HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı
Eğitimde Ölçme ve Değerlendirme Programı

GRUP İÇİ VARYANSLAR HETEROJEN OLDUĞUNDA ÇOK DÜZEYLİ MADDE TEPKİ MODELİNİN BAYES YAKLAŞIMI İLE MODELLENMESİ

Yusuf KARA

Doktora Tezi

Ankara, 2018

Liderlik, araştırma, inovasyon, kaliteli eğitim ve değişim ile

Daha ileriye... En İyiye...

HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı
Eğitimde Ölçme ve Değerlendirme Programı

GRUP İÇİ VARYANSLAR HETEROJEN OLDUĞUNDA ÇOK DÜZEYLİ MADDE TEPKİ MODELİNİN BAYES YAKLAŞIMI İLE MODELLENMESİ

BAYESIAN MODELING OF MULTILEVEL ITEM RESPONSE MODEL WITH HETEROGENEOUS WITHIN-GROUP VARIANCE

Yusuf KARA

Doktora Tezi

Ankara, 2018

Kabul ve Onay

Eğitim Bilimleri Enstitüsü Müdürlüğüne,

Yusuf KARA'nın hazırladığı "Grup İçi Varyanslar Heterojen Olduğunda Çok Düzeyli Madde Tepki Modelinin Bayes Yaklaşımı ile Modellenmesi" başlıklı bu çalışma jürimiz tarafından **Eğitim Bilimleri Ana Bilim Dalı, Eğitimde Ölçme ve Değerlendirme Bilim Dalında Doktora Tezi** olarak kabul edilmiştir.

Jüri Başkanı	Prof. Dr. Şener BÜYÜKÖZTÜRK
Jüri Üyesi (Danışman)	Prof. Dr. Hülya KELECİOĞLU
Jüri Üyesi	Prof. Dr. Selahattin GELBAL
Jüri Üyesi	Doç. Dr. Nuri DOĞAN
Jüri Üyesi	Yard. Doç. Dr. Murat AKYILDIZ

İkinci Tez Danışmanı Prof. Dr. Akihito KAMATA

Enstitü Yönetim Kurulunun
..../.../.... Tarihli ve ......
sayılı kararı.

Bu tez Hacettepe Üniversitesi Lisansüstü Eğitim, Öğretim ve Sınav Yönetmeliği'nin ilgili maddeleri uyarınca yukarıdaki jüri üyeleri tarafından 22 / 01 / 2018 tarihinde uygun görülmüş ve Enstitü Yönetim Kurulunca ..... / ..... / .... tarihinde kabul edilmiştir.

Prof. Dr. Ali Ekber ŞAHİN
Eğitim Bilimleri Enstitüsü Müdürü

Öz

Bireylerin okul, şehir vb. birimlerde kümelendiği hiyerarşik veri yapıları eğitim ve davranış bilimlerinde oldukça yaygındır. Geniş ölçekli sınavlarda hiyerarşik verilerle çalışmak kaçınılmazdır. Bu koşullarda geleneksel regresyon yerine hiyerarşik doğrusal modellerin kullanılması daha etkili olmaktadır. Ancak bu modeller, bağımlı değişkenin gözlenebilen olduğu durumlar için uygundur. Dolayısıyla, öğrenci başarısı gibi gizil bağımlı değişkenlerin varlığında çok düzeyli madde tepki modellerinin (ÇDMTM) kullanılması daha etkilidir. Geleneksel ÇDMTM, diğer hiyerarşik modellerde olduğu gibi grup içi varyansların homojen olduğunu varsaymaktadır. Ancak her bir okulun veya sınıfın eş yetenek varyansına sahip olması, karşılanması zor bir varsayımdır. Öte yandan, okulların eş varyansa sahip olmaması bir varsayım ihlalinden ziyade, okul etkililiği açısından önemli bir bilgi kaynağı olarak görülmelidir. Bu çalışmada ÇDMTM, her grup için ayrı bir varyans bileşeni kestirecek şekilde yeniden modellenmiştir. Modelin parametre kestirimleri Bayes yaklaşımıyla gerçekleştirilmiş, değişen grup içi heterojenlik koşullarındaki parametre kestirim performansı simülasyon çalışması ile incelenmiştir. Simülasyonun ardından, önerilen ve geleneksel model 2009 ÖBBS (Öğrenci Başarılarının Belirlenmesi Sınavı) Türkçe testi verilerine uygulanmıştır. Her iki modelin gerçek veriye uyumları ve parametre kestirimleri karşılaştırılmış, ayrıca okul içi varyanslar önerilen model ile koşullu olarak modellenmiştir. Simülasyon sonucunda, parametre kestirimlerinin en iyi orta heterojenlik düzeyinde gerçekleştirildiği; yüksek heterojenlik düzeyinde ise kabul edilebilir bir performansla kestirim yapılabildiği görülmüştür. Düşük heterojenlik düzeyinde özellikle uç değerlerdeki okul içi varyansların ve varyansların varyans parametresinin kestirimi yeterince etkili olmamıştır. Ancak diğer parametrelerin kestirimlerinde belirgin bir sorun gözlenmemiştir. Dolayısıyla, grup içi varyans kestirimlerinin birbirlerine oldukça yakın olması halinde geleneksel ÇDMTM'nin tercih edilebileceği önerilmiştir. Gerçek veri analizleri sonucunda, önerilen modelin veriye geleneksel ÇDMTM'den daha iyi uyum sağladığı görülmüştür. Koşullu varyans modelinin analizi sonucunda, sınavla öğrenci alan liselerdeki okul içi yetenek dağılımlarının, genel liselerdeki dağılımlardan daha düşük varyansa sahip olduğu sonucuna ulaşılmıştır. Ayrıca, sınavla alım yapan okulların düşük değişkenlik ile beraber daha

yüksek yetenek parametrelerine sahip oldukları, dolayısıyla okul etkililiği bakımından genel liselere göre daha iyi oldukları belirlenmiştir.

Anahtar sözcükler: hiyerarşik modeller, çok düzeyli madde tepki modelleri, grup içi varyans heterojenliği, Bayes yaklaşımı, okul etkililiği

Abstract

In educational and behavioral sciences, it is very common to work with hierarchical data structures where individuals are nested in upper-level units such as schools, neighborhoods etc. Hierarchical data are even inevitable in large-scale assessments. Employing hierarchal linear models would be more efficient in such multilevel data conditions. However, this model family is appropriate only if the dependent variable is manifest. Thus, it would be more efficient to employ multilevel item response models (MIRMs), if the relevant variable is latent like “ability levels of examinees”. Conventional MIRMs assume homogeneity of within-group variances, just as conventional multilevel models in general. Nonetheless, assuming a constant variance for ability distributions of different classes or schools does not sound realistic. Moreover, heterogeneity of within-group variances should be regarded as an important source of information for school effectiveness studies rather than a violation of a model assumption. In this study, three-level conventional MIRM was re-modeled in order to estimate unique within-group variances. Bayesian approach was adopted for estimation, and parameter recovery was evaluated under varying levels of within-group heterogeneity through a simulation study. The proposed model and its conventional counterpart then were applied to a real dataset that were obtained from 2009 Turkish test of the Student Achievement Determination Exam administered in Turkey. Model-data fits, as well as, parameter estimates were compared across models. Lastly, conditional extension of the proposed model was applied to the data in order to explain possible sources of within-group variance heterogeneity. Depending on results of the simulation study, it was concluded that the proposed model’s parameter estimation was mostly efficient under moderate level heterogeneity, while it was acceptable under high level. When heterogeneity was low, especially the highest within-school variance parameters and variance of the within-school variances were not recovered efficiently. Nonetheless, there was no significant problems in recovery of the other model parameters. In the light of the mentioned simulation results, it is recommended to employ the conventional homogeneous MIRM in case of low within-group heterogeneity. Analyses on real data showed that the proposed model had a better model fit than the conventional MIRM. Moreover, schools that require an entry exam for admission had lower within-school variances and higher school-level abilities compared to the other ones. Such

an inference should be regarded as a good sign of better school effectiveness for examination-required high schools.

Keywords: hierarchical models, multilevel item response models, within-group variance heterogeneity, Bayesian approach, school effectiveness

Teşekkür

Lisansüstü öğrenim serüvenimin ilk aşamasından, tezimin tamamlanmasına dek her zaman sunduğu destek ve motivasyon için değerli hocam ve tez danışmanım Prof. Dr. Hülya KELECİOĞLU'na,

Tezimin gelişimi için önemli katkılar sunan hocalarım Prof. Dr. Selahattin GELBAL, Prof. Dr. Şener BÜYÜKÖZTÜRK, Doç. Dr. Nuri DOĞAN ve Yard. Doç. Dr. Murat AKYILDIZ'a,

Doktora sonrası kariyer planlarımda bana destek olan hocalarım Doç. Dr. Alper Tolga KUMTEPE ve Doç. Dr. Burcu ATAR'a,

Bana güvenerek başarılı olacağıma inanan ve buna koşulsuz kefil olan değerli arkadaşlarım Dr. Güler YAVUZ, Dr. Sümeyra SOYSAL, Beyza HİMMETOĞLU, Seray TANYER ve Zafer SUSOY'a,

Tüm bu süreç boyunca bana her zaman cesaret ve güç veren biricik eşim Tuğçe KARA'ya,

Hayatımın her anında emekleri olan başta annem ve babam olmak üzere ailemin tüm fertlerine,

Doktora eğitimim süresince yurtiçi ve yurt dışında sağladığı burs imkânları için TÜBİTAK'a,

En önemlisi, bu çalışmanın ortaya çıkmasında ve gelişerek son haline gelmesinde en büyük payı olan eş danışmanım ve değerli hocam Prof. Dr. Akihito KAMATA'ya sonsuz teşekkürlerimi sunuyorum.

İçindekiler

Tablolar Dizini

Şekiller Dizini

Simgeler ve Kısaltmalar Dizini

1P-HGLLM: 1-Parameter Hierarchical Generalized Linear Logistic Model

ARB: Absolute Relative Bias

BGR: Brooks-Gelman-Rubin

ÇDMTM: Çok Düzeyli Madde Tepki Modelleri

DIC: Deviance Information Criterion

DMF: Değişen Madde Fonksiyonu

EMA: Ecological Momentary Assessment

ESS: Effective Sample Size

GLM: Generalized Linear Models

HDI: Highest Density Interval

HGLM: Hierarchical Generalized Linear Models

HLM: Hierarchical Linear Models

INVALSI: Italian National Institute for the Evaluation of the School System

JAGS: Just Another Gibbs Sampler

KTK: Klasik Test Kuramı

LLTM: Lineer Lojistik Test Model

LYS: Lisans Yerleştirme Sınavı

MC: Monte Carlo

MCMC: Markov Chain Monte Carlo

ML: Maximum Likelihood

MLSM: Mixed-Effects Location Scale Model

MTK: Madde Tepki Kuramı

ÖBBS: Öğrenci Başarılarının Belirlenmesi Sınavı

PISA: Programme for International Student Assessment

PQL: Penalized Quasi Likelihood

RMSE: Root Mean Squared Error

SE: Standard Error

TEOG: Temel eğitimden Ortaöğretime Geçiş Sınavı

TIMMS: The Trends in International Mathematics and Science Study

YGS: Yüksek Öğretime Geçiş Sınavı

Bölüm 1
Giriş

Bu bölümde sırasıyla araştırmanın problem durumuna, amacı ve önemine ve son olarak dayandığı kuramsal temele ilişkin alt bölümler sunulmuştur.

Problem Durumu

Eğitim bilimleri araştırmalarında kullanılan verilerin genelde hiyerarşik bir yapı sergilediği bilinmektedir. Hiyerarşik yapıdan kasıt, birey bazında toplanan verilerin değişik düzeylerde yuvalanmış, yani iç içe geçmiş olmasıdır. Örneğin il genelindeki öğrenci başarısı ile ilgili bir araştırma yapmak amacıyla toplanan veriler, belirli okulların belirli sınıflarından elde edilmiş olacaktır. Dolayısıyla, aynı sınıftaki öğrencilerden; daha geniş kapsamlı kümelenme düşünüldüğünde ise aynı okuldaki sınıflardan toplanan verilerin ilişkili olması muhtemeldir. Öğrencilerin belirli bir okula kabul edilme süreçleri düşünüldüğünde, aynı okuldaki öğrencilerin sosyoekonomik düzeylerinin birbirlerine yakın olması bu tür ilişkilere örnek olarak verilebilir (Hox, 2010). Hiyerarşik verilere YGS (Yüksek Öğretime Geçiş Sınavı), LYS (Lisans Yerleştirme Sınavı) ve TEOG (Temel Eğitimden Ortaöğretime Geçiş Sınavı) gibi ulusal geniş ölçekli sınavlarda ve PISA (Programme for International Student Assessment: Uluslararası Öğrenci Değerlendirme Programı), TIMMS (The Trends in International Mathematics and Science Study: Uluslararası Matematik ve Fen Eğilimleri Araştırması) gibi uluslararası değerlendirme araştırmalarında rastlamak kaçınılmazdır.

Hiyerarşik veri setlerinin geleneksel yöntemler ile analiz edilmesi beraberinde birtakım sınırlılıkları getirmektedir. Bu veri setleri ile ilgili en belirgin sorun, yapılması düşünülen geleneksel analizlerin temel varsayımlarının ihlal edilmesidir. Örneğin, eğitim bilimleri araştırmalarında yaygın olarak kullanılan çoklu doğrusal regresyon analizinin varsayımlarından biri olan gözlemlerin bağımsızlığı, hiyerarşik verilerde ihlal edilebilmektedir. Bir okulun farklı sınıflarından toplanan verilerin analizinde, aynı sınıftaki öğrencilerin başarı puanlarının birbirinden tamamen bağımsız olduğu varsayılmaktadır. Oysaki aynı sınıfta yer alan öğrenciler benzer eğitim koşullarına sahip oldukları için puanları birbirleri ile ilişkili olabilir. Böyle bir durumda sınıf içindeki gözlemler birbirinden bağımsız olmayacaktır. Gözlemlerin bağımsızlığı varsayımının ihlal edilmesi, regresyon katsayılarına ait standart hataların yanlı kestirilmesine, dolayısıyla analizin istatistiksel olarak güç kaybetmesine yol açmaktadır (Snijders & Bosker, 1999; Raudenbush & Bryk, 2002; Stevens, 2009; Atar, 2010; Hox, 2010).

Gözlemlerin bağımsız olmama durumu, araştırmacıları verinin hiyerarşik yapısını dikkate alan ve bunun yanında üst düzeydeki birimlere (ör: sınıf, okul) ait değişkenleri de analize dâhil edebilme esnekliğine sahip olan bir modelleme yaklaşımı geliştirmeye itmiştir. Bağımlı değişkenin sürekli olduğu durum için hiyerarşik doğrusal modeller (hierarchical linear models: HLM); sıralı veya kesikli olduğu durum için hiyerarşik genelleştirilmiş doğrusal modeller (hierarchical generalized linear models: HGLM) geliştirilmiştir (Goldstein, 1986, 2011; Longford, 1987; Bryk & Raudenbush, 1992). HLM ve HGLM ile sadece öğrenci düzeyindeki değişkenler değil; sınıf, okul, şehir gibi daha üst birimlerin özelliklerine ait değişkenler de analizlerde dikkate alınabilmektedir.

Hiyerarşik modellerin uygulama bulduğu alanlardan biri de Madde Tepki Kuramı (MTK) olmuştur. Adams, Wilson ve Wu (1997), MTK modellerini maddelerin bireyler içerisinde kümelendiği iki düzeyli hiyerarşik modeller şeklinde formüle ederek, modele birey bazında yordayıcı değişkenlerin alınmasına olanak sağlamışlardır. Madde-birey şeklindeki veri yapısına ek olarak bireylerin kümelendiği sınıf, okul ve şehir gibi üçüncü düzeydeki yapılanmayı da dikkate alabilen model ise ilk olarak Kamata (1998) tarafından geliştirilmiştir. Bu model, geleneksel Rasch modelin HGLM yaklaşımı ile formüle edilmesi ile ortaya çıkmış olup, Kamata bu formülasyonun matematiksel olarak Rasch modele eşit olduğunu göstermiştir. Bu model, anılan diğer modellerden farklı olarak üçüncü düzey yapılanmayı da dikkate alabilmekte ve dolayısıyla modele sınıf veya okul düzeyinde yordayıcı değişkenler alınmasına olanak sağlamaktadır.

Kamata (1998) önerdiği modeli bir parametreli genelleştirilmiş aşamalı doğrusal lojistik model (1-parameter hierarchical generalized linear logistic model: 1P-HGLLM) şeklinde isimlendirmiştir. Özellikle burada vurgulanması gereken önemli noktalardan biri, HLM ve HGLM ile MTK modellerinin bütünleştirildiği hiyerarşik modellerin, alanyazında çok farklı şekillerde isimlendirilmiş olmasıdır. Örneğin Adams, Wilson ve Wu (1997) geliştirdikleri modeli çok düzeyli madde tepki modeli (multilevel/hierarchical IRT model) biçiminde isimlendirmişlerdir. Bu isimlendirmeyi yaygın olarak çoğu araştırmacı kullanmıştır (ör: Fox & Glas, 2001;

Pastor, 2001, 2003; Natesan, 2007; Zheng, 2009; Atalay-Kabasakal, 2014). Yapılan bazı çalışmalarda ise (ör: Maier, 2000, 2001; Brune, 2011; Moyer, 2013) anılan modeller çok düzeyli ölçme modelleri (multilevel measurement models) şeklinde ele alınmıştır. Bu çalışmada daha yaygın olarak kullanılan çok düzeyli madde tepki modelleri (ÇDMTM) ifadesi tercih edilmiştir.

Maddelere verilen yanıtların bireylerde yuvalandığı ÇDMTM iki düzeyli olup, geleneksel MTK modellerine göre birçok avantaj sağlamaktadır. Örneğin Natesan (2007) ve Brune (2011) yaptıkları çalışmalar sonucunda etkili parametre kestirimleri için ÇDMTM'nin, geleneksel MTK modellerine göre daha küçük örneklemlere ihtiyaç duyduğunu belirtmişlerdir. Ancak ÇDMTM'nin en önemli avantajı, eş zamanlı olarak birey bazında yordayıcı değişkenlerin de ölçme modeline dâhil edilebilmesidir. Bu durum, yetenek (Adams, Wilson, & Wu, 1997) ve madde parametrelerinin (Mislevy, 1987) kestirim hassasiyetini modele sağlanan fazladan bilgi sayesinde olumlu yönde etkilemektedir. Öte yandan Kamata (1998) tarafından geliştirilen ÇDMTM ile değişen madde fonksiyonu (DMF) analizleri yapmak geleneksel DMF belirleme yöntemlerine göre birçok avantaj sağlamaktadır (Chu, 2002). Maddelere verilen yanıtların bireylerde; bireylerin de okullarda yuvalandığı yapıdaki ÇDMTM, daha önce anılan iki düzeyli modele ek olarak birey-grup formundaki kümelenmeyi de dikkate almakta, ayrıca grup düzeyindeki yordayıcı değişkenlerin modele alınabilmesine olanak sağlamaktadır. Böylece üç düzeyli bu model kullanılarak, birey düzeyindeki DMF'nin gruplar arasında değişim göstermesine izin verilen (Kamata & Binici, 2003; Binici, 2007) ve hem birey hem de grup düzeyi kaynaklı iki yönlü DMF analizleri (Patarapichayatham, Kamata, & Kanjanawasee, 2012) gibi daha karmaşık DMF modellerinin kurulması mümkün olmaktadır.

Yukarıda belirtilen avantajlarının yanında, üç düzeyli modelin en önemli özelliği, doğrudan maddelere verilen 1-0 formundaki yanıt örüntülerini kullanarak her iki düzeydeki (birey ve grup) yetenek parametrelerini ve madde güçlüklerini kestirmesi, ayrıca eşzamanlı olarak tıpkı HLM'de olduğu gibi her iki düzeyde yordayıcı değişkenlerin alınabildiği modeller oluşturabilmesidir. Bu eşzamanlı kestirim özelliği, araştırmacılar tarafından uygulana gelen "öncelikle geleneksel bir MTK modeli ile yetenek parametrelerinin kestirimi ve bu parametrelerin sürekli bir bağımlı değişken olarak alınıp HLM ile modellenmesi" şeklindeki iki aşamalı uygulamayı gereksiz kılmaktadır. Bu iki aşamalı uygulama her ne kadar gizil değişkendeki ölçme hatalarını dikkate almış gibi görünse de, MTK'daki yetenek parametrelerine ait hataların heterojen olmasından dolayı ikinci aşamadaki analizlerde (HLM analizi kısmında) sorunlara yol açmaktadır (Adams, Wilson, & Wu, 1997; Kamata, 1998; Fox, 2001, 2004; Pastor, 2001, 2003; Dowling, 2006; Schmitt, 2007). Bu durumda ÇDMTM'nin, eşzamanlı kestirim avantajıyla iki aşamalı bu uygulamaya göre daha güvenilir sonuçlar vereceği açıktır.

ÇDMTM'nin yukarıda anılan tüm avantajlarından faydalanabilmek için, her istatistiksel modelde olduğu gibi öncelikle modelin varsayımlarının karşılanması gerekmektedir. Daha önce de anıldığı gibi üç düzeyli model, tıpkı HLM gibi birey-grup şeklindeki kümelenmeyi dikkate almakta, dolayısıyla HLM'nin iyi bilinen bir varsayımı olan grup içi varyansların homojenliği koşulunu gerektirmektedir. Bu varsayım, HLM'de tam olarak grup içi artıkların homojen varyans göstermesi şeklinde ifade edilmektedir (Goldstein, 1986; 2011; Raudenbush & Bryk, 2002; Hox, 2010). HLM'de kimi koşullar altında bu varsayımın karşılanamaması durumunda bazı parametrelerin kestirimlerinde sorunlar olabileceği yapılan araştırmalarla (Kasim & Raudenbush 1998; Raudenbush & Bryk, 2002; Vallejo, Ato, Fernandez, & Livacic-Rojas, 2013) ortaya konulmuştur. Benzer biçimde, ÇDMTM'de de homojenlik varsayımının ihlal edilmesi halinde birtakım parametre ve standart hata kestirim sorunlarının baş göstermesi olasıdır.

Bu araştırmada madde-birey-grup formundaki ÇDMTM, grup içi varyansların homojenliği varsayımının karşılanmasına gerek duyulmadan, her bir grup için ayrı bir varyans bileşeni kestirecek şekilde yeniden modellenmiştir. Öncelikle önerilen modelin matematiksel eşitlikleri ve istatistiksel yapısı sunulmuştur. Modelin tüm parametre kestirimleri tamamen Bayes yaklaşımı (fully Bayesian approach) çerçevesinde yapılmıştır. Önerilen modelin değişen grup içi heterojenlik düzeylerindeki parametre kestirim performansını ortaya koymak amacıyla bir simülasyon çalışması yürütülmüştür. Yürütülen simülasyon çalışmasının ardından önerilen model ve geleneksel ÇDMTM, 2009 ÖBBS (Öğrenci Başarılarının Belirlenmesi Sınavı) Türkçe testi verilerine uygulanmıştır. Böylece önerilen modelin gerçek verideki parametre kestirimi incelenmiş ve her iki modelin kestirimleri ve model-veri uyumları karşılaştırılmıştır. Ayrıca okul içi varyansların koşullu olarak modellendiği heterojen model gerçek veriye uygulanarak okul etkililiği hakkında daha somut çıkarımlar elde edilmiştir.

Araştırmanın Amacı ve Önemi

Madde-birey-grup formundaki ÇDMTM'de grup içi artıklar aslında bireylerin okul ortalamalarından arındırılmış yetenek parametrelerine karşılık gelmektedir (Kamata, 1998). Bu durumda ÇDMTM için karşılanması gereken varsayım kavramsal olarak, her bir grup içerisindeki yetenek dağılımlarının eş varyansa sahip olması şeklindedir. Diğer bir ifadeyle, bu varsayım ile her bir okul içerisindeki yetenek parametrelerine ait varyansın birbirine eşit olduğu kabul edilmektedir. Ancak Raudenbush ve Bryk (2002) teorik olarak yapılan bu varsayımın eğitim bilimleri ile ilgili gerçek verilerin analizinde çoğu zaman sağlanamadığını vurgulamıştır.

Gerçek veri koşulları düşünüldüğünde, her bir sınıf veya okul içerisindeki bireylere ait yetenek dağılımlarının farklı varyans gösterebileceği muhtemel bir durumdur. Şöyle ki, bir okul içerisinde yer alan farklı sınıflardaki öğrenci başarısına ait dağılımlar, başarı düzeyi oldukça yüksek olan öğrencilerden oluşan bir sınıfta daha az varyans gösterirken, daha heterojen (değişen başarı düzeyine sahip öğrencilerden oluşan) bir sınıfta ise bu dağılımın daha yüksek bir varyansa sahip olması beklenir. Aynı durum bir il genelindeki farklı türdeki okullar için de düşünülebilir. Örneğin fen liseleri gibi başarı düzeyi yüksek olan okullardaki yetenek dağılımları ile öğrenci çeşitliliğinin daha fazla olduğu genel liselerdeki yetenek dağılımlarının eş varyansa sahip olduğu varsayımı pek gerçekçi olmayacaktır. Kim (2005) benzer biçimde farklı altyapılara ve öğrenci akademik geçmişine sahip okullarda varyansların homojen olmasının zor olduğunun altını çizmiştir. Öte yandan Nunnaly (1994), madde sayısı fazla olsa da test puanlarının nadir olarak normal dağıldığı gerçeğini vurgulamış; Raudenbush ve Bryk (2002) ise grup içi varyans homojenliğinin ihlal edilmesinde verilerin normalden daha basık bir dağılım göstermesi veya aşırı uç değerlerin olmasının da etkili olabileceğini belirtmiştir. Gerçek veri setlerinde bu gibi koşullarla karşılaşılması birçok araştırmacı tarafından sıkça yaşanan bir durumdur. Dolayısıyla, gerçek veri setlerinde normalliğin ciddi derecelerde ihlal edilmesi halinde ÇDMTM'de grup içi varyansların homojenliği varsayımının karşılanamaması olası bir durumdur.

İstatistiksel bir modelin en önemli özelliği, gerçek verilere uygulandığında modelin veriyi mümkün olduğunca iyi bir şekilde açıklayabilmesi, yani modelin veriye uyum sağlamasıdır. Yukarıda somut örneklerle de ifade edildiği üzere, geleneksel

ÇDMTM homojen grup içi varyans varsayımı ile bazı durumlarda gerçek veri setlerine uyum konusunda yeterli esnekliği sağlayamayabilir. Bu durum, model tarafından kestirilecek parametrelerin ve standart hataların etkili olamamasına neden olabilir. Öte yandan her bir grup için ayrı olarak kestirilecek varyans parametreleri sayesinde ortalama yeteneğin (grup düzeyi yetenek) yanında, grup içi yetenek değişkenliğinin de her bir grup için ayrı olarak değerlendirilmesi mümkün olacaktır. Bu bağlamda, araştırmanın üç temel amacı bulunmaktadır. Bu amaçlar sırasıyla:

1) Madde-birey-grup formundaki ÇDMTM'yi grup içi varyansların homojenliği varsayımına gerek kalmayacak şekilde genişleterek, her grup için farklı bir varyans bileşeni kestirecek yeni bir model önermek,
2) Önerilen modelin değişen heterojenlik koşullarındaki parametre kestirim performansını simülasyon çalışmasıyla belirlemek,
3) Önerilen modelin gerçek veriye olan uyumunu değerlendirerek uygulamada kullanıcılara sağlayacağı avantajları somut olarak irdelemektir.

ÇDMTM'nin geleneksel MTK modelleri ile HLM'ye; hatta bu modellerin yanıt örüntülerine (1-0 şeklindeki veriye) iki aşamalı şekilde uygulanmasına karşı sağladığı avantajlar, modelin bu tür hiyerarşik verilerde araştırmacılar açısından ne kadar kullanışlı olabileceğini yeterince ortaya koymaktadır. Hâlihazırda güçlü olan bu modelin, homojenlik varsayımına ihtiyaç duymayarak gerçek veri setlerine olan uyumunu artıracak bir biçimde genişletilmesinin, başta alanyazın olmak üzere modeli kullanacak diğer araştırmacılara da önemli katkılar sunacağı düşünülmektedir. Öte yandan, önerilen modelin farklı grup içi heterojenlik düzeyleriyle ne derece başa çıkabildiğinin ve gerçek veri analizindeki performansının ortaya konulması, araştırmacılara model hakkında bilgi vermesi açısından önem arz etmektedir.

Önerilen modelin geleneksel ÇDMTM'ye göre en önemli avantajı, çok düzeyli madde-yanıt örüntülerinde ortalamanın yanı sıra varyans bileşenlerinin değişkenliğinin de dikkate alınabilmesidir. Bağımlı değişkenin sürekli olduğu HLM'de varyans heterojenliğinin bir varsayım ihlalinden ziyade önemli bir bilgi kaynağı olarak değerlendirilebileceği alanyazında sıkça vurgulanmaktadır. Örneğin Kim ve Choi (2008) HLM kapsamında okulların sadece ortalama bağlamında değil; değişkenlik (varyans) bağlamında da incelenmesi gerektiğini ve okul başarısının, yüksek düzeydeki okul ortalamasıyla beraber daha düşük okul içi değişkenlik ile açılanabileceğini savunmaktadırlar. Öte yandan Kim (2005) HLM'de heterojen okul içi varyansların okul düzeyi değişkenlerle yordanabileceğini belirtmiş; örneğin öğretmen özellikleri ve okulun ortalama sosyo-ekonomik düzeyi gibi değişkenlerin bu amaçlarla kullanılabileceğini belirtmiştir. HLM için yapılan bu öneriler, bu çalışma kapsamında önerilen heterojen ÇDMTM için de geçerlidir. Dolayısıyla önerilen model kapsamında, geleneksel ÇDMTM'den farklı olarak, ortalamaların yanında okul içi yetenek değişkenliklerinin de yordanabilmesi koşulsuz (unconditional) heterojen modelde yapılacak genişletmeler ile mümkün olmaktadır. Bu durum okul başarılarının, yüksek ortalamayla birlikte "daha düşük okul içi değişkenlik" bağlamında ele alınmasında önerilen modelin oldukça kullanışlı sonuçlar sunacağını ortaya koymaktadır.

Araştırma Problemi

"Grup içi varyanslar heterojen olduğunda, her bir grup için ayrı bir varyans bileşeni kestirmeye olanak sağlayacak şekilde yeniden modellenen ÇDMTM'nin değişik heterojenlik düzeylerindeki ve gerçek veri analizindeki parametre kestirim performansı nasıldır?"

Alt problemler. Bu çalışma kapsamında homojen ÇDMTM'ye alternatif olarak önerilen modelin,

1) Parametre kestirim performansı, grup içi varyansların farklı düzeydeki heterojenliğinden nasıl etkilenmektedir?

a) Farklı heterojenlik düzeyine göre modelin yakınsama durumu nasıl değişmektedir?

b) Farklı heterojenlik düzeyine göre aşırı parametrelerin kestirimi nasıl değişmektedir?

c) Farklı heterojenlik düzeyine göre temel parametrelerin kestirimi nasıl değişmektedir?

2) Gerçek verilerdeki parametre kestirimleri ve model-veri uyumu nasıldır, ayrıca uygulamada sağladığı avantajlar nelerdir?

Sayıltılar

1) Önerilen modelde yeteneklerin gösterdiği dağılımlar gibi istatistiksel sayıltılara, araştırmanın yöntem bölümünde “Modelin Tanımlanması: Matematiksel Eşitlikler ve Varsayımlar” başlığı altında yer verilmiştir.

2) Araştırmanın simülasyon çalışması kısmına ait temel sayıltı, her bir koşul için üretilecek veri setlerinin, gerçek veri setlerini temsil ettiği şeklinde ifade edilebilir.

Sınırlılıklar

Yeniden modellenen ÇDMTM, üç düzeyli (madde-birey-grup) ve madde parametrelerinden yalnızca güçlüklerin kestirildiği 1-0 şeklindeki iki kategorili yanıtları ele alan model ile sınırlıdır. Simülasyon çalışmasında ele alınan koşullar ise grup içi varyans bileşenlerinin heterojenlik düzeyleri ile sınırlandırılmıştır.

Bölüm 2

Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar

Bu bölümde geleneksel MTK modelleri ve ÇDMTM hakkında temel kuramsal bilgilere yer verilecektir. Özellikle ÇDMTM ile geleneksel MTK modellerinin benzeştiği ve farklılaştığı noktalara değinilerek, araştırmanın temel amacı kapsamında genişletilen madde-birey-grup formundaki geleneksel ÇDMTM detaylı olarak ele alınacaktır. Bu bölümde ayrıca, önerilen modelin parametrelerinin kestirilmesinde kullanılan Bayes yaklaşımı hakkında detaylı bilgilere yer verilecektir. Kuramsal temele ilişkin içeriğin ardından, ilgili araştırmalar sunulacaktır.

Geleneksel MTK Modelleri

Bilindiği gibi MTK, Klasik Test Kuramı'nın (KTK) ardından ölçme ve değerlendirme ile psikometri alanlarında oldukça benimsenmiş ve kuram çerçevesinde önerilen modeller günümüze değin geçerliklerini büyük oranda korumaya devam etmiştir. MTK'ya dayalı ölçme modellerinin temelinde, bireylere ait gözlenemeyen yeteneklerin, bu yetenekleri ölçtüğü varsayılan maddeler aracılığıyla kestirilmesi fikri yatmaktadır. Yetenek parametrelerine ek olarak maddelere ait parametrelerin de eşzamanlı kestirimi söz konusudur. Maddelere verilen yanıtların 1-0 şeklinde iki kategorili olduğu durumlar için kullanılabilecek en yaygın MTK modelleri, Rasch Model ile bir, iki ve üç parametreli modeller olarak adlandırılmaktadır. Bir parametreli model Rasch Model ile benzerlik göstermekte olup, bu iki model ayırıcılık parametrelerinin ele alınması bağlamında farklılaşmaktadır. Bir parametreli modelde Rasch Model'den farklı olarak tüm maddeler için tek bir ayırıcılık parametresi kestirilmektedir. Diğer bir ifadeyle, tüm madde ayırıcılıklarının eşit olduğu varsayılmaktadır. Rasch modelde de tüm madde ayırıcılıklarının eşit olduğu varsayılır; ancak bu değer kestirilmez ve bu değerin 1 olduğu kabul edilir.

Rasch model. Rasch modelde bireyler için yetenek; maddeler için ise yalnızca güçlük parametresi kestirilir. Modelin matematiksel eşitliği,

$$P\left(X_{ij} = 1 \middle| \theta_j\right) = \frac{e^{(\theta_j - b_i)}}{1 + e^{(\theta_j - b_i)}} \tag{1}$$

şeklindedir. Bu eşitliğin daha sade hali alanyazında şu biçimde de kullanılmaktadır:

$$P\left(X_{ij} = 1 \middle| \theta_j\right) = \frac{1}{1 + e^{-(\theta_j - b_i)}} \tag{2}$$

(1) ve (2) numaralı eşitliklerde yer alan $P\left(X_{ij} = 1 \middle| \theta_j\right)$ terimi, θ yetenek düzeyine sahip j bireyinin, güçlük düzeyi b olan i maddesini doğru yanıtlama olasılığını ifade etmektedir. Bu eşitlik, yukarıda ifade edilen olasılık değerine logit bağlantı fonksiyonu uygulandığında ise şu hali alır:

$$\ln\left(\frac{P\left(X_{ij} = 1 \middle| \theta_j\right)}{1 - P\left(X_{ij} = 1 \middle| \theta_j\right)}\right) = \theta_j - b_i \tag{3}$$

(1) ve (2) numaralı eşitliklerin (3) numaralı eşitlikteki gibi ifade edilmesiyle, yetenek ve güçlük parametrelerinin aynı ölçek üzerinde olduğu daha açık şekilde görülmektedir. Buna ek olarak bu eşitlik, Rasch modelin logit bağlantı fonksiyonu kullanılarak doğrusal bir model şeklinde ifade edilebileceğini göstermektedir.

İki parametreli model. İki parametreli MTK modelinde, madde güçlük parametrelerine ek olarak madde ayırıcılık parametreleri de kestirilir. Böylece, Rasch modeldeki ayırıcılıkların eşit olması sınırlılığı da kaldırılmış olur. İki parametreli modelin eşitliği ise şu şekildedir:

$$P\left(X_{ij} = 1 \middle| \theta_j\right) = \frac{e^{a_i(\theta_j - b_i)}}{1 + e^{a_i(\theta_j - b_i)}} \tag{4}$$

Yukarıda görüldüğü gibi Rasch model eşitliğine ek olarak, (4) numaralı eşitlikte a_i ile sembolize edilen madde ayrıcılık parametrelerinin de yer aldığı görülmektedir.

Üç parametreli model. Çoktan seçmeli test maddelerinin yanıtlanmasında bireylerin yetenekleri dışında tahmin yoluyla maddelere doğru yanıt verme olasılıkları da söz konusudur. Bir ve iki parametreli modellerde şans başarısı kontrol edilmezken, üç parametreli modelde ise bu durum üçüncü bir madde parametresinin (şans parametresi) iki parametreli modele eklenmesi ile dikkate alınır. Üç parametreli modelin matematiksel eşitliği şu şekildedir:

$$P\left(X_{ij} = 1 \middle| \theta_j\right) = c_i + (1 - c_i)\frac{e^{a_i(\theta_j - b_i)}}{1 + e^{a_i(\theta_j - b_i)}} \tag{5}$$

Eşitlikte yer alan c_i parametresi, ilgili maddenin şans parametresini, yani madde bilgi fonksiyonunun düşey asimptotunu ifade etmektedir.

Çok Düzeyli Madde Tepki Modelleri (ÇDMTM)

ÇDMTM, maddelere ait yanıt örüntülerinin bireylerin içerisinde kümelendikleri yaklaşımını benimsemektedir. İki düzeyli ve koşulsuz olan, yani sadece madde-birey şeklindeki kümelenmeyi dikkate alan ve birey düzeyinde yordayıcı değişken içermeyen model aslında geleneksel MTK modellerine eşdeğer olmaktadır. Bu eşdeğerliğe rağmen, ÇDMTM'nin geleneksel MTK modellerine göre avantajlı olduğu noktalar vardır. Bunlardan ilki, ÇDMTM'nin geleneksel MTK modellerinin aksine yetenek parametrelerini seçkisiz (random) olarak ele alması ve dolayısıyla toplamda daha az parametre kestirmesidir. Bu durum aynı zamanda geleneksel MTK modellerinde çok bilinen bir problem olan Neyman-Scott sorununu, yani madde ve yetenek parametrelerinin eşzamanlı kestirim zorluğunu önlemektedir. Öte yandan ÇDMTM ile uygun hiyerarşik kestirim algoritmaları (ör: PQL: Penalized Quasi Likelihood) kullanıldığında, yine geleneksel MTK modellerinin başa çıkamadığı tamamen doğru veya yanlış yanıt durumlarında bile kestirimlerin sorunsuz yapılabilmesi mümkün olmaktadır (Kamata, 1998).

Madde-birey formundaki koşulsuz ÇDMTM'ye birey düzeyinde yordayıcı değişkenler alındığında, geleneksel MTK modellerinden farklı olarak daha gelişmiş ölçme modelleri elde edilebilmektedir. Koşullu modeller olarak nitelendirilen bu modeller ile DMF analizleri daha kolay ve anlaşılır bir biçimde yapılabilmektedir. Madde-birey-grup formundaki ÇDMTM'de ise madde-birey yapısına ek olarak bireylerin kümelendiği sınıf veya okul gibi üst birimler de dikkate alınır. Burada vurgulanması gereken noktalardan biri de, üç düzeyli bu modelin aslında HLM bağlamında iki düzeyli bir model olduğudur. Üç düzeyli modele birey düzeyi değişkenlere ek olarak grup düzeyinde yordayıcı değişkenlerin alınması da mümkün olmaktadır. Bu araştırmada sadece koşulsuz modeller ele alındığından, aşağıda yer alan bölümlerde sırasıyla koşulsuz madde-birey ve madde-birey-grup formundaki ÇDMTM'ye ilişkin kuramsal bilgilere yer verilecektir.

Madde-birey formundaki ÇDMTM. İki düzeyli bu modelin gelişimi, Fischer (1973) tarafından önerilen Lineer Lojistik Test Model'e (LLTM) dayanmaktadır. LLTM'de tıpkı normal regresyon modellerinde olduğu gibi, geleneksel bir MTK modeline birey düzeyi yordayıcı değişkenlerin alınması şeklinde bir yaklaşım benimsenmiştir (Akt. Kamata, 1998). Bireylere ait yordayıcı değişkenlerin geleneksel MTK modellerine dâhil edildiği diğer çalışmalar ise Mislevy (1987) ve

Zwinderman (1991) tarafından yapılmıştır. Benzer şekilde Mellenbergh (1994) MTK modellerini genelleştirilmiş doğrusal modeller (GLM: generalized linear models) bağlamında incelemiştir; ancak geleneksel MTK modelleri ilk olarak Adams, Wilson ve Wu (1997) tarafından hiyerarşik modeller olarak ele alınmıştır. Buradaki hiyerarşik kavramı ile vurgulanmak istenen durum, önerilen modellerin madde-birey şeklinde iki düzeyli olarak ele alınmasıdır. Bu araştırmalardan sonra Kamata (1998) geleneksel Rasch modelini HGLM yaklaşımı çerçevesinde bağımlı değişkenin kategorik olduğu iki düzeyli doğrusal bir regresyon şeklinde modellemiştir. Kamata ayrıca iki düzeyli olarak nitelendirilen bu modele ait HGLM temelli parametre kestirimlerinin matematiksel olarak geleneksel Rasch modeldeki madde ve yetenek kestirimlerine eşit olduğunu göstermiştir.

Kamata (1998) tarafından geliştirilen iki düzeyli koşulsuz modelde birinci düzey madde düzeyi iken, ikinci düzey birey düzeyi olarak ele alınır. Modelin ilk adımında (3) numaralı eşitlikteki geleneksel Rasch modelde olduğu gibi, bireylerin bir maddeyi doğru cevaplama olasılıkları olan $P(X_{ij} = 1|\theta_j)$ terimi kısaca p_{ij} olarak gösterilip bu değerlere logit dönüşümü (6) numaralı eşitlikte gösterildiği şekilde uygulanır.

$$\ln\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \eta_{ij} \tag{6}$$

Logit dönüşümü ile beraber, doğrusal model için p_{ij} olasılıkları yerine η_{ij} değerleri bağımlı değişken olarak kullanılır. Burada logit dönüşümünün kullanılmasındaki amaç, sadece 0-1 aralığında değer alabilen olasılıkların, $-\infty$ ile $+\infty$ arasında değer alabilecek şekilde yeni bir değişkene dönüştürülmesidir. Böylece, yukarıda yer alan Rasch modelin (3) numaralı eşitliği gibi doğrusal bir model formuna geçiş yapılır. Bu dönüşümün ardından madde düzeyi eşitlikleri şu şekilde ifade edilmiştir:

$$\begin{aligned} \eta_{ij} &= \beta_{0j} + \beta_{1j}X_{1ij} + \beta_{2j}X_{2ij} + \cdots + \beta_{(k-1)j}X_{(k-1)ij} \\ &= \beta_{0j} + \sum_{q=1}^{k-1} \beta_{qj}X_{qij} \end{aligned} \tag{7}$$

(7) numaralı eşitliklerde j indisi bireyleri $(j = 1,2,\dots,n)$; i indisi ise maddeleri $(i = 1,2,\dots,k)$ temsil etmektedir. β_{0j}, j bireyi için sabit (intercept) terimini, X_{qij} ise β_{qj} terimi ile ilişkili olan yapay değişkeni göstermektedir. X_{qij} yapay değişkeni, $q = i$

olduğunda −1; $q \neq i$ olduğunda ise 0 değerini alır (Kamata, 1998). Bu yapay değişken kodlaması aslında geleneksel regresyon modellerinde kategorik bir yordayıcı için yapılan yapay kodlama ile aynı mantıktadır. ÇDMTM'de toplam k maddeden biri referans madde olarak alınır ve böylece madde güçlüklerinin değerleri bu referans maddeye göre tayin edilir (Kamata, 1998). Öte yandan β_{qj} terimi ise bu yapay değişkene ait olan katsayı olup q maddesinin j bireyi üzerindeki etkisi ile ilişkilidir (Brune, 2011; Moyer, 2013). Referans olarak alınan k. madde için $\beta_{qj} = \beta_{kj} = 0$ olarak alınır. Sonuç olarak yapay kodlama ile beraber eşitliklerin sadeleştirilmiş hali Kamata (1998) tarafından

$$\ln\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \eta_{ij} = \beta_{0j} - \beta_{qj} \tag{8}$$

şeklinde ifade edilmiştir. Sonraki yıllarda Kamata (2001) $q = i$ olduğunda −1 yerine 1 şeklindeki kodlamanın yapılmasını önermiş ve bu eşitliği

$$\ln\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \eta_{ij} = \beta_{0j} + \beta_{qj} \tag{9}$$

şeklinde güncellemiştir. Birey düzeyi eşitlikler ise, yukarıda verilen madde düzeyi eşitliklerinin sağ tarafındaki terimlerin bağımlı değişken olarak ele alınması ile ifade edilmektedir. Birey düzeyi model eşitlikleri şu şekildedir:

$$\begin{cases} \beta_{0j} = \gamma_{00} + u_{0j} \\ \beta_{1j} = \gamma_{10} \\ \quad . \\ \quad . \\ \quad . \\ \beta_{(k-1)j} = \gamma_{(k-1)0} \end{cases} \tag{10}$$

Eşitlikteki u_{0j} terimi, β_{0j} teriminin artık değerlerini ifade eder ve bu artık değerlerin $u_{0j} \sim N(0, \tau)$ şeklinde ortalaması 0 ve varyansı τ olan normal bir dağılım gösterdiği varsayılır (Kamata, 1998). Bu artık değerler aslında bireylere ait yetenek kestirimleridir ve iki düzeyli modelde, yetenek parametrelerinin sabit değerli (fixed) birer parametre olarak kestirilmesinden ziyade normal dağılan seçkisiz parametreler şeklinde kestirildiği açıkça görülmektedir. Maddelere ait güçlük parametreleri her bir madde için $-\gamma_{q0} - \gamma_{00}$ terimine eşit olmaktadır. Burada dikkat edilmesi gereken nokta, referans olarak alınan k. madde için yukarıda belirtilen $\beta_{qj} = \beta_{kj} = 0$

eşitliğinden dolayı $-\gamma_{q0}$ teriminin de sıfıra eşit olacağı; dolayısıyla bu maddeye ait güçlük parametresinin $-\gamma_{00}$ değerine eşit olacağıdır.

Özetlemek gerekirse, HGLM yaklaşımına göre modellenen madde-birey formundaki ÇDMTM'de madde güçlükleri modelin sabit parametre kestirimlerine eşit olurken, yetenek kestirimleri ise modelin artıklarına, yani seçkisiz parametrelerine eşit olmaktadır. Bu matematiksel eşitlik Kamata (2001) tarafından aşağıda görüldüğü gibi daha açık biçimde gösterilmiştir.

$$p_{ij} = \frac{1}{1 + e^{\{-[u_{0j} - (-\gamma_{q0} - \gamma_{00})]\}}} \tag{11}$$

Bu eşitlik, aynı formda olan (2) numaralı geleneksel Rasch model eşitliği ile karşılaştırıldığında, $\theta_j = u_{0j}$ ve $b_i = (-\gamma_{q0} - \gamma_{00})$ olduğu açık bir biçimde görülebilir.

Madde-birey-grup formundaki ÇDMTM. Üç düzeyli bu model HGLM yaklaşımı çerçevesinde yine Kamata (1998) tarafından geliştirilmiş olup, iki düzeyli modele ek olarak modelde birey-grup şeklindeki kümelenme de dikkate alınır. Üç düzeyli modelin madde düzeyi eşitlikleri iki düzeyli model ile aynı olup, tek farklılık (12) numaralı eşitliklerde görüldüğü gibi grup düzeyini ifade eden *m* indislerinin kullanılmasıdır.

$$\begin{gathered} \ln\left(\frac{p_{ijm}}{1 - p_{ijm}}\right) = \eta_{ijm} \\ \eta_{ijm} = \beta_{0jm} + \beta_{1jm}X_{1ijm} + \beta_{2jm}X_{2ijm} + \cdots + \beta_{(k-1)jm}X_{(k-1)ijm} \end{gathered} \tag{12}$$

(12) numaralı eşitlikteki p_{ijm} terimi m okulunda $(m = 1,2,\ldots,r)$ yer alan j bireyinin i maddesine doğru yanıt verme olasılığını ifade etmektedir. (Yorumlama kolaylığı açısından, bundan sonraki bölümlerde grup düzeyi, "okul" olarak ifade edilecektir). X_{qjm} terimi iki düzeyli modelde olduğu gibi yapay değişkenleri, β_{0jm} terimi referans maddenin etkisini ve son olarak β_{qjm} terimi ise q. maddenin referans maddeye göre belirlenen etkisini göstermektedir (Kamata, 2001). Üç düzeyli modelin ikinci düzeyinde ise birey düzeyi eşitlikler yer alır.

$$\left\{ \begin{aligned} \beta_{0jm} &= \gamma_{00m} + u_{0jm} \\ \beta_{1jm} &= \gamma_{10m} \\ &\;\;. \\ &\;\;. \\ &\;\;. \\ \beta_{(k-1)jm} &= \gamma_{(k-1)0m} \end{aligned} \right. \tag{13}$$

(13) numaralı eşitliklerde yer alan tüm terimler iki düzeyli model eşitliği ile aynı olup, tek farklılık okul göstergesi olan m indislerine yer verilmesidir. Üç düzeyli modelde β_{0jm} teriminin artık değerleri için $u_{0jm} \sim N(r_{00m}, \tau_\gamma)$ şeklinde bir dağılım varsayımı yapılmaktadır. u_{0jm} değerleri, m okulunda yer alan j bireyinin yetenek parametresinin okul ortalaması olan r_{00m} değerinden sapmasını ifade etmektedir. Dolayısıyla her bir okuldaki bireylere ait yetenek parametrelerinin, o okula ait ortalama (r_{00m}) ile normal dağıldığı varsayılır. Burada dikkat edilmesi gereken nokta τ_γ ile gösterilen bu dağılıma ait varyansın tüm okullar için aynı olduğudur. Diğer bir ifadeyle, her bir okul içerisindeki yetenek dağılımı varyansının tüm okullarda eşit olduğu varsayımı ile tek bir grup içi varyans parametresinin kestirilmesi söz konusudur.

Modelin son düzeyinde, birey düzeyindeki eşitliklerin bağımsız değişkenleri bu sefer bağımlı değişkenler olarak ele alınır. Okul düzeyi eşitlikleri şu şekildedir:

$$\begin{cases} \gamma_{00m} = \pi_{000} + r_{00m} \\ \gamma_{10m} = \pi_{100} \\ \quad\vdots \\ \gamma_{(k-1)0m} = \pi_{(k-1)00} \end{cases} \tag{14}$$

(14) numaralı eşitlikler de aslında (10) numaralı iki düzeyli modelin ikinci düzey eşitliklerine benzemektedir. π_{000} terimi, γ_{00m} teriminin sabiti iken; r_{00m} terimi ise artıkları ifade etmektedir. Artık değerlerinin $r_{00m} \sim N(0, \tau_\pi)$ şeklinde ifade edilen sıfır ortalama ve τ_π varyans ile normal dağılım gösterdikleri varsayılmaktadır. (13) numaralı ikinci düzey eşitliklerindeki $u_{0jm} \sim N(r_{00m}, \tau_\gamma)$ varsayımından hatırlanacağı üzere, r_{00m} değerleri her bir okul içerisindeki bireylerin yetenek ortalamalarını ifade etmektedir. Dolayısıyla üçüncü düzeydeki $r_{00m} \sim N(0, \tau_\pi)$ varsayımı ile okul ortalamalarının genel ortalamasının sıfır olduğu ve varyansının ise τ_π olup, gruplar arası varyans bileşeni olarak model tarafından kestirileceği açıktır. Üç düzeyli model için tüm düzey eşitlikleri bir araya getirildiğinde

$$p_{ijm} = \frac{1}{1 + e^{\{-[(r_{00m} + u_{0jm}) - (-\pi_{q00} - \pi_{000})]\}}} \tag{15}$$

eşitliği elde edilmektedir (Kamata, 2001). Bu eşitlik, (2) numaralı geleneksel Rasch model eşitliği ile karşılaştırıldığında $\theta_j = (r_{00m} + u_{0jm})$ ve $b_i = (-\pi_{q00} - \pi_{000})$ olduğu görülmektedir. Bu durumda referans olarak alınan k. maddenin güçlüğü

$-\pi_{000}$ iken, geriye kalan $q = 1,2,\ldots,(k-1)$ maddenin güçlükleri ise $(-\pi_{q00} - \pi_{000})$'e eşit olmaktadır. Öte yandan j bireyine ait yetenek parametresi ise $(r_{00m} + u_{0jm})$'e eşit olmaktadır.

$\theta_j = (r_{00m} + u_{0jm})$ eşitliğinde görüldüğü gibi, üç düzeyli modelde iki düzeyli modelden farklı olarak bireylere ait yetenek parametresi, bireyin bulunduğu okulun ortalama yeteneği ile bu bireyin okul ortalamasından sapmasının toplamı şeklinde elde edilmektedir. Böylece üç düzeyli model, birey-okul şeklindeki kümelenmeyi dikkate alarak hem birey hem de okul düzeyi için ayrı yetenek parametreleri kestirebilmektedir. Özellikle geniş ölçekli değerlendirme çalışmalarında, okul başarılarının yorumlanması büyük önem kazandığından, üç düzeyli modelin kestirdiği okul yetenek parametrelerinin büyük bir işlevselliğe sahip olacağı açıktır.

Kamata (1998) üç düzeyli modeli maddelerin bireylerde ve bireylerin de gruplarda kümelendiği hiyerarşik bir model olarak ele almış ve modellemede HGLM yaklaşımını benimseyerek parametreleri en çok olabilirlik (Maximum Likelihood: ML) temelli yöntemleri kullanan HLM programı (Raudenbush, Bryk, Cheong, & Congdon, 2005) ile kestirmiştir. Fox (2001) ise üç düzeyli modeli ölçme modeli ve bireylerin yeteneklerinin modellendiği iki düzeyli (birey-grup) yapısal modelin birleşimi olarak ele almıştır. Bu yaklaşımda ÇDMTM'nin ölçme modeli, Y_{ijk} j okulundaki i bireyinin k maddesine verdiği 0-1 şeklinde yanıt olmak üzere,

$$P\left(Y_{ijk} = 1 \middle| \theta_{ij}, a_k, b_k\right) = \phi(a_k\theta_{ij} - b_k) \qquad (16)$$

eşitliği ile ifade edilmektedir. Burada $P\left(Y_{ijk} = 1 \middle| \theta_{ij}, a_k, b_k\right)$ terimi, j okulundaki i bireyinin k maddesini doğru yanıtlama olasılığını, ϕ kümülatif standart normal dağılım fonksiyonunu, θ_{ij} j okulundaki i bireyinin yetenek parametresini, a_k ve b_k sırasıyla k maddesinin ayırıcılık ve güçlük parametrelerini temsil etmektedir. Eşitlikte görüldüğü gibi Fox (2001), ölçme modelinde lojistik yerine normal ogive bağlantı fonksiyonunu kullanmaktadır. Ölçme modelinin ardından koşulsuz ÇDMTM'nin çok düzeyli yapısal modeline ait eşitlikler,

$$\theta_{ij} = \beta_{0j} + e_{ij} \qquad (17)$$

$$\beta_{0j} = \gamma_{00} + u_{0j} \qquad (18)$$

şeklinde ifade edilmektedir. Birinci düzey yapısal model eşitliğinde yer alan β_{0j} j okuluna ait sabiti terimi; e_{ij} ise okul içi hata terimini ifade etmektedir. İkinci düzey eşitlikte β_{0j} sabit terimi okulların genel ortalamasını gösteren γ_{00} ve j okulunun genel ortalamadan sapmasını gösteren u_{0j} terimlerine ayrıştırılmaktadır. Birinci ve ikinci düzey eşitliklerde yer alan artık terimlerin sıfır ortalama ve σ_{θ}^{2} ve τ_{00}^{2} varyans değerleri ile normal dağılım gösterdikleri varsayılmaktadır. Bu varyanslar aslında sırasıyla okul içi ve okullar arası varyans bileşenleri olup, Kamata (1998) tarafından modellenen ÇDMTM'deki τ_{γ} ve τ_{π} terimlerine karşılık gelmektedir. Özellikle çok düzeyli yapısal model eşitlikleri dikkatli incelendiğinde, bunların aslında iki düzeyli koşuluz HLM eşitlikleriyle tamamen aynı olduğu fark edilecektir. Öte yandan Fox (2001), geliştirdiği ÇDMTM'de Bayes yaklaşımına dayalı parametre kestirim yöntemlerini kullanmaktadır.

Her iki yaklaşımla modellenen ÇDMTM, birey ve okul düzeyindeki yetenekleri normal dağılan seçkisiz parametreler olarak kestirdiğinden, birey düzeyinde (grup içi) ve okul düzeyinde (gruplar arası) olmak üzere iki varyans parametresi kestirmektedir. Bu araştırma kapsamında yeniden modellenen ÇDMTM ile grup içi varyansların homojenliği varsayımı ile tek bir varyans bileşeni yerine, her bir grup için ayrı grup içi varyans bileşeni kestirilmesi mümkün olacaktır. Bu amaç doğrultusunda parametre kestirimleri için önemli avantajlar sağlaması bakımından Bayes yaklaşımı benimsenmiştir. Aşağıda yer alan bölümlerde Bayes yaklaşımının kuramsal temelleri ve Bayes yaklaşımı ile parametre kestirimi hakkında temel bilgilere yer verilmiştir. Önerilen modelin matematiksel eşitlikleri, sayıltıları ve Bayes yaklaşımına dayalı parametre kestirim sürecine ilişkin bilgiler araştırmanın yöntem bölümünde ilgili başlıklar altında detaylı olarak ele alınmıştır.

Bayes Yaklaşımı ile Parametre Kestirimi

Bayes yaklaşımını frekansa dayalı yaklaşım (frequentist approach) olarak da anılan geleneksel parametre kestirim yaklaşımından ayıran temel nokta, model parametrelerine olan bakış açısıdır. Frekansa dayalı yaklaşımla yapılan kestirimlerde parametrelerin sabit birer değer olduğu varsayılır ve istatistiksel modelin eldeki veri setine uygulanması ile bu sabit değerlerin kestirilmesi söz konusudur. Örneğin yaygın olarak kullanılan frekansa dayalı kestirim yöntemlerinden biri olan ML yönteminde bir parametreye ait kestirim, modelin olabilirlik fonksiyonunu maksimize eden değerdir. Bayes yaklaşımında ise model parametreleri sabit değil, aksine birer seçkisiz değişken olarak görülürler. Bir değişkenin seçkisiz olması, bu değişkene ait birden çok değerin olacağını ve bu değişkenin olasılıksal bir dağılım göstereceğini belirtir (Maier, 2000). Geleneksel yaklaşımdan farklı olarak parametrelerin olasılıksal dağılımlara sahip olma özelliği ile veri setinden gelecek temel bilgi dışında parametre hakkında fazladan bir ön bilginin de analize dâhil edilebilmesi mümkün olmaktadır. Analiz sürecinde veri seti dışında parametre hakkında ön bir bilgi sağlanabilmesi, Bayes yaklaşımını geleneksel yaklaşımdan ayıran en belirgin uygulama olmaktadır.

Bayes yaklaşımının geleneksel yaklaşıma göre sahip olduğu temel farklılıklar, başta parametre ve hata kestirimleri olmak üzere hipotez ve model uyumluluk testlerinde de önemli avantajlar sağlamaktadır. Öte yandan geleneksel yaklaşımlarla parametreleri kestirilemeyen karmaşık modeller, Bayes yaklaşımı ile kolayca analiz edilip tüm model parametreleri etkili bir biçimde kestirilebilmektedir. Bunlara ek olarak Bayes yaklaşımı ile parametre kestirimlerinde kayıp veriler geleneksel yaklaşımdaki yöntemlere göre oldukça kolay biçimde kontrol edilebilmekte ve çok sıkı parametrik varsayımların karşılanmasına gerek kalmamaktadır. Bayes yaklaşımının sunduğu tüm bu avantajlar özellikle son yıllarda birçok araştırmacı tarafından yaygın olarak kullanılmaktadır.

Bayes yaklaşımı ile parametre kestiriminin didaktik açıdan da geleneksel yaklaşıma göre daha avantajlı olduğunu söylemek yanlış olmaz. Şöyle ki, Bayes yaklaşımı ile kestirimde araştırmacı verinin istatistiksel yapısını ve bu veriye uygulamak istediği modeli iyi tanımalı ve modelin matematiksel eşitliklerine hâkim olmalıdır. Bayes yaklaşımı ile yürütülen analizler sonucunda parametrelere ilişkin elde edilen çıktıların yorumlanması, araştırmacının genel istatistik bilgisinin iyi olmasını gerektirmekle beraber bu çıktıların sunduğu bilgiler ile oldukça öğretici olmaktadır. Dolayısıyla araştırmacı, alışılagelen geleneksel veri analizlerinde olduğu gibi sadece parametre kestirimlerine ulaşmaz, bunun yanında uygulamaya çalıştığı modelin dayandığı istatistiksel teorinin de farkına analiz sürecinde daha iyi varır. Aşağıda yer alan bölümlerde Bayes yaklaşımı ile parametre kestirim sürecine ilişkin bilgiler gerekli teorik altyapı ile beraber sunulmuştur.

Bayes Teoremi ve istatistiksel modellere uygulanması. Bayes yaklaşımı ile parametre kestiriminin temeli Bayes Teoremi'ne dayanmaktadır. Bayes Teoremi temel olarak koşullu olasılıkların formüle edilmesi şeklinde nitelendirilebilir. A ve B iki olay olmak üzere, Bayes Teoremi eşitliği

$$p(A|B) = \frac{p(B|A)\,p(A)}{p(B)} \propto p(B|A)p(A) \tag{19}$$

şeklinde formüle edilir. $p(A)$ ve $p(B)$ terimleri, sırasıyla A ve B olaylarının marjinal (bağımsız olarak) gerçekleşme olasılıklarını ifade etmektedir. $p(A|B)$ terimi ise B olayının gerçekleştiğinin bilinmesi halinde, A olayının gerçekleşme olasılığını belirtir. Diğer bir ifadeyle $p(A|B)$ terimi, A olayının B olayına bağlı koşullu olasılığıdır. Benzer şekilde $p(B|A)$ terimi ise B olayının A olayına bağlı koşullu olasılığını belirtir. $\propto$ sembolü, $p(A|B)$ koşullu olasılığının oransal olarak $p(B|A)\,p(A)$ çarpımına eşit olduğunu ifade etmektedir. Sonuç olarak (19) numaralı eşitlik, bir olaya ait koşullu bir olasılığın, bu olayın marjinal olasılığı ve bu koşullu olasılığın tersi ile elde edilebileceğini göstermektedir (Ntzoufras, 2009). Koşullu olasılıklar bağlamında değerlendirilen Bayes eşitliğini, hipotez doğrulama süreci bağlamında açıklamak daha anlaşılır olacaktır. Örneğin burada A doğrulanmaya çalışılan bir hipotez, B ise bu hipotezi doğrulamak için kullanılacak bir kanıt olsun. Bayes eşitliği yardımıyla $p(A|B)$'ye, yani bilinen B kanıtı ışığında bilinmeyen A hipotezi ile ilgili sonuca varılır (Kéry, 2010).

İlk bakışta basit gibi görünen Bayes eşitliği, aslında istatistiksel modellemeler için çok önemli bir temel oluşturmaktadır. Bilindiği gibi istatistiksel modellerin temel amacı, eldeki veri seti tarafından sağlanan bilgi kullanılarak model parametrelerini kestirmek ve bu kestirimler yardımıyla araştırma problemlerine yanıt aramaktır. Bayes yaklaşımı ile bu amacın gerçekleştirilmesinde veri seti ile sağlanan bilginin yanında, verilerden bağımsız olarak parametrelerin gerçek değerlerine ilişkin ön tahminler de kullanılmaktadır. Örneğin eldeki veri seti y ve bu veri setine uygulanacak model ile kestirilmeye çalışılan parametre θ ile sembolize edilsin. Bu istatistiksel model durumunda Bayes eşitliği,

$$f(\theta|y) = \frac{f(y|\theta)\,f(\theta)}{f(y)} \propto f(y|\theta)\,f(\theta) \tag{20}$$

şeklinde ifade edilir. Buradaki $f(\theta)$ terimi θ parametresine ait ön tahminlerimizi içeren önsel (prior) dağılımının fonksiyonunu, $f(y|\theta)$ terimi modelin olabilirlik fonksiyonunu ifade etmektedir. (20)'deki olabilirlik fonksiyonu, i her bir gözlemi ve n örneklem sayısını ifade etmek üzere,

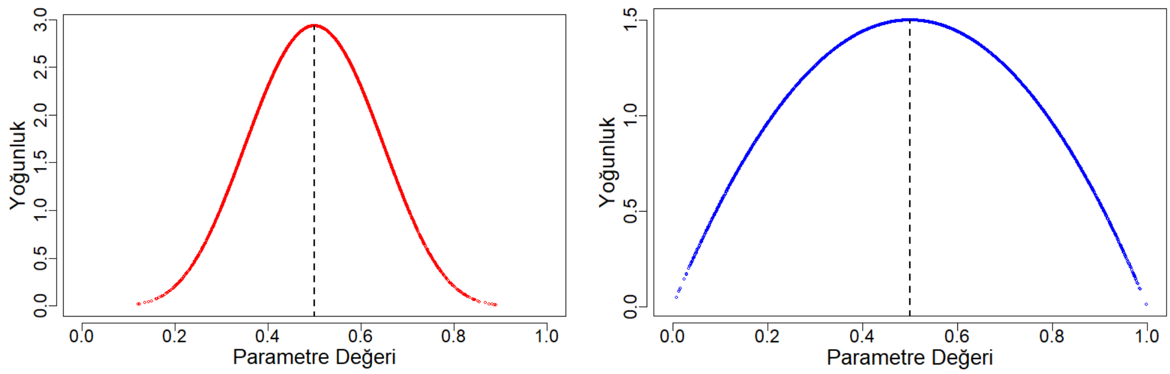
$$f(y|\theta) = \prod_{i=1}^{n} f(y_i|\theta) \tag{21}$$

eşitliği ile hesaplanır. Olabilirlik fonksiyonunun temel rolü, “hangi θ parametre değeri için gözlenen y_i değeri elde edilir?” sorusuna yanıt aramaktır. Öte yandan (20) numaralı eşitlikteki $f(\theta|y)$ terimi ise θ parametresine ait sonsal (posterior) dağılımın fonksiyonunu ifade eder. Parametrelere ait sonsal dağılım Bayes eşitliği ile ulaşılmaya çalışılan terim olup, aslında parametre kestirimlerini içeren dağılım olarak nitelendirilebilir.

Özetlemek gerekirse, Bayes yaklaşımı (20) numaralı temel eşitliği kullanarak parametre hakkındaki ön bilgi ve veri seti ile gelen bilgiyi harmanlar ve sonuçta parametre hakkındaki son bilgiye, yani parametre kestirimleri ve bu kestirimlere ait hatalara ulaşmaya çalışır. Aşağıda yer alan bölümde, Bayes yaklaşımındaki önsel ve sonsal dağılım kavramları hakkında detaylı bilgilere yer verilmiştir.

Bayes yaklaşımında önsel dağılım kavramı. Bayes yaklaşımı çerçevesinde parametrelerin seçkisiz değişkenler olarak ele alınmasıyla, her bir parametrenin analiz öncesinde önsel bir dağılıma sahip olduğu kabul edilir. Diğer bir ifadeyle, henüz veri seti ile gelen bilgiler değerlendirilmeden, istatistiksel modelin her bir parametresinin gerçek değerine ait ön tahminler vardır ve bu tahminlere ait kesinlik durumu da önsel dağılımların özellikleri ile ifade edilir (Fox, 2010). Önsel dağılımın ortalaması, ilgili parametrenin gerçek değeri için en olası tahmini ifade ederken, bu nokta tahmini hakkındaki bilgimizin kesinliği de dağılımın varyansı ile ifade edilir. Varyansın düşük olması halinde önsel olarak ortaya atılan parametre tahmini daha dar aralıkta olup, kesinlik daha fazladır. Ancak önsel dağılımın varyansı büyüdükçe, ilgili parametre için yaptığımız değer tahmini geniş bir aralıktadır ve bu tahminin kesinliği de düşük olur.

Aşağıda yer alan Şekil 1a ve 1b’deki aynı parametreye ait iki önsel dağılım örneği incelendiğinde, her iki dağılımın ortalamasının 0.5 civarında olduğu görülmektedir. Dolayısıyla ilgili parametre için en olası önsel tahminimiz her iki durumda 0.5’tir. Ancak 1b’deki dağılımın varyansı, 1a’daki dağılımdan daha büyüktür. Dolayısıyla bu nokta tahminimizin, 1b’deki önsel dağılımda daha muğlak olduğunu söyleyebiliriz.

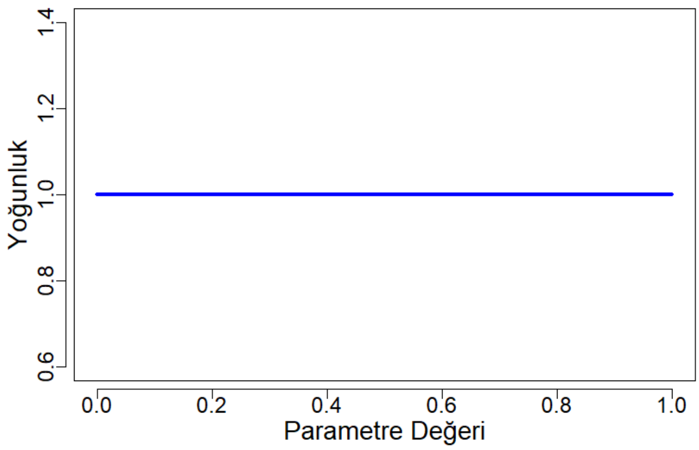


Şekil 1a ve 1b. Farklı varyansa sahip önsel dağılımlara örnekler.

Veri analizi öncesi belirlenen bu önsel dağılımlar, aslında birer olasılık yoğunluk fonksiyonu (probability density function) şeklindedirler. Burada vurgulanması gereken noktalardan biri, grafiklerin y eksenindeki değerlerin olasılık değil, yoğunluk olduğudur. Olasılıkların elde edilmesi için, ilgili değerler arasındaki alana tekabül eden alan hesaplanır. Örneğin parametre için önsel tahminimizin 0.4 ile 0.6 arasında olmasına ilişkin olasılık, fonksiyonun bu değerler altında kalan alanının integral yoluyla hesaplanmasıyla elde edilir. Eğrinin altında kalan toplam alan ise, x değişkeninin tüm değerleri için olasılıkların toplamı olan 1'e eşit olmaktadır. Diğer bir ifadeyle eğri altında kalan tüm alan olasılıkların toplamı olan 1'dir.

Parametreler için belirlenecek önsel dağılımlar, olasılık yoğunluk fonksiyonları matematiksel olarak tanımlanan Normal, Beta, Gamma, Tekbiçimli (Uniform) vb. istatistiksel dağılımlar olmaktadır. Bayes yaklaşımı ile parametre kestirim sürecinde, önsel dağılımların bu şekilde bilinen olasılık yoğunluk dağılımları olarak ifade edilmesi gerekir. Örneğin Şekil 1a ve 1b'de yer alan iki önsel dağılım, şekil parametreleri (shape parameters) ile tanımlanan ve her iki şekil parametresi sırasıyla 7 ve 2 olan Beta dağılımıdır. İlgili parametre θ ile sembolize edildiğinde, θ parametresi için belirlenen bu iki önsel dağılım sırasıyla $\theta \sim Beta(7,7)$ ve $\theta \sim Beta(2,2)$ şeklinde ifade edilir.

Model parametrelerine ilişkin ön tahminler geçmiş araştırmaların bulgularına veya uzman görüşlerine dayanarak şekillendirilebilir (Kruschke, 2014). Bu gibi durumlarda kullanılan ve nispeten düşük varyansa sahip önsel dağılımlar bilgi verici (informative) önsel dağılımlar şeklinde nitelendirilmektedir. Aslında Bayes yaklaşımının geleneksel yaklaşımı benimseyen araştırmacılar tarafından en çok eleştirildiği nokta, veri analizi öncesi önsel dağılımlar ile modele bilgi içeren tahminlerin dâhil edilmesidir. Geleneksel yaklaşımı benimseyen araştırmacılar, parametre kestirimi öncesi önsel dağılımlar ile sağlanan bilgilerin yanlılığa yol açacağını savunmaktadırlar. Bayes yaklaşımını benimseyen araştırmacılar ise bu durumu olasılıksal olayları açıklamada önceki deneyim ve bilgilerin kullanılması gerektiği düşüncesi ile savunmaktadırlar (Kruschke, 2014). Ancak yine de Bayes yaklaşımı bu eleştiriyi bertaraf edecek bir çözümü de sunmaktadır. Şöyle ki, ilgili parametre için elde herhangi bir bilgi bulunmadığında veya bu tür bir bilgi analize dâhil edilmek istenmediğinde, bu durumlara uygun önsel dağılımların tanımlanması mümkündür. Örneğin Şekil 1a ve 1b'de yer alan dağılımların varyansları oldukça yüksek değerlere (örneğin 100) çekildiğinde, parametre hakkındaki tahminlerimizin kesinliği de oldukça düşecektir. Bu tür dağılımlar Bayes alanyazınında bilgi vermeyen (non-informative) önsel dağılımlar şeklinde tanımlanmaktadır.



Şekil 2. Belirsiz önsel dağılıma örnek.

Öte yandan, Şekil 1a ve 1b'deki dağılımlarda olduğu gibi her bir değere değişen olasılıklarda şans vermek yerine, her değere de eş olasılık verilebilecek bilgi vermeyen önsel dağılımlar da tanımlanabilir. Özellikle tekbiçimli dağılımların örnek gösterilebileceği bu tür dağılımların integralleri 1'e eşit olmadığından aynı zamanda belirsiz (improper) önsel dağılımlar olarak da adlandırılmaktadırlar. Şekil 2'deki dağılım $\theta \sim Beta(1,1)$ şeklinde ifade edilen tekbiçimli bir dağılım olup, belirsiz önsel dağılımlara iyi bir örnektir. Bu önsel dağılımda 0 ile 1 arasında değer alabilen parametreye ait tüm olasılık yoğunluklarının eşit olduğu görülmektedir. Dolaysıyla herhangi bir değere daha yüksek bir olasılık atfedilmemekte, böylece parametre kestirimlerinde önsel dağılımın etkisi devreden çıkmaktadır. Beta dağılımı ile ifade edilebilen bu önsel dağılımın ayrıca $\theta \sim Uniform(0,1)$şeklinde ifade edilen ve 0-1 aralığında değer alabilen tekbiçimli dağılıma eşdeğer olduğunu belirtmekte fayda vardır.

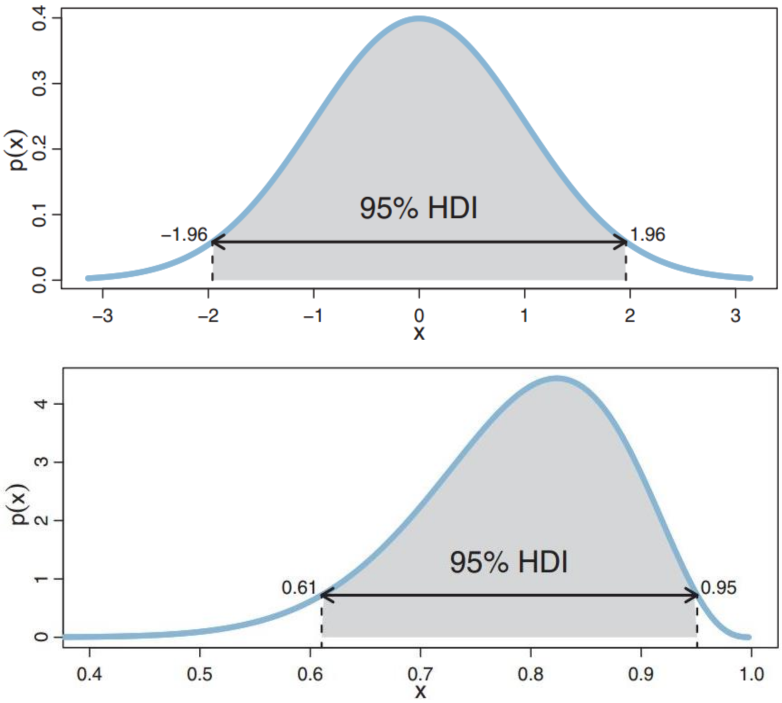
Hatırlanacağı üzere, Şekil 1a ve 1b'de yer alan iki önsel dağılımda en yüksek olasılığa sahip olan 0.5 değeri parametre için en olası ön tahmin idi. Şekil 2'deki önsel dağılımda da bu parametrenin önsel olarak 0.5 değerini alması mümkündür. Ancak 0-1 arası her değerin olasılığı eşit olduğundan, bu durumda 0.5 için en olası tahmin kavramı kullanılamaz. Öte yandan, buradaki dağılımın ranjı 0-1 arasındadır; ancak araştırmacılar kendi modelleri dâhilinde uygun önsel dağılımlar seçerek bu ranjı daha da büyütebilme esnekliğine sahiptirler (ör: $\theta \sim Uniform(0, 100)$ veya negatif değerleri de kapsayan $\theta \sim Uniform(-10, 10)$ gibi). Bu tür belirsiz önsel dağılımlar ile yapılan parametre kestirimleri, frekansa dayalı yaklaşımlar (örneğin ML) ile yapılan parametre kestirimlerine oldukça yakın çıkmaktadır.

Bayes yaklaşımında sonsal dağılım kavramı. Yukarıda vurgulandığı gibi, Bayes yaklaşımında parametreler seçkisiz değişkenlerdir; dolayısıyla bir parametrenin sonsal dağılımı, aslında o parametreye ait kestirimlerin dağılımını ifade etmektedir. Sonsal dağılım, parametrenin önsel dağılımı ve modelin olabilirlik fonksiyonu kullanılarak (20) numaralı Bayes eşitliği yardımıyla elde edilir. Tıpkı önsel dağılımlarda olduğu gibi, sonsal dağılımın en olası değeri (yani ortalaması) o parametrenin nokta kestirimi; dağılımın sahip olduğu standart sapma da bu nokta kestiriminin standart hatasını ifade etmektedir.

Bilindiği gibi geleneksel yaklaşımda standart hatalar nokta kestirimlerinin güven aralıklarının belirlenmesinde kullanılmaktadır. Normal dağılımın özelliği ve standart hata değeri kullanılarak, kestirilen parametre değeri için %95 güven aralığı belirlenir. Aslında Bayes yaklaşımında da benzer durum söz konusudur; ancak güven aralığı sayısal bir değer olarak kestirilmekten ziyade, sonsal dağılımın kendisinden elde edilir. Diğer bir ifadeyle, sonsal dağılım zaten parametre için farklı olasılıklarla kestirilmiş değerlerden meydana gelir; dolayısıyla yapılan işlem, bu dağılımın 2.5 ve 97.5 yüzdeliklerine denk gelen parametre değerlerinin belirlenmesidir. Belirlenen bu değerler, ilgili parametre için elde edilen nokta kestiriminin %95 güven aralığının uç değerleri olur.

Bayes yaklaşımında genellikle normal dağılım referans alınarak sonsal dağılımın 2.5 ve 97.5 yüzdelikleri ile sınırlanan %95 güven aralığı elde edilir.

Belirlenen bu aralık, Bayes alanyazınında Bayes güven aralığı (Bayesian Credible Interval) olarak da anılmaktadır. Ancak Kruschke (2014), sonsal dağılımların simetrik olmaması halinde bu yaklaşımın doğru olmayacağını belirterek Highest Density Interval (HDI) olarak adlandırılan en yüksek yoğunluk aralığı kavramının kullanılmasını önermiştir. Aşağıda Kruschke'den (2014, s.88) alınan iki sonsal dağılıma ait grafikler incelendiğinde, HDI kavramı daha iyi anlaşılacaktır.



Şekil 3a ve 3b. Farklı sonsal dağılımların %95 HDI değerlerinin gösterimi.

Şekil 3a'da görülen dağılım standart normal bir dağılımdır, dolayısıyla %95 güven aralığının alt ve üst limiti simetrik olmaktadır. Şekilden de anlaşılacağı üzere sonsal dağılımın simetrik olması halinde, geleneksel güven aralığı ile HDI aynı olacaktır. Ancak Şekil 3b'de yer alan çarpık dağılım için geleneksel yöntemle simetrik bir güven aralığı belirlenmesi halinde, dağılımın sol ucundaki bazı değerlere ait olasılıkların düşük olmasına rağmen güven aralığına girmesi; buna karşın olasılıkların daha yüksek olduğu sağ uçtaki bazı değerlerin ise bu aralığa girememesi söz konusu olacaktır. Bu çarpık dağılım için geleneksel güven aralığı yerine %95 HDI kullanıldığında, eğri altında kalan toplam alanda (1'e eşit olan toplam olasılık) en yüksek olasılık değerlerini içeren %95'lik bölüm belirlenir ve bu bölümü sınırlayan alt ve üst parametre değerleri güven aralığının limitlerini gösterir. Sonuç olarak bu %95'lik alan içerisinde yer alan her bir parametre değerine ait olasılık, alanın dışında yer alan herhangi bir parametre değerine ait olasılıktan büyük olmaktadır (Kruschke, 2014).

Sonsal dağılımın ortalaması ve varyansı, önsel dağılım ile modelin uygulandığı veri setinden gelen bilginin niteliğine göre değişkenlik göstermektedir. Örneğin bilgi içermeyen bir önsel dağılım kullanıldığında, parametreye ait sonsal dağılımın belirlenmesinde sadece veri seti tarafından sağlanan bilgi kullanılır ve daha önce vurgulandığı gibi parametre kestirimleri frekansa dayalı yaklaşımlar ile elde edilecek kestirimlere oldukça yakın olur. Ancak aynı veri seti söz konusu iken, parametrenin önsel dağılımı bilgi verici nitelikte olduğunda, sonsal dağılımın ortalaması ve varyansında önsel dağılımdan gelen bu bilgi ile bir değişkenlik olacaktır. Bu değişkenlik; parametrenin nokta kestiriminin önsel dağılımın ortalamasına doğru kayması, kestirime ilişkin hatayı ifade eden varyansın ise azalması şeklinde olacaktır. Öte yandan önsel dağılımlar aynı iken, modelin uygulandığı verinin artırılması halinde, parametrenin nokta kestirimi olabilirlik dağılımının ortalamasına doğru bir çekilme yaşarken, yine modele sağlanan fazladan bilgi sayesinde sonsal dağılımın varyansında bir düşüş olacaktır (Kruschke, 2014).

Sonuç olarak Bayes yaklaşımında sonsal dağılımlar ile ifade edilen parametre kestirimleri, veri seti ve önsel dağılım ile gelen bilgiler ile dengelenmektedir. Dolayısıyla veri seti veya önsel dağılımda bir bilgi artışı olması halinde, kestirimler de o yönde bir hareket gösterecek ve kestirimlerin hatalarında bir düşüş yaşanacaktır. Aşağıda yer alan bölümde, Bayes yaklaşımı ile kestirimin en önemli aşaması olan sonsal dağılımların elde edilmesine ilişkin bilgiler sunulmuştur.

Sonsal dağılımların elde edilmesi. (20) numaralı Bayes eşitliği basit gibi görünse de, önsel dağılım ve olabilirlik fonksiyonunun eşitlikte yerlerini alması ile sonsal dağılımın elde edilmesi zorlaşmaktadır. Bu durum özellikle çok sayıda kestirilecek parametreye sahip karmaşık modellerde daha da belirgin olmaktadır.

Parametrenin nokta kestirimi ve bu kestirime ait güven aralığının elde edilebilmesi için, ilgili parametrenin sonsal dağılımının ortalaması ve varyansına ulaşmak gerekmektedir. Sonsal dağılımın matematiksel formunun belli bir olasılık dağılımı olması halinde bu dağılıma ait ortalama ve varyansın bulunması kolaydır.

Bu tür dağılımlar matematiksel olarak çözümlenebilir (tractable) dağılımlar olarak nitelendirilmektedir (Ntzoufras, 2009). Matematiksel olarak çözümlenebilir sonsal dağılımlar elde edebilmek için, eldeki modele uygun olan özel önsel dağılımların kullanılması gerekmektedir. Matematiksel çözüme olanak sağlayan bu tür özel önsel dağılımlara eşlenik (conjugate) önsel dağılımlar adı verilmektedir.

Uygun olabilirlik fonksiyonları ile eşlenik olan bir önsel dağılım kullanıldığında elde edilen sonsal dağılımın türü önsel dağılım ile aynı olmakta ve matematiksel çözüm mümkün olmaktadır. Örneğin verinin Binom dağılımı gösterdiği modellerde (olabilirlik dağılımı da Binom olur) önsel dağılım olarak Beta dağılımı kullanıldığında, ilgili parametrenin sonsal dağılımı da Beta dağılımı olur ve bu dağılımın ortalaması ile varyansının matematiksel olarak hesaplanması mümkündür (Ntzoufras, 2009). Ancak eşlenik önsel dağılımlar çok sınırlı türde ve parametre sayısı az olan modellerde kullanılabilmekte, dolayısıyla her türlü modele uyum sağlayacak eşlenik önsel dağılım bulunamamaktadır. Ayrıca belirli olabilirlik dağılımları için kısıtlı sayıda eşlenik önsel dağılımların olması, araştırmacıların önsel dağılım seçiminde esnek davranmalarına engel olmaktadır.

Bilindiği gibi, genellikle istatistiksel modellerde kestirilmesi gereken birden çok parametre bulunmaktadır. Parametre sayısı birden fazla olduğunda Bayes eşitliği ile tüm parametrelerin bileşke (joint) sonsal dağılımı elde edilmekte, bu bileşke dağılımdan her bir parametrenin marjinal sonsal dağılımına ulaşılması gerekmektedir. Örneğin geleneksel Rasch modelde θ yetenek parametrelerini, β ise madde parametrelerini temsil etmek üzere, bu iki parametre setine ait bileşke sonsal dağılım

$$p(\theta,\beta|y) = \frac{p(y|\theta,\beta)\, p(\theta)\, p(\beta)}{p(y)} \tag{22}$$

olarak ifade edilmektedir. Dikkat edilirse bu eşitlik, (19) numaralı temel Bayes eşitliğinin aynısıdır. Buradaki farklılık, kestirilecek iki farklı parametre seti olduğundan sonsal dağılımların $p(\theta,\beta|y)$ şeklinde sembolize edilerek bileşke olarak verilmesidir. Bu bileşke sonsal dağılımdan her bir madde ve yetenek parametresine ait marjinal sonsal dağılıma ulaşılması gerekmektedir. Bunun için bileşke sonsal dağılımın integrali alınmalıdır (Fox, 2010). Örneğin yetenek parametrelerinin marjinal sonsal dağılımları,

$$p(\theta|y) = \int p(y|\theta, \beta)\, p(\theta)p(\beta)\, /p(y)d(\beta) \qquad (23)$$

integrali ile elde edilir. Burada $p(\theta|y)$ yetenek parametrelerinin marjinal sonsal dağılımını, $p(\theta)$ yetenek parametreleri için önsel dağılımı, $p(\beta)$ madde parametreleri için önsel dağılımı ve $p(y|\theta, \beta)$ ise modelin olabilirlik dağılımını ifade etmektedir. (23) numaralı eşitlikteki gibi integraller, özellikle modeller daha da karmaşık olduğunda matematiksel olarak çözülemez bir hal almaktadır. İntegralin çözülememesi durumunda parametre kestirimlerine de ulaşmak mümkün değildir.

Yukarıda bahsedilen eşlenik önsel dağılımların kısıtlı uygulamaları ve karmaşık integrallerin matematiksel çözümlerinin olanaksızlığı, Bayes yaklaşımı ile kestirimlerin uzun yıllar boyunca yaygın olarak kullanılmasına engel oluşturmuştur (Kruschke, 2014). Ancak ilerleyen yıllarda sonsal dağılımların elde edilmesinde kullanılabilecek simülasyona dayalı yöntemlerin oldukça gelişmesi ile bu kısıtlılık ortadan kalkmıştır. Özellikle yüksek işlem kapasiteli bilgisayarların daha ucuz ve ulaşılabilir hale gelmesiyle, simülasyona dayalı kestirimler daha kolay ve hızlı olarak elde edilmeye başlanmıştır. Tüm bu gelişmeler ile beraber Bayes yaklaşımına dayalı parametre kestiriminin kullanımı daha da yaygınlaşmıştır. Aşağıda yer alan bölümde sonsal dağılımların simülasyona dayalı yöntemler ile elde edilmesine ilişkin bilgilere yer verilmiştir.

Simülasyona dayalı yöntemler ve Gibbs örneklemesi. Simülasyona dayalı Bayes çözümlerinde sonsal dağılımların matematiksel formlarına ulaşılmasından ziyade, bu dağılımlardan geniş örneklemler üretilmesi söz konusudur. Her bir parametreye ait sonsal dağılımı yeterince temsil edecek bir örneklem üretilebilmesi halinde, bu dağılımın ortalaması ve varyansı başta olmak üzere diğer istatistiksel özellikleri hakkında bilgi sahibi olunabilir. Genel anlamda simülasyona dayalı veri üretiminde en çok kullanılan tekniklerin başında Monte Carlo (MC) yöntemi gelmektedir. MC yöntemi ile simülasyon direkt simülasyon olarak da anılmakta olup, tek bir parametre için veri üretmede etkilidir; ancak yukarıda belirtildiği gibi birbirlerine bağımlı olan birden fazla parametre olması durumunda MC yöntemi yerine Markov Chain Monte Carlo (MCMC) simülasyon yöntemleri kullanılmaktadır (Ntzoufras, 2009).

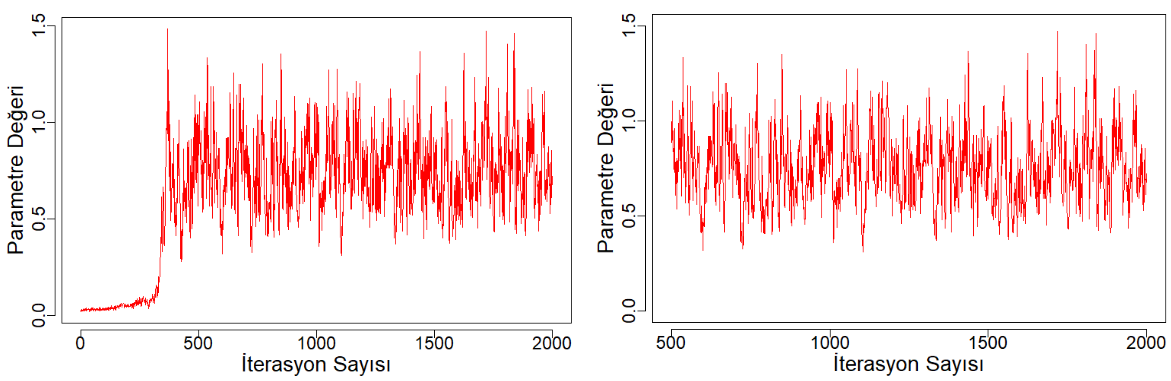
MCMC yöntemlerinin Bayes yaklaşımındaki temel rolü, her bir parametre için Bayes eşitliğinden elde edilecek sonsal dağılımı yeterince temsil edecek büyüklük ve özellikte bir örnekleme ulaşmaktır. Markov zinciri adı verilen iteratif bir algoritma ile işleyen MCMC yönteminde, peş peşe üretilen veriler sonucunda elde edilecek örneklemin ilgili sonsal dağılımı yeterince temsil edeceği; diğer bir ifadeyle iterasyonların yakınsayacağı (convergence) varsayılmaktadır. Alanyazında çok sayıda MCMC algoritması yer almaktadır. Bayes yaklaşımı ile parametre kestirimlerinde en çok kullanılan MCMC algoritması Metropolis-Hastings algoritmasının özel bir türü olan Gibbs örneklemesidir (Gibbs sampling) (Fox, 2010).

Gibbs örneklemesinde modelin her bir parametresi için koşullu sonsal dağılımdan iteratif olarak değerler üretilir. Süreç bir zincir şeklinde ilerleyerek her adımda üretilen değerler bir sonraki adımda güncellenir. Örneğin y ile sembolize edilen veri setinden bir model ile kestirilecek toplam k adet parametre $\theta_1, \theta_2, \dots, \theta_k$ olsun. İlk aşamada bu k adet parametre için sahip oldukları önsel dağılımların ranjı içerisinden rastgele seçilen $\theta_1^{(0)},\ \theta_2^{(0)}, \dots, \theta_k^{(0)}$ başlangıç değerleri (initials) ile zincir başlatılır. Üst indisteki değerler iterasyon sayısını göstermekte olup, sıfır indis değeri zincirin başlangıcını ifade etmektedir. Ardından $\theta_1^{(1)},\ p\left(\theta_1 \middle| \theta_2^{(0)},\ \theta_3^{(0)} \dots, \theta_k^{(0)},\ y\right)$ koşullu dağılımından; $\theta_2^{(1)},\ p\left(\theta_2 \middle| \theta_1^{(1)},\ \theta_3^{(0)} \dots, \theta_k^{(0)},\ y\right)$ koşullu dağılımından ve bu işlem k. parametre için $\theta_k^{(1)}, p\left(\theta_k \middle| \theta_1^{(1)},\ \theta_2^{(1)}, \theta_3^{(1)} \ \dots, \theta_{k-1}^{(1)},\ y\right)$ dağılımından üretilerek $^{(1)}$ indisi ile belirtilen zincirin birinci iterasyonu tamamlanmış olur. Burada dikkat edilmesi gereken nokta, ilk iterasyon içerisinde her bir parametre için yeni bir değer üretildiğinde, bu değer ilgili iterasyonda sıradaki parametrenin üretilecek değeri için koşullu bir bilgi olarak ele alınmakta, yani elde edilen yeni bilgi ile iterasyon güncellenmektedir. Bu şekilde toplam m tane iterasyon tamamlandığında $\theta_1^{(m)}$, $\theta_2^{(m)}, \dots, \theta_k^{(m)}$ şeklinde sembolize edilen k tane parametre için zincirin son değerleri üretilir. m iterasyon sayısı sonsuza yaklaştığında, her bir θ_k parametresi için üretilen toplam m tane değerin dağılımının, ilgili parametrenin Bayes eşitliği ile ulaşılmaya çalışılan marjinal sonsal dağılımına yakınsayacağı belirtilmektedir (Dowling, 2006; Fox, 2010). Diğer bir ifadeyle, m tane iterasyon sonrasında yakınsama gerçekleşmesi halinde, elimizdeki k tane parametrenin her biri için üretilen m büyüklüğünde bir örneklem olacaktır. Bu örneklemler ilgili parametrenin marjinal sonsal dağılımını temsil eden özellikte olup parametreye ait kestirim ve %95 HDI başta olmak üzere her türlü istatistiğin elde edilmesinde kullanılacaktır.

Bayes yaklaşımı ile kestirim sürecinin son aşamasında MCMC iterasyonlarının yakınsama durumlarının ve yakınsamanın gerçekleşmesi halinde elde edilen sonsal dağılımların özelliklerinin incelenmesi gerekmektedir. Gerek MCMC'nin doğru şekilde tamamlanması, gerekse örneklemenin tamamlanmasından sonra parametre kestirimlerinin doğru şekilde yorumlanabilmesi için birtakım koşulların sağlanması gerekmektedir. Aşağıda yer alan bölümde MCMC sürecinin değerlendirilmesi ve elde edilen çıktıların yorumlanmasına ilişkin bilgilere yer verilmiştir.

MCMC sürecinin değerlendirilmesi. m iterasyon sayısı sonsuza doğru gittiğinde yakınsamanın kesinlikle gerçekleşeceği ve elde edilecek bu sonsuza yaklaşan büyüklükteki örneklem ile sonsal dağılımın mükemmel bir şekilde temsil edileceği bilinmektedir (Ntzoufras, 2009; Kruschke, 2014). Ancak elbette iterasyonların sonsuz sayıda yapılması mümkün değildir ve dolayısıyla m tane iterasyon sonucunda elde edilecek örneklemin, tıpkı geleneksel istatistik teorisindeki örneklem-evren ilişkisinde olduğu gibi sonsal dağılımı temsil edeceği varsayılır. MCMC iterasyonlarının ne kadar uzunlukta olması gerektiği, yani hangi m sayısının yeterli olacağı, aslında birden fazla koşula bağlı olmaktadır. Ancak öncelikli olarak yakınsamanın gerçekleşip gerçekleşmediği kontrol edilmelidir.

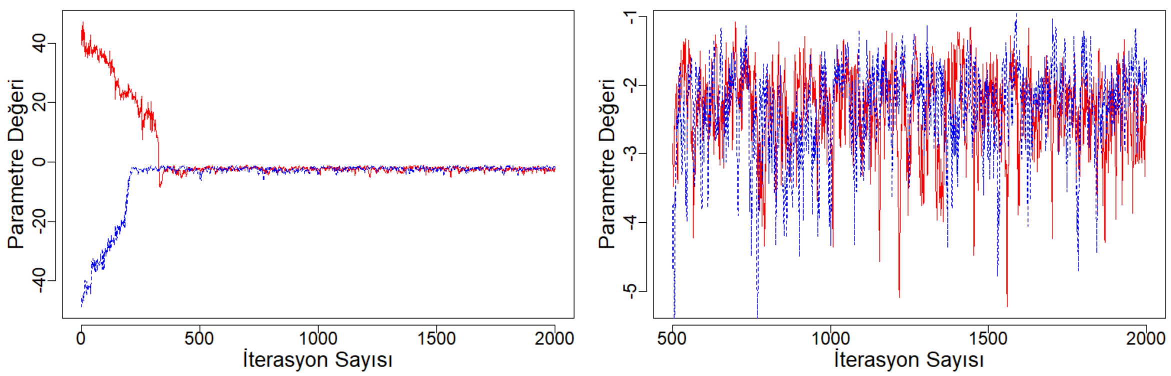
Yakınsamanın kontrol edilmesi. Yakınsamanın kontrolü görsel ve sayısal olarak yapılmaktadır (Kruschke, 2014). Görsel kontrolde, MCMC ile üretilen değerlerin her bir iterasyonda gösterdiği değişimin grafiği incelenmektedir. Bayes alanyazınında bu grafiklere iz grafiği (trace plot) adı verilmektedir. İz grafiklerinde başlangıç değeri ($\theta_k^{(0)}$) ile başlayan MCMC simülasyonunun, ilerleyen iterasyon sayısı sonrasında gösterdiği değişim görsel olarak incelenir. İstenen durum, belirli bir iterasyon sayısından sonra iz grafiğinin istikrarlı olarak ilerlemesidir.



Şekil 4a ve 4b. İz grafiği örneği.

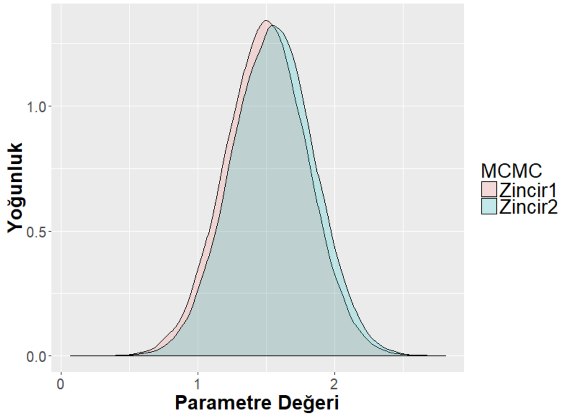
Şekil 4a'da yer alan örnek iz grafiği incelendiğinde, ilgili parametre için önsel değerin 0 civarında olduğu ve iterasyon sayısının yaklaşık olarak 500'ü geçmesinden sonra iz grafiğinin 0.4 ile 1.5 değerleri arasında istikrarlı bir biçimde ilerlediği görülmektedir. Bu durumda, zincirin 500'den önceki değerlerinin sonsal dağılımı temsil etmeyen nitelikte olduğu söylenebilir. MCMC zincirindeki bu ilk iterasyonlara ısınma (burn-in) periyodu adı verilmekte ve başlangıç değerlerinin etkisini içeren bu kısmın sonuçlara dâhil edilmemesi gerekmektedir (Ntzoufras, 2009). Şekil 4b'de görülen iz grafiği, Şekil 4a'daki iz grafiğinden ısınma periyodunun çıkarılmış halidir. Görüldüğü gibi ısınma periyodunun çıkarılmasının ardından, 500-2000 arası iterasyonlarda zincirin yapısı daha istikrarlıdır.

Yakınsamanın daha iyi değerlendirilmesinde farklı başlangıç noktalarından başlatılacak birden fazla zincirin kullanılması daha verimli olmaktadır. Bu şekilde başlangıç değerlerinin kontrolü daha iyi yapılabilir ve ulaşılmaya çalışılan sonsal dağılım aynı olduğundan, tüm zincirlerin yakınsama halinde birbirleriyle örtüşmesi beklenir. Örneğin Şekil 5a ve 5b'de, bir parametre kestirimi için farklı başlangıç noktalarından başlatılan iki zincire ait iz grafikleri yer almaktadır. Şekil 5a'daki grafikte görüldüğü gibi, yine yaklaşık 500 iterasyon sonrasında her iki zincir de istikrarlı bir hale gelmekte ve birbirleriyle örtüşmektedir. Bu iz grafiğinden ısınma periyodu olarak ilk 500 değer çıkarıldığında Şekil 5b'deki iz grafiği elde edilir. Isınma periyodunun çıkarıldığı bu iz grafiğinde, zincirlerin birbiriyle örtüştüğü daha iyi gözlenebilir.



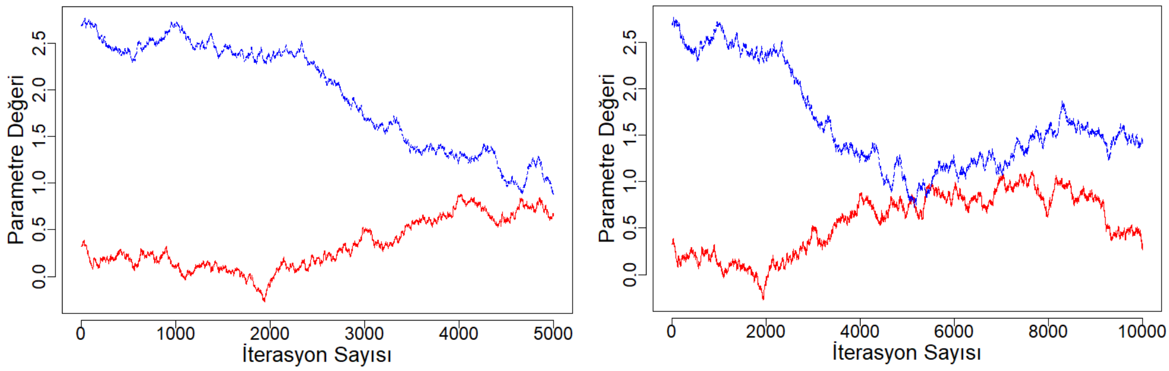
Şekil 5a ve 5b. Birden çok zincire sahip iz grafiği örneği.

Zincirlerin örtüşme durumu, her bir zincire ait aşağıda yer alan Şekil 6'daki dağılım grafiklerinden de gözlenebilir. Görüldüğü gibi her iki zincirden elde edilen değerlere ait dağılımlar oldukça benzerdir. Bu durumda farklı başlangıç noktalarından başlatılan her iki zincirin de aynı sonsal dağılımı temsil eden örneklemleri oluşturduğu yorumu yapılır. Elde edilen bu görsel kanıtlar ışığında, ilgili parametre için yakınsamanın gerçekleştiği söylenebilir.



Şekil 6. İki zincire ait dağılım grafiği.

MCMC'de birden fazla zincir kullanılması, görsel kanıtların yanında yakınsamanın sayısal göstergesi olan Brooks-Gelman-Rubin (BGR) (Brooks & Gelman, 1998) istatistiğinin hesaplanmasına da olanak sağlamaktadır. BGR, her bir zincir içindeki ve zincirler arasındaki değişkenliği dikkate alarak ANOVA temelli bir istatistik sunar ve 1.0 ile 1.1 arasındaki BGR değerleri yakınsamanın olduğuna işaret eder (Ntzoufras, 2009). BGR istatistiğinin 1.1 değerini aşması halinde, yakınsamanın gerçekleşmediği ve zincirlerden en az birinin istikrarlı bir yapıya kavuşmadığı yorumu yapılır. Zincirin fazla sayıda iterasyona rağmen istikrarlı bir yapıda olmaması, üretilen değerlerin sonsal dağılımı temsil etmediğini ifade eder. Şekil 7a ve 7b'de yer alan iz grafikleri, yakınsamanın olmadığı bu gibi durumlara bir örnek teşkil etmektedir.



Şekil 7a ve 7b. Yakınsamayan zincirlere ait iz grafiği örneği.

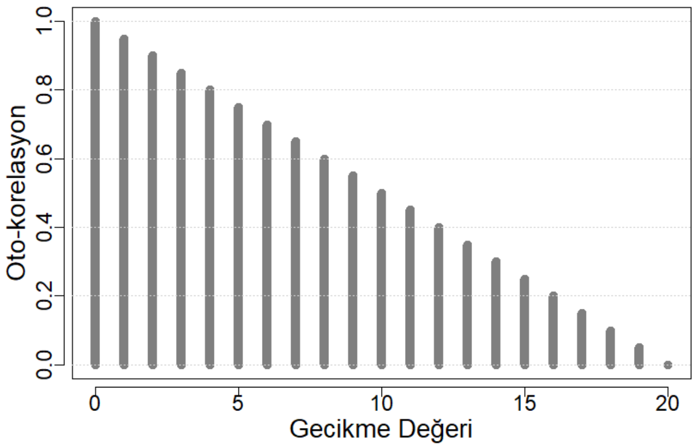
Şekil 7a'da yer alan grafikteki senaryoda, iterasyon sayısı 5000'den 10000'e çıkarılmasına rağmen Şekil 7b'de görüldüğü gibi her iki zincir de kararlı bir yapıya ulaşamamakta ve birbirleriyle örtüşememektedirler. Bu gibi durumlarda iterasyon sayısının daha da artırılması denenebilir; ancak tıpkı bu örnekte olduğu gibi oldukça fazla sayıda iterasyona rağmen durumun değişmemesi halinde sorun modelin kendisinde veya MCMC kodlamasında aranmalıdır. Diğer bir ifadeyle, bu gibi durumlarda problem modelin kendisiyle veya kodlama hatasıyla ilgili olduğundan, iterasyon sayısının sonsuza kadar artırılması halinde bile yakınsama gerçekleşmeyecektir.

Oto-korelasyon (autocorrelation). MCMC iterasyonlarının yakınsama durumları görsel ve sayısal olarak teyit edildikten sonra, her bir parametreye ait sonsal dağılımın etkili bir biçimde temsil edilebilmesi için gerekli m iterasyon sayısına karar verilmesi gerekmektedir. Elbette m iterasyon sayısı ne kadar fazla olursa, elde edilecek Gibbs örneklemesi sonsal dağılımı daha iyi temsil edecektir. Ancak yakınsamanın gerçekleştiğine kanaat getirilmesinin ardından, yine birtakım ölçütlere göre m sayısının yeterliliğine karar verilebilir. Bu ölçütlerden bir tanesi, zincirlerin oto-korelasyon değerlerinin incelenmesidir.

Oto-korelasyon kavramı, bir zincirdeki iterasyonların ilerlemesi ile beraber sonsal dağılımdan üretilen parametre değerlerinin birbirlerinden farklılaşmasının bir ölçüsü olarak ifade edilebilir. Oto-korelasyon değerinin yüksek olması, ilerleyen iterasyonlara rağmen üretilen parametre değerlerinin sonsal dağılımın tüm bölgelerini yeterince temsil etmediğini gösterir. Dolayısıyla yüksek oto-korelasyon durumlarında, sonsal dağılımın yeterince temsil edilebilmesi için daha fazla iterasyona ihtiyaç duyulacaktır. Yüksek oto-korelasyon değeri ayrıca zincirlerin yavaş yakınsadığını ve örtüşmenin de yavaş bir biçimde gerçekleştiğini ifade etmektedir (Kruschke, 2014). Bu açıdan, yüksek oto-korelasyon değerlerine rağmen, iterasyon sayısının artırılması halinde sonsal dağılımı yeterince temsil edecek bir örneklem elde edilebileceğini söylemek yanlış olmaz.

Oto-korelasyon değeri, ısınma periyodunun çıkarılmasının ardından kalan zincir içerisindeki parametre değerlerinde k gibi bir gecikme (lag) değeri dikkate alınarak hesaplanmaktadır. Örneğin $k = 5$ için hesaplanacak oto-korelasyon değeri, bir parametreye ait zincirin uzunluğu $m = 1000$ olmak üzere, zincirin birinci değerinden 995. değerine kadar olan set ile; zincirin 5. değerinden 1000. değerine kadar olan set arasındaki korelasyon değeridir. Bu korelasyon değerinin yüksek çıkması, ilgili parametre için iterasyonların ilerlemesine rağmen üretilen değerlerin çok değişmediğini, yani yakınsamanın yavaş bir şekilde gerçekleştiğini ifade eder.

Daha yüksek gecikme değerleri alındığında oto-korelasyon değeri daha da düşecektir. Çünkü iterasyonun daha ilerisinden başlanarak alınacak parametre değerlerinin, ilk değerler ile daha düşük korelasyon göstermesi beklenen bir durumdur. Artan gecikme değerlerine bağlı olarak oto-korelasyon değerlerinin gösterdiği değişim, oto-korelasyon grafikleri ile daha iyi gözlenebilir.



Şekil 8. Oto-korelasyon grafiği örneği.

Şekil 8'de yer alan örnek oto-korelasyon grafiği incelendiğinde, gecikme değerinin 20 olması ile beraber oto-korelasyonun sıfıra yaklaştığı görülmektedir. Bu grafikten hareketle, bir zincirdeki ardışık tüm iterasyonların alınması yerine, her 20. değerin alınmasıyla elde edilecek örneklemde oto-korelasyonun sıfır olacağı sonucuna varılır. Bu şekilde her k. değerin alınması işlemine Bayes alanyazınında inceltme (thinning) adı verilmektedir. İnceltme işlemi ile her k. değer alındığından, elde edilecek örneklem inceltme yapılmamış durumdaki örneklemden daha küçük olacaktır. Ancak oto-korelasyonun yüksek olması, ardışık iterasyonlar ile üretilen parametre değerlerinin birbirlerine oldukça yakın olduğunu göstermektedir. Dolayısıyla inceltme ile yapılan işlem, birbirine oldukça yakın olan bu ardışık değerlerin alınması yerine, gecikme değerinin belirttiği iterasyondaki farklılaşan değerin alınmasıdır.

İnceltme işlemiyle, üretilen her k tane parametre değerinden biri alınıp oto-korelasyon değeri sıfıra çekilse de, toplamda gerçekleştirilen iterasyon sayısı aynı kalmaktadır. Dolayısıyla inceltme işlemi ile zaman kazanmaktan ziyade, daha az sayıda değer saklandığından, sadece bilgisayar belleği kullanımı açısından avantaj sağlanmaktadır (Kruschke, 2014). Ancak Link ve Eaton (2012) inceltme işlemi ile üretilen tüm değerlerin alınması yerine, her k. değerin alınması ile kestirimlerde bir bilgi kaybının yaşanacağını belirtmektedirler. Dolaysıyla bellek sıkıntısı olmadığı durumlarda, inceltme yoluna gitmek yerine iterasyon sayısının daha da artırılıp üretilen tüm değerlerin dikkate alınması daha verimli olacaktır.

Etkili örneklem büyüklüğü (effective sample size: ESS). İterasyonlarda yüksek bir oto-korelasyonun olması durumunda, ardışık olarak üretilen değerlerin yavaş bir biçimde farklılaşacağı daha önce belirtilmişti. Dolayısıyla bu gibi durumlarda m büyüklüğünde üretilen bir örneklemin tüm değerlerinin birbirlerinden tamamen bağımsız olmayacağı açıktır (Kruschke, 2014). Oto-korelasyon değerinin daha da büyük olması ile beraber m büyüklüğündeki örneklemin düşük bir kısmı sonsal dağılımı temsil edici nitelikte olur. Bu gibi durumlarda, elde edilen örneklemin sonsal dağılımı temsil etme gücü ESS değeri ile incelenir.

ESS, yapılan m iterasyon sayısı ve her bir zincirdeki oto-korelasyon değeri ile hesaplanır. Örneğin, bir parametre için m uzunluğunda bir iterasyon gerçekleştirildiğinde, Şekil 8'deki gibi bir oto-korelasyon durumu söz konusu ise, ESS m değerinden daha düşük olacaktır. Dolayısıyla, etkili parametre kestirimleri için önceden hedeflenen iterasyon sayısından daha fazla iterasyona ihtiyaç duyulacaktır. Kruschke (2014), özellikle %95 HDI'nin mümkün olduğunca hassas bir biçimde kestirilebilmesi için etkili örneklem büyüklüğünün en az 10000 olması gerektiğini önermektedir. Ancak araştırmacılar için uygulamalarında %95 HDI'nin hassasiyeti çok önemli değilse, daha düşük ESS değerlerinin de yeterli olacağını vurgulamıştır.

MC hatası (MC error). MCMC iterasyonları sonrasında elde edilen örneklemin ortalaması, ilgili parametre için nokta kestirimi olmaktadır. İterasyon sayısının sonsuza yaklaşmasıyla birlikte, bir parametre için üretilecek değerler sonsal dağılımı mükemmel bir biçimde temsil eder. Böylece örneklem ortalaması, sonsal dağılımın ortalamasına tam olarak eşit olur ve dolaysıyla parametrenin gerçek değerine ulaşılır. Ancak MCMC simülasyonu ile sonsal dağılımdan m büyüklüğünde bir örneklem elde edilir ve bu örneklemin ortalaması, sonsal dağılımın ortalamasını belirli bir hata ile temsil eder. İşte kestirimlerdeki simülasyondan kaynaklı değişimleri ölçen bu hataya MC hatası adı verilmektedir (Ntzoufras, 2009).

MC hatası, m iterasyon sayısının artması ile düşüş gösterir ve dolayısıyla etkili bir kestirim için minimum iterasyon sayısının ne olması gerektiği konusunda araştırmacıya bir fikir verebilir. Kéry (2010), MC hatasının parametreye ait sonsal dağılımın standart sapmasının %5'inden küçük olması gerektiğini belirtmektedir. Dolayısıyla bu şartın sağlanmasına dek iterasyon sayısının artırılması, etkili parametre kestirimleri için makul bir hareket olacaktır. Daha önce belirtildiği gibi MCMC iterasyonlarında yüksek oto-korelasyonların olması, elde edilecek etkili örneklem büyüklüğünün düşmesine neden olmaktadır. Bu açıdan özellikle yüksek oto-korelasyon durumlarında MC hatasının düşürülmesi için iterasyon sayısının daha da artırılması gerekecektir.

MCMC sürecinin değerlendirilmesinde dikkat edilecek hususları özetlemek gerekirse, ilk olarak MCMC simülasyonlarının yakınsama durumları görsel olarak değerlendirilmeli ve ısınma periyodunun hangi iterasyondan itibaren dikkate alınması gerektiğine karar verilmelidir. Zincirlerin yeterince örtüştüğüne görsel olarak kanaat getirilmesinin ardından; BGR istatistiği, ESS ve MC hatası incelenerek iterasyonların ne kadar uzunlukta olması gerektiğine karar verilmelidir. İterasyonların tamamlanmasının ardından, her bir parametre için üretilen örneklemin ortalaması nokta kestirimi iken; bu örnekleme ait standart sapma ise standart hata görevi görür. Ayrıca parametre kestiriminin daha etkili olan güven aralığı % 95 HDI ile ifade edilmektedir.

Bayes yaklaşımı ile kestirimin aşamaları. Bu bölüme kadar Bayes yaklaşımı ile kestirim ile ilgili aktarılan teorik bilgilerin daha iyi yapılandırılabilmesi için, basit bir modele ait parametre kestirimlerinin nasıl elde edildiğinin baştan sona aşamalı olarak açıklanması faydalı olacaktır.

Önceki bölümlerde vurgulandığı gibi, parametre kestirimlerini elde edebilmek için önsel dağılım ve modelin olabilirlik fonksiyonu kullanılarak Bayes eşitliğiyle parametrelerin sonsal dağılımlarına ulaşılması gerekmektedir. Bunun için ilk olarak veriye uygulanması planlanan model ile kestirilecek parametreler belirlenmelidir. Sonraki aşamada, eldeki verinin yapısı ve istatistiksel model dikkate alınarak veri gözlemlerinin gösterdiği dağılım belirlenmeli ve modelin olabilirlik fonksiyonu yazılmalıdır. Son aşamada ise modele ait tüm parametreler için önsel dağılımlar belirlenmelidir. Örneğin sürekli olan bir y değişkeninin, yine sürekli olan bir x değişkeni ile yordandığı basit doğrusal regresyon modelini düşünelim. Bu doğrusal regresyon modelinin matematiksel eşitliği,

$$y_i = a + bx_i + e_i \tag{24}$$

şeklinde ifade edilir. Doğrusal regresyonun varsayımı gereği, artıklar $e_i \sim N(0, \sigma^2)$ şeklinde ifade edilen, 0 ortalama ve σ^2 varyans ile normal dağılım gösterir. Bu modelden kestirilecek olan parametreler a, b ve σ^2'dir. Her bir i bireyi için beklenen bağımlı değişken değeri $\mu_i = a + bx_i$ şeklinde ifade edildiğinde, önerilen model için olabilirlik fonksiyonu, regresyon modelinin dayandığı temel gereği $y_i \sim N(\mu_i, \sigma^2)$ şeklinde ifade edilir.

Olabilirlik fonksiyonunun ardından a, b ve σ^2 parametreleri için önsel dağılımlar belirlenmelidir. Bu parametreler için bilgi vermeyen önsel dağılımlar seçilmek istenirse, sabit (a) ve regresyon katsayısı (b) için alanyazında sıkça kullanılan ortalaması 0 ve varyansı 100 gibi çok yüksek olan normal dağılımlar seçilebilir ve $a \sim N(0, 100)$, $b \sim N(0, 100)$ şeklinde gösterilir. Böylece analiz öncesinde sabit ve regresyon katsayılarına ait önsel tahminimizin ortalaması sıfırdır; ancak önsel dağılım çok büyük bir varyansa sahip olduğundan olası değerlerin ranjı oldukça geniş olmaktadır. Artıkların varyans parametresi (σ^2) için önsel dağılım belirlenirken dikkatli olmak gerekir; çünkü bir varyans değerinin pozitif olması şarttır. Bu durumda yine bilgi vermeyen bir önsel dağılım olarak, alanyazında varyans parametreleri için sıkça kullanılan tek biçimli dağılım kullanılabilir. Örneğin 0 ile 100 değerleri arasında tanımlanan (dolayısıyla sadece pozitif değerler alabilen) ve $\sigma^2 \sim Uniform(0, 100)$ şeklinde gösterilen bir dağılım uygun bir seçim olacaktır. Burada vurgulanması gereken bir diğer nokta da önsel dağılım seçimlerinde ilgili modeller için daha önce başka araştırmacılar tarafından önerilen ve etkililiği bilinen dağılımların tercih edilmesinin yararlı olacağıdır.

Modelin olabilirlik fonksiyonunun yazılması ve kestirilecek model parametrelerinin önsel dağılımlarının belirlenmesinin ardından, MCMC yöntemleri ile her bir parametrenin sonsal dağılımından örneklemler üretilmesi aşamasına geçilir. Bu aşamada MCMC algoritmaları araştırmacılar tarafından programlanacağı gibi, bunu otomatik olarak yapan ve ücretsiz olarak sunulan çeşitli istatistiksel yazılımların da kullanılabilmesi mümkündür. Örneğin WinBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000) ve JAGS – Just Another Gibbs Sampler (Plummer, 2003, 2015) gibi ücretsiz yazılımlar, MCMC yöntemlerinden Gibbs örneklemesi ile parametre kestirimlerinde yaygın olarak kullanılmaktadır. Yukarıdaki regresyon modelinin anılan bu yazılımlarda analiz edilmesinde kullanılabilecek model kodu şu şekildedir:

```
model {
  for (i in 1:n) {
  y[i] ~ dnorm(mu[i], tau)
  mu[i] <- a + b*x[i]
}
  tau <- 1/sigma
  a ~ dnorm(0, 0.01)
  b ~ dnorm(0, 0.01)
  sigma ~ dunif(0, 100)
  }
```

Model kodu incelendiğinde, aslında bunun yukarıda sıralanan aşamaların bir yansıması olduğu sezilecektir. Diğer bir ifadeyle, yazılan kod ile yukarıda sözel olarak açıklanan modelin olabilirlik dağılımının ve her bir parametreye ait önsel dağılımların ifade edilmesi söz konusudur. Model kodunda vurgulanması gereken noktalardan biri de, modelin olabilirlik dağılımı ve regresyon katsayılarının önsel dağılımları için kullanılan normal dağılımlarda, Bayes alanyazınında ele alındığı gibi varyans yerine varyansın tersinin (1/varyans) kullanılmış olmasıdır. İlgili veri setinin de programa tanıtılması ile birlikte, gereken Gibbs örneklemesi yazılımlar tarafından yapılacak ve sonuçta her bir parametre için sonsal dağılımlar elde edilecektir.

Elde edilen sonsal dağılımların yorumlanmasından önce, yazılımlar tarafından sunulacak çıktılardan yakınsamanın görsel ve sayısal kontrolü yapılmalıdır. Ayrıca yine yazılımlar tarafından sunulan ve önceki bölümde detaylı olarak açıklanan istatistikler ve ölçütler kullanılarak yakınsama sonrasında yeterli sayıda MCMC iterasyonunun yapıldığından emin olunmalıdır. Gerekli kontrollerin ardından elde edilen sonsal dağılımların ortalama değerleri parametrelerin nokta kestirimleri olacak; dağılımın standart sapmaları kestirimlerin standart hataları olup, yine yazılımlar tarafından elde edilecek olan HDI değerleri %95 güven aralığını ifade edecektir.

Hiyerarşik Bayes modellemesi. Bayes yaklaşımı ile kestirilmek istenen model parametreleri için kullanılan önsel dağılımlar da birtakım parametreler ile ifade edilmektedir. Örneğin yukarıda regresyon katsayısı parametresi için $b \sim N(0, 100)$ şeklinde gösterilen, ortalaması 0 ve varyansı 100 olan normal bir dağılım kullanılmıştır. Bu önsel dağılımdaki ortalama ve varyans değerleri de aslında bu normal dağılımı tanımlayan birer parametredirler. Hiyerarşik Bayes modellemede önsel dağılımların parametreleri bu örnekte olduğu gibi sabit değerler (0 ve 100) olarak alınmak yerine, bu parametreler için de birer önsel dağılım tanımlanır ve bunlar aşırı önsel dağılımlar (hyperprior) olarak ifade edilir. Önsel dağılımların parametreleri ise aşırı parametreler (hyperparameters) olarak isimlendirilmektedir. Hiyerarşik Bayes modellemesinde, normal Bayes modellemesinden farklı olarak iki aşamalı hiyerarşik bir yapı söz konusudur (Ntzoufras, 2009). Bu yapıyı daha somut biçimde görebilmek için, yine basit regresyon modeli üzerinden ilerlemek faydalı olacaktır.

Hiyerarşik Bayes modelleme ile regresyon katsayısının önsel dağılımındaki ortalama ve varyans değeri sabit birer sayı olarak alınmak yerine, bu parametreler için uygun aşırı önsel dağılımlar tanımlanır. Örneğin ortalama parametresi için ortalaması 0 ve varyansı 100 olan normal; varyans parametresi için ise 0-100 arasında değer alan tekbiçimli aşırı önsel dağılımlar tanımlanabilir. Bu durumda regresyon katsayısı için önsel dağılım $b \sim N(\mu, \sigma^2)$ şeklinde gösterilir. Burada yer alan μ ve σ^2 parametreleri, regresyon katsayısına ait önsel dağılımın aşırı parametreleridir. Bu aşırı parametrelerin önsel dağılımları ise $\mu \sim N(0,\ 100)$ ve $\sigma^2 \sim Uniform(0,\ 100)$ şeklinde gösterilir. Burada görüldüğü gibi, normal modellemede b için tek aşamada bir önsel dağılım tanımlanırken, hiyerarşik Bayes modellemede ise bu süreç iki aşamada tamamlanır.

Hiyerarşik Bayes modellemesinin temel avantajı, bir model parametresi için tanımlanan önsel dağılıma ilişkin bilginin, normal önsel dağılımlara göre daha muğlak olmasıdır. Çünkü önsel dağılımın kendi parametreleri de sabit değerlerden ziyade, seçkisiz birer değişken olmaktadırlar. Öte yandan, Hiyerarşik Bayes modelleme ile kestirilecek parametre değerinin, aşırı parametre değerine doğru çekilmesi (shrinkage) söz konusu olmaktadır (Kruschke, 2014). Özellikle küçük örneklemler ile analiz yapılması söz konusu olduğunda, bu çekilmenin parametre kestiriminde bir avantaj sağlayacağı belirtilmektedir (Ntzoufras, 2009). Bu avantajlara ek olarak, hiyerarşik Bayes modelleme ile parametre kestirimlerindeki öznelliğin azalacağı belirtilmektedir (Robert, 2007).

Bu çalışma kapsamında önerilen ÇDMTM'de, madde güçlük parametreleri için hiyerarşik bir modelleme yoluna gidilmiştir. Yani güçlük parametreleri için öncelikle bir önsel dağılım tanımlanmış; bu önsel dağılımı ifade eden parametreler için de ayrıca aşırı önsel dağılımlar tanımlanmıştır. Fox'un (2010) belirttiği gibi, madde parametreleri için aşırı önsel dağılımların kullanılması ile her bir güçlük parametresinin kestiriminde, diğer güçlük parametrelerinin kestirimlerinden gelen bilgiler de etkili olmakta ve kestirim hatalarında bir düşüş gözlenmektedir.

İlgili Araştırmalar

Yapılan alanyazın taraması sonucunda, doğrudan ÇDMTM'de grup içi varyansların heterojenliğinin ele alındığı herhangi bir çalışmaya rastlanmamıştır. ÇDMTM'nin model varsayımları bağlamında bir tarama yapıldığında, genel olarak farklı koşullar altındaki parametre kestirimlerinin incelendiği ve kestirim yöntemlerinin karşılaştırıldığı çalışmalara (ör: Dowling, 2006; Kim, 2007; Schmitt, 2007; Brune, 2011; Moyer, 2013) rastlanmıştır. Öte yandan alanyazında ÇDMTM'nin uygulamaya yönelik araştırma desenlerinde kullanıldığı çalışmalar (ör: Pastor, 2001, 2003; Fox, 2004; Natesan, 2007; Zheng, 2009; Sun, 2011) ile yanlılık ve test eşitleme gibi temel psikometri uygulamalarının yapıldığı çalışmalar (ör: Cheong, 2006; Turhan, 2006; Vaughn, 2006; Acar, 2008; Acar & Kelecioğlu, 2010; Patarapichayatham, Kamata, & Kanjanawasee, 2012; Atalay-Kabasakal, 2014) yoğun olarak yer almaktadır. Alanyazında bu çalışmayla ilişkili olan araştırmaların, çok düzeyli modellerdeki grup içi varyansların ve tekrarlı verilere dayalı boylamsal modellerdeki birey-içi varyansların heterojenliği bağlamında olduğu görülmüş ve aşağıda bu çalışmalara ilişkin özet bilgiler sunulmuştur.

HLM'de grup içi varyansların heterojenliği ile ilgili araştırmalar. Kasim ve Raudenbush (1998) iki düzeyli HLM'de grup içi varyans heterojenliğinin modellenebilmesi için MCMC yöntemlerinden Gibbs örneklemesinin uygulanmasına ilişkin bir çalışma yürütmüşlerdir. Bayes yaklaşımı çerçevesinde heterojen grup içi varyansların kestirildiği model ile homojenlik varsayımı yapan geleneksel HLM, yürütülen simülasyon çalışması ile karşılaştırılmıştır. Elde edilen sonuçlar ışığında araştırmacılar, geleneksel HLM'de ikinci düzey regresyon katsayılarının homojenlik varsayımının ihlaline karşı dayanıklı olduğunu belirtmişlerdir. Ancak heterojenliğin yüksek olduğu durumlarda geleneksel HLM'nin ortak (pooled) bir varyans bileşeni kestirmesinden dolayı, varyans bileşeni kestirimlerinde hatalı sonuçların söz konusu olabileceğini ve çok düzeyli modellerin parametre kestirimlerinde Bayes yaklaşımının avantajlı olacağını vurgulamışlardır.

Kim (2005), geleneksel HLM'nin grup içi varyans homojenliği varsayımının özellikle eğitim bilimleri ile ilgili araştırmalarda çoğu zaman karşılanamayacağını belirterek, deneysel ve yarı deneysel desenlerde HLM'nin heterojen grup içi varyanslar kestirecek şekilde modellenmesi üzerinde çalışmıştır. Yayılmanın modellenmesi (dispersion modeling) şeklinde nitelendirdiği yaklaşımda Kim, grup içi heterojen varyansları yalnızca modellemekle kalmamış, ayrıca bu varyansları bir bağımlı değişken olarak ele alıp, öğretmen özellikleri ve okul politikaları gibi grup düzeyinde bağımsız değişkenler ile yordamaya çalışmıştır. Kim, yeni modelin veriye geleneksel HLM'den daha çok uyum sağlayacağının altını çizmiştir. Parametre kestirimlerinde Bayes yaklaşımını kullanmış ve önerdiği modelin diğer araştırmacılar tarafından genişletilebilmesine ilişkin öneriler sunmuştur.

HLM'deki grup içi varyansların heterojenliğinin ele alındığı bir diğer çalışma Kim ve Choi (2008) tarafından yapılmıştır. Bu araştırmada özellikle okullar içerisindeki başarı düzeyi farklılığına odaklanılmış ve okul içi varyansların, okulların etkililiği açısından önemli bilgiler vereceği vurgulanmıştır. Okulların etkililiğinin yalnızca yüksek öğrenci başarısı ile değil, okul içerisindeki başarı farklılıklarının da az olması ile ifade edilebileceği belirtilmiştir. Bu bağlamda, okullar içerisindeki varyansların heterojen olduğu HLM, TIMMS-R veri setine uygulanmış ve heterojen okul içi varyanslar ile ilişkili olabilecek okul düzeyi değişkenler modele alınmıştır. Parametre kestiriminde Bayes yaklaşımına dayalı olarak MCMC yöntemi kullanılmıştır. Veri analizi sonucunda varyans homojenliği varsayımının karşılanmadığı görülmüş ve heterojen varyans modeline göre ortalama okul başarısı ve okullarda tahtanın sık kullanılmasına ilişkin değişkenlerin okul içi varyansları açıkladığı görülmüştür. Ayrıca, okul ortalaması kontrol edildiğinde tahta kullanımının okullar arasındaki başarı farklılıklarını azaltacağına ilişkin sonuçlara ulaşılmıştır.

Birey-okul formundaki iki düzeyli modellerde varyans heterojenliğinin uygulamaya yönelik olarak kullanıldığı bir diğer çalışma Sani ve Grilli (2011) tarafından yapılmıştır. Bu araştırmacılar da Kim ve Choi (2008) ile paralel biçimde, okul içi varyansların değişkenlik göstermesinin öğretimin etkililiği ve buna ek olarak eğitimde fırsat eşitliği açısından istenen bir durum olmadığını vurgulamışlardır. Dolayısıyla, her bir okul içerisindeki artık varyansın eş olduğunu varsayan geleneksel HLM yerine, varyansların heterojenliğine izin veren ve bunun birtakım değişkenler ile açıklanmasına olanak sağlayan modeli kullanmışlardır. INVALSI (Italian National Institute for the Evaluation of the School System) tarafından sağlanan verilerin kullanıldığı analizler sonucunda, öğrencilerin matematik başarılarının bağımlı değişken olarak alındığı modelde okul içi artık varyansların cinsiyete göre değişkenlik gösterdiği bulgusuna ulaşılmıştır. Öte yandan, okul içi artık varyansların ayrıca okul düzeyi değişkenler ile açıklanabildiği ve özellikle güneydeki bölgelerde yer alan okulların daha büyük artık varyansa sahip olduğu, yani bu okullardaki başarı farklılıklarının daha fazla olduğu belirtilmiştir.

HLM'de varyans heterojenliğinin eğitim ve davranış bilimlerinde önemli bir olgu olduğunu vurgulayan Leckie, French, Charlton ve Browne (2014), heterojenliği modelleyebilen ve güncel alanyazında yer alan iki düzeyli genişletilmiş HLM yaklaşımlarını incelemişlerdir. Araştırmacılar ayrıca varyans heterojenliğinin olası sonuçlarına ilişkin bir simülasyon çalışması yürütmüş ve son olarak farklı heterojen modellemelere ilişkin olanakları gerçek bir veri seti üzerinden açıklamışlardır. Yapılan simülasyon çalışması sonucunda, birinci düzeydeki heterojenliğin ihmal edilmesinin özellikle birinci düzey varyans fonksiyonundaki regresyon katsayılarının sorunlu kestirimine yol açacağını belirtmişlerdir. Uygulamaya yönelik sonuçlarda ise, okulların akademik başarının değişkenliği bağlamında farklılık gösterdiği ve okul sosyo-ekonomik düzeyinin kontrol edilmesine rağmen bu değişkenliğin devam ettiği rapor edilmiştir. Son olarak heterojen modellerin analizinde ML tabanlı kestirimlerde veri uyumunun zorluğuna değinilmiş ve alternatif olarak Bayes yaklaşımına dayalı kestirimlerin daha esnek olacağı vurgulanmıştır.

Son olarak Kuppens ve Yzerbyt (2014) de diğer araştırmacılar gibi heterojenliğin sadece bir varsayım ihlali olarak görülmesinden ziyade, genişletilmiş HLM'ler ile ele alınabileceğini ve bu tür modellerin uygulamaya yönelik önemli bilgiler vereceğini belirtmişlerdir. Araştırmacılar özellikle heterojen HLM'lerin değişik paket programlar ile nasıl analiz edilebileceğine dair detaylı bilgiler vermişler ve uygulamada heterojen modelin, homojen modele göre gerçek verilere daha iyi uyum sağladığına ilişkin kanıtlar sunmuşlardır.

Boylamsal modellerde varyansların heterojenliği ile ilgili araştırmalar. Heterojen varyans bileşenlerinin modellenmesine bireylere ilişkin tekrarlı verilerin ele alındığı boylamsal çalışmalarda da rastlanmaktadır. Bu boylamsal çalışmalar aslında ölçüm-birey formunda iki düzeyli hiyerarşik modellerdir. Heterojen boylamsal modellerde bireylerin tekrarlı ölçümlerine ait artık varyansların tüm bireylerde eşit olarak kabul edilmesi yerine, her bir birey için ayrı bir artık varyansın kestirilmesi yoluna gidilmektedir.

Zhang ve Weiss (2000) tekrarlı ölçümlere dayalı çok düzeyli modellerde, birinci ve ikinci düzey varyans heterojenliği ile çok değişkenli varyans heterojenliğine ilişkin bir çalışma yapmışlardır. Araştırmacılar, önceki bölümde değinilen yaygın yaklaşımı benimseyerek, varyans heterojenliğinin bir varsayım ihlalinden ziyade yordayıcı değişkenler tarafından açıklanabilecek bir durum olduğunu vurgulamışlardır. Üç türdeki (birinci düzey, ikinci düzey ve çok değişkenli) varyans heterojenliğinin tespit edilebilmesi için istatistiksel yöntemler sunan araştırmacılar, ayrıca artık grafikleri ile heterojenliğin irdelenmesine ilişkin bilgiler sunmuşlardır.

Öte yandan Browne, Draper, Goldstein ve Rashbash (2002) her bir birey için ayrı bir varyans bileşeninin yalnızca modellenmesiyle yetinmeyip, bu bileşenlerin giriş puanı, cinsiyet gibi değişkenler tarafından yordanmasına ilişkin bir çalışma yürütmüşlerdir. Araştırmacılar yaklaşımlarını “karmaşık birinci düzey değişkenlik (complex level-1 variation)” şeklinde isimlendirmişlerdir. Frekansa ve Bayes yaklaşımına dayalı 4 farklı kestirim yönteminin simülasyon çalışmasıyla karşılaştırılmasının ardından, genel anlamda tüm yöntemlerin iyi sonuçlar verdiğini rapor etmişlerdir. Ancak heterojenliğin polinomal veya logaritma ölçeğinde modellenmesinde Bayes yaklaşımı yöntemlerinden adaptif Metropolis-Hastings örneklemesinin, adaptif red (adaptive rejection) örneklemesinden daha verimli olduğunu vurgulamışlardır.

Boylamsal çok düzeyli modellerde varyans heterojenliğini ele alan bir diğer çalışma Hoffman (2007) tarafından yapılmıştır. Bu çalışmada bireylerin ruh hallerine ilişkin tekrarlı verilerden oluşan iki düzeyli modelde öncelikle varyans homojenliği varsayımının ihlali, ilgili testler ile ortaya konulmuştur. Daha önce Zhang ve Weiss’ın (2000) da vurguladığı gibi, Hoffman da heterojenliğin bir varsayım ihlali ve istatistiksel modeller için düzeltilmesi gereken bir sorun olarak görülmesinden ziyade, grup içi değişkenliğin başka değişkenler tarafından kaynaklandığının üzerinde durulması gerektiğini ve bunun araştırmacılar için oldukça önemli bilgiler sağlayacağını belirtmiştir. Çalışma kapsamında Hoffman, bireylerin ruh hallerine ilişkin tekrarlı ölçümlerine ait heterojen varyans değerlerini, bireylerin zihinsel işleyişlerindeki değişkenlik ile açıklamaya çalışmıştır. Araştırma sonucunda, zihinsel işleyiş bakımından yüksek değerler gösteren bireylerin, pozitif ruh hali bağlamında daha az birey-içi değişkenlik gösterdiğini raporlamıştır.

Boylamsal modellerde heterojenliğin hem birey içi hem de bireyler arasında modellenebilmesine olanak sağlayan bir başka model Hedeker, Mermelstein ve Demirtaş (2008) tarafından önerilmiştir. Araştırmacılar bu modeli mixed-effects location scale model (MLSM) şeklinde isimlendirmişlerdir. Önerilen bu model, daha önceki araştırmalarda olduğu gibi, bireylerden alınan tekrarlı ölçümlerin birey içi artık varyanslarının geleneksel modellerden farklı olarak heterojen olmasına olanak sağlamaktadır. Bu avantajının yanında model, bireyler arasındaki varyansın da heterojen olmasına ve her iki düzeydeki heterojenliğin açıklanması için ölçüm ve birey düzeylerinde açıklayıcı değişkenlerin modele alınabilmesine olanak tanımaktadır. Model ayrıca boylamsal çalışmalarda hem ortalama hem de varyans yapısının eşzamanlı olarak ele alınmasına da olanak sağlamakta ve özellikle bu yönüyle diğer heterojen boylamsal modellerden daha gelişmiş bir yapıya sahip olmaktadır. Araştırmacılar bu modeli bireylerden 30-40 gibi çok fazla sayıda anlık ölçümlerin alındığı ve alanyazında “ecological momentary assessment (EMA)” şeklinde nitelendirilen boylamsal veri toplama uygulamaları için uygun olduğunu belirtmişlerdir. Çalışmanın akışı içerisinde araştırmacılar, önerdikleri modelin istatistiksel temelini sunmuş ve kestirimlerin SAS yazılımı (SAS Institute, 2000) yardımıyla ML ile yapılabileceğini belirtmişlerdir. Çalışmanın son aşamasında önerilen model sigara içme davranışı ile ilgili EMA uygulamasıyla toplanan gerçek verilere uygulanmıştır. Gerçek veri analizi sonucunda, negatif ruh hali ölçümlerinde birey içi varyansların oldukça değişken olduğu ve yine negatif ruh hali bakımından daha yüksek ortalamaya sahip bireylerin daha değişken bir yapıya sahip olduğu gibi birçok somut bulgu rapor edilmiştir.

Alanyazında MLSM’nin ele alındığı birçok çalışma yer almaktadır. Jahng (2008) yaptığı araştırmada MLSM’nin iki aşamalı biçimini önermiştir. İki aşamalı bu modelin geleneksel MLSM’den temel farkı, ortalama ve varyans yapılarının eşzamanlı olarak değil, iki farklı aşamada modellenmesidir. Jahng yaptığı simülasyon çalışması sonucunda, iki aşamalı modelin daha iyi bir yakınsama performansı gösterdiğini ve yansız kestirimler sunduğunu belirtmiştir. Li ve Hedeker (2012) ise iki düzeyli MLSM’yi, tekrarlı ölçümlerin günler içerisinde yuvalandığı ölçüm-gün-birey olacak şekilde üçüncü düzeye genişletmişlerdir. Araştırmacılar bu üç düzeyli modelin iki düzeyli modele göre daha esnek olduğunu ve günlere göre olası sistematik değişkenliği dikkate alabildiğini vurgulamışlardır. Modelin ML tabanlı parametre kestirim sürecine ilişkin detaylı bilgilerin verilmesinin ardından, kestirimlerin SAS yazılımı (SAS Institute, 2000) ile yapılabileceğini belirtmişlerdir. Yürütülen simülasyon çalışması sonrasında modelin, önerilen ML algoritması ile etkili kestirimler yapabildiği rapor edilmiş ve çalışmanın son bölümünde gerçek EMA veri seti üzerinden modelin uygulamasına ilişkin bilgiler sunulmuştur. Pugach (2012) ise iki düzeyli MLSM'yi iki bağımlı değişkenin eşzamanlı olarak modellenebileceği şekilde genişleterek iki değişkenli MLSM'yi geliştirmiştir. Önerilen model öncelikle gerçek veri setine uygulanmış ve modelin kullanışlılığı gerçek bulgular yardımıyla somutlaştırılmıştır. Yürütülen simülasyon çalışması sonucunda, iki değişkenli veri durumlarında önerilen modelin geleneksel MLSM'ye göre daha avantajlı olduğu rapor edilmiştir.

MLSM ile uygulama yapan Rast ve Zimprich (2011) yaptıkları çalışmada bireylerin reaksiyon sürelerindeki değişkenliği incelemişlerdir. Analizleri SAS yazılımı (SAS Institute, 2000) yardımıyla yürüten araştırmacılar, MLSM'nin birey-içi varyansların heterojenliğinin modellenmesindeki etkililiğini teyit etmişlerdir. Analizler sonucunda özellikle yaşlı katılımcıların reaksiyon süreleri bağlamında daha değişken olduğunu raporlamışlardır. Öte yandan MLSM'nin diğer heterojen boylamsal modellerden farklı olarak hem ortalama hem de varyans yapısını eşzamanlı olarak kestirip hem de bunların ilişkili olmasına izin vermesinin önemine değinmişlerdir. Modelin bu avantajını kullanarak ortalama birey reaksiyon süresi ile birey içi reaksiyon süresi değişkenliğinin yüksek korelasyon gösterdiğine ilişkin bulgular elde etmişlerdir. Böylece daha düşük reaksiyon sürelerinin, yüksek varyanslarla ilişkili olduğunu ve bu durumun heterojenlikten kaynaklı olduğunu vurgulamışlardır.

Bir diğer MLSM uygulaması ise Bayes yaklaşımı çerçevesinde Rast, Hofer ve Sparks (2012) tarafından yapılmıştır. Araştırmacılar MLSM'yi, bireylerin duygu durumlarına ilişkin alınan yedi günlük tekrarlı ölçümlere uygulamışlardır. Modeldeki birey içi artık varyanslar log-normal dağılım ile modellenmiş ve hiyerarşik Bayes yaklaşımı kullanılarak parametreler MCMC yöntemiyle kestirilmiştir. Heterojenliğin açıklanması için bağımsız değişkenlerin de alındığı koşullu modellerin analizi sonucunda, günlük stres düzeylerinin fazla olduğunu belirten bireylerin, duygu durumu bakımından daha değişken oldukları bulgusuna ulaşılmıştır. Ancak bu durumun, sadece düşük düzeydeki ortalama duygu durumuna sahip ve düşük ortalama negatif duygu durumu gösteren bireyler için geçerli olduğunu belirtmişlerdir. Son olarak araştırmacılar Bayes yaklaşımı ile kestirimlerin yapılmasının ML tabanlı kestirimler yapan yazılımlarda sıkça yaşanan sorunların önüne geçtiğini vurgulamışlardır.

Boylamsal modellerdeki varyans heterojenliğinin ele alındığı çalışmalara sağlık bilimleri alanında da rastlanmaktadır. Myles, Price, Hunter, Day ve Duffy (2003) bireylerin günlük retinol alımıyla ilgili tekrarlı ölçümlerinin olduğu çok düzeyli verilerin analizinde, her bir birey için ayrı bir varyans bileşeni kestirmişlerdir. Araştırmacılar, bireylerin günlük değişimlerini ifade eden birinci düzey varyansların tüm bireyler için eşit olmasının anlamlı olmayacağını ve heterojen varyansların geleneksel çok düzeyli modeller ile kestirilemeyeceğini vurgulamışlardır. Bu bağlamda geleneksel çok düzeyli modeli genişleterek varyans heterojenliğine izin veren karmaşık bir model önermişler ve parametre kestirimlerini MCMC yöntemleri ile yapmışlardır. Araştırma sonucunda, yeni modelin daha açıklayıcı kestirimler sunduğunu ve özellikle önerilen maksimum retinol alımını aşan günlerin yüzdesini geleneksel modele göre daha iyi tespit ettiğini vurgulamışlardır.

Çok düzeyli modellerde varyans heterojenliğine ilişkin yukarıda özetlenmeye çalışılan araştırmaların çoğunda, varyans heterojenliğinin bir varsayım ihlali olarak görülmesinin doğru bir yaklaşım olmadığının vurgulandığı görülmektedir. Bunun aksine heterojenliğin de modellenmesi, eğitim ve davranış bilimleri açısından “grup içindeki bireylerin değişkenliği” veya “bireylerin durumlarındaki değişkenlik” gibi aslında özellikle üzerinde durulması gereken noktaların dikkate alınabilmesine olanak sağlamaktadır. Alanyazında yer alan bu araştırmalar da genel anlamda bu yaklaşımı benimseyerek farklı grup ve birey içi varyansları modelleme ve ayrıca bu varyans bileşenlerinin birtakım değişkenler tarafından açıklanmaya çalışılması şeklinde bir yaklaşım izlemişlerdir. Bu çalışmada da aynı yaklaşım benimsenerek, geleneksel ÇDMTM’deki grup içi varyansların homojen olduğu varsayımına gerek kalmaksızın her bir grup için ayrı bir varyans bileşenin kestirilmesi için yeni bir model önerilmiştir. Böylece, yetenek kestirimleri bakımından grupların kendi içlerindeki değişkenliklerinin incelenebilmesine ve bazı durumlarda bu değişkenliğin çeşitli yordayıcı değişkenler ile açıklanabilmesine olanak sağlanmıştır.

Bölüm 3
Yöntem

Araştırmanın Türü

Bu çalışmada birey-madde-grup formundaki geleneksel ÇDMTM, heterojen grup içi varyanslar kestirecek şekilde yeniden modellenmiştir. Önerilen modelin matematiksel eşitlikleri verilmiş, parametre kestirim yöntemi hakkında gerekli bilgiler sunulmuş ve modelin değişen grup içi heterojenlik düzeylerindeki parametre kestirim performansı yürütülen simülasyon çalışması ile test edilmiştir. Simülasyon çalışmasının ardından önerilen ve geleneksel model gerçek bir veri setine uygulanarak, modellerin parametre kestirimleri ve model-veri uyumları karşılaştırılmıştır. Dolaysıyla bu çalışma kurama katkı sağlayacak nitelikte olup temel bir araştırma özelliği taşımaktadır.

Modelin Tanımlanması: Matematiksel Eşitlikler ve Varsayımlar

Çalışma kapsamında önerilen yeni model, madde-birey-okul şeklindeki yapıya sahip ÇDMTM'nin, heterojen okul içi yetenek varyansları kestirecek şekilde genişletilmesi ilkesine dayanmaktadır. HLM çerçevesinde ele alındığında aslında üç düzeyli geleneksel model, yeteneklerin bağımlı değişken olduğu ve bireylerin okullarda yuvalandığı iki düzeyli bir model olarak yorumlanabilir. Bu bağlamda burada sunulacak model eşitliklerinde HGLM yaklaşımını benimseyen Kamata'dan (1998) ziyade Fox (2001) tarafından tercih edilen modelleme yaklaşımı benimsenmiştir. Araştırmanın kuramsal temelinde belirtildiği gibi, bu yaklaşımda genel model, ölçme modeli ve yapısal çok düzeyli model şeklinde iki kısımda ele alınmaktadır.

Ölçme modeli. Üç düzeyli modelde bağımlı değişkenler, diğer bir ifadeyle ele alınan veriler, bireylerin maddelere verdikleri 1-0 şeklindeki yanıtlardır. Bu durumda ilk aşama, genel modelin ölçme modeli kısmının tanımlanmasıdır. Bu çalışmada önerilen modelde, ölçme modeli olarak geleneksel Rasch model kullanılmıştır. Rasch modelde bireylerin maddelere verdikleri yanıtların Bernuolli dağılımı gösterdiği varsayılır. Bu durumda yanıt örüntülerinin dağılımı, y_{ijk} k okulundaki j bireyinin i maddesine verdiği yanıt olmak üzere, $y_{ijk} \sim Bernuolli(p_{ijk})$ şeklinde gösterilir. Yanıt örüntülerine ilişkin Bernuolli dağılımının beklenen değeri olan p_{ijk} terimi ise k okulundaki j bireyinin i maddesine doğru yanıt verme olasılığıdır. Rasch modelin logit bağlantı fonksiyonu kullanılarak doğrusal şekilde ifade edildiği (3) numaralı eşitlikteki yaklaşım kullanıldığında,

$$\ln\left(\frac{p_{ijk}}{1-p_{ijk}}\right) = \theta_{jk} - b_i \tag{25}$$

eşitliği elde edilir. Bu eşitlikte yer alan θ_{jk} terimi, k okulundaki j bireyinin yetenek parametresidir. b_i terimi ise i maddesinin güçlük parametresini ifade etmektedir. Güçlük parametresinde birey veya okula ait herhangi bir indis olmaması, madde güçlüklerinin MTK'nın temel varsayımına uyacak şekilde değişmez olduğunu belirtmektedir. Ölçme modelinin tanımlanabilmesi için, bireylerin yetenek parametrelerinin (θ_{jk}) ortalaması sıfıra sabitlenmiştir.

Çok düzeyli yapısal model. Önerilen modelin ölçme modeli kısmının tamamlanmasının ardından, bu sefer yapısal model ile yetenek parametrelerinin birey ve okul düzeylerinde ayrıştırılması gerekmektedir. Bunun için Fox'un (2001) çalışmasında olduğu gibi iki düzeyli koşulsuz HLM çerçevesinde bir yol izlenmiştir. Tıpkı HLM'de olduğu gibi, θ_{jk} sürekli bağımlı değişken olmak üzere ilk düzey doğrusal eşitlik,

$$\theta_{jk} = u_k + e_{jk} \tag{26}$$

şeklinde ifade edilir. Burada u_k, k okulunun yetenek parametresi iken, e_{jk} terimi ise k okulunda yer alan j bireyinin kendi okul ortalamasından sapmasını (birey düzeyindeki artık değeri) ifade eder. Ölçme modeli bölümünde, modelin tanımlanabilmesini sağlamak için bireylerin yeteneklerinin ortalamasının sıfır olarak alındığı belirtilmişti. Dolayısıyla Fox'un (2001) çalışmasında olduğu gibi u_k teriminin bir üst düzey eşitlik ile sabit ve seçkisiz bileşenlere ayrılmasına gerek duyulmaz. Diğer bir ifadeyle, tüm birey düzeyi yeteneklerin ortalaması sıfıra sabitlendiğinden, okul yetenek ortalamaları da sıfıra eşit olacaktır. Bu açıdan okul yeteneklerine ait ikinci düzey doğrusal bir eşitlik yerine, okul düzeyi yeteneklerin doğrudan gösterdikleri dağılım ele alınır. Burada, Kamata (1998) tarafından modellenen geleneksel ÇDMTM'de olduğu gibi okul düzeyi yeteneklerin sıfır ortalama ve belirli bir varyans değeri ile normal dağılım gösterdiği varsayılmıştır. Bu durum $u_k \sim N(0, \sigma_u^2)$ şeklinde ifade edilir.

Bu çalışmada önerilen yeni modelin geleneksel ÇDMTM'den farklılaştığı nokta, bireylerin okul ortalamalarından olan sapmalarını ifade eden e_{jk} terimleriyle ilgilidir. Bu artık yetenek terimlerinin (okul ortalamasından arındırılmış yetenekler) yine sıfır ortalama; ancak her bir okul için ayrı bir varyans ile normal dağılım gösterdiği varsayılmıştır ve bu durum $e_{jk} \sim N(0, \sigma^2_{e_k})$ şeklinde ifade edilir. Okul içi artıklara ait varyans teriminde (k) indisinin olması, birinci düzey varyansın homojen olarak kabul edilip tüm okullar için eşit olmasından ziyade, ayrı birer parametre olarak kestirileceğini ifade etmektedir.

Ölçme ve yapısal model eşitliklerinin birleştirilmesi ile genel model eşitliği $u_k \sim N(0, \sigma^2_u)$ ve $e_{jk} \sim N(0, \sigma^2_{e_k})$ olmak üzere,

$$\ln\left(\frac{p_{ijk}}{1 - p_{ijk}}\right) = \theta_{jk} - b_i = \left(u_k + e_{jk}\right) - b_i \tag{27}$$

şeklinde ifade edilir.

Varyansların açıklandığı koşullu model. Önceki bölümde sunulan çok düzeyli yapısal modelde, her bir okul için ayrı olarak kestirilecek okul içi varyans bileşenleri ($\sigma^2_{e_k}$) söz konusudur. Kestirilen okul içi varyansların okullar arasında belirgin bir değişkenlik göstermesi halinde, bu değişkenliğin nedenlerinin araştırılması için okul veya birey düzeyinde bağımsız değişkenler varyans modeline dâhil edilebilir. Aslında burada belirtilen durum, tıpkı geleneksel ÇDMTM'de olduğu gibi, farklılaşan okul yeteneklerinin açıklanması için modele bağımsız değişkenlerin alındığı koşullu model ile benzer mantıktadır. Aradaki farklılık okul yetenek ortalamalarından ziyade, okul içi yetenek varyanslarının bağımsız değişkenler ile açıklanmaya çalışılmasıdır.

Anılan koşullu varyans modelinin kurulabilmesi için, öncelikle kestirilen okul içi varyans bileşenlerinin her zaman pozitif olduğunun göz önünde bulundurulması gerekir. Dolayısıyla, doğrusal bir model kurulması halinde bağımlı değişken olacak okul içi varyans bileşenlerinin pozitifliğinin korunması şarttır. Kim ve Choi (2008), karekök dönüşümünü kullanarak, varyanslar yerine standart sapmaları doğrusal olarak modellemişlerdir. Raudenbush ve Bryk (2002) ise modelleme için logaritma dönüşümünü önermişlerdir ve bu çalışmada da bu dönüşüm kullanılmıştır. Önceki bölümde $\sigma_{e_k}^2$ terimi ile ifade edilen okul içi varyans bileşenleri logaritma dönüşümü ile aşağıdaki şekilde koşullu olarak modellenebilir.

$$\ln\left(\sigma_{e_k}^2\right) = \alpha_0 + \sum \alpha_m C_m \tag{28}$$

Eşitlikte yer alan C_m terimi, koşullu varyans modeline alınabilecek m sayıda birey veya okul düzeyindeki bağımsız değişkenleri ifade etmektedir. Sonuç olarak koşullu varyans modeli ile birlikte, her bir okul için ayrı bir varyans parametresi kestirilmesinin yanında, bu varyansların heterojenliğinin olası kaynakları da belirlenmeye çalışılır.

Modelin parametre kestirim yaklaşımı. Önerilen model kapsamında farklı okul içi varyans parametrelerinin varlığının kabul edilmesi, bu parametrelerin frekansa dayalı yaklaşımlarda olduğu gibi sabit değerler olarak görülmesi anlayışıyla tam olarak uyuşmamaktadır. Dolaysıyla bu modelde farklı okul içi varyans değerlerinin belli bir dağılım gösteren seçkisiz parametreler olarak ele alınması ve bu anlayışa göre kestirim yapan Bayes yaklaşımının benimsenmesi uygun görülmüştür. Burada okul içi varyansların dağılımları, Kasim ve Raudenbush'un (1998) çalışmasında olduğu gibi Ters-Gamma (Inverse-Gamma) olarak belirlenmiştir. Önceki bölümlerde açıklandığı gibi, Bayes yaklaşımında varyansların tersinin modellenmesi genel bir uygulama olduğundan, bu durum okul içi varyansların tersinin Gamma dağılım göstermesine tekabül eder. Burada vurgulanması gereken bir nokta da, Gamma dağılımı ile varyansların terslerinin, dolayısıyla da varyansların mutlaka pozitif değerler almasının garantilenmiş olmasıdır.

Gamma dağılımı sırasıyla şekil (shape) ve ölçek (scale) parametreleri ile tanımlanmaktadır. Bu çalışmada okul içi varyans bileşenlerinin terslerine ilişkin dağılım Kasim ve Raudenbush'un (1998) çalışmasında olduğu gibi $\sigma_{e_k}^{-2} \sim \mathrm{Gamma}(\frac{1}{2\tau}, \frac{2\tau}{\mu})$ şeklinde ifade edilmiştir. Dağılımının şekil parametresi $\frac{1}{2\tau}$ ve ölçek parametresi $\frac{2\tau}{\mu}$ değerine eşit iken μ ve τ terimleri sırasıyla okul içi varyansların ortalama ve varyans değerlerine eşit olmaktadır. Burada okul içi varyansların heterojenlik derecesini τ terimi temsil etmektedir. τ'nın yüksek olması halinde, okul içi varyansların büyük bir değişkenlik gösterdiği, dolaysıyla heterojenliğin fazla olduğu yorumu yapılır. Bunun aksine τ'nın sıfıra yaklaşması ise okul içi varyansların birbirlerine (ve dolayısıyla μ değerine) yakın ve homojen oluğunu belirtir. Şekil ve ölçek parametrelerinin bu şekilde ifade edilmesiyle Gamma dağılımının ortalama ve varyans değerlerinin tıpkı normal dağılımda olduğu gibi pratik bir şekilde tespit edilmesi sağlanır. Diğer bir ifadeyle, okul içi varyansların eğilim ve yayılım ölçülerinin birer parametre ile ifade edilebilmesi için Gamma dağılımının parametreleri bu şekilde ele alınmaktadır.

Önerilen model kapsamında kestirilebilecek tüm parametreler (koşulsuz varyans modeli durumunda) sırasıyla i tane madde güçlük parametresi (b_i), okul içi varyansların ortalama değeri (μ), okul içi varyansların varyans değeri (τ), k tane okul içi varyans değeri $(\sigma_{e_k}^2)$, okul yeteneklerinin varyans değeri (σ_u^2), k tane okul yetenek parametresi (u_k) ve son olarak j tane birey yetenek parametresi (θ_{jk}) şeklindedir.

Bayes yaklaşımı ile kestirimlerde model parametreleri için önsel dağılımların ve hiyerarşik Bayes modellemesinde ayrıca aşırı önsel dağılımların belirlenmesi gerekmektedir. Bu modellemede madde güçlükleri için $b_i \sim N(\mu_b, \tau_b)$ şeklinde hiyerarşik normal bir dağılım tanımlanmıştır. Bu dağılıma ait ortalama ve varyansın tersi aşırı parametreleri için sırasıyla $\mu_b \sim N(0, 0.01)$ ve $\tau_b \sim Gamma(0.1, 0.1)$ şeklinde bilgi vermeyen aşırı önsel dağılımlar tanımlanmıştır. Okul yeteneklerinin varyansının tersi için yine bilgi vermeyen $\sigma_u^{-2} \sim Gamma(0.1, 0.1)$ önsel dağılımı tanımlanmıştır. Okul içi varyans değerlerinin tersi için yukarıda da ifade edilen hiyerarşik yapıya sahip $\sigma_{e_k}^{-2} \sim Gamma(\frac{1}{2\tau}, \frac{2\tau}{\mu})$ şeklindeki Gamma dağılımı tanımlanmıştır. Bu dağılıma ait ortalama ve varyans aşırı parametreleri için sırasıyla $\tau \sim Uniform(0, 100)$ ve $\mu \sim Uniform(0, 100)$ şeklinde bilgi verici olmayan tekbiçimli önsel dağılımlar tanımlanmıştır.

Yukarıda sunulan parametre kestirimlerine ilişkin bilgiler, okul içi varyansların herhangi bir değişken tarafından yordanmadığı koşulsuz varyans modeline ilişkindir. Ancak daha önceki bölümde belirtildiği gibi, olası bir heterojenliğin nedenlerinin incelenmesi için grup içi varyansları açıklamak üzere bağımsız değişkenlerin devreye sokulduğu koşullu varyans modelleri kurulabilir. Bu durumda $\sigma_{e_k}^2$ ile ifade edilen okul içi varyans değerlerinin Ters-Gamma dağılımı ile koşulsuz olarak modellenmesi yerine, (28) numaralı eşitlikte ifade edilen biçimde doğrusal bir model genel modele dâhil dilebilir. Böylece, kestirilecek parametrelere doğrusal modeldeki sabit katsayısı olan α_0 ile m tane yordayıcıya ilişkin katsayılar olan α_m regresyon katsayıları da dâhil olacaktır. Bu katsayılar için de uygun önsel dağılımların belirlenmesi gerekir. Bu çalışmada gerçek veri uygulamasında kurulan koşullu varyans modeli için kullanılan önsel dağılımlar $\alpha_0 \sim Uniform(-2.5, 0)$ ve okul düzeyinde alınan tek yordayıcı değişken için $\alpha_1 \sim Uniform(-0.5, 0)$ olarak belirlenmiştir. Modellemede logaritma dönüşümü kullanıldığı için, koşullu varyans modelindeki bu regresyon katsayılarına ilişkin önsel dağılımların belirlenmesinde dikkatli olunması gerektiğini vurgulamakta fayda vardır. Uygun olmayan önsel dağılımların seçilmesi halinde, aşırı uç değerlerin logaritmalarındaki hesaplama zorluklarının yol açtığı yakınsama problemleri ile karşılaşılabilmektedir.

Önerilen modelin tüm parametreleri Bayes yaklaşımı çerçevesinde MCMC yöntemlerinden Gibbs örneklemesi ile kestirilmektedir. Bu işlem için Plummer (2003, 2015) tarafından geliştirilen ve ücretsiz olarak sunulan JAGS yazılımı kullanılmış ve bu yazılım runjags (Denwood, 2016) isimli R (R Core Team, 2016) paketi ile kontrol edilmiştir. JAGS yazılımı modele ilişkin eşitliklerin, dağılım varsayımlarının ve tüm önsel ve aşırı önsel dağılımların belirlenmesinin ardından bu süreci betimleyen bir kod ile çalışmaktadır. Varyansların koşulsuz ve koşullu olarak kestirildiği modellere ait JAGS kodları EK-A ve EK-B'de sunulmuş olup, bu kodlar ayrıca WinBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000) yazılımında da kullanılabilmektedir. JAGS yazılımı, modelin olabilirlik fonksiyonunu, önsel dağılımlarını ve kendisine tanıtılan veri setini ele alarak her bir parametrenin marjinal sonsal dağılımından Gibbs örneklemesi ile değerler üretir. Çalışmanın kuramsal bölümünde detaylı olarak açıklandığı gibi, her bir parametre için elde edilen sonsal dağılımlar üzerinden parametre kestirimi, standart hata ve %95 HDI değerleri elde edilir.

Simülasyon Çalışması

Önerilen modelin değişen grup içi heterojenlik düzeylerindeki parametre kestirim performansının incelenebilmesi için, sadece heterojenlik düzeyinin değişimlendiği bir simülasyon çalışması planlanmıştır. Simülasyon çalışmasında temel model olan grup içi varyansların koşulsuz olarak kestirildiği model dikkate alınmıştır. Bunun için Kasim ve Raudenbush (1998) tarafından kullanılan τ değerleri ile düşük ($\tau = 0.02$), orta ($\tau = 0.1$) ve yüksek ($\tau = 0.2$) grup içi heterojenlik düzeyleri gösteren koşullar için veriler üretilmiştir. Yorumlama kolaylığı açısından gruplar, bireylerin kümelendiği okullar olarak ele alınmıştır. Yalnızca heterojenliğin etkisinin incelenebilmesi için madde sayısı, okul sayısı ve her bir okul içerisinde yer alan birey sayısı optimum değerler olarak belirlenmeye çalışılmıştır. Burada ifade edilen optimum değerler, etkili parametre kestirimleri için diğer araştırmacılar tarafından önerilen/benimsenen en uygun değerlerdir. Bu bağlamda madde sayısı, eğitimde ölçme ve değerlendirme alanında yaygın optimum değer olan 30 olarak belirlenmiştir. Diğer bir ifadeyle, MTK uygulamalarında genelde madde sayısının 30 civarında olmasının etkili parametre kestirimleri için yeterli olacağı yaygın olarak kullanılan bir koşuldur. Öte yandan, geniş ölçekli sınavlarda test başına düşen madde sayısının genellikle 30 civarında olduğu bilinmektedir. Okul sayısı ve her bir okul içerisindeki öğrenci sayısının belirlenmesinde Maas ve Hox (2004a) tarafından HLM'nin parametre kestirim performansının ortaya konulduğu çalışma dikkate alınmıştır. Bu çalışmada özellikle varyans bileşeni parametrelerinin yansız kestirimi için grup sayısının 50 ve üzerinde olması gerektiği, grup içi birey sayısının da 30 ve üzerinde olmasının verimli olacağı belirtilmiştir. Araştırmacılar ayrıca parametre kestirimlerinde grup sayısının, grup içi birey sayısından daha önemli olduğunu vurgulamışlardır. Browne ve Draper (2000) de benzer şekilde grup sayısının 50 civarı olmasının yansız varyans bileşeni kestirimleri için yeterli olacağını belirtmiştir. Bu öneriler ışığında okul içi öğrenci sayısı 50, okul sayısının parametre kestirimlerinde daha etkili olduğu önerisine dayanarak okul sayısı 100 olarak belirlenmiştir.

Simülasyon çalışması için hem seçkisiz model parametreleri, hem de bu parametrelere dayanan 0-1 şeklindeki yanıt örüntüleri R (R Core Team, 2016) ile üretilmiştir. İlk olarak madde güçlük parametreleri (b_i) standart normal dağılımdan (ortalaması 0, varyansı 1 olan normal dağılım) üretilmiştir. Ardından okul yetenek parametreleri (u_k) ortalaması 0, varyansı 0.2 olan normal dağılımdan üretilmiştir. Her bir okula ait okul içi varyans parametrelerinin $(\sigma^2_{e_k})$ üretilmesinde öncelikle $\mu = 0.8$ ve ilgili heterojenlik koşulundaki τ değeri alınarak varyansların tersleri (1/varyans) önceki bölümde sunulan Gamma dağılımından üretilmiştir. Sonrasında bu değerlerin tersleri alınarak her bir okul için okul içi varyans değeri elde edilmiştir. Elde edilen bu okul içi varyans değerleri kullanılarak, her bir okul içerisindeki öğrencilerin okul ortalamalarından sapma değerleri (e_{jk}) sıfır ortalama ve ilgili okul içi varyans değeriyle normal dağılımdan üretilmiştir. Tüm parametrelerin üretilmesinin ardından, (27) numaralı model eşitliği ile bireylerin maddelere doğru cevap verme olasılıkları elde edilmiş ve bu olasılıklar kullanılarak 0-1 yanıt örüntüleri Bernuolli dağılımından üretilmiştir.

Simülasyon çalışması kapsamında her bir heterojenlik koşulu için 50 tekrarlı veri seti üretilip analiz edilmiştir. Bayes yaklaşımı kapsamında MCMC ile karmaşık modellerin parametre kestirimlerinde yüksek bilgisayar belleğine ihtiyaç duyulduğu ve kestirim zamanının frekansa dayalı yöntemlerden oldukça fazla olduğu bilinen bir durumdur. Özellikle hiyerarşik modellerde kestirim süreleri katlanarak artmaktadır. Simülasyon çalışmasına ilişkin yapılan denemelik analizlerde, her bir birey için θ_{jk} yetenek parametresinin kestirilmesi halinde toplam 50x100=5000 tane parametrenin ciddi boyutlarda bellek ve iterasyon süresine ihtiyaç duyduğu gözlenmiştir. Bu açıdan yapılan simülasyon çalışmasında, zaman ve bellek kısıtlamalarından dolayı birey düzeyi yetenekler kestirilmemiştir. Ancak burada vurgulanması gereken nokta, modelin kesinlikle bu yetenek parametrelerini kestirmeye olanak sunduğudur. Daha gelişmiş işlemci ve bellek ile beraber bu parametreler de makul sürelerde kestirilebilir.

Gerçekleştirilen denemelik analizler sonucunda ayrıca, her bir koşul için yakınsamaların sağlanması ve MCMC süreçlerinin verimli bir biçimde tamamlanması için gereken iterasyon sayılarına da karar verilmiştir. Bu bağlamda elde edilen bulgular ışığında orta ve yüksek heterojenlik koşulları için ısınma periyodu olarak 5000; ısınma sonrasında ise 30000 iterasyon 3 ayrı zincir ile yapılmıştır. Böylece bu koşullarda her bir parametrenin sonsal dağılımından 3x30000=90000 değer üretilmiştir. Düşük heterojenlik koşulunda ise yine 3 zincir halinde 5000 ısınma iterasyonu yapılmıştır. Ancak bu koşul için oto-korelasyonların daha yüksek olmasından dolayı ısınma sonrası 50000 iterasyon yapılarak toplamda her bir parametre için 3x50000=150000 değer üretilmiştir. Tek bir veri seti için tipik bir analiz, 64 bit işletim sistemine, 2.80 GHz hızındaki Intel Core i7 2640M işlemciye ve 4gb belleğe sahip bilgisayar ile 3 çekirdekte paralel işlem yapılmasıyla yaklaşık olarak 4 saat sürmüştür.

Simülasyon çalışması kapsamında tüm koşullar için veri analizleri yapılıp madde güçlükleri, okul yetenek parametrelerinin varyans değeri, okul yetenek parametreleri, okul içi varyansların ortalama ve varyans değerleri ile her bir okula ait okul içi varyans parametresi kestirilmiştir. Modelin, anılan parametreleri Bayes yaklaşımı çerçevesinde kestirim performansının incelenmesi için sistematik hatayı ifade eden yanlılık (bias), seçkisiz hatayı ifade eden standart hata (random/standard error: SE) ve toplam hatayı ifade eden hataların karelerinin ortalamasının karekökü (root mean squared error: RMSE) indeksleri hesaplanmıştır. Bu yaygın indekslere ek olarak, yanlılığı gerçek parametre değerine göre oransal olarak karşılaştıran göreceli mutlak yanlılık (absolute relative bias: ARB) indeksleri de elde edilmiştir. Hesaplanan bu indekslere ait formüller şu şekildedir:

$$Bias(\hat{\theta}) = \overline{\hat{\theta}_m} - \theta \tag{29}$$

$$RMSE(\hat{\theta}) = \sqrt{\frac{\sum_m^M (\hat{\theta}_m - \theta)^2}{M}} \tag{30}$$

$$SE(\hat{\theta}) = \sqrt{\frac{\sum_m^M (\hat{\theta}_m - \overline{\hat{\theta}_m})^2}{M}} \tag{31}$$

$$ARB(\hat{\theta}) = |Bias(\hat{\theta})/\theta| \tag{32}$$

Formüllerde yer alan θ terimi parametrenin gerçek değerini, $\hat{\theta}_m$ simülasyonun m. tekrarında kestirilen parametre değerini, M tekrar sayısını ve $\overline{\hat{\theta}_m}$ M tekrar sonucunda kestirilen parametre değerlerinin ortalamasını ($\hat{\theta}_m$ değerlerinin ortalamasını) ifade etmektedir. Bu indeksler arasındaki ilişki ise,

$$RMSE^2 = Bias^2 + SE^2 \tag{33}$$

şeklindedir.

Anılan indeksler model tarafından kestirilen tüm parametreler için hesaplanmıştır. Ancak modelin aşırı parametreleri dışında kalan madde güçlükleri, okul yetenekleri ve okul içi varyans bileşenleri için hesaplanan indeks değerlerinin ortalamaları rapor edilmiştir. Yanlılık indekslerinin negatif değerler de alabildiği göz önünde bulundurularak, ilgili parametreler için yanlılıkların mutlak değerlerinin ortalamaları hesaplanarak yorumlanmıştır. Ele alınan bu değerler ortalama mutlak yanlılık olarak ifade edilmiştir. Öte yandan bu parametrelerin gerçek değerleri ile olan korelasyonları her bir tekrar için hesaplanmıştır. Simülasyon sonunda elde edilen bu 50 korelasyon değerinin ortalama ve standart sapmaları da hesaplanarak, anılan parametrelerin kestirim performansına ilişkin ek bulgular diğer indekslerle beraber sunulmuştur.

Modelin Gerçek Verilere Uygulanması

Yürütülen simülasyon çalışmasının ardından, önerilen model 2009 yılı 9. sınıf ÖBBS verilerine uygulanmıştır. A kitapçığında yer alan 15 maddelik Türkçe testi verileri ilk olarak koşulsuz varyans modeli ile analiz edilmiştir. Öğrenciler tarafından yanıtlanmayan maddelere ilişkin gözlemler kayıp veri olarak ele alınmış ve Bayes yaklaşımının avantajı ile tüm parametreler mevcut kayıp veriler ile kestirilmiştir. Veri setinde yer alan toplam 300 okuldan 50 ve üzerinde öğrenciye sahip olan 127 okul analiz için seçilmiştir. Seçilen bu okullardan yaklaşık üçte birinin öğrenci sayısı 100 ve üzerinde olup, bir okuldaki en fazla öğrenci sayısı 205; tüm okullardaki toplam öğrenci sayısı ise 12289 olmuştur. Analiz için JAGS (Plummer, 2003, 2015) yazılımı kullanılarak 3 ayrı zincir şeklinde MCMC iterasyonları Gibbs örneklemesi yöntemiyle gerçekleştirilmiştir. Simülasyon çalışmasına ilişkin bölümde belirtildiği gibi, toplam 12289 öğrencinin tek tek yetenek parametrelerinin kestirilmesi ciddi bir zaman ve bilgisayar belleği gerektirdiğinden, gerçek veri analizlerinde de sadece okul düzeyi yetenekler kestirilmiştir. Anılan veri seti aynı zamanda geleneksel ÇDMTM ile analiz edilmiş ve bu modele ait parametreler de kestirilmiştir. Böylece kestirilen parametre değerleri ile birlikte önerilen model ile geleneksel modelin gerçek veriye olan uyumları da karşılaştırılmıştır.

Bayes yaklaşımı çerçevesinde model uyumlarının değerlendirilmesi, temelde her bir iterasyonda kestirilen parametreler ile analiz edilen verinin yeniden üretilme kalitesine dayanmaktadır. Böylece her bir iterasyon sonrasında kestirilen parametrelerle üretilen veri, analiz edilen gerçek veri ile karşılaştırılarak aradaki sapma değerleri elde edilmektedir. Analiz edilen verinin bu veriden kestirilen parametreler ile mükemmel düzeyde üretilebilmesi halinde, sapma değerleri de sıfıra oldukça yakın olacaktır. Sonuç olarak, kestirilen parametrelerin analiz edilen veriyi üretebilme düzeyleri, model-veri uyumu hakkında araştırmacılara bilgi vermektedir. Tüm iterasyonların tamamlanmasının ardından model-veri uyumu, bu sapmaların genel eğilimlerine dayalı olarak yorumlanmaktadır. Bu çalışmada model-veri uyumlarının değerlendirilmesinde "sapma bilgi ölçütü" olarak ifade edilen

Deviance Information Criterion (DIC: Spiegelhalter, Best, Carlin, & van der Linde, 2002) indeksi kullanılmıştır.

DIC, örneklem sayısı ve kestirilen parametre sayısının artması gibi durumlara dayanıklı olan ve Bayes yaklaşımı çerçevesinde yaygın olarak kullanılan uyum indeksi olup,

$$DIC = \overline{D(\theta)} + pD \tag{34}$$

eşitliği ile hesaplanmaktadır. Eşitlikte yer alan $\overline{D(\theta)}$ terimi, Bayes yaklaşımında model uyumuna ilişkin bir ölçüm olan sapmanın sonsal ortalamasıdır. pD terimi ise modeldeki serbest parametre sayısıdır. Daha düşük DIC değerine sahip olan modelin veriye olan uyumu daha yüksektir. Formülde yer alan pD terimi, daha karmaşık olan modeller için bir düzeltme rolüne sahiptir (Kang & Cohen, 2005). Diğer bir ifadeyle bu terim, daha fazla parametreye sahip olan modelin DIC indeksinin düşük olup basit modele göre avantajlı olmasını engellemek için bir düzeltme uygular. Böylece daha az parametreye sahip olan modelin veriye uyum konusunda dezavantajlı duruma düşmesinin önüne geçilir.

Gerçek veri analizinin ikinci aşamasında, okul içi varyansların heterojenliklerinin modellendiği koşullu varyans modeli kullanılmıştır. Kurulan koşullu modelde okul içi yetenek varyanslarının, okulların sınavla öğrenci alma durumuna göre farklılaşma durumu test edilmiştir. Bu amaçla, 2009 ÖBBS raporunda belirtilen akademik lise türleri içerisinden sınavla öğrenci alımı yapan liseler olan fen liseleri, Anadolu liseleri, Anadolu öğretmen liseleri ve sosyal bilimler liseleri analize dâhil edilmiş; sınavsız alım yapan akademik liseler kategorisinde ise genel liseler yer almıştır. Sınavla öğrenci alan liseler 1, sınavsız öğrenci alımı yapan genel liseler ise 0 şeklinde kodlanarak okul düzeyi açıklayıcı değişken olan “okul türü” değişkeni oluşturulmuştur. Sınavla alım yapan liselerdeki öğrenci sayılarının, genel liselerdeki öğrenci sayılarına göre düşük olmasından dolayı, daha önceki analizlerde 50 olarak belirlenen minimum öğrenci sayısı bu model için 10 olarak belirlenmiştir. Böylece, koşullu varyans modeli için analiz edilen veri setinde toplam 211 okul yer almıştır. Bu okulların 141 tanesi genel lise iken, 70 tanesi ise sınavla alım yapan liseler olmuştur. Analiz edilen bu veri setindeki toplam öğrenci sayısı ise 11946 olmuştur.

Koşullu varyans modeli için kurulan eşitlik,

$$\ln\left(\sigma_{e_k}^2\right) = \alpha_0 + \alpha_1 * OKL_k \quad (35)$$

şeklinde olup, OKL_k terimi k tane okulun sınavsız (0 şeklinde kodlanmış) veya sınavla (1 şeklinde kodlanmış) öğrenci alma durumunu gösteren iki kategorili bağımsız değişkeni ifade etmektedir. Kurulan doğrusal eşitlikte logaritma dönüşümü olduğundan, ortalama okul içi varyans kestirimi sınavsız öğrenci alan liseler için e^{α_0}, sınavla öğrenci alan liseler için $e^{\alpha_0+\alpha_1}$ olmaktadır. Okul türüne göre ortalama okul içi varyans kestirimlerinin birbirlerinden farklılaşma durumları, $e^{\alpha_0}/e^{\alpha_0+\alpha_1}$ şeklinde iki varyans bileşeninin oranlarının yeni bir parametre olarak kestirilip, bu parametrenin 1 değerinden ne kadar farklılaştığı ile belirlenmiştir. Sonuç olarak, test edilen bu koşullu varyans modeli ile okul türlerine göre ortalama okul içi varyansların farklılık gösterme durumu test edilmiştir.

Bölüm 4
Bulgular ve Yorumlar

Bu bölümde ilk olarak araştırma kapsamında önerilen modelin değişen heterojenlik düzeylerindeki parametre kestirim performansına ilişkin simülasyon çalışmasının bulgularına yer verilmiştir. Simülasyon çalışmasına ilişkin bulguların ardından, önerilen model ve geleneksel ÇDMTM'nin 2009 yılı ÖBBS Türkçe testi verilerine uygulanması ile elde edilen parametre kestirimleri ve model-veri uyumları karşılaştırılmıştır. Son olarak, uygulanan koşullu varyans modelinin analizine ilişkin bulgulara yer verilmiştir.

Simülasyon Çalışmasına İlişkin Bulgular

Yürütülen simülasyon çalışmasının temel amacı, heterojen grup içi varyans bileşenleri kestirecek şekilde yeniden modellenen ÇDMTM'nin parametre kestirim performansının değişen heterojenlik düzeylerinde incelenmesidir. Bu açıdan simülasyon çalışmasında değişimlenen tek koşul grup içi varyans bileşenlerinin heterojenlik düzeyi olmuştur. Aşağıda yer alan bölümlerde, araştırma problemlerine paralel olarak öncelikle simülasyon çalışması kapsamında yapılan analizlerin yakınsama durumlarına ilişkin bulgulara yer verilmiştir. Daha sonra sırasıyla aşırı ve temel model parametrelerinin kestirilme performanslarına ilişkin bulgular sunulmuştur.

Yakınsama durumlarına ilişkin bulgular. Simülasyon çalışması kapsamında yürütülen tekrarlı veri analizlerinde MCMC zincirlerinin yakınsama durumları, yapılan denemelik analizler üzerinden görsel ve istatistiksel olarak test edilmiştir. Her bir heterojenlik koşulu için yürütülen denemelik analizlerden elde edilen yakınsama bulguları EK-C, EK-Ç ve EK-D'de sunulmuştur. Sunulan bilgilerde iz, oto-korelasyon, "shrink factor" olarak da ifade edilen BGR ve sonsal dağılım grafiklerinin yanı sıra, ESS ve MC hatası kestirimleri de yer almaktadır. Modelin temel parametreleri olan madde güçlükleri, okul yetenekleri ve okul içi varyanslar fazla sayıda olduğundan, eklerde bu parametrelerin sadece en büyük ve en küçük değerlerine ait grafikler sunulmuştur.

EK-C, EK-Ç ve EK-D'deki ilgili grafikler incelendiğinde, tüm heterojenlik koşullarında iz grafiklerinin yeterli ölçüde karıştığı ve kararlı hale geldiği görülmektedir. Bu durum belirlenen ısınma periyotlarının yeterli olduğunu da doğrulamaktadır. BGR değerlerinin 1.1'in altında olması, ayrı zincirlerden elde edilen sonsal dağılım grafiklerinin birbirleriyle örtüşmesi ve MC hatası değerlerinin sıfıra oldukça yakın olması tüm koşullarda yakınsamanın gerçekleştiğine ilişkin güçlü kanıtlar sunmaktadır. EK-Ç ve EK-D'deki oto-korelasyon grafikleri incelendiğinde, orta ve yüksek heterojenlik düzeylerinde tüm parametrelere ait oto-korelasyonların küçük gecikme değerlerinde sıfıra yaklaştığı görülmektedir. Buna bağlı olarak parametrelerin ESS değerleri Kruschke (2014) tarafından önerilen 10000 değerini aşmış ve bu durum sonsal dağılımların etkili bir biçimde temsil edildiğini göstermiştir.

Ancak EK-C'deki grafiklerde görüldüğü gibi, düşük heterojenlik düzeyinde özellikle okul içi varyansların ortalama (μ) ve varyans (τ) parametreleri için oto-korelasyonlar yüksek gecikme değerleriyle azalmakta, dolayısıyla ESS değerleri gerek bu koşuldaki diğer parametrelere, gerekse diğer koşullarda kestirilen aynı parametrelere göre düşük kalmaktadır. Düşük heterojenlik koşulunda τ değerinin sıfıra; dolayısıyla okul içi varyans parametrelerinin birbirlerine oldukça yakın olması bu duruma neden olarak gösterilebilir. Kasim ve Raudenbush (1998) heterojen HLM'de grup içi varyansların varyans parametresinin sonsal dağılımının şekil ve değişkenliğinin, grup sayısı ve parametrenin gerçek değerine bağlı olduğunu belirtmişlerdir. Bu çalışmadaki grup sayısı daha önceki araştırmaların önerilerine dayanarak belirlendiği için, 100 grubun yeterli olduğu düşünülmektedir. Dolayısıyla düşük heterojenlik düzeyinde τ parametresine ait sonsal dağılımın ESS değerinin düşük olmasının, parametrenin gerçek değerinin sıfıra yakın (0.02) olmasından kaynaklandığı savı ön plana çıkmaktadır. Öte yandan okul içi varyans bileşenlerinin terslerine ait Gamma önsel dağılımındaki şekil ve ölçek parametrelerinin her ikisi de τ terimini içerdiğinden, bu durumun μ parametresine ait ESS değerini de etkilemiş olabileceği düşünülmektedir.

Sonuç olarak bütün heterojenlik düzeyleri için simülasyon öncesi yürütülen denemelik analizlerde MCMC iterasyonlarının tüm parametreler için yakınsadığı görülmektedir. Ancak tüm denemelik analizlerde MCMC zincirleri yakınsamış olsa da, düşük heterojenlik düzeyinde özellikle τ ve μ parametreleri için oto-korelasyonların yüksek; dolayısıyla ESS değerlerinin düşük olduğu gözlenmiştir. Bu durumun daha detaylı incelenmesi için simülasyon çalışmasının tamamlanmasının ardından tüm heterojenlik koşullarının her bir tekrarındaki parametre kestirimlerine ait ESS değerleri incelenmiştir. Düşük heterojenlik düzeyinde, tıpkı denemelik analizde olduğu gibi τ ve μ parametrelerinin ESS değerleri 50 tekrarın hepsinde 10000'den oldukça düşük çıkmıştır. Öte yandan okul içi varyans parametrelerinin çoğu değeri için bazı tekrarlarda ESS değerlerinin yine 10000'den düşük olduğu görülmüştür. Ancak diğer parametrelerin tümünün tüm tekrarlarda ESS değerleri 10000'i oldukça aşmıştır.

Orta heterojenlik düzeyinde sadece uç değerlerdeki birkaç okul içi varyans parametresine ait ESS değerleri bazı tekrarlarda 6000 civarında kalmış; diğer tüm parametreler için tüm tekrarlarda etkili örneklem büyüklükleri 10000'i aşmıştır. Yüksek heterojenlik düzeyinde ise orta heterojenlik düzeyinde olduğu gibi yalnızca uç değerdeki birkaç okul içi varyans parametresine ait ESS değerlerinin bazı tekrarlarda 3800 civarında kaldığı gözlenmiştir. Daha önceki bölümlerde belirtildiği gibi Kruschke (2014), 10000 ESS değerini özellikle parametrelerin standart hatalarının hassas bir biçimde kestirilmesi için önermiştir. Buna ek olarak araştırmacıların sadece parametrenin nokta kestirimi ile ilgilenmeleri halinde, daha düşük ESS değerlerinin de yeterli olacağını vurgulamıştır. Simülasyon çalışması kapsamında yapılan analizlerden elde edilen ESS değerlerine ilişkin bulguların parametre ve standart hata kestirimlerini etkileme durumları ilerleyen bölümler içerisinde tartışılmıştır.

Aşırı parametrelerin kestirimine ilişkin bulgular. Aşağıda yer alan Tablo 1'de modelin aşırı parametreleri olan okul içi varyansların varyans değeri (τ), okul içi varyansların ortalama değeri (μ) ve okul yeteneklerinin varyans değeri (σ_u^2) için 50 tekrarlı kestirim sonrasında hesaplanan göreceli mutlak yanlılık, yanlılık, RMSE ve SE indeksleri yer almaktadır.

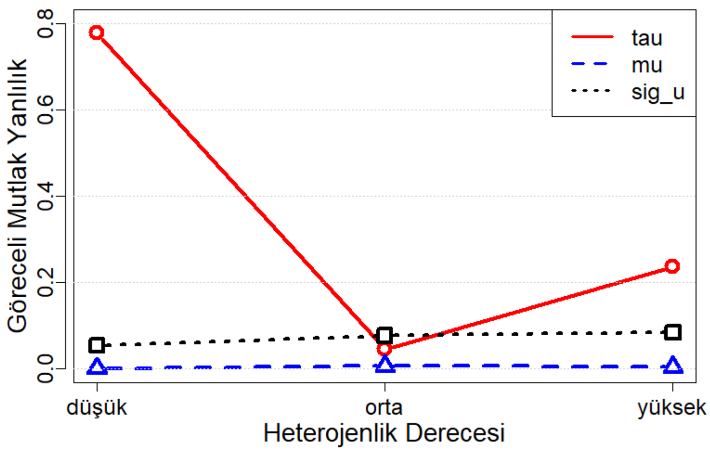
Hoogland ve Boomsma (1998) göreceli mutlak yanlılığın 0.05 ve altında olmasının kestirimlerin yanlılığı bakımından kabul edilebilir olduğunu belirtmiştir. Dolayısıyla bu ölçüte göre Tablo 1'deki göreceli mutlak yanlılık değerleri incelendiğinde düşük ve yüksek heterojenlik düzeyinde τ (tau) kestirimlerinin yanlı olduğu söylenebilir. Özellikle düşük heterojenlik düzeyindeki indeks değeri oldukça yüksek çıkmıştır. Ancak orta heterojenlik düzeyinde ise göreceli mutlak yanlılık belirtilen sınır değerin altında kalmıştır. μ (mu) kestirimleri için tüm koşullarda göreceli mutlak yanlılık 0.05'in oldukça altında kalmıştır. σ_u^2 (sig_u) kestirimlerinde ise heterojenlik düzeyi arttıkça, göreceli mutlak yanlılık değerlerinin de artış gösterdiği görülmüştür. Ancak yine de bu değerler 0.1'i aşmamıştır.

Tablo 1

Aşırı Parametrelere İlişkin İndeks Değerleri

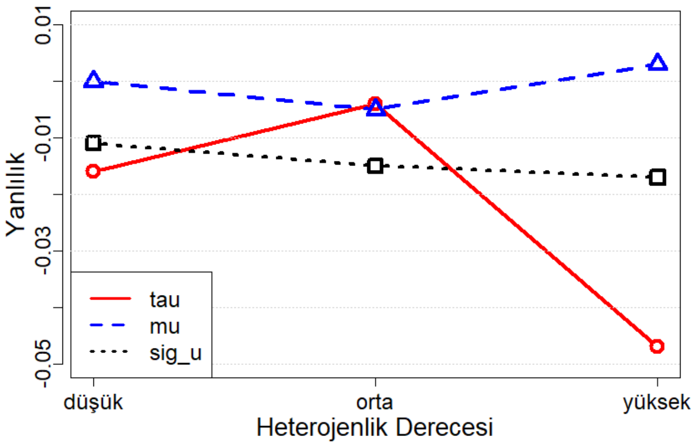
Aşırı Parametreler	*İndeks*	*Heterojenlik Derecesi* *Düşük* ($\tau = 0.02$)	*Orta* ($\tau = 0.1$)	*Yüksek* ($\tau = 0.2$)
Okul içi varyansların varyans değeri (τ)	Gör. Mut. Yanlılık	0.778	0.044	0.236
	Yanlılık	-0.016	-0.004	-0.047
	RMSE	0.016	0.008	0.048
	SE	0.002	0.007	0.009
Okul içi varyansların ortalama değeri (μ)	Gör. Mut. Yanlılık	0.000	0.006	0.003
	Yanlılık	0.000	-0.005	0.003
	RMSE	0.014	0.018	0.017
	SE	0.014	0.017	0.017
Okul yeteneklerinin varyans değeri (σ_u^2)	Gör. Mut. Yanlılık	0.053	0.076	0.083
	Yanlılık	-0.011	-0.015	-0.017
	RMSE	0.012	0.017	0.018
	SE	0.006	0.006	0.007

Anılan parametrelere ait göreceli mutlak yanlılık değerlerinin heterojenlik düzeylerine göre değişimleri aşağıda yer alan Şekil 9'daki grafikten de gözlenebilir.



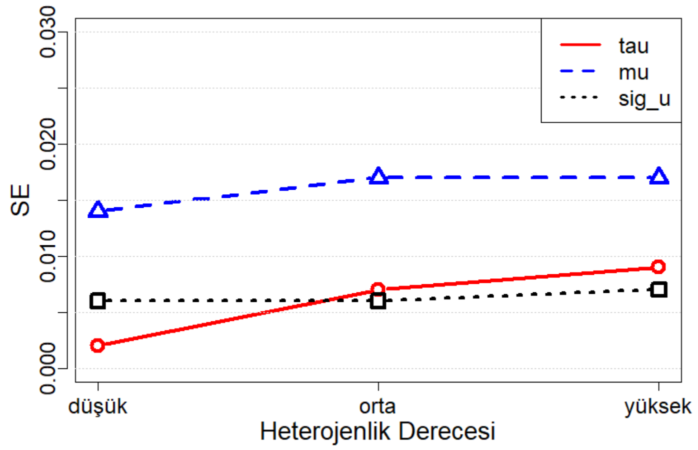
Şekil 9. Aşırı parametrelere ait göreceli mutlak yanlılık değerlerinin değişimi.

Tablo 1'de yer alan yanlılık indekslerinin genelde negatif eğilimde oldukları gözlenmiştir. Yanlılık indeksi bir modelin parametre kestirimindeki sistematik hatayı ifade etmektedir (Walther & Moore, 2005). Elde edilen yanlılık değerlerinin genellikle negatif olması, modelin aşırı parametreleri gerçek değerinden daha düşük kestirdiğini ifade etmektedir. Bu alt kestirim, göreceli mutlak yanlılık değerlerinden de hatırlanacağı üzere parametrenin kendi değerine göre en fazla τ parametresinde gözlenmiştir. Aşağıda yer alan Şekil 10'da aşırı parametrelere ilişkin yanlılık değerlerinin heterojenlik düzeylerine göre değişimleri görülmektedir. Grafikten de görüldüğü gibi μ (mu) ve σ_u^2 (sig_u) parametreleri için yanlılık değerleri 0 değerinden fazla uzaklaşmayarak benzer bir eğilim göstermektedir. τ (tau) parametresine ait yanlılık değeri orta heterojenlik düzeyine doğru sıfıra yaklaşırken, yüksek heterojenlik düzeyinde -0.05 değerine yaklaşmıştır.



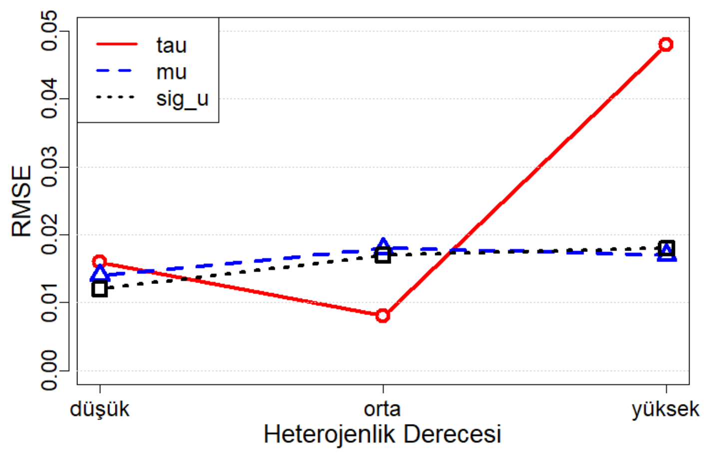
Şekil 10. Aşırı parametrelere ait yanlılık değerlerinin değişimi.

SE indeksi, parametre kestirimlerini gerçek parametre değerinden bağımsız olarak değerlendirir ve model ile yapılan tekrarlı kestirimler sonucunda elde edilen değerlerin kararlılığına ilişkin bilgi verir. Bu açıdan SE indeksi kestirimlerin hassasiyetini (precision) değerlendiren bir ölçüt olarak nitelendirilmektedir (Walther & Moore, 2005). Tablo 1'deki SE değerleri incelendiğinde, tüm koşullarda tüm aşırı parametreler için bu değerlerin düşük olduğu görülmektedir. SE değerleri göz önünde bulundurulduğunda, modelin aşırı parametreleri kestirme konusunda tüm koşullarda hassas olduğu görülmektedir. Diğer bir ifadeyle, tekrarlı analizler sonucunda elde edilen aşırı parametre kestirimlerinin birbirlerine oldukça yakın oldukları, nerdeyse sıfır olan SE değerleri ile teyit edilmiştir. Aşağıda yer alan Şekil 11'de tüm SE değerlerinin artan heterojenlik düzeyi ile birlikte hafif bir artış gösterdiği görülmektedir. Ancak bu artış çok düşük düzeyde olup, tüm heterojenlik düzeyleri için SE değerlerinin 0.02'den küçük olduğu göze çarpmaktadır. Bu bulgular ışığında modelin aşırı parametreleri heterojenlik düzeyinden etkilenmeden hassas bir biçimde kestirdiği söylenebilir. Diğer bir ifadeyle, tüm heterojenlik koşullarında tekrarlı analizler sonucunda kestirilen parametre değerlerinin çok fazla değişkenlik göstermediği söylenebilir.



Şekil 11. Aşırı parametrelere ait SE değerlerinin değişimi.

RMSE indeksi (33) numaralı eşitlikte ifade edildiği üzere, yanlılık ve SE indekslerinin sunduğu bilgileri harmanlayarak kestirimlerdeki toplam hatayı ifade etmektedir. RMSE, parametre kestirimleri için bir hatasızlık (accuracy) ölçütü olarak nitelendirilmekte ve kestirim ile gerçek değer arasındaki toplam uzaklık ölçütü olarak tanımlanmaktadır. Dolayısıyla düşük yanlılığa ve düşük SE'ye sahip olan kestirimlerin RMSE değerleri de düşük olacaktır (Walther & Moore, 2005). RMSE değerlerinin heterojenlik koşullarına göre değişimleri Şekil 12'de sunulmuştur.



Şekil 12. Aşırı parametrelere ait RMSE değerlerinin değişimi.

Grafikte görüldüğü gibi μ (mu) ve σ_u^2 (sig_u) parametrelerine ait RMSE değerleri artan heterojenlik düzeyleri ile çok fazla bir değişim göstermeyerek 0.01-0.02 değerleri arasında kalmıştır. τ (tau) parametresine ait RMSE değerinin orta heterojenlik düzeyinde hafif bir düşüş gösterdiği; ancak yüksek heterojenlik düzeyinde ise 0.05 sınırına yaklaştığı dikkat çekmektedir. Bu durum Şekil 10 ve 11'deki grafiklerden görüldüğü gibi SE değerinin çok değişmeyip, yanlılık değerinin negatif yönde benzer bir eğilim göstermesinden kaynaklanmaktadır. Bu açıdan okul içi varyansların varyans parametresi (τ) yüksek heterojenlik düzeyinde kararlı olarak kestirilmiştir; ancak tekrarlı analizlerde kestirimler gerçek değerinden daha düşük olmuştur. τ parametresinin kendi değeri de heterojenlik koşullarına göre değiştiğinden, göreceli mutlak yanlılık üzerinden bir karşılaştırma yapmak daha sağlıklı olacaktır. Dolayısıyla her ne kadar RMSE değeri τ için yüksek heterojenlik düzeyinde daha fazla olsa da, aslında kestirimlerdeki farklılıklar oransal olarak düşük heterojenlik düzeyinde daha fazla olmuştur. Bu durum τ için kestirilen güven aralıklarının hassasiyetinde de etkili olmuş ve bu duruma ilişkin bulgular aşağıda aktarılmıştır.

Modelin aşırı parametrelerinin nokta kestirimlerine ilişkin performansının ardından, bu kestirimlere ilişkin güven aralıklarının kestirim performansı da incelenmiştir. Bunun için simülasyonun her bir tekrarı sonucunda her bir parametreye ait %95 HDI değeri belirlenmiş ve bu aralığın ilgili parametrenin gerçek değerini içerme durumları toplam 50 tekrar için saptanmıştır. Son aşamada 50 tekrarda ilgili parametrenin gerçek değerinin güven aralığı tarafından kapsanma yüzdesi hesaplanmıştır. Bu oran alanyazında kapsanma (coverage) oranı olarak da anılmaktadır (Asparouhov & Muthén, 2014). Elde edilen oranlar incelendiğinde, düşük heterojenlik düzeyinde τ parametresi yalnızca %2 oranında kapsanabilmiştir. μ ve σ_u^2 parametreleri ise %100 oranında mükemmel biçimde kapsanmıştır. Önceki bölümde yer alan bulgulardan hatırlanacağı üzere, düşük heterojenlik düzeyinde ESS değerleri hem τ hem de μ için düşük çıkmıştı. Ancak buna rağmen μ parametresinin standart hata kestirimlerinin bu durumdan etkilenmediği %100 olan kapsanma oranından görülmektedir. Orta heterojenlik düzeyinde tüm aşırı parametrelerin kapsanma oranları %98 ve üzerinde çıkmış, herhangi bir olumsuz durum saptanmamıştır. Yüksek heterojenlik düzeyinde ise τ parametresi %70; μ ve σ_u^2 parametreleri ise %100 oranında kapsanmıştır.

Yanlılık değerlerinden de hatırlanacağı üzere, özellikle düşük heterojenlik düzeylerinde τ parametreleri gerçek değerlerinden daha düşük olarak kestirilmişlerdir. Bu durum nokta kestirimlerinin yanında, standart hatanın ve dolayısıyla %95 HDI aralıklarının hassasiyetine de yansımıştır. Düşük heterojenlik düzeyinde μ parametresine ait ESS değerinin de düşük olmasına rağmen kapsanma oranının mükemmel olması, τ parametresine ait kapsanma yüzdelerindeki problemin düşük ESS değerinden ziyade parametrenin gerçek değerinin sıfıra oldukça yakın olmasından kaynaklandığı düşünülmektedir. Bu durum ayrıca göreceli mutlak yanlılık değerinin de diğer koşullar ve parametrelere göre oldukça yüksek olmasına neden olmuştur. Nitekim düşük heterojenlik düzeyinde bu durumu destekleyen açıklamalar, yakınsama durumlarına ilişkin bulgularda belirtildiği gibi Kasim ve Raudenbush (1998) tarafından da raporlanmıştır. Öte yandan yüksek heterojenlik düzeyinde τ parametresine ait kapsanma değerinin %70 civarında olması Li ve Hedeker (2012) tarafından raporlanan bulgularla paralellik göstermektedir. Araştırmacılar boylamsal 3 düzeyli (ölçüm-gün-birey) heterojen model için yaptıkları simülasyon çalışmasında, birinci düzeydeki varyansların varyans parametresine ait nokta kestirimlerinin iyi olmasına rağmen, yüksek heterojenlik düzeyinde kapsanma oranının %86'ya düştüğünü belirtmişlerdir. Ancak %70'lik kapsanma oranı, düşük heterojenlik düzeyindeki %2'lik orana göre karşılaştırıldığında kabul edilebilir bir değer olarak yorumlanabilir.

Temel parametrelerin kestirimine ilişkin bulgular. Modelin aşırı parametrelerinin ardından, temel parametreler olan madde güçlükleri (b_i), okul yetenekleri (u_k) ve okul içi varyans bileşenleri ($\sigma^2_{e_k}$) için hesaplanan ortalama indeks değerleri ile gerçek ve kestirilen değerler arasındaki korelasyonların ortalama ve standart sapma değerleri Tablo 2'de sunulmuştur.

Tablo 2

Temel Parametrelere İlişkin Korelasyon ve İndeks Değerleri

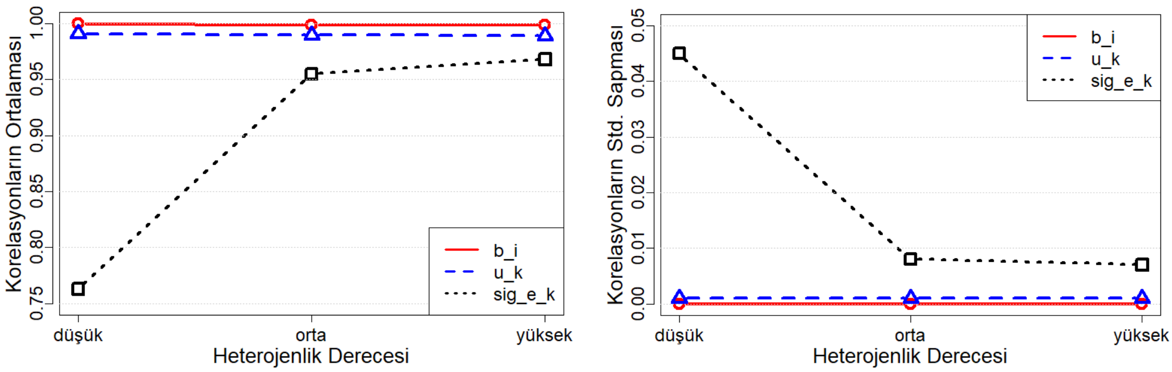
		Sonuçlar					
Temel Parametreler	*Het. Derecesi*	*Kor. Ort.*	*Kor. SS.*	*Ort. Gör. Mut. Yanlılık.*	*Ort. Mut. Yanlılık*	*Ort. RMSE*	*Ort. SE*
Madde güçlükleri ($\boldsymbol{b_i}$)	Düşük	1.000	0.000	0.056	0.004	0.033	0.033
	Orta	0.999	0.000	0.058	0.004	0.034	0.033
	Yüksek	0.999	0.000	0.021	0.006	0.035	0.034
Okul yetenekleri ($\boldsymbol{u_k}$)	Düşük	0.991	0.001	0.118	0.034	0.068	0.055
	Orta	0.99	0.001	0.159	0.042	0.073	0.054
	Yüksek	0.989	0.001	0.154	0.046	0.076	0.053
Okul içi varyans bileşenleri ($\boldsymbol{\sigma^2_{e_k}}$)	Düşük	0.763	0.045	0.142	0.110	0.114	0.021
	Orta	0.955	0.008	0.110	0.084	0.148	0.112
	Yüksek	0.968	0.007	0.114	0.089	0.173	0.136

Tablo 2'deki korelasyonların ortalama ve standart sapmaları incelendiğinde, b_i parametrelerinin gerçek ve kestirilen değerleri arasındaki korelasyonların tüm heterojenlik düzeylerinde mükemmel olduğu görülmektedir. Bu korelasyonlara ait standart sapma değerlerinin sıfır olması, madde güçlükleri için her bir iterasyon sonunda hesaplanan korelasyon değerinin 1'e yakınlık konusunda neredeyse hiç değişim göstermediğine işaret etmektedir. u_k parametrelerine ilişkin korelasyon değerleri tüm koşullarda 0.99 civarında olmuş, korelasyonların standart sapmaları da sıfıra oldukça yakın çıkmıştır. Ortalama korelasyonların yüksek çıkıp standart sapmaların sıfıra oldukça yakın olması okul yeteneklerinin de ardışık iterasyonlar boyunca kararlı bir biçimde kestirildiğine işaret etmektedir.

$\sigma_{e_k}^2$ parametrelerine ait ortalama korelasyon değeri orta ve yüksek heterojenlik düzeyinde 0.95'in üzerinde iken, düşük heterojenlik düzeyinde ise 0.76 olarak hesaplanmıştır. Daha önce belirtildiği gibi hiyerarşik Bayes modellerinde kestirimlerin ortalamaya doğru çekilmeleri sıkça yaşanan bir durumdur (Fox, 2010; Ntzoufras, 2009; Kruschke, 2014). Bu çekilme etkisinin, düşük heterojenlik düzeyinde zaten büyük bir çoğunluğu birbirlerine yakın olan okul içi varyans parametrelerini μ değerine (varyansların ortalamasına) daha fazla yaklaştırmış olması muhtemeldir. Bu durum sonucunda meydana gelen ranj daralması ile beraber kestirimlerin gerçek parametre değerleriyle olan korelasyonlarının azaldığı düşünülmektedir. Buna ek olarak, daha önceki bölümde belirtildiği gibi okul içi varyansların varyans parametresi (τ) gerçek değerinden daha düşük kestirilmiştir. Bu durum kestirilen okul içi varyans parametrelerini, gerçek değerlerindeki dağılımlarına göre birbirlerine daha da fazla yaklaştırmıştır. Dolayısıyla, ortalamaya çekilme etkisi ile beraber daha düşük olan varyans parametresi kestiriminin de anılan korelasyon değerinin düşük heterojenlik düzeyinde diğer iki koşula göre daha düşük olmasına neden olduğu düşünülmektedir.

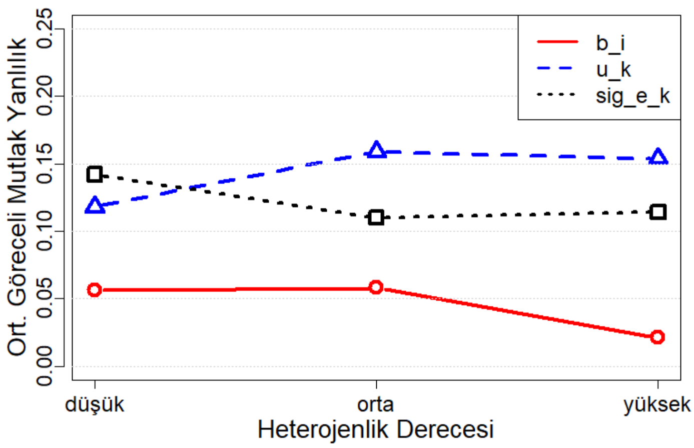
Temel parametrelere ilişkin korelasyonların ortalama ve sapmalarının heterojenlik düzeylerine göre değişimleri Şekil 13a ve 13b'de yer alan grafiklerden daha iyi gözlenebilir. İlgili grafikler incelendiğinde, b_i (b_i) ve u_k (u_k) parametrelerine ait korelasyonların ortalama ve standart sapmalarının heterojenlik düzeyinden etkilenmediği görülmektedir. Ancak $\sigma_{e_k}^2$ (sig_e_k) parametrelerine ait korelasyon değerlerinin ortalaması düşük heterojenlik düzeyinde nispeten düşük iken, orta ve yüksek heterojenlik düzeylerinde daha yüksek değerlere ulaşmıştır.

Orta ve yüksek heterojenlik düzeyleri için korelasyonların standart sapmaları da sıfıra oldukça yakın çıkmıştır. Bu durum yüksek korelasyonların orta ve yüksek heterojenlik düzeylerinde hemen hemen her iterasyon için elde edildiğini göstermektedir.



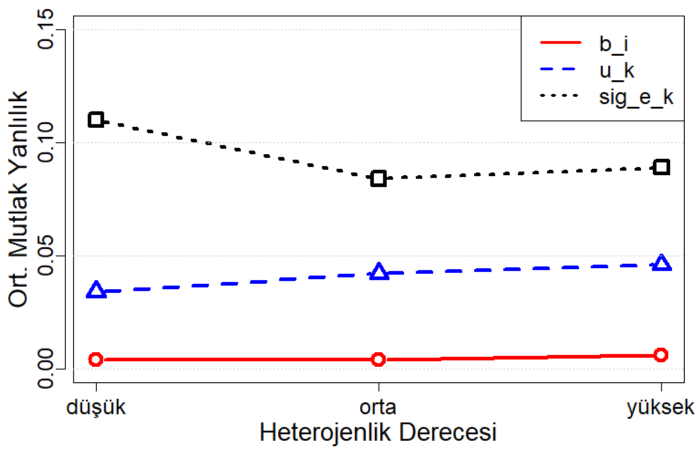
Şekil 13a ve 13b. Temel parametrelere ait korelasyonların ortalama ve standart sapmalarının değişimi.

Tablo 2 ve aşağıda yer alan Şekil 14'teki ortalama göreceli mutlak yanlılık değerleri incelendiğinde, b_i (b_i) parametreleri için bu değerlerin düşük ve yüksek heterojenlik düzeylerinde 0.05'i önemsenmeyecek düzeyde aştıkları; ancak yüksek heterojenlik düzeyinde ise sıfıra yaklaştıkları görülmektedir. $\sigma^2_{e_k}$ (sig_e_k) parametreleri için ortalama göreceli mutlak yanlılık değeri düşük heterojenlik düzeyinde 0.15 civarında iken, orta ve yüksek heterojenlik düzeylerinde 0.1 civarına gerilemiştir. u_k (u_k) parametrelerine ait ortalama indeks değerleri düşük heterojenlik düzeyinde 0.1 civarında iken, orta ve yüksek heterojenlik düzeylerinde ise 0.15 civarına yükselmiştir.



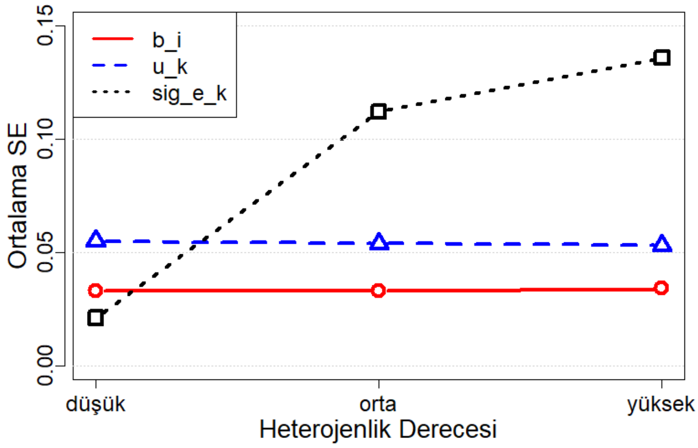
Şekil 14. Temel parametrelere ait ortalama göreceli mutlak yanlılık değerlerinin değişimi.

Tablo 2'de yanlılık değerlerinin ortalamaları yerine, yanlılıkların mutlak değerlerinin ortalamaları rapor edilmiştir. Böylece zıt işaretli yanlılık değerlerinin toplanmasıyla, sıfır ortalama yanlılık değeri elde etme olasılığının önüne geçilmiştir. Tablodaki ilgili değerler incelendiğinde, b_i parametreleri için ortalama mutlak yanlılıkların tüm koşullarda sıfıra oldukça yakın olduğu görülmektedir. u_k ve $\sigma^2_{e_k}$ parametreleri için yanlılıkların heterojenlik koşullarına göre değişim gösterdiği görülmektedir. Aşağıda yer alan Şekil 15'teki grafik incelendiğinde, b_i (b_i) parametreleri için ortalama mutlak yanlılık değerlerinin değişkenlik göstermeyip sıfıra oldukça yakın kaldığı görülmektedir. u_k (u_k) parametreleri için ortalama mutlak yanlılık değerleri artan heterojenlik düzeyi ile hafif bir artış göstermiştir; ancak yine tüm koşullarda 0.05 değerini aşmamıştır. $\sigma^2_{e_k}$ (sig_e_k) parametrelerine ait yanlılık değerleri ise düşük heterojenlik düzeyinde 0.1 değerini aşarken, orta ve yüksek heterojenlik düzeylerinde hafif bir düşüş göstermiştir.



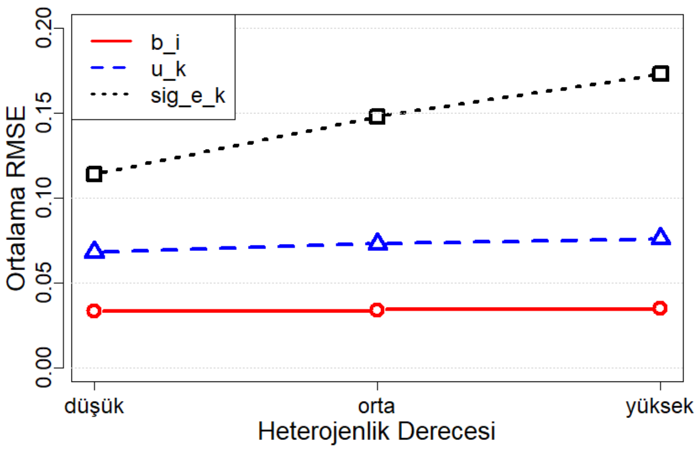
Şekil 15. Temel parametrelere ait ortalama mutlak yanlılık değerlerinin değişimi.

Ortalama mutlak yanlılık değerlerinin ardından, Tablo 2'deki ortalama SE değerleri incelendiğinde, bu değerlerin b_i ve u_k parametreleri için 0.1'in altında kaldığı görülmektedir. $\sigma^2_{e_k}$ parametreleri için düşük heterojenlik düzeyinde ortalama SE 0.05'ten küçük iken, orta ve yüksek heterojenlik düzeylerinde 0.1'in üstünde olmuştur. Bu değişimler aşağıda yer alan Şekil 16'da görselleştirilmiştir. Grafik incelendiğinde, b_i (b_i) ve u_k (u_k) parametrelerine ait ortalama SE değerlerinin heterojenlik düzeylerine göre belirgin bir değişim göstermediği görülmektedir. $\sigma^2_{e_k}$ (sig_e_k) parametrelerine ait ortalama SE değeri düşük heterojenlik düzeyinde sıfıra oldukça yakın iken, yüksek heterojenlik düzeyinde 0.15 değerine doğru artış göstermiştir. Dolayısıyla, okul içi varyans parametrelerine ilişkin kestirimlerin hassasiyeti, artan heterojenlik düzeyiyle beraber hafif bir düşüş göstermiştir.



Şekil 16. Temel parametrelere ait ortalama SE değerlerinin değişimi.

Tablo 2'deki ortalama RMSE değerleri incelendiğinde, b_i parametreleri için tüm değerlerin sıfıra oldukça yakın olduğu görülmektedir. Dolayısıyla madde güçlüklerinin tüm heterojenlik düzeylerindeki kestirimlerinin toplam hatasızlığının iyi düzeyde olduğu teyit edilmiştir. u_k parametrelerine ait ortalama RMSE değerleri 0.05 ile 0.1 arasında değişkenlik göstermiştir. $\sigma_{e_k}^2$ parametrelerine ait ortalama RMSE değerlerinin ise 0.1'i aştığı görülmektedir. Anılan değişimlere ait grafik Şekil 17'deki grafikte sunulmuştur.



Şekil 17. Temel parametrelere ait ortalama RMSE değerlerinin değişimi.

Şekil 17'deki RMSE değişim grafiği incelendiğinde, b_i (b_i) parametrelerine ait ortalama RMSE değerlerinin tüm heterojenlik düzeylerinde herhangi bir değişim göstermediği ve 0.05'ten düşük olduğu görülmektedir. u_k (u_k) parametrelerine ait değerler de heterojenlik düzeylerine göre belirgin bir değişim göstermemiş ve ortalama RMSE indeksleri 0.05 ile 0.1 değerleri arasında kalmıştır. $\sigma_{e_k}^2$ (sig_e_k) parametrelerine ait RMSE değerleri ise artan heterojenlik düzeyi ile beraber sürekli bir artış göstererek 0.15 değerini aşmıştır. Dolayısıyla okul içi varyans parametrelerinin $(\sigma_{e_k}^2)$ toplam kestirim hatasızlığının artan heterojenlik düzeyi ile artış gösterdiği söylenebilir. Bu artışın temel nedeninin, önceki bölümlerde anılan Bayes yaklaşımı ile kestirimlerdeki ortalamaya çekilme etkisinin olduğu düşünülmektedir. Nitekim yüksek heterojenlik düzeyinde okul içi varyans değerlerinin dağılımındaki uç değerlere ait RMSE'ler incelendiğinde, bu değerlerin oldukça yüksek olduğu saptanmıştır. Örneğin yüksek heterojenlik düzeyinde dağılımın en sağ ucunda yer alan okul içi varyans değeri yaklaşık olarak 4.01 olup, bu parametrenin kestirimine ait RMSE değeri ise 0.83 olmuştur. Benzer şekilde yüksek olan diğer uç değerlere ait yüksek RMSE'ler, yüksek heterojenlik düzeyinde okul içi varyans parametrelerinin kestirimine ilişkin ortalama RMSE değerinin yükselmesine yol açmıştır. Sonuç olarak ortalamanın oldukça uzağında olan (yüksek heterojenlik düzeyinde varyansların ortalaması 0.8 ve varyansı 0.2 idi) bu gibi değerlere ilişkin kestirimlerde ortalamaya çekilme etkisi çok daha belirgin olmuştur. Bu durum da okul içi varyans parametrelerinin kestirimine ilişkin ortalama RMSE değerinin yükselmesini tetiklemiştir. Ancak daha önce de vurgulandığı gibi özellikle okul içi varyans parametrelerinin varyans değerleri ve buna bağlı olarak üretilen tekil varyans değerleri de heterojenlik düzeyine bağlı olarak değişmektedir. Dolayısıyla bu parametrelere ilişkin değerlendirmelerin göreceli mutlak yanlılık ile ortalama göreceli mutlak yanlılık indeksleri üzerinden yapılması daha sağlıklı olacaktır.

Temel parametrelerin güven aralıkları ile kapsanma durumlarına ait bulgular incelendiğinde, düşük heterojenlik düzeyinde tüm madde güçlük ve okul yetenek parametrelerinin %98 ve üzerindeki oranlarla kapsandığı görülmüştür. Ancak düşük heterojenlik düzeyinde her bir okula ait varyans parametrelerinden özellikle uç değerlerde olanların kapsanma yüzdeleri sıfır olmuştur. Bu uç değerlere ait RMSE'ler de diğer okul içi varyans değerlerine göre daha yüksek olmuştur. Dolayısıyla düşük heterojenlik düzeyinde uç değerlerde yer alan okul içi varyans bileşenlerinin hem nokta kestirimlerinin hem de güven aralıklarının verimli bir biçimde elde edilemediği söylenebilir.

Orta heterojenlik düzeyinde tüm temel parametrelerin kapsanma oranları %98 ve üzerinde olmuştur. Yüksek heterojenlik düzeyinde ise madde güçlük ve okul yetenek parametreleri %98 ve üstünde kapsanma oranlarına sahip olmuştur. Ancak yüksek heterojenlik düzeyinde okul içi varyans parametrelerinin geneli oldukça yüksek kapsanma oranına sahipken, en düşük değere sahip iki okul içi varyans parametresinin kapsanma oranları ise %84 ve %56 olmuştur.

Parametrelerin güven aralıklarına ait kestirimlere ilişkin genel bir değerlendirme yapıldığında, modelin en iyi performansı orta heterojenlik düzeyinde gösterdiği açıkça görülmektedir. Yüksek heterojenlik düzeyinde de birkaç istisna haricinde bulgular oldukça iyi çıkmıştır. Ancak düşük heterojenlik düzeyinde özellikle τ ve buna bağlı olarak çoğu okul içi varyans parametresinin güven aralıklarının etkili bir biçimde kestirilemediği görülmüştür.

Modelin Gerçek Verilere Uygulanmasına İlişkin Bulgular

2009 ÖBBS verilerinin geleneksel ve önerilen koşulsuz model ile analizlerinin (en az 50 ve üzerinde öğrenciye sahip okullar ile yapılan analizler) yakınsama durumlarını gösteren grafikler EK-E ve EK-F'de sunulmuştur. Simülasyon çalışmasının yakınsama sonuçlarında olduğu gibi, gerçek veri analizleri için de aşırı parametreler dışında kalan parametrelerden sadece uç değerde olanların grafiklerine yer verilmiştir.

EK-E ve EK-F'de yer alan tüm iz grafiklerinin mükemmel biçimde karışıp ısınma periyodu sonrasında kararlı hale geldikleri görülmektedir. Her iki modelin tüm parametreleri için BGR değerleri 1.1'in altında bulunmuş ve MC hataları sıfıra oldukça yakın çıkmıştır. Ayrı zincirlere dayalı olarak çizdirilen sonsal dağılım grafikleri de birbirleriyle mükemmel düzeyde örtüşmüştür. Bu bulgular ışığında, gerçek veri analizlerinde her iki model için tüm parametre kestirimlerine ait zincirlerin yakınsadığı açıktır. Öte yandan heterojen modelin okul içi varyansların varyans parametresi (τ) haricinde, her iki modelin tüm parametrelerine ait oto-korelasyon değerleri düşük olmuş, ESS değerleri ise 10000'i oldukça aşmıştır. τ parametresine ait ESS değeri 8501 olmuş; ancak bu değer 10000'e oldukça yakın olup, simülasyon çalışmasında aynı parametre için elde edilen değerlerden oldukça büyük çıkmıştır. Bu bulgular, her iki model ile yapılan analizler sonucunda elde edilen sonsal dağılımların temsiliyet düzeylerinin yüksek olduğunu göstermektedir.

Aşağıda yer alan Tablo 3'te her iki model tarafından kestirilen madde güçlük parametreleri ve bu kestirimlere ait %95 HDI değerleri yer almaktadır. İlgili değerler incelendiğinde, her iki modelin madde güçlük parametrelerini nerdeyse aynı olarak kestirdikleri görülmektedir. Nokta kestirimlerine ilişkin güven aralıkları da neredeyse eşit olmuş; yalnızca bazı maddelerde virgülden sonraki ikinci basamak düzeyinde küçük farklılıklar gözlenmiştir. Madde güçlük kestirimlerinin homojen ve heterojen modellerde farklılık göstermemesi aslında beklenen bir bulgudur. Çünkü her iki model, yalnızca okul içi yeteneklerin heterojenliğinin dikkate alınmasında farklılaşmaktadır. Öte yandan Shieh (1999) ile Maas ve Hox (2004b) HLM'de artıkların normal dağılmaması durumunda modelin sabit parametrelerinin bu durumdan etkilenmediğini rapor etmişlerdir. Madde-birey-okul formundaki ÇDMTM'de modelin sabit parametreleri madde güçlükleridir; dolayısıyla bu açıklamalara dayanarak madde güçlüklerinin, modelin ikinci düzey artıkları olan yetenek parametrelerinin dağılım ihlallerinden etkilenmeyeceğini söylemek yanlış olmaz.

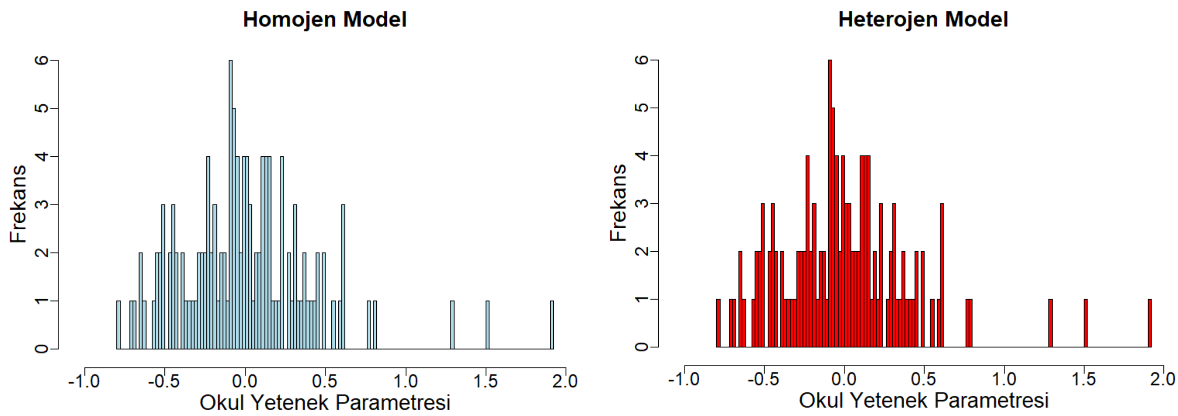
Tablo 3

Madde Güçlük Kestirimleri

	Homojen Model		*Heterojen Model*	
Madde	*Kes.*	*%95 HDI*	*Kes.*	*%95 HDI*
1	0.329	[0.243, 0.415]	0.328	[0.24, 0.414]
2	0.323	[0.236, 0.409]	0.322	[0.235, 0.409]
3	-0.873	[-0.958, -0.785]	-0.873	[-0.962, -0.787]
4	0.062	[-0.022, 0.151]	0.062	[-0.026, 0.147]
5	-0.239	[-0.326, -0.154]	-0.240	[-0.328, -0.155]
6	-0.175	[-0.26, -0.088]	-0.176	[-0.261, -0.089]
7	-0.36	[-0.447, -0.274]	-0.361	[-0.448, -0.274]
8	-0.497	[-0.585, -0.412]	-0.498	[-0.585, -0.411]
9	0.376	[0.289, 0.464]	0.375	[0.287, 0.463]
10	-0.573	[-0.659, -0.485]	-0.573	[-0.661, -0.487]
11	-0.007	[-0.092, 0.08]	-0.007	[-0.095, 0.079]
12	-0.544	[-0.63, -0.457]	-0.545	[-0.633, -0.459]
13	0.119	[0.032, 0.206]	0.119	[0.031, 0.205]
14	0.511	[0.425, 0.597]	0.510	[0.425, 0.599]
15	-1.083	[-1.171, -0.995]	-1.083	[-1.172, -0.996]

Yürütülen koşulsuz model analizleri sonucunda okul yetenek parametrelerinin varyans parametresi (σ_u^2), parantez içindeki değerler %95 HDI olmak üzere, homojen modelde 0.188[0.140, 0.240]; heterojen modelde ise 0.188[0.142, 0.241] olarak kestirilmiştir. Görüldüğü gibi σ_u^2 için nokta kestirimleri her iki modelde aynı olmuş, %95 HDI değerleri ise yalnızca virgülden sonraki üçüncü basamakta farklılaşmıştır. Okul yeteneklerinin varyansı da her iki modelde okul içi yeteneklerin heterojenliğinden bağımsız olarak modellendiğinden, bu benzerlik de beklenen bir bulgudur.

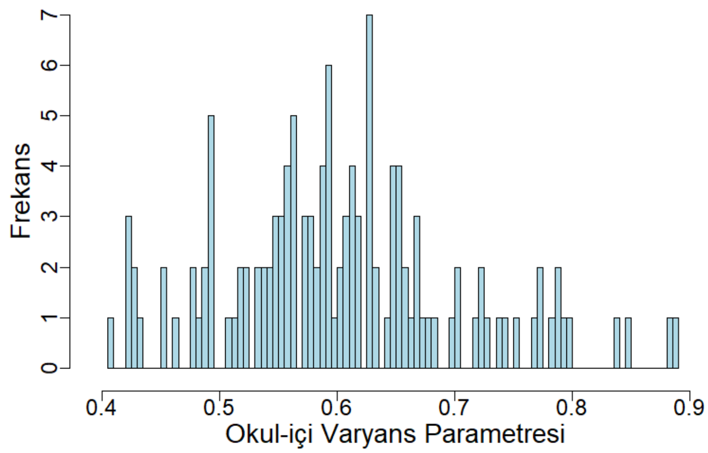
Her iki model tarafından kestirilen 127 adet okul yetenek parametresi ve bu kestirimlere ait %95 HDI değerleri karşılaştırmalı olarak EK-G'de sunulmuştur. Okul yetenek parametreleri de her iki modelde neredeyse aynı olarak kestirilmiş; ancak kestirimlere ilişkin güven aralıkları genellikle heterojen modelde biraz daha geniş olmuştur. Dolayısıyla, okul yeteneklerine ilişkin nokta kestirimlerdeki hataların heterojen modelde biraz daha yüksek olduğu söylenebilir. Her iki model tarafından kestirilen okul yetenek parametreleri arasındaki korelasyon 1; kestirimler arasındaki mutlak farkların ortalaması ise 0 olmuştur. Aşağıda yer alan Şekil 18a ve 18b'de kestirilen okul yeteneklerine ait dağılımların histogramları incelendiğinde, her iki histogramın nerdeyse eş olduğu ve okul yeteneklerinin model tanımında belirtildiği biçimde sıfır ortalama etrafında yığıldığı görülmektedir. En düşük okul yetenek parametresi -1 ile -0.5 arasında yer alırken, en yüksek okul yetenek parametresi ise 2'ye oldukça yakındır.



Şekil 18a ve 18b. Okul yeteneklerine ait histogramlar.

Homojen modelde okul içi yeteneklerin eş varyansa sahip olduğu varsayımına dayalı olarak tek bir okul içi varyans parametresi (σ_e^2) kestirilmektedir. σ_e^2, parantez içindeki değerler %95 HDI olmak üzere, 0.597[0.571, 0.623] olarak kestirilmiştir. Öte yandan heterojen model ile her bir okul için ayrı bir varyans parametresi kestirilmektedir. Analiz sonucunda okul içi varyansların ortalama (μ) ve varyans (τ) değerleri sırasıyla 0.571[0.535, 0.609] ve 0.03[0.015, 0.045] olarak kestirilmiştir. Görüldüğü gibi homojen modelin kestirdiği ortak okul içi varyans bileşeni (σ_e^2), heterojen model tarafından kestirilen μ değerinden biraz daha yüksek olmuştur. 0.03 olarak kestirilen τ değeri, gerçek veride okul içi varyansların heterojenliğinin düşük ile orta düzey arasında olduğunu göstermektedir. Bu açıdan homojen model tarafından kestirilen σ_e^2 değeri ile heterojen model tarafından kestirilen μ değeri arasındaki fark çok fazla olmamıştır. Benzer bir durum Kasim ve Raudenbush (1998) tarafından rapor edilmiştir. Araştırmacılar, yüksek heterojenlik koşulu için ürettikleri verinin geleneksel HLM ile analizi sonucunda kestirdikleri grup içi varyans bileşeninin, verinin üretilmesinde kullanılan grup içi varyansların ortalamasından oldukça yüksek olduğunu belirtmişlerdir. Ancak bu durumun düşük heterojenlik düzeyinde söz konusu olmadığını vurgulamışlardır.

Heterojen model tarafından kestirilen okul içi varyansların dağılımına ait histogram Şekil 19'da sunulmuştur. Histogramda görüldüğü gibi okul içi varyans değerleri 0.571 olarak kestirilen ortalama değerleri etrafında yığılmıştır. Okul içi varyans bileşenlerine ilişkin elde edilen bu bulgular, okulların öğrenci yeteneklerinin dağılımları bakımından az da olsa heterojenlik gösterdiğini doğrular niteliktedir. Her bir okul için ayrı olarak kestirilen okul içi varyans parametreleri ve kestirimlere ilişkin %95 HDI değerleri EK-Ğ'de yer almaktadır. Kestirilen parametreler incelendiğinde, en düşük ve en yüksek okul içi varyansın sırasıyla 0.409 ve 0.886 olduğu görülmektedir.

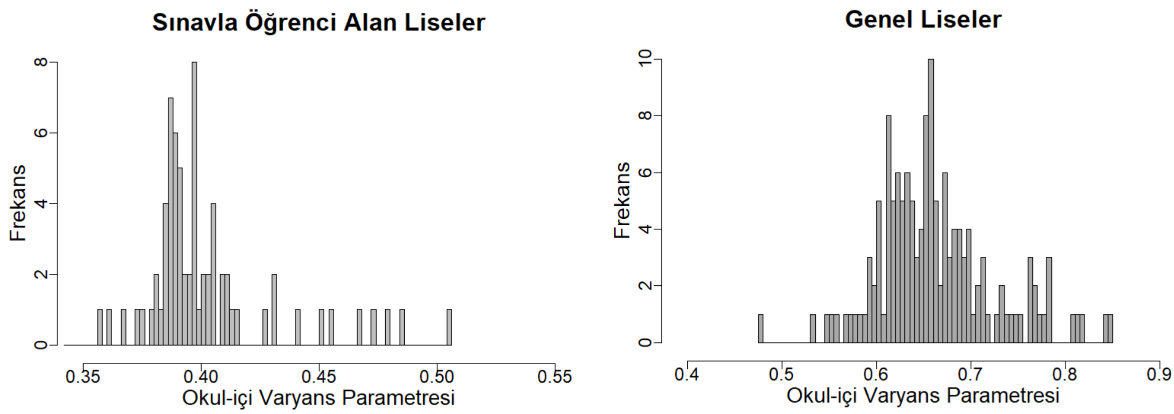


Şekil 19. Okul içi varyansların histogramı.

Gerçek veriler ile yürütülen analizler sonucunda homojen model için hesaplanan DIC indeksi 223455; heterojen model içen hesaplanan DIC indeksi ise

223410 olmuştur. Homojen ve heterojen model arasındaki DIC indeks farkı $223455 - 223410 = 45$ olmuştur. Lunn, Jackson, Best, Thomas ve Spiegelhalter (2012), 5 ve üzerindeki DIC farkının model uyum karşılaştırmaları için güçlü bir değer olduğunu belirtmişlerdir. İki model ile yürütülen analizler sonucunda 45 çıkan DIC farkı, önerilen 5 değerini oldukça aştığından, heterojen modelin gerçek veriye homojen modelden çok daha iyi uyum sağladığı görülmektedir. Öte yandan heterojen modelin veriye daha iyi uyum sağlaması, gerçek veri koşullarında okulların yetenek dağılımları bakımından heterojenlik gösterebileceği hipotezini de destekler niteliktedir.

Gerçek veri analizlerinin ikinci aşamasında, en az 10 ve üzerinde öğrenciye sahip toplam 211 akademik türdeki lisenin verileri ele alınmıştır. Bu verilere uygulanan koşullu varyans modelinin analizi sonucunda, sınavla öğrenci alan liselerin ortalama okul içi varyans değeri ($e^{\alpha_0+\alpha_1}$) 0.401[0.373, 0.431] olarak kestirilmiştir. Sınavsız öğrenci alan genel liselerin ortalama okul içi varyans değeri (e^{α_0}) ise 0.661[0.632, 0.691] olarak kestirilmiştir. Elde edilen bu bulgulara dayanarak, sınavla öğrenci alan akademik türdeki liselerin Türkçe yetenekleri bakımından daha homojen oldukları söylenebilir. Bunun aksine, genel liselerdeki okul içi yetenekler daha heterojen bir dağılım sergilemiştir. Diğer bir ifadeyle, sınavla öğrenci alan liselerdeki öğrenciler Türkçe yeteneği bakımından daha benzeşik iken, genel liselerde yetenekler bakımından daha değişken bir yapının olduğu söylenebilir. Aşağıda yer alan histogramlarda, iki farklı lise türüne ait okul içi varyansların dağılımları görselleştirilmiştir.



Şekil 20a ve 20b. Lise türlerine göre okul içi varyansların histogramları.

Şekil 20a'daki grafikte görüldüğü gibi, sınavla öğrenci alan liselerin okul içi varyans değerleri 0.55 değerini aşmamış; bu liseler içerisinde en büyük okul içi varyans değeri dahi, genel liselerin ortalama okul içi varyans değerinden düşük kalmıştır. Şekil 20b incelendiğinde, genel liselerin okul içi varyans değerlerinin büyük bir çoğunluğunun 0.5'in üzerinde olduğu; 0.8 ve üzerinde okul içi varyans değerine sahip olan okulların da olduğu görülmektedir. İki farklı lise türüne ait ortalama okul içi varyansların oranı ise ($e^{\alpha_0}/e^{\alpha_0+\alpha_1}$) 0.623[0.607, 0.655] olarak kestirilmiştir. Görüldüğü gibi bu oran yaklaşık olarak %62 civarında olup güven aralığı 1 değerini kapsamamaktadır. Dolayısıyla, her iki okul türü için hesaplanan ortalama okul içi varyans değerlerinin eşit olduğu şeklindeki hipotezin istatistiksel olarak (%95 anlamlılık düzeyinde) manidar olmadığı söylenebilir. Buna dayalı olarak, okul türü değişkeninin okul içi varyansların değişkenliğini anlamlı düzeyde açıkladığı yorumu yapılabilir.

Okulların kestirilen yetenek parametrelerinin lise türlerine göre ortalamaları hesaplanmış, genel liselerin ortalama Türkçe yetenek parametresi -0.504, sınavla alım yapan liselerin ise 1.018 olarak bulunmuştur. Bu bulgulara göre, sınavla öğrenci alımı yapan akademik liselerin hem yüksek ortalama okul başarı düzeyi; hem de düşük okul içi öğrenci yeteneği değişkenliği ile okul etkililiği bakımından genel liselerden daha iyi oldukları söylenebilir. Okul içi varyansların homojen varsayılarak her bir okul için tekil olarak kestirilmemesi, okul etkililiği hakkında ÇDMTM ile daha anlamlı bilgiler edinilmesine sınırlama getirmektedir. Önerilen model gerçek duruma daha uygun analizler yaparak, araştırma sonuçlarından daha etkili bir biçimde yararlanmaya olanak sağlamaktadır.

Bölüm 5

Sonuç, Tartışma ve Öneriler

Çalışmanın bu bölümünde, elde edilen bulgulara dayalı olarak ulaşılan sonuçlara ve bu sonuçlar ışığında geliştirilen önerilere yer verilmiştir.

Sonuçlar ve Tartışma

Araştırma kapsamında madde-birey-grup formundaki geleneksel ÇDMTM heterojen grup içi varyans bileşenleri kestirecek biçimde yeniden modellenmiş ve önerilen modelin değişen heterojenlik düzeylerindeki parametre kestirim performansı simülasyon çalışması ile ortaya konulmuştur. Yürütülen simülasyon çalışmasının ardından, önerilen model gerçek bir veri setine uygulanarak kestirilen parametreler ve model-veri uyumu geleneksel ÇDMTM ile karşılaştırılmış ve aynı zamanda koşullu varyans modeli çerçevesinde olası okul içi heterojenliğin kaynakları açıklanmaya çalışılmıştır. İlerleyen bölümlerde öncelikle modelin parametre kestirim performansına ilişkin simülasyon çalışmasına yönelik sonuçlara yer verilmiştir. Sonrasında gerçek veri analizine ilişkin sonuçlar rapor edilmiştir.

Simülasyon çalışmasına ilişkin sonuçlar. Yürütülen simülasyon çalışmasından elde edilen sonuçlar, araştırma problemlerine paralel biçimde aşırı ve temel parametreler için ayrı olarak ele alınmış ve aşağıda sunulmuştur.

Aşırı parametrelerin kestirimine ilişkin sonuçlar. Simülasyon çalışması sonucunda aşırı parametrelerin kestirimlerine ait yakınsama bulguları incelendiğinde, tüm MCMC iterasyonlarının tüm heterojenlik düzeylerinde yakınsadığı görülmüştür. Ancak düşük heterojenlik düzeyinde grup içi varyansların ortalama ve varyans parametrelerine ait MCMC zincirlerindeki oto-korelasyonların nispeten yüksek kaldığı görülmüştür. Zincirlerdeki yüksek oto-korelasyonlar, bu aşırı parametrelere ait ESS değerlerinin düşük olmasına neden olmuştur. Düşük heterojenlik düzeyinde, okul yeteneklerinin varyans parametresine ait ESS değerlerinde herhangi bir problem gözlenmemiştir. Öte yandan orta ve yüksek heterojenlik düzeylerinde tüm aşırı parametreler için oto-korelasyon değerleri düşük çıkmış ve buna bağlı olarak ESS değerleri oldukça yüksek olmuştur.

Aşırı parametrelerin kestirim kalitesine ilişkin bulgular incelendiğinde, tüm heterojenlik düzeylerinde okul içi varyansların ortalama değerinin neredeyse yansız olarak kestirildiği sonucuna ulaşılmıştır. Toplam hata değeri olan RMSE'ler ise artan heterojenlik düzeyi ile hafif bir artış göstermiş olsa da, bu artış belirgin olmamıştır. Öte yandan bu parametreye ilişkin kapsanma oranları tüm heterojenlik koşullarında oldukça yüksek olmuştur. Dolayısıyla okul içi varyansların ortalamasının nokta kestirimlerinin ve standart hatalarının tüm heterojenlik koşullarında sorunsuz bir biçimde elde edildiği söylenebilir.

Okul içi varyansların varyans parametresinin kestirimlerinin özellikle düşük ve yüksek heterojenlik düzeylerinde yanlı olduğu sonucuna ulaşılmıştır. Güven aralıklarının kestirimine ilişkin sonuçlar, yanlılık bulguları ile paralellik göstermiştir. Okul içi varyansların varyans parametresinin kestirim kalitesi için genel bir değerlendirme yapıldığında, orta heterojenlik düzeyinde nokta kestirimi ve güven aralığının oldukça iyi olduğu söylenebilir. Yüksek heterojenlik düzeyinde ise parametrenin nokta kestirimi yanlı olmuş, güven aralıklarının kestirimleri de kabul edilebilir bir seviyede gerçekleşmiştir. Düşük heterojenlik düzeyinde ise hem nokta kestirimi hem de güven aralıkları yeterince etkili olmamıştır.

Okul yeteneklerinin varyans parametresine ilişkin nokta kestirimleri, artan heterojenlik düzeyi ile birlikte gerçek değerlerinden hafif bir sapma göstermiştir. Ancak yine de gerek yanlılık, gerekse toplam hata değerleri heterojenliğin artması ile aşırı bir artış göstermemiştir. Parametrenin tüm koşullardaki kapsanma oranları mükemmel olmuştur. Dolayısıyla, okul yeteneklerinin varyans parametresinin hem nokta kestirimleri hem de güven aralıkları tüm heterojenlik düzeylerinde iyi bir performans ile kestirilmiştir.

Önerilen modelin aşırı parametreleri kestirme kalitesi ile ilgili genel bir değerlendirme yapıldığında, okul yeteneklerinin varyansı ve okul içi varyansların ortalamalarının tüm heterojenlik koşullarında etkili bir biçimde kestirildiği söylenebilir. Ancak okul içi varyansların varyans parametresinin kestirimleri yalnızca orta heterojenlik düzeyinde iyi olmuştur. Bu parametre için yüksek heterojenlik düzeyinde kabul edilebilir bir kestirim performansı söz konusu iken, düşük heterojenlik düzeyinde ise kestirimler ve güven aralıkları etkili olarak elde edilememiştir.

Temel parametrelerin kestirimine ilişkin sonuçlar. Temel parametrelere ilişkin bulgular değerlendirildiğinde, madde güçlük parametrelerinin tüm heterojenlik düzeylerinde nerdeyse yansız ve hatasız olarak kestirildiği açıkça görülmüştür. Yine tüm koşullarda madde güçlüklerinin kapsanma oranları yüksek olmuştur. Dolayısıyla, önerilen model ile madde güçlük parametrelerinin nokta kestirimleri ve bu kestirimlerin güven aralıkları tüm heterojenlik koşullarında neredeyse mükemmel bir biçimde elde edilmiştir.

Okul yeteneklerinin nokta kestirimlerine ait yanlılık ve hatasızlık değerleri, artan heterojenlik ile birlikte artış göstermiş; ancak bu artış dikkat çekici boyutlarda olmamıştır. Öte yandan tüm heterojenlik düzeylerinde, tüm okul yetenek parametrelerinin kapsanma oranları yüksek olmuştur. Genel bir değerlendirme yapıldığında önerilen modelin, okul yetenek parametrelerinin nokta kestirimlerini ve bu kestirimlerin güven aralıklarını tüm heterojenlik düzeylerinde kabul edilebilir bir performansla kestirebildiği söylenebilir.

Okul içi varyans bileşenlerinin, düşük heterojenlik düzeyindeki kestirimlerinin yanlı olduğu sonucuna ulaşılmıştır Yine düşük heterojenlik düzeyinde özellikle uç değerlerin kapsanma oranları oldukça düşük olmuştur. Orta ve yüksek heterojenlik düzeylerinde ise nokta kestirimlerin yanlılıkları daha düşük olmuş ve kapsanma oranları genel olarak yüksek olmuştur. Okul içi varyans bileşenlerinin kestirimlerine ilişkin genel bir değerlendirme yapıldığında, özellikle düşük heterojenlik düzeyinde nokta kestirimleri ve güven aralıklarının çok tutarlı sonuçlar vermediği sonucuna ulaşılmıştır. Orta ve yüksek heterojenlik düzeylerinde ise hem nokta kestirimlerinin hem de standart hataların genel anlamda kabul edilebilir nitelikte oldukları görülmüştür.

Modelin gerçek verilere uygulanmasına ilişkin sonuçlar. 2009 ÖBBS 9. sınıf A kitapçığı Türkçe testi verilerinin, geleneksel ve önerilen model ile analizleri sonucunda her iki modelin tüm parametrelerine ait MCMC zincirlerinin yakınsadığı ve ESS değerlerinin yeterince yüksek olduğu görülmüştür. Önerilen model ile yürütülen gerçek veri analizinde yakınsamaya ilişkin herhangi bir problem olmaması, modelin uygulamadaki performansı için olumlu bir sonuç olmuştur. Nitekim bu gibi karmaşık ve özellikle çok düzeyli modellerin gerçek verilere uygulanmasında yakınsama problemlerinin yaşanması olası bir durumdur. Öte yandan, her iki model ile kestirilen madde güçlük parametreleri neredeyse aynı olmuştur. Tıpkı madde güçlüklerinde olduğu gibi, her iki model tarafından kestirilen okul yeteneklerinin varyansı ve her bir okula ait yetenek parametresi de neredeyse aynı olmuştur. Her iki modelin, anılan bu parametreleri birbirlerine oldukça yakın kestirmesi beklenen bir sonuçtur.

Geleneksel ve önerilen modelin farklılaştığı kısım olan okul içi varyans bileşenlerinin modellenmesinde, geleneksel model tek bir okul içi varyans bileşeni kestirirken, önerilen model okul içi varyansların düşük ile orta düzey arasında bir heterojenlik gösterdiğini ortaya koymuştur. Gerçek verilerin analizinden elde edilen bulgulara dayanarak, okul içi yetenek dağılımlarının varyanslarının okullar arasında değişkenlik gösterdiği sonucuna ulaşılmıştır. Dolayısıyla, bazı okulların öğrencilerin Türkçe yetenekleri bakımından daha homojen, bazı okulların ise daha heterojen oldukları tespit edilmiştir. Elde edilen bu sonuçlar, 0-1 şeklindeki hiyerarşik verilerde geleneksel ÇDMTM yerine önerilen modelin kullanılmasının okul etkililiğinin “okul içi yetenek değişkenliği” bağlamında da ortaya konulması bakımından sağladığı avantajı göstermektedir.

Yapılan uyum iyiliği testinde, önerilen modelin gerçek veriye geleneksel modelden daha iyi uyum sağladığı sonucuna ulaşılmıştır. Bu sonuç, kestirilen diğer model parametrelerinin kalitesi ile birlikte, önerilen modelin gerçek verilere rahatlıkla uyum sağlayabilecek nitelikte olduğunu doğrulamaktadır. Böylece okul içi varyansların heterojen olması halinde, önerilen modelin homojenlik varsayımına dayanan geleneksel ÇDMTM’den daha uygun olacağı söylenebilir. Önerilen modelin veriye geleneksel modelden daha iyi uyum sağlaması, kestirilen varyans parametreleri ile birlikte değerlendirildiğinde, okulların Türkçe yetenekleri bakımından heterojen olduklarına dair güçlü bir kanıt sunmaktadır. Diğer bir ifadeyle geleneksel modelin varsaydığının aksine, her bir okul içerisindeki Türkçe yetenek dağılımının eş varyansa sahip olmadığı sonucuna ulaşılmıştır.

Koşulsuz varyans modelinin ardından, akademik lise türlerindeki heterojen okul içi varyans değerlerinin, okulların sınavla öğrenci alma durumuna göre değişiminin incelendiği koşullu model analiz edilmiştir. Bu analizin sonucunda, akademik lise türleri içerisinde sınavla öğrenci alımı yapan fen liseleri, Anadolu öğretmen liseleri, Anadolu liseleri ve sosyal bilimler liselerinin, sınavsız öğrenci alımı yapan genel liselere göre daha düşük okul içi varyans değerlerine sahip oldukları tespit edilmiştir. Dolayısıyla, sınavla öğrenci alımı yapan akademik liseler, öğrencilerin Türkçe yetenekleri bakımından daha benzeşik bir yapı göstermektedir.

Buna karşın genel liselerdeki öğrenciler ise Türkçe yetenekleri bakımından daha fazla değişkenlik göstermektedirler.

Çalışmanın giriş bölümünde vurgulandığı gibi, öğrencilerin bir sınav ile seçilip geldiği liselerdeki yetenek dağılımlarının, genel liselerdeki dağılımlara göre varyans bakımından farklılık göstermesi aslında beklenen bir durumdur. Önerilen heterojen ÇDMTM ile birlikte koşullu varyans modeliyle yapılan analiz sonucunda bu beklentinin doğrulandığı görülmektedir. Bu sonuç, istatistiksel olarak modelin gerçek verilere uygunluğu konusunda önemli bir dayanak olarak görülebilir. Ancak bunun yanında, modelin öğretim faaliyetlerine yönelik uygulamalar için önemli bir araç olarak da kullanılabileceği söylenebilir. Nitekim Kim ve Choi (2008) tarafından vurgulandığı gibi okul etkililiği, yalnızca yüksek olan okul ortalamaları bağlamında değil, buna ek olarak düşük okul içi değişkenlik bağlamında da değerlendirilmelidir. Diğer bir ifadeyle, etkili olan bir okul hem yüksek başarı düzeyine, hem de mümkün olduğunca düşük okul içi yetenek değişkenliğine sahip olmalıdır. Yapılan analizler sonucunda sınavla öğrenci alımı yapan akademik liselerin hem ortalama okul başarısı; hem de düşük okul içi öğrenci yeteneği değişkenliği ile okul etkililiği bakımından genel liselerden daha iyi oldukları açıkça görülmüştür.

Öneriler

Modelin kullanımına ilişkin öneriler. Simülasyon çalışmasından elde edilen sonuçlar ışığında, önerilen modelin düşük grup içi varyans heterojenliği koşullarında grup içi varyansları yeterince etkili bir biçimde kestiremediği görülmüştür. Bu açıdan, araştırmacıların analiz sonuçlarına dikkat ederek heterojenlik düzeyinin düşük $(\tau \leq 0.02)$ çıkması halinde önerilen modelden ziyade geleneksel ÇDMTM'yi kullanmaları ve ortak bir okul içi varyans bileşeni kestirmeleri önerilebilir. Açıkçası kavramsal açıdan düşünüldüğünde, düşük heterojenlik düzeyinde bu tercih makul olarak kabul edilebilir. Diğer bir ifadeyle, heterojenlik düzeyinin oldukça düşük çıkması, grup içi varyansların ortalama değerine ve dolayısıyla birbirlerine oldukça yakın olduğunu ifade eder. Ancak burada vurgulanması gereken nokta, çalışma kapsamında heterojenlik düzeylerinin ifade edilmesinde kullanılan τ değerlerinin evrensel ölçütler olmadığıdır. Nitekim simülasyon çalışmasında düşük heterojenlik düzeyi için üretilen grup içi varyans parametrelerinin maksimum değeri, ortalamadan yaklaşık olarak 3.3 standart sapma ileride olmuştur. Dolayısıyla, yalnızca τ değerlerine bakılarak dağılımın homojene yakın ve dolayısıyla ranjının dar olduğu şeklinde bir karara varmak her zaman doğru olmayacaktır. Özellikle yüksek uç değerlerin olması, düşük heterojenlik düzeylerinde okul içi varyansların varyans parametresinin ve buna bağlı olarak tekil varyans değerlerinin kestirimlerini etkilemektedir.

Öte yandan, heterojenliğin düşük ile orta düzey arasında çıktığı gerçek veri analizi sonuçlarında, kestirilen maksimum okul içi varyans bileşeninin ortalamadan uzaklığı yaklaşık olarak 1.8 standart sapma olmuştur. Dolayısıyla simülasyon çalışmasında düşük heterojenlik düzeyi için gerçek veri koşullarından daha kötü senaryolara uygun veri setleri üzerinden çalışıldığı söylenebilir. Bu durum aslında grup içi varyansların üretildiği Gamma dağılımının sağa çarpık bir özellik göstermesinden kaynaklanmaktadır. Bu açıdan modelin aslında düşük heterojenlik gösteren gerçek veri setlerindeki parametre kestirimlerinin, bu koşul için yapılan simülasyon çalışmasından biraz daha etkili olacağı düşünülmektedir. Ancak yine de yukarıda belirtildiği gibi heterojenliğin oldukça düşük olması halinde, özellikle herhangi bir parametreye ait MCMC zincirlerinde olumsuz bir durum gözlenirse homojen modelin kullanılması daha sağlıklı olacaktır.

Orta heterojenlik düzeyinde, önerilen model tüm parametreleri genel anlamda iyi sayılabilecek bir performansla kestirebilmiştir. Uç değerlerdeki grup içi varyans parametrelerinin nokta kestirimleri Bayes yaklaşımı dolayısıyla ortalamaya doğru çekilme etkisi gösterse de, bu değerler için güven aralıkları oldukça etkili bir biçimde kestirilmiştir. Dolayısıyla araştırmacıların orta düzeyde grup içi varyans heterojenliği elde etmeleri halinde, özellikle model ile kestirdikleri yüksek grup içi varyans parametrelerini standart hataları ile birlikte değerlendirmeleri daha sağlıklı olacaktır. Yüksek grup içi heterojenliğin tespit edilmesi halinde de benzer öneriler geçerlidir.

Modelin kullanımı ile ilgili olarak vurgulanması gereken bir diğer nokta da, çalışma kapsamında yürütülen simülasyonda, heterojenlik düzeyinin dışındaki koşulların sabit tutulmuş olmasıdır. Dolayısıyla modeli kullanacak araştırmacıların kendi veri koşullarını da dikkate almaları önerilmektedir. Özellikle okul içi en düşük öğrenci sayısı ve toplam okul sayısının, önerilen modelin parametre kestirim performansını etkileyebileceği düşünülmektedir.

İleriye dönük araştırmalara ilişkin öneriler. Önerilen model ile ilişkili olarak yapılabilecek bazı araştırma önerileri aşağıda sunulmuştur.

1) Bu çalışmada modelin parametre kestirim performansı, diğer tüm koşullar optimum değerlerde tutularak, yalnızca değişen grup içi heterojenlik düzeyleri altında incelenmiştir. Grup sayısı, grup içi birey sayısı ve madde sayısının da değişimlendiği daha geniş simülasyon çalışmaları ile modelin çaprazlanmış koşullar altındaki parametre kestirim performansının ortaya konulmasının, modeli kullanacak araştırmacılar açısından oldukça faydalı sonuçlar üreteceği düşünülmektedir.

2) Önerilen modelde Bayes yaklaşımı çerçevesinde Kasim ve Raudenbush'ta (1998) olduğu gibi grup içi varyansların tersinin Gamma dağılımı gösterdiği kabul edilmiştir. Ancak yapılacak diğer çalışmalarda, varyans bileşenlerinin pozitif olmasını garantileyen ve alanyazında başka araştırmacılar tarafından kullanılan log-normal gibi dağılımlar da kullanılabilir. Bu şekilde farklı dağılımların kullanıldığı simülasyon çalışmaları yapılarak, değişen heterojenlik düzeylerinde hangi dağılımın kullanılmasının daha etkili olacağı ortaya konulabilir. Bu durumun özellikle düşük heterojenlik koşullarında modelin gösterdiği düşük kestirim performansının düzeltilmesi konusunda önemli bulgular sağlayacağı düşünülmektedir.

3) Modelin parametreleri için bu çalışmada kullanılanların dışında belirlenecek farklı önsel dağılımlar kullanılarak elde edilen sonuçlar karşılaştırılabilir. Farklı önsel dağılımlar kullanılarak, parametre kestirimlerinin bu durumdan ne şekilde etkilendiklerinin araştırılması, Bayes alanyazınında oldukça önemli olup hassasiyet analizi (sensitivity analysis) olarak anılmaktadır.

4) Bu çalışmada Bayes yaklaşımı ile kestirimler için MCMC yöntemlerinden Gibbs örneklemesi kullanılmıştır. Özellikle son yıllarda daha etkili kestirimler yaptığı belirtilen Hamiltonian MC örneklemesi ile modelin parametre kestirim performansı incelenerek Gibbs örneklemesi ile bir karşılaştırma yapılabilir. Anılan kestirimleri yapabilen popüler programlardan biri de Stan'dır (Stan Development Team, 2017).

5) Önerilen model, birey ve/veya okul düzeyinde açıklayıcı değişkenlerin içerildiği koşullu ortalama modelleri şeklinde de genişletilebilir. Örneğin okula ait sosyo-ekonomik düzey gibi bir açıklayıcı değişken alınarak ortalama okul yeteneğinin bu değişken ile açıklandığı ve aynı zamanda ayrı okul içi varyansların kestirildiği

heterojen bir model kurulabilir. Böylece okul etkililiği ile ilgili daha kapsamlı araştırma problemleri üzerinde çalışılabilir.

6) Bu çalışmada gerçek veri analizlerinde 2009 ÖBBS verilerinin A kitapçığındaki Türkçe testi ele alınmıştır. Önerilen model diğer derslerin verilerine uygulanarak okulların yetenek bakımından heterojenlikleri incelenebilir. Ayrıca koşullu varyans modeli yine okul türlerine göre oluşturularak, sınavla öğrenci alımı yapan lise türlerinin diğer derslerdeki etkililikleri de ortaya konulabilir.

7) Önerilen model ayrıca PISA ve TIMSS gibi geniş ölçekli değerlendirme programlarının 1-0 şeklindeki hiyerarşik ham verilerine uygulanarak okulların etkililiği hakkında yapısal eşitlik modelleri, geleneksel ÇDMTM ve HLM’ye göre daha detaylı bulgular elde edilebilir.

Kaynaklar

Acar, T. (2008). *Maddenin farklı fonksiyonlaşmasını belirlemede kullanılan genelleştirilmiş aşamalı doğrusal modelleme, lojistik regresyon ve olabilirlik oranı tekniklerinin karşılaştırılması.* Yayımlanmamış Doktora Tezi. Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.

Acar, T. ve Kelecioğlu, H. (2010). Maddenin farklı fonksiyonlaşmasını belirleme tekniklerinin karşılaştırılması: GADM, LR ve MTK-OO. *Kuram ve Uygulamada Eğitim Bilimleri, 10*(2), 625-649.

Adams, R. J., Wilson, M, & Wu, M. (1997). Multilevel item response models: An approach to errors in variables regression. *Journal of Educational and Behavioral Statistics, 22*(1), 47-76.

Asparouhov, T., & Muthèn, B. (2014). Auxiliary variables in mixture modeling: Three step approaches using Mplus. *Structural Equation Modeling: A Multidisciplinary Journal, 21*, 329-341.

Atalay-Kabasakal, K. (2014). *Değişen madde fonksiyonunun test eşitlemeye etkisi.* Yayımlanmamış Doktora Tezi. Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.

Atar, B. (2010). Basit doğrusal regresyon analizi ile hiyerarşik doğrusal modeller analizinin karşılaştırılması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme, 1*(2), 78-84.

Binici, S. (2007). *Random-effect differential item functioning via hierarchical generalized linear model and generalized linear latent mixed model: A comparison of estimation methods.* Unpublished Doctoral Dissertation. Florida State University.

Brooks, S. P., & Gelman, A. (1998). Alternative methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics, 7*, 434-455.

Browne, W. J., & Draper, D. (2000). Implementation and performance issues in the Bayesian and likelihood fitting of multilevel models. *Computational Statistics 15*, 391–420.

Browne, W. J., Draper, D., Goldstein, H., & Rasbash, J. (2002). Bayesian and likelihood methods for fitting multilevel models with complex level-1 variation. *Computational Statistics & Data Analysis, 39*, 203–225.

Brune, K. D. (2011). *An evaluation of item difficulty and person ability estimation using the multilevel measurement model with short tests and small sample sizes*. Unpublished Doctoral Dissertation, University of Texas at Austin.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park: Sage Publications.

Cheong, Y. F. (2006). Analysis of school context effects on differential item functioning using hierarchical generalized linear models. *International Journal of Testing, 6*(1), 57-79.

Chu, K. L. (2002). *Equivalent group test equating with the presence of differential item functioning.* Unpublished Doctoral Dissertation. Florida State University.

Denwood, M. J. (2016). runjags: An R package providing interface utilities, model templates, parallel computing methods and additional distributions for MCMC models in JAGS. *Journal of Statistical Software, 71*(9), 1-25.

Dowling, N. M. (2006). *An investigation of the robustness of multilevel item response theory models to violations of distributional assumptions*. Unpublished Doctoral Dissertation. University of Wisconsin-Madison.

Fischer, G. H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica, 37*, 359-374.

Fox, J.-P. (2001). Multilevel IRT: *A Bayesian perspective on estimating parameters and testing statistical hypotheses.* PhD dissertation. University of Twente.

Fox, J.-P. (2004). Applications of multilevel IRT modeling. *School Effectiveness and School Improvement, 15*, 261-280.

Fox, J. -P. (2010). *Bayesian item response modeling: Theory and applications*. New York: Springer.

Fox, J. -P., & Glas, C. (2001). Bayesian estimation of a multilevel IRT model using Gibbs sampling. *Psychometrika, 66*, 271-288.

Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika, 73*, 43-56.

Goldstein, H. (2011). *Multilevel statistical models (4th edition)*. Chicester: Wiley.

Hedeker, D., Mermelstein, R. J., & Demirtas, H. (2008). An application of a mixed-effects location scale model for analysis of ecological momentary assessment (EMA) data. *Biometrics, 64*(2), 627–634.

Hoffman, L. (2007). Multilevel models for examining individual differences in within-person variation and covariation over time. *Multivariate Behavioral Research, 42*(4), 609-629.

Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling: An overview and a meta-analysis. *Sociological Methods & Research, 26*, 329-367.

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications (2nd edition).* New York: Routledge.

Jahng, S. (2008). *Multilevel models for intensive longitudinal data with heterogeneous error structure: Covariance transformation and variance function models.* Unpublished Doctoral Dissertation. University of Missouri.

Kamata, A. (1998). *Some generalizations of the Rasch model: An application of the hierarchical generalized linear model.* Unpublished Doctoral Dissertation. Michigan State University.

Kamata, A. (2001). Item analysis by the hierarchical generalized linear model. *Journal of Educational Measurement, 38*(1), 79-93.

Kamata, A., & Binici, S. (2003, June). *Random-effect DIF analysis via hierarchical generalized linear models.* Paper presented at the International Meeting of the Psychometric Society: The 68st annual meeting of the Psychometric Society, Sardinia, Italy.

Kang, T.-H., & Cohen, A.S. (2005, April). *IRT model selection methods for polytomous items*. Paper presented at the annual meeting of the National Council on Measurement in Education, Montreal, Canada.

Kasim, R. M., & Raudenbush, S. W. (1998). Application of Gibbs sampling to nested variance components models with heterogeneous within-group variance. *Journal of Educational and Behavioral Statistics, 23*(2), 93-116.

Kéry, M. (2010). *Introduction to WinBUGS for Ecologists: Bayesian approach to regression, ANOVA, mixed models and related analyses.* Burlington, MA: Academic Press.

Kim, I. (2007). *A comparison of a Bayesian and maximum likelihood algorithms for estimation of a multilevel IRT model.* Unpublished Doctoral Dissertation. University of Georgia, Athens.

Kim, J. -O. (2005). *Examining residual variance heterogeneity in experimental and quasi-experimental settings: Modeling mean and dispersion structures in a hierarchical modeling framework.* Unpublished Doctoral Dissertation. University of California, Los Angeles.

Kim, J., & Choi, K. (2008). Closing the gap: Modeling within-school variance heterogeneity in school effect studies. *Asia Pacific Education Review, 9*, 206–220.

Kruschke, J. K. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan (2nd edition).* New York, Academic Press.

Kuppens, T., & Yzerbyt, V. Y. (2014). Predicting variability: Using multilevel modeling to assess differences in variance. *European Journal of Social Psychology, 44*(7), 691-700.

Leckie, C., French, R., Charlton, C., & Browne, W. (2014). Modeling heterogeneous variance-covariance components in two-level models. *Journal of Educational and Behavioral Statistics, 39*(5), 307-332.

Li, X., & Hedeker, D. (2012). A three-level mixed-effects location scale model with an application to ecological momentary assessment (EMA) data. *Stat Med., 31*(26), 3192-3210.

Link, W. A., & Eaton, M. J. (2012). Forum on thinning of chains in MCMC. *Methods in Ecology and Evolution, 3*, 112-115.

Longford, N. T. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika, 74*, 817-827.

Lunn, D., Jackson, C., Best, N., Thomas, A., & Spiegelhalter, D. (2012). *The BUGS book: A practical introduction to Bayesian analysis*. Boca Raton, FL: Chapman & Hall/CRC.

Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS-A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing, 10*, 325–337.

Maas, C. J. M., & Hox, J. J., (2004a). Robustness issues in multilevel regression analysis. *Statistica Neerlandica, 58*(2), 127-137.

Maas, C. J. M. & Hox, J. J. (2004b). The influence of violations of assumptions on multilevel parameter estimates and their standard errors. *Computational Statistics and Data Analysis, 46*(3), 427-440.

Maier, K. S. (2000). *Applying Bayesian methods to hierarchical measurement models.* Unpublished Doctoral Dissertation, University of Chicago.

Maier, K. S. (2001). A Rasch hierarchical measurement model. *Journal of Educational and Behavioral Statistics, 26*(3), 307-330.

Mellenbergh, G. J. (1994). Generalized linear item response theory. *Psychological Bulletin, 115*(2), 300-307.

Mislevy, R. J. (1987). Exploiting auxiliary information about examinees in the estimation of item parameters. *Applied Psychological Measurement, 5*(3), 369-376.

Moyer, E. (2013). *Examining the invariance of item and person parameters estimates from multilevel measurement models when distribution of person parameters are non-normal.* Unpublished Doctoral Dissertation. The University of Texas at Austin.

Myles, J. P., Price, G. M., Hunter, N., Day, M., & Duffy, S. W. (2003). A potentially useful distribution model for dietary intake data. *Public Health Nutrition 6*, 513–519.

Natesan, P. (2007). *Two-parameter multilevel item-response models: A simulation study of test length and sample size influences*. Unpublished Doctoral Dissertation. Texas A&M University, College Station.

Ntzoufras, I. (2009). *Bayesian modeling using WinBUGS*. Hoboken, NJ: Wiley.

Nunnally, J. C. (1994). *Psychometric theory (3rd edition).* New York: McGraw Hill.

Pastor, D. A. (2001). *Application of multilevel IRT modeling to the study of self-esteem in adolescents.* Unpublished Doctoral Dissertation. University of Texas at Austin.

Pastor, D. A. (2003). The use of multilevel item response theory modeling in applied research: An illustration. *Applied Measurement in Education, 16*(3), 223-243.

Patarapichayatham, C., Kamata, A., & Kanjanawasee, S. (2012). Evaluation of model selection strategies for cross-level two-way differential functioning analysis. *Educational and Psychological Measurement, 72*(1), 44-51.

Plummer, M. (2003). *JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling*. In Proceedings of the 3rd international workshop on distributed statistical computing (dsc 2003), Vienna, Austria.

Plummer, M. (2015). *JAGS version 4.0.0 user manual* [Computer software manual]. (October 1, 2015).

Pugach, O. (2012). *A bivariate location-scale mixed-effects model*. Unpublished Doctoral Dissertation. University of Illinois at Chicago.

Rast, P., & Zimprich, D. (2011). Modeling within-person variance in reaction time data of older adults. *GeroPsych, 24*(4), 169-176.

Rast, P., Hofer, S. M., & Sparks, C. (2012). Modeling individual differences in within-person variation of negative and positive affect in a mixed effects location scale model using BUGS/JAGS. *Multivariate Behavioral Research, 47*, 177–200.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage Publications, Inc.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, R. T. (2005). HLM 6:Hierarchical linear and nonlinear modeling. Lincolnwood, IL: Scienctific software.

R Core Team (2016). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Robert, C. (2007). *The Bayesian choice (2nd edition)*. New York: Springer-Verlag.

Sani, C., & Grilli, L. (2011). Differantial variability of test scores among schools: A multilevel analysis of the fifth-grade INVALSI test using heteroscedastic random effects. *Journal of Applied Quantitative Methods, 6*(4), 88-99.

SAS Institute Inc. (2000). SAS/STAT User's Guide, Version 8. Cary, NC: SAS Publishing.

Schmitt, T. A. (2007). *An evaluation of parameter recovery using the 1-parameter hierarchical generalized linear logistic model for dichotomous responses.* Unpublished Doctoral Dissertation. University of Wisconsin-Milwaukee.

Shieh, Y. Y. (1999). *An evaluation of mixed effects multilevel modeling under conditions of error term nonnormality.* Unpublished doctoral dissertation, The University of Texas, Austin, Texas.

Snijders, T. A. B., & Bosker, R. J. (1999). Multilevel *analysis: An introduction to basic and advanced multilevel modeling.* London: Sage.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B-Statistical Methodology, 64*, 583-616.

Stan Development Team (2017). *Stan Modeling Language Users Guide and Reference Manual*, Version 2.17.0. `http://mc-stan.org`.

Stevens, J. P. (2009). *Applied multivariate statistics for the social sciences (5th edition)*. New York: Routledge.

Sun, M. (2011). *Exploring how school intra-organizational mechanisms mediate the effects of external interventions on improving teaching and learning.* Unpublished Doctoral Dissertation. Michigan State University.

Turhan, A. (2006). *Multilevel 2PL item response model vertical equating with the presence of differential item functioning.* Unpublished Doctoral Dissertation. Florida State University.

Vallejo, G., Ato, M., Fernandez, M. P., & Livacic-Rojas, P. E. (2013) Multilevel bootstrap analysis with assumptions violated. *Psicothema, 25*, 520–528.

Vaughn, B. K. (2006). *A hierarchical generalized linear model of random differential item functioning for polytomous items: A Bayesian multilevel approach.* Unpublished Doctoral Dissertation. Florida State University.

Walther, B. A. & Moore, J. L. (2005). The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. *Ecography, 28*, 815-829.

Zhang, F., & Weiss, R. E. (2000). Diagnosing explainable heterogeneity of variance in random-effects models. *The Canadian Journal of Statistics, 28*(1), 3-18.

Zheng, X. (2009). *Multilevel item response modeling: Applications to large-scale assessment of academic achievement.* Unpublished Doctoral Dissertation. University of California, Berkeley.

Zwinderman, A. H. (1991). A generalized Rasch model for manifest predictors. *Psychometrika, 56*(4), 589-600.

EK-A: Koşulsuz Varyans Modeline Ait JAGS Kodu

```
model{

    for(n in 1:N){
    for (i in 1:I){
    response[n,i]~dbern(prb[n,i]) #0-1 şeklindeki veri
    logit(prb[n,i]) <- ejk[n] + uk[sch[n]] - dif[i] #logit dönüşümü ile doğrusal model
    }

    ejk[n]~dnorm(0, prec[n]) #okul ortalamasından arındırılmış öğrenci yetenekleri
    prec[n] <- prc[sch[n]]
    }

    for(k in 1:K){
    uk[k]~dnorm(0, prec.sch) #okul düzeyi yetenekler
    prc[k]~dgamma(shp, rate) #okul içi varyansların tersi
    sig.ek[k] <- 1/prc[k]
    }

    for (i in 1:I){
    dif[i]~dnorm(mu.dif, prec.dif) #madde güçlükleri
    }

    prec.sch~dgamma(0.1, 0.1)
    sig.u <- 1/prec.sch #okul yeteneklerinin varyansı
    shp <- 1/(2*tau)
    tau ~ dunif(0, 100)
    rate <- 1/scl
    scl <- (2*tau)/mu
    mu~dunif(0, 100)
    mu.dif~dnorm(0, 0.01)
    prec.dif~dgamma(0.1,0.1)
    }
```

NOT: JAGS ile analiz sırasında “glm” modülünün aktif olması, zincirlerdeki oto-korelasyonların daha düşük olmasını ve böylece yakınsamanın daha hızlı gerçekleşmesini sağlamaktadır. Öte yandan paralel analiz seçeneği ile birden fazla çekirdeğin çalıştırılması, toplam analiz süresini oldukça düşürmektedir.

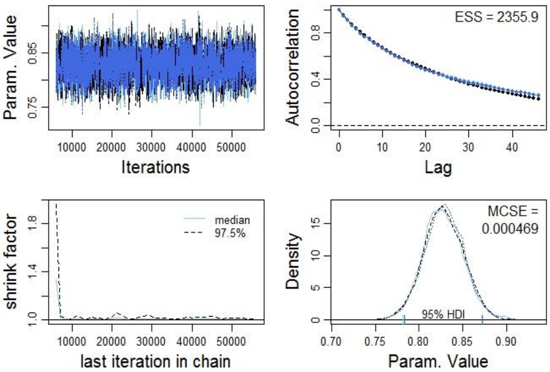
EK-B: Koşullu Varyans Modeline Ait JAGS Kodu

```
model{

      for(n in 1:N){
      for (i in 1:I){
      response[n,i]~dbern(prb[n,i]) #0-1 şeklindeki veri
      logit(prb[n,i]) <- ejk[n] + uk[sch[n]] - dif[i] #logit dönüşümü ile doğrusal model
      }

      ejk[n]~dnorm(0, prec[n]) #okul ortalamasından arındırılmış öğrenci yetenekleri
      prec[n] <- prc[sch[n]]
      }

      for(k in 1:K){
      uk[k]~dnorm(0, prec.sch) #okul düzeyi yetenekler
      prc[k] <- 1/sig.ek[k]
      log(sig.ek[k]) <- log_sig.e[k]
      log_sig.e[k]~dnorm(zeta[k], omega)
     zeta[k] <- alpha_zero + alpha_one * okul_turu[k] #koşullu varyans modeli
      }

      for (i in 1:I){
      dif[i]~dnorm(mu.dif, prec.dif) #madde güçlükleri
      }

      prec.sch~dgamma(0.1, 0.1)
      sig.u <- 1/prec.sch #okul yeteneklerinin varyansı
      alpha_zero~dunif(-2.5, 0)
      alpha_one~dunif(-0.5, 0)
      mean_var_zero <- exp(alpha_zero) #kodu 0 olan okulların ortalama varyansı
      mean_var_one <- exp(alpha_zero + alpha_one) #kodu 1 olan okulların
ortalama varyansı
      mean_var_ratio <- mean_var_one/mean_var_zero #okulların ortalama
varyanslarının oranı
      mu.dif~dnorm(0, 0.01)
      prec.dif~dgamma(0.1,0.1)
      omega~dgamma(0.1, 0.1)
      }
```

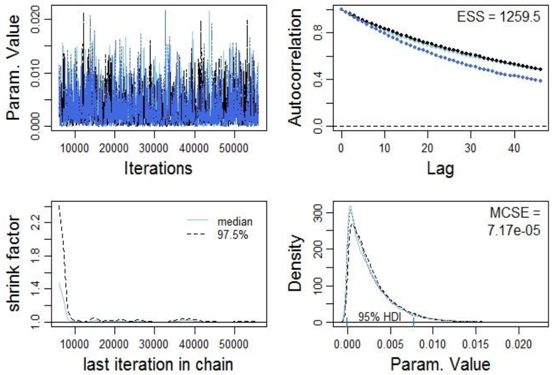
NOT: JAGS ile analiz sırasında "glm" modülünün aktif olması, zincirlerdeki oto-korelasyonların daha düşük olmasını ve böylece yakınsamanın daha hızlı gerçekleşmesini sağlamaktadır. Öte yandan paralel analiz seçeneği ile birden fazla çekirdeğin çalıştırılması, toplam analiz süresini oldukça düşürmektedir.

EK-C: Düşük Heterojenlik Düzeyi İçin Yakınsama Bilgi Grafikleri

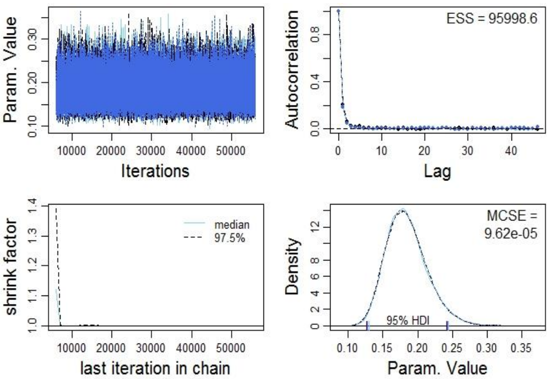
Okul içi varyansların ortalaması (μ)



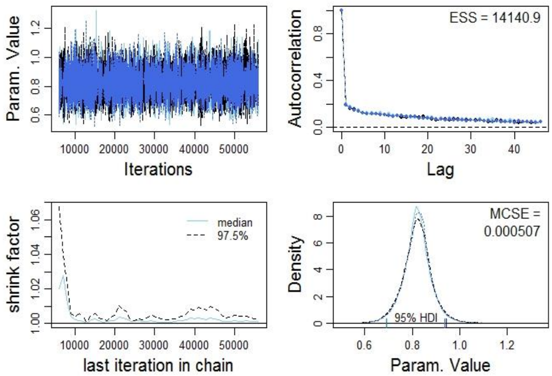
Okul içi varyansların varyansı (τ)



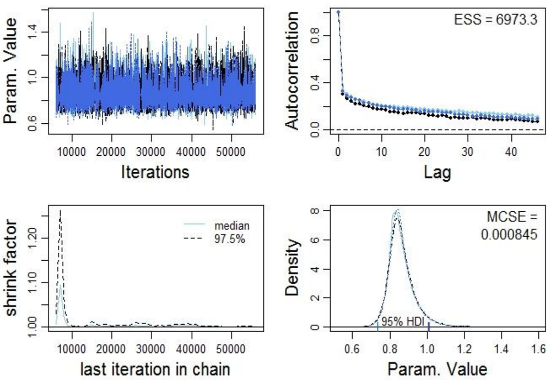
Okul yeteneklerinin varyansı (σ_u^2)



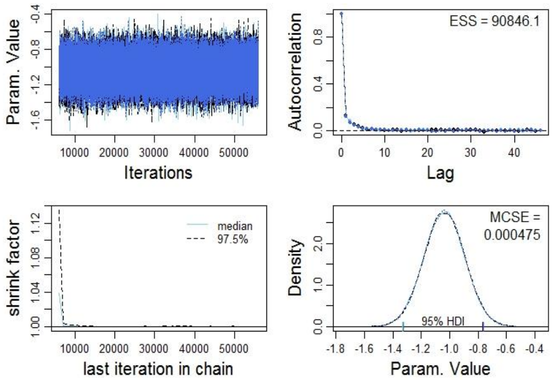
Minimum okul içi varyans $(\sigma_{e_k}^2\ min)$



Maksimum okul içi varyans $(\sigma_{e_k}^2\ max)$

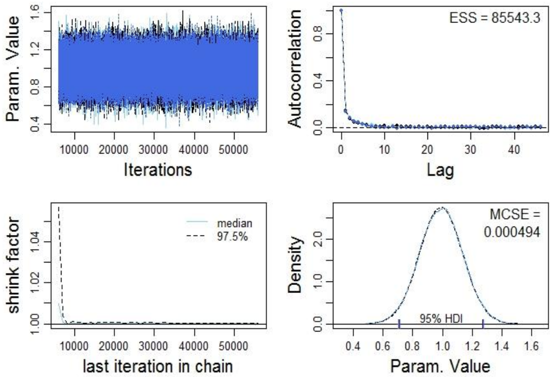


Minimum okul yeteneği $(u_k\ min)$

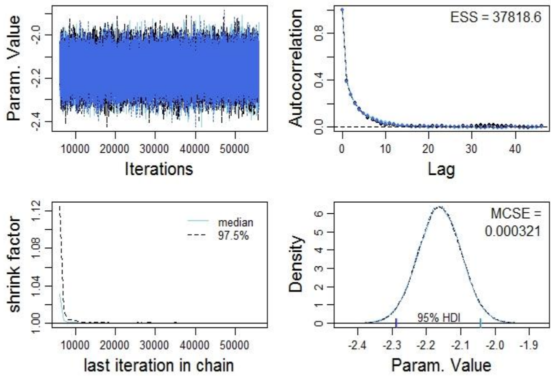


EK-C Devam,

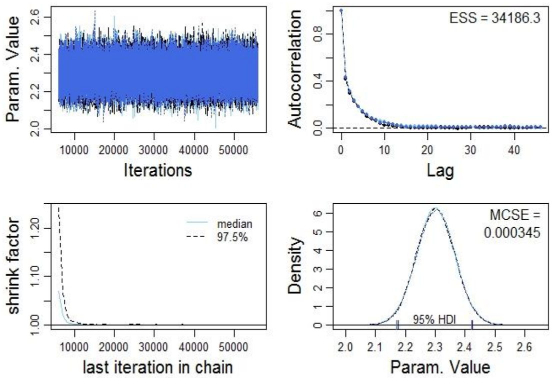
Maksimum okul yeteneği $(u_k\ max)$



Minimum madde güçlüğü $(b_i\ min)$

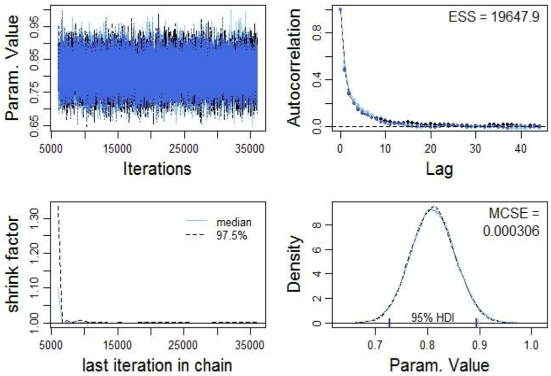


Maksimum madde güçlüğü $(b_i\ max)$

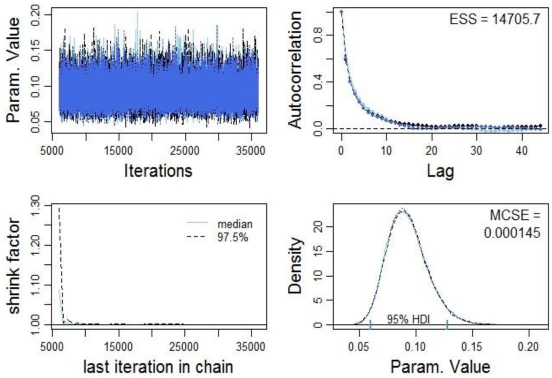


EK-Ç: Orta Heterojenlik Düzeyi İçin Yakınsama Bilgi Grafikleri

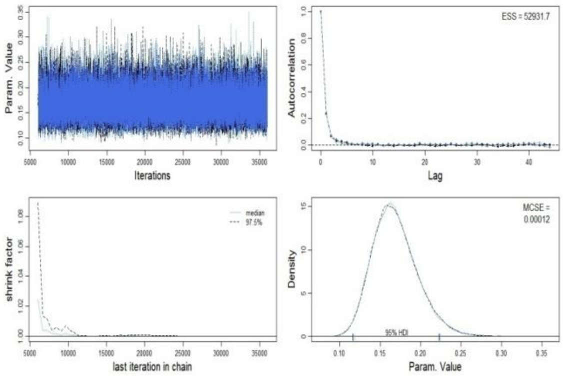
Okul içi varyansların ortalaması (μ)



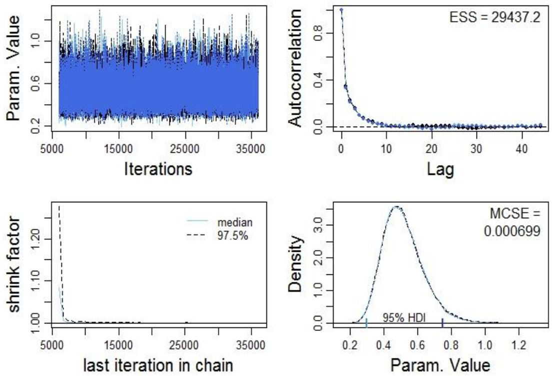
Okul içi varyansların varyansı (τ)



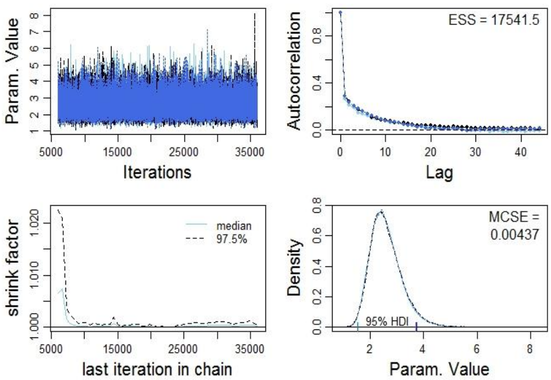
Okul yeteneklerinin varyansı (σ_u^2)



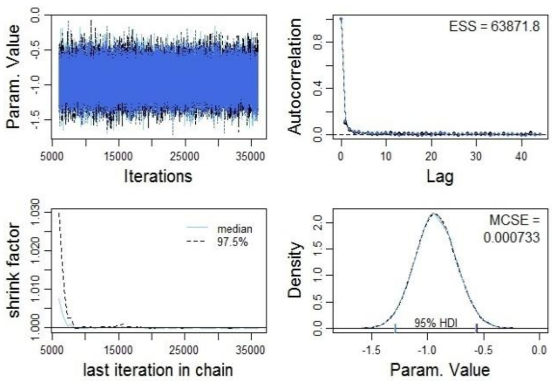
Minimum okul içi varyans $(\sigma_{e_k}^2\ min)$



Maksimum okul içi varyans $(\sigma_{e_k}^2\ max)$

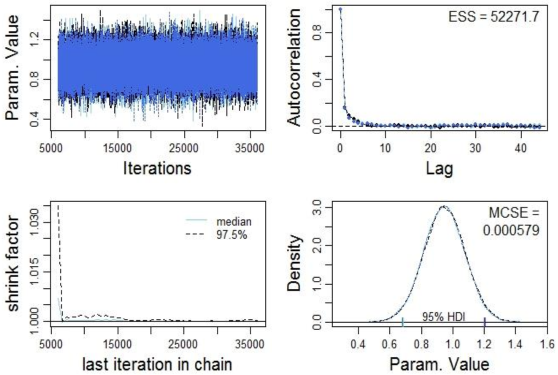


Minimum okul yeteneği $(u_k\ min)$

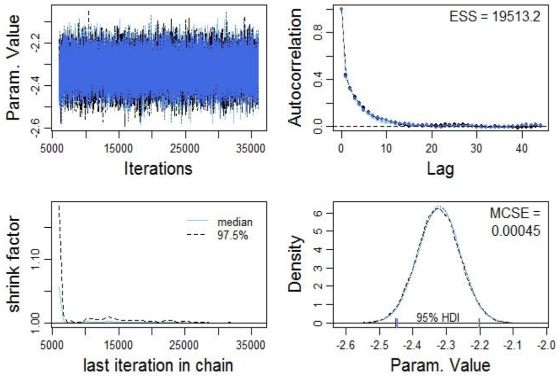


EK-Ç Devam,

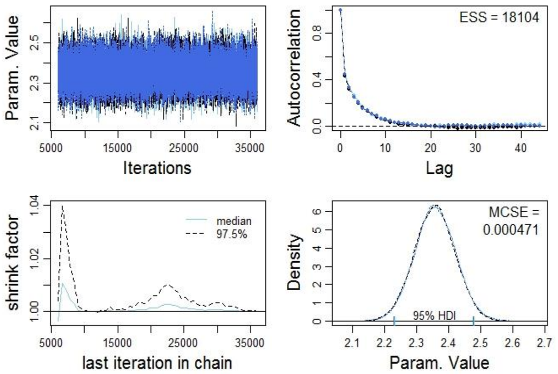
Maksimum okul yeteneği $(u_k\ max)$



Minimum madde güçlüğü $(b_i\ min)$

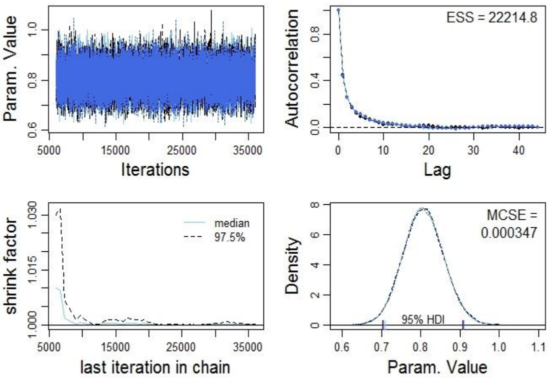


Maksimum madde güçlüğü $(b_i\ max)$

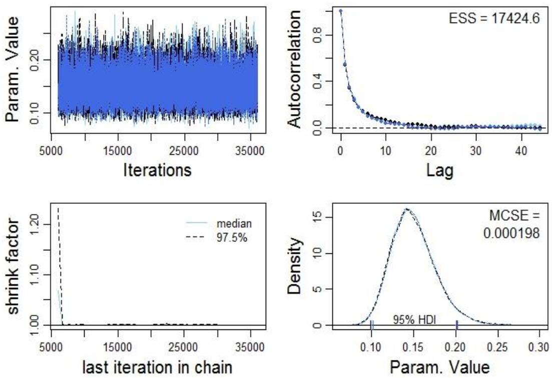


EK-D: Yüksek Heterojenlik Düzeyi İçin Yakınsama Bilgi Grafikleri

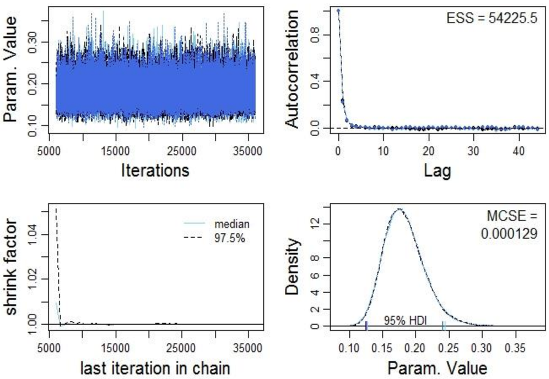
Okul içi varyansların ortalaması (μ)



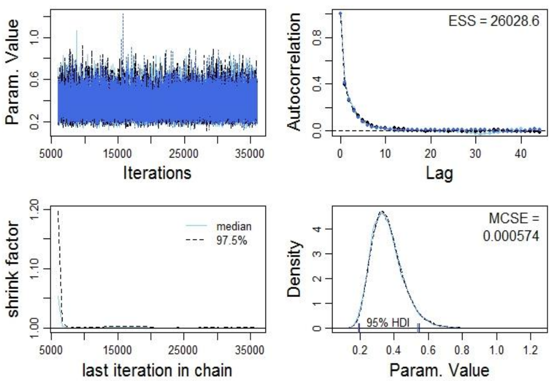
Okul içi varyansların varyansı (τ)



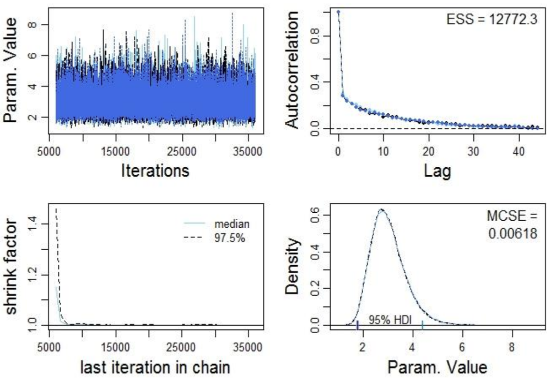
Okul yeteneklerinin varyansı (σ_u^2)



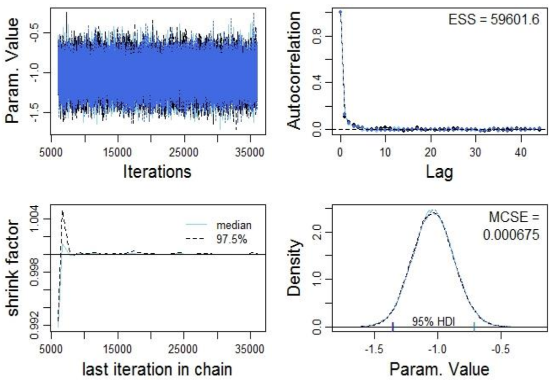
Minimum okul içi varyans ($\sigma_{e_k}^2\ min$)



Maksimum okul içi varyans ($\sigma_{e_k}^2\ max$)

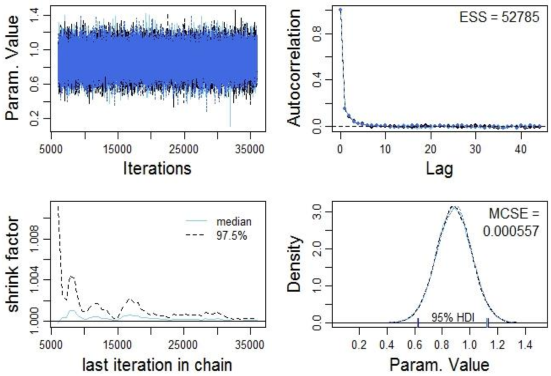


Minimum okul yeteneği ($u_k\ min$)

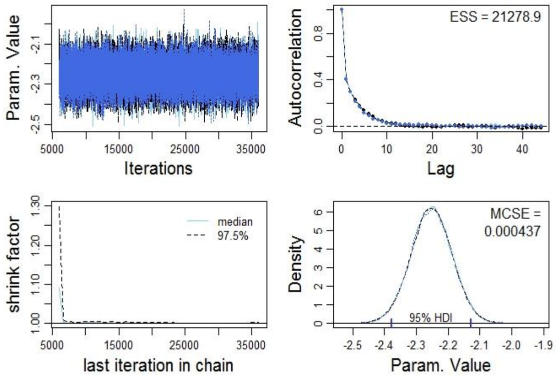


EK-D Devam,

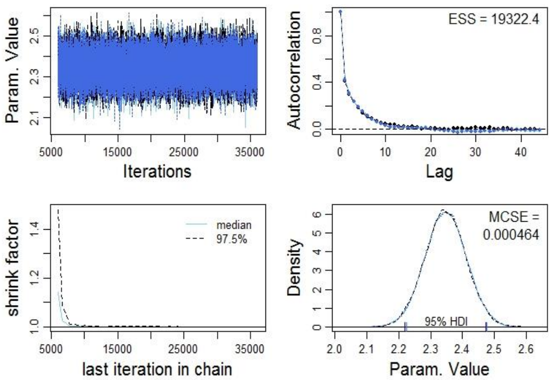
Maksimum okul yeteneği $(u_k\ max)$



Minimum madde güçlüğü $(b_i\ min)$

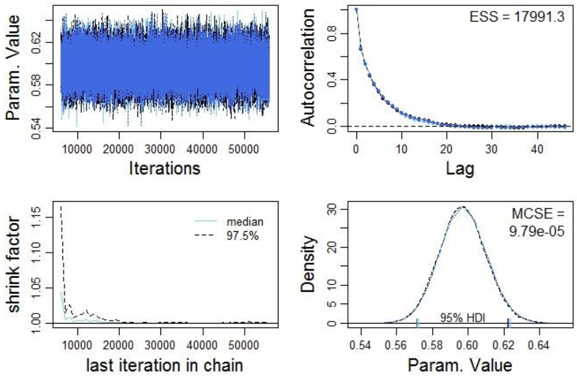


Maksimum madde güçlüğü $(b_i\ max)$

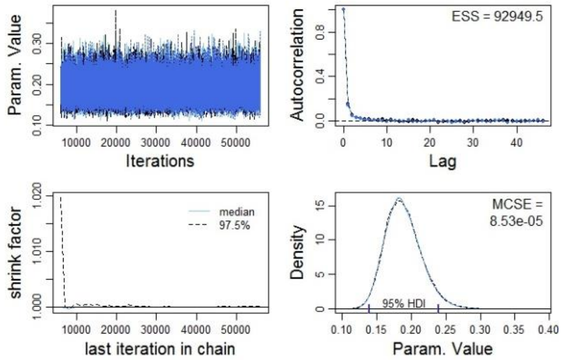


EK-E: Gerçek Verinin Geleneksel Model İle Analizine İlişkin Yakınsama Bilgi Grafikleri

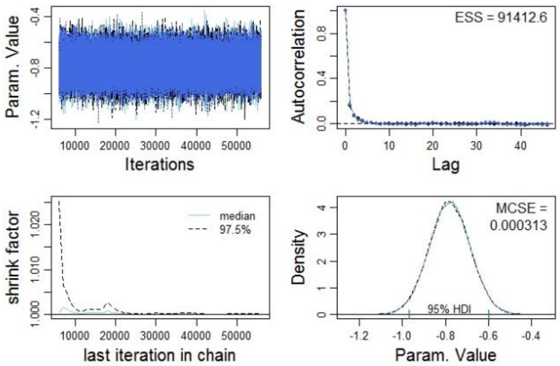
Okul içi varyans (σ_e^2)



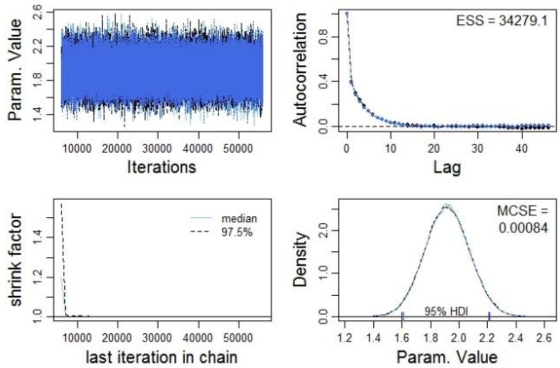
Okul yeteneklerinin varyansı (σ_u^2)



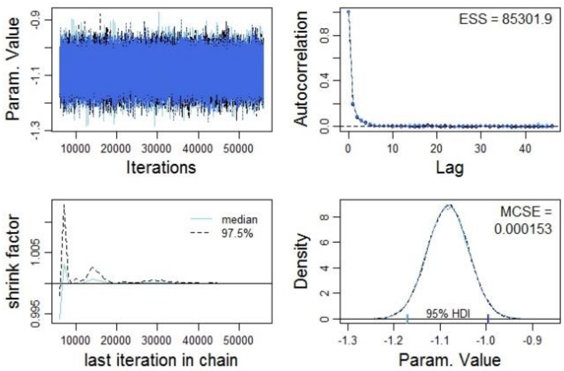
Minimum okul yeteneği ($u_k\ min$)



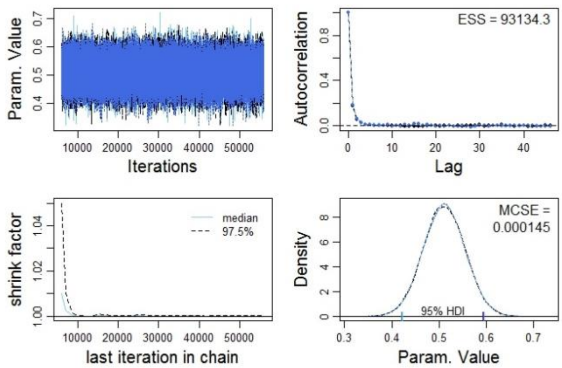
Maksimum okul yeteneği ($u_k\ max$)



Minimum madde güçlüğü ($b_i\ min$)

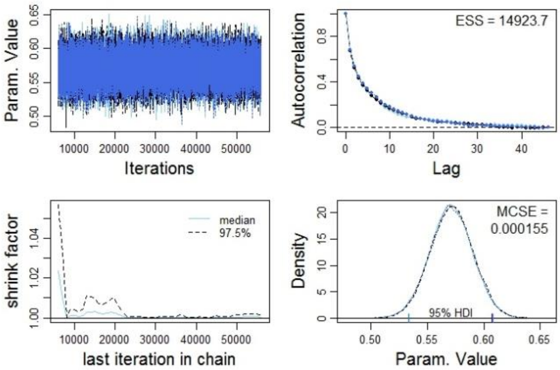


Maksimum madde güçlüğü ($b_i\ max$)

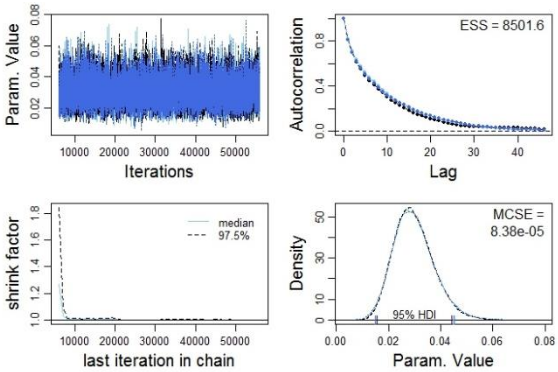


EK-F: Gerçek Verinin Önerilen Model ile Analizine İlişkin Yakınsama Bilgi Grafikleri

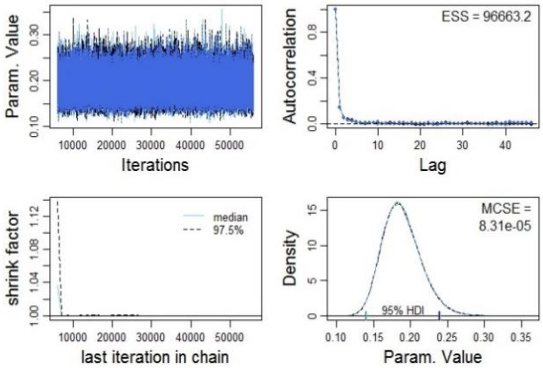
Okul içi varyansların ortalaması (μ)



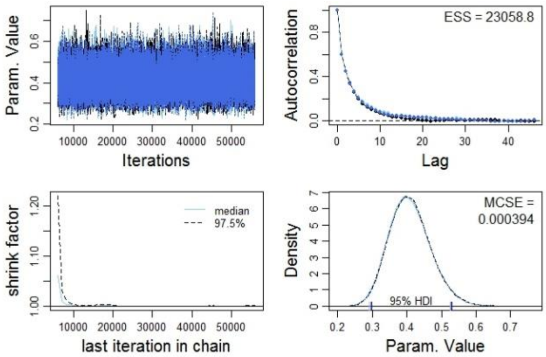
Okul içi varyansların varyansı (τ)



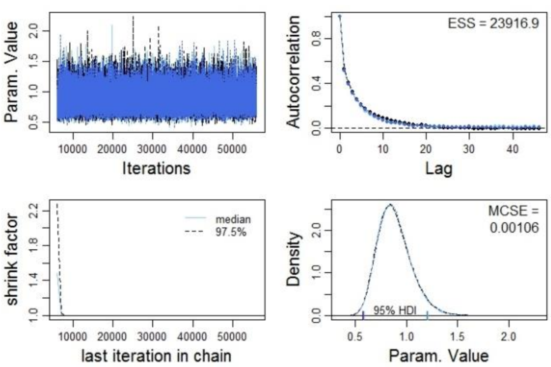
Okul yeteneklerinin varyansı (σ_u^2)



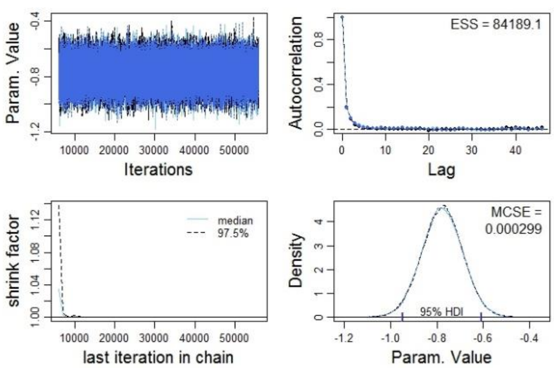
Minimum okul içi varyans ($\sigma_{e_k}^2\ min$)



Maksimum okul içi varyans ($\sigma_{e_k}^2\ max$)

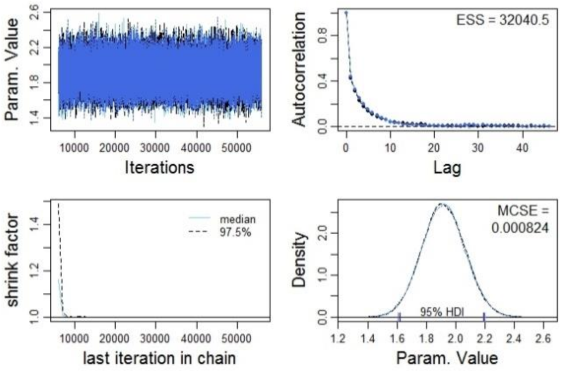


Minimum okul yeteneği ($u_k\ min$)

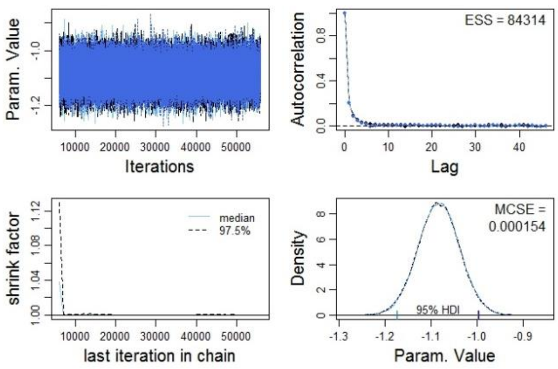


EK-F Devam,

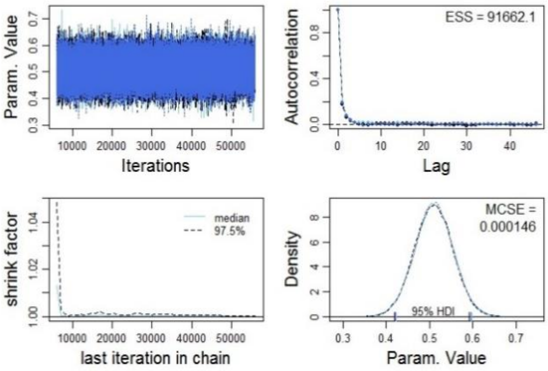
Maksimum okul yeteneği $(u_k\ max)$



Minimum madde güçlüğü $(b_i\ min)$



Maksimum madde güçlüğü $(b_i\ max)$



EK-G: 2009 ÖBBS Türkçe Testi Okul Yetenek Kestirimleri

Okul	*Homojen Model*		*Heterojen Model*	
No	*Kes.*	*%95 HDI*	*Kes.*	*%95 HDI*
1	0.805	[0.639, 0.971]	0.797	[0.633, 0.959]
2	0.484	[0.272, 0.698]	0.483	[0.268, 0.694]
3	0.04	[-0.174, 0.262]	0.044	[-0.19, 0.281]
4	0.013	[-0.193, 0.212]	0.02	[-0.207, 0.248]
5	0.383	[0.17, 0.598]	0.384	[0.159, 0.603]
6	-0.017	[-0.213, 0.188]	-0.015	[-0.22, 0.197]
7	-0.013	[-0.219, 0.197]	-0.011	[-0.223, 0.202]
8	-0.289	[-0.51, -0.074]	-0.29	[-0.514, -0.058]
9	0.586	[0.33, 0.848]	0.584	[0.336, 0.841]
10	0.46	[0.198, 0.729]	0.459	[0.198, 0.725]
11	0.111	[-0.09, 0.315]	0.116	[-0.104, 0.334]
12	0.139	[-0.096, 0.376]	0.14	[-0.108, 0.379]
13	0.613	[0.414, 0.818]	0.62	[0.404, 0.839]
14	0.173	[0.012, 0.334]	0.175	[0.009, 0.34]
15	0.003	[-0.162, 0.168]	0.003	[-0.168, 0.17]
16	-0.027	[-0.183, 0.126]	-0.027	[-0.183, 0.129]
17	0.265	[0.013, 0.513]	0.271	[0.011, 0.547]
18	0.075	[-0.113, 0.257]	0.074	[-0.117, 0.261]
19	-0.051	[-0.243, 0.147]	-0.047	[-0.266, 0.169]
20	-0.045	[-0.265, 0.178]	-0.044	[-0.278, 0.184]
21	0.031	[-0.181, 0.25]	0.036	[-0.196, 0.272]
22	-0.005	[-0.25, 0.245]	-0.004	[-0.252, 0.254]
23	0.101	[-0.141, 0.333]	0.101	[-0.138, 0.337]
24	-0.091	[-0.333, 0.149]	-0.092	[-0.335, 0.149]
25	-0.048	[-0.301, 0.208]	-0.049	[-0.304, 0.202]
26	0.08	[-0.077, 0.236]	0.079	[-0.079, 0.236]
27	0.15	[-0.059, 0.357]	0.15	[-0.066, 0.357]
28	-0.136	[-0.36, 0.089]	-0.138	[-0.366, 0.086]
29	-0.025	[-0.243, 0.194]	-0.028	[-0.237, 0.177]
30	-0.243	[-0.391, -0.092]	-0.243	[-0.382, -0.102]
31	0.202	[0.011, 0.389]	0.2	[0.017, 0.386]
32	-0.551	[-0.721, -0.386]	-0.549	[-0.713, -0.393]
33	-0.158	[-0.335, 0.017]	-0.157	[-0.347, 0.027]
34	0.353	[0.138, 0.566]	0.349	[0.14, 0.557]
35	-0.249	[-0.446, -0.058]	-0.251	[-0.432, -0.076]
36	0.287	[0.066, 0.514]	0.287	[0.052, 0.513]
37	0.618	[0.406, 0.837]	0.619	[0.401, 0.839]
38	0.224	[0.045, 0.408]	0.218	[0.052, 0.389]
39	0.023	[-0.165, 0.218]	0.023	[-0.167, 0.212]
40	1.917	[1.608, 2.214]	1.918	[1.625, 2.204]
41	-0.084	[-0.341, 0.171]	-0.083	[-0.344, 0.182]
42	-0.068	[-0.314, 0.181]	-0.069	[-0.31, 0.178]
43	-0.429	[-0.58, -0.279]	-0.428	[-0.564, -0.288]
44	-0.531	[-0.724, -0.344]	-0.53	[-0.708, -0.344]
45	0.406	[0.145, 0.666]	0.404	[0.146, 0.663]
46	-0.468	[-0.621, -0.309]	-0.468	[-0.625, -0.318]
47	-0.34	[-0.505, -0.17]	-0.34	[-0.503, -0.175]
48	-0.083	-0.285, 0.132]	-0.087	[-0.289, 0.115]
49	-0.382	[-0.564, -0.197]	-0.382	[-0.568, -0.196]
50	1.282	[1.026, 1.546]	1.282	[1.036, 1.519]
51	1.509	[1.217, 1.784]	1.504	[1.229, 1.781]
52	-0.719	[-0.925, -0.524]	-0.718	[-0.911, -0.529]

EK-G Devam,

Okul	Homojen Model		Heterojen Model	
No	Kes.	%95 HDI	Kes.	%95 HDI
53	-0.326	[-0.494, -0.16]	-0.326	[-0.5, -0.149]
54	0.123	[-0.045, 0.299]	0.124	[-0.053, 0.301]
55	0.275	[0.067, 0.477]	0.283	[0.059, 0.497]
56	-0.196	[-0.45, 0.058]	-0.195	[-0.459, 0.058]
57	0.017	[-0.16, 0.194]	0.023	[-0.171, 0.228]
58	-0.659	[-0.83, -0.484]	-0.659	[-0.829, -0.491]
59	0.057	[-0.134, 0.262]	0.054	[-0.139, 0.245]
60	-0.054	[-0.245, 0.133]	-0.054	[-0.249, 0.142]
61	-0.183	[-0.408, 0.034]	-0.183	[-0.407, 0.049]
62	0.159	[-0.081, 0.398]	0.157	[-0.077, 0.39]
63	0.154	[-0.025, 0.337]	0.155	[-0.027, 0.339]
64	0.31	[0.121, 0.504]	0.306	[0.124, 0.489]
65	-0.081	[-0.275, 0.105]	-0.081	[-0.277, 0.113]
66	-0.457	[-0.627, -0.286]	-0.457	[-0.615, -0.297]
67	-0.504	[-0.755, -0.246]	-0.504	[-0.756, -0.255]
68	0.318	[0.083, 0.552]	0.318	[0.089, 0.538]
69	0.436	[0.229, 0.648]	0.436	[0.222, 0.652]
70	0.108	[-0.115, 0.335]	0.108	[-0.119, 0.337]
71	0.095	[-0.095, 0.289]	0.095	[-0.1, 0.287]
72	0.184	[-0.007, 0.37]	0.184	[-0.017, 0.39]
73	-0.166	[-0.341, 0.011]	-0.166	[-0.354, 0.028]
74	0.481	[0.243, 0.723]	0.481	[0.231, 0.74]
75	-0.129	[-0.386, 0.12]	-0.129	[-0.377, 0.142]
76	0.766	[0.515, 1.012]	0.766	[0.516, 1.008]
77	0.118	[-0.053, 0.289]	0.118	[-0.062, 0.292]
78	0.557	[0.345, 0.764]	0.557	[0.349, 0.773]
79	0.323	[0.086, 0.564]	0.323	[0.1, 0.544]
80	-0.442	[-0.613, -0.268]	-0.442	[-0.604, -0.277]
81	0.006	[-0.212, 0.224]	0.006	[-0.212, 0.226]
82	0.605	[0.393, 0.824]	0.605	[0.394, 0.806]
83	-0.519	[-0.729, -0.318]	-0.519	[-0.72, -0.32]
84	0.156	[-0.083, 0.389]	0.156	[-0.085, 0.39]
85	-0.697	[-0.949, -0.45]	-0.697	[-0.943, -0.441]
86	-0.078	[-0.293, 0.14]	-0.078	[-0.294, 0.129]
87	0.377	[0.116, 0.648]	0.377	[0.108, 0.632]
88	-0.112	[-0.311, 0.089]	-0.112	[-0.314, 0.087]
89	-0.07	[-0.242, 0.102]	-0.07	[-0.24, 0.093]
90	-0.096	[-0.324, 0.134]	-0.096	[-0.335, 0.151]
91	-0.072	[-0.316, 0.176]	-0.072	[-0.319, 0.18]
92	-0.083	[-0.317, 0.15]	-0.083	[-0.322, 0.148]
93	0.09	[-0.099, 0.294]	0.09	[-0.116, 0.282]
94	-0.444	[-0.718, -0.193]	-0.444	[-0.699, -0.204]
95	0.371	[0.109, 0.629]	0.371	[0.111, 0.631]
96	0.126	[-0.101, 0.352]	0.126	[-0.103, 0.349]
97	-0.509	[-0.69, -0.334]	-0.509	[-0.675, -0.348]
98	0.235	[-0.03, 0.499]	0.235	[-0.034, 0.508]
99	-0.226	[-0.477, 0.024]	-0.226	[-0.477, 0.021]
100	0.449	[0.236, 0.656]	0.449	[0.245, 0.662]
101	-0.062	[-0.265, 0.14]	-0.062	[-0.285, 0.163]
102	0.134	[-0.107, 0.376]	0.134	[-0.123, 0.389]
103	-0.781	[-0.964, -0.595]	-0.781	[-0.95, -0.608]
104	0.236	[0.054, 0.417]	0.236	[0.052, 0.416]
105	-0.523	[-0.701, -0.34]	-0.523	[-0.698, -0.349]
106	-0.559	[-0.799, -0.331]	-0.559	[-0.784, -0.337]
107	-0.198	[-0.383, -0.016]	-0.198	[-0.382, -0.015]
108	-0.236	[-0.474, 0.008]	-0.236	[-0.476, -0.005]
109	-0.276	[-0.474, -0.086]	-0.276	[-0.47, -0.087]
110	-0.222	[-0.408, -0.034]	-0.222	[-0.407, -0.041]
111	-0.362	[-0.554, -0.175]	-0.362	[-0.563, -0.172]
112	-0.563	[-0.77, -0.362]	-0.563	[-0.764, -0.36]

EK-G Devam,

Okul No	*Homojen Model*		*Heterojen Model*	
	Kes.	*%95 HDI*	*Kes.*	*%95 HDI*
113	-0.384	[-0.547, -0.216]	-0.384	[-0.554, -0.217]
114	-0.466	[-0.69, -0.237]	-0.466	[-0.695, -0.245]
115	-0.621	[-0.828, -0.413]	-0.621	[-0.826, -0.419]
116	0.316	[0.106, 0.52]	0.316	[0.093, 0.545]
117	-0.144	[-0.397, 0.117]	-0.144	[-0.4, 0.118]
118	-0.425	[-0.688, -0.166]	-0.425	[-0.673, -0.182]
119	0.233	[0.081, 0.39]	0.233	[0.077, 0.38]
120	-0.209	[-0.408, -0.01]	-0.209	[-0.416, 0.002]
121	-0.235	[-0.403, -0.066]	-0.235	[-0.401, -0.071]
122	-0.004	[-0.236, 0.227]	-0.004	[-0.233, 0.213]
123	-0.261	[-0.468, -0.054]	-0.261	[-0.474, -0.056]
124	-0.646	[-0.843, -0.449]	-0.646	[-0.843, -0.455]
125	-0.297	[-0.505, -0.088]	-0.297	[-0.518, -0.087]
126	-0.202	[-0.456, 0.052]	-0.202	[-0.452, 0.051]
127	-0.312	[-0.516, -0.11]	-0.312	[-0.529, -0.089]

EK-Ğ: 2009 ÖBBS Türkçe Testi Okul İçi Varyans Kestirimleri

Okl. No	*Kes.*	*%95 HDI*	*Okl. No*	*Kes.*	*%95 HDI*	*Okl. No*	*Kes.*	*%95 HDI*
1	0.513	[0.361, 0.672]	44	0.492	[0.337, 0.658]	87	0.574	[0.369, 0.812]
2	0.595	[0.396, 0.816]	45	0.577	[0.372, 0.811]	88	0.635	[0.434, 0.857]
3	0.787	[0.514, 1.094]	46	0.548	[0.398, 0.704]	89	0.506	[0.356, 0.669]
4	0.886	[0.603, 1.225]	47	0.548	[0.39, 0.714]	90	0.718	[0.47, 1.001]
5	0.682	[0.459, 0.935]	48	0.52	[0.345, 0.704]	91	0.63	[0.405, 0.882]
6	0.703	[0.479, 0.949]	49	0.63	[0.436, 0.837]	92	0.602	[0.392, 0.838]
7	0.656	[0.44, 0.895]	50	0.431	[0.275, 0.603]	93	0.613	[0.419, 0.824]
8	0.697	[0.461, 0.962]	51	0.543	[0.335, 0.767]	94	0.494	[0.312, 0.686]
9	0.558	[0.355, 0.781]	52	0.489	[0.333, 0.663]	95	0.565	[0.363, 0.794]
10	0.613	[0.387, 0.862]	53	0.725	[0.528, 0.952]	96	0.593	[0.389, 0.819]
11	0.771	[0.512, 1.045]	54	0.667	[0.476, 0.881]	97	0.425	[0.296, 0.562]
12	0.676	[0.434, 0.939]	55	0.782	[0.519, 1.07]	98	0.616	[0.387, 0.866]
13	0.751	[0.502, 1.025]	56	0.626	[0.398, 0.875]	99	0.593	[0.38, 0.828]
14	0.664	[0.485, 0.86]	57	0.882	[0.621, 1.166]	100	0.571	[0.381, 0.777]
15	0.652	[0.471, 0.848]	58	0.56	[0.394, 0.739]	101	0.837	[0.559, 1.144]
16	0.605	[0.444, 0.779]	59	0.542	[0.375, 0.731]	102	0.725	[0.465, 1.021]
17	0.744	[0.473, 1.061]	60	0.654	[0.45, 0.87]	103	0.421	[0.288, 0.562]
18	0.629	[0.443, 0.843]	61	0.675	[0.449, 0.932]	104	0.592	[0.414, 0.789]
19	0.849	[0.578, 1.145]	62	0.555	[0.362, 0.768]	105	0.524	[0.363, 0.692]
20	0.67	[0.436, 0.917]	63	0.598	[0.416, 0.793]	106	0.492	[0.32, 0.676]
21	0.796	[0.528, 1.095]	64	0.48	[0.328, 0.643]	107	0.582	[0.406, 0.773]
22	0.646	[0.417, 0.91]	65	0.628	[0.431, 0.834]	108	0.565	[0.372, 0.789]
23	0.589	[0.376, 0.812]	66	0.455	[0.32, 0.595]	109	0.563	[0.39, 0.759]
24	0.593	[0.383, 0.826]	67	0.579	[0.374, 0.818]	110	0.553	[0.388, 0.745]
25	0.588	[0.373, 0.815]	68	0.518	[0.34, 0.716]	111	0.67	[0.468, 0.897]
26	0.612	[0.448, 0.789]	69	0.657	[0.441, 0.899]	112	0.588	[0.397, 0.793]
27	0.647	[0.431, 0.879]	70	0.607	[0.397, 0.831]	113	0.631	[0.449, 0.821]
28	0.592	[0.394, 0.818]	71	0.619	[0.426, 0.836]	114	0.575	[0.379, 0.797]
29	0.481	[0.316, 0.652]	72	0.789	[0.549, 1.06]	115	0.556	[0.379, 0.758]
30	0.425	[0.307, 0.544]	73	0.793	[0.562, 1.045]	116	0.766	[0.509, 1.056]
31	0.531	[0.365, 0.708]	74	0.737	[0.475, 1.034]	117	0.618	[0.389, 0.862]
32	0.491	[0.347, 0.643]	75	0.646	[0.412, 0.914]	118	0.486	[0.31, 0.674]
33	0.727	[0.512, 0.957]	76	0.579	[0.377, 0.819]	119	0.548	[0.396, 0.702]
34	0.525	[0.348, 0.718]	77	0.652	[0.463, 0.853]	120	0.701	[0.477, 0.95]
35	0.429	[0.293, 0.578]	78	0.61	[0.407, 0.829]	121	0.563	[0.403, 0.737]
36	0.646	[0.427, 0.898]	79	0.455	[0.296, 0.629]	122	0.537	[0.353, 0.741]
37	0.626	[0.414, 0.858]	80	0.48	[0.335, 0.629]	123	0.63	[0.422, 0.854]
38	0.427	[0.295, 0.567]	81	0.609	[0.41, 0.841]	124	0.551	[0.374, 0.74]
39	0.582	[0.399, 0.777]	82	0.491	[0.327, 0.669]	125	0.644	[0.433, 0.877]
40	0.464	[0.282, 0.654]	83	0.533	[0.356, 0.721]	126	0.564	[0.364, 0.793]
41	0.651	[0.407, 0.914]	84	0.613	[0.392, 0.85]	127	0.771	[0.518, 1.045]
42	0.56	[0.357, 0.776]	85	0.589	[0.373, 0.818]	-	-	-
43	0.409	[0.298, 0.53]	86	0.537	[0.358, 0.736]	-	-	-

EK-H: Etik Komisyonu Onay Bildirimi

Form: 40

Tez Çalışması Etik Komisyon İzin Muafiyeti Formu

10 / 02 / 2018

Hacettepe Üniversitesi
Eğitim Bilimleri Enstitüsü
Eğitim Bilimleri Anabilim Dalı Başkanlığı'na

Tez Başlığı / Konusu:	GRUP İÇİ VARYANSLAR HETEROJEN OLDUĞUNDA ÇOK DÜZEYLİ MADDE TEPKİ MODELİNİN BAYES YAKLAŞIMI İLE MODELLENMESİ

Yukarıda başlığı/konusu gösterilen tez çalışmam:

1. İnsan ve hayvan üzerinde deney niteliği taşımamaktadır,
2. Biyolojik materyal (kan, idrar vb. biyolojik sıvılar ve numuneler) kullanılmasını gerektirmemektedir.
3. Beden bütünlüğüne müdahale içermemektedir.
4. Gözlemsel ve betimsel araştırma (anket, ölçek/skala çalışmaları, dosya taramaları, veri kaynakları taraması, sistem-model geliştirme çalışmaları) niteliğinde değildir.

Hacettepe Üniversitesi Etik Kurullar ve Komisyonlarının Yönergelerini inceledim ve bunlara göre tez çalışmamın yürütülebilmesi için herhangi bir Etik Komisyondan/Kuruldan izin alınmasına gerek olmadığını; aksi durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini saygılarımla arz ederim.

Yusuf KARA
(Öğrencinin Adı Soyadı, İmzası)

Öğrenci Bilgileri

Adı Soyadı	**Yusuf KARA**
Öğrenci No	N10126330
Anabilim Dalı	Eğitim Bilimleri
Programı	Eğitimde Ölçme ve Değerlendirme
Statüsü	☐ Yüksek Lisans ☐ Doktora ☒ Bütünleşik Dr.

Danışman Görüşü ve Onayı

Araştırmada simülasyon yöntemi ile veri türetilmiş ve MEB tarafından uygulanmış olan ÖBBS verileri kullanılmıştır. ÖBBS verilerinin sadece cevap örüntüleri alınıp, öğrencilerin kişisel bilgileri kullanılmadığından etik kurul iznine gerek yoktur.

Prof. Dr. Hülya KELECİOĞLU
(İmza)
(Danışmanın Ünvanı, Adı ve Soyadı)

EK-I: Etik Beyanı

Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmasında,

- tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- görsel, işitsel ve yazılı bütün bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- atıfta bulunduğum eserlerin bütününü kaynak olarak gösterdiğimi,
- kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- bu tezin herhangi bir bölümünü bu üniversitede veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

10/02/2018

Yusuf KARA

EK-İ: Doktora Tez Çalışması Orijinallik Raporu

10/02/2018

HACETTEPE ÜNİVERSİTESİ
Eğitim Bilimleri Enstitüsü
Eğitim Bilimleri Ana Bilim Dalı Başkanlığına,

Tez Başlığı: GRUP İÇİ VARYANSLAR HETEROJEN OLDUĞUNDA ÇOK DÜZEYLİ MADDE TEPKİ MODELİNİN BAYES YAKLAŞIMI İLE MODELLENMESİ

Yukarıda başlığı verilen tez çalışmamın tamamı (kapak sayfası, özetler, ana bölümler, kaynakça) aşağıdaki filtreler kullanılarak **Turnitin** adlı intihal programı aracılığı ile kontrol edilmiştir. Kontrol sonucunda aşağıdaki veriler elde edilmiştir:

Rapor Tarihi	Sayfa Sayısı	Karakter Sayısı	Savunma Tarihi	Benzerlik Oranı	Gönderim Numarası
10/02/2018	99	170.105	22/01/2018	%1	913985800

Uygulanan filtreler:
1. Kaynaklar hariç
2. Alıntılar dâhil
3. 5 kelimeden daha az örtüşme içeren metin kısımları hariç

Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Uygulama Esasları'nı inceledim ve çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan eder, gereğini saygılarımla arz ederim.

Ad Soyadı: Yusuf KARA
Öğrenci No.: N10126330
Ana Bilim Dalı: Eğitim Bilimleri
Programı: Eğitimde Ölçme ve Değerlendirme
Statüsü: ☐ Y.Lisans ☐ Doktora ☒ Bütünleşik Dr.

10/02/2018

DANIŞMAN ONAYI

UYGUNDUR.
Prof. Dr. Hülya KELECİOĞLU

EK-J: Dissertation Originality Report

10/02/2018

HACETTEPE UNIVERSITY
Graduate School Of Educational Sciences
To The Department Of Educational Sciences

Dissertation Title: BAYESIAN MODELING OF MULTILEVEL ITEM RESPONSE MODEL WITH HETEROGENEOUS WITHIN-GROUP VARIANCE

The whole dissertation that includes the *title page, introduction, main chapters, conclusions and bibliography section* is checked by using **Turnitin** plagiarism detection software take into the consideration requested filtering options. According to the originality report obtained data are as below.

Time Submitted	Page Count	Character Count	Date of Dissertation Defense	Similarity Index	Submission ID
10/02/2018	99	170,105	22/01/2018	%1	913985800

Filtering options applied:

1. Bibliography excluded
2. Quotes included
3. Match size up to 5 words excluded

I declare that I have carefully read Hacettepe University Graduate School of Educational Sciences Guidelines for Obtaining and Using Dissertation Originality Reports; that according to the maximum similarity index values specified in the Guidelines, my dissertation does not include any form of plagiarism; that in any future detection of possible infringement of the regulations I accept all legal responsibility; and that all the information I have provided is correct to the best of my knowledge.

I respectfully submit this for approval.

Name Lastname: Yusuf KARA
Student No.: N10126330
Department: Educational Sciences
Program: Educational Measurement and Evaluation
Status: ☐ Masters ☐ Ph.D. ☒ Integrated Ph.D.

10/02/2018

ADVISOR APPROVAL

APPROVED
Prof. Dr. Hülya KELECİOĞLU

EK-K: Yayımlama ve Fikrî Mülkiyet Hakları Beyanı

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kâğıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe Üniversitesine verdiğimi bildiririm. Bu izinle Üniversite'ye verilen kullanım hakları dışındaki bütün fikrî mülkiyet haklarım bende kalacak, tezimin tamamının veya bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinleri yazılı izin alarak kullandığımı ve istenildiğinde suretlerini Üniversite'ye teslim etmeyi taahhüt ederim.

☐ **Tezimin/Raporumun tamamı dünya çapında erişime açılabilir ve bir kısmı veya tamamının fotokopisi alınabilir.**

(Bu seçenekle teziniz arama motorlarında indekslenebilecek, daha sonra tezinizin erişim statüsünün değiştirilmesini talep etseniz ve kütüphane bu talebinizi yerine getirse bile, teziniz arama motorlarının ön belleklerinde kalmaya devam edebilecektir)

☒ **Tezimin/Raporumun 01.01.2019 tarihine kadar erişime açılmasını ve fotokopi alınmasını (İç Kapak, Özet, İçindekiler ve Kaynakça hariç) istemiyorum.**

(Bu sürenin sonunda uzatma için başvuruda bulunmadığım takdirde, tezimin/raporumun tamamı her yerden erişime açılabilir, kaynak gösterilmek şartıyla bir kısmı veya tamamının fotokopisi alınabilir).

☐ **Tezimin/Raporumun .... tarihine kadar erişime açılmasını istemiyorum ancak kaynak gösterilmek şartıyla bir kısmı veya tamamının fotokopisinin alınmasını onaylıyorum.**

☐ **Serbest Seçenek/Yazarın Seçimi:**

....

....

10/02/2018

Yusuf KARA