ANKARA ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

EĞİTİM BİLİMLERİ ANABİLİM DALI
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME PROGRAMI

KLASİK VE BAYESÇİ YAKLAŞIMLARA GÖRE MADDE TEPKİ KURAMI PARAMETRE KESTİRİMLERİNİN FARKLI SİMÜLASYON KOŞULLARINDA KARŞILAŞTIRILMASI

DOKTORA TEZİ

ERAY SELÇUK

ANKARA
ŞUBAT, 2023


ANKARA ÜNİVERSİTESİ
1946

ANKARA ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

EĞİTİM BİLİMLERİ ANABİLİM DALI
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME PROGRAMI

KLASİK VE BAYESÇİ YAKLAŞIMLARA GÖRE MADDE TEPKİ KURAMI PARAMETRE KESTİRİMLERİNİN FARKLI SİMÜLASYON KOŞULLARINDA KARŞILAŞTIRILMASI

DOKTORA TEZİ

ERAY SELÇUK

DANIŞMAN: DOÇ. DR. ERGÜL DEMİR

ANKARA
ŞUBAT, 2023

Ankara Üniversitesi Eğitim Bilimleri Enstitüsü Müdürlüğüne,

Eray Selçuk adlı öğrencinin hazırladığı "Klasik ve Bayesçi Yaklaşımlara Göre Madde Tepki Kuramı Parametre Kestirimlerinin Farklı Simülasyon Koşullarında Karşılaştırılması" başlıklı bu çalışma Eğitim Bilimleri Anabilim Dalı / Eğitimde Ölçme ve Değerlendirme Doktora Programı'nda jüri üyelerince oy birliği ile **Doktora Tezi** Olarak Kabul Edilmiştir.

	Jüri Üyeleri	İmza
Başkan	Prof. Dr. Duygu ANIL	……...
Üye	Doç. Dr. Ergül DEMİR (Danışman)	……...
Üye	Doç. Dr. H. Deniz GÜLLEROĞLU	……...
Üye	Prof. Dr. Ömay ÇOKLUK BÖKEOĞLU	……..
Üye	Prof. Dr. Burcu ATAR	……...

ONAY

Bu tez Ankara Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliği'nin ilgili maddeleri uyarınca jüri üyeleri tarafından 09/02/2023 tarihinde, Enstitü Yönetim Kurulunca …/02/2023 tarihinde kabul edilmiştir.

Prof. Dr. İlhan YALÇIN

Eğitim Bilimleri Enstitüsü Müdürü

ETİK İLKELERE UYGUNLUK BİLDİRİMİ

Tez içindeki bütün bilgileri akademik yazım kurallarına uygun biçimde raporlaştırdığımı ve bunları etik ilkelere (atıfta bulunulan tüm yapıtlara kaynaklarda yer verilmesi, tezde kullanılan bilgi ve belgelere resmi yollarla ulaşılması ve bunların aslı bozulmadan kullanılması vb.) uygun olarak elde ettiğimi ve sunduğumu bildiririm.

Eray SELÇUK

ÖZET

KLASİK VE BAYESÇİ YAKLAŞIMLARA GÖRE MADDE TEPKİ KURAMI PARAMETRE KESTİRİMLERİNİN FARKLI SİMÜLASYON KOŞULLARINDA KARŞILAŞTIRILMASI

SELÇUK, Eray
Doktora Tezi, Eğitim Bilimleri Anabilim Dalı
Tez Danışmanı: Doç. Dr. Ergül DEMİR
Şubat, 2023, xv + 149 Sayfa

Bu araştırmanın amacı, farklı simülasyon koşularında Klasik ve Bayesçi yaklaşımlara göre Madde Tepki Kuramı (MTK)'nın yetenek ve madde parametre kestirimlerinin karşılaştırılmasıdır. Bu amaçla; önsel yetenek (theta) dağılım biçiminin, örneklem büyüklüğünün, madde sayısının ve modelin kestirilen parametre sayısının değişimlerine bağlı olarak; klasik en çok olabilirlik ve Bayes yöntemine göre kestirilen yetenek ve madde parametreleri ile bu parametrelerin RMSE değerlerindeki farklılaşmalar incelenmiştir. Verilerin üretilmesinde Monte Carlo Simülasyon tekniği kullanılmıştır. Buna göre; parametre kestirim yöntemleri sabit tutularak, farklı önsel yetenek dağılım biçimleri, örneklem büyüklüğü, madde sayısı ve modelin kestirilen parametre sayısı değişimlenerek simülasyon koşulları oluşturulmuştur. Önsel yetenek dağılımı (normal, sola çarpık, sağa çarpık, sivri ve basık), madde sayısı (10, 20, 40), örneklem büyüklüğü (100, 500, 1000), model parametre sayısı (2 PL, 3 PL) olmak üzere 5x3x3x2=90 adet simülasyon koşulu 100 tekrarla gerçekleştirilmiştir. Verilerin üretilmesinde, parametrelerin ve RMSE değerlerinin kestiriminde, RMSE değerlerindeki farklılaşmaların incelenmesinde ve grafiklerin çiziminde R programlama dili kullanılmıştır. Simülasyon koşullarına göre üretilen verilerin MTK varsayımlarına göre tek boyutluluk, yerel bağımsızlık ve model-veri uyumu analizleri yapılmıştır. Ayrıca simülasyon verilerinin geçerliği için istatistiksel ve mantıksal kontroller yapılmıştır. Simülasyon koşullarının her birinde klasik ve Bayesçi yaklaşımlara göre MTK'nın yetenek ve madde parametre kestirimlerindeki RMSE değerlerinin farklılaşmaları incelenmiştir. Farklılaşmaların belirlenmesinde karma model ANOVA analizi

gerçekleştirilmiştir. Buna göre önsel yetenek dağılımı normal, sola çarpık, sağa çarpık, sivri ya da basık olan verilerde klasik (en çok olabilirlik) ve Bayes kestirimler arasında hem yetenek hem de madde parametrelerinin RMSE değerlerinde manidar farklılaşmalar olduğu görülmüştür. 2 PL modelde kestirilen yetenek parametresi ve RMSE değerlerinin farklılaşmasında önsel yetenek dağılım biçimi, madde sayısı ve kestirim yöntemi; 3 PL modelde kestirilen yetenek parametresi ve RMSE değerlerindeki farklılaşmalarda önsel yetenek dağılım biçimi ve madde sayısı etkili olmuştur. 2 PL modelde kestirilen madde ayırt edicilik parametresi RMSE değerlerinde önsel yetenek dağılım biçimi, örneklem büyüklüğü ve kestirim yöntemi manidar farklılaşma yaratırken madde güçlük parametresi RMSE değerlerinde hiçbir koşul manidar bir farklılaşma oluşturmamıştır. 3 PL modelde kestirilen madde ayırt edicilik RMSE değerlerinde önsel yetenek dağılım biçimi, örneklem büyüklüğü ve kestirim yöntemi; şans parametresi RMSE değerlerinde önsel yetenek dağılım biçimi ve kestirim yöntemi manidar farklılaşma oluşturmuştur. Ancak madde güçlük parametreleri RMSE değerlerinde hiçbir koşul manidar farklılaşma oluşturmamıştır. Yetenek ve madde parametreleri için manidar olan durumlar önsel yetenek dağılımı sola ve sağa çarpık olduğu, örneklem büyüklüğünün küçüldüğü ve madde sayısının azaldığı durumlarda Bayes kestirim yöntemi lehinedir. Bayes kestirim yöntemi klasik en çok olabilirlik yöntemine göre daha düşük hata değerleri kestirmektedir. Buna göre önsel yetenek dağılımı normalden dağılımdan uzaklaştıkça, örneklem küçüldükçe ve madde sayısı azaldıkça yetenek ve madde parametrelerinin kestiriminde Bayes yönteminin kullanılması önerilmektedir.

Anahtar Sözcükler: MTK parametre kestirimi, En çok olabilirlik kestirim, Bayes kestirim, MCMC, R programlama dili

ABSTRACT

COMPARISON OF ITEM RESPONSE THEORY PARAMETER ESTIMATIONS ACCORDING TO CLASSICAL AND BAYESIAN APPROACHES IN DIFFERENT SIMULATION CONDITIONS

SELÇUK, Eray

Dissertation, Department of Educational Sciences

Supervisor: Assoc. Prof. Dr. Ergül DEMİR

February, 2023, xv + 149 Pages

The purpose of this research is to compare the ability and item parameter estimations of Item Response Theory (IRT) according to the Classical and Bayesian approaches in different simulation conditions. For this purpose, normal, skewed left, skewed right, leptokurtic, platykurtic five different type including the distribution sample size, item count (test length), and estimated parameters of the model depending on the number of cases of classical maximum likelihood method and the Bayesian method to differentiation according to the differences in ability and item parameters and the RMSE value were investigated. Monte Carlo Simulation technique was used to generate the data. According to this; by keeping the estimation methods constant, simulation conditions were created by manipulating the sample size, item count and parameters of the model in different types of distribution. Prior distribution (normal, skewed left, skewed right, leptokurtic, platykurtic), item count (10, 20, 40), sample size (100, 500, 1000), model parameters (2 PL, 3 PL) and 5x3x3x2=90 simulation conditions were carried out with 100 replications. The R program was used to generate the data, estimate the parameters and draw the graphs. The unidimensionality, local independence and model-data fit analyzes of the data produced according to the simulation conditions were made according to the assumptions of the IRT. Also, statistical and logical validity checks were made for the simulation data. In each of the simulation conditions, the differences in the RMSE values in the ability and item parameter estimations of the IRT according to the maximum likelihood and Bayesian approaches were examined. Mixed model ANOVA analysis was performed to determine the differentiations. Accordingly, it was observed

that there were significant differences in the RMSE values of both ability and item parameters between classical (maximum likelihood) and Bayesian estimations in data with normal, left skewed, right skewed, leptokurtic or platykurtic a priori ability distribution. The a priori ability distribution type, number of items and estimation method were significant in the differentiation of the estimated ability parameter and RMSE values in the 2 PL model. The a priori ability distribution type and the number of items were significant in the differences in the ability parameter and RMSE values estimated in the 3 PL model. While a priori ability distribution type, sample size and estimation method created significant differences in the RMSE values of the item discrimination parameter estimated in the 2PL model, none of conditions created a significant difference in the RMSE values of the item difficulty parameter. A priori ability distribution type, sample size and estimation method created significant differences in the item discrimination RMSE values predicted in the 3PL model. Prior ability distribution and estimation method created significant differentiation in the RMSE values of the lower asymptote parameter. However, none of the conditions created a significant difference in the RMSE values of item difficulty parameters. Significant conditions for ability and item parameters are in favor of Bayes estimation method when the a priori ability distribution is skewed to the left and right, the sample size is smaller and the number of items decreases. Bayesian estimation method estimates lower error values than classical maximum likelihood method. Accordingly, as the a priori ability distribution moves away from the normal distribution, as the sample gets smaller and the number of items decreases, it is recommended to use the Bayesian method in the estimation of ability and item parameters.

Key Words: IRT Techniques of Parameter Estimation, Maximum Likelihood Estimation, Bayesian Estimation, MCMC, R Program Language

ÖNSÖZ

Ölçme ve değerlendirme, bir disiplin ya da bir uzmanlık alanı olmaktan çok bireyin yaşama uyum sağlamasında önemli bir davranış olan karar verme mekanizmasının kendisidir. Çevreyi öğrenirken, dostluklar kurarken, alışveriş yaparken her süreçte bir ölçme ve değerlendirme vardır. İşte bu çalışma, bu sürecin bir ürünüdür.

Bu süreçte, istatistik bilgimin sağlamlaşmasında, doktora sürecinde farklı araştırma konularıyla yön bulmamı sağlayan, doktora eğitimimin başlangıcından bitişine kadar her soruya, her kafa karışıklığına ve her kayboluşuma rehber olan ve kişilik özellikleriyle de örnek aldığım çok değerli danışmanım Doç. Dr. Ergül DEMİR'e teşekkürden fazlasını bir borç bilirim. Ve çalışma yoldaşlığımızın uzun yıllar devam etmesini dilerim.

Doktora çalışmam süresince hem bilimsel hem de psikolojik desteğini esirgemeyen, tezimin konusuyla ilgili nokta atış kaynaklara yönlendiren ve farklı sorularıyla araştırmamın yönlenmesine katkıda bulunan aynı zamanda enerjisine hayranlık duyduğum hocam, Prof. Dr. Duygu ANIL'a teşekkür ederim.

Diğer bir TİK üyesi olan ve tezin her aşamasında dönütleriyle katkı sağlayan, tezimin eksikliklerini görmem için yönlendirici sorular soran, her karşılaşmamızda bir merhabadan fazlasını paylaştığımız değerli hocam, Doç. Dr. H. Deniz GÜLLEROĞLU'na teşekkürlerimi sunarım.

Değerli görüş ve önerileriyle tezime katkı sağlayan değerli jüri üyeleri ve saygıdeğer hocalarım Prof. Dr. Burcu ATAR'a ve Prof. Dr. Ömay ÇOKLUK BÖKEOĞLU'na ayrıca teşekkür ederim.

Sürecin en başından beri maddi ve manevi destekçim, yol arkadaşım, kader ortağım ve kıymetli eşim Gülbörteçine SELÇUK'a teşekkürden fazlasını bir ömür boyu vermek isterim.

Eğitim hayatım boyunca her zaman yanımda olan, varlıklarını her zaman hissettiğim, onlara verdiğim değeri sözcüklerin anlatamayacağı kadar kıymetli olan anneme, babama, kardeşime ve BERTİZ ailesine teşekkür ederim.

Bu çalışmayı ocak ayında doğan biricik kızım Melisa SELÇUK'a atfediyorum.

Eray SELÇUK

Şubat, 2023

İÇİNDEKİLER

Sayfa

ETİK İLKELERE UYGUNLUK BİLDİRİMİ ... iii
ÖZET ... iv
ABSTRACT ... vi
ÖNSÖZ ... viii
İÇİNDEKİLER ... ix
TABLOLAR DİZİNİ ... xii
ŞEKİLLER DİZİNİ ... xiv
KISALTMALAR ... xv
BÖLÜM 1 ... 1
GİRİŞ ... 1
Problem ... 1
Amaç ... 6
Önem ... 7
Sayıltılar ... 9
Sınırlılıklar ... 9
Tanımlar ... 10
BÖLÜM 2 ... 11
KURAMSAL ÇERÇEVE VE İLGİLİ ARAŞTIRMALAR ... 11
Klasik Test Kuramı (KTK) ... 11
Madde Tepki Kuramı (MTK) ... 13
Klasik MTK Parametreleri Kestirim Yöntemleri ... 19
Klasik MTK Yetenek Parametreleri Kestirim Yöntemleri ... 20
Klasik MTK Madde Parametreleri Kestirim Yöntemleri ... 23
Klasik MTK'da Yetenek ve Madde Parametrelerinin Birlikte Kestirim Yöntemleri ... 24
Klasik En Çok Olabilirlik Kestirimi (Joint MLE) ... 24
Koşullu En Çok Olabilirlik Kestirimi (Conditional MLE). ... 26
Marjinal En Çok Olabilirlik Kestirimi (Marginal MLE). ... 28
Bayesçi MTK Parametreleri Kestirim Yöntemleri ... 31
Bayesçi MTK Yetenek Parametreleri Kestirim Yöntemleri ... 31
Bayesçi MTK Yetenek ve Madde Parametrelerinin Birlikte Kestirim Yöntemleri 32
Markov Zinciri ve Monte Carlo Simülasyonu ve Gibbs Örneklemesi (Markov Chain Monte Carlo and Gibbs Sampling-MCMC) ... 36
İlgili Araştırmalar ... 37
BÖLÜM 3 ... 45
YÖNTEM ... 45
Araştırmanın Modeli ... 45
Verilerin Üretilmesi ... 46
Simülasyon Koşulları ... 49

Sabit Tutulan Koşullar 50
Parametre Kestirim Yöntemleri 50
Değişimlenen Koşullar 50
Örneklem Büyüklüğü 50
Madde Sayısı (Test Uzunluğu). 50
Önsel Yetenek (Theta) Dağılımı 51
MTK Modeli 52
Verilerin Çözümlenmesi 54
Verilerin MTK Varsayımlarının İncelenmesi 54
Parametre Kestirimlerini Değerlendirme Ölçütleri 63
Simülasyonun Geçerliği 65
Simülasyonun Mantıksal Geçerliği 65
Simülasyonun İstatistiksel Geçerliği 68
BÖLÜM 4 71
BULGULAR VE YORUMLAR 71
Önsel Yetenek Dağılımı Normal Olan ve Olmayan 2 PL Modeldeki Verilerde Örneklem Büyüklüğüne, Madde Sayısına Göre Klasik (En Çok Olabilirlik) ve Bayes Yöntemleri ile Kestirilen Yetenek Parametrelerinin RMSE Değerlerinin İncelenmesi 71
Önsel Yetenek Dağılımı Normal Olan ve Olmayan 3 PL Modeldeki Verilerde Örneklem Büyüklüğüne, Madde Sayısına Göre Klasik (En Çok Olabilirlik) ve Bayes Yöntemleri ile Kestirilen Yetenek Parametrelerinin RMSE Değerlerinin İncelenmesi 77
Sonuçlar 103
Önsel Yetenek Dağılımı Normal Olan ve Olmayan 2 PL Verilerde Kestirilen Yetenek Parametrelerinin RMSE Değişim Sonuçları 103
Önsel Yetenek Dağılımı Normal Olan ve Olmayan 3 PL Verilerde Kestirilen Yetenek Parametrelerinin RMSE Değişim Sonuçları 104
Önsel Yetenek Dağılımı Normal Olan ve Olmayan 2 PL Verilerde Kestirilen Madde Parametrelerinin RMSE Değişim Sonuçları 104
Önsel Yetenek Dağılımı Normal Olan ve Olmayan 3 PL Verilerde Kestirilen Madde Parametrelerinin RMSE Değişim Sonuçları 105
Öneriler 106
Uygulayıcılara Öneriler 106
Araştırmacılara Öneriler 107
KAYNAKLAR 109
EKLER 121
EK 1. Verilerin Üretilmesinde Kullanılan Örnek R Kodları 122
EK 2. Verilerin En Çok Olabilirlik Yöntemle Kestiriminde Kullanılan Örnek R Kodları 123
EK 3. Verilerin Bayesçi Yöntemle Kestiriminde Kullanılan Örnek R kodları 124
EK 4. Önsel Yetenek Dağılımı Normal Olan ve Olmayan Verilerin En Çok Olabilirlik ve Bayes Yöntemlerine Göre Kestirilen Yetenek Parametrelerinin Yanlılık (Bias) ve RMSE Değerleri 125

EK 5. Önsel Yetenek Dağılımı Normal Olan ve Olmayan Verilerin En Çok Olabilirlik ve Bayes Yöntemlerine Göre Kestirilen Yetenek Parametrelerinin Yanlılık (Bias) ve RMSE Değişimleri 127
EK 6. Önsel Yetenek Dağılımı Normal Olan ve Olmayan Verilerin En Çok Olabilirlik ve Bayes Yöntemlerine Göre Kestirilen Madde Parametrelerinin Yanlılık (Bias) ve RMSE Değerleri 130
EK 7. Önsel Yetenek Dağılımı Normal Olan ve Olmayan Verilerin En Çok Olabilirlik ve Bayes Yöntemlerine Göre Kestirilen Madde Parametrelerinin Yanlılık (Bias) ve RMSE Değişimleri 133
EK 8. Önsel Yetenek Dağılımı Normal Olan ve Olmayan Verilerin En Çok Olabilirlik ve Bayes Yöntemleriyle Kestirilen Yetenek Parametreleri ve Standart Hata Değerleri 141
EK 9. Önsel Yetenek Dağılımı Normal Olan ve Olmayan Verilerin En Çok Olabilirlik ve Bayes Yöntemleriyle Kestirilen Madde Parametreleri ve Standart Hata Değerleri 143
EK 10. Etik Kurul Onayı 146
BENZERLİK BİLDİRİMİ 147
ÖZGEÇMİŞ 149

TABLOLAR DİZİNİ

Tablo **Sayfa**

Tablo 1. Araştırmanın Koşulları 53
Tablo 2. Önsel Yetenek Dağılımı Normal Olan Verilere Ait Tek Boyutluluk Analizi Sonuçları 54
Tablo 3. Önsel Yetenek Dağılımı Normal Olmayan Verilere Ait Tek Boyutluluk Analizi Sonuçları 55
Tablo 4. Önsel Yetenek Dağılımı Normal Olan Verilerin Yerel Bağımsızlığı Karşılama Oranları 57
Tablo 5.Önsel Yetenek Dağılımı Normal Olmayan Verilerin Yerel Bağımsızlığı Karşılama Oranları 58
Tablo 6. Önsel Yetenek Dağılımı Normal Olan Verilerin Model-Veri Uyumuna Ait M2 Test İstatistiği Sonuçları 60
Tablo 7. Önsel Yetenek Dağılımı Normal Olmayan Verilerin Model-Veri Uyumuna Ait M2 Test İstatistiği Sonuçları 61
Tablo 8. Önsel Yetenek Dağılımı Normal Olan Verilere Ait Çarpıklık ve Basıklık Katsayıları 65
Tablo 9. Önsel Yetenek Dağılımı Normal Olmayan Verilere Ait Çarpıklık ve Basıklık Katsayıları 66
Tablo 10. Önsel Yetenek Dağılımı Normal Olan ve Olmayan 2 PL Modeldeki Verilerde Yetenek Parametreleri RMSE Değerleri İçin Karma Model ANOVA Sonuçları 72
Tablo 11. 2 PL Modelde Yetenek Parametresi RMSE Değerlerinin Kestirim Yöntemlerine Göre İkili Karşılaştırmaları 73
Tablo 12. 2 PL Modelde Yetenek Parametresi RMSE Değerlerinin Madde Sayısına Göre İkili Karşılaştırmaları 74
Tablo 13. 2 PL Modelde Yetenek Parametresi RMSE Değerlerinin Önsel Yetenek Dağılımına Göre İkili Karşılaştırmaları 75
Tablo 14. Önsel Yetenek Dağılımı Normal Olan ve Olmayan 3 PL Modeldeki Verilerde Yetenek Parametreleri RMSE Değerleri İçin Karma Model ANOVA Sonuçları 78
Tablo 15. 3 PL Modelde Yetenek Parametresi RMSE Değerlerinin Madde Sayısına Göre İkili Karşılaştırmaları 79
Tablo 16. 3 PL Modelde Yetenek Parametresi RMSE Değerlerinin Önsel Yetenek Dağılımına Göre İkili Karşılaştırmaları 80
Tablo 17. Önsel Yetenek Dağılımı Normal Olan ve Olmayan 2 PL Modeldeki Verilerde Madde Parametreleri RMSE Değerleri İçin Karma Model ANOVA Sonuçları 83
Tablo 18. 2 PL Modelde Madde Ayırt Edicilik Parametresi RMSE Değerlerinin Kestirim Yöntemine Göre İkili Karşılaştırmaları 84

Tablo 19. 2 PL Modelde Madde Ayırt Edicilik Parametresi RMSE Değerlerinin Örneklem Büyüklüğüne Göre İkili Karşılaştırmaları 85
Tablo 20. 2 PL Modelde Madde Ayırt Edicilik Parametresi RMSE Değerlerinin Önsel Yetenek Dağılım Biçimine Göre İkili Karşılaştırmaları 87
Tablo 21. Önsel Yetenek Dağılımı Normal Olan ve Olmayan 3 PL Modeldeki Verilerde Madde Parametreleri RMSE Değerleri İçin Karma Model ANOVA Sonuçları 89
Tablo 22. 3 PL Modelde Madde Ayırt Edicilik Parametresi RMSE Değerlerinin Kestirim Yöntemine Göre İkili Karşılaştırmaları 91
Tablo 23. 3 PL Modelde Madde Ayırt Edicilik Parametresi RMSE Değerlerinin Örneklem Büyüklüğüne Göre İkili Karşılaştırmaları 92
Tablo 24. 3 PL Modelde Madde Ayırt Edicilik Parametresi RMSE Değerlerinin Önsel Yetenek Dağılım Biçimine Göre İkili Karşılaştırmaları 93
Tablo 25. 3 PL Modelde Şans Parametresi RMSE Değerlerinin Kestirim Yöntemine Göre İkili Karşılaştırmaları 95
Tablo 26. 3 PL Modelde Şans Parametresi RMSE Değerlerinin Önsel Yetenek Dağılım Biçimine Göre İkili Karşılaştırmalar 96
Tablo 27. 2 PL ve 3 PL Modeldeki Verilerde Yetenek Parametreleri RMSE Değerleri Koşullara Bağlı Özet Bilgiler 97
Tablo 28. 2 PL ve 3 PL Modeldeki Verilerde Madde Parametreleri RMSE Değerleri Koşullara Bağlı Özet Bilgiler 98

ŞEKİLLER DİZİNİ

Şekil **Sayfa**

Şekil 1. Madde Karakteristik Eğrisinin (MKE) Genel Görünümü 14
Şekil 2. Sonsal Bir Dağılımın Üretilmesi İçin Önsel Bir Dağılıma Olabilirlik Fonksiyonunun Eklenmesi 34
Şekil 3. 2 PL Modelde Yetenek Parametresi RMSE Değerlerinin Kestirim Yöntemlerine Göre Değişimi 73
Şekil 4. 2 PL Modelde Yetenek Parametresi RMSE Değerlerinin Madde Sayısına Göre Değişimi 74
Şekil 5. 2 PL Modelde Yetenek Parametresi RMSE Değerlerinin Önsel Yetenek Dağılım Biçimlerine Göre Değişimi 76
Şekil 6. 3 PL Modelde Yetenek Parametresi RMSE Değerlerinin Kestirim Yöntemlerine Göre Değişimi 79
Şekil 7. 3 PL Modelde Yetenek Parametresi RMSE Değerlerinin Önsel Yetenek Dağılım Biçimlerine Göre Değişimi 81
Şekil 8. 2 PL Modelde Madde Ayırt Edicilik Parametresi RMSE Değerlerinin Kestirim Yöntemlerine Göre Değişimi 85
Şekil 9. 2 PL Modelde Madde Ayırt Edicilik Parametresi RMSE Değerlerinin Örneklem Büyüklüğüne Göre Değişimi 86
Şekil 10. 2 PL Modelde Madde Ayırt Edicilik Parametresi RMSE Değerlerinin Önsel Yetenek Dağılım Biçimlerine Göre Değişimi 87
Şekil 11. 3 PL Modelde Madde Ayırt Edicilik Parametresi RMSE Değerlerinin Kestirim Yöntemlerine Göre Değişimi 91
Şekil 12. 3 PL Modelde Madde Ayırt Edicilik Parametresi RMSE Değerlerinin Örneklem Büyüklüğüne Göre Değişimi 92
Şekil 13. 3 PL Modelde Madde Ayırt Edicilik Parametresi RMSE Değerlerinin Önsel Yetenek Dağılım Biçimlerine Göre Değişimi 94
Şekil 14. 3 PL Modelde Şans Parametresi RMSE Değerlerinin Kestirim Yöntemlerine Göre Değişimi 95
Şekil 15. 3 PL Modelde Şans Parametresi RMSE Değerlerinin Önsel Yetenek Dağılım Biçimlerine Göre Değişimi 96

KISALTMALAR

KTK	Klasik Test Kuramı
MTK	Madde Tepki Kuramı
MKE	Madde Karakteristik Eğrisi
MAP	Modal A Posterior (Maksimum Sonsal)
EAP	Expected A Posterior (Beklenen Sonsal)
MCMC	Markov Chain Monte Carlo (Markov Zinciri Monte Karlo)
ML	Maximum Likelihood (En Çok Olabilirlik)
Joint ML	Klasik En Çok Olabilirlik
Conditional ML	Koşullu En Çok Olabilirlik
Marginal ML	Marjinal En Çok Olabilirlik
SE	Standard Error (Standart Hata)
RMSE	Root Mean Square Error (Ortalama Hata Kareleri Karekökü)
1 PL	Bir Parametreli Lojistik Model
2 PL	İki Parametreli Lojistik Model
3 PL	Üç Parametreli Lojistik Model
Θ	Yetenek (ilgili özellik) parametresi
a_i	Madde ayırt edicilik parametresi
b_i	Madde güçlük parametresi
c_i	Şans (düşük asimptot) parametresi
Θ_{SE}	Yetenek parametresi kestirim hatası
a_{SE}	Madde ayırt edicilik parametresi kestirim hatası
b_{SE}	Madde güçlük parametresi kestirim hatası
c_{SE}	Şans (düşük asimptot) parametresi kestirim hatası

BÖLÜM 1

GİRİŞ

Bu bölümde araştırmanın amacına, problemine, alt problemlerine, önemine, sınırlılıklarına ve araştırma boyunca kullanılan kısaltmalara yer verilmiştir.

Problem

Psikolojik yapılar, soyut ve dolaylı ölçülebilir bir durumda olmasından dolayı ölçme işlemi için bir araç geliştirmeyi zorunlu kılar. Ölçme aracı geliştirme işlemi psikometrinin ilgilendiği konular arasındadır. Psikometri, insan davranışlarının temelindeki psikolojik yapıların ölçülebilmesiyle ilgilenmektedir. Bireylerin sahip olduğu örtük özelliklerin ortaya çıkarılması, açıklanması, analiz edilebilmesi ve yorumlanması için çeşitli kuramlar geliştirilmiştir. Bu kuramlara bağlı olarak oluşturulan ölçme araçlarından biri de testlerdir.

Testler, eğitim ve psikolojide önemli kararların verilmesi, öğrenme sürecindeki eksiklerin belirlenmesi, öğrencilerin güçlü ve zayıf yönlerinin ortaya çıkarılması, başarı durumunun ölçülmesi ya da seçme ve yerleştirme gibi amaçlarla kullanılmaktadır (Baykul, 2015). Aynı zamanda testler, bireyin özelliklerini standart koşullar altında belirlemeye yarayan ölçme araçlarıdır (Anastasi, 1988). Testlerin bireylerle ilgili önemli ve kritik kararlar almada kullanılması için bazı psikometrik özellikler taşıması gerekmektedir. Çeşitli amaçlarla ölçme işlemini gerçekleştiren testler doğru bir şekilde uygulandıklarında elde edilen sonuçlar, eğitim programlarının kazanımlarına ulaşmada ve bireyler hakkında kararlar sağlamada önemli bilgiler ortaya koymaktadır (Kolen ve Brennan, 2014).

Test geliştirme art arda gelen birtakım faaliyetlerden oluşan bir süreçtir (Thorndike, 1982). Bu test geliştirme süreçleri, hatayı en aza indirgemeyi amaçlayan çeşitli kuramlar çerçevesinde yürütülmektedir. Bu kapsamda test kuramları, ölçme işleminin güvenirlik ve geçerliğini sağlamak üzere çeşitli yöntem ve modeller kullanırlar. Test kuramları, gözlenen değişkenleri, gizil değişkenlere bağlayan genel bir bakıştır. Test kuramlarının amacı, tanımlanan psikolojik yapılara ilişkin ölçme sonuçlarının ölçme hatalarından ne derece etkilendiğini belirlemek ve bu hataları en aza indirecek yöntemleri

ve uygulamaları keşfetmektir. Test teorisinin başka bir amacı da uzmanların test kullanımında ve yapımında standart uygulamaların altında yatan mantık ve matematiksel modellerin farkında olmasına yardımcı olmaktır (Crocker ve Algina, 1986).

Test geliştirme süreçleri ve bunların psikometrik nitelikleriyle ilgili sorunların ele alınmasını konu edinen psikometri biliminin tarihsel sürecinde yaygın iki ölçme kuramı kullanılmaktadır. Bunlar, ilk geliştirilen Klasik Test Kuramı (KTK) ve giderek kullanımı yaygınlaşan Örtük Özellikler Kuramı olarak da adlandırılan Madde Tepki Kuramı (MTK)'dır.

Bu kuramlardan ilki olan ve gerçek puan teorisi olarak da tanımlanan KTK, ölçme sonuçlarını anlamamıza, değiştirmemize ve yorumlamamıza izin veren bir ölçüm çerçevesidir. KTK'nın temelinde, her ölçümün hata içerdiği ve tüm gözlemlerin kusurlu olması varsayımı yatar (Desjardins ve Bulut, 2018). KTK, çeşitli ölçek ve test geliştirme gibi konularda ölçme ve değerlendirme uzmanlarına uzun yıllarca hizmet etmesine rağmen, test eşitleme, yanlılık analizleri, madde ve birey parametrelerinin gruptan bağımsız kestirilmesi gibi konularda tatmin edici yanıtlar bulmakta kısıtlı kalmıştır (Hambleton ve Swaminathan, 1985; Hambleton, Swaminathan ve Rogers, 1991; Hambleton ve Jones, 1993). Bu tür sorunlara çözüm olarak 1930'lu yıllardan başlayan, bilgisayar teknolojilerinin gelişimiyle hız kazanan ve günümüzde hala ölçme ve değerlendirme uzmanları için yeni çözüm önerileri sunan Madde Tepki Kuramı (MTK) geliştirilmiştir. Ölçme problemlerine getirdiği pratik çözüm önerileri bakımından Madde Tepki Kuramı günümüzde araştırmacılar tarafından daha çok tercih edilmektedir (Embretson ve Reise, 2000).

MTK'ya göre yetenek ya da gizil özellik, test maddeleri üzerinde gösterilen performanstır. Aslında MTK, gizil değişkenler ile bu değişkenlerden elde edilen bulgular arasındaki uyumluluğu tesis etmek için izlenecek yolu gösteren bir model olarak tanımlanmaktadır. MTK, kuramsal bir teori olarak görülmemelidir. Çünkü bu kuram, bir kişinin bir maddeye neden belirli bir yanıtı verdiğini ya da yanıtlamaya nasıl karar verdiğini açıklar bir özellikte değildir. MTK, daha çok istatistiksel kestirimlere dayalı bir modeldir. MTK, gözlemlenen yanıtların kestirilebilmesi için bireylerin ve maddelerin gizil karakteristik özelliklerini kullanmaktadır (Hambleton, Swaminathan ve Rogers, 1991; Hambleton ve Jones, 1993). Başka bir deyişle, MTK, incelenen madde ve test performansının testteki maddeler tarafından ölçülen yeteneklerle nasıl ilişkili olduğu hakkında genel bir istatistiksel teoridir (Hambleton and Jones, 1993).

Madde tepki kuramı modellerinin sahip oldukları avantajlar, yalnızca model ve test verileri arasındaki uyum tatmin edici olduğunda elde edilebilir. MTK, KTK'ya oranla birçok problem durumuna çözüm bulmada avantajlı görünse de (Embretson ve Reise, 2000), temelinde karşılanmasını gerektiren birtakım sayıltıları bulunmaktadır. Bunlar; tek boyutluluk (unidimensionality), yerel bağımsızlık (local independence) ve model-veri uyumu (model-data fit)'dur (Baker, 2001; Baker ve Kim, 2004; De Ayala, 2009; Hambleton ve Swaminathan, 1985; Hambleton, Swaminathan ve Rogers, 1991; Hambleton ve Jones, 1993). MTK, madde ve birey parametrelerini kestirirken temel aldığı varsayımların yanı sıra çeşitli uygulamaları da beraberinde getirir. Bu varsayımların ve uygulamaların sağladığı avantajlarla birlikte bazı sınırlılıklarda mevcuttur.

MTK'nın sınırlılıklarından biri geniş örneklem büyüklüklerine ihtiyaç duymasıdır. Alanyazında birçok çalışma yapılmasına rağmen üzerinde uzlaşılmış standart bir örneklem büyüklüğü değeri yoktur (De Ayala, 2009; Reise ve Yu, 1990). Bununla ilgili olarak DeMars (2010), Rasch veya 1 PL modellerle 100 ya da 200; 2 PL modelle 500; 3 PL modelle 2000 kadar büyük örneklemlerle çalışılması gerektiğini belirtmektedir. Aynı durum test uzunluğu (madde sayısı) konusunda da çeşitlilik göstermektedir. 3 PL model ile çalışılması durumunda 40 ila 60 arasında madde kullanılması gerektiği vurgulanmıştır (Hulin, Lissak ve Drasgow, 1982).

Bu konuda yapılan ilk çalışmada, Lord (1968), VSAT verileri üzerinden joint maximum likelihood yöntemini kullanarak 3 PL model üzerinde parametre kestirimleri gerçekleştirmiştir. Araştırmasının sonucunda örneklem büyüklüğünün 1000'e ve madde sayısının 50'ye ulaşmadığı durumlarda örnekleme hatasının yüksek değerler aldığını belirtmiştir.

Alanyazındaki diğer çalışmalar incelendiğinde MTK'nın parametre kestirimlerinde test uzunluğu ve örneklem büyüklüğü bakımından artış gösterdikçe standart hata oranın düştüğü ve model veri uyumunun daha iyi sağlandığı rapor edilmiştir. Bu çalışmaların birçoğu Lord (1968)'un araştırmasına atıf yapmış ve kestirim yöntemi olarak joint maximum likelihood yöntemini kullanmışlardır. Bunun yanı sıra hem Klasik hem de Bayes kestiriminin kullanıldığı çalışmalarda ise üretilen veri setlerinin önsel yetenek dağılım biçimleriyle ilgili bir araştırmaya rastlanmamıştır (Swaminathan ve Gifford, 1986; Harwell ve Janosky, 1991; Gao ve Chen, 2005).

Parametre kestirimlerinin standart hata oranının miktarı, örneklem büyüklüğüne ve test uzunluğuna bağlı olmakla birlikte kestirim yöntemleri de bu standart hata miktarını

etkilemektedir. MTK'da iki farklı yaklaşımı temel alan parametre kestirim yöntemleri mevcuttur. Bunlardan ilki frekansçı yaklaşımı kullanan en çok olabilirlik (maximum likelihood) yöntemleri, ikincisi Bayesçi yaklaşımı kullanan yöntemlerdir.

Parametre kestirim yöntemlerinin etkili bir şekilde çalışabildiği bazı varsayımlar mevcuttur. Veriler açısından bu varsayımların dikkate alınmaması ve ihmal edilmesi durumunda kestirimlerdeki standart hata oranları artış göstermektedir (Hambleton ve Swaminathan, 1985; Hambleton and Jones, 1993).

MTK çerçevesinde, madde (güçlülük, ayırt edicilik, şans) ve yetenek (birey) parametrelerinin kestiriminde altı farklı yöntem mevcuttur. Bunlar; (1) En Çok Olabilirlik Yöntemi (Maximum Likelihood [ML], Lord, 1980), (2) Bayesçi-Moda Göre Kestirim (Bayesian Modal Estimation [BME], Samejima, 1980) ya da diğer adıyla Moda Göre Sonsal Dağılım (Modal a Posteriori [MAP], Samejia, 1969), (3) Owen'ın Ardışık Bayesçi Yaklaşımı (Owen's Sequential Bayesian [OSB], Owen, 1975), (4) Bayesçi Beklenen Sonsal Dağılım (Expected a Posteriori [EAP] Bock ve Mislevy, 1982), (5) Marjinal En Çok Olabilirlik (Marginal Maximum Likelihood [MML], Bock ve Aitkin, 1981), (6) Ağırlıklandırılmış En Çok Olabilirlik (Weight Maximum Likelihood [WML], Warm, 1989)'dır. Bu yöntemlerin birçoğu temelinde en çok olabilirlik fonksiyonun (maximum likelihood) hesaplanmasına bağlı olarak şekillenmektedir. En çok olabilirlik fonksiyonu, gözlenen veriler üzerinden madde ve yetenek parametrelerini maksimize eden olasılık değerlerinin kestirilmesiyle bulunmaktadır. Bu kestirim yöntemleri, iteratif (ötelemeli/yakınsamalı) bir süreçle çözüm gerçekleştirmektedir. En çok olabilirlik fonksiyonlarının genel anlamdaki en önemli sınırlılığı, bir testten tam puan ya da sıfır puan alan bireylerin yetenek parametrelerinin kestiriminin ya da herkes tarafından doğru veya yanlış yapılmış maddelerin parametre kestirimlerinin yapılamamasıdır (Lord, 1983; Samejima, 1993a, 1993b). Bu durumun çözümüne ilişkin çeşitli öneriler bulunmaya çalışılsa da (marjinal en çok olabilirlik kestirimi gibi) verilerin dağılım özelliklerinden ya da biçimlerinden dolayı istenilen titiz (accurate) sonuçlara tam olarak ulaşılamamıştır.

Normallik varsayımı MTK'nın sayıltıları arasında yer almaz fakat dikkate alınması gereken bir durumdur (DeMars, 2010). Normallik varsayımının dikkate alınmadığı durumlarda MTK modellerinin gücünün düştüğü görülmektedir. MTK modelleri için normallik varsayımının gerekli olmadığı belirtilse de bu varsayımın sağlanamaması durumu, yetenek kestiriminde sorunlara yol açabileceği ifade edilmektedir. En çok olabilirlik kestirim yöntemlerinin kullanıldığı modellerde ilgilenilen özelliğin (Θ) evrende normal dağıldığı varsayılır. Verilerin dağılımı konusunda dikkat

edilmesi gereken nokta, parametrelerin doğru kestirilebilmesi için örneklemin ilgilenilen özellik (Ɵ) bakımından geniş bir aralığı kapsaması gerektiğidir. Örneğin, düşük ve yüksek Ɵ değerlerinin fazla temsil edilmesi dağılımın çarpıklığını arttırmakta; madde ve yetenek parametreleri kestirimlerinin standart hatalarının yükselmesine neden olmaktadır. Aynı şekilde, Ɵ düzeyi madde güçlük parametresi (b_i) civarında olan birey sayısı arttıkça b_i kestiriminin standart hatası azalmaktadır. Madde ayırt edicilik parametresi (a_i) ise, testi alan bireylerin Ɵ düzeylerinin ranjı geniş olduğunda en iyi şekilde kestirilmektedir (Baker & Kim, 2004; DeMars, 2010). Ayrıca, madde parametre kestirimlerinin önsel yetenek dağılımından etkilenmediği, ancak özellikle sağa ya da sola çok çarpık verilerden elde edilen madde parametre kestirimlerinin etkilendiği belirtilmektedir (Xu ve Jia, 2011). Buna göre, önsel yetenek dağılım biçimi (normal, sağa çarpık, sola çarpık, sivri, basık), madde ve yetenek parametrelerinin kestiriminde ve bu kestirimlerin hata miktarlarının belirlenmesinde etkili olmaktadır. Dağılımın çarpıklaştığı durumlarda ya da en çok olabilirlik yöntemlerinin yukarıda belirtilen genel anlamdaki sorunuyla karşılaşıldığında kestirimlerin titiz (kestirim hatası düşük) bir şekilde yapılabilmesi için "Bayes Kestirim Yöntemleri" önerilmektedir (Bock ve Mislevy, 1982; De Ayala, 2009; Hambleton ve Swaminathan, 1985; Hambleton, Swaminathan ve Rogers, 1991; Hambleton ve Jones, 1993).

Bayes kestirim yaklaşımlarının temeli Bayes teoremine dayanmaktadır. Bayes yaklaşımı, önsel bilgilerin (prior) çalışmaya dahil edilmesi ve olabilirlik (likelihood) kestirim yönteminin kullanılmasıyla sonsal (posterior) bilginin elde edilmesi olarak açıklanabilmektedir. Bu süreçte elde edilen sonsal dağılım sonuçları Bayes hipotezi için gerekli olan bileşenleri sağlamaktadır. Temel fikir, yetenek parametreleri hakkında daha önceden sahip olduğumuz bilgileri de kullanarak olabilirlik fonksiyonunu yeniden düzenlemektir (Kaplan ve Depaoli, 2012).

Gao ve Chen (2005) tarafından yapılan araştırmada 3 PL model üzerinden geniş ölçekli bir simülasyon çalışması gerçekleştirilmiştir. Çalışmalarında test uzunluğu olarak 10, 20 ve 60 maddelik, örneklem büyüklüğü olarak 100, 200 ve 500 kişilik, tek-biçimli (uniform) dağılan veri setleri kullanmışlardır. Bu veriler üzerinden marjinal en çok olabilirlik yöntemi (MML) ile Bayes kestirim yöntemlerini karşılaştırmışlardır. Araştırmalarının sonucunda marjinal en çok olabilirlik yönteminin küçük örneklemlerde gerçek madde parametre değerlerinin dışında kestirim yapma eğiliminde olduğu sonucuna varmışlardır. Ayrıca, örneklem büyüklüğünün 100 kişiye kadar düştüğü durumlarda, Bayes kestirimlerinin marjinal en çok olabilirlik kestirimlerine göre daha

titiz sonuçlar verdiğini belirtmişlerdir. Kestirimlerin standart hatası (RMSE) bakımından Bayes kestirimlerinin sonuçlarının daha kabul edilebilir düzeyde olduğunu, ancak test uzunluğu ve örneklem büyüklüğünün artması durumunda marjinal en çok olabilirlik yöntemiyle kestirim sonuçlarının da hata miktarlarının düzelme eğiliminde olduğunu vurgulamışlardır. Buna göre Bayesçi MTK kestirim yöntemleri kullanmanın klasik MTK kestirim yöntemleri kullanımına göre avantajlar sağladığı görülmektedir.

Bayes yaklaşımının özünde, herhangi bir veri elde etmeden önce bireyin bir özellik bakımından dağılımdaki yeri hakkında bilgi sahibi olunmasına dayanır. Bu dağılıma önsel (prior) dağılım denir. Önsel dağılım kullanarak parametre kestirimlerinin bu şekilde belirli aralıklar içinde kısıtlamak MTK'nın Bayes kestirimleri için önemli bir noktadır (Gao ve Chen, 2005).

Parametre kestiriminde Bayes yönteminin kullanılması, en çok olabilirlik yöntemlerine göre bazı durumlarda avantajlar sağlayabilir. En çok olabilirlik yöntemleri genel olarak küçük örneklemlerde, madde sayısının az olduğu testlerde ve özellikle örnekleme ait verilerin çarpık dağılım gösterdiği durumlarda madde ve yetenek parametrelerini titiz (standart hata oranı düşük bir şekilde) kestirememektedir. Aynı şekilde MTK modelindeki (2 PL ve 3 PL modellerdeki gibi) parametre sayısının artması da bu kestirimlerdeki standart hata miktarını arttırmaktadır. Alanyazında bu tür sorunlar için Bayes yaklaşımıyla parametre kestirimi yapılması önerilmektedir. Ancak, bu yaklaşımın farklı önsel yetenek dağılım biçimlerinde ortaya koyacağı parametre kestirimlerinin titizliği bilinmemektedir.

Buna göre MTK'ya dayalı parametre kestirimi yapılırken veriler, önsel yetenek dağılım biçimi, örneklem büyüklüğü, madde sayısı ve model parametre sayısı bakımından değişimlendiğinde kestirimlerin RMSE değerleri arasında klasik ve Bayesçi MTK'ya göre bir farklılaşma olup olmayacağı bu araştırmanın problemini oluşturmaktadır.

Amaç

Bu araştırmanın amacı, klasik (joint maximum likelihood) ve Bayesçi yaklaşımlara göre Madde Tepki Kuramı (MTK) yetenek ve madde parametrelerinin kestirimlerinin farklı simülasyon koşularında karşılaştırılmasıdır. Bu amaçla; önsel yetenek (theta) dağılım biçiminin, örneklem büyüklüğünün, madde sayısının ve modelin kestirilen parametre sayısının değişimlenmesine bağlı olarak; klasik en çok olabilirlik ve

Bayes yöntemine göre kestirilen yetenek ve madde parametreleri ile bu parametrelerin kestirimlerinin hata değerlerindeki (RMSE) farklılaşmalar incelenmiştir.

Buna göre simülasyon koşullarına göre üretilen veriler üzerinden aşağıdaki sorulara yanıtlar aranmıştır:

1. Önsel yetenek dağılımı normal ve normal olmayan (sola çarpık, sağa çarpık, sivri ve basık) 2 PL modele göre üretilmiş veri setlerinde farklı örneklem büyüklüğü (N=100, 500 ve 1000) ve madde sayısı (i= 10, 20 ve 40) kullanıldığında yetenek parametrelerine yönelik Θ_{RMSE} değerleri arasında klasik ve Bayes kestirim yöntemlerine göre manidar bir farklılık var mıdır?

2. Önsel yetenek dağılımı normal ve normal olmayan (sola çarpık, sağa çarpık, sivri ve basık) 3 PL modele göre üretilmiş veri setlerinde farklı örneklem büyüklüğü (N=100, 500 ve 1000) ve madde sayısı (i= 10, 20 ve 40) kullanıldığında yetenek parametrelerine yönelik Θ_{RMSE} değerleri arasında klasik ve Bayes kestirim yöntemlerine göre manidar bir farklılık var mıdır?

3. Önsel yetenek dağılımı normal ve normal olmayan (sola çarpık, sağa çarpık, sivri ve basık) 2 PL modele göre üretilmiş veri setlerinde farklı örneklem büyüklüğü (N=100, 500 ve 1000) ve madde sayısı (i= 10, 20 ve 40) kullanıldığında madde parametrelerine yönelik a_{RMSE}, b_{RMSE} ve c_{RMSE} değerleri arasında klasik ve Bayes kestirim yöntemlerine göre manidar bir farklılık var mıdır?

4. Önsel yetenek dağılımı normal ve normal olmayan (sola çarpık, sağa çarpık, sivri ve basık) 3 PL modele göre üretilmiş veri setlerinde farklı örneklem büyüklüğü (N=100, 500 ve 1000) ve madde sayısı (i= 10, 20 ve 40) kullanıldığında madde parametrelerine yönelik a_{RMSE}, b_{RMSE} ve c_{RMSE} değerleri arasında klasik ve Bayes kestirim yöntemlerine göre manidar bir farklılık var mıdır?

Önem

MTK, araştırmacılara madde ve birey parametrelerinin birbirinden ve gruba bağlı olmaksızın kestirme imkânı sunmuştur. Yani, maddeler MTK ile

ölçeklendirilebilmektedir. Bu ölçekleme işleminin yapılabilmesi için MTK'nın avantajlarından biri olan parametre değişmezliğinin sağlanması gerekmektedir. Bunun için alanyazında örneklem büyüklüğünün ne kadar olması gerektiği konusundan birçok farklı görüş bulunmaktadır. Bu görüşlerin ortak yanı, modelde kestirilmesi gereken parametre sayısı arttıkça test uzunluğunun (madde sayısının) (Reise ve Yu, 1990) ve örneklem büyüklüğünün (Hambleton, 1989; Orlando, 2004) de artması gerekliliğidir. Amaç, parametre kestirimlerinin hata değerlerini en aza indirmektir. Ancak, önsel yetenek dağılım biçiminin değişmesi bu kestirimlerin hata miktarını etkileyecektir. Alanyazında, önsel yetenek dağılım biçimine bağlı olarak yapılmış parametre kestirim yöntemlerine değinen çalışmalar çok az sayıdadır. Özellikle bu çalışmanın amacını oluşturan önsel yetenek dağılım biçiminin farklılaşmasına bağlı olarak parametre kestiriminde kullanılan klasik ve Bayes yaklaşımının farklı koşullar altında incelendiği bir çalışmaya rastlanmamıştır.

Kestirim yöntemleri, testi alan bireylerin yeteneklerinin ve madde parametrelerinin dağılım biçimlerinden etkilenmektedir. Ayrıca birçok özelliğin (Ɵ) evrende normal dağıldığı varsayılmaktadır. Bu varsayım MTK'nın güçlü yanlarını ortaya çıkarmakta ve parametre kestirimlerini etkilemektedir. Bu nedenle, çarpık dağılımlar parametre kestirimlerinde bazı sorunlara yol açmaktadırlar. Çünkü parametrelerin sağlıklı kestirimi, belirli Ɵ düzeylerinde varyansın yeteri kadar geniş olmasına bağlıdır. Eğer belirtilen dağılım varsayımları sağlanamazsa MTK'nın en çok olabilirlik yöntemlerine dayalı olan parametre kestirimlerinin titizliği tartışılmaktadır. Sonuç olarak, bu araştırma normal dağılımın dışındaki farklı önsel yetenek dağılım biçimlerinde (sola ve sağa çarpık, basık ve sivri) kestirilen parametrelerin yüksek hata değerlerine sahip olduklarını kabul ederek bu hata miktarının en aza indirilmesi konusunda alternatif bir kestirim yöntemi önermesi ve Bayesçi MTK yaklaşımının klasik MTK yaklaşımına göre parametre kestirimlerinde avantajlar sağlaması açısından önemlidir.

Araştırma sonucunda, farklı biçimlerdeki önsel yetenek dağılım biçimleri için örneklem büyüklüğünün ve madde sayısının yeterli olmadığı durumlarda Bayes kestirim yöntemlerinin kullanılmasının parametre (yetenek ve madde) kestirimlerinde düşük RMSE değerleri elde edilmesini öngörmektedir. Bu açıdan, araştırmanın MTK'da kullanılan parametre kestirimlerine yönelik farklı bir bakış açısı getireceği ve ilgili alanyazına katkı sağlayacağı düşünülmektedir.

Sayıltılar

1. Bayesçi ve klasik MTK parametre kestirimleri için kullanılan önsel yetenek dağılımın ilgilenilen özellik bakımından evrende normal dağıldığı varsayılmıştır.

Sınırlılıklar

Bu bölümdeki 1., 2. ve 3. maddeler araştırmanın sınırlılıklarıdır. Ancak daha sonra verilen 4., 5., 6. ve 7. maddeler ise araştırmanın sınırlarıdır ve araştırmacı tarafından oluşturulmuştur.

1. En çok olabilirlik yöntemine göre parametre kestiriminde kullanılan "irtplay" paketi, yetenek parametreleri kestirimlerini -7 ile +7 arasında yapabilmektedir. Bu nedenle yetenek parametreleri belirtilen değerler arasında sınırlandırılmıştır.
2. Araştırmada üretilen verilerde madde parametrelerinin dağılımı alanyazında belirtilen dağılım biçimleriyle sınırlandırılmıştır.
3. Araştırmada Bayes MCMC örneklemi 3000, yakma (burning) işlemi 1000 ile sınırlandırılmıştır.
4. Bu araştırma simülasyon koşullarıyla (örneklem büyüklüğü, madde sayısı, önsel yetenek dağılım biçimi, modelin kestirilen parametre sayısı ve kestirim yöntemleriyle) sınırlandırılmıştır.
5. Araştırma problem durumunda belirtilen beş koşulla değişimlenerek belirtilen modellerle sınırlandırılmıştır.
6. ML, MCMC, RMSE, Bias, EAP, MAP gibi birçok kısaltmanın Türkçe alanyazındaki karşılıklarının farklı şekillerde kullanılması nedeniyle araştırma boyunca İngilizce karşılılıkları kullanılmıştır. Bu kavramların Türkçe karşılıklarının kullanılmaması bir sınırlılık olarak kabul edilmiştir.
7. Alanyazında parametre kestirimlerinin doğruluğu için kullanılan "accurate" sözcüğü kestirimlerin titizliği anlamında sınırlandırılmıştır.

Tanımlar

Bu bölümde araştırma kapsamında kullanılan kavramsal yapılar hakkında tanımlamalara yer verilmiştir:

Normal Dağılım: Aritmetik ortalaması (μ) 0, standart sapması (δ) 1, çarpıklık ($Z_ç$) ve basıklık (Z_b) katsayıları -1.00 ile +1.00 arasında olan dağılımdır.

Sola Çarpık Dağılım: Aritmetik ortalamanın (μ), ortanca değerinden küçük olduğu, çarpıklık katsayısının ($Z_ç$) -1.5 ve daha küçük olduğu dağılımdır.

Sağa Çarpık Dağılım: Aritmetik ortalamanın (μ), ortanca değerinden büyük olduğu, çarpıklık katsayısının ($Z_ç$) +1.5 ve daha yüksek olduğu dağılımdır.

Sivri Dağılım: Standart sapması (δ) 1'den küçük, basıklık değeri (Z_b) +1.5 ve daha büyük olan dağılımlardır.

Basık Dağılım: Standart sapması (δ) 1'den büyük, basıklık değeri (Z_b) -1.5 ve daha küçük olan dağılımlardır.

BÖLÜM 2

KURAMSAL ÇERÇEVE VE İLGİLİ ARAŞTIRMALAR

Bu bölümde, Klasik Test Kuramı (KTK), Madde Tepki Kuramı (MTK) ve Parametre Kestirim Yöntemleri ile ilgili kuramsal açıklamalara ve ilgili araştırmalara yer verilmiştir.

Klasik Test Kuramı (KTK)

Klasik Test Kuramı (KTK)'nın temel ilkeleri İngiliz psikolog Charles Spearman tarafından 1905 yılında ortaya atılmıştır. Bu kuramda hata içeren test sonuçlarıyla gerçek puanlar arasındaki ilişkiyi açıklamaya çalışmıştır. Bu kuramın özünde gözlenen test puanları kuramsal olarak gerçek puan ve tesadüfi hata şeklinde iki bileşenden oluşmaktadır (Baykul, 2015).

KTK, doğru yanıtlardan elde edilen puan modeline dayanmaktadır. Bu model bireyin gözlemlenen puanını, onun gizil değişken üzerindeki bulunduğu konum ile ilişkilendirmektedir.

KTK, bireyin gözlenen puanı içerisinde gerçek ve hata puanları bulunduğunu kabul eder. Bir başka deyişle; gerçek puanı, ölçme sonuçlarından elde edilen puanla kestirmeye çalışan bir kuramdır. Bu kuramda ölçmedeki hataların seçkisiz olduğu ve bu hatalara ait puanların, bu hataların her ölçme için gerçek puanlardan ve bir başka ölçmedeki hata puanlarından bağımsız olduğu kabul edilir. Bu kuramın temel denklemi aşağıdaki gibidir (Crocker ve Algina, 1986, Gulliksen, 1950):

$$\mathbf{X = T + E}$$

X: Gözlenen Puan

T: Gerçek Puan

E: Hata Puanı

Formel küme ile ölçülmek istenen yapı arasındaki durumu açıklamaya yönelik bu eşitlik regresif modeldir. Buna göre bu doğrusal bağıntı, gözlenen puanlar (X) ve gerçek

puanlar (T) arasındaki doğrusal bir eşitlikle modellenmektedir. Bir başka ifadeyle KTK, $f(T \mid X)$ gibi doğrusal regresif fonksiyonun kestiriminden oluşmaktadır. $f(T \mid X)$ fonksiyonu ise gözlenen veriler (X) üzerine kurulmaktadır (Lord ve Novick, 1968).

Buna göre yukarıdaki eşitlikte tanımlanan gerçek puan modeline ilişkin temel sayıltılar aşağıdaki gibidir (Gulliksen, 1950; Lord, 1980; Crocker ve Algina, 1986; Desjardins ve Bulut, 2018):

1. Hataların evrendeki dağılımının ortalaması sıfırdır, $\varepsilon(E)=0$
2. Gözlenen puanın beklenen değeri gerçek puandır, $\varepsilon(X)=T$.
3. Gerçek ve hata puanı arasındaki kovaryans ya da korelasyon sıfırdır, $Cov(T, E)=0$ veya $\rho(T, E)=0$.
4. Farklı hata puanları arasındaki kovaryans ya da korelasyon sıfırdır, $Cov(E_1, E_2)=0$ veya $\rho(E_1, E_2)=0$.

KTK uygulamalarının diğer ölçme kuramlarına göre süreç ve işlemler bakımından daha kolay olması dolayısıyla yıllarca yaygın olarak kullanılmıştır. Eğitim ve psikolojinin bilim olarak kabul edilmesinde bu kuramın etkisi büyüktür. Aynı zamanda birçok eleştiriye de maruz kalmıştır. Yapılan eleştirilerin temelinde KTK'nın çok fazla sayıltıya sahip olması ve bu sayıltıların çoğu zaman karşılanamaması yatmaktadır.

KTK'nın en büyük dezavantajlarından biri bireylerin özellikleri ile test maddelerinin birbirinden ayrılmaz oluşudur. Uygulanan test zor olduğunda birey ilgilenilen yetenek bakımından düşük düzeyde, uygulanan test kolay olduğunda ise yüksek düzeyde görülmektedir. Testi alan bireylerin ölçülen özellik bakımından değişmesi test ve madde istatistiklerini değiştirmektedir. Buna bağlı olarak aynı yapıyı ölçen bir testi alan bireyin özellikleri aldığı testin içeriği farklılaştıkça değişiklik gösterecektir. Bu nedenle, farklı testler alan bireylerin ve farklı bireyler kullanılarak elde edilen madde istatistiklerinin karşılaştırılması oldukça zor olacaktır (Hambleton, Swaminathan ve Rogers, 1991; Hambleton ve Jones, 1993). Bunlara ek olarak, KTK'da testi alan tüm bireylerin yetenek ranjı için tek bir standart hata değerinin hesaplanması kuramın sınırlılıkları arasında görülmektedir. Buna karşın, Madde Tepki Kuramında (MTK) birey ve madde parametreleri değişmezliğe sahiptir (De Ayala, 2009).

Madde Tepki Kuramı (MTK)

KTK'nın temelinde eleştirilen ölçme problemlerine getirdiği çözüm önerileri bakımından Madde Tepki Kuramı günümüzde araştırmacılar tarafından daha çok kullanılmaktadır (Embretson ve Reise, 2000). Madde Tepki Kuramı (MTK), büyük ölçekli değerlendirme (Hambleton ve Swaminathan, 1985; Lord, 1980), insan kaynakları ve pazar araştırmaları (Drasgow ve Hulin, 1990) ve tıp alanındaki testler (Jenkinson, Fitzpatrick, Garratt, Peto ve Stewart-Brown, 2001; Hambleton, 2002) de dahil olmak üzere çeşitli ölçme problemlerine getirdiği çözümlerle güçlü bir psikometrik yaklaşımdır.

1950'lerde, Dr. F. Lord, bireylerin gözlenen ve gerçek puanlarının yetenek puanlarıyla eş değer olmadığını belirtmiştir. Gözlenen ve gerçek puanların testlere bağlı olduğunu ancak yeteneğin testlerden bağımsız olduğunu belirtmiştir. Kişiler, zor testlerden daha düşük ve kolay testlerden daha yüksek gerçek puanlar alırken ancak aynı yapıyı ölçmek için geliştirilmiş testlerde bireylerin yetenek puanları sabit kalmaktadır. Bu sayede, testlerden bağımsız olarak kestirilen yetenek puanları norm referanslı karşılaştırmalara olanak sağlamaktadır (Hambleton ve Jones, 1993).

1970'lerden itibaren MTK, uzmanlar için önemli bir çalışma alanı olmuştur. MTK'nın tarihi 1930'lu yıllara kadar uzanmaktadır. Bu süreçte, MTK'nın temel kavramlarından biri "Madde Karakteristik Eğrisi" terimi Tucker'a atfedilmiştir. 1950'lerde, konunun matematiksel karmaşıklığı ve bilgisayar programlarının yetersiz oluşu nedeniyle MTK çalışmaları yavaş ilerlemiştir. Alanyazına birçok katkıya rağmen MTK, 1960'lara kadar önemli bir gelişme gösterememiştir. Ancak MTK'ya olan ilgi gerçek puan teorisinin etkisiyle artmaya başlamıştır (Hambleton ve Swaminathan, 1985).

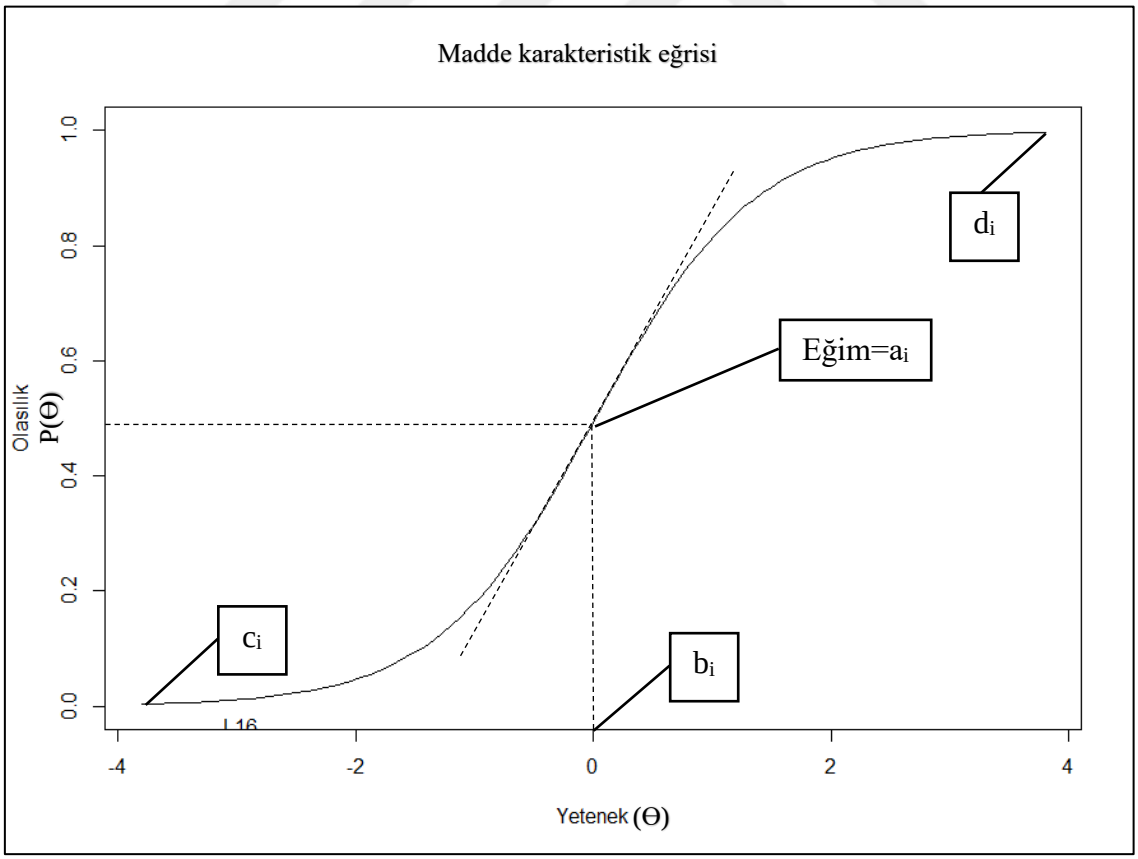
MTK, bireyin testten aldığı toplam puandan daha çok her bir maddeye verdiği yanıtın doğruluğu ya da yanlışlığıyla ilgilenir. Çünkü MTK'nın parametreleri, testten alınan toplam puana değil testin her bir maddesine dayalıdır. Bir test maddesini yanıtlayan her bir bireyin maddenin arka planında yatan gizil özelliğe bir miktar sahip olduğunu varsaymak mantıklı bir yaklaşımdır. Buna göre, her bir bireyin, yetenek ölçeği üzerinde kendisine ait bir puana sahip olduğu düşünülebilir. Her bir yetenek puanında, bu yetenek düzeyine sahip bir bireyin maddeyi doğru yanıtlamasına ilişkin belirli bir olasılık değeri bulunmaktadır (Baker, 2001).

Hambleton ve Swaminathan (1985), MTK'nın üzerinde şekillendiği iki temel postulatı şu şekilde belirtmişlerdir:

1. Bireylerin test performansı; özellik, gizil özellik ve yetenek (Θ) adı verilen faktörler tarafından açıklanabilir.
2. Bireyin madde üzerinde gösterdiği performans ile bu performansını etkilediği düşünülen özellikler arasındaki ilişki, monoton bir şekilde artmaktadır. Bu durum madde karakteristik eğrisi (MKE) olarak tanımlanan bir fonksiyonla tanımlanabilir.

Özellikle ikinci postülata göre MTK'nın temelinde Madde Karakteristik Eğrisi (MKE) bulunmaktadır. MKE, bir test performansının altında yatan gizil özelliğin (Θ), maddeyi doğru yanıtlama olasılığının fonksiyonu olarak tanımlanmaktadır (Crocker ve Algina, 1986). MKE, doğru yanıt verme olasılığı Θ'nın bir fonksiyonu olarak ifade edilmiştir. Θ'nın belirli bir değeri için olasılık hesaplandığında, bu olasılık belirlenen Θ yetenek düzeyindeki bir grup bireyden seçkisiz olarak seçilen bir bireyin doğru yanıt verme olasılığı P(Θ) olarak yorumlanır (DeMars, 2010).

P(Θ) değerlerinin, yetenek düzeyinin bir fonksiyonu olarak çizilmesi durumunda, Şekil 1'deki gibi S şeklinde bir eğri elde edilir. Bu eğriye, Madde Karakteristik Eğrisi (MKE) ya da Madde Tepki Fonksiyonu (MTF)'da denir.



Şekil 1. Madde Karakteristik Eğrisinin (MKE) Genel Görünümü

Şekil 1'de bir örneği görülen S şeklindeki bu eğri, yetenek ölçeği (Ɵ) ile maddeyi doğru yanıtlama olasılığı P(Ɵ) arasındaki ilişkiyi ifade eder (Baker, 2001). Bu eğri, gizil özellik puanı arttıkça maddenin doğru yanıtlanma olasılığının artması şeklinde yorumlanmaktadır (Crocker ve Algina, 1986). MKE'de eğriyi tanımlamak için çeşitli teknik özellikler ya da parametreler vardır. Bunlar, madde güçlüğü (b_i), madde ayırt ediciliği (a_i), şans (düşük asimptot) (c_i) ve yüksek asimptot (d_i)'tur. Madde güçlüğü (b_i), maddenin yetenek ölçeği üzerinde bulunan bir noktadır. P(Ɵ) değerinin 0.5 değerine karşılık gelen Ɵ değerine eşittir. Madde ayırt ediciliği (a_i), madde güçlülüğünün bulunduğu noktanın altındaki yetenek düzeylerine sahip bireyleri ne kadar iyi ayırt edebildiğini yansıtır. Madde ayırt ediciliği aslında madde karakteristik eğrisinin orta noktasındaki eğimi ifade eder. Düşük (c_i) ve yüksek (d_i) asimptot, alt ve üst yetenek düzeylerindeki MKE'nin durumunu gösteren parametrelerdir (Baker, 2001; Barton ve Lord, 1981).

Madde karakteristik eğrileri, fonksiyonda yer alan parametrelere göre değişmektedir. Bu değişim ya da farklılaşma MTK'nın modellerini oluşturur. MTK modelleri ile bir bireyin yeteneği doğrultusunda bir maddeye nasıl bir tepki vereceğinin olasılığı hesaplanmaktadır.

MTK modelleri, tek boyutlu ve çok boyutlu modeller olmak üzere ikiye ayrılmaktadır. Tek boyutlu modeller, maddelerin puanlanma şekline göre ikili (dichotomous) puanlanan ve çoklu (polytomous) puanlanan maddeler olarak ikiye ayrılmaktadır. Tek boyutlu modellerin arasındaki temel ayırım parametre sayısıdır. Tek boyutlu ve ikili puanlanan maddelerde kullanılan MTK modelleri, ogive ve lojistik modellerdir. Normal ogive fonksiyonun madde yanıt modellerinde kullanılması lojistik fonksiyonlara göre daha öncedir. Normal ogive modelin dezavantajı integral almayı gerektirmesidir. Yani, doğru yanıt verme olasılığı $a_i(Ɵ-b_i)$ değerinin altında kalan eğrinin alanının hesaplanarak bulunmasıdır. Bu nedenle lojistik fonksiyonlarla çalışılması daha kolaydır. Bu yüzden 1, 2, 3 ve 4 parametreli lojistik modellerde yer alan ve yaklaşık değeri 1.7 olan "D" sabiti lojistik model parametrelerini normal ogive model parametreleri ile aynı metriğe yerleştirir. Bu katsayının kullanılması genel bir kanıdır ancak evrensel değildir (DeMars, 2010).

İşlemlerin kolaylığı ve yorumlanabilmesi açısından lojistik modeller daha çok tercih edilmektedir. Alanyazında sıklıkla kullanılan lojistik modeller; Rasch model, bir parametreli lojistik model (1 PL), iki parametreli lojistik model (2 PL) ve üç parametreli lojistik (3 PL) modeldir (Hambleton ve Swaminathan, 1985; Hambleton, Swaminathan

ve Rogers, 1991; Hambleton ve Jones, 1993). Barton ve Lord (1981), üç parametreli model için yüksek asimptot adlı çalışmalarında dört parametreli lojistik modeli (4 PL) önermişlerdir.

1 PL model, test maddelerinin madde ayırt edicilik (a_i) parametrelerinin sabit bir değerde tutulduğu ve sadece madde güçlük (b_i) parametresindeki değişime göre madde yanıtlama olasılıklarının hesaplandığı modeldir. 1 PL modelde, madde güçlük (b_i) parametresi maddeyi doğru yanıtlama olasılığının P(Ɵ), 0.5 olduğu yetenek düzeyindeki değeri ifade eder. Bu modelin Rasch modelden farkı madde ayırt edicilik (a_i) parametrelerinin sabit bir değerde tutulmasıdır. Rasch modelde ise madde ayırt edicilik (a_i) parametresi 1'e sabitlenir. Ayrıca, Rasch modelde b_i yerine δ, Ɵ yerine de β sembolleri kullanılır. Sembol değişimlerinin matematiksel olarak hiçbir etkisi yoktur (DeMars, 2010).

2 PL model, birikimli normal dağılıma benzemektedir. Bu bakımdan normal ogive olarak da adlandırılır. Bu modelde, madde güçlük (b_i) ve madde ayırt edicilik (a_i) parametresi değişimlenmekte ancak düşük asimptot değeri (c_i) sıfır kabul edilmektedir.

3 PL model, Birnbaum (1968)'un 2PL modele doğru yanıt verme olasılığına şans başarısının etkisini gösteren üçüncü bir parametre (c_i) eklemesiyle geliştirilmiştir. Bu model, 2 PL modelin özel bir halidir.

Yüksek yetenekli bireyler test maddelerine her zaman doğru cevap vermeyebilirler. Bazen bu bireyler maddeyi yanıtlarken dikkatsiz davranabilirler ya da madde yazarının varsaydığından daha fazla bilgiye sahip olabilirler. Bu yüzden de maddeyi yanlış yanıtlamış olabilirler. Bu durumun çözümü için Barton ve Lord (1981), dört parametreli lojistik modeli (4 PL) önermişlerdir. Bu modelde, 3 PL modele ek olarak yüksek asimptot (d_i) parametresi eklenmiştir.

MTK'da, MKE'nin standart matematiksel modeli, lojistik fonksiyonunun yığılmalı halidir. MKE modelinin ilk kullanımı 1950'lilerin sonlarına karşılık gelmektedir. Modele D=1.7 çarpanının girmesi normal ogive modelden lojistik modele bir dönüşüm sağlar (Baker, 2001; Lord ve Novick, 1968). Normal ogive ve lojistik fonksiyon, birimlerin (1,7 lojit = 1 probit) farklı olması sebebiyle kullanılan formülü farklılaştırsa da model değişimine gereklilik oluşturmamaktadır (De Ayala, 2009).

MTK, KTK'ya oranla birçok problem durumuna çözüm bulmada avantajlı görünse de (Embretson ve Reise, 2000), temelinde karşılanmasını gerektiren varsayımları bulunmaktadır. Bunlar; tek boyutluluk ve yerel bağımsızlıktır (Baker, 2001; Baker ve

Kim, 2004; De Ayala, 2009; Hambleton ve Swaminathan, 1985; Hambleton, Swaminathan ve Rogers, 1991; Hambleton ve Jones, 1993).

Tek boyutluluk, bir madde üzerinde bireyin performansını etkileyen tek bir yetenek türü olması durumudur. Yani testin ölçtüğü özellik bakımından tek boyutlu olması gerektiğidir. Maddenin doğru yanıtlanma olasılığında eğer birden fazla özellik etkinse, tek boyutluluk varsayımı ihlal edilir. Ancak testlerde tek boyutluluğu sağlamak zor olduğundan ölçülen özellik bakımından başat bir faktör olması durumunda bu varsayımın yerine getirildiği kabul edilir (Hambleton ve Swaminathan, 1985; Embreston ve Reise, 2000).

Yerel bağımsızlık, test performansı üzerinde etkili olan yetenek düzeyi sabit tutulduğunda, bireylerin maddelere vereceği yanıtların birbirinden bağımsız veya ilişkisiz olmasıdır. Yerel bağımsızlık varsayımında, maddeler arasındaki ilişki veya bağımlılık sabit bir yetenek koşulu altında incelenmektedir. Yerel bağımsızlık varsayımının karşılanabilmesi için tek boyutluluk varsayımı önsel bir şarttır (Hambleton, Swaminathan ve Rogers, 1991; Hambleton ve Jones, 1993). Alanyazında yerel bağımsızlık varsayımının sağlanması için tek boyutluluk varsayımının sağlanmış olmasının yeterli olduğunu belirten çalışmalar (Embretson ve Reise, 2000; Hambleton ve Swaminathan, 1985; Lord, 1980; Morizot, Ainsworth ve Reise, 2007) olmasına rağmen yerel bağımsızlık varsayımının test edilmesi gerektiği önerilmektedir (DeMars, 2010). Yen'in (1984) Q_3 istatistiği, yerel bağımsızlık varsayımını test etmede kullanılan yöntemlerden birisidir. Bu yöntem madde çiftlerinin arasındaki artık korelasyonlar arasındaki ilişkiyi dikkate almaktadır. Bu yöntemin kritik değeri 0.20 olarak kabul edilmektedir (DeMars, 2010). Çeşitli çalışmalarda bu değer farklı olarak ele alınmıştır. Marais ve Andrich (2008), çalışmalarında 0.10 kritik değerini dikkate almışlardır. Ancak bu değerin 0.30 olarak alındığı (das Nair, Moreton ve Lincoln, 2011; La Porta, Franceschini, Caselli, Cavallini, Susassi, Tennnat, 2011; Ramp, Khan, Misajon ve Pallant, 2009; Røe, Damsgard, Fors ve Anke, 2014); 0.50 olarak alındığı (Davidson, Keating ve Eyres, 2004; Ten Klooster, Taal ve van der Laar, 2008); 0.70 olarak alındığı (Gonzalez-de paz, Kostov, Lopez-Pina, Solans-Julian, Navvaro-Rubio, Siso-Almirall, 2015) çalışmalar da mevcuttur. Bu kritik değerin seçiminde testin ya da ölçeğin ölçtüğü yapının özellikleri dikkate alınmalıdır (Røe ve diğerleri, 2014).

Bu varsayımların karşılanması MTK'nın avantajlarını ortaya çıkarmaktadır. Bu avantajlardan biri madde ve birey parametrelerinin gruptan bağımsız bir şekilde kestirilebilmesidir. Yani MTK, bireylerin testin ölçtüğü özellik bakımından sabit bir

yetenek düzeyine sahip olduğunu varsayar ve testin zorluğuna ya da kolaylığına bağlı olarak değişmeyeceğini kabul eder.

Değişmezlik, kestirimin yapıldığı gruptan seçkisiz olarak seçilen bireylerin yetenek parametrelerinin ya da belirli gruplardan elde edilen madde parametrelerinin sabit kalması anlamına gelmemektedir. Değişmezlik, madde ve yetenek parametrelerinin doğrusal dönüştürülebilirliği ve farklı gruplardan kestirilen parametreler arasındaki yüksek korelasyon elde edilmesidir. Bu durum, MTK'nın çok geniş gruplarda olası tüm yetenek düzeylerinde parametrelerinin kestirilmesiyle sağlanmaktadır. Aynı durum madde parametreleri içinde geçerlidir. Yani madde parametrelerinin gruba göre farklı değerler almayacağını varsayar. Daha önce de bahsedildiği gibi bu avantajların temelinde MTK'nın temel varsayımları bulunmaktadır (Hambleton, Swaminathan ve Rogers, 1991).

Maddelerin ve testlerin tanımlanması, test maddelerinin seçilmesi ve testlerin karşılaştırılması işlemlerinin gerçekleştirilebilmesi için güçlü bir yöntem, MTK tarafından sağlanmaktadır. Söz konusu yöntem, $I_i(\Theta)$ biçiminde ifade edilen madde bilgi fonksiyonunun kullanılmasıdır. Madde bilgisi fonksiyonunda a_i, b_i ve c_i parametrelerinin rolüne ilişkin olarak aşağıdaki çıkarımlarda bulunulmaktadır:

a) Elde edilen bilgi düzeyi b_i değeri Θ'a yakın olduğunda, b_i değerinin Θ'dan uzak olduğu duruma göre daha yüksektir.

b) Elde edilen bilgi düzeyi, a_i parametre değeri yüksek olduğunda genellikle yüksek olmaktadır.

c) c_i parametre değeri sıfıra yaklaştıkça elde edilen bilgi düzeyi artış gösterecektir.

Θ yetenek düzeyine uygun oluşturulmuş bir testten sağlanan bilgi, Θ düzeyindeki madde bilgi fonksiyonlarının toplamına eşittir (Hambleton, Swaminathan ve Rogers, 1991). MTK'da ölçmenin standart hatasını ve güvenirliğini hesaplamak için bilgi fonksiyonu kullanılır. Ölçmenin standart hatası da test bilgisinin karekökünün tersidir. Dolayısıyla testten elde edilen bilginin büyüklüğü standart hata ile ters orantılıdır. Yani bilgi ne kadar artarsa güvenirlik de o kadar yüksek olur (DeMars, 2010).

Bilgi fonksiyonu kavramı, maksimum olasılık kestiriminin varyansı ile ilgili genel bir kavramdır. Yetenek parametresinin maksimum olasılık kestirimi elde edildiğinde bu kestirimin varyansı, ilgili bilgi fonksiyonunun yerini tutan karşılığı olarak verilir. Benzer şekilde, madde parametrelerinin maksimum olasılık kestirimleri elde edildiğinde, kestirimlerden oluşturulan varyans-kovaryans matrisi, madde parametre kestirimlerine ait bilgi matrisinin tersi olarak ifade edilir.

Marjinal maksimum olabilirlik kestirimlerinin varyans-kovaryans matrisi için basit ifadeler mevcut değildir, ancak bunları elde etme işleminin bir açıklaması Mislevy (1984) ve Mislevy (1986)'de verilmiştir. Madde parametre kestirimlerinin varyans-kovaryans matrisi, yanlılık ve DMF çalışmalarında ortaya çıkan bir problem olmak üzere iki gruptaki madde parametrelerini karşılaştırırken önemlidir (Hambleton, Swaminathan ve Rogers, 1991; Hambleton ve Jones, 1993).

MTK, madde ve birey parametrelerini kestirirken temel aldığı varsayımların yanı sıra çeşitli yöntemleri de beraberinde getirir. MTK, temelinde bir yapısal kuram olmasından ziyade istatistiksel ve matematiksel hesaplamalara dayalı bir madde ölçekleme kuramı olarak da kabul edilebilir (Hambleton ve Jones, 1993). Bu ölçekleme ya da kalibrasyon adı verilen işlem yetenek ve madde parametrelerinin kestirilmesiyle başlayan süreçtir.

Klasik MTK Parametreleri Kestirim Yöntemleri

Madde Tepki Kuramını, test verisine uygulamak için ilk ve en önemli adım, seçilen madde tepki modelini tanımlayan parametrelerin kestirilmesidir. Aslında, MTK'nın başarılı bir şekilde uygulanması, model parametrelerini tahmin etmek için yeterli prosedürlerin bulunmasına bağlıdır. MTK modellerinde, bireyin kendisine yöneltilen maddeye doğru yanıt verme olasılığı, testi alan bireyin yeteneğine (Θ) ve maddeyi tanımlayan parametrelere bağlıdır. Hem bireyin yeteneğine hem de madde parametrelerine ilişkin herhangi bir bilgi mevcut değilse, bireylerin maddelere verdikleri yanıtlar hakkında da herhangi bir değerlendirmede bulunmak mümkün değildir. MTK'da bağımsız değişken olan yetenek parametreleri (Θ) gözlemlenebilir değildir. Yetenek parametreleri (Θ) gözlemlenebiliyor ya da biliniyor olması, madde parametrelerinin kestirilmesini kolaylaştıracaktır. Aynı şekilde, madde parametrelerinin biliniyor olması, yetenek tahminini basitleştirecektir. Bu durumda bilinmeyen parametreleri kestirebilmek için madde parametreleri kadar, bireylerden elde edilen toplam test puanlarına da ihtiyaç duyulmaktadır (Hambleton, Swaminathan ve Rogers, 1991; Hambleton ve Jones, 1993). Buna göre, MTK'da yetenek ve madde parametrelerinin nasıl kestirileceğine yönelik teknikler aşağıda verilmiştir.

Klasik MTK Yetenek Parametreleri Kestirim Yöntemleri

MTK'ya göre geliştirilen testlerin amacı, bireylerin yetenek ölçeği üzerindeki yerlerinin tespit edilmesidir. Gizil bir özelliği ölçmek için kullanılan bir testteki her bir madde gizil özelliğin bir başka yönünü ölçmektedir. Testi alan bir birey testteki n tane maddenin her birini yanıtlar ve bireyin yanıtları ikili (1-0) olarak puanlanır. Bireyin testteki her bir maddeye verdiği yanıtlardan tepki örüntüsü (yanıt deseni, yanıt vektörü) oluşturulur. Madde tepki örüntüsüyle birlikte madde parametreleri kullanılarak testi alan bireyin yetenek parametresine ilişkin kestirimleri yapılır (Baker, 2001).

Yetenek kestirimi için maddeler, hedeflenen evrenden seçkisiz olarak seçilen örnekleme uygulanarak bireylerin maddelere verdikleri tepkiler alınır. Yetenek ve madde parametreleri bu tepkilerden kestirilir (Erkuş, Sünbül, Ömür, Yormaz ve Aşiret, 2017).

Klasik MTK'da yetenek parametrelerini kestirmek için en çok olabilirlik yöntemi kullanılmaktadır. Bu yöntemin süreçlerini daha kolay anlayabilmek için rastgele seçilen bir yanıtlayıcının, kendisine sunulan n tane maddeyi yanıtladığını ve elde edilen yanıtların U_1, U_2, …, U_j, …, U_n şeklinde, doğru yanıt 1 ve yanlış yanıt 0 olacak biçimde oluşturulduğu bir yanıt deseni elde edildiğini varsayalım. Yerel bağımsızlık varsayımı göz önüne alındığında, yanıt deseninin gözlemlenme olasılığı, maddelerin her birine verilen yanıtların gözlemlenme olasılıklarının bir ürünü olarak elde edilecektir. Yani,

$$\mathbf{P(U_1, U_2, \dots, U_J, \dots, U_n|\theta) = P(U_1|\theta).P(U_2|\theta) \dots P(U_J|\theta) \dots P(U_n|\theta) = \prod_{j=1}^{n} P(U_J|\theta)}$$

biçiminde gösterilebilir.

Testi alan birey tarafından verilen yanıt, U_j, 1 ya da 0 değerini almaktadır. Bu durum göz önünde bulundurulduğunda olasılık fonksiyonu,

$$\mathbf{P(U_1, U_2, \dots, U_n|\theta) = \prod_{j=1}^{n}(U_J|\theta)^{U}.(U_J|\theta)^{1-U_J} = \prod_{j=1}^{n} P_J^{U_J} - Q_J^{1-U_J}}$$

şeklinde ifade edilebilir.

Bu eşitlik, yanıt desenine ilişkin ortak olasılık değerini ifade etmektedir. Eğer bu yanıt deseni gözlenirse yani eşitlikteki U_1, U_2, …, U_n; değişkenleri U_i'nin 1 ya da 0 olduğu değerlere dönüşürse bu eşitlik artık olasılık olmaktan çıkar, olabilirlik fonksiyonuna dönüşür (Hambleton, Swaminathan ve Rogers, 1991).

$$\mathbf{L(U_1, U_2, \dots, U_n|\theta) = \prod_{J=1}^{n} P_J^{U_J} - Q_J^{1-U_J}}$$

Pj ve Qj, Θ ve madde parametrelerinin fonksiyonları olduğundan, olabilirlik fonksiyonu da bu parametrelerin bir fonksiyonudur. Olabilirlik fonksiyonunu maksimum yapan Θ'nın değeri yetenek kestirimi olarak kabul edilir (Hambleton ve Swaminathan, 1985).

Ayrıca, Θ yetenek değerine ait bir olabilirlik fonksiyonu grafiğinin çizilmesiyle de Θ'nın en çok olabilirlik kestirimi belirlenebilmektedir. Olabilirlik fonksiyonu L(u|θ), her biri 0 ile 1 arasında değişen olasılık ifadelerinin bir sonucu olduğu için değeri çok küçük olacaktır. Olabilirlik fonksiyonunun daha iyi ölçeklendirilmesi, bu fonksiyonun logaritma kullanılarak dönüştürülmesi ile sağlanabilir. Bu sayede, ln L(u|θ), (-∞, 0) arasında değer alabilecektir. Ayrıca, logaritmalar toplam olarak da ifade edilebilmektedir. Logaritma kullanımı, hesaplamaları basitleştirir. Olasılık fonksiyonunun logaritması için oluşturulan genel ifade (kısaca, log olasılığı) şu şekilde yazılabilir:

$$\mathbf{\ln\ L(u|\theta) = \sum_{J=1}^{n} [u_j . \ln P_j + (1 - u_j) . \ln(1 - P_j)]}$$

Burada tanımlanan u, madde yanıtlarının vektörüdür. Ancak, fazla sayıda bireyin ve maddenin var olduğu hesaplamalar için bu fonksiyon kullanışlı olmayacaktır. Daha etkin hesaplamalarda, fonksiyonun maksimuma ulaştığı noktada, fonksiyonun eğiminin (birinci türevinin) sıfır olduğu durum kullanılır. Böylece, maksimum olabilirlik tahmini, olabilirlik ya da log olabilirlik fonksiyonunun birinci türevinin sıfıra eşitlenmesiyle elde edilen denklemin çözülmesi sayesinde gerçekleştirilebilir. Tekrar söylemek gerekirse, bu denklem doğrudan çözülemez ve çözüm için yakınsama (iteratif) yöntemleri kullanılmalıdır. Bu çözümlerden bir tanesi de Newton-Raphson yöntemidir (Hambleton, Swaminathan ve Rogers, 1991). Bu yöntemde yetenek kestiriminde kullanılan eşitlik aşağıdaki gibidir:

$$\mathbf{\theta_{m+1} \equiv \theta_m - \underbrace{\left[\frac{d}{d\theta} \ln L\,(u|\theta)\right]_m / \left[\frac{d^2}{d\theta^2} \ln L\,(u|\theta)\right]_m}_{\Delta\Theta}}$$

θ_{m+1}: Bireyin gerçek Θ'sı

θ_m: Bireyin yetenek düzeyine ilişkin yapılan kestirim

ΔΘ: Düzeltme faktörü

m: İterasyon değeri

Öncelikle eşitliğin sağ tarafındaki Θ_m keyfi bir değer olarak belirlenir. MKE modelindeki parametreler kullanılarak testte yer alan n maddenin her birine belirlenen yetenek düzeyinde doğru yanıt verme olasılığı hesaplanır. Daha sonra eşitliğin sağ tarafındaki ikinci terim ($\Delta\Theta$) değerlendirilir. Bu düzeltme faktörüdür ve $\Delta\Theta$ ile gösterilir. Eşitliğin sol tarafındaki Θ_{m+1} değeri, Θ_m'e $\Delta\Theta$'nın eklenmesiyle elde edilir. Θ_{m+1} bir sonraki tekrarlama işlemindeki Θ_m olur. Düzeltme faktörü gerçekleştirilen işlemin esasını oluşturur. Bu işlem, Θ_{m+1} ile Θ_m arasındaki fark önceden belirlenen bir değerin (ölçüt) altında kalıncaya kadar devam ettirilir. Yakınsamış değer Θ'nın en çok olabilirlik kestirimi olarak alınmakta ve $\hat{\theta}$ ile ifade edilmektedir. Newton-Raphson yöntemi, Hambleton ve Swamimnathan (1985) ve Baker ve Kim (2004)'de ayrıntılı olarak açıklanmıştır.

Yetenek kestirimine ilişkin yapılan kestirimin duyarlılığını gösteren bir standart hata elde edilebilir. Standart hata (SE), $\hat{\theta}$'nın bireyin bilinmeyen Θ yetenek parametresi etrafındaki değişimin ölçüsüdür. Bu durumda standart hataya ilişkin kestirim aşağıdaki eşitlik yardımıyla hesaplanabilir:

$$\mathbf{SE}(\hat{\boldsymbol{\theta}}) = \frac{\mathbf{1}}{\sqrt{\sum_{\mathbf{j=1}}^{\mathbf{N}} \mathbf{a_j^2 P}(\hat{\boldsymbol{\theta}})\mathbf{Q}(\hat{\boldsymbol{\theta}})}}$$

Dolayısıyla, standart hata bireyin yetenek düzeyine ilişkin yapılan kestirimin bir yan ürünü olarak elde edilmektedir (Baker, 2001). Ancak testi alan bir birey tüm maddeleri doğru ya da yanlış olarak yanıtladığında, olabilirlik (ya da log olabilirlik) fonksiyonu maksimum olduğu noktada sınırlı bir değere sahip olmayabilir. Bu nedenle maksimum olabilirlik değeri -∞ ile +∞ arasında değer almaktadır. Bazı özel yanıt desenleri sonlu mutlak maksimum değere sahip olmayan bir olabilirlik fonksiyonunun oluşmasına yol açabilir. Dolayısıyla, her iki durum için de maksimum olabilirlik kestirimleri mevcut olmayacaktır. Bu durumun nedeni, testi alan bireylerin oluşturdukları yanıt desenlerinin normal olmamasıdır. Testi alan bireyler nispeten zor ve seçici bazı maddeleri doğru ve bazı kolay maddeleri ise yanlış olarak yanıtlamış olabilirler. Bu tip durumlarda, fonksiyonun maksimum değer aldığı noktayı bulmak için kullanılan yöntemler de değişiklik gösterecektir. Bu tür durumlarda maksimum olabilirlik kestirim

işleminin gerçekleştirilememesi halinde, Bayes kestirim işlemi kullanılarak söz konusu bu problem çözüme ulaştırılmaktadır (Lord, 1980; Hambleton, Swaminathan ve Rogers, 1991).

Klasik MTK Madde Parametreleri Kestirim Yöntemleri

Ɵ yetenek düzeylerinin belirlenmesinde madde parametrelerinin bilindiği varsayılmaktadır. Daha önceki kısımda anlatıldığı üzere madde parametreleri bilindiği takdirde testi alan bir bireyin yetenek düzeyinin kestirimi için bireye çok sayıda test maddesi yöneltilmekte ve bireyin bu maddelere verdiği yanıtlar kullanılarak bir olabilirlik fonksiyonu elde edilmektedir. Bu olabilirlik fonksiyonundan hareketle bireylerin yetenek düzeyleri kestirilebilmektedir. Bu durumun tam tersi olarak bireylerin yetenek düzeyleri bilindiğinde madde parametrelerinin kestirilmesi de mümkündür.

Testi alan her bir bireyin Ɵ yetenek düzeyinin bilindiği bir durumda madde parametreleri kestirilmek isteniyorsa, ilgilenilen yeteneğe ilişkin maddeler çok sayıda bireye uygulanır ve bireylerin maddelere verdikleri yanıtlar doğrultusunda olabilirlik fonksiyonu elde edilir. Buna göre 3 PL model için aşağıdaki eşitlik kurulur:

$$\mathbf{L(u_1, u_2, \dots, u_n|\theta, a, b, c)} = \prod_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{P_i^{u_j}.Q_i^{1-u_j}}$$

Testi alan bir birey ile testte yer alan bir madde için oluşturulan olabilirlik fonksiyonları arasındaki fark, madde için elde edilecek olabilirlik fonksiyonu için yerel bağımsızlık varsayımına ihtiyaç duyulmamasıdır. Ɵ yetenek düzeyine ilişkin değerler bilindiği takdirde madde parametrelerinin kestirimi, önceki kısımda anlatılan işlemlere benzer biçimde yapılmaktadır. Aradaki fark, testi alan her birey için oluşturulan olabilirlik fonksiyonunun aksine, madde için oluşturulan olabilirlik fonksiyonunun üç parametreden oluşması, yani çok boyutlu olmasıdır. Bu yüzden a_i, b_i ve c_i parametrelerine ait maksimum olabilirlik fonksiyonunu elde etmek için üç boyutlu bir düzlemde a_i, b_i ve c_i değerlerine karşılık gelen maksimum değerler bulunmalıdır. Bunun için, olabilirlik fonksiyonunun birinci dereceden türevinin a_i, b_i ve c_i parametrelerinin her biri için ayrı ayrı bulunması, elde edilen türevlerin sıfıra göre ayarlanması ve elde edilen üç denklemin eş zamanlı olarak üç bilinmeyenli ve doğrusal olmayan bir eşitlik halinde çözümlenmesi gerekmektedir. Bu yüzden, kestirim işlemleri her madde için bir kez olmak üzere n madde için n kez tekrarlanmalıdır (Hambleton, Swaminathan ve Rogers, 1991).

Klasik MTK'da Yetenek ve Madde Parametrelerinin Birlikte Kestirim Yöntemleri

Bazı durumlarda hem yetenek hem de madde parametreleri bilinmemektedir. Bu tür durumlarda klasik MTK'da sıklıkla kullanılan kestirim yöntemleri şunlardır (Warm, 1989; Baker ve Kim, 2004):

- Klasik En Çok Olabilirlik Kestirimi (Joint MLE)
- Koşullu En Çok Olabilirlik Kestirimi (Conditional MLE)
- Ağırlıklandırılmış En Çok Olabilirlik Kestirimi (Weighted MLE)
- Marjinal En Çok Olabilirlik Kestirimi (Marginal MLE)

Klasik En Çok Olabilirlik Kestirimi (Joint MLE). Bir noktada hem Ɵ yetenek düzeyinin hem de madde parametrelerinin bilinemeyeceği bir durum söz konusu olabilmektedir. Böyle bir durumda testi alan tüm bireylerin tüm test maddelerine verdikleri yanıtlar eş zamanlı olarak değerlendirilmektedir. Testi alan n birey, maddelerin tümüne yanıt verdiğinde yerel bağımsızlık varsayımı kullanılarak elde edilen olabilirlik fonksiyonu,

$$\mathbf{L(u_1, u_2, \dots, u_n|\Theta, a, b, c)} = \prod_{\mathbf{i=1}}^{\mathbf{N}} \prod_{\mathbf{j=1}}^{\mathbf{n}} \mathbf{P_i^{u_j}.Q_i^{1-u_j}}$$

biçiminde ifade edilebilir. Bu bağıntıda u_i, testi alan bireyin n maddeye verdiği yanıt desenini; Ɵ, N adet yetenek parametresine ait vektörü; a, b ve c ise, n maddeli test için madde parametrelerine ait vektörleri temsil etmektedir. Madde parametrelerinin sayısı, üç parametreli model için 3n'dir (iki parametreli model için 2n ve bir parametreli model için ise n). Yetenek parametrelerinin sayısı N'dir ve bu yüzden üç parametreli model için toplamda 3n+N adet parametre tahmin edilmektedir. Bununla birlikte kestirimin yapılabilmesi için belirsizlik sorunu da ele alınmalıdır.

Yukarıda verilen olabilirlik fonksiyonunda, madde ve yetenek parametreleri tek bir biçimde belirlenmemiştir. Örneğin, üç parametreli modeli baz alacak olursak, eğer Ɵ yerine Ɵ'=αƟ+β, b yerine b'=αb+β ve a yerine a'=a/α alınacak olursa doğru yanıt verme olasılığında herhangi bir değişme olmayacaktır.

$$\mathbf{P(\theta) = P(\theta^*)}$$

α ve β, keyfi ölçeklendirme sabitleri olduğundan, olabilirlik fonksiyonu tek bir maksimum değere sahip olmayacaktır. Maksimum olabilirlik fonksiyonunu elde etmek

için kullanılan herhangi bir algoritma, söz konusu bu belirsizlik nedeniyle başarısız olacaktır. Belirsizlik sorunu, yetenek değerleri (veya b_i değerleri) için keyfi bir ölçek seçilerek ortadan kaldırılabilmektedir. Genellikle N tane yetenek değerinin (ya da n adet maddenin güçlük değerinin) ortalaması ve standart sapması sırasıyla 0 ve 1 olarak ayarlanır. Bu ölçeklendirme iki veya daha fazla grup için madde parametrelerine ilişkin kestirimleri karşılaştırılırken dikkate alınmalıdır.

Belirsizlik ortadan kaldırıldıktan sonra, olabilirlik fonksiyonunu maksimum yapan madde ve yetenek parametrelerinin değerleri belirlenebilmektedir. Eşzamanlı veya klasik maksimum olabilirlik kestirim algoritmalarında, bu belirleme iki aşamada yapılmaktadır. İlk aşamada yetenek parametreleri için başlangıç değerleri seçilir. Testi alan her bir birey için doğru yanıt sayısının yanlış yanıt sayısına oranının logaritma değeri, iyi başlangıç değerleri sağlamaktadır. Bu değerler daha sonra standartlaştırılır (belirsizliği ortadan kaldırmak için) ve yetenek değerlerinin bilinen şekilde ele alınmasıyla madde parametreleri kestirilir. İkinci aşamada, madde parametrelerinin bilinen şekilde ele alınması ile yetenek parametrelerinin kestirimi gerçekleştirilmektedir. Bu işlemler, kestirim değerleri ardışık iki kestirim arasında belirli bir yakınsama ölçüsüne kadar tekrarlanır. Bu tekrar sürecinde başvurulan yöntem Newton-Raphson ya da Fisher yöntemidir.

Klasik maksimum olasılık işlemi, kavramsal anlamda cazip görünmekle birlikte bazı dezavantajlara sahiptir. Birincisi, tam ve sıfır puanlara sahip yeteneklerin kestirimi mevcut değildir. İkincisi, testi alan tüm bireyler tarafından doğru ya da yanlış olarak yanıtlanan maddeler için madde parametre kestirimleri yapılamamaktadır. Bu tür yanıt örüntülerine sahip maddeler ve test alan bireyler, kestirim işlemi gerçekleştirilmeden önce elenmelidirler. Üçüncüsü, iki ve üç parametreli modellerde klasik maksimum olabilirlik işlemleri, madde ve yetenek parametreleri için tutarlı kestirimler sağlamamaktadır. Swaminathan ve Gifford (1985), ampirik olarak hem testi alan birey sayısı hem de madde sayısı büyük olursa, madde ve yetenek parametreleri için tutarlı tahminlerin elde edilebileceğini göstermiştir. Dördüncüsü, üç parametreli modelde, madde ve yetenek parametrelerinin aldığı değerlere sınırlamalar getirilmediği takdirde, kestirimleri elde etmek için gerçekleştirilen sayısal işlemler başarısız olabilir (Hambleton, Swaminathan ve Rogers, 1991).

Koşullu En Çok Olabilirlik Kestirimi (Conditional MLE). Koşullu olasılıklara dayanan en çok olabilirlik parametre tahmini Andersen (1972, 1973) tarafından geliştirilmiştir (akt: Baker ve Kim, 2004). Andersen (1972, 1973), madde parametrelerinin en çok olabilirlik kestirimlerinde sabit olmadığını öne sürmüştür. Bunun nedeni kestiricilerdeki yanlılığın birey sayısının artmasıyla ile ortadan kaybolmadığıdır. Bu yüzden madde parametrelerinin koşullu olarak kestirilmesi gerektiğini öne sürmüştür (Hambleton ve Swamimnathan, 1985).

Bu kestirim yönteminde, bilinmeyen yetenek düzeyleri ve yetenek düzeyi parametreleri dâhil edilmeyen yanıt örüntülerinin olasılıkları kullanılmaktadır. Bu yöntemin kullanılabilmesi için yetenek düzeyinin kestirilmesinde yeterli istatistik bilgisinin bulunması gerekmektedir. Yeterli istatistik demek, parametreleri kestirmek için verilerden başka bilgiye ihtiyaç duyulmadığı anlamına gelmektedir. Örneğin, Rasch modelde kişilerin toplam puanları yetenek düzeyini kestirmede yeterli bir istatistik olarak görülmektedir. Buna göre, hangi maddelerden aldığına bakılmaksızın aynı toplam puana sahip bireyler, aynı yetenek düzeyine sahip olacaklardır. Ancak, 2 PL ya da 3 PL modellerde toplam puan yetenek düzeyini kestirmede yeterli bir istatistik değildir. Bu nedenle koşullu en çok olabilirlik yöntemi yeterli istatistiğe ihtiyaç duyulduğundan sadece Rasch ve uzantılarında uygulanabilmektedir (Embretson ve Reise, 2000).

Baker ve Kim (2004), yapısal parametreler (madde parametreleri) ve tesadüfi parametreler (yetenek parametreleri) sorunu nedeniyle, bu yöntemin sadece madde güçlüğü parametrelerinin kestiriminde kullanılabileceğini belirtmiştir.

Koşullu en çok olabilirlik yöntemi eşitlikleri, klasik en çok olabilirlik yöntemi eşitliklerine maddenin doğru yanıtlanma olasılığın eklenmiş hali gibidir. Madde parametrelerinin tahmininde Newton-Raphson yönteminde olduğu gibi iteratif bir süreç uygulanır. Madde parametreleri iteratif olarak kestirildikten sonra, bunların madde parametrelerinin değerleri olarak kabul edilir ve birey parametreleri kestirilir (Embretson ve Reise, 2000).

Koşullu en çok olabilirlik yöntemi çeşitli avantajlara sahiptir. Birincisi, yetenek düzeyi için bir dağılım varsaymaz. Bilinmeyen yetenek parametreleri, madde kestirimi için kullanılan eşitliklerin dışında tutulur. İkincisi, madde parametresi kestirimleri, madde parametrelerinin değişmezlik ilkesini yansıtır. Başka bir deyişle, kestirimler değişmezdirler, çünkü örneklemin yetenek düzeyinden doğrudan etkilenmezler. Üçüncüsü, birçok koşul altında istenen özelliklere sahiptir. Yani tutarlı, normal dağılmış ve etkilidirler. Dördüncüsü, kestirimlerin standart hatası, klasik en çok olabilirlik

yöntemine göre daha kabul görülebilir seviyededir. Beşincisi, verilerin en çok olabilirlik (log-likelihood) değerleri hipotez testleri için kullanılabilir. Örneğin, Rasch model ile madde parametreleri üzerinde sınırlandırılmış bir model ki-kare olabilirlik oran testi ile karşılaştırılabilir.

Koşullu en çok olabilirlik yöntemi çeşitli dezavantajlara sahiptir. İlk olarak en göze çarpanı sadece Rasch model ve onun uzantılarında kullanılabilmesidir. Daha önce belirtildiği gibi 2 PL ve 3 PL modellerde kullanılamaz, çünkü bu modellerde toplam puan özellik düzeyini tahmin etmede yeterli bir istatistik değildir. İkincisi, tam puan alan bireyler ve herkes tarafında doğru yanıtlanan maddeler için kestirim gerçekleştirilemez. Üçüncüsü, test uzunluğu arttıkça verilerdeki karmaşıklıklar, yanıt örüntülerindeki farklılaşmalar ve çoklu puanlanan maddeler bu yöntemin kullanılmasında bir kısıtlılıktır. Çünkü kestirimlerin hassasiyeti etkilenmekte ve hataların oranı artmaktadır. Sonuncusu, bu yöntem hakkında genel olarak bir eksiklik düşüncesinin hâkim olmasıdır. Çünkü Rasch modele göre yapılan bu kestirimlerde yetenek dağılımlarının yetenek düzeyinde etkili olmaması psikometristler tarafından bir eksiklik olarak görülmektedir (Embretson ve Reise, 2000).

Ağırlıklandırılmış En Çok Olabilirlik Kestirimi (Weighted MLE). Warm (1989) tarafından önerilen ağırlıklandırılmış en çok olabilirlik yöntemi, ikili puanlanan maddelerde bir Ɵ kestirim yöntemi olarak kullanılmaktadır. Warm (1989), 2 PL MTK modeli için, düzeltilmiş ya da ağırlıklı bir log-olasılık denklemini çözerek en çok olabilirlik (ML) kestirimine bir düzeltme sağlayan bir ağırlıklandırılmış en çok olabilirlik (WML) kestirimi önermiştir. Bu yeni kestirim yöntemi madde parametrelerinin gerçek değerlerinin bilindiği varsayımıyla hareket etmektedir (Warm, 1989). Daha sonra çoklu puanlanan maddeler içinde kullanılmıştır (Samejima, 1998, Wang ve Wang, 2001). Bu yöntemin klasik en çok olabilirlik yöntemine göre daha az yanlı kestirimler yaptığı iddia edilmektedir (Kim ve Nicewander, 1993; Warm, 1989). Buna göre ağırlıklandırılmış en çok olabilirlik yönteminin eşitliği aşağıdaki gibidir:

$$\frac{\sum_j \mathbf{P'}_j\left(\mathbf{u}_j - \mathbf{P}_j\right)}{\mathbf{P}_j\mathbf{Q}_j} + \frac{\mathbf{P'}_j\mathbf{P''}_j}{\mathbf{2}\sum_j\left(\mathbf{P'}_j\right)^2} = \mathbf{0}$$

Bu eşitlikte,

$$\mathbf{P'_j} = \frac{\partial \mathbf{P_j}}{\partial \Theta_j}$$

yerine yazılırsa

$$\mathbf{P''_j} = \frac{\partial^2 \mathbf{P_j}}{\partial^2 \Theta}$$

elde edilir. Bu değer, Θ'nın ağırlıklandırılmış en çok olabilirliği eşitliği maksimize eden değerdir.

Marjinal En Çok Olabilirlik Kestirimi (Marginal MLE). Koşullu en çok olabilirlik yönteminde belirtildiği gibi, yeterli istatistik mevcut olduğunda, madde parametreleri yetenek parametrelerine bağlı olarak koşullu bir şekilde kestirilebilmektedir. Ancak bu durum sadece Rasch modelde mümkün olabilmektedir. Ancak 2 PL modelde ağırlıklandırılmış yanıt $(r = \sum a_i u_i)$ kestirilmemiş ayırt edicilik (a_i) parametresine bağlı olduğundan koşullu bir kestirim mümkün olmamaktadır. Alternatif olarak, olabilirlik fonksiyonu yetenek parametrelerine değinmeden ifade edilebilirse, madde parametreleri kestirilebilmektedir. Eğer yetenek parametreleri sürekliyse ya da bu parametreler toplanabilirse bu parametrelerin (yetenek dağılımının) integralinin alınmasıyla olabilirlik fonksiyonundan yetenek parametresinin çıkarımı sağlanmış olur. Elde edilen bu olabilirlik fonksiyonu marjinal olabilirlik fonksiyonudur. Madde parametrelerinin marjinal olabilirlik kestiricileri, marjinal olabilirlik fonksiyonunu maksimize eden değerlerdir (Hambleton ve Swamimnathan,1985).

Madde parametreleri marjinal en çok olabilirlik yöntemi kullanılarak kestirildiğinde daha doğru sonuçlar vermektedir. Çünkü marjinal en çok olabilirlik yönteminde yetenek parametresi kestirimleri, madde parametresi kestirimlerine dahil edilmezler. Yetenek parametrelerinin bu sürece dahil edilmemesiyle madde parametreleri daha tutarlı bir şekilde kestirilir. Bu tutarlılık, geniş bir örneklem ve sonlu bir testte kestirilen madde parametrelerinin iyileşmesine bağlıdır (Gao ve Chen, 2015).

N bireye, n maddeden oluşan tek boyutlu bir test uygulandığını varsayalım. Buna göre yetenek düzeyi Θ_i olan bir bireyin doğru (X_{ij}=1) ve diğer (X_{ij}=0) yanıtlarından oluşan yanıt puan vektörü $X_i=(X_{i1}, X_{i2}, \ldots, X_{in})$ olacaktır. Buradan bu puan matrisi,

$$\mathbf{X} = [\mathbf{X_i}]_{\mathbf{i=1}}^{\mathbf{N}}$$

gibi olacaktır. Buna göre Θ yetenek düzeyine sahip bir bireyin maddeyi doğru yanıtlama olasılığı $P_j(\Theta)$ fonksiyonu ile modellenir ve J maddesine ait parametre vektörü de δ_j ile gösterilirse tüm testin için madde parametre vektörü $\delta=(\delta_1, \delta_2, \ldots,\delta_n)$ olacaktır. Yerel bağımsızlık varsayımı dikkate alındığında ve $Q_j(\Theta)=1-P_j(\Theta)$ olarak alındığında, Θ yetenek düzeyine sahip bir bireyin gözlenen puan vektörü $X_i=(X_{i1}, X_{i2}, \ldots, X_{in})$'in olabilirlik fonksiyonu,

$$\mathbf{f}_{(\mathbf{X_i}|\boldsymbol{\theta},\boldsymbol{\delta})}(\mathbf{X_i}|\boldsymbol{\theta},\boldsymbol{\delta}) = \prod_{\mathbf{j=1}}^{\mathbf{n}} \mathbf{P_j}(\boldsymbol{\theta})^{\mathbf{X_{ij}}}\mathbf{Q}(\boldsymbol{\theta})^{\mathbf{1-X_{ij}}}$$

Buna göre bireylerin (Θ_s), g(Θ) yoğunluk fonksiyonuna sahip bir dağılımdan gelen bir örneklem olduğu varsayılırsa, gözlenen X_i puan vektörüne ait marjinal olasılık fonksiyonu,

$$\mathbf{f}_{(\mathbf{X_i}|\,\boldsymbol{\delta})}(\mathbf{X_i}|\boldsymbol{\delta}) = \int \mathbf{f}_{(\mathbf{X_i}|\boldsymbol{\theta},\boldsymbol{\delta})}(\mathbf{X_i}|\boldsymbol{\theta},\boldsymbol{\delta})\,.\,\mathbf{g}(\boldsymbol{\theta})\mathbf{d}\boldsymbol{\theta}$$

olarak yazılabilir. X'in yanıt puan matrisinin marjinal olasılığı,

$$\mathbf{f}_{(\mathbf{X}|\boldsymbol{\delta})}(\mathbf{X}|\boldsymbol{\delta}) = \prod_{\mathbf{i=1}}^{\mathbf{N}} \mathbf{f}_{(\mathbf{X_i}|\boldsymbol{\delta})}(\mathbf{X_i}|\boldsymbol{\delta})$$

olarak yazılabilir. X'in gözlenmesinden sonra puan matrisinin marjinal olasılığının logaritması ya da $\hat{\delta}$'yı maksimize eden olabilirlik fonksiyonu

$$\mathbf{L_1} = \log \mathbf{f}_{(\mathbf{X}|\boldsymbol{\delta})}(\mathbf{X}|\boldsymbol{\delta})$$

şeklinde ifade edilir. Marjinal en çok olabilirlik, model veri uyumu sağlandığında sonlu testler için tutarlı madde parametreleri üretebilmektedir. Ancak küçük örneklemlerde, marjinal en çok olabilirlik yöntemiyle doğru madde parametreleri üretmek zordur. Bu durum özellikle c_i parametresinin kestirildiği 3 PL modelde daha zordur.

Bock ve Lieberman (1970), normal ogive model için marjinal en çok olabilirlik kestirimlerini belirlemişlerdir. Başlangıçta Bock ve Lieberman (1970), bu durum için yetenek parametreleri ile eş zamanlı kestirimlerinden dolayı madde parametrelerinin kestirimine koşullu kestirim adını vermişlerdir. Bunun aksine, bireyler seçkisiz bir örneklem olarak alındığında ve olabilirlik fonksiyonunun yetenek dağılımı üzerine integrali alındığında madde parametrelerinin kestiricilerinin koşulsuz kestiriciler olarak adlandırılması gerektiğini önermişlerdir. Bu terminoloji, Andersen (1972) ve Wright ve Panchapakesan (1969) gibi araştırmacılar tarafından koşullu ve koşulsuz gibi

adlandırmalarla karışıklığa neden olmuştur. Andersen ve Madsen (1977), bu karışıklığın çözümü için marjinal kestiriciler terimini önermişlerdir (akt: Hambleton ve Swamimnathan,1985).

Marjinal en çok olabilirlik yönteminde ilk olarak yetenek parametrelerinin marjinal dağılımı belirlenmektedir. Bock ve Lieberman (1970), 2 PL model için yetenek dağılımının ortalaması 0, varyansı 1 olan normal dağılımın Ɵ'nın üzerinde integrali olduğunu varsaymışladır. Bock ve Lieberman'ın önerdiği hesaplamalar kısa testler üzerine uygulanabilirdir (akt: Baker ve Kim, 2004).

Bock ve Aitkin (1981), marjinal en çok olabilirlik hesaplamalarını bir adım daha öteye taşımışladır. Belirli koşullar altında, EM (expectation-maximization) algoritmasının bir modifikasyonunu uygulayarak yetenek dağılımını deneysel olarak nitelendirmişlerdir. Genel olarak EM algoritması, modeldeki gözlenemeyen seçkisiz değişkenlerin varlığında modelin parametrelerinin en çok olabilirlik kestiricilerini bulmak için kullanılan iteratif bir süreçtir. "E", beklenti aşamasını, "M", maksimizasyon aşamasını temsil eder. Beklenti aşamasında (E), her bir özellik düzeyindeki (quadrature noktaları/q parselleri) maddeyi yanıtlayan beklenen kişi sayısıyla birlikte her bir maddeyi doğru yanıtlayan beklenen kişi sayısı hesaplanır. Ardından, beklenti aşamasındaki kestirimler kullanılarak madde parametresi kestirimlerini maksimum yapan olabilirlikler bulunur. İkinci adımda yeniden beklenti aşamasında bir önceki madde parametreleri kullanılarak yeni beklentiler hesaplanır. Yeni beklenti değerleri kullanılarak doruklaştırma (maksimizasyon) aşaması yeniden gerçekleştirilir. Bu durum değerler arasında yakınsama işleminin beklenilen düzeye inmesine kadar tekrar edilir (Embretson ve Reise, 2000).

Marjinal en çok olabilirlik yönteminin çeşitli avantajları vardır. Birincisi, çok boyutlu modeller de dâhil olmak üzere birçok MTK modelinde kullanılabilir. İkincisi hem uzun hem de kısa testler için kullanılabilir. Üçüncüsü, kestirimlerin standart hataları beklenen örneklem varyansının kestirimleri için kabul edilebilir seviyede yakınsama gösterir. Dördüncüsü, tam puanlar için kestirimler gerçekleştirebilir. Bireyler için bilgi kaybı yaşanmaz. Her bir maddenin dağılımı iyi bir şekilde belirlenirse, madde ayıklanma işlemi gerektirmez. Beşincisi, marjinal en çok olabilirlik kestirimleri, hipotez testleri ve uyum iyiliği istatistikleri için kullanılabilir. Eğer bir model başka bir modelin alt kümesiyse (nested) modeller arası ki kare olabilirlik oran testi ile manidarlıkları test edilebilir.

Marjinal en çok olabilirlik yönteminin çeşitli dezavantajları vardır. Birincisi, bu yöntem için bir algoritma programlamak zordur. Buna rağmen birçok MTK kestirimleri bu yöntemi sıklıkla kullanmaktadır. İkinci olarak, parametre kestirimleri dağılma uygun olarak yapılabilmesi için özellik (Θ) bir dağılım varsayılmalıdır (Embretson ve Reise, 2000). Marjinal en çok olabilirlik yöntemi, yetenek ve madde parametresini kestirim teknikleri açısından önemli bir ilerleme olduğu açıktır (Baker ve Kim, 2004).

Bayesçi MTK Parametreleri Kestirim Yöntemleri

MTK'da kullanılan en çok olabilirlik yöntemleri klasik (frekansçı) yaklaşıma dayanmaktadır. Ancak klasik yaklaşımların tek bir değere bağlı kalması ve dağılımlar hakkında yeterli bilgi sağlamaması bakımından eksiklikleri bulunmaktadır. Bayes yaklaşımı, dağılım hakkında önsel bilgileri de işin içine katarak kestirimlerin yapılmasına olanak vermektedir. Bayes yaklaşımında önsel dağılımın varyansı, parametre kestirimlerinin belirsizlik seviyesini yansıtmaktadır. Önsel dağılımın varyansı küçükse parametre kestirimlerinin hata oranları daha düşük çıkacaktır (van de Schoot ve Depaoli, 2014).

Bayesçi MTK Yetenek Parametreleri Kestirim Yöntemleri

Önceki deneyimlerden ya da ampirik çalışmalardan yola çıkarak, özelliğin (Θ), belli bir ortalama (μ) ve standart sapma (σ) ile normal dağılım gösterdiğini belirleyebiliriz. Bu durumda, önceki edinilen bilgiler f(Θ) biçiminde bir yoğunluk fonksiyonu olarak ifade edilebilir.

Bayes teoremi, B olayının gerçekleşme durumda A olayının meydana gelme olasılığını,

$$\mathbf{P(A\backslash B) \propto P(B\backslash A).P(A)}$$

bağıntısı ile ifade eder. Burada P(A), A olayının meydana gelmesine ilişkin önsel olasılık değeridir. Yukarıdaki ilişki aynı zamanda yoğunluk fonksiyonları için de geçerlidir. Burada A, yetenek düzeyini ve B ise gözlemlenen madde yanıt deseni olan U'yu temsil etmektedir. Buna göre Bayes teoremi şu şekilde de yazılabilir:

$$\mathbf{f(\theta|u) \propto f(u|\theta).f(\theta)}$$

Aslında $f(u|\theta)$, bir olasılık fonksiyonudur ve bu nedenle,

$$\mathbf{f(\theta|u) \propto L(u|\theta).f(\theta)}$$

posterior (sonsal) $\propto$ *likelihood (en çok olabilirlik) X prior (önsel)*

biçiminde ifade edilebilir.

Yeniden düzenlenmiş bu yeni olasılık fonksiyonu $f(\theta|u)$ sonsal yoğunluk fonksiyonu olarak adlandırılır ve bu fonksiyondan Ɵ değeri için elde edilecek değer, en muhtemel değerdir ve bu değer Ɵ'nın bir kestirimi olarak değerlendirilebilir. Ɵ için genel bir yoğunluk dağılımı bulunduğunu varsayarsak (yani, f(Ɵ)=k, a bir sabit olmak üzere) bu durum,

$$\mathbf{f(\theta|u) \propto L(u|\theta)}$$

bağıntısı kullanılarak ifade edilebilir. Bu durumda, Bayes kestirimi sayısal olarak maksimum olasılık kestirimi ile özdeştir. Bayesci bir yaklaşım kullanmak, maksimum olabilirlik yaklaşımı ile karşılaşılan bazı zorlukların çözümünü sağlayacaktır. Ɵ yetenek düzeyi için yapılan Bayes kestirimleri, sıfır doğru yanıt madde desenleri, tamamı doğru yanıtlanmış madde desenleri ve anormal yanıt desenleri için elde edilebilir.

Bayesçi MTK Yetenek ve Madde Parametrelerinin Birlikte Kestirim Yöntemleri

Klasik en çok olabilirlik yöntemiyle ilgili bazı problem durumları söz konusudur. Maddelerin tamamının doğru ya da yanlış yanıtlandığı durumlarda yetenek ve madde parametrelerinin kestirim hesaplamalarında eksiklikler ya da yetersizlikler bulunmaktadır. Asıl sorun, klasik en çok olabilirlik yönteminin sonlu uzunluktaki testlerde istatistiksel olarak tutarlı olmayan kestirimlerle sonuçlanmasıdır. Rasch modelin yetenek parametresi sorunu, ham test puanı yeterli bir istatistik olarak kullanılmasıyla çözülmektedir. Buna da koşullu en çok olabilirlik yöntemi denilmiştir. Madde parametrelerinin tutarsız kestirimleri için daha uygulanabilir bir çözüm olan marjinal en çok olabilirlik kullanılmıştır. Marjinal en çok olabilirlik yöntemi, klasik en çok olabilirlik yönteminden bireylerin yeteneklerinin bir evrendeki dağılıma sahip olduğu varsayımıyla ayrılır. Bu dağılım bilindiğinde ya da kestirildiğinde bunun olabilirlik fonksiyonuyla bütünleştirilebilmesine izin verir. Bu durum madde parametrelerini, yetenek parametrelerine bağımlılıktan kurtarır. Aynı zamanda sonlu uzunluktaki testler için madde parametrelerinin tutarlı en çok olabilirlik kestirimi ile sonuçlanmasını sağlar. Her

ne kadar marjinal en çok olabilirlik yöntemi/EM algoritması tutarsız madde parametreleri kestirimlerine çözüm üretse de yetenek kestirimi için kullanılan alışılmadık yanıt örüntülerinin oluşturduğu madde parametrelerindeki sapmaya tam olarak bir çözüm bulamamaktadır. Bu durum karşısında önerilen yöntem Bayes yaklaşımıdır. Bayes yaklaşımı, madde parametre kestirimlerinin yetenek parametreleri üzerine marjinalleştirme sonrasında yapılması ya da herhangi bir marjinalleştirme yapılmaması olarak ikiye ayrılmaktadır. Bayesçi yaklaşımla çalışan Mislevy (1986), kendi yöntemini marjinal en çok olabilirlik yöntemine dayandırmakta olup marjinalleştirilmiş Bayes tekniğini kullanmıştır. Çünkü yaklaşımı marjinal en çok olabilirlik yönteminin parametre kestirimlerinin sapkın değer almasını önlemektir. Bununla birlikte marjinalleştirilmiş çözümlerin marjinal olmayan çözümlerden daha üstün olduğuna dair bulgular elde edilmiştir. Ayrıca, bireyin yetenek kestirimi rastlantısal parametreler problemi gerektirmediğinden bu durumun çözümünde marjinal olmayan Bayes kestirim süreci kullanılabilir. PC-BILOG programında MTK parametrelerinin kestirimi için gerekli olan Bayes prosedürleri bulunmaktadır. Ancak bu programın tam anlamıyla kullanılabilmesi için Bayes temelindeki süreçlerin tam olarak anlaşılması gerekmektedir. Bayes kestirim süreçleri marjinal en çok olabilirlik yönteminin EM algoritmasının uzantısı olarak düşünülebilir. (Kim ve Baker, 2004).

Bayes yaklaşımında kestirilen parametrenin δ belirli bir olasılık dağılımını (önsel dağılım) takip eden bir seçkisiz vektör olduğu kabul edilmektedir. Bayes kestirimini bulmak için sonsal dağılım fonksiyonu maksimize edilir. $g_{\Theta}(\Theta)$, Θ'nın önsel dağılımını göstermektedir. Bayes teoremine göre, X'in puan matrisinden üretilen Θ'nın sonsal dağılımı $f_{X|\theta}(X|\Theta)$ olarak ifade edilmektedir. Bu ifade bundan sonra

$$\mathbf{f}_{\mathbf{X}|\theta}(\mathbf{X}|\Theta)\mathbf{g}_{\Theta}(\Theta)$$

ifadesi ile orantılıdır. Bayes kestiricisi $\widehat{\Theta}$ yukarıdaki ifadenin mod değeridir. Aynı zamanda bu kestirici bu ifadeyi maksimize eden değerdir. Bu değerin logaritması

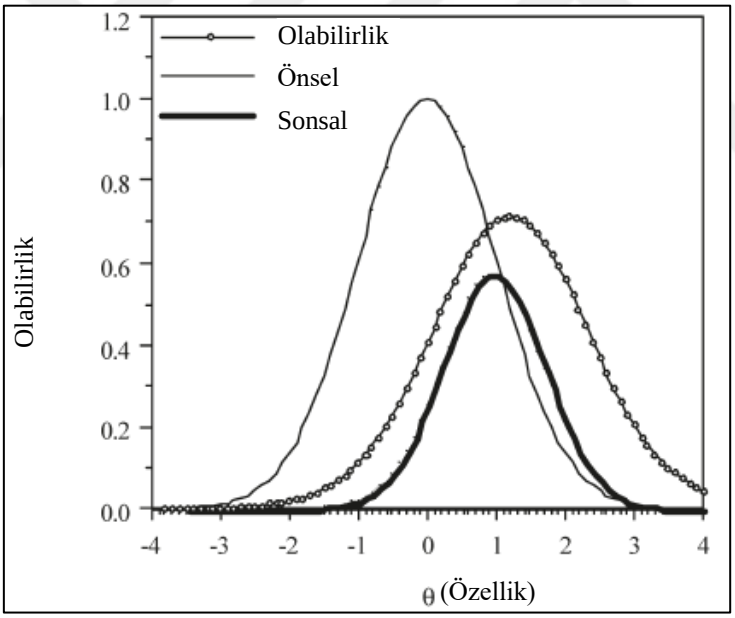
$$\mathbf{L_2} = \mathbf{log}\, \mathbf{f}_{\mathbf{X}|\theta}(\mathbf{X}|\Theta)\mathbf{g}_{\Theta}(\Theta)$$

$$\mathbf{L_2} = \mathbf{log}\, \mathbf{f}_{\mathbf{X}|\theta}(\mathbf{X}|\Theta) + \mathbf{log}\, \mathbf{g}_{\Theta}(\Theta)$$

olarak ifade edilir. Buradan görüleceği üzere $\mathbf{L_2} = \mathbf{L_1} + \mathbf{log}\, \mathbf{g}_{\Theta}(\Theta)$'dir. Böylece marjinal en çok olabilirlik kestirimini maksimize eden eşitliğe ek olarak Bayesçi kestirimde parametrelerin önsel bilgilerinin bulunduğu önsel dağılım süreci de işe koşulur.

Marjinal en çok olabilirlik ile Bayesçi kestirimler arasındaki farkın büyüklüğü Θ'nın önsel dağılımının L_2'yi maksimize etme etkisine bağlıdır. Yani $\log g_\Theta(\theta)$'nin görece olarak L_2'ye katkısıdır. Eğer L_2 eşitliğinde tek-biçimli (uniform) bir önsel dağılım kullanılırsa, o zaman L_2'yi maksimize etmek L_1'i maksimize etmekle aynı olacaktır. Yani tek-biçimli (uniform) biçime sahip önsel dağılım parametre kestirimlerini belirli aralıklarla sınırlandıracaktır.

Test uygulandıktan sonra elde edilen veriler bu önsel dağılım bilgisi ile birleştirilir. İki dağılım bilgisinin birleşmesiyle sonsal (posterior) dağılım elde edilir. Bireyin bir özellik bakımından dağılım içerisindeki yerini belirlememiz sonsal dağılım bilgisiyle mümkün olmaktadır. Ancak önsel ve sonsal dağılımın ikisi de görelidir. Yani sonsal dağılımın ikinci bir kestirim sürecinde önsel dağılım olarak işlev görebileceği de unutulmamalıdır. Şekil 2'de gözlenen veri ile önsel dağılımın birleştirilmesi sonucunda sonsal dağılımın grafiği elde edilmiştir.



Şekil 2. Sonsal Bir Dağılımın Üretilmesi İçin Önsel Bir Dağılıma Olabilirlik Fonksiyonunun Eklenmesi (De Ayala, 2009, s.76.'dan uyarlanmıştır.)

Şekil 2'deki durumda en çok olabilirlik için (x=110) yanıt deseni kullanılmaktadır. Birim normal dağılımı önsel dağılım olduğunu varsayarsak ve bu dağılım bilgisini en çok olabilirlik fonksiyonuna dâhil edersek bireylerin sonsal dağılım üzerindeki yerleri konusundaki belirsizliği azaltmış olacaktır. Önsel dağılım, normal dağılımın dışında başka tanımlanmış dağılımlar olarak da belirlenebilir. Sonsal dağılım

sonucunda bireyin yeri konusunda dağılımın modu ya da ortalaması kullanılabilmektedir. Genel olarak, dağılımın ortalamasının kullanılması standart hatayı da en aza indirecektir. Burada elde edilen standart hata, en çok olabilirlik kestirimlerinden elde edilen hatalardan daha düşüktür. Ancak elde edilen ortalama değeri (bireyin yeri) önsel dağılıma doğru yaklaşacağından elde edilen kestirimlerin yanlı olduğu söylenmektedir. Şekil 1'de bireyin gerçek yerini Θ=1.0 olarak kabul edilirse; bununla birlikte şekilde bireyin en çok olabilirlik kestirimi sonucundaki yeri Θ=1.20 olacaktır. Bayes yaklaşımı sonucu sonsal dağılımdaki yeri ise $\hat{\theta}$=0.90 olarak hesaplanmıştır. Buna göre sonsal dağılımın ortalaması önsel dağılımın ortalaması tarafından çekilmiştir. Kestirim yanlılığı değişkendir ve önsel dağılım ile en çok olabilirlik ortalaması arasındaki farka bağlıdır. Bu fark arttıkça yanlılık miktarı da artacaktır. Bu nedenle, Swaminathan ve Gifford (1982) önsel dağılım kullanırken bu dağılımın ne spesifik ne de müphem bir dağılım olması gerektiği önermiş; bu dağılımın geniş bir ranja sahip olmasının yanlılık miktarını azaltacağını belirtmiştir.

Sonsal dağılımın ortalamasını ya da modunu kullanmaya bağlı olarak iki farklı Bayes yöntemi mevcuttur. Bunlar, MAP (Maximum A Posterior) ve EAP (Expected A Posterior)'tır. Sonsal dağılımın değişkenliği, kestirimlerin hatasının bir göstergesi olarak kullanılır. EAP ve MAP, en çok olabilirlik yöntemine göre, tam ve sıfır puan bireyler için ve tüm bireyler tarafından doğru yapılan ya da yanlış yapılan maddeler için kestirim yapmada kullanışlıdır. Bayes yaklaşımını kullanmanın avantajı tüm bireyler için sonlu bir Θ kestirimi sağlamasıdır. Dezavantajı ise bireyin Θ'sı ile önsel dağılımın ortalaması arasındaki farkın artması sonucu sonsal dağılımda meydana gelecek yanlılığın artmasıdır (De Ayala, 2009).

MAP, EAP'tan süreç olarak farklılık gösterir. MAP, en çok olabilirlik fonksiyonunda olduğu gibi iteratif bir süreçtir. Ancak EAP, iteratif değildir ve marjinal en çok olabilirlik yönteminde olduğu gibi özellik düzeylerine (quadrature noktaları) bağlı metotlara dayanır. MAP, sürekli bir önsel dağılım gerektirirken, EAP, ayrık bir önsel dağılım gerektirir. EAP kestirimlerinin standart hatası MAP kestirimlerine göre daha düşüktür. EAP'ın matematiksel eşitlikleri MAP'a göre daha basittir (Embretson ve Reise, 2000).

Markov Zinciri ve Monte Carlo Simülasyonu ve Gibbs Örneklemesi (Markov Chain Monte Carlo and Gibbs Sampling-MCMC)

Çalışmalarında bayesçi çıkarımı kullanan araştırmacılar, son 15 yılda Markov Zinciri Monte Carlo yöntemlerinin sağladığı olanaklardan giderek faydalanmaya başlamışlardır. Özellikle, Bayes yaklaşım çok boyutlu verilerden önsel dağılımın elde edilmesinde karmaşık integral hesaplamaları gerektirmektedir. İşte bu noktada, MCMC yöntemleri sonsal bir dağılımdan rastgele bir örneklem çekilmesi için basit bir yol sağlamaktadır (Martin ve Quinn, 2006).

MCMC, herhangi bir teorik çok değişkenli dağılımdan özellikle Bayes çıkarımının odak noktası olan çok değişkenli sonsal dağılımdan rastgele bir örneklemi simüle etme yöntemidir. Böylece teorik dağılımın özellikleri, rastgele örneklemin özelliklerinden kestirilebilir (Patz ve Junker, 1999).

Bayes istatistikleri çerçevesinde, model karmaşık olduğunda model parametrelerinin sonsal dağılımının hesaplanması çok zordur. Bu problem, madde parametrelerinin ve gizli değişkenlerin sonsal dağılımlarından simüle etmek için kullanılabilen MCMC yöntemleri kullanılarak çözülebilmektedir. Buna göre MCMC, karmaşık sonsal dağılımlardan örneklemler çekmek ve sonsal momentleri hesaplamak için kullanılmaktadır. MCMC yöntemlerinin ana özelliği, modelin karmaşıklığı arttığında basit ve uygulanması kolay olmalarıdır (Meng, Tao ve Chang, 2015).

MCMC yaklaşımının temelinde bir kestirim zincirinin oluşturulması ve her zincirden elde edilen değerlerin bir önceki zincire bağlanması yatmaktadır. Buradaki amaç kestirim zincirinden istenilen sonsal dağılıma yaklaşmaktır. Zincir oluşturulurken ilk adıma yakma (burn) adı verilir ve bu adım zincirden çıkarılır. Daha sonra belirtilen yineleme sayısı kadar örneklemler üretilir ve parametre kestirimi gerçekleştirilir (Patz ve Junker, 1999).

MCMC'nin basit olarak algoritması

$$\int_{\mathbf{S}} \mathbf{g}(\boldsymbol{\theta})\mathbf{p}(\boldsymbol{\theta})\mathbf{d}\boldsymbol{\theta} \cong \frac{\mathbf{1}}{\mathbf{n}} \sum_{\mathbf{t=1}}^{\mathbf{n}} \mathbf{g}(\boldsymbol{\theta}^{\mathbf{t}})$$

şeklindedir. $\mathbf{g}(\boldsymbol{\theta})$'nın ilgili fonksiyon kendisi ve $\boldsymbol{\theta}^{\mathbf{t}}$'nin *S* üzerinden sağlanan $\mathbf{p}(\boldsymbol{\theta})$'dan çekilen bir örneklem olduğudur. Buradaki amaç $\mathbf{g}(\boldsymbol{\theta})$'nın beklenen değerine yaklaştırmaktır (SAS Institute, 2020).

MCMC yöntemlerinden en çok kullanılan yöntemlerinden bir tanesi Gibbs örneklemesidir. Gibbs örneklemesi, sırayla karmaşık dağılımlardan örneklemler çekerek sonsal dağılıma uygun olarak parametre kestirimlerini gerçekleştirir. Buradaki sıradan kasıt sürecin yinelemeli (iteratif) olarak devam etmesidir. MCMC yöntemindeki Gibbs örneklemesi MTK'daki EM algoritmasına benzemektedir. Yani her adımda bir önceki adımda ürettiği parametreleri kullanarak bir sonraki adıma geçer ve bu parametre değerlerini günceller. Bu şekilde MCMC'nin oluşturduğu tek bir uzun zincirden sonsal dağılım parametrelerine ulaşılır (Geman ve Geman, 1984).

Gibbs örnekleme algoritması şu şekilde çalışmaktadır:$\boldsymbol{\theta} = (\boldsymbol{\theta_1}, \dots, \boldsymbol{\theta_k})'$ bir parametre vektörü, $\mathbf{p}(\mathbf{y}|\boldsymbol{\theta})$ olabilirlik ve $\boldsymbol{\pi}(\boldsymbol{\theta})$ önsel dağılım olarak kabul edelim. $\pi(\theta_i|\theta_j, i \neq j, y)$ ifadesinin koşullu sonsal dağılımının tamamı klasik sonsal yoğunluğu ile orantılıdır. Bu da $\boldsymbol{\pi}(\boldsymbol{\theta_i}|\boldsymbol{\theta_j}, \mathbf{i} \neq \mathbf{j}, \mathbf{y}) \propto \mathbf{p}(\mathbf{y}|\boldsymbol{\theta})\boldsymbol{\pi}(\boldsymbol{\theta})$ olarak ifade edilir. Örneğin, $\boldsymbol{\theta_j} = \boldsymbol{\theta_j^*}, \mathbf{2} \leq \mathbf{j} \leq \mathbf{k}$ şeklinde verilen $\boldsymbol{\theta_1}$'in tek boyutlu koşullu dağılımı şu şekilde hesaplanır:

$$\boldsymbol{\pi}(\boldsymbol{\theta_1}|\boldsymbol{\theta_j} = \boldsymbol{\theta_j^*}, \mathbf{2} \leq \mathbf{j} \leq \mathbf{k}, \mathbf{y}) = \mathbf{p}(\mathbf{y}|(\boldsymbol{\theta} = (\boldsymbol{\theta_1}, \boldsymbol{\theta_2^*}, \dots, \boldsymbol{\theta_k^*})')\boldsymbol{\pi}(\boldsymbol{\theta} = (\boldsymbol{\theta_1}, \boldsymbol{\theta_2^*}, \dots, \boldsymbol{\theta_k^*})')$$

Buradan Gibbs örneklemleyicisi de şu şekilde çalışır:

1. **t=0** olarak ayarlanır ve keyfi bir başlangıç noktası seçilir.

   $$\boldsymbol{\theta^{(0)}} = \left\{\boldsymbol{\theta_1^{(0)}}, \dots, \boldsymbol{\theta_k^{(0)}}\right\}$$

2. θ'nın her bir bileşenini aşağıdaki gibi oluşturun.
   - $\boldsymbol{\pi}\left(\boldsymbol{\theta_1}\middle|\boldsymbol{\theta_2^{(t)}}, \dots, \boldsymbol{\theta_k^{(t)}}, \mathbf{y}\right)$'den $\boldsymbol{\theta_1^{(t+1)}}$'i çekiniz.
   - $\boldsymbol{\pi}\left(\boldsymbol{\theta_2}\middle|\boldsymbol{\theta_1^{(t+1)}}, \boldsymbol{\theta_3^{(t)}}, \dots, \boldsymbol{\theta_k^{(t)}}, \mathbf{y}\right)$'den $\boldsymbol{\theta_2^{(t+1)}}$'i çekiniz.
   - ...
   - $\boldsymbol{\pi}\left(\boldsymbol{\theta_k}\middle|\boldsymbol{\theta_1^{(t+1)}}, \dots, \boldsymbol{\theta_{k-1}^{(t+1)}}, \mathbf{y}\right)$'den $\boldsymbol{\theta_k^{(t+1)}}$'i çekiniz.
3. $\mathbf{t} = \mathbf{t} + \mathbf{1}$ ayarlanır. Eğer $\mathbf{t} < \mathbf{T}$, istenilen örneklem sayısına ulaşılmışsa, 2. adıma geri dönülür aksi taktirde süreç durdurulur (SAS Institute, 2020).

İlgili Araştırmalar

Bu kısımda araştırma kapsamı boyunca değinilen çalışmalara ayrıntılı olarak yer verilmiştir. Buna göre ilgili çalışmalar, araştırmanın problem durumuna kaynaklık etmekte ve alanyazındaki eksikliğe dikkat çekmektedir. İlgili çalışmalarda parametre

kestirim yöntemlerine değinilirken bu kestirimleri etkileyen değişkenler dikkatle incelenmiş ve buna göre araştırmanın problemi ve sınırlılıkları belirlenmiştir. Bu konuda yurtiçindeki ve yurtdışındaki alanyazında bulunan çalışmalar kronolojik sırayla verilmiştir.

Lord (1968), VSAT verileri üzerinden klasik en çok olabilirlik yöntemini kullanarak 3 PL model üzerinde parametre kestirimleri gerçekleştirmiştir. Araştırmasının sonucunda örneklem büyüklüğünün 1000'e ve madde sayısının 50'ye ulaşmadığı durumlarda örnekleme hatasının yüksek değerler aldığını belirtmiştir.

Hulin, Lissak ve Drasgow (1982), 2 PL ve 3 PL modeller üzerinden Monte Carlo simülasyonu gerçekleştirmişlerdir. Örneklem büyüklüğü 200, 500, 1000 ve 2000 olan; test uzunluğu 15, 30 ve 60 maddelik veri setleri üretmişlerdir. 2 PL için 500 kişilik ve 30 maddelik bir grubun kestirim için yeterli büyüklük olduğunu ancak 3 PL modelde daha yüksek titizlikte kestirimler elde edebilmek için örneklem büyüklüğünün 1000 ve madde sayısının 60 olması gerektiğini vurgulamışlardır.

Goldman ve Raju (1986), araştırmalarında 1 PL ve 2 PL modeller kullanarak yetenek parametrelerinin kestirimlerinin doğruluğu üzerinde çalışmışlardır. Bunun için daha önceki bir tutum ölçeğinden elde edilen madde parametrelerini kullanmışlardır. Ayrıca, 250, 500 ve 1000 kişilik örneklemler üretmişler ve bu örneklemler üzerinden madde ve yetenek parametrelerini kestirmişlerdir. Tutum ölçeğinden elde ettikleri parametreler ile simülasyonlardan elde edilen parametreler kıyaslandıklarında 0.96 ile 1.00 arasında korelasyon değerleri bulmuşlardır. Buna göre araştırmalarının sonucunda 1 PL model için 250, 2 PL model için 500 kişilik örneklem büyüklüğünün yeterli olduğu sonucuna varmışlardır.

Swaminathan ve Gifford (1986), 3 PL model için parametre kestiriminde Bayes kestirim yöntemini önermişlerdir. Bu yöntemi, simülasyonla üretilmiş 100, 200 ve 400 kişilik örneklemlerde ve 25 ile 35 maddelik testlerde kullanmışlardır. Araştırmalarının sonucunda klasik en çok olabilirlik yöntemine göre ilerleme kaydettiklerini belirtmişlerdir.

Warm (1989), çalışmasında yetenek kestirimleri için en çok olabilirlik yöntemlerine dayandırdığı ağırlıklandırılmış en çok olabilirlik yöntemini kullanmıştır. Ağırlıklandırılmış en çok olabilirlik yönteminin normal dağılım altında en çok olabilirlik yöntemlerine göre yetenek parametresi kestirimlerinde daha az yanlı sonuçlar verdiğini belirlemiştir. Madde sayısı 10 ile 60 arasında değişen ve madde ayırt edicilik parametresi 1 ya da 2 olan 12 geleneksel test üzerinden en çok olabilirlik ve ağırlıklandırılmış en çok

olabilirlik yöntemlerini kullanarak kestirimler gerçekleştirmiştir. Buna göre ağırlıklandırılmış en çok olabilirlik yönteminde yetenek ölçeğinde daha küçük varyans değeri elde etmiştir. Ayrıca ağırlıklandırılmış en çok olabilirlik yönteminin, klasik en çok olabilirlik yöntemine göre parametre kestiriminde yanlılığı ve varyansı azalttığını belirtmiştir.

Seong (1990), araştırmasında önsel yetenek dağılımı sola çarpık, sağa çarpık ve normal olan 45 maddeden oluşan 35 veri seti üretmiştir. Bu veriler üzerinden örneklem büyüklüğüne ve kuadrat noktası sayısına göre marjinal en çok olabilirlik kestirimleri gerçekleştirmiş ve kestirimlerin titizliğini incelemiştir. Buna göre örneklem büyüklüğü artıkça ve önsel dağılım ile gerçek dağılım örtüştüğünde parametre kestirimlerinin titizliğinin arttığını belirtmiştir. Ancak aynı durumun küçük örneklemlerde madde parametrelerinin kestiriminde aynı titizliği sağlamamıştır. Bununla birlikte kuadrat noktalarının sayısı arttıkça örneklem büyüklüğünden ve önsel yetenek dağılımın özelliğinden bağımsız olarak kestirimlerin titizliği artış göstermiştir.

Harwell ve Janosky (1991), 2 PL model kullanarak 75 ile 1000 kişilik örneklemler ve 15 ile 75 maddelik test uzunluklarında veri setleri üretmişlerdir. Bayes kestirim yöntemini kullanarak gerçekleştirdikleri araştırmalarının sonucunda 15 maddelik ve 250 kişilik örneklemlerin yeterli olduğu sonucuna varmışlardır. Bu sonucun Drasgow (1989)'un sonuçlarıyla benzer olduğunu vurgulamışlardır. Drasgow (1989) çalışmasında ayrıca marjinal en çok olabilirlik yönteminin kullanımı için en az 200 kişilik örneklem büyüklüğünün gerekli olduğunu belirtmiştir.

Stone (1992), araştırmasında 10, 20 ve 40 maddeden, 250, 500 ve 1000 kişilik örneklemlerden oluşan ve gerçek yetenek dağılımı normal, çarpık ve basık olan 100 tekrarlı simülatif veriler üretmiştir. Çalışmasının amacı bu veri setleri üzerinden marjinal ve klasik en çok olabilirlik kestirimlerini, örneklem büyüklükleri ve test uzunlukları açısından 2 PL model üzerinde karşılaştırmaktır. Araştırmasının sonucunda küçük örneklemlerde, kısa testlerde ve farklı yetenek dağılım biçimlerinde madde güçlük parametrelerinin marjinal en çok olabilirlik kestiriminde daha tutarlı ve kesin sonuçlar verdiğini belirtmiştir. Ayrıca yetenek dağılımının normal olduğu durumlarda madde ayırt edicilik parametrelerinin marjinal en çok olabilirlik kestiriminde daha tutarlı ve kesin sonuçlar ürettiğini söylemiştir. Ancak dağılım çarpıklaştıkça, örneklem büyüklüğünden bağımsız olarak madde sayısı azaldıkça parametre kestirimlerinin yanlılığının arttığını belirtmiştir.

Kirisci, Hsu ve Yu (2001), araştırmalarında testlerin boyutluluğu (1ve 3 boyut), yetenek dağılımının biçimi (normal, pozitif kayışlı ve basık) ve üç farklı kestirim programı (Bilog, Multilog ve Xcalibre) kullanarak parametre kestirimlerinin titizliğini incelemişlerdir. Veri seti olarak 40 maddelik ve 1000 kişilik bir örneklem kullanmışlardır. Araştırmalarının sonucunda kestirim programı ile testin boyutluluğu arasında bir etkileşim olduğunu belirtmişlerdir. Bununla birlikte örneklem büyüklüğünün, test uzunluğunun ve kestirim programının yetenek dağılımın altında yatan dağılım biçiminde bir etkileşim göstermediğini bulmuşlardır. Ayrıca Bilog programının kestirimlerinin daha az hatalı olduğunu, madde ayırt edicilik parametresinin kestirimi dışında tek boyutluluğun ihlal edildiği durumlarda Multilog ve Xcalibre programlarının daha az varyans ürettiğini söylemişlerdir.

Doğan (2002), araştırmasında gerçek veri seti kullanarak normal, sola çarpık, sağa çarpık ve basık dağılımdan oluşan beş farklı örneklem üzerinden MTK'nın KTK'ya üstün olduğu varsayılan parametre değişmezliğinin sağlanıp sağlanmadığını incelemiştir. Araştırmasının sonucunda MTK'nın parametre değişmezliğinin dağılım biçimlerinden etkilendiğini ve hem MTK'nın hem de KTK'nın parametrelerindeki farklılaşmanın en fazla çarpık dağılımlarda oluştuğu sonucuna varmıştır.

Gao ve Chen (2005), tarafından yapılan araştırmada 3 PL model üzerinden geniş ölçekli bir simülasyon çalışması gerçekleştirilmiştir. Çalışmalarında örneklem büyüklüğü, 100, 200 ve 500 kişilik, test uzunluğu 10, 20 ve 60 maddelik tek biçimli (uniform) dağılan veri setleri kullanmışlardır. Veriler üzerinden marjinal en çok olabilirlik yöntemi ile Bayes kestirim yöntemlerini karşılaştırmışlardır. Araştırmalarının sonucunda marjinal en çok olabilirlik yönteminin küçük örneklemlerde gerçek madde parametre değerlerinin dışında kestirim yapma eğiliminde olduğu sonucuna varmışlardır. Ayrıca, örneklem büyüklüğünün 100 kişiye kadar düştüğü durumlarda, Bayes kestirimlerinin marjinal en çok olabilirlik kestirimlerine göre daha doğru sonuçlar verdiğini belirtmişlerdir. Kestirimlerin standart hatası (RMSE) bakımından Bayes kestirimlerinin sonuçlarının daha kabul edilebilir düzeyde olduğunu, ancak test uzunluğu ve örneklem büyüklüğünün artması durumunda marjinal en çok olabilirlik yöntemiyle kestirim sonuçlarının ve hata miktarlarının azalma eğiliminde olduğunu vurgulamışlardır.

Sass, Schimitt ve Walker (2005), çalışmalarında normal ve normal olmayan koşullar altında gizil özellik dağılımının kestirim hatalarını karşılaştırmışlardır. Simülatif olarak 1000 kişilik, 30 maddelik ve 2 PL modelde veri üretmişlerdir. Parametre kestirim yöntemleri olarak en çok olabilirlik (ML), MAP ve EAP kullanmışlardır. Ayrıca, örtük

özellik kestirim hatalarından madde parametresi kestirim hatalarını ayrıştırmak için hem gerçek hem de kestirilmiş madde parametrelerini incelemişlerdir. Araştırmanın sonucunda normal olmayan gizil özelik dağılımlarının normal olanlara göre daha yüksek kestirim hatası ürettiğini belirtmişlerdir.

Chuah, Drasgow ve Luecht (2006), çalışmalarında 300, 500 ve 1000 kişilik örneklem büyüklüklerinde 3 PL model üzerinden simülasyon çalışması gerçekleştirmişlerdir. Bilog programı gerçekleştirilerek yapılan araştırmanın sonucunda madde başına 300 katılımcıya kadar yetenek ve madde parametrelerinin kestirimi için yeterli büyüklükte olduğunu belirtmişlerdir. Ancak aynı durumun ayırt edicilik parametreleri için geçerli olmadığını, bu parametrenin kestirimlerinde bir şişkinlik oluştuğunu vurgulamışlardır.

Köse (2010), araştırmasında tek boyutlu ve çok boyutlu MTK modellerini, parametre kestirimleri ve model veri uyumu açısından test uzunluğu (12, 24) ve örneklem büyüklüğü (500, 1000, 1500) ile karşılaştırmıştır. Araştırma sonucunda çok boyutlu modellerin tek boyutlu modellere göre parametre kestirimlerinde daha az hata değerine sahip sonuçlar verdiğini belirtmiştir. Bununla birlikte örneklem büyüklüğünün madde parametreleri kestirimlerini etkileyen en önemli değişken olduğunu belirtmiştir. Ayrıca yetenek kestiriminde en önemli faktörün test uzunluğu olduğu bir diğer sonucu olarak bulunmuştur. Buna göre test uzunluğunun tek boyutlu MTK modellerinde model veri uyumunu sağlamada etkili olduğu sonucuna varmıştır. Örneklem büyüklüğü açısından tek boyutlu MTK modellerinde parametre kestirimlerinin çok boyutlu MTK modellerine göre daha fazla hata barındırdığını belirtmiştir.

Şahin (2012), Şahin ve Anıl (2017) araştırmalarında MTK'da test uzunluğunun ve örneklem büyüklüğünün model veri uyumu, madde parametreleri ve standart hata kestirimlerine etkisini incelemişlerdir. Bu amaçla, 150, 250, 350, 500, 750, 1000, 2000, 3000 ve 5000 örneklem büyüklükleri ile 10, 20 ve 30 maddelik testlerden madde parametre kestirimleri gerçekleştirmişlerdir. Araştırmanın sonuncunda, 10, 20 ve 30 maddelik testlerde 1 PL modelde 150 kişilik örneklemin; 2 PL modelde; 10 maddelik bir test için en az 750; 20 maddelik bir test için en az 500 ve 30 maddelik bir test için en az 250 kişilik örneklemlerin; 3 PL modelde 10 maddelik bir test için en az 750; 20 maddelik bir test için en az 750 ve 30 maddelik bir test için en az 350 kişilik örneklemlerin yeterli olabileceğini belirtmişlerdir. 10, 20 ve 30 maddelik testlerde en yüksek hata kestirimlerinin "c_i" parametresinde olduğunu bulmuşlardır. Tüm parametrelere ait ortalama standart hata kestirimlerinin örneklem büyüklüğünün artışından olumlu

etkilendiği ve örneklem büyüklüğünün artışından en olumlu etkilenen parametrenin "c_i" parametresi olduğu belirtilmiştir.

Koğar (2014), araştırmasında tek ve iki boyutlu yapılarda, farklı örneklem büyüklükleri, test uzunlukları ve boyutlar arası korelasyon değerlerinde, Tek Boyutlu Parametrik Olmayan Madde Tepki Kuramı (TBPOMTK), Tek Boyutlu Parametrik Madde Tepki Kuramı (TBPMTK) ve Çok Boyutlu Madde Tepki Kuramı (ÇBMTK)'ndan elde edilen madde parametreleri, maddelere ait model veri uyumları ve teste ait model veri uyumlarını belirlemiş ve sonuçları karşılaştırmıştır. Araştırmasının sonucunda örneklem büyüklüğü ve test uzunluğundaki artışın, testin model veri uyumunu arttığını belirtmiştir.

Finch ve Edwards (2015), araştırmalarında 2 PL modelde 3 farklı madde güçlük parametresi dağılımı, 5 farklı parametre kestirim yöntemi, 4 farklı örneklem büyüklüğü, 3 farklı test uzunluğu kullanarak 1000 tekrarlı simülatif veri oluşturmuşlardır. Araştırmanın amacı gizil özellik ve madde güçlük parametresi dağılımlarının normal olmadığı durumlarda marjinal en çok olabilirlik (MML), EAP, MAP, RC-IRT ve Bayes MCMC kestirimlerinin karşılaştırılmasıdır. Sonuç olarak simülasyonun tüm koşullarında gizil özelliğin ve madde güçlük parametresinin normal dağılmadığı durumlarda Bayes kestirim yönteminin daha titiz sonuçlar verdiğini belirtmişlerdir.

Reyhanlıoğlu Keçeoğlu (2018), araştırmasında farklı boyutluluk ve örneklem büyüklüklerinde MTK'da kestirilen parametrelerin değişmezliğini incelemiştir. Araştırmasının sonucunda tek boyutlu testin Tek Boyutlu Parametrik Olmayan Madde Tepki Kuramı (TBPOMTK)'nda madde parametresi kestirim ortalamalarının örneklem büyüklüğüne göre büyük değişimler göstermediğini belirtmiştir. Bununla birlikte örneklem büyüklüğü 200 olduğunda madde parametresi değişmezliği sağlanmıştır. Tek Boyutlu Parametrik Madde Tepki Kuramı (TBPMTK)'nda evren değere yakın madde parametresi kestirimleri için örneklem büyüklüğünün en az 1000 olması gerektiği bulunmuştur.

Çelikten ve Çakan (2019), araştırmalarında MTK'da, farklı örneklem koşullarında Bayesçi olmayan kestirim yöntemlerinden en çok olabilirlik, ağırlıklandırılmış en çok olabilirlik ve Bayesçi kestirim yöntemlerinden MAP ve EAP kullanılarak kestirilen yetenek puanlarına göre bireylerin sınıflandırılması sonucu elde edilen sınıflama doğruluğu ve sınıflama tutarlılığı indekslerini karşılaştırmışlardır. Araştırmanın sonucunda hem büyük hem de küçük örneklemlerde Bayesçi olmayan kestirim yöntemlerinden en çok olabilirlik yöntemi ile elde edilen yetenek kestirimlerine

dayalı olarak yapılan sınıflamanın diğer yöntemlere göre daha doğru ve tutarlı olduğu belirtilmiştir.

Karadavut (2019), çalışmasında önsel yetenek dağılımının normal ya da tek biçimli olduğu durumlarda parametre kestirimlerini incelenmiştir. Bunun için, 600 ile 2000 kişilik örneklem büyüklükleri ve 15 ile 30 maddelik test uzunlukları kullanarak 25 veri setini simülatif yöntemle üretmiştir. Sonuç olarak, 2 PL ve 3 PL modeller için yeteneğin gerçek dağılımı bilinmediğinde yetenek parametresinin doğru kestirimleri için önsel dağılım olarak tek biçimli dağılımın kullanılması gerektiğini belirtmiştir. Ancak Rasch modelde daha doğru parametre kestirimi için bir model önerilememiştir.

Kıbrıslıoğlu Uysal (2020), araştırmasında parametrik ve parametrik olmayan MTK modellerini farklı gizil özellik dağılımlarına, madde güçlük dağılımlarına, madde sayısına ve örneklem büyüklüğüne göre karşılaştırmıştır. Kestirim yöntemi olarak en çok olabilirlik (ML), maksimum sonsal (MAP), Davidian eğrileri, Ampirik histogram ve parametrik olmayan Bayesçi yöntemler kullanılmıştır. Araştırmasının sonucunda kestirim yöntemi, örneklem büyüklüğü, madde sayısı ve örtük özellik dağılımı ve bu koşulların birlikte etkileşimleri hem madde hem de yetenek parametreleri üzerinde etkili olduğunu belirtmiştir. Buna göre, gizil özellik dağılımı çarpıklaştığında en az hatalı kestirimler Davidian eğrileri yöntemi ile elde edilmiştir. Genel olarak parametrik olmayan Bayesçi yöntem en yüksek hatalı kestirimleri üretmiştir. En az hatalı kestirimler ise Davidian eğrileri ve Ampirik histogram yöntemleri ile elde edilmiştir. Ayrıca, b parametresinin dağılım biçimi yetenek kestirimleri üzerinde anlamlı olmadığı bulunmuş ve gizil özelliğin çarpık dağıldığı durumlarda madde parametre kestirimleri etkilenmiştir.

İlgili araştırmalar incelendiğinde alanyazında MTK parametre kestirimleri üzerinde farklı kestirim yöntemleri farklı koşullar altında karşılaştırılmıştır. En çok karşılaştırmanın örneklem büyüklüğü ve test uzunluğu koşulları altında gerçekleştiği, kestirim yöntemi olarak ise en çok olabilirlik (ML), MAP ve EAP'ın kullandığı görülmektedir. Gizil özelliğin ya da madde parametrelerinin dağılımlarının etkisinin araştırıldığı çalışmalar da mevcuttur. Çalışmalar genellikle simülatif veriler üzerinden gerçekleştirilmiştir.

Bu araştırmanın amacı değişimlenen koşulların etkisiyle yukarıdaki çalışmalarda olduğu gibi farklı örneklem büyüklükleri, farklı test uzunlukları, farklı gizil özellik dağılımları ve farklı parametre kestirim yöntemlerini karşılaştırmaktır. Araştırma bu yönüyle diğer çalışmalarla benzerlik göstermektedir. Ancak bu araştırmayı diğer araştırmalardan farklı kılan en önemli durum önsel yetenek dağılımının beş farklı

biçiminin üretilmesi ve buna bağlı olarak toplam test puanlarında da bu dağılım biçimine sahip olmasıdır. Bununla birlikte gizil dağılımın çarpık olduğu durumların incelendiği bazı çalışmalar mevcuttur. Bu araştırmada ise çarpıklık durumu ikili yönlü (sola ve sağa çarpık) olarak ele alınmış ve ek olarak sivri ve basık dağılım durumları da incelenmiştir. Ayrıca çalışmaların bazılarında bayes kestirimler genellikle Rasch ya da 2 PL modelde incelenmiştir. Bu araştırmada ek olarak 3 PL modelde Bayes MCMC parametre kestirimin sonuçları da incelenmiştir.

BÖLÜM 3

YÖNTEM

Bu bölümde araştırmanın modeline, verilerin üretilmesine, simülasyon koşullarına, verilerin çözümlenmesine ve simülasyonun geçerliği ile ilgili ayrıntılı bilgilere değinilmiştir.

Araştırmanın Modeli

Bu araştırmada, simülasyon koşullarına uygun olarak farklı önsel yetenek dağılım biçimlerine sahip veri setleri oluşturulmuş ve bu veri setleri üzerinden klasik (joint MLE) en çok olabilirlik ve (standart) Bayesçi (MCMC) yöntemler kullanılarak yetenek ve madde parametrelerinin kestirimleri yapılmıştır. Bu amaçla; normal ve normal olmayan (sola çarpık, sağa çarpık, sivri ve basık) olmak üzere beş farklı önsel yetenek dağılım biçiminde, kestirim yöntemleri sabit tutularak; örneklem büyüklüğü, madde sayısı ve modelin kestirilen parametre sayısı değişimlenerek simülasyon koşulları oluşturulmuştur. Araştırma, kuramsal katkı sağlaması açısından belirli koşulların parametre kestirimleri üzerindeki etkisini belirlemek amacıyla simülatif veriler üzerinde yürütülmüştür.

Bazı araştırmacılar, simülasyon çalışmalarını gerçek dünyanın, bir sistemin ya da bir sürecin matematiksel olarak modellenmesi ve bilinen bir gerçek ile ilgili olarak çeşitli istatistiksel yöntemlerin performansını değerlendirmek olarak tanımlanmaktadır. Bu yönüyle simülasyon çalışmalarının bazıları temel araştırma olarak değerlendirilmektedir (Harwell, Kohli ve Peralta, 2017). Ayrıca simülasyon araştırmalarının kuramsal problemlerin çözümünde yöntemsel olarak güçlü çalışmalar oldukları belirtilmektedir (Morris, White ve Crowther, 2019). İlgili alanyazında, simülasyon çalışmalarının temel araştırma olarak tanımlandığı çalışmalar bulunmaktadır (Çakıcı Eser, 2015; Kilmen, 2010; Koğar, 2014; Köse, 2010; Sayın, 2014; Sünbül, 2011; Şahin, 2012; Şengül Avşar, 2015).

Aynı zamanda simülasyon çalışmalarında belirli değişkenlerin araştırılmasında simülatif ortamlarda üretilmiş veriler kullanabilmektedirler. Simülasyon yaklaşımı, içinde ilgili bilgilerin ve verilerin üretilebileceği yapay bir ortamın oluşturulmasını içerir.

Bu, kontrollü koşullar altında bir sistemin (veya alt sisteminin) dinamik davranışını gözlemlemenizi sağlar (Fraenkel ve Wallen, 2009; Kothari, 2004). Alanyazında, simülasyon çalışmalarının birer ampirik deney olduğu (Morris ve diğerleri, 2017) ve belirlenmiş araştırma tasarımı ve veri analizi ilkelerine bağlı olarak istatistiksel örnekleme deneyleri olarak ele alınması gerektiği savunulmaktadır (Hoaglin ve Andrews, 1975). Buna göre bu araştırmada, simülatif üretilmiş veriler aracılığıyla çeşitli koşullar değişimlenerek farklı kestirim yöntemlerinin karşılaştırılması açısından istatistiksel anlamda deneysel bir yöntem izlenmektedir. Bu yönüyle bu araştırma deneysel bir çalışmadır.

Verilerin Üretilmesi

Bu çalışma kapsamında veriler, araştırma problemlerinde belirtilen; önsel yetenek dağılım biçimi, örneklem büyüklüğü, madde ve model parametre sayısına göre farklı şekillerde değişimlenen koşullara uygun bir şekilde Monte Carlo Simülasyonu ile üretilmiştir. Monte Carlo (MC) simülasyonu, MTK parametre kestiriminde yeni yöntemlerin değerlendirilmesi, farklı madde analiz programlarının kıyaslanması, çok boyutlu verilerde parametre kestirimi gibi birçok uygulamada kullanılmaktadır. Buna göre, Monte Carlo simülasyon tekniği ile gerçekleştirilen MTK uygulamaları aşağıda verilenlerden en az birini içermelidir (Harwell ve diğerleri, 1996):

1. Parametre iyileştirme veya parametre kestirim yöntemlerinin değerlendirilmesi,
2. MTK temelli istatistiklerin özelliklerinin değerlendirilmesi,
3. Farklı MTK uygulamalarını birleştirerek yöntemsel karşılaştırma yapılması.

Simülasyon çalışmaları, bilgisayar teknolojilerindeki ilerlemelerle, istatistiksel problemleri çözmeye ilişkin popüler ve formal bir araştırma yöntemi haline gelmiştir (Harwell ve diğerleri, 1996). Veri üretiminde simülasyon çalışmaları için 1970'lerden bu yana birçok bilgisayar uygulaması geliştirilmiştir. DATAGEN (Hambleton ve Rovinelli, 1973), GENIRV (Baker, 1989), RESGEN (Muraki, 1992) bunlardan bazılarıdır (akt: Han, 2007).

Bu araştırmada, simülasyon koşullarına bağlı olarak verilerin üretilmesinde R yazılım dili kullanılmıştır. R içerisinde ise *mirt* (Chalmers, 2012), *e1071* (Meyer, 2022), *psych* (Revelle, 2022) ve *lattice* (Sarkar, 2022) paketleri çalıştırılmıştır. *Mirt* (Chalmers, 2012) paketi içerisinde *simdata* fonksiyonu kullanılarak simülasyon koşullarına göre ikili

(1-0) puan matrisleri araştırmacı tarafından yazılan *önsel* fonksiyonuyla üretilmiştir. İkili puan matrisleri üretilirken *simdata* fonksiyonu içerisinde bulunan gizil dağılım argümanına, dağılım biçimlerine uygun olarak üretilen önsel yetenek puanları yerleştirilmiştir. Veri dağılımlarının araştırma problemlerinde belirtilen çarpıklık ve basıklık değerlerine uygun olup olmadığı *f.skew.kurt* (Demir, 2019) fonksiyonu ile belirlenmiştir. Önsel yetenek puanları dağılım durumuna göre üretilirken çarpık veya basık dağılım üreten bir R paketi bulunmamaktadır. Bu nedenle araştırmacı kendi oluşturduğu kodlar yardımıyla iki farklı (normal ve tek-biçimli) dağılım türünü bir araya getirerek çarpık ve basık dağılımları üretmiştir. Bu dağılımları üretirken kullanılan ve araştırmacı tarafından yazılan *önsel** fonksiyonu şu şekildedir:

```
#Simülasyon koşullarına göre gerekli önsel dağılımların ve veri setlerinin üretilmesi

library(psych)
library(e1071)
library(mirt)

#I: madde sayısı
#N: birey sayısı
#M: parametre sayısı
#D: dağılım durumu

önsel <- function (I, N, M=c("2PL", "3PL"), D=c("normal","sola çarpık", "sağa çarpık","sivri","basık"))
{
  a <- rlnorm(I, meanlog = 0.3, sdlog = 0.2)
  b <- rnorm(I, mean = 0, sd = 1)
  c <- runif(I, min = 0.01, max = 0.25)

     if (D=="normal")

     {k <- as.matrix(rnorm(N, mean = 0, sd = 1))}

 else if (D=="sola çarpık")

     {k <- as.matrix(c(rnorm(N*86/100, 2, 1), runif(N*7/100, min = -5, max = -4), runif(N*7/100, min = -4, max = -3)))}

else if (D=="sağa çarpık")

     {k <- as.matrix(c(rnorm(N*86/100, -2, 1), runif(N*7/100, min = 3, max = 4), runif(N*7/100, min = 4, max = 5)))}

else if (D=="sivri")

     {k <- as.matrix(c(rnorm(N*3/100, -1, 100), rnorm(N*94/100, 0, 0.00001), rnorm(N*3/100, 1, 100)))}

else if (D=="basık")
```

```
     {k <- as.matrix(c(rnorm(N*40/100, 0, 1), runif(N*30/100, min = -3,
     max = -1), runif(N*30/100, min = 1, max = 3)))}

     if (M=="2PL")

{dat <- as.data.frame(simdata(a = a, d = b, N = N, itemtype = "dich",
Theta = k))

  model2pl <- mirt(data = dat, 1, itemtype = "2PL", SE = TRUE, verbose
= FALSE, technical = list(NCYCLES = 10000))

  irt.parameters <- as.data.frame(coef(model2pl, simplify = TRUE)$items)
  bias.a <- mean(irt.parameters[,1]-a)
  bias.b <- mean(irt.parameters[,2]-b)
  rmse.a <- sqrt(mean((irt.parameters[,1]-a)^2))
  rmse.b <- sqrt(mean((irt.parameters[,2]-b)^2))

  fit2pl <- M2(model2pl)
  M2 <- fit2pl$M2
  p <- fit2pl$p

  data <- list(dat, bias.a, rmse.a, bias.b, rmse.b, M2, p, k)

  print(data)}

  else if (M=="3PL")

  {
  dat <- as.data.frame(simdata(a = a, d = b, guess = c, N = N, itemtype
= "dich", Theta = k))

  model3pl <- mirt(data = dat, 1, itemtype = "3PL", SE = TRUE, verbose
= FALSE, technical = list(NCYCLES = 10000))

  parameters <- as.data.frame(coef(model3pl, simplify = TRUE)$item)
  bias.a <- mean(parameters[,1]-a)
  bias.b <- mean(parameters[,2]-b)
  bias.c <- mean(parameters[,3]-c)
  rmse.a <- sqrt(mean((parameters[,1]-a)^2))
  rmse.b <- sqrt(mean((parameters[,2]-b)^2))
  rmse.c <- sqrt(mean((parameters[,3]-c)^2))

  fit3pl <- M2(model3pl)
  M2 <- fit3pl$M2
  p <- fit3pl$p

  data <- list(dat, bias.a, rmse.a, bias.b, rmse.b, bias.c, rmse.c, M2,
p, k)

  print(data)}}
```

**Bu fonksiyonun bazı kısımlarında Bulut ve Sünbül (2017)'ün kodlarından faydalanılmıştır.*

Alanyazın incelendiğinde yetenek ve madde parametrelerinin üretilmesinde birçok dağılım biçiminin kullanıldığı görülmektedir. Bu araştırmada, araştırmacı tarafından üretilen *önsel* fonksiyonunda veriler üretilirken başlangıç madde parametreleri bu çalışmalara göre; madde ayırıcılık parametresi log-normal dağılım [$a \sim lnN(0.3, 0.2)$];

madde güçlük parametresi standart normal dağılım [$b \sim N(0, 1)$], madde şans parametresi (düşük asimptot) tek-biçimli dağılım [$c \sim U(0.01, 0.25)$] kullanılmıştır (Baker, 2001; Feinberg ve Rubright, 2016; Soysal, 2017; Bulut ve Sünbül, 2017; Pekmezci, 2018).

Önsel yetenek parametreleri üretilirken birden fazla ve farklı (normal ve tek-biçimli) dağılım türü birleştirilmiştir. Bunun nedeni, dağılımların normalden farklılaşmasında uç değerlerin etkili olmasıdır. Uç değerler, normal olmayan dağılımların bir parçası olarak dağılımın kuyruk kısımlarında toplanmaktadırlar (Tabachnick ve Fidell, 2014). Eğer bir dağılım, normal dağlıyorsa, z puanları ±3 arasında dağılım göstermektedir. Çünkü bunun nedeni puanların %99'nun bu aralıkta yer almasıdır. Ancak, küçük örneklemlerde (N≤100) puanlar ±3 aralığının dışında yer alıyorsa bu puanlar uç değer oluşturmaktadır. Bununla birlikte büyük örneklemlerde (N>100) bu aralık değerleri uç değerlerinin belirlenmesinde yeterli olmamaktadır. Bu nedenle büyük örneklemlerde (N>100) verilerin normal dağılım için Z puan aralığı ±4 arasında kabul edilmektedir (Stevens, 2002). Buna göre, normal dağılımın dışındaki çarpık, sivri ve basık dağılımların oluşturulmasında ±4'ün üzerindeki Z puanlarında uç değerler üretilmiştir. Buna göre dağılım normalse $\Theta \sim N(0, 1)$; sola çarpıksa $\Theta \sim N(2, 1)$, $\Theta \sim U(-5.0, -4.0)$ ve $\Theta \sim U(-4.0, -3.0)$; sağa çarpıksa $\Theta \sim N(-2, 1)$, $\Theta \sim U(3.0, 4.0)$ ve $\Theta \sim U(4.0, 5.0)$; sivriyse $\Theta \sim N(-1, 100)$, $\Theta \sim N(1, 100)$ ve $\Theta \sim N(0, 0.00001)$; basıksa $\Theta \sim N(0, 1)$, $\Theta \sim U(-3.0, -1.0)$ ve $\Theta \sim U(1.0, 3.0)$ dağılım türleri kullanılmıştır. Aynı zamanda istenilen basıklık ve çarpıklık durumlarına göre örneklem büyüklüğü, madde sayısı ve model parametre koşulları dikkate alınarak bu dağılımların fonkisyonda bulunan oranlarında değişiklikler yapılmıştır.

Simülasyon Koşulları

Simülasyon koşullarının belirlenmesinde alanyazındaki diğer çalışmaların incelenmesi ve kuramsal alana katkı sağlaması açısından hangi faktörlerin seçilmesi gerektiği önemli bir durumdur. Bu araştırmada problemlerin çözümü için oluşturulan simülasyon modelinde bazı koşullar sabit tutulurken bazıları değişimlenmiştir. Araştırmada yer alan koşulların seçiminde alanyazındaki benzer çalışmalar incelenerek belirlenmiştir. Sabit tutulan ve değişimlenen koşullar aşağıda verilmiştir:

Sabit Tutulan Koşullar

Parametre Kestirim Yöntemleri. Yetenek ve madde parametrelerinin kestiriminde klasik en çok olabilirlik (Joint ML) ve Bayesçi MCMC yöntemleri işe koşulmuştur. Bu kestirim yöntemleri tüm simülasyon koşulları ve tekrarları için ayrı ayrı kullanılmıştır. Ayrıca bu koşul araştırmanın yanıt aradığı en önemli problemlerden biridir.

Değişimlenen Koşullar

Örneklem Büyüklüğü. Simülasyonun her koşulu için 100, 500 ve 1000 kişilik olmak üzere üç farklı örneklem büyüklüğü belirlenmiştir. MTK kestirimleri için örneklem büyüklüğü önemli bir faktör olarak kabul edilmektedir (Hambleton, 1989; Orlando, 2004). MTK'nın güçlü yanlarının ortaya çıkarılabilmesi örneklem büyüklüğüne bağlıdır ve geniş örneklemlerde çalışılması gerektiği önerilmektedir (DeMars, 2010). Linacre (1994), modeldeki parametre sayısı az olduğunda küçük örneklemlere, daha karmaşık modellerin ise daha büyük örneklemlere ihtiyacı olduğunu belirtmiştir. İlgili alanyazında, 1PL model için 200 (Wright ve Stone, 1979) ya da 500 (Hulin ve diğerleri, 1982), 2PL model için 1000 (Ree ve Jensen, 1980) ve 3PL model için 1000 (Lord, 1968) ya da 10000 ve daha fazla (Thissen ve Wainer, 1983) kişiden oluşan örneklem büyüklüklerinin yeterli olduğunu belirten çalışmalar yapılmıştır. Ayrıca, De Ayala (2009), 250 veya 500 kişilik örneklem büyüklüklerinin parametre kestirimi için yeterli olduğunu belirtirken, Hulin ve diğerleri (1982), genel olarak en çok olabilirlik yöntemlerinin kullanıldığı parametre kestirimleri için örneklem büyüklüğünün 2000'den fazla olmasının gereksiz olduğu sonucuna varmıştır. Mislevy (1986), Bayes yaklaşımla parametre kestirimi yaptığı çalışmasında 1000 kişilik bir örneklem kullanmıştır. Bu çalışmada farklı örneklem büyüklükleri kullanılarak Bayes yaklaşımının avantajlarından faydalanılmak istenmektedir. Bu nedenle küçük bir örneklem büyüklüğü için 100; orta büyüklükte bir örneklem için 500 ve büyük bir örneklem büyüklüğü için 1000 kişilik veri setleri kullanılmıştır.

Madde Sayısı (Test Uzunluğu). Simülasyonun her koşulu için 10, 20 ve 40 maddelik olmak üzere üç farklı test uzunlukları belirlenmiştir. Farklı sayıda madde

kullanılması madde yanıt desenlerindeki örüntülerin de farklılaşmasına neden olmaktadır. Bu farklılaşma özellikle madde parametrelerindeki kestirimlerin doğruluğu açısından önemlidir (Hulin, Lissak ve Drasgow, 1982). Madde sayısı artığında, Ɵ kestirimlerinin titizliği artmaktadır. Buna göre örneklem büyüklüğünün ve madde sayısının artırılması madde parametrelerinin titizliğinin artmasına ve de böylece Ɵ kestirimlerinin hassasiyetinin artmasına neden olacaktır (Reise ve Yu, 1990). DeMars (2010), 2 PL ve 3 PL modeller için 500 kişilik bir örneklem kullanıldığında madde sayısının 20; 1000 kişilik bir örneklem kullanıldığında madde sayısının 40, 2000-3000 kişilik örneklem kullanıldığında madde sayısının 50 ila 80 olması gerektiğini belirtmiştir. Hulin ve diğerleri (1982), 2 PL modellerde 500 kişilik bir örneklemde 30 maddelik, 3 PL modellerde 1000 kişilik bir örneklemde 60 maddelik bir testin kullanılmasının parametre kestirimlerinin doğruluğu açısından yeterli olacağını söylemektedir. Hambleton ve Cook (1983), 3 PL modelde 500 kişilik bir örneklemde 20 maddelik bir testin parametre kestiriminde yeterli olduğunu belirtmişlerdir. Ancak madde sayısı 40'a çıktığında kestirimlerin hatasının olumsuz etkilendiğini söylemişlerdir. Akour ve Al-Omari (2013), 200 kişilik örneklemde 15 maddelik test uzunluğunun 3 PL modelde parametre kestirim için yeterli bir büyüklük olduğunu belirtmişlerdir. Mislevy (1986), Bayes yaklaşımla parametre kestirimi yaptığı çalışmasında test uzunluğu olarak 20 ve 40 madde kullanmıştır. Bu araştırmada kısa testler için 10 madde, orta uzunluktaki testler için 20 madde ve daha uzun testler için 40 maddelik veri setleri üretilmiştir. Kısa testler, çoğunlukla sınıf içi uygulamalarda öğretmen yapımı testler olmakla birlikte artık bu testler Türkiye'de yapılan ortaöğretime geçiş sınavlarında da kullanılmaktadır. Bu sınavlarda, TC. İnkılap Tarihi ve Atatürkçülük, Din Kültürü ve Ahlak Bilgisi ve Yabancı Dil alt testlerinde madde sayısı 10'dur (MEB LGS Kılavuzu, 2022). Bu amaçla simülasyon koşullarından biri olan madde sayısı (test uzunluğu) olarak 10 madde tercih edilmiştir.

Önsel Yetenek (Theta) Dağılımı. Simülasyonun her koşulu için standart sapma değerleri sabit tutularak beş farklı dağılım biçimi kullanılmıştır. İlgili koşullar normal ve normal olmayan (sola çarpık, sağa çarpık, sivri ve basık) dağılım olacak şekilde seçilmiştir. Çarpıklık katsayısının mutlak değerinin; 1.00'den büyük olduğu durumda örneklemlerin dağılım biçimlerinin yüksek düzeyde çarpık, 0.50 ve 1.00 arasında orta düzeyde çarpık, 0.50'den küçük olduğu durumlarda ise yaklaşık olarak simetrik olduğunu

belirtilmektedir. Basıklık için ise katsayının 3 olması durumunda dağılımın normal, 3'ten büyük olması durumunda sivri, 3'ten küçük olması durumunda ise basık olduğu söylenmektedir (Bulmer, 1979). Ancak formüle -3'ün eklenmesiyle bu referans değer 0 olmaktadır. Yani basıklık katsayısının 0 olması dağılımın normal, 0'dan büyük olması dağılımın sivri, 0'dan küçük olması dağılımın basık olduğunu göstermektedir. Tabachnick and Fidell (2014), çarpıklık ve basıklık değerlerinin -1.50 ile +1.50 aralığında olduğu zaman dağılımın normal olduğu kabul edildiğini belirtmiştir. Çarpık dağılımların normal dağılım varsayımlarıyla değerlendirilmesi ise, yanlış sonuçlara yol açmaktadır (Kolen, 1985). Normal dağılımdan sapmaların en çok olabilirlik kestirim yöntemleri ile parametre kestirimi yaparken çeşitli sorunlara neden olduğu bilinmektedir (Hambleton ve Swamimnathan, 1985). Bu nedenle farklı önsel yetenek dağılım biçimlerinin parametre kestirim yöntemleri üzerinde nasıl bir etki yaratacağı araştırmanın problemlerini oluşturmaktadır.

MTK Modeli. Bu çalışmada parametre kestirimlerine yönelik olarak 2 PL ve 3 PL modelleri kullanılmıştır. Hulin ve diğerleri (1982)'ye göre bu modeller güçlü ve en yaygın kullanılan modellerdir.

Buna göre simülasyon koşulları belirlenirken iki farklı kaynak göz önüne alınmıştır. Bunlardan birincisi her bir koşul belirlenirken alanyazındaki benzer çalışmalardan faydanılması, ikincisi ise araştırmanın amacına bağlı olarak Bayes kestirim yönteminin sağlacağı avantajların dikkate alınmasıdır.

Birinci kaynakta alanyazınla ilgili araştırmalar, her bir koşulun başlığı altında detaylı bir şekilde açıklanmıştır. İkinci kaynakta ise en çok olabilirlik kestirimlerinin dezavantajları ve Bayes kestirimin avantajları birlikte düşünüldüğünde bu koşullar belirlenmiştir. Özellikle örneklemin küçülmesi, madde sayısının azalması ve önsel yetenek dağılımın çarpıklaşması durumlarında klasik ve Bayes kestirim sonuçlarının nasıl bir değişim göstereceği bu araştırmanın amacı olduğundan simülasyon koşullarının bu şekilde seçilmesinde diğer bir önemli etkendir.

Buna göre araştırmada kullanılan veri setleri için araştırma koşulları Tablo 1'de verilmiştir.

Tablo 1

Araştırmanın Koşulları

Araştırmanın Koşulları				
Sabit tutulan koşullar	Değişimlenen koşullar			
Parametre kestirim yöntemleri (x2)	Örneklem büyüklüğü (x3)*	Madde sayısı (x3)*	Önsel yetenek dağılım biçimi (x5)*	Model türü (x2)*
Klasik en çok olabilirlik (ML) Bayes (MCMC)	100 500 1000	10 20 40	Normal Sola çarpık Sağa çarpık Sivri Basık	2PL 3PL

*Her koşul 100 defa tekrar edilmiştir.

Tablo 1 incelendiğinde, araştırma koşullarının sabit ve değişimlenen koşullardan oluştuğu görülmektedir. Sabit koşullar, kestirim yöntemleri; değişimlenen koşullar, örneklem büyüklüğü, madde sayısı, önsel yetenek dağılım biçimi ve MTK modeli olarak belirlenmiştir. Buna göre, örneklem büyüklüğü (100 x 500 x 1000), madde sayısı (10 x 20 x 40), önsel yetenek dağılım biçimi (normal x sola çarpık x sağa çarpık x sivri x basık), model parametre sayısı (2 PL x 3 PL) olmak üzere 90 adet simülasyon koşulu 100 tekrarla gerçekleştirilmiştir.

Alanyazın incelendiğinde benzer simülasyon çalışmalarında veriler üretilirken farklı tekrar (replikasyon) sayıları kullanıldığı görülmektedir. Bu durumu etkileyen iki faktör bulunmaktadır. Bunlardan birincisi, az sayıda tekrar sonucu üretilen verilerin titizlik derecesinin yetersiz olması, ikincisi ise çok sayıda tekrar sonucu simülasyon programının yetersizliği ve zamansal yönden maliyetli olması durumlarıdır (Bulut ve Sünbül, 2017).

Harwell ve diğerleri (1996), MTK parametre değerlerinin değişimlendiği araştırmalarda en az 25 replikasyon kullanılmasını önermişlerdir. Bununla birlikte, ilgili araştırmalarda verilen çeşitli simülasyon çalışmalarında Seong (1990) 5 tekrar, Stone (1992) 100 tekrar, Kirisci ve diğerleri (2001) 10 tekrar, Sass ve diğerleri (2008) 100 tekrar, Finch ve Edwards (2015) 1000 tekrar, Bulut ve Sünbül (2017) 100 tekrar, Karadavut (2019) 25 tekrar, Kıbrıslıoğlu Uysal (2020) 100 tekrar kullanmışlardır.

Buna göre alanyazın dikkate alınarak hem titiz sonuçların üretilmesi hem de simülasyon süresinin artmaması açısından bu araştırmada 100 tekrar sayısının kullanılması uygun görülmüştür.

Verilerin Çözümlenmesi

Araştırma problemlerine yanıt ararken her bir soru tek tek ele alınmış ve elde edilen sonuçlar, bulgular ve yorum bölümünde başlıklar halinde sunulmuştur. Verilerin üretilmesinde olduğu gibi verilerin çözümlenmesinde de R programlama dili kullanılmıştır. Kullanılan R program sürümü R Studio, Version: 2022.12.0+353'dur.

Verilerin MTK Varsayımlarının İncelenmesi

İlk olarak üretilen veri setlerinin MTK parametre kestirimlerine uygunluğunu belirlemek amacıyla temel varsayımlar test edilmiştir. Bu varsayımlar tek boyutluluk, yerel bağımsızlık ve model-veri uyumudur (Baker, 2001; Baker ve Kim, 2004; De Ayala, 2009; Hambleton ve Swaminathan, 1985; Hambleton, Swaminathan ve Rogers, 1991; Hambleton ve Jones, 1993). Temel varsayımlar test edilirken *psych* (Revelle, 2022), *sirt* (Robitzsch, 2022) ve *mirt* (Chalmers, 2012) paketleri kullanılmıştır. Tek boyutluluğun test edilmesi için veriler üzerinden tetrakorik korelasyon matrisleri oluşturulmuş ve bu matrisler üzerinden açımlayıcı faktör analizi gerçekleştirilmiştir. Açımlayıcı faktör analizi sonuçlarına göre açıklanan varyans oranı, birinci özdeğerlerin ortalamaları ve birinci özdeğerin ikinci özdeğere oranı hesaplanmıştır. Başat bir faktörün bulunması durumunda bu varsayımın karşılandığı kabul edilmiştir. Tablo 2'de önsel yetenek dağılımı normal olan verilere ait açımlayıcı faktör analizi sonuçları verilmiştir.

Tablo 2

Önsel Yetenek Dağılımı Normal Olan Verilere Ait Tek Boyutluluk Analizi Sonuçları

Önsel Yetenek Dağılımı	Madde Sayısı	Örneklem Büyüklüğü	MTK Modeli	Açıklanan Varyans Ortalaması*	Birinci Özdeğerlerin Ortalaması	λ_1/λ_2**
Normal	10	100	2PL	0.380	3.796	3.420
		500		0.393	3.934	4.971
		1000		0.385	3.854	5.196
	20	100		0.377	7.536	4.608
		500		0.388	7.754	7.493
		1000		0.390	7.803	8.542
	40	100		0.353	14.139	6.177
		500		0.386	15.447	11.331
		1000		0.387	15.470	13.581
	10	100	3PL	0.302	3.018	2.592
		500		0.296	2.962	3.518
		1000		0.304	3.043	3.958

(Devam ediyor)

Tablo 2 (Devam)

Önsel Yetenek Dağılımı Normal Olan Verilere Ait Tek Boyutluluk Analizi Sonuçları

Önsel Yetenek Dağılımı	Madde Sayısı	Örneklem Büyüklüğü	MTK Modeli	Açıklanan Varyans Ortalaması*	Birinci Özdeğerlerin Ortalaması	λ_1/λ_2**
Normal	20	100		0.294	5.876	3.421
		500		0.306	6.114	5.503
		1000		0.300	5.990	6.045
	40	100		0.276	11.048	4.504
		500		0.298	11.931	8.127
		1000		0.297	11.896	9.557

*Birinci faktörün açıkladığı varyans oranı
**Birinci özdeğerin ikinci özdeğere oranı

Tablo 2 incelendiğinde önsel yetenek dağılımı normal olan verilerde, birinci faktöre göre açıklanan varyans oranlarının 0.276 ile 0.393; birinci özdeğerlerin ortalamaları 2.962 ile 15.470; birinci özdeğerin ikinci özdeğere oranlarının 2.592 ile 13.581 arasında değiştiği görülmektedir. Buna göre önsel yetenek dağılımı normal olan verilerde tek boyutluluk varsayımı karşılanmıştır.

Tablo 3'te önsel yetenek dağılımı normal olmayan verilere ait açımlayıcı faktör analizi sonuçları verilmiştir.

Tablo 3

Önsel Yetenek Dağılımı Normal Olmayan Verilere Ait Tek Boyutluluk Analizi Sonuçları

Önsel Yetenek Dağılımı	Madde Sayısı	Örneklem Büyüklüğü	MTK Modeli	Açıklanan Varyans Ortalaması*	Birinci Özdeğerlerin Ortalaması	λ_1/λ_2**
Sola çarpık	10	100	2PL	0.846	8.458	17.909
		500		0.848	8.479	22.641
		1000		0.849	8.490	24.362
	20	100		0.821	16.415	22.402
		500		0.822	16.446	31.187
		1000		0.825	16.500	34.763
	40	100		0.783	31.339	27.194
		500		0.805	32.206	45.574
		1000		0.809	32.362	51.267
	10	100	3PL	0.775	7.750	12.290
		500		0.763	7.633	15.458
		1000		0.762	7.622	16.359
	20	100		0.733	14.657	14.703
		500		0.736	14.717	21.422
		1000		0.735	14.710	23.287
	40	100		0.698	27.937	18.640
		500		0.719	28.773	31.903
		1000		0.722	28.891	36.374

(Devam ediyor)

Tablo 3 (Devam)

Önsel Yetenek Dağılımı Normal Olmayan Verilere Ait Tek Boyutluluk Analizi Sonuçları

Önsel Yetenek Dağılımı	Madde Sayısı	Örneklem Büyüklüğü	MTK Modeli	Açıklanan Varyans Ortalaması*	Birinci Özdeğerlerin Ortalaması	λ_1/λ_2**
Sağa çarpık	10	100	2PL	0.846	8.456	18.015
		500		0.849	8.487	22.873
		1000		0.849	8.488	24.085
	20	100		0.813	16.268	21.201
		500		0.824	16.476	31.537
		1000		0.820	16.399	32.982
	40	100		0.785	31.396	27.923
		500		0.805	32.214	45.106
		1000		0.808	32.308	51.062
	10	100	3PL	0.682	6.823	8.964
		500		0.686	6.862	11.552
		1000		0.686	6.858	12.014
	20	100		0.582	11.641	9.344
		500		0.599	11.972	14.244
		1000		0.613	12.254	16.261
	40	100		0.548	21.932	12.276
		500		0.574	22.952	19.723
		1000		0.569	22.742	21.477
Sivri	10	100	2PL	0.206	2.064	1.739
		500		0.209	2.094	2.427
		1000		0.207	2.075	2.600
	20	100		0.203	4.051	2.293
		500		0.276	5.523	4.739
		1000		0.265	5.310	5.179
	40	100		0.136	5.450	2.090
		500		0.143	5.704	3.591
		1000		0.138	5.502	4.140
	10	100	3PL	0.172	1.728	1.476
		500		0.154	1.544	1.884
		1000		0.172	1.728	2.215
	20	100		0.180	3.602	2.033
		500		0.159	3.183	2.777
		1000		0.223	4.468	3.802
	40	100		0.106	4.236	1.644
		500		0.089	3.566	2.367
		1000		0.086	3.445	2.724
Basık	10	100	2PL	0.736	7.364	11.672
		500		0.738	7.381	16.064
		1000		0.734	7.341	17.154
	20	100		0.713	14.259	15.152
		500		0.722	14.443	23.394
		1000		0.727	14.535	27.224
	40	100		0.684	27.349	19.772
		500		0.722	28.888	37.011
		1000		0.723	28.903	42.966
	10	100	3PL	0.628	6.276	7.567
		500		0.644	6.437	10.871
		1000		0.649	6.495	11.878
	20	100		0.592	11.835	9.558
		500		0.596	11.922	14.699
		1000		0.598	11.966	16.319
	40	100		0.550	22.003	12.282
		500		0.578	23.115	21.445
		1000		0.576	23.037	24.259

* Birinci faktörün açıkladığı varyans oranı
** Birinci özdeğerin ikinci özdeğere oranı

Tablo 3 incelendiğinde önsel yetenek dağılımı normal olmayan verilerde, birinci faktöre göre açıklanan varyans oranlarının 0.086 ile 0.849; birinci özdeğerlerin ortalamaları 1.544 ile 32.362; birinci özdeğerin ikinci özdeğere oranlarının 1.476 ile 51.267 arasında değiştiği görülmektedir.

Tek boyutluluğa kanıt sağlaması açısından birinci faktörün açıkladığı varyans oranının 0.30'un altına düşmemesi beklenir. Reckase (1979) açıklanan varyans oranının 0.20 olmasını tek boyutluluk için alt sınır olarak kabul etmiş ancak bu durumda madde parametrelerinin stabil olmayacağını belirtmiştir. Buna göre sivri dağılımların bazı koşulları hariç diğer dağılımlara ait tüm veri setlerinde tek boyutluluk için elde edilen varyans oranı 0.20 üzerindedir. Ayrıca, Lord (1980), eğer birinci özdeğer, ikinciye kıyasla büyükse ve ikincisi, diğer özdeğerlerin herhangi birinden çok daha büyük değilse, bunun tek boyutluluğun bir kanıtı olarak alınabileceğini öne sürmektedir. Buna göre önsel yetenek dağılımı normal olmayan veriler tek boyutluluk varsayımının karşılandığı görülmektedir.

Yerel bağımsızlık varsayımının test edilmesi için Yen (1984)'in Q_3 istatistiği kullanılmıştır. Q_3 istatistiğinde madde parametrelerinin kestirimlerine dayalı olarak her birey için her bir maddeye verdiği tepkinin artıkları ve daha sonra maddelerden elde edilen artıklar arasındaki doğrusal korelasyon hesaplanır. Yen (1984), test geliştiricilerin 0.20 ve daha yüksek korelasyon değerlerine sahip madde çiftlerine problemli olarak yaklaşmalarını tavsiye etmektedir (DeMars, 2010). Önsel yetenek dağılımı normal olan verilerde yerel bağımsızlığın test edilmesi sonucu elde edilen değerler Tablo 4'te verilmiştir.

Tablo 4

Önsel Yetenek Dağılımı Normal Olan Verilerin Yerel Bağımsızlığı Karşılama Oranları

Önsel Yetenek Dağılımı	Madde Sayısı	Örneklem Büyüklüğü	MTK Modeli	Yerel Bağımsızlık (%)
Normal	10	100	2PL	100
		500		100
		1000		100
	20	100		95
		500		100
		1000		100
	40	100		46
		500		100
		1000		100

(Devam ediyor)

Tablo 4 (Devam)

Önsel Yetenek Dağılımı Normal Olan Verilerin Yerel Bağımsızlığı Karşılama Oranları

Önsel Yetenek Dağılımı	Madde Sayısı	Örneklem Büyüklüğü	MTK Modeli	Yerel Bağımsızlık (%)
Normal	10	100	3PL	96
		500		100
		1000		100
	20	100		84
		500		100
		1000		100
	40	100		36
		500		100
		1000		100

Tablo 4 incelendiğinde önsel yetenek dağılımı normal olan verilerin büyük bir kısmında yerel bağımsızlık varsayımının sağlandığı görülmektedir. Ancak, örneklem büyüklüğü 100'e düştüğünde yerel bağımsızlık varsayımının karşılanma durumunun azaldığı görülmektedir.

Önsel yetenek dağılımı normal olmayan verilerde yerel bağımsızlığın test edilmesi sonucu elde edilen değerler Tablo 5'te verilmiştir.

Tablo 5

Önsel Yetenek Dağılımı Normal Olmayan Verilerin Yerel Bağımsızlığı Karşılama Oranları

Önsel Yetenek Dağılımı	Madde Sayısı	Örneklem Büyüklüğü	MTK Modeli	Yerel Bağımsızlık (%)
Sola çarpık	10	100	2PL	30
		500		100
		1000		100
	20	100		2
		500		100
		1000		100
	40	100		2
		500		97
		1000		100
	10	100	3PL	51
		500		100
		1000		100
	20	100		10
		500		99
		1000		100
	40	100		2
		500		100
		1000		100

(Devam ediyor)

Tablo 5 (Devam)

Önsel Yetenek Dağılımı Normal Olmayan Verilerin Yerel Bağımsızlığı Karşılama Oranları

Sağa çarpık	10	100	2PL	32
		500		99
		1000		99
	20	100		2
		500		100
		1000		100
	40	100		2
		500		94
		1000		100
	10	100	3PL	82
		500		100
		1000		100
	20	100		47
		500		100
		1000		100
	40	100		2
		500		100
		1000		100
Sivri	10	100	2PL	100
		500		100
		1000		100
	20	100		93
		500		100
		1000		100
	40	100		54
		500		100
		1000		100
	10	100	3PL	97
		500		100
		1000		100
	20	100		93
		500		100
		1000		100
	40	100		50
		500		100
		1000		100
Basık	10	100	2PL	90
		500		100
		1000		100
	20	100		54
		500		100
		1000		100
	40	100		2
		500		100
		1000		100
	10	100	3PL	94
		500		100
		1000		100
	20	100		65
		500		100
		1000		100
	40	100		5
		500		100
		1000		100

Tablo 5 incelendiğinde önsel yetenek dağılımı normal olmayan verilerde büyük bir oranda yerel bağımsızlık varsayımının sağlandığı görülmektedir. Ayrıca, yerel bağımsızlık varsayımının, dağılımın örneklem büyüklüğünden etkilendiği görülmektedir. Örneklem büyüklüğü azaldıkça yerel bağımsızlığı ihlal eden veri seti oranı artmaktadır. Örneğin örneklem büyüklüğü 100'e düştükçe yerel bağımsızlık varsayımını ihlal eden veri seti oranı artış göstermektedir. Buna göre önsel yetenek dağılımı normal olan ve olmayan veriler için yerel bağımsızlık varsayımının büyük bir oranda karşılandığı belirlenmiştir.

Verilerin model-veri uyumu varsayımının test edilmesi için M2 değerleri incelenmiştir. M2 sınırlı bilgi uyum iyiliği değerleri, p değeriyle birlikte yorumlanmalıdır. Tek ve çift yönlü marjinal tablolarına göre hesaplanan M2 istatistiği, model-veri uyumunun ne kadar yüksek olduğunu göstermektedir. Bu istatistikte, momentleri eşleştirerek asimptotik p değerleri elde edilir ve parametre tahminlerinin asimptotik kovaryans matrisinin hesaplanmasını gerektirir. Diğer uyum iyiliği indekslerinde olduğu gibi M2 evrende mükemmel model-veri uyumu olduğu varsayımına dayanmaktadır (Maydeu-Olivares ve Joe, 2006). Uyum göstergesi olarak M2 istatistiğinin manidar olmaması beklenmektedir. Önsel yetenek dağılımı normal olan verilere ait M2 değerleri ve manidarlık durumları Tablo 6'da verilmiştir.

Tablo 6

Önsel Yetenek Dağılımı Normal Olan Verilerin Model-Veri Uyumuna Ait M2 Test İstatistiği Sonuçları

Önsel Yetenek Dağılımı	Madde Sayısı	Örneklem Büyüklüğü	MTK Modeli	Model Veri Uyum İyiliği			
				M2 min	M2 max	sd	Ort. p*
Normal	10	100	2PL	18.153	49.777	35	0.494
		500		14.441	49.563		0.542
		1000		15.751	49.481		0.524
	20	100		124.373	201.357	170	0.428
		500		136.681	200.899		0.515
		1000		131.487	198.907		0.553
	40	100		647.353	801.075	740	0.417
		500		623.636	801.548		0.447
		1000		657.574	804.057		0.483
	10	100	3PL	9.450	37.055	25	0.571
		500		7.352	37.613		0.558
		1000		9.488	35.397		0.541
	20	100		110.713	176.258	150	0.532
		500		106.731	178.820		0.484
		1000		120.379	179.544		0.483
	40	100		599.351	757.984	700	0.499
		500		606.412	762.245		0.511
		1000		624.574	762.193		0.556

*p<0.05, **p<0.001

Tablo 6 incelendiğinde önsel yetenek dağılımı normal olan verilerin model-veri uyumunun sağlandığı görülmektedir ($M2_{min}$= 7.352; $M2_{max}$= 804.057; p>0.05).

Önsel yetenek dağılımı normal olmayan verilere ait M2 değerleri ve manidarlık durumları Tablo 7'de verilmiştir.

Tablo 7

Önsel Yetenek Dağılımı Normal Olmayan Verilerin Model-Veri Uyumuna Ait M2 Test İstatistiği Sonuçları

Önsel Yetenek Dağılımı	Madde Sayısı	Örneklem Büyüklüğü	MTK Modeli	Model Veri Uyum İyiliği			
				M2 min	M2 max	sd	Ort. p*
Sola çarpık	10	100	2PL	13.534	48.178	35	0.728
		500		10.650	48.504		0.642
		1000		17.412	49.008		0.564
	20	100		79.776	169.729	170	0.962
		500		102.282	167.951		0.770
		1000		96.172	194.506		0.999
	40	100		390.874	651.422	740	0.999
		500		430.573	602.534		0.999
		1000		497.795	671.956		0.807
	10	100	3PL	6.604	35.582	25	0.816
		500		6.782	26.571		0.839
		1000		9.233	27.853		0.992
	20	100		67.426	131.531	150	0.995
		500		70.565	131.326		0.994
		1000		79.989	129.607		0.999
	40	100		332.034	551.934	700	0.999
		500		364.802	512.077		0.999
		1000		356.281	511.431		0.999
Sağa çarpık	10	100	2PL	12.525	40.679	35	0.818
		500		9.039	49.567		0.633
		1000		14.415	48.577		0.562
	20	100		80.706	182.694	170	0.954
		500		96.877	171.321		0.933
		1000		113.545	200.194		0.746
	40	100		389.962	644.047	740	0.999
		500		429.738	610.889		0.999
		1000		492.886	685.902		0.998
	10	100	3PL	8.670	35.251	25	0.606
		500		10.327	36.283		0.606
		1000		13.857	37.538		0.483
	20	100		115.519	179.097	150	0.446
		500		111.548	178.639		0.449
		1000		112.512	179.334		0.421
	40	100		650.893	762.493	700	0.332
		500		633.890	762.497		0.351
		1000		586.364	761.978		0.354

(Devam ediyor)

Tablo 7 (Devam)

Önsel Yetenek Dağılımı Normal Olmayan Verilerin Model-Veri Uyumuna Ait M2 Test İstatistiği Sonuçları

Önsel Yetenek Dağılımı	Madde Sayısı	Örneklem Büyüklüğü	MTK Modeli	Model Veri Uyum İyiliği			
				M2 min	M2 max	sd	Ort. p*
Sivri	10	100	2PL	18.633	49.508	35	0.483
		500		20.335	49.724		0.446
		1000		20.146	49.543		0.331
	20	100		136.173	200.821	170	0.313
		500		144.032	201.249		0.189
		1000		153.966	201.406		0.152
	40	100		675.114	803.785	740	0.343
		500		648.991	803.567		0.372
		1000		662.044	800.567		0.380
	10	100	3PL	10.056	37.162	25	0.619
		500		11.245	36.387		0.560
		1000		12.856	37.242		0.481
	20	100		122.581	178.015	150	0.456
		500		119.780	178.488		0.429
		1000		113.008	178.097		0.500
	40	100		630.296	758.815	700	0.392
		500		627.895	757.972		0.472
		1000		614.915	752.788		0.537
Basık	10	100	2PL	18.090	49.437	35	0.569
		500		19.071	48.814		0.576
		1000		15.354	49.770		0.558
	20	100		122.760	195.959	170	0.651
		500		112.843	192.551		0.750
		1000		119.883	195.198		0.745
	40	100		514.921	776.424	740	0.931
		500		553.310	718.812		0.982
		1000		563.294	726.958		0.970
	10	100	3PL	15.275	54.745	25	0.524
		500		14.407	52.061		0.577
		1000		17.571	56.692		0.491
	20	100		131.659	221.752	150	0.545
		500		125.914	226.722		0.555
		1000		126.867	222.681		0.443
	40	100		585.216	858.696	700	0.602
		500		594.909	917.087		0.477
		1000		539.962	742.728		0.857

*p<0.05, **p<0.001

Tablo 7 incelendiğinde önsel yetenek dağılımı normal olmayan verilerin; sola çarpık dağılım ($M2_{min}$= 6.604; $M2_{max}$= 671.956; p>0.05), sağa çarpık dağılım ($M2_{min}$= 8.670; $M2_{max}$= 762.497; p>0.05), sivri dağılım ($M2_{min}$= 10.056; $M2_{max}$= 803.785; p>0.05), basık dağılım ($M2_{min}$= 14.407; $M2_{max}$= 917.087; p>0.05) için model-veri uyumunun sağlandığı görülmektedir. Buna göre önsel yetenek dağılım türlerinin tümünde model-veri uyumunun sağlandığı görülmektedir. Ayrıca, bu varsayımın

karşılanmasındaki en önemli neden, veriler üretilirken simülasyonun mantıksal geçerliğini sağlamak amacıyla *önsel* fonksiyonuna bu kriterin yerleştirilmiş olmasıdır.

Veriler üzerinden MTK'nın varsayımlarının karşılanıp karşılanmadığı belirlendikten sonra simülasyonun geçerliği üzerine analizler yapılmıştır. Daha sonra madde ve yetenek parametrelerinin kestirimine geçilmiştir. Simülasyon koşullarına göre üretilen veriler üzerinden kestirilen yetenek ve madde parametreleri ile başlangıç parametreleri arasındaki yanlılık (bias) ve ortalama hata kareleri karekök (RMSE) değerleri hesaplanmıştır. İlk olarak *irtplay* (Lim ve Wells, 2020) paketiyle klasik yaklaşımlara göre en çok olabilirlik (ML) kestirimi, daha sonra *bairt* (Martinez, 2017) ve *sirt* (Robitzsch, 2022) paketleriyle Bayesçi yaklaşımlara göre Monte Carlo Markov Zinciri (MCMC) yöntemlerinden standart Bayes kestirimleri gerçekleştirilmiştir. Bayes kestirimlerde yakma (burning) değeri 1000, öteleme (iterasyon) değeri 3000 olarak belirlenmiştir. Klasik (en çok olabilirlik) kestirimlerine yönelik olarak kullanılan örnek R kodları EK.2'de ve Bayes MCMC kestirimlerine yönelik örnek R kodları ise EK.3'te verilmiştir.

Parametre Kestirimlerini Değerlendirme Ölçütleri

Parametre kestirimlerin sonuçları yorumlanırken kestirilen ve başlangıç parametrelerinin nasıl karşılaştırılacağına ve yorumlanacağına karar verilmelidir. Araştırmacılar genellikle korelasyon, kovaryans, yanlılık (bias), mutlak yanlılık, kestirimin standart hatası (SE), ortalama hata karesi (MSE) ve ortalama hata karelerinin karekökü (RMSE) gibi ölçümleri kullanırlar. Hangi istatistiklerin kullanılacağı ve nasıl yorumlanacağı araştırmanın problemlerine bağlıdır. Alanyazın incelendiğinde yanlılık (bias), standart hata (SE) ve ortalama hata kareleri karekökü (RMSE) en yangın olarak kullanılanlarıdır. Daha önceki simülasyon çalışmalarında kullanılan istatistikleri göz önünde bulundurmak da önemlidir. Bu şekilde çalışmaların benzer bağımlı değişkenleri karşılaştırabilir ve araştırmanın sonuçları önceki çalışmaların üzerine inşa edilebilir (Feinberg ve Rubright, 2016).

Bu araştırmada parametre kestirimlerinin değerlendirilmesinde yanlılık (bias) ve RMSE değerleri hesaplanmıştır. Yanlılık (bias) ve RMSE değerlerinin elde edilmesinde kullanılan bağıntılar aşağıda verilmiştir:

$$\boldsymbol{BIAS} = \frac{\mathbf{1}}{\boldsymbol{K}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{K}}\left(\widehat{\boldsymbol{X}}_{\boldsymbol{i}} - \boldsymbol{X}_{\boldsymbol{i}}\right) \qquad \boldsymbol{RMSE}_{\boldsymbol{i}} = \sqrt{\frac{\mathbf{1}}{\boldsymbol{K}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{K}}\left(\widehat{\boldsymbol{X}}_{\boldsymbol{i}} - \boldsymbol{X}_{\boldsymbol{i}}\right)^{\mathbf{2}}}$$

K; toplam test uzunluğunu, $\widehat{\boldsymbol{X}}_{\boldsymbol{i}}$; i maddesi için kestirilen madde parametresinin değeri, $\boldsymbol{X}_{\boldsymbol{i}}$; madde parametresinin gerçek değeridir. Bu eşitlikler simülasyonun her tekrarında üretilen parametreler için uygulanmıştır.

Bu iki istatistik değerleri öncelikle simülasyonun istatistiksel geçerliğine kanıt olarak kullanılmıştır. Daha sonra araştırma problemlerine yanıt olarak sadece RMSE değerleri işe koşulmuştur. Çünkü yanlılık (bias) değerleri hem pozitif hem de negatif değerler alabilmektedir. Bu durum yanlılık ortalamalarını etkilemektedir. Ayrıca RMSE ile yanlılık (bias) arasında bir ilişki bulunmaktadır. Bu ilişki aşağıdaki eşitlikte verilmiştir (Atar, 2007; Bilir, 2009; Feinberg ve Rubright, 2016):

$$\boldsymbol{RMSE^2 = Bias^2 + SE^2}$$

Bu eşitlikte yanlılığın karesi ile standart hatanın karesinin toplamı RMSE'nin karesine eşittir. Buna göre yanlılığın yarattığı negatif ve pozitif kayışlılık ortadan kalkmıştır.

Parametrelere ait RMSE değerleri arasındaki farkların anlamlılığı örneklem büyüklüğü, madde sayısı, MTK modeli, önsel dağılım ve kestirim yöntemine göre karma model ANOVA ile test edilmiştir. Bu analiz gerçekleştirilirken *afex* (Singmann, 2022) ve *emmeans* (Lenth, 2022) paketleri kullanılmıştır. Karma model ANOVA için ana etkiler (between) değişkeni olarak simülasyonun değişimlenen (örneklem büyüklüğü, madde sayısı, önsel yetenek dağılım biçimi) ve sabit tutulan (parametre kestirim yöntemleri) koşulları; etkileşimler (within) değişkeni olarak simülasyon tekrar (replikasyon) sayısı atanmıştır.

Analiz sonuçlarına göre manidar bulunan koşulların etki büyükleri (genelleştirilmiş eta-kare katsayısı) hesaplanmış ve Cohen (1988)'in önerisine göre değerlendirilmiştir. Buna göre etki değerinin büyüklüğü; 0.0099'dan küçük olması durumunda zayıf, 0.0588 olması durumunda orta, 0.1379'dan büyük olması durumunda güçlü şeklinde yorumlanmıştır. Aynı zamanda genelleştirilmiş eta-kare katsayısı 0 ile 1 arasında bir değer aldığından, bu değer 100 ile çarpıldığında bağımlı değişkene ait varyansın ne kadarının bağımsız değişkenler tarafından açıklandığını göstermektedir (Lakens, 2013).

Analiz sonuçlarına göre manidar koşulların grafiksel gösteriminde *ggplot2* (Wickham, 2016) ve *ggbeeswarm* (Clarke, 2022) paketleri kullanılmıştır.

Simülasyonun Geçerliği

Bu bölümde veri üretiminde kullanılan simülasyonun geçerliği hakkında bilgiler verilmiştir. Simülasyonun geçerliği birbirini tamamlayan iki farklı yolla ortaya konmuştur. Bunlardan birincisi, simülasyonun mantıksal geçerliği, ikicisi simülasyonun istatistiksel geçerliğidir.

Simülasyonun Mantıksal Geçerliği

Araştırmanın problem durumlarında belirtildiği üzere amaç belirli çarpıklık ve basıklık değerlerine sahip önsel yetenek dağılımları üretmektir. Araştırmacı, *önsel* adı verilen veri üretme fonksiyonunu kullanırken bu fonksiyonun içerisine mantıksal argümanlar yerleştirmiştir. Araştırmanın yöntem bölümünde verilen önsel fonksiyon *(bkz. sf. 46-47)* dikkatle incelendiğinde her bir dağılım türü için gizil dağılımları oluşturan kodlar yazılmıştır. Dağılımların üretilmesinde kullanılan örnek R kodları EK.1'de verilmiştir. Ancak bu durum yeterli görülmemiş, önsel dağılımlar üretilirken fonksiyonun her çalışma döngüsünde *(looping)* üretilen veri seti üzerinden çarpıklık ve basıklık değerleri hesaplanmıştır. Elde edilen çarpıklık ve basıklık değeri istenilen aralıktaysa veri seti araştırmaya dahil edilmiştir. Buna göre önsel yetenek dağılımı normal olan verilere ait çarpıklık ve basıklık katsayıları Tablo 8'de verilmiştir.

Tablo 8

Önsel Yetenek Dağılımı Normal Olan Verilere Ait Çarpıklık ve Basıklık Katsayıları

Önsel Yetenek Dağılımı	Madde Sayısı	Örneklem Büyüklüğü	MTK Modeli	Çarpıklık		Basıklık	
				$Z_ç$ min	$Z_ç$ max	Z_b min	Z_b max
Normal	10	100	2PL	-0.604	0.498	-0.998	-0.351
		500		-0.623	0.424	-0.997	-0.315
		1000		-0.420	0.473	-0.999	-0.587
	20	100		-0.478	0.690	-0.992	-0.420
		500		-0.343	0.347	-0.997	-0.681
		1000		-0.225	0.329	-0.998	-0.747
	40	100		-0.450	0.364	-0.999	-0.587
		500		-0.264	0.328	-0.999	-0.717
		1000		-0.367	0.276	-0.999	-0.735
	10	100	3PL	-0.517	0.425	-0.998	-0.238
		500		-0.509	0.376	-0.980	-0.324
		1000		-0.720	0.372	-0.948	-0.320

(Devam ediyor)

Tablo 8 (Devam)

Önsel Yetenek Dağılımı Normal Olan Verilere Ait Çarpıklık ve Basıklık Katsayıları

Önsel Yetenek Dağılımı	Madde Sayısı	Örneklem Büyüklüğü	MTK Modeli	Çarpıklık		Basıklık	
				$Z_ç$ min	$Z_ç$ max	Z_b min	Z_b max
		100		-0.521	0.346	-0.999	-0.145
	20	500		-0.380	0.227	-0.999	-0.499
		1000		-0.377	0.214	-0.989	-0.607
		100		-0.422	0.337	-0.998	-0.378
	40	500		-0.252	0.154	-0.991	-0.684
		1000		-0.295	0.188	-0.992	-0.683

Tablo 8 incelendiğinde önsel yetenek dağılımı normal olan veriler için çarpıklık ($Z_ç$) katsayıları -0.720 ile 0.690; basıklık (Z_b) katsayıları -0.999 ile -0.145 aralığındadır. Buna göre önsel yetenek dağılımı normal olan veriler istenilen çarpıklık ve basıklık katsayılarına sahiptir.

Önsel yetenek dağılımı normal olmayan verilere ait çarpıklık ve basıklık katsayıları Tablo 9'da verilmiştir.

Tablo 9

Önsel Yetenek Dağılımı Normal Olmayan Verilere Ait Çarpıklık ve Basıklık Katsayıları

Önsel Yetenek Dağılımı	Madde Sayısı	Örneklem Büyüklüğü	MTK Modeli	Çarpıklık		Basıklık	
				$Z_ç$ min	$Z_ç$ max	Z_b min	Z_b max
		100		-1.616	-1.502	0.657	0.999
	10	500		-1.611	-1.501	0.735	0.998
		1000		-1.599	-1.500	0.742	0.999
		100		-1.613	-1.500	0.715	0.992
	20	500	2PL	-1.592	-1.501	0.748	0.998
		1000		-1.587	-1.501	0.762	0.997
		100		-1.606	-1.501	0.696	0.996
	40	500		-1.610	-1.501	0.729	0.999
Sola çarpık		1000		-1.585	-1.500	0.780	0.998
		100		-1.583	-1.500	0.746	0.999
	10	500		-1.569	-1.500	0.852	0.999
		1000		-1.557	-1.500	0.867	0.999
		100		-1.584	-1.500	0.743	0.999
	20	500	3PL	-1.576	-1.500	0.790	0.999
		1000		-1.561	-1.501	0.873	0.999
		100		-1.596	-1.501	0.737	0.999
	40	500		-1.577	-1.501	0.815	0.996
		1000		-1.570	-1.501	0.821	0.999

(Devam ediyor)

Tablo 9 (Devam)

Önsel Yetenek Dağılımı Normal Olmayan Verilere Ait Çarpıklık ve Basıklık Katsayıları

Önsel Yetenek Dağılımı	Madde Sayısı	Örneklem Büyüklüğü	MTK Modeli	Çarpıklık		Basıklık	
				$Z_ç$ min	$Z_ç$ max	Z_b min	Z_b max
Sağa çarpık	10	100	2PL	1.500	1.615	0.727	0.999
		500		1.501	1.594	0.707	0.994
		1000		1.500	1.593	0.717	0.998
	20	100		1.501	1.597	0.670	0.999
		500		1.501	1.598	0.748	0.999
		1000		1.500	1.587	0.770	0.999
	40	100		1.500	1.609	0.739	0.999
		500		1.502	1.592	0.768	0.998
		1000		1.502	1.589	0.767	0.999
	10	100	3PL	1.500	1.544	0.829	0.998
		500		1.500	1.545	0.893	0.999
		1000		1.500	1.544	0.914	0.999
	20	100		1.500	1.558	0.788	0.999
		500		1.500	1.552	0.868	0.999
		1000		1.501	1.549	0.899	0.999
	40	100		1.500	1.586	0.754	0.998
		500		1.500	1.570	0.829	0.997
		1000		1.501	1.556	0.829	0.994
Sivri	10	100	2PL	-0.948	0.943	1.500	1.598
		500		-0.912	0.938	1.500	1.599
		1000		-0.963	0.955	1.501	1.599
	20	100		-0.985	0.991	1.500	1.599
		500		-0.760	0.788	1.501	1.597
		1000		-0.859	0.769	1.501	1.595
	40	100		-0.806	0.847	1.501	1.598
		500		-0.401	0.500	1.500	1.599
		1000		-0.386	0.455	1.502	1.600
	10	100	3PL	-0.968	0.907	1.501	1.598
		500		-0.910	0.924	1.500	1.599
		1000		-0.920	0.884	1.502	1.597
	20	100		-0.876	0.991	1.500	1.599
		500		-0.520	0.668	1.502	1.599
		1000		-0.762	0.835	1.500	1.600
	40	100		-0.864	0.989	1.502	1.599
		500		-0.526	0.616	1.500	1.600
		1000		-0.556	0.435	1.502	1.597
Basık	10	100	2PL	-0.274	0.262	-1.706	-1.503
		500		-0.200	0.210	-1.659	-1.500
		1000		-0.198	0.200	-1.667	-1.500
	20	100		-0.240	0.337	-1.726	-1.501
		500		-0.171	0.188	-1.633	-1.502
		1000		-0.233	0.192	-1.643	-1.504
	40	100		-0.204	0.241	-1.777	-1.506
		500		-0.118	0.232	-1.675	-1.526
		1000		-0.105	0.140	-1.688	-1.542
	10	100	3PL	-0.261	0.153	-1.679	-1.501
		500		-0.234	0.101	-1.611	-1.501
		1000		-0.211	0.112	-1.571	-1.500
	20	100		-0.288	0.219	-1.714	-1.500
		500		-0.183	0.133	-1.595	-1.500
		1000		-0.183	0.123	-1.581	-1.500
	40	100		-0.230	0.221	-1.721	-1.500
		500		-0.206	0.136	-1.620	-1.502
		1000		-0.220	0.066	-1.616	-1.501

Tablo 9 incelendiğinde önsel yetenek dağılımı normal olmayan verilerde; sola çarpık dağılımlar için ($Z_ç$) katsayıları -1.616 ile -1.500; basıklık (Z_b) katsayıları 0.657 ile 0.999, sağa çarpık dağılımlar için çarpıklık ($Z_ç$) katsayıları 1.500 ile 1.615; basıklık (Z_b) katsayıları 0.670 ile 0.999, sivri dağılımlar için çarpıklık ($Z_ç$) katsayıları -0.985 ile 0.991; basıklık (Z_b) katsayıları 1.500 ile 1.600, basık dağılımlar için çarpıklık ($Z_ç$) katsayıları -0.288 ile 0.337; basıklık (Z_b) katsayıları -1.777 ile -1.500 aralığındadır. Buna göre önsel yetenek dağılımların tümünde istenilen çarpıklık ve basıklık değerlerine sahip veriler üretilmiştir. Ayrıca önsel yetenek dağılımlarının çarpıklık ve basıklık katsayıları incelendiğinde sadece tek bir dağılım biçiminin özelliğini göstermesine dikkat edilmiştir. Örneğin çarpık dağılımların sadece çarpıklık değerleri değişimlenmiş, basıklık değerleri normal sınırlar içerisinde kontrol altında tutulmuştur. Aynı şekilde sivri ve basık dağılımlarda sadece çarpıklık değerleri normal sınırlar içerisinde kontrol altında tutulmuş, basıklık değerleri değişimlenmiştir.

Simülasyonun mantıksal geçerliliği için yapılan çalışmalara göre veri üretme sürecinin titiz bir şekilde yürütüldüğü, veri setinin çarpıklık ve basıklık değerlerine göre eksiksiz ve hatasız, sonuçların istenildiği gibi olduğu belirlenmiştir. Buna göre simülasyonun mantıksal geçerliği sağlanmıştır.

Simülasyonun İstatistiksel Geçerliği

Simülasyonun istatistiksel geçerliğine kanıt olarak kestirimlerin yanlılık (bias) ve RMSE değerleri üretilmiştir. Buna göre veriler üretilirken elde edilen madde parametrelerinin başlangıç madde parametrelerine olan benzeşikliği incelenmiştir. Bu süreçte, madde parametrelerinin yanlılık (bias) ve ortalama hata kareleri karekökü (RMSE) değerlerinin her bir önsel yetenek dağılımına göre nasıl değiştiği gözlenmiştir.

Üretilen yetenek ve madde parametrelerine ait yanlılık (bias) ve ortalama hata kareleri karekökü (RMSE) ortalama sonuçları EK 4 ve EK 6'da tablolar halinde ve bu parametrelere ait değişimler ise EK 5 ve EK 7'de verilmiştir.

EK 4'teki tablo ve EK 5'teki değişim grafikleri birlikte incelendiğinde önsel yetenek dağılımı normal ve normal olmayan (sol çarpık, sağa çarpık, sivri ve basık) 2 PL ve 3 PL modelde yetenek parametreleri için yanlılık ve RMSE değerleri bayes kestirimde klasik kestirime göre daha düşük elde edilmiştir. Yani Bayes kestirim klasik kestirime göre daha az yanlıdır.

EK 6'daki tablo ve EK 7'deki değişim grafikleri birlikte incelendiğinde önsel yetenek dağılımı normal olan ve normal olmayan (sol çarpık, sağa çarpık, sivri ve basık) 2 PL ve 3 PL modelde madde ayırt edicilik (a_i) ve madde güçlük (b_i) parametreleri için yanlılık ve RMSE değerleri bayes kestirimde klasik kestirime göre daha düşüktür. Ancak madde şans (c_i) parametresi için önsel yetenek dağılım biçimi ya da model fark etmeksizin yanlılık ve RMSE değerleri açısından bir farklılık gözlenmemiştir.

Simülasyonun istatistiksel geçerliğine ilişkin bulgular verildikten sonra araştırma problemlerine göre yetenek ve madde parametrelerinin kestirimleri gerçekleştirilmiştir. Her bir araştırma problemine yönelik bulgular başlıklar halinde sunulmuştur.

BÖLÜM 4

BULGULAR VE YORUMLAR

Bu bölümde araştırmanın problemlerine yanıt olarak verilerin çözümlenmesi sonucunda elde edilen bulgulara ve bunlara ilişkin yorumlara yer verilmiştir.

Önsel Yetenek Dağılımı Normal Olan ve Olmayan 2 PL Modeldeki Verilerde Örneklem Büyüklüğüne, Madde Sayısına Göre Klasik (En Çok Olabilirlik) ve Bayes Yöntemleri ile Kestirilen Yetenek Parametrelerinin RMSE Değerlerinin İncelenmesi

Önsel yetenek dağılımı normal olan verilerde klasik ve Bayesçi parametre kestirim yöntemleri sonucu elde edilen yetenek (Θ) parametrelerinin ve kestirimin standart hatalarının (Θ_{SE}) ortalamaları EK 8'de verilmiştir.

Araştırmanın birinci probleminde belirtilen önsel yetenek dağılımı normal olan ve olmayan (sola çarpık, sağa çarpık, sivri ve basık) 2 PL modeldeki verilerde örneklem büyüklüğü, madde sayısı ve kestirim yöntemine göre yetenek parametreleri RMSE değişimleri karma model ANOVA ile analiz edilmiştir. Analiz sürecinde kurulan modeller ve bu modellere ait örnek R kodları aşağıda verilmiştir:

```
#####2PL Yetenek Parametreleri RMSE Değerleri İçin Karma Model ANOVA####
#2PL modelde yetenek parametreleri RMSE değerleri üzerinde örneklem büyüklüğünün ve kestirim yönteminin etkisi
data.2PL.theta.model1 <- aov_ez(data = data.2PL.theta, dv = "theta.2PL.rmse", id = "theta.id.2PL",between = "theta.sample.size",
                    within = "theta.replication")

print(ascii(nice(data.2PL.theta.model1), include.rownames = FALSE, caption = "ANOVA 1"), type = "org")

summary(as.glht(pairs(emmeans(data.2PL.theta.model1,
~theta.sample.size))))

data.2PL.theta.model1.1 <- aov_ez(data = data.2PL.theta, dv = "theta.2PL.rmse", id = "theta.id.2PL",
      between = c("theta.sample.size","theta.estimation.method"),
       within = "theta.replication")
```

```
print(ascii(nice(data.2PL.theta.model1.1), include.rownames = FALSE,
caption = "ANOVA 1"), type = "org")

afex_plot(data.2PL.theta.model1.1,x="theta.replication",trace="theta.s
ample.size", panel = "theta.estimation.method", mapping =
c("color","fill"),
error = "within", data_geom = geom_boxplot, data_arg = list(width=0.4),
point_arg = list(size=1.5), line_arg = list(size=1))
```

Örnek R kodları, 2 PL modelde sadece yetenek parametreleri RMSE değerleri üzerinde örneklem büyüklüğünün ve kestirim yönteminin etkisi için verilmiştir. Diğer koşullar için aynı şekilde kodlar üretilmiştir.

Buna göre önsel yetenek dağılımı normal olan ve olmayan (sola çarpık, sağa çarpık, sivri ve basık) 2 PL modeldeki verilerde örneklem büyüklüğü, madde sayısı ve kestirim yöntemine göre yetenek parametreleri RMSE değerleri için gerçekleştirilen karma model ANOVA sonuçları Tablo 10'da verilmiştir.

Tablo 10

Önsel Yetenek Dağılımı Normal Olan ve Olmayan 2 PL Modeldeki Verilerde Yetenek Parametreleri RMSE Değerleri İçin Karma Model ANOVA Sonuçları

Bağımsız değişkenler	Hata kareleri ortalaması	Serbestlik derecesi	F	p	Genelleştirilmiş η^2
Kestirim yöntemi (K)	72.05	1	7.83	.006**	.078
Örneklem büyüklüğü (Ö)	79.36	2	0.00	.997	.001
Madde sayısı (M)	65.24	2	9.42	.001**	.171
Önsel yetenek dağılım biçimi (D)	42.74	4	1.88	.001**	.456
K * Ö	75.46	2	0.01	.994	.001
K * M	58.50	2	1.69	.191	.037
K * D	31.90	4	4.05	.005**	.155
Hata	0.30	198			
Toplam	425.55				

*p<.05 **p<.01

Tablo 10 incelendiğinde önsel yetenek dağılımı normal olan ve olmayan 2 PL modeldeki verilerde yetenek parametreleri RMSE değerleri için karma model ANOVA sonuçlarına göre bağımsız değişkenlerin ana etkileri olarak kestirim yönteminin *($F_{(1, 88)} = 7.83; p<.01, \eta^2 = .078$)*, madde sayısının *($F_{(2, 84)} = 9.42; p<.01, \eta^2 = .171$)* ve önsel yetenek dağılım biçiminin *($F_{(4, 80)} = 1.88; p<.01, \eta^2 = .456$)* manidar etkiye sahip oldukları görülmektedir. Ancak örneklem büyüklüğü *($F_{(2, 87)} = 0.00; p>.05, \eta^2 = .001$)* manidar bir etkiye sahip değildir. Manidar olan kestirim yöntemi orta, madde sayısı yüksek ve önsel yetenek dağılım biçimi yüksek etki büyüklüğüne sahiptir.

Koşulların ikili etkileşimleri incelendiğinde kestirim yöntemi ile önsel yetenek dağılım biçimi ortak etkileşimi manidar bulunmuştur *($F_{(4, 80)} = 4.05; p<.01, \eta^2 = .155$).* Ortak etkileşimin etki büyüklüğü ise yüksek düzeydedir.

Bu sonuçlar genel hatlarıyla incelediğinde Finch ve Edwards (2015)'ın sonuçlarıyla benzerlik göstermektedir. Aynı şekilde 2 PL modelde gizil özelliğin normal dağılmadığı durumlarda Bayes kestirimler daha titiz sonuçlar vermektedir. Madde sayısı açısından benzer durum Köse (2010)'nin araştırmasında da görülmektedir. Madde sayısındaki değişim yetenek parametrelerindeki kestirim sonuçlarını etkilemektedir.

2 PL modelde yetenek parametresi kestirim yöntemlerine göre ikili karşılaştırmalar Tablo 11'de verilmiştir.

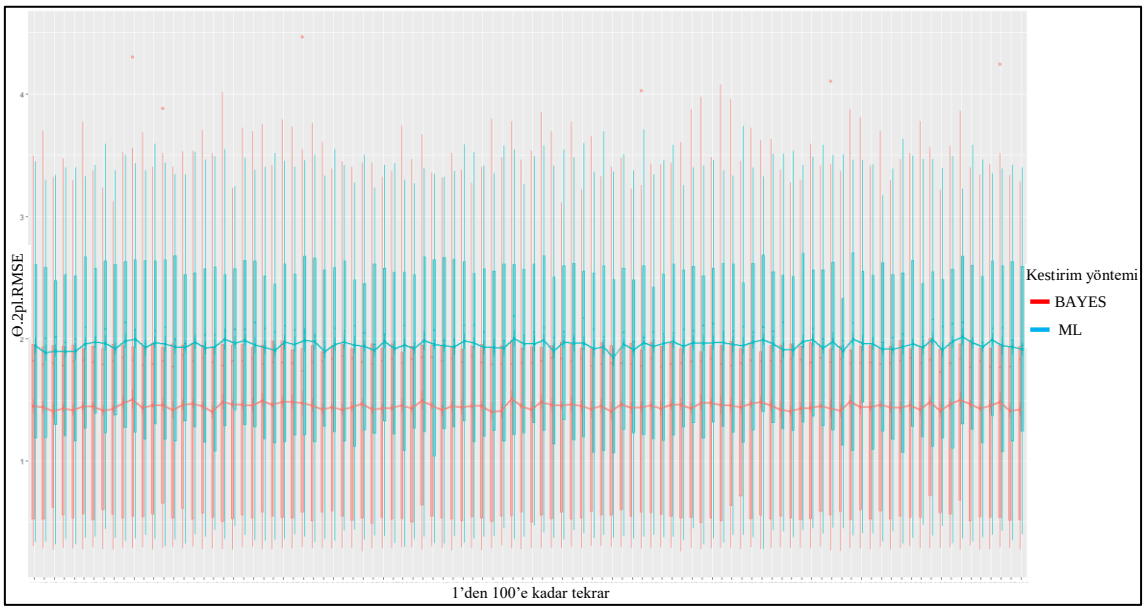
Tablo 11

2 PL Modelde Yetenek Parametresi RMSE Değerlerinin Kestirim Yöntemlerine Göre İkili Karşılaştırmaları

Kestirim yöntemi	Fark	Standart hata	t	p
BAYES-ML	-0.501	0.179	-2.799	.001**

*p<.05 **p<.01

Tablo 11 incelendiğinde 2 PL modelde yetenek parametresi kestirim yöntemlerinden bayes kestirim klasik kestirime göre daha düşük ve manidar RMSE değerleri üretmiştir *($t=-2.799; p<.01$).* 2 PL modelde yetenek parametresi kestirim yöntemlerine ait RMSE değişimleri Şekil 3'te verilmiştir.



Şekil 3. 2 PL Modelde Yetenek Parametresi RMSE Değerlerinin Kestirim Yöntemlerine Göre Değişimi

Şekil 3 incelendiğinde araştırmanın koşullarından bağımsız olarak 2 PL modeldeki tüm veri setlerinden elde edilen yetenek parametreleri RMSE değerlerinin tamamı verilmiştir. Buna göre 2 PL modelde yetenek parametresi kestirilirken Bayes yöntemi klasik yönteme göre daha düşük ve manidar RMSE değerleri üretmiştir.

2 PL modelde yetenek parametresi üzerinde madde sayısına göre ikili karşılaştırmalar Tablo 12'de verilmiştir.

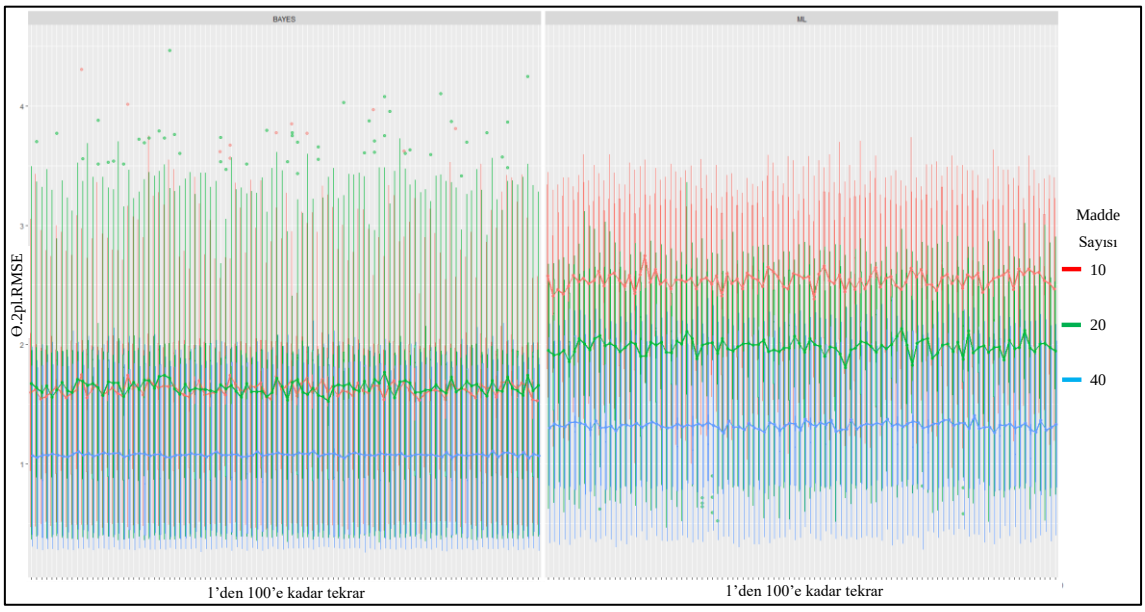
Tablo 12

2 PL Modelde Yetenek Parametresi RMSE Değerlerinin Madde Sayısına Göre İkili Karşılaştırmaları

Madde sayısı	Fark	Standart hata	t	p
10 – 20	0.272		1.304	.397
10 – 40	0.884	0.209	4.236	.001**
20 – 40	0.612		2.933	.012*

*p<.05 **p<.01

Tablo 12 incelendiğinde 2 PL modelde yetenek parametresi RMSE değerleri üzerinde madde sayısı 10 ile 40 *(t=4.236; p<.01)*, 20 ile 40 *(t=2.933; p<.05)* arasında manidar farklılıklar bulunmaktadır. 2 PL modelde yetenek parametresi üzerinde madde sayısına göre RMSE değişimi Şekil 4'te verilmiştir.



Şekil 4. 2 PL Modelde Yetenek Parametresi RMSE Değerlerinin Madde Sayısına Göre Değişimi

Şekil 4 incelendiğinde 2 PL modelde yetenek parametresi kestirimleri üzerinde madde sayısı artıkça RMSE değerleri azalma göstermektedir. Klasik yöntemle yapılan kestirim sonucunda yetenek parametreleri RMSE değerleri testin madde sayısı artıkça azalma göstermektedir. Aynı durum Bayes kestirim yönteminde de görülmektedir. Bayes kestirim yönteminde madde sayısının 10 ile 20 olması durumunda bir farklılaşma olmamakla birlikte madde sayısının 40 çıkması durumunda daha düşük RMSE değerleri elde edilmiştir. Bununla birlikte Bayes kestirimde madde sayısına göre elde edilen yetenek parametreleri RMSE değerleri klasik kestirime göre daha düşük değerler almıştır.

2 PL modelde yetenek parametresi üzerinde önsel yetenek dağılım biçimine göre ikili karşılaştırmalar Tablo 13'te verilmiştir.

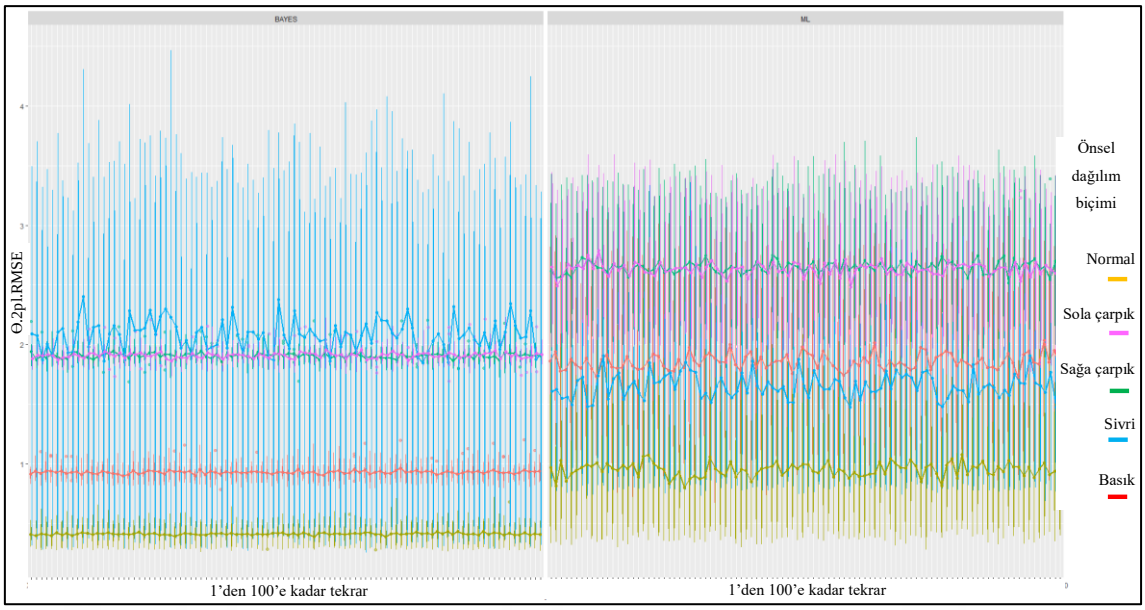
Tablo 13

2 PL Modelde Yetenek Parametresi RMSE Değerlerinin Önsel Yetenek Dağılımına Göre İkili Karşılaştırmaları

Önsel yetenek dağılımı	Fark	Standart hata	t	p
Normal – Sola çarpık	-1.589		-7.292	.000**
Normal – Sağa çarpık	-1.595		-7.321	.000**
Normal – Sivri	-1.184		-5.434	.000**
Normal – Basık	-0.712		-3.267	.013*
Sola çarpık – Sağa çarpık	-0.006	0.218	-0.029	.999
Sola çarpık – Sivri	0.405		1.858	.347
Sola çarpık – Basık	0.877		4.026	.001**
Sağa çarpık – Sivri	0.411		1.887	.332
Sağa çarpık – Basık	0.884		4.054	.001**
Sivri – Basık	0.472		2.267	.202

*p<.05 **p<.01

Tablo 13 incelendiğinde 2 PL modelde yetenek parametresi RMSE değerleri üzerinde önsel yetenek dağılım biçimi normal ile sola çarpık *(t=-7.292; p<.01)*, normal ile sağa çarpık *(t=-7.321; p<.01)*, normal ile sivri *(t=-5.434; p<.01)*, normal ile basık *(t=-3.267; p<.05)*, sola çarpık ile basık *(t=4.026; p<.01)*, sağa çarpık ile basık *(t=4.054; p<.01)* arasında manidar farklılıklar bulunmuştur. Bu farklılıklar Bayes kestirim yöntemi lehindedir. 2 PL modelde önsel yetenek dağılım biçimi normalden farklılaştıkça Bayes kestirim daha düşük RMSE değerleri üretmektedir. 2 PL modelde yetenek parametresi üzerinde önsel yetenek dağılım biçimine göre RMSE değişimi Şekil 5'te verilmiştir.



Şekil 5. 2 PL Modelde Yetenek Parametresi RMSE Değerlerinin Önsel Yetenek Dağılım Biçimlerine Göre Değişimi

Şekil 5 incelendiğinde 2 PL modelde önsel yetenek dağılımı normalden çarpıklaştıkça (sola çarpık, sağa çarpık, sivri ve basık) klasik kestirimde yetenek parametreleri RMSE değerleri artış göstermektedir. Klasik kestirimde yetenek parametreleri üzerinde en düşük RMSE değerleri önsel yetenek dağılım biçimi normal olduğunda elde edilmiştir. Dağılım çarpıklaştıkça hata değerlerinde artış görülmektedir. Dağılımın sola ve sağa çarpık durumlarında RMSE değerleri en yüksek, sivri ve basık olma durumlarında daha düşüktür. Dağılım normalleştikçe bu değerler daha da azalma göstermektedir.

2 PL modelde yetenek parametreleri kestirimi Bayes yöntemi ile gerçekleştirildiğinde RMSE değerleri klasik kestirime göre daha düşük değerler almaktadır. Benzer şekilde en düşük RMSE değerleri sırasıyla normal, basık, sola ve sağa çarpık ve en son sivri dağılımdadır. Burada dikkat edilmesi gereken durum sivri dağılım haricindeki diğer tüm önsel yetenek dağılım türlerinde, RMSE değerleri bayes kestirimde azalma gösterirken sivri dağılımda klasik kestirime göre daha yüksek değerler almıştır. Önsel yetenek dağılımı üretilirken sivri dağılımın normal dağılıma göre daha düşük standart sapmaya sahip olması ve dağılım aralığı olarak nispeten -1 ile +1 arasında kalması nedeniyle önsel (prior) dağılımla karşılaştırıldığında sonsal dağılım olarak daha geniş ranj aralığında bir şekle bürünecektir. Bu nedenle başlangıç yetenek parametreleri ile kestirilen yetenek parametreleri arasında RMSE farkları artış gösterecektir. Buna göre 2 PL modelde yetenek parametreleri kestirilirken önsel yetenek dağılımının sivri olduğu

durumlar hariç diğer dağılım biçimlerinde Bayes kestirim yönteminin kullanılmasında daha düşük RMSE değerleri üretilmesini sağlayacaktır.

Önsel Yetenek Dağılımı Normal Olan ve Olmayan 3 PL Modeldeki Verilerde Örneklem Büyüklüğüne, Madde Sayısına Göre Klasik (En Çok Olabilirlik) ve Bayes Yöntemleri ile Kestirilen Yetenek Parametrelerinin RMSE Değerlerinin İncelenmesi

Araştırmanın ikinci probleminde belirtilen önsel yetenek dağılımı normal olan ve olmayan (sola çarpık, sağa çarpık, sivri ve basık) 3 PL modeldeki verilerde örneklem büyüklüğü, madde sayısı ve kestirim yöntemine göre yetenek parametreleri RMSE değişimleri karma model ANOVA ile analiz edilmiştir. Analiz sürecinde kurulan modeller ve bu modellere ait örnek R kodları aşağıda verilmiştir:

```
#3PL modelde yetenek parametreleri RMSE değerleri üzerinde örneklem büyüklüğünün ve kestirim yönteminin etkisi

data.3PL.theta.model1 <- aov_ez(data = data.3PL.theta, dv = "theta.3PL.rmse", id = "theta.id.3PL", between = "theta.sample.size",
                      within = "theta.replication")

print(ascii(nice(data.3PL.theta.model1), include.rownames = FALSE, caption = "ANOVA 1"), type = "org")

summary(as.glht(pairs(emmeans(data.3PL.theta.model1,
~theta.sample.size))))

data.3PL.theta.model1.1 <- aov_ez(data = data.3PL.theta, dv = "theta.3PL.rmse", id = "theta.id.3PL",
                    between = c("theta.sample.size", "theta.estimation.method"),
                    within = "theta.replication")

print(ascii(nice(data.3PL.theta.model1.1), include.rownames = FALSE, caption = "ANOVA 1"), type = "org")

afex_plot(data.3PL.theta.model1.1,x="theta.replication",trace="theta.sample.size", panel = "theta.estimation.method", mapping = c("color","fill"),error = "within", data_geom = geom_boxplot, data_arg = list(width=0.4), point_arg = list(size=1.5), line_arg = list(size=1))
```

Örnek R kodları, 3 PL modelde sadece yetenek parametreleri RMSE değerleri üzerinde örneklem büyüklüğünün ve kestirim yönteminin etkisi için verilmiştir. Diğer koşullar için aynı şekilde kodlar üretilmiştir.

Buna göre önsel yetenek dağılımı normal olan ve olmayan (sola çarpık, sağa çarpık, sivri ve basık) 3 PL modeldeki verilerde örneklem büyüklüğü, madde sayısı ve

kestirim yöntemine göre yetenek parametreleri RMSE değerleri için gerçekleştirilen karma model ANOVA sonuçları Tablo 14'te verilmiştir.

Tablo 14

Önsel Yetenek Dağılımı Normal Olan ve Olmayan 3 PL Modeldeki Verilerde Yetenek Parametreleri RMSE Değerleri İçin Karma Model ANOVA Sonuçları

Bağımsız değişkenler	Hata kareleri ortalaması	Serbestlik derecesi	F	p	Genelleştirilmiş η^2
Kestirim yöntemi (K)	2769.62	1	0.27	.607	.003
Örneklem büyüklüğü (Ö)	2747.44	2	0.99	.376	.022
Madde sayısı (M)	2488.40	2	5.62	.005**	.111
Önsel yetenek dağılım biçimi (D)	2315.05	4	5.15	.001**	.189
K * Ö	2836.73	2	0.00	.999	.001
K * M	2568.23	2	0.00	.996	.001
K * D	2441.93	4	0.07	.991	.003
Hata	0.96	198			
Toplam	18168.36				

*p<.05 **p<.01

Tablo 14 incelendiğinde önsel yetenek dağılımı normal olan ve olmayan 3 PL modeldeki verilerde yetenek parametreleri RMSE değerleri için karma model ANOVA sonuçlarına göre bağımsız değişkenlerin ana etkileri olarak madde sayısı *($F_{(2, 84)} = 5.62$; $p<.01$, $\eta^2 = .111$)* ve önsel yetenek dağılım biçimi *($F_{(4, 80)} = 5.15$; $p<.01$, $\eta^2 = .189$)* manidar olarak bulunmuştur. Kestirim yöntemi *($F_{(1, 88)} = 0.27$; $p>.05$, $\eta^2 = .003$)* ve örneklem büyüklüğü *($F_{(2, 87)} = 0.99$; $p>.05$, $\eta^2 = .022$)* manidar bir etkiye sahip değildir. Manidar etkiye sahip olan madde sayısı orta ve önsel yetenek dağılım biçimi yüksek etki büyüklüğüne sahiptir. İkili etkileşimler incelendiğinde herhangi bir koşul manidar bulunamamıştır.

Bu problem için elde edilen genel sonuçlar Swaminathan ve Gifford (1986)'un çalışmasının sonuçlarıyla benzerlik göstermektedir. Çalışmalarında 3 PL model için klasik kestirim yerine Bayes kestirimin kullanılması gerektiğini önermişlerdir. Ayrıca, 3 PL modelde yetenek parametresi için kestirim yapılırken önsel yetenek dağılım biçiminin bilinememesinin hatalı kestirimlere neden olacağını Karadavut (2019) çalışmasında belirtmiştir. Benzer durum bu çalışma için önsel yetenek dağılımın farklılaşması olarak görülebilir.

3 PL modelde yetenek parametresi üzerinde madde sayısına göre ikili karşılaştırmalar Tablo 15'te verilmiştir.

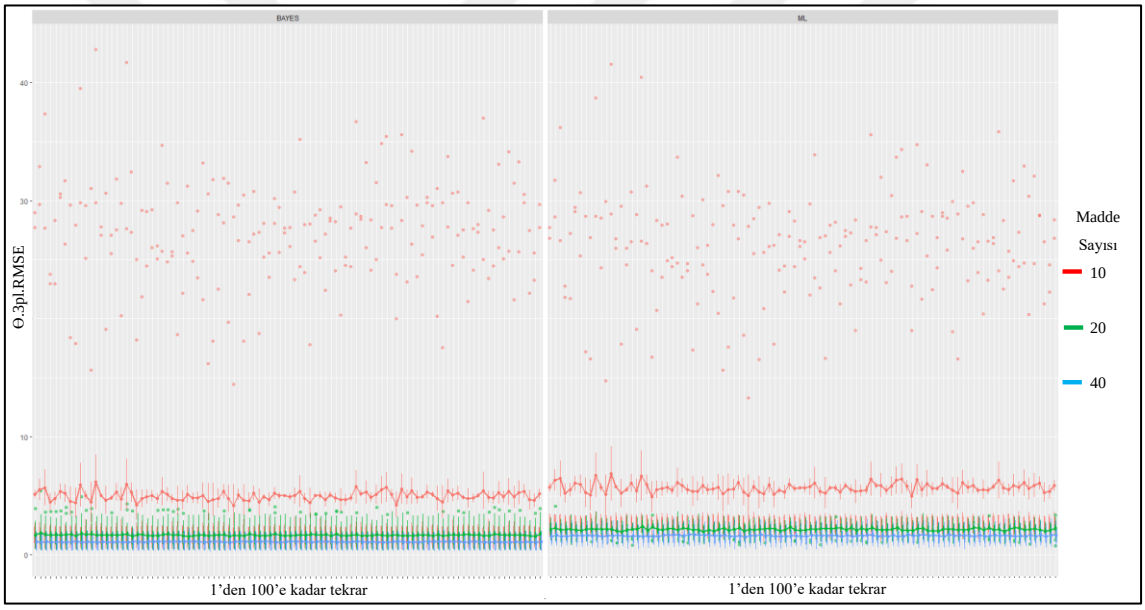
Tablo 15

3 PL Modelde Yetenek Parametresi RMSE Değerlerinin Madde Sayısına Göre İkili Karşılaştırmaları

Madde sayısı	Fark	Standart hata	t	p
10 – 20	3.429		2.663	.025*
10 – 40	3.988	1.288	3.096	.007**
20 – 40	0.558		0.434	.902

*p<.05 **p<.01

Tablo 15 incelendiğinde 3 PL modelde yetenek parametresi RMSE değerleri üzerinde madde sayısı 10 ile 20 *(t=2.663; p<.05)*, 10 ile 40 *(t=3.096; p<.01)* arasında manidar farklılıklar bulunmaktadır. 3 PL modelde yetenek parametresi üzerinde madde sayısına göre RMSE değişimi Şekil 6'da verilmiştir.



Şekil 6. 3 PL Modelde Yetenek Parametresi RMSE Değerlerinin Kestirim Yöntemlerine Göre Değişimi

Şekil 6 incelendiğinde 3 PL modelde yetenek parametresi kestirimleri üzerinde madde sayısı artıkça RMSE değerleri azalma göstermektedir. Aynı zamanda madde sayısının azaldığı durumlarda Bayes kestirim yöntemi klasik kestirime göre daha düşük değerler almıştır. Ancak bu durum manidar bulunmamıştır. Bu nedenle 3 PL modelde yetenek parametreleri kestirilirken klasik ya da Bayes yöntemlerin kullanılması bir fark oluşturmamaktadır. Ancak hangi yöntem kullanılırsa kullanılsın madde sayısındaki artış yetenek parametrelerindeki RMSE değerlerinde azalmaya neden olmaktadır. Özellikle madde sayısı 10'a düştüğünde yetenek parametrelerindeki RMSE değerleri manidar bir

şekilde artış göstermiştir. Buna göre 3 PL modelde yetenek parametreleri için düşük RMSE değerleri madde sayısının 20 ve 40 olduğu durumlarda gözlenmiştir.

3 PL modelde yetenek parametresi üzerinde önsel yetenek dağılım biçimine göre ikili karşılaştırmalar Tablo 16'da verilmiştir.

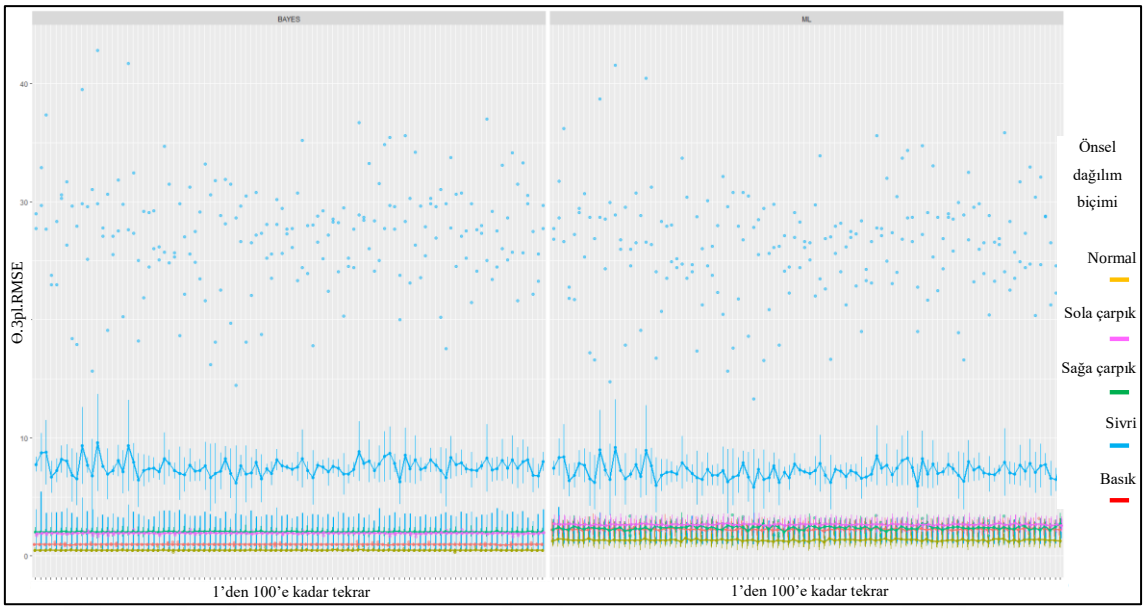
Tablo 16

3 PL Modelde Yetenek Parametresi RMSE Değerlerinin Önsel Yetenek Dağılımına Göre İkili Karşılaştırmaları

Önsel yetenek dağılımı	Fark	Standart hata	t	p
Normal – Sola çarpık	-1.376		-0.858	.911
Normal – Sağa çarpık	-1.300		-0.811	.926
Normal – Sivri	-6.463		-4.030	.001**
Normal – Basık	-0.697		-0.434	.992
Sola çarpık – Sağa çarpık	0.076	1.604	0.047	.999
Sola çarpık – Sivri	-5.088		-3.172	.017*
Sola çarpık – Basık	0.679		0.424	.993
Sağa çarpık – Sivri	-5.163		-3.219	.015*
Sağa çarpık – Basık	0.604		-0.376	.996
Sivri – Basık	5.767		3.595	.004**

*p<.05 **p<.01

Tablo 16 incelendiğinde 3 PL modelde yetenek parametresi RMSE değerleri üzerinde önsel yetenek dağılım biçimi normal ile sivri *(t=-4.030; p<.01)*, sola çarpık ile sivri *(t=-3.172; p<.05)*, sağa çarpık ile sivri *(t=-3.219; p<.05)*, basık ile sivri *(t=3.595; p<.01)* arasında manidar farklılıklar bulunmuştur. 3 PL modelde yetenek parametreleri üzerinde önsel yetenek dağılımı sivrileştikçe RMSE değerleri artış göstermektedir. 3 PL modelde yetenek parametresi üzerinde önsel yetenek dağılım biçimine göre RMSE değişimi Şekil 7'de verilmiştir.



Şekil 7. 3 PL Modelde Yetenek Parametresi RMSE Değerlerinin Önsel Yetenek Dağılım Biçimlerine Göre Değişimi

Şekil 7 incelendiğinde 3 PL modelde önsel yetenek dağılım biçimi sivrileştikçe yetenek parametreleri RMSE değerleri daha yüksek değerler almaktadır. Ancak yetenek parametreleri kestirim yöntemine göre önsel yetenek dağılım biçimleri üzerinde sivri dağılım hariç diğer dağılım biçimleri bir farklılaştırma oluşturmamıştır.

2PL modelde olduğu gibi yetenek parametrelerinin kestirimleri üzerinde önsel yetenek dağılımın sivri olması RMSE değerlerini manidar bir şekilde etkilemektedir. Bu durum hem klasik hem de Bayes kestirim yönteminde görülmektedir. Buna göre önsel yetenek dağılımın sivrileşmesi model (2 PL ya da 3 PL) fark etmeksizin yetenek parametrelerinin RMSE değerlerinde olumsuz bir etki oluşturmaktadır. Bu durumun nedeni olarak sivri dağılımdan (çok düşük standart sapma ve dar bir ranj değeri) ve simülasyon gerçekleştirilirken üretilen verilerin yapısında kaynaklandığı düşünülmektedir. Bu nedenle önsel yeteneğin sivri olduğu durumlar, MTK parametre kestirimleri çerçevesinde daha detaylı incelenmelidir.

Önsel Yetenek Dağılımı Normal Olan ve Olmayan 2 PL Modeldeki Verilerde Örneklem Büyüklüğü, Madde Sayısına Göre Klasik (En Çok Olabilirlik) ve Bayes Yöntemleri ile Kestirilen Madde Parametrelerinin RMSE Değerlerinin İncelenmesi

Önsel yetenek dağılımı normal olan verilerde klasik ve Bayesçi parametre kestirim yöntemleri sonucu elde edilen madde parametrelerinin (a_i, b_i ve c_i) ve kestirimin standart hatalarının (a_{SE}, b_{SE} ve c_{SE}) ortalamaları EK 9'da verilmiştir.

Araştırmanın üçüncü probleminde belirtilen önsel yetenek dağılımı normal olan ve olmayan (sola çarpık, sağa çarpık, sivri ve basık) 2 PL modeldeki verilerde örneklem büyüklüğü, madde sayısı ve kestirim yöntemine göre madde parametreleri RMSE değişimleri karma model ANOVA ile analiz edilmiştir. Analiz sürecinde kurulan modeller ve bu modellere ait örnek R kodları aşağıda verilmiştir:

```
#######a parametresi için...
#2PL modelde madde ayırt edicilik parametresi RMSE değerleri üzerinde örneklem büyüklüğünün ve kestirim yönteminin etkisi

data.a.2PL.item.model1 <- aov_ez(data = data.2PL.item, dv = "item.a.2PL.rmse", id = "item.id.2PL", between = "item.sample.size",
                   within = "item.replication")

print(ascii(nice(data.a.2PL.item.model1), include.rownames = FALSE, caption = "ANOVA 1"), type = "org")

summary(as.glht(pairs(emmeans(data.a.2PL.item.model1, ~item.sample.size))))

data.a.2PL.item.model1.1 <- aov_ez(data = data.2PL.item, dv = "item.a.2PL.rmse", id = "item.id.2PL",
                   between = c("item.sample.size", "item.estimation.method"),
                   within = "item.replication")

print(ascii(nice(data.a.2PL.item.model1.1), include.rownames = FALSE, caption = "ANOVA 1"), type = "org")

afex_plot(data.a.2PL.item.model1.1,x="item.replication",trace="item.sample.size", panel = "item.estimation.method", mapping = c("color","fill"),error = "within", data_geom = geom_boxplot, data_arg = list(width=0.4), point_arg = list(size=1.5), line_arg = list(size=1))
```

Örnek R kodları, 2 PL modelde sadece madde ayırt edicilik parametresi RMSE değerleri üzerinde örneklem büyüklüğünün ve kestirim yönteminin etkisi için verilmiştir. Diğer koşullar için aynı şekilde kodlar üretilmiştir.

Buna göre önsel yetenek dağılımı normal olan ve olmayan (sola çarpık, sağa çarpık, sivri ve basık) 2 PL modeldeki verilerde örneklem büyüklüğü, madde sayısı ve

kestirim yöntemine göre madde parametreleri RMSE değerleri için gerçekleştirilen karma model ANOVA sonuçları Tablo 17'de verilmiştir.

Tablo 17

Önsel Yetenek Dağılımı Normal Olan ve Olmayan 2 PL Modeldeki Verilerde Madde Parametreleri RMSE Değerleri İçin Karma Model ANOVA Sonuçları

Bağımsız değişkenler	Hata kareleri ortalaması	Serbestlik derecesi	F	p	Genelleştirilmiş η^2
Madde ayırt edicilik (a_{RMSE})					
Kestirim yöntemi (K)	5385.60	1	8.17	.005**	.045
Örneklem büyüklüğü (Ö)	4935.31	2	8.97	.001**	.090
Madde sayısı (M)	5939.41	2	0.10	.905	.001
Önsel yetenek dağılım biçimi (D)	5141.71	4	3.93	.006**	.083
K * Ö	3891.39	2	7.52	.001**	.070
K * M	5621.35	2	0.05	.952	.001
K * D	4210.31	4	3.34	.014**	.069
Hata	53.68	198			
Toplam	35178				
Madde güçlük (b_{RMSE})					
Kestirim yöntemi (K)	5827.21	1	1.26	.264	.002
Örneklem büyüklüğü (Ö)	5706.21	2	2.08	.131	.007
Madde sayısı (M)	5900.27	2	0.58	.562	.002
Önsel yetenek dağılım biçimi (D)	5606.80	4	1.94	.111	.014
K * Ö	5597.01	2	1.69	.191	.006
K * M	5931.07	2	0.65	.523	.003
K * D	5244.57	4	2.37	.060	.017
Hata	311.54	198			
Toplam	40124.68				

*p<.05 **p<.01

Tablo 17 incelendiğinde önsel yetenek dağılımı normal olan ve olmayan 2 PL modeldeki verilerde madde ayırt edicilik parametresi RMSE (a_{RMSE}) değerleri için karma model ANOVA sonuçlarına göre bağımsız değişkenlerin ana etkileri olarak kestirim yöntemi *($F_{(1, 88)} = 8.17; p<.01, \eta^2 = .045$)*, örneklem büyüklüğü *($F_{(2, 87)} = 8.97; p<.01, \eta^2 = .090$)* ve önsel yetenek dağılım biçimi *($F_{(4, 85)} = 3.93; p<.01, \eta^2 = .083$)* manidar etkiye sahiptir. Madde sayısı *($F_{(2, 87)} = 0.10; p>.05, \eta^2 = .001$)* manidar bir farklılık oluşturmamıştır. Manidar etkiye sahip olan kestirim yöntemi düşük, örneklem büyüklüğü orta ve önsel yetenek dağılım biçimi orta düzeyde etki büyüklüğüne sahiptir.

2 PL modelde madde ayırt edicilik parametresi RMSE (a_{RMSE}) değerleri üzerinde ikili etkileşimlerden kestirim yöntemi ile örneklem büyüklüğü *($F_{(2, 84)} = 7.52; p<.01, \eta^2 = .070$)* ve kestirim yöntemi ile önsel yetenek dağılım biçimi *($F_{(4, 80)} = 3.34; p<.01, \eta^2 = .069$)* manidar farklılıklar oluşturmuştur. Bu ikili etkileşimler orta düzeyde etki büyüklüğüne sahiptirler.

Önsel yetenek dağılımı normal olan ve olmayan 2 PL modeldeki verilerde madde güçlük parametresi RMSE değerleri (b_{RMSE}) için karma model ANOVA sonuçlarına göre bağımsız değişkenlerin (koşulların) hiçbiri manidar farklılık oluşturmamıştır.

Bu sonuçlar, Harwell ve Janosky (1991) çalışmasındaki sonuçlarla benzerlik göstermektedir. Buna göre 2 PL modelde küçük örneklem ve kısa testler, Bayes kestirim için yeterli görülmektedir. Madde ayırt edicilik parametresi için önsel dağılım çarpıklaştıkça klasik kestirimde yanlılığın arttığı Stone (1992)'un çalışmasında belirtilmektedir. Bu çalışmada da madde ayırt edicilik parametresi için RMSE değerleri önsel yetenek dağılımın çarpıklaşmasından etkilenmektedir. Bu yönüyle bu iki çalışma benzer sonuçlar ortaya koymuştur. Madde parametrelerinin önsel dağılımdan etkilenmesi ve yüksek hata değerleri üretmesi Sass, Schimitt ve Walker (2005)'ın çalışmasında da görülmektedir.

2 PL modeldeki veriler için madde ayırt edicilik parametresi RMSE değerleri (a_{RMSE}) üzerinde manidar etkiye sahip olan kestirim yönteminin ikili karşılaştırmaları Tablo 18'de verilmiştir.

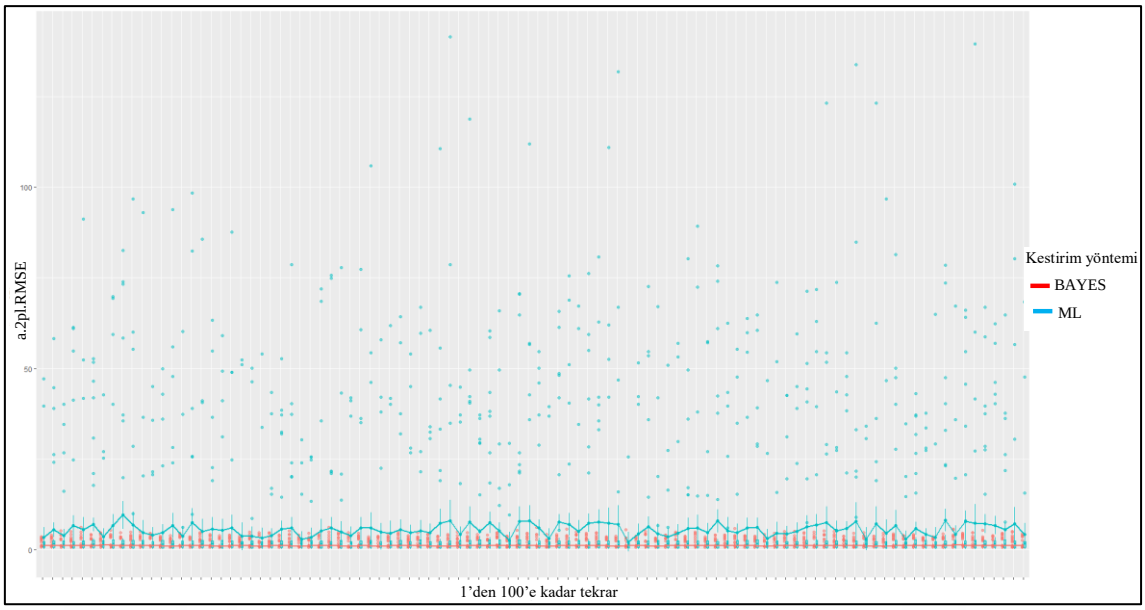
Tablo 18

2 PL Modelde Madde Ayırt Edicilik Parametresi RMSE Değerlerinin Kestirim Yöntemine Göre İkili Karşılaştırmaları

Kestirim yöntemi	Fark	Standart hata	t	p
BAYES-ML	-4.421	1.547	-2.858	.005**

*p<.05 **p<.01

Tablo 18 incelendiğinde önsel yetenek dağılım biçimi normal olan ve olmayan 2 PL modeldeki verilerde madde ayırt edicilik parametresi RMSE değerleri üzerinde kestirim yöntemi, bayes kestirim yöntemi lehine ve manidardır *(t=-2.858; p<.01).* Bayes kestirim klasik kestirime göre daha düşük RMSE değerleri üretmiştir. 2 PL modelde madde ayırt edicilik parametresi kestirim yöntemlerine ait RMSE değişimleri Şekil 8'de verilmiştir.



Şekil 8. 2 PL Modelde Madde Ayırt Edicilik Parametresi RMSE Değerlerinin Kestirim Yöntemlerine Göre Değişimi

Şekil 8 incelendiğinde 2 PL modelde araştırmanın tüm koşullarından bağımsız olarak madde ayırt edicilik parametresi RMSE değerlerinin klasik kestirime göre Bayes kestirimde daha düşük değerler aldığı görülmektedir. Klasik yöntemde gerçekleştirilen kestirim sonuçlarına göre madde ayırt edicilik parametresi RMSE (a_{RMSE}) değerleri bir saçılma gösterirken Bayes kestirimde bu değerler daha doğrusal ve stabildir. Buna göre madde ayırt edicik parametresinin kestiriminde Bayes yöntem klasik yönteme göre avantajlar sağlamaktadır.

2 PL modeldeki veriler için madde ayırt edicilik parametresi üzerinde manidar etkiye sahip örneklem büyüklüğünün ikili karşılaştırmaları Tablo 19'da verilmiştir.

Tablo 19

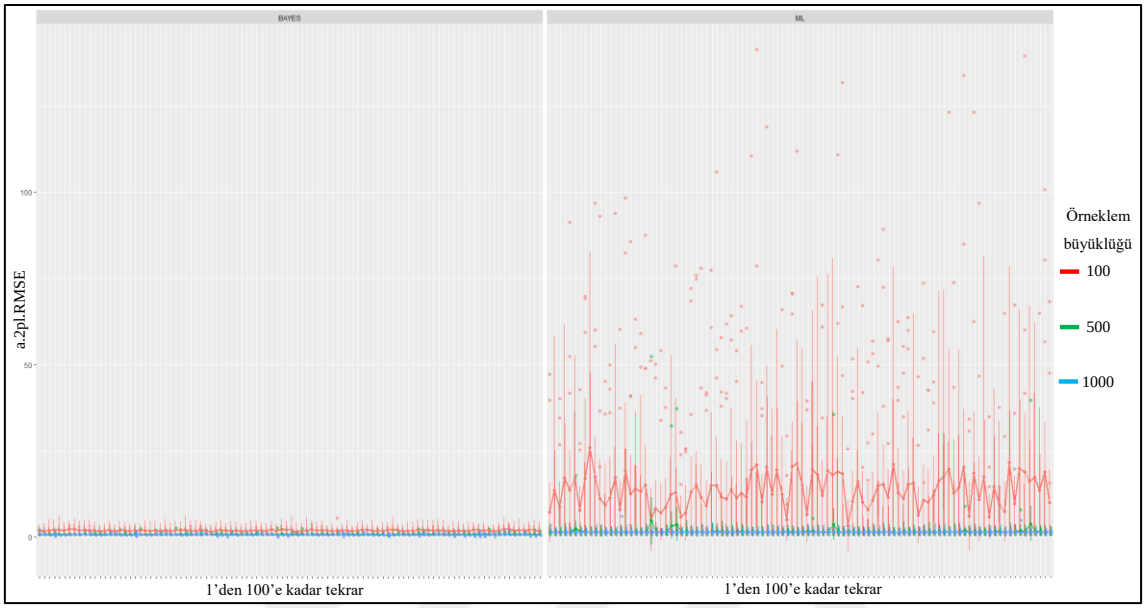
2 PL Modelde Madde Ayırt Edicilik Parametresi RMSE Değerlerinin Örneklem Büyüklüğüne Göre İkili Karşılaştırmaları

Örneklem büyüklüğü	Fark	Standart hata	t	p
100 – 500	6.578		3.626	.001**
100 – 1000	6.727	1.814	3.709	.001**
500 – 1000	0.149		0.082	.996

*p<.05 **p<.01

Tablo 19 incelendiğinde 2 PL modelde madde ayırt edicilik parametresi RMSE (a_{RMSE}) değerleri üzerinde örneklem büyüklüğü 100 ile 500 *(t=3.626; p<.01)*, 100 ile 1000 *(t=3.709; p<.01)* arasında manidar farklılıklar bulunmuştur. Buna göre örneklem

büyüklüğü artıkça madde ayırt edicilik parametresi RMSE (a_{RMSE}) değerleri azalmaktadır. 2 PL modelde madde ayırt edicilik parametresi üzerinde örneklem büyüklüğüne göre RMSE değişimi Şekil 9'da verilmiştir.



Şekil 9. 2 PL Modelde Madde Ayırt Edicilik Parametresi RMSE Değerlerinin Örneklem Büyüklüğüne Göre Değişimi

Şekil 9 incelendiğinde örneklem büyüklüğü azaldığında madde ayırt edicilik parametresi RMSE (a_{RMSE}) değerleri üzerinde Bayes kestirim yöntemi klasik kestirime göre daha düşük değerler üretmiştir. Özellikle klasik kestirimde örneklem büyüklüğü 100'e düştüğünde madde ayırt edicilik parametresi RMSE değerleri aşırı artış göstermiş ve saçılma oluşturmuştur. Bu durumda Bayes kestirim yöntemi kullanıldığında RMSE değerleri azalama eğilimindedir ve doğrusal bir dağılım göstermiştir. Örneklem büyüklüğü 500 ya da 1000 olduğunda RMSE değerleri kestirim yöntemine göre manidar bir farklılaşma göstermemiştir. Buradan anlaşıldığı üzere 2 PL modelde klasik kestirim yöntemi kullanıldığında madde ayırt edicilik parametresi RMSE değerlerini düşürmek için en az 500 kişilik bir örneklemin kullanılması gerekmektedir. Örneklem sayısı 100'e düştüğünde ise kestirim yöntemi olarak Bayes kullanılmalıdır. Aynı sonuçları Gao ve Chen (2005)'in çalışmasında da görmek mümkündür.

2 PL modelde madde ayırt edicilik parametresi üzerinde önsel yetenek dağılım biçimine göre ikili karşılaştırmalar Tablo 20'de verilmiştir.

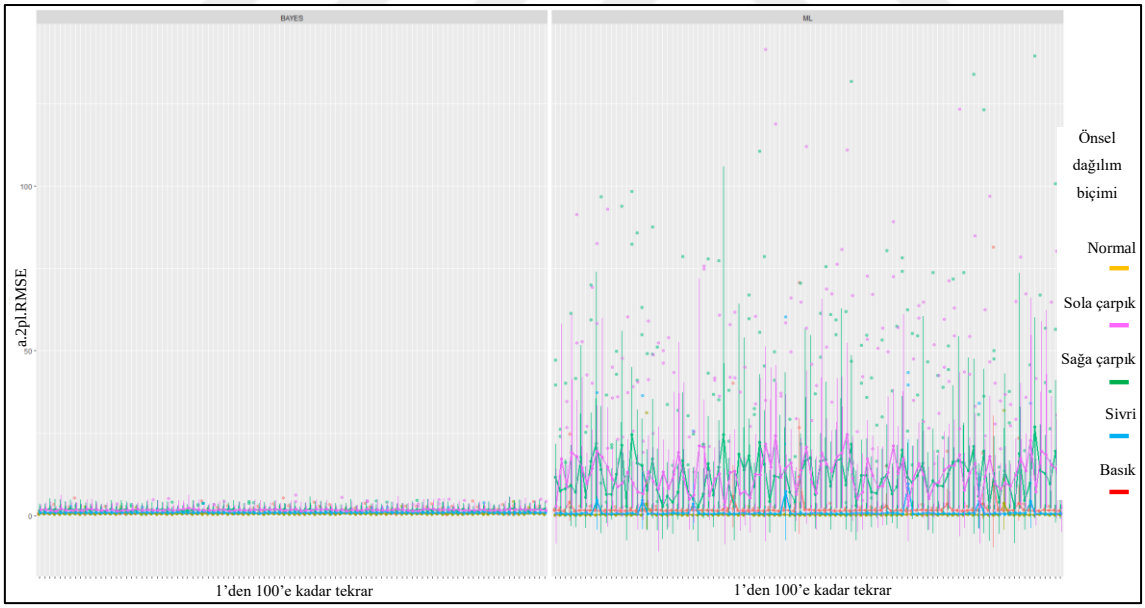
Tablo 20

2 PL Modelde Madde Ayırt Edicilik Parametresi RMSE Değerlerinin Önsel Yetenek Dağılım Biçimine Göre İkili Karşılaştırmaları

Önsel yetenek dağılımı	Fark	Standart hata	t	p
Normal – Sola çarpık	-6.746	2.390	-2.822	.046*
Normal – Sağa çarpık	-6.310		-2.640	.050*
Normal – Sivri	-0.494		-0.207	.999
Normal – Basık	-0.823		0.344	.999
Sola çarpık – Sağa çarpık	0.437		0.182	.999
Sola çarpık – Sivri	6.252		-2.616	.077
Sola çarpık – Basık	5.923		2.478	.105
Sağa çarpık – Sivri	5.816		2.433	.117
Sağa çarpık – Basık	5.487		0.344	.997
Sivri – Basık	-0.328		0.137	.999

*p<.05 **p<.01

Tablo 20 incelendiğinde 2 PL modelde madde ayırt edicilik parametresi RMSE (a_{RMSE}) değerleri üzerinde önsel yetenek dağılım biçimi normal ile sola çarpık *(t=-2.822; p<.05)* ve normal ile sağa çarpık *(t=-2.640; p<.05)* arasında manidar farklılıklar bulunmaktadır. 2 PL modelde madde ayırt edicilik parametresi üzerinde önsel yetenek dağılım biçimine göre RMSE değerleri Şekil 10'da verilmiştir.



Şekil 10. 2 PL Modelde Madde Ayırt Edicilik Parametresi RMSE Değerlerinin Önsel Yetenek Dağılım Biçimlerine Göre Değişimi

Şekil 10 incelendiğinde 2 PL modelde önsel yetenek dağılımı çarpıklaştıkça madde ayırt edicilik parametresi RMSE (a_{RMSE}) değerleri daha yüksek değerler almaktadır. Klasik kestirim yönteminde önsel yetenek dağılımı sola ya da sağa

çarpıklaştıkça madde ayırt edicilik parametresi RMSE (a_{RMSE}) değerleri artış göstermektedir. Ancak dağılımın sivri ya da basık olması madde ayırt edicilik parametresi RMSE değerleri üzerinde artırıcı bir etki yaratmamıştır.

Bununla birlikte önsel yetenek dağılımın sola ya da sağa çarpık olması durumunda madde ayırt edicilik parametresinin yüksek RMSE değerleri Bayes kestirim yöntemiyle daha aşağılara çekilmiştir. Buna göre 2 PL modelde madde ayırt edicilik parametresi RMSE (a_{RMSE}) değerleri Bayes kestirim yöntemi kullanıldığında önsel yetenek dağılım biçimimin farklılaşmasında etkilenmekte ve düşük değerler üretmektedir.

2 PL modelde madde güçlük parametresi RMSE (b_{RMSE}) değerleri koşulların hiçbirinde manidar bir farklılaşma göstermemiştir. Bu nedenle üçüncü araştırma probleminde yalnızca madde ayırt edicilik parametresi RMSE (a_{RMSE}) değerleri üzerindeki manidar koşullar verilmiştir.

Önsel Yetenek Dağılımı Normal Olan ve Olmayan 3 PL Modeldeki Verilerde Örneklem Büyüklüğüne, Madde Sayısına Göre Klasik (En Çok Olabilirlik) ve Bayes Yöntemleri ile Kestirilen Madde Parametrelerinin RMSE Değerlerinin İncelenmesi

Araştırmanın dördüncü probleminde belirtilen önsel yetenek dağılımı normal olan ve olmayan (sola çarpık, sağa çarpık, sivri ve basık) 3 PL modeldeki verilerde örneklem büyüklüğü, madde sayısı ve kestirim yöntemine göre madde parametreleri RMSE değişimleri karma model ANOVA ile analiz edilmiştir. Analiz sürecinde kurulan modeller ve bu modellere ait örnek R kodları aşağıda verilmiştir:

```
#######a parametresi için...
#3PL modelde madde ayırt edicilik parametresi RMSE değerleri üzerinde örneklem büyüklüğünün ve kestirim yönteminin etkisi

data.a.3PL.item.model1 <- aov_ez(data = data.3PL.item, dv = "item.a.3PL.rmse", id = "item.id.3PL", between = "item.sample.size",
                    within = "item.replication")

print(ascii(nice(data.a.3PL.item.model1), include.rownames = FALSE, caption = "ANOVA 1"), type = "org")

summary(as.glht(pairs(emmeans(data.a.3PL.item.model1,
~item.sample.size))))

data.a.3PL.item.model1.1 <- aov_ez(data = data.3PL.item, dv = "item.a.3PL.rmse", id = "item.id.3PL",
                    between = c("item.sample.size", "item.estimation.method"),
                    within = "item.replication")
```

```
print(ascii(nice(data.a.3PL.item.model1.1), include.rownames = FALSE,
caption = "ANOVA 1"), type = "org")

afex_plot(data.a.3PL.item.model1.1,x="item.replication",trace="item.sa
mple.size", panel = "item.estimation.method", mapping =
c("color","fill"),error = "within", data_geom = geom_boxplot, data_arg
= list(width=0.4), point_arg = list(size=1.5), line_arg = list(size=1))
```

Örnek R kodları, 3 PL modelde sadece madde ayırt edicilik parametresi RMSE değerleri üzerinde örneklem büyüklüğünün ve kestirim yönteminin etkisi için verilmiştir. Diğer koşullar için aynı şekilde kodlar üretilmiştir.

Buna göre önsel yetenek dağılımı normal olan ve olmayan (sola çarpık, sağa çarpık, sivri ve basık) 3 PL modeldeki verilerde örneklem büyüklüğü, madde sayısı ve kestirim yöntemine göre madde parametreleri RMSE değerleri için gerçekleştirilen karma model ANOVA sonuçları Tablo 21'de verilmiştir.

Tablo 21

Önsel Yetenek Dağılımı Normal Olan ve Olmayan 3 PL Modeldeki Verilerde Madde Parametreleri RMSE Değerleri İçin Karma Model ANOVA Sonuçları

Bağımsız değişkenler	Hata kareleri ortalaması	Serbestlik derecesi	F	p	Genelleştirilmiş η^2
Madde ayırt edicilik (a_{RMSE})					
Kestirim yöntemi (K)	98140.76	1	28.61	.001**	.203
Örneklem büyüklüğü (Ö)	119088.76	2	4.55	.013*	.078
Madde sayısı (M)	129940.17	2	0.53	.588	.010
Önsel yetenek dağılım biçimi (D)	103482.30	4	6.40	.001**	.192
K * Ö	79972.48	2	5.22	.007**	.085
K * M	99643.32	2	0.64	.530	.012
K * D	44745.66	4	13.46	.001**	.295
Hata	273.87	198			
Toplam	675287.32				
Madde güçlük (b_{RMSE})					
Kestirim yöntemi (K)	1149170.54	1	1.82	.180	.001
Örneklem büyüklüğü (Ö)	1132417.18	2	2.08	.132	.001
Madde sayısı (M)	117180079	2	0.54	.582	.001
Önsel yetenek dağılım biçimi (D)	1106887.16	4	2.06	.093	.001
K * Ö	1095077.60	2	2.03	.138	.001
K * M	1173689.34	2	0.54	.586	.001
K * D	1037641.31	4	2.16	.081	.001
Hata	797927.50	198			
Toplam	8664611.42				
Şans (c_{RMSE})					
Kestirim yöntemi (K)	0.06	1	9.10	.003**	.074
Örneklem büyüklüğü (Ö)	0.06	2	2.49	.089	.043
Madde sayısı (M)	0.07	2	0.50	.606	.009
Önsel yetenek dağılım biçimi (D)	0.04	4	13.00	.001**	.306
K * Ö	0.06	2	2.11	.127	.037
K * M	0.06	2	0.32	.727	.006
K * D	0.03	4	4.11	.004**	.117
Hata	0.00	198			
Toplam	0.38				

*p<.05 **p<.01

Tablo 21 incelendiğinde önsel yetenek dağılımı normal olan ve olmayan 3 PL modeldeki verilerde madde ayırt edicilik parametresi RMSE (a_{RMSE}) değerleri için karma model ANOVA sonuçlarına göre bağımsız değişkenlerin ana etkileri olarak kestirim yöntemi *($F_{(1, 88)} = 28.61; p<.01, \eta^2 = .203$)*, örneklem büyüklüğü *($F_{(2, 87)} = 4.55; p<.05, \eta^2 = .078$)* ve önsel yetenek dağılım biçimi *($F_{(4, 85)} = 6.40; p<.01, \eta^2 = .192$)* manidar farklılıklar oluşturmuştur. Madde sayısı *($F_{(2, 87)} = 0.53; p>.05, \eta^2 = .010$)* manidar bir farklılık oluşturmamıştır. Manidar etkiye sahip olan kestirim yöntemi yüksek, örneklem büyüklüğü orta ve önsel yetenek dağılım biçimi yüksek düzeyde etki büyüklüğüne sahiptir.

3 PL modelde madde ayırt edicilik parametresi RMSE değerleri üzerinde ikili etkileşimlerden kestirim yöntemi ile örneklem büyüklüğü *($F_{(2, 84)} = 5.22; p<.01, \eta^2 = .085$)* manidar ve orta düzeyde etki büyüklüğüne, kestirim yöntemi ile önsel yetenek dağılım biçimi *($F_{(4, 80)} = 13.46; p<.01, \eta^2 = .295$)* manidar ve yüksek düzeyde etki büyüklüğüne sahiptir.

Önsel yetenek dağılımı normal olan ve olmayan 3 PL modeldeki verilerde madde güçlük parametresi RMSE (b_{RMSE}) değerleri için karma model ANOVA sonuçlarına göre bağımsız değişkenlerin hiçbiri manidar farklılık oluşturmamıştır.

Önsel yetenek dağılımı normal olan ve olmayan 3 PL modeldeki verilerde şans parametresi RMSE (c_{RMSE}) değerleri için karma model ANOVA sonuçlarına göre bağımsız değişkenlerin ana etkileri olarak kestirim yöntemi *($F_{(1, 88)} = 9.10; p<.01, \eta^2 = .074$)* ve önsel yetenek dağılım biçimi *($F_{(4, 80)} = 13.00; p<.01, \eta^2 = .306$)* manidar farklılık oluşturmuştur. Örneklem büyüklüğü *($F_{(2, 87)} = 2.49; p>.05, \eta^2 = .043$)* ve madde sayısı *($F_{(2, 87)} = 0.50; p>.05, \eta^2 = .009$)* manidar farklılık oluşturmamıştır. Manidar farklılık oluşturan kestirim yöntemi orta düzeyde ve önsel yetenek dağılım biçimi yüksek düzeyde etki büyüklüğüne sahiptir.

3 PL modelde şans parametresi RMSE değerleri üzerinde ikili etkileşimlerden kestirim yöntemi ile önsel yetenek dağılım biçimi *($F_{(4, 80)} = 4.11; p<.01, \eta^2 = .117$)* manidar ve orta düzeyde etki büyüklüğüne sahiptir.

Bu sonuçlar incelendiğinde Swaminathan ve Gifford (1986)'ın önerisinin doğruluğu görülmektedir. Buna göre bu ilgili araştırmada, 3 PL model için parametre kestiriminde Bayes yöntem önerilmiştir. Bu çalışmada da 3 PL modeldeki madde parametrelerinin kestiriminde, özellikle madde ayırt edicilik parametresinde Bayes kestirim yöntemi kullanmanın avantaj sağladığı görülmektedir.

Aynı şekilde Gao ve Chen (2005)'in çalışmasında olduğu gibi örneklem büyüklüğünün 100'e kadar düştüğü durumlarda Bayes kestirimin madde parametrelerinin kestirimlerinde daha titiz sonuçlar verdiği bu çalışmada da görülmüştür.

3 PL modeldeki veriler için madde ayırt edicilik parametresi üzerinde manidar etkiye sahip kestirim yönteminin ikili karşılaştırmaları Tablo 22'de verilmiştir.

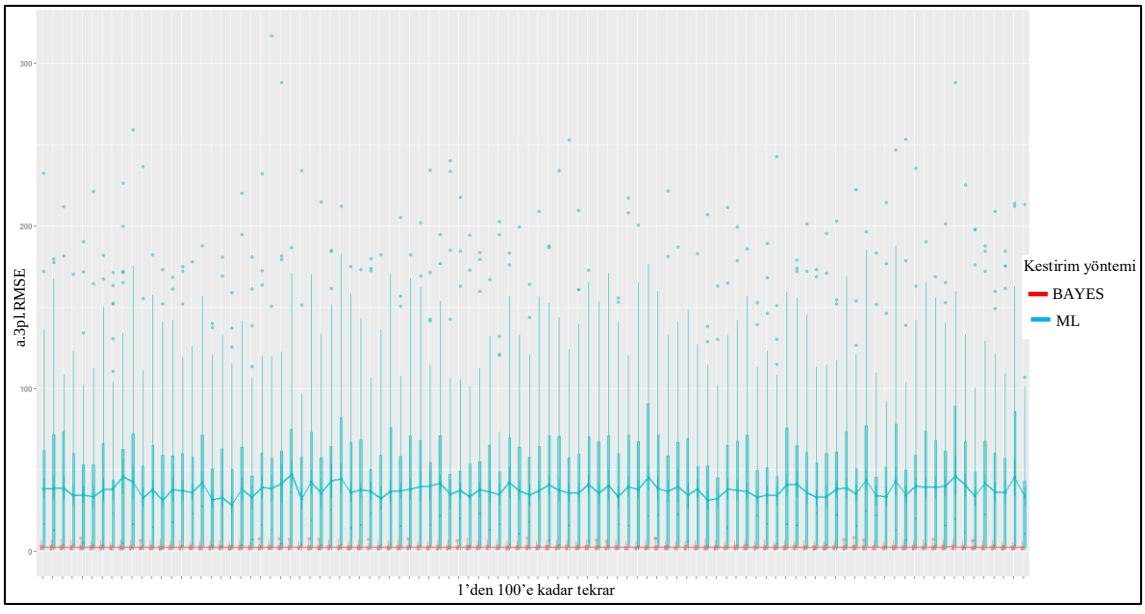
Tablo 22

3 PL Modelde Madde Ayırt Edicilik Parametresi RMSE Değerlerinin Kestirim Yöntemine Göre İkili Karşılaştırmaları

Kestirim yöntemi	Fark	Standart hata	t	p
BAYES-ML	-35.323	6.604	-5.348	.001**

*p<.05 **p<.01

Tablo 22 incelendiğinde önsel yetenek dağılım biçimi normal olan ve olmayan 3 PL modeldeki verilerde madde ayırt edicilik parametresi RMSE (a_{RMSE}) değerleri üzerinde bayes kestirim yöntemi lehine ve manidardır *(t=-2.858; p<.01)*. Bayes kestirim yöntemi klasik kestirim yöntemine göre daha düşük RMSE değerleri üretmiştir. 2 PL modelde madde ayırt edicilik parametresi kestirim yöntemlerine ait RMSE (a_{RMSE}) değişimleri Şekil 11'de verilmiştir.



Şekil 11. 3 PL Modelde Madde Ayırt Edicilik Parametresi RMSE Değerlerinin Kestirim Yöntemlerine Göre Değişimi

Şekil 11 incelendiğinde 3 PL modelde araştırmanın tüm koşullarından bağımsız olarak madde ayırt edicilik parametresi RMSE (a_{RMSE}) değerlerinin klasik kestirime göre Bayes kestirimde daha düşük değerler aldığı görülmektedir.

3 PL modeldeki veriler için madde ayırt edicilik parametresi RMSE (a_{RMSE}) değerleri üzerinde manidar etkiye sahip örneklem büyüklüğünün ikili karşılaştırmaları Tablo 23'te verilmiştir.

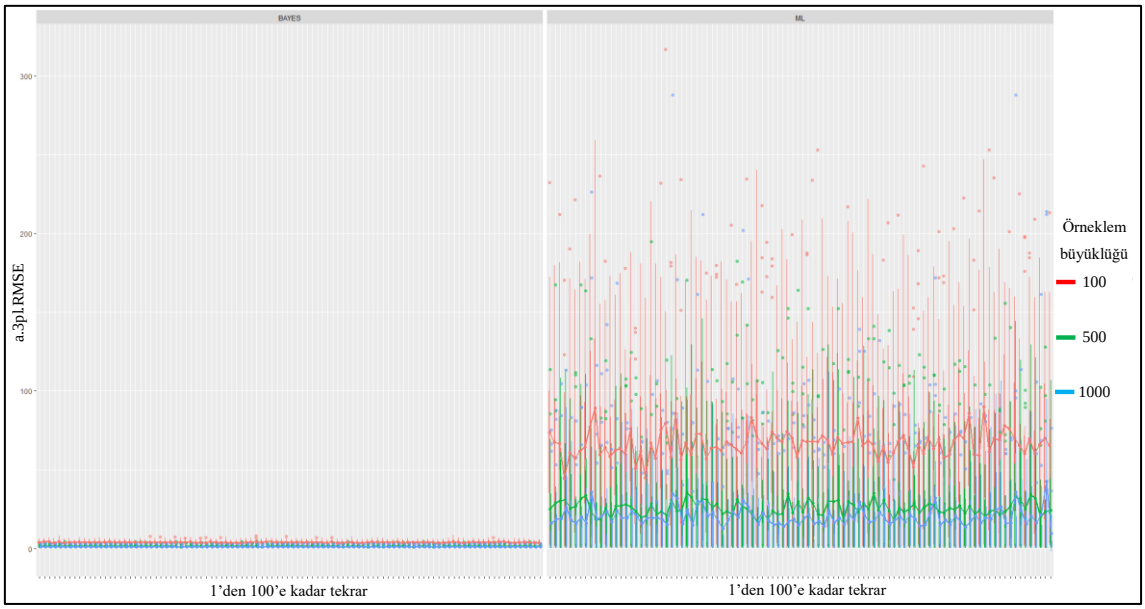
Tablo 23

3 PL Modelde Madde Ayırt Edicilik Parametresi RMSE Değerlerinin Örneklem Büyüklüğüne Göre İkili Karşılaştırmaları

Örneklem büyüklüğü	Fark	Standart hata	t	p
100 – 500	21.666		2.432	.044*
100 – 1000	24.598	8.910	2.761	.019*
500 – 1000	2.932		0.329	.924

*p<.05 **p<.01

Tablo 23 incelendiğinde 3 PL modelde madde ayırt edicilik parametresi RMSE (a_{RMSE}) değerleri üzerinde örneklem büyüklüğü 100 ile 500 *(t=2.432; p<.05)*, 100 ile 1000 *(t=2.761; p<.05)* arasında manidar farklılıklar bulunmuştur. Buna göre örneklem büyüklüğü artıkça madde ayırt edicilik RMSE (a_{RMSE}) değerleri azalmaktadır. 3 PL modelde madde ayırt edicilik parametresi üzerinde örneklem büyüklüğüne göre RMSE değişimi Şekil 12'de verilmiştir.



Şekil 12. 3 PL Modelde Madde Ayırt Edicilik Parametresi RMSE Değerlerinin Örneklem Büyüklüğüne Göre Değişimi

Şekil 12 incelendiğinde örneklem büyüklüğü azaldığında madde ayırt edicilik parametresi üzerinde Bayes kestirim yöntemi klasik kestirime göre daha düşük RMSE değerleri üretmiştir. Klasik kestirimde örneklem büyüklüğü azaldıkça madde ayırt edicilik RMSE değeri artış göstermektedir. Ayrıca bu değerler saçılma göstermektedir. Bu durum 2 PL modelde elde edilen sonuçlarla benzerdir. Örneklem büyüklüğü artış gösterdikçe bu değer azalmaktadır. Ancak Bayes yönteminin klasik yönteme göre madde ayrıt edicilik parametresi RMSE değerlerini azaltma eğiliminde olduğu görülmektedir. Bayes kestirimde örneklem büyüklüğünün artışı madde ayırt edicilik parametresi RMSE (a_{RMSE}) değerleri üzerinde bir farklılaşma oluşturmamıştır. Örneklem büyüklüğüne göre elde edilen RMSE değerleri doğrusal şekildedir. Yani Bayes yöntemi klasik kestirime göre madde ayırt edicilik parametresi RMSE (a_{RMSE}) değerlerini hem düşürmüş hem de stabil hale getirmiştir.

Benzer sonuçlar Chuah, Drasgow ve Luecht (2006)'ın çalışmasında da görülmektedir. İlgili çalışmada örneklem büyüklüğü 300'ün altına düştüğünde madde ayırt edicilik parametreleri aşırı şişkinlik göstermiştir. Aynı durumla bu çalışmada da karşılaşılmıştır. Bu duruma çözüm Bayes kestirim yöntemi ile bulunmuştur.

3 PL modelde madde ayırt edicilik parametresi RMSE (a_{RMSE}) değerleri üzerinde önsel yetenek dağılım biçimine göre ikili karşılaştırmalar Tablo 24'te verilmiştir.

Tablo 24

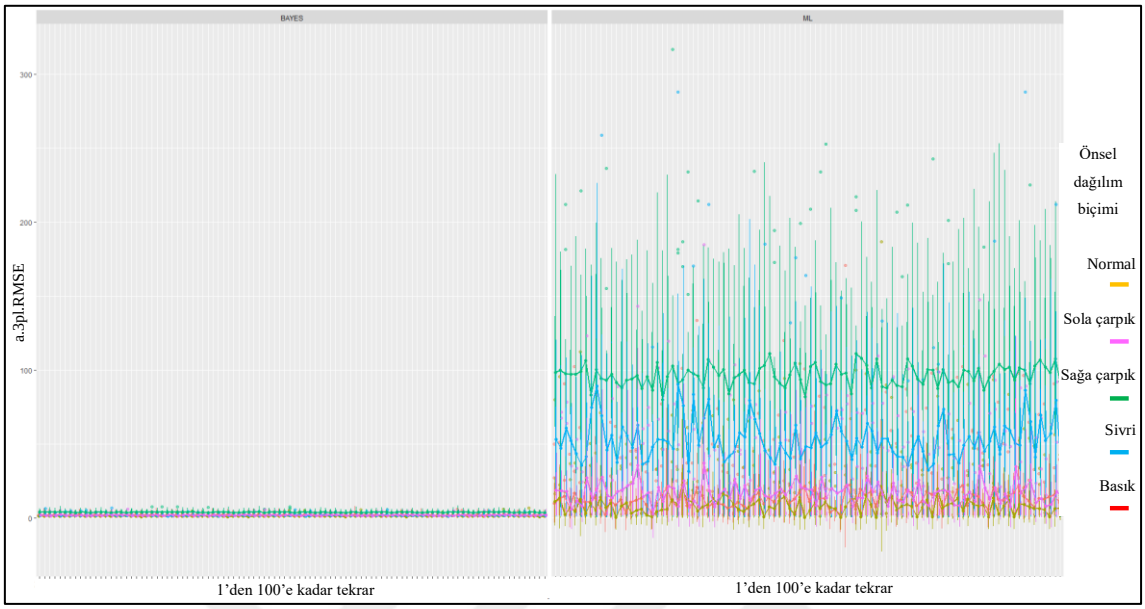
3 PL Modelde Madde Ayırt Edicilik Parametresi RMSE Değerlerinin Önsel Yetenek Dağılım Biçimine Göre İkili Karşılaştırmaları

Önsel yetenek dağılımı	Fark	Standart hata	t	p
Normal – Sola çarpık	-4.715		-0.440	.992
Normal – Sağa çarpık	-45.442		-4.238	.000**
Normal – Sivri	-22.701		-2.117	.222
Normal – Basık	-2.580		-0.241	.999
Sola çarpık – Sağa çarpık	-40.727	10.723	-3.798	.002**
Sola çarpık – Sivri	-17.986		-1.677	.453
Sola çarpık – Basık	2.135		0.199	.999
Sağa çarpık – Sivri	22.741		2.121	.221
Sağa çarpık – Basık	42.862		3.997	.001**
Sivri – Basık	20.120		1.876	.338

*p<.05 **p<.01

Tablo 24 incelendiğinde 3 PL modelde madde ayırt edicilik parametresi RMSE (a_{RMSE}) değerleri üzerinde önsel yetenek dağılım biçimi normal ile sağa çarpık *(t=-4.238; p<.01)*, sola çarpık ile sağa çarpık *(t=-3.798; p<.01)* ve sağa çarpık ile basık *(t=3.997; p<.01)* arasında manidar farklılıklar bulunmaktadır. 3 PL modelde madde ayırt edicilik

parametresi üzerinde önsel yetenek dağılım biçimine göre RMSE değerleri Şekil 13'te verilmiştir.



Şekil 13. 3 PL Modelde Madde Ayırt Edicilik Parametresi RMSE Değerlerinin Önsel Yetenek Dağılım Biçimlerine Göre Değişimi

Şekil 13 incelendiğinde 3 PL modelde önsel yetenek dağılımı çarpıklaştıkça madde ayırt edicilik parametresi RMSE değerleri daha yüksek değerler almaktadır. Bu yüksek RMSE değerleri Bayes kestirim yöntemiyle daha aşağılara çekilmiştir. 3 PL modeldeki klasik kestirimde madde ayırt edicilik parametresi RMSE değerleri önsel yetenek dağılımı sağa çarpık olduğunda en yüksek sonuçları vermiştir. Bu durumu daha sonra sivri, sola çarpık, basık ve normal dağılım izlemiştir. Önsel yeteneğin sağa çarpık olması durumunda madde ayırt edicilik parametresi RMSE (a_{RMSE}) değerlerinin en yüksek değerleri almasında modelin 3 PL olması önemli bir etkendir. 2 PL modelden farklı olarak bu modelde üçüncü bir parametre olan şans parametresi (c_i) eklenmesiyle önsel yetenek dağılımlarının başlangıç noktasını değişmektedir. Bu nedenle bu durumdan en çok etkilenen sağa çarpık olarak üretilen önsel yetenek parametreleridir.

Bayes kestirim kullanıldığında madde ayırt edicilik parametresi RMSE (a_{RMSE}) değerleri klasik kestirime göre tüm önsel dağılım biçimlerinde düşük RMSE değerleri üretmiş ve bu değerler stabil bir şekilde dağılmıştır.

3 PL modelde madde güçlük parametresi RMSE (b_{RMSE}) değerleri bağımsız değişkenlerin (koşulların) hiçbirinde manidar bir farklılaşma göstermemiştir.

3 PL modeldeki veriler için şans parametresi RMSE (c_{RMSE}) değerleri üzerinde manidar etkiye sahip kestirim yönteminin ikili karşılaştırmaları Tablo 25'te verilmiştir.

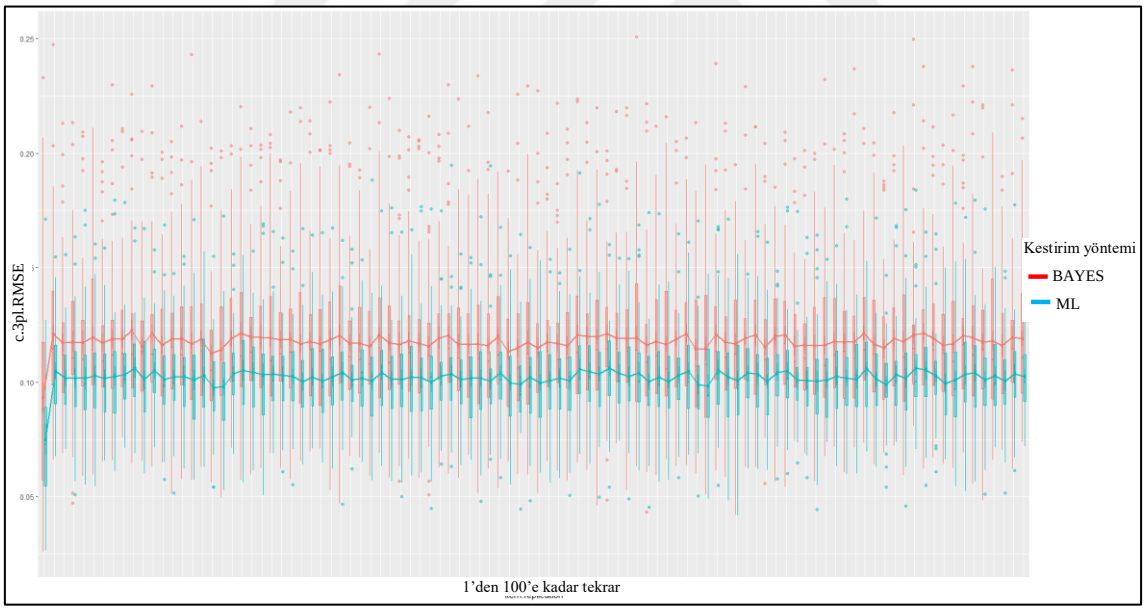
Tablo 25

3 PL Modelde Şans Parametresi RMSE Değerlerinin Kestirim Yöntemine Göre İkili Karşılaştırmaları

Kestirim yöntemi	Fark	Standart hata	t	p
BAYES-ML	0.016	0.005	3.016	.003**

*p<.05 **p<.01

Tablo 25 incelendiğinde önsel yetenek dağılım biçimi normal olan ve olmayan 3 PL modeldeki verilerde şans parametresi RMSE değerleri üzerinde klasik kestirim yöntemi lehine ve manidardır *(t=3.016; p<.01).* Bayes kestirim yöntemi klasik kestirim yöntemine göre daha yüksek RMSE değerleri üretmiştir. 3 PL modelde şans parametresi kestirim yöntemlerine ait RMSE değişimleri Şekil 14'te verilmiştir.



Şekil 14. 3 PL Modelde Şans Parametresi RMSE Değerlerinin Kestirim Yöntemlerine Göre Değişimi

Şekil 14 incelendiğinde 3 PL modelde araştırma koşullarından bağımsız olarak şans parametresi RMSE değerlerinin Bayes kestirimde daha yüksek değerler aldığı görülmektedir. Diğer parametrelerin aksine şans parametrelerinin RMSE değerlerinin düşmesinde klasik kestirim Bayes kestirime göre daha etkili olmuştur. Alanyazında şans parametresi ile ilgili çok fazla çalışma bulunmamaktadır. Sonucun bu şekilde çıkmasında

önsel fonksiyon oluşturulurken şans parametresi için tanımlanan dağılım biçiminden kaynakladığını düşünülmektedir.

3 PL modeldeki veriler için madde şans parametresi üzerinde manidar etkiye sahip önsel yetenek dağılım biçimi ikili karşılaştırmaları Tablo 26'da verilmiştir.

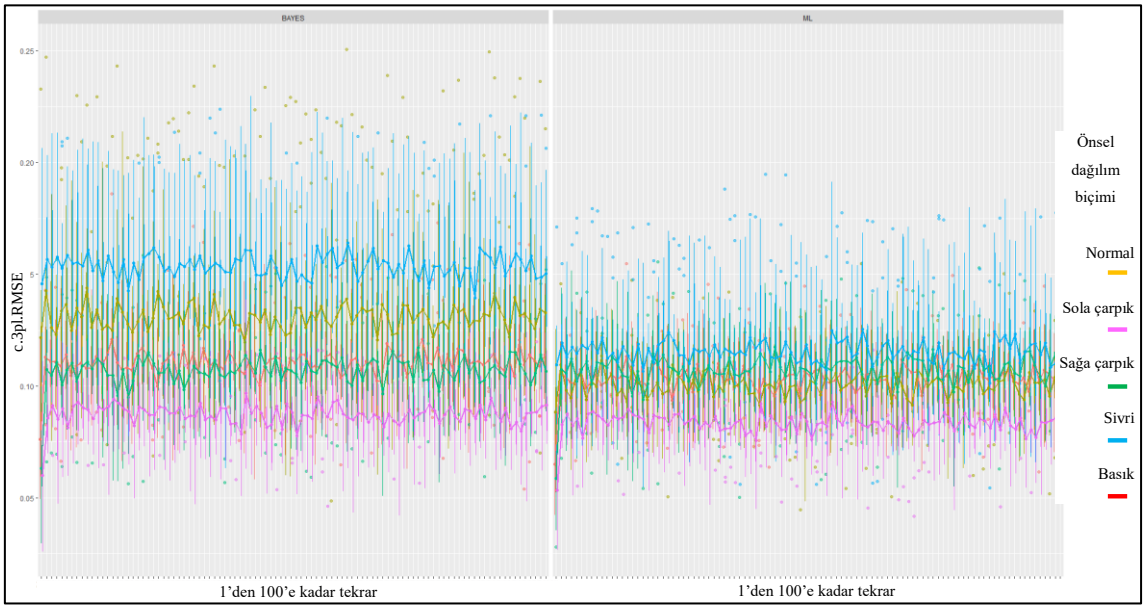
Tablo 26

3 PL Modelde Şans Parametresi RMSE Değerlerinin Önsel Yetenek Dağılım Biçimine Göre İkili Karşılaştırmalar

Önsel yetenek dağılımı	Fark	Standart hata	t	p
Normal – Sola çarpık	0.030	0.006	4.351	.000**
Normal – Sağa çarpık	0.008		1.255	.719
Normal – Sivri	0.019		-2.712	.060
Normal – Basık	0.008		1.265	.713
Sola çarpık – Sağa çarpık	-0.002		-3.097	.021*
Sola çarpık – Sivri	-0.004		-7.063	.000**
Sola çarpık – Basık	-0.002		-3.086	.022*
Sağa çarpık – Sivri	-0.002		-3.967	.001**
Sağa çarpık – Basık	0.000		0.010	.999
Sivri – Basık	0.002		3.977	.001**

*p<.05 **p<.01

Tablo 26 incelendiğinde 3 PL modelde şans parametresi RMSE değerleri üzerinde önsel yetenek dağılım biçimi normal ile sola çarpık *(t=4.351; p<.01)*, sola çarpık ile sağa çarpık *(t=-3.097; p<.05)*, sola çarpık ile sivri *(t=-7.063; p<.01)*, sola çarpık ile basık *(t=-3.086; p<.05)*, sağa çarpık ile sivri *(t=-3.967; p<.01)*, sivri ile basık *(t=3.977; p<.01)* arasında manidar farklılıklar bulunmaktadır. 3 PL modelde şans parametresi üzerinde önsel yetenek dağılım biçimine göre RMSE değerleri Şekil 15'te verilmiştir.



Şekil 15. 3 PL Modelde Şans Parametresi RMSE Değerlerinin Önsel Yetenek Dağılım Biçimlerine Göre Değişimi

Şekil 15 incelendiğinde 3 PL modelde önsel yetenek dağılımı çarpıklaştıkça şans parametresi RMSE değerleri daha yüksek değerler almaktadır. Ancak önsel yetenek dağılım biçimi sivrileştikçe şans parametresi RMSE değerleri Bayes kestirimde diğer madde parametrelerinin aksine artış göstermiştir. Buna göre 3 PL modelde önsel yetenek dağılımı sivrileştikçe klasik kestirim yöntemi daha düşük RMSE değerleri üretmiştir.

Ayrıca tüm önsel yetenek dağılımları için klasik kestirimdeki elde edilen RMSE değerleri Bayes kestirime göre daha dar bir alanda dağılım göstermiştir. Bayes kestirim sonucu elde edilen şans parametresi RMSE (c_{RMSE}) değerlerinin daha geniş bir alana yayılmasında başlangıç parametre değerlerinin normal dağılımdan farklı bir dağılım türüyle üretilmesinin etkili olduğu düşünülmektedir. Standart Bayes kestirimleri bilindiği üzere sonsal dağılımı normalleştirme eğilimdedir.

Şahin ve Anıl (2017) çalışmasında en yüksek hata değerlerini şans parametresinde elde etmişlerdir. Bu hata değerinin azalmasının ancak örneklem büyüklüğünün artmasıyla mümkün olduğunu bulmuşlardır. Bu çalışmada ise örneklem büyüklüğünün değişmesi şans parametresi RMSE (c_{RMSE}) değerleri üzerinde manidar farklılaşma oluşturmamıştır.

Genel olarak araştırmanın tüm problem durumlarındaki araştırılan koşullar göz önüne alındığında aşağıdaki özet tablolar ortaya çıkmaktadır. Yetenek parametreleri RMSE değerlerinin incelendiği birinci ve ikinci araştırma problemlerinde verilen tüm koşulların manidarlık durumları Tablo 27'de verilmiştir.

Tablo 27

2 PL ve 3 PL Modeldeki Verilerde Yetenek Parametreleri RMSE Değerleri Koşullara Bağlı Özet Bilgiler

Araştırma problemleri		Koşullar						
		Kestirim yöntemi (K)	Örneklem büyüklüğü (Ö)	Madde sayısı (M)	Önsel yetenek dağılımı (D)	K*Ö	K*M	K*D
2 PL	Θ_{RMSE}	+	-	+	+	-	-	+
3 PL	Θ_{RMSE}	-	-	+	+	-	-	-

+: Manidar farklılaşma VAR
-: Manidar farklılaşma YOK

Tablo 27 incelendiğinde 2 PL modeldeki verilerde yetenek parametreleri RMSE (Θ_{RMSE}) değerleri üzerinde kestirim yöntemi, madde sayısı, önsel yetenek dağılım biçimi ve kestirim yöntemi ile önsel yetenek dağılım biçimi ortak etkileşimi manidar farklılaşma

oluşturmuştur. 3 PL modeldeki verilerde yetenek parametreleri RMSE değerleri üzerinde madde sayısı ve önsel yetenek dağılım biçimi manidar farklılıklar yaratmıştır.

Yetenek parametrelerinin ve RMSE değerlerinin kestiriminde kestirim yöntemi sadece 2 PL modelde manidar bir farklılık yaratmıştır. Bu manidarlık Bayes kestirim yöntemi lehinedir. Çünkü klasik yöntemdeki elde edilen çok yüksek hata değerlerini daha düşük değerlere indirmiştir. Alanyazındaki çalışmalarda (Swaminathan ve Gifford, 1986; Harwell ve Janosky, 1991; Gao ve Chen, 2005; Finch ve Edwards, 2015) olduğu gibi benzer sonuçlar elde edilmiştir.

Madde parametreleri RMSE değerlerinin incelendiği üçüncü ve dördüncü araştırma problemlerinde verilen tüm koşulların manidarlık durumları Tablo 28'de verilmiştir.

Tablo 28

2 PL ve 3 PL Modeldeki Verilerde Madde Parametreleri RMSE Değerleri Koşullara Bağlı Özet Bilgiler

Araştırma problemleri		Koşullar						
		Kestirim yöntemi (K)	Örneklem büyüklüğü (Ö)	Madde sayısı (M)	Önsel yetenek dağılımı (D)	K*Ö	K*M	K*D
2 PL	a_{RMSE}	+	+	-	+	+	-	+
	b_{RMSE}	-	-	-	-	-	-	-
3 PL	a_{RMSE}	+	+	-	+	+	-	+
	b_{RMSE}	-	-	-	-	-	-	-
	c_{RMSE}	+	-	-	+	-	-	+

+: Manidar farklılaşma VAR
-: Manidar farklılaşma YOK

Tablo 28 incelendiğinde 2 PL modeldeki verilerde madde ayırt edicilik parametresi RMSE (a_{RMSE}) değerleri üzerinde kestirim yöntemi, örneklem büyüklüğü, önsel yetenek dağılım biçimi, kestirim yöntemi ile örneklem büyüklüğü ortak etkileşimi ve kestirim yöntemi ile önsel yetenek dağılım biçimi ortak etkileşimi manidar farklılıklar oluşturmuştur. 2 PL modeldeki verilerde madde güçlük parametresi RMSE (b_{RMSE}) değerleri üzerinde herhangi bir koşul manidar bir farklılaşmaya neden olmamıştır. 3 PL modeldeki verilerde madde ayırt edicilik parametresi RMSE (a_{RMSE}) değerleri üzerinde kestirim yöntemi, örneklem büyüklüğü, önsel yetenek dağılım biçimi, kestirim yöntemi ile örneklem büyüklüğü ortak etkileşimi ve kestirim yöntemi ile önsel yetenek dağılım biçimi ortak etkileşimi manidar farklılıklar oluşturmuştur. 3 PL modeldeki verilerde madde güçlük parametresi RMSE (b_{RMSE}) değerleri üzerinde hiçbir koşul manidarlık

yaratmamıştır. 3 PL modeldeki verilerde şans parametresi RMSE (c_{RMSE}) değerleri üzerinde kestirim yöntemi, önsel yetenek dağılım biçimi ve kestirim yöntemi ile önsel yetenek dağılım biçimi ortak etkileşimi manidar farklılıklar yaratmıştır.

Madde parametrelerinin ve RMSE değerlerinin kestiriminde kestirim yöntemi genellikle manidar bir farklılaşma ortaya koymuştur. Bu farklılaşma madde ayırt ediciliği (a_{RMSE}) ve şans (c_{RMSE}) parametreleri için model fark etmeksizin manidardır. Bu manidar farklılaşma madde ayırt edicilik parametresi için Bayes kestirim yöntemi; şans parametresi için klasik kestirim yöntemi lehinedir. Ancak madde güçlülük (b_{RMSE}) parametresi için kestirim yöntemine göre hem 2 PL hem de 3 PL modelde bir farklılaşmadan söz edilememektedir. Madde güçlük parametresindeki durum Kıbrıslıoğlu Uysal (2020)'in çalışmasıyla benzer sonuçlar vermiştir.

Örneklem büyüklüğü yetenek parametresinin kestiriminde manidar bir farklılık oluşturmazken madde sayısı, önsel yetenek dağılım biçimi ve kestirim yöntemi (sadece 2 PL modelde) RMSE değerleri üzerinde manidar farklılıklar oluşturmuştur. Örneklem büyüklüğünün yetenek parametresini kestirimi ve hata değerleri üzerinde etkili olmamasının nedenini kestirilen parametre sayısının tek bir adet olmasından kaynaklanmaktadır. Bu durum Goldman ve Raju (1986) ve Harwell ve Janosky (1991)'nin çalışmasıyla da benzerlik göstermektedir. Goldman ve Raju (1986)'nun çalışmasında kestirilen parametre sayısı 1'e düştüğünde örneklem büyüklüğünün 250 olmasının yeterli olacağını belirtmiştir. Harwell ve Janosky (1991)'de 15 maddelik ve 250 kişilik örneklemlerin yeterli olduğu sonucuna varmışlardır. Benzer durum Şahin (2012) çalışmasında da görülmektedir. Şahin (2012) 1 PL modelde parametre kestirimi yapabilmek için 150 kişilik örneklemin yeterli olacağı sonucuna varmıştır.

Örneklem büyüklüğü, sadece madde ayırt edicilik parametresi RMSE kestirimlerinde etkili olmuştur. Bu durum hem 2 PL hem de 3 PL model kullanıldığında geçerlidir. Madde ayırt edicilik parametresi RMSE (a_{RMSE}) değerleri örneklem büyüklüğünün artmasından olumlu etkilenmiş ve bu değerler azalma göstermiştir. Ancak örneklem büyüklüğü azaldıkça özellikle madde ayırt edicilik parametresi RMSE değerleri şişkinlik göstermiştir. Klasik yöntemle kestirim sonucu madde ayırt edicilik parametresi RMSE (a_{RMSE}) değerlerinde oluşan şişkinlik Chuah, Drasgow ve Luecht (2006)'in çalışmlarında da görülmüştür. Bu şişkinliğin azalmasında Bayes kestirim yöntemi önemli bir rol oynamıştır. Benzer durum Gao ve Chen (2005)'in araştırmasında da görülmektedir. Bu araştırmada örneklem sayısının 100'e düştüğü durumlarda Bayes

kestirimlerin marjinal en çok olabilirlik kestirimlerine göre daha doğru sonuçlar verdiği belirtilmiştir.

Madde sayısının artması sadece yetenek parametresi RMSE (Θ_{RMSE}) değerlerinin azalmasını sağlamıştır. Hatta madde sayısının 40 olduğu bazı durumlarda klasik ve Bayes yöntemle yetenek kestirim sonuçları birbirlerine yakın değerler almışlardır. Benzer şekilde, Gao ve Chen (2005)'in araştırmasında test uzunluğunun ve örneklem büyüklüğünün artması kestirimlerin standart hatalarını düzeltme eğiliminde olduğunu vurgulamışlardır. Ancak madde sayısı 10'a düştüğünde klasik kestirimde yetenek parametrelerinin RMSE değerlerinde şişkinliğe neden olmuş ancak bu durum Bayes kestirim yöntemiyle azaltılmıştır. Madde ayırt edicilik (a_i), madde güçlük (b_i) ve şans (c_i) parametreleri RMSE değerleri madde sayısının değişiminden hiçbir şekilde etkilenmemişlerdir.

Önsel yetenek dağılım biçimi yetenek parametreleri RMSE değerleri üzerinde manidar farklılıklar oluşturmuştur. Önsel yetenek dağılım biçimi yetenek parametrelerinde modelin parametre sayısına göre de manidar bir farklılaşma göstermemiştir. Hem 2 PL modelde hem de 3 PL modelde önsel dağılım biçimi madde ayırt edicilik (a_i) ve şans (c_i) parametreleri RMSE değerlerinde manidar bir farklılaşma göstermiştir. 2 PL ve 3 PL modelde madde güçlük parametresi RMSE (b_{RMSE}) değerleri üzerinde önsel yetenek dağılım biçimine göre bir farklılaşma görülmemiştir. Madde parametrelerinin önsel yetenek dağılım biçimine göre farklılaşması sola ve sağa çarpık dağılım üzerinde diğer dağılım türlerine oranla daha manidardır. Doğan (2002)'ın yaptığı benzer çalışmada dağılım biçimleri (çarpık ya da basık) MTK'nın parametre değişmezliği durumunu etkilemiştir. Özellikle farklılaşmanın çarpık dağılımlarda daha fazla olduğu belirtilmiştir. Benzer durum Seong (1990), Stone (1992), Kirisci ve diğerleri (2001), Sass ve diğerleri (2005) ve Karadavut (2019)'un çalışmalarında da görülmektedir.

Modelin parametre sayısının, yetenek ve madde parametreleri ve RMSE değerleri üzerinde etkili olduğu görülmüştür. 3 PL model, 2 PL modele göre daha yüksek kestirim RMSE değerleri oluşturmuştur. Bu değerlerdeki yüksekliğin düşürülmesinde Bayes kestirim yöntemi klasik yönteme göre daha etkili olmuştur.

Araştırmanın temel amacını oluşturan farklı koşullarda parametre kestirim yöntemi, yetenek ve madde parametreleri RMSE değerleri üzerinde manidar bir farklılaşma oluşturmuştur. Araştırmanın tüm koşullarında Bayes kestirim yönteminin klasik en çok olabilirlik kestirim yöntemine göre daha düşük RMSE değerleri elde ettiğini

göstermiştir. Ancak bu RMSE değerlerinin manidarlığı araştırmanın tüm koşullarında görülmemiştir.

RMSE, bir parametre kestiriminin kesinliğinin ve kestirim yanlılığının toplam hata göstergesidir (Thissen ve Wainer, 1983). Alanyazın incelendiğinde yaygın olarak kullanılan modeller (Rasch, 1 PL, 2 PL ve 3 PL) için parametre kestirimleri standart hataları için kapsamlı bir şekilde değinilmemiştir (Lord, 1980). Araştırmanın sonuçlarında belirtildiği üzere Bayes yöntemi yetenek ve madde parametreleri RMSE değerlerini klasik en çok olabilirlik yöntemine oranla daha düşük seviyelere indirmiştir.

Buna göre Bayes kestirim yöntemi klasik (en çok olabilirlik) kestirim yöntemine göre daha düşük parametre RMSE değerleri ürettiği için avantajlı görülmektedir. Özellikle klasik (en çok olabilirlik) kestirim yöntemi kullanıldığında küçük örneklemlerde ve kısa testlerde oluşan parametre RMSE değerlerindeki aşırı yükselişi azaltma eğiliminde olduğu görülmüştür.

BÖLÜM 5

SONUÇLAR VE ÖNERİLER

Bu bölümde araştırmanın bulgularından elde edilen sonuçlara ve buna göre uygulayıcılara ve araştırmacılara yönelik önerilere yer verilmiştir.

Sonuçlar

Bu çalışmanın sonuçları araştırmanın amacına bağlı olarak oluşturulan her bir problem için başlıklar halinde sunulmuştur. İlk iki başlık yetenek parametrelerinin kestirimi üzerine sonuçları açıklarken üçüncü ve dördüncü başlıklar ise madde parametrelerinin kestirim sonuçlarına değinmektedir.

Önsel Yetenek Dağılımı Normal Olan ve Olmayan 2 PL Verilerde Kestirilen Yetenek Parametrelerinin RMSE Değişim Sonuçları

Önsel yetenek dağılımı biçimi normal olan ve olmayan (sola çarpık, sağa çarpık, sivri ve basık) 2 PL modeldeki verilerde yetenek parametrelerinin kestirimleri üzerinde kestirim yöntemi olarak klasik ve Bayesçi yaklaşımlar kullanılmıştır. Bayes kestirim yöntemi klasik kestirim yöntemine göre bazı durumlarda avantajlar sağlamıştır. Bunlar örneklem büyüklüğü küçüldükçe ve madde sayısı azaldıkça kestirim hatalarında oluşan şişkinlik azalmıştır. Ancak bu değişkenler içerisinden sadece kestirim yöntemi, madde sayısı ve önsel yetenek dağılım biçimi RMSE değerlerindeki değişimlerde manidar farklılıklar oluşturmuştur. Özellikle madde sayısı azaldıkça (kısa testler) konusunda klasik kestirim yöntemin sonuçlarına bağlı olarak artış gösteren RMSE değerleri, Bayes kestirim yöntemi sayesinde daha düşük seviyelere çekilmiştir. Ayrıca önsel yetenek dağılımının normalden uzaklaşması (çarpıklaşması, sivrileşmesi ya da basıklaşması) durumunda klasik yöntem yüksek RMSE değerleri üretmektedir. Yetenek parametreleri RMSE değerleri önsel yetenek dağılım biçimindeki farklılaşma çoktan aza doğru; sağa çarpık, sola çarpık, sivri ve basık şeklinde sıralanmaktadır. Bayes kestirim yöntemi önsel dağılımdaki bu durumdan dolayı oluşan yüksek RMSE değerlerini daha düşük değerlere

indirmiştir. Bayes kestirim yöntemi bu açılardan klasik yönteme göre avantajlar sağladığı görülmektedir.

Önsel Yetenek Dağılımı Normal Olan ve Olmayan 3 PL Verilerde Kestirilen Yetenek Parametrelerinin RMSE Değişim Sonuçları

Örneklem büyüklüğündeki ve madde sayısındaki değişimin yanı sıra kullanılan lojistik modeldeki parametre sayısının artması da RMSE değerlerine etki etmektedir. Özellikle 2 PL modelden 3 PL modele geçişte yetenek parametrelerindeki kestirim değerleri ve bu kestirime bağlı oluşan RMSE değerleri değişim göstermektedir. Araştırma sonuçları bu değişimin genel anlamada RMSE değerlerinde artışa neden olduğunu göstermektedir. 3 PL modelde 2 PL modele göre yetenek parametreleri RMSE değerlerindeki manidar farklılaşma madde sayısının ve önsel yetenek dağılım biçiminin değişimlerinde görülmüştür. Kestirim yöntemi, 3 PL modelde yetenek parametreleri RMSE değerlerinde bir azalmaya neden olsa da manidar bir farklılaşma oluşturmamıştır. Aynı şekilde örneklem büyüklüğü de 2 PL modelde olduğu gibi bir RMSE değerlerinde farklılaşma oluşturmamıştır. 3 PL modeldeki yetenek parametrelerinde RMSE değişimleri madde sayısı azaldıkça (test kısaldıkça) artış göstermektedir. Bu artış aynı şekilde manidar bir şekilde olmasa da Bayes kestirim yöntemiyle azalma eğilimindedir. Önsel yetenek dağılım biçimindeki değişim ise özellikle dağılım sivrileştiğinde yetenek parametreleri RMSE değerlerinde aşırılık göstermektedir.

Önsel Yetenek Dağılımı Normal Olan ve Olmayan 2 PL Verilerde Kestirilen Madde Parametrelerinin RMSE Değişim Sonuçları

Önsel yetenek dağılımı normal olan ve olmayan 2 PL modeldeki verilerde madde parametrelerinden sadece madde ayırt edicilik parametresi RMSE (a_{RMSE}) değerleri araştırmanın bazı koşullarından etkilenmiştir. Madde güçlük parametresi RMSE (b_{RMSE}) değerleri bu koşulların hiçbirinden etkilenmemiştir. Madde ayırt edicilik parametresi RMSE (a_{RMSE}) değerleri kestirim yöntemine, örneklem büyüklüğüne ve önsel yetenek dağılım biçimlerine göre manidar farklılaşmalar göstermiştir. Kestirim yöntemi açısından Bayes kestirim yöntemi madde ayırt edicilik parametresi RMSE (a_{RMSE}) değerleri üzerinde klasik kestirime göre daha düşük değerler üretmiştir. Madde ayırt edicilik

parametresi RMSE değerlerindeki aşırılıklar Bayes kestirim yöntemi ile daha aşağılara çekilmiştir. Örneklem açısından örneklem azaldıkça madde ayırt edicilik parametresi RMSE (a_{RMSE}) değerleri daha da yükselmektedir. Ancak örneklem büyüklüğü artıkça bu değerler düşmekte ve manidar farklılaşma ortadan kalkmaktadır. Küçük örneklemlerde madde ayırt edicilik parametresi RMSE (a_{RMSE}) değerlerindeki aşırılıklar Bayes kestirim yöntemiyle klasik kestirime göre daha düşük değerler almıştır. Önsel yetenek dağılım biçiminin sola ya da sağa çarpıklaşması madde ayırt edicilik parametresi RMSE (a_{RMSE}) değerlerini manidar bir şekilde etkilemiştir. Dağılımın sivri ya da basık olma durumu madde ayırt edicilik parametresi RMSE (a_{RMSE}) değerlerini etkilememiştir. Dağılımdaki çarpıklaşmadan oluşan yüksek RMSE (a_{RMSE}) değerleri ise aynı şekilde Bayes kestirim yönteminin etkisiyle daha aşağılara çekilmiştir.

Önsel Yetenek Dağılımı Normal Olan ve Olmayan 3 PL Verilerde Kestirilen Madde Parametrelerinin RMSE Değişim Sonuçları

Önsel yetenek dağılımı normal olan ve olmayan 3 PL modeldeki verilerde madde şans parametresinin işe koşulmasıyla madde parametrelerindeki RMSE değişimleri de 2 PL modele göre farklılıklar göstermiştir. Madde ayırt edicilik parametresi RMSE (a_{RMSE}) değerleri kestirim yönteminden, örneklem büyüklüğünden ve önsel yetenek dağılım biçiminden manidar bir şekilde etkilenmiştir. Daha önceki durumlarda olduğu gibi RMSE değerleri kestirim yöntemi Bayes olduğunda daha düşük değeler almıştır. Örneklem durumunda ise büyüklük azaldıkça madde ayırt edicilik parametresi RMSE (a_{RMSE}) değerleri artış göstermektedir. Örneklem büyüklüğü artıkça RMSE (a_{RMSE}) değerleri azalmaktadır. Küçük örneklemlerdeki RMSE değerlerindeki aşırılıklar Bayes kestirim yöntemiyle daha aşağılara çekilmektedir. Madde ayırt edicilik parametresi RMSE (a_{RMSE}) değerlerini etkileyen bir başka değişken ise önsel yetenek dağılım biçimidir. Dağılım sağa çarpıklaştıkça diğer dağılım biçimleriyle manidar farklılaşmalar ortaya koymaktadır. Önsel yetenek dağılımından kaynaklanan bu farklılaşma aynı şekilde kestirim yöntemi olarak Bayes kullanılmasıyla yüksek RMSE (a_{RMSE}) değerlerinin düşürülmesini sağlamıştır. 2 PL modelde olduğu gibi 3 PL modelde de madde güçlük parametresi RMSE (b_{RMSE}) değerleri hiçbir koşulda manidar bir farklılaşma göstermemiştir. Ancak madde şans parametresi RMSE (c_{RMSE}) değerleri kestirim yönteminden madde ayırt edicilik parametresine RMSE (a_{RMSE}) değerlerine göre tam tersi bir şekilde farklılaşma

göstermiştir. Yani şans parametresi RMSE (c_{RMSE}) değerleri klasik kestirimde Bayes kestirime göre daha düşük değerler almıştır. Ayrıca şans parametresi RMSE (c_{RMSE}) değerleri önsel yetenek dağılım biçiminden etkilenmiştir. Buna göre dağılım çarpıklaştıkça, sivrileştikçe ya da basıklaştıkça şans parametresi RMSE (c_{RMSE}) değerleri farklılaşmıştır. Ancak madde şans parametresi RMSE (c_{RMSE}) değerlerindeki bu farklılaşma Bayes kestirimde klasik kestirime göre daha fazladır. Buna göre şans parametresi kestirimlerinde daha düşük RMSE (c_{RMSE}) değerleri isteniyorsa klasik kestirim kullanılması daha uygun görülmektedir.

Buna göre parametre kestiriminde Bayes yöntemini kullanmanın avantajları seçilen araştırma koşullarda açık bir şekilde görülmüştür. Alanyazında klasik kestirim yönteminin bazı durumlarda yüksek hata oranlarına sahip olması Bayesçi yöntemi ön plana çıkarmaktadır. Günümüzde artık sınıf içi başarı testlerinin geliştirilmesinde MTK'yı kullanmak mümkündür. Ancak ilk akla gelen durumlar küçük örneklemler ve kısa testlerde bunun nasıl gerçekleşeceğidir. Bayes yaklaşım bunu mümkün kılmakta ve kestirim hatalarını makul seviyelere indirmektedir. Ayrıca üzerinde çalışılan dağılımın her zaman normal olması mümkün değildir. Böyle bir durumda da klasik kestirim yöntemi titiz sonuçlar vermemektedir. Bu noktada Bayes kestirimin avantajlarından faydanılmaktadır. Bu çalışmanın sonuçları, klasik kestirimin titiz sonuçlar alamadığı noktalarda bir çözüm önerisi olarak Bayes kestirimin sunulabileceğini göstermiştir.

Öneriler

Araştırmanın sonuçlarına göre uygulayıcılara ve araştırmacılara yönelik öneriler aşağıda verilmiştir.

Uygulayıcılara Öneriler

1. Klasik en çok olabilirlik yönteminde elde edilecek yetenek ve madde parametreleri için düşük RMSE kestirimleri bekleniyorsa örneklem büyüklüğü ve madde sayısı uygun şekilde belirlenmelidir. Bu araştırma da bu açıdan alanyazındaki bu konudaki çalışmalarla benzerlik göstermiştir. Örneğin, 3 PL model için en az 1000 kişilik örneklem ve 20 maddelik bir test uzunluğu önerilmektedir.

2. Önsel yetenek dağılımın normal olduğu verilerde klasik (en çok olabilirlik) ve Bayes kestirimler benzer sonuçlar verse de çarpıklık ve basıklığın artığı önsel yetenek dağılımlarında yetenek ve madde parametreleri için özellikle 3 PL modelde Bayes kestirimin kullanılması önerilmektedir.

3. Bayes yöntemle parametre kestirimi sınıf içi küçük örneklemlerde, kısa testlerde ve farklı önsel yetenek dağılım biçimlerinde düşük RMSE değerleri ürettiği için MTK'ya dayalı test geliştirmelere olanak sağlayabilir.

4. Verilerde çarpıklık, sivrilik ya da basıklık arttıkça klasik (en çok olabilirlik) yöntemle parametre kestirimi yapılmaması önerilmektedir. Bunun yerine örneklem büyüklüğü, madde sayısı ve model dikkate alınarak Bayes kullanımı önerilmektedir.

5. Bazı R paketlerindeki (ltm, mirt gibi) MAP ve EAP gibi kestirimler sadece yetenek parametrelerini Bayes yöntemle kestirmekte ancak madde parametrelerini Bayes yöntemle kestirmemektedir. Bayes yöntemle kestirim yapan paketler hakkında daha ayrıntılı bilgi için Bürkner (2017, 2018) incelenebilir.

Araştırmacılara Öneriler

1. Bu araştırma tek boyutlu MTK modellerinde gerçekleştirilmiştir. Çok boyutlu MTK modellerinde da Bayes yöntemle parametre kestirimi yapılabilir. Klasik (en çok olabilirlik) yöntemle kestirimler arasındaki farklar incelenebilir.

2. Araştırmanın örneklem büyüklüğü, madde sayısı ve model koşulları farklılaştırılabilir ya da veri üretim aşamasında replikasyon sayısı değişimlenerek tekrar yapılabilir.

3. Bu araştırmada sadece önsel yetenek dağılımı farklı biçimlerde değişimlenmiştir. Madde parametreleri alanyazında belirtilen dağılım biçimleriyle sınırlandırılmıştır. Bir başka araştırmada madde parametrelerinin dağılım biçimleri de değişimlenerek çalışma tekrarlanabilir.

4. Önsel yetenek dağılım biçimi sivri olduğunda hem klasik hem de Bayes kestirim yönteminde elde edilen RMSE değerlerinin saçılım göstermesinin nedenleri derinlemesine araştırılabilir.

5. Madde şans (c_i) parametresinde, diğer parametre kestirimlerinin aksine klasik kestirimin daha düşük RMSE değerleri vermesinin nedenleri farklı şans parametresi (c_i) dağılım biçimleri açısından incelenebilir.

6. Araştırmada başka paket programlar ya da farklı yazılım dilleri kullanılarak parametre kestirimlerinde bir farklılaşma olup olmadığı incelenebilir.

7. Bayes MCMC kestirimdeki yakma (iterasyon) ve örnekleme değerleri farklı sayılarda alınarak analizler yeniden gerçekleştirilebilir.

8. Araştırma farklı çarpıklık ve basıklık değerlerine sahip önsel yetenek dağılımlar için standart Bayes yöntemiyle gerçekleştirilmiştir. Bunun yanı sıra çarpıklığın ya da basıklığın durumuna göre yapılandırılmış Bayes yöntemiyle de gerçekleştirilebilir.

9. Küçük örneklemlerde parametrik olmayan MTK modelleri ile Bayes kestirim yöntemleri parametre kestirimleri açısından karşılaştırılabilir.

KAYNAKLAR

Akour, M., & Al-Omari, H. (2013). Empirical Investigation of the Stability of IRT item-parameters estimation. *International Online Journal of Educational Sciences, 5*(2), 291-301. Retrieved from https://iojes.net/?mod=tammetin&makaleadi=&makaleurl=IOJES_980.pdf&key=41119

Anastasi, A. (1988). *Psychological testing.* New york: Mac Millan Pub. Co. Inc.

Atar, B. (2007). *Differential item functioning analyses for mixed response data using IRT likelihood-ratio test, logistic regression, and GLLAMM procedures* (Unpublished doctoral dissertation). Florida State University, USA. Retrieved from http://purl.flvc.org/fsu/fd/FSU_migr_etd-0248

Baker, F. B. (2001). *The basics of item response theory* (2nd ed.). College Park, (MD): ERIC Clearinghouse on Assessment and Evaluation.

Baker, F. B., & Kim, S. (2004). *Item response theory: Parameter estimation techniques* (2nd ed.). New York: Marcel Dekker.

Barton, M. A., & Lord, F. M. (1981). An upper asymptote for the three-parameter logistic item-response model. *ETS Research Report Series*, 1-8. doi: https://doi.org/10.1002/j.2333-8504.1981.tb01255.x

Baykul, Y. (2015). *Eğitimde ve Psikolojide Ölçme: Klasik Test Teorisi ve Uygulaması.* (3. Baskı). Ankara: Pegem Akademi.

Bilir, M. K. (2009). *Mixture item response theory-mimic model: Simultaneous estimation of differential item functioning for manifest groups and latent classes* (Unpublished doctoral dissertation). Florida State University, USA. Retrieved from http://diginole.lib.fsu.edu/islandora/object/fsu:182011/datastream/PDF/view

Bock, R. D., & Aitkin M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46*(4), 443-459. doi: https://doi.org/10.1007/BF02293801

Bock, R. D., & Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement, 6*, 431-444. doi: https://doi.org/10.1177/014662168200600405

Bürkner, P., C. (2017). brms: An R package for Bayesian multilevel models using stan. *Journal of Statistical Software.* doi: https://doi.org/10.18637/jss.v080.i01

Bürkner, P., C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal, 10*(1), 395-411. doi: https://doi.org/10.32614/RJ-2018-017

Bulmer, M. G. (1979). *Principles of statistics*. New York: Dover Publications.

Bulut, O. ve Sünbül, Ö. (2017). R Programlama Dili ile Madde Tepki Kuramında Monte Carlo Simülasyon Çalışmaları. *Journal of Measurement and Evaluation in Education and Psychology*, *8*(3), 266-287. doi: https://doi.org/10.21031/epod.305821

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, *48*(6), 1–29. doi: https://doi.org/10.18637/jss.v048.i06

Chuah, S. C., Drasgow F., & Luecht, R. (2006). How big is big enough? Sample size requirements for cast item parameter estimation. *Applied Measurement in Education, 19*(3), 241-255. doi: https://doi.org/10.1207/s15324818ame1903_5

Clarke, E. (2022). ggbeeswarm: Categorical scatter (violin point) plots. https://cran.r-project.org/web/packages/ggbeeswarm/index.html Erişim Tarihi: Aralık 2022.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, (NJ): Lawrence Erlbaum Associates, Publishers.

Crocker, L., & Algina, J. (1986). *Introduction to clasical and modern test theory.* Belmont: Wadsworth Group.

Çakıcı-Eser, D. (2015). *Çok Boyutlu Madde Tepki Kuramının Farklı Modellerinden Çeşitli Koşullar Altında Kestirilen Parametrelerin İncelenmesi* (Yayımlanmamış doktora tezi). Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara. http://hdl.handle.net/11655/1817

Çelikten, S. ve Çakan, M. (2019). Bayesian ve NonBayesian Kestirim Yöntemlerine Dayalı Olarak Sınıflama İndekslerinin TIMSS-2015 Matematik Testi Üzerinde İncelenmesi. *Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi, 13*(1), 105-124. doi: 10.17522/balikesirnef.566446

Das Nair, R., Moreton, B. J., & Lincoln, N. B. (2011). Rasch analysis of the Nottingham extended activities of daily living scale. *Journal of rehabilitation medicine*, *43*(10), 944–950. doi: https://doi.org/10.2340/16501977-0858

Davidson, M., Keating, J. L., & Eyres, S. (2004). A low back-specific version of the SF-36 Physical Functioning scale. *Spine*, *29*(5), 586–594. doi: https://doi.org/10.1097/01.brs.0000103346.38557.73

De Ayala, R. J. (2009). *The theory and practice of item response theory.* New York, (NY): The Guilford Press.

DeMars, C. (2010). *Item response theory: understanding statistics measurement.* New York: Oxford University Press.

Demir, E. (2019). *R Diliyle İstatistik Uygulamaları* (1. Baskı). Ankara: Pegem Akademi.

Desjardins, C. D., & Bulut, O. (2018). *Handbook of educational measurement and psychometrics using R.* Boca Raton, (FL): CRC Press.

Doğan, N. (2002). *Klasik Test Kuramı ve Örtük Özellikler Kuramının Örneklem Büyüklükleri Bağlamında Karşılaştırılması* (Yayınlanmamış doktora tezi). Hacettepe Üniversitesi, Ankara.

Drasgow, F. (1989). An evaluation of marginal maximum likelihood estimation for the two-parameter logistic model. *Applied Psychological Measurement, 13*(1), 77-90. doi: https://doi.org/10.1177/014662168901300108

Drasgow, F., & Hulin, C. L. (1990). Item response theory. In M. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (pp. 577-636). Palo Alto, (CA): Consulting Psychologists Press.

Embreston, S. E., & Reise, S.P. (2000). *Item response theory for psychologists*. Lawrence: Erlbaum Ass., Inc.

Erkuş, A., Sünbül, Ö., Önder Ömür, S., Yormaz, S. ve Aşiret, S. (2017). *Psikolojide Ölçme ve Ölçek Geliştirme II.* Ankara: Pegem Akademi.

Feinberg, R.A., & Rubright, J.D. (2016), Conducting simulation studies in psychometrics. *Educational Measurement: Issues and Practice, 35*: 36-49. doi: https://doi.org/10.1111/emip.12111

Finch, H., & Edwards, J. M. (2016). Rasch model parameter estimation in the presence of a nonnormal latent trait using a nonparametric Bayesian approach. *Educational and Psychological Measurement, 76*(4), 662-684. doi: https://doi.org/10.1177/0013164415608418

Fraenkel, J. R., & Wallen, E (2009). *How to design and evaluate research in education.* New York, NY: McGraw-Hils Companies.

Gao, F., & Chen, L. (2005). Bayesian or non-Bayesian: A comparison study of item parameter estimation in the three-parameter logistic model. *Applied Measurement in Education, 18*(4), 351-380. doi: https://psycnet.apa.org/doi/10.1207/s15324818ame1804_2

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 6,* 721–741. doi: https://doi.org/10.1109/TPAMI.1984.4767596

Goldman, S.H., & Raju, N. S. (1986). Recovery of one- and two-parameter logistic item parameters: An empirical study. *Educational and psychological measurement, 46*(1), 11-21. doi: https://doi.org/10.1177/0013164486461002

González-de Paz, L., Kostov, B., López-Pina, J. A., Solans-Julián, P., Navarro-Rubio, M. D., & Sisó-Almirall, A. (2015). A Rasch analysis of patients' opinions of primary health care professionals' ethical behaviour with respect to communication issues. *Family practice, 32*(2), 237–243.doi: https://doi.org/10.1093/fampra/cmu073

Gulliksen, H. (1950). *Theory of mental tests.* New York: Wiley.

Hambleton, R. K. (1989). *Principles and selected applications of item response theory.* In R.L. Linn (Ed.), *Educational Measurement,* (pp.147-200). Washington, (DC): American Council of Education.

Hambleton, R. K. (2002). Emergence of item response modeling in instrument development and data analysis. *Medical Care*, *38*, 1160-1165. doi: https://doi.org/10.1097/00005650-200009002-00009

Hambleton, R.K., & Cook, L.L. (1983). Robustness of ítem response models and effects of test length and sample size on the precision of ability estimates. In D.J. Weiss (Ed.), *New horizons in testing: Latent trait test theory and computerized adaptive testing* (pp. 31-49). New York: Vancouver.

Hambleton, R., & K., Swaminathan H. (1985). *Item response theory: Principals and applications.* Norwell, (MA): Kluwer Academic Publishers.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. California: Sage Publications Inc.

Hambleton, R. K., & Jones R.W. (1993). An NCME instructional module on comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice, 12*(3), 38-47. doi: https://doi.org/10.1111/j.1745-3992.1993.tb00543.x

Han, K. T. (2007). *WinGen2: Windows software that generates IRT parameters and item responses [computer program]*. Amherst, (MA): University of Massachusetts, Center for Educational Assessment. Retrieved from https://www.umass.edu/remp/software/simcata/wingen/homeF.html

Harwell, M., Stone, C. A., Hsu, T.-C., & Kirisci, L. (1996). Monte carlo studies in item response theory. *Applied Psychological Measurement, 20*(2), 101-125. doi: https://doi.org/10.1177/014662169602000201

Harwell, M., Kohli, N., & Peralta, Y. (2017). Experimental design and data analysis in computer simulation studies in the behavioral sciences. *Journal of Modern Applied Statistical Methods, 16*(2), 3-28. doi: https://doi.org/10.22237/jmasm/1509494520

Harwell, M., & Janosky, J. (1991). An empirical study of the effects of small datasets and varying prior variances on item parameter estimation in BILOG. *Applied Psychological Measurement. 15*, 279-291. doi: https://doi.org/10.1177/014662169101500308

Hoaglin, D. C., & Andrews, D. F. (1975). The reporting of computation-based results in statistics. *The American Statistician, 29*, 122-126. doi: https://doi.org/10.2307/2683438

Hulin, C. L., Lissak, R. I., & Drasgow, F. (1982). Recovery of two and three-parameter logistic item characteristic curves: A monte carlo study. *Applied Psychological Measurement, 6,* 249-260. doi: https://psycnet.apa.org/doi/10.1177/014662168200600301

Jenkinson, C., Fitzpatrick, R., Garratt, A., Peto V., & Stewart-Brown, S. (2001). Can item response theory reduce patient burden when measuring health status in neurological disorders? Results of the Rasch analysis of the SF-36 physical functioning scale (pf-10). *Journal of Neurology, Neurosurgery & Psychiatry, 71,* 220-224. doi: https://doi.org/10.1136/jnnp.71.2.220

Kaplan, D., & Depaoli, S. (2013). Bayesian statistical methods. In T. D. Little (Ed.), *Oxford library of psychology. The Oxford handbook of quantitative methods (Vol. 1): Foundations* (pp. 407-437). New York, USA: Oxford University Press.

Karadavut, T. (2019). The uniform prior for bayesian estimation of ability in item response theory models. *International Journal of Assessment Tools in Education, 6*(4), 568-579. doi: https://dx.doi.org/10.21449/ijate.581314

Kıbrıslıoğlu Uysal, N. (2020). *Parametrik ve Parametrik Olmayan Madde Tepki Modellerinin Kestirim Hatalarının Karşılaştırılması* (Yayımlanmamış doktora tezi). Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara. http://hdl.handle.net/11655/22495

Kılıç, A. F., Uysal, İ. ve Doğan, N. (2018, Nisan). *The effect of replication number on generated data sets in simulations studies*. 27. Uluslararası Eğitim Bilimleri Kongresi, Antalya, Türkiye.

Kilmen, S. (2010). *Madde Tepki Kuramı'na Dayalı Test Eşitleme Yöntemlerinden Kestirilen Eşitleme Hatalarının Örneklem Büyüklüğü ve Yetenek Dağılımına Göre Karşılaştırılması* (Yayımlanmamış doktora tezi). Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.

Kim, J. K., & Nicewander, W. A. (1993). Ability estimation for conventional tests. *Psychometrika, 58,* 587-599. doi: https://psycnet.apa.org/doi/10.1007/BF02294829

Kirisci, L., Hsu, T.-c., & Yu, L. (2001). Robustness of item parameter estimation programs to assumptions of unidimensionality and normality. *Applied Psychological Measurement, 25*(2), 146-162. doi: https://doi.org/10.1177/01466210122031975

Koğar, H. (2014). *Madde Tepki Kuramının Farklı Uygulamalarından Elde Edilen Parametrelerin ve Model Uyumlarının Örneklem Büyüklüğü ve Test Uzunluğu Açısından Karşılaştırılması* (Yayınlanmamış doktora tezi). Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.

Kolen, M. J. (1985). Standard errors of Tucker equating. *Applied Psychological Measurement, 9*(2), 209-223. doi: https://doi.org/10.1177/014662168500900209

Kolen, M.J., & Brennan, R.L. (2014). Nonequivalent groups: linear methods. In: *Test Equating, Scaling, and Linking. Statistics for Social and Behavioral Sciences.* (pp. 103-133). New York: Springer.

Kothari, C. R. (2004), *Research methodology: methods and techniques* (2nd. ed.). New Delhi: New Age International Publishers.

Köse, İ. A. (2010). *Madde Tepki Kuramına Dayalı Tek Boyutlu ve Çok Boyutlu Modellerin Test Uzunluğu ve Örneklem Büyüklüğü Açısından Karşılaştırılması* (Yayınlanmamış doktora tezi). Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *frontiers in Psychology*, *4*(863), 1-12. doi: https://doi.org/10.3389/fpsyg.2013.00863

La Porta, F., Franceschini, M., Caselli, S., Susassi, S., Cavallini, P., & Tennant, A. (2011). Unified Balance Scale: classic psychometric and clinical properties. *Journal of rehabilitation medicine*, *43*(5), 445–453. https://doi.org/10.2340/16501977-0800

Lenth, R. V. (2022). emmeans: Estimated marginal means, aka Least-Squares Means. https://cran.r-project.org/web/packages/emmeans/index.html Erişim Tarihi: Aralık 2022.

Lim, H., & Wells, C. S. (2020). irtplay: An R package for online item calibration, scoring, evaluation of model fit, and useful functions for unidimensional IRT. *Applied psychological measurement*, *44*(7-8), 563–565. https://doi.org/10.1177/0146621620921247

Linacre, J. M. (2008). *A user's guide to winsteps ministep: rasch-model computer programs.* Retrieved from https://www.winsteps.com/winman/copyright.htm

Lord, F. M. (1968). An analysis of the verbal scholastic aptitude test using Birnbaum's three-parameter logistic model. *Educational and Psychological Measurement, 28*, 989-1020. doi: https://doi.org/10.1177/001316446802800401

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, (NJ): Lawrence Erlbaum Associates.

Lord, F. M. (1983). Unbiased estimators of ability parameters, of their variance, and of their parallelforms reliability. *Psychometrika*, *48*, 233-245. doi: https://doi.org/10.1007/BF02294018

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading MA: Addison-Welsley Publishing Company.

Marais, I., & Andrich, D. (2008). Effects of varying magnitude and patterns of response dependence in the unidimensional Rasch model. *Journal of applied measurement, 9*(2), 105-124.

Martin, A. D., & Quinn, K. M. (2006). Applied Bayesian inference in R using MCMCpack. *The Newsletter of the R Project*, *6*(1), 2-7.

Martinez, J. (2017). Bairt: Bayesian analysis of item response theory models. http://cran.nexr.com/web/packages/bairt/index.html Erişim Tarihi: Ağustos 2022.

Maydeu-Ovivares, A., & Joe, H. (2006). Limited information goodness-of-fit testing in multidimensional contingency tables. *Psychometrica, 71*, 713-732. doi: https://doi.org/10.1007/s11336-005-1295-9

MEB (2022). *Sınavla Öğrenci Alacak Ortaöğretim Kurumlarına İlişkin Merkezî Sınav Başvuru ve Uygulama Kılavuzu.* Ankara: MEB https://www.meb.gov.tr/2022-lgs-kapsamindaki-merkez-sinav-kilavuzu-yayimlandi/haber/25705/tr Erişim tarihi: Ağustos, 2022.

Meng, X-B., Tao, J., & Chang, H-H. (2015). A conditional joint modeling approach for locally dependent item responses and response times. *Journal of Educational Measurement*, *52*(1), 1-27. doi: https://doi.org/10.1111/jedm.12060

Meyer, D. (2022). e1071: Misc functions of the department of statistics, Probability Theory Group (Formerly: E1071), TU Wien. Retrieved from https://CRAN.R-project.org/package=e1071.

Mislevy, R. J. (1984). Estimating latent distributions. *Psychometrika, 49*, 359–381. doi: https://doi.org/10.1007/BF02306026

Mislevy, R. J. (1986). Bayes modal estimation in item response models. *Psychometrika, 51*, 177-195. doi: https://doi.org/10.1007/BF02293979

Morizot, J., Ainsworth, A. T., & Reise, S. P. (2007). Toward modern psychometrics. Application of item response theory models in personality research. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personalty psychology*. (pp. 407-423). New York: The Guilford Press.

Morris, T. P., White, I. R., & Crowther, M. J. (2017). Using simulation studies to evaluate statistical methods. *Tutorial in Biostatistics, 38*(11), 2074-2102. doi: https://doi.org/10.1002/sim.8086

Orlando, M. (2004, June). Critical issues to address when applying item response theory models. *Paper presented at the conference on improving health outcomes assessment*, National Cancer institute, Bethesda, MD, USA.

Owen, R. J. (1975). A Bayesian sequential procedure for quantal response in the context of adaptive mental testing. *Journal of the American Statistical Association, 70*(350), 351–356. https://doi.org/10.2307/2285821

Patz, R. J., & Junker, B. R. (1999). A straightforward approach to markov chain monte carlo methods for ıtem response models. *Journal of Educational and Behavioral Statistics, 24*(2), 146-178. doi: https://doi.org/10.2307/1165199

Pekmezci Barış, F. (2018). *İki Faktör Modelde (Bifactor) Diklik Varsayımının Farklı Koşullar Altında Sınanması* (Yayımlanmamış doktora tezi). Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.

Ramp, M., Khan, F., Misajon, R. A., & Pallant, J. F. (2009). Rasch analysis of the Multiple Sclerosis Impact Scale MSIS-29. *Health and quality of life outcomes, 7*, 58. doi: https://doi.org/10.1186/1477-7525-7-58

Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics, 4*(3), 207–230. doi: https://doi.org/10.3102/10769986004003207

Ree, M. J., & Jensen, H. E. (1980). Effects of sample size on linear equating of item characteristic curve parameters. In D. J. Weiss (Ed.), *Proceedings of the 1979 computerized adaptive testing conference.* Minneapolis: University of Minnesota. doi: https://doi.org/10.1016/B978-0-12-742780-5.50017-2

Reise, S. P., & Yu, J. (1990). Parameter recovery in the graded response model using MULTILOG. *Journal of Educational Measurement, 27*, 133–144. doi: https://doi.org/10.1111/j.1745-3984.1990.tb00738.x

Revelle, W. (2022). psych: Procedures for psychological, psychometric, and personality research. doi: https://cran.r-project.org/web/packages/psych/index.html Erişim Tarihi: Ekim 2022.

Reyhanlıoğlu Keçeoğlu, Ç. (2018). *Parametrik ve Parametrik Olmayan Madde Tepki Kuramında Farklı Örneklem Büyüklüklerine ve Boyutluluklarına Göre Parametre Değişmezliğinin İncelenmesi* (Yayınlanmamış doktora tezi). Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara. http://hdl.handle.net/11655/5279

Røe, C., Damsgård, E., Fors, T., & Anke, A. (2014). Psychometric properties of the pain stages of change questionnaire as evaluated by rasch analysis in patients with chronic musculoskeletal pain. *BMC Musculoskeletal Disorders, 15*, 95. doi: https://doi.org/10.1186/1471-2474-15-95

Robitzsch, A. (2022). sirt: Supplementary item response theory models. https://cran.r-project.org/web/packages/sirt/index.html Erişim Tarihi: Ekim 2022.

Samejima, F. (1969). Estimation of latent ability using a response patterns of graded scores. *Psychometrika Monograph, 17*, 1-169. doi: https://doi.org/10.1002/j.2333-8504.1968.tb00153.x

Samejima, F. (1980). *Is Bayesian estimation proper for estimating the individual's ability? Research Report 80-3.* University of Tennessee, Department of Psychology, Knoxville, TN.

Samejima, F. (1993a). An approximation for the bias function of the maximum likelihood estimate of a latent variable for the general case where the item responses are discrete. *Psychometrika, 58,* 119-138. doi: https://doi.org/10.1007/BF02294476

Samejima, F. (1993b). The bias function of the maximum likelihood estimate of ability for the dichotomous response level. *Psychometrika, 58,* 195-209. doi: https://doi.org/10.1007/BF02294573

Samejima, F. (1998, April). *Some considerations for eliminating biases in ability estimation in computerized adaptive testing.* Paper presented at the annual meeting of the American Educational Association, San Diego.

Sarkar D (2022). *lattice: Trellis graphics for R.* R package version 0.20-45, URL http://CRAN.R-project.org/package=lattice. Erişim tarihi: Ekim 2022.

SAS Institute (2020). Introduction to Bayesian analysis procedures. In *User's Guide Introduction to Bayesian Analysis Procedures.* (pp. 127-161). SAS Institute Inc., Cary, (NC), USA.

Sass, D., Schmitt, T., & Walker, C. (2008). Estimating non-normal latent trait distributions within item response theory using true and estimated item parameters. *Applied Measurement in Education, 21*(1), 65-88. doi: https://doi.org/10.1080/08957340701796415

Sayın, A. (2014). *Klasik Test Kuramı ve Madde Tepki Kuramına Göre Kestirilen Parametrelerle Sınırlandırılan Yapısal Eşitlik Modellerinin Uyum İndekslerinin Karşılaştırılması* (Yayımlanmamış doktora tezi). Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.

Seong, T.-J. (1990). Sensitivity of marginal maximum likelihood estimation of item and ability parameters to the characteristics of the prior ability distributions. *Applied Psychological Measurement, 14*(3), 299-311. doi: https://psycnet.apa.org/doi/10.1177/014662169001400307

Singmann, H. (2022). afex: Analysis of factorial experiments. https://cran.r-project.org/web/packages/afex/afex.pdf Erişim Tarihi: Aralık 2022.

Soysal, S. (2017). *Toplam Test Puanı ve Alt Test Puanlarının Kestiriminin Hiyerarşik Madde Tepki Kuramı Modelleri ile Karşılaştırılması* (Yayımlanmamış doktora tezi). Hacettepe Üniversitesi. Eğitim Bilimleri Enstitüsü, Ankara.

Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences.* Mahwah, (NJ): Lawrence Erlbaum.

Stone, C. A. (1992). Recovery of marginal maximum likelihood estimates in the twoparameter logistic response model: An evaluation of MULTILOG. A*pplied Psychological Measurement, 16*(1), 1-16. doi: https://doi.org/10.1177/014662169201600101

Sünbül, Ö. (2011). *Çeşitli Boyutluluk Özelliğine Sahip Yapılarda, Madde Parametrelerinin Değişmezliğinin Klasik Test Teorisi, Tek Boyutlu Madde Tepki Kuramı ve Çok Boyutlu Madde Tepki Kuramı Çerçevesinde İncelenmesi* (Yayımlanmamış doktora tezi). Mersin Üniversitesi, Eğitim Bilimleri Enstitüsü, Mersin.

Swaminathan, H., & Gifford, J. A. (1982). Bayesian estimation in the Rasch model. *Journal of Educational Statistics, 7*, 175–191. doi: https://doi.org/10.2307/1164643

Swaminathan, H., & Gifford, J. A. (1985). Bayesian estimation in the two-parameter logistic model. *Psychometrika, 50*, 349–364. doi: https://psycnet.apa.org/doi/10.1007/BF02294110

Swaminathan, H., & Gifford, J. A. (1986). Bayesian estimation in the three-parameter logistic model. *Psychometrika, 51*, 589–601. doi: https://doi.org/10.1007/BF02295598

Şahin, A. (2012). *Madde Tepki Kuramı'nda Test Uzunluğu ve Örneklem Büyüklüğünün Model Veri Uyumu, Madde Parametreleri ve Standart Hata Değerlerine Etkisinin İncelenmesi* (Yayınlanmamış doktora tezi). Eğitim Bilimleri Enstitüsü, Hacettepe Üniversitesi, Ankara.

Şahin, A. ve Anıl, D. (2017). The effects of test length and sample size on item parameters in item response theory. *Educational Sciences: Theory & Practice, 17*, 321–335. http://dx.doi.org/10.12738/estp.2017.1.0270

Şengül Avşar, A. (2015). *Çok Kategorili Puanlanan Maddelerin Psikometrik Özelliklerinin Farklı Test Koşullarında Parametrik Olmayan Madde Tepki Kuramı Modellerine Göre İncelenmesi* (Yayımlanmamış doktora tezi). Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.

Tabachnick, B. G., & Fidell, L. S. (2014). *Using multivariate statistics* (6th ed.). Boston: Pearson New International Edition.

Ten Klooster, P.M., TAAL, E., & van de Laar, M.A.F.J. (2008), Rasch analysis of the Dutch health assessment questionnaire disability index and the health assessment questionnaire II in patients with rheumatoid arthritis. *Arthritis & Rheumatism, 59*, 1721-1728. doi: https://doi.org/10.1002/art.24065

Thissen, D., & Wainer, H. (1983). Some standart errors in item response theory. *Psychometrika*, *47*, 397-412. doi: https://doi.org/10.1007/BF02293705

Thorndike, L. R. (1982). *Applied Psychometrics.* Boston: Houghton Mifflin Co.

Van de Schoot, R., & Depaoli, S. (2014). Bayesian analyses: Where to start and what to report. *The European Health Psychologist, 16*(2), 75-84. doi: https://www.ehps.net/ehp/index.php/contents/article/view/ehp.v16.i2.p75/26

Wang, S., & Wang, T. (2001). Precision of Warm's weighted likelihood estimates for a polytomous model in computerized adaptive testing. *Applied Psychological Measurement*, *25*, 317-331. doi: https://doi.org/10.1177/01466210122032163

Warm, A. W. (1989). Weighted likelihood estimation of ability in item response theory with tests of finite length. *Psychometrika, 54*, 427-450. doi: https://doi.org/10.1007/BF02294627

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis.* New York: Springer-Verlag. doi: 10.1007/978-0-387-98141-3

Wright, B. D., & Stone, M. H. (1979). *Best test design.* Chicago: Mesa Press

Xu, X., & Jia, Y. (2011). The sensitivity of parameter estimates to the latent ability distribution. *ETS Research Report Series, 2*. doi: https://doi.org/10.1002/j.2333-8504.2011.tb02276.x

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*, 125-145. doi: https://doi.org/10.1177/014662168400800201

EKLER

EK 1. Verilerin Üretilmesinde Kullanılan Örnek R Kodları

```
#Önsel Yetenek Dağılımı Normal Olan 2 PL Modelde 10 Maddelik 100
Bireylik Veri Üretmede Kullanılan Örnek R Kodları
set.seed(1234567890)
L1 <- vector("list",100000)
L1.bias.rmse.ab <- vector("list", 100)
L1.a <- vector("list", 100)
L1.b <- vector("list", 100)
L1.k <- vector("list", 100)

for (i in 1:100000) {
  D <- önsel(10, 100, M="2PL", D="normal")
  S <- rowSums(D[[1]])
  if(abs(skewness(S))<1    &    abs(kurtosis(S))<1)    if(D[[7]]>0.05)
{L1[[i]] <- D[[1]]
  L1.bias.rmse.ab[[i]] <- cbind(D[[2]], D[[3]], D[[4]], D[[5]])
  L1.a[[i]] <- D[[8]]
  L1.b[[i]] <- D[[9]]
  L1.k[[i]] <- D[[10]]}
  if(length(L1[L1 != "NULL"]) == 100) {break}
}
L1 <- L1[L1 != "NULL"]
L1.bias.rmse.ab <- L1.bias.rmse.ab[L1.bias.rmse.ab != "NULL"]
L1.a <- L1.a[L1.a != "NULL"]
L1.b <- L1.b[L1.b != "NULL"]
L1.k <- L1.k[L1.k != "NULL"]

round(range(lapply(lapply(L1,rowSums), skewness)), 3)
round(range(lapply(lapply(L1,rowSums), kurtosis)), 3)
#Önsel Yetenek Dağılımı Normal Olan 3 PL Modelde 10 Maddelik 100
Bireylik Veri Üretmede Kullanılan Örnek R Kodları
L10 <- vector("list",100000)
L10.bias.rmse.abc <- vector("list", 100)
L10.a <- vector("list", 100)
L10.b <- vector("list", 100)
L10.c <- vector("list", 100)
L10.k <- vector("list", 100)

for (i in 1:100000) {
  D <- önsel(10, 100, M="3PL", D="normal")
  S <- rowSums(D[[1]])
  if(abs(skewness(S))<1    &    abs(kurtosis(S))<1)    if(D[[9]]>0.05)
{L10[[i]] <- D[[1]]
  L10.bias.rmse.abc[[i]]  <-  cbind(D[[2]],  D[[3]],  D[[4]],  D[[5]],
D[[6]], D[[7]])
  L10.a[[i]] <- D[[10]]
  L10.b[[i]] <- D[[11]]
  L10.c[[i]] <- D[[12]]
  L10.k[[i]] <- D[[13]]}
  if(length(L10[L10 != "NULL"]) == 100) {break}
}
L10 <- L10[L10 != "NULL"]
L10.bias.rmse.abc <- L10.bias.rmse.abc[L10.bias.rmse.abc != "NULL"]
L10.a <- L10.a[L10.a != "NULL"]
L10.b <- L10.b[L10.b != "NULL"]
L10.c <- L10.c[L10.c != "NULL"]
L10.k <- L10.k[L10.k != "NULL"]

round(range(lapply(lapply(L10,rowSums), skewness)), 3)
round(range(lapply(lapply(L10,rowSums), kurtosis)), 3)
```

EK 2. Verilerin En Çok Olabilirlik Yöntemle Kestiriminde Kullanılan Örnek R Kodları

```
library(irtplay)
######################## MAXIMUM LIKELIHOOD KESTİRİMLERİ ############
#Önsel Yetenek Dağılımı Normal Olan 2 PL Modelde 10 Maddelik 100
Bireylik Verilerde En Çok Olabilirlik Kestiriminde Kullanılan Örnek R
Kodları
ml.L1 <- vector("list", 100)
ml.L1.person <- vector("list", 100)
ml.L1.per <- vector("list", 100)
ml.L1.per.se <- vector("list", 100)
ml.L1.a <- vector("list", 100)
ml.L1.a.se <- vector("list", 100)
ml.L1.b <- vector("list", 100)
ml.L1.b.se <- vector("list", 100)
for (i in 1:100) {ml.L1[[i]] <- est_irt(data=L1[[i]], D=1,
model="2PLM", cats=2, MaxE = 10000)
ml.L1.person[[i]] <- est_score(ml.L1[[i]],data=L1[[i]], D=1,
method="MLE", range = c(-7, 7), se=TRUE, ncor=2)
ml.L1.per[[i]] <- ml.L1.person[[i]]$est.theta
ml.L1.per.se[[i]] <- ml.L1.person[[i]]$se.theta

ml.L1.a[[i]] <- ml.L1[[i]]$estimates$par.1
ml.L1.a.se[[i]] <- ml.L1[[i]]$estimates$se.1
ml.L1.b[[i]] <- ml.L1[[i]]$estimates$par.2
ml.L1.b.se[[i]] <- ml.L1[[i]]$estimates$se.2}
#####################################################################
#Önsel Yetenek Dağılımı Normal Olan 3 PL Modelde 10 Maddelik 100
Bireylik Verilerde En Çok Olabilirlik Kestiriminde Kullanılan Örnek R
Kodları
ml.L10 <- vector("list", 100)
ml.L10.person <- vector("list", 100)
ml.L10.per <- vector("list", 100)
ml.L10.per.se <- vector("list", 100)
ml.L10.a <- vector("list", 100)
ml.L10.a.se <- vector("list", 100)
ml.L10.b <- vector("list", 100)
ml.L10.b.se <- vector("list", 100)
ml.L10.c <- vector("list", 100)
ml.L10.c.se <- vector("list", 100)
for (i in 1:100) {ml.L10[[i]] <- est_irt(data=L10[[i]], D=1,
model="3PLM", cats=2, MaxE = 10000)
ml.L10.person[[i]] <- est_score(ml.L10[[i]],data=L10[[i]], D=1,
method="MLE", range = c(-7, 7), se=TRUE, ncor=2)
ml.L10.per[[i]] <- ml.L10.person[[i]]$est.theta
ml.L10.per.se[[i]] <- ml.L10.person[[i]]$se.theta

ml.L10.a[[i]] <- ml.L10[[i]]$estimates$par.1
ml.L10.a.se[[i]] <- ml.L10[[i]]$estimates$se.1
ml.L10.b[[i]] <- ml.L10[[i]]$estimates$par.2
ml.L10.b.se[[i]] <- ml.L10[[i]]$estimates$se.2
ml.L10.c[[i]] <- ml.L10[[i]]$estimates$par.3
ml.L10.c.se[[i]] <- ml.L10[[i]]$estimates$se.3}
#####################################################################
```

EK 3. Verilerin Bayesçi Yöntemle Kestiriminde Kullanılan Örnek R kodları

```
library(sirt)
library(bairt)
cprior <- select.c.prior(5)
############################ STANDART BAYES KESTİRİMLERİ #############
#Önsel Yetenek Dağılımı Normal Olan 2 PL Modelde 10 Maddelik 100
Bireylik Verilerde Bayes Kestiriminde Kullanılan Örnek R Kodları
sb.L1 <- vector("list", 100)
sb.L1.a <- vector("list", 100)
sb.L1.a.se <- vector("list", 100)
sb.L1.b <- vector("list", 100)
sb.L1.b.se <- vector("list", 100)
sb.L1.per <- vector("list", 100)
sb.L1.per.se <- vector("list", 100)
for (i in 1:100) {sb.L1[[i]] <- mcmc.2pnob(L1[[i]], burning = 1000,
iter = 3000, thin = 1, parts = 4)
sb.L1[[i]] <- sb.L1[[i]]$diagnostic$diag
sb.L1.a[[i]] <- sb.L1[[i]][1:10,5]
sb.L1.a.se[[i]] <- sb.L1[[i]][1:10,6]
sb.L1.b[[i]] <- sb.L1[[i]][11:20,5]
sb.L1.b.se[[i]] <- sb.L1[[i]][11:20,6]
sb.L1.per[[i]] <- sb.L1[[i]][21:120,5]
sb.L1.per.se[[i]] <- sb.L1[[i]][21:120,6]}
#####################################################################
#Önsel Yetenek Dağılımı Normal Olan 3 PL Modelde 10 Maddelik 100
Bireylik Verilerde Bayes Kestiriminde Kullanılan Örnek R Kodları
sb.L10 <- vector("list", 100)
sb.L10.a <- vector("list", 100)
sb.L10.a.se <- vector("list", 100)
sb.L10.b <- vector("list", 100)
sb.L10.b.se <- vector("list", 100)
sb.L10.c <- vector("list", 100)
sb.L10.c.se <- vector("list", 100)
sb.L10.per <- vector("list", 100)
sb.L10.per.se <- vector("list", 100)
for (i in 1:100) {sb.L10[[i]] <- mcmc.3pnob(L10[[i]], c.prior = cprior,
burning = 1000, iter = 3000, thin = 1, parts = 4)
sb.L10[[i]] <- sb.L10[[i]]$diagnostic$diag
sb.L10.a[[i]] <- sb.L10[[i]][1:10,5]
sb.L10.a.se[[i]] <- sb.L10[[i]][1:10,6]
sb.L10.b[[i]] <- sb.L10[[i]][11:20,5]
sb.L10.b.se[[i]] <- sb.L10[[i]][11:20,6]
sb.L10.c[[i]] <- sb.L10[[i]][21:30,5]
sb.L10.c.se[[i]] <- sb.L10[[i]][21:30,6]
sb.L10.per[[i]] <- sb.L10[[i]][31:130,5]
sb.L10.per.se[[i]] <- sb.L10[[i]][31:130,6]}
#####################################################################
```

EK 4. Önsel Yetenek Dağılımı Normal Olan ve Olmayan Verilerin En Çok Olabilirlik ve Bayes Yöntemlerine Göre Kestirilen Yetenek Parametrelerinin Yanlılık (Bias) ve RMSE Değerleri

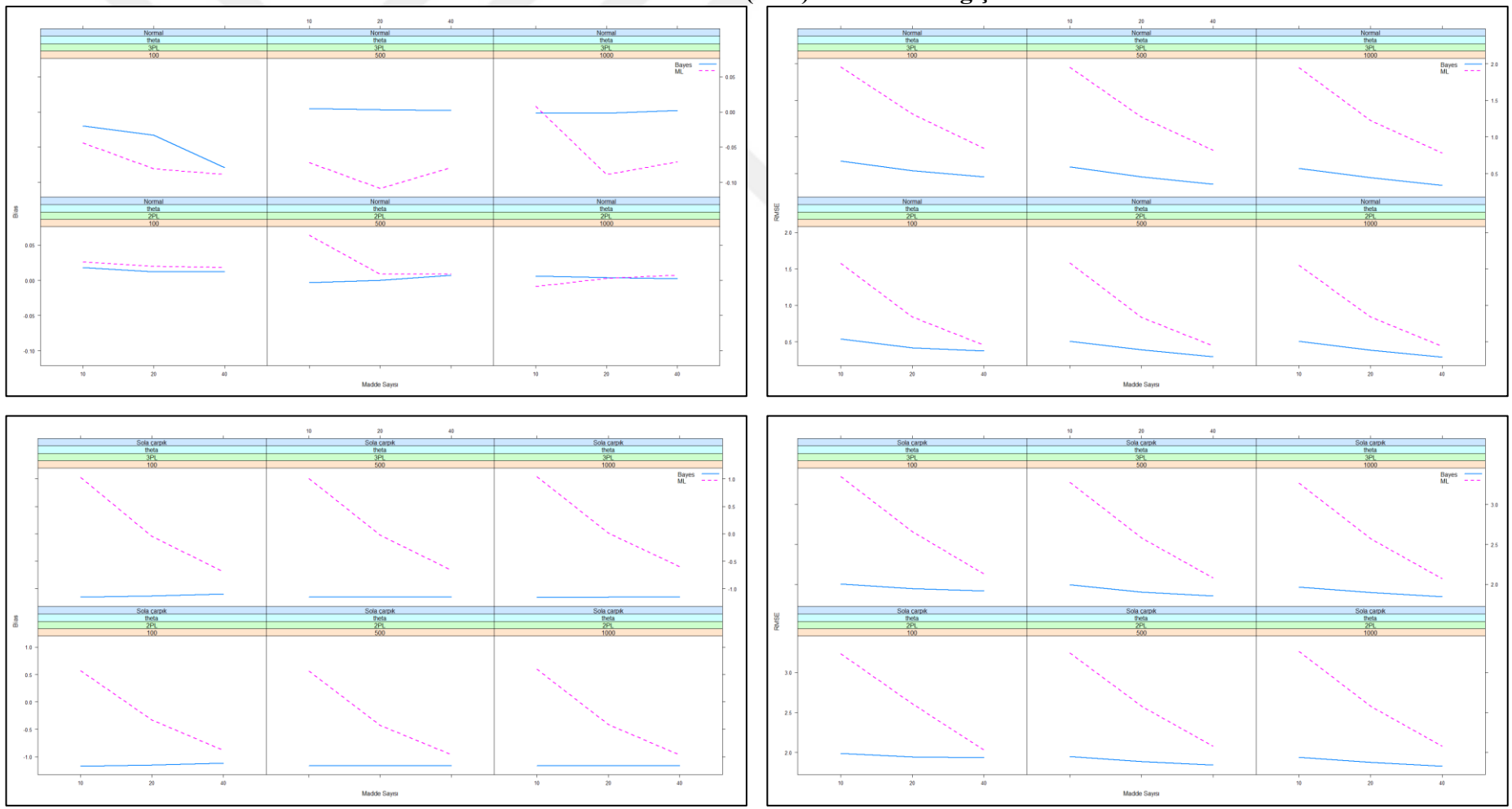
Önsel Yetenek Dağılımı	Madde Sayısı	Örneklem Büyüklüğü	MTK Modeli	En çok olabilirlik kestirimi		Bayes kestirimi	
				Bias.Ɵ	RMSE.Ɵ	Bias.Ɵ	RMSE.Ɵ
Normal	10	100	2PL	0.026	1.577	0.018	0.538
		500		0.064	1.580	-0.003	0.508
		1000		-0.009	1.551	0.006	0.510
	20	100		0.020	0.838	0.012	0.419
		500		0.009	0.833	0.000	0.391
		1000		0.003	0.842	0.004	0.388
	40	100		0.018	0.455	0.012	0.375
		500		0.009	0.444	0.007	0.296
		1000		0.007	0.437	0.002	0.292
	10	100	3PL	-0.044	1.959	-0.020	0.671
		500		-0.072	1.956	0.005	0.589
		1000		0.008	1.949	-0.002	0.570
	20	100		-0.081	1.314	-0.033	0.536
		500		-0.109	1.274	0.003	0.451
		1000		-0.089	1.222	-0.002	0.444
	40	100		-0.089	0.846	-0.079	0.454
		500		-0.079	0.815	0.002	0.351
		1000		-0.071	0.780	0.002	0.337
Sola Çarpık	10	100	2PL	0.569	3.233	-1.169	1.987
		500		0.560	3.244	-1.161	1.948
		1000		0.601	3.262	-1.162	1.941
	20	100		-0.330	2.607	-1.149	1.942
		500		-0.428	2.580	-1.163	1.886
		1000		-0.411	2.581	-1.161	1.876
	40	100		-0.881	2.032	-1.116	1.936
		500		-0.963	2.078	-1.155	1.840
		1000		-0.968	2.078	-1.160	1.829
	10	100	3PL	1.028	3.349	-1.153	2.004
		500		1.009	3.275	-1.153	1.996
		1000		1.045	3.266	-1.160	1.965
	20	100		-0.044	2.658	-1.134	1.945
		500		-0.021	2.585	-1.153	1.904
		1000		0.014	2.575	-1.154	1.899
	40	100		-0.686	2.132	-1.096	1.917
		500		-0.663	2.081	-1.150	1.856
		1000		-0.595	2.070	-1.151	1.844

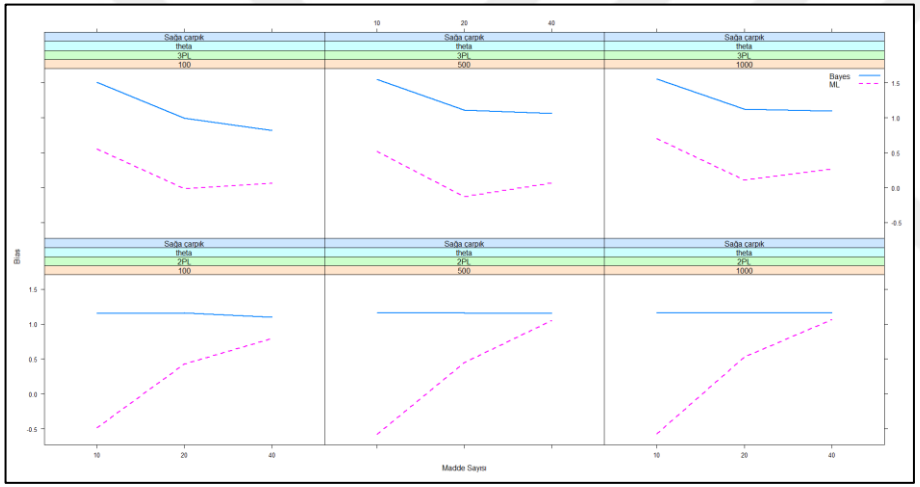
(Devam ediyor)

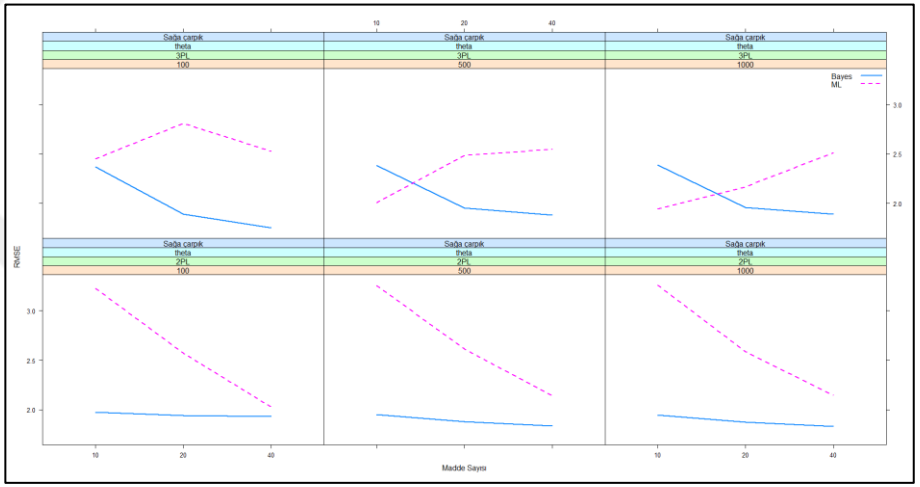
EK 4 (Devam)

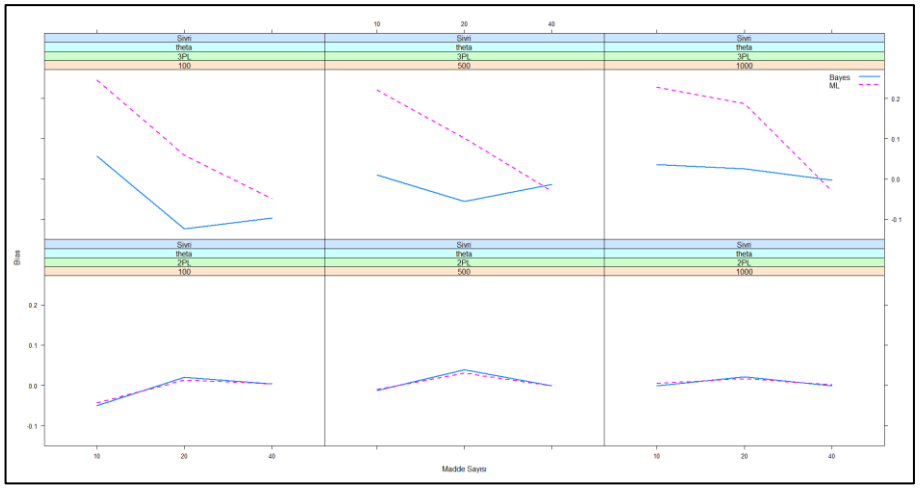
Önsel Yetenek Dağılımı	Madde Sayısı	Örneklem Büyüklüğü	MTK Modeli	En çok olabilirlik kestirimi		Bayes kestirimi	
				Bias.Θ	RMSE.Θ	Bias.Θ	RMSE.Θ
Sağa Çarpık	10	100	2PL	-0.483	3.226	1.155	1.974
		500		-0.578	3.257	1.161	1.949
		1000		-0.572	3.258	1.166	1.946
	20	100		0.429	2.572	1.158	1.942
		500		0.451	2.615	1.159	1.881
		1000		0.528	2.585	1.158	1.873
	40	100		0.793	2.029	1.101	1.930
		500		1.049	2.143	1.151	1.838
		1000		1.066	2.146	1.162	1.830
	10	100	3PL	0.554	2.451	1.510	2.371
		500		0.519	2.007	1.549	2.384
		1000		0.704	1.945	1.556	2.387
	20	100		-0.012	2.811	0.998	1.893
		500		-0.127	2.486	1.107	1.953
		1000		0.107	2.167	1.123	1.960
	40	100		0.058	2.527	0.822	1.751
		500		0.067	2.544	1.065	1.880
		1000		0.268	2.513	1.097	1.893
Sivri	10	100	2PL	-0.043	2.017	-0.050	2.730
		500		-0.011	2.005	-0.013	2.729
		1000		0.005	1.998	-0.001	2.719
	20	100		0.013	1.851	0.020	2.633
		500		0.031	2.211	0.039	3.290
		1000		0.017	2.267	0.021	3.353
	40	100		0.004	0.897	0.003	0.345
		500		-0.001	0.820	-0.001	0.449
		1000		0.001	0.813	-0.001	0.463
	10	100	3PL	0.246	25.570	0.057	26.722
		500		0.222	2.228	0.010	2.532
		1000		0.228	26.851	0.036	27.865
	20	100		0.060	2.141	-0.124	2.955
		500		0.102	2.238	-0.056	2.983
		1000		0.187	2.568	0.025	3.435
	40	100		-0.048	1.157	-0.097	0.513
		500		-0.030	1.067	-0.013	0.496
		1000		-0.030	1.035	-0.003	0.502
Basık	10	100	2PL	0.044	2.613	0.008	1.002
		500		-0.001	2.645	-0.003	0.951
		1000		0.017	2.654	0.001	0.945
	20	100		0.014	1.780	0.001	0.997
		500		-0.013	1.767	0.007	0.887
		1000		-0.016	1.796	-0.002	0.877
	40	100		-0.030	1.122	-0.004	1.039
		500		-0.008	1.176	0.007	0.856
		1000		0.004	1.167	0.002	0.835
	10	100	3PL	0.079	2.924	-0.002	1.054
		500		0.133	2.978	0.006	1.004
		1000		0.122	2.944	0.002	0.985
	20	100		-0.139	2.194	-0.034	1.029
		500		-0.136	2.197	0.002	0.393
		1000		-0.098	2.160	0.002	0.924
	40	100		-0.286	1.643	-0.083	1.036
		500		-0.250	1.651	-0.004	0.906
		1000		-0.228	1.613	-0.002	0.876

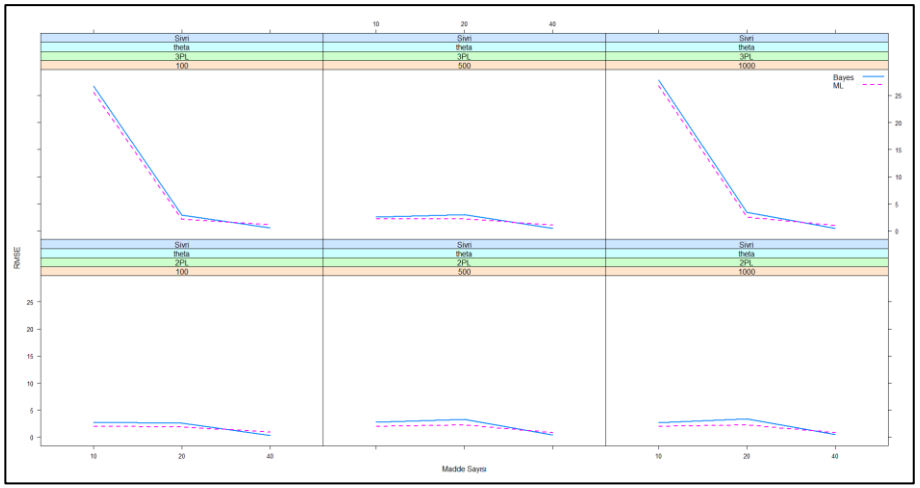
EK 5. Önsel Yetenek Dağılımı Normal Olan ve Olmayan Verilerin En Çok Olabilirlik ve Bayes Yöntemlerine Göre Kestirilen Yetenek Parametrelerinin Yanlılık (Bias) ve RMSE Değişimleri

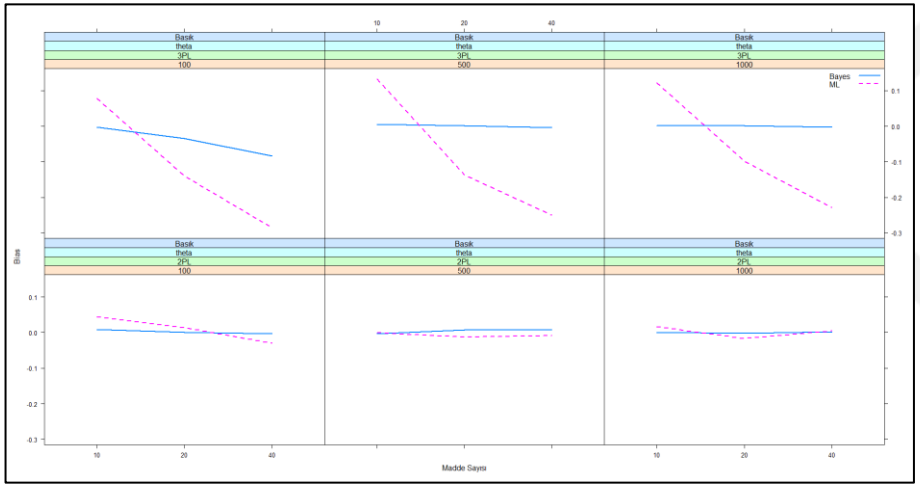


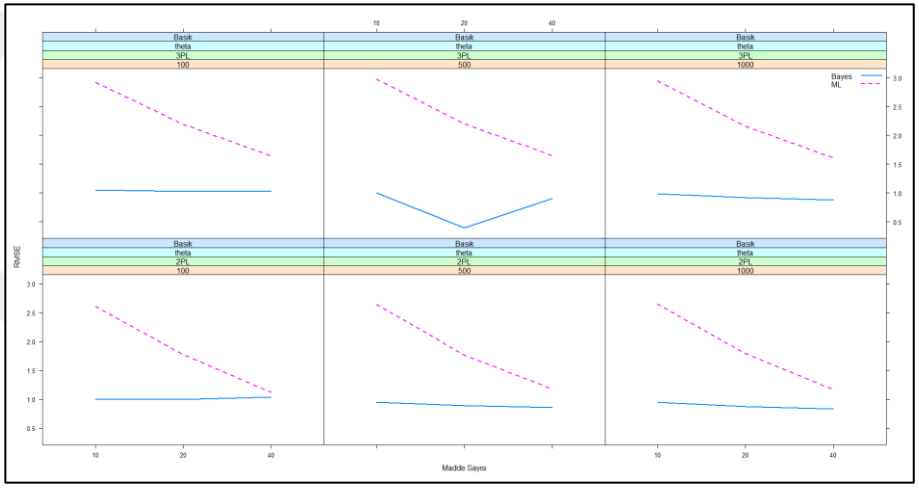

Sağa çarpık
theta
3PL
2PL
100
500
1000
Bayes
ML
Bias
Madde Sayısı


Sağa çarpık
theta
3PL
2PL
100
500
1000
Bayes
ML
RMSE
Madde Sayısı


Sivri
theta
3PL
2PL
100
500
1000
Bayes
ML
Bias
Madde Sayısı


Sivri
theta
3PL
2PL
100
500
1000
Bayes
ML
RMSE
Madde Sayısı


Basik
theta
3PL
100
500
1000
2PL
Bayes
ML
Bias
Madde Sayısı


Basik
theta
3PL
2PL
100
500
1000
Bayes
ML
RMSE
Madde Sayısı

EK 6. Önsel Yetenek Dağılımı Normal Olan ve Olmayan Verilerin En Çok Olabilirlik ve Bayes Yöntemlerine Göre Kestirilen Madde Parametrelerinin Yanlılık (Bias) ve RMSE Değerleri

Önsel Yetenek Dağılımı	Madde Sayısı	Örneklem Büyüklüğü	MTK Modeli	En çok olabilirlik kestirimi						Bayes kestirimi					
				Bias.a	RMSE.a	Bias.b	RMSE.b	Bias.c	RMSE.c	Bias.a	RMSE.a	Bias.b	RMSE.b	Bias.c	RMSE.c
Normal	10	100	2PL	0.157	0.807	-0.022	1.856	-	-	-0.311	0.759	-0.004	1.674	-	-
		500		0.016	0.204	-0.170	1.800	-	-	-0.545	0.566	-0.151	1.632	-	-
		1000		0.002	0.137	0.033	1.816	-	-	-0.556	0.572	0.033	1.627	-	-
	20	100		0.054	0.449	-0.312	3.125	-	-	-0.334	0.596	-0.052	1.755	-	-
		500		0.011	0.176	0.004	1.774	-	-	-0.537	0.557	0.005	1.590	-	-
		1000		0.007	0.125	0.026	1.819	-	-	-0.557	0.574	0.023	1.636	-	-
	40	100		0.082	0.725	0.028	1.917	-	-	-0.213	0.427	0.013	1.671	-	-
		500		-0.001	0.166	0.050	1.762	-	-	-0.532	0.552	0.041	1.581	-	-
		1000		0.002	0.115	0.025	1.789	-	-	-0.551	0.567	0.017	1.597	-	-
	10	100	3PL	9.608	24.516	1.991	7.717	0.078	0.111	1.907	3.480	1.646	3.982	0.178	0.205
		500		0.676	1.804	0.164	1.774	0.069	0.102	-0.153	0.624	0.425	1.942	0.123	0.156
		1000		0.166	0.338	-0.073	1.706	0.034	0.084	-0.375	0.465	0.043	1.641	0.061	0.100
	20	100		6.801	23.586	-1.279	8.306	0.062	0.096	1.383	2.940	1.361	3.552	0.097	0.128
		500		0.216	0.423	0.114	1.685	0.054	0.105	-0.296	0.464	0.287	1.724	0.080	0.122
		1000		0.162	0.300	0.109	1.738	0.024	0.092	-0.404	0.457	0.189	1.676	0.040	0.098
	40	100		4.915	21.788	0.125	2.111	0.071	0.107	1.308	2.743	1.095	3.207	0.111	0.140
		500		0.213	0.402	0.095	1.757	0.054	0.106	-0.287	0.428	0.259	1.774	0.075	0.119
		1000		0.147	0.265	0.079	1.757	0.040	0.103	-0.408	0.450	0.158	1.677	0.056	0.110
Sola çarpık	10	100	2PL	15.190	34.622	-0.791	1.341	-	-	2.628	3.592	-2.797	3.850	-	-
		500		3.208	4.340	-0.906	1.399	-	-	1.124	1.398	-2.058	2.727	-	-
		1000		2.679	2.914	-0.897	1.382	-	-	0.934	1.099	-1.935	2.531	-	-
	20	100		12.237	34.301	-0.528	1.286	-	-	2.356	3.307	-1.889	2.884	-	-
		500		2.298	2.487	-0.605	1.339	-	-	0.746	0.889	-1.415	2.160	-	-
		1000		2.270	2.419	-0.578	1.312	-	-	0.624	0.720	-1.376	2.093	-	-
	40	100		8.511	28.461	-0.239	1.295	-	-	2.545	3.386	-1.075	2.189	-	-
		500		2.060	2.188	-0.404	1.330	-	-	0.604	0.691	-1.058	1.906	-	-
		1000		2.039	2.136	-0.368	1.294	-	-	0.494	0.545	-1.042	1.858	-	-
	10	100	3PL	19.129	43.017	-0.755	1.301	0.031	0.078	3.032	4.076	-2.711	3.863	0.059	0.093
		500		2.697	3.593	-0.721	1.258	-0.005	0.074	0.925	1.296	-1.850	2.580	0.007	0.075
		1000		2.319	2.707	-0.726	1.255	-0.022	0.080	0.694	0.968	-1.753	2.415	-0.013	0.077
	20	100		16.44	46.435	-0.611	1.380	0.038	0.088	2.682	3.802	-1.915	3.074	0.063	0.100
		500		2.082	2.398	-0.485	1.282	-0.004	0.077	0.622	0.868	-1.365	2.181	0.008	0.076
		1000		1.974	2.374	-0.536	1.306	-0.029	0.086	0.434	0.633	-1.363	2.149	-0.024	0.082

(Devam ediyor)

EK 6 (Devam)

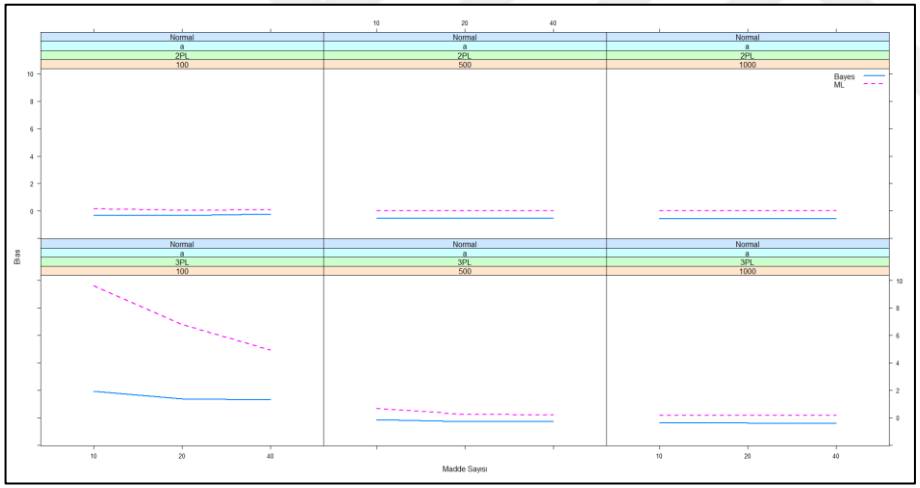
Önsel Yetenek Dağılımı	Madde Sayısı	Örneklem Büyüklüğü	MTK Modeli	En çok olabilirlik kestirimi						Bayes kestirimi					
				Bias.a	RMSE.a	Bias.b	RMSE.b	Bias.c	RMSE.c	Bias.a	RMSE.a	Bias.b	RMSE.b	Bias.c	RMSE.c
Sola Çarpık	40	100		14.842	48.649	-0.392	1.287	0.044	0.090	2.918	4.097	-1.135	2.390	0.070	0.106
		500		1.915	2.406	-0.369	1.296	0.006	0.088	0.519	0.693	-1.080	1.935	0.017	0.088
		1000		1.743	1.886	-0.376	1.321	-0.015	0.090	0.327	0.445	-1.099	1.939	-0.009	0.088
Sağa çarpık	10	100	2PL	14.213	33.351	0.872	1.408	-	-	2.564	3.407	2.733	3.683	-	-
		500		3.031	3.791	0.938	1.450	-	-	1.133	1.395	2.087	2.776	-	-
		1000		2.680	2.908	0.905	1.376	-	-	0.953	1.118	1.928	2.506	-	-
	20	100		10.846	30.001	0.561	1.337	-	-	2.284	3.157	1.885	2.889	-	-
		500		2.287	2.596	0.692	1.362	-	-	0.742	0.882	1.420	2.148	-	-
		1000		2.166	2.324	0.675	1.356	-	-	0.611	0.716	1.345	2.073	-	-
	40	100		8.827	27.069	0.192	1.260	-	-	2.622	3.481	1.114	2.149	-	-
		500		2.014	2.145	0.565	1.387	-	-	0.624	0.711	1.088	1.924	-	-
		1000		1.958	2.057	0.546	1.389	-	-	0.489	0.539	1.055	1.901	-	-
	10	100	3PL	160.172	182.833	0.615	1.120	0.020	0.084	5.807	6.279	6.083	6.659	0.022	0.083
		500		80.392	101.936	0.662	1.140	0.004	0.103	3.586	3.753	4.759	5.109	0.006	0.100
		1000		54.992	74.027	0.706	1.155	0.047	0.105	3.268	3.382	4.541	4.862	0.044	0.100
	20	100		125.224	160.446	0.492	1.230	0.074	0.128	5.219	5.866	4.416	5.224	0.084	0.132
		500		57.438	78.866	0.683	1.207	0.067	0.114	2.967	3.182	3.938	4.441	0.059	0.106
		1000		46.212	64.394	0.759	1.232	0.046	0.117	2.622	2.753	3.740	4.179	0.038	0.108
	40	100		90.350	135.737	0.216	1.274	0.045	0.101	4.701	5.511	2.557	3.724	0.067	0.115
		500		22.797	44.507	0.356	1.190	0.032	0.106	2.420	5.645	2.831	3.487	0.041	0.111
		1000		11.694	25.110	0.572	1.276	0.035	0.105	2.057	2.202	2.642	3.262	0.036	0.105
Sivri	10	100	2PL	0.174	2.874	-3.935	33.508	-	-	-0.628	1.274	-0.066	2.158	-	-
		500		-0.509	0.659	0.005	2.326	-	-	-0.852	0.914	0.006	1.876	-	-
		1000		-0.525	0.645	-0.087	2.214	-	-	-0.868	0.922	-0.086	1.790	-	-
	20	100		-0.289	1.341	0.021	2.086	-	-	-0.647	0.825	0.005	1.539	-	-
		500		-0.170	0.405	-0.040	1.644	-	-	-0.612	0.679	-0.040	1.534	-	-
		1000		-0.215	0.388	-0.020	1.484	-	-	-0.656	0.712	-0.031	1.363	-	-
	40	100		-0.597	1.105	-2.825	67.166	-	-	-0.758	0.864	-0.007	1.672	-	-
		500		-0.675	0.706	0.042	2.458	-	-	-0.937	0.962	0.029	1.580	-	-
		1000		-0.686	0.713	0.001	2.449	-	-	-0.954	0.978	0.003	1.566	-	-
	10	100	3PL	31.922	67.673	-238.014	866.012	0.101	0.129	0.968	3.246	1.879	5.133	0.138	0.157
		500		17.906	38.176	-0.535	2.517	0.076	0.103	-0.317	1.068	0.265	2.886	0.128	0.147
		1000		29.934	52.170	-0.525	2.319	0.048	0.084	-0.414	0.986	0.152	2.647	0.089	0.115

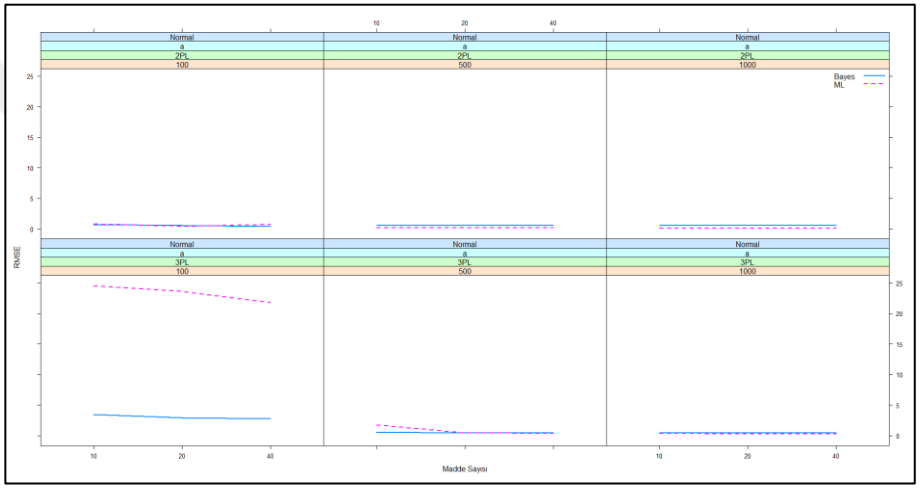
(Devam ediyor)

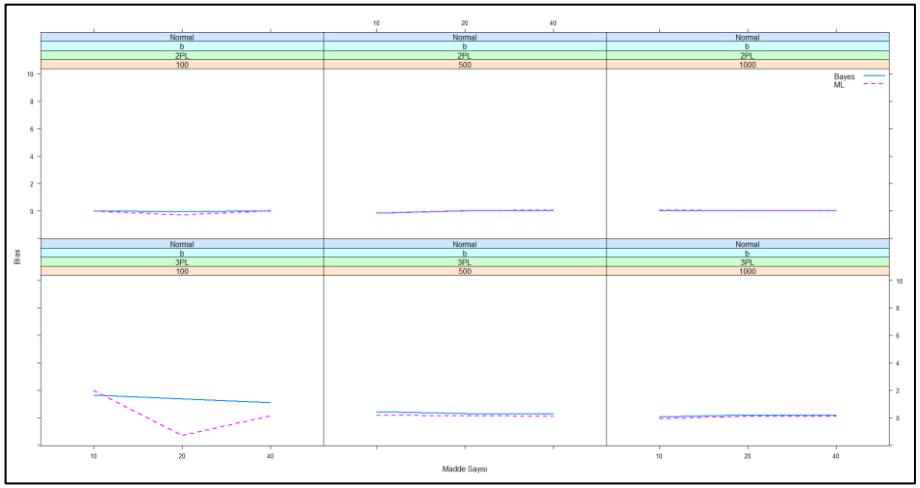
EK 6(Devam)

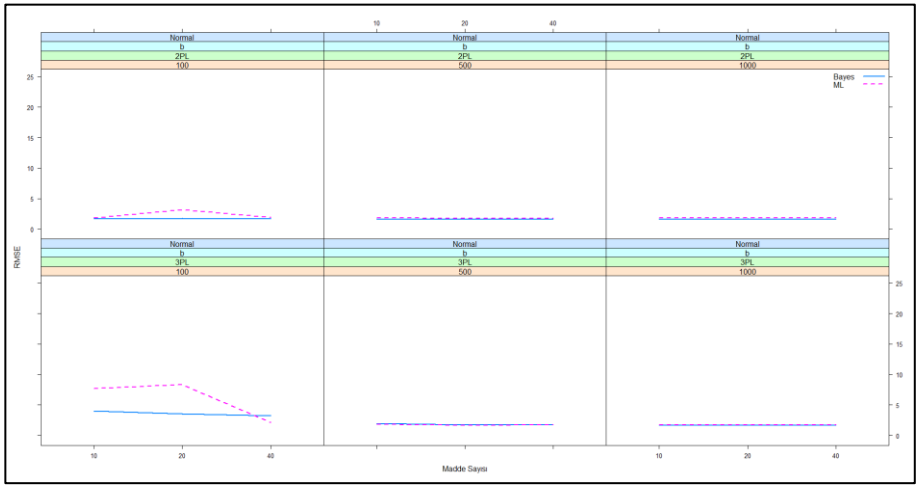
Önsel Yetenek Dağılımı	Madde Sayısı	Örneklem Büyüklüğü	MTK Modeli	En çok olabilirlik kestirimi						Bayes kestirimi					
				Bias.a	RMSE.a	Bias.b	RMSE.b	Bias.c	RMSE.c	Bias.a	RMSE.a	Bias.b	RMSE.b	Bias.c	RMSE.c
Basık	20	100	2PL	32.056	77.272	-0.152	4.029	0.066	0.100	1.628	3.274	2.225	4.826	0.122	0.151
		500		39.615	88.255	0.320	1.885	0.116	0.162	0.164	1.234	1.468	3.214	0.162	0.197
		1000		28.792	66.214	0.065	2.031	0.078	0.141	0.071	1.155	0.962	3.135	0.080	0.132
	40	100		18.707	64.360	-52.334	557.929	0.083	0.110	0.491	2.659	2.298	4.865	0.165	0.198
		500		3.016	17.236	0.202	3.276	0.068	0.101	-0.628	0.911	0.728	2.432	0.103	0.128
		1000		0.730	7.313	0.205	2.795	0.073	0.109	-0.761	0.888	0.562	2.187	0.123	0.158
	10	100		2.385	3.914	-0.026	1.254	-	-	1.042	1.668	-0.030	1.583	-	-
		500		1.563	1.653	-0.002	1.239	-	-	0.347	0.434	-0.006	1.418	-	-
		1000		1.515	1.581	-0.028	1.213	-	-	0.286	0.337	-0.035	1.379	-	-
	20	100		1.783	2.684	-0.007	1.344	-	-	0.819	1.196	-0.030	1.589	-	-
		500		1.349	1.417	0.001	1.367	-	-	0.219	0.286	0.005	1.534	-	-
		1000		1.357	1.411	0.022	1.347	-	-	0.182	0.224	0.026	1.505	-	-
	40	100		1.445	1.911	0.030	1.448	-	-	0.910	1.183	0.040	1.685	-	-
		500		1.253	1.313	0.046	1.408	-	-	0.176	0.236	0.052	1.563	-	-
		1000		1.229	1.275	-0.017	1.387	-	-	0.114	0.154	-0.019	1.530	-	-
	10	100	3PL	19.485	45.586	0.139	1.161	0.079	0.110	3.530	4.687	0.909	2.370	0.109	0.132
		500		2.686	3.539	0.141	0.979	0.025	0.101	1.053	1.350	0.340	1.256	0.031	0.093
		1000		2.162	2.334	0.152	0.967	0.038	0.101	0.721	0.868	0.274	1.177	0.041	0.100
	20	100		10.224	28.862	0.150	1.276	0.065	0.106	2.990	4.133	0.923	2.363	0.089	0.123
		500		1.930	2.097	0.098	1.221	0.023	0.100	0.728	0.904	0.280	1.471	0.034	0.104
		1000		1.792	1.970	0.091	1.204	0.028	0.103	0.517	0.614	0.202	1.416	0.033	0.102
	40	100		7.928	26.089	0.160	1.389	0.051	0.102	3.007	4.086	0.920	2.368	0.082	0.124
		500		1.752	1.889	0.091	1.330	0.026	0.101	0.669	0.797	0.307	1.606	0.039	0.106
		1000		1.600	1.685	0.083	1.308	0.049	0.107	0.425	0.490	0.207	1.531	0.057	0.108

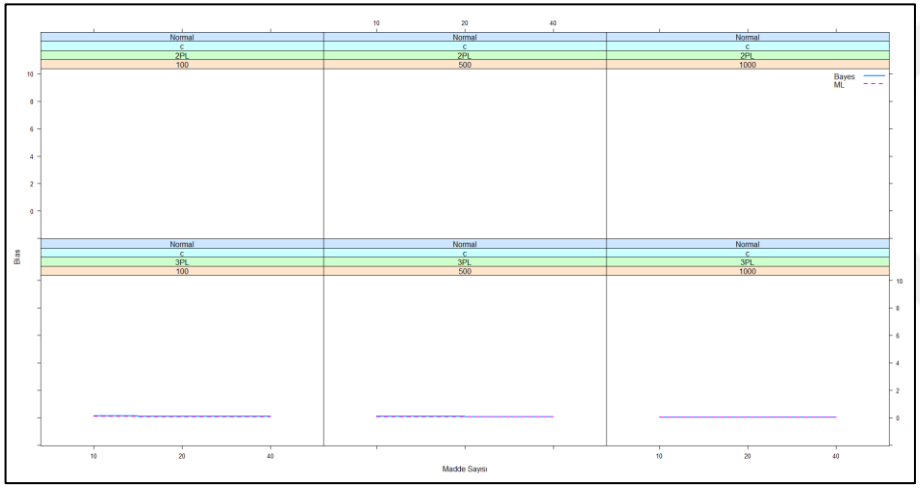
EK 7. Önsel Yetenek Dağılımı Normal Olan ve Olmayan Verilerin En Çok Olabilirlik ve Bayes Yöntemlerine Göre Kestirilen Madde Parametrelerinin Yanlılık (Bias) ve RMSE Değişimleri

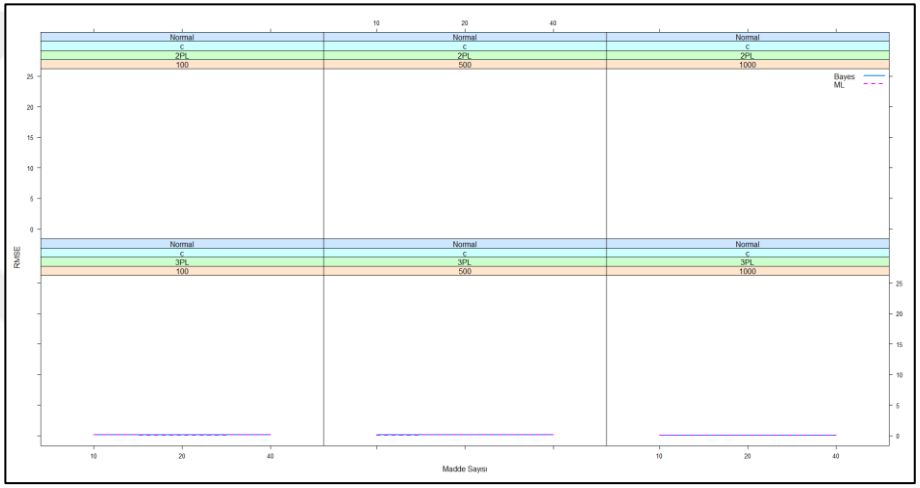


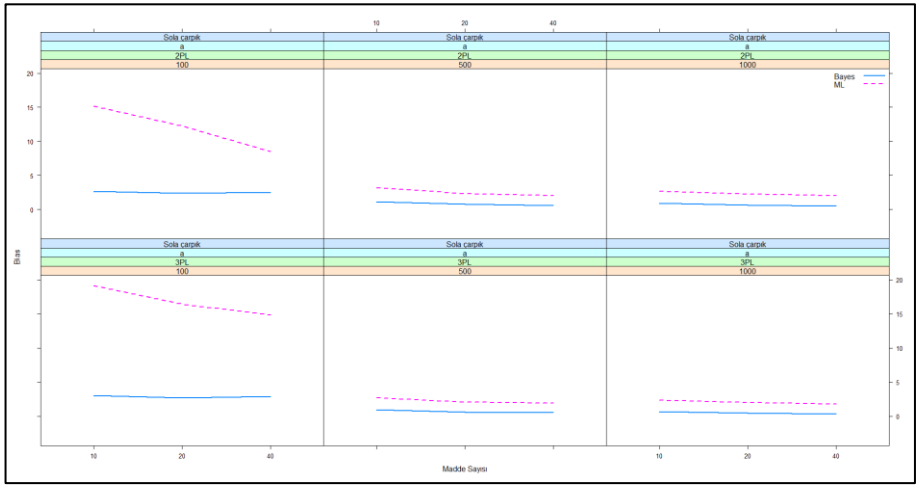


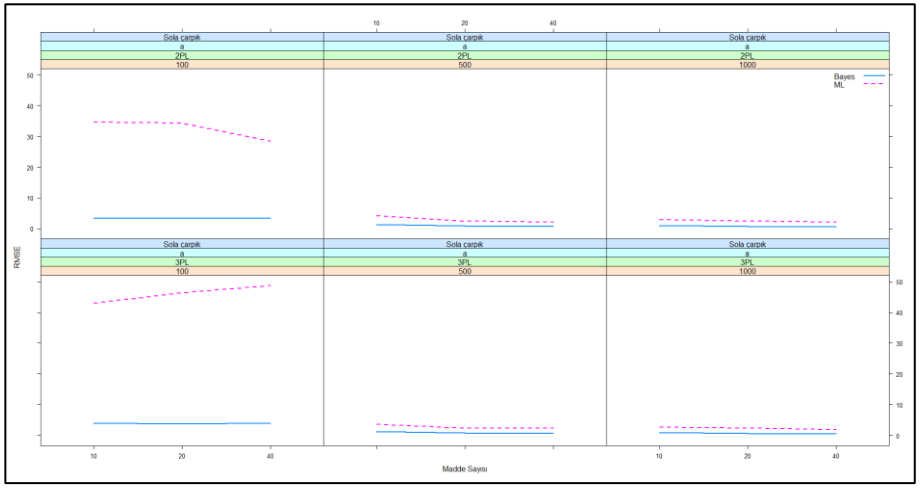


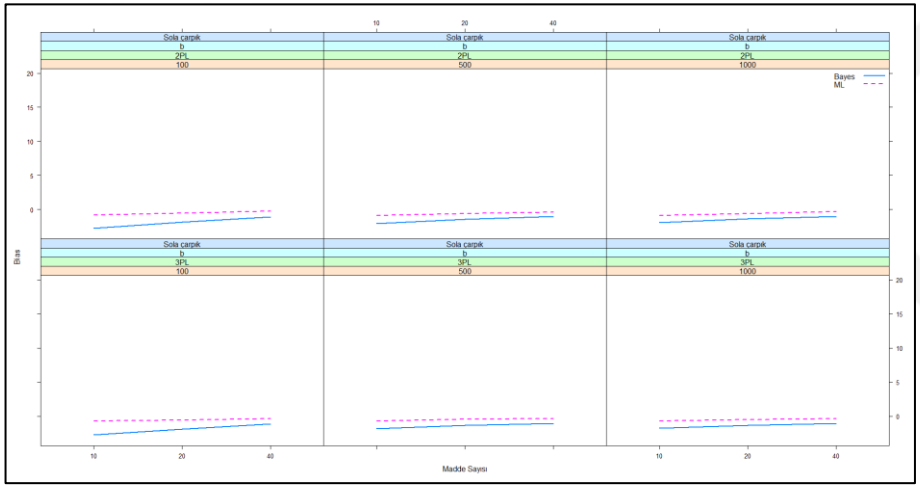


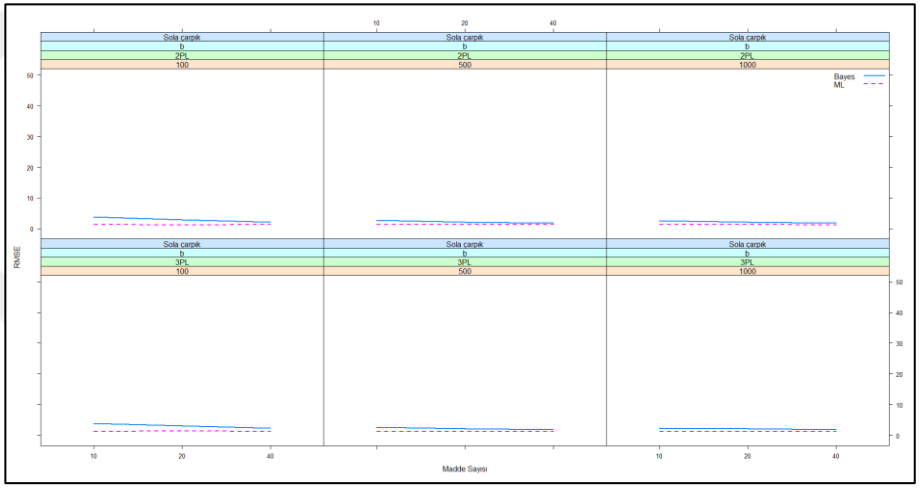

Normal
c
2PL
100
500
1000
3PL
Bayes
ML
Bias
Madde Sayısı

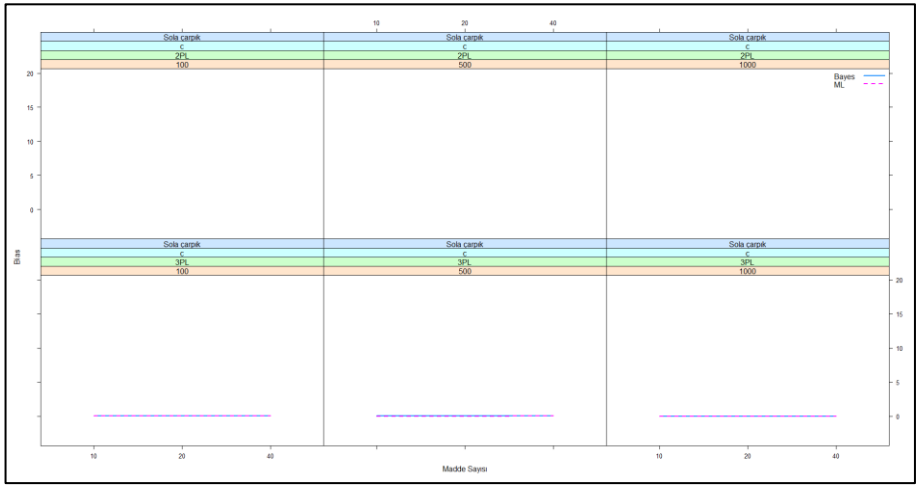

Normal
c
2PL
100
500
1000
3PL
Bayes
ML
RMSE
Madde Sayısı

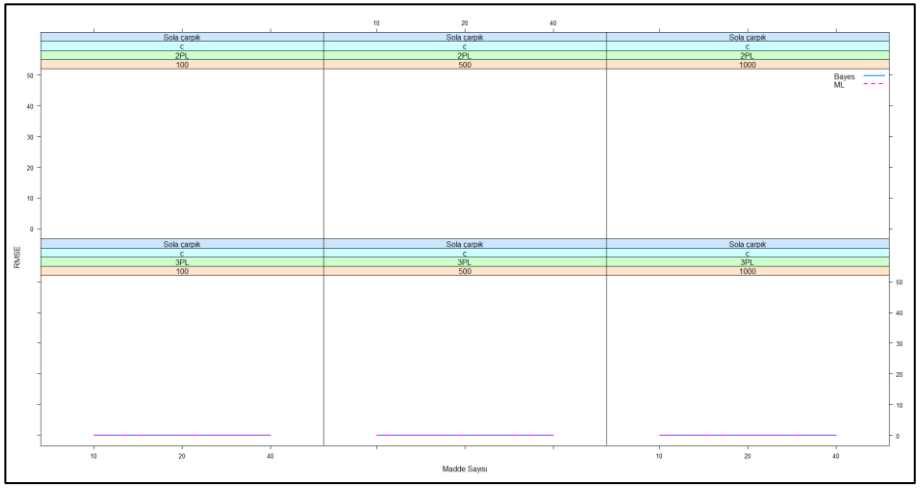

Sola çarpık
a
2PL
100
500
1000
3PL
Bayes
ML
Bias
Madde Sayısı

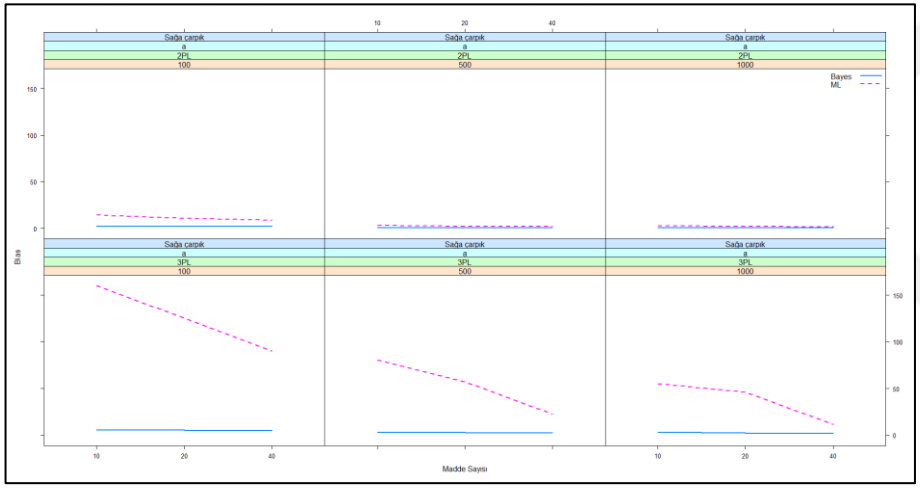

Sola çarpık
a
2PL
100
500
1000
3PL
Bayes
ML
RMSE
Madde Sayısı

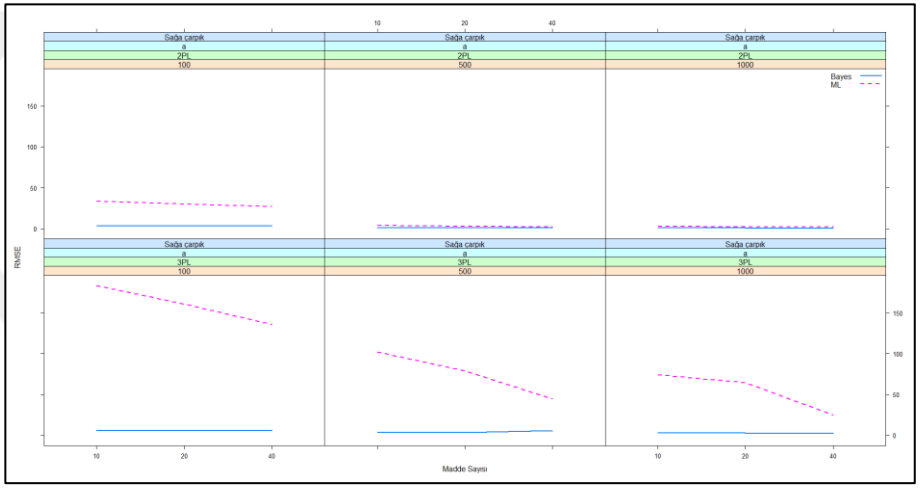

Sola carpık
b
2PL
100
500
1000
3PL
Bayes
ML
Bias
Madde Sayısı

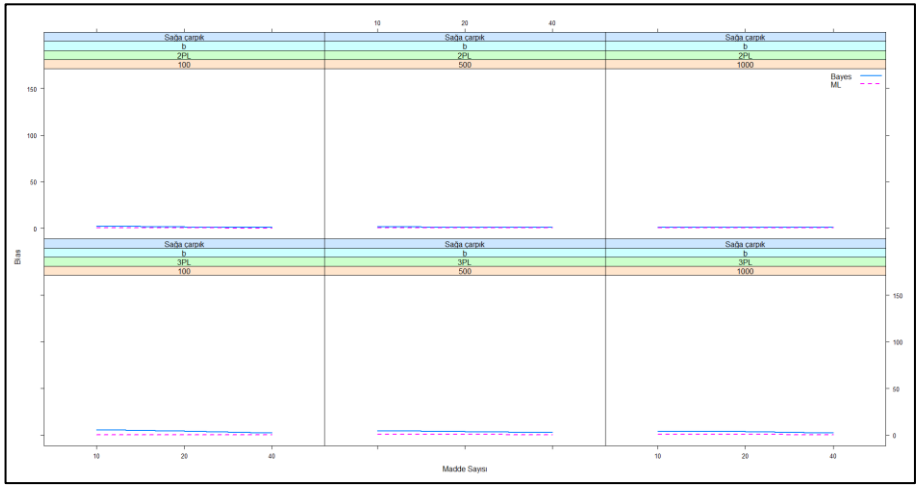

Sola carpık
b
2PL
100
500
1000
3PL
Bayes
ML
RMSE
Madde Sayısı

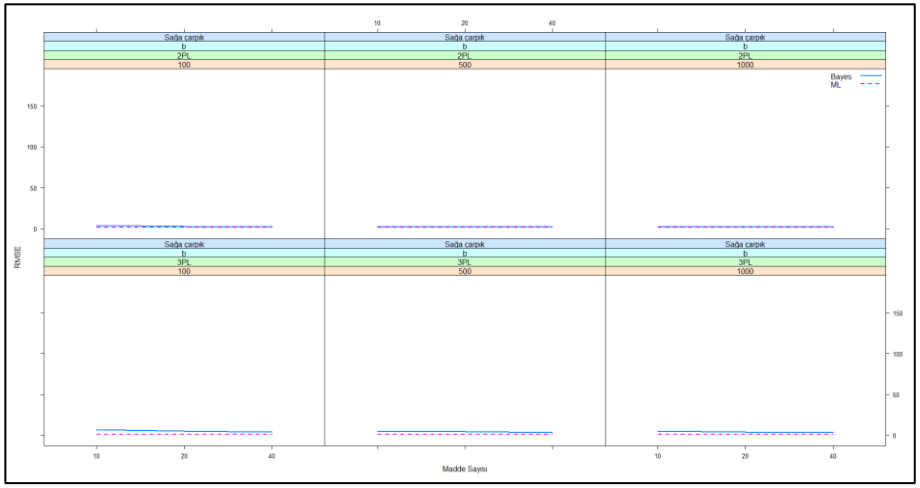

Sola carpık
c
2PL
100
500
1000
3PL
Bayes
ML
Bias
Madde Sayısı

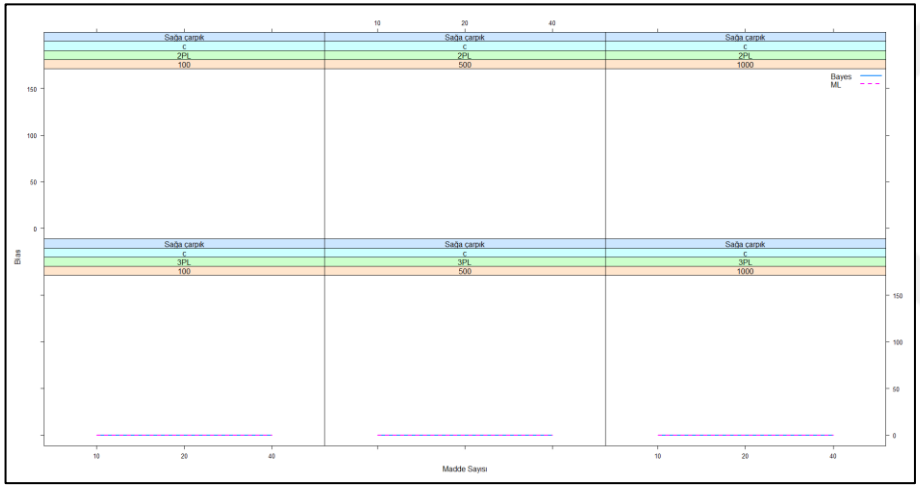

Sola carpık
c
2PL
100
500
1000
3PL
Bayes
ML
RMSE
Madde Sayısı

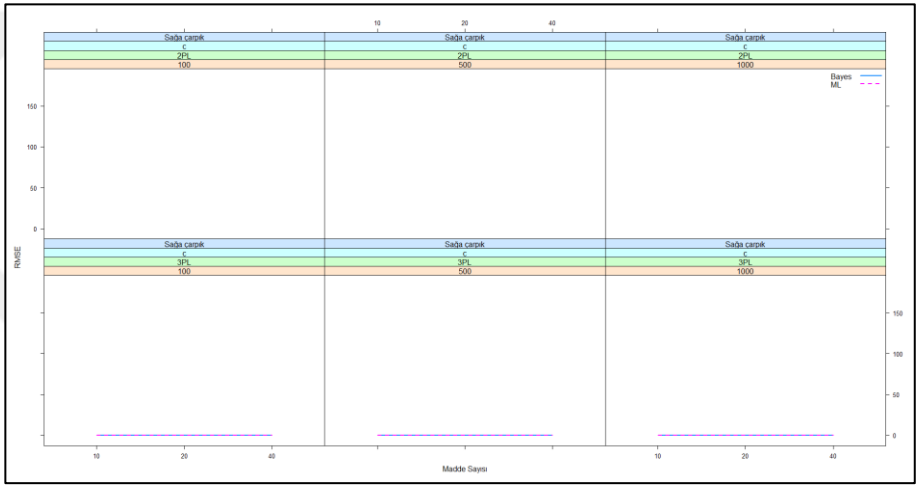

Sağa çarpık
a
2PL
100
500
1000
3PL
Bayes
ML
Bias
Madde Sayısı

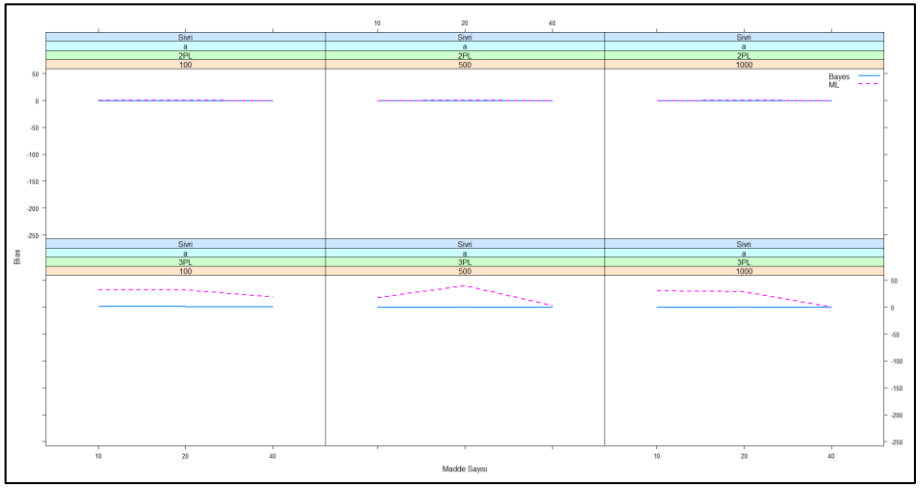

Sağa çarpık
a
2PL
100
500
1000
3PL
Bayes
ML
RMSE
Madde Sayısı

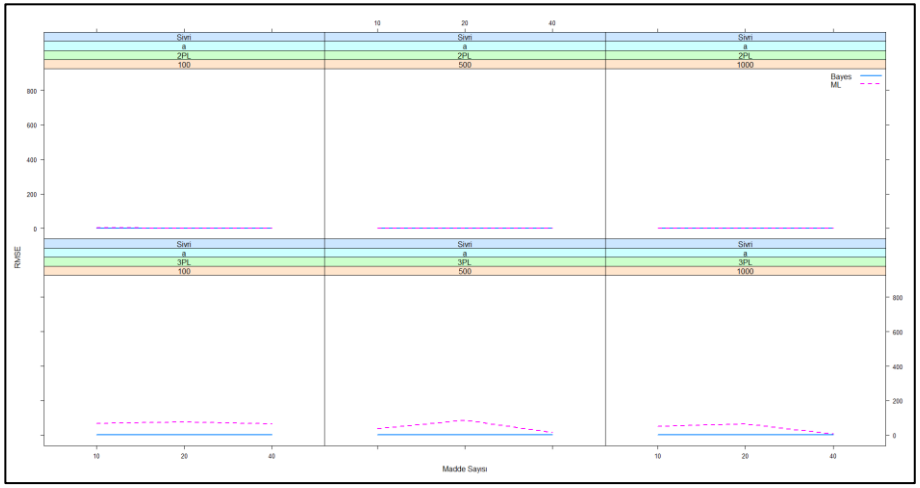

Sağa çarpık
b
2PL
100
500
1000
3PL
Bayes
ML
Bias
Madde Sayısı

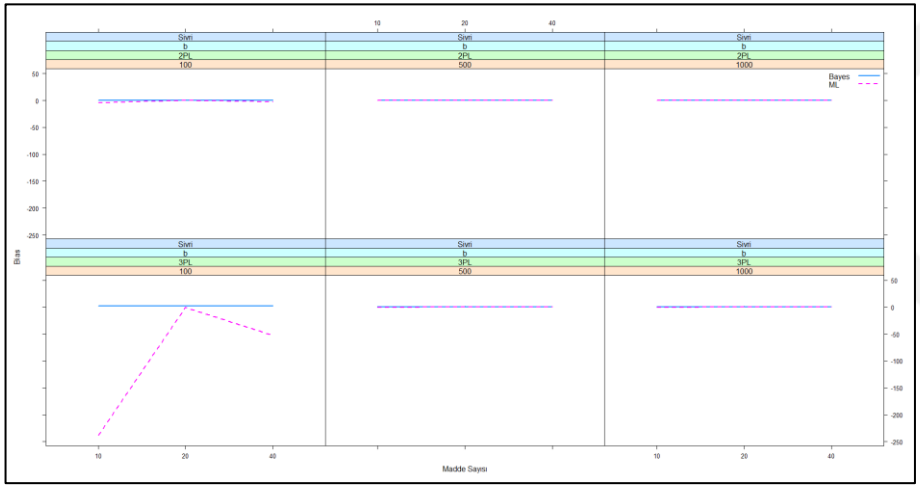

Sağa çarpık
b
2PL
100
500
1000
3PL
Bayes
ML
RMSE
Madde Sayısı

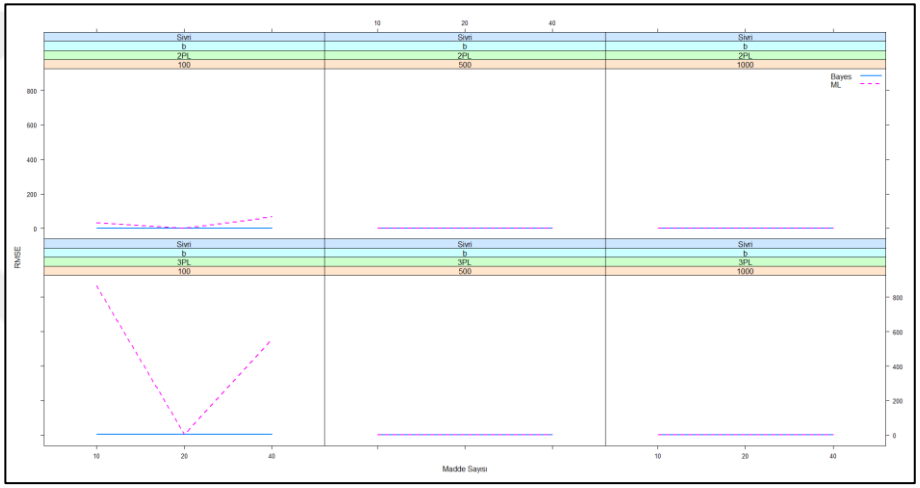

Sağa çarpık
c
2PL
100
500
1000
3PL
Bayes
ML
Bias
Madde Sayısı

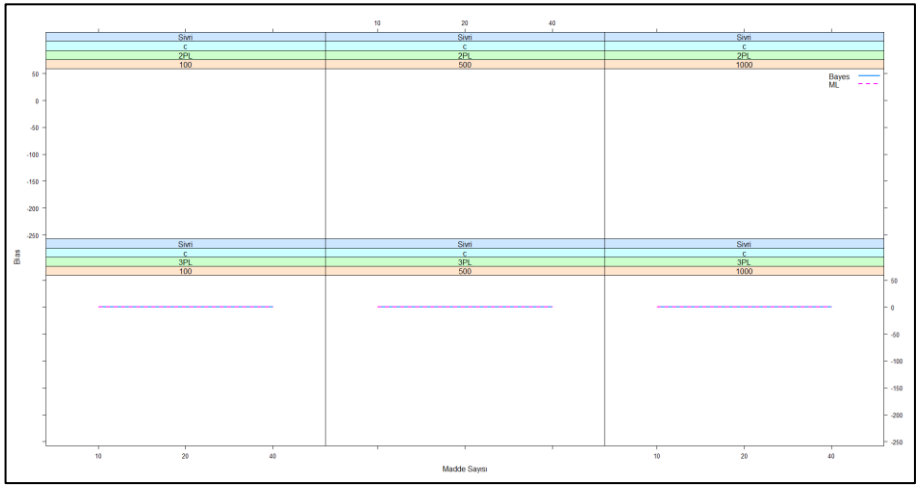

Sağa çarpık
c
2PL
100
500
1000
3PL
Bayes
ML
RMSE
Madde Sayısı

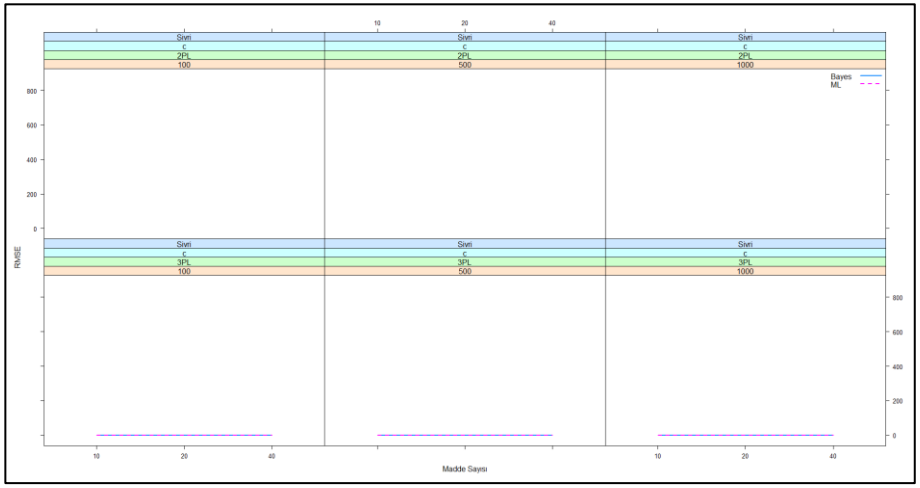

Sivri
a
2PL
100
500
1000
3PL
Bayes
ML
Bias
Madde Sayısı

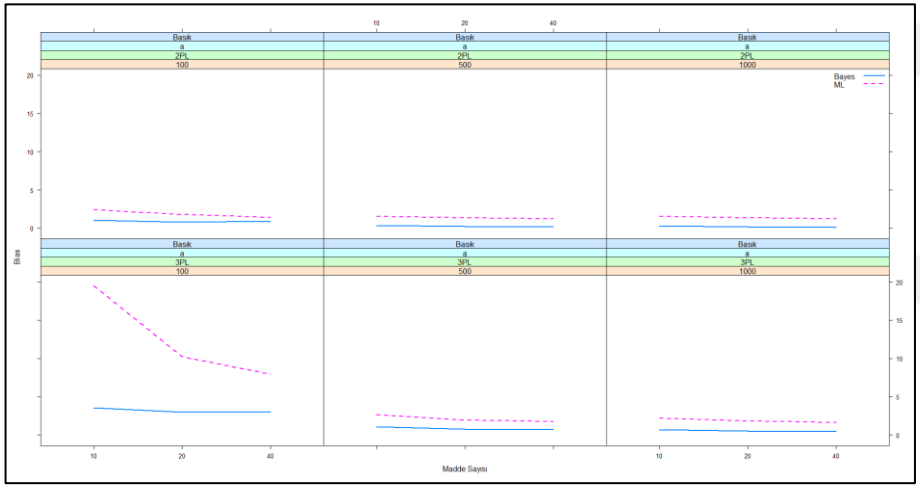

Sivri
a
2PL
100
500
1000
3PL
Bayes
ML
RMSE
Madde Sayısı

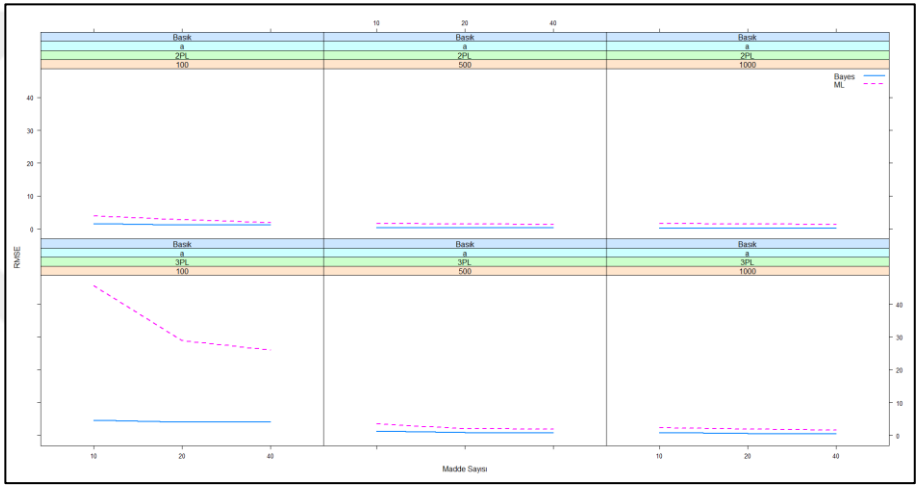

Sivri
b
2PL
100
500
1000
3PL
Bayes
ML
Bias
Madde Sayısı

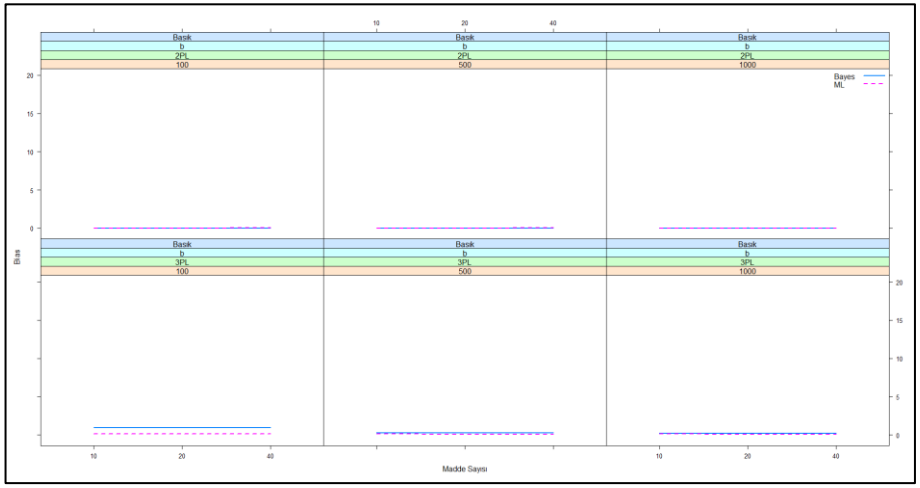

Sivri
b
2PL
3PL
100
500
1000
Bayes
ML
RMSE
Madde Sayısı

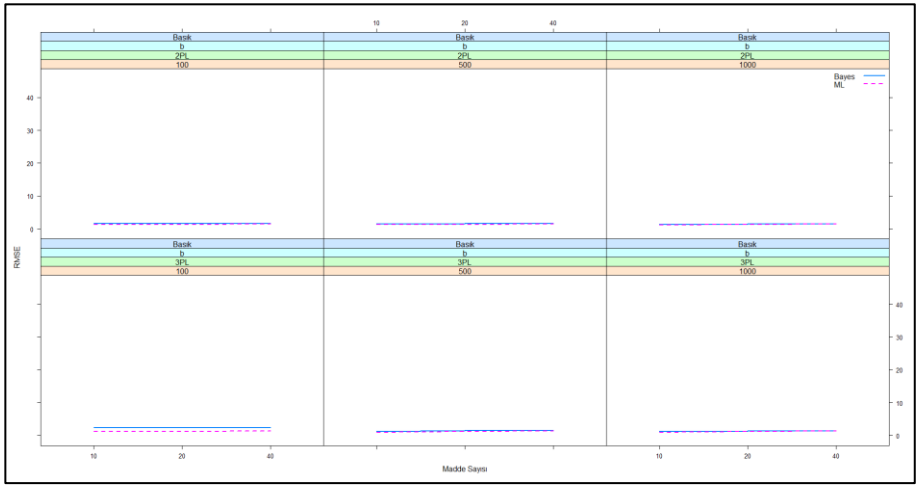

Sivri
c
2PL
3PL
100
500
1000
Bayes
ML
Bias
Madde Sayısı

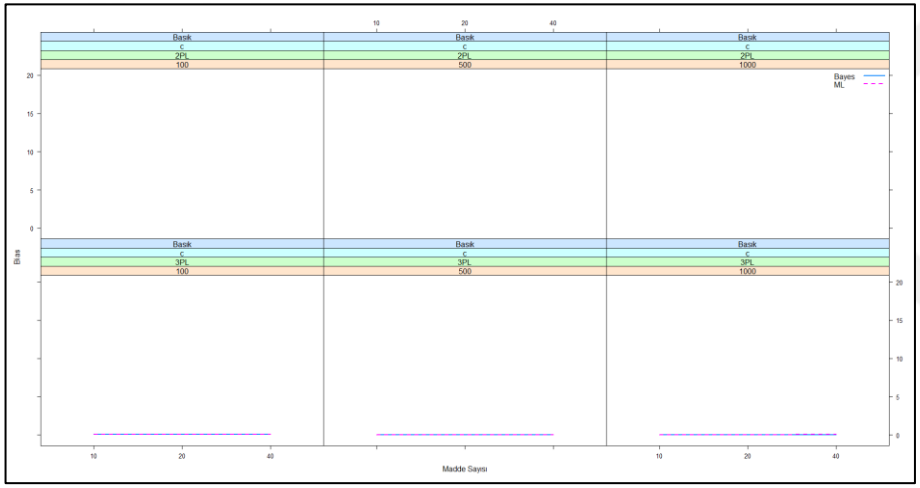

Sivri
c
2PL
3PL
100
500
1000
Bayes
ML
RMSE
Madde Sayısı

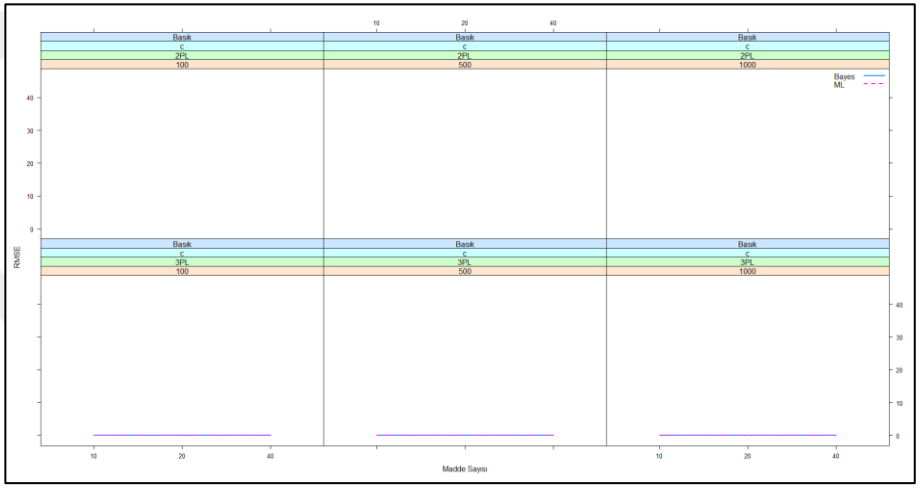

Basik
a
2PL
100
500
1000
3PL
Bayes
ML
Bias
Madde Sayısı


Basik
a
2PL
100
500
1000
3PL
Bayes
ML
RMSE
Madde Sayısı


Basik
b
2PL
100
500
1000
3PL
Bayes
ML
Bias
Madde Sayısı


Basik
b
2PL
100
500
1000
3PL
Bayes
ML
RMSE
Madde Sayısı


Basik
c
2PL
100
500
1000
3PL
Bayes
ML
Bias
Madde Sayısı


Basik
c
2PL
100
500
1000
3PL
Bayes
ML
RMSE
Madde Sayısı

EK 8. Önsel Yetenek Dağılımı Normal Olan ve Olmayan Verilerin En Çok Olabilirlik ve Bayes Yöntemleriyle Kestirilen Yetenek Parametreleri ve Standart Hata Değerleri

Önsel Yetenek Dağılımı	Madde Sayısı	Örneklem Büyüklüğü	MTK Modeli	En çok olabilirlik kestirimi		Bayes kestirimi	
				Θ	$Θ_{SH}$	Θ	$Θ_{SH}$
Normal	10	100	2PL	0.008	8.174	0.000	0.477
		500		0.067	8.370	0.000	0.494
		1000		-0.015	8.014	0.000	0.500
	20	100		0.008	2.514	0.000	0.349
		500		0.009	2.568	0.000	0.374
		1000		-0.002	2.640	-0.001	0.376
	40	100		0.008	0.746	0.003	0.240
		500		0.002	0.771	-0.001	0.272
		1000		0.005	0.761	0.001	0.276
	10	100	3PL	-0.043	11.059	-0.019	0.509
		500		-0.080	11.510	-0.002	0.561
		1000		0.010	11.631	0.000	0.560
	20	100		-0.096	5.242	-0.048	0.397
		500		-0.112	5.367	0.000	0.432
		1000		-0.087	4.945	0.000	0.435
	40	100		-0.095	2.347	-0.084	0.279
		500		-0.082	2.342	-0.001	0.323
		1000		-0.073	2.154	-0.001	0.326
Sola çarpık	10	100	2PL	1.755	51.339	0.018	0.422
		500		1.721	51.442	0.000	0.435
		1000		1.763	51.906	-0.001	0.436
	20	100		0.845	33.789	0.025	0.310
		500		0.735	32.356	0.000	0.330
		1000		0.752	32.957	0.001	0.332
	40	100		0.290	17.735	0.054	0.207
		500		0.194	18.097	0.002	0.239
		1000		0.193	18.637	0.000	0.424
	10	100	3PL	2.215	69.175	0.035	0.442
		500		2.165	47.835	0.003	0.466
		1000		2.205	46.802	0.000	0.467
	20	100		1.124	31.195	0.034	0.333
		500		1.134	28.850	0.002	0.364
		1000		1.168	28.050	0.000	0.368
	40	100		0.476	17.230	0.066	0.229
		500		0.492	16.341	0.004	0.271
		1000		0.556	15.478	0.000	0.276
Sağa çarpık	10	100	2PL	-1.657	50.825	-0.019	0.422
		500		-1.741	51.524	-0.002	0.435
		1000		-1.738	51.732	0.000	0.436
	20	100		-0.744	32.221	-0.015	0.309
		500		-0.707	32.742	0.001	0.331
		1000		-0.631	31.816	-0.001	0.331
	40	100		-0.366	18.518	-0.057	0.206
		500		-0.103	18.222	-0.002	0.239
		1000		-0.095	18.182	0.001	0.241
	10	100	3PL	-1.078	99999.999	-0.122	0.530
		500		-1.083	99999.999	-0.053	0.605
		1000		-0.891	99999.999	-0.039	0.610

(Devam ediyor)

EK 8 (Devam)

Önsel Yetenek Dağılımı	Madde Sayısı	Örneklem Büyüklüğü	MTK Modeli	En çok olabilirlik kestirimi		Bayes kestirimi	
				Θ	Θ_{SH}	Θ	Θ_{SH}
Sağa çarpık	20	100	2PL	-1.220	99999.999	-0.210	0.426
		500		-1.307	99999.999	-0.073	0.516
		1000		-1.069	99999.999	-0.053	0.526
	40	100		-1.132	92.727	-0.369	0.327
		500		-1.102	99999.999	-0.104	0.415
		1000		-0.902	16745.25	-0.073	0.423
Sivri	10	100	2PL	0.007	9.242	0.000	0.624
		500		0.001	9.129	0.000	0.663
		1000		0.007	9.301	0.000	0.667
	20	100		-0.008	8.002	-0.001	0.433
		500		-0.008	12.235	0.000	0.401
		1000		-0.004	12.479	0.000	0.409
	40	100		0.002	0.888	0.001	0.356
		500		0.001	0.732	0.001	0.435
		1000		0.002	0.722	0.000	0.450
	10	100	3PL	0.165	8.874	-0.024	0.596
		500		0.207	8.033	-0.006	0.728
		1000		0.187	205.443	-0.005	0.720
	20	100		0.105	8.481	-0.079	0.450
		500		0.136	8.487	-0.023	0.567
		1000		0.145	10.002	-0.016	0.502
	40	100		-0.050	2.116	-0.099	0.398
		500		-0.029	1.427	-0.012	0.554
		1000		-0.033	1.351	-0.005	0.567
Basık	10	100	2PL	0.035	27.496	-0.001	0.359
		500		0.002	28.234	0.000	0.364
		1000		0.015	28.226	-0.001	0.365
	20	100		0.005	12.681	-0.008	0.275
		500		-0.018	12.383	0.001	0.279
		1000		-0.015	12.911	-0.001	0.279
	40	100		-0.027	3.775	-0.001	0.196
		500		-0.018	4.322	-0.002	0.205
		1000		0.001	4.207	0.000	0.206
	10	100	3PL	0.063	134.175	-0.018	0.417
		500		0.126	85.032	-0.002	0.436
		1000		0.121	34.491	0.001	0.432
	20	100		-0.143	18.773	-0.038	0.331
		500		-0.138	18.971	0.000	0.352
		1000		-0.101	18.646	-0.001	0.350
	40	100		-0.284	9.611	-0.080	0.243
		500		-0.246	10.327	0.000	0.273
		1000		-0.227	9.831	-0.001	0.273

EK 9. Önsel Yetenek Dağılımı Normal Olan ve Olmayan Verilerin En Çok Olabilirlik ve Bayes Yöntemleriyle Kestirilen Madde Parametreleri ve Standart Hata Değerleri

Önsel Dağılım	Madde Sayısı	Örneklem Büyüklüğü	MTK Modeli	En çok olabilirlik kestirimi						Bayes kestirimi					
				a	a_{SH}	b	b_{SH}	c	c_{SH}	a	a_{SH}	b	b_{SH}	c	c_{SH}
Normal	10	100	2PL	1.514	100.602	-0.016	100.380	-	-	1.046	0.362	0.002	0.222	-	-
		500		1.400	0.203	-0.079	0.132	-	-	0.839	0.115	-0.060	0.083	-	-
		1000		1.369	0.140	0.017	0.087	-	-	0.811	0.079	0.017	0.057	-	-
	20	100		1.428	0.554	-0.274	34.787	-	-	1.040	0.303	-0.014	0.205	-	-
		500		1.379	0.183	-0.003	0.121	-	-	0.832	0.100	-0.001	0.079	-	-
		1000		1.385	0.127	0.017	0.084	-	-	0.821	0.070	0.015	0.056	-	-
	40	100		1.454	25.971	0.026	25.820	-	-	1.159	0.313	0.011	0.200	-	-
		500		1.378	0.181	0.028	0.125	-	-	0.847	0.096	0.019	0.079	-	-
		1000		1.373	0.121	0.015	0.085	-	-	0.819	0.065	0.007	0.055	-	-
	10	100	3PL	10.962	6702.032	1.977	6432.308	0.202	0.084	3.261	1.339	1.632	1.119	0.268	0.076
		500		2.045	400.598	0.149	400.248	0.189	0.072	1.217	0.322	0.411	0.299	0.220	0.065
		1000		1.540	0.310	0.023	0.181	0.177	0.066	0.999	0.188	0.140	0.182	0.199	0.059
	20	100		8.158	4052.060	-1.270	4030.537	0.203	0.083	2.740	1.091	1.370	0.883	0.259	0.075
		500		1.600	0.375	0.112	0.206	0.182	0.067	1.088	0.241	0.285	0.237	0.206	0.061
		1000		1.535	0.264	0.102	0.166	0.176	0.059	0.968	0.152	0.182	0.162	0.192	0.053
	40	100		6.291	2552.593	0.127	2526.697	0.200	0.085	2.684	1.086	1.098	0.745	0.250	0.074
		500		1.581	0.359	0.107	0.207	0.184	0.065	1.081	0.218	0.270	0.220	0.205	0.058
		1000		1.517	0.239	0.094	0.160	0.175	0.056	0.962	0.138	0.172	0.150	0.190	0.050
Sola çarpık	10	100	2PL	16.584	8751.019	-0.634	8559.388	-	-	4.022	1.149	-2.640	0.733	-	-
		500		4.603	301.186	-0.711	300.091	-	-	2.519	0.352	-1.863	0.243	-	-
		1000		4.083	0.518	-0.710	0.063	-	-	2.338	0.229	-1.747	0.160	-	-
	20	100		13.629	7912.819	-0.465	8029.305	-	-	3.749	1.031	-1.826	0.522	-	-
		500		3.687	0.635	-0.565	0.098	-	-	2.134	0.274	-1.375	0.170	-	-
		1000		3.662	0.416	-0.535	0.066	-	-	2.016	0.179	-1.333	0.115	-	-
	40	100		9.891	45161.52	-0.257	42604.000	-	-	3.926	1.118	-1.093	0.418	-	-
		500		3.439	0.594	-0.440	0.108	-	-	1.983	0.241	-1.094	0.139	-	-
		1000		3.420	0.374	-0.414	0.068	-	-	1.874	0.159	-1.088	0.095	-	-

(Devam ediyor)

EK 9 (Devam)

Önsel Dağılım	Madde Sayısı	Örneklem Büyüklüğü	MTK Modeli	En çok olabilirlik kestirimi						Bayes kestirimi					
				a	a_{SH}	b	b_{SH}	c	c_{SH}	a	a_{SH}	b	b_{SH}	c	c_{SH}
Sola çarpık	10	100	3PL	20.531	10106.180	-0.668	9927.982	0.167	0.074	4.434	1.392	-2.624	0.840	0.199	0.071
		500		4.118	200.977	-0.671	200.118	0.114	0.048	2.347	0.403	-1.800	0.270	0.131	0.045
		1000		3.735	0.741	-0.687	0.086	0.091	0.036	2.111	0.248	-1.713	0.171	0.101	0.034
	20	100		17.823	7655.042	-0.570	7539.154	0.170	0.075	4.066	1.307	-1.875	0.628	0.199	0.069
		500		3.474	0.860	-0.523	0.131	0.121	0.046	2.013	0.323	-1.404	0.191	0.134	0.043
		1000		3.361	50.532	-0.554	50.091	0.104	0.036	1.822	0.203	-1.382	0.129	0.112	0.034
	40	100		16.217	13649.85	-0.399	13432.500	0.171	1000.067	4.293	1.387	-1.142	0.512	0.197	0.068
		500		3.290	75.706	-0.434	75.149	0.130	0.047	1.894	0.295	-1.146	0.159	0.142	0.042
		1000		3.126	0.415	-0.436	0.099	0.116	0.036	1.709	0.185	-1.160	0.108	0.122	0.033
Sağa çarpık	10	100	2PL	15.606	6504.670	0.670	6500.278	-	-	3.957	1.160	2.532	0.718	-	-
		500		4.436	200.824	0.731	154.103	-	-	2.538	0.352	1.880	0.245	-	-
		1000		4.078	0.517	0.729	0.064	-	-	2.351	0.234	1.751	0.159	-	-
	20	100		12.238	7069.110	0.499	7050.799	-	-	3.675	1.053	1.823	0.521	-	-
		500		3.675	50.625	0.652	15.834	-	-	2.130	0.274	1.380	0.168	-	-
		1000		3.550	0.399	0.644	0.067	-	-	1.995	0.179	1.314	0.113	-	-
	40	100		10.209	48113.310	0.188	44844.580	-	-	4.003	1.135	1.111	0.411	-	-
		500		3.388	0.589	0.579	0.110	-	-	1.998	0.245	1.102	0.139	-	-
		1000		3.331	0.361	0.584	0.070	-	-	1.862	0.156	1.093	0.095	-	-
	10	100	3PL	161.572	87537.600	0.421	86146.690	0.149	300.041	7.026	1.725	5.890	1.569	0.155	0.037
		500		81.808	58312.770	0.464	48164.920	0.139	400.019	5.003	0.798	4.561	0.806	0.137	0.019
		1000		56.417	40105.520	0.491	28483.710	0.137	700.009	4.693	0.600	4.326	0.604	0.133	0.014
	20	100		126.616	69598.400	0.312	68734.770	0.177	200.051	6.611	1.721	4.236	1.355	0.188	0.041
		500		58.835	45874.270	0.368	40698.640	0.172	400.022	4.364	0.733	3.624	0.687	0.170	0.022
		1000		47.611	34356.930	0.418	26698.090	0.168	450.013	4.021	0.535	3.398	0.509	0.161	0.016
	40	100		91.725	49594.71	0.141	49121.560	0.175	1150.052	6.075	1.677	2.482	1.105	0.191	0.043
		500		24.182	16832.34	0.247	14929.460	0.171	525.024	3.805	0.661	2.722	0.558	0.175	0.023
		1000		13.079	5577.728	0.488	4777.787	0.174	75.019	3.442	0.474	2.557	0.413	0.173	0.017

(Devam ediyor)

EK 9 (Devam)

Önsel Dağılım	Madde Sayısı	Örneklem Büyüklüğü	MTK Modeli	En çok olabilirlik kestirimi						Bayes kestirimi					
				a	a_{SH}	b	b_{SH}	c	c_{SH}	a	a_{SH}	b	b_{SH}	c	c_{SH}
Sivri	10	100	2PL	1.558	531.144	-3.881	1662.164	-	-	0.756	0.422	-0.012	0.275	-	-
		500		0.869	0.238	0.002	0.299	-	-	0.527	0.120	0.002	0.082	-	-
		1000		0.857	0.163	-0.036	0.195	-	-	0.513	0.080	-0.035	0.055	-	-
	20	100		1.091	100.679	0.029	101.160	-	-	0.734	0.305	0.013	0.181	-	-
		500		1.205	0.283	-0.010	0.167	-	-	0.762	0.116	-0.010	0.076	-	-
		1000		1.167	0.190	0.003	0.110	-	-	0.725	0.078	-0.009	0.051	-	-
	40	100		0.777	26.233	-2.814	1267.085	-	-	0.615	0.280	0.005	0.167	-	-
		500		0.710	0.172	0.024	0.365	-	-	0.447	0.086	0.012	0.067	-	-
		1000		0.693	0.111	0.000	0.234	-	-	0.425	0.059	0.001	0.047	-	-
	10	100	3PL	33.304	19019.720	-237.904	21502.540	0.213	0.085	2.350	1.284	1.989	1.219	0.282	0.079
		500		19.270	12600.760	-0.199	11687.680	0.209	0.078	1.048	0.352	0.601	0.423	0.253	0.076
		1000		31.3110	17502.150	-0.195	15967.340	0.202	0.067	0.963	0.213	0.482	0.285	0.230	0.066
	20	100		33.444	17404.400	-0.168	17323.050	0.219	150.082	3.016	1.206	2.209	1.119	0.286	0.076
		500		40.995	21964.490	0.265	21188.610	0.246	0.065	1.544	0.394	1.413	0.512	0.278	0.063
		1000		30.169	17801.590	0.104	16630.370	0.200	0.048	1.447	0.247	1.001	0.314	0.212	0.048
	40	100		20.086	9483.158	-52.358	12412.270	0.214	0.090	1.870	1.158	2.274	1.117	0.286	0.079
		500		4.392	2575.917	0.201	2463.294	0.212	0.081	0.749	0.257	0.727	0.388	0.255	0.074
		1000		2.109	1075.588	0.182	1028.198	0.210	0.075	0.617	0.165	0.539	0.281	0.241	0.069
Basık	10	100	2PL	3.780	301.021	-0.003	300.164	-	-	2.436	0.666	-0.008	0.346	-	-
		500		2.970	0.351	-0.001	0.070	-	-	1.753	0.196	-0.006	0.119	-	-
		1000		2.908	0.239	-0.004	0.049	-	-	1.679	0.130	-0.011	0.082	-	-
	20	100		3.167	200.976	0.003	200.191	-	-	2.203	0.538	-0.020	0.307	-	-
		500		2.724	0.298	0.006	0.074	-	-	1.594	0.155	0.010	0.111	-	-
		1000		2.745	0.208	0.007	0.051	-	-	1.570	0.107	0.010	0.077	-	-
	40	100		2.821	51.624	0.009	50.372	-	-	2.285	0.507	0.020	0.298	-	-
		500		2.635	0.288	0.023	0.078	-	-	1.559	0.141	0.028	0.108	-	-
		1000		2.610	0.193	-0.002	0.053	-	-	1.495	0.095	-0.004	0.074	-	-
	10	100	3PL	20.889	10713.510	0.134	10229.340	0.180	0.067	4.933	1.532	0.904	0.838	0.210	0.060
		500		4.168	300.957	0.087	247.293	0.151	0.045	2.535	0.476	0.286	0.245	0.163	0.039
		1000		3.669	0.579	0.075	0.076	0.135	0.034	2.228	0.301	0.197	0159	0.142	0.031
	20	100		11.616	5402.914	0.127	5165.586	0.183	0.069	4.382	1.374	0.900	0.737	0.213	0.060
		500		3.341	0.653	0.085	0.109	0.163	0.045	2.139	0.367	0.267	0.221	0.175	0.040
		1000		3.195	0.431	0.076	0.080	0.154	0.035	1.919	0.234	0.187	0.148	0.160	0.032
	40	100		9.310	3758.184	0.143	3712.992	0.186	0.076	4.389	1.364	0.903	0.671	0.216	0.059
		500		3.126	0.591	0.094	0.114	0.168	0.045	2.043	0.324	0.309	0.213	0.181	0.039
		1000		2.973	0.377	0.084	0.082	0.162	0.034	1.797	0.020	0.208	0.143	0.169	0.032

EK 10. Etik Kurul Onayı

ANKARA ÜNİVERSİTESİ
SOSYAL BİLİMLER ALT ETİK KURULU
KARAR ÖRNEĞİ

Karar Tarihi : **04/11/2019**

Toplantı Sayısı : **13**

Karar Sayısı : **399**

399- Üniversitemiz Eğitim Bilimleri Enstitüsü Eğitim Bilimleri Anabilim Dalı doktora öğrencisi **Eray Selçuk**'un "Klasik ve Bayesçi Madde Tepki Kuramı Parametre Kestirim Yöntemlerinin Farklı Simülasyon Koşullarında Karşılaştırılması" başlıklı tezi ile ilgili 11/10/2019 tarihli "İnsan Üzerinde Yapılan Klinik Dışı Araştırmalar Başvuru Formu" Etik Kurulumuzca incelendi.

Üniversitemiz Eğitim Bilimleri Enstitüsü Eğitim Bilimleri Anabilim Dalı doktora öğrencisi **Eray Selçuk**'un "Klasik ve Bayesçi Madde Tepki Kuramı Parametre Kestirim Yöntemlerinin Farklı Simülasyon Koşullarında Karşılaştırılması" başlıklı tezinin, insan üzerinde yapılan bir çalışma olmaması nedeniyle etik kurul onayına ihtiyaç duymadığına oy birliği ile karar verildi.

BENZERLİK BİLDİRİMİ

"Klasik ve Bayesçi Yaklaşımlara Göre Madde Tepki Kuramı Parametre Kestirimlerinin Farklı Simülasyon Koşullarında Karşılaştırılması" başlıklı doktora tezimin ana bölümü (ön bölüm, kaynaklar ve ekler hariç) Turnitin İntihali Engelleme Programı aracılığıyla incelenmiş ilgili rapor tez danışmanım tarafından da kontrol edilmiştir. Kontrol sırasında (1) "Kapak, Önsöz, Özet, İçindekiler, Kısaltmalar/Simgeler Sayfası, Tablolar Dizini, Şekiller Dizini, Dipnotlar, Kaynaklar, Ekler, Özgeçmiş" (2) "Tırnak içinde ve blok olarak verilen (doğrudan) alıntılar" (3) Yedi sözcüğe kadar olan benzerlik/örtüşme içeren metin kısımları" (4) "Öğrencinin lisansüstü teziyle ilgili yapmış olduğu yayınlar ve yararlanılan mevzuat metinleri" dışarıda tutulmuştur. Benzerlik kontrolüne ilişkin rapordan elde edilen sonuçlar aşağıdaki tabloda sunulmuştur.

Rapor Tarihi	:20/02/2023
Gönderim Numarası	: 1993526759
Sayfa Sayısı	: 109
Sözcük Sayısı	: 29193
Karakter Sayısı	: 182965
Benzerlik Oranı	: % 4
Savunma Tarihi	: 09/02/2023

Yukarıda belirtilen sonuçları gösteren Turnitin İntihali Engelleme Programı'na ilişkin orijinal raporu, sonuçlarda herhangi bir değişiklik yapmaksızın bu beyanım ekinde Enstitüye teslim ettiğimi, tezimin **%10**'dan fazla benzerlik oranı içerdiğinin ve tek bir kaynakla eşleşme oranının **%2**'yi geçtiğinin belirlenmesi durumunda, bundan doğabilecek tüm yasal sorumluluğu kabul ettiğimi bildirir, saygılarımı sunarım.

Öğrencinin Adı ve Soyadı: Eray SELÇUK

Tarih: 20/02/2023

İmza:

ÖZGEÇMİŞ

Kişisel Bilgiler

Adı ve Soyadı : Eray SELÇUK

E-Posta Adresi :

İş Deneyimi :

Unvanı	Görev Yeri	Yıl
Öğretmen	Millî Eğitim Bakanlığı (Hatay, Mersin)	2010-2015
Ölçme ve Değerlendirme Uzmanı	Millî Eğitim Bakanlığı (Ölçme, Değerlendirme ve Sınav Hiz. Gen. Müd.)	2015-Halen

Akademik Bilgiler

Öğrenim Durumu:

Derece	Bölüm/Program	Üniversite	Yıl
Lisans	Fen Bilgisi Öğretmenliği	Gazi Üniversitesi	2003-2007
Yüksek Lisans	Eğitimde Ölçme ve Değerlendirme	Mersin Üniversitesi	2008-2012
Doktora	Eğitimde Ölçme ve Değerlendirme	Ankara Üniversitesi	2016-2022

Yayınlar:

Selçuk, E. (2019). Öğretmenlerin Okul Müdürlerine Yönelik Likert Tipi Tutum Ölçeğinin Geliştirilmesi: Güvenirlik ve Geçerlik Çalışması. *Ihlara Eğitim Araştırmaları Dergisi, 4*(1), 112-132. Retrieved from http://ihead.aksaray.edu.tr/tr/pub/issue/42161/554516

Selçuk, E. & Tezbaşaran, A. A. (2022). PISA Fen Okuryazarlığı Yeterlik Düzeylerinin Ölçme ve Değerlendirme ve Fen Bilimleri Alan Uzmanları Açısından Değerlendirilmesi. *Anadolu Öğretmen Dergisi, 6*(1), 18-30. DOI: 10.35346/aod.1096600