063	-0,005
m2	m18	-0,005	-0,016	0,071
m2	m19	-0,019	-0,090	-0,008
m2	m20	-0,054	-0,053	-0,114
m2	m21	-0,090	-0,085	-0,045
m2	m22	-0,011	-0,086	-0,049
m2	m23	-0,049	-0,009	-0,042
m2	m24	-0,071	-0,037	-0,066
m2	m25	-0,042	0,023	0,002
m2	m26	-0,033	-0,006	-0,082
m2	m27	-0,037	-0,065	-0,044
m2	m28	-0,058	-0,039	-0,066
m2	m29	-0,053	0,003	-0,068
m2	m30	-0,022	-0,046	-0,036
m3	m4	-0,067	-0,066	0,032
m3	m5	-0,021	-0,049	0,011
m3	m6	-0,003	-0,106	-0,024
m3	m7	-0,030	-0,095	0,007
m3	m8	0,037	-0,065	-0,047
m3	m9	-0,025	-0,030	-0,074
m3	m10	-0,084	0,006	-0,071
m3	m11	-0,061	-0,023	-0,041
m3	m12	0,005	-0,065	-0,058
m3	m13	-0,070	-0,067	-0,063
m3	m14	-0,038	-0,056	-0,018
m3	m15	-0,039	-0,117	0,004
m3	m16	-0,069	-0,046	0,010
m3	m17	-0,079	-0,063	-0,051
m3	m18	0,000	-0,037	0,035
m3	m19	-0,074	-0,014	-0,014
m3	m20	-0,038	-0,062	-0,111
m3	m21	-0,015	-0,052	-0,076
m3	m22	-0,008	-0,033	-0,070
m3	m23	-0,024	-0,030	0,030
m3	m24	-0,005	-0,005	-0,045
m3	m25	-0,011	-0,026	0,073
m3	m26	0,000	0,003	-0,068
m3	m27	0,005	-0,041	-0,041
m3	m28	-0,024	-0,019	-0,055
m3	m29	-0,034	0,045	-0,067
m3	m30	-0,039	-0,017	-0,033
m4	m5	-0,068	-0,017	-0,068
m4	m6	-0,049	0,032	-0,002
m4	m7	-0,088	0,015	-0,021
m4	m8	-0,022	0,011	-0,087
m4	m9	-0,085	-0,023	-0,061
m4	m10	-0,026	-0,039	-0,040
m4	m11	-0,025	-0,070	-0,070

m4	m12	-0,085	-0,080	-0,078
m4	m13	-0,054	0,011	-0,044
m4	m14	-0,020	-0,044	-0,036
m4	m15	-0,098	-0,036	-0,007
m4	m16	0,009	-0,037	-0,020
m4	m17	-0,014	0,001	0,039
m4	m18	-0,025	-0,028	-0,078
m4	m19	-0,052	-0,040	-0,030
m4	m20	-0,063	0,003	-0,009
m4	m21	-0,030	-0,051	-0,040
m4	m22	-0,047	-0,040	-0,046
m4	m23	0,007	-0,028	0,019
m4	m24	-0,003	0,010	-0,076
m4	m25	-0,045	0,008	-0,022
m4	m26	-0,040	-0,080	-0,055
m4	m27	-0,004	-0,056	-0,033
m4	m28	-0,060	-0,015	-0,071
m4	m29	0,004	0,036	-0,037
m4	m30	-0,002	-0,052	-0,053
m5	m6	-0,007	-0,022	-0,047
m5	m7	-0,049	0,025	-0,061
m5	m8	-0,013	0,002	-0,016
m5	m9	-0,059	-0,045	-0,087
m5	m10	-0,099	0,008	-0,031
m5	m11	-0,048	-0,054	-0,060
m5	m12	-0,098	0,002	0,007
m5	m13	-0,021	-0,055	-0,053
m5	m14	0,012	-0,079	-0,063
m5	m15	0,030	-0,008	-0,045
m5	m16	-0,056	-0,076	0,006
m5	m17	-0,038	-0,031	-0,024
m5	m18	-0,030	0,002	-0,046
m5	m19	0,006	-0,021	-0,072
m5	m20	-0,007	-0,031	-0,044
m5	m21	-0,023	0,005	-0,031
m5	m22	-0,062	-0,027	-0,095
m5	m23	-0,082	-0,002	-0,053
m5	m24	-0,023	-0,041	-0,033
m5	m25	-0,096	0,014	-0,027
m5	m26	0,002	-0,039	-0,056
m5	m27	-0,032	-0,008	0,006
m5	m28	-0,028	-0,007	0,013
m5	m29	-0,046	-0,017	-0,078
m5	m30	-0,056	-0,088	-0,030
m6	m7	-0,023	-0,005	-0,003
m6	m8	-0,002	-0,047	-0,015
m6	m9	-0,028	0,039	-0,046
m6	m10	0,023	-0,020	-0,022

m6	m11	-0,057	-0,106	-0,070
m6	m12	-0,041	0,019	-0,041
m6	m13	-0,076	0,037	-0,086
m6	m14	-0,056	-0,116	-0,115
m6	m15	-0,001	-0,027	0,004
m6	m16	0,015	-0,046	0,009
m6	m17	-0,049	-0,021	-0,013
m6	m18	-0,055	-0,019	-0,071
m6	m19	-0,063	0,008	-0,003
m6	m20	-0,028	0,034	-0,025
m6	m21	-0,022	-0,005	-0,046
m6	m22	-0,037	0,055	0,037
m6	m23	0,022	-0,014	0,017
m6	m24	-0,008	-0,031	-0,065
m6	m25	-0,074	-0,053	-0,066
m6	m26	-0,013	-0,093	-0,013
m6	m27	-0,036	0,043	-0,052
m6	m28	-0,069	-0,131	-0,059
m6	m29	0,030	-0,056	0,016
m6	m30	-0,082	-0,132	-0,055
m7	m8	-0,025	0,009	-0,004
m7	m9	-0,041	0,007	-0,065
m7	m10	0,031	0,008	0,044
m7	m11	-0,044	-0,042	-0,114
m7	m12	-0,022	0,012	-0,074
m7	m13	-0,074	-0,023	-0,048
m7	m14	-0,118	-0,073	-0,031
m7	m15	-0,050	-0,026	0,004
m7	m16	-0,056	-0,036	0,056
m7	m17	-0,079	-0,012	-0,080
m7	m18	-0,014	-0,031	0,041
m7	m19	-0,024	-0,050	0,026
m7	m20	-0,020	-0,008	-0,077
m7	m21	-0,024	-0,068	0,028
m7	m22	0,003	-0,033	-0,043
m7	m23	-0,051	-0,059	-0,063
m7	m24	-0,003	-0,038	0,004
m7	m25	-0,071	-0,014	-0,012
m7	m26	-0,048	-0,066	-0,012
m7	m27	-0,085	-0,042	-0,033
m7	m28	-0,021	-0,117	-0,026
m7	m29	-0,028	-0,012	-0,056
m7	m30	-0,034	-0,059	-0,032
m8	m 9	0,015	-0,033	-0,060
m8	m10	0,015	0,014	-0,050
m8	m11	-0,016	0,042	-0,047
m8	m12	-0,030	-0,006	-0,060
m8	m13	-0,050	-0,049	-0,054

m8	m14	-0,078	-0,068	-0,121
m8	m15	-0,022	0,043	0,048
m8	m16	-0,070	-0,109	0,002
m8	m17	-0,089	-0,031	-0,004
m8	m18	-0,006	-0,065	-0,018
m8	m19	-0,014	-0,050	-0,073
m8	m20	-0,080	-0,010	-0,014
m8	m21	-0,007	-0,002	-0,067
m8	m22	-0,031	-0,056	-0,080
m8	m23	0,008	-0,023	-0,040
m8	m24	-0,054	-0,052	-0,067
m8	m25	-0,012	-0,026	-0,045
m8	m26	-0,029	-0,042	-0,045
m8	m27	0,003	-0,059	0,028
m8	m28	-0,026	-0,023	-0,043
m8	m29	-0,043	-0,072	-0,019
m8	m30	0,002	-0,045	0,000
m9	m10	-0,056	-0,006	-0,037
m9	m11	-0,042	-0,126	0,006
m9	m12	-0,013	0,016	-0,048
m9	m13	0,010	0,000	-0,046
m9	m14	0,008	-0,093	-0,019
m9	m15	-0,058	-0,020	-0,089
m9	m16	0,011	-0,025	-0,058
m9	m17	-0,030	0,003	-0,065
m9	m18	-0,074	-0,029	-0,049
m9	m19	-0,053	-0,060	-0,068
m9	m20	-0,087	-0,019	0,005
m9	m21	-0,005	0,008	-0,021
m9	m22	-0,030	-0,054	-0,038
m9	m23	-0,038	-0,045	-0,040
m9	m24	-0,077	-0,075	-0,033
m9	m25	0,029	-0,014	-0,014
m9	m26	-0,013	-0,027	-0,046
m9	m27	-0,031	0,014	-0,010
m9	m28	-0,027	-0,095	-0,088
m9	m29	0,018	0,002	0,009
m9	m30	-0,072	-0,095	-0,078
m10	m11	-0,053	-0,120	-0,105
m10	m12	0,055	-0,031	-0,043
m10	m13	-0,043	-0,073	-0,053
m10	m14	-0,054	-0,035	-0,038
m10	m15	-0,095	-0,068	-0,019
m10	m16	0,005	-0,038	-0,032
m10	m17	-0,055	-0,030	-0,050
m10	m18	0,024	-0,019	-0,058
m10	m19	-0,047	-0,092	-0,001
m10	m20	-0,019	-0,007	-0,030

m10	m21	0,012	-0,032	-0,022
m10	m22	-0,103	-0,024	-0,030
m10	m23	-0,111	0,017	-0,013
m10	m24	-0,034	0,040	-0,075
m10	m25	-0,024	0,003	-0,081
m10	m26	-0,017	-0,043	-0,022
m10	m27	0,025	-0,043	0,042
m10	m28	-0,097	-0,027	-0,077
m10	m29	-0,130	-0,080	0,022
m10	m30	-0,041	-0,039	-0,026
m11	m12	-0,047	-0,010	-0,009
m11	m13	-0,036	-0,100	-0,070
m11	m14	0,005	-0,019	-0,043
m11	m15	-0,080	-0,099	-0,039
m11	m16	-0,099	-0,065	0,036
m11	m17	-0,044	0,008	-0,041
m11	m18	0,022	-0,009	-0,049
m11	m19	-0,052	-0,059	-0,092
m11	m20	0,007	-0,086	-0,098
m11	m21	0,007	-0,082	-0,025
m11	m22	-0,042	-0,060	-0,109
m11	m23	-0,013	0,013	-0,039
m11	m24	-0,020	-0,032	-0,040
m11	m25	-0,066	-0,017	-0,070
m11	m26	0,031	-0,022	-0,015
m11	m27	-0,068	-0,055	-0,002
m11	m28	-0,013	0,021	0,013
m11	m29	-0,031	-0,036	-0,061
m11	m30	-0,106	-0,038	-0,056
m12	m13	-0,026	-0,031	-0,080
m12	m14	-0,058	0,002	-0,077
m12	m15	-0,024	0,002	-0,077
m12	m16	0,013	-0,014	-0,009
m12	m17	-0,021	0,004	0,015
m12	m18	-0,035	0,020	0,015
m12	m19	-0,038	-0,007	-0,031
m12	m20	-0,055	0,005	0,000
m12	m21	-0,022	0,016	-0,057
m12	m22	0,030	0,026	-0,063
m12	m23	-0,086	-0,022	-0,098
m12	m24	-0,054	-0,003	-0,010
m12	m25	-0,024	-0,056	-0,041
m12	m26	0,022	-0,001	-0,058
m12	m27	-0,102	-0,044	-0,031
m12	m28	-0,053	-0,023	0,005
m12	m29	-0,086	-0,018	-0,047
m12	m30	-0,067	-0,037	-0,097
m13	m14	-0,049	-0,028	-0,047

m13	m15	-0,041	-0,044	-0,027
m13	m16	-0,052	-0,056	0,008
m13	m17	-0,037	-0,042	-0,058
m13	m18	-0,062	-0,035	-0,033
m13	m19	-0,101	-0,051	0,071
m13	m20	-0,016	-0,064	-0,028
m13	m21	-0,053	-0,069	0,008
m13	m22	-0,059	-0,028	-0,052
m13	m23	-0,013	-0,030	-0,021
m13	m24	-0,060	-0,070	-0,039
m13	m25	-0,019	0,013	-0,018
m13	m26	-0,015	-0,090	-0,046
m13	m27	-0,009	0,011	0,011
m13	m28	-0,008	-0,070	-0,040
m13	m29	-0,058	-0,035	-0,065
m13	m30	-0,016	-0,018	-0,026
m14	m15	0,054	-0,025	-0,004
m14	m16	-0,025	0,012	-0,050
m14	m17	-0,016	-0,049	-0,094
m14	m18	-0,009	-0,121	-0,050
m14	m19	-0,058	-0,034	-0,055
m14	m20	0,013	0,019	0,033
m14	m21	-0,119	-0,059	-0,071
m14	m22	-0,044	-0,112	0,005
m14	m23	-0,001	-0,033	-0,024
m14	m24	-0,078	-0,116	-0,067
m14	m25	-0,042	0,002	-0,041
m14	m26	0,000	-0,024	0,029
m14	m27	-0,077	-0,034	0,013
m14	m28	-0,080	0,014	0,019
m14	m29	-0,018	-0,034	-0,018
m14	m30	-0,053	-0,017	-0,005
m15	m16	-0,003	-0,043	-0,016
m15	m17	-0,108	-0,029	-0,009
m15	m18	-0,067	-0,028	-0,011
m15	m19	0,003	-0,040	0,006
m15	m20	-0,046	0,032	0,022
m15	m21	-0,080	-0,075	0,033
m15	m22	-0,027	-0,025	0,009
m15	m23	-0,010	-0,063	-0,002
m15	m24	-0,097	-0,065	-0,018
m15	m25	-0,016	-0,048	-0,065
m15	m26	0,038	-0,056	-0,021
m15	m27	-0,059	-0,068	0,105
m15	m28	-0,010	-0,053	0,007
m15	m29	-0,001	-0,077	-0,016
m15	m30	-0,011	-0,035	-0,056
m16	m17	-0,053	-0,062	-0,006

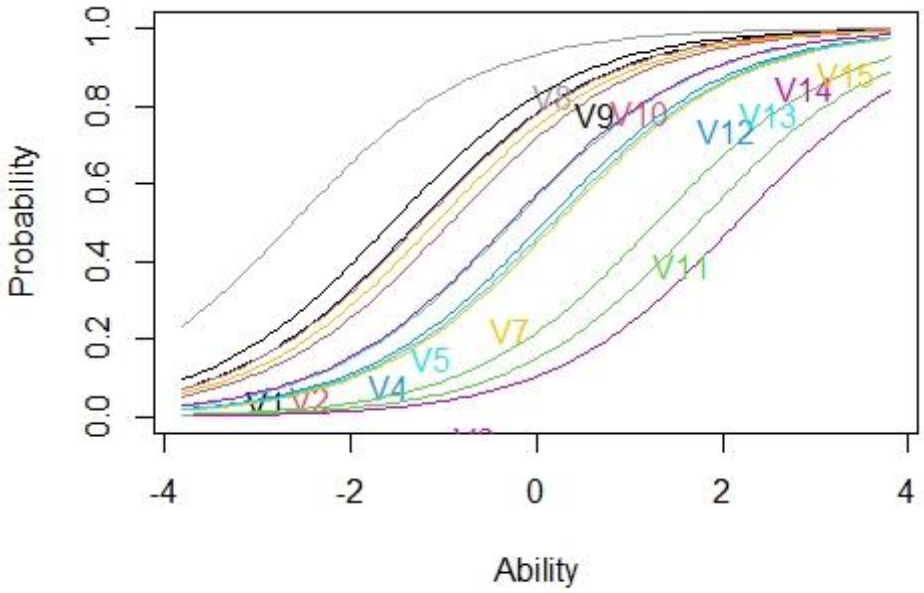
m16	m18	-0,001	0,003	-0,041
m16	m19	-0,066	-0,048	0,009
m16	m20	-0,062	-0,042	-0,019
m16	m21	0,004	-0,086	0,005
m16	m22	-0,061	-0,081	-0,052
m16	m23	-0,084	0,000	0,020
m16	m24	0,006	-0,025	-0,062
m16	m25	-0,035	-0,062	-0,020
m16	m26	0,011	0,001	-0,035
m16	m27	0,014	-0,041	0,007
m16	m28	-0,011	-0,089	-0,036
m16	m29	-0,064	-0,046	-0,064
m16	m30	-0,007	-0,020	-0,007
m17	m18	0,031	0,014	-0,049
m17	m19	-0,042	0,019	0,015
m17	m20	-0,038	-0,026	-0,017
m17	m21	-0,037	-0,026	0,007
m17	m22	-0,080	0,012	-0,078
m17	m23	-0,028	-0,008	-0,047
m17	m24	-0,006	0,007	0,088
m17	m25	-0,026	0,010	-0,046
m17	m26	-0,099	-0,046	-0,041
m17	m27	-0,048	-0,006	-0,012
m17	m28	-0,035	0,015	-0,004
m17	m29	0,002	0,014	-0,002
m17	m30	-0,016	-0,010	0,002
m18	m19	-0,015	-0,001	-0,036
m18	m20	-0,038	-0,042	-0,121
m18	m21	0,001	0,007	-0,088
m18	m22	-0,084	0,004	-0,090
m18	m23	-0,020	0,099	-0,037
m18	m24	0,015	0,031	-0,006
m18	m25	-0,030	-0,046	-0,014
m18	m26	-0,045	0,005	-0,111
m18	m27	0,020	-0,018	-0,040
m18	m28	-0,035	-0,050	-0,052
m18	m29	-0,056	-0,050	-0,016
m18	m30	0,005	-0,075	-0,079
m19	m20	-0,040	0,002	-0,047
m19	m21	-0,008	-0,016	0,030
m19	m22	0,024	0,060	-0,005
m19	m23	0,005	0,004	-0,027
m19	m24	-0,064	-0,082	-0,052
m19	m25	-0,018	-0,016	-0,072
m19	m26	-0,013	-0,110	-0,038
m19	m27	-0,132	0,014	0,011
m19	m28	-0,019	-0,048	-0,001
m19	m29	-0,007	-0,040	0,042

m19	m30	-0,075	-0,071	-0,066
m20	m21	-0,008	-0,061	0,024
m20	m22	-0,027	0,004	0,064
m20	m23	-0,026	0,001	-0,070
m20	m24	-0,033	-0,069	0,008
m20	m25	-0,091	0,015	-0,001
m20	m26	-0,043	-0,041	-0,026
m20	m27	0,003	-0,025	0,030
m20	m28	-0,067	-0,039	-0,029
m20	m29	-0,065	-0,019	-0,061
m20	m30	-0,045	-0,059	-0,010
m21	m22	-0,088	-0,003	0,022
m21	m23	-0,005	0,013	0,015
m21	m24	-0,011	-0,033	-0,022
m21	m25	-0,031	-0,019	-0,083
m21	m26	-0,028	-0,057	0,008
m21	m27	-0,021	0,002	-0,001
m21	m28	-0,049	-0,049	-0,012
m21	m29	-0,023	-0,015	-0,033
m21	m30	-0,014	-0,047	0,042
m22	m23	0,025	-0,043	-0,017
m22	m24	-0,016	-0,019	-0,088
m22	m25	-0,031	-0,039	-0,031
m22	m26	-0,100	-0,074	-0,025
m22	m27	-0,033	0,011	0,035
m22	m28	-0,099	-0,107	-0,118
m22	m29	0,013	-0,061	-0,015
m22	m30	-0,012	-0,066	0,028
m23	m24	-0,038	-0,037	-0,139
m23	m25	-0,072	0,001	-0,023
m23	m26	-0,090	-0,115	-0,065
m23	m27	-0,059	-0,018	0,024
m23	m28	-0,006	-0,007	-0,107
m23	m29	-0,008	-0,073	-0,014
m23	m30	0,012	-0,061	-0,040
m24	m25	-0,056	-0,001	-0,035
m24	m26	-0,024	-0,082	-0,041
m24	m27	-0,043	0,029	-0,075
m24	m28	-0,028	-0,042	-0,045
m24	m29	-0,014	0,028	-0,061
m24	m30	-0,006	-0,033	-0,047
m25	m26	0,014	-0,041	-0,049
m25	m27	-0,035	-0,043	-0,030
m25	m28	-0,031	0,034	0,022
m25	m29	0,007	-0,024	-0,058
m25	m30	0,012	-0,034	-0,090
m26	m27	-0,001	-0,054	-0,059
m26	m28	-0,002	-0,030	-0,081

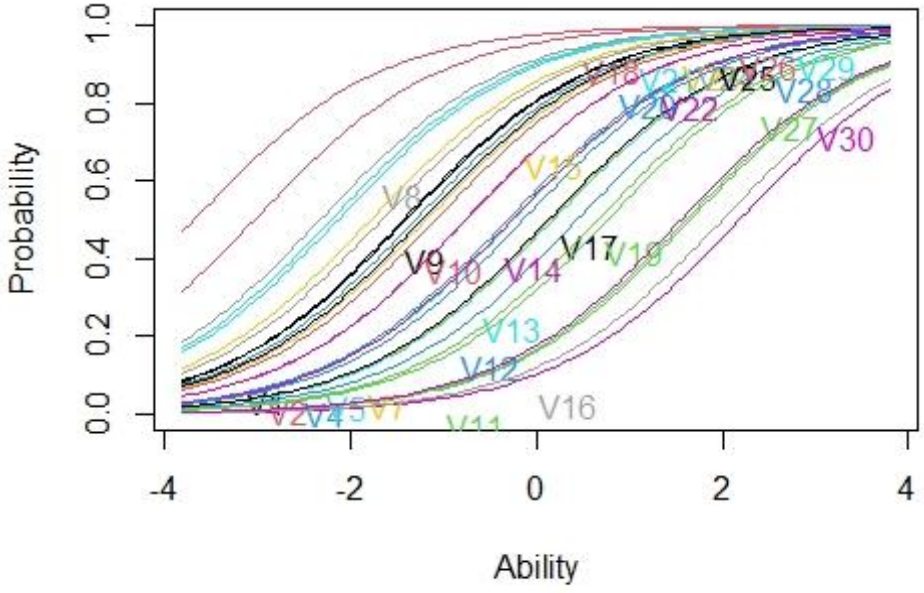
m26	m29	-0,043	-0,061	-0,029
m26	m30	0,015	-0,016	-0,091
m27	m28	-0,041	-0,032	-0,041
m27	m29	-0,049	-0,119	-0,033
m27	m30	-0,052	-0,075	-0,017
m28	m29	-0,066	-0,045	-0,122
m28	m30	-0,045	-0,042	-0,027
m29	m30	0,007	-0,041	-0,063

EK 3. Madde Karakteristik Eğrileri

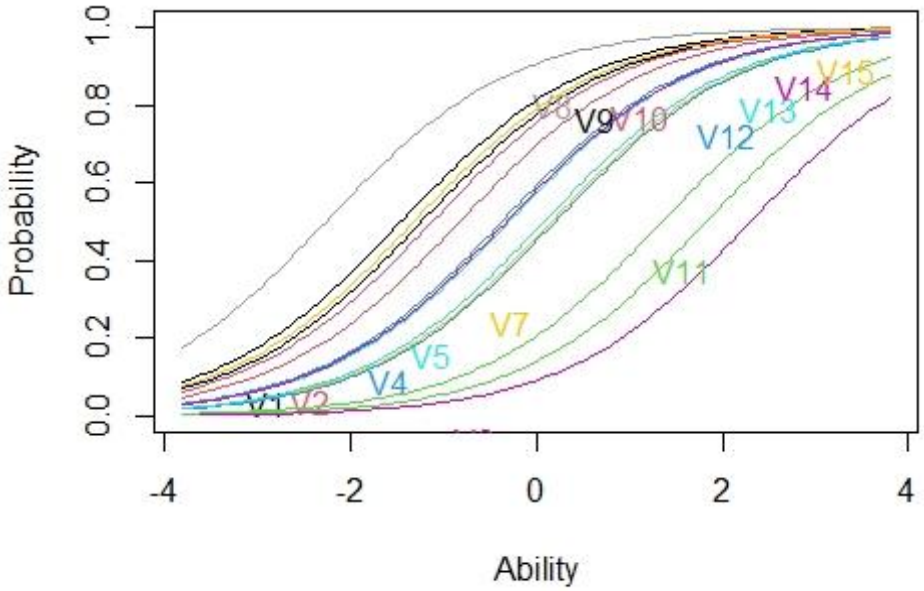
1. Simülasyon Koşulu



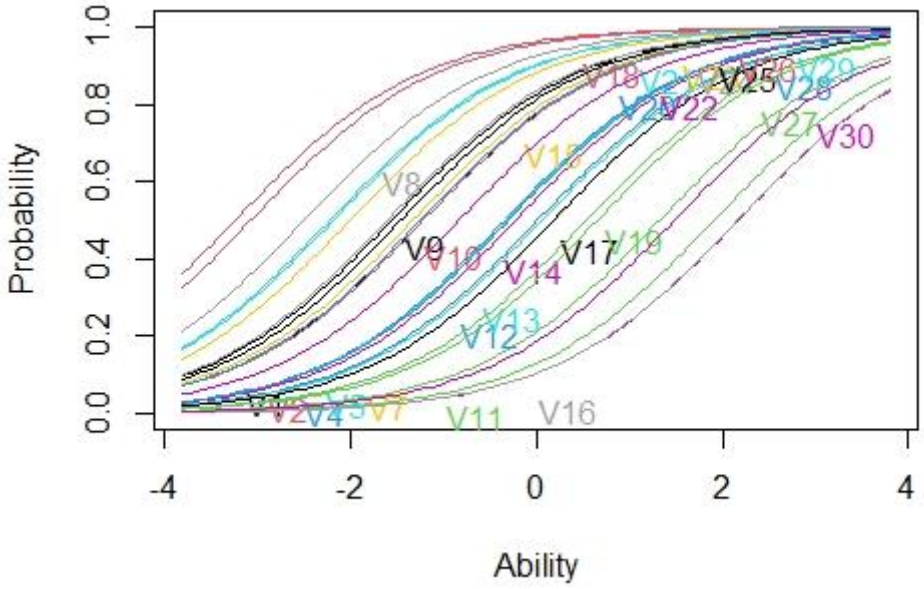
2. Simülasyon Koşulu



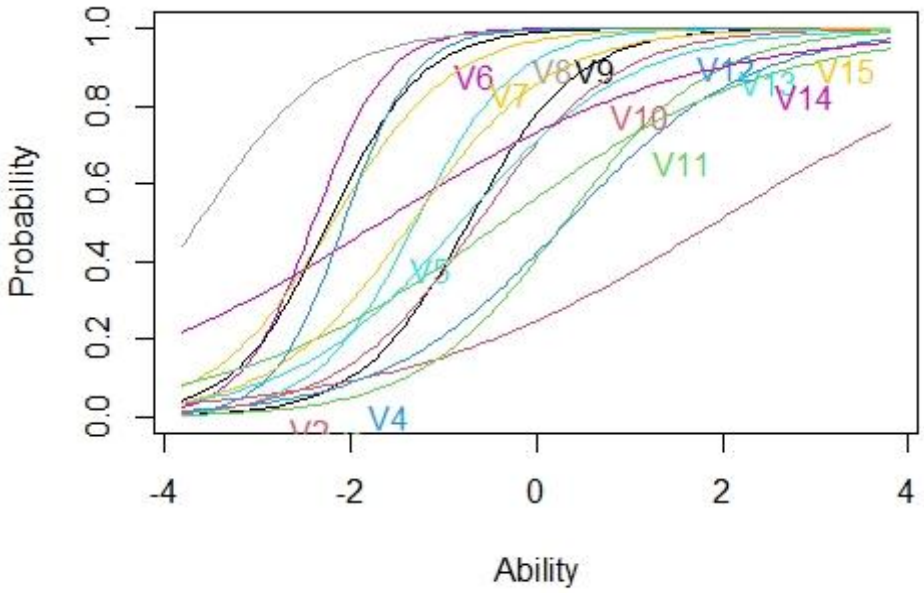
3. Simülasyon Koşulu



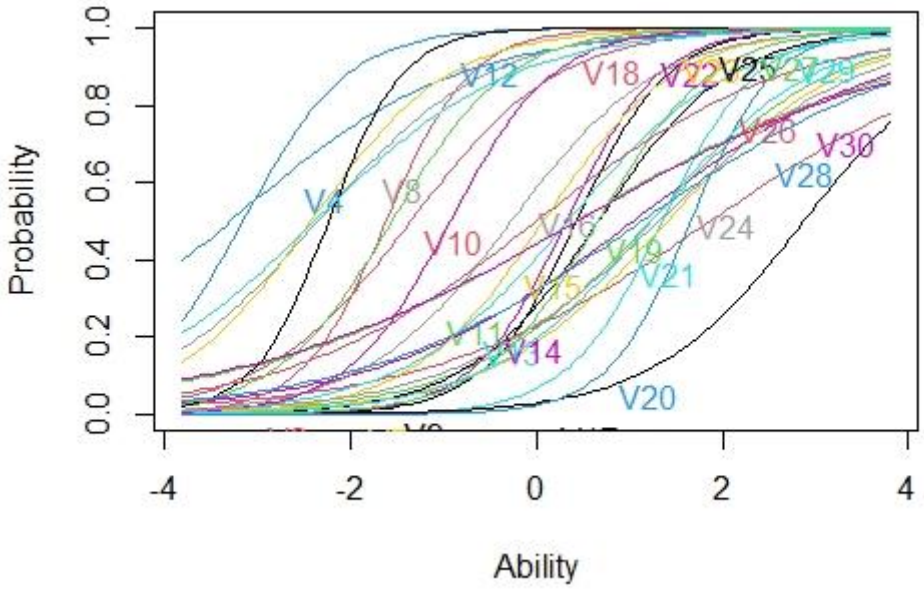
4. Simülasyon Koşulu



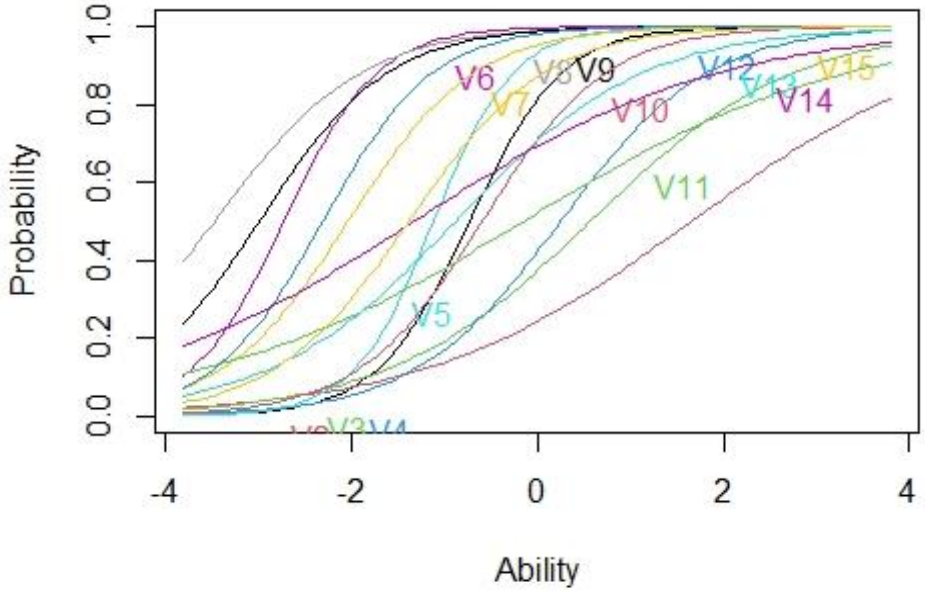
5. Simülasyon Koşulu



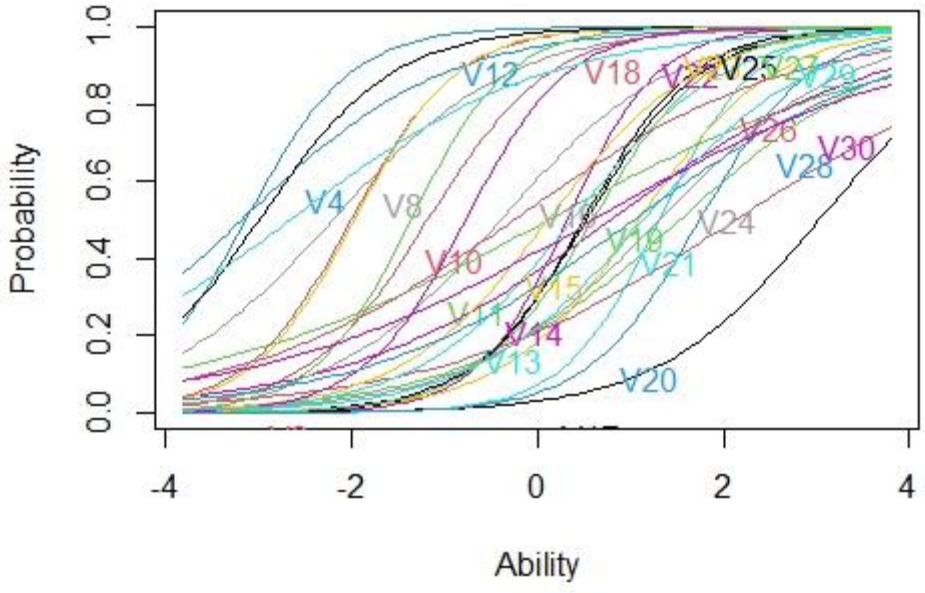
6. Simülasyon Koşulu



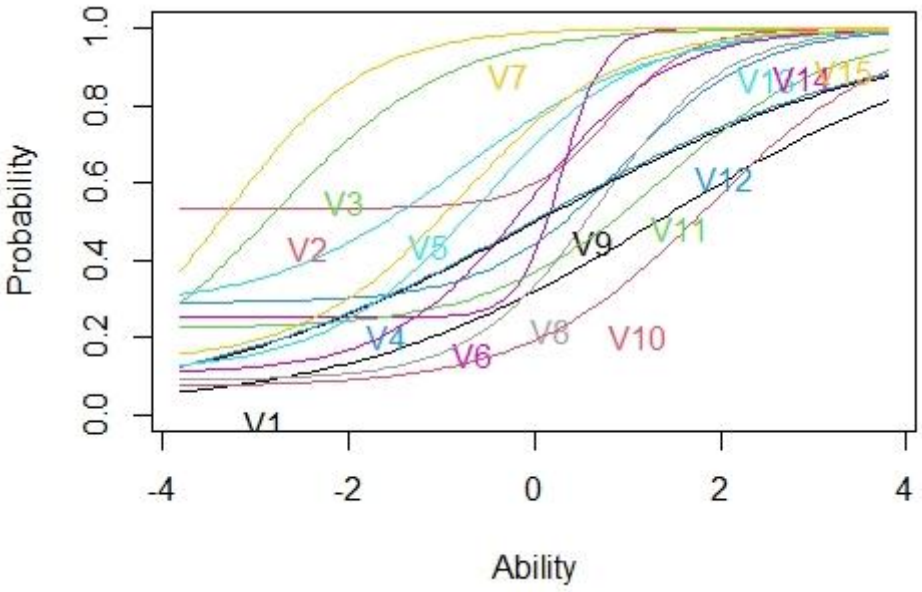
7. Simülasyon Koşulu



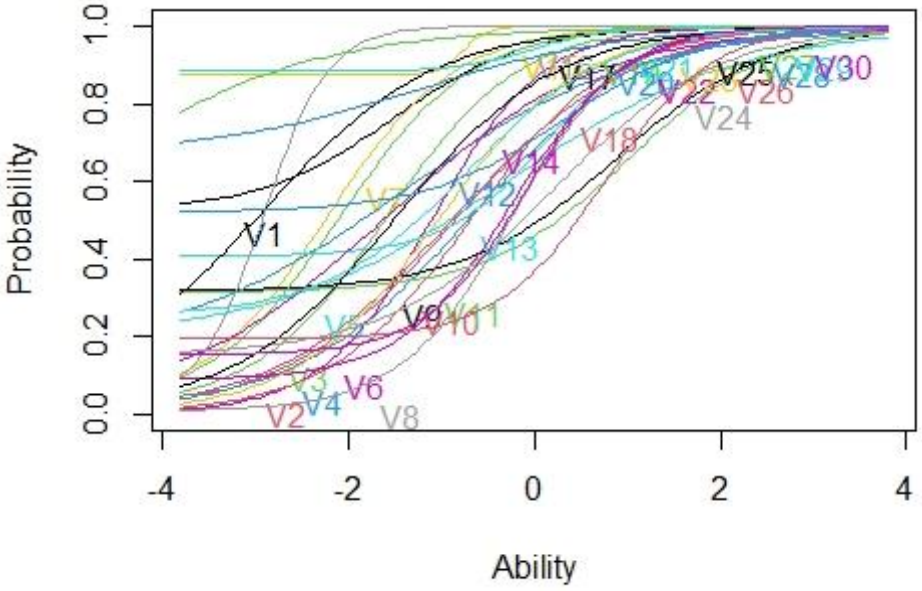
8. Simülasyon Koşulu



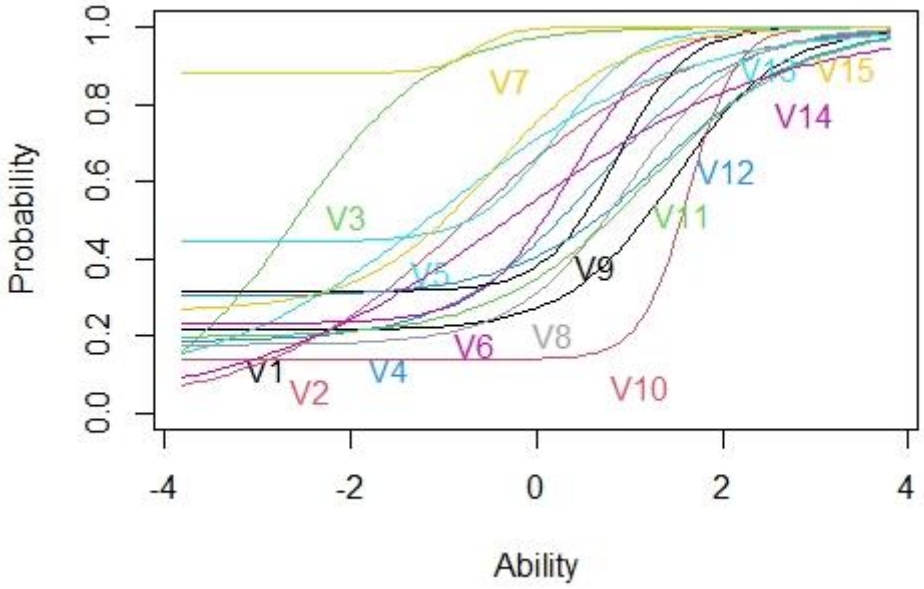
9. Simülasyon Koşulu



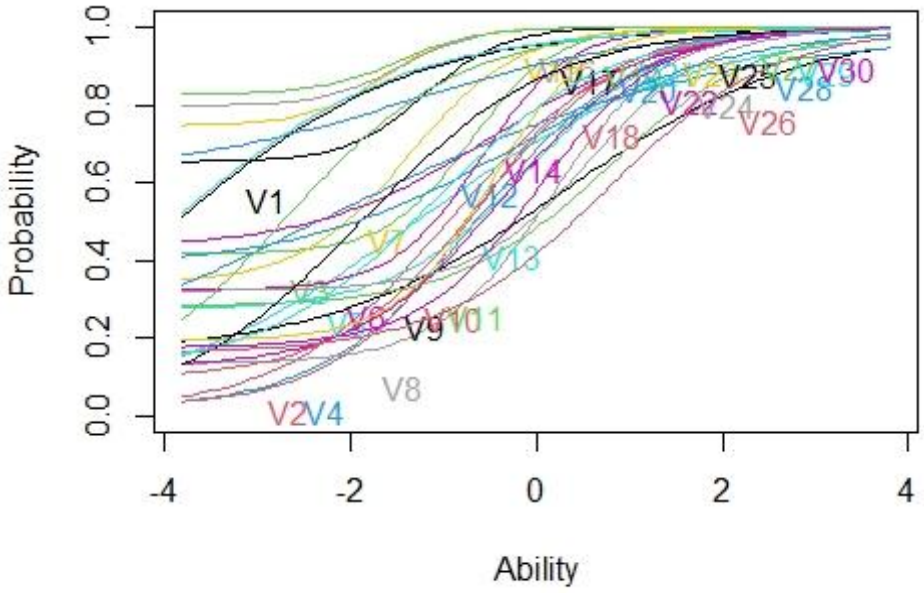
10. Simülasyon Koşulu



11. Simülasyon Koşulu



12. Simülasyon Koşulu



EK 4. Üretilen Veri Setlerinin 1PLM, 2PLM ve 3PLM Altında Test Edilmesine İlişkin Madde Uyum İndeksleri

1. Simülasyon Koşulu Altında Üretilen Veri Setinin 1PLM, 2PLM ve 3PLM Altında Test Edilmesine İlişkin Madde Uyum İndeksleri

	1PLM		2PLM		3PLM	
	χ^2	p	χ^2	p	χ^2	p
m1	6,867	0,651	21,3797	0,0062	14,5770	0,0418
m2	4,837	0,902	15,9828	0,0426	14,1703	0,0482
m3	7,014	0,636	21,8146	0,0053	14,9985	0,036
m4	6,686	0,670	14,8000	0,0632	16,4698	0,0212
m5	12,464	0,188	22,4555	0,0041	18,0148	0,0119
m6	6,657	0,574	13,2483	0,1036	15,7630	0,0274
m7	14,017	0,122	22,0178	0,0049	18,0934	0,0116
m8	12,519	0,130	14,1135	0,0789	10,7207	0,1513
m9	15,123	0,088	13,5061	0,0956	13,1090	0,0695
m10	8,906	0,541	10,5813	0,2266	11,4731	0,1193
m11	13,265	0,209	17,3422	0,0267	15,4306	0,0309
m12	10,923	0,281	19,9278	0,0106	19,7270	0,0062
m13	7,780	0,556	16,1021	0,0409	15,5349	0,0297
m14	13,987	0,123	21,4900	0,006	24,0283	0,0011
m15	5,062	0,829	15,3758	0,0522	21,8653	0,0027

2. Simülasyon Koşulu Altında Üretilen Veri Setinin 1PLM, 2PLM ve 3PLM Altında Test Edilmesine İlişkin Madde Uyum İndeksleri

	1PLM		2PLM		3PLM	
	χ^2	*p*	χ^2	*p*	χ^2	*p*
m1	8,3655	0,4978	9,2640	0,3205	14,5850	0,0417
m2	12,1520	0,2049	11,9583	0,1531	11,7901	0,1077
m3	6,1308	0,7268	7,3587	0,4985	8,0096	0,3317
m4	19,1568	0,0239	9,4410	0,3065	9,6272	0,2107
m5	15,3438	0,0819	12,1661	0,1439	16,7731	0,0189
m6	9,3052	0,4096	11,5856	0,1707	10,3868	0,1677
m7	12,5674	0,1832	14,5342	0,0689	16,4607	0,0212
m8	6,9755	0,6397	8,0644	0,4272	9,5310	0,2167
m9	8,0511	0,529	10,0040	0,2647	6,7538	0,455
m10	1,8402	0,9937	5,9921	0,6481	7,1456	0,4139
m11	6,9581	0,6415	8,6414	0,3734	6,1753	0,5194
m12	10,5173	0,3103	14,8422	0,0623	13,0710	0,0704
m13	13,8800	0,1267	11,3972	0,1802	11,1756	0,1311
m14	14,2126	0,115	13,0122	0,1114	7,8164	0,3491
m15	20,1606	0,0169	15,2100	0,0552	10,1461	0,1804
m16	8,0447	0,5297	10,7617	0,2156	11,8823	0,1045
m17	12,8602	0,169	6,8594	0,5519	6,1282	0,5249
m18	7,1310	0,6235	5,0039	0,7572	5,2720	0,6268
m19	7,6921	0,5655	11,8116	0,1598	10,2551	0,1746
m20	12,4195	0,1907	9,3891	0,3105	7,7757	0,3528
m21	4,9509	0,8386	4,6418	0,7951	4,7358	0,6922
m22	8,0629	0,5278	9,0383	0,3391	12,4210	0,0875
m23	16,9274	0,0499	9,1896	0,3266	10,2091	0,177
m24	16,2923	0,061	10,9316	0,2056	11,1051	0,1341
m25	15,7242	0,0729	18,1646	0,02	22,6285	0,002
m26	12,5204	0,1855	13,9520	0,083	10,2901	0,1727
m27	11,6084	0,2363	11,6148	0,1692	9,3381	0,2293
m28	12,2069	0,2019	10,5949	0,2257	9,9639	0,1906
m29	8,8978	0,4468	4,7336	0,7856	8,0227	0,3306
m30	23,9646	0,0044	16,2030	0,0396	16,9255	0,0179

3. Simülasyon Koşulu Altında Üretilen Veri Setinin 1PLM, 2PLM ve 3PLM Altında Test Edilmesine İlişkin Madde Uyum İndeksleri

	1PLM		2PLM		3PLM	
	χ^2	*p*	χ^2	*p*	χ^2	*p*
m1	11,657	0,390	24,5588	0,0018	18,3867	0,0103
m2	6,620	0,761	20,3754	0,009	22,9240	0,0018
m3	11,107	0,268	19,4037	0,0128	18,9204	0,0084
m4	9,289	0,505	20,1015	0,01	26,8264	0,0004
m5	5,637	0,845	22,7472	0,0037	25,2315	0,0007
m6	10,757	0,293	18,5875	0,0172	22,9948	0,0017
m7	5,139	0,882	25,9114	0,0011	30,1935	0,0001
m8	11,687	0,307	27,2053	0,0007	25,6487	0,0006
m9	6,661	0,672	65,1326	<0,0001	66,4763	<0,0001
m10	5,059	0,887	20,8110	0,0077	28,4932	0,0002
m11	4,852	0,901	14,8814	0,0615	26,1716	0,0005
m12	10,099	0,432	47,3332	<0,0001	45,5180	<0,0001
m13	4,137	0,941	32,5150	0,0001	38,5800	<0,0001
m14	15,316	0,121	24,8334	0,0017	37,8250	<0,0001
m15	6,923	0,805	23,0676	0,0033	24,8947	0,0008

4. Simülasyon Koşulu Altında Üretilen Veri Setinin 1PLM, 2PLM ve 3PLM Altında Test Edilmesine İlişkin Madde Uyum İndeksleri

	1PLM		2PLM		3PLM	
	χ^2	*p*	χ^2	*p*	χ^2	*p*
m1	22,6840	0,0069	18,3608	0,0187	18,0929	0,0116
m2	20,2974	0,0162	14,0473	0,0805	13,6723	0,0573
m3	13,8448	0,128	15,2624	0,0542	12,0127	0,1001
m4	11,4059	0,2489	10,0890	0,2588	11,4963	0,1184
m5	9,7067	0,3747	16,7064	0,0333	11,9095	0,1036
m6	18,0675	0,0344	18,7166	0,0165	17,8844	0,0125
m7	18,9217	0,0259	16,8986	0,0312	16,7665	0,019
m8	9,5935	0,3844	7,6939	0,4639	7,3488	0,3935
m9	10,1428	0,3391	10,0957	0,2584	8,5030	0,2903
m10	7,3476	0,601	17,1135	0,0289	19,6293	0,0064
m11	19,2021	0,0235	16,3582	0,0375	16,2210	0,0232
m12	29,8682	0,0005	23,2524	0,0031	22,8063	0,0018
m13	14,0889	0,1192	13,6769	0,0906	12,4738	0,086
m14	23,6312	0,0049	10,9509	0,2045	10,0819	0,184
m15	17,1254	0,0468	13,4915	0,096	13,7154	0,0565
m16	17,1808	0,046	15,0600	0,058	16,4721	0,0211
m17	12,7206	0,1757	15,1203	0,0568	11,6391	0,1131
m18	7,1240	0,6242	13,1613	0,1064	2,6303	0,917
m19	32,0893	0,0002	23,0997	0,0032	22,3078	0,0022
m20	4,6949	0,8601	12,1223	0,1458	17,4280	0,0148
m21	15,9551	0,0678	9,8343	0,2768	12,4778	0,0859
m22	24,4399	0,0037	18,7700	0,0161	18,7941	0,0089
m23	8,3715	0,4972	10,9856	0,2025	9,1295	0,2435
m24	6,4733	0,6918	7,8606	0,4472	6,0352	0,5356
m25	9,2227	0,417	7,5568	0,4779	8,5121	0,2896
m26	9,9168	0,3573	9,3956	0,31	9,6987	0,2063
m27	27,9230	0,001	31,1886	0,0001	26,0140	0,0005
m28	11,2295	0,2603	12,3326	0,137	11,4318	0,1209
m29	18,3659	0,0312	17,4074	0,0261	15,7131	0,0279
m30	24,3177	0,0038	19,6094	0,119	18,5645	0,0097

5. Simülasyon Koşulu Altında Üretilen Veri Setinin 1PLM, 2PLM ve 3PLM Altında Test Edilmesine İlişkin Madde Uyum İndeksleri

	1PLM		2PLM		3PLM	
	χ^2	*p*	χ^2	*p*	χ^2	*p*
m1	44,1517	<0,0001	28,6711	0,0004	31,4972	0,0001
m2	8,3145	0,216	8,1220	0,4216	10,6218	0,156
m3	47,8708	<0,0001	26,3335	0,0009	28,8785	0,0002
m4	32,9217	<0,0001	18,3418	0,0188	20,7773	0,0041
m5	48,0974	<0,0001	17,1168	0,0289	18,5914	0,0096
m6	19,6113	0,0032	8,0309	0,4305	13,9078	0,0528
m7	16,0953	0,0133	19,2636	0,0135	11,1334	0,1329
m8	6,4910	0,3705	7,3996	0,4942	4,8248	0,6813
m9	15,3592	0,0176	7,3536	0,499	7,4017	0,3883
m10	37,6180	<0,0001	17,3050	0,0271	22,7072	0,0019
m11	16,0905	0,0133	8,2732	0,4072	8,3301	0,3044
m12	25,0181	0,0003	6,6723	0,5724	17,1482	0,0165
m13	24,9466	0,0003	20,3384	0,0091	19,1372	0,0078
m14	4,4180	0,6203	13,8398	0,086	11,7122	0,1104
m15	35,9367	<0,0001	21,3250	0,0063	14,6668	0,0405

6. Simülasyon Koşulu Altında Üretilen Veri Setinin 1PLM, 2PLM ve 3PLM Altında Test Edilmesine İlişkin Madde Uyum İndeksleri

	1PLM		2PLM		3PLM	
	χ^2	*p*	χ^2	*p*	χ^2	*p*
m1	43,0616	<0,0001	71.128	0,5245	7,1128	0,4646
m2	20,4559	0,0153	151.382	0,0565	15,1382	0,027
m3	9,1195	0,4263	68.213	0,556	6,8213	0,4095
m4	8,3725	0,4971	116.615	0,167	11,6615	0,3451
m5	33,6993	0,0001	113.146	0,1845	11,3146	0,1582
m6	47,4233	<0,0001	107.873	0,214	10,7873	0,0736
m7	9,5128	0,3913	68.430	0,5537	6,8430	0,3431
m8	12,8450	0,1697	118.201	0,1594	11,8201	0,0497
m9	25,0556	0,0029	121.261	0,1457	12,1261	0,0175
m10	27,6646	0,0011	90.562	0,3376	9,0562	0,3085
m11	7,6525	0,5695	101.407	0,2553	10,1407	0,1403
m12	9,3459	0,406	107.687	0,2152	10,7687	0,1702
m13	14,4127	0,1084	234.945	0,0028	23,4945	0,0016
m14	8,9545	0,4415	122.008	0,1425	12,2008	0,1807
m15	26,0011	0,002	86.376	0,3738	8,6376	0,2619
m16	20,1576	0,017	81.060	0,4232	8,1060	0,2733
m17	8,5087	0,4838	74.149	0,4926	7,4149	0,5371
m18	29,7868	0,0005	75.830	0,4752	7,5830	0,3165
m19	30,3834	0,0004	70.340	0,533	7,0340	0,3015
m20	41,4516	<0,0001	104.988	0,2317	10,4988	0,1586
m21	8,7822	0,4576	110.488	0,199	11,0488	0,0602
m22	36,2118	<0,0001	101.963	0,2515	10,1963	0,1962
m23	20,2485	0,0164	130.692	0,1095	13,0692	0,0798
m24	8,3903	0,4953	54.053	0,7135	5,4053	0,4164
m25	26,0079	0,002	145.413	0,0687	14,5413	0,0317
m26	8,1192	0,5222	65.781	0,5828	6,5781	0,4771
m27	20,0262	0,0178	110.967	0,1963	11,0967	0,0618
m28	1,7433	0,9949	64.466	0,5973	6,4466	0,3629
m29	6,1849	0,7213	77.954	0,4537	7,7954	0,4326
m30	16,4699	0,0577	94.392	0,3066	9,4392	0,2034

7. Simülasyon Koşulu Altında Üretilen Veri Setinin 1PLM, 2PLM ve 3PLM Altında Test Edilmesine İlişkin Madde Uyum İndeksleri

	1PLM		2PLM		3PLM	
	χ^2	*p*	χ^2	*p*	χ^2	*p*
m1	89,3517	<0,0001	5,046	0,753	38,2433	<0,0001
m2	24,1545	0,0011	12,893	0,116	23,5699	0,0014
m3	29,3379	0,0001	11,053	0,199	31,7448	<0,0001
m4	53,1150	<0,0001	13,258	0,103	58,0623	<0,0001
m5	84,7931	<0,0001	5,180	0,638	32,7864	<0,0001
m6	19,2611	0,0074	15,491	0,017	8,4685	0,2931
m7	24,1766	0,0011	11,480	0,244	19,9920	0,0056
m8	10,2475	0,175	8,771	0,362	9,3941	0,2256
m9	11,4500	0,1202	11,393	0,180	16,8586	0,0183
m10	61,9438	<0,0001	7,241	0,511	41,8778	<0,0001
m11	7,0177	0,427	7,386	0,689	17,4844	0,0145
m12	30,3753	0,0001	2,963	0,937	12,8102	0,0769
m13	22,2518	0,0023	7,956	0,539	28,9826	0,0001
m14	6,7550	0,4548	8,624	0,568	20,5708	0,0045
m15	37,0273	<0,0001	16,354	0,090	20,9411	0,0039

8. Simülasyon Koşulu Altında Üretilen Veri Setinin 1PLM, 2PLM ve 3PLM Altında Test Edilmesine İlişkin Madde Uyum İndeksleri

	1PLM		2PLM		3PLM	
	χ^2	*p*	χ^2	*p*	χ^2	*p*
m1	57,4758	<0,0001	12,6974	0,1227	13,0225	0,0716
m2	14,4422	0,1074	15,3890	0,052	11,0888	0,1348
m3	6,9876	0,6384	10,4612	0,2341	10,4910	0,1624
m4	7,7510	0,5594	6,3428	0,6089	4,2403	0,7517
m5	62,9460	<0,0001	17,1357	0,0287	18,1834	0,0112
m6	83,2676	<0,0001	20,4907	0,0086	20,7482	0,0042
m7	30,1953	0,0004	11,8349	0,1587	13,4791	0,0613
m8	16,0803	0,0652	17,3946	0,0263	15,8823	0,0262
m9	63,9727	<0,0001	15,0496	0,0582	12,2428	0,0929
m10	44,6692	<0,0001	18,2529	0,0194	14,4680	0,0435
m11	20,1559	0,017	7,4424	0,4897	9,3013	0,2317
m12	9,4883	0,3935	9,3117	0,3167	4,8603	0,677
m13	27,0142	0,0014	17,3299	0,0269	16,2212	0,0232
m14	8,9899	0,4382	6,8702	0,5507	7,3357	0,3948
m15	26,2716	0,0018	10,7880	0,214	10,3851	0,1678
m16	16,5454	0,0563	12,6645	0,1239	12,1653	0,0953
m17	7,4355	0,5919	7,9856	0,4349	7,5957	0,3696
m18	35,0061	0,0001	15,9934	0,0425	14,9718	0,0364
m19	50,0040	<0,0001	15,9251	0,0435	14,2102	0,0476
m20	44,0956	<0,0001	14,5759	0,0679	14,3784	0,0448
m21	15,7493	0,0723	8,8516	0,355	11,1176	0,1336
m22	70,8940	<0,0001	20,4016	0,0089	19,2310	0,0075
m23	39,9465	<0,0001	15,7790	0,0457	18,2049	0,0111
m24	9,3463	0,4059	16,2361	0,0391	15,4600	0,0305
m25	21,0749	0,0123	13,3093	0,1016	19,3025	0,0073
m26	6,7836	0,6596	5,6063	0,6912	5,2699	0,6271
m27	58,8694	<0,0001	24,6051	0,0018	31,4060	0,0001
m28	11,8057	0,2245	11,5206	0,1739	12,5086	0,085
m29	9,5477	0,3883	11,2053	0,1903	9,6420	0,2098
m30	13,9647	0,1236	17,8978	0,022	16,4378	0,0214

9. Simülasyon Koşulu Altında Üretilen Veri Setinin 1PLM, 2PLM ve 3PLM Altında Test Edilmesine İlişkin Madde Uyum İndeksleri

	1PLM		2PLM		3PLM	
	χ^2	*p*	χ^2	*p*	χ^2	*p*
m1	7,6613	0,4672	16,2095	0,0395	6,671	0,671
m2	8,5685	0,38	14,0491	0,0805	8,641	0,471
m3	9,1077	0,3333	8,9804	0,3439	6,046	0,642
m4	19,0869	0,0144	18,4622	0,018	12,607	0,181
m5	10,5051	0,2313	24,2171	0,0021	5,786	0,761
m6	53,1225	<0,0001	36,7479	<0,0001	12,275	0,139
m7	8,3343	0,4015	15,9691	0,0428	18,061	0,006
m8	27,8484	0,0005	30,5380	0,0002	10,827	0,212
m9	9,9854	0,266	13,2263	0,1043	11,770	0,227
m10	8,7076	0,3676	17,0009	0,0301	4,677	0,861
m11	8,7926	0,3601	20,1420	0,0098	8,165	0,613
m12	9,8950	0,2725	14,3239	0,0737	9,888	0,360
m13	19,4279	0,0127	21,1573	0,0067	5,715	0,768
m14	23,8606	0,0024	25,1725	0,0015	3,665	0,886
m15	21,1185	0,0068	20,0202	0,0103	7,217	0,513

10. Simülasyon Koşulu Altında Üretilen Veri Setinin 1PLM, 2PLM ve 3PLM Altında Test Edilmesine İlişkin Madde Uyum İndeksleri

	1PLM		2PLM		3PLM	
	χ^2	*p*	χ^2	*p*	χ^2	*p*
m1	9,8437	0,3633	7,6432	0,4691	5,1775	0,6383
m2	11,7271	0,2291	10,5242	0,2301	14,9372	0,0368
m3	20,9698	0,0128	14,8440	0,0623	13,0176	0,0717
m4	19,6257	0,0204	13,8057	0,087	16,9072	0,018
m5	13,5901	0,1377	8,6340	0,3741	9,5846	0,2134
m6	23,1402	0,0059	12,3836	0,1349	7,1966	0,4087
m7	12,8422	0,1699	10,7311	0,2174	10,2566	0,1745
m8	41,2496	<0,0001	15,2785	0,054	16,1866	0,0235
m9	9,8647	0,3615	12,2858	0,1389	8,8758	0,2617
m10	30,7672	0,0003	12,3456	0,1364	6,7710	0,4531
m11	22,6666	0,007	8,7318	0,3654	8,3951	0,299
m12	8,8753	0,4489	12,6496	0,1245	12,1803	0,0948
m13	22,5245	0,0074	12,3504	0,1362	8,8998	0,2599
m14	10,1109	0,3416	8,1065	0,4231	9,0207	0,2512
m15	9,4300	0,3986	9,5393	0,2989	1,7798	0,971
m16	4,7401	0,8564	4,8361	0,7749	10,3329	0,1705
m17	7,8731	0,547	9,7191	0,2853	4,6251	0,7056
m18	13,6059	0,1371	7,4504	0,4889	5,8716	0,5548
m19	11,7058	0,2304	6,2707	0,6169	6,3735	0,4969
m20	4,5769	0,8695	5,8046	0,6691	4,1461	0,7628
m21	1,4706	0,9974	12,2258	0,1414	9,5705	0,2142
m22	31,5505	0,0002	13,6301	0,0919	13,9287	0,0525
m23	16,7453	0,0529	10,2174	0,2501	15,2864	0,0325
m24	18,3235	0,0316	14,9595	0,0599	11,7736	0,1083
m25	11,2800	0,257	9,1146	0,3327	9,8421	0,1977
m26	25,9410	0,0021	16,8256	0,032	10,5262	0,1607
m27	16,7966	0,052	8,3703	0,3982	8,6388	0,2796
m28	10,9349	0,2802	7,4930	0,4845	5,5093	0,5981
m29	8,9419	0,4427	7,8930	0,444	7,5533	0,3736
m30	35,7520	<0,0001	13,3210	0,1013	15,6739	0,0283

11. Simülasyon Koşulu Altında Üretilen Veri Setinin 1PLM, 2PLM ve 3PLM Altında Test Edilmesine İlişkin Madde Uyum İndeksleri

	1PLM		2PLM		3PLM	
	χ^2	*p*	χ^2	*p*	χ^2	*p*
m1	9,6562	0,2089	25,6531	0,0012	3,796	0,956
m2	32,0363	<0,0001	31,2094	0,0001	6,489	0,690
m3	40,1573	<0,0001	44,1794	<0,0001	27,284	0,001
m4	32,5287	<0,0001	37,7444	<0,0001	12,854	0,232
m5	23,6017	0,0013	31,6298	0,0001	7,334	0,694
m6	61,1821	<0,0001	52,8540	<0,0001	13,581	0,138
m7	19,0870	0,0079	23,2488	0,0031	3,243	0,862
m8	29,6811	0,0001	41,5062	<0,0001	9,988	0,351
m9	30,8591	0,0001	36,6742	<0,0001	14,751	0,141
m10	14,1520	0,0485	22,5040	0,0041	16,066	0,066
m11	17,6281	0,0138	29,4117	0,0003	18,232	0,051
m12	18,4400	0,0101	25,3996	0,0013	22,424	0,013
m13	46,3941	<0,0001	37,1413	<0,0001	9,361	0,405
m14	14,6374	0,0409	19,8760	0,0108	24,530	0,006
m15	50,8515	<0,0001	41,9654	<0,0001	7,973	0,537

12. Simülasyon Koşulu Altında Üretilen Veri Setinin 1PLM, 2PLM ve 3PLM Altında Test Edilmesine İlişkin Madde Uyum İndeksleri

	1PLM		2PLM		3PLM	
	χ^2	*p*	χ^2	*p*	χ^2	*p*
m1	27,4099	0,0012	14,6936	0,0654	8,3291	0,3045
m2	28,5609	0,0008	15,6408	0,0478	20,6301	0,0044
m3	43,1115	<0,0001	14,8212	0,0627	9,9065	0,1939
m4	34,4637	0,0001	24,9623	0,0016	20,9240	0,0039
m5	19,7314	0,0196	18,7551	0,0162	13,7619	0,0556
m6	47,3816	<0,0001	20,2449	0,0094	14,8437	0,0381
m7	33,4644	0,0001	12,1068	0,1465	14,0701	0,0499
m8	46,2002	<0,0001	21,2095	0,0066	18,1894	0,0111
m9	4,7336	0,8569	17,1701	0,0284	12,3317	0,0902
m10	45,4181	<0,0001	19,3998	0,0129	24,9089	0,0008
m11	11,5125	0,2422	11,8178	0,1595	10,1029	0,1828
m12	6,9285	0,6446	13,7674	0,088	14,2414	0,0471
m13	24,4974	0,0036	19,7277	0,0114	20,4705	0,0046
m14	12,2169	0,2014	7,4527	0,4887	10,8560	0,145
m15	21,0375	0,0125	10,8800	0,2086	7,5930	0,3698
m16	18,3024	0,0318	9,6914	0,2874	6,6656	0,4645
m17	11,1323	0,2668	19,6588	0,0117	17,6057	0,0139
m18	35,4374	<0,0001	17,8383	0,0225	23,0701	0,0017
m19	10,6750	0,2986	6,9839	0,5384	13,5422	0,0599
m20	5,0645	0,8286	7,8482	0,4484	8,3027	0,3067
m21	3,2441	0,9538	8,1745	0,4166	9,6780	0,2076
m22	29,2565	0,0006	17,9177	0,0219	15,5706	0,0293
m23	53,2818	<0,0001	24,4819	0,0019	19,0277	0,0081
m24	24,6730	0,0034	14,6652	0,066	13,9963	0,0512
m25	13,7652	0,1309	11,5062	0,1746	13,5665	0,0595
m26	25,6856	0,0023	30,2138	0,0002	21,4258	0,0032
m27	11,4136	0,2484	8,0452	0,4291	5,1258	0,6446
m28	10,9105	0,2819	8,1290	0,421	10,5084	0,1615
m29	17,7830	0,0378	9,2806	0,3192	13,0336	0,0713
m30	36,0245	<0,0001	24,6514	0,0018	16,2083	0,0233

EK 5. Farklı Simülasyon Koşulunda Üretilen Her Bir Veri Setinin Farklı Kesme Puanları ve Yöntemler İçin Üretildiği Model Altında Test Edilmesi Durumunda Sınıflama İndekslerinin Değerleri

Simülasyon Koşulu	Kesme Puanı	Rudner'ın İndeksi					Guo'nun İndeksi					Lee'nin İndeksi				
		CA	FP	FN	CC	K	CA	FP	FN	CC	K	CA	FP	FN	CC	K
1	-0,75	0,901	0,474	0,051	0,858	0,580	0,855	0,019	0,126	0,824	0,595	0,848	0,046	0,106	0,788	0,464
1	0	0,859	0,073	0,067	0,799	0,599	0,855	0,081	0,064	0,796	0,592	0,818	0,081	0,101	0,753	0,451
1	0,75	0,885	0,086	0,030	0,842	0,519	0,807	0,186	0,007	0,813	0,578	0,901	0,690	0,030	0,859	0,321
2	-0,75	0,931	0,041	0,028	0,903	0,723	0,902	0,024	0,074	0,866	0,675	0,893	0,035	0,072	0,848	0,622
2	0	0,899	0,042	0,059	0,855	0,711	0,881	0,053	0,066	0,832	0,664	0,863	0,062	0,075	0,810	0,593
2	0,75	0,914	0,060	0,026	0,882	0,648	0,863	0,126	0,011	0,851	0,640	0,919	0,031	0,051	0,897	0,484
3	-0,75	0,895	0,045	0,060	0,851	0,561	0,840	0,019	0,140	0,814	0,576	0,837	0,041	0,122	0,780	0,457
3	0	0,864	0,077	0,060	0,806	0,612	0,857	0,086	0,057	0,799	0,599	0,829	0,078	0,094	0,766	0,475
3	0,75	0,890	0,083	0,028	0,850	0,545	0,817	0,176	0,006	0,825	0,599	0,904	0,064	0,031	0,864	0,352
4	-0,75	0,924	0,033	0,043	0,892	0,696	0,885	0,020	0,094	0,858	0,658	0,888	0,035	0,077	0,847	0,604
4	0	0,904	0,060	0,037	0,865	0,730	0,886	0,071	0,043	0,841	0,681	0,870	0,086	0,071	0,818	0,619
4	0,75	0,915	0,050	0,035	0,880	0,650	0,875	0,109	0,016	0,854	0,652	0,914	0,037	0,049	0,886	0,498
5	-0,75	0,904	0,034	0,062	0,857	0,607	0,868	0,020	0,112	0,831	0,599	0,865	0,043	0,092	0,808	0,503
5	0	0,837	0,067	0,096	0,771	0,539	0,855	0,083	0,063	0,800	0,601	0,756	0,041	0,203	0,742	0,338
5	0,75	0,859	0,081	0,061	0,813	0,405	0,798	0,193	0,010	0,816	0,601	0,872	0,064	0,064	0,842	0,194
6	-0,75	0,923	0,034	0,042	0,893	0,689	0,894	0,018	0,088	0,868	0,674	0,895	0,026	0,079	0,857	0,585
6	0	0,903	0,051	0,046	0,865	0,729	0,890	0,052	0,059	0,846	0,693	0,852	0,033	0,115	0,817	0,634
6	0,75	0,937	0,038	0,026	0,910	0,740	0,900	0,085	0,016	0,877	0,693	0,910	0,055	0,036	0,871	0,612
7	-0,75	0,905	0,029	0,066	0,861	0,609	0,870	0,017	0,113	0,841	0,615	0,863	0,044	0,094	0,808	0,513
7	0	0,836	0,065	0,099	0,769	0,536	0,857	0,081	0,062	0,803	0,605	0,746	0,039	0,215	0,742	0,305
7	0,75	0,854	0,081	0,065	0,807	0,389	0,796	0,193	0,111	0,811	0,591	0,879	0,061	0,060	0,854	0,164
8	-0,75	0,916	0,036	0,048	0,885	0,669	0,885	0,018	0,096	0,864	0,670	0,890	0,030	0,081	0,851	0,567
8	0	0,907	0,049	0,044	0,871	0,741	0,893	0,050	0,057	0,851	0,702	0,856	0,031	0,113	0,821	0,641

8	0,75	0,938	0,036	0,026	0,912	0,742	0,901	0,083	0,015	0,878	0,695	0,909	0,056	0,036	0,870	0,617
9	-0,75	0,870	0,509	0,079	0,822	0,409	0,773	0,009	0,218	0,781	0,527	0,832	0,016	0,153	0,782	0,296
9	0	0,819	0,099	0,082	0,747	0,485	0,832	0,049	0,119	0,795	0,589	0,775	0,071	0,154	0,710	0,416
9	0,75	0,863	0,079	0,058	0,804	0,445	0,869	0,115	0,017	0,851	0,649	0,832	0,107	0,061	0,768	0,397
10	-0,75	0,920	0,034	0,046	0,889	0,667	0,886	0,017	0,097	0,863	0,661	0,879	0,024	0,092	0,845	0,582
10	0	0,885	0,047	0,067	0,834	0,668	0,890	0,053	0,057	0,845	0,690	0,859	0,047	0,094	0,810	0,600
10	0,75	0,881	0,059	0,060	0,839	0,511	0,870	0,113	0,016	0,858	0,666	0,927	0,027	0,046	0,916	0,260
11	-0,75	0,847	0,049	0,104	0,796	0,347	0,738	0,009	0,253	0,772	0,518	0,801	0,014	0,185	0,751	0,269
11	0	0,798	0,122	0,080	0,726	0,442	0,838	0,056	0,106	0,786	0,572	0,769	0,082	0,149	0,705	0,409
11	0,75	0,858	0,096	0,046	0,792	0,475	0,869	0,105	0,026	0,832	0,600	0,867	0,092	0,041	0,811	0,463
12	-0,75	0,913	0,038	0,048	0,875	0,620	0,884	0,014	0,102	0,861	0,659	0,881	0,025	0,094	0,842	0,568
12	0	0,875	0,055	0,070	0,819	0,637	0,886	0,057	0,056	0,839	0,679	0,849	0,049	0,103	0,799	0,571
12	0,75	0,882	0,060	0,058	0,838	0,499	0,868	0,117	0,015	0,855	0,656	0,937	0,027	0,035	0,925	0,233

EK 6. Farklı Simülasyon Koşulunda Üretilen Her Bir Veri Setinin Farklı Kesme Puanları ve Yöntemler İçin Yanlış Model Altında Test Edilmesi Durumunda Sınıflama İndekslerinin Değerleri

Simülasyon Koşulu	Test Edilen Model	Kesme Puanı	Rudner'ın İndeksi					Guo'nun İndeksi					Lee'nin İndeksi				
			CA	FP	FN	CC	K	CA	FP	FN	CC	K	CA	FP	FN	CC	K
1	2PLM	-0,75	0,902	0,044	0,054	0,859	0,588	0,854	0,019	0,128	0,823	0,592	0,845	0,036	0,120	0,789	0,466
1	2PLM	0	0,857	0,070	0,073	0,799	0,597	0,853	0,078	0,069	0,796	0,592	0,807	0,066	0,127	0,752	0,450
1	2PLM	0,75	0,887	0,065	0,049	0,843	0,522	0,832	0,153	0,015	0,816	0,584	0,899	0,062	0,040	0,859	0,316
2	2PLM	-0,75	0,931	0,028	0,041	0,902	0,720	0,892	0,016	0,092	0,867	0,677	0,893	0,040	0,068	0,847	0,617
2	2PLM	0	0,898	0,051	0,051	0,855	0,711	0,882	0,062	0,056	0,834	0,668	0,860	0,055	0,085	0,810	0,592
2	2PLM	0,75	0,914	0,046	0,040	0,881	0,644	0,876	0,106	0,018	0,853	0,646	0,927	0,045	0,029	0,897	0,476
3	2PLM	-0,75	0,896	0,045	0,059	0,852	0,562	0,841	0,019	0,140	0,815	0,577	0,836	0,041	0,123	0,779	0,455
3	2PLM	0	0,862	0,074	0,063	0,804	0,608	0,857	0,083	0,060	0,799	0,599	0,818	0,065	0,117	0,764	0,471
3	2PLM	0,75	0,891	0,068	0,040	0,851	0,543	0,832	0,156	0,012	0,825	0,599	0,905	0,064	0,032	0,865	0,350
4	2PLM	-0,75	0,924	0,034	0,043	0,893	0,689	0,884	0,019	0,098	0,859	0,659	0,886	0,036	0,077	0,844	0,596
4	2PLM	0	0,904	0,049	0,047	0,865	0,731	0,887	0,059	0,054	0,842	0,683	0,866	0,050	0,084	0,816	0,614
4	2PLM	0,75	0,916	0,047	0,037	0,882	0,649	0,876	0,108	0,016	0,855	0,653	0,917	0,041	0,042	0,887	0,494
1	3PLM	-0,75	0,895	0,042	0,063	0,850	0,531	0,840	0,011	0,150	0,825	0,597	0,840	0,034	0,126	0,786	0,457
1	3PLM	0	0,850	0,076	0,074	0,788	0,576	0,856	0,068	0,076	0,799	0,597	0,808	0,057	0,134	0,761	0,472
1	3PLM	0,75	0,884	0,064	0,051	0,838	0,527	0,849	0,132	0,019	0,822	0,590	0,900	0,062	0,037	0,860	0,361

2	3PLM	-0,75	0,925	0,035	0,039	0,894	0,678	0,892	0,015	0,093	0,870	0,679	0,891	0,040	0,069	0,846	0,613
2	3PLM	0	0,892	0,051	0,057	0,847	0,694	0,883	0,052	0,065	0,836	0,672	0,862	0,053	0,085	0,814	0,602
2	3PLM	0,75	0,913	0,048	0,039	0,879	0,641	0,881	0,100	0,019	0,857	0,651	0,926	0,044	0,031	0,896	0,490
3	3PLM	-0,75	0,878	0,049	0,072	0,833	0,480	0,822	0,012	0,166	0,816	0,585	0,828	0,040	0,132	0,774	0,440
3	3PLM	0	0,843	0,094	0,062	0,780	0,559	0,859	0,075	0,066	0,804	0,608	0,824	0,061	0,115	0,773	0,492
3	3PLM	0,75	0,887	0,066	0,048	0,840	0,527	0,852	0,132	0,016	0,831	0,608	0,901	0,054	0,044	0,864	0,384
4	3PLM	-0,75	0,918	0,035	0,046	0,887	0,660	0,881	0,016	0,103	0,861	0,661	0,885	0,036	0,079	0,843	0,592
4	3PLM	0	0,899	0,047	0,054	0,858	0,717	0,888	0,051	0,061	0,843	0,686	0,868	0,050	0,081	0,819	0,620
4	3PLM	0,75	0,915	0,048	0,037	0,880	0,647	0,879	0,104	0,017	0,858	0,657	0,917	0,039	0,044	0,887	0,505
5	1PLM	-0,75	0,891	0,033	0,076	0,843	0,551	0,834	0,012	0,153	0,816	0,582	0,848	0,039	0,113	0,787	0,443
5	1PLM	0	0,826	0,109	0,066	0,762	0,524	0,840	0,115	0,045	0,789	0,577	0,735	0,033	0,232	0,745	0,323
5	1PLM	0,75	0,851	0,110	0,038	0,801	0,407	0,758	0,237	0,005	0,796	0,565	0,898	0,083	0,019	0,850	0,204
6	1PLM	-0,75	0,910	0,049	0,040	0,875	0,642	0,884	0,021	0,095	0,853	0,655	0,895	0,035	0,070	0,851	0,554
6	1PLM	0	0,893	0,070	0,038	0,848	0,696	0,882	0,070	0,048	0,834	0,668	0,848	0,038	0,114	0,805	0,608
6	1PLM	0,75	0,928	0,046	0,026	0,896	0,716	0,881	0,105	0,014	0,859	0,672	0,907	0,058	0,035	0,866	0,582
7	1PLM	-0,75	0,897	0,035	0,068	0,849	0,568	0,845	0,013	0,142	0,821	0,589	0,845	0,037	0,118	0,784	0,455
7	1PLM	0	0,827	0,043	0,130	0,763	0,526	0,846	0,046	0,109	0,788	0,575	0,730	0,031	0,239	0,751	0,308
7	1PLM	0,75	0,850	0,113	0,037	0,799	0,406	0,754	0,241	0,005	0,793	0,558	0,906	0,082	0,012	0,859	0,185
8	1PLM	-0,75	0,909	0,049	0,041	0,873	0,642	0,882	0,021	0,097	0,852	0,657	0,892	0,037	0,071	0,848	0,547

8	1PLM	0	0,897	0,066	0,037	0,854	0,708	0,885	0,068	0,047	0,837	0,675	0,850	0,038	0,111	0,808	0,615
8	1PLM	0,75	0,930	0,042	0,029	0,898	0,716	0,885	0,099	0,015	0,861	0,672	0,907	0,059	0,034	0,866	0,590
5	3PLM	-0,75	0,897	0,043	0,060	0,848	0,556	0,871	0,017	0,112	0,837	0,610	0,866	0,044	0,090	0,809	0,501
5	3PLM	0	0,832	0,071	0,097	0,765	0,529	0,858	0,074	0,068	0,804	0,608	0,756	0,037	0,207	0,747	0,360
5	3PLM	0,75	0,856	0,067	0,076	0,809	0,406	0,831	0,153	0,016	0,822	0,607	0,869	0,061	0,071	0,839	0,217
6	3PLM	-0,75	0,923	0,036	0,042	0,892	0,684	0,895	0,018	0,086	0,869	0,675	0,895	0,028	0,077	0,857	0,585
6	3PLM	0	0,902	0,053	0,045	0,863	0,725	0,890	0,053	0,058	0,847	0,693	0,853	0,034	0,112	0,818	0,635
6	3PLM	0,75	0,936	0,036	0,028	0,909	0,738	0,902	0,081	0,017	0,877	0,695	0,910	0,053	0,037	0,872	0,613
7	3PLM	-0,75	0,903	0,029	0,068	0,858	0,602	0,870	0,016	0,114	0,841	0,615	0,862	0,045	0,094	0,806	0,512
7	3PLM	0	0,836	0,066	0,099	0,770	0,539	0,856	0,079	0,065	0,801	0,602	0,749	0,038	0,214	0,747	0,322
7	3PLM	0,75	0,854	0,083	0,062	0,805	0,397	0,802	0,186	0,012	0,810	0,585	0,882	0,061	0,057	0,854	0,184
8	3PLM	-0,75	0,910	0,040	0,050	0,878	0,637	0,885	0,017	0,098	0,866	0,673	0,887	0,034	0,080	0,847	0,558
8	3PLM	0	0,901	0,051	0,048	0,861	0,722	0,894	0,046	0,060	0,852	0,705	0,857	0,032	0,111	0,822	0,644
8	3PLM	0,75	0,937	0,039	0,024	0,910	0,738	0,901	0,084	0,015	0,879	0,697	0,911	0,054	0,035	0,873	0,625
9	1PLM	-0,75	0,901	0,027	0,073	0,858	0,519	0,815	0,009	0,176	0,799	0,524	0,853	0,033	0,114	0,792	0,384
9	1PLM	0	0,843	0,070	0,087	0,778	0,556	0,832	0,078	0,090	0,766	0,533	0,765	0,048	0,187	0,712	0,422
9	1PLM	0,75	0,900	0,056	0,043	0,857	0,510	0,835	0,152	0,013	0,809	0,548	0,837	0,124	0,039	0,774	0,396
10	1PLM	-0,75	0,920	0,030	0,050	0,887	0,683	0,876	0,018	0,106	0,852	0,648	0,876	0,018	0,107	0,845	0,578
10	1PLM	0	0,885	0,065	0,050	0,836	0,672	0,880	0,081	0,039	0,837	0,673	0,845	0,041	0,114	0,798	0,572

10	1PLM	0,75	0,882	0,047	0,071	0,840	0,508	0,868	0,113	0,019	0,849	0,655	0,923	0,023	0,054	0,923	0,206
11	1PLM	-0,75	0,885	0,040	0,074	0,839	0,468	0,807	0,014	0,179	0,791	0,517	0,827	0,040	0,133	0,765	0,366
11	1PLM	0	0,854	0,081	0,065	0,790	0,580	0,843	0,088	0,069	0,778	0,555	0,774	0,054	0,171	0,720	0,439
11	1PLM	0,75	0,902	0,067	0,031	0,862	0,534	0,828	0,163	0,009	0,819	0,568	0,853	0,114	0,033	0,796	0,416
12	1PLM	-0,75	0,927	0,028	0,045	0,894	0,694	0,881	0,018	0,101	0,852	0,642	0,881	0,019	0,100	0,845	0,579
12	1PLM	0	0,882	0,062	0,057	0,829	0,657	0,876	0,078	0,046	0,827	0,655	0,832	0,038	0,130	0,790	0,551
12	1PLM	0,75	0,886	0,048	0,067	0,844	0,500	0,864	0,118	0,018	0,843	0,635	0,933	0,023	0,043	0,928	0,214
9	2PLM	-0,75	0,884	0,047	0,069	0,837	0,500	0,818	0,016	0,167	0,800	0,551	0,837	0,019	0,143	0,785	0,339
9	2PLM	0	0,840	0,083	0,077	0,775	0,548	0,845	0,089	0,066	0,786	0,572	0,765	0,060	0,175	0,702	0,402
9	2PLM	0,75	0,883	0,061	0,056	0,838	0,486	0,828	0,160	0,012	0,818	0,593	0,828	0,111	0,061	0,761	0,345
10	2PLM	-0,75	0,924	0,035	0,041	0,893	0,697	0,889	0,023	0,089	0,859	0,658	0,882	0,020	0,098	0,849	0,591
10	2PLM	0	0,888	0,042	0,070	0,835	0,667	0,890	0,058	0,052	0,844	0,688	0,847	0,043	0,110	0,800	0,579
10	2PLM	0,75	0,882	0,054	0,064	0,841	0,500	0,864	0,123	0,013	0,858	0,673	0,937	0,032	0,031	0,924	0,174
11	2PLM	-0,75	0,872	0,051	0,077	0,822	0,474	0,807	0,019	0,174	0,787	0,529	0,817	0,021	0,161	0,765	0,338
11	2PLM	0	0,841	0,082	0,077	0,776	0,550	0,844	0,088	0,068	0,782	0,564	0,765	0,069	0,165	0,701	0,401
11	2PLM	0,75	0,885	0,069	0,047	0,841	0,500	0,820	0,169	0,012	0,813	0,581	0,845	0,111	0,044	0,784	0,355
12	2PLM	-0,75	0,926	0,027	0,047	0,893	0,702	0,884	0,018	0,098	0,856	0,654	0,885	0,021	0,094	0,848	0,587
12	2PLM	0	0,882	0,045	0,073	0,828	0,654	0,885	0,061	0,054	0,837	0,674	0,840	0,046	0,114	0,789	0,551
12	2PLM	0,75	0,883	0,056	0,060	0,840	0,495	0,858	0,130	0,012	0,850	0,656	0,945	0,030	0,025	0,932	0,169

EK 7. Farklı Simülasyon Koşulunda Üretilen Her Bir Veri Setinin Farklı Kesme Puanları ve Yöntemler İçin Üretildiği Model Altında Test Edilmesi Durumunda Sınıflama İndekslerinin Mutlak Yanlılık Değerleri

Simülasyon Koşulu	Kesme Puanı	Rudner'ın İndeksi					Guo'nun İndeksi					Lee'nin İndeksi				
		CA	FP	FN	CC	κ	CA	FP	FN	CC	κ	CA	FP	FN	CC	κ
1	-0,75	0,001	0,402	0,025	0,142	0,420	0,047	0,053	0,100	0,176	0,405	0,040	0,006	0,034	0,212	0,536
1	0	0,013	0,019	0,005	0,201	0,401	0,009	0,011	0,002	0,204	0,408	0,028	0,011	0,039	0,247	0,549
1	0,75	0,017	0,050	0,066	0,158	0,481	0,061	0,150	0,089	0,187	0,422	0,037	0,608	0,024	0,141	0,679
2	-0,75	0,037	0,021	0,016	0,097	0,277	0,008	0,038	0,030	0,134	0,325	0,011	0,011	0,000	0,152	0,378
2	0	0,013	0,044	0,031	0,145	0,289	0,005	0,033	0,038	0,168	0,336	0,027	0,008	0,019	0,190	0,407
2	0,75	0,014	0,026	0,040	0,118	0,352	0,037	0,092	0,055	0,149	0,360	0,025	0,037	0,013	0,103	0,516
3	-0,75	0,025	0,040	0,015	0,149	0,439	0,030	0,066	0,095	0,186	0,424	0,013	0,003	0,010	0,220	0,543
3	0	0,042	0,003	0,044	0,194	0,388	0,035	0,012	0,047	0,201	0,401	0,007	0,004	0,010	0,234	0,525
3	0,75	0,017	0,059	0,075	0,150	0,455	0,056	0,152	0,097	0,175	0,401	0,035	0,004	0,032	0,136	0,648
4	-0,75	0,024	0,028	0,004	0,108	0,304	0,015	0,041	0,055	0,142	0,342	0,006	0,004	0,002	0,153	0,396
4	0	0,033	0,008	0,040	0,135	0,270	0,015	0,019	0,034	0,159	0,319	0,001	0,034	0,006	0,182	0,381
4	0,75	0,028	0,007	0,035	0,120	0,350	0,012	0,066	0,054	0,146	0,348	0,036	0,042	0,006	0,114	0,502
5	-0,75	0,000	0,018	0,018	0,143	0,393	0,036	0,032	0,068	0,169	0,401	0,005	0,005	0,010	0,192	0,497
5	0	0,015	0,017	0,002	0,229	0,461	0,033	0,001	0,031	0,200	0,399	0,046	0,033	0,079	0,258	0,662
5	0,75	0,025	0,017	0,041	0,187	0,595	0,036	0,129	0,092	0,184	0,399	0,038	0,062	0,024	0,158	0,806
6	-0,75	0,017	0,016	0,002	0,107	0,311	0,012	0,032	0,044	0,132	0,326	0,007	0,022	0,015	0,143	0,415
6	0	0,011	0,007	0,004	0,135	0,271	0,002	0,006	0,009	0,154	0,307	0,028	0,019	0,047	0,183	0,366
6	0,75	0,029	0,006	0,022	0,090	0,260	0,008	0,041	0,032	0,123	0,307	0,002	0,005	0,006	0,129	0,388
7	-0,75	0,007	0,042	0,035	0,139	0,391	0,028	0,054	0,082	0,159	0,385	0,003	0,002	0,004	0,192	0,487
7	0	0,006	0,016	0,010	0,231	0,464	0,027	0,000	0,027	0,197	0,395	0,062	0,026	0,088	0,258	0,695
7	0,75	0,009	0,027	0,036	0,193	0,611	0,049	0,139	0,010	0,189	0,409	0,054	0,052	0,002	0,146	0,836
8	-0,75	0,011	0,023	0,012	0,115	0,331	0,020	0,041	0,060	0,136	0,330	0,010	0,018	0,029	0,149	0,433
8	0	0,029	0,010	0,019	0,129	0,259	0,015	0,009	0,006	0,149	0,298	0,014	0,020	0,034	0,179	0,359

8	0,75	0,016	0,006	0,022	0,088	0,258	0,021	0,053	0,033	0,122	0,305	0,005	0,015	0,009	0,130	0,383
9	-0,75	0,046	0,395	0,017	0,178	0,591	0,051	0,105	0,156	0,219	0,473	0,014	0,070	0,057	0,218	0,704
9	0	0,027	0,023	0,004	0,253	0,515	0,040	0,073	0,033	0,205	0,411	0,025	0,025	0,050	0,290	0,584
9	0,75	0,493	0,833	0,001	0,942	0,196	0,445	0,085	0,063	0,149	0,351	0,058	0,780	0,049	0,939	0,232
10	-0,75	0,330	0,872	0,006	0,954	0,111	0,245	0,031	0,057	0,137	0,339	0,021	0,831	0,028	0,908	0,155
10	0	0,005	0,003	0,007	0,166	0,332	0,000	0,003	0,003	0,155	0,310	0,023	0,027	0,050	0,190	0,400
10	0,75	0,023	0,013	0,036	0,161	0,489	0,012	0,067	0,080	0,142	0,334	0,075	0,033	0,042	0,084	0,740
11	-0,75	0,020	0,079	0,059	0,204	0,653	0,089	0,119	0,208	0,228	0,482	0,014	0,054	0,068	0,249	0,731
11	0	0,025	0,021	0,004	0,274	0,558	0,015	0,045	0,030	0,214	0,428	0,047	0,008	0,039	0,295	0,591
11	0,75	0,009	0,038	0,029	0,208	0,525	0,002	0,047	0,049	0,168	0,400	0,013	0,019	0,032	0,189	0,537
12	-0,75	0,327	0,849	0,003	0,952	0,125	0,275	0,050	0,061	0,139	0,341	0,012	0,821	0,022	0,906	0,158
12	0	0,308	0,816	0,007	0,930	0,181	0,242	0,002	0,006	0,161	0,321	0,024	0,770	0,001	0,897	0,201
12	0,75	0,008	0,024	0,016	0,162	0,501	0,022	0,081	0,059	0,145	0,344	0,053	0,018	0,036	0,075	0,767

EK 8. Farklı Simülasyon Koşulunda Üretilen Her Bir Veri Setinin Farklı Kesme Puanları ve Yöntemler İçin Yanlış Model Altında Test Edilmesi Durumunda Sınıflama İndekslerinin Mutlak Yanlılık Değerleri

Simülasyon Koşulu	Test Edilen Model	Kesme Puanı	Rudner'ın İndeksi					Guo'nun İndeksi					Lee'nin İndeksi				
			CA	FP	FN	CC	κ	CA	FP	FN	CC	κ	CA	FP	FN	CC	κ
1	2PLM	-0,75	0,004	0,032	0,028	0,141	0,412	0,044	0,057	0,102	0,177	0,408	0,043	0,004	0,048	0,211	0,534
1	2PLM	0	0,013	0,026	0,013	0,201	0,403	0,009	0,018	0,009	0,204	0,408	0,039	0,026	0,065	0,248	0,550
1	2PLM	0,75	0,023	0,007	0,029	0,157	0,478	0,032	0,095	0,063	0,184	0,416	0,035	0,020	0,014	0,141	0,684
2	2PLM	-0,75	0,047	0,052	0,005	0,098	0,280	0,008	0,064	0,056	0,133	0,323	0,011	0,006	0,004	0,153	0,383
2	2PLM	0	0,012	0,025	0,013	0,145	0,289	0,004	0,014	0,018	0,166	0,332	0,030	0,001	0,029	0,190	0,408
2	2PLM	0,75	0,012	0,000	0,012	0,119	0,356	0,026	0,060	0,034	0,147	0,354	0,033	0,023	0,009	0,103	0,524
3	2PLM	-0,75	0,026	0,040	0,014	0,148	0,438	0,029	0,066	0,095	0,185	0,423	0,014	0,003	0,011	0,221	0,545
3	2PLM	0	0,038	0,002	0,037	0,196	0,392	0,033	0,007	0,040	0,201	0,401	0,004	0,009	0,013	0,236	0,529
3	2PLM	0,75	0,023	0,029	0,053	0,149	0,457	0,036	0,117	0,081	0,175	0,401	0,036	0,004	0,031	0,135	0,650
4	2PLM	-0,75	0,028	0,033	0,006	0,107	0,311	0,012	0,048	0,061	0,141	0,341	0,004	0,003	0,002	0,156	0,404
4	2PLM	0	0,040	0,019	0,021	0,135	0,269	0,023	0,009	0,014	0,158	0,317	0,005	0,002	0,007	0,184	0,386
4	2PLM	0,75	0,029	0,004	0,033	0,118	0,351	0,011	0,065	0,054	0,145	0,347	0,039	0,038	0,001	0,113	0,506
1	3PLM	-0,75	0,005	0,052	0,047	0,150	0,469	0,050	0,083	0,134	0,175	0,403	0,048	0,006	0,054	0,214	0,543
1	3PLM	0	0,002	0,028	0,026	0,212	0,424	0,008	0,036	0,028	0,201	0,403	0,038	0,035	0,072	0,239	0,528
1	3PLM	0,75	0,022	0,002	0,021	0,162	0,473	0,013	0,066	0,053	0,178	0,410	0,036	0,020	0,017	0,140	0,639

2	3PLM	-0,75	0,043	0,051	0,007	0,106	0,322	0,010	0,071	0,061	0,130	0,321	0,009	0,006	0,003	0,154	0,387
2	3PLM	0	0,014	0,043	0,029	0,153	0,306	0,005	0,042	0,037	0,164	0,328	0,028	0,001	0,029	0,186	0,398
2	3PLM	0,75	0,019	0,020	0,001	0,121	0,359	0,013	0,032	0,019	0,143	0,349	0,032	0,024	0,007	0,104	0,510
3	3PLM	-0,75	0,020	0,054	0,033	0,167	0,520	0,036	0,091	0,127	0,184	0,415	0,022	0,002	0,020	0,226	0,560
3	3PLM	0	0,025	0,007	0,033	0,220	0,441	0,041	0,012	0,029	0,196	0,392	0,002	0,013	0,011	0,227	0,508
3	3PLM	0,75	0,026	0,014	0,039	0,160	0,473	0,009	0,080	0,071	0,169	0,392	0,032	0,014	0,019	0,136	0,616
4	3PLM	-0,75	0,023	0,038	0,014	0,113	0,340	0,014	0,057	0,071	0,139	0,339	0,003	0,003	0,000	0,157	0,408
4	3PLM	0	0,038	0,033	0,005	0,142	0,283	0,027	0,029	0,002	0,157	0,314	0,003	0,002	0,004	0,181	0,380
4	3PLM	0,75	0,028	0,005	0,033	0,120	0,353	0,008	0,061	0,053	0,142	0,343	0,039	0,040	0,001	0,113	0,495
5	1PLM	-0,75	0,003	0,037	0,040	0,157	0,449	0,060	0,058	0,117	0,184	0,418	0,012	0,001	0,011	0,213	0,557
5	1PLM	0	0,024	0,035	0,058	0,238	0,476	0,038	0,041	0,079	0,211	0,423	0,067	0,041	0,108	0,255	0,677
5	1PLM	0,75	0,019	0,062	0,082	0,199	0,593	0,074	0,189	0,115	0,204	0,435	0,064	0,043	0,021	0,150	0,796
6	1PLM	-0,75	0,022	0,001	0,024	0,125	0,358	0,004	0,027	0,031	0,147	0,345	0,007	0,013	0,006	0,149	0,446
6	1PLM	0	0,013	0,018	0,030	0,152	0,304	0,002	0,018	0,020	0,166	0,332	0,032	0,014	0,046	0,195	0,392
6	1PLM	0,75	0,020	0,004	0,016	0,104	0,284	0,027	0,055	0,028	0,141	0,328	0,001	0,008	0,007	0,134	0,418
7	1PLM	-0,75	0,013	0,045	0,032	0,151	0,432	0,039	0,067	0,106	0,179	0,411	0,015	0,005	0,020	0,216	0,545
7	1PLM	0	0,014	0,092	0,078	0,237	0,474	0,033	0,089	0,057	0,212	0,425	0,078	0,034	0,112	0,249	0,692
7	1PLM	0,75	0,007	0,078	0,085	0,201	0,594	0,089	0,206	0,117	0,207	0,442	0,081	0,031	0,050	0,141	0,815
8	1PLM	-0,75	0,009	0,001	0,011	0,127	0,358	0,018	0,027	0,045	0,148	0,343	0,008	0,011	0,019	0,152	0,453

8	1PLM	0	0,027	0,015	0,042	0,146	0,292	0,015	0,017	0,032	0,163	0,325	0,020	0,013	0,032	0,192	0,385
8	1PLM	0,75	0,016	0,001	0,016	0,102	0,284	0,029	0,058	0,030	0,139	0,328	0,007	0,018	0,011	0,134	0,410
5	3PLM	-0,75	0,001	0,017	0,016	0,152	0,444	0,025	0,043	0,068	0,163	0,390	0,006	0,006	0,012	0,191	0,499
5	3PLM	0	0,012	0,021	0,009	0,235	0,471	0,038	0,018	0,020	0,196	0,392	0,046	0,037	0,083	0,253	0,640
5	3PLM	0,75	0,024	0,015	0,010	0,191	0,594	0,001	0,071	0,070	0,178	0,393	0,035	0,065	0,031	0,161	0,783
6	3PLM	-0,75	0,017	0,014	0,002	0,108	0,316	0,011	0,032	0,042	0,131	0,325	0,007	0,020	0,013	0,143	0,415
6	3PLM	0	0,008	0,005	0,003	0,137	0,275	0,004	0,005	0,010	0,153	0,307	0,027	0,018	0,044	0,182	0,365
6	3PLM	0,75	0,030	0,012	0,018	0,091	0,262	0,004	0,033	0,029	0,123	0,305	0,002	0,003	0,005	0,128	0,387
7	3PLM	-0,75	0,003	0,042	0,039	0,142	0,398	0,030	0,055	0,085	0,159	0,385	0,002	0,003	0,004	0,194	0,488
7	3PLM	0	0,009	0,019	0,011	0,230	0,461	0,029	0,006	0,023	0,199	0,398	0,059	0,027	0,087	0,253	0,678
7	3PLM	0,75	0,009	0,028	0,038	0,195	0,603	0,043	0,131	0,088	0,190	0,415	0,057	0,052	0,005	0,146	0,816
8	3PLM	-0,75	0,009	0,025	0,016	0,122	0,363	0,016	0,048	0,064	0,134	0,327	0,013	0,014	0,028	0,153	0,442
8	3PLM	0	0,027	0,018	0,009	0,139	0,278	0,020	0,023	0,003	0,148	0,295	0,013	0,019	0,032	0,178	0,356
8	3PLM	0,75	0,021	0,007	0,028	0,090	0,262	0,015	0,052	0,037	0,121	0,303	0,003	0,013	0,010	0,127	0,375
9	1PLM	-0,75	0,063	0,091	0,029	0,142	0,481	0,023	0,109	0,132	0,201	0,476	0,035	0,053	0,018	0,208	0,616
9	1PLM	0	0,043	0,026	0,017	0,222	0,444	0,032	0,018	0,014	0,234	0,467	0,035	0,048	0,083	0,288	0,578
9	1PLM	0,75	0,022	0,042	0,065	0,143	0,490	0,043	0,138	0,095	0,191	0,452	0,053	0,072	0,019	0,226	0,604
10	1PLM	-0,75	0,020	0,018	0,002	0,113	0,317	0,024	0,030	0,054	0,148	0,352	0,024	0,030	0,055	0,155	0,422
10	1PLM	0	0,009	0,033	0,042	0,164	0,328	0,004	0,049	0,053	0,163	0,327	0,037	0,033	0,070	0,202	0,428

10	1PLM	0,75	0,030	0,013	0,017	0,160	0,492	0,016	0,053	0,069	0,151	0,345	0,071	0,037	0,034	0,077	0,794
11	1PLM	-0,75	0,048	0,076	0,027	0,161	0,532	0,030	0,102	0,132	0,209	0,483	0,012	0,028	0,016	0,235	0,634
11	1PLM	0	0,038	0,007	0,045	0,210	0,420	0,027	0,014	0,041	0,222	0,445	0,042	0,020	0,061	0,280	0,561
11	1PLM	0,75	0,044	0,043	0,087	0,138	0,466	0,030	0,139	0,109	0,181	0,432	0,001	0,041	0,040	0,204	0,584
12	1PLM	-0,75	0,034	0,032	0,002	0,106	0,306	0,012	0,042	0,054	0,148	0,358	0,012	0,041	0,053	0,155	0,421
12	1PLM	0	0,012	0,015	0,026	0,171	0,343	0,006	0,031	0,037	0,173	0,345	0,041	0,041	0,082	0,210	0,449
12	1PLM	0,75	0,002	0,003	0,004	0,156	0,500	0,020	0,073	0,053	0,157	0,365	0,049	0,022	0,028	0,072	0,786
9	2PLM	-0,75	0,060	0,053	0,007	0,163	0,500	0,006	0,084	0,091	0,200	0,449	0,019	0,067	0,047	0,215	0,661
9	2PLM	0	0,030	0,011	0,041	0,225	0,452	0,035	0,017	0,052	0,214	0,428	0,035	0,036	0,071	0,298	0,598
9	2PLM	0,75	0,003	0,035	0,038	0,162	0,514	0,052	0,134	0,082	0,182	0,407	0,062	0,059	0,003	0,239	0,655
10	2PLM	-0,75	0,020	0,005	0,015	0,107	0,303	0,015	0,017	0,033	0,141	0,342	0,018	0,028	0,046	0,151	0,409
10	2PLM	0	0,002	0,000	0,002	0,165	0,333	0,004	0,016	0,020	0,156	0,312	0,035	0,031	0,066	0,200	0,421
10	2PLM	0,75	0,020	0,010	0,030	0,159	0,500	0,002	0,079	0,081	0,142	0,327	0,085	0,028	0,057	0,076	0,826
11	2PLM	-0,75	0,036	0,047	0,011	0,178	0,526	0,029	0,079	0,108	0,213	0,471	0,002	0,047	0,044	0,235	0,662
11	2PLM	0	0,016	0,012	0,028	0,224	0,450	0,019	0,018	0,037	0,218	0,436	0,051	0,005	0,055	0,299	0,599
11	2PLM	0,75	0,023	0,031	0,053	0,159	0,500	0,042	0,131	0,088	0,187	0,419	0,009	0,038	0,029	0,216	0,645
12	2PLM	-0,75	0,032	0,029	0,003	0,107	0,298	0,010	0,038	0,048	0,144	0,346	0,008	0,039	0,047	0,152	0,413
12	2PLM	0	0,003	0,007	0,004	0,172	0,346	0,006	0,009	0,015	0,163	0,326	0,033	0,033	0,066	0,211	0,449
12	2PLM	0,75	0,015	0,024	0,010	0,160	0,505	0,040	0,098	0,058	0,150	0,344	0,061	0,015	0,046	0,068	0,831



GAZİLİ OLMAK AYRICALIKTIR...