



HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı
Eğitimde Ölçme ve Değerlendirme Programı

PARAMETRİK OLMAYAN BAYES YÖNTEMİYLE ORTAK DEĞİŞKENLERE
GÖRE YAPILAN TEST EŞİTLEMELERİNİN KARŞILAŞTIRILMASI

Meltem YURTÇU

Doktora Tezi

Ankara, 2018



Liderlik, araştırma, inovasyon, kaliteli eğitim ve değişim ile

Daha ileriye... En İyiyeye...



HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı
Eğitimde Ölçme ve Değerlendirme Programı

PARAMETRİK OLMAYAN BAYES YÖNTEMİYLE ORTAK DEĞİŞKENLERE
GÖRE YAPILAN TEST EŞİTLEMELERİNİN KARŞILAŞTIRILMASI

THE COMPARISON OF TEST EQUATING WITH COVARIATES USING
BAYESIAN NONPARAMETRIC METHOD

Meltem YURTÇU

Doktora Tezi

Ankara, 2018

Kabul ve Onay

Eđitim Bilimleri Enstitüsü M¼d¼rl¼đ¼ne,
Meltem YURTÇU'nun hazırladıđı "Parametrik Olmayan Bayes Y¼ntemiyle Ortak Deđiřkenlere G¼re Yapılan Test Eřitlemelerinin Karřılařtırılması" bařlıklı bu alıřma j¼rimiz tarafından **Eđitim Bilimleri Anabilim Dalı, Eđitimde ¼lme ve Deđerlendirme Bilim Dalı'nda Doktora Tezi** olarak kabul edilmiřtir.

J¼ri Bařkanı

Prof. Dr.řeref TAN



J¼ri Üyesi (Danıřman)

Prof. Dr. H¼lya KELECİOđLU



J¼ri Üyesi

Prof. Dr. Selahattin GELBAL



J¼ri Üyesi

Prof. Dr. Nuri DOđAN



J¼ri Üyesi

Do. Dr. Kaan Z¼lfikar DENİZ



İkinci Tez Danıřmanı

Do.Dr. Edward L. BOONE

Enstit¼ Y¼netim Kurulunun
02/02/2016 Tarihli ve 6
sayılı kararı.

Bu tez Hacettepe Üniversitesi Lisans¼st¼ Eđitim, ¼đretim ve Sınav Y¼netmeliđi'nin ilgili maddeleri uyarınca yukarıdaki j¼ri ¼yeleri tarafından 01 / 06 / 2018 tarihinde uygun g¼r¼lm¼ř ve Enstit¼ Y¼netim Kurulunca / / tarihinde kabul edilmiřtir.

Prof. Dr. Ali Ekber řAHİN
Eđitim Bilimleri Enstit¼s¼ M¼d¼r¼

Öz

Bu araştırma ortak değişkenler kullanılarak parametrik olmayan bayes modeliyle eşitlenmiş puanların elde edilmesini temel almaktadır. Araştırmada ortak değişken olarak; cinsiyet, matematik öz yeterlik puanları ve ortak madde puanları kullanılmıştır. Ortak maddelerin ve ortak değişkenlerin kullanılması ile elde edilen eşitlenmiş puanlara ait dağılımların hedef teste olan uzaklıkları hesaplanmıştır. Bu uzaklıklar, madde tepki kuramına dayalı yöntemlerden elde edilen eşitlenmiş puanların dağılımlarının hedef teste olan uzaklıkları ile karşılaştırılmıştır. Araştırma, PISA 2012 uygulamasındaki Kanada ve İtalya örneklemi üzerinde yürütülmüştür. Çalışma kapsamında; madde parametrelerinin kestiriminde PARSCALE, ölçek dönüştürme işleminde IRTEQ programları, parametrik olmayan Bayes modelinde R yazılımı kullanılmıştır. Parametrik olmayan Bayes yaklaşımında elde edilen eşitlenmiş puanlara ait dağılımların ortak değişkenlere göre değiştiği sonucuna ulaşılmıştır. Cinsiyet, matematik öz yeterlik puanları ve ortak madde puanları modelde ortak değişken olarak kullanıldığında elde edilen eşitlenmiş puanların hedef teste olan uzaklık değerleri birbirlerine yakınlık gösterse de dağılımları farklılık göstermiştir. Hedef teste en yakın dağılımın modelde cinsiyet ve matematik öz yeterlik puanlarının birlikte ortak değişken olarak kullanıldığında, en uzak dağılımların ise madde tepki kuramı yöntemlerinde elde edildiği sonucuna ulaşılmıştır. Araştırma sonucunda modelin klasik yöntemlere kıyasla daha bilgilendirici olduğu, ortak değişkenlerin ortak maddeler yerine kullanılabileceği hatta bazı durumlarda daha iyi sonuç verebileceği ve model ile elde edilen eşitlenmiş puanların hedef teste daha yakın sonuçlar verebileceği belirlenmiştir.

Anahtar sözcükler: Test eşitleme, parametrik olmayan bayes modeli, ortak değişkenler,

Abstract

This research is based on the use of covariates to obtain equated scores with Bayesian nonparametric model. As covariates in the study; gender, mathematics self-efficacy scores and common item scores were used. The distances of equated scores obtained by using common items and covariates are calculated from distances to target test. These distances were compared with distances of equated scores obtained from methods based on Item Response Theory to target test. The study was conducted on Canadian and Italian samples of PISA 2012 application. The scope of work, PARSCALE was used for estimation of material parameters, IRTEQ was used for scale transformation, and R software was used for Bayesian nonparametric model. Distributions of equated scores obtained in the Bayesian nonparametric approach changed according to the covariates. When gender, mathematics self-efficacy scores, and common item scores were used as covariates in the model, distance values of obtained equated scores to target test are close to each other but their distributions are different. The closest distribution to target test was achieved when gender and mathematics self-efficacy scores were used together as covariates in the model and the farthest distributions was obtained from item response theory methods. As a result of research, it was determined that model is more informative than the classical methods. Covariates can be used instead of common items and even better in some cases, and that equated scores obtained with model can give closer results to target test.

Keywords: Test equating, Bayesian nonparametric model, covariates.

Teşekkür

Doktora tez çalışmamın her aşamasında yanımda olan, gece gündüz, uzak yakın fark etmeksizin desteklerini esirgemeyen, heyecanımı paylaşan, değerli danışman hocam Prof. Dr. Hülya KELECİOĞLU'na;

Tezimin gelişimine görüş ve önerileri ile katkı sağlayan, anlayışlı ve güler yüzlü tavırları ile motive olmamı sağlayan değerli jüri üyelerim Prof. Dr. Selahattin GELBAL'a, Prof. Dr. Şeref TAN'a, Prof. Dr. Nuri DOĞAN'a ve Doç. Dr. Kaan Zülfikar DENİZ'e;

Tezimin her aşamasında karşılaştığım sorunlara çözüm yolları sunarak yardımcı olan, bana benden çok inanarak cesaret ve ilham veren değerli eş danışmanım Doç. Dr. Edward L. BOONE'e;

Her koşulda yanımda olan, benimle birlikte heyecanlanıp benimle birlikte üzülen, kaygılarımı, stresimi ve sevincimi paylaşan sevgilerini tariflere sığdıramayacağım canım annem ve babama; kolum kanadım olan canım abim ve kardeşime;

Beni hiçbir zaman yalnız bırakmayan, sevincimi üzüntümü paylaşarak her konuda desteği ile yanımda olan, hayatımda ki en büyük şanslarımdan biri can dostum, Semiha GÖKTEKİN'e ve eşi Harun GÖKTEKİN'e; bana koşulsuz kefil olan abim ve ablam dediğim ve her zaman öyle kalacak olan canım arkadaşlarım Çağdan - Levent ZORLUOĞLU'na; motive ve enerji kaynağım olarak hep güç veren ve ne zaman ihtiyaç duysam yanımda olan canım arkadaşım Merve Lütfiye ŞENTÜRK'e; Hacettepe Üniversitesi'nde görev yaptığım süre boyunca destek ve katkılarıyla yanımda olan birlikte olmaktan hep mutluluk duyduğum başta sevgili arkadaşlarım Mine ZORLU, Sümeyra SOYSAL, Eda DEMİRDÜZEN-DEMİREL, Esin YILMAZ-KOĞAR ve Haydar KARAMAN olmak üzere Eğitim Bilimleri Bölümü'nde görev yapan tüm arkadaşlarıma;

Süreç içerisinde karşılaştığımız problemlerde yapıcı ve ılımlı tavırları ile yanımda olan güler yüzünü esirgemeyen tüm Eğitim Bilimleri Enstitüsü personellerine;

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

çok teşekkür ederim..

İçindekiler

Öz.....	i
Abstract	ii
Teşekkür.....	iii
Tablolar Dizini.....	vi
Şekiller Dizini.....	vii
Bölüm 1 Giriş.....	1
Problem Durumu	1
Araştırmanın Amacı ve Önemi	3
Araştırma Problemi	4
Sayıtlılar	6
Sınırlılıklar	6
Bölüm 2 Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar.....	7
Bağlama ve Eşitleme	7
Eşitleme Desenleri	9
Eşitleme Yöntemleri	17
Bayes Yaklaşımı	25
Parametrik olmayan Bayes Yaklaşımı.....	29
Parametrik olmayan Bayes Yaklaşımı ile Eşitleme	29
İlgili Araştırmalar.....	38
Bölüm 3 Yöntem.....	44
Araştırmanın Türü	44
Araştırmanın Evreni ve Örneklemi	44
Veri Toplama Araçları	45
Verilerin Analizi	46
Bölüm 4 Bulgular ve Yorumlar.....	55
Alt Problem 1'e İlişkin Bulgular ve Yorumlar	55

Alt Problem 2'ye İlişkin Bulgular ve Yorumlar	57
Alt Problem 3'e İlişkin Bulgular ve Yorumlar	63
Alt Problem 4'e İlişkin Bulgular ve Yorumlar	71
Alt Problem 5'e İlişkin Bulgular ve Yorumlar	86
Bölüm 5 Sonuç, Tartışma ve Öneriler	92
Sonuçlar ve Tartışma	92
Öneriler	96
Kaynaklar	98
EK-A: Yöntemlere Göre Kestirilen Puanlar	110
EK-B: Her Yöntemden Elde Edilen Eşitlenmiş Puanların Standart Sapması Ve Ortalaması.....	111
EK-C: Ülkelere Ait Ayrı Kalibrasyon Kestirimleri için Kullanılan PARSCALE kodu	112
EK-Ç: MCMC Kodu (MATHEFF için).....	113
EK-D: Eşitlenmiş Puanların Elde Edilmesi (Cinsiyet için)	115
EK-E: Modelde İki Ortak Değişken Kullanıldığında Elde Edilen Parametreler....	116
EK-F: Etik Komisyonu Onay Bildirimi	118
EK-G: Etik Beyanı	119
EK-Ğ: Doktora Tez Çalışması Orijinallik Raporu	120
EK-H: Dissertation Originality Report	121
EK-I: Yayımlama ve Fikrî Mülkiyet Hakları Beyanı	122

Tablolar Dizini

Tablo 1 <i>Tek Grup Deseni</i>	11
Tablo 2 <i>Dengelenmiş Grup Deseni</i>	11
Tablo 3 <i>Eşit Gruplar Deseni</i>	12
Tablo 4 <i>Denk Olmayan Gruplar Ortak Test Deseni (DoGOM)</i>	12
Tablo 5 <i>Kitapçıklara ait bilgiler</i>	44
Tablo 6 <i>Kitapçık Türlerine Göre Soruların Dağılımı</i>	45
Tablo 7 <i>5.Kitapçığa Ait Açıklanan Varyans</i>	47
Tablo 8 <i>6.Kitapçığa Ait Açıklanan Varyans</i>	48
Tablo 9 <i>Alt ve Üst Gruplar için maddeler arası korelasyon</i>	49
Tablo 10 <i>Model Veri Uyumunun İncelenmesi</i>	50
Tablo 11 <i>MTK eşitleme yöntemlerine göre elde edilen RMSE değerleri</i>	56
Tablo 12 <i>Matematik Öz Yeterliği Düşük Öğrencilerin Matematik Puanlarının 5. ve 6. Kitapçığındaki Frekans Dağılımları</i>	73

Şekiller Dizini

Şekil 1. Eşitleme desenleri.	10
Şekil 2. Eşitleme yaklaşımları.	23
Şekil 3. 5. kitapçığı alan İtaya örnekleminin puan dağılımı.	55
Şekil 4. 6. kitapçığı alan Kanada örnekleminin puan dağılımı.	55
Şekil 5. MTK yöntemlerinden elde edilen puanların dağılımı ve hedef teste olan uzaklıkları.	56
Şekil 6. Kız öğrencilerin kitapçıklara göre puan dağılımı ve güven aralığı.	58
Şekil 7. Kız öğrencilerin 5. kitapçık ve 6. kitapçık puanlarının PoB modeline göre eşitlenmesi.	58
Şekil 8. Erkek öğrencilerin kitapçıklarda göstermiş oldukları puan dağılımı.	59
Şekil 9. Erkek öğrencilerin kitapçıklara göre puan dağılımı ve güven aralıkları. ...	60
Şekil 10. Erkek öğrencilerin 5. Kitapçık ve 6. kitapçık puanlarının PoB yöntemine göre eşitlenmesi.	61
Şekil 11. Cinsiyet ortak değişkenine göre PoB yönteminden elde edilen puanların ve hedef testin puanlarının dağılımı.	62
Şekil 12. Matematik öz yeterlik düzeylerine göre öğrencilerin kitapçıklardaki puan dağılımları.	63
Şekil 13. Yüksek matematik öz yeterlik düzeyine sahip öğrencilerin kitapçıklarda göstermiş oldukları puan dağılımı ve güven aralıkları.	64
Şekil 14. Orta Düzey matematik öz yeterlik düzeyine sahip öğrencilerin puan dağılımı ve güven aralıkları.	65
Şekil 15. Düşük düzey matematik öz yeterlik düzeyine sahip öğrencilerin kitapçıklarda göstermiş oldukları puan dağılımı ve güven aralıkları.	65
Şekil 16. Farklı matematik öz yeterlik düzeylerinde PoB modeli ile yapılan eşitleme.	66
Şekil 17. 5. kitapçıkta MATHEFF düzeylerine göre kestirilen puanların dağılımı..	67
Şeki 18. 5. kitapçıkta MATHEFF düzeylerine göre puanlarına kümülatif dağılımı.	68
Şeki 19. 5. kitapçıktaki kümülatif dağılımın perspektif görünüşü.	68
Şekil 20. 6. kitapçıkta MATHEFF düzeylerine göre kestirilen puanların dağılımı..	69
Şekil 21. 6 kitapçıkta MATHEFF düzeylerine göre puanlarına kümülatif dağılımı.	69
Şekil 22. MATHEFF ortak değişkenine göre PoB yönteminden elde edilen puanların ve hedef testin puanlarının dağılımı.	70

Şekil 23. Matematik öz yeterlik düzeylerine göre 5. Kitapçıkta kız öğrencin puan dağılımlar ve güven aralıkları.	72
Şekil 24. Matematik öz yeterlik düzeylerine göre 5. Kitapçıkta erkek öğrencin puan dağılımları ve güven aralıkları.	72
Şekil 25. Matematik öz yeterlik düzeylerine göre 6. Kitapçıkta kız öğrencilerin puan dağılımları ve güven aralıkları.	74
Şekil 26. Matematik öz yeterlik düzeylerine göre 6. Kitapçıkta erkek öğrencilerin puan dağılımları ve güven aralıkları.	75
Şekil 27. Matematik öz-yeterliği düzeylerine göre kız öğrencilerin kitapçıklardaki puan dağılımı.....	76
Şekil 28. Matematik öz-yeterliği düzeylerine göre erkek öğrencilerin kitapçıklardaki puan dağılımı.....	77
Şekil 29. Matematik öz yeterlik düzeylerine göre öğrencilerin 5. ve 6. kitapçık puanlarının PoB modeli ile eşitlenmesi.....	78
Şekil 30. Erkek öğrencilerin 5. Kitapçıkta MATHEFF düzeylerine göre kestirilen puanlarına ait dağılımlar.....	80
Şekil 31. Erkek öğrencilerin 5. kitapçıkta MATHEFF düzeylerine göre kestirilen puanlarına ait kümülatif dağılımlar.....	80
Şekil 32. Erkek öğrencilerin 5. kitapçıkta kümülatif dağılımlarını perspektift görünüşü	80
Şekil 33. Erkek öğrencilerin 6. kitapçıkta MATHEFF düzeylerine göre kestirilen puanlarına ait dağılımlar.....	81
Şekil 34. Erkek öğrencilerin 6. kitapçıkta MATHEFF düzeylerine göre kestirilen puanlarına ait kümülatif dağılımlar.....	81
Şekil 35. Erkek öğrencilerin 6. kitapçıkta kümülatif dağılımlarını perspektift görünüşü.	81
Şekil 36. Kız öğrencilerin 5. kitapçıkta MATHEFF düzeylerine göre kestirilen puanlarına ait dağılımlar.....	83
Şekil 37. Kız öğrencilerin 5. kitapçıkta MATHEFF düzeylerine göre kestirilen puanlarına ait kümülatif dağılımlar.....	83
Şekil 38. Kız öğrencilerin 5. kitapçıkta kümülatif dağılımlarını perspektift görünüşü.	83
Şekil 39. Kız öğrencilerin 6. kitapçıkta MATHEFF düzeylerine göre kestirilen puanlarına ait dağılımlar.....	84

Şekil 40. Kız öğrencilerin 6. kitapçıkta MATHEFF düzeylerine göre kestirilen puanlarına ait kümülatif dağılımlar.....	84
Şeki 41. Kız öğrencilerin 6. kitapçıktaki kümülatif dağılımlarını perspektift görünüşü.	84
Şekil 42. İki ortak değişkenine göre PoB yönteminden elde edilmiş olan puanların ve hedef testteki puanların dağılımları.....	86
Şekil 43. 5. Kitapçıkta ortak maddelerden elde edilen puanlara göre öğrencilerin puan dağılımları.....	87
Şekil 44. 6. Kitapçıkta ortak maddelerden elde edilen puanlara göre öğrencilerin puan dağılımları.....	88
Şekil 45. Ortak madde puanlarının ortak değişken olarak ele alınması ile PoB yönteminden elde edilen puanların ve hedef testteki puanların dağılımları.	90

Simgeler ve Kısaltmalar Dizini

BBPS = Bağımlı Bernstein Polinom Süreci (**DBPP** = Depent Berntein Polynominal Process)

BDÖ = Bernstein-Dirichlet Önseli (**BDP** = Bernstein-Dirichlet prior)

BDS = Bağımlı Dirichlet Süreci (**DDP** = Dependent Dirichlet Process)

DoGOD = Denk Olmayan Gruplarda Ortak Değişken (**NEC** = Non-equivalent Groups with Covariates)

DoGOM = Denk Olmayan Gruplarda Ortak Madde (**NEAT** = Non-equivalent groups with ancor test)

DS = Dirichlet Süreci (**DP** = Dirichlet Process)

KDS = Karma DS (**DPM** = DP Mixture)

KTK= Klasik Test Kuramı

MCMC = Markov Chain Monte Carlo

MTK= Madde Tepki Kuramı

PoB = Parametrik Olmayan Bayes (**BNP**/ Bayesian Nonparametric)

RMSE = Root Mean Square Error

Bölüm 1

Giriş

Eğitim sisteminde politikaların belirlenmesi veya yöneticilerin karar vermesinde bireylere uygulanan sınavlardan alınan puanlar önemli bir yer tutmaktadır. Testler bireylerin ve ülkelerin ne tür eksiklikleri olduğunu veya ne kadar yeterliliğe sahip olduğunu görmesinde yardımcı olmaktadır. Ekonomik İşbirliği ve Kalkınma Örgütü'nün (Organisation for Economic Co-operation and Development- OECD) projesi olan Uluslararası Öğrenci Değerlendirme Programı (The Programme for International Student Assessment – PISA); ülkelerin kendi eğitim sistemlerini değerlendirmesini, ülkeler arasındaki karşılaştırmalar ile eğitimdeki gelişimsel faaliyetleri ve politik kaynaklar arasındaki ilişkiyi ortaya çıkarmayı amaçlayan geniş ölçekli sınavlardan bir tanesidir (OECD, 2014). PISA uygulamalarında öğrencilerin kendi yeterliklerini, ailelerin ve okul ortamlarının avantajlı ve/veya dezavantajlı durumlarını ve öğrencilerin öğrenme deneyimlerini ortaya koymak için çeşitli anketler kullanılmaktadır. Uygulama, bireylerin eğitim ile elde ettikleri bilgileri günlük yaşantılarında kullanabilmelerini de değerlendirmeye almaktadır. Bu uygulamada öğrencilerinin matematik ve fen okuryazarlığı ile okuma becerilerine ilişkin performanslarını ölçmeye yönelik bilişsel testler kullanılmaktadır. Bu testlerde güvenliği sağlamak için eşdeğer olacak şekilde farklı formları kullanılmaktadır. Ancak gerçekte bir testin farklı formlarını bire bir elde etmek oldukça zordur. Test formlarında yer alan soruların eşit güçlükte olmaması, birbirlerinden içerik veya psikometrik özellikler bakımından farklılaşması testlerin paralelliğinin bozulmasına neden olmaktadır. Bu farklılık ise test sonuçlarına göre değerlendirilecek olan bireyler arasında haksız bir kıyaslamaya sebebiyet verecektir. Bu olumsuzluğu en aza indirmek ve puanlar arasında karşılaştırılabilirliği sağlamak amacıyla eşitleme yapılması gerekmektedir.

Problem Durumu

Test puanlarına göre değerlendirilecek olan bireylerin puanları arasında karşılaştırma yapabilmek oldukça önemlidir. Aynı amaca yönelik farklı test formlarından elde edilen puanların karşılaştırılabilir duruma getirilmesi için eşitleme yapılması gerekmektedir. Eşitleme yapılabilmesi için ise en önemli adım en az hatayı verecek olan desenin ve yöntemin seçilmesidir.

Eşitleme desenleri; bireylerin test formlarını alma durumuna ve bireylerin yetenekleri arasındaki farkı kontrol altına alma durumlarına göre farklılık göstermektedir. Bu desenler genel olarak; tek grup deseni, eşit gruplar deseni ve denk olmayan gruplarda ortak madde deseni olarak adlandırılmaktadır. Testlerin farklı formlarının aynı aynı yetenek seviyesine sahip bireyler tarafından alınmış olması her zaman mümkün olmamaktadır. Bu durumlar için ise iki testi de temsil eden ve incelenen gruplar arasındaki farkı ortaya çıkaracak olan ortak maddeler kullanılmaktadır. Ancak ortak testler, bu tür özellikleri her zaman sağlayamayabilir. Bu durum, testlerin eşitlenmesi için güvenilirliği ve ortak testlere bağlı diğer süreçleri etkilemektedir (Wiberg & von Davier, 2017; Wei, 2010). Ayrıca birçok standart testte olduğu gibi büyük uygulamaları gerektiren testlerde ortak madde veya ortak test bulunmamaktadır. Bu durumda da puanların kestirim sürecine yardımcı ek bilgilerin eklenmesi veya ortak testlerin yerine bu bilgileri verecek olan ortak değişkenlerin kullanılması, kestirimin doğruluğunu arttırabileceği gibi eşitleme çalışmalarının birçok yönden incelenmesine katkı sağlayacaktır (Livingston & Lewis, 2009; Branberg & Wiberg, 2011). Ortak değişkenlerin kullanılmasını temel alan “denk olmayan gruplarda ortak değişken” deseni de son çalışmalarla birlikte alanyazında yerini almaktadır (Wiberg & Branberg, 2015).

Test puanlarının eşitlenmesinde kullanılan eşitleme yöntemleri farklı kuram ve varsayımlara dayalıdır. Literatürde eşitleme yöntemleri kurama dayalı olarak, Klasik Test Kuramı ve Madde Tepki Kuramı şeklinde sınıflandırılmaktadır. Ancak son yıllarda Bayes yaklaşımı da test eşitleme çalışmalarında öne çıkmaktadır. Özellikle Parametrik olmayan Bayes yaklaşımı ortak değişkenlerin modele eklenmesini olası hale getirmektedir.

Standart yöntemlerden Kernel ve Local Eşitleme yöntemleri, standart olmayan yöntemlerden ise Parametrik olmayan Bayes eşitleme yöntemleri alanyazın için oldukça yeni kavramlardır. Ülkemizde kernel eşitleme yönteminin uygulamaları son yıllarda çalışılmış olup (Akın-Arıkan, 2017), local eşitleme ve Parametrik olmayan Bayes modeli kullanılarak yapılan eşitleme ile ilgili herhangi bir çalışmaya rastlanılmamıştır. Bu eşitleme yönteminden elde edilen eşitlenmiş puanların; Madde Tepki Kuramına dayalı eşitleme yöntemlerinden elde edilen eşitlenmiş puanlardan göstereceği farklılıklar merak konusudur.

Ülkemiz alanyazında aynı içeriği ölçmeye yönelik olarak hazırlanmış ve yıl içerisinde iki uygulaması olup ortak madde içermeyen standart sınavlardan elde edilen puanların eşitlenmesinde genellikle tek grup deseni kullanılmıştır. Tek grup deseni yıl içerisinde düzenlenen ve eşitlenecek olan testlerin iki formunu da almış olan bireyleri ihtiva etmektedir. Ancak her iki formun da alınmasının zorunluluğu olmadığı durumda, örnekleme ulaşma sıkıntı oluşturabilmektedir. Bu çalışma ile alanyazında yeni bir desen olan “Denk olmayan Gruplar için Ortak değişken (Non-equivalent Groups with Covariates / NEC) Deseni” kullanılarak ortak değişkenlerin eşitleme sürecine eklenmesi ile eşitleme yapılabilmesi mümkün olacaktır.

Alanyazında eksikliği görülen bu durumların cevapları çalışmaya gerek duyulan bir konu olmakla birlikte bu cevaplara ulaşmak ülkemizde test eşitleme çalışmalarına ve alanyazına oldukça büyük katkılar sağlayacaktır.

Araştırmanın Amacı ve Önemi

Ülkemizde Ölçme, Seçme ve Yerleştirme Merkezi (ÖSYM) tarafından birçok büyük ölçekli sınav uygulanmaktadır. Kamu Personeli Seçme Sınavı (KPSS), Yabancı Dil Sınavı (YDS), Üniversiteler Arası Dil Sınavı (ÜDS), Akademik Personel ve Lisansüstü Eğitim Giriş Sınavı (ALES), ÖSYM tarafından hazırlanan sınavlar arasındadır. Bu tür sınavların güvenilirliklerinin sağlanması için yıl içerisinde farklı zamanlarda farklı formları ile tekrarları yapılmaktadır. Ancak bu sınavların birçoğu ortak bireylere uygulanmadığı gibi ortak madde de içermemektedir. Ortak olmayan gruplarda ortak testin / maddenin olmadığı durumlarda ortak değişkenlerin gruplar arasındaki yanlılığı ortadan kaldırarak eşitleme yapılabileceğinin gösterilmesi bakımından önem arz etmektedir. Bu durumda eşitleme yapılmasını olanaklı hale getirecek, ülkemizdeki ölçme ve değerlendirme alanına büyük bir avantaj sağlayacaktır. Aynı zamanda Bayes yönteminin diğer istatistiksel kestirim yöntemlerine kıyasla daha kesin kestirimler yapması da yeni bir yaklaşımın eğitimde kullanılması açısından önemlidir.

Alanyazında da yeni olan bu konu üzerinde yapılandırılmış olan bu çalışmada iki amaç yer almaktadır. İlk amaç doğrultusunda eşitleme desenlerinde yeni yaklaşım olan ve ortak olmayan gruplarda ortak test / madde yerine ortak değişkenler kullanılarak oluşturulan “Denk olmayan Gruplar için Ortak değişken (DoGOD / Non-equivalent Groups with Covariates / NEC) Deseni” ülkemiz

alanyazınında ilk bu çalışma ile ele alınmaktadır. Ortak madde / ortak test olmadığına veya ortak maddelerin / testlerin uygun özellikleri sağlamadığında bu desene yer verilecektir. İkinci amaç doğrultusunda standart olmayan yöntemlerden Parametrik olmayan Bayes modeli kullanılarak eşitlenmiş puanlar elde edilmektedir. Bayes yönteminin sınırsız sayıda parametre kestirime sahip olması durumunda ortaya çıkan Parametrik olmayan Bayes (PoB) yaklaşımının eşitlemede kullanılması ile elde edilen puan dağılımları; kitapçıklar ve gruplar için detaylı bilgiler vermektedir. Dolayısı ile alt gruplar dikkate alınarak kestirimler elde edilmektedir. Bu yöntemlerin eşitlemede kullanılması ile elde edilen kestirimlerin kesinliğinin incelenmesi amaçlanmaktadır. Ayrıca eşitleme yöntemleri için sağlanması gereken bazı varsayımların karşılanma zorunluluğu olmaması araştırmanın önemini arttırmaktadır.

Araştırma Problemi

PISA 2012 yılı matematik alt testi puanlarının denk olmayan gruplarda ortak madde deseninde MTK eşitleme yöntemleri ile ve denk olmayan gruplarda ortak değişken deseninde PoB modeli ile eşitlenmesi sonucunda elde edilen eşitlenmiş puanların dağılımları nasıldır?

Alt problemler

1. PISA 2012 yılı matematik alt testi puanlarının denk olmayan gruplarda ortak madde deseninde MTK eşitleme yöntemleri ile eşitlenmesi sonucunda elde edilen eşitlenmiş puanların dağılımları ile hedef teste ait dağılım arasında nasıl bir ilişki bulunmaktadır?
2. PISA 2012 yılı matematik alt testi puanlarını PoB modeli ile eşitlemek için öğrencilerin cinsiyeti ortak değişken olarak alındığında;
 - 2.1. Cinsiyet değişkeninin kategorilerine göre elde edilen eşitlenmiş puanlar arasında nasıl bir ilişki bulunmaktadır?
 - 2.2. Elde edilen eşitlenmiş puanların dağılımı ile hedef teste ait dağılım arasında nasıl bir ilişki bulunmaktadır?
3. PISA 2012 yılı matematik alt testi puanlarını PoB modeli ile eşitlemek için ortak değişken olarak alınan öğrencilerin matematik özyeterlik puanları;

- 3.1. Düşük, orta ve yüksek düzeyi temsil edecek şekilde üç kategorili olarak ele alındığında kategorilere göre elde edilen eşitlenmiş puanların dağılımları nasıldır?
- 3.2. Sürekli değişken olarak ele alındığında eşitlenmiş puanların dağılımının hedef teste ait dağılıma olan uzaklığı nasıldır?
4. PISA 2012 yılı matematik alt testi puanlarını PoB modeli ile eşitlemek için öğrencilerin cinsiyetleri ve matematik özyeterlik puanları birlikte ortak değişken olarak alındığında;
 - 4.1. Üç kategorili (düşük, orta, yüksek) olacak şekilde dönüştürülen MATHEFF değişkeninin kategorilerine göre elde edilen eşitlenmiş puanların dağılımı nasıldır?
 - 4.2. Sürekli olarak ele alındığında eşitlenmiş puanların dağılımının hedef teste ait dağılıma olan uzaklığı nasıldır?
5. PISA 2012 yılı matematik alt testi puanları PoB modeli ile eşitlemek için ortak maddelerin puanları ortak değişken olarak alındığında elde edilen eşitlenmiş puanların dağılımının hedef teste ait dağılıma olan uzaklığı nasıldır?

Sayıtlılar

Bu arařtırmada;

Arařtırmada kullanılan DoGOD deseninin özelliğinden dolayı; test puanlarının ortak deęiřkenlere ait kategorilerdeki daęılımı iki grupta da aynı olduđu varsayılmaktadır.

Matematik öz yeterlięi ortak deęiřkeninde yer alan maddelerin; cinsiyet, kùltür, sosyo-ekonomik durum, vb. faktörlerden etkilenmeksizin, her öğrenci için aynı anlařıldığı varsayılmıřtır.

Sınırlılıklar

1. Arařtırma PISA 2012 yılı sonuçlarına göre yetenek düzeyleri birbirinden farklı olan Kanada ve İtalya'dan elde edilen veriler ile sınırlıdır.
2. Arařtırma PISA 2012 yılında uygulanan 5. ve 6. kitapçıklardaki matematik sorularına, matematik öz yeterlięi anketine ve cinsiyete verilen cevaplarla sınırlıdır.

Bölüm 2

Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar

Bağlama ve Eşitleme

Bireyler arasında doğru bir değerlendirme yapmak için, bir testin farklı formlarından elde edilen puanların karşılaştırılabilir durumda olması gerekmektedir (Tian, 2011). Ancak bireylerin istatistiksel özelliklerine ve örnekleme bağlı olarak elde ettikleri puanlar doğrudan karşılaştırılmazlar (Weeks, 2010). Bir testin farklı formlarından elde edilen puanların karşılıklı olarak birbiri yerine kullanılabilmesi ortak bir ölçeğe yerleştirilmeleri durumunda mümkündür. Bu testleri karşılaştırılabilir yapmak ya da farklı test formlarından dönüşmüş puanların prosedürlerini tanımlamak için “equating (eşitleme)” ve “linking (bağlama)” terimleri kullanılmaktadır (Tian, 2011). Bağlama, test puanlarının karşılaştırılmasını tanımlamak için kullanılan temel bir terimdir (Kim, 2015). Holland ve Dorans (2006) da “bağlama” terimini bir testten alınan puanlar ile diğer testten alınan puanlar arasında dönüşüm kümesi olarak ifade ederek bağlamanın temel 3 kategorisini; kestirim / tahmin(predicting), ölçek bağlama / ölçekleme(scale aligning) ve eşitleme (equating) olarak ele almaktadır.

- Kestirim (predicting), bireyin diğer bir testten aldığı puanı / puanları veya demografik özellikleri kullanarak bireyin yeni test puanını ifade etmek anlamına gelen bir doğrusal regresyon yöntemidir (Holland, 2007).
- Ölçekleme, farklı test formlarından elde edilen puanların ortak bir ölçeğe dönüştürülmesi sürecidir (Dorans & Walker, 2007).
- Eşitleme, iki test formundan elde edilen puanlar arasında bir ilişki oluşturarak her bir test formundaki puanların aynı testten gelmiş gibi kullanılmasını amaçlamaktadır (Dorans, Moses & Eignor, 2010). Bağlamanın en güçlü türü olup standart testler için en önemli süreç olarak değerlendirilmektedir (Wiberg, 2013).

Eşitleme bağlamanın özel bir türüdür (Dorans, Moses & Eignor, 2010). Eşitleme, aynı yapıyı ölçmek için tasarlanmış test formlardan elde edilen puanların birbiri yerine kullanılmasını sağlayan bir dönüşüm işlemi olarak tanımlanmaktadır.

(Kolen & Brennan, 2014, s.2; von Davier, Holland & Thayer 2003; Kolen 1988; Liu, Feigenbaum & Dorans, 2005; Wiberg, 2016; Wei, 2010; Dorans, Moses & Eignor, 2010).

Bireyler veya gruplar arasında karşılaştırılma süreci, testin güçlük düzeyine ve uygulandığı grubun yetenek dağılımına göre iki kısımda incelenmektedir. Uygulanan test formlarının güçlük düzeyleri ve bu testleri alan grubun yetenek dağılımları benzerlik gösteriyorsa yatay eşitleme; farklılık gösteriyor ise dikey ölçekleme olarak adlandırılmaktadır (Hambleton & Swaminathan, 1985; Kolen, 1988). Yatay eşitlemeye genellikle bir testin güvenlik sebebi ile farklı test formlarının kullanıldığı durumlarda başvurulmaktadır (Hambleton & Swaminathan, 1985). Dikey ölçekleme ise, daha çok standart olmayan öğretmen yapımı ve boylamsal olarak gelişimi izlemede kullanılan testlerden alınan puanları karşılaştırmak için kullanılmaktadır. Bu araştırmada kullanılan testlerin eşitlenmesinde yatay eşitlemeye başvurulmuştur.

Eşitleme çalışmasını yürütebilmek için testlerin bir takım koşulları sağlaması gerekmektedir. Lord (1980) bu koşulları; aynı yapıyı ölçme, eşitlik, simetriklik ve grup değişmezliği olarak ele alırken Dorans ve Holland (2000) buna eşit güvenilirlik özelliğini de eklemiş ve en çok öne çıkan beş eşitleme koşulunu; aynı yapıyı ölçme, güvenilirliklerin aynı olması, simetriklik, eşitlik ve gruptan bağımsızlık olarak belirlemiştir.

- Eşitlenecek iki testin mutlaka aynı yapıyı, özelliği veya yeteneği ölçmesi eşitleme için en temel koşuldur.
- Eşitlemede hatanın düşük olması testlerin güvenilirliklerinin eşit olması ile sağlanabilmektedir.
- Simetriklik eşitlenecek olan testlerin formları arasında referans ve hedef test ayrımı yapılmaksızın puanların birbirleri yerine kullanılabilmesini sağlamaktadır.
- Eşitlik, testi alan bireyler için eski ve yeni formlar arasında dönüştürülmüş puanların dağılımının aynı olmasıdır.

- Eşitlemenin gruptan bağımsızlığı, eşitleme ilişkisinin eşitlemenin yürütüldüğü alt gruplar için aynı olmasıdır. (Kolen & Brennan, 2014, s.12).

Bu özelliklerin bazıları test eşitleme işlemleri için kontrol edilebilmesine rağmen, uygulamada eşitlik, gruptan bağımsızlık ve eşit güvenirlik gibi varsayımların sağlamak oldukça zordur (Andersson, 2014).

Eşitleme Adımları

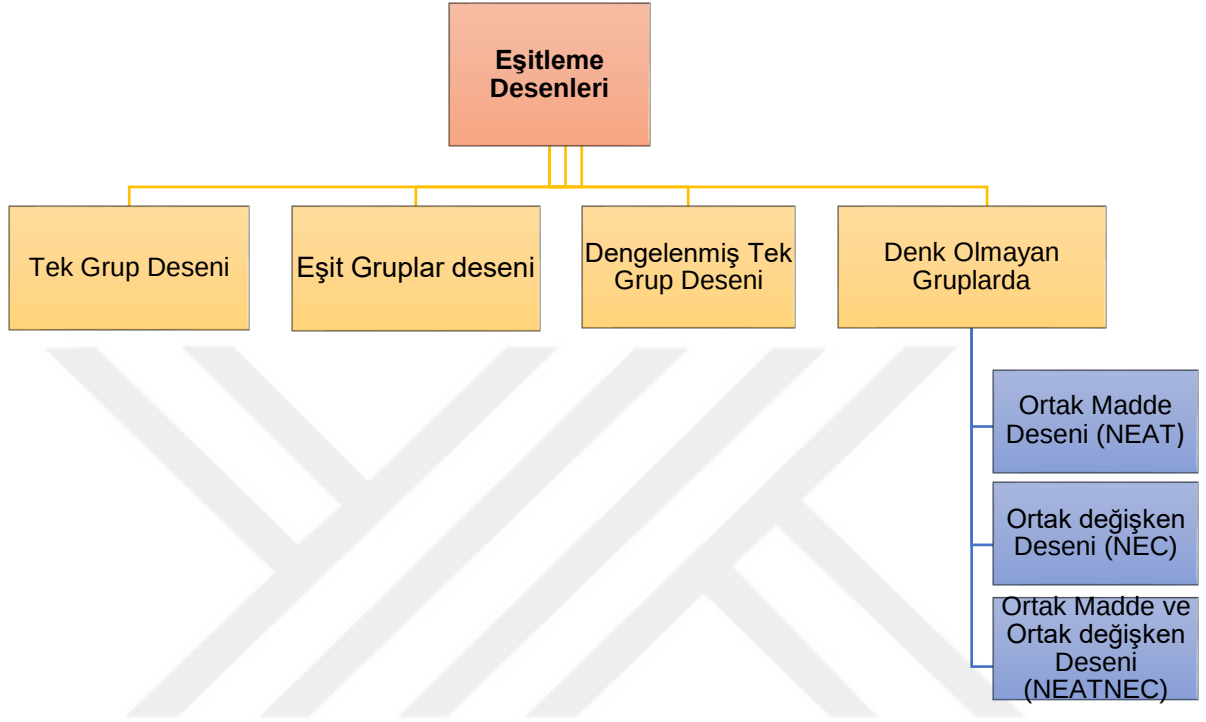
Kolen ve Brennan'a (2014, s.7) göre, test formlarının eşitlenebileceği kararı verildikten sonra izlenecek adımlar şöyle sıralanabilir:

- 1) Eşitlemenin amacına karar vermek,
- 2) Alternatif formlar oluşturmak,
- 3) Veri toplama desenini seçmek
- 4) Seçilen veri toplama desenine uygun olarak verileri toplamak,
- 5) Eşitlemenin işlevsel tanımlarını seçmek,
- 6) Eşitleme ilişkilerini kestirecek istatistiksel yöntemleri seçmek,
- 7) Eşitleme sonuçlarını değerlendirmek.

Eşitleme Desenleri

Bir eşitleme çalışmasında veri toplama desenini seçerken uygulanabilirlik ve istatistiki özellikler göz önünde bulundurulmalıdır. Veri toplama deseni olarak ortak maddelerin veya ortak bireylerin kullanıldığı farklı yaklaşımlar vardır. Ortak bireylerin kullanıldığı desenler; tek grup deseni (single group design), dengelenmiş gruplar (counterbalanced design) deseni ve eşit grup deseni (equivalent group design) olarak sınıflandırılabilir. Ortak maddelerin kullanıldığı desen de denk olmayan gruplarda ortak madde / test (DoGOM / Non-equivalent groups with ancor test / NEAT) deseni olarak ele alınmaktadır (Kolen & Brennan, 2014, p.18; Branberg & Wiberg, 2011). Son yıllardaki çalışmalarda ortak maddelerin olmadığı durumda ortak değişkenlerin kullanılması ile “Denk Olmayan Gruplarda Ortak değişken (DoGOD / Non-equivalent Groups with Covariates / NEC)” (Branberg & Wiberg, 2011; Wiberg & Branberg, 2015) ve hem ortak madde hem de ortak değişkenlerin kullanılması ile NEATNEC deseni literatüre eklenmiştir (Wiberg & Branberg, 2015).

En son güncellenen hali ile eşitleme desenlerini Şekil.1 deki gibi göstermek mümkündür.



Şekil 1. Eşitleme desenleri.

Tek Grup Deseni

Tek grup deseninde, iki veya daha fazla test formu aynı gruba uygulanmaktadır. Aynı cevaplayıcıların her iki testi de alması nedeniyle testlerin zorluk düzeyleri cevaplayıcıların yetenek düzeylerinden etkilenmez. Tek bir gruba testler uygulandığı için ölçme hatası diğer desenlere göre daha küçüktür (Livingston, 2004). Ancak aynı gruba iki uygulama yapmanın etkisi eşitleme sürecini etkileyebilir. Bir başka deyişle, testlerin veriliş sırası bir hata kaynağı oluşturabilir (Kolen & Brennan, 2014, s.14).

Tablo 1

Tek Grup Deseni

Evren	Örneklem	X	Y
P	1	✓	✓

Dengelenmiş Tek Grup Deseni

Tek grup deseniindeki bireylerin yarısına eşitlenecek testlerden X formu önce verilirken, diğer yarısına da Y formunun önce verilmesi ile elde edilen desendir. Böylelikle testlerin veriliş sırasından kaynaklanacak hatalar (sıra etkisi) giderilmiş olur. Bu desenin dezavantajı ise yorgunluk nedeniyle ikinci formdan elde edilecek performansın düşük olmasıdır.

Tablo 2

Dengelenmiş Grup Deseni

Evren	Örneklem	X1	Y1	X2	Y2
P	1	✓			✓
P	2		✓	✓	

Eşit Gruplar Deseni (Random Grup Deseni)

Random grup desenine eşdeğer / eşit gruplar deseni (equivalent-groups design) de denilmektedir. Bu desende, X ve Y formlarındaki bütün puan düzeylerini temsil edecek yeterlikte büyük bir heterojen grup seçilir. Grup, random olarak ikiye ayrılır. Bu desende grupların iki formdaki performans düzeyleri arasındaki fark doğrudan formların güçlük düzeyleri arasındaki farkı göstermektedir (Kolen & Brennan, 2014,s.13-14). Ancak direkt eşit gruplar deseni için, grupların yeteneklerinin eşit olarak kabul edilmesi varsayımı yanlış olacaktır. Bu varsayım incelenmeden eşitleme sürecine devam edilmesi puanların kullanımlarının geçerliğini düşürecek ve test programını riske atacaktır (Lyren & Hambleton, 2011).

Tablo 3

Eşit Gruplar Deseni

Evren	Örneklem	X	Y
P	1	✓	
P	2		✓

Denk Olmayan Gruplarda Ortak Madde/Test (DoGOM) Deseni

Denk olmayan gruplarda ortak test deseni, uygulamada birden fazla formun uygulanma şansının olmadığı ve gruplar arasındaki farkı ortaya çıkarmak için gerekli verilerin farklı formlarda yer alan ortak madde / testlerden elde edildiği desendir (Liou, Chen & Li, 2001; Moses, Deng & Zhang, 2010). Desen için ortak testin seçimi oldukça önemli olup bu testin bazı özelliklere sahip olması gerekmektedir. Ortak test; eşitlenecek olan testler ile benzer ortalama, madde zorluğuna sahip olmalı ve bu testleri içerik olarak temsil etmelidir (Wei, 2010; Mittelhaeuser, Beguin & Sijtsma, 2011; Wiberg & von Davier, 2017; Kolen, 1988; Sinharay & Holland, 2006; Dorans, Moses & Eignor, 2010). Ortak testten elde edilen toplam puan eğer bireylerin testlerden elde ettikleri puanına katkı sağlıyorsa iç ortak test; katkı sağlamıyorsa dış ortak test olarak adlandırılmaktadır (Kolen, 1988; Kolen & Brennan, 2014, p.18). Mittelhaeuser, Beguin ve Sijtsma (2011) ise çalışmasında iç ve dış ortak teste ek olarak ön-test desenini eklemiştir. Ön-test deseni; maddelerin istatistiksel özelliklerini incelemek için bireylerin farklı örnekleme uygulanması sürecini içermektedir. Desende kullanılacak olan ön-test, eşitlenecek olan testlerden seçilerek oluşturulabilir. Ön test maddeleri iki ardışık yıl için ele alınmış ise, dış ortak test desenine benzer olarak testler arasında bağlantı kurulmaktadır.

Tablo 4

Denk Olmayan Gruplar Ortak Test Deseni (DoGOM)

Evren	Örneklem	X	A	Y
P	1	✓	✓	
Q	2		✓	✓

DoGOM deseninde birden fazla ortak test kullanılması ile gruplardaki farklılıklarının test puanları üzerindeki etkisini daha doğru olarak incelemek mümkündür (Moses, Deng & Zhang, 2010). Moses, Deng ve Zhang (2010), birden fazla ortak test puanı ile bireylerin test puanlarını birlikte ele almasında kolaylık sağlayan ve benzer varsayımlara sahip üç yöntemi ele almıştır. Bu yöntemler;

Son tabakalama Eşitleme (Post-stratification equating-PSE): Hem X testi hem de Y testinin; P ve Q evreninin bir karışımı olan yapay bir evrene uygulanan A test formu ile ilişkisine bakılarak elde edilir (Andersson, Branberg & Wiberg, 2013). Bu eşitleme yöntemi DoGOM deseni için eşit yüzdelikli eşitleme yöntemine karşılık gelmektedir ve aynı zamanda frekans kestirimi olarak da adlandırılmaktadır (Bramley & Vidal Roderio, 2014).

Yerine Atama (Imputation); bu yaklaşımda yine yapay evren kullanılarak X testinde eksik veriler yerine Q popülasyonunun, Y testindeki eksik veriler yerine P popülasyonunun bireyler atanarak X ve Y dağılımlarının kestirilmesi sonucu elde edilmektedir. Atama varsayımı, Q da kayıp X verileri ve P de kayıp Y verilerinin random olması durumuna bağlıdır (Moses, Deng & Zhang, 2010).

Eğilim puan eşleşmesi (Propensity score matching); bireylerin puanları ilgili karakteristik özelliklerine bağlı olarak değişeceğinden dolayı direkt olarak karşılaştırılamayacağı temeline dayanmaktadır. Gruplar arasındaki sistematik farklılığı kontrol etmek için gözlenen özelliklere bağlı olarak alt gruplar birim olarak kabul edilir ve bu alt gruplar arasında karşılaştırma yapılır (Rosenbaum & Rubin's, 1984). Bu puanlamada genellikle lojistik regresyon ve ayırma analizi uygulamalarına yer verildiği için sürekli, kesikli ve karma değişkenlerin kullanımına müsaade etmektedir. Eğilim puanları geçmiş değişkenlerin tek bir boyuta toplanmasını temel almakta ve denge puanı olarak nitelendirilmektedir (Longford, 2015).

Bu yöntemler klasik yöntemlerde oldukça tercih edilmekte olup, ortak madde / test puanlarının birden fazla olması durumunda bireylerin puanları ile bağlanmasında esneklik sağlamaktadır. Yapısal olarak benzer bir çerçeve göstermesi ile bu yöntemlerin DoGOD deseninin ortaya çıkmasında ve yorumlanmasında kolaylık sağlayacaktır.

Denk Olmayan Gruplarda Ortak deęiřken Deseni (DoGOD)

Denk olmayan gruplar deseninde gruplar arasında sistematik farklılıęı kontrol etmek ortak testlerin kullanımı ile mümkündür. Ancak standart testlerin birçoęunda ortak test yer almayabilir veya bu ortak testler için dikkat edilmesi gereken bazı durumlar ihlal edilmiř olabilir. Bu durumlar; ortak testlerin tek boyutlu olmaması, dięer testlerdeki puanlarla yüksek oranda iliřki vermemesi, tam olarak test formlarındaki yapıyı ölçmede yetersiz kalması (Willian & Wiberg, 2017, p.316) veya uygulamasından kaynaklı hataların oluřması (Liou, Chen & Li, 2001) olarak ele alınabilir. Ayrıca sadece zaman ierisindeki eğilimleri ele alan ortak testlerin DoGOM deseninde kullanılması, bu testleri alanların sadece sınırlı sayıdaki bireyleri için uygun olabilir ki bu durumda eşitleme için bir yanlılık oluřturabilir. Bu da testlerin güvenirliklerini olumsuz yönde etkileyecektir (Wiberg & Branberg, 2015; Wiberg & von Davier, 2017; Wei, 2010). Gruplar arasındaki yetenek farklılıęını açıklayan ortak deęiřkenlerin dikkatli seilerek sürece dahil edilmesi ile bu olumsuz etki en aza indirilebilir.

İki test arasında eşitlemede ortak test yerine ortak deęiřkenler kullanıldığında, desen “Denk Olmayan Gruplarda Ortak deęiřken (Non-equivalent Groups with Covariates) Deseni” olarak adlandırılıp, NEC kısaltması ile gösterilmektedir (Branberg & Wiberg, 2011; Wiberg & Branberg, 2015). Bu desen alıřmada DoGOD kısaltması ile ele alınmaktadır. Bu desende, X testini alan P örneklemini ile Y testini alan Q örneklemindeki bireylerin test puanları iliřkili olan deęiřkenler ile birbirine baęlanır. Böylece iki grup arasında yetenek farkı kontrol etmek için bir ortak test yerine ortak deęiřkenlerden elde edilen bilgiler kullanılarak iřlem yapılır (Wiberg & Branberg, 2015).

Klasik yöntemlerden DoGOM de kullanılan son tabakalama eşitleme (NEAT-PSE); testlerden alınan puan daęılımlarını grupların aęırlıklandırılması ile oluřturulan yapay bir evrende tanımlamaktadır (Andersson, Branberg & Wiberg, 2013; Andersson & Wiberg, 2017). DoGOD deseni ise ortak testin yerini ortak deęiřkenler alması ve ortak deęiřkenlerin kategorilerine ait daęılımlar için bir yapay evren oluřturulmasından dolayı DoGOM son tabakalama eşitleme yöntemi ile yapı olarak yakınlık göstermekte ve benzer sonuçlar vermektedir. Ayrıca DoGOD deseninde kullanılan ortak deęiřkenlerin testlerle iliřkisi arttıęında, DoGOM

deseninde zincirleme ve son tabakalama eşitleme yöntemlerinden elde edilen sonuçlarla benzerlik daha da artacaktır (Wiberg & Branberg, 2015).

DoGOD deseninin en önemli varsayımı, ortak değişkenlerin gruplar arasında ölçülen değişken bakımından farklılığını açıklayabildiğidir. Test puanlarının durumsal dağılımlarının, ortak değişkenlerin kategorilerine göre her iki grupta da aynı olması bu desen için en önemli adımdır. Bu durum da bireylerin grup başarısından ziyade, sahip oldukları kategorik özelliklere göre değerlendirildiğinin bir göstergesidir. Ancak eşitlenecek olan test puanları farklı zaman aralıklarında elde edilmişse (yani çok eski bir test ile yeni bir test, LES-ALES gibi); test puanları ve ortak değişkenler arasındaki ilişki zamanla değişmiş ise bireylere ait özellikler zamanla değişim göstereceğiden bu varsayım geçerli olmayabilir (Wiberg & Branberg, 2015).

Branberg (2010, p.20) çalışmasında DoGOM deseninde geçmiş değişkenlerin kullanılmasının kestiricilerin ortalama standart hatalarının azaltılmasında oldukça önemli bir rol oynadığını ifade etmektedir. Ayrıca geçmiş değişkenleri kullanmanın DoGOM deseninde elde edilen kestirimlerin kesinliğini arttırabileceğini ifade ederek hem ortak madde hem de ortak değişkenlerin kullanılması ile oluşan desenin temellerini ortaya atmıştır. Bu desen, Wiberg ve Branberg (2015)'in çalışmasında NEATNEC deseni olarak adlandırılmaktadır.

Ortak değişken Kullanarak Eşitleme

Eşitleme sürecinde toplanan verilerin yanında bireylere ait değişkenler ortak değişken olarak işlemlere dahil edilebilir. Ortak değişkenler gruplar arasındaki farkı ortaya çıkarmak için kullanılan ek bilgilerdir. Ortak test / madde uygulamasının bulunmadığı veya kullanılmasının uygun olmadığı sınavlarda, test puanları ile ilişkili ortak değişkenler bireyler için uygun olarak kabul görüyorsa; bu değişkenler de eşitleme işlemine dahil edilebilir. (Branberg & Wiberg, 2011; Wiberg & Branberg, 2015). Doğru ortak değişkenlerin kullanımı ile kestirimlerin eşitleme hatası azalabilmekte ve eşitlemenin kesinliği / doğruluğu artabilmektedir (Branberg & Wiberg, 2011; Wiberg & Branberg, 2015; Kim, Livingston & Lewis, 2009; 2011; Oh, Guo, & Walker, 2009).

Ortak deęiřkenlerin kullanılmasında en önemli nokta bu deęiřkenlerin seçimidir. Gözlenen veride doğru ortak deęiřkenlerin seçilmesi ile yanlılık ve ortalama standart hatanın azaltılması amaçlandıęından, bu deęiřkenlerin gruplar arasındaki farkları açıklayabilmesi ve test puanları ile ilişkili olması gerekmektedir (Oh, Guo & Walker, 2009; Liou, Cheng & Li, 2001; Branberg & Wiberg, 2011; Wiberg & Branberg, 2015; Wiberg, 2015). Eęer test puanları ile yüksek korelasyon veren tek bir ortak deęiřken kullanılırsa, bir ortak testin kullanımı ile benzer sonuçlar elde edilmiř olur (Wiberg & Branberg, 2015); hatta bireylerin geęmiř bilgilerinden seçilen deęiřkenler bazen ortak testlerden daha iyi sonuç verebilir (Liou, Chen & Li, 2001). Gonzalez, Barrientos ve Quintana (2015b) alıřmasında; eřitleme modellerinin ortak deęiřkenler içermesinin bir avantajının da; ortak deęiřkenler kullanılmasıyla řekillenen alt gruplara ait daęılımlara bakarak bu alt gruplar arasında karřılařtırma yapılabileceęini ifade etmiřtir. Örneęin; cinsiyet ve sosyo-ekonomik düzey ortak deęiřken olarak alındıęında, düşük sosyo-ekonomik düzeye sahip olan X formunu almıř olan bir kız öęrencinin puanı ile sosyo-ekonomik düzeyi düşük olan Y formunu almıř olan bir erkek öęrencinin puanları arasında bir baęlantı kurulabilmektedir. Ancak gruplara göre yanlı olan herhangi bir deęiřkenin kullanılması katılımcılar arasında yapılan deęerlendirmeyi de yanlı olarak etkileyecek ve adil bir deęerlendirme yapılmasını engeleyecektir.

Denk olmayan gruplarda ortak test yerine ortak deęiřkenlerin kullanılmasının ortaya ıkıřı Kolen (1990)'in "ortak testin iyi iřlemedięi zaman eski ve yeni formu alan grupların eřleneceęi deęiřkenler vardır" ifadesine dayanmaktadır (Wiberg & Branberg, 2015). Dorans ve Holland (2000) alıřmasında grupların deęiřmezlik varsayımının saęlanmadıęı durumlarda eřitleme fonksiyonlarının modellenmesinde ortak deęiřkenlerin kullanımının yararlı olabileceęini ifade etmiřtir. Birok arařtırmacı ortak deęiřkenleri farklı terimlerle ifade etmiř olsa da, kullanımlarının yanlılıęı azaltacaęı ve alt gruplara ait deęiřkenlerden yararlanarak farklı alt gruplar arasında iliřki kurulabileceęi řeklinde ortak bir vurgu yapmıřtır (Livingston, Dorans & Wright, 1990; Wright & Dorans,1993; Liou, 1998; Liou, Cheng & Li, 2001; Kim, Livingston & Lewis,2009). Okula devam, cinsiyet gibi geęmiř deęiřkenler kadar test puanları ile ilişkili ortak test puanlarını da ortak deęiřken olarak kullanan Liou, Cheng ve Li (2001) bu ortak deęiřkenleri, "vekil (surrogate)" deęiřkenler olarak adlandırırken; Kim, Livingston ve Lewis (2009;2011) ile Oh, Guo, ve Walker (2009)

de “tanımlayıcı (collateral)” değişkenler olarak adlandırmaktadır. Birçok araştırmacı tarafından ise bu değişken “ortak değişken (covariate)” olarak nitelendirmiştir (Branberg, 2010; Gonzalez, Branberg & Quintana, 2015a; 2015b; Wiberg & Branberg, 2015; Gonzalez & Wiberg, 2017). Bu araştırmada “ortak değişken” terimi kullanılmıştır.

Alanyazında ortak değişken olarak genellikle yaş, cinsiyet, eğitim durumu gibi değişkenlerin yer aldığı görülmektedir (Wiberg & von Davier, 2017; Liou, Cheng & Li, 2001; Branberg & Wiberg, 2011; Gonzalez, Barrientos & Quintana, 2015a; Karabatsos & Walker, 2008; Wiberg & Branberg, 2015).

Ortak testin kullanımından daha iyi sonuç vermesi için ortak değişkenlerin sayısı artırılabilir. Çalışmaya eklenen ortak değişkenlerin sayısı önemli olduğu kadar ortak değişkenlerin kategorilerinin sayısı da önemlidir. Ortak değişken sayısı arttıkça, bu değişkenlerin kategorilerine düşen birey sayısı azalacağından dolayı bu kategori sayılarının sınırlandırılması daha uygun sonuçlar verecektir (Wiberg & Branberg, 2015; Wallin & Wiberg, 2017, p.309). Çalışmada ortak değişkenlerin kombinasyonu ile ortaya çıkan kategoriler bireylerin sahip oldukları profilleri temsil etmektedir.

Eşitleme Yöntemleri

Test eşitleme yöntemleri Klasik Test Kuramına dayalı yöntemler ve Madde Tepki Kuramına dayalı yöntemler olmak üzere iki ana başlık altında sınıflandırılmaktadır (Hambleton & Swaminathan, 1985; Kolen, 1988). Bu sınıflandırmayı kapsayan eşitleme yöntemleri standart modeller olarak değerlendirilirken, standart yöntemlerin geliştirilmesi ile ortaya çıkan daha ileri ve teorik eşitleme yöntemleri standart olmayan yöntemler olarak karşımıza çıkmaktadır. Standart yöntemlere; ortalama, doğrusal, eşit yüzdelikli, kernel, Tucker, MTK'ya göre eşitleme yöntemleri; standart olmayan yöntemlere de local ve Parametrik olmayan Bayes eşitleme yöntemleri örnek olarak verilebilir (Gonzalez, 2014)

Klasik Test Kuramına Dayalı Yöntemler

Klasik Test Kuramı (KTK), zayıf teorik varsayımlara sahip ve kolay uygulanabilir olması nedeniyle büyük avantaja sahiptir. Bu teori örnekleme bağı olduğu gibi madde parametre kestirimleri bireylere ve bireyin parametre kestirimleri teste bağlıdır. Bu durum ise model için eşitleme ve bağlamanın uygulanabilirliğinde bazı kısıtlamalar getirmekte; grup değişmezliği, simetriklik ve eşitlik gibi varsayımların tam olarak sağlanmasına engel oluşturmaktadır (Hambleton, Swaminathan & Rogers, 1991). Gözlenen puanların eşitlenmesinde kullanılan bazı yöntemleri; ortalama, doğrusal, eşit yüzdelikli, kernel ve local eşitleme yöntemi olarak sıralamak mümkündür (Gonzalez, 2014).

Ortalama Eşitleme Yöntemi

Ortalama eşitlemede, eşitlenecek olan testlerin güçlükleri arasında fark olduğu fakat bu farkın puan ölçeği boyunca sabit olduğu kabul edilmektedir. Bu yöntem, eşitlenecek olan testlerdeki puanların ortalamaya olan uzaklıklarının eşit kabul edilmesi esasına dayanmaktadır (Kolen, 1988; Kolen & Brennan, 2014,s.30). Bu eşitleme yöntemine dayalı olarak kullanılan olan eşitleme fonksiyonu;

$$x - \mu_x = y - \mu_y \quad \Rightarrow \quad y = \varphi(x; \mu_x, \mu_y) = x - \mu_x + \mu_y$$

şeklindedir.

Doğrusal Eşitleme Yöntemi

Doğrusal eşitleme, paralel testlerin ortalama ve standart sapma haricindeki tüm özelliklerinin eşit olması varsayımına dayanmaktadır (Woldbeck, 1998, s.11; Crocker & Algina, 1986, s.458). Standart puanlara dayalı bir yöntem olan doğrusal eşitlemede, aynı standart puanlara karşılık gelen test puanların eşit olduğu kabul edilir (Angoff, 1971, s.564). Doğrusal eşitleme yönteminde sonuçlar gruba bağımlı olduğu için oldukça büyük ve heterojen bir grubun seçilmesi gerekmektedir (Kolen & Brennan, 2014, s.31). Bu eşitleme yöntemine dayalı olarak kullanılan eşitleme fonksiyonu;

$$\frac{x - \mu_x}{\sigma_x} = \frac{y - \mu_y}{\sigma_y} \quad \Rightarrow y = \varphi(x; \mu_x, \mu_y, \sigma_x, \sigma_y) = \left(\frac{\sigma_y}{\sigma_x} \right) (x - \mu_x) + \mu_y$$

şeklindedir.

Ortalama ve doğrusal dönüşümler parametrelere bağlı olduğundan, parametrik yöntemler olarak nitelendirilmektedir (Gonzalez, 2014).

Eşit Yüzdelikli Eşitleme Yöntemi

Bir puan dağılımında aynı yüzdelik sıraya denk gelen puanları belirlemeye dayanan eşitleme yöntemidir. Eşit yüzdelikli eşitleme yöntemi sadece ilk iki nokta arasındaki farkın yanında tüm puan dağılımı arasındaki farkı da hesaba kattığı için lineer ve ortalama eşitlemenin genel halidir.

Eşitlenecek olan X ve Y testlerinden elde edilen puanlar X ve Y olarak ele alındığında; F(x) ve G(y) yığılmalı fonksiyonları göstermek üzere, eşitleyecek olan fonksiyon;

$$\varphi(x) = e_y(x) = G^{-1}(F(x))$$

şeklinde gösterilir.

Eşit yüzdelikli eşitleme iki aşamalı bir süreçten oluşmaktadır. İlk aşamada eşitlenecek iki formun puan dağılımının yığılmalı frekans dağılımları tablolaştırılırken ikinci aşamada ise elde edilen yığılmalı frekanslarda aynı yüzdelik sıraya denk gelen puanlar eşitlenir (Kolen, 1988). Ancak pratikte iki formdan elde edilen ham puanlar nadiren aynı yüzdelik sıraya gelmektedir. Bu sorunu ortadan kaldırarak dağılımı sürekli hale getirmek için öteleme (Interpolation) ve düzgünleştirme (Smoothing) kullanılmaktadır (Cook & Petersen, 1987).

Kernel Eşitleme

Bu yöntem yapısı itibari ile kesikli yığılmalı puan dağılımını sürekli hale getirmeyi amaçlamaktadır. Eşitlenecek olan iki testin yığılmalı dağılımları sürekli hale getirildikten sonra bu testlerde yer alan iki puan arasındaki uzaklığı temel alarak eşitleme yapılmaktadır. (von Davier, Holland & Thayer, 2003; Silverman, 1986). Kernel düzgünleştirme ile süreklileştirmenin kontrolünü sağlamak için genellikle Gaussian dağılımını, log-lineer ve uniform kernel yöntemleri kullanılmaktadır (Andersson & Wiberg, 2017).

F ve G eşitlenecek olan testlerin yığılmalı dağılımlarını göstermek üzere, orijinalde kesikli olan bu fonksiyonlar sürekli hale getirilerek, bu iki fonksiyon arasında dönüşüm tanımlanmaktadır. Eşitleme fonksiyonu

$$\varphi(x) = G_{hy}^{-1}(F_{hx}(x; r; s)) = G_{hy}^{-1}(F_{hx}(x))$$

olarak tanımlanır. Burada s ve r parametreleri kestirilen puanların olasılıklarının vektörleri olarak tanımlanmakta ve veri toplama deseninin seçilmesini hesaplamak için adlandırılan desen fonksiyonları (DF) kullanılarak elde edilmektedir. $\varphi(x)$ kestirimlerinin doğruluğuna yine SEE ile bakılmaktadır. φ nin kestirimleri hem s , r vektörlerini hem de F ve G fonksiyon parametrelerini içerdiğinden dolayı yarı parametrik olarak değerlendirilmektedir (Gonzalez, 2014).

Kernel yönteminin diğer yöntemlere göre 2 önemli avantajı vardır;

Daha küçük standart hataya sahip olup örneklemin büyüklüğünden daha az etkilenmektedir. Bu yüzden küçük örneklem içeren uygulamalarda da kullanılabilir.

Yaygın kullanılan veri toplama desenlerinin hepsi ile birlikte benzer bir yol kullanarak standart hataların kestirimini ve eşitleme fonksiyonlarının geliştirilmesini sağlar (Bramley & Vidal Roderio, 2014).

Local Eşitleme Yöntemi

Local Eşitleme, grup değişmezliği ve eşitlik kriterlerinin sağlanmadığı durumlarda kullanılmak üzere ortaya atılmış bir eşitleme yöntemidir (van der Linden, 2010). Eşit yüzdelikli eşitlemedeki gibi puan dağılımlarının kullanılmasının aksine puanların yeteneklerdeki koşullu dağılımdan yararlanır. φ yi elde etmek için yetenek bağımlı eşitleme dönüşümlerinin bir ailesi kullanılmaktadır. Dönüşüm formu aşağıdaki gibi ele alınmaktadır:

$$\varphi(x; \theta) = G_{Y/\theta}^{-1}(F_{X/\theta}(x)) \quad \theta \in R$$

Yeteneğin kestirimleri için Maksimum Sonsal Kestirim (Maximum A Posterior / MAP), Beklenen Sonsal Kestirimi (Expected A Posteriori / EAP) veya En Çok Olabilirlik (Maksimum Likelihood / MLE) kestirim yöntemleri kullanılabilir ve MTK modellerinde bireylerin tepkilerinden yararlanılabilir. Koşullu dağılımın süreklileştirilmesinde öteleme (interpolation) kullanılmaktadır. Bu eşitleme yöntemi de hem vektörleri hem de fonksiyonları içerdiğinden dolayı yarı-parametrik bir yöntemdir (Gonzalez, 2014).

Eşitleme için tek bir dönüşüm kullanma varsayımı bulunmakla birlikte bu durum zorunlu değildir. Bireylerin sınavlardan belirlenen yetenek düzeylerine göre gruplandırılması önemli rol oynamaktadır. Ancak sınava giren her bir birey için θ yi

temsil edecek bir değerin bulunmaması bu yöntemin olumsuz tarafını oluşturmaktadır (Xin & Zhang, 2015).

Madde Tepki Kuramına Dayalı Yöntemler

Madde Tepki Kuramı (MTK), farklı yetenek düzeyindeki bireylerin ölçülen bir özellik için bir maddeye verebilecekleri cevabın matematiksel bir model ile açıklamasını sunmaktadır. Madde parametreleri testin uygulandığı gruptan bağımsız olarak elde edilmektedir. Bireylerin θ ile gösterilen yeteneği de maddelere verdikleri cevaplardan bağımsızdır. Bu yüzden bir bireyin aynı θ yeteneğini ölçen iki test formundan herhangi birini alması, testlerin güçlüğü farklı olsa bile yetenek kestirimini değiştirmeyecektir (Hambleton & Swaminathan, 1985). Madde tepki Kuramına dayalı modeller, hem θ hem de madde karakteristik vektörüne (a,b,c,..) bağlı bir test maddesine verilen doğru cevabın olasılığı ile belirlenmektedir (Gonzalez, 2014).

MTK modelleri ölçtükleri örtük özelliğe göre tek boyutlu ve çok boyutlu olarak ele alınmaktadır. Tek boyutlu MTK modelleri iki kategorili ve çok kategorili modeller; çok boyutlu modeller de telafi edici ve telafi edici olmayan modeller olarak iki kısımda incelenmektedir. Tek boyutlu MTK modellerinden iki kategorili modeller; 1 parametrelili lojistik modeller (1 PLM), 2 parametrelili lojistik modeller (2 PLM) ve 3 parametrelili lojistik modeller (3 PLM) dir (Embretson, 1996, 1997; Embretson & Reise, 2000; Hambleton & Swaminathan, 1985).

MTK modellerinden uygun olanı seçildikten sonra madde parametreleri kestirilir. Testleri bağlamanın temel amacı 2 veya daha fazla testten elde edilen madde parametreleri ve / veya yetenek kestirimlerini ortak bir ölçeğe yerleştirmektir (Weeks, 2010). İki test formunda puanları birbirleri yerine kullanılabilir ve karşılaştırılabilir için aynı ölçeğe yerleştirme süreci; madde parametrelerinin ayrı olarak veya eş zamanlı olarak kestirilmesine göre iki durumda incelenmektedir (Kim & Cohen, 1998; Petersen, Cook, & Stocking, 1983; Hanson & Béguin, 2002). Eş zamanlı kalibrasyon farklı formlarda yer alan bütün maddelerin birlikte tek bir dosyada kestirilmesi sürecini içermektedir (Tian, 2011, p.8; Kim & Cohen,1998). Eş zamanlı kalibrasyon ile ortalama ve standart sapmalar grupları eşitlemek için özelleştirilmiş olur. Böylece parametre kestirimleri aynı ölçekte yer alır (Brossman,

2010; Kolen & Brennan, 2014, s.182). Ayrı kalibrasyonda farklı formlar ayrı ayrı kestirildikten sonra parametreler bir bağlama süreci ile aynı ölçek üzerinde bir araya getirilmektedir. Bu araştırmada farklı formları alan grupların ortalama ve standart sapmaları farklı olduğu için parametreler ayrı ayrı kestirilmektedir.

Ayrı kalibrasyon ile madde ve yetenek parametrelerini ortak metriğe yerleştirmek için ölçeklemede en sık kullanılan yöntemler moment yöntemleri için ortalama-ortalama, ortalama-sapma; test karakteristik eğrisi yöntemleri için Heabera ve Stocking-Lord metodlarıdır. (Haebara, 1980; Stocking & Lord, 1983). Çalışmada kestirilen parametreleri aynı metriğe yerleştirmek için MTK modeline dayalı olarak 3 Parametrelili Lojistik modeller altında bu dört ölçek dönüştürme yöntemi kullanılmıştır.

Eşitleme Hatası

Random ve *sistematiik* olmak üzere iki çeşit eşitleme hatası vardır. Eşitleme hatasının olabildiğince küçük olması seçilen eşitleme yönteminin uygunluğunun bir göstergesidir (Kolen, 1988; Kolen & Brennan, 2014; Wang, 2006). *Random* eşitleme hatası, eşitleme ilişkisini belirlemede yer alan ortalama, standart sapma ve yüzdelik sıra gibi parametrelerin kestirildiği örnekleme bağılı olarak ortaya çıkan hatadır. Eşitlemede *random* hatanın bulunması, örneklemedeki verilerin, evrenden farklılık gösterdiği anlamına gelmektedir. Örneklem büyüklüğünün artırılması ile bu sorun bir miktar da olsa azaltılabilmektedir (Kolen & Brennan, 2014, s.29; Liu, Schulz and Yu, 2007; Wang, 2006).

Sistematiik hatanın oluşma nedeni eşitleme ilişkisinin incelendiği yöntemin yanlışlık içermesi ya da eşitlemenin varsayımlarının ihlal edilmesidir. Sistematiik hata, eşitlemenin yürütüldüğü gruptaki bireylerin eşitlenen formu alan bireylerden önemli ölçüde farklılık gösterdiği durumlarda da oluşabilmektedir. Bu yüzden, eşitleme çalışmasının yürütüldüğü örneklemin büyüklüğünün artması *random* hatayı azaltırken sistematiik hatayı azaltmamaktadır (Kolen & Brennan, 2014,s.29).

Ortak maddelerin gruptaki farklılıkları yansıtmadığı veya testleri temsil etmediği durumlarda, bu koşul ihlal edilmiş olup, eşitleme sonuçlarının geçerliğini etkilemekle birlikte sistematiik hataya yol açmaktadır (Wei, 2010).

Eşitleme yöntemlerinden elde edilen eşitlenmiş puanların hata miktarlarını belirlemek için genellikle ağırlıklandırılmış hata kareleri ortalamaları (WMSE),

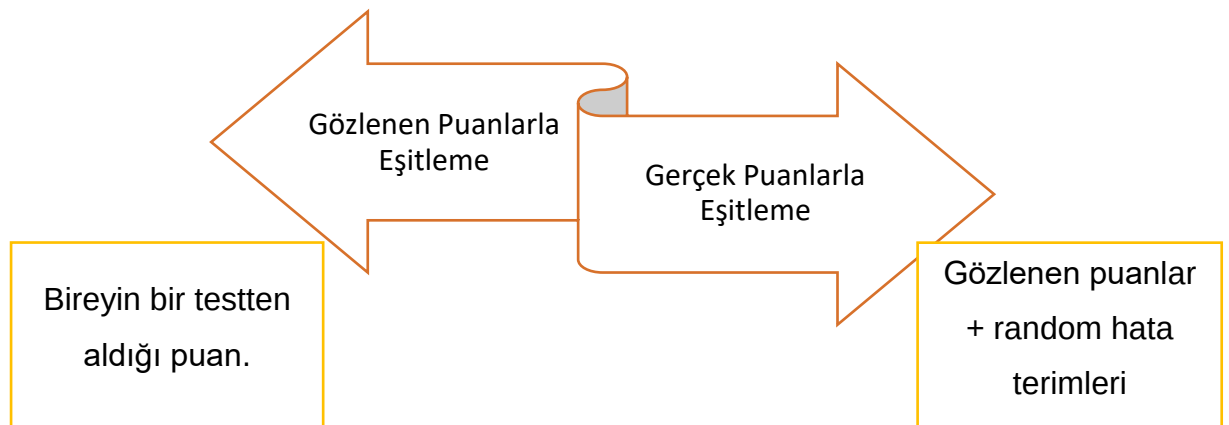
standart hata (Standart Error - SE) veya Eşitleme hatası (Root mean square error- RMSE) kullanılan katsayılarıdır.

Gözlenen ve Gerçek Puan

Farklı psikometrik özelliklere sahip olan çeşitli puan eşitleme yöntemleri bulunmaktadır. Bunlar puanların dağılımına bağlı olarak gözlenen puanlar ve gerçek puanlara odaklanan yöntemlerdir (Brossman, 2010; p.1).

Gözlenen puanlarda eşitlemenin amacı, paralel formlardaki puanların dağılım özelliklerini olabildiğince benzer olarak düzenlemek ve bu puanlar arasında bir eşitlik ilişkisi kurmaktır (Brossman, 2010; p.1). Dolayısı ile temel testteki gözlenen puanlardan yola çıkarak bu puanlara hedef testte karşılık gelen puanı bulmaktır (Andersson, 2014).

Gerçek puanı eşitleme sürecinin amacı ise, bir formdaki gerçek puanı, gözlenen puanlara eşlemek yerine, başka bir formdaki (klasik test teorisinde ya da madde tepki kuramı teorisinde gerçek puanın tanımı yoluyla) gerçek puanlara eşlemektir (Brossman, 2010; p.2). Gerçek puanlarla eşitleme yapıldığı zaman eşitleme dönüşümünden elde edilen puanın beklenen değerinin hedef (Y) testindeki puanlara eşit olması gerekmektedir. Bu durumda temel (X) testindeki gerçek puanlara uygulanacak bir dönüşüm bulmak amaçlanmaktadır (Andersson, 2014).



Şekil 2. Eşitleme yaklaşımları.

Lord ve Wingersky (1984) çalışmasında bu iki puan dönüşümünün madde tepki kuramında benzer sonuçlar verdiğini göstermiştir.

MTK'ya dayalı kestirilen yeteneklere karşılık gelen gerçek puanlar;

$$\xi_x = \sum_{i=1}^k P_i(\theta) \text{ ve } \xi_y = \sum_{j=1}^k P_j(\theta)$$

eşitlikleri ile hesaplanmaktadır.

ξ_x = Bireylerin X testindeki gerçek puanını,

ξ_y = Bireylerin Y testindeki gerçek puanını,

$P_i(\theta)$ = θ yeteneğindeki bireyin i maddesini doğru cevaplama olasılığını,

$P_j(\theta)$ = θ yeteneğindeki bireyin j maddesini doğru cevaplama olasılığını temsil etmektedir. Araştırmada MTK ya dayalı yöntemlerden gerçek puan eşitlemesi için model olarak 3 PLM kullanılmıştır. 3 parametrelili modele göre θ düzeyinde her bir bireyin i maddesini yanıtlama olasılığını veren formül;

$$P_i(\theta) = c_i + (1 - c_i) \frac{e^{a_i(\theta - b_i)}}{1 + e^{a_i(\theta - b_i)}}$$

şeklindedir. Bu formülde;

a_i : i maddesinin ayırtıcılık gücü indeksi

b_i : i maddesinin güçlük indeksi

c_i : i maddesini şans parametresi

Formüle göre bireylerin cevaplama olasılıklarından yola çıkılarak gerçek puanlar elde edilmektedir.

Grup Değişmezliği;

Ölçme aracı ile ölçülmek istenen bir yapının, alt gruplar arasında matematiksel olarak birbirine denk olmasıdır (Little, 1997; Huh & Lee, 2009). Farklı alt gruplarda bağlama (linking) fonksiyonunun bir uzantısı olarak tanımlanan grup değişmezliği, bir testin farklı formlarındaki bu bağlama ilişkisini karakterize etmek için kullanılmaktadır (Huh & Lee, 2009; Liu, Feigenbaum & Dorans, 2005). Eşitleme fonksiyonunu hesaplamak için kullanılan alt grupların seçimi önemli olmadığı zaman

eşitleme fonksiyonunun grup değişmezliği özelliğine sahip olduğu ifade edilir (Moses, 2006).

Grup değişmezliği özelliğinin incelenmesi adil bir değerlendirme yapmada önemli bir rol oynamaktadır (Huh & Lee, 2009; Meredith, 1993). Dorans ve Holland (2000) da, grup değişmezliğini birbiri yerine kullanılabilir olması beklenen puanların aslında değiştirilebilir olup olmadığını değerlendirmek için bir kriter olması gerektiğini ve test eşitleme için en önemli kriter olduğunu ifade etmektedir. İki test arasındaki ilişki bireylerin kız veya erkek olmalarına bağlı olarak değişiyorsa testler aynı şeyi ölçmüyor anlamına gelmektedir. Bu durum da güvenilirliğin derecesini etkileyecektir (Liu, Feigenbaum & Dorans, 2005).

MTK modelleri gerekli koşullara sahipse gerçek puanlar bütün alt gruplar için eşitlenebilmektedir. Bireylerin gruplardaki MTK parametrelerinin değişmezliği sağlandığında, modeller için de grup değişmezliğinin / gruptan bağımsızlığın sağlandığı anlamına gelmektedir. Gözlenen puanlar için ise bu özelliğin herhangi bir teorik çerçevede doğrulanamadığı görülmektedir (Lord & Wingersky, 1984).

Büyük bir örnekleme büyük madde sayısına sahip sonuçlar, küçük örnekleme daha az sayıda maddeye sahip olan durumlara kıyasla daha az grup bağımlılığı göstermektedir (Liu, Feigenbaum & Dorans, 2005, Huh & Lee, 2009). Dolayısı ile çalışmamızda örneklem sayısının yeterli düzeyde tutulup, MCMC örnekleme sürecinin yüksek alınması ile evrene yakın bir dağılım elde etmek amaçlanmıştır.

Bayes Yaklaşımı

Klasik yaklaşımlarda olasılık sınama kriteri olan p anlamlılık değerinin temel alınması her zaman güvenilir sonuçlar vermemektedir. Bu p değeri araştırmacının amacına, örneklem büyüklüğüne göre farklılaşabilmekte ve sınır değerlerde araştırmacıyı kararsızlığa götürebilmektedir (Rounder, Morey, Speckman & Province, 2012; Berger, Baukai & Wang, 1997; Kruschke, 2012; Kruschke, 2010b). Klasik yöntemlerde sıfır hipotezinin (yokluk hipotezinin) doğrulanması temel alınmaktadır. Bu yüzden bu yöntemlere dayalı olarak analizlerden hesaplanan t ve F gibi sabit değerler sonuçların olabilirliğinin maksimum olduğunu ya da verinin model ile uyumunu göstermektedir. Ancak bu değer bize parametreler hakkında hiçbir bilgi sağlamamaktadır (Kruschke, 2010a; 2012). Araştırmalara eklenecek

parametre sayısı arttığında bu parametreler için oluşturulacak modellerin hepsi için yokluk hipotezlerinin incelenmesi zor olmakla birlikte hesaplanan p değerleri yanıltıcı sonuçlar vermektedir (Lee & Boone, 2011). Bayes yaklaşımı bu p değerindeki belirsizliği ortadan kaldırmak için çalışmaya önsel bilgileri (prior) dahil etmektedir. Ön bilgilerin araştırmaya dahil edilmesi ve olabilirlik (likelihood) kestirim yönteminin kullanılması ile sonsal (posterior) bilginin elde edilmesi klasik yaklaşıma alternatif olarak kullanılacak olan bu yaklaşımın ortaya çıkmasını sağlamıştır (Kuschke, Agunis & Joo, 2012; van de Schoot, vd., 2013; van de Schoot & Depaoli, 2014). Sonsal dağılımın sonuçları Bayes hipotez testi için gerekli bileşenleri sağlamaktadır (Kaplan & Depaoli, 2012).

Bayes ve geleneksel (klasik) yaklaşımların ikisi de istatistiksel bir modelde bilinmeyen parametrelerle ilgilenmektedir (van de Schoot & Depaoli, 2014; Kaplan & Depaoli, 2012; 651). Geleneksel yaklaşımın analizinde sadece bir tane parametre bilinmeyen olarak ele alınırken Bayes yaklaşımında bütün bilinmeyen parametreler belirsiz olarak ele alınmaktadır (van de Schoot vd.; 2013). Dolayısı ile klasik yöntemlere göre parametre kavramına yaklaşımda, parametrelerin sabit olduğu kabul edilirken; Bayes yaklaşımına göre parametrelerin her biri bir olasılık değişkeni olarak kabul edilmekte ve daha önceki bilgilerden yararlanılarak çıkarımsal bilgilerin elde edilmesini temel alınmaktadır (Boone, Ye & Smith, 2009, Kaplan & Depaoli, 2012; Oh, Guo & Walker, 2009). Ayrıca Bayes çalışmaları teorik olarak modellerin daha iyi performans göstermesini sağlamaktadır (MacEachern, 2000; Kim, Livingston & Lewis, 2009; 2011; Oh, Guo & Walker, 2009).

Eğitim alanında yapılan çalışmalar genellikle klasik yöntemlere göre değerlendirmeye alınmaktadır. Ancak klasik yöntemlerin, sadece tek bir değere bağlı kalması, yeterli düzeyde gruplar hakkında bilgi vermesi vb. gibi eksiklikleri bulunmaktadır. Bu eksikleri gidermek ve bu yöntemlerden kaynaklanan hataları azaltmak üzere birçok yeni yöntem geliştirilmeye çalışılmaktadır. Yeni yöntemler için bir çok kriter eklenerek yeni kısıtlamalar, varsayımlar ve sayıtlar getirilmektedir. Gerçek hayatta ise bu kriterlerin sağlanması genellikle mümkün olmamaktadır. Bayes istatistiksel yöntemi ise esnek bir yapıya sahip olup bir toplumun gelişimine en büyük katkı sağlayacak olan eğitim alanında yapılan çalışmalarda kullanılması ile alanın gelişmesine katkı sağlayacaktır.

Bayes yaklaşımında önsel bilgilerin dağılımının varyansı ve kesinliği ilgilenilen parametre değerleri hakkındaki belirsizlik seviyesini yansıtmaktadır. Önsel bilgilerin varyansı ne kadar az ise, parametre değerleri önsellerden yanlış şekilde etkilenebilme olasılığı da azaldığından daha kesin sonuçlar vermektedir (van de Schoot & Depaoli, 2014; Livingston & Lewis, 2009; Kim, Livingston & Lewis, 2009; 2011; Oh, Gou & Walker, 2009; Branberg, 2010). Önsel bilgi için dağılımlar bu özelliğinden dolayı içerdiği bilgi yoğunluğuna göre üç farklı sınıfa ayrılmaktadır. Bu sınıflama alanyazında farklı isimlerde yer almakla birlikte genelde; 1. Bilgi vermeyen önsel (Non-informative prior), 2. zayıf bilgi veren önsel (Weakly informative prior) ve 3. bilgilendirici önsel (informative prior) olarak adlandırılmaktadır. Sonsal dağılım, bilgi vermeyen ve zayıf bilgi veren önsel bilgilerin seçiminden çok etkilenebilir (van de Schoot & Depaoli, 2014).

Bayes yaklaşımı, analizlerde birçok avantaj sağlamaktadır. Bu avantajlardan bazıları aşağıda ifade edilmiştir:

- 1-Klasik yaklaşımlarla kestirimi yapılamayan kompleks modellerde kullanılabilme,
- 2- Güven aralığından olasılık hesaplayabilme,
- 3- Gerek duyulduğunda, araştırmalarda geçmiş bilgileri analize dahil edebilme,
- 4- Büyük örnekleme sahip olmadığımız durumlarda analiz yapabilme,

Bu gibi en önemli avantajların yanına bazı analizler için gerekli varsayımların (bir çok analiz için örneklem büyüklüğü, eşitleme için süreklileştirme veya düzgünleştirme gibi) karşılanma zorunluluğu Bayes yöntemi ile ortadan kalkmaktadır (Kruschke, 2012; Kruschke, Aguinis & Joo, 2012; Schoot, vd., 2013)

Bayes yaklaşımı klasik yaklaşımdan yaklaşık 150 yıl önce ortaya atılmıştır. Ancak uygun yazılımların yetersizliğinden ve Bayes istatistiğinin frekans istatistiğine karşı tutumundan dolayı 20. yüzyıla kadar tercih edilmemiştir (Kaplan & Depaoli, 2013;p.407). Kruschke (2012) ise artık 21. yüzyılın istatistiği olarak Bayes yaklaşımının ön plana çıktığını belirtmektedir. Ülkemizde de Bayes Yaklaşımı ile birçok alanda uygulamalar yapılmıştır (Ekici, 2005; Karabey, 2007; Can, 2008 Altuntaş, 2011; Öztürk, 2011; Cimşit, 2013; Eroğlu, 2013; Gasım,2013; Kara, 2018).

Bayes teoremi temel itibari ile koşullu olasılık teoreminden ortaya çıkmaktadır. Bir olayın olma olasılığı diğer olayın olmuş olma olasılığına bağlı olarak elde edilmektedir. Dolayısıyla Bayes yönteminde önsel olarak alınan kısım iki olaydan gerçekleşmiş olandır.

X ve Y gibi iki olayı ele aldığımızda koşullu dağılım;

$$p(Y, X) = p(Y | X)p(X) \text{ veya } p(X, Y) = p(X | Y)p(Y)$$

şeklinde ifade edilmektedir. Yukarıdaki formüller simetrik olduğundan

$$p(Y | X)p(X) = p(X | Y)p(Y)$$

ifadesi formülde yerine yazıldığında Bayes teoreminin olasılık ifadesi;

$$p(X | Y) = \frac{p(Y | X) p(X)}{p(Y)}$$

olarak karşımıza çıkmaktadır.

Bayes yaklaşımında parametrelerin olasılıkları ele alındığından olasılık dağılımları ile ilgilenilmektedir. Bir olasılık dağılımında her değer random olduğundan dolayı Bayes Teoremi;

$$p(\theta | Y) = \frac{p(Y | \theta) p(\theta)}{p(Y)}$$

olup bu eşitlikte;

$p(\theta | Y)$: Y si belli olan bir θ nın sonsal dağılımını

$p(\theta)$: θ nın önsel dağılımını

$p(Y)$: verinin marjinal dağılımını

$p(Y | \theta)$: olabilirlik fonksiyonunu temsil etmektedir (Raftery, 1995; Kruschke, 2010a)

$$p(\theta | Y) \propto p(Y | \theta) p(\theta)$$

$p(Y)$ değeri sabit bir değer olduğundan sonsal dağılımın, önsel dağılım ve olabilirlik fonksiyonu ile doğrudan ilişkili olduğu görülmektedir.

Parametrik olmayan Bayes Yaklaşımı

Bayes yaklaşımları gözlenen veri ve bilinmeyen değer için iyi tanımlanmış olasılık modelleri oluşturmaktadır. Parametrik Bayes modeli sınırlı sayıda parametre kullanabilirken, Parametrik olmayan Bayes modelleri için uygulamalarda model oluşturabilmek için böyle bir sınırlama yoktur. Bayes modellerinde parametre sayısının esnek olarak kullanılması ve sonsuz boyutlu bir parametre uzayının kullanılması Parametrik olmayan Bayes modellerinin temelini oluşturmaktadır (Müller & Quintana, 2004; Orbanz & Teh, 2010; Shah & Ghahramani, 2013). Ayrıca bu parametre uzayı verilen bir problem için bütün olası çözümlerin kümesini de içermektedir. Gonzalez, Barrientos ve Quintana (2015b, p.213-226) ise parametrik olmayan yaklaşımların dağılım fonksiyonlarının uzaya odaklanması ile başladığını belirtmiştir. Bu sebepten dolayı belirsizlik F olasılık dağılım fonksiyonu üzerinde kendi kendine açıklanabilmektedir. Random ve sabit etkilerden elde edilen bilgiler, F olasılık dağılım fonksiyonu üzerinde bir Parametrik olmayan önsel dağılıma yerleşmiş olur ve random etkilerin dağılımı bireysel etkiler hakkında çıkarım yapar (MacEachern, 2000; 2001). Bu şartlar sağlandığında ve X sınırsız bir küme olarak ele alındığında; F olasılık fonksiyonu da sınırsız boyutta olur ve F üzerinde önsel model olarak $p(F)$ tanımlanır (Gonzalez, Barrientos, & Quintana, 2015b, p.215-216; Gonzalez, & Wiberg, 2017; p.168-169; MacEachern, 2000; 2011). Bir dağılım fonksiyonundaki belirsizliği tanımlamak için kullanılan olasılık ölçümlerinin kümesi Rassal Olasılık Ölçümleri (random probability measures / RPMs) olarak adlandırılmaktadır (Müller & Quintana, 2004; Narrete, Quinana & Müller, 2008).

Parametrik olmayan Bayes Yaklaşımı ile Eşitleme

Eğitim alanında en önemli yapı taşı öğrencidir. Öğrencilerin performanslarını incelemek ise oldukça zordur. Öğrencilerin performansları herhangi bir durumdan etkilendiğinde motivasyonlarında, enerjilerinde ve dikkatlerinde değişiklik olmakta ve ölçülmek istenen özellikler için aynı araçlar kullanılsa dahi aynı performansı gösterememektedir. Uygulanan sınavların fiziki ortamı dahi bireylerin performanslarını, dolayısı ile sınavdan aldıkları başarı puanlarını etkilemektedir. Bu yüzden bireylerin önceki göstermiş oldukları performanslar kullanılarak testlerden almış oldukları puanların değerlendirilmesi daha makul bir değerlendirme yapılmasına katkı sağlayabilir.

Bayes istatistiklerinde önsel olarak bilinen bilgilerin çoğunu ortak değişkenler oluşturmaktadır (Boone, Ye & Smith, 2009). İki veya daha fazla testi birbirine dönüştürmede genellikle parametrik ve yarı parametrik eşitleme yöntemleri kullanılmaktadır (Gonzalez, Barrientos & Quintana, 2015a; 2015b). Ancak geleneksel eşitleme yöntemleri geçmiş değişkenlerin tanımlayıcı değişkenler olarak eşitleme dönüşümlerinde kullanılmasında sınırlı kalmakta iken, parametrik olmayan Bayes modeli eşitleme dönüşümleri için fonksiyonların kestiriminde “ortak değişkenleri” kullanılmasına müsaade etmektedir. Bayes yöntemi dağılım fonksiyonlarının şeklini ve konumlarını, lineer veya nonlineer regresyon modelleri gibi sadece ortalamalar veya dağılıma ait fonksiyonlar ile değil, ortak değişkenler (z) ile de değiştirebilmektedir. Dolayısıyla bu değişkenlerin işleme dahil edilmesi ile dönüşüm fonksiyonlarında değişiklikler oluşmakta ve ortak değişkenlere ait kategoriler arasında eşitleme olası hale gelmektedir (Gonzalez, Barrientos & Quintana, 2015a; Karabatsos & Walker, 2009).

Parametrik olmayan Bayes yaklaşımına göre ilk testten alınan puanın ikinci testte karşılığı daha önceki eşitleme yöntemlerinde olduğu gibi birebir ve örten özelliğine sahip değildir. Örneğin; eşitlemesi yapılacak olan testler için ilk testten 60 puan almış bir erkek ve bir kız öğrencinin yeni testten alacakları puanlar farklılaşabileceği gibi benzer şekilde matematiği sevme düzeylerine göre öğrencilerin yeni testten alacakları puanlar farklılaşabilir. İlk testten farklı puanlar almış matematiği çok seven bir kız öğrenci ile matematiği orta düzeyde seven bir erkek öğrencinin yeni testteki puanları aynı olabilir. Dolayısı ile parametrik olmayan Bayes ile klasik yöntemlerdeki gibi eşitlenmiş puanlar arasındaki ilişki incelenmemektedir. Ortak değişkenler ile grup değişmezliği varsayımı kontrol altına alındıktan sonra, bu değişkenlerin kategorileri için elde edilen eşitlenmiş puanları incelenmektedir. Dolayısı ile bireylerin geçmiş bilgilerine göre şekillenen durumlar için puanlar elde edildiğinde daha doğru sonuçlar gözlenebilecektir.

Parametrik olmayan Bayes modeli ile grup değişmezliği kontrol altında tutulduğu gibi dağılımların düzensizliği ve süreksizliği de giderilmiş olduğundan; diğer eşitleme yöntemlerinde kullanılan ön-düzgünleştirme, bant genişliği parametresinin seçimi ve eşitlemenin standart hatasının türetilmesine ihtiyaç duyulmamaktadır (Gonzalez, Barrientos & Quintana, 2015b, p.223). Bu durum ise modelin önemliliğinin bir göstergesidir (Karabatsos & Walker, 2009).

Parametrik olmayan Bayes modellerinde ortak değişkenler, bağımlı rassal (random) olasılık değişkeni olarak işleme alınmaktadır. Sürece önsel olarak dahil edilecek olan ortak değişkenler / kategorik ise Bernoulli dağılımı (Boone, Ye & Smith, 2009), sürekli ve bir aralığa sahip ise Dirichlet dağılımı kullanılabılır (Gonzalez, Barrientos & Quintana, 2015a). Dirichlet süreci, dirichlet dağılımının sürekli uzayda bir uzantısı olarak oluşmaktadır. Model olarak Dirichlet süreçleri (DS / Dirichlet Process-DP) birçok alana uygulandığı gibi parametrik olmayan Bayes çıkarımlarında da merkezi bir role sahiptir (Gonzalez, Barrientos & Quintana, 2015b; De Iorio, Müller, Rosner & MacEachern, 2004, Petrone, 1999a). Bu merkezi rolün en büyük sebeplerinden biri Dirichlet süreçlerinin birçok gösteriminin olmasıdır. Dirichlet dağılımının gösterim şekillerinden biri de Sethuraman'ın Dirichlet süreci tanımlamasıdır. Sethuraman kullandığı bu gösterimlerde temel nokta, her bir fonksiyonun yeri ile ilgili kesikli bir değişkenin tanımlanmasıdır (MacEachern, 2000). Dirichlet sürecinin Sethuraman (1994) tarafından tanımlanan formülü *stick-breaking* olarak adlandırılmaktadır. Ancak DS kesikli bir yapıya sahip olup, sürekli bir kernel kıvrımın eklenmesi ile Karma DS (KDS / DP Mixture-DPM) yapısı haline gelerek kesiklilik sürecinden farklılaşmaktadır (Müller & Quintana, 2004; De Iorio, Müller, Rosner & MacEachern, 2004).

Test puanlarını etkileyecek olan değişkenlerin puan dağılımları üzerindeki random etkisi ise bir bağımlılık durumunu ortaya çıkarmaktadır. Bağımlı Parametrik olmayan süreçlerin gelişmesi, önsel dağılım için esnek bir kullanım oluşturması ile modelin Bağımlı Dirichlet Süreci (BDS / Dependent Dirichlet Process-DDP) ile açıklanmasını sağlamaktadır (MacEachern; 1999, 2000). Yani, BDS de bağımlılığı sağlayan temel nokta kümelerdeki her elementin *stick-breaking gösterimidir*. Bu süreçte yer alacak dağılım fonksiyonlarının Dirichlet süreci ile bağımlılık kümesindeki her elemanın stick-breaking gösteriminin modifiye edilmesi durumunda BDS oluşturulmaktadır (Barrieston, Jara & Quintana, 2016). BDS nin arkasındaki temel yargı, her olası kestirici değer için DS ölçümlerinin random olarak yapılandırılmasıdır (Barrieston, Jara & Quintana, 2012).

Bağımlı Dirichlet Süreci için ilk olarak ağırlıkların kullanıldığı bir formül MacEachern'ın(1999,2000) yaptığı çalışma ile ortaya atılmış, daha sonra ise aynı formülün bağımsız kestirici ağırlıkları ile sürecin alternatif formunu oluşturmuştur.

MacEachern in alternatif olarak ifade ettiği $\{G_z : z \in Z\}$ ortak değişken-bağımlı sürekli dağılımlara ait formül aşağıdaki gibidir;

$$G_z(\cdot) = \sum_{i=1}^{\infty} w_i \delta_{\theta_i(z)}(\cdot)$$

Burada $\delta_{\theta}(\cdot)$ Dirac ölçümü olup $\theta, \theta_j(x)$ üzerinde bir yığılma noktası, $j = 1, \dots$ değerleri X index kümesi ile bağımsız skolastik süreçleri göstermektedir.

$\beta(1, M)$ marjinal dağılım ve X index kümesi ile bağımsız skolastik süreçler $j = 1, \dots$ ile gösterilmek üzere $V_j(x)$ ile birlikte $w_j = V_j(x) \prod_{i < j} [1 - V_i(x)]$ den elde edilen ağırlıkları göstermektedir.

Parametrik olmayan olasılık dağılımları ile ilişkili çeşitli olasılık modelleri bulunmaktadır (Barrieston, Jara & Quintana, 2012). Çalışmamızın amacı doğrultusunda, puanların eşitleme için sürekli olarak ele alınmasından kaynaklı olarak Dirichlet dağılımı yerine kapalı bir aralığa sahip ve sayı dizisinin yoğunluğunun değişmezliğini sağlayan Bernstein polinomu kullanılması daha uygun olacaktır.

Random Bernstein Polinomları

Parametrik olmayan Bayes yaklaşımı için önsellerin seçiminde oldukça zorluk yaşanmaktaydı. Bu durumda Petrone (1999a), yoğunlukların bazı uygun dönüşümlerini sağlamak ya da yoğunluk fonksiyonlarının kestirim sürecinde model oluşturmak için olasılık doğasına dayalı olan polinaminal yaklaşımların en basit örneği olan Bernstein polinomlarını kullanmıştır. Bernstein polinomlarında, birim aralıktaki bütün dağılım fonksiyonlarının bir olasılık ölçümü olan Bernstein önselleri (Bernstein Prior) kullanılmaktadır (Petrone & Wasserman, 2002). Böylece dağılım fonksiyonları için önsellerin seçimi daha makul hale gelmiştir.

$$b(; k, w_{1,k}, w_{2,k}, \dots, w_{k,k}) = \sum_{j=1}^k w_{j,k} \beta(x; j, k - j + 1)$$

Bu eşitlikte k parametresi düzleştirme parametresidir. k nın seçimi yoğunluk kestiriminde literatüre bağlıdır. Burada ise k bir random değişken olup k ve veri üzerinde Bayes yoğunluklarının kestiriminin koşulsal olarak hesaplanmasından ve daha sonra ise k nın sonsal kestirimi ile ilişkilidir. Yani k terimi random olarak ele

alındığında Bernstein polinomları da random olur (Petrone,1999a). β dağılımı ise yoğunlukların toplamı için temel sağlamaktadır.

T bir kümülatif dağılımı temsil etmek üzere $[0,1]$ aralığında bir tanım kümesinden R ya tanımlı, k . dereceye sahip ve ortak değişkenlerin uygulanmasında kolaylık sağlayan bir fonksiyondur. Barrientos, Jara ve Quintana (2016), çalışmasında Bernstein fonksiyonlarında, kümülatif dağılım için $T(0) = 0$ ise:

$w_{j,k} = T(j/k) - T((j-1)/k)$, ve a ve b parametreleri ile bir Beta yoğunluk dağılımı $\beta(\cdot | a, b)$ olarak gösterilmek üzere, bu yoğunluk fonksiyonu aşağıdaki şekilde göstermiştir;

$$bp(y|k, T) = \sum_{j=1}^k w_{j,k} \beta(y|j, k-j+1)$$

Tek bir yoğunluk kestirim problemi için bu formülü Petrone (1999a,b) tekrar düzenleyerek önce Bernstein polinom önseli (Bernstein polynomial prior-BPP) yoğunluğu ve daha sonra ise w_j ağırlıklarının Dirichlet sürecinden geldiğini ortaya koyarak Bernstein-Dirichlet önseli (BDÖ / Bernstein-Dirichlet prior-BDP) modelini ortaya atmıştır. Petrone (1999a,b) çalışmalarında MCMC algoritmasını bu modele dahil etmiştir.

Daha sonraki süreçte Barrienston, Jara ve Quintana (2012; 2016) çalışmalarında MacEachern in (1999) ortak değişkenlere bağlı olarak oluşturduğu BDS yi ve Petrone (1999a,b) nın BDÖ yerine KDS modelini kullanarak *Bağımlı Bernstein Polinom Süreci (DBPP-Depent Berntein polynominal process)* modelini oluşturmuştur. Bu model normal koşullar ile sınırlanma probleminin olmadığı koşullarda bile daha iyi bir performans göstermektedir. Ferguson (1973)'ün KDS modellerinin DS önselleri olduğunu ifade etmiş olması da modelin geliştirilmesinde yararlanılan noktalardandır.

$$g_x(\cdot) = \sum_{j=1}^{\infty} w_j(\mathbf{x}) \beta(\cdot | \lfloor k\theta_j(\mathbf{x}) \rfloor, k - \lfloor k\theta_j(\mathbf{x}) \rfloor + 1),$$

Bu formülde

$$\theta_j(x) = h_x(z_j(x)), \quad H = \{H_z: z \in Z\}$$

$$w_j(x) = v_x\{n_j(x)\} \prod_{l < j} [1 - v_x\{n_l(x)\}]$$

olarak verilmektedir. Ayrıca $[\cdot]$ ceiling fonksiyonunu temsil etmektedir. (Ceiling fonksiyonu aralıkta en küçük tamsayıya eşitler ($x \leq$))

Bu ifade ile $G = \{G_x: x \in X\}$ fonksiyonu **Bağımlı Bernstein Polinom Süreci (Dependent Bernstein Polynomial process)** olarak adlandırılmaktadır ve DBPP olarak gösterilmektedir. Çalışmada BBPS olarak kısaltılmıştır.

BBPS nin iyi tanımlanmış olması eşlemelerin ölçülebilirliğine bağlıdır. Eşlemelerin ölçülebilirliği ise formülün oluşumunu sağlayan parametre değerlerinin olasılık dağılımlarını ilişkili koleksiyonu olan $G = \{G_x: x \in X\}$ e götürmesi süreci ile mümkündür (Barrientos, Jara & Quintana, 2016)

Barrientos, Jara ve Quintana (2012; 2016) çalışmasında, BBPS nin iki özel durumundan bahsetmektedir. BBPS nin özel durumlarından biri olarak “ortak ağırlıklar” ve “kestirici bağımlı destek noktaları” ile bağımlı stick-breaking süreci bir arada kullanıldığında “*tekil (single) ağırlıklı BBPS (wDBPP)*” modelini oluşturmaktadır. Genel formülde tanımlanan tekil ağırlıklı BBPS (wBBPS) yi elde etmek için gerçek değerli skolastik süreçler ($[0,1]$ aralığındaki bağımsız random değişkenler olan v_j ler, dağılım fonksiyonu olan F_0 in ortak ağırlıkları ile yer değiştirerek tekrar oluşturulmaktadır.) yeniden düzenlenmiştir.

Barrientos, Jara ve Quintana (2016) un formülazsında ortak değişkenler z ile gösterildiğinde;

$\forall z \in Z$ için F_z ortak değişken bağımlı puan dağılımını temsil edecek ve $\{F_z: z \in Z\}$ olarak gösterilecektir. wBBPS nin formülü;

$$f_{(z)}(\cdot) = \sum_{j=1}^{\infty} w_j \beta(z | [k\theta_j(z)], k - [k\theta_j(z)] + 1)$$

olarak son halini almaktadır. Beta dağılımlarının sonsuz bir kümesini temsil eden bu model, test puanlarının ortak değişken bağımlı örneklem yoğunluklarına sahip olduklarını ifade eder.

$\{F_z; z \in Z\} \sim \text{wBBPS}(\alpha, \lambda, \psi, H)$ olarak gösterebiliriz. Burada $H = \{h_z; z \in Z\}$;

$v_1, v_2, \dots$, $\alpha > 0$ olmak üzere $\beta(1, \alpha)$ dağılımı ile tanımlı bağımsız ve tanımlı random değişkenleri; k , sonlu boyutlu bir λ değişkeninden indekslenen kesikli bir random değişken; $\theta_{j(z)} = h_z(r_j(z))$, $r_1, r_2, \dots$; ψ parametresi ile indekslenen bağımsız ve tanımlı scolastik süreçleri temsil etmektedir. Bu model puanlar için sadece esnek bir model olmak ile kalmayıp ortak değişken bağımlı eşitleme dönüşümünü vermektedir (Gonzalez, Barrieston & Quantiana, 2015b).

Önsel dağılım (Prior distributuion).

wBBPS modelinin sayısal uygulaması MCMC metoduna dayanmaktadır. wBBPS modeli için $h_z(\cdot) = \exp\{\cdot\} / (1 + \exp\{\cdot\})$, $r_j(z) = z^T \gamma_j$ ve $\gamma_j | \mu, S \sim^{iid} N_p(\mu, S)$, $j = 1, 2, \dots$ kullanılmaktadır.

Burada;

$$v_j | \alpha \sim \beta(1, \alpha),$$

$$k | \lambda \sim \text{Poisson}(\lambda) \mathbb{I}_{\{k > 1\}},$$

$$\mu | m_0, S_0 \sim N_p(m_0, S_0),$$

$$S | v, \psi \sim IW_p(v, \psi)$$

$IW_p(v, A)$; ölçek matrisi A , serbestlik derecesi v olan p boyutlu inverted-Wishart dağılımını göstermektedir. Gonzalez, Barrieston ve Qantiana (2015a) de yapmış olduğu çalışmada model ile uyumlu olarak bulunan değerleri 2015b'deki çalışmasında da alması üzerine prior dağılımının oluşturulma aşamasında $\lambda = 25$, $m_0 = \mathbf{0}_p$, $S_0 = 2.25 * I_p$, $v = p + 2$ ve $\alpha = 1$ değerleri kullanılmıştır. wBBPS modelinin sonsal dağılımını açıklamak; dolayısı ile bütün model parametrelerinin sonsal örneklemelerini elde etmek için MCMC algoritması uygulanmıştır.

Sonsal çıkarım (Posterior inference).

Sonsal çıkarım tüm sabitler ve daha önceden model için uygun olduğu görülen katsayılar ile son haline getirilerek aşağıdaki formül ile ele alınmıştır;

$$\begin{aligned}
& p(v, k, w, \gamma | y, z) \\
& \propto \prod_{i=1}^n \left[\sum_{j=1}^{10} w_j \beta \left(y_i \left| k \frac{e^{z_i^T \gamma_j}}{1 + e^{z_i^T \gamma_j}} \right|, k - \left| k \frac{e^{z_i^T \gamma_j}}{1 + e^{z_i^T \gamma_j}} \right| \right. \right. \\
& \quad \left. \left. + 1 \right) \right] \left[\prod_{j=1}^{10} \beta(v_j | 1, 1) \right] \\
& \times \left[\frac{25^k e^{-25}}{k! (1 - e^{-25})} \right] \left[\prod_{j=1}^{10} (2\pi) |S|^{-\frac{1}{2}} e^{-0.5(\gamma_j - \mu)^T S^{-1}(\gamma_j - \mu)} \right] (2\pi) |S_0|^{-\frac{1}{2}} e^{-0.5(m_0)^T S_0^{-1}(m_0)} \\
& \times \frac{|\psi|^2}{2^2 \Gamma_2(2)} |S|^{\frac{7}{2}} e^{-\frac{1}{2} \text{tr}(\psi S^{-1})}.
\end{aligned}$$

Bu formülasyonu takiben sonsal bir kestirici dağılım üretmek için aşağıdaki formül kullanılmıştır;

$$p(T|y_i, z_i) = \int p(v, k, w, \gamma | y, z) L(T|v, k, w, \gamma) dv dk dw d\gamma$$

Burada;

$$L(T|v, k, w, \gamma) = \sum_{j=1}^{10} w_j \beta \left(T \left| k \frac{e^{z_i^T \gamma_j}}{1 + e^{z_i^T \gamma_j}} \right|, k - \left| k \frac{e^{z_i^T \gamma_j}}{1 + e^{z_i^T \gamma_j}} \right| + 1 \right).$$

T puan dağılımımızı göstermek üzere, X testinden elde edilen puanlar t_x , Y testinden elde edilen puanlar t_y olarak; z de ortak değişkenlerin bir gösterimi olarak ifade edilirse genel ve ortak değişken eklenmesi ile elde edilen dönüşüm fonksiyonlarımız aşağıdaki gibi olmaktadır;

$$t_x = F^{x^{-1}}(\cdot)$$

$$t_y = F^{y^{-1}}(\cdot) \text{ için } t_y = \varphi(t_x) = F^{y^{-1}}(F^x(\cdot)) \text{ olarak gösterilirken}$$

$$t_{z_x} = F^{z_x^{-1}}(\cdot)$$

$$t_{z_y} = F^{z_y^{-1}}(\cdot) \text{ için } t_{z_y} = \varphi(t_{z_x}) = F^{z_y^{-1}}(F^{z_x}(\cdot)) \text{ olarak ifade edilebilir.}$$

Model için gerekli tüm adımlar R 3.2.1 istatistiksel paket programında kod olarak yazılmış ve analiz edilmiştir. Gonzalez, Berrienston ve Quantiana (2015a,b) in yapmış olduğu çalışmalarda uygun sonuç veren sabitler kullanılarak analiz adımları başlatılmıştır. Her bir veri seti için sonsal çıkarımlarda 150000 MCMC örnekleme alınmış ve iterasyonlar değerlerin en iyi olduğu duruma kadar devam ettirilmiştir. Daha sonraki aşama için ise eşitleme dönüşümü yapılarak eşitlenmiş puanlar elde edilmiştir.



İlgili Araştırmalar

Yang (1997) tarafından yapılan çalışmada, denk olmayan gruplar eşitleme deseni kullanılarak ortak maddelerin güçlüğüünün testin güçlüğüünü yansıttığı ve ortak maddelerin orta güçlükte olduğu iki farklı yapıdaki ortak testin eşitlemeye etkisi incelenmiştir. Araştırmada, ortak maddelerin güçlüğüünün testin güçlüğüünü yansıttığı durumlarda eşitlemenin daha doğru yapıldığı sonucuna ulaşılmıştır.

Wei (2010) çalışmasında, DoGOM deseninde kullanılan ortak maddeler için gerekli olan temsiliyetlerin karşılanmaması durumunda eşitleme sonuçlarını ve ölçek puanlarının istikrarının nasıl etkileneceğini araştırmıştır. Ortak maddelerin içerik olarak ve istatistiksel olarak testleri temsil edip etmediği durumlar koşul olarak alınarak simülasyon verisi üzerinde çalışmasını yürütmüştür. Araştırma sonucunda ortak maddelerin testleri içerik olarak temsil durumlarının ihlalinin, istatistiksel olarak temsil etme durumunun ihlalinin daha etkili olduğu sonucu bulunmuştur. Hem içerik hem de istatistiksel özelliklerin bütün testi temsil etmesi için ortak maddelerin zorluklarının düzenlenmesi ve dengeli bir içerik hazırlanmasının önemine dikkat çekmiştir.

Moses, Deng ve Zhang (2010) çalışmasında, ortak testler zayıf olduğunda veya gruplar oldukça farklı olduğunda tek bir ortak testin kullanılmasının eşit olmayan gruplar arasındaki farkı ortaya çıkarmakta yeterli olmadığını ve yanlış sonuçlar verebileceğini ifade ederek birden çok ortak test kullanmıştır. DoGOM deseninde birden fazla ortak test kullanmak için son tabakalama, yerine atama ve eğilim puan eşlemesi yaklaşımlarını kullanmıştır. Yerine atama ve son tabakalama yöntemi için log lineer öndüzgünleştirme metodu kullanılmıştır. Araştırmanın genel sonucunda ise son tabakalama yaklaşımının diğer yaklaşımlardan daha iyi sonuçlar verdiği görülmektedir.

Kim, Livingston ve Lewis (2011), çalışmasında küçük örneklemin kullanıldığı durumlarda tanımlayıcı bilgilerin eşitleme çalışmalarında kullanılması ile deneysel bir Bayes prosedürünü ele almıştır. Yeniden örnekleme aynı testin iki farklı formunda yapmıştır. Zincirleme lineer ve zincirleme ortalama eşitleme metodlarının deneysel ve deneysel olmayan Bayes versiyonları ortak maddeler için toplanan veri setinden tekrarlanan küçük örneklemlere uygulanmış ve büyük örneklem kullanılması durumunda zincirleme lineer eşitleme sonucunu karşılaştırılmıştır. Aynı

testin farklı formlarının eşitlenmesi tamamlayıcı bilgiler aracılığıyla sağlamıştır. Temel formdaki örneklem sayısı sabit kalırken hedef form için örneklem sayısının 10 dan 200'e değişimi ve testlerin zorlukları arasındaki farklılıklar ele alınmıştır. Hedef formda 50 veya daha az örnekleme sahip olunması durumunda deneysel Bayesin eşitlemenin doğruluğunu arttırdığı görülürken; hedef testin temel testen daha zor olması durumunda daha az doğru sonuçlar verdiği görülmüştür.

Kim, Livingston, Lewis (2009), küçük örnekleme yapılan eşitleme çalışmalarında, eşitlemenin doğruluğunu arttırmada tamamlayıcı bilgilerin kullanılmasının etkililiğini incelemiştir. Tamamlayıcı bilgiler olarak; benzer test uzunluğu ve test zorluğuna sahip diğer testlerden elde edilmiş olan bilgiler kullanılmıştır. Bu bilgiler eşitlenmiş puanların direkt olarak elde edilebileceği Bayes kestirimlerinden Deneysel Bayes metoduna uygulanmıştır. Küçük grupların eşitlenmesini değerlendirmek için büyük grupların eşitlemeleri kriter olarak kullanılmıştır. Hedef testin 5 farklı örneklem büyüklüğünde ele alınması ile örneklem büyüklüğünün 25 ve daha az olduğu durumda, tamamlayıcı bilgilerin, eşitlemenin kesinliğini artırma eğiliminde olduğu görülmüştür.

Livingston ve Lewis (2009) yapmış olduğu çalışmada örneklem sayısının az olması durumunda test eşitlemesi yapıldığında oluşabilecek hataları önlemeye yönelik olarak ampirik / deneysel Bayes yaklaşımından yararlanarak karşılaşılabilecek eşitleme sorunlarının önüne geçilebileceğini göstermiştir. Her bir puan için ayrı ayrı eşitlenmiş puanların elde edilebileceğini ve puan dağılımları veya eşitleme dönüşümlerinin modelde kullanılmasının gerekli olmadığını ifade etmektedir. Bireylerin gerçek puanları kullanılarak yapılan kestirimlerde önsel bilgi olarak diğer test formlarının eşitlenmesinden elde edilen bilgiler kullanmıştır. Uygulamayı basit bir şekilde yürütebilmek adına replikasyon serilerini kullanmıştır.

Wiberg ve von Davier (2017) çalışmasında ortak değişkenleri kullanarak zamanla veya farklı değişkenlerden etkilenerek değişiklik gösteren ortak maddeleri incelemiştir. Çalışmada, ortak maddeler kümesinde (maddelerinde veya içeriğinde herhangi bir değişiklik olmaksızın birçok defa uygulamasında) nitelikli bir kontrol süreci sağlayarak, ortak maddelere cevapların zamanla değişip değişmediğini ve farklı gruplar için değişmezliği sağlayıp sağlamadığını göstermek amaçlanmıştır. Bu sayede ortak maddelerin kalitesinin izlenmesi ve dolaylı olarak da temel testlerin izlenmesi sağlanacaktır. Araştırmada betimsel istatistikler, ANOVA ve karma lineer

etki modelleri kullanılarak ortak testlerde ortak deęiřkenlerin etkisi incelenmiřtir. Arařtırma sonucunda farklı uygulamalarda, testi alanlara ait ortak deęiřkenlerin kullanılmasının incelenmesi adil testlerin ortaya ıkmasına yardımcı olabileceęi sonucu ortaya konulmuřtur.

Liou, Cheng ve Li (2001) yapmıř oldukları alıřmada, denk olmayan gruplarda grupların seiminde rneklemden kaynaklanabilecek olan yanlılıęı nlemeye ynelik olarak ortak maddeler yerine ortak deęiřkenlerin kullanılması durumunu incelemiřtir. Ortak deęiřkenler ile btn daęılımın kestirimi iin genel bir model oluřturulmuřtur. Oluřturulan model parametreleri EM algoritması kullanılarak kestirilmiřtir. Model sonucunda elde edilen eřitlenmiř puanların karřılařtırılmasında standart hata kullanılmıřtır.

Wiberg ve Branberg (2015) te yapmıř olduęu alıřmada, DoGOD desenin de test puanları ile iliřkili olan deęiřkenlerden yararlanarak kernel eřitleme ve denk olmayan gruplarda ortak maddenin kullanıldıęı desende son dzgnleřtirme (post-stratification) iin geliřtirilen modelden yararlanarak eřitleme iřlemini gerekleřtirmiřtir. Desenin kullanımını gstermek iin bir kolejden elde edilmiř verilerden yararlanılmıřtır. Bu alıřmada eřdeęer gruplarda yapılan eřitleme desenindeki eřitlenen puanlar ile DoGOD deseninden elde edilen eřitlenmiř puanların karřılařtırması yapmıřtır. Sonular, eřit gruplardaki tasarımı ile karřılařtırıldıęında DoGOD tasarımının daha dřk standart hatalar retebileceęini gstermektedir. Aynı zamanda ortak test ile birlikte ortak deęiřkenler kullanıldıęında, test puanlarının byk ranjı zerinde en kk standart hata elde etmiřtir.

Cho (2007) tarafından yapılan bir arařtırmada, oklu kategorilerde puanlanan yazma testleri Klasik Test Kuramına dayalı eřit yzdelikli eřitleme yntemine ve rtk zellikler Teorisi'ne dayalı yntemlere gre leklenerek eřitlenmiř puanlar elde edilmiřtir. Arařtırmada ortak maddeli denk olmayan gruplar deseni kullanılmıřtır. Arařtırmada Klasik Test Kuramına dayalı eřit yzdelikli eřitleme ynteminin, Madde Tepki Kuramına dayalı eřitleme yntemlerine gre daha hatasız eřitlemeler yaptıęı saptanmıřtır. Elde edilen bu sonucun daha nce zetlenen alıřmaların sonularıyla ters dřtę grlmektedir.

Oh, Guo ve Walker (2009) da çalışmasında küçük gruplarda yapılan kestirimlerin kesinliğini az olacağını ifade etmiş ve küçük örneklemelerin varlığı durumunda Deneysel Bayes yaklaşımına dayalı güvenilirlik kestirimlerini kullanarak, bir grubun alt gruplarının güvenilirlik kestirimleri için kesinliğin artıp artmadığını incelemiştir. Bayes yaklaşımını kullanırken tamamlayıcı bilgilerden yararlanmıştır. Çalışmada birkaç Bayes kestirimi diğer metotlardan elde edilen kestirimler ile varyans ve yanlılıkları bakımından karşılaştırılmıştır. Aynı zamanda çalışmada kullanılan farklı deneysel Bayes yaklaşımları da farklı örneklem büyüklüklerinde karşılaştırılmıştır. Çalışma sonucunda tamamlayıcı bilgiler kullanılarak elde edilen Bayes kestirimlerinin birkaç varsayım altında daha büyük yanlılık, daha düşük hata ürettiği görülmektedir. Örneklem sayısı arttıkça deneysel Bayes ve geleneksel yöntemlerden elde edilen dağılımlarının şekli birbirine yaklaşmakta ve RMSE, ortalama varyans ve yanlılık karelerinin ortalaması azalmaktadır. Deneysel Bayes temelli güvenilirlik kestirimlerinin küçük örneklemelerde bile umut verici sonuçlar verdiği sonucunu elde etmiştir. Ancak küçük gruplarda alt gruplara ait kestirimlerde örneklem daha da azalacağından dolayı deneysel Bayes temelli güvenilirlik kestirimleri çok iyi işlev göremeyebilir. Bu yüzden tamamlayıcı bilgilerin özelliklerinin dikkatli bir şekilde ele alınması önerilmektedir.

Karabatsos ve Walker (2009) da yapmış olduğu çalışmada test puanlarını eşitlemek için bir Parametrik olmayan Bayes modeli tanımlamıştır. Parametrik olmayan Bayes modelinin bütün eşitleme desenlerine uygulanabileceğini ve bu modelin ilk eşitleme modellerine göre daha avantajlı olduğunu ifade etmiştir. İlk eşitleme modellerinin aksine Parametrik olmayan Bayes modeli iki testten elde edilen puanların dağılımları arasındaki pozitif bağımlılığı hesapladığını ifade etmiştir. Ayrıca araştırma sonuçlarına göre eşit yüzdelikli, doğrusal ve ortalama eşitleme yöntemlerinin önsel bilgileri içeren Bayes modelinin özel bir hali olduğu ifade edilmiştir.

Lee ve Ban (2010) tarafından yapılan araştırmada eş zamanlı kalibrasyon ile eşitleme yöntemi, Stocking-Lord yöntemi ve Haebara yöntemi karşılaştırılmıştır. Araştırma sonucunda, eş zamanlı eşitleme yönteminin diğer yöntemlere nazaran daha fazla eşitleme hatası verdiği saptanmıştır. Haebara ve Stocking-Lord yöntemleri karşılaştırıldığında, Haebara yönteminin Stocking-Lord yöntemine nazaran daha az eşitleme hatasına sahip olduğu saptanmıştır.

Branberg ve Wiberg (2011) eşit gruplar ve eşit olmayan gruplarda ortak değişken deseninde gözlenen puanlarda doğrusal eşitleme yöntemini ele almıştır. Çalışma kapsamında ortak değişkenlerin eşitleme sürecine dahil edilmesi ile eşitlenmiş puanların kestirimi üzerindeki etkisi incelenmiştir. İlk olarak İsveç Akademik Değerlendirme Testinin iki defa uygulaması sonucu elde edilen puanlarda ikinci olarak da bir simülasyon verisinde cinsiyet ve eğitim gibi ortak değişkenlerin alınması ile doğrusal eşitleme yöntemi kullanılarak eşitleme yapılmıştır. Bu modellerin kullanılması ile kestirimlerin ortalama hata karelerinin azalması ve eşitlemenin kararlılığı incelenmiştir. Eşitleme sürecinde ortak değişkenlerin kullanılması ile eşitlemenin kesinliğinin arttığı ve ortalama hata karelerinde azalma olduğu gözlemlenmiştir.

Branberg (2010) yapmış olduğu tez çalışmasında ortak değişkenleri kullanarak eşitleme çalışmasını sürdürmüştür. Çalışmasını farklı aşamalarda ele almıştır. İlk aşamada ele aldığı ortak değişkenler ile kullandığı test puanlarındaki formların kesikli değişkenli dağılımlarını tanımlamak için polinomial log lineer modelinden yararlanmıştır. İkinci aşamada hedef gruptaki puan dağılımlarının kestirilmesi aşaması olup bu aşamada, ortak değişkenlerin verilmesi halinde X ve Y testlerinin durumsal dağılımını aynı popülasyondan geldiği varsayımına dayandırarak DoGOM deseni için son tabakalama (post stratification Equation) son düzgünleştirmesi kullanmıştır. Üçüncü aşamada Kernel eşitlemenin düzgünleştirme adımlarını uygulamış ve son aşamada ise δ -metodu kullanılarak eşitlemenin standart hatasını hesaplamıştır. Tüm bu durumlar için teorik bir çerçeve oluşturmuş ve ortak değişkenlerin ortak maddeler yerine kullanılabileceğini ifade etmiştir.

Gonzalez, Barrientos ve Quintana (2015a) yapmış olduğu çalışmada, eşitleme dönüşümleri için kullanılan puan dönüşümlerinin kestiriminde ortak değişkenlerin Parametrik olmayan Bayes modelinde kullanımını temel almıştır. Bu model ile puanların dağılım şekillerinin ortak değişkenlerle değiştiğini ifade ederek her bir yöntem için dönüşüm fonksiyonu oluşturmuştur. Oluşturulan modelleri hem gerçek veriler ile hem de simülatif veriler kullanarak diğer eşitleme yöntemleri ile karşılaştırmıştır. Parametrik olmayan Bayes modeli kullanılarak yapılan eşitleme ile test eşitleme için sağlanması gereken birçok varsayımın sağlanmasına gerek kalmadığını ve ortak madde olmadığı durumlar için de eşitlemenin mümkün olduğunu ifade etmiştir.

Huh ve Lee (2009) çalışmasında simülasyon ile yapılandırılan birkaç uyum durumu altında grup değişmezliğini incelemiştir. Lineer, paralel lineer, düzgünleştirilmemiş eşit yüzdelikli ve son düzgünleştirilmiş eşit yüzdelikli metotları karşılaştırılmıştır. Araştırma sonucunda madde sayısı ve örneklem sayısı daha fazla olan testlerin daha az grup bağımlı oldukları; kolay testleri zor testlere bağlama puanlarının ve uzun testi kısa teste bağlama puanlarının aksi durumundan daha az grup bağımlılığı gösterdiği görülmektedir.

Powers, Turhan, ve Binici, (2012) çalışmasında okuma ve matematikteki durumlarını değerlendirmek için dikey ölçeklemenin grup duyarlılığını incelemiştir. Çalışmada alt gruplar kız ve erkekler olarak belirlenmiştir. Ortak madde deseninin kullanıldığı çalışmada matematik testi için alt grupların okuma testinden daha çok benzerlik gösterdiği görülmüştür.

İlgili Araştırmalar Özet

İlgili araştırmaları genel olarak özetlediğimizde; denk olmayan gruplarda ortak madde veya ortak testlerin eşitenecek testleri temsil etme durumlarının ihlali sonucunda anlamlı sonuçlar elde edilmemeyeceği gözlemlenmiştir. Bazı araştırmacılar bu gibi durumların önüne geçmek için birden fazla ortak test veya bireylerin özelliklerini iyi temsil eden ortak değişkenler kullanmıştır. Ortak değişkenlerin dikkatli seçilerek denk grup veya tek grup deseninde kullanılmasında eşitlemenin yanlılığını ve ortalama hata karelerinde azalma olacağı gözlemlenmiştir. Denk olmayan gruplarda sadece ortak değişkenin kullanılması durumunda DoGOD deseni kullanılarak iki farklı test arasında eşitleme çalışmalarının yapılabileceği gösterilmiştir. Eşitleme sürecinde ortak değişkenlerin modele dahil edilmesinde kolaylık sağlayan Bayes modelinin kullanılması ile daha kesin kestirimler elde edilebileceği gibi eşitleme sorunlarının önüne geçilebileceği vurgulanmıştır. Ayrıca daha küçük örneklerde Bayes modelinin avantajlı olduğu ve Parametrik olmayan Bayes modelinin eşitlemede kullanılması ile test eşitleme için sağlanması gereken birçok varsayımın sağlanmasına gerek olmadığı ifade edilmiştir.

Bölüm 3

Yöntem

Bu bölümde araştırmanın türü, araştırmanın evren ve örnekleme, veri toplama süreci ve verilerin analizi ile ilgili bilgiler yer almaktadır.

Araştırmanın Türü

Araştırmada ortak maddelerin bulunmadığı DoGOD deseninde farklı ortak değişkenler ile Parametrik olmayan Bayes modeli kullanılmıştır. Modellere göre elde edilmiş olan puan dağılımları ve eşitlenmiş puanlar incelenmiştir. Araştırma gerçek veri üzerinde yürütülmüş olup PoB modeline göre elde edilen eşitlenmiş puanlara ait dağılımlar ile madde tepki kuramına dayalı olarak ölçekleme yöntemlerinden elde edilen eşitlenmiş puanların dağılımları karşılaştırılmıştır. Dolayısı ile kurama katkı sağlayacak nitelikte olan bu araştırma temel bir araştırma niteliği taşımaktadır (Karasar, 2006).

Araştırmanın Evreni ve Örnekleme

Araştırmanın amacı denk olmayan gruplar arasında eşitleme yapmak olup; bu koşulu karşılayacak gruplara PISA 2012 verilerinden ulaşılmıştır. PISA 2012 matematik sonuçlarına göre 13. sırada 518 ortalama puan ile yer alan Kanada verisi ile 32. sırada 485 ortalama puan ile yer alan İtalya verisine OECD nin kendi sitesinde yayınladığı veri tabanından (<http://www.oecd.org/pisa/data>) ulaşılmıştır. Kitapçıklarda kayıp ve eksik veriler temizlendikten sonra, 5. kitapçık için 908 kişilik İtalya verisi, 6. kitapçık için 931 kişilik Kanada verisi kullanılmıştır. Bu kitapçıklara ait bilgiler Tablo 5 de yer almaktadır;

Tablo 5:

Kitapçıklara ait bilgiler;

	Ortak olmayan madde sayısı	Ortak madde sayısı	Testi alan kişi sayısı
Kanada 6. Kitapçık	24	12	931
İtalya 5. Kitapçık	24	12	908

İki ülkenin matematik bilişsel testinde göstermiş oldukları performanslar arasında fark olduğunu yapılan sıralamalardan görmek mümkündür.

Veri Toplama Araçları

PISA 2012 kapsamında öğrencilere uygulanan matematik okuryazarlığını ölçen bilişsel testten ve öğrenci anketinden yararlanılmıştır.

Araştırmanın verilerini,

- İtalya öğrencilerinin 5. kitapçıkta matematik alt testine verdikleri cevaplar,
- Kanada öğrencilerinin 6. kitapçıkta matematik alt testine verdikleri cevaplar;

Aynı öğrencilerin

- Cinsiyet bilgileri ve
- Matematik öz yeterlilik anketine verdikleri cevaplar oluşturmaktadır.

Denk olmayan gruplarda yapılan eşitleme çalışmalarında, eşitleme hatasını azaltmak için testlerde soru sayısının en az %20 si kadar ortak madde içermesi gerekmektedir (Angoff, 1971). Matematik sorularının eşit sayıda olması ve ortak madde sayısının fazla olması nedeniyle 5. ve 6. kitapçıklar seçilmiştir.

Kitapçıklarda ikili ve kısmi puanlanan maddeler bulunmaktadır. Farklı bir kriter olarak modeli etkilemesine engel olmak için tüm maddeler aynı kategori sayısında ele alınmıştır. Kısmi puanlanan maddeler 0 ve 1 kategorileri 0, 2 kategorisi 1 şeklinde yeniden kodlanarak iki kategoriye dönüştürülmüştür. Ortak maddelerin sırası ve kısmi puanlanan maddelerin kitapçıklardaki dağılımları Tablo 6'de gösterilmiştir.

Tablo 6.

Kitapçık Türlerine Göre Soruların Dağılımı

5.Kitapçık	6.Kitapçık	5.Kitapçık	6.Kitapçık	5.Kitapçık	6.Kitapçık	5.Kitapçık	6.Kitapçık
2	1	20	13	53	28	78*	51
4	4	24*	14	54	30	79	52
5	5	26	16	68	31	80	55
6	6	35	17	69	32	81	56
7*	7*	41	19	73	33	82	57
8*	8*	42	20	74	35*	83	58
9	9	45	24*	75*	39*	84*	105
16	10	46	26	76	40	108	106
17	12	47	27	77	50	109	107

Gri renkli olan sorular ortak maddeleri temsil etmektedir.

* Kısmi puanlanan sorular.

Denk olmayan gruplarda ortak test deseninde Tablo 6’de gösterilen 12 ortak madde kullanılmıştır.

Denk olmayan gruplarda ortak maddelerin kullanılmadığı desende ise bireyler hakkında bilgi edinmek için cinsiyet ve matematik öz yeterlik puanı ortak değişken olarak kullanılmıştır. Ayrıca, bu desende ortak madde puanları da ortak değişken olarak alınıp, elde edilen sonuçlar diğer ortak değişkenlerden elde edilen sonuçlarla karşılaştırılmıştır.

PISA 2012’de, Matematiğe ilişkin duyuşsal özellikleri ölçen maddelerden 10 tane indeks oluşturulmuştur. Öğrenci anketinde yer alan bu indekslerden matematik öz yeterliğini gösteren MATHEFF ortak değişken olarak seçilmiş ve DoGOD deseninde sürekli ortak değişken olarak kullanılmıştır. Bu değişken PISA uygulamalarında matematik başarısını en fazla açıklayan değişken olarak nitelendirilmektedir (Thien & Darmawan, 2016, s.94; Koğar, 2015). Öğrenci anketinde ST37Q01-08 aralığında yer alan maddeler ile temsil edilen değişken, öğrencilerin matematiğe olan özgüvenini göstermektedir. Maddeler 1-4 aralığında puanlanmıştır. Yapılan analizlerde MATHEFF değişkene ait ölçek güvenirliği Kanada verisi için 0,89; italya verisi için 0,902 olarak elde edilmiştir. Bu değişken için öğrencilerin puanı, madde puanlarının toplamından elde edilmiştir. Yüksek puan düşük öz yeterliği göstermektedir.

DoGOD deseni için Parametrik olmayan Bayes modelinde kullanılacak bir diğer ortak değişken ise kategorik olan cinsiyet değişkenidir. Literatürde DoGOD deseni için cinsiyeti ortak değişken olarak ele alan birçok çalışma yer almaktadır (Branberg & Wiberg, 2011; Gonzalez, Barrientos & Quintana, 2015a,b; Gonzalez & Wiberg, 2017; p.171; Liou, Cheng & Li, 2001). Ayrıca birçok çalışmada gruplar arasında farklılıkların ortaya çıkaran bir değişken olduğuna ilişkin bulgular yer almaktadır (Martin, Mullis, Foy & Stanco, 2012; Yıldırım, Yıldırım, Yetişir & Ceylan, 2013).

Verilerin Analizi

Araştırmada MTK kuramına dayalı ölçek dönüştürme yöntemleri ve Parametrik olmayan Bayes modeli için analizler ayrı ayrı sürdürülmüştür. İlk olarak MTK varsayımlarından tek boyutluluk ve yerel bağımsızlık test edilmiştir.

Tek boyutluluk

Kitapçıklarda yer alan matematik maddelerinin tek boyutluluğunu incelemek için faktör 10.3 analiz programı kullanılmıştır. Kitapçıkların tek boyutluluğu 36 madde üzerinden incelenmiştir. 36 madde için tek boyutlu olarak elde edilmesi 24 maddelik yapının da tek boyutlu olduğunun göstergesi olarak ele alınmıştır.

PISA 2012 de 5. Kitapçığı alan İtalya verisi

Faktör analizi sonucunda 5 numaralı kitapçığa ait olarak elde edilen KMO değeri 0,949 olup bu değer örneklemin analiz için yeterli olduğunu ifade etmektedir. Bartlett's istatistiği değeri 7086,6 (df=630; p:0,000) olarak elde edilmiştir. Bu değer de bize bu veri setinin faktörlenebilir olduğunu ifade etmektedir. Temel bileşenler analizi dikkate alınarak faktörlerin öz değer yüklerine bağlı açıkladığı varyanslar Tablo 7 de verilmiştir.

Tablo 7

5.Kitapçığa Ait Açıklanan Varyans

Faktör	Özdeğer	Varyans Yüzdesi	Kümülatif Yüzde
1	8,489	23,5	23,5
2	1.49	4,45	
3	1.125	3,1	
4	1.101	3,0	
5	1.048	2,9	
6	1.014	2,8	
7	1.001	2,7	

Tablo 7'de öz değeri 1 ve üzeri olan 7 faktöre yer verilmiştir. İlk faktörün öz değerinin ikinci faktörün öz değerinin 5 katından fazla olması, ikinci faktörün öz değerinin ise diğer faktörlerin öz değerlerine yakın olması kitapçığın baskın olarak ifade edilecek tek boyuta sahip olduğunu göstermektedir. Paralel analiz sonuçları da testin tek boyutlu olduğunu desteklemektedir. Elde edilen değerlerden ve programın verdiği sonuçtan testin tek boyutlu olduğu ifade edebiliriz.

PISA 2012 de 6. Kitapçığı alan Kanada verisi

Faktör analizi sonucu incelendiğinde; KMO değeri 0,944, Bartlett istatistiği değeri ise 6427, (df=630, p=0,000) olarak elde edilmiştir. Bu kitapçık için de örneklemin yeterli olduğunu ve veri setinin faktörlenebilir olduğunu ifade etmektedir. Temel bileşenler analizi dikkate alınarak faktörlerin öz değer yüklerine bağlı olarak açıkladığı varyanslar Tablo 8’de verilmiştir.

Tablo 8

6.Kitapçığa Ait Açıklanan Varyans

Faktör	Özdeğer	Varyans Yüzdesi	Kümülatif Yüzde
1	7.902	21,9	21,9
2	1.507	4,1	
3	1.188	3,3	
4	1.167	3,2	
5	1.065	2,9	
6	1.03	2,86	
7	1.013	2,81	

Tablo 8’de öz değeri 1 ve üzeri olan 7 faktöre yer verilmiştir. İlk faktörün öz değeri ikinci faktörün öz değerinin 5 katından fazla olması, ikinci faktörün öz değerinin ise diğer faktörlerin öz değerlerine yakın olması kitapçığın tek boyutlu olduğunu göstermektedir. Paralel analizi sonuçları da testin tek boyutlu olduğunu desteklemektedir. Elde edilen değerlerden ve programın verdiği sonuçtan testin tek boyutlu olduğu ifade edebiliriz.

Yerel Bağımsızlık

Bir testteki maddeler evrenin tamamı için istatistiksel olarak bağımlı ise test tek boyutludur. Tek bir boyut varsa, gizil yetenek bakımından homojen olan her bir alt evrende maddeler istatistiksel olarak bağımsızdır. Bu bağımsızlık gizil yetenek ölçeğindeki belli bir yetenek düzeyinde bulunan katılımcılardan oluşan bir alt evren için tanımlanır ve yerel bağımsızlık olarak adlandırılır (Crocker & Algina, 1986). İncelenen kitapçıkların tek boyutlu olması yerel bağımsızlık varsayımının da

sağlandığı hakkında fikir vermektedir. Ancak bu durum, varsayımın tamamen karşılanması için yeterli değildir. Yerel bağımsızlık varsayımını test etmek için veri setinin %27 lik alt ve üst grupları için maddeler arası korelasyon hesaplanmıştır. İki kitapçık için de alt ve üst gruptan ve grubun tamamından elde edilen maddeler arası korelasyon dağılımına ilişkin bilgiler Tablo 9 da yer almaktadır;

Tablo 9

Alt ve Üst Gruplar için maddeler arası korelasyon

	5.Kitapçık		6.Kitapçık	
	Minimum	Maksimum	Minimum	Maksimum
Alt Grup	-0,124	0,125	-0,145	0,196
Üst Grup	-0,133	0,251	-0,180	0,258
Tüm Grup	-0,024	0,402	-0,020	0,388

Tablo 9 a göre alt, üst ve tüm grubun maddeleri arasındaki korelasyon için hesaplanan minimum ve maksimum değerleri 0,5 değerinden düşüktür. Elde edilen değerlerin yerel bağımsızlık için kabul edilebilir olduğu ifade edilebilir (Ackerman, 1987). Alt ve üst gruplardaki korelasyonun, toplam gruptaki korelasyondan düşük olması da yerel bağımsızlığını destekler nitelikte olduğu ifade edilebilir.

Parametre Kestirimi

Çalışmamızda ölçeklemesi yapılacak olan 2 test formu paralel olup, bu formlardan elde edilen parametreler farklı bireyler üzerinde kestirildiğinden ve grupların ortalama ile standart sapmaları farklı olduğundan ayrı kalibrasyon yöntemi ile kestirimler yapılmıştır.

DoGOM deseni için eşitleme yetenek parametreleri üzerinden yapılmaktadır. Modellerden hangisinin veri seti ile uyum gösterdiğini incelemek için -2loglikelihood değerleri karşılaştırılmıştır. Bu iki modele ait değerler Tablo 10 de yer almaktadır.

Tablo 10

Model Veri Uyumunun İncelenmesi

	5. Kitapçık		6. Kitapçık	
	2 PLM	3 PLM	2 PLM	3 PLM
-2loglikelihood	31517,55	31460,59	34909,96	34838,85
Parametre Sayısı	72	108	72	108
Fark	56,95		71,11	

2 PLM ve 3 PLM için elde edilen olabilirlik değerlerinin farkı alınarak ki-kare istatistiği ile test edilmiştir. Elde edilen değerler, 36 serbestlik derecesindeki X^2 kritik değerinden ($p=0,01$) büyük olduğundan 3 PLM model anlamlı bulunmuştur. Parametreler 3 PLM yöntemine göre kestirilmiştir.

Madde parametrelerinin kestirimi için Parscale 4.1 programından yararlanılmıştır. Kalibre aşamasında Bayes modellerini temel alan modellerden Expected A Posteriori (EAP) yöntemi kullanılmıştır.

Ölçek Dönüşümü

DoGOD deseninde ortak değişkenler ortak madde yerine kullanılarak 24 madde üzerinden analizleri gerçekleştirilecektir. DoGOM deseninde de ortak maddeler, DoGOD deseni ile karşılaştırmayı sağlayabilmek için, dış ortak madde olarak alınmıştır. Parscale programından elde edilen parametreleri aynı ölçeğe getirmek için IRTEQ programı kullanılmıştır. Araştırmada 6. kitapçık hedef test olarak belirlenmiştir. 5. kitapçık temel test olarak alınmış ve gerçek puan hesaplanmıştır.

Parametrik olmayan Bayes (PoB) yaklaşımına göre Test Eşitleme

PoB yöntemi kullanılarak yapılan eşitleme çalışmaları ile eski ve yeni test puanları arasında kurulabilecek ilişki ortak değişkenlerin sürece katılması ile şekillenmektedir. Modelde yer alan parametrelerin (k, γ, w) kestirimlerinde uygun sonuçlar elde edebilmek için MCMC yöntemi kullanılmıştır. MCMC örnekleme yönteminde amacımız, doğru istatistiksel kestirimler yapabilmek için evreni temsil edecek örneklem elde etmektir (StataCorp, 2015, 78; Kruschke, 2015,s.178; 287).

Ortak deęişkenler yardımı ile gemiő bilgilerden yararlanılarak rneklem hakkında genel bilgi edinilmektedir. MCMC rnekleme sreci ile hazırlanan dosyalarda BBPS modeli kullanılarak veriye uygun parametreler ve ortak deęişkenler birleőtirilmektedir. Kanada ve İtalya veri setleri iin ayrı ayrı MCMC sreleri yrtlmőtr. Ortak deęişkenlerin kategorilerine gre ortak profillerde elde edilen daęılımların aynı yzdelik dilime yerleőtirilmesi ile eőtikleme sreci tamamlanmıştır.

Araőtırmada kesikli olarak *cinsiyet*, srekli olarak *MATHEFF* (Matematik z yeterlik Dzeyi) ortak deęişkenleri kullanılmıştır. MATHEFF deęişkeninin lek puanları 8-32 aralığında deęişmektedir. Ancak modelde puanlar birim aralıkta alındığından, lek puanları bir birimdeki deęişimi gsterecek őekilde 0-1 aralığında ele alınmış ve 100 puan zerinden yorumlanmıştır. PoB yntemi ile bireylerin aldığı puanlara gre profil oluőturulmuş ve eőtilenmiş puanlar elde edilmiştir.

Srekli olan MATHEFF deęişkeninin analizi sonucunda  boyutlu grafikler elde edilmiştir.  boyutlu grafikleri yorumlamaya yardımcı olması amacıyla MATHEFF deęişkeni nce kategorik olarak ele alınmış ve her kategoriye temsil eden bireylere ait daęılımlar ve eőtilenmiş puan grafikleri oluőturulmuőtr; daha sonraki alt problemlerde srekli deęişken iin  boyutlu grafiklere yer verilmiştir. Kategoriler iin kesme puanları MATHEFF deęişkeninin  boyutlu grafiklerinde daęılımların deęişimleri dikkate alınarak, araőtırmanın baőtlangıcında; 0-35 puan aralığı “Matematik z-yeterlięi Yksek”, 35-65 puan aralığı “Matematik z-Yeterlięi Orta” ve 65-100 puan aralığı “Matematik z-yeterlięi Dők” őeklinde belirlenmiştir. Kategori dzeyleri iin verilmiş olan daęılımlar bu aralığı temsil eden bireyler iin elde edilmiştir.

Araőtırmada kullanılan bu iki ortak deęişken bilgi vermeyen nseller olarak modele eklenmiştir. Bu durum, bu deęişkenlerden kaynaklanabilecek yanlılıęın puanların sonsal daęılımlarını etkilemesini engellemekte ve daha objektif olarak deęerlendirme yapılabilmesini saęlamaktadır.

Oluőturulan dosyalarda uygun parametreleri denemek zere iterasyon sayısı ilk olarak 5000 alınmıştır. Daha sonraki srete MCMC sayısı 150000 alınarak analizler yrtlmőtr. Bu dosyaların her biri iin daha dz daęılımlar elde etmek amacıyla 10 ar tekrar kullanılmıştır. Modele en uygun olarak elde edilen parametre deęerleri Ek- 3 de (cinsiyet ve MATHEFF’in ayrı ayrı ve birlikte ele alındığı durumlar

için) yer almaktadır. Test formlarının analizleri eş zamanlı olarak yürütülmüş ve analizler her bir dosya için yaklaşık olarak 10 saat 23 dakika sürmüştür. Eşitlenmiş puanları elde etmek için yürütülen analizler ise yaklaşık 7 gün 6 saat sonunda tamamlanmıştır.

MCMC dosyalarında model ile değişkenleri birbirlerine bağlayarak elde edilen sonsal dağılımını açıklamak için kullanılan Gibbs ve Metropolis-Hastings örnekleme yöntemindeki algoritma aşağıdaki gibidir;

Başlangıç bir $v^* \sim p(v|v^{(i)})$ değeri elde edilir ve Metropolis-Hastings oranı kullanılarak ilk verilen değer makul bir değer ise kabul edilir veya red edilerek en uygun değer elde edilene kadar süreç devam eder, (araştırmada 10 tane v değeri bulunmaktadır)

Başlangıç bir $\gamma^* \sim p(\gamma|\gamma^{(i)})$ değeri elde edilir ve Metropolis-Hastings oranı kullanılarak ilk verilen değer makul bir değer ise kabul edilir veya red edilerek en uygun değer elde edilene kadar süreç devam eder, (araştırmada 20 tane γ değeri bulunmaktadır)

Başlangıç bir $k^* \sim p(k|k^{(i)})$ değeri elde edilir ve Metropolis-Hastings oranı kullanılarak ilk verilen değer makul bir değer ise kabul edilir veya red edilerek en uygun değer elde edilene kadar süreç devam eder, (araştırmada 1 tane k değeri bulunmaktadır)

Bu aşamadan sonra ise ortak değişkenler ile birbirlerine bağlanmış olan test puanlarının kümülatif dağılımlardan yararlanılarak puan dönüşümleri ile eşitlenmiş puanlar elde edilmiştir.

BBPS modeli, sürekli dağılım fonksiyonlarını birim aralıklarda tanımlamaktadır. Bu yüzden puanların kestirimi; Gonzaalez, Barrientos ve Quintana (2015b) nın da çalışmasında yaptığı gibi birim aralıklarda elde edilmiştir. Ayrıca çalışmada, eşitlemeden sonra puanlar en yüksek puan 100 olacak şekilde dönüştürülmüştür. Aralıkları farklı olan puanlar arasında yapılan eşitleme çalışmalarında puanların yüzdeliklerinin alınması en iyi yöntem olarak kabul edilmektedir (Livingston, 2004). Dolayısı ile dağılımlar için kullanılan sürekli değişkenler birim aralıklarda değerlendirilmiş, eşitlenmiş puanlar için elde edilen grafikler ve dağılımlar yüzölçümüne dönüştürülerek yorumlanmıştır.

Puan dağılımları kestirilirken %95 güven aralıkları kullanılmış ve bu aralıklar grafiklerde gri renk ile gösterilmiştir. Güven aralığı, bir p olasılık değeri ve nokta kestiriminden daha fazla bilgi taşımaktadır. Klasik yöntemlerde anlamlılık testi için kullanılan p değeri güven aralıklarının oluşmasında dezavantajlı bir durum ortaya koymaktadır. Hatta ki-kare gibi sıfır hipotezi için sadece bir tane p değeri üreten testlerde güven aralıkları da yer almamaktadır. Bayes yöntemi ile üretilen sonsal puan dağılımları, ortalama ve %95 *en yüksek güven aralığı (highest density interval / HDI)* gibi değerlerle özetlenebilir. Bu aralığın içinde yer alan değerler, dışına düşen noktalardan daha kesin değerler almaktadır (Kruschke, 2010b). Çalışmada kestirilen puan dağılımları ile birlikte güven aralıklarına da yer verilmiştir. Güven aralıklarının hesaplanmasında çeyrek dilimler (0.025, 0.5 ve 0.975) kullanılmıştır. Puan dağılımları, güven aralıklarına göre yorumlanmıştır.

Parametrik olmayan Bayes modeli için, Gonzalez, Barrientos ve Quintana (2015a; 2015b)'ın çalışmalarında kullanmış olduğu formüllerden yararlanılarak R 3.2.1 programında kodlar oluşturularak analizler gerçekleştirilmiştir.

Karşılaştırma Kriteri

Geleneksel eşitleme yöntemlerinde, parametre kestirimlerinin hatasını değerlendirmesi için standart olarak RMSE (root mean square error), MSE (mean square error), yanlılık, standart hatalar (SE) gibi ölçütler kullanılmaktadır. Ancak MTK ve PoB gibi farklı modelleri temel alan yöntemlerden elde edilen sonuçları karşılaştırmak oldukça zordur (Wiberg & Gonzalez, 2016). Bu nedenle bu çalışmada iki yöntemle elde edilen sonuçlar RMSE, MSE gibi ölçütler kullanılarak karşılaştırma yapmak mümkün olamamıştır. Çalışmamızda, MTK yöntemleri ile Parametrik olmayan Bayes Yöntemini karşılaştırmak için istatistiksel bilgi veren ve eşitlenmiş puanlara ait dağılımların hedef teste olan uzaklıklarını inceleyen **Hellinger Uzaklığı** kullanılmıştır. Bu uzaklık, dağılımların birbirine noktasal uzaklıklarının toplamıdır. Hellinger uzaklığının birçok formu bulunmaktadır. Araştırmamıza uygun olarak; f ve g gibi herhangi iki dağılım arasındaki uzaklık için kullanılan Hellinger Uzaklığı formülü;

$$\hat{H}(f, g) = \left[\frac{1}{2} \int \left(\sqrt{\hat{f}(x)} - \sqrt{\hat{g}(x)} \right)^2 dx \right]^{1/2}$$

$$\approx \left[\frac{1}{2} \sum_{l=1}^k \left(\sqrt{\hat{f}(x)} - \sqrt{\hat{g}(x)} \right)^2 (x_l - x_{l-1}) \right]^{\frac{1}{2}}$$

şeklindedir (Boone, Merrick & Krachey, 2012). Elde edilen eşitlenmiş puanlar için oluşturulan dağılımların her birinin hedef testteki dağılıma olan uzaklığını gösteren grafiklere yer verilmiştir.

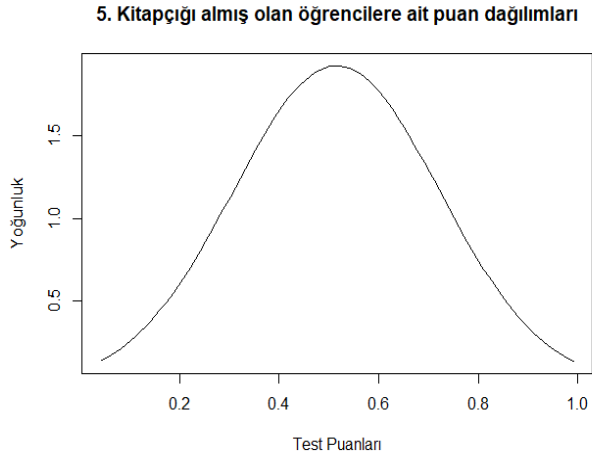


Bölüm 4

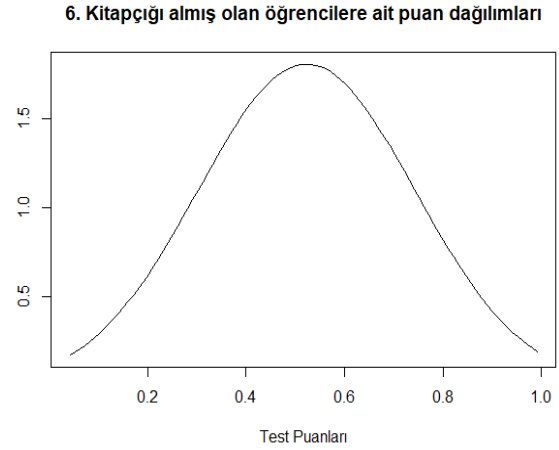
Bulgular ve Yorumlar

Bu bölümde, alt problemlere göre verilmiş araştırma bulguları ve bu bulgulara yönelik yorumlar yer almaktadır.

Alt problemlere ilişkin bulgulara geçmeden önce PISA 2012 için kitapçıkları alan ülkelerin betimsel istatistiklerine ait dağılımlar yer verilmiştir.



Şekil 3. 5. kitapçığı alan İtalya örnekleminin puan dağılımı.



Şekil 4. 6. kitapçığı alan Kanada örnekleminin puan dağılımı.

PISA 2012 uygulamasında 5. kitapçığı alan İtalya verisine ait 908 öğrencinin puanlarının ortalaması 51,514 standart sapması ise 20,72'dir. 6. Kitapçığı alan Kanada verisine ait 931 öğrencinin puanlarının ortalaması 52,27, standart sapması ise 22,06'dır. Grafikte görüldüğü gibi kitapçıklardaki bireylerin almış olduğu puanlar orta düzey puanlarda yoğunlaşmaktadır ve dağılımlar benzerlik göstermektedir.

Alt Problem 1'e İlişkin Bulgular ve Yorumlar

“2012 PISA matematik testi puanlarının DoGOM deseninde MTK eşitleme yöntemlerine göre eşitlenmesi sonucunda elde edilen eşitleme hataları ve gerçek puan dağılımları nasıldır?”

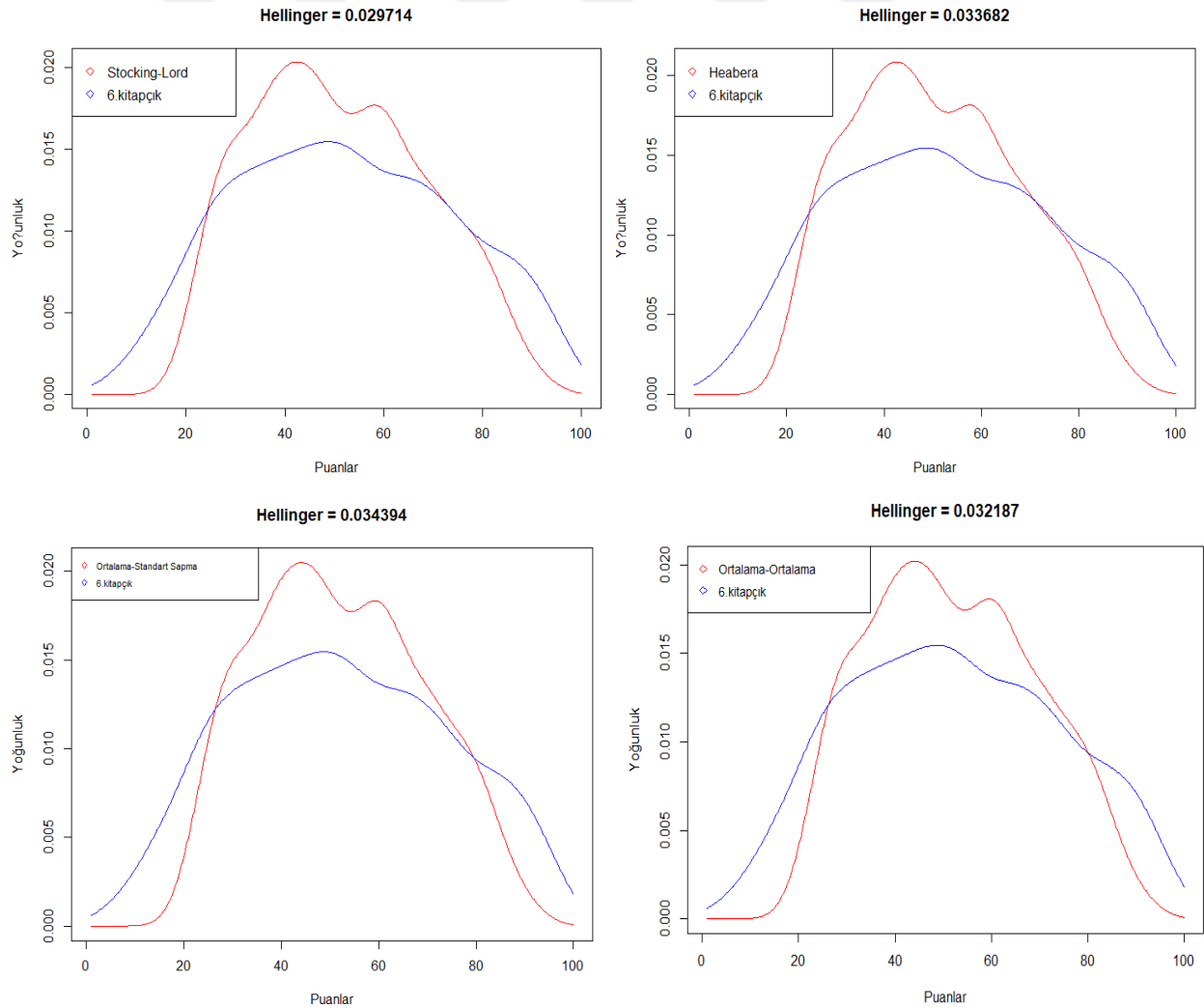
2012 PISA uygulamasında denk olmayan iki grubun almış olduğu iki kitapçık arasında MTK yöntemine göre ölçek dönüştürme yapılması için ortak maddeler kullanılmıştır. MTK ölçek dönüştürme yöntemlerine göre elde edilen eşitleme hatalarını gösteren RMSE değerleri aşağıdaki gibidir:

Tablo 11.

MTK eşitleme yöntemlerine göre elde edilen RMSE değerleri

Ortalama - Ortalama	Ortalama-Standart Sapma	Stocking-Lord	Heabera
0,149	0,132	0,201	0,180

Dış ortak maddeler ile MTK modeline göre yapılan ölçek dönüşümü sonucunda, en düşük hata ortalama-standart sapma yönteminden, en yüksek hata Stocking-Lord yönteminden elde edilmiştir. Ölçek dönüştürme metotları sonucunda elde edilen yeni yetenek parametreleri ve hedef kitapçığın madde parametreleri kullanılarak gerçek puanlar elde edilmiştir. Her bir yöntem ve hedef teste ait olasılık yoğunluk dağılımları Hellinger uzaklıkları ile incelenmiştir.



Şekil 5. MTK yöntemlerinden elde edilen puanların dağılımı ve hedef teste olan uzaklıkları.

Kestirilen puanların olasılık yoğunluk değerleri incelendiğinde, gerçek puanlara ait dağılımların benzerlik gösterdiği ve Hellinger uzaklığına göre hedef test dağılımına ise yaklaşık olarak benzer uzaklıklarda oldukları gözlemlenmektedir. En düşük RMSE değerini ortalama-standart sapma yöntemi vermiş olmasına rağmen, gerçekte puanlar elde edildiğinde karakteristik eğri yöntemlerinden elde edilen kestirimlerin hedef testin dağılımına daha yakın olduğu görülmektedir. Hellinger uzaklığına göre en yakın uzaklığı veren dağılım 0,029714 birim ile Stocking-Lord yöntemidir. Bu durum Stocking-Lord yönteminin diğer MTK yöntemlerine göre daha hatalı kestirim yapmış olsa dahi gerçek puan olarak hedef teste daha yakın değerler ürettiği şeklinde ifade edilebilir. Li, Jiang ve von Davier de (2012) araştırmasında MTK gerçek puan eşitleme ile elde edilen puanların daha doğru ve kesin olduğunu vurgulamaktadır.

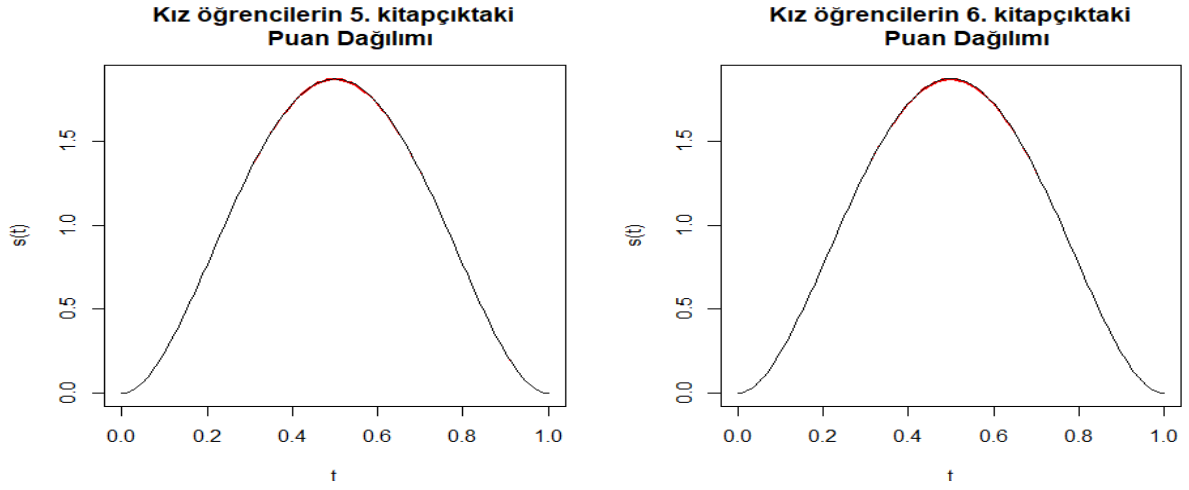
Alt Problem 2'ye İlişkin Bulgular ve Yorumlar

“PISA 2012 yılı matematik alt testi puanlarını PoB modeli ile eşitlemek için öğrencilerin cinsiyeti ortak değişken olarak alındığında;

2.1 Cinsiyet değişkeninin kategorilerine göre elde edilen eşitlenmiş puanlar arasında nasıl bir ilişki bulunmaktadır?”

Denk olmayan iki grup için cinsiyet ortak değişken alınmış ve bireylerin puanları bu değişken ile bağlanmıştır. Her bir birey için cinsiyet değişkenine göre evreni en uygun olarak yansıtacak şekilde puan dağılımları elde edilmiş ve bu dağılımlar aracılığı ile PoB modeline göre eşitlenmiş puanlar bulunmuştur.

Cinsiyet ortak değişken olarak modele alındığında 5. ve 6. kitapçıkları almış kız ve erkek öğrencilerin dağılımları incelenerek dağılımları hakkında bilgi edinilebildiği veya karşılaştırmalar yapılabildiği gibi ortak değişkenin her bir alt kategorisine ait ayrı ayrı eşitleme de yapılabilmektedir. Ayrıca, ortak değişkene ait alt kategorilerin gruplarda benzer dağılım gösterdiği varsayımı altında kitapçıklar arasındaki farklılık incelenebilmektedir. Kız öğrencilerin kitapçıklarda gösterdikleri dağılımlar Şekil 6'daki gibidir.

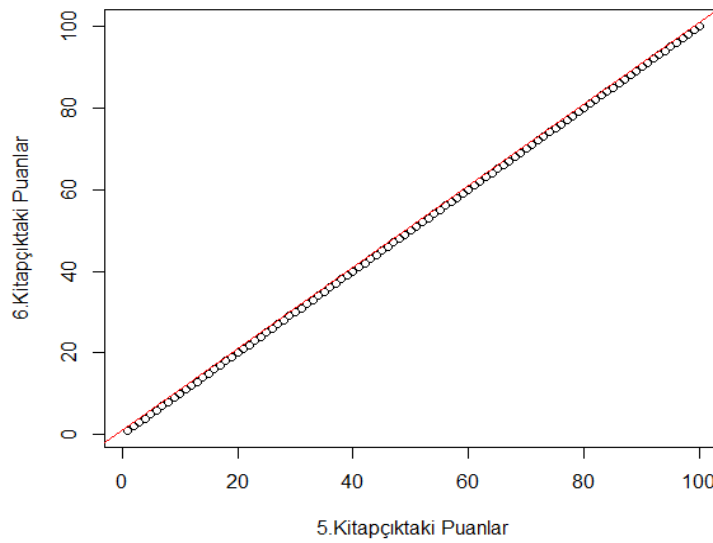


***Güven aralığı oldukça dar olduğu için kırmızı renk ile gösterilmiştir.

Şekil 6. Kız öğrencilerin kitapçıklara göre puan dağılımı ve güven aralığı.

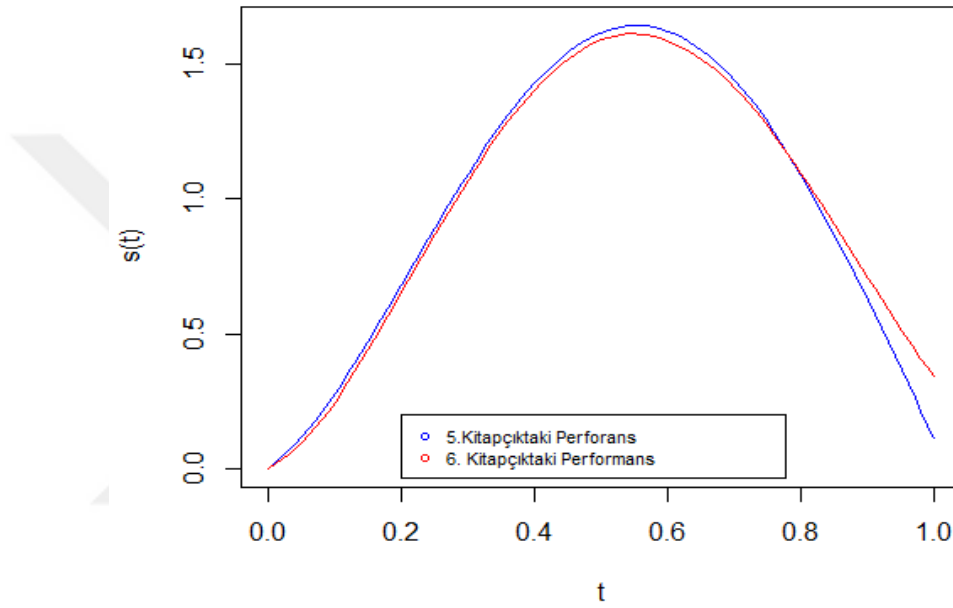
İki kitapçıkta da kız öğrencilerin gösterdikleri dağılımlar ve güven aralıkları benzerlik göstermektedir. 6.kitapçıkta orta düzeyde puan alan kız öğrencilerin yoğunluğu diğer kitapçığa kıyasla nispeten daha fazladır. Cinsiyet değişkeni ile puanların bağlanması sonucu kestirilen puanlar için elde edilen güven aralıkları birçok puan dilimi için dağılım ile örtüşmekte ve en yoğun olarak alınan puanlarda dar bir şekilde görülmektedir. Güven aralıklarının dar olması kız öğrencilere ait puanların kestirimlerinin tutarlı olduğunu göstermektedir.

Kız öğrencilerin 5. Kitapçıktaki puanlarının 6. Kitapçığa eşitlenmesi sonucu elde edilen puan dağılımının grafiği Şekil 7’de yer almaktadır.



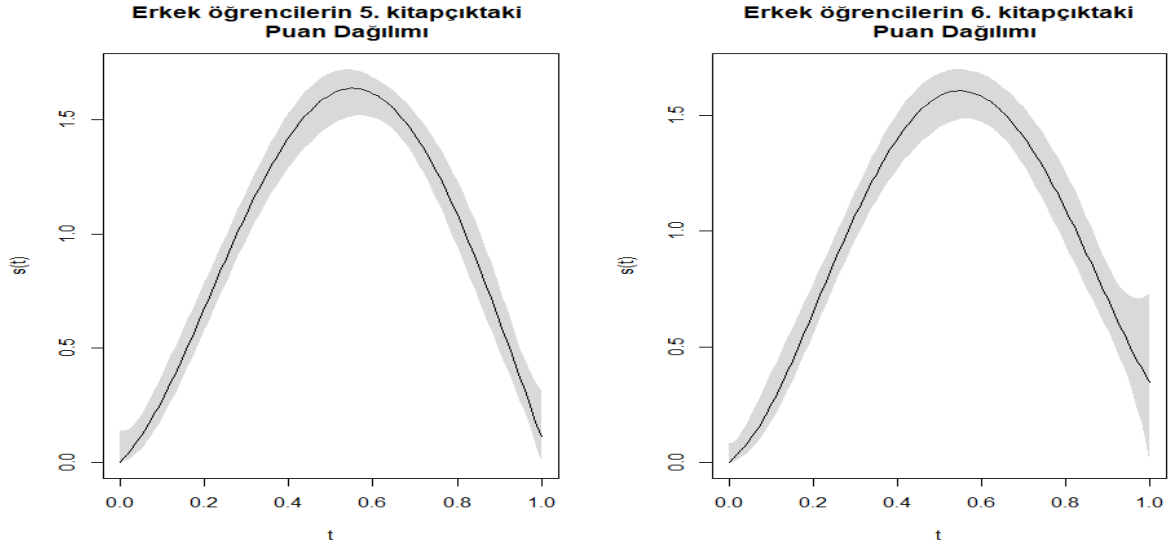
Şekil 7. Kız öğrencilerin 5. kitapçık ve 6. kitapçık puanlarının PoB modeline göre eşitlenmesi.

Cinsiyet değişkeninin bir alt kategorisi olan kızlara ait eşitlenmiş puanları arasında doğrusal bir ilişki görülmektedir. Dağılımlardan da görüldüğü gibi kızların kitapçıklarda göstermiş oldukları performanslar oldukça benzerdir. Bu durum eşitlenmiş puanlarda görülmektedir. Dağılımların yakınlık göstermesi ve eşitlenmiş puanların doğrusal olması, değişkene ait kategorinin kitapçıklar arasındaki farkı net olarak ortaya çıkarmadığı veya kitapçıkların aynı düzeyde oldukları şeklinde ifade edilebilir. Kız öğrenciler için dağılımlar neredeyse örtüşmekte iken erkek öğrencilerin kitapçıklarda gösterdiği dağılımların farklılıkları daha belirgindir.



Şekil 8. Erkek öğrencilerin kitapçıklarda göstermiş oldukları puan dağılımı.

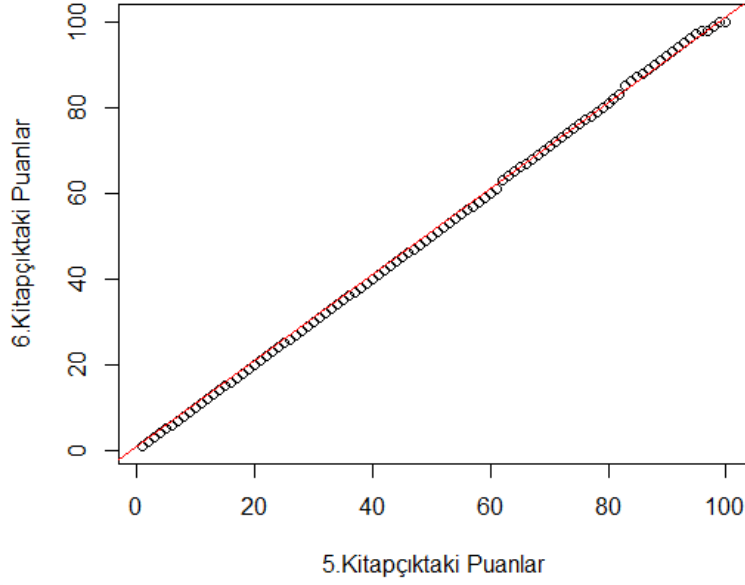
Erkek öğrencilerin kitapçıklardaki dağılımlar incelendiğinde; yaklaşık olarak 77 puana kadar puanlardaki yoğunluğun 5. kitapçıkta, bu puandan sonraki puanlarda ise 6. kitapçıkta fazla olduğu görülmektedir. DoGOD deseninin en önemli adımı olan ortak kategorilere ait dağılımların her iki grup için aynı olması varsayımı ile erkek öğrenciler için 6. kitapçığın 5. kitapçıktan daha kolay sorular ihtiva ettiğini söylenebilir. Erkek öğrenciler için kestirilen dağılımların kesinliğini incelemek için güven aralıkları incelenmiştir.



Şekil 9. Erkek öğrencilerin kitapçıklara göre puan dağılımı ve güven aralıkları.

5. kitapçık için kestirilen puanların dağılımına ait güven aralığı uç değerlere doğru azalırken, ortalama puanlarda artmaktadır. 6. kitapçık için de benzer bir durumdan bahsetmenin yanı sıra yüksek puanlarda da güven aralığının diğer puan dilimlerindeki aralıklardan nispeten geniş olduğu görülmektedir. Dağılımda geniş olan güven aralıkları, yapılan kestirimlerde kesinliğin azaldığını veya bu noktalardaki kestirimler için belirsizliklerin olduğunu göstermektedir. İki kitapçık için güven aralıklarında karşılaştırma yapıldığında 6. kitapçığındaki aralığın nispeten daha geniş olduğu görülmektedir. Kesinliğin azalması, model için puanların kestirilmesinde kullanılan veride yeterli örneklem olmamasından veya örneklemede bu profile sahip bireylerin geniş ranjda bir puan aralığında değerler almasından kaynaklandığı şeklinde yorumlanabilir.

Erkek öğrencilere ait dağılımlardan yola çıkarak eşitlenmiş puanlar elde edilmiştir.



Şekil 10. Erkek öğrencilerin 5. Kitapçık ve 6. kitapçık puanlarının PoB yöntemine göre eşitlenmesi.

Erkek öğrenciler için elde edilen puan dağılımları arasında eşitleme yapıldığında kız öğrencilerinkine benzer bir grafik ortaya çıkmaktadır. Ancak kızlara ait eşitlenmiş puanlardan çok küçük bir farklılık göstererek, erkek öğrencilere ait yüksek puanlar için 6. kitapçıkta yüksek puan aldıkları görülmektedir. Kitapçıklar arasında doğrusala yakın bir ilişkinin olması erkek öğrencilerin kitapçıklarda benzer performans sergilediklerini göstermektedir. Bu durum cinsiyet değişkenine ait erkek kategorisinin kitapçıklar arasındaki farkı ortaya çıkarmada yetersiz kaldığı veya kitapçıkların oldukça benzer özellikler gösterdiği şeklinde ifade edilebilir.

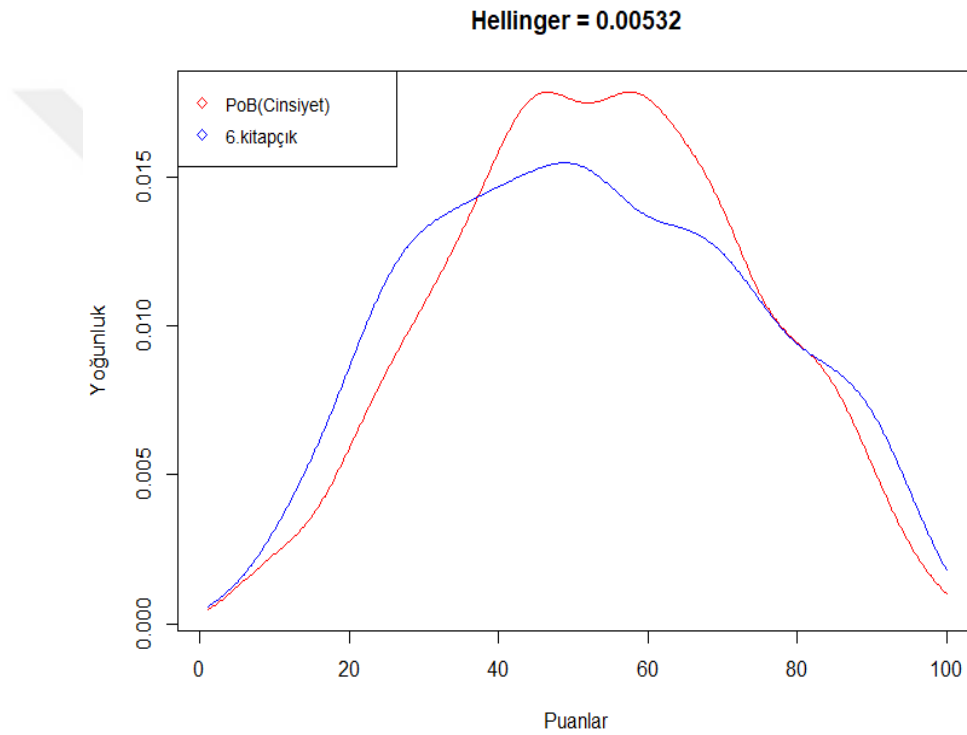
Cinsiyet ortak değişkeninin kategorilerine göre kitapçıklar arasındaki fark incelendiğinde, kızlar için bir fark oluşmadığı, erkekler için ise küçük bir farklılık olduğu görülmektedir. Bu farklılık 6. Kitapçığın daha kolay sorular içerdiği yönündedir. Dolayısı ile kitapçıklar arasındaki farkı açıklamada cinsiyet değişkeninin büyük bir etkisi olmadığı söylenilebilir.

Ortak değişken olarak cinsiyetin kullanıldığı araştırmaları literatürde görmek mümkündür (Branberg & Wiberg, 2011; Gonzalez, Barrientos & Quintana, 2015a,b; Gonzalez & Wiberg, 2017; p.171; Liou, Cheng & Li, 2001). Aynı kitapçığı almış olan kız ve erkek öğrenciler için güven aralıkları farklılık gösterse de dağılımları oldukça benzer olup cinsiyetin matematik performansı üzerinde önemli bir etkisinin

olmadığını göstermektedir. Literatürde bu durumu destekleyen benzer çalışmaların yer aldığını görmek mümkündür (Thien & Darmawan, 2016, s.94; Lindberg, Hyde, Petersen & Linn, 2010; Hall & Hoff).

2.2. Elde edilen eşitlenmiş puanların dağılımının hedef teste ait dağılıma olan uzaklıkları nasıldır?

Her bir öğrencinin cinsiyet ortak değişkeni kullanılarak PoB modeline göre eşitlenmiş puanları elde edilmiştir. Eşitlenmiş puanlara ve hedef testteki puanlara ait dağılımlar birlikte incelenmiştir



Şekil 11. Cinsiyet ortak değişkenine göre PoB yönteminden elde edilen puanların ve hedef testin puanlarının dağılımı.

Her bir bireyin cinsiyet kategorisi ve test puanları birleştirilerek PoB modelinde eşitlenmiş puanlar elde edilmiştir. Eşitlenmiş puanların dağılımı, hedef teste ait puanların dağılımdan daha sivridir. Noktasal olarak her bir puan için Hellinger uzaklığı hesaplandığında iki dağılım arasındaki uzaklık 0,00532 birim olarak elde edilmiştir. Bu uzaklık MTK yöntemlerinden elde edilen uzaklığının yaklaşık olarak beşte biri olup, PoB modeli ile yapılan eşitlemede cinsiyet ortak değişkeni kullanılması ile hedef teste oldukça yakın puanlar kestirildiğini göstermektedir.

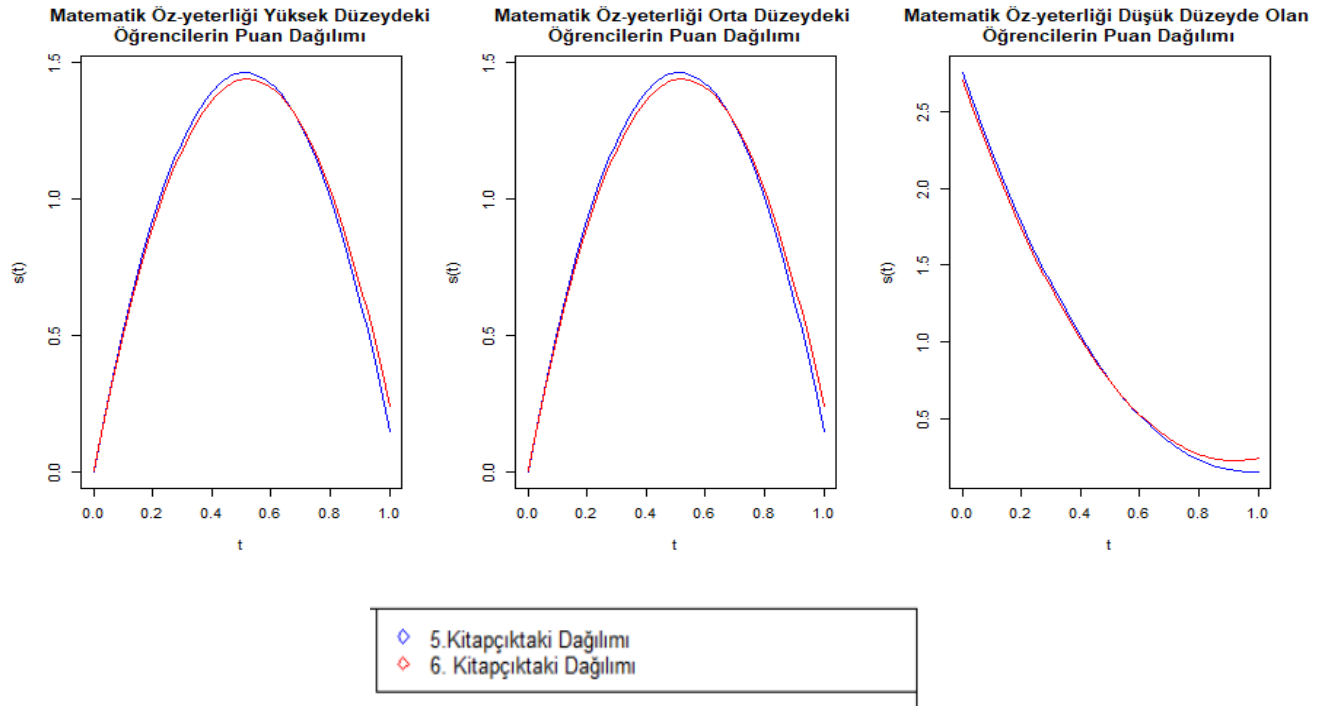
Alt Problem 3'e İlişkin Bulgular ve Yorumlar

PISA 2012 yılı matematik alt testi puanlarını PoB modeli ile eşitlemek için ortak değişken olarak alınan öğrencilerin matematik özyeterlik puanları

3.1. Düşük, orta ve yüksek düzeyi temsil edecek şekilde üç kategorili olarak ele alındığında kategorilere göre elde edilen eşitlenmiş puanların dağılımı nasıldır?

Araştırmanın bir bileşeni olarak ele alınmak istenen diğer bir ortak değişken ise sürekli olarak seçilmiştir. Sürekli değişken puanları MATHEFF değişkenine bireylerin vermiş olduğu cevaplardan elde edilmiştir.

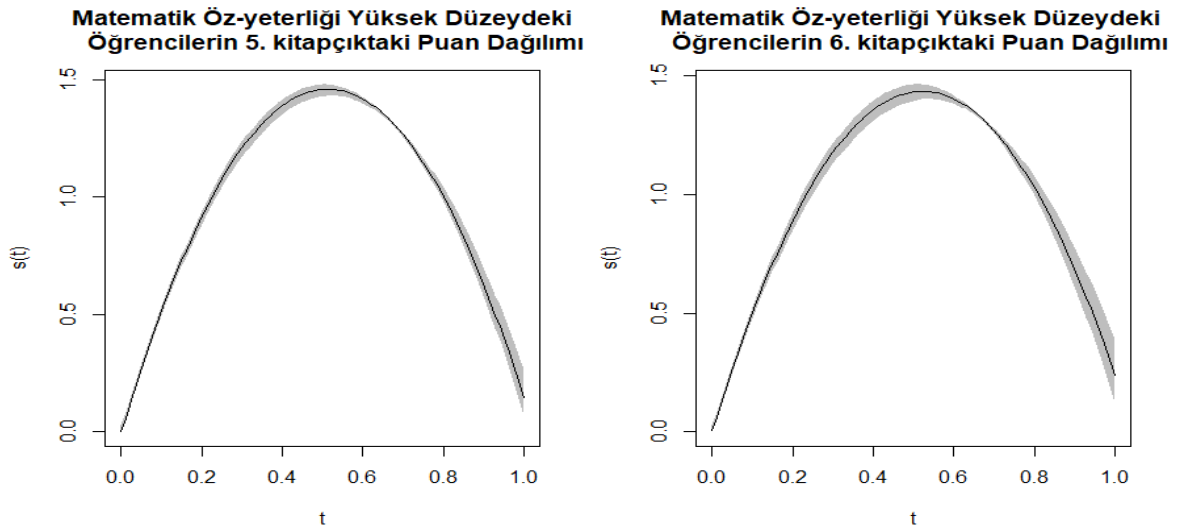
Ortak değişken olarak modelde yer alan MATHEFF değişkeni yorumlamada kolaylık sağlaması ve bir sonraki problemde ele alınan üç boyutlu grafik için çıkarımların yapılabilmesi için önce üç kategori altında incelenmiştir. Matematik öz yeterliğinin yüksek, orta ve düşük olarak adlandırılan her bir kategorisi için puan dağılımları ve eşitlenmiş puanlar elde edilmiştir.



Şekil 12. Matematik öz yeterlik düzeylerine göre öğrencilerin kitapçıklardaki puan dağılımları.

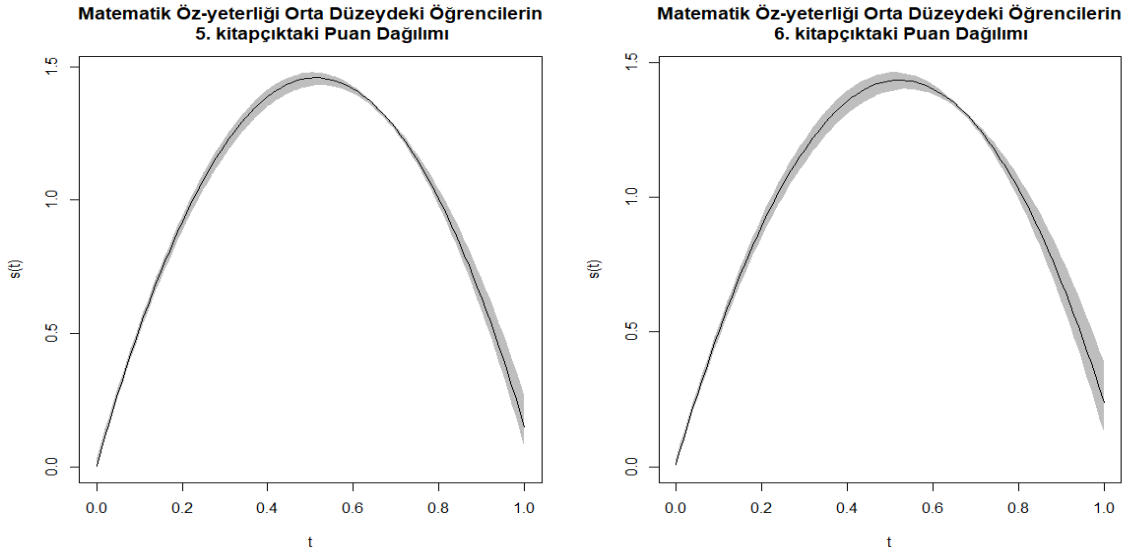
Matematik öz-yeterlik düzeylerine göre bireylerin kitapçıklardan aldıkları puanların dağılımı yukarıda verilmiştir. Matematik öz yeterlik düzeyleri yüksek ve

orta düzeyde olan bireyler için birbirine yakın dağılımlar elde edilmiştir. Bu seviyelere sahip bireyler ortalama puanlarda yoğunluk göstermektedir. Matematik öz yeterlik düzeyi düşük olan bireylerin ise yüksek puan alma olasılıklarının azaldığı görülmektedir. Her düzey için yüksek puan alma olasılığının 6. kitapçıkta daha fazla olduğu görülmektedir. Dolayısı ile bu grafiklerden de, 6. Kitapçıkta daha kolay soruların yer aldığı ifade edilebilir. Kitapçıklara göre elde edilen dağılımlara ait kestirimlerin doğruluğunu görmek üzere güven aralıkları incelenmiştir.



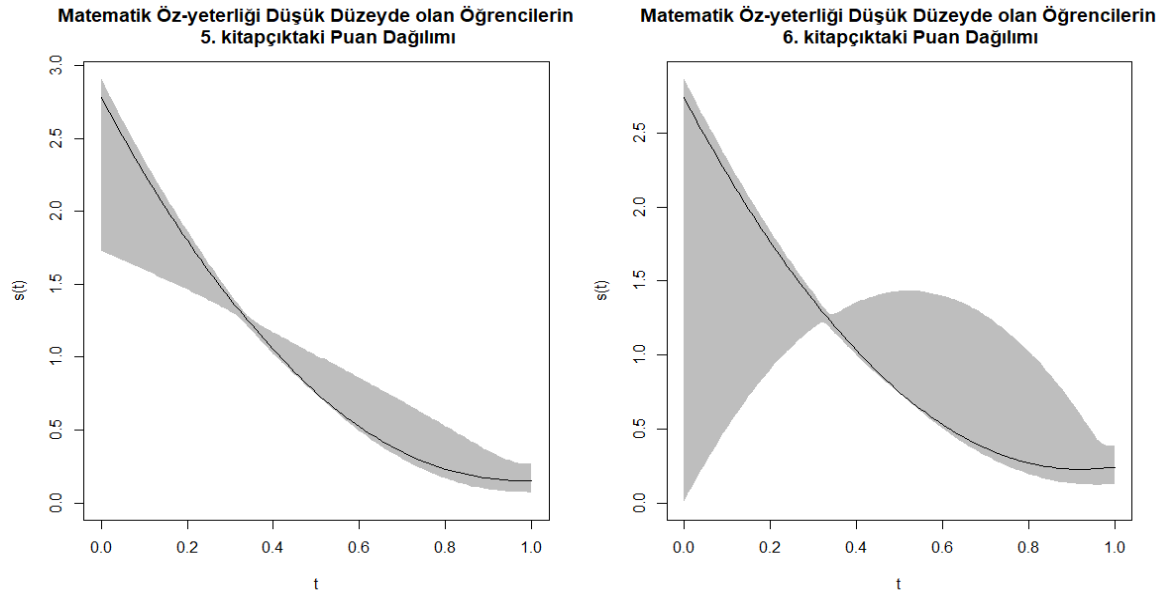
Şekil 13. Yüksek matematik öz yeterlik düzeyine sahip öğrencilerin kitapçıklarda göstermiş oldukları puan dağılımı ve güven aralıkları.

Yüksek matematik öz yeterlik düzeyine sahip bireylerin dağılımlarının kitapçıklardaki güven aralıklarına göre ayrı ayrı ele alındığında, dağılımların büyük bir farklılığa sahip olmadığını, ancak 6.kitapçıkta güven aralığının nispeten daha geniş olduğunu söylemek mümkündür. Bu durumda matematik öz yeterlik düzeyi yüksek olan öğrencilerin kitapçılarda benzer performans sergiledikleri, fakat 6. kitapçıkta kestirimleri için yeterli ön bilginin olmadığı veya aynı düzeydeki bireylerin farklı puanlar almış olması güven aralıklarındaki belirsizliği arttırdığı ifade edilebilir.



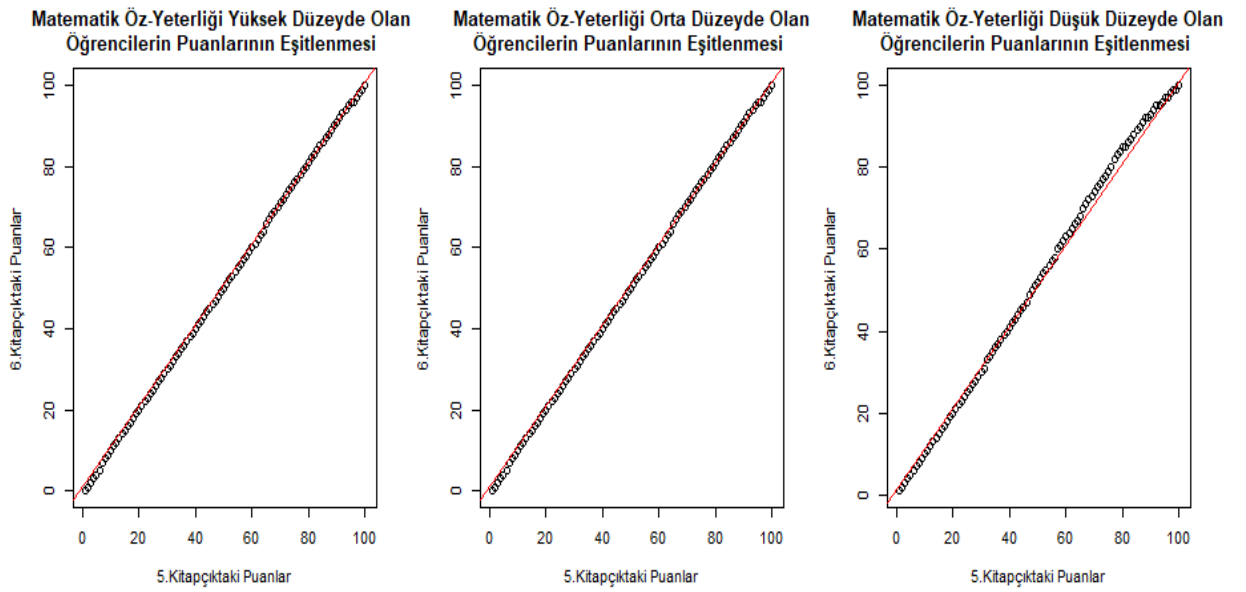
Şekil 14. Orta Düzey matematik öz yeterlik düzeyine sahip öğrencilerin puan dağılımı ve güven aralıkları.

Orta düzeyde matematik öz yeterliğine sahip bireyler için de dağılımlar benzerlik gösterirken, güven aralıklarındaki farklılık biraz daha aşıkardır. Güven aralıklarının 6. kitapçıkta daha arttığı görülmektedir.



Şekil 15. Düşük düzey matematik öz yeterlik düzeyine sahip öğrencilerin kitapçıklarda göstermiş oldukları puan dağılımı ve güven aralıkları.

Düşük öz yeterliğe sahip bireylerin kitapçıklarda göstermiş oldukları dağılım şeklindeki gibidir. Bu dağılımlarda görüldüğü üzere matematik öz yeterlik düzeyinin düşük olması durumunda yüksek puan alma yoğunluğu düşmektedir. Örneklem sürecinde çeyrek dilimler için oluşturulan dağılımların özellikle 6. kitapçıkta farklılaştığı görülmektedir. 6. Kitapçıkta güven aralığının geniş olması, kestirimlerde karışabilecek hatanın 5. kitapçığa oranla daha fazla olduğu göstermektedir. Güven aralıklarının geniş olması, kestirilecek puanları temsil edecek önsel bilgilerin yetersizliğinden veya aynı profildeki bireylerin birbirinden oldukça farklı puanlar almış olmasından kaynaklanabilir.

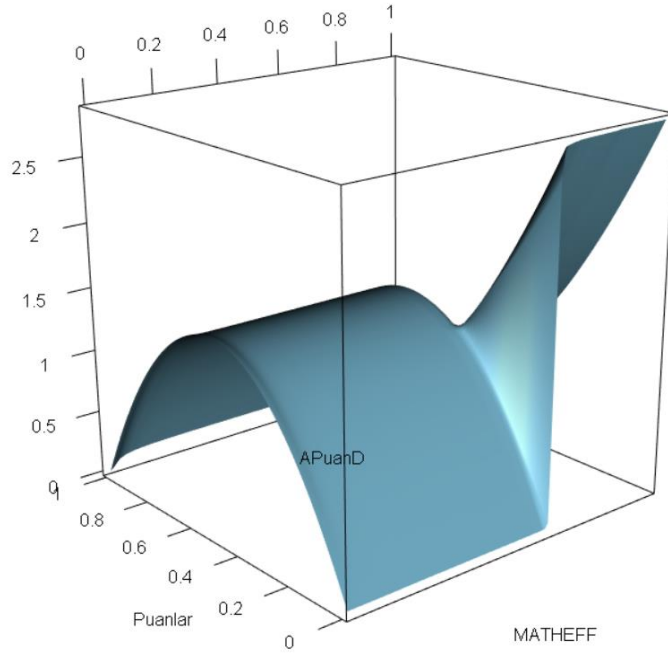


Şekil 16. Farklı matematik öz yeterlik düzeylerinde PoB modeli ile yapılan eşitleme.

Matematik öz yeterlik düzeylerine göre oluşturulan profiller temel alındığında 5. kitapçığın 6. kitapçığa eşitlenmesi ile elde edilen grafikler yer almaktadır. İlk iki profile öğrencilerin düzeylerine göre kitapçıklardan aldıkları puanlar dağılımlarda olduğu gibi benzerlik göstermektedir. MATHEFF in kategorik alındığı durumda dağılımlardan ve eşitlenmiş puanlardan da görülebileceği gibi; değişken, kitapçıklar arasındaki farkı açıklamada yeterli olmamış veya kitapçıklar birbirleri ile benzerlik göstermiştir. Ancak yine de 6. Kitapçığın nispeten daha kolay sorular içerdiği aşıkardır. Özellikle matematik öz yeterliği düşük olan bireyler için kitapçıkların eşitlenmesi ile elde edilen grafikte bu sonucu görmek mümkündür.

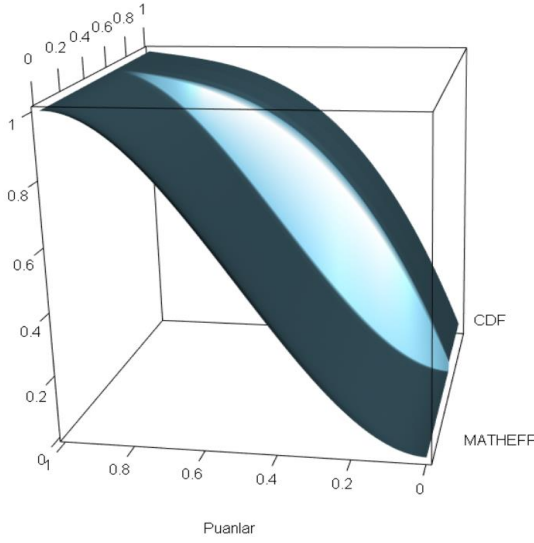
3.2. Sürekli değişken olarak ele alındığında eşitlenmiş puanların dağılımının hedef teste ait dağılıma olan uzaklığı nasıldır?

Sürekli değişkenin her bir puanı için dağılımlar ele alınarak, kitapçıklar için üç boyutlu grafikler elde edilmiştir. Grafiğin boyutlarını; test puanları, sürekli olarak ele aldığımız MATHEFF değişkenine ait puanlar ve olasılık yoğunluk değerleri oluşturmaktadır. Dolayısı ile bu grafikler ile herhangi bir matematik öz yeterlik puanına sahip bireylerin kitapçıklara göre dağılımını ve bu dağılımların hangi noktada değişim gösterdiğini görmek mümkündür.

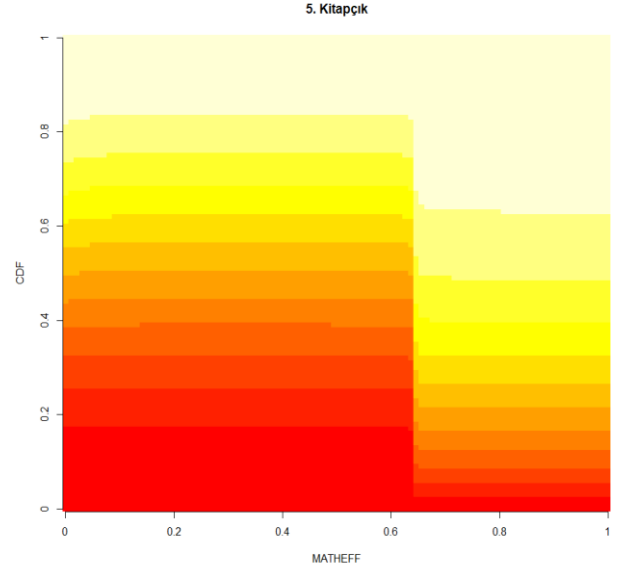


Şekil 17. 5. kitapçıkta MATHEFF düzeylerine göre kestirilen puanların dağılımı.

Şekil17’de 5. kitapçıkta öğrencilerin matematik öz yeterlilik puanlarına göre matematik puan dağılımları gözlemlenmektedir. Birim aralıkta değerlendirilen MATHEFF ortak değişkeninin yüksek değerler alması, yani bireylerin öz yeterlik düzeylerinin azalması durumunda, bireylerin yüksek puan alma yoğunluğunun düştüğü görülmektedir. Eşitlemede kullanılacak olan kümülatif dağılımları da birim aralıktaki her bir değer için Şekil 18 ve 19’daki gibi görmek mümkündür.

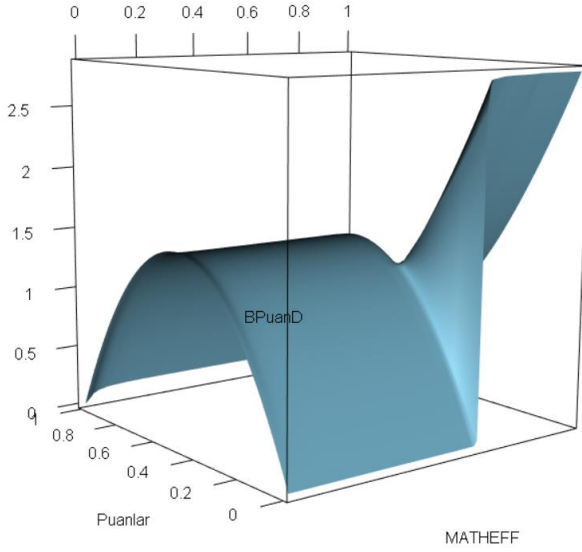


Şeki 18. 5. kitapçıkta MATHEFF düzeylerine göre puanlarına kümülatif dağılımı.

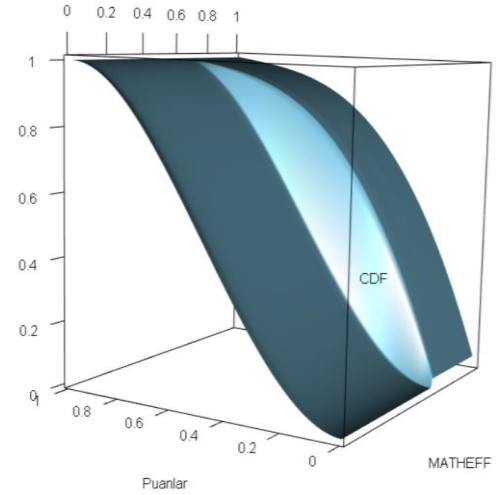


Şeki 19. 5. Kitapçıkta kümülatif dağılımın perspektif görünüşü.

Şekil18 de her bir matematik öz yeterlik düzeyinde sahip bireylerin dağılımlarından yararlanılarak 5. kitapçığa ait kümülatif dağılımlar elde edilmiştir. Elde edilen grafikte MATHEFF puanındaki değişiklik, bireylere ait puanların kümülatif dağılımlarının şeklinde değişikliğe sebep olmaktadır. MATHEFF in yaklaşık olarak 65 ve üzeri puan olması durumunda dağılımlar monoton artan şeklindedir. Dolayısı ile 65 ve üzeri MATHEFF puanına sahip bireylerin yüksek puan alma yoğunluğunun azaldığı ifade edilebilir. Şekil19 da ise bu kümülatif dağılımların perspektif görünüşü ile hangi aralıklarda nasıl bir artış gösterdiğini detaylı bir şekilde görmek mümkündür. Grafikte kırmızı renkler koyuluklarına göre artışta olan yoğunluğu, açık renkler ise artışın azaldığını göstermektedir. Yatay şeritlerin genişliği ise renk düzeyinde grafikleri artış miktarını ifade etmektedir. Bu grafikte de yaklaşık olarak 65 ve üzeri MATHEFF puanlarında dağılımların değişimini ve bu puandan sonra yoğunlukların artış miktarının azaldığı görülmektedir.



Şekil 20. 6. kitapçıkta MATHEFF düzeylerine göre kestirilen puanların dağılımı



Şekil 21. 6 kitapçıkta MATHEFF düzeylerine göre puanlarına kümülatif dağılımı.

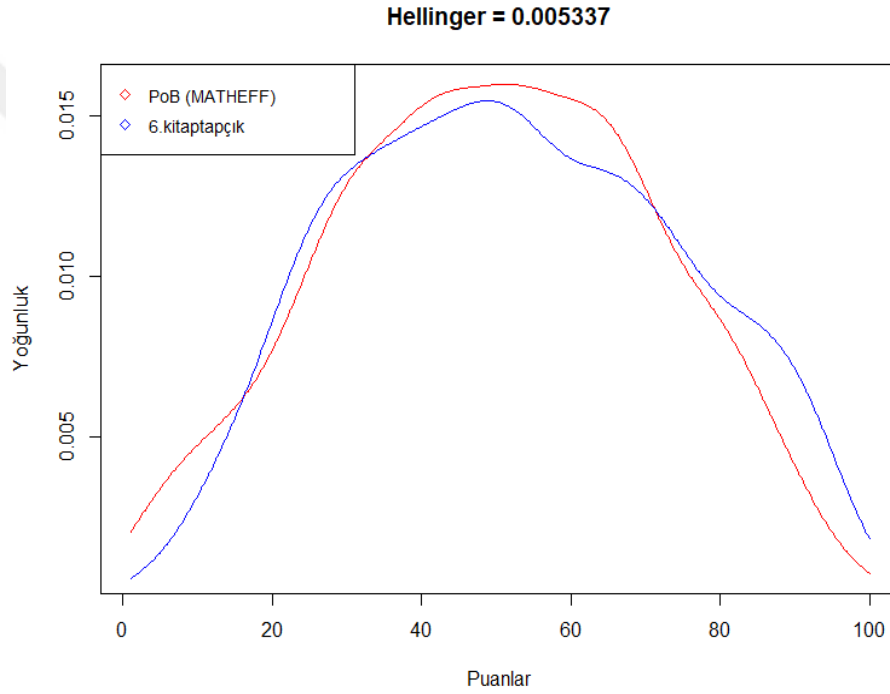
6.kitapçık için elde edilen dağılımların 5. kitapçığa benzer olduğu görülmektedir. Bu kitapçık için de bireylere ait dağılımların hangi puan diliminde değiştiğini ve bu değişimin nasıl bir etki uyandırdığını görmek mümkündür.

Çalışmada her iki kitapçık için, matematik öz yeterlik düzeyinin değişmesi bireylere ait dağılımları değiştirmektedir. Dolayısı ile MATHEFF ortak değişkeninin matematik başarısını açıkladığı sonucuna ulaşılabilir. Literatürde MATHEFF değişkeninin matematik başarısını açıkladığını gösteren çalışmalar yer almaktadır (Thien & Darmawan, 2016, s.94; Ding, 2016,Koçar, 2015). Geleneksel yöntemle yapılan eşitleme çalışmalarında bireylere ait önsel bilgilere yer verilmemesi durumunda her birey için eşitleme dağılımları aynı olarak alınacaktır. Ancak bu çalışma ile MATHEFF puanına göre alt gruplarda değişimin kontrol altına tutulmakta ve bu alt gruplara göre eşitleme yapılmaktadır. MATHEFF değişkeni sürekli olduğundan bireyler farklı düzeylere, dolayısı ile farklı profillere sahip olacaktır. Bu da her bir birey için bir eşitleme grafiği elde etmek anlamına gelmektedir. Çalışmada bu grafikler verilmemiş, her bir birey için eşitlenmiş puanları gösteren dağılım kullanılmıştır.

DoGOD deseninde en önemli varsayım ortak değişkenlerden elde edilen dağılımlara ait kategorilerin alt gruplar için aynı olmasıdır (Wiberg & Branberg,

2015). Bu varsayımdan yararlanılarak kitapçıklar arasındaki farklar gözlenebilmektedir. MATHEFF değişkeninin her iki kitapçıkta da benzer dağılımlar vermiş olması ile kitapçıklar arasındaki farkı tam olarak açıklayamadığını veya kitapçıkların birbirlerine oldukça benzer oldukları söylenebilir. Fakat bu durumda dahi, bu alt problem için elde edilen sonuçlarda 5.kitapçığın, 6.kitapçığa kıyasla zor sorular içerdiği ifade edilebilir.

Her bir öğrencinin matematik öz yeterlik düzeylerine göre eşitenmiş puanları elde edilmiştir. Bu puanlara ve ait hedef teste ait olasılık dağılım birlikte incelenmiştir.



Şekil 22. MATHEFF ortak değişkenine göre PoB yönteminden elde edilen puanların ve hedef testin puanlarının dağılımı.

PoB modelinde MATHEFF'in ortak değişken olarak kullanılması ile elde edilen dağılım, hedef teste ait dağılım ile oldukça yakınlık göstermektedir. Noktasal olarak her bir puan için Hellinger uzaklığı hesaplandığında 0,005337 birim elde edilmiştir. Bu uzaklık MTK yöntemlerinden elde edilen uzaklıklardan oldukça düşüktür. Cinsiyet değişkeninin kullanıldığı modelden elde edilen uzaklık ile yakın değerler vermiştir. Cinsiyete göre elde edilen dağılım ile karşılaştırıldığında,

dağılımların şekilsel olarak farklı noktalarda yaklaştığı ve uzaklaştığı görülmektedir. MATHEFF değişkeninin kullanıldığı modelde dağılım uç değerlere doğru, cinsiyet değişkeninin kullanıldığı modelde ise dağılım ortalama değerlerde hedef testten uzaklaşmaktadır.

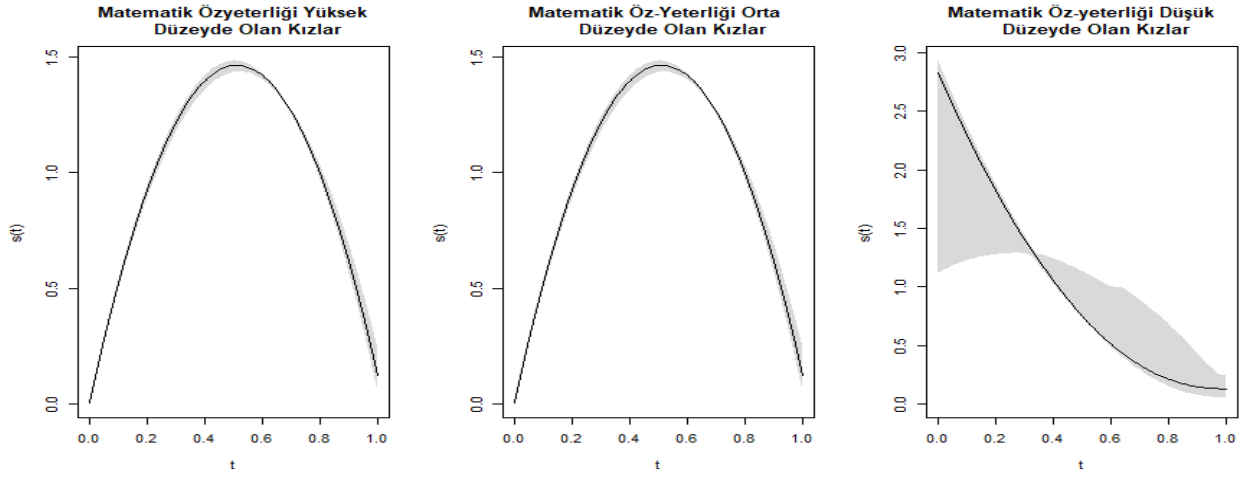
Alt Problem 4'e İlişkin Bulgular ve Yorumlar

PISA 2012 yılı matematik alt testi puanlarını PoB modeli ile eşitlemek için öğrencilerin cinsiyetleri ve matematik özyeterlik puanları birlikte ortak değişken olarak alındığında;

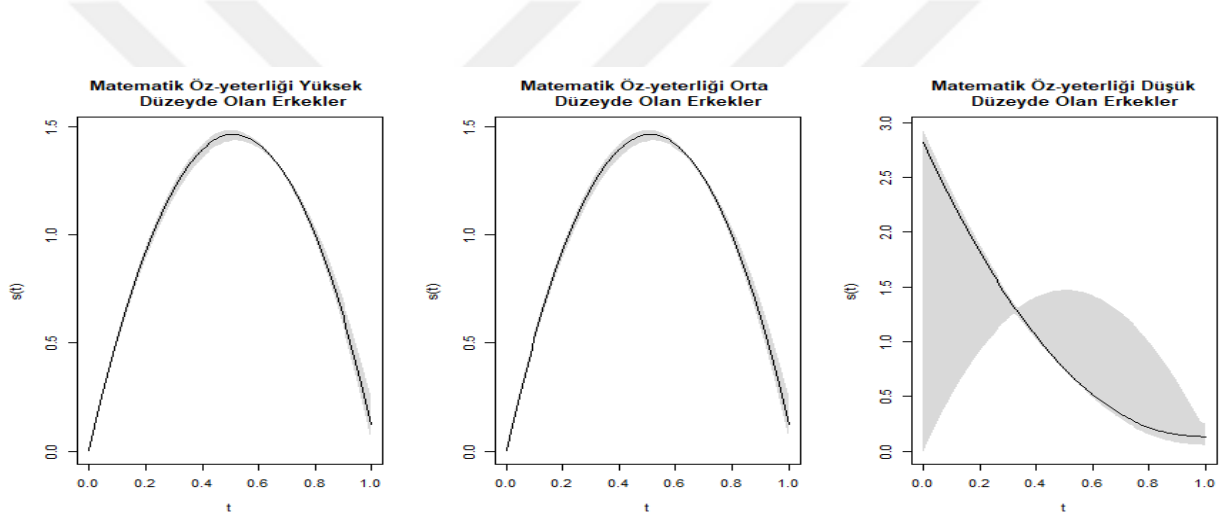
4.1. Üç kategorili (düşük, orta, yüksek) olacak şekilde dönüştürülen MATHEFF değişkeninin kategorilerine göre elde edilen eşitlenmiş puanların dağılımı nasıldır?

MTK ölçekleme yöntemlerinde birden fazla ortak madde kullanılabildiği (Moses, Deng & Zhang, 2011) gibi PoB modeli ile yapılan eşitleme için de birden fazla ortak değişken sisteme dahil edilebilir. 2012 PISA matematik alt bilişsel testi puanlarının DoGOD deseninde cinsiyet kesikli değişkeninin ve MATHEFF sürekli değişkeninin ortak değişken olarak alınması durumu iki kısımda incelenecektir;

Cinsiyet değişkeni ile MATHEFF değişkeni birlikte incelendiğinde, ilk olarak yine yorumlamasında kolaylık olması ve ortak değişkenlerin kategorilerine göre eşitlenmiş puanları daha yakından incelemek amacıyla sürekli değişken üç kategori altında incelenmiştir. İki değişkenin birlikte ele alınması ile profil / kategori sayısı; 2 kategori (cinsiyet) * 3 kategori (Matematik öz-yeterlik düzeyleri) 6 dır. Kız ve erkek öğrencilerin matematik öz yeterlik düzeylerine göre puan dağılımları ve bu puan dağılımlarına ait güven aralıkları ayrı elde edilmiştir. Dağılımlara ait grafikler Şekil 23 ve 24'te yer almaktadır.



Şekil 23. Matematik öz yeterlik düzeylerine göre 5. Kitapçıkta kız öğrencin puan dağılımlar ve güven aralıkları.



Şekil 24. Matematik öz yeterlik düzeylerine göre 5. Kitapçıkta erkek öğrencin puan dağılımları ve güven aralıkları.

Şekil 23 ve 24 da 5. kitapçıkta kız ve erkek öğrencilerin değişkenlerimizin kategorilerine göre göstermiş oldukları dağılımlar ve güven aralıkları yer almaktadır. Şekillerde puan dağılımları benzerlik gösterse de, puan dağılımlarının kestirimlerindeki kesinliği veya belirsizliği gösteren güven aralıklarında cinsiyete göre farklılıklar bulunmaktadır.

Örnekleme işlemi sonucunda düşük düzeyde matematik öz yeterliğe sahip kız öğrencilerin puan dağılımları ve bu dağılımların çeyrek dilimlerindeki dağılımlar göz önüne alınarak oluşturdukları güven aralıkları diğer düzeylerden farklılaşmaktadır. Matematik öz yeterlik düzeyleri yüksek ve orta olması durumunda

güven aralığının dar ve daha kesin kestirimler verdiği; matematik öz yeterlik düzeyi düşük kız öğrenciler için ise puan dağılımlarının kestirimindeki hatanın daha fazla olduğu görülmektedir. Bu düzeydeki öğrencilerin puanlarının kestiriminde belirsizlik olduğunu ifade edebiliriz.

5. kitapçıkta erkek öğrencilerin matematik öz yeterlik düzeylerine göre puan dağılımları kız öğrenciler ile benzerlik göstermektedir. Güven aralıkları da kız öğrencilerde olduğu gibi düşük matematik öz yeterlik düzeyine sahip erkekler için artmaktadır. Ayrıca bu düzeydeki güven aralığının kızlara göre daha fazla olduğu görülmektedir. Güven aralığının geniş olması, kestirilen puanlar için önsel bilginin yetersizliğinden veya aynı profile sahip bireylerin matematik puan ranjlarının geniş olmasından kaynaklanabilir.

5. kitapçıkta matematik öz yeterliği düşük düzeyde olan kız ve erkek öğrencilerin sayısal değerleri Tablo 12’de yer almaktadır.

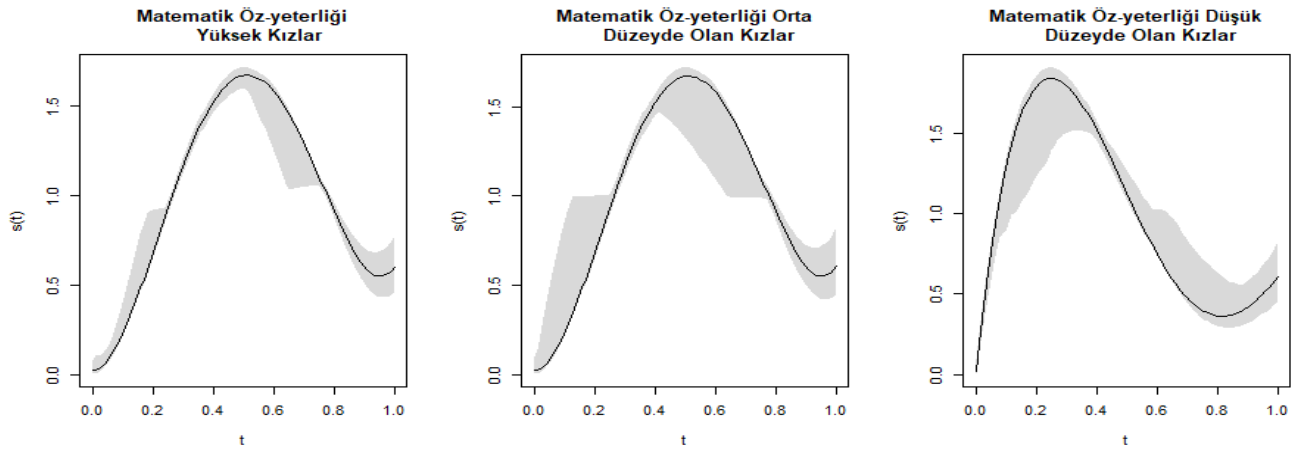
Tablo 12

Matematik Öz Yeterliği Düşük Öğrencilerin Matematik Puanlarının 5. ve 6. Kitapçıkta Frekans Dağılımları

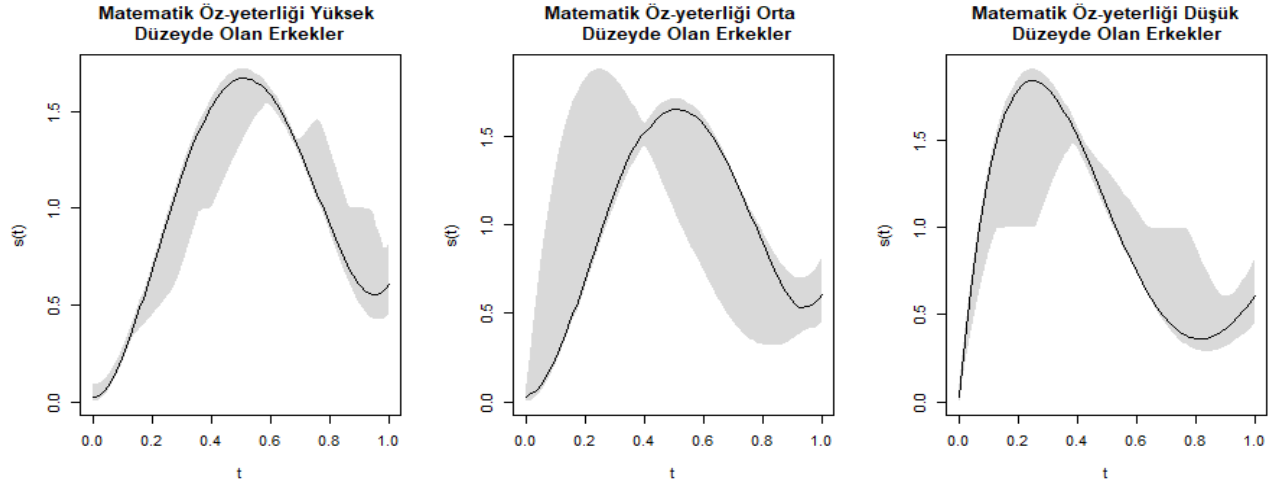
5.Kitapçık				6.Kitapçık			
Erkeklere ait Puan Frekansları		Kızlara ait Puan Frekansları		Erkeklere ait Puan Frekansları		Kızlara ait Puan Frekansları	
Frekans	Puan	Frekans	Puan	Frekans	Puan	Frekans	Puan
1	4,16	1	8,33	1	4.16	2	4.16
2	12,50	2	12,50	1	12.50	1	8.33
1	16,66	1	16,66	2	16.66	1	12.5
1	20,83	2	20,83	2	20.83	5	16.66
1	29,16	1	25	1	25	3	20.83
2	37,50	2	29,16	2	29.16	5	25
1	41,66	1	33,33	3	33.33	2	29.16
1	70,83	3	37,5	4	37.5	4	33.33
		1	45,83	3	41.66	4	37.5
		1	50	4	45.8	2	41.66
		1	54,16	1	54.16	1	75
				1	83.33		

Bu düzeyde 5. kitapçıkta 10'u erkek, 16'sı kız olmak üzere toplam 26 öğrenci bulunmaktadır. Öğrencilerin puanlarına bakıldığında 5. kitapçıkta sadece 1 tane erkek öğrencinin 50 puanın üstünde aldığı, yine sadece 1 tane erkek öğrencinin puanının 12,5 puandan düşük olduğu görülmektedir. Bu düzeyde kız öğrencilerin sayısının daha fazla olması güven aralığını oluşturan dağılımları birbirine nispeten yakın hale getirmiştir. Güven aralıklarındaki genişlik, örnekleme profili temsil eden önsel bilgiye ait öğrenci sayısının azlığından kaynaklanmış olabilir. Aynı şekilde 6. kitapçık için de, matematik öz yeterliği düşük öğrencileri incelediğimizde güven aralıklarının erkeklerde biraz daha fazla olması erkek öğrencilerin sayısının az olmasından kaynaklanmaktadır. Bununla birlikte, aynı MATHEFF puanına sahip kız veya erkeklerin almış oldukları puanların farklılaşması da güven aralıklarının genişlemesine sebep olmuştur.

6. kitapçığı almış öğrencilerin tüm kategorilere göre dağılımı Şekil 25 ve 26'da yer almaktadır.



Şekil 25. Matematik öz yeterlik düzeylerine göre 6. Kitapçıkta kız öğrencilerin puan dağılımları ve güven aralıkları.

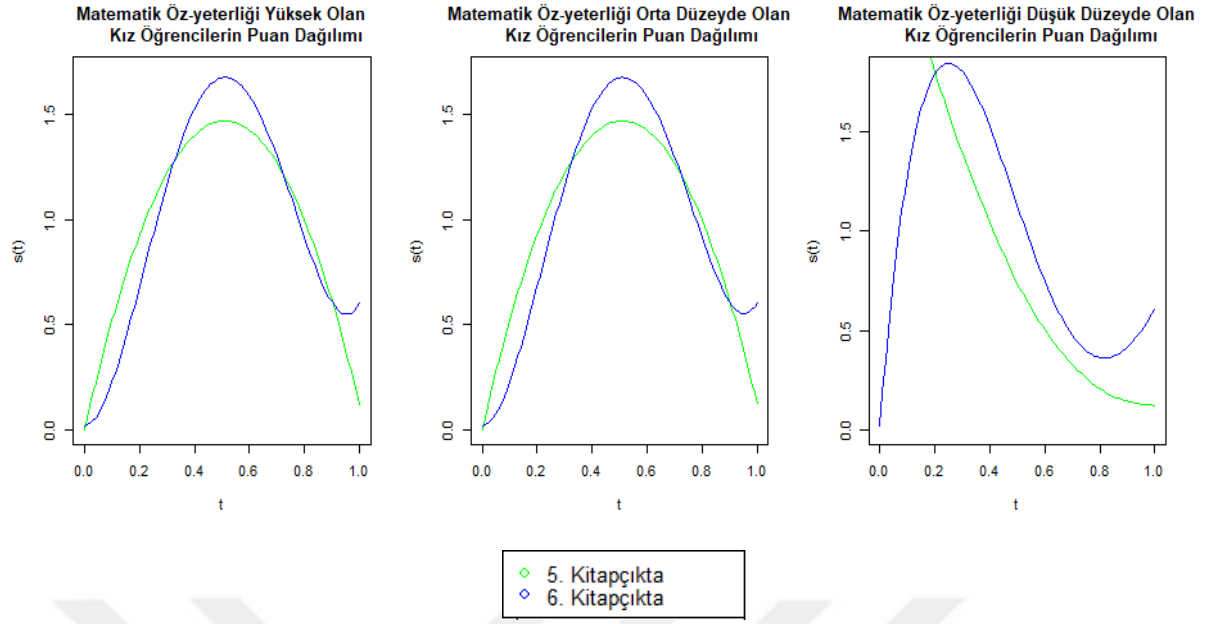


Şekil 26. Matematik öz yeterlik düzeylerine göre 6. Kitapçıkta erkek öğrencilerin puan dağılımları ve güven aralıkları.

Şekil 25 ve 26'da 6. kitapçıkta kız ve erkek öğrencilerin ortak değişkenlerin kategorilerine ait olarak göstermiş oldukları dağılımlar yer almaktadır. Kızlara ve erkeklere göre elde edilen puan dağılımları benzerlik göstermekte olup güven aralıkları farklılaşmaktadır. Erkeklerin dağılımlarından elde edilen güven aralıkları kızların dağılımlarındaki güven aralığından daha geniştir. 5. kitapçığa kıyasla bu kitapçıkta bireylerin matematik öz yeterlik düzeylerinin düşük olması durumundaki dağılımdan farklılaşmakta ve yüksek puan alabilme yoğunluklarının arttığı görülmektedir. Ayrıca modelde sadece MATHEFF değişkeni kullanıldığında kitapçıklardaki dağılım ile karşılaştırma yapıldığında, cinsiyet değişkeninin kitapçıklar arasındaki farkı açıklamıyor olsa da sisteme eklenmesi halinde 6. kitapçıktaki dağılımlarda farklılığa sebep olduğu gözlemlenmektedir. Bu durum cinsiyetin tamamen etkisiz bir değişken olmadığına bir göstergesi olarak ifade edilebilir.

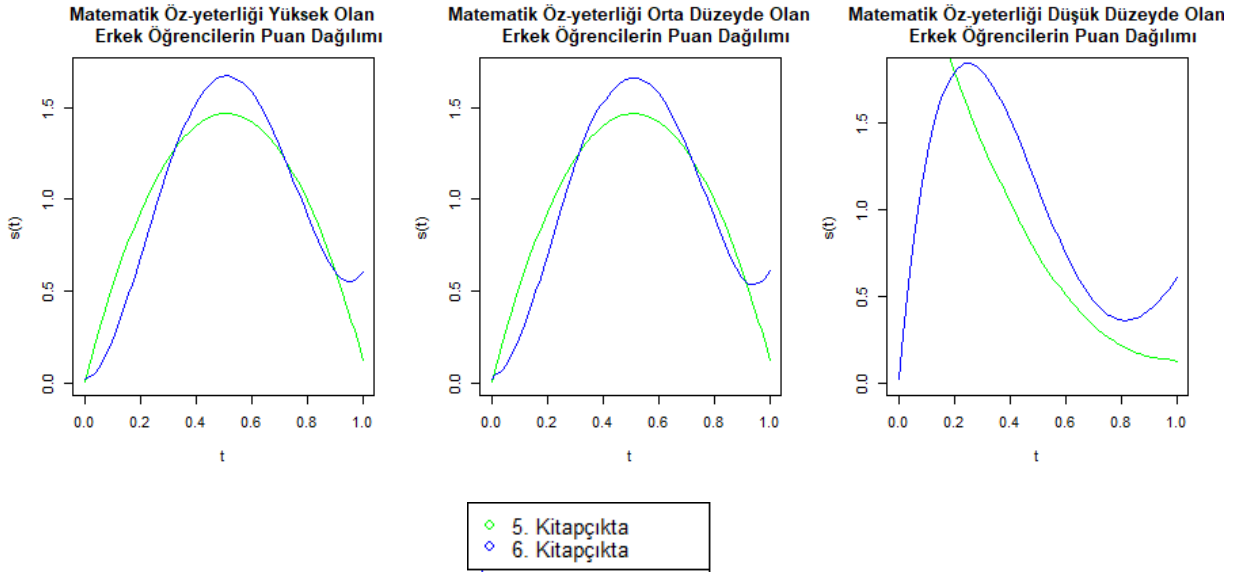
Aynı profile sahip bireylerin farklı kitapçıklarda farklı dağılımlar göstermesi bu kitapçıklar arasında karşılaştırma imkanı vermektedir. Şekil 23-25 ve Şekil 24-26 karşılaştırıldığında, 6.kitapçığın 5. Kitapçığa göre daha kolay olduğunu ifade etmek mümkündür.

Kız ve erkek öğrencilerin dağılımlarını kitapçıklara göre birlikte görmek karşılaştırma açısından önemli sonuçlar verecektir.



Şekil 27. Matematik öz-yeterliği düzeylerine göre kız öğrencilerin kitapçıklardaki puan dağılımı.

Kız öğrencilerin matematik öz-yeterlik düzeylerine göre kitapçıklardaki dağılımları Şekil 27’de verilmiştir. Bu dağılımlar için ortak olan profil her bir matematik öz yeterlik düzeyi için kız öğrencilerin puanlarıdır. Aynı profile sahip kızların farklı kitapçıklarda aldığı puanların dağılımları incelenmektedir. Matematik öz-yeterliği yüksek ve orta düzeyde olan kızların kitapçıklardaki dağılımları benzer olup, 6. kitapçıkta dağılımların daha sivri olduğu görülmektedir. Matematik öz yeterlik düzeyi düşük olan kız öğrencilerin yüksek puan alma yoğunluklarının kitapçıkların ikisi için de azaldığı ancak 6. kitapçıkta bu durumun biraz dışına çıkılarak 80 ve üzeri puan alan bireylerin sayısında da artış olduğu görülmektedir. Dolayısı ile her düzeydeki dağılımlar için yüksek puan alan bireylerin sayısı 6. kitapçıkta daha fazladır. Bu dağılımların kitapçıklara göre karşılaştırılmasını ele alarak, 6. kitapçığın 5 kitapçıktan daha kolay olduğunu görmek mümkündür. Aynı şekilde erkek öğrenciler için de dağılımlar incelenebilir.

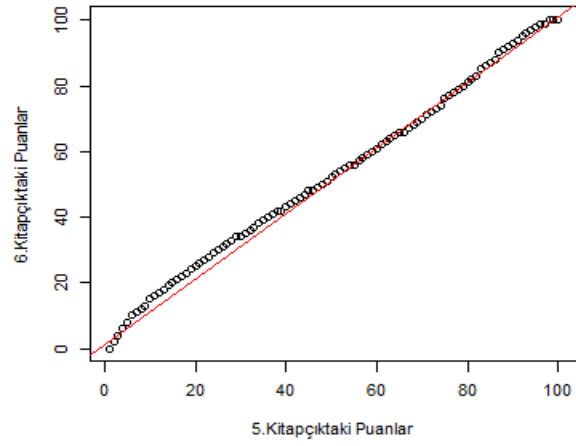


Şekil 28. Matematik öz-yeterliği düzeylerine göre erkek öğrencilerin kitapçıklardaki puan dağılımı.

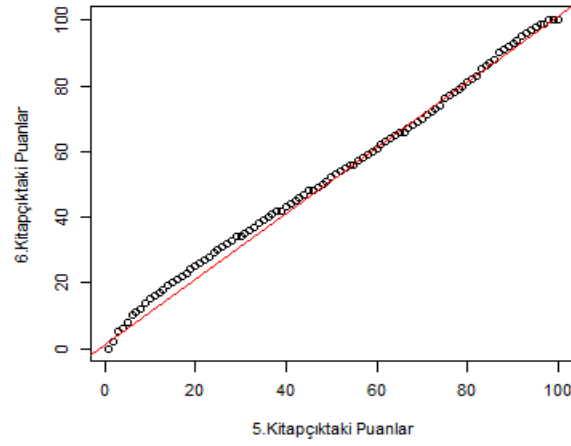
Erkek öğrenciler için matematik öz yeterlik düzeylerine göre oluşturulan kitapçıklardaki dağılımlar, kız öğrencilerin dağılımına benzerlik göstermektedir. Erkek öğrenciler için de yüksek ve orta düzeyde matematik öz yeterliğe sahip dağılımlar benzerlik göstermektedir. Aynı zamanda matematik öz yeterlik düzeylerinin düşük olması durumunda da dağılımlar kitapçıklara göre farklılık göstermekte ve 6. kitapçıkta yüksek puan alma yoğunluğu artmaktadır. Genellikle kız ve erkeklere ait dağılımlar örtüşmektedir. Fakat bazı düzeylerde ise bu dağılımlar küçük farklılıklar göstermektedir.

Matematik öz yeterlik düzeylerine göre kız ve erkek öğrenciler için oluşan profillerden 5. ve 6. kitapçıklar arasında yapılan eşitleme sonuçlarına ait grafikler elde edilmiş ve Şekil 29'da verilmiştir.

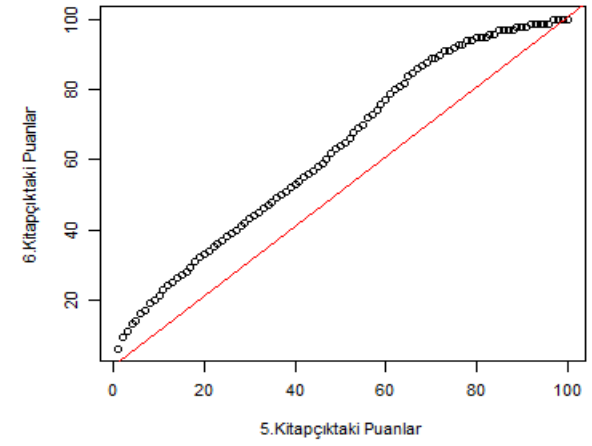
Matematik Öz-Yeterliği Yüksek Düzeyde olan Kız Öğrencilerin Kitapçıklardaki Puanlarının Eşitlenmesi



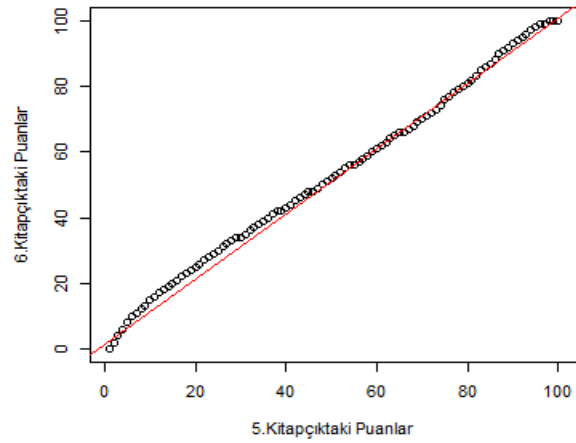
Matematik Öz-Yeterliği Orta Düzeyde olan Kız Öğrencilerin Kitapçıklardaki Puanlarının Eşitlenmesi



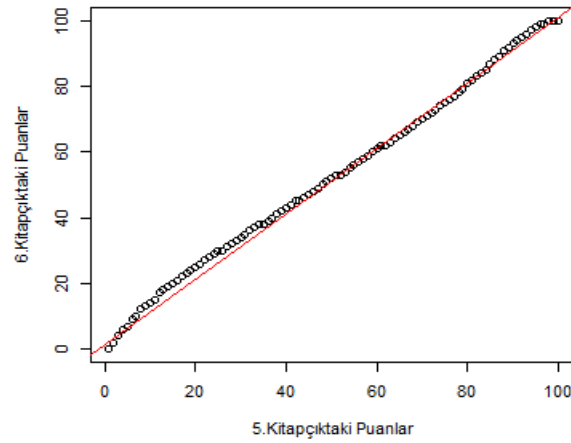
Matematik Öz-Yeterliği Düşük Düzeyde olan Kız Öğrencilerin Kitapçıklardaki Puanlarının Eşitlenmesi



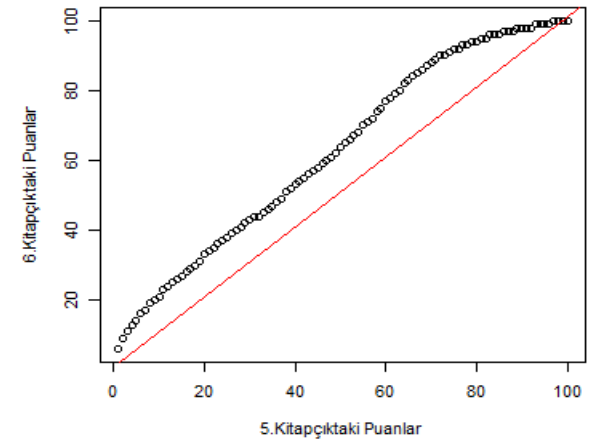
Matematik Öz-Yeterlik Düzeyi Yüksek olan Erkek Öğrencilerin Kitapçıklardaki Puanlarının Eşitlenmesi



Matematik Öz-Yeterliği Orta Düzeyde olan Erkek Öğrencilerin Kitapçıklardaki Puanlarının Eşitlenmesi



Matematik Öz-Yeterliği Düşük Düzeyde olan Erkek Öğrencilerin Kitapçıklardaki Puanlarının Eşitlenmesi



Şekil 29. Matematik öz yeterlik düzeylerine göre öğrencilerin 5. ve 6. kitapçık puanlarının PoB modeli ile eşitlenmesi.

Kategoriler için puan dağılımları arasında eşitleme dönüşümü yapılmıştır. Yüksek matematik öz yeterliğine sahip bireyler (kız / erkek öğrenciler) ile orta düzey matematik öz yeterliğine sahip bireylerin dağılımlarında gözlendiği gibi eşitlenmiş puanların grafiklerinde de benzerik görülmektedir. Bu grafiklerde görüldüğü üzere yaklaşık 60 puana kadar bireylerin 5. kitapçıkta daha az puan aldığı, bu puandan sonra ise kitapçıklarda aynı puanı aldıkları görülmektedir. Örneğin, matematik öz yeterlik düzeyi yüksek olan bir erkek öğrenci, 5. kitapçıkta 27 puan aldığında 6. kitapçıkta 32 puan almakta; 5. kitapçıkta 74 puan aldığında ise 6. kitapçıkta da 74 puan almaktadır. Matematik öz yeterlik düzeyi düşük kız ve erkekler için elde edilen eşitlenmiş puanlar arasındaki ilişki kendi aralarında benzerlik gösterse de diğer düzeylerde elde edilen ilişkiden biraz farklılık göstermektedir. Bu düzeyde eşitlenmiş puanlar karşılaştırıldığında bireylerin 6. kitapçıkta daha yüksek puanlar aldığı daha net bir şekilde görülmektedir. Örneğin, matematik öz yeterlik düzeyi düşük ve 5. kitapçıktan 64 almış bir kız öğrencinin 6. kitapçıkta alacağı puanı 82 dir.

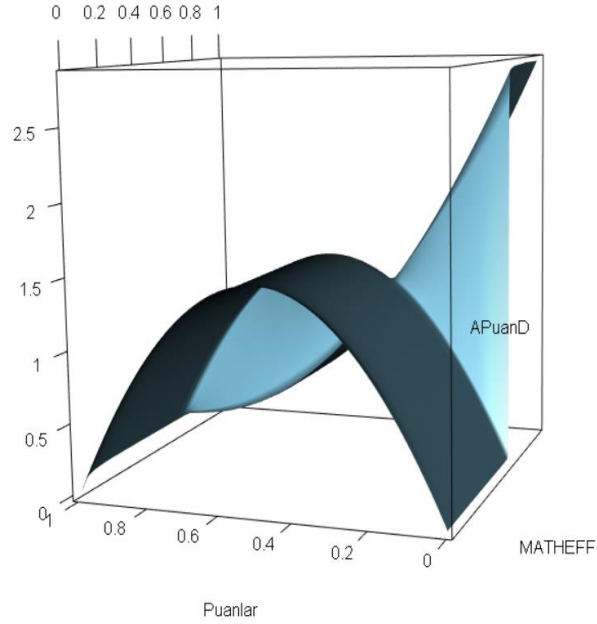
Parametrik olmayan Bayes modelinin eşitlemede kullanılması, dağılımlar hakkında detaylı bilgi edinilmesini ve profillere ait dağılımlar aracılığı ile her türlü karşılaştırmanın yapılmasını sağlamaktadır.

PoB modeli ile yapılan eşitleme çalışmalarında grup değişmezliği varsayımı kontrol altında tutulduğundan ortak profillerden yararlanarak kitapçıklar için eşitlenmiş puanlar elde edilebildiği gibi aynı kitapçığı almış kız ve erkek öğrenciler arasında veya matematik öz yeterlik düzeyleri arasında da eşitleme veya dönüşüm işlemi gerçekleştirilebilir. Böylece matematik öz yeterlik düzeyi düşük olan öğrenciler ile matematik öz yeterlik düzeyleri yüksek olan öğrenciler arasında, bu öğrencilerin göstermiş oldukları performanslar ve almış oldukları kitapçıklar hakkında yorum yapmak, aralarındaki ilişkiyi incelemek mümkün olacaktır.

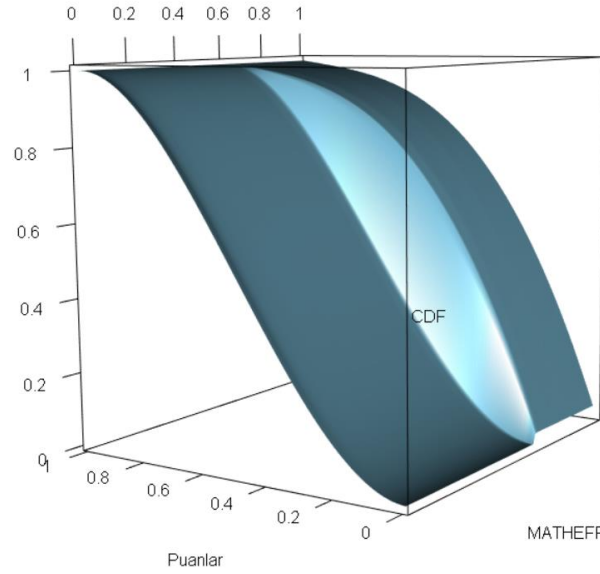
4.2. Sürekli olarak ele alındığında eşitlenmiş puanların dağılımının hedef teste ait dağılıma olan uzaklığı nasıldır?

Araştırma problemi için öğrencilerin sahip oldukları her bir MATHEFF değeri cinsiyetlere göre incelenmiştir.

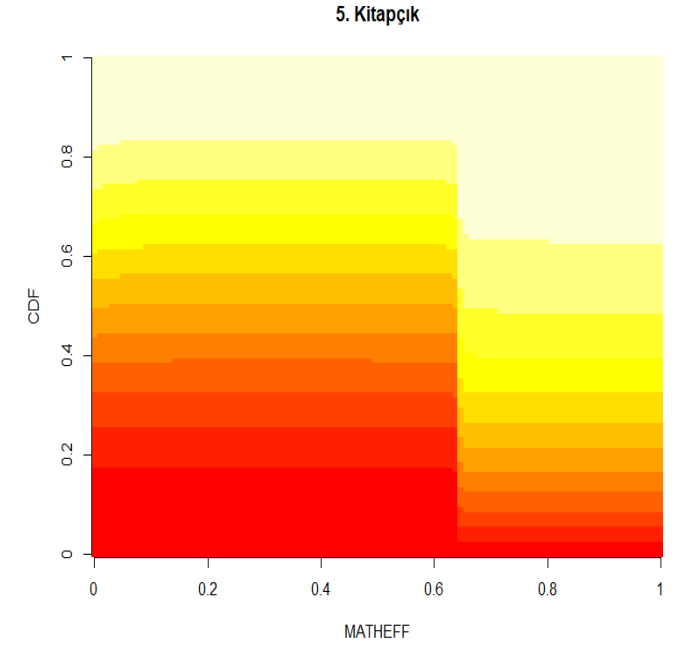
Erkek öğrencilerin 5. Kitapçıkta dağılımları Şekil 30, 31 ve 32'de yer almaktadır.



Şekil 30. Erkek öğrencilerin 5. Kitapçıkta MATHEFF düzeylerine göre kestirilen puanlarına ait dağılımlar.

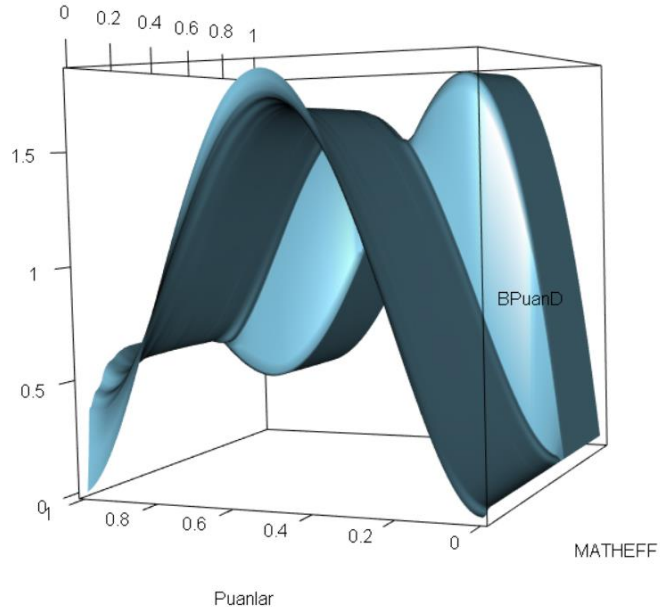


Şekil 31. Erkek öğrencilerin 5. kitapçıkta MATHEFF düzeylerine göre kestirilen puanlarına ait kümülatif dağılımlar.

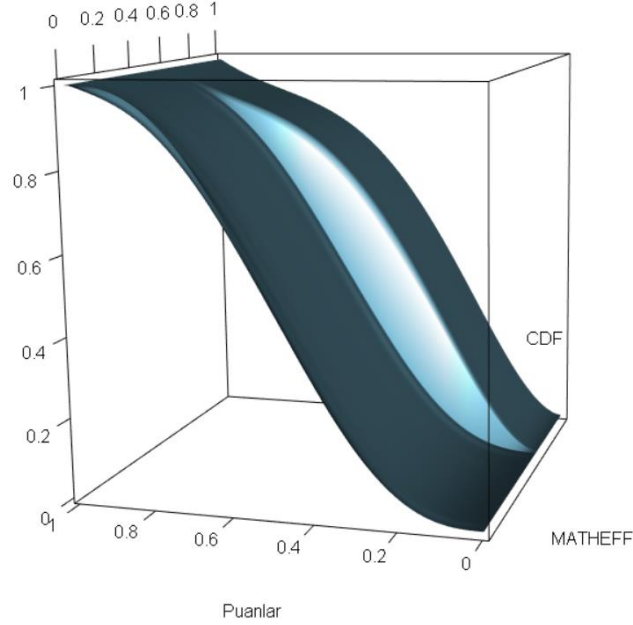


Şekil 32. Erkek öğrencilerin 5. kitapçıkta kümülatif dağılımlarını perspektift görünüşü

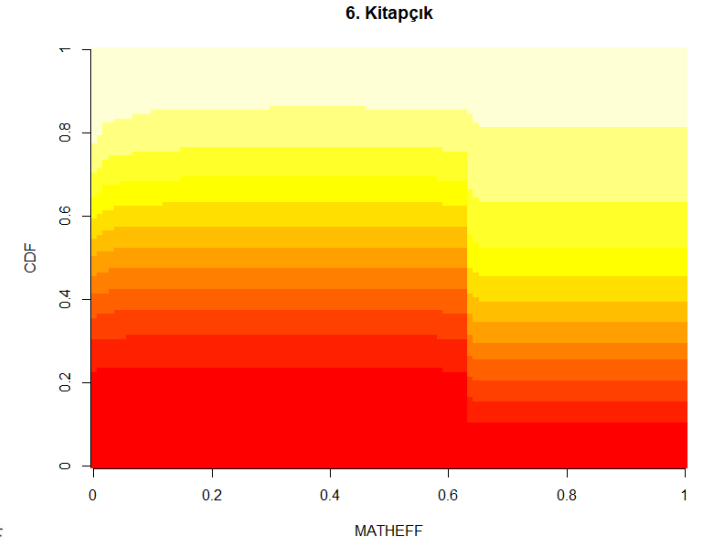
Erkek öğrencilerin 6. Kitapçığındaki dağılımları Şekil 33, 34 ve 35’de yer almaktadır.



Şekil 33. Erkek öğrencilerin 6. kitapçıkta MATHEFF düzeylerine göre kestirilen puanlarına ait dağılımlar.



Şekil 34. Erkek öğrencilerin 6. kitapçıkta MATHEFF düzeylerine göre kestirilen puanlarına ait kümülatif dağılımlar.

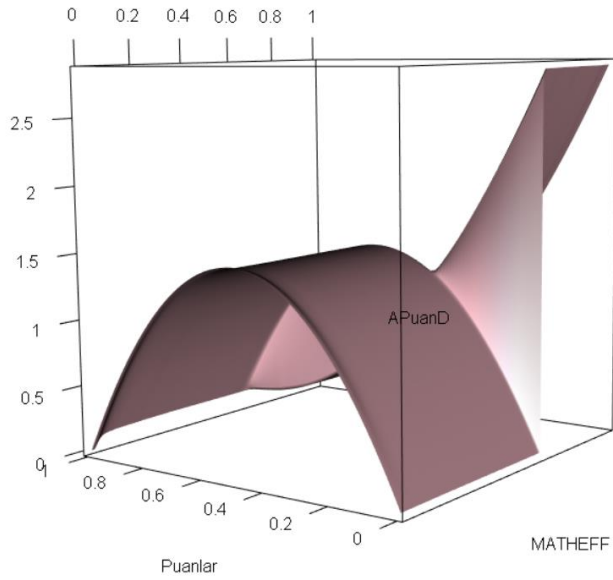


Şeki 35. Erkek öğrencilerin 6. kitapçığındaki kümülatif dağılımlarını perspektift görünüşü.

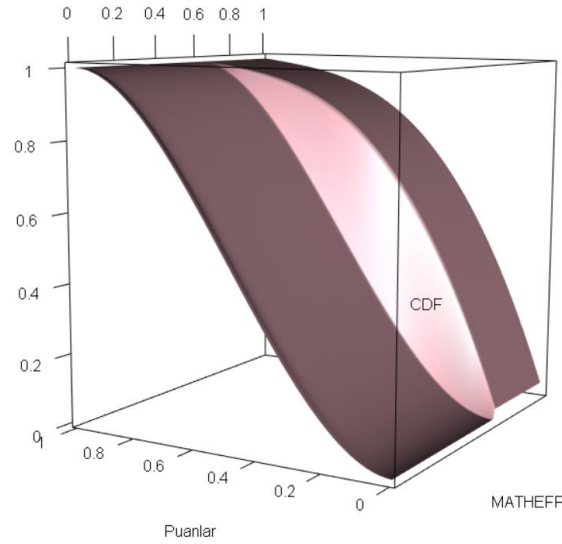
Şekil 30 ve Şekil 33’de erkek öğrencilerin 5. ve 6. kitapçıkta tüm matematik öz yeterlik düzeyinde göre alacakları puan dağılımları elde edilmiştir. Bu dağılımlarda görüldüğü üzere MATHEFF puanı arttıkça, yani matematik öz yeterlik düzeyi azaldıkça, bireylerin yüksek puan alma yoğunlukları da azalmaktadır. Düşük düzeyde matematik öz yeterliğe sahip erkek öğrenciler 5. kitapçıkta en düşük puanlarda yoğunluk gösterirken, 6. kitapçıkta ise yaklaşık olarak 20 puan civarında yoğunluk göstermektedir. Dolayısı ile 6. kitapçığın erkek öğrenciler için daha kolay olduğunu bu alt problem için de söylemek mümkündür. Ayrıca erkek öğrenciler için dağılımların kitapçıklarda farklılaşması, bu iki ortak değişkenin kullanılmasının kitapçıklardaki farklılığı ortaya çıkartmada etkili olduğu şeklinde ifade edilebilir.

Kümülatif dağılımlardan da MATHEFF puanındaki artışın dağılımların şeklinde değişikliğe sebep olduğu görülmektedir. Kümülatif dağılımlarının perspektif görünüşlerini incelediğimizde iki kitapçıkta da yaklaşık olarak 65 ve üzeri MATHEFF puanına sahip erkekler için dağılımların şekil değiştirdiği görülmektedir. 5. kitapçıkta yüksek MATHEFF puanına sahip erkeklerin 0,6 CDF değerinden sonra, 6. kitapçıkta ise 0,8 CDF değerinden sonra artışın azaldığı görülmektedir. Bu durum da yine 5. kitapçıkta yüksek puan alma durumunun zorlaştığını göstermektedir.

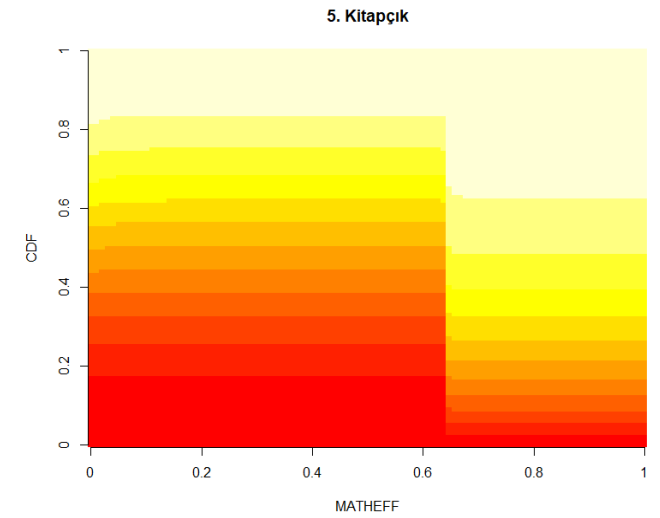
Kız öğrencilerin 5. kitapçıkta dağılımları Şekil 36, 37 ve 38'de görülmektedir.



Şekil 36. Kız öğrencilerin 5. kitapçıkta MATHEFF düzeylerine göre kestirilen puanlarına ait dağılımlar.

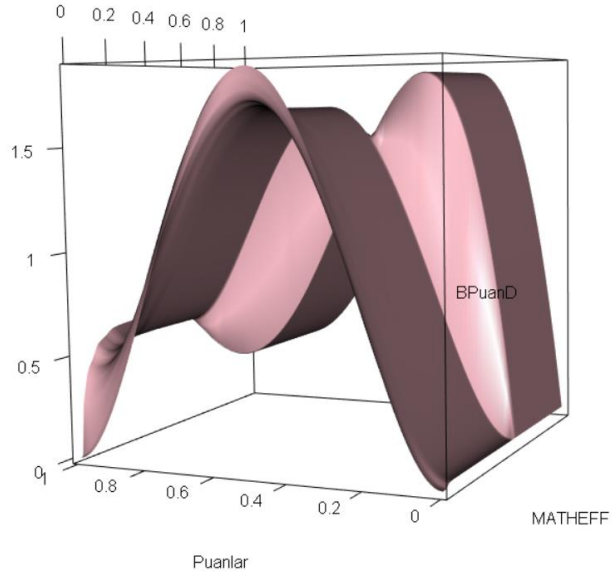


Şekil 37. Kız öğrencilerin 5. kitapçıkta MATHEFF düzeylerine göre kestirilen puanlarına ait kümülatif dağılımlar.

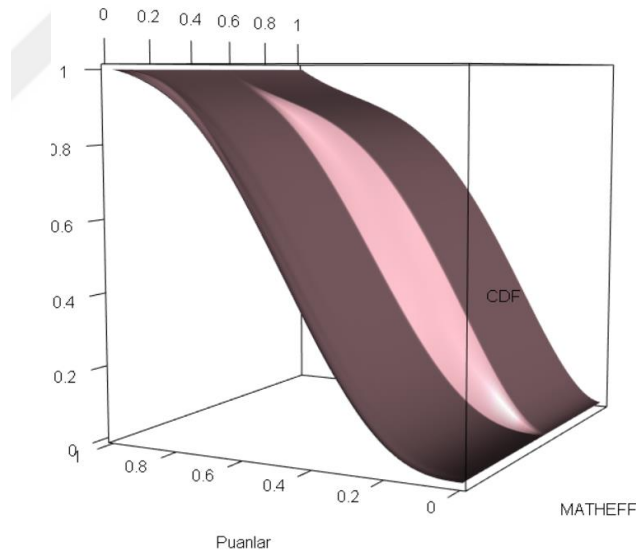


Şekil 38. Kız öğrencilerin 5. kitapçıkta kümülatif dağılımlarını perspektift görünüşü.

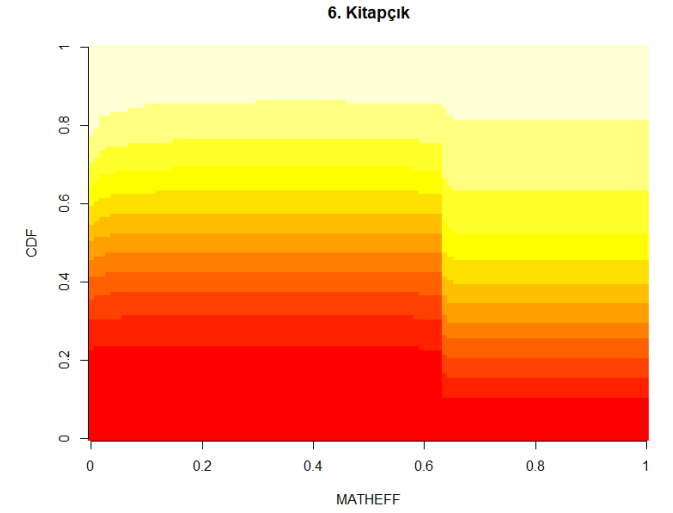
Kız öğrencilerin 6. kitapçıktaki dağılımları Şekil 39, 40 ve 41'de görülmektedir.



Şekil 39. Kız öğrencilerin 6. kitapçıkta MATHEFF düzeylerine göre kestirilen puanlarına ait dağılımlar.



Şekil 40. Kız öğrencilerin 6. kitapçıkta MATHEFF düzeylerine göre kestirilen puanlarına ait kümülatif dağılımlar.

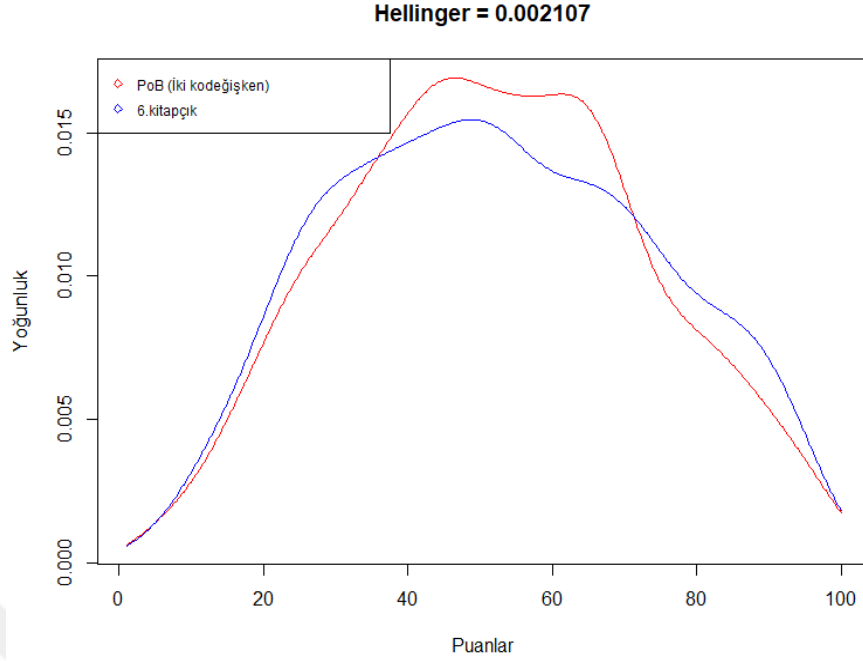


Şeki 41. Kız öğrencilerin 6. kitapçıktaki kümülatif dağılımlarını perspektift görünüşü.

Kız öğrenciler için elde edilen dağılımlar erkek öğrencilerden elde edilen dağılımlar ile benzerlik göstermektedir. PoB modeli kullanılarak oluşturulan eşitleme sürecinde her bir bireye özgü ortak değişkenler, farklı bir özellik olarak o bireylerin puanları ile birleştirilmekte ve en son hali ile eşitlenmiş puanlar elde edilmektedir. MATHEFF değişkeni kategorik olarak alındığından, ortak değişkenlerin kategorilerine göre eşitlenmiş puanlar elde edilmiştir. Ancak MATHEFF değişkeni sürekli olarak alındığında her bir öğrenci için farklı düzey MATHEFF puanı olduğu için eğer aynı cinsiyetten aynı öz yeterlik düzeyine sahip bireyler yoksa, birey sayısı kadar kategori / profil oluşacaktır. Dolayısı ile kitapçıklar arasındaki eşitlenmiş puanları gösteren grafiklerin her bir öğrencinin sahip olduğu profillere göre oluşturulması gerekmektedir. PoB modeli ile eşitlenmiş puanlar olarak elde edilmiş olup, kitapçıklar arasındaki eşitlenmiş puanları gösteren grafiklere yer verilmemiştir. Ancak eşitlenmiş puanları elde etmede kullanılan kümülatif dağılımlardan çıkarımları yapmak mümkündür. Her bir bireyin sahip olduğu profillere yönelik olarak modelden elde edilen eşitlenmiş puanların dağılımına yer verilmiştir.

MATHEFF ve cinsiyet değişkeni birlikte ortak değişken olarak alındığında daha önceki alt problemlere kıyasla model daha detaylı bilgi vermektedir. Bu alt problem ile hangi değişkenin bireylerin matematik başarısına ait dağılımlarını ne kadar değiştirdiğini görmek mümkündür. İki değişkenin birlikte kullanılması kitapçıklar arasındaki fark ortaya çıktığından bu değişkenlerin birlikte, kitapçıklardaki matematik performansı için bir açıklayıcı olduğunu ifade etmek mümkündür. Değişkenler ayrı ayrı ele alındığında, kız ve erkekler için elde edilen dağılımlar oldukça benzerlik gösterdiğinden tek başına cinsiyetin matematik performansını açıklamada yeterli olmadığı ifade edilebilir. Ancak sadece MATHEFF değişkeninin kullanıldığında elde edilen dağılımlar ile iki değişkenin kullanılması ile elde edilen dağılımlar karşılaştırıldığında özellikle 6. kitapçık için dağılımda değişiklik olduğu görülmektedir. MATHEFF değişkeninin ise matematik başarısını açıklamada tek başına etkili olabileceğini ifade etmek mümkündür.

Her bir öğrencinin cinsiyet ve matematik öz yeterlik düzeylerine göre eşitlenmiş puanları elde edilmiştir. Bu puanlara ve hedef teste ait olasılık dağılım birlikte incelenmiş ve bu dağılıma Şekil 42’de yer verilmiştir.



Şekil 42. İki ortak değişkenine göre PoB yönteminden elde edilmiş olan puanların ve hedef testteki puanların dağılımları.

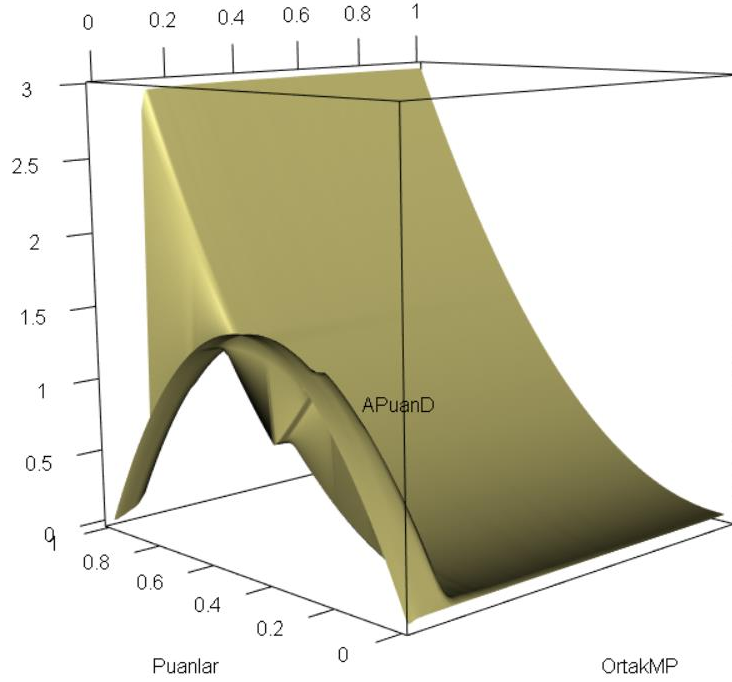
İki değişken birlikte ele alınarak modele dahil edildiğinde eşitlenmiş puanlara ait dağılım hedef teste oldukça yakındır. Bu yakınlığı temsil edecek Hellinger uzaklığının da diğer modellerdekine kıyasla oldukça küçük değer verdiği (0,002107) görülmektedir. Bu değer ve grafik ile birlikte değerlendirildiğinde, modelde bu ortak değişkenlerin kullanımının eşitlenmiş puanları hedef teste yaklaştırdığı ifade edilebilir. Özellikle uç değerlere doğru ise dağılımların birbirine yaklaştıkları ve genellikle uç değerlerde oluşan hatayı tolere etmede de kullanılabilecek bir yöntem olarak ele alınabileceği öngörüsünü de getirmektedir.

Alt Problem 5'e İlişkin Bulgular ve Yorumlar

2012 PISA matematik alt bilişsel testi puanlarını PoB modeli ile eşitlemek için ortak maddelerin puanları ortak değişken olarak alındığında elde edilen eşitlenmiş puan dağılımının hedef teste ait dağılıma olan uzaklığı nasıldır?

Araştırmanın ilk kısmında MTK ölçekleme yöntemlerine göre ortak maddeler üzerinden eşitlenmiş puanlar elde edilmiştir. Bu alt problemi için ise ortak maddelerinden elde edilen puanlar, ortak değişken olarak kullanıldığında PoB modelinden elde edilecek eşitleme sonuçları incelenmiştir.

Öğrencilerin 5. kitapçıktaki dağılımları Şekil 43 te görülmektedir.



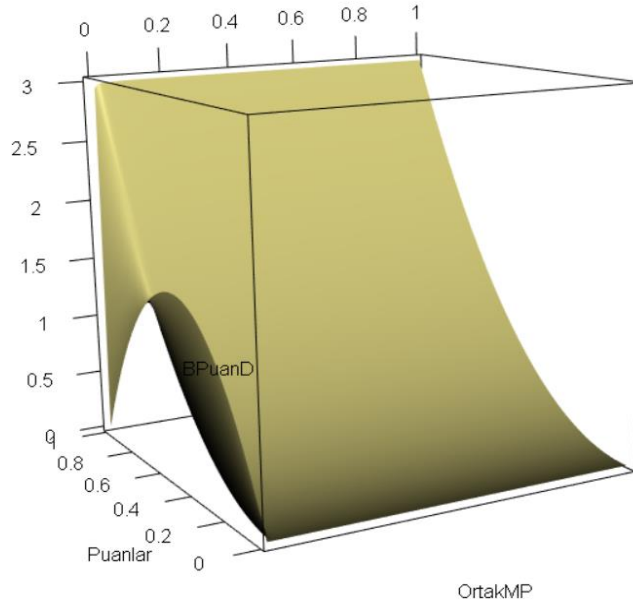
Şekil 43. 5. Kitapçıkta ortak maddelerden elde edilen puanlara göre öğrencilerin puan dağılımları.

Ortak maddelerdeki toplam puanlara göre öğrencilerin 5. kitapçıkta gösterdikleri dağılımlar elde edilmiştir. 5. kitapçıktaki test puanları ile bu kitapçıktaki ortak madde puanlarından oluşturulan test puanları arasındaki korelasyon 0,793 olarak elde edilmiştir.

Şekil 43 incelendiğinde, ortak madde puanlarının ortak değişken olarak modelde yer almaması veya çok düşük puan ile model katkıda bulunması durumunda 5. kitapçıkta öğrencilerin ortalama puanlarda yoğunluk gösterdiği ve uç puanlara doğru yoğunlukların azaldığı görülmektedir. Ortak madde puanlarından elde edilen puanlar arttıkça, ilk dağılımlara kıyasla dağılımların şekilleri farklılaşmakta; düşük puan alma yoğunluklarının azaldığı ve yüksek puan alma

yoğunluklarının arttığı görülmektedir. Ortak madde puanlarının yüz tam puan üzerinden 13 ve üzeri puan olması durumunda bireylerin alabilecekleri puan dağılımları benzerlik göstermektedir. Ortak madde puanlarından 20 almış birey ile 80 veya 100 almış bireyin 5. kitapçıkta göstermiş oldukları dağılımlar aynı olup, ortak madde puanlarının farklı olması bireylerin matematik performansları arasındaki farkı açıklamada yetersiz kalmıştır.

Öğrencilerin 6. kitapçıktaki dağılımları Şekil 44 te verilmiştir.



Şekil 44. 6. Kitapçıkta ortak maddelerden elde edilen puanlara göre öğrencilerin puan dağılımları.

Bireylerin ortak maddelerden almış oldukları puanlara göre 6.kitapçıktaki puan dağılımları görülmektedir. 6.kitapçıktaki test puanları ile bu kitapçıktaki ortak madde puanları arasındaki korelasyon 0,750 olarak elde edilmiştir.

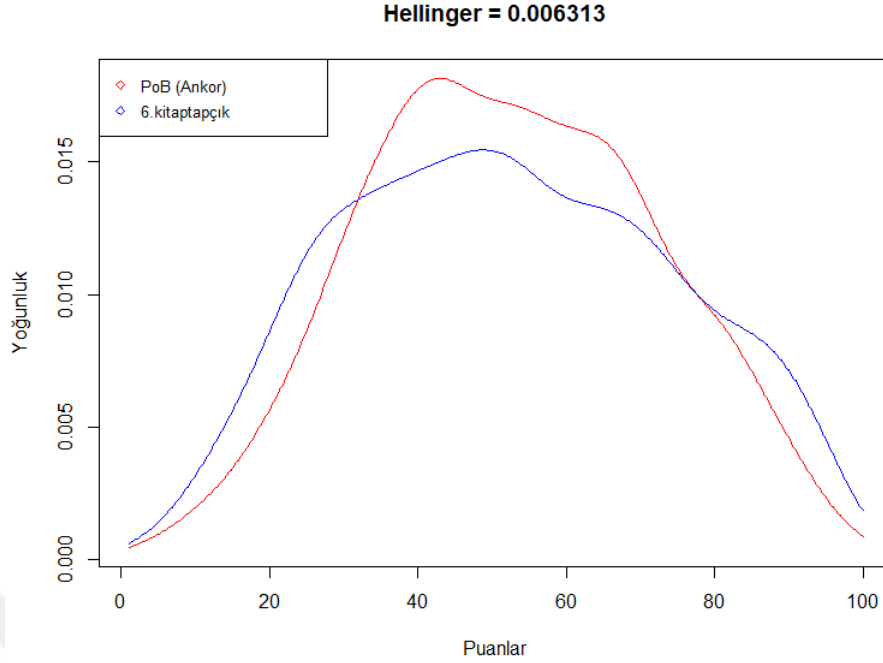
Ortak madde puanlarının ortak değişken olarak modelde yer almaması veya düşük puan ile modelde bulunması durumunda bu kitapçıkta da öğrencilerin ortalama puanlarda yoğunluk gösterdiği görülmektedir. Ortak maddelerden alınan puanların yüz tam puan üzerinden 2 puan ve üzeri olması durumunda bireylerin gösterdikleri dağılımlar oldukça benzerdir. Bu durumda ortak maddelerden alınan puanlar bireylerin dağılımlarını değiştirirse de, 2 ve üzeri ortak madde puanına sahip bireyleri birbirinden ayıramadığı ifade edilebilir.

Ortak madde puanlarının ortak deęiřken olarak alındığında elde edilen daęılımların bireylerin matematik testi puanlarından elde edilen daęılımı deęiřtirdięi ancak bu deęiřimin sadece dūřuk ortak madde puanlarında olduęu g r lmektedir. Bu da farklı d zey ortak madde puanına sahip  ęrencilerin matematik bařarıları arasında net bir ayırım yapılamadıęını g stermektedir. İki kitap ıkta bu durum benzer řekilde olmakla birlikte 5. kitap ıkta ortak test puanlarının dięer kitap ıęa kıyasla daha fazla ayırım yapabildięi ifade edilebilir. Dolayısı ile ortak maddelerin aynı d zeyde, aynı zorluk seviyesinde olduęu ve tek bir ortak maddenin bile bu maddelerin g revini  stenebileceęi ifade edilebilir. Ayrıca ortak maddelerindeki puan artışı daęılımlarda deęiřiklik yapmadıęı i in, ortak maddelerin kitap ıklardaki soruları i erik olarak yansıtmadıęını s ylemek m mk nd r. Dolayısı ile ortak madde puanlarının 5. kitap ıęı kısmen temsil ettięi ve dięer testten 5. testin i erięine daha  ok benzedięi ifade edilebilir. (korelasyonda da bu sonu  g r lmektedir).

Ortak maddelerin istatistiksel ve i erik olarak eřitlenecek testleri temsil etmesi gerekmektedir. İstatistiksel olarak bu temsiliyetin, korelasyonlardan elde edilen sonuca g re karřılandıęını s ylemek m mk nd r.

Ortak madde puanları ile her iki kitap ıkta benzer daęılımlar elde edilmiř olsa da, d řuk ortak test puanlarında daęılımlar kitap ıklara g re farklılařmaktadır. Dięer alt problemlerden elde edilen sonu lara benzer olarak, bu alt problemde de 6. kitap ıęın 5. kitap ıktan daha kolay olduęu ve bu kitap ıkta bireylerin y ksek puan olma yoęunluęunun fazla olduęu g r lmektedir.

Her bir  ęrencinin ortak maddelerden aldıkları puanların ortak deęiřken olarak ele alınması ile eřitenmiř puanlara ve hedef teste ait olasılık daęılımları birlikte incelenmiřtir. Elde edilmiř daęılımlara řekil 45 te yer verilmiřtir.



Şekil 45. Ortak madde puanlarının ortak değişken olarak ele alınması ile PoB yönteminden elde edilen puanların ve hedef testteki puanların dağılımları.

Ortak maddelerden elde edilen puanların ortak değişken olarak kullanılması ile elde edilen dağılımın hedef teste uzaklığını 0,006313 olarak elde edilmiştir. Hellinger uzaklığına göre elde edilen bu değer MTK yöntemlerinden elde edilen değerlerden daha küçük iken, PoB modelinin kullanıldığı durumda elde edilen değerlerden büyüktür. Dağılım incelendiğinde ise cinsiyet değişkeninin kullanıldığı PoB modeli ile benzerlik göstermekte olan bu dağılımı uç değerlere doğru hedef test dağılımından uzaklaşmaktadır.

Noktasal olarak mesafeyi hesaplayan Hellinger uzaklıklarının sayısal değerleri yetersiz olsa da şekilsel olarak bu dağılımlar hakkında verilen bilgileri desteklemektedir. Dolayısı ile ortak maddelerin Parametrik olmayan Bayes yöntemi ile ortak değişken olarak ele alınması, MTK modelinde yapılan analizden hedef teste daha yakın sonuç üretmiştir. Bayes yönteminin klasik yöntemlerden daha iyi kestirim yaptığını ve daha yararlı bilgiler için de kullanılabileceğini ifade eden çalışmalar bu sonucu desteklemektedir (Karabatsos & Walker, 2009; Kruschke, 2012; Kruschke, Aguinis & Joo, 2012; Schoot, vd., 2013).

Ortak maddelerin eşitleme için gereken özellikleri taşımadığı veya test eşitleme için varsayımların ihlal edildiği durumlar için, ortak değişkenlerin kullanılmasının eşitleme sürecinde daha uygun sonuçlar vereceğini destekleyen çalışmalar literatürde yer almaktadır (Liou, Chen & Li, 2001; Wright & Dorans,1993; Dorans & Holland,2000).



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

- Bu araştırmada, klasik yöntemlerden farklı bir bakış açısı getirecek olan PoB modeli ile eşitleme yapılmıştır. Araştırmada ortak değişken olarak öğrencilerin cinsiyetleri, matematik öz yeterlik puanları ve ortak maddelerden aldıkları puanlar kullanılmıştır. Farklı ortak değişkenlerin ele alınması ile eşitlenmiş puanlar elde edilmiş ve bu puanların dağılımlarının hedef testin dağılımına olan uzaklıkları incelenmiştir. PoB modelinin kullanılması ile değişkenlerin kitapçıklar arasındaki farka ve matematik başarısına etkisi hakkında yorumlara ulaşılmıştır. MTK ve PoB modeli ile elde edilen bulgu ve yorumlara ait sonuçlar aşağıda sunulmuştur.
- Araştırmada ortak maddelerden elde edilen puanlar dış ortak test olarak alınmıştır. Ortak maddelerin parametreleri üzerinden yapılan ölçekleme sonucunda en düşük hata ortalama-standart sapma; en yüksek hata ise Stocking-Lord yönteminden elde edilmiştir. Bu durum dış ortak maddelerinin, testlere ait madde karakteristik eğriler arasındaki farkı azaltmada moment yöntemlerine kıyasla daha fazla hataya sebep olduğu; gruplara ortak testin uygulanması ile elde edilen ayırıcılık parametreleri arasındaki farkın, karakteristik eğrilerdeki farktan daha düşük olduğu şeklinde ifade edilebilir. Kitapçıklardaki ele alınan değişkenlere (cinsiyet ve MATHEFF) ait bireylerin sayılarındaki farklılıkta RMSE değerlerin etkilenmesine sebebiyet verebileceğini söylemek mümkündür. MTK ölçek dönüştürme yöntemi ile elde edilen gerçek puanların hedef teste olan uzaklıkları incelendiğinde; yöntemlerden elde edilen dağılımların benzer olduğu ancak en yakın dağılımın Stocking- Lord yönteminden elde edildiği sonucuna ulaşılmıştır.
- Modelde ortak değişken olarak sadece cinsiyet ele alındığında, kız ve erkek öğrenciler için kitapçıklarda benzer dağılımlar elde edilmiştir. Cinsiyet değişkeninin kitapçıklar arasındaki farkı göstermede yetersiz olduğu sonucu görülse de DoGOD desenine ait varsayımlardan kaynaklı olarak 6.kitapçığın 5.

kitapçıktan daha kolay sorular içerdği sonucu elde edilmiştir. Aynı kitapçığı almış olan kız ve erkek öğrenciler için dağılımlar oldukça benzer olup cinsiyetin matematik performansı üzerinde önemli bir etkisinin olmadığı sonucuna ulaşılmıştır. Her bir bireyin cinsiyeti ve test puanlarından yararlanılarak elde edilen eşitlenmiş puanların dağılımı hedef teste ait puanların dağılımından daha sivri olarak elde edilmiştir. Dolayısı ile dağılımların birbirlerine MTK yöntemine kıyasla yakınlığı ve bu model kullanılarak daha keskin kestirimler elde edildiği sonucuna ulaşılmıştır.

- Modelde ortak değişken olarak MATHEFF alındığında, ilk olarak belirlenen kategorilere göre bireylerin dağılımları, daha sonra ise tüm düzeylerdeki bireylere yönelik dağılım grafiği verilmiştir. Ele alınan kategorilere göre orta ve yüksek düzeydeki profile sahip bireylere ait dağılımlar benzerlik göstermiş, düşük düzeydeki profile sahip bireylere ait dağılımlar ise bu iki düzeydeki dağılımlardan farklılaşmıştır. Bu değişim üç boyutlu dağılım grafiklerinde de görülmektedir. Kitapçıkların her ikisi için de dağılımlar MATHEFF puan düzeyinde göre değişim gösterdiğinden, MATHEFF değişkeninin matematik performansında bireyler arasındaki farkı ortaya koyduğu sonucuna ulaşılmaktadır. Bireylere ait puan dağılımlarının alt gruplarda farklılaşması kontrol altında tutularak, alt gruplara göre eşitlenmiş puanlar elde edilmiştir. MATHEFF değişkeninin modelde kullanılması ile eşitlenmiş puanlardan elde edilen dağılımın, hedef testteki puanlara yaklaştığı sonucunu ortaya çıkarmaktadır. Ayrıca her iki kitapçıkta da benzer bir dağılım gözlemlenmiştir. Bu durum ise değişkenin kitapçıklar arasındaki farkı ortaya çıkarmada yeterli olmadığını göstermektedir.
- MATHEFF ve cinsiyet birlikte ortak değişken olarak alındığında daha önceki alt problemlere kıyasla modelde daha detaylı bilgiler elde edilmiştir. Bu değişkenler birlikte ele alındığında, her kitapçık ve MATHEFF değişkeninin her puan düzeyi için farklı dağılımlar oluşturduğundan bu değişkenlerin birlikte hem kitapçıklar arasındaki farkı hem de matematik performansı için bireyler arasındaki farkı ortaya çıkartabildikleri sonucuna ulaşılmıştır. İki ortak değişken kullanımı ile elde edilen uzaklık değeri ve dağılımlar incelendiğinde, eşitlenmiş puanlar ve hedef test puanlarına ait dağılımların birbirlerine yaklaştığı sonucuna ulaşılmıştır.

- Ortak madde puanları ortak değişken olarak alındığında bireylere ait elde edilen puan dağılımları sadece düşük puanlarda ve çok az bir ranjda değişmektedir. Dolayısı ile ortak maddelerden yüksek puan alan bireyler ile düşük puan alan bireylerin puan dağılımları benzerlik göstermektedir. Bu durum ortak maddelerin aynı düzey ve tek tip olduğu veya direkt test puanlarına etki ederek dağılımlarını değiştirecek nitelikte sorular olduğu sonucunu ortaya çıkarmaktadır. İki kitapçık için de bu durum benzer şekildedir. Ancak ortak madde puanlarının 5. kitapçıkla daha yüksek korelasyon vermesi ve bu kitapçıktaki dağılımlarda daha çok değişim yapmış olması, ortak maddelerin 5.kitapçıktaki sorulara daha çok benzediği ve bu kitapçıktaki farklı puan almış alt gruplar arasında daha fazla ayırım yaptığını göstermektedir. Ortak madde puanlarından elde edilen dağılımların kitapçıklara göre büyük bir farklılık göstermemesi, ortak maddelerin matematik performansında bireyler arasındaki farkı yeterli düzeyde açıklamadığı sonucunu ortaya çıkarmıştır. Ortak madde puanlarının kullanılması ile elde edilen eşitlenmiş puanlar ile hedef teste ait dağılım arasındaki uzaklık yöntemin etkili olduğunu ancak iki ortak değişken kullanılmasının ortak maddelerden daha etkili olduğu sonucunu ortaya çıkarmıştır.
- Sadece MATHEFF ve sadece cinsiyet değişkeni kullanıldığında dağılımlar kitapçıklara göre aşırı bir farklılık göstermemektedir. İki ortak değişkenin kullanıldığı modelde dağılımların kitapçıklara göre farklılıkları açık bir şekilde görülmekte; ortak madde puanlarının kullanıldığı modelde ise düşük ortak madde puanlarında kitapçıklara göre dağılımların farklılaştığı görülmektedir. Bu durum PoB modellerinde; ortak madde puanlarının kitapçıklar arasındaki farkı, sadece MATHEFF değişkeni kullanıldığı modelden daha çok açıkladığı sonucunu ortaya çıkarmaktadır.
- Bütün PoB modellerinde farklı ortak değişkenler kullanılsa dahi 6. kitapçığın 5. kitapçıktan daha kolay olduğu ve bu kitapçıkta bireylerin yüksek puan olma yoğunluğunun fazla olduğu sonucu ortaya çıkmaktadır. Aynı şekilde yine her model için 5. kitapçıktaki soruların alt gruplar için daha ayırıcı olduğunu söylemek mümkündür. Bu sonuçlardan görüldüğü üzere PoB modelinin eşitleme çalışmalarında kullanılması ile ortak değişkenlerin, ortak madde

puanlarının; kitapçıklar ve değişkenlerin alt gruplar arasındaki farkı açıklamada yeterli olup olmadığı konusunda da çıkarımlar yapılabilmektedir.

- PoB modeli ile elde edilmiş olan eşitlenmiş puanlar için hesaplanan Hellinger Uzaklığı, MTK ölçek dönüştürme yöntemlerine göre oldukça düşük olup, bu dağılımlar hedef teste daha yakındır. Bu dağılımlardan en yakın uzaklığı iki ortak değişkenin kullanıldığı PoB modeli vermiştir. Eşitlenmiş puanlara ait dağılımlar ve bu dağılımların hedef teste uzaklıkları incelendiğinde MTK yöntemleri ve PoB modelleri arasında karşılaştırma yapmak kolaydır. Ancak farklı ortak değişkenler kullanılarak PoB modellerinden elde edilen eşitlenmiş puanların hedef teste olan uzaklıklarını veren Hellinger uzaklıkları arasında karşılaştırma yapmak oldukça zordur. Hellinger uzaklığından elde edilen sayısal değerler karar almada yetersiz olsa da, dağılımların grafikleri hedef teste uzaklıkları hakkındaki bilgileri destekler niteliktedir.
- PoB modeli dağılımlar üzerinden eşitleme yaptığından, ilk testteki birey sayısı ile hedef testteki birey sayılarının eşit olması gibi bir sınırlama gerektirmez. Hatta testlerde yer alan alt gruplardaki birey sayısı için de bir sınırlamaya gerek yoktur (A kitapçığındaki 50 kız öğrenci ile B kitapçığındaki 70 kız öğrenci arasında eşitleme yapılabilmektedir). Araştırmada bazı alt gruplarda birey sayısının az olması da güven aralıklarının geniş olmasına sebep olmuştur. Ayrıca ortak değişkenlerin bilgi vermeyen önseller olarak modele eklenmesinin güven aralıklarında genişliğe sebep olduğu söylenebilir. Ancak güven aralıklarının geniş olması halinde dahi kestirimler klasik yöntemlerden dikkate değer bir şekilde hedef teste yakınlık göstermiştir. Bu durumda modelin kullanışlı olduğu sonucunu ortaya çıkartmaktadır.

Öneriler

Araştırmadan elde edilen sonuçlar ile gelecek araştırmalara yönelik olarak sunulan öneriler iki başlık altında verilmiştir.

Araştırmanın Bulgularına Yönelik Öneriler

- Araştırmada ele alınan model için kullanılan sürekli ortak değişkenlerin seçiminde matematik performansına etkisi dikkate alınmıştır. Kullanılacak değişkenin ilgili performans ile yüksek oranda ilişkili olması yanında gruplar için yanlılık oluşturmamasına dikkat edilmelidir.
- Araştırmada ele alınan modelde iki ortak değişken kullanılmıştır. Ortak değişkenden fazla kullanılması ortak kategorilere düşen birey sayısında azalmaya sebep olacak ve puanların kestirimlerinde hataya sebep olacağından çok fazla ortak değişken kullanılmamalıdır.
- Sürekli değişkenin kategorilerinin belirlenmesinde dağılımlar dikkate alınmış ve üç kategori belirlenmiştir. Kullanılacak sürekli değişkenin kategorilerinin oluşturulmasında puan dağılımlarımdaki değişim göz önüne alınmalıdır.
- Araştırmada MCMC örnekleme 150000 olarak alınmıştır. Dağılımların evrene yaklaşması için örnekleme sürecinin yüksek alınmasına dikkat edilmeli ve modeldeki parametrelere ait iz grafiğinde sonuçlar birbirine yaklaşıp daha sabit dağılımlar elde edene kadar analizler tekrar edilmelidir.
- Eşitlenecek testlere ait puanlar birim aralıklarda alınarak modele dâhil edilmelidir.

Gelecek Araştırmalara Yönelik Öneriler

- Kategorileri temsil edecek bireylerin sayısı kontrol altında tutularak uygulamada elde edilecek güven aralıkları karşılaştırılabilir.
- MTK yöntemleri için bootstrap süreçleri aynı düzeyde tutularak, bu yöntem ile karşılaştırmalar yapılabilir.
- Ortak değişken kullanılmadan sadece modelin etkililiğini incelemek için analizler yapılabilir.
- Kitapçıklarda alt gruplar arasındaki değişim incelenerek DIF li maddeler veya kitapçıklar için çalışma yürütülebilir.

- Araştırmada model olarak DDBP kullanılmıştır. Bu modelde ortak değişkenleri veri ile bağlamak için modelde kullanılan k , w ve γ parametreleri sabitlenerek kestirimler tekrardan yapılabilir.
- Ortak maddelerin ortak değişken olarak alınması durumunda bu maddelerin zorluk / ayıricılık seviyeleri kontrol altına alınarak, farklı düzeyde zorluklarda / ayıricılıklardaki maddeler modele dahil edilerek eşitlenmiş puanlar incelenebilir.
- Eşitlemesi yapılacak olan grupların ortak değişkenlerin kategorilerinde yer alan örneklem sayısı arasında yüksek bir fark olduğu durumda güven aralıkları incelenebilir.
- Sosyo-ekonomik durum, yaş, bireye ait geçmiş başarı puanları gibi sürekli ve kesikli farklı ortak değişkenler kullanılarak eşitlenmiş puanlar incelenebilir.
- Model, sürekli ölçeklere ait kesme puanların belirlenmesinde kullanılabilir.

Kaynaklar

- Ackerman, T. A. (1987). *The robustness of LOGIST and BILOG IRT estimation programs to violations of local independence*. Paper presented at the annual meeting of the American Educational Research Association, Washington, DC. (ERIC Document Reproduction Service No. ED284902).
- Akın-Arıkan, Ç. (2017). *Kernel eşitleme ve madde tepki kuramına dayalı eşitleme yöntemlerinin karşılaştırılması*, (Yayınlanmamış doktora tezi). Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü. Ankara.
- Altuntaş, M. (2011). *İstatistiksel model seçmede bayesci yaklaşımlar & bayes faktörü*. Yayınlanmamış Yüksek Lisans Tezi. Sinop Üniversitesi, Fen Bilimleri Enstitüsü. Sinop.
- Andersson, B. (2014). Contributions to Kernel Equating. Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences 106. 24 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-554-9089-8.
- Andersson, B., & Wiberg, M. (2017). Item response theory observed-score kernel equating. *Psychometrica*, Vol.82; No:1, 48-66.
- Andersson, B., Branberg, K., & Wiberg, M. (2013). Performing the kernel method of test equating with the package kequate. *Journal of Statistical Software*, 55(6), 1-25.
- Angoff, W. H. (1971). Scales, norms and equivalent scores. In R. L. Thorndike (Ed.) *Educational measurement* (pp.508–600). Washington, DC: American Council on Education.
- Barrientos, A.F., Jara, A., & Quintana, F., (2012). On the support of MacEachern's dependent dirichlet processes and extensions. *Bayesian Analysis*, 7, 2, pp.277-310.
- Barrientos, A.F., Jara, A., & Quintana, F., (2016). *Fully nonparametric regression for bounded data using Bernstein polynomials*. Tech. rep.. Department of Statistics, Pontificia Universidad Catolica de Chile.
- Berger, J.O., Boukai, B., & Wang, Y. (1997). Unied frequentist and bayesian testing of a precise hypothesis. *Statistical Science*. 1997, Vol. 12, No. 3, 133.

- Birbiri, D. (2014). *Pisa 2003 Ve Pisa 2012 sınav sonuçlarının problem çözme becerilerine yönelik değişkenlerinin türkiye açısından incelenmesi*, Yayınlanmamış Yüksek Lisans Tezi, Atatürk Üniversitesi, Eğitim Bilimleri Enstitüsü, Erzurum.
- Bramley, T., & Vidal Rodeiro, C.L. (2014). *Using statistical equating for standard maintaining in GCSEs and A levels*. Cambridge Assessment Research Report. Cambridge, UK: CambridgeAssessment.
- Branberg, B. (2010). Observed score equating with covariates. Statistical Studies No. 41). Umea: Umea University, Department of Statistics.
- Branberg, K., & Wiberg, M. (2011). Observed score linear equating with covariates. *Journal of Educational Measurement*, 48(4), 419-440
- Boone, E. L, Ye, K., & Smith, E. P. (2009). Using data augmentation via the Gibbs Sampler to incorporate missing covariate structure in linear models for ecological assessments. *Environ Ecol Stat*, 16, 75-87.
- Boone, E L. Merrick, J.R.W., & Krachey, M. J. (2012). A Hellinger distance approach to MCMC diagnostics, *Journal of Statistical Computation and Simulation*, 84, 833-849. DOI:10.1080 / 00949655.2012.729588
- Can, A. (2008). *Bayesçi kestirim teknikleri ile hedef takibi*. Yayınlanmamış Yüksek Lisans Tezi. İstanbul Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.
- Cho, Y. (2007). *Comparision of bootstrap standart errors of equating using irt and equipercentile methods with polytomously scored items under the common item nonequivalent groups design*. Unpublished doctoral dissertation, University of Iowa, Iowa City.
- Cimşit, İ. (2013). *Değerlendirmecilerin iki ve ikiden fazla olduğu durumlarda uyum için bir bayesci yaklaşım*. Yüksek Lisans Tezi. Ondokuz Mayıs Üniversitesi, Fen Bilimleri Enstitüsü, Samsun.
- Cook, L. L., & Petersen, N. S. (1987). Problems related to the use of conventional and item response theory equating methods in less than optimal circumstances. *Applied Psychological Measurement*, 11, 225–244.

- Cook, L. L., Eignor, D. R., & Schmitt, A. P. (1990). Equating achievement tests using samples matched on ability (College Board Report No. 90-2). New York, NY: College Entrance Examination Board.
- Crocker, L., & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. New York: Holt, Rinehart and Winston.
- De Iorio, M., Müller, P., Rosner, G., L., & MacEachern, S. N. (2004) An ANOVA model for dependent random measures, *Journal of the American Statistical Association*, 99:465, 205-215, DOI: 10.1198 / 016214504000000205
- Ding, Y. (2016). How do students' mathematics self-efficacy, mathematics self-concept and mathematics anxiety influence mathematical literacy? — *A Comparison between Shanghai-China and Sweden in PISA 2012*. University of Gothenburg, Faculty of Education.
- Dorans, J. N., & Holland, P. W. (2000). Population invariance and the equitability of tests: Basic theory and the linear case. *Journal of Educational Measurement*, 37, 281-306.
- Dorans, N. J., & Walker, M. E. (2007). Sizing up linkage. In Dorans, N. J., Pommerich, M., & Holland, P. W. (2007). *Linking and aligning scores and scales* (pp. 179-199). New York: Springer.
- Dorans, N. J., Moses, T. P., & Eignor, D. R. (2010). principles and practices of test score equating. Research Report. ETS RR-10-29
- Ekici, O. (2005). *Bayesyen regresyon ve winBUGS ile bir uygulama*. Yüksek Lisans Tezi, İstanbul Üniversitesi, Sosyal Bilimler Enstitüsü, İstanbul.
- Embretson, S. (1996) The New Rules of Measurement. *Psychological Assessment*, 8 (4), 341-349.
- Embretson, S. (1997) Multicomponent Response Models. In W. van der Linden and R. Hambleton (Eds), *Handbook of Modern Item Response Theory* (p. 305-321). Springer-Verlag: New York. Embretson, S. E. & Reise, S. P. (2000). *Item Response Theory for Psychologists*. London: Lawrence Elbaum Associates, Publishers.
- Haebara, T. (1980). Equating logistic ability scales by a weighted least squares method. *Japanese Psychological Research*, 22(3), 144-149.

- Hall, C. W., & Hoff, C. (1988). Gender differences in mathematical performance. *Educational Studies in Mathematics* 19 (1988) 395-401.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer-Nijhoff Publishing.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.
- Hanson, B. A., & Beguin, A. A. (2002). Obtaining a common scale for IRT item parameters using separate versus concurrent estimation in the common item nonequivalent groups equating design. *Applied Psychological Measurement*, 26(1), 3-24.
- Holland, P. W. (2007). A framework and history for score linking. In Dorans, N.J., Pommerich, M. & Holland, P. W. (2007). *Linking and aligning scores and scales* (pp. 5-30). New York: Springer.
- Holland, P. W., & Dorans, N. J. (2006). Linking and equating. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed., pp. 187-220). Westport, CT: Praeger.
- Huh, N., & Lee, W-C. (2009). The effect of different factors on group invariance in a concordance context with a single group design, Center for Advanced Studies in Measurement and Assessment (CASMA)
- Gasım, N. (2013). *Bayesyen model ile doğrusal regresyon modellerinin karşılaştırılması üzerine bir uygulama*. Yüksek Lisans Tezi. Dokuz Eylül Üniversitesi, Sosyal Bilimler Enstitüsü, İzmir.
- Gonzalez, J. (2014). SNSequate: standard and nonstandard statistical models and methods for test equating. *Journal of Statistical Software*, 59(7)
- González, J., Barrientos, A. F., & Quintana, F. A. (2015a). Bayesian nonparametric estimation of test equating functions with covariates. *Computational Statistics and Data Analysis* 89 (2015) 222–244.
- González, J., Barrientos, A. F., & Quintana, F. A. (2015b). A dependent Bayesian nonparametric model for test equating. *In Quantitative Psychology Research* (pp.213-226). Springer Cham Heidelberg New York Dordrecht London.

- González J., & Wiberg M. (2017) Recent Developments in Equating. *In: Applying Test Equating Methods. Methodology of Educational Measurement and Assessment (pp.157-178)* Springer, Cham
- Kaplan D., & Depaoli, S. (2012). *Handbook of structural equation modeling*. Edited by Rick H. Hoyle. The Guilford Press.
- Kaplan, D., & Deapaoli, S.(2013). Bayesian statistical estimation. In T. D. Little (ed.) *The Oxford Handbook of Quantitative Methods*. Vol. 2 Statistical Analysis (pp. 407-453). New York: Oxford University.
- Kara, Y. (2018). Grup İçi Varyanslar Heterojen Olduğunda Çok Düzeyli Madde Tepki Modelinin Bayes Yaklaşımı ile Modellenmesi. Yayınlanmamış Doktora Tezi, Hacettepe Üniversitesi, Ankara.
- Karabatsos, G., & Walker, S. G. (2009). *A bayesian nonparametric approach to test equating*. Psychometrika. DOI: 10.1007 / S11336-008-9096-6
- Karabey. U. (2007). *Yol kazalarının bayesçi yaklaşımla analizi*. Yüksek Lisans Tezi.Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.
- Karasar, N. (2006). *Bilimsel Araştırma Yöntemi (16. Baskı)*. Ankara: Nobel Akademik Yayıncılık.
- Kim, S., & Cohen, A.S. (1992). Effects of linking methods on detection of DIF. *Applied Psychological Measurement*, 29(1), 51-56.
- Kim, S-H., & Cohen, A. S. (1998). A comparison of linking and concurrent calibration under item response theory. *Applied Psychological Measurment*, Vol. 22, No.2, 131-143.
- Kim, S., Livingston, S. A., & Lewis, C. (2008). Investigating the effectiveness of collateral information on small-sample *Equating*. Educational Testing Service, Princeton NJ
- Kim, S., Livingston, S. A., & Lewis, C. (2009). Effectiveness of collateral information for improving equating in small samples. Educational Testing Service, Princeton NJ
- Kim, S., Livingston, S. A., & Lewis, C. (2011). Collateral Information for Equating in Small Samples: A Preliminary Investigation. *Applied Measurement In Education*, 24: 302–323

- Kim, S. (2015). Linking with Planned Missing Data: Concurrent Calibration with Multiple Imputation. (Unpublished doctoral dissertation). University of Kansas
- Koğar, H. (2015). PISA 2012 Matematik okuryazarlığını etkileyen faktörlerin aracılık modeli ile incelenmesi. *Eğitim ve Bilim Cilt 40 (2015) Sayı 179* 45-55.
- Kolen, M. J. (1988). An NCME instructional module on traditional equating methodology. *Educational Measurement: Issues and Practice*, 7, 29-36.
- Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices* (3nd. ed.). New York: Springer.
- Kruschke, J. K. (2010a). What to believe: Bayesian methods for data analysis. *Trends in Cognitive Sciences* Vol.14 No.7
- Kruschke, J.K. (2010b). Bayesian data analysis. Wiley Interdisciplinary Reviews; *Cognitive Science*, 1(5), 658-676, <http://dx.doi.org/10.1002/wcs.72>
- Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General*, Vol. 142, No. 2, 573–603
- Kruschke, J. K., Aguinis, H., & Joo, H. (2012). The Time has come: bayesian methods for data analysis in the organizational sciences. *Organizational Research Methods*, 15(4) 722-752
- Kruschke, J. K. (2015). Doing Bayesian Data Analysis: Second Editin: A Tutorial with R, JAGS, and Stan. Waltham, MA: Academic Press / Elsevier.
- Lee, W., & Ban, J. (2010). A comparison of IRT linking procedures. *Applied Measurement in Education*, 23, 23–48.
- Lee, A. H., & Boone, E. L. (2011). A frequentist assessment of bayesian inclusion probabilities for screening predictors. *Journal of Statistical Computation and Simulation*, 81(9), 1111-1119.
- Li, D., Jiang, Y., & von Davier, A. A. (2012). The accuracy and consistency of series of IRT true score equatings. *Journal of Educational Measurment*, 49(2), 167-189.
- Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: A meta-analysis. *Psychological Bulletin*, 136, 1123–1135. doi:10.1037/a0021276

- Liou, M., Cheng, P. E., & Li, M. (2001). Estimating comparable scores using surrogate variables. *Applied Psychological Measurement*, 25, 197-207.
- Liou, M. (1998). Establishing score comparability in heterogeneous populations. *Statistica Sinica*, 8, 669-690.
- Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research*, 32(1), 53-76.
- Liu, J., Feigenbaum, M., & Dorans, N. L. (2005). SAT Reasoning Test™ to the SAT I: reasoning test across gender groups.
- Liu, Y., Schulz, E. M., & Yu, L. (2007). Standard error estimation of 3PL IRT true score equating with an MCMC method. *Journal of Educational and Behavioral Statistics*. DOI: 10.3102 / 1076998607306076.
- Livingston, S. A. (2004). *Equating test scores (Without IRT)*. Educational Testing Service.
- Livingston, S. A., & Lewis, C. (2009). Small-Sample Equating with prior information. (ETS Research Rep. No. RR-09-25). Princeton, NJ: ETS.
- Longford, N. T. (2015). Equating without an anchor for nonequivalent groups of examinees. *Journal of Educational and Behavioral Statistics*, Vol. 40, No. 3, pp. 227–253
- Lord, F. M., & Wingersky, M. S. (1984). Comparison of IRT true-score and equipercentile observedscore “equatings.” *Applied Psychological Measurement*, 8, 452-461.
- Lord, F. M. (1980). *Applications of Item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum.
- Lyren, P.-E., & Hambleton, R. K. (2011) Consequences of violated equating assumptions under the equivalent groups design, *International Journal of Testing*, 11:4,308-323, DOI: 10.1080 / 15305058.2011.585535
- MacEachern, S.N., (1999). Dependent nonparametric processes. In: ASA Proceedings of the Section on Bayesian Statistical Science. pp. 50–55.

- MacEachern, S.N., (2000). Dependent Dirichlet processes. Tech. rep.. Department of Statistics, The Ohio State University.
- MacEachern, S.N. (2001). Decision theoretic aspects of dependent nonparametric processes. In *Bayesian Methods with Applications to Science, Policy and Official Statistics* (Edited by E. George), 551–560. Eurostat.
- Martin, M. O., Mullis, I. V. S., Foy, P., & Stanco, G. M. (2012). *TIMSS 2011 international results in science*. International Study Center, Boston College, MA, USA.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525–543.
- Mittelhaeuser, M.-A., Beguin, A. A., & Sijtsma, K. (2011). Comparing the effectiveness of different linking design: the internal anchor versus the external anchor and pre-test data. *Measurement and Research Department Reports*, 1. Cito Arnhem.
- Moses, T. (2006). Using the kernel method of test equating for estimating the standard errors of population invariance measures, Research Report.
- Moses, T., Deng, W., & Zhang, Y.-L. (2010). The use of two anchors in nonequivalent groups with anchor test (NEAT) equating. ETS Princeton, New Jersey, research report. RR-10-23
- Moses, T., Deng, W., & Zhang, Y.-L. (2011). Two approaches for using multiple anchors in neat equating: A description and demonstration. *Applied Psychological Measurement*, 35, 362–379. doi: 10.1177 / 0146621611405510
- Müller, P., & Quintana, F. A. (2004). Nonparametric bayesian data analysis. *Statistical Science*, 19(1), 95-110
- Narrete, C., & Quintana, F. A., & Müller, P. (2008). Some Issues on nonparametric bayesian modeling using species sampling models. *Statistical Modelling International Journal*, 8(1), 3-21.
- Ogasawara, H. (2001). Item response theory true score equating and their standard errors. *Journal of Educational Behavioral Statistics*, 26(1), 31-50.

- Oh, H. J., Guo, H., & Walker, M. E. (2009). Improved reliability estimates for small samples using empirical Bayes techniques. ETS Princeton, New Jersey, research report, RR-09-46.
- Orbanz, P., & Teh, Y. W. (2010). Bayesian nonparametric models. *In Encyclopedia of Machine Learning*.
- Öztürk, O. (2011). *Bayesian semiparametric models for nonignorable missing data mechanisms in logistic regression*, Degree of Master. Middle of Technical University. The Graduate School of Natural and Applied Sciences, Ankara.
- Petersen, N. S., Cook, L. L., & Stocking, M. L. (1983). IRT versus conventional equating methods: A comparative study of scale stability. *Journal of Educational Statistics*, 8(2), 137-156.
- Petrone, S. (1999a). Bayesian density estimation using Bernstein polynomials. *The Canadian Journal of Statistics* 27 105–126.
- Petrone, S. (1999b). Random Bernstein polynomials. *Scandinavian Journal of Statistics* 26, 373–393.
- Petrone, S., & Wasserman, L. (2002). Consistency of Bernstein polynomial posteriors. *J. R. Statist. Soc. B, Part 1*, pp. 79-100.
- Powers, S., Turhan, A., & Binici, S. (2012). Population invariance of vertical scaling results. National Council of Measurement in Education Vancouver, BC
- Raftery, A.E. 1995. Bayesian model selection in social research. *Sociological methodology*, Marsden, P.V. Cambridge, Mass., Blackwells: 111-196.
- Rounder, J. N., Morey, R. D., Speckmanc, P.L., & Province, M. (2012). Default bayes factors for ANOVA designs. *Journal of Mathematical Psychology*. 56 (2012) 356–374
- Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*, 79(387), 516–524.
- Thien, L., R., & Darmawan, I. G. N. (2016). Factors Associated with Malaysian Mathematics Performance in PISA 2012. In What can PISA 2012 data tell us? Performance and challenges in five participating Southeast Asian countries. Sense Publishers, pp.81-105,

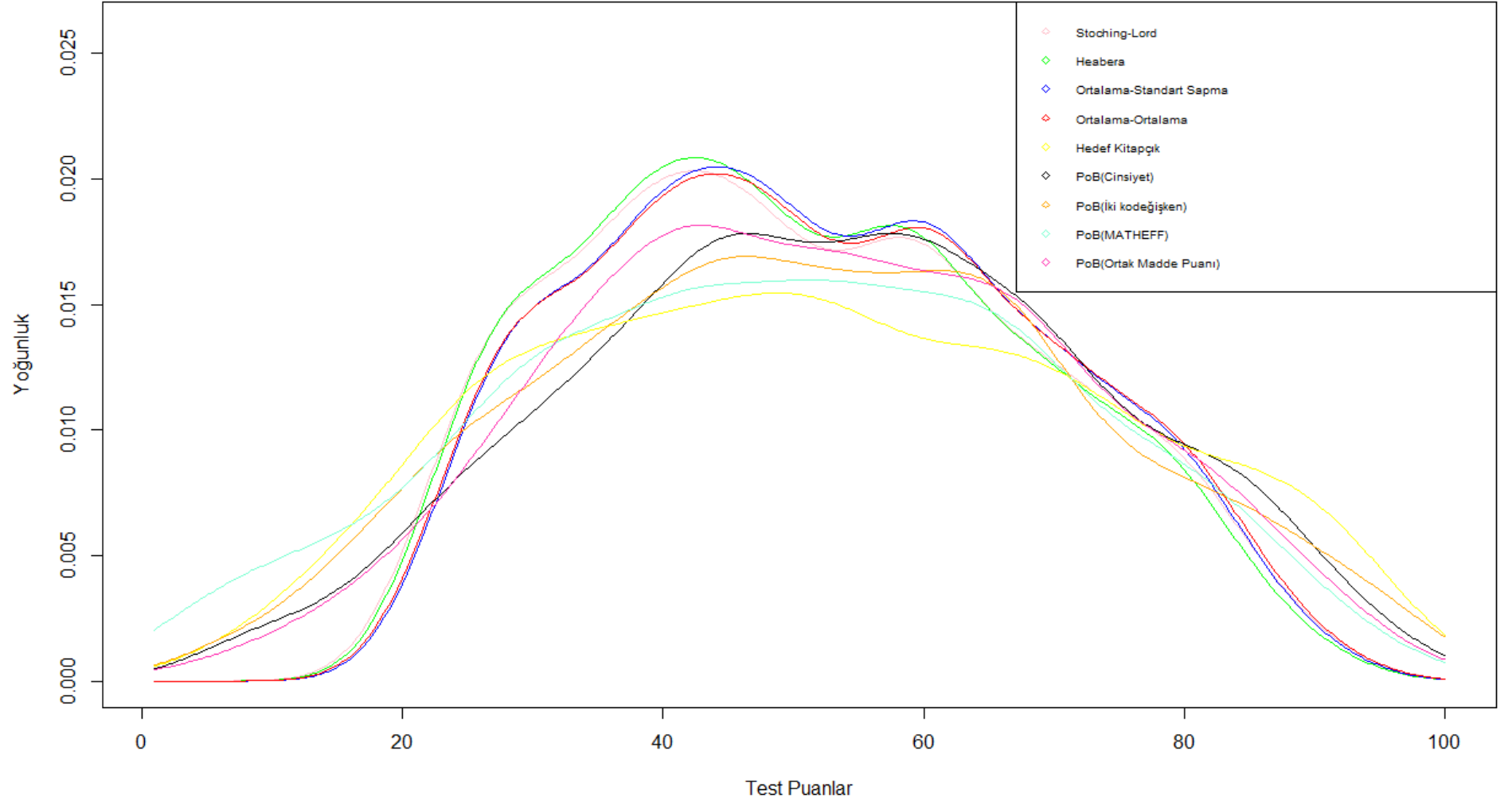
- Tian, F. (2011). A comparison of equating / linking using the Stocking-Lord method and concurrent calibration with mixed-format tests in the non-equivalent groups common-item design under IRT . (Unpublished doctoral dissertation). Boston College, Chestnut Hill, MA
- Sethuraman, J., (1994). A constructive definition of Dirichlet priors. *Statist. Sinica* 4, 639–650.
- Shah, A., & Ghahramani, Z. (2013). *Determinantal clustering process- A nonparametric bayesian approach to kernel based semi-supervised clustering*
- Sinharay, S., & Holland, P. W. (2006). Choice of anchor test in equating ETS Research. Princeton, NJ: Educational Testing Service
- StataCorp. (2015). Stata: Release 14. Statistical Software. College Station, TX: StataCorp LLC. <https://www.stata.com/manuals14/bayes.pdf>.
- Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement*, 7(2), 201-210. doi: 10.1177 / 014662168300700208
- van de Schoot R., Kaplan D., Denissen J., Asendorpf J. B., Neyer F. J., & van Aken M. A. G. (2013). A gentle introduction to bayesian analysis: applications to developmental research. *Child Development*, Volume 00, Number 0, Pages 1–19.
- van de Schoot, R., & Depaoli, S. (2014). Bayesian analyses: Where to start and what to report. *EHP*, 16(2), 75-84.
- von Davier, A. A., Holland, P. W., & Thayer, D. T. (2003). *Statistics for social science and public Policy*. Springer.
- van der Linden, W. J. (2010). Local observed-score equating. In A. A. von Davier (Ed.), *Statistical models for equating, scaling, and linking* (pp. 201-223). New York, NY: Springer.
- Wallin, G., & Wiberg, M. (2017). Non-equivalent groups with covariates design using propensity scores for kernel equating. In L. A. van der Ark, M. Wiberg, S. A. Culpepper, J. A. Douglas, & W.-C. Wang (Eds.), *Quantitative*

psychology – 81st annual meeting of the psychometric society, Asheville, North Carolina, 2016. New York: Springer

- Wang, T. (2006). Standard errors of equating for equipercentile equating with log-linear pre-smoothing using the delta method (CASMA Research Report, No. 14). Center for Advanced Studies in Measurement and Assessment, Iowa
- Weeks, J. P. (2010). plink: an r package for linking mixed-format tests using IRT-based methods. *Journal of Statistical Software. Volume 35, Issue 12.*
- Wei, H. (2010). Impact of Non-representative anchor items on scale stability. Paper presented at the annual conference of the National Council on Measurement in Education, Denver, CO.
- Wiberg, M. (2013). Alina A. von Davier (Ed.) (2011) Statistical models for test equating, scaling, and linking. *Psychometrika*, 78(1), 185-187.
- Wiberg, M. (2015). A note on equating test scores with covariates. Festschrift in Honor of Hans Nyquist on the Occasion of His 65th Birthday, 96-99. Department of Statistics, Stockholm University, Stockholm, Sweden, Edited by Ellinor Frackle-Fornius. ISBN:978-91-87355-19-6
- Wiberg, M., & Branberg, K. (2015). Kernel equating under the non-equivalent groups with covariates design. *Applied Psychological Measurement, Vol. 39(5) 349–361*
- Wiberg, M. (2016). Alternative linear item response theory observed-score equating methods. *Applied Psychological Measurement, Vol. 40(3) 180–199*
- Wiberg, M., & Gonzalez, J (2016). Statistical assessment of estimated transformations in observed-score equating. *Journal of Educational Measurement. Vol. 53, No. 1, pp. 106–125*
- Wiberg, M., & von Davier, A. A. (2017). Examining the impact of covariates on anchor tests to ascertain quality over time in a college admissions test, *International Journal of Testing, 17:2, 105-126, DOI: 10.1080 / 15305058.2016.1277357*
- Wright, N. K., & Dorans, N. J. (1993). Using the selection variable for matching or equating (Research Report No. RR-93–04). Princeton, NJ: Educational Testing Service

- Woldbeck, T. (1998). Basic concepts in modern methods of test equating. Paper presented at the annual meeting of the Southwest Psychological Association, New Orleans.
- Xin, T., & Zhang, J. (2015). Local Equating of Cognitively Diagnostic Modeled Observed Scores. *Applied Psychological Measurement*, Vol. 39(1) 44–61
- Yang, W. L. (1997). *The Effects of Content Homogeneity and Equating Method on the Accuracy of Common Item Test Equating*. Unpublished doctoral dissertation, Michigan State University, Michigan.
- Yıldırım, H. H., Yıldırım, S., Yetişir , M. İ., & Ceylan, E. (2013). *PISA 2012 ulusal ön raporu*. MEB Yenilik ve Eğitim Teknolojileri Genel Müdürlüğü (YeğiTek), Ankara.

EK-A: Yöntemlere Göre Kestirilen Puanlar



EK-B: Her Yöntemden Elde Edilen Eşitlenmiş Puanların Standart Sapması Ve Ortalaması

	Ortalama- Ortalama	Ortalama- Standart Sapma	Heabera	Staoocking- Lord	PoB Sadece Cinsiyet	PoB Sadece MATHEFF	PoB İki Değişken	6. Kitapçıkta ki Ham Puan	PoB Ortak Madde Puanları
Ortalama	52,1918	52,10785	50,87776	51,14383	53,4163	49,52093	52,0859	52,26907	52.55066
Standart Sapma	17,1692	16,97196	16,99018	17,35477	19,89888	21,48748	20,89495	22,05677	19,50667
Birim standart sapmada ortalamanı n değişimi	3,03985	3,070231	2,99454	2,946961	2,684387	2,304641	2,492751	2, 369751	2,693984

**EK-C: Ülkelere Ait Ayrı Kalibrasyon Kestirimleri için Kullanılan
PARSCALE kodu**

```
>FILE DFNAME='italya5.DAT',  
    SAV;  
>SAVE PAR='a.par',  
score='a.sco';  
>INPUT NID=3,  
NTOTAL=36,  
NTEST=1;  
(3A1,1X,36A1)  
>TEST1 TNAME=bayes,ITEM=(1(1)36),NBLOCK=1;  
>BLOCK1 BNAME=bayes,NITEMS=36,NCAT=2,  
ORIGINAL=(0,1), GUESSING=(2,ESTIMATE);  
>CALIB LOGISTIC,DIAGNOSIS=0,  
    CYCLES=(100,5,5,5,5),CRIT=0.001,  
    TPRIOR,GPRIOR,SPRIOR,RIDGE;  
>SCORE EAP;
```

EK-Ç: MCMC Kodu (MATHEFF için)

```
setwd("C: / ")
source('C: / testequatefn.R')
library(LaplacesDemon)
data1 <- read.table( 'C: / .....', header=TRUE, sep='\\t')
z1a <- cbind( data1$MATHEFF )
x1a <- data1$Score
n1a <- length( x1a)

nMCMC <- 150000

j_max <- 10

p1 <- ncol(z1a)*j_max
m0 <- rep( 0, p1 )
S0 <- diag( rep(2.25, p1 ))

beta1 <- 5
lambda1 <- 25
v1 <- p1+2
Phi1 <- diag( rep(1, p1 ))
mu1 <- rep( 0, p1 )
S1 <- diag( rep(1, p1 ))
gamma_j <- rep( 0.2, p1 )
v_j <- rep(0.5, j_max)
w_j <- w_j_fn( v_j )
k1 <- 6
posterior1( x1a, z1a, n1a, w_j, beta1, v_j, v1, Phi1, k1, lambda1, gamma_j, S1, mu1,
m0, S0 )

v_j_step1 <- 0.01
gamma_j_step1 <- 0.01

v_j_out1 <- matrix( 0, ncol = j_max , nrow = nMCMC )
w_j_out1 <- matrix( 0, ncol = j_max, nrow = nMCMC )
gamma_j_out1 <- matrix( 0, ncol = ncol(z1a)*j_max, nrow = nMCMC )
k1_out1 <- matrix( 0, ncol = 1, nrow = nMCMC )
x1new_score <- matrix( 0, ncol = nrow( z1a ), nrow = nMCMC )
posthold1 <- posterior1( x1a, z1a, n1a, w_j, beta1, v_j, v1, Phi1, k1, lambda1,
gamma_j, S1, mu1, m0, S0 )

for( iter1 in 1:nMCMC){
  v_j_new <- v_j + v_j_step1*rnorm(length(v_j))
  w_j_new <- w_j_fn( v_j_new)
```

```

postnew1 <- posterior1( x1a, z1a, n1a, w_j_new, beta1, v_j_new, v1, Phi1, k1,
lambda1, gamma_j, S1, mu1, m0, S0 )
Utest1 <- log( runif(1, 0, 1) )
if( (postnew1 - posthold1) > Utest1){
  v_j <- v_j_new
  w_j <- w_j_new
  posthold1 <- postnew1
}

gamma_j_new <- gamma_j + gamma_j_step1*rnorm(length(gamma_j))
postnew1 <- posterior1( x1a, z1a, n1a, w_j, beta1, v_j, v1, Phi1, k1, lambda1,
gamma_j_new, S1, mu1, m0, S0 )
Utest1 <- log( runif(1, 0, 1) )
if( (postnew1 - posthold1) > Utest1){
  gamma_j <- gamma_j_new
  posthold1 <- postnew1
}

k1_new <- k1 + ifelse( runif(1) < 0.5, 1, -1 )
if( k1_new > 0 ){
  postnew1 <- posterior1( x1a, z1a, n1a, w_j, beta1, v_j, v1, Phi1, k1_new,
lambda1, gamma_j, S1, mu1, m0, S0 )
  Utest1 <- log( runif(1, 0, 1) )
  if( (postnew1 - posthold1) > Utest1){
    k1 <- k1_new
    posthold1 <- postnew1
  }
}
gamma_jrow <- matrix( gamma_j, ncol = ncol(z1a), nrow = j_max, byrow = TRUE
)
for( student1 in 1:nrow(z1a) ){
  x1samp_index <- sample( 1:j_max, size=1, prob = w_j )
  a1_temp <- ceiling( k1*theta_j_fn( z1a[student1,], gamma_jrow[x1samp_index,] )
)
  temp_score <- rbeta( 1, a1_temp, k1 - a1_temp + 1 )
  x1new_score[ iter1, student1 ] <- temp_score
}

v_j_out1[iter1, ] <- v_j
w_j_out1[iter1, ] <- w_j
gamma_j_out1[iter1, ] <- gamma_j
k1_out1[iter1, ] <- k1
if( iter1 %% 100 == 0){
25=0mod(25) /
  print(iter1)
}
}

```

EK-D: Eşitlenmiş Puanların Elde Edilmesi (Cinsiyet için)

```
studentprofile1 <- matrix( 0, ncol = ..., nrow = n1a )
for(student1 in 1:n1a){

  z1a1 <- matrix( ....., ncol = 2 )
  z2a1 <- matrix( ....., ncol = 2 )
  x1a1 <- seq( 0, 1, by = 0.01)
  scoreA <- matrix( 0, ncol = length(.....), nrow = nrow(gamma_j_outA) )
  scoreB <- matrix( 0, ncol = length(.....), nrow = nrow(gamma_j_outB) )

  for(i in 1:nrow(.....)){
    gamma_jrowA <- matrix( gamma_j_outA[i,], ncol = ....., nrow = ....., byrow =
TRUE )
    scoreA[i,] <- s_z_fn( x1a1, w_j_outA[i,], k1_outA[i], z1a1, gamma_jrowA )
    gamma_jrowB <- matrix( gamma_j_outB[i,], ncol = ....., nrow = ....., byrow =
TRUE )
    scoreB[i,] <- s_z_fn( x1a1, w_j_outB[i,], k1_outB[i], z2a1, gamma_jrowB )
  }

  cdfA <- cumsum( apply( scoreA,.....)*0.01 ) / sum( apply( scoreA, .....)*0.01 )
  cdfB <- cumsum( apply( scoreB, 2, ..... )*0.01 ) / sum( apply( scoreB, ... )*0.01 )

  x1atest <- x1a[student1]
  x1acdf <- cdfA[floor(x1atest*100)]

  x1ind <- ifelse( x1acdf < cdfB, 1, 0 )
  x1acdfB <- min((0:100)[x1ind == 1])
  studentprofile1[student1,] <- c( .....x1acdfB, ..... )
}
```

**EK-E: Modelde İki Ortak Değişken Kullanıldığında Elde Edilen
Parametreler**

	5.Kitapçığa ait parametreler			6.kitapçığa ait parametreler		
Parametre	Ortalama	Standart Sapma	(0,025 0,975)	Ortalama	Standart Sapma	(0,025 0,975)
k	2.992453	0.123274	3 3	4.922147	0.513829	4 5
w_1	0.8976394	0.1106958	0.5606 0.9758	0.0302526	0.04064264	0.0006722403 0.1607421232
w_2	0.0135385	0.01372039	0.00037 0.04569	0.008979508	0.05031254	0,0000717033 0,01328980
w_3	0.04807361	0.0940087	0.00053359 0.339224	0.02370609	0.02601168	0.0006172119 0.0751133016
w_4	0.01473744	0.02411569	0.0002156 0.087819	0.02584554	0.02795822	0.0006665706 0.0793513462
w_5	0.007404908	0.00899944	0.000138 0.029769	0.02521606	0.02336812	0.0007483099 0.0758558661
w_6	0.007120595	0.01649756	0,00006674 0,043754	0.02303124	0.01881318	0.0006163447 0.0708729870
w_7	0.00403168	0.004687399	0,0000688904 0,01719980	0.8466467	0.1084118	0.6856078 0.8996638
w_8	0.002816302	0.00504885	0,0000367688 0,0168926	0.002579654	0.00294082 2	0.000686212 0,01064023
w_9	0.001276043	0.001690909	0,00002369063 0,005610182	0.007177533	0.00788916 9	0.0001438858 0.0286006978
w_{10}	0.0009064241	0.00109425	0,0000114 0,0038806	0.002693434	0.00434047 6	0.0000333623 0.0147338591
γ_1	0.005410013	0.05070485	-0.055757001 0.179012063	-0.3935725	0.07754092	-0.564485313 -0.276349999
γ_2	-0.01023734	0.008422349	-0.011097795 - 0.009734963	0.2213	0.1770208	-0.009833342 0.527368079
γ_3	-0.1423921	0.2906572	-0.593632076 0.510086085	-0.3524219	0.1129432	-0.588578898 -0.193334683
γ_4	0.3476062	0.185098	0.056620184 0 .673770254	-0.3490834	0.1110784	-0.547055018 -0.220795563
γ_5	-0.1202295	0.2023838	-0.521702116 0.208197987	-0.01484485	0.1215677	-0.172557712 0.225025865
γ_6	0.1234675	0.1223492	-0.030199494 0.362772702	0.1329002	0.05388825	0.052031783 0.218791778
γ_7	0.1554316	0.1388001	-0.128984335 0.415025839	0.3191902	0.1074288	0.086612258 0.536331235
γ_8	0.1776026	0.2790494	-0.069448797 0.810123649	0.900378	0.0880464	0.723576822 1.070456624
γ_9	0.3875055	0.1645709	0.153681570 0 .734508730	0.7073089	0.1382168	0.388291811 0.878828415
γ_{10}	0.5879142	0.1541645	0.305151362 0 .832271055	0.9523576	0.1849331	0.695576443 1.226742552
γ_{11}	0.6242532	0.09724444	0.441650271 0 .812074754	0.7783203	0.1180862	0.582765647 1.009610121

γ_{12}	0.0659296	0.1139994	-0.099980622 0.303571791	0.5168871	0.176268	0.257988602 0.921808131
γ_{13}	0.2573592	0.1333155	0.031007141 0 .459199560	0.07112908	0.05406791	0.010209736 0.241210677
γ_{14}	0.2237701	0.1115575	-0.009397796 0.379840064	-0.00658358	0.01765983	-0.007623856 -0.005258608
γ_{15}	-0.004437639	0.1459134	-0.308084333 0.216677013	-0.1468707	0.3000974	-0.609316085 0.217693376
γ_{16}	0.06252445	0.1054492	-0.165413300 0.222756478	-0.4702557	0.1141376	-0.660915411 -0.233472753
γ_{17}	0.04405279	0.1065333	-0.126159065 0.245569817	-0.07874709	0.09984755	-0.295507479 0.084726599
γ_{18}	0.2287154	0.1879547	-0.094911019 0.469661956	0.28266	0.09448105	0.103911635 0.462732701
γ_{19}	0.5504439	0.2227428	0.118427166 0 .916779754	0.3406615	0.1003429	0.172133263 0.537187849
γ_{20}	-0.2297639	0.2196895	-0.546214732 0.139635726	0.7437543	0.1055174	0.542053436 0.885192447

EK-F: Etik Komisyonu Onay Bildirimi



T.C.
HACETTEPE ÜNİVERSİTESİ
Rektörlük

Sayı : 35853172/

433-1711

18 Nisan 2018

EĞİTİM BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜNE

İlgi: 19.03.2018 tarih ve 775 sayılı yazınız.

Enstitünüz Eğitim Bilimleri Anabilim Dalı Eğitimde Ölçme ve Değerlendirme Bilim Dalı bütünlük doktora programı öğrencisi Meltem YURTÇU'nun Prof. Dr. Hülya KELECİOĞLU danışmanlığında yürüttüğü "Klasik, Örtük Özellikler ve Nonparametrik Bayes Eşitleme Yöntemlerinin Eşit Olmayan Gruplarda Kodeğişken Deseni ile Karşılaştırılması" başlıklı tez çalışması, Üniversitemiz Senatosu Etik Komisyonunun 03 Nisan 2018 tarihinde yapmış olduğu toplantıda incelenmiş olup, etik açıdan uygun bulunmuştur.

Bilgilerinizi ve gereğini rica ederim.

Prof. Dr. Rahime M. NOHUTCU
Rektör a.
Rektör Yardımcısı

EK-G: Etik Beyanı

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

- 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 bütününe 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.

09 / 07 / 2018



Meltem YURTÇU

EK-Ğ: Doktora Tez Çalışması Orijinallik Raporu

09 / 07 / 2018

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

Tez Başlığı: Parametrik Olmayan Bayes Yöntemiyle Ortak Değişkenlere Göre Yapılan Test Eşitlemelerinin Karşılaştırılması

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:

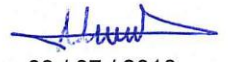
Rapor Tarihi	Sayfa Sayısı	Karakter Sayısı	Savunma Tarihi	Benzerlik Oranı	Gönderim Numarası
09 / 07 / 2018	111	160,311	01 / 06 / 2018	%6	981368700

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: Meltem YURTÇU
Öğrenci No.: N12162398
Ana Bilim Dalı: Eğitim Bilimleri
Programı: Eğitimde Ölçme ve Değerlendirme
Statüsü: ☐ Y.Lisans ☐ Doktora ☒ Bütünleşik Dr.


09 / 07 / 2018

DANIŞMAN ONAYI


UYGUNDUR.

Prof. Dr. Hülya KELECİOĞLU

EK-H: Dissertation Originality Report

09/07/2018

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

Thesis Title : The Comparison Of Test Equating With Covariates Using Bayesian Nonparametric Method

The whole thesis 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 Thesis Defense	Similarity Index	Submission ID
09 / 07 / 2018	111	160,311	01 / 06 / 2018	%6	981368700

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 Thesis Originality Reports; that according to the maximum similarity index values specified in the Guidelines, my thesis 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: Meltem YURTÇU
Student No.: N12162398
Department: Educational Science
Program: Educational Measurement and Evaluation
Status: ☐ Masters ☐ Ph.D. ☒ Integrated Ph.D.


09 / 07 / 2018

ADVISOR APPROVAL


APPROVED
Prof. Dr. Hülya KELECIOĞLU

EK-I: Yayınlama 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 etmeniz ve kütüphane bu talebinizi yerine getirse bile, teziniz arama motorlarının ön belleklerinde kalmaya devam edebilecektir)

☒ **Tezimin / Raporumun 05.05.2021 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:**

.....
.....
.....

09 / 07 / 2018



Meltem YURTÇU

