

**KARADENİZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

İSTATİSTİK VE BİLGİSAYAR BİLİMLERİ ANABİLİM DALI

**FARKLI ÖLÇEK TİPLERİNDE DEĞİŞEN MADDE FONKSİYONUNUN
BELİRLENMESİ VE YÖNTEMLERİN KARŞILAŞTIRILMASI ÜZERİNE
BİR ÇALIŞMA**

DOKTORA TEZİ

Burcu HASANÇEBİ

**TEMMUZ 2021
TRABZON**



KARADENİZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

İSTATİSTİK VE BİLGİSAYAR BİLİMLERİ ANABİLİM DALI

**FARKLI ÖLÇEK TİPLERİNDE DEĞİŞEN MADDE FONKSİYONUNUN
BELİRLENMESİ VE YÖNTEMLERİN KARŞILAŞTIRILMASI ÜZERİNE
BİR ÇALIŞMA**

Burcu HASANÇEBİ

Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsünde
DOKTOR (İSTATİSTİK)
Unvanı Verilmesi İçin Kabul Edilen Tezdir.

Tezin Enstitüye Verildiği Tarih : 23 / 06 / 2021

Tezin Savunma Tarihi : 28 / 07 / 2021

Tez Danışmanı : Prof. Dr. Zafer KÜÇÜK

İkinci Danışman Prof. Dr. Yüksel TERZİ

Trabzon 2021

ÖNSÖZ

Doktora tezi olarak sunulan bu çalışmada, öncelikle tarafımızca geliştirilen İstatistiğe Giriş Okuryazarlık Testi'nin kullanılabilirliği araştırılmış, alınan olumlu sonuçların ardından Karadeniz Teknik Üniversitesi ve Ondokuz Mayıs Üniversitesi bünyesinde İstatistik dersi almış öğrenciler belirlenerek geliştirilen ölçek bu öğrencilere uygulanmıştır. Buna ek olarak bu çalışmada, Karadeniz Teknik Üniversitesi Yabancı Uyruklu Öğrenci Sınavı (KTÜYÖS-2019) verileri de kullanılmıştır. Bu verilerin kullanılmasında amaç, ilgili ölçeklerin yanlılıklarına ilişkin Değişen Madde Fonksiyonu içerip içermediklerini belirlemektir. Ayrıca, model uyumları, madde ve kişi tahminleri de elde edilmiştir. Değişen Madde Fonksiyonunun elde edilmesinde Klasik Test Kuramına ve Madde Tepki Kuramına dayalı yöntemler kullanılmış ve bu yöntemlerin karşılaştırmaları yapılmıştır.

Tez danışmanlığımı üstlenerek, konu seçimi ve çalışmaların yürütülmesi sırasında bilgisini ve tecrübesini benimle paylaşan ve yardımlarını esirgemeyen değerli hocalarım Sayın Prof. Dr. Yüksel TERZİ'ye ve Sayın Prof. Dr. Zafer KÜÇÜK'e şükranlarımı sunarım. Ayrıca yapıcı eleştirileri ve yöntem önerileri ile tezime büyük katkıda bulunan hocam Sayın Prof. Dr. Türkan ERBAY DALKILIÇ'a, her zaman beni dinleyip çözüme ulaşmamda yardımcı olan hocam Sayın Prof. Dr. Tülay KESEMEN'e teşekkürü bir borç bilirim. Bu çalışmada, her soru ve sorunumda yanımda olan hocam Sayın Dr. Öğr. Üyesi Tolga BERBER'e ve yardımını göz ardı edemeyeceğim doktora dönem arkadaşım Dr. Öğr. Üyesi Erdinç KARAKULLUKÇU'ya da çok teşekkür ederim.

Son olarak, hayatımın her döneminde arkamda olan aileme, her daim beni dinleyen, moral veren ve yapabileceğimi hatırlatan canım arkadaşlarıma teşekkür ederim.

Burcu HASANÇEBİ
Trabzon 2021

TEZ ETİK BEYANNAMESİ

Doktora Tezi olarak sunduđum “Farklı Ölçek Tiplerinde Deđişen Madde Fonksiyonunun Belirlenmesi ve Yöntemlerin Karşılaştırılması Üzerine Bir Çalışma” başlıklı bu çalışmayı baştan sona kadar danışmanlarım Prof. Dr. Yüksel TERZİ ve Prof. Dr. Zafer KÜÇÜK’ün sorumluluğunda tamamladığımı, verileri/örnekleri kendim topladığımı, deneyleri/analizleri ilgili laboratuvarlarda yaptığımı/yaptırdığımı, başka kaynaklardan aldığım bilgileri metinde ve kaynakçada eksiksiz olarak gösterdiğimi, çalışma sürecinde bilimsel araştırma ve etik kurallara uygun olarak davrandığımı ve aksinin ortaya çıkması durumunda her türlü yasal sonucu kabul ettiğimi beyan ederim. 28/07/2021

Burcu HASANÇEBİ

İÇİNDEKİLER

	<u>Sayfa No</u>
ÖNSÖZ.....	III
TEZ ETİK BEYANNAMESİ.....	IV
İÇİNDEKİLER.....	V
ÖZET	VIII
SUMMARY	IX
ŞEKİLLER DİZİNİ	X
TABLolar DİZİNİ.....	XII
KISALTMALAR VE SEMBOLLER DİZİNİ	XVI
1. GENEL BİLGİLER.....	1
1.1. Giriş.....	1
1.2. Tezin Amacı ve Yapısı	11
2. YAPILAN ÇALIŞMALAR	12
2.1. Madde Analizi	12
2.2. Yetenek Parametresinin Tahmini	14
2.3. Lojistik Regresyon	16
2.3.1. Regresyon Katsayılarının Tahmini.....	17
2.3.1.1. En Çok Olabilirlik Tahmin Yöntemi.....	17
2.3.1.2. Bayes Yöntemi	18
2.4. Güvenilirlik ve Geçerlilik.....	18
2.4.1. Güvenilirlik	19
2.4.1.1. Test- Tekrar Test Yöntemi	20
2.4.1.2. Paralel Formlar Yöntemi	20
2.4.1.3. Test Yarılama Yöntemi	20
2.4.1.4. İç Tutarlılık Katsayısı.....	21
2.4.1.5. Cronbach Alfa Katsayısı	22
2.4.2. Geçerlilik.....	23
2.4.2.1. Kapsam Geçerliliği.....	24
2.4.2.2. Ölçüt Geçerliliği.....	24
2.4.2.3. Yapı Geçerliliği	25

2.5.	Klasik Test Kuramı	25
2.6.	Madde Tepki Kuramı	27
2.6.1.	Madde Tepki Kuramı Varsayımları	30
2.6.2.	Madde Karakteristik Eğrisi	31
2.6.3.	Madde Tepki Kuramı Modelleri.....	35
2.6.3.1.	1 Parametrelı Lojistik Model.....	35
2.6.3.2.	2 Parametrelı Lojistik Model.....	35
2.6.3.3.	3 Parametrelı Lojistik Model.....	36
2.6.3.4.	4 Parametrelı Lojistik Model.....	37
2.6.3.5.	Aşamalı Tepki Modeli.....	38
2.6.3.6.	Rasch Model.....	39
2.6.3.7.	Kısmi Kredi Modeli.....	40
2.6.3.8.	Çok Yüzeyle Rasch Model	42
2.7.	Uyum Kriterleri	43
2.7.1.	Sapma Bilgi Kriteri	44
2.7.2.	Akaike Bilgi Kriteri.....	45
2.7.3.	Bayesci Bilgi Kriteri.....	46
2.7.4.	LogLikelihood Testi	47
2.8.	Madde Yanlılıđı.....	48
2.8.1.	Breslow-Day İstatistiđi.....	49
2.8.2.	Lojistik Regresyon Yöntemi	51
2.8.3.	Lord'un χ^2 İstatistiđi.....	52
2.8.4.	Raju'nun Alan Ölçümleri	53
2.8.5.	Olabilirlik Oranı	54
2.9.	Etkili Örneklem Büyüklüğü	55
2.10.	Test Uzunluđu	56
3.	BULGULAR VE İRDELEME.....	57
3.1.	İstatistiđe Giriş Okuryazarlık Testi	57
3.1.1.	İstatistiđe Giriş Okuryazarlık Testi İçin Cinsiyet Açısından Deđişen Madde Fonksiyonu Analizi	62
3.1.2.	İstatistiđe Giriş Okuryazarlık Testi İçin Üniversiteler Açısından Deđişen Madde Fonksiyonu Analizi.	69
3.2.	Yabancı Uyruklu Öğrenci Sınavı	75
3.2.1.	KTÜYÖS-2019 Cinsiyet Açısından Deđişen Madde Fonksiyonu Analizi.....	82

3.2.1.1.	Türkiye-İran Karşılaştırması	82
3.2.1.2.	Türkiye-Azerbaycan Karşılaştırması.....	85
3.2.1.3.	Almanya- Suriye Karşılaştırılması	87
3.2.1.4.	Hollanda-İran Karşılaştırması	89
3.2.2.	KTÜYÖS-2019 Uyrak Açısından Değişen Madde Fonksiyonu Analizi.....	91
3.2.2.1.	Türkiye-İran Karşılaştırması	91
3.2.2.2.	Türkiye- Azerbaycan Karşılaştırması.....	96
3.2.2.3.	Almanya- Suriye Karşılaştırması	99
3.2.2.4.	Hollanda-İran Karşılaştırması	101
3.2.3.	KTÜYÖS-2019 Bölgeler Açısından Değişen Madde Fonksiyonu Analizi.....	104
3.2.3.1.	Güney Asya-Avrupa Karşılaştırması.....	104
3.2.3.2.	Amerika-Afrika Karşılaştırması	113
3.2.3.3.	Avrupa-Orta Doğu Karşılaştırması	116
3.2.3.4.	Batı Asya-Avrupa Karşılaştırması.....	122
3.2.3.5.	Orta Asya-Afrika Karşılaştırması.....	127
4.	SONUÇLAR	133
5.	ÖNERİLER	142
6.	KAYNAKLAR.....	143
7.	EKLER	156
ÖZGEÇMİŞ		

Doktora Tezi

ÖZET

FARKLI ÖLÇEK TİPLERİNDE DEĞİŞEN MADDE FONKSİYONUNUN BELİRLENMESİ VE YÖNTEMLERİN KARŞILAŞTIRILMASI ÜZERİNE BİR ÇALIŞMA

Burcu HASANÇEBİ

Karadeniz Teknik Üniversitesi

Fen Bilimleri Enstitüsü

İstatistik ve Bilgisayar Bilimleri Anabilim Dalı

Danışman: Prof. Dr. Zafer KÜÇÜK-Prof. Dr. Yüksel TERZİ

2021, 155 Sayfa, 3 Sayfa Ek

Tarihte ilk ölçme, karış, kulaç, adım gibi insan uzuvlarının uzunluklarından yola çıkılarak yapılmıştır. Ölçmeye duyulan bu ihtiyaç, uygarlıklar geliştikçe, doğrudan gözlemlenmesi mümkün olmayan zeka, başarı, davranış, öğrenme gibi kavramların ölçülmesini gündeme getirmiştir. Örtük özellikleri ölçmek için geliştirilen çeşitli ölçme araçlarının ölçmeyi ne kadar doğru yaptığı sorusu ise tüm ölçekler için cevaplanması gereken öncelikli sorudur. Geliştirilen ölçeklerin ölçeceği alan kişilerin belirli özelliklerinden dolayı onlara avantaj ya da dezavantaj sağlaması kabul edilemez bir durumdur. Sunulan bu tezde, öncelikle öğrenme örtük özelliğini ölçmek üzere İstatistiğe Giriş Okuryazarlık Testi geliştirilmiştir. Geliştirilen testin uygulanabilirliği ortaya konmuştur. Ayrıca bu tezde, 2019 yılı Karadeniz Teknik Üniversitesi Yabancı Uruklu Öğrenci Sınavı (KTÜYÖS-2019) verileri kullanılmıştır. Her iki ölçek için de Madde Tepki Kuramı'nın varsayımları sağlanmış ve kuramdaki en uygun model belirlenmiştir. Mevcut ölçeklerin ikisi için de madde ve kişi parametreleri kestirilmiştir. Ardından ölçek içinde yanlılık oluşturabilecek maddelerin olup olmadığını varsa bu maddelerin hangi maddeler olduğunu bulabilmek için Değişen Madde Fonksiyonu analizi; Breslow-Day, Lojistik Regresyon, Lord'un χ^2 İstatistiği, Olabilirlik Oranı ve Raju'nun Alan Ölçümleri yöntemleri ile cinsiyet, okul ve sosyokültürel durumu için araştırılmıştır. Bu analizler yardımı ile hem mevcut Değişen Madde Fonksiyonu içeren maddeler belirlenmiş hem de yöntemlerin karşılaştırılmaları yapılmıştır. Yeterli örneklem büyüklüğünün ve uygun test uzunluğunun sağlandığı durumlarda dahi Klasik Test Kuramı'na bağlı yöntemlerin birbirinden uzakta ve tutarsız sonuçlar verdiği görülmüştür. Bunun yanında Madde Tepki Kuramına bağlı yöntemlerden Olabilirlik Oranının da gereken şartlar sağlandığında birebir örtüştüğü sonucuna varılmıştır.

Anahtar Kelimeler: Ölçme, Ölçek, Klasik Test Kuramı, Madde Tepki Kuramı, Değişen Madde Fonksiyonu, Madde Yanlılığı

PhD. Thesis

SUMMARY

A STUDY ON DETERMINING DIFFERENTIAL ITEM FUNCTION IN DIFFERENT SCALE
TYPES AND COMPARISON OF METHODS

Burcu HASANÇEBİ

Karadeniz Technical University
The Graduate School of Natural and Applied Sciences
Statistics and Computer Sciences Graduate Program
Supervisor: Prof. Dr. Zafer KÜÇÜK- Prof. Dr. Yüksel TERZİ
2021, 155 Pages 3 Pages Appendix

The first measurement in history was made based on the lengths of human limbs such as span, fathom, step. This need for measurement has brought to the agenda the measurement of concepts such as intelligence, success, behavior, learning that cannot be directly observed as civilizations develop. The question of how accurately the various measurement tools developed to measure implicit properties are the primary question that needs to be answered for all scales.

It is unacceptable for the developed scales to provide advantage or disadvantage to the people who take the scale due to their specific characteristics. In this thesis presented, firstly, Introduction to Statistics Literacy Test was developed to measure learning latent feature. The applicability of the developed test has been demonstrated. In addition, in this thesis, the data of 2019 Karadeniz Technical University Foreign Student Exam (KTUYOS-2019) were used. For both scales, the assumptions of the Item Response Theory were provided and the most suitable model in the theory was determined. Item and person parameters were estimated for both available scales. Then, a Differential Item Function analysis to find out whether there are items that may cause bias in the scale, if any, which items are; Breslow-Day was investigated for gender, school and sociocultural status with Logistic Regression, Lord's Statistics of χ^2 , Likelihood Ratio and Raju's Field Measurements. With the help of these analyzes, both the items containing the current Differential Item Function were determined and the methods were compared. It has been observed that even in cases where sufficient sample size and appropriate test length are provided, methods based on Classical Test Theory give distant and inconsistent results. In addition, it was concluded that the Likelihood Ratio, which is one of the methods based on the Item Response Theory, coincides exactly when the necessary conditions are met.

Key Words: Measurement, Scale, Classical Test Theory, Item Response Theory, Differential Item Function, Item Bias

ŞEKİLLER DİZİNİ

Sayfa No

Şekil 1.	Testlerin sınıflandırılması (Tekin, 1984; Gömleksiz ve Erkan, 2016).....	5
Şekil 2.	Madde karakteristik eğrisi	32
Şekil 3.	Ayırt edicilikleri aynı 3 maddeye ait MKE	32
Şekil 4.	Güçlükleri aynı 3 maddeye ait MKE	33
Şekil 5.	Mükemmel ayırt ediciliğe sahip maddeye ait MKE	34
Şekil 6.	Lord'un χ^2 istatistiği ile belirlenen cinsiyet için DMF'li maddeler	66
Şekil 7.	Cinsiyet açısından Raju'nun Alan Ölçümleri yöntemi	67
Şekil 8.	Cinsiyet açısından Olabilirlik Oranı yöntemi	68
Şekil 9.	Lord'un χ^2 istatistiği ile belirlenen üniversiteler için DMF'li maddeler	72
Şekil 10.	Üniversiteler açısından Raju yöntemi ile belirlenen DMF içeren maddeler	73
Şekil 11.	Üniversiteler için Olabilirlik Oranı yöntemiyle bulunan DMF'li maddeler.....	74
Şekil 12.	Güney Asya ve Avrupa için Breslow-Day Yöntemi	105
Şekil 13.	Güney Asya ve Avrupa için Lojistik Regresyon Yöntemi	107
Şekil 14.	Güney Asya ve Avrupa için Lord'un χ^2 İstatistiği	109
Şekil 15.	Güney Asya ve Avrupa için Raju'nun Alan Ölçümleri	111
Şekil 16.	Güney Asya ve Avrupa için Olabilirlik Oranı yöntemi.....	113
Şekil 17.	Amerika ve Afrika için Lord'un χ^2 İstatistiği	114
Şekil 18.	Amerika ve Afrika için Raju'nun Alan Ölçümleri.....	115
Şekil 19.	Amerika ve Afrika için Olabilirlik Oranı	115
Şekil 20.	Avrupa ve Orta Doğu için Breslow-Day Yöntemi	116
Şekil 21.	Avrupa ve Orta Doğu için Lojistik Regresyon Yöntemi.....	118
Şekil 22.	Avrupa ve Orta Doğu için Lord'un χ^2 İstatistiği	119
Şekil 23.	Avrupa ve Orta Doğu için Raju'nun Alan Ölçümleri	120
Şekil 24.	Avrupa ve Orta Doğu için Olabilirlik Oranı Yöntemi.....	122
Şekil 25.	Batı Asya ve Avrupa için Breslow-Day	123
Şekil 26.	Batı Asya ve Avrupa için Lojistik Regresyon Yöntemi	124
Şekil 27.	Batı Asya ve Avrupa için Lord'un χ^2 İstatistiği.....	125
Şekil 28.	Batı Asya ve Avrupa için Raju'nun Alan Ölçümleri	126
Şekil 29.	Batı Asya ve Avrupa için Olabilirlik Oranı	127

Şekil 30. Orta Asya ve Afrika için Breslow-Day Yöntemi	128
Şekil 31. Orta Asya ve Afrika için Lojistik Regresyon yöntemi	129
Şekil 32. Orta Asya ve Afrika için Lord'un χ^2 İstatistiği.....	130
Şekil 33. Orta Asya ve Afrika için Raju'nun Alan Ölçümleri	131
Şekil 34. Orta Asya ve Afrika için Olabilirlik Oranı	131



TABLolar DİZİNİ

	<u>Sayfa No</u>
Tablo 1. Madde güçlük indeksi için referans aralığı.....	13
Tablo 2. Ayırt edicilik indeksi için düzeyler.....	14
Tablo 3. Cronbach alfa için referans değerleri.....	23
Tablo 4. Odak ve referans grup için çapraz tablo	50
Tablo 5. Toplam puana ilişkin betimleyici istatistikler.....	57
Tablo 6. Verilerin tek boyutluluk ölçüsü	58
Tablo 7. Model seçimi için uyum kriterleri	59
Tablo 8. Verilen cevapların oranı ve logit değerler	60
Tablo 9. Her bir madde için güvenilirlik katsayıları.....	61
Tablo 10. Madde istatistiklerinin tahmini	62
Tablo 11. Cinsiyet açısından BD İstatistiği yöntemi ile DMF analizi	63
Tablo 12. Cinsiyet açısından Lojistik Regresyon yöntemi ile DMF analizi.....	64
Tablo 13. Cinsiyet açısından Lord'un χ^2 İstatistiği yöntemi ile DMF analizi.....	65
Tablo 14. Cinsiyet açısından Raju'nun alan ölçümleri yöntemi ile DMF analizi	66
Tablo 15. Cinsiyet açısından olabilirlik oranı yöntemi ile DMF analizi	68
Tablo 16. Üniversiteler açısından BD İstatistiği yöntemi ile DMF analizi	69
Tablo 17. Üniversiteler açısından lojistik regresyon yöntemi ile DMF analizi.....	70
Tablo 18. Üniversiteler açısından Lord'un χ^2 İstatistiği yöntemi ile DMF analizi	71
Tablo 19. Üniversiteler açısından Raju'nun alan ölçümleri yöntemi ile DMF analizi.....	72
Tablo 20. Üniversiteler açısından olabilirlik oranı yöntemi ile DMF analizi.....	74
Tablo 21. 2018-2019 KTÜ yabancı uyruklu öğrenci sayıları (İlk 5 ülke).....	76
Tablo 22. 2019 KTÜYÖS'e giren ülkeler ve cinsiyete göre öğrenci sayıları	76
Tablo 23. KTÜYÖS-2019 verisi için tek boyutluluk kriteri.....	76
Tablo 24. Model-Veri uyum kriterleri	77
Tablo 25. Verilen cevapların oranı ve logit değerler	78
Tablo 26. Güvenilirlik katsayıları	78
Tablo 27. Madde istatistiklerinin tahmini	80
Tablo 28. Türkiye-İran cinsiyete göre BD istatistiği sonuçları.....	82

Tablo 29.	Türkiye-İran cinsiyete göre lojistik regresyon yöntemi sonuçları.....	83
Tablo 30.	Türkiye-İran cinsiyete göre Lord'un χ^2 istatistiği yöntemi	83
Tablo 31.	Türkiye-İran cinsiyete göre Raju'nun alan ölçümleri yöntemi.....	84
Tablo 32.	Türkiye-İran cinsiyete göre olabilirlik oranı yöntemi	84
Tablo 33.	Türkiye-Azerbaycan cinsiyete göre BD yöntemi sonuçları.....	85
Tablo 34.	Türkiye-Azerbaycan cinsiyete göre lojistik regresyon yöntemi.....	85
Tablo 35.	Türkiye-Azerbaycan cinsiyete göre Lord'un χ^2 istatistiği yöntemi	86
Tablo 36.	Türkiye-Azerbaycan cinsiyete göre Raju'nun alan ölçümleri yöntemi.....	86
Tablo 37.	Türkiye-Azerbaycan cinsiyete göre olabilirlik oranı yöntemi.....	86
Tablo 38.	Almanya-Suriye cinsiyete göre BD yöntemi.....	87
Tablo 39.	Almanya-Suriye cinsiyete göre lojistik regresyon yöntemi.....	87
Tablo 40.	Almanya-Suriye cinsiyete göre Lord'un χ^2 istatistiği	88
Tablo 41.	Almanya-Suriye cinsiyete göre Raju'nun alan ölçümleri.....	88
Tablo 42.	Almanya-Suriye cinsiyete göre olabilirlik oranı.....	88
Tablo 43.	Hollanda-İran cinsiyete göre BD yöntemi.....	89
Tablo 44.	Hollanda-İran cinsiyete göre lojistik regresyon yöntemi.....	89
Tablo 45.	Hollanda-İran cinsiyete göre Lord'un χ^2 istatistiği	90
Tablo 46.	Hollanda-İran cinsiyete göre Raju'nun alan ölçümleri.....	90
Tablo 47.	Hollanda-İran cinsiyete göre olabilirlik oranı.....	90
Tablo 48.	Türkiye-İran uyruk için BD istatistiği sonuçları.....	91
Tablo 49.	Türkiye-İran uyruk için lojistik regresyon yöntemi.....	92
Tablo 50.	Türkiye-İran uyruk için Lord'un χ^2 istatistiği yöntemi	93
Tablo 51.	Türkiye-İran uyruk için Raju'nun alan ölçümleri yöntemi.....	94
Tablo 52.	Türkiye-İran uyruk için olabilirlik oranı yöntemi	95
Tablo 53.	Türkiye-Azerbaycan uyruk için BD yöntemi sonuçları	96
Tablo 54.	Türkiye-Azerbaycan uyruk için lojistik regresyon yöntemi.....	96
Tablo 55.	Türkiye-Azerbaycan uyruk için Lord'un χ^2 istatistiği yöntemi	97
Tablo 56.	Türkiye-Azerbaycan uyruk için Raju'nun alan ölçümleri yöntemi.....	98
Tablo 57.	Türkiye-Azerbaycan uyruk için olabilirlik oranı.....	98
Tablo 58.	Almanya-Suriye uyruk için BD yöntemi.....	99
Tablo 59.	Almanya-Suriye uyruk için lojistik regresyon yöntemi.....	99
Tablo 60.	Almanya-Suriye uyruk için Lord'un χ^2 istatistiği	100
Tablo 61.	Almanya-Suriye uyruğa göre Raju'nun alan ölçümleri.....	100

Tablo 62.	Almanya-Suriye uyruk için olabilirlik oranı yöntemi	100
Tablo 63.	Hollanda-İran Uyruk için BD yöntemi	101
Tablo 64.	Hollanda-İran Uyruk için Lojistik regresyon yöntemi	101
Tablo 65.	Hollanda-İran Uyruk için Lord'un χ^2 istatistiği yöntemi	102
Tablo 66.	Hollanda-İran Uyruk için Raju'nun alan ölçümleri.....	103
Tablo 67.	Hollanda-İran uyruğa göre olabilirlik oranı.....	103
Tablo 68.	Güney Asya ve Avrupa için BD yöntemi	105
Tablo 69.	Güney Asya ve Avrupa için lojistik regresyon yöntemi.....	106
Tablo 70.	Güney Asya ve Avrupa için Lord'un χ^2 istatistiği.....	108
Tablo 71.	Güney Asya ve Avrupa için Raju'nun alan ölçümleri	110
Tablo 72.	Güney Asya ve Avrupa için olabilirlik oranı yöntemi.....	112
Tablo 73.	Amerika ve Afrika için Lord'un χ^2 istatistiği	114
Tablo 74.	Amerika ve Afrika için Raju'nun alan ölçümleri.....	114
Tablo 75.	Avrupa ve Orta Doğu için BD yöntemi.....	116
Tablo 76.	Avrupa ve Orta Doğu için lojistik regresyon yöntemi.....	117
Tablo 77.	Avrupa ve Orta Doğu için Lord χ^2 İstatistiği	118
Tablo 78.	Avrupa ve Orta Doğu için Raju'nun alan ölçümleri	120
Tablo 79.	Avrupa ve Orta Doğu için olabilirlik oranı.....	121
Tablo 80.	Batı Asya ve Avrupa için BD yöntemi	122
Tablo 81.	Batı Asya ve Avrupa için lojistik regresyon yöntemi.....	123
Tablo 82.	Batı Asya ve Avrupa için Lord'un χ^2 istatistiği	124
Tablo 83.	Batı Asya ve Avrupa için Raju'nun alan ölçümleri	125
Tablo 84.	Batı Asya ve Avrupa için olabilirlik oranı.....	126
Tablo 85.	Orta Asya ve Afrika için BD yöntemi	128
Tablo 86.	Orta Asya ve Afrika için lojistik regresyon yöntemi.....	129
Tablo 87.	Orta Asya ve Afrika için Lord χ^2 istatistiği.....	129
Tablo 88.	Orta Asya ve Afrika için Raju'nun alan ölçümleri	130
Tablo 89.	Orta Asya ve Afrika için olabilirlik oranı.....	131
Tablo 90.	Cinsiyet için alınan sonuçların karşılaştırılması.....	134
Tablo 91.	Üniversiteler için alınan sonuçların karşılaştırılması	134
Tablo 92.	Türkiye-İran cinsiyet açısından karşılaştırma.....	136
Tablo 93.	Türkiye-Azerbaycan cinsiyet açısından karşılaştırma	136
Tablo 94.	Almanya-Suriye cinsiyet açısından karşılaştırma.....	137

Tablo 95. Hollanda-İran cinsiyet açısından karşılaştırma.....	137
Tablo 96. Türkiye-İran Uyruk açısından karşılaştırma.....	137
Tablo 97. Türkiye-Azerbaycan uyruk açısından karşılaştırma	138
Tablo 98. Almanya-Suriye uyruk açısından karşılaştırma.....	138
Tablo 99. Hollanda-İran uyruk açısından karşılaştırma.....	138
Tablo 100. Güney Asya-Avrupa karşılaştırması.....	139
Tablo 101. Amerika-Afrika karşılaştırması	139
Tablo 102. Avrupa-Orta Doğu karşılaştırması.....	139
Tablo 103. Batı Asya- Avrupa karşılaştırması.....	140
Tablo 104. Orta Asya-Afrika karşılaştırması.....	140



KISALTMALAR VE SEMBOLLER DİZİNİ

a	: Ayırt Edicilik Parametresi
AIC	: Akaike Bilgi Kriteri
ALES	: Akademik Personel Lisansüstü Eğitim Sınavı
ATM	: Aşamalı Tepki Modeli
b	: Madde Güçlük İndeksi
BD	: Breslow-Day İstatistiği
BIC	: Bayesian Bilgi Kriteri
c	: Şans Parametresi (Alt Asimptot)
ÇYRM	: Çok Yüzeyle Rasch Model
d	: Dikkatsizlik Parametresi (Üst Asimptot)
DIC	: Sapma Bilgi Kriteri
DMF	: Değişen Madde Fonksiyonu
EKK	: En Küçük Kareler
EÖB	: Etkili Örneklem Büyüklüğü
KKM	: Kısmi Kredi Modeli
KPSS	: Kamu Personeli Seçme Sınavı
KR-20	: Kuder-Richardson 20 Yöntemi
KR-21	: Kuder-Richardson 21 Yöntemi
KTK	: Klasik Test Kuramı
KTÜ	: Karadeniz Teknik Üniversitesi
KTÜYÖS-2019	: 2019 Yılı Karadeniz Teknik Üniversitesi Yabancı Uyruklu Öğrenci Sınavı
L	: Olabilirlik Fonksiyonu
LGS	: Lise Geçiş Sınavı
LL	: LogLikelihood
LR	: Lojistik Regresyon Yöntemi
MCMC	: Markov Zinciri Monte Carlo Yöntemi
MH	: Mantel-Haenszel Yöntemi
MKE	: Madde Karakteristik Eğrisi
MTK	: Madde Tepki Kuramı

OMÜ	: Ondokuz Mayıs Üniversitesi
OO	: Olabilirlik Oranı
ÖSS	: Öğrenci Seçme Sınavı
ÖSYM	: Öğrenci Seçme ve Yerleştirme Merkezi
P_{ij}	: KTK için Madde Güçlük İndeksi
r_{jx}	: KTK için Ayırt Edicilik İndeksi
SBS	: Seviye Belirleme Sınavı
YDS	: Yabancı Dil Bilgisi Seviye Belirleme Sınavı
YÖS	: Yabancı Uyruklu Öğrenci Sınavı
θ	: Yetenek Parametresi
1PLM	: 1 Parametrelili Lojistik Model
2PLM	: 2 Parametrelili Lojistik Model
3PLM	: 3 Parametrelili Lojistik Model
4PLM	: 4 Parametrelili Lojistik Model

1. GENEL BİLGİLER

1.1. Giriş

Ölçme ile doğrudan anlaşılması mümkün olmayan örtük özelliklerin var oluşu ve var oluş dereceleri sorgulanır. Bu anlamda ölçme, var olan özelliklerin tanımlanmasında önemli bir rol oynar. Ölçme ile ilgilenilen örtük özelliğin var oluş dereceleri sayı ve semboller ile ifade edilir (Tekin, 2004). Ölçmeye konu olması mümkün örtük özellikler birçok disipline ait olabilir. Ölçme teorisinde ölçülmek istenen özelliklerin her biri örtük ya da gizil özellik olarak ifade edilirler. Bu özellikler hakkında, ölçekler üzerinden yapılan gözlemlere bağlı olarak çıkarımlar yapılabilmektedir.

Sağlıklı bir ölçme işleminden beklenen, objektif, hassas ve tutarlı olmasıdır. Herhangi bir nesneye ait özellikler ile ilgili elde edilen bilgiler kesin, açık ve doğru bir şekilde ifade edilmelidir. Bilimsel gelişmede ölçmenin yeri ve önemi tartışmasız çok büyüktür. Doğru yapılan ölçümler sonucunda alınan objektif, hassas ve tutarlı sonuçlar ile literatüre katkı sağlanarak bilimin ilerlemesi mümkün kılınmaktadır. İstenen özelliklerde ölçme yapabilmek için birtakım koşulların sağlanması gerekir, bunlardan ilki; ölçme aracının gruptan bağımsız olmasıdır diğeri ise ölçülen özelliklerin kullanılan ölçme aracından bağımsız olarak ölçeklenebilmesidir. Ölçmeye konu olan özelliklerin türlerine göre yapılması gereken ölçüm farklılık gösterebilmektedir. Literatürde 3 çeşit ölçme türü mevcuttur. Bunlardan ilki olan doğrudan ölçmedir. Bu ölçme türü adından da anlaşılacağı gibi örtük özelliğin doğrudan gözlemlenebildiği durumlarda kullanılmaktadır. Bu tür ölçmelerde ölçmeye konu olan özellik duyu organları ile tespit edilebilir ve kullanılan ölçme aracı ile eşleştirilerek yapılır. Boy ve kilo ölçümü doğrudan ölçme için ilk verilecek örnekler arasındadır. Diğer bir ölçme türü ise dolaylı ölçmedir. Bu tip ölçümde de araştırılan özellik doğrudan gözlemlenememekte ölçümler ancak bir ölçme aracı vasıtasıyla yapılabilmektedir. Ölçülen kişilerin ölçme aracına verdikleri tepkiler yoluyla ölçüm sonuçları elde edilmektedir. Eğitim bilimleri ve psikoloji gibi birçok disipline ölçülmek istenen özellikler soyut kavramlar olduklarından doğrudan ölçülmeleri mümkün olmamaktadır. Bu nedenle bu disiplinlerde belli bir özellik araştırılırken bu özelliği ölçmek için geliştirilmiş ölçeklere ihtiyaç duyulmaktadır. Dolaylı ölçmeye zeka, başarı, özyeterlilik ve tutum ölçekleri örnek olarak verilebilir. Son ölçme türü ise türetilmiş

ölçmedir. Bu ölçme tipinde ise araştırılan özelliği ölçmek için başka özelliklerden oluşan bir bağıntı kullanılır. Ölçülmek istenen özelliğe o özellik için gereken formüllerden elde edilen sonuçlar yoluyla ulaşılmaktadır. Türetilmiş ölçmeye dönem sonu notu, yoğunluk hesaplanması gibi birçok örnek verilebilir (Arıkan, vd., 2017).

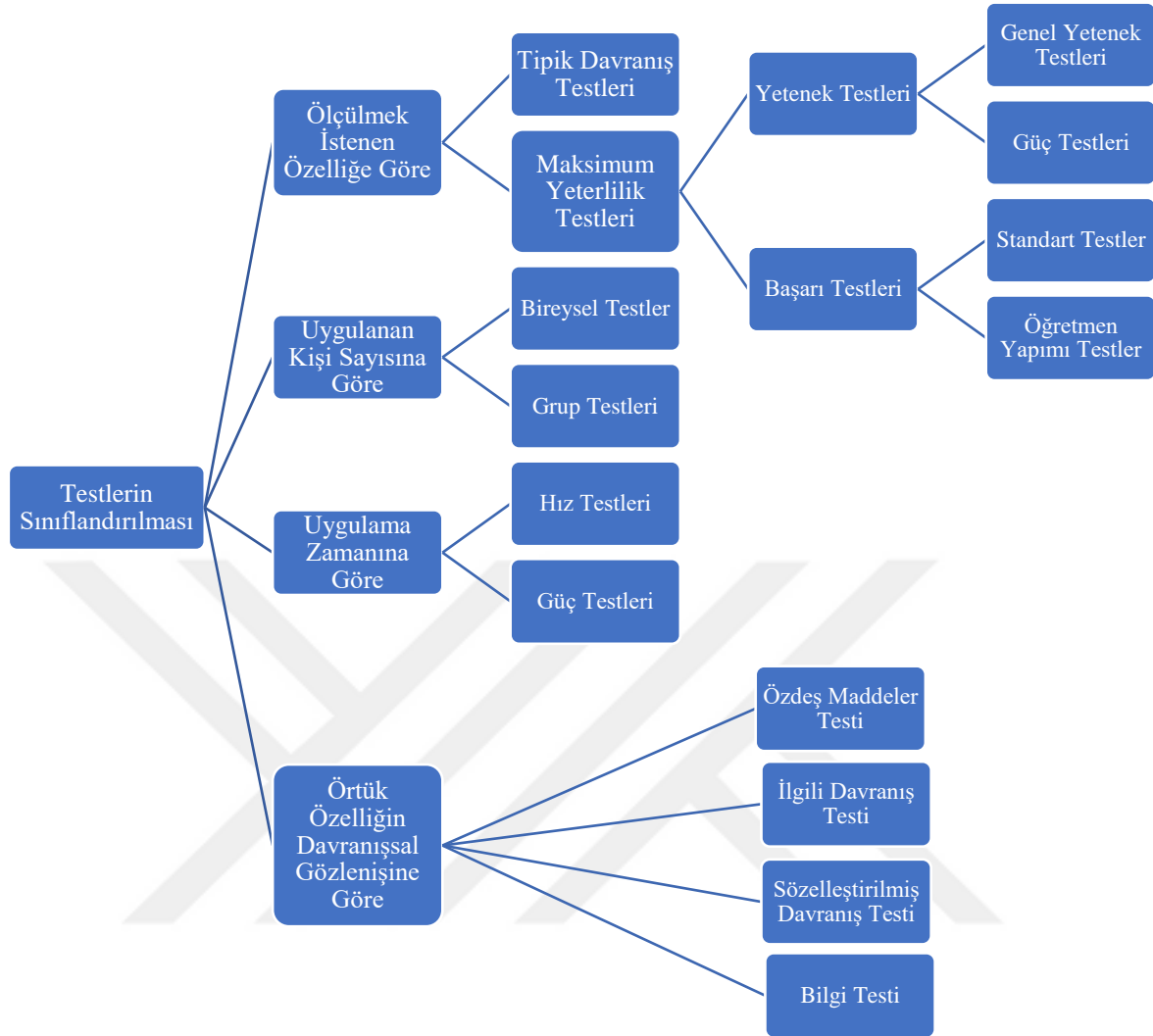
Ölçülecek bir örtük özelliğin doğrudan gözlemlenememesi veya gözlemlenebilse bile, yeterince hassas sonuçlar alınamaması durumunda ölçme devreye girmektedir. Ölçme sonuçlarının elde edilebilmesi için kullanılan araç ve gereçlere ölçek denir. Ölçekler sayesinde araştırmacı ölçülen özellik ile ilgili bilgi edinebilir. Bireylerin çeşitli alanlardaki örtük özelliklerini ölçerken değişik ölçek tiplerine ve ölçme yöntemlerine başvurulur. Hangi ölçek tipine başvurulması gerektiği, ölçülecek örtük özelliğin niteliği, denek sayısı ve maliyet gibi birçok bileşenden etkilenmektedir. Kullanılan ölçekler, ölçmenin amacına uygun belirli sayıda madde içermeli ve verilen yanıtlara göre cevaplayıcılar hakkında bilgi verebilecek yapıda olmalıdır. Yapılan ölçme işleminin hassasiyetinin artırılması, yanlılığın azaltılması, ölçeğin tekrar edilebilmesi, genellenebilmesi, doğrulanabilmesi gibi özellikleri sağlaması ile mümkündür. Literatürde dört çeşit ölçek türüne rastlanmaktadır. Bunlardan ilki, sınıflama ölçekleri olarak adlandırılır. Bu ölçeklerde araştırılan özellik bakımından benzerlik gösteren kişilerden gruplar oluşturulması söz konusudur. Cinsiyet, medeni durum, eğitim durumu gibi birçok özellik sınıflama ölçeği ile ölçülebilir. Bu ölçekte gruplara birbirlerinden ayrılabilmeleri için belirli sayılar verilebilir. Ancak bu sayıların matematiksel hiçbir anlamı yoktur. Bu sayılar üzerinden herhangi bir işlem yapılması söz konusu değildir. Bu ölçekten elde edilen sonuçlardan sadece frekans bilgisinin kullanılması mümkündür (Arıkan, vd., 2017). Sıralama ölçek türünde ise araştırılan özellik bakımından kişilerin sıralanması mantığına dayalı bir sistem mevcuttur. Öğrencilerin not ortalamalarına göre sıralanmaları, bir yarışmada kazananlardan itibaren kişilerin sıralanması, maaşlarına göre kişilerin sıralanması bu ölçek türüne örnek olarak verilebilir. Burada da sınıflama ölçeğine benzer şekilde verilmiş sayılar ile işlemler yapılamaz. Çünkü burada kişilere atanmış sayılar birer sıra sayısını belirtir. Sayılar arasındaki farklar kişiler arasındaki farkı yansıtmadığı gibi iki sıra sayısının toplanmasıyla üçüncü kişiye atanacak sıra sayısına da ulaşamaz (Gömleksiz ve Erkan, 2016). Bir diğer ölçek türü ise eşit aralıklı ölçektir. Burada, belirli bir başlangıç noktasına göre kişiler ölçülen özelliğe göre eşit aralıklarla sıralanırlar. Başlangıç noktası keyfi olarak atanır ve yokluk ifade etmez. Bu ölçme tipinde toplama ve çıkarma işlemleri, sıralama işlemleri, korelasyon, ortalama, standart sapma hesaplamaları yapılabilir ancak çarpma ve bölme işlemleri yapılamamaktadır. Eğitim

bilimlerinde kullanılan ölçme araçlarından elde edilen puanlar, hava sıcaklık değerleri eşit aralıklı ölçek türüne örnek olarak verilebilir (Arıkan vd., 2017). Son ölçek türü ise eşit oranlı ölçektir. Bu ölçek tipinde eşit aralıklı ölçekteki keyfi sıfır noktası artık gerçek sıfır noktasıdır. Mutlak sıfır ifadesi ölçek ile araştırılan özelliğin ölçüm sonucunun sıfır olması durumunda araştırılan özelliğin mevcut olmadığı anlamına gelmektedir. Diğer ölçek türlerinden farklı olarak sıfır noktasının mutlak sıfır olması sebebiyle çarpma ve bölme işlemleri de yapılabilmektedir. Ölçekten elde edilen sonuçlar oranlanabilir. En çok bilgi sağlayan ölçek tipi eşit oranlı ölçeklerdir. Ağırlık, uzunluk ve gelir eşit oranlı ölçek tipine örnek verilebilir (Arıkan vd., 2017).

Bir testin en temel birimi test maddesidir. Madde geliştirme, test gelişim sürecinin en sancılı dönemidir. Nihai teste alınacak özellikte maddeleri ortaya çıkarmak sürecin en zaman alan parçasıdır. Bir ölçek geliştirilirken, deneklerin bilgi, beceri ve yetenekleri hakkında bilgi sağlayacak maddeler geliştirmek için oldukça çaba sarf edilmelidir (Altun, 2014). Geliştirilen ölçeğin özellikleri, ölçme aracını meydana getiren maddelerin özelliklerine bağlıdır. İstenen özellikte bir ölçek geliştirmenin yolu, istenen özellikte madde oluşturmaktan geçmektedir. Teste konu olan özelliği gerektiği gibi ölçülebilmek için ve teste tabi olan deneklerin özellikleri hakkında yorumda bulunabilmek için belli miktarda maddeye ihtiyaç duyulur. Testte kullanılacak madde tipi de ölçülecek olan özelliğe uygun olmalıdır. Maddeler, ölçeğin güvenilirlik ve geçerliliğini arttıracak şekilde hazırlanmalı ve hangi ölçek tipi kullanılacaksa içeriğindeki maddeler ona uygun şekilde yazılmalıdır.

Test, araştırılan örtük özelliği ölçmek amacıyla düzenlenen, içerdiği her bir maddenin ilgilenilen örtük özelliğin farklı bir yönünü ölçtüğü, elde edilen sonuçlar ile testi alan kişiler arasındaki farkların ortaya konduğu ölçme araçlarıdır. Bireylerin test performansları değerlendirilerek test ile ölçülmek istenen özellik bakımından birtakım yargılara varılır. İyi bir test oluşturmak demek testin uygulandığı tüm disiplinlerde doğru karar vermek demektir. Bu kararlar bazı durumlarda insan hayatını etkileyecek öneme sahip olduklarından doğru verilmelidir. Testler türlerine göre doktorlara, psikologlara, öğretmenlere, velilere ve öğrencilere geri bildirimler vermektedir. Doğru karar vermenin tek yolu ise kaliteli bir test hazırlamaktan geçmektedir. İyi bir test hazırlamak için öncelikle araştırmacının hangi örtük özelliği ölçmek istediği net bir şekilde belirlenmelidir. Bu adım testin amaca hizmet etmesi için çok önemlidir. Testin amacı ve kapsamı belirlenmeden ölçek geliştirmek anlamsız olacaktır. Testin neyi ölçmek için geliştirileceğinin belirlenmesinin ardından ölçeğin hangi tür maddelerden oluşacağına

karar verilmelidir (Çoktan seçmeli, ikili, açık uçlu gibi). Madde türüne karar verildikten sonra nihai teste seçilmek üzere aday maddelerden oluşan madde havuzu oluşturulmalıdır. Oluşturulan bu madde havuzundan testin son hali için amaca en iyi hizmet eden maddeler seçilecektir. Bu sürecin tamamlanmasıyla test uzunluğuna karar verme süreci başlayacaktır. Testin uzunluğu yapılacak tüm değerlendirmelerin doğruluğu ve hassaslığı için oldukça önemli bir konudur. İyi bir test örtük özelliğin tüm kazanımlarını sorgulayacak kadar madde içermelidir. Ayrıca güvenilirliği ve geçerliliği olumsuz yönde etkilemeyecek sayıda maddeye sahip olmalıdır. Bütün bu aşamalardan geçen test için bir sonraki adım nihai testin oluşturulmasıdır. Madde havuzundan ilgilenilen örtük özelliği en iyi ölçecek maddeler nihai teste seçilerek teste son hali verilir. Son halini alan test için uygulama ve değerlendirme adımları ile test hazırlama ve geliştirme süreci tamamlanmaktadır. Testler pek çok açıdan sınıflandırılabilir.



Şekil 1. Testlerin sınıflandırılması (Tekin, 1984; Gömleksiz ve Erkan, 2016).

Şekil 1 ile testlerin sınıflandırılma şeması gösterilmektedir. Testler genel itibariyle dört ana başlıkta sınıflandırılabilir. Bunlardan ilki ölçülmek istenen özelliğe göre sınıflanan testlerdir. Bu tür testlerde kendi aralarında ikiye ayrılmaktadır. Tipik davranış testlerinde testi alan bireyler kendilerine has özellikler ile maddelere cevap verirler. Kendi yaşamlarının olağan akışında durumlara ve olaylara verdikleri tepkiler bu testteki maddelerin yanıtları olacaktır. Bu durumda kesin bir doğru ya da yanlış yoktur. Cevaplar kişisel özelliklere göre değişkenlik gösterir. Burada önemli olan nokta kişinin kendisiyle ilgili verdiği cevapların doğru olmasıdır (Özgüven, 1994; Gömleksiz ve Erkan, 2016). Maksimum yeterlilik testlerinde testi alan bireyin bir işi ne kadar doğru yapabildiği ölçülmektedir. Kendi arasında yetenek testleri ve başarı testleri olarak ikiye ayrılır. Yetenek testlerinde, ölçülen bireyin kendisine verilecek bir yapıyı ne ölçüde

öğrenebileceği tahmin edilir (Tekin, 1984). Yetenek testleri de kendi arasında genel yetenek testleri ve özel yetenek testleri olarak ikiye ayrılmaktadır. Başarı testlerinde ise kişinin kendisine verilen bir yapıyı ne kadar öğrendiği ölçülür. Burada, başarı testlerini yetenek testlerinden ayıran nokta, başarı testinde geçmişte verilen bir yapının ne ölçüde öğrenildiği ölçülürken yetenek testlerinde ileride verilecek bir yapının ne kadar öğrenilebileceğinin ölçülmesidir. Başarı testleri testi hazırlayan kişi bakımından standart testler ve öğretmen yapımı testler olarak kendi içinde ikiye ayrılmaktadır (Tekin, 1984; Gömleksiz ve Erkan, 2016).

Testlerin sınıflandırılmasında ikinci ana başlık uygulanan kişi sayısına göre testlerdir. Bu kategoride bireysel testler ve grup testleri bulunmaktadır. Bireysel testler, daha çok psikolojik teşhis koymak için kullanılan testlerdir. Bu test türünde uygulayan kişi ve uygulanan kişi birebir muhatap olurlar. Bireysel testler tek seferde tek kişiye yapılabilir. Grup testlerinde ise bireysel testlerin aksine uygulayıcı tek seferde birçok kişiye testi uygulayabilir. Sınıflarda öğretmenler tarafından uygulanan testler bu tip test türüne örnek gösterilebilir. Bireysel testlere göre her anlamda daha kullanışlı testlerdir (Tekin, 1984; Gömleksiz ve Erkan, 2016).

Sınıflandırmada bir diğer kategori uygulama zamanına göre ayrılan testlerdir. Bu sınıfta hız ve güç testleri mevcuttur. Hız testlerinde adından da anlaşılacağı üzere önemli olan verilen sürede maksimum maddeye doğru yanıt vermektir. Bu testlerden alınacak puan kişinin hızıyla birebir alakalıdır. Güç testlerinde ise hız testlerinin aksine zamana karşı bir yarış yoktur. Burada, ilerledikçe gücünü artan maddelerden hazırlanmış bir test söz konusudur ve yanıtlamak için yeterli zaman verilmektedir (Tekin, 1984; Özgüven, 1994; Gömleksiz ve Erkan, 2016).

Son kategori ise örtük özelliğin davranışsal olarak gözlemlenmesine göre ayrılan testlerden oluşmaktadır. Bu tip testler kendi içlerinde dörde ayrılmaktadır. Bunlardan ilki olan özdeş maddeler testi ölçülmek istenen örtük özelliği doğrudan ölçen testlerdir. Performans testleri buna örnek olarak verilebilir. İlgili davranış testinde araştırılan örtük özellik oluşturulan yapay ortamlarda gözlenir, bu gözlemler yardımıyla gerçekte özelliğin var oluşu ile ilgili tahminler yapılır. Sözcüklerle oluşturulmuş davranış testlerinde ise belli bir dönemde verilen ancak o dönem içerisinde ölçülemeyen davranışları ölçen test tipidir. Öğrencilerin seçimlerde oy kullanmaları bilgisi dersin içinde öğrenciye verilir ancak öğrenciler o dönem içerisinde oy kullanma yetkisine sahip olmadıklarından bu bilgiyi davranışa dönüştürüp dönüştürmedikleri sadece sözel olarak sorgulanabilir. Bilgi

testlerinde de kişinin sözelleştirilmiş davranış testine konu olan özelliklere ilişkin bilgi düzeyi sorgulanır (Tekin, 1984; Gömleksiz ve Erkan, 2016).

Ölçme araçlarından elde edilen sonuçlar, bireyin ölçmeye konu olan özelliği ile ilgili karar vermek için kullanılır. Verilen kararın doğruluğu, karar vericinin adaletinin yanı sıra ölçeğin kalitesiyle de oldukça ilişkilidir. Bir ölçeğin kalitesinden bahsederken akla ilk gelmesi gereken kavramlar, güvenilirlik ve geçerlilik kavramlarıdır. Bu iki kavram bir ölçeğin vazgeçilmez özelliklerindedir. Bireyler hakkında verilen kararların doğruluğu her ne kadar uygun ölçek kullanımı ile ilişkilendirilse de düşük geçerlilik ve güvenilirlikteki bir uygun ölçekle alınan sonuçlar, karar vericileri yanı sıra sürükleyecektir (Gelbal, 1994).

Araştırmacı ölçekten elde ettiği sonuçların ölçmenin rastgele hatasından ne kadar uzakta ölçüm yaptığını merak etmek durumundadır. Çünkü alacağı sonuçların doğruluğu bununla birebir ilişkilidir. Hatasız ölçme yapmak mümkün değildir ancak, ölçekten beklenen, ölçtüğü örtük özelliği mümkün olan en az hatayla ölçebilmesidir. Bunu başarabilen ölçme araçları için güvenilirlik sağlanmış demektir. Güvenirlik, yapılan ölçümün tekrarlanabilmesi özelliği ile ilişkilidir. Güvenilir bir ölçekten alınan sonuçlar; aynı şartlarda, aynı cevaplayıcılara, değişik zamanlarda veya kişiler tarafından uygulandığında birbirine benzer olmalıdır. Ancak hiçbir zaman aynı olamaz çünkü her uygulamada ölçmeye farklı hata kaynakları karışacaktır. Her tekrar sonrası birbirine benzer olmasını beklediğimiz sonuçların dağılımının standart hatası ölçmenin standart hatasıdır. Hazırlanan ölçeğin bahsedilen hatalardan arınmışlık derecesi de ölçeğin güvenilirliğini vermektedir. Düşük standart hata yüksek güvenilirlik anlamına gelir. Güvenirlik kavramı kesin olarak hesaplanmaz tahmin edilir ve çeşitli tahmin yöntemleri mevcuttur. Bu yöntemler tek ya da tekrar uygulamalar ile yapılabilir. Güvenirliği kestirmek için kullanılan bazı yöntemler; Test-Tekrar Test Yöntemi, Eş Değer Formlar Yöntemi, Test-Yarı Yöntemi, İç Tutarlılık Katsayısı (KR-20 ve KR-21), Cronbach Alfa Güvenirlik Katsayısı şeklinde sıralanabilir (Gömleksiz ve Erkan, 2016).

Ölçeğin kalitesi için olmazsa olmaz bir diğer kavram olan geçerlilik, ölçmeye konu olan örtük özelliğin geliştirilen ölçek tarafından ölçülebilme seviyesidir. Bir test ölçmek istediği özelliği doğru ve diğer özelliklerden ayırıştırarak ölçebiliyor ise bu test geçerlidir. Bu anlamda geçerlilik, ölçeğin kesinliğinin bir ölçüsü ya da ölçeğin amaca hizmet edip etmediğinin bir ölçüsü olarak da ifade edilebilir. Geçerlilik türleri; Kapsam Geçerliliği, Ölçüt Dayanıklı Geçerlilik ve Yapı Geçerliliği olarak ifade edilebilir. Bir ölçeğe geçerlidir diyebilmek için öncelikle güvenilirirdir denilebilmelidir. Güvenirlik, geçerlilik için gerek

şart ancak yeter şart değildir. Güvenilir bir ölçek her zaman geçerli olmayabilir. Ölçeğin geçerliliği yüksek ise güvenilirliğinin de yüksek olması beklenirken tersi her zaman mümkün değildir (Ercan ve Kan, 2004).

Güvenilirlik ve geçerlilik kavramları ölçeğin ayakları olarak düşünülebilir. Bir ölçme aracının gücü bu iki ayağın gücüne bağlıdır. Ölçek için geçerlilik ve güvenilirlikten ayrı tutulamayacak bir diğer kavram ise madde yanlılığıdır. Yanlılık, ilgilenilen maddenin belirli kişi ya da kişilere bazı durumlarından dolayı avantaj ya da dezavantaj getirmesidir. Bu ölçekler için istenmeyen bir durumdur. İdeal ölçekte kişilerin ölçmek istenen özellikleri hariç diğer hiçbir özellik ölçmeye karışmamalıdır. Heterojen bir kitleye uygulanan herhangi bir ölçek için yanlılık analizi ölçüm sürecinin en önemli bileşeni haline gelir. Buradaki en önemli hedef denekler için adil, hassas ve duyarlı ölçümler yapmaktır. Bir maddenin yanlı olabilmesi için gerek şart Değişen Madde Fonksiyonu (DMF) içermesidir ancak bu yeter şart değildir. Bu şekilde idealize olmuş ölçek hazırlamak ölçme yapmayı amaçlayan araştırmacılar için oldukça zorlu bir süreçtir. Yanlılık için yapılan çalışmalar incelendiğinde bu literatürün 1900'lü yıllara kadar uzandığı görülmektedir. İlk olarak, anadili İngilizce olmayan çocuklara uygulanan mental testlerde anadili İngilizce olan çocukların lehine çalışan maddelerin tespitiyle günümüze kadar bu alanda pek çok çalışma yapılmış ve yapılmaya devam etmektedir. Yapılan bu tespitin ardından takip eden yıllarda dilin, kültürel yapının ve ırksal farklılıkların yanlılığa etkisinin araştırıldığı çalışmalar yapılmıştır. Kültürel farklılıkların yanlılığa etkisi olduğu kabul gören bir gerçek haline geldikten sonra Eells v.d. (1951), kültürel farklılıklardan bağımsız ölçekler geliştirmeye çalışmışlardır. Green (1974), okullarda uygulanan başarı testlerinde mevcut olan etnik yanlılığın ortadan kaldırılması için 2 öneri ortaya koymuştur. Green, norm referanslı testlere bazı prosedürler ekleyerek ve standartlaştırılmış testler yerine kriter referanslı testler kullanarak etnik kökenden kaynaklanan yanlılığın ortadan kalkacağını belirtmiştir. Scheuneman (1979), test maddelerinin yanlılığını belirlemek için toplam puana dayalı bir yöntem sunmuştur. Yöntemi Metropolitan hazırlık testi üzerinde denemiş ve yöntemin literatürde mevcut olan bazı yöntemler ile çeliştiğini ortaya koymuştur. Shepard, Camilli ve Averil (1981), Lorge-Thorndike zeka testini ırksal farklılıklar açısından incelemişler ve DMF analizi için dönüştürülmüş madde güçlükleri, ayırt edicilik, madde karakteristik eğrisi ve χ^2 İstatistiği yöntemlerini karşılaştırmışlardır. Seong, Tae-Je, Subkoviak (1987), Linn ve Harnisch'in Madde Tepki Kuramı (MTK)(Z) yöntemi, Camilli'nin χ^2 tekniği ve Angoff'un dönüştürülmüş madde güçlük indeksi yöntemlerini karşılaştırmışlar ve χ^2

teknikinin yanlılığı tespit etmede (MTK) (Z) yöntemi ile ilişkili olduğunu ancak χ^2 teknikinin daha gerçeğe yakın doğru sonuçlar verdiğini ve Angoff'un dönüştürülmüş madde güçlük indeksi yönteminin diğer yöntemlere göre zayıf olduğunu ortaya koymuşlardır. Raju (1988), MTK modelleri için iki Madde Karakteristik Eğrisi (MKE) arasında kalan işaretli ve işaretsiz alanların hesaplanmasında kullanılmak üzere bazı formüller ortaya koymuş, buna ek olarak şans parametresinin özel durumları için MKE altındaki alanların durumlarını belirtmiştir. Raju, Drasgow ve Slinde (1993), etnik köken ve cinsiyet açısından DMF analizi yapmış ve bu analizlerde Lord'un χ^2 istatistiği, İşaretili ve İşaretsiz alanlar ve son olarak Mantel- Haenszel (MH) yöntemlerini karşılaştırmışlardır. Sadece cinsiyet açısından bakıldığında MH yöntemi ile MTK tabanlı diğer yöntemler benzer sonuçlar verirken etnik köken açısından MTK tabanlı yöntemler benzeşmiş MH tekniği farklı sonuçlar üretmiştir. Dancer, Anderson ve Derlin (1994), log lineer modelleri kullanarak ruh sağlığına etnik köken ve cinsiyet değişkenlerinin etkisi olup olmadığını belirlemek için DMF analizi yapmışlar ve DMF içeren maddeler saptamışlardır. Buna göre; ruh sağlığı problemlili olan kişiler cinsiyet ve etnik kökene göre farklılaşmaktadır. Allalouf, Hambleton ve Sireci (1999), İsrail Psikometrik Giriş Testi'ni farklı dillere çevirerek uygulamışlar ve diller arası DMF analizi yapmışlardır. Maddelerin %34'ünün diller açısından DMF içerdiğini ortaya koymuşlardır. Gelin ve Zumbo (2003), depresyon ölçeği kullanarak maddeleri puanlamada kullanılan sistemlerin DMF'ye neden olup olmadığını araştırmışlar ve puanlama sistemlerinin DMF'ye etkisi olduğu sonucuna varmışlardır. Gök, Kelecioğlu ve Doğan (2010), Mantel-Haenszel ve Lojistik regresyon (LR) yöntemlerinin karşılaştırmasını yapmak için Ortaöğretim Kurumları Sınavı fen ve matematik alt testlerini kullanmışlardır. Cinsiyet ve okul türü açısından yapılan DMF analizinde iki yöntemin az miktarda uyum gösterdiklerini ortaya koymuşlardır. Bakan Kalaycıoğlu ve Kelecioğlu (2011), 2005 Öğrenci Seçme Sınavı'nda (ÖSS) yer alan maddelere cinsiyet açısından DMF analizi yapmış ve DMF saptanan maddeler için uzman görüşüne başvurmuşlardır. Atalay Kabasakal, Gök, Kelecioğlu ve Arsan (2014), MTK Olabilirlik Oranı, SIBTEST ve Mantel-Haenszel yöntemlerinden elde ettikleri sonuçları karşılaştırmıştır. Ayrıca, bu yöntemlerin I. tip hata ve güç değerleri için de karşılaştırmalar yapmışlardır. Özmen (2014), PISA 2009 okuma becerileri testini Türkiye ile Amerika ve Birleşik Krallık için karşılaştırmış ve ülkeler arası DMF içeren maddelere rastlanmış maddelerin DMF içerme nedenleri araştırılmıştır. Türkan (2014), 2012 Seviye Belirleme Sınavı (SBS) maddelerinin cinsiyet değişkenine göre yanlılık içerip içermediğini araştırmış

ve DMF saptanan maddeler bulmuştur. İlgili maddeler için uzman görüşü alındığında maddelerin yanlı olmadıkları belirlenmiştir. Akın Arıkan (2015), PISA 2012 matematik çalışma etiği tutum maddeleri için DMF analizi, MTK Olabilirlik Oranı, Ordinal Lojistik Regresyon ve Poly-SIBTEST yöntemleri kullanılarak gerçekleştirilmiş olup cinsiyete göre yapılan bu analizlerde DMF içeren maddelere rastlanmıştır. Sünbül ve Ömür Sünbül (2016), çalışmasında DMF belirlenmesinde kullanılan yöntemler için I. tip hata ve güç karşılaştırmaları yapmışlardır. Queensoap ve Orluwune (2019), Kimya başarı testinin dört farklı etnik grup açısından yanlı olup olmadığını incelemiştir. Analizlerde Scheuneman'ın modifiye χ^2 yöntemi kullanılmış ve DMF içeren maddeler saptanmıştır. Çalışmada ayrıca kullanılan yöntemin başarısından bahsedilmekte ve uygulamalarda kullanılması önerilmektedir. Hasaıçebi, Terzi ve Küçük (2019), Karadeniz Teknik Üniversitesi İstatistik ve Bilgisayar Bilimleri Bölümü 3. ve 4. sınıf öğrencilerinin olasılık teorisi okuryazarlık düzeylerini, geliştirdikleri olasılık teorisi okuryazarlık testi yardımıyla ölçmüş ve öğrencilerin cinsiyet ve sınıf seviyelerine göre yanlılık analizini yapmışlardır. DMF analizi Lord'un χ^2 istatistiği ve Raju'nun Alan Ölçümleri yöntemleri kullanılarak yapılmıştır. Elde edilen sonuçlardan yola çıkarak her iki yöntemin de aynı maddeleri DMF'li olarak işaretlediğini ortaya koymuşlardır. Sırgancı ve Çakan (2020), PISA 2006 verilerini kullanarak sıralı lojistik regresyon ve poly-SIBTEST yöntemleri ile kültürel ve dil açısından DMF analizleri yapmışlardır. Avustralya, Yeni Zelanda, ABD ve Türkiye'nin dahil olduğu bu karşılaştırmalarda ülkeler arasındaki farklılıklar arttıkça DMF içeren madde sayısında artış olduğunu göstermişlerdir. Her iki yöntemin DMF belirlemede uyumlu olduğu; ancak poly-SIBTEST yönteminin daha hassas sonuçlar verdiğini ortaya koymuşlardır. Başman ve Kutlu (2020), PISA 2012 Matematik okuryazarlık testi üzerinde cinsiyet açısından yanlılık analizi yapmışlar ve cinsiyet açısından yanlı maddeler bulmuşlardır. Ancak, matematik başarısındaki cinsiyet farklılıklarının nedenlerini anlamak için cinsiyet ve içsel motivasyon, araçsal motivasyon, öz yeterlik ve kaygı değişkenlerinin matematik test maddeleri üzerindeki etkileşimi DMF açısından incelenmiş ve bazı duyuşsal etkileşimlerin DMF saptanan maddelerde değişim sağladığını ortaya koymuşlardır.

1.2. Tezin Amacı ve Yapısı

İnsanlar hayatları boyunca pek çok önemli sınava girerler. Anadolu ve Fen liseleri için, üniversite için, İngilizce yeterliliği alabilmek için, devlet kurumlarına atanabilmek için ya da akademik kariyer yapabilmek için. Özellikle geniş kitlelerin girdiği farklı coğrafyalardan katılımcıların olduğu bu sınavlarda en önemli konu sınav sonuçlarının belirli kişilere belirli durumlarından dolayı bir yarar ya da zarar sağlayıp sağlamamasıdır. İşte bu durum ölçme teorisinde madde yanlılığı olarak adlandırılır. Sunulan bu tezde amaç, tarafımızca geliştirilen İstatistiğe Giriş Okuryazarlık Testi ve Karadeniz Teknik Üniversitesi (KTÜ) tarafından geliştirilen ve uygulanan Yabancı Uruklu Öğrenci Sınavı (YÖS) ölçeklerinin içerdiği maddeler için ölçme teorisinin farklı iki kuramına dayalı beş yöntemle DMF analizini yapmak, yeterli test uzunluğu ve etkili örneklem büyüklüğünün sağlandığı ve sağlanmadığı durumlarda KTK ve MTK'ya bağlı yöntemlerin tutarlılıklarını hem birbirleriyle hem de kendi aralarında karşılaştırmak ve DMF analizlerinden elde edilen sonuçlara göre maddeler ve ölçeklerle ilgili yorumlarda bulunmaktır.

2. YAPILAN ÇALIŞMALAR

2.1. Madde Analizi

Geliştirilen ölçeğin güvenilir ve geçerli olmasının yanı sıra ölçekte yer alan maddelerin her biri için analizler yapılarak maddelerin özelliklerinin belirlenmesi de ölçme teorisinin önemli bir adımıdır. Böylelikle maddelerin amaca hizmet edip etmediği belirlenir ve beklenen sonucu vermeyen maddeler için düzeltme yoluna gidilir. Bu süreç madde analiz sürecidir. Madde analizi, test maddeleri ve maddelerin seçenekleri üzerinden yapılan bir çalışmadır ve bu çalışmada maddelerin zorluk dereceleri, bilenle bilmeyeni ayırabilme yetenekleri, çeldiricilerin işlevselliği, test süresinin ve cevaplayıcı kitlesinin uygunluğu gibi birçok öge araştırılır. Elde edilen sonuçlar yardımıyla da en iyi teste ulaşılmaya çalışılır. Çünkü unutulmamalıdır ki; istenen özellikteki ölçüğe giden yol istenen özellikteki maddeleri seçebilmekten geçmektedir. Madde analizinde maddelerin özelliklerini ortaya koymak için maddeyi betimleyen istatistikler hesaplanır. Bunlardan ilki madde güçlük indeksidir. Bu indeks Eşitlik 1'deki gibi elde edilir.

$$P_j = \frac{N_a}{N} \quad (1)$$

N_a , doğru cevaplayanların sayısı, N ve tüm cevaplayıcıların sayısı olmak üzere, P_j 0 ile 1 arasında değer alır ve maddenin zor ya da kolay bir madde olduğu hakkında bilgi verir. P_j değeri 1'e yaklaştıkça madde kolay, 0'a yaklaştıkça zordur denir.

Alt ve üst gruplar yöntemiyle P_j hesaplanırken öncelikle cevaplayıcının verdiği doğru cevaplar için 1 puan yanlış cevaplar ve boş bıraktığı sorular için 0 puan verilerek ham skorlar elde edilir. Bu skorlar en yüksekten düşüğe doğru sıralanır ve en başarılı %27'lik grup üst grup, en başarısız %27'lik grup alt grup olarak belirlenir. Arada kalan %46'luk dilim analiz dışı bırakılır. Madde Güçlük İndeksi b ile gösterilir ve Eşitlik 2'deki gibi hesaplanır (Gömleksiz ve Erkan, 2016).

$$b = \frac{\frac{\text{Maddeyi üst grupta doğru cevaplayanların sayısı}}{\text{Üst gruptaki öğrenci sayısı}} + \frac{\text{Maddeyi alt grupta doğru cevaplayanların sayısı}}{\text{Alt gruptaki öğrenci sayısı}}}{2} \quad (2)$$

İdeal maddeden beklenen davranış b indeksinin 0.50 olmasıdır fakat, bir ölçekteki tüm maddelerin güçlüklerinin 0.50 olması ölçeğin kalitesini düşürecektir. Bunun yerine test maddelerinin güçlüklerinde çeşitlilik olması istenen bir durumdur. Aslında burada amaçlanan maddelerin hepsinin güçlüğü'nün 0.50 olması yerine testin ortalama güçlüğü'nün 0.50 olmasıdır. Böylelikle geliştirilen ölçek daha da güvenilir olacaktır.

Tablo 1. Madde güçlük indeksi için referans aralığı

Madde Güçlük İndeksi	Maddenin Değerlendirilmesi
0.29 ve altında	Zor
0.30-0.49	Orta Güçlükte
0.50-0.69	Kolay
0.70-1	Çok Kolay

Tablo 1'de b parametresi hesaplandıktan sonra çıkan değere göre maddenin güçlük düzeylerinin belirlenmesi için referans aralıkları görülmektedir.

Madde ayırt edicilik indeksi, madde analizi için güçlükten sonra gelen bir diğer istatistiktir. Ayırt edicilik, ilgili maddenin ölçülen örtük özelliğe sahip olan cevaplayıcı ile olmayan cevaplayıcıyı birbirinden ayırt etme gücünü temsil eder. -1 ile +1 arasında değer alır. İndeksin 0'a yaklaşması ilgili maddenin alt ve üst grubu ayırt ediciliği düşük olarak yorumlanırken, 1'e yaklaşması durumunda ise maddenin alt ve üst grubu ayırt ediciliği yüksek yorumu yapılabilir. Aynı zamanda a parametresi madde analizinin yapı taşlarından biri olan MKE'nin eğimi olarak yorumlanır. Ayrıca, yine bu indeks geçerliliğin bir göstergesidir. İndeksin negatif değer alması maddede bir belirsizlik olduğuna işaret edebilir. Bu durumda ilgili madde nihai teste alınmadan önce tekrar incelenmelidir. a parametresi alt ve üst gruplar yöntemiyle Eşitlik 3'deki gibi hesaplanır (Gömlüksiz ve Erkan).

$$a = \frac{\text{Üst grupta doğru cevaplayan sayısı} - \text{Alt grupta doğru cevaplayan sayısı}}{\text{Grupların herhangi birindeki öğrenci sayısı}} \quad (3)$$

Tablo 2. Ayırt edicilik indeksi için düzeyler

Madde Ayırt Edicilik İndeksi	Maddenin Değerlendirilmesi
0.40 ve daha büyük	Çok iyi madde
0.30-0,39	Oldukça iyi ama yine de geliştirilebilir
0.20-0.29	Düzeltilmesi ve geliştirilmesi gerekir
0.19 ve daha küçük	Çok zayıf mutlaka çıkarılmalı

Tablo 2’de a parametresi için referans değer aralığı verilmiştir. Bu aralıklara göre maddenin ayırt edicilik gücü ile ilgili yorumda bulunulabilir

Madde istatistiklerinden madde varyansı, ilgili maddenin ölçülen özellik açısından cevaplayıcılar arasındaki farklılıkları ne kadar ortaya koyabildiğinin göstergesidir. Varyans ne kadar büyükse cevaplayıcılar arasındaki fark o kadar fazla açığa çıkmaktadır. Madde en yüksek 0.25 varyans değeri alabilir (Cronbach, 1984; MacDonald ve Paunonen, 2002; Ural ve Kılıç, 2005). Madde varyansı Eşitlik 4’ deki gibi hesaplanır.

$$S^2 = P_j * q_j \quad (4)$$

Eşitlik 4’ de P_j , madde güçlüğü ve q_j , $(1 - P_j)$ dir.

2.2. Yetenek Parametresinin Tahmini

Ölçme teorisinde amaçlardan biri de ölçme konusu olan özelliğin en doğru şekilde ölçülmesidir. Bu sebeple de bir maddenin nihai teste seçilebilmesi için madde analizi sürecinden başarıyla çıkması gerekmektedir. Ancak, ölçme yapılırken sadece iyi bir test oluşturmakla analiz süreci noktalanmaz. Ölçeği geliştirdikten sonra cevaplayıcıya doğru bir şekilde aktarmak ve buna bağlı olarak cevaplayıcıların özelliklerini belirlemek de ölçmenin ana amaçlarından biridir. Böylelikle cevaplayıcıların ilgili örtük özellik ile ilgili yeteneklerinin ölçek üzerindeki konumları belirlenmektedir. Kişinin yetenek parametresinin tahmin edilmesi demek o kişiyle ilgili örtük özellik açısından yorum

yapabilmek demektir. Ayrıca, özellikle eğitim bilimlerinde, yetenek parametresinin tahmininden yola çıkılarak kişilerin birtakım kazanımlar elde etmesi de sağlanabilir. Bu anlamda madde parametrelerinin tahmini kadar yetenek parametreleri de ölçme teorisinde önem kazanmıştır. Yetenek parametresi belirlenirken, madde parametrelerinin bilindiği varsayılır. Bu varsayım araştırmacılara yetenek tahminleri ile madde parametre tahminlerini aynı ölçek üzerinde değerlendirme imkanı sunmaktadır. Bireylere herhangi bir örtük özelliği ölçmesi için hazırlanmış belirli sayıda madde içeren bir test uygulandığında, cevaplayıcılardan beklenen maddelerin her birini yanıtlamalarıdır. İkili ya da çoktan seçmeli testlerde cevaplayıcının verdiği yanıtlar 0-1 şeklinde kodlanır. Bu kodlama doğru yanıt için 1 değerinin atanması, yanlış ya da boş bırakılmış maddeler için 0 değerinin atanması şeklindedir. Bu kodlamanın ardından cevaplayıcıların yanıtları araştırmacıya madde tepki vektörünü vermektedir. İşte bu madde tepki vektörü ve önceden tahmin edilmiş madde parametreleri kullanılarak kişiler hakkında yetenek tahminleri yapılmaktadır. Yetenek tahmini yapılırken genellikle en çok olabilirlik yöntemi kullanılmaktadır.

$$\theta_{t+1} = \theta_t + \frac{\sum_{i=1}^N -a_i [y_i - P_i(\hat{\theta}_t)]}{\sum_{i=1}^N a_i^2 P_i(\hat{\theta}_t) Q_i(\hat{\theta}_t)} \quad (5)$$

Eşitlik 5'te yetenek parametresi tahmininin hesaplanması görülmektedir. $\hat{\theta}_t$, cevaplayıcının yetenek kestirimidir. a_i , i . maddenin ayırt edicilik düzeyidir. y_i , cevaplayıcının i . maddeye verdiği 0-1 şeklinde kodlanmış yanıtıdır. $P_i(\theta_t)$, i . maddeye doğru yanıt verme olasılığı iken, $Q_i(\hat{\theta}_t)$, i . maddeye yanlış yanıt verme olasılığıdır.

Kişinin yetenek kestirimi yinelemeli işlemlerden oluşmaktadır. İlk olarak, θ_t 'ye keyfi bir başlangıç değeri atanarak süreç başlar ve testteki tüm maddelere ilişkin atanan bu yetenek değerinde doğru yanıt verme olasılıkları hesaplanır. Eşitlik 5'in sağ kısmındaki ikinci terim ise bir düzeltme terimidir ve $\Delta\theta$ olarak değerlendirilir. Burada amaç, $y_i - P_i(\hat{\theta}_t)$ farkının en küçüklenmesidir. Yapılan tahmin gerçek değere yaklaştıkça bu fark küçülür ve bir noktada $\Delta\theta$ o kadar küçülür ki bir sonraki adımda elde edilen değer ile bir önceki adımdan gelen değer arasındaki fark ihmal edilebilir düzeyde olur. İşte bu noktada hesaplanan değer kişinin yetenek kestirimi olarak kullanılır (Baker, 2001).

2.3. Lojistik Regresyon

Araştırmacılar çalışmalarında bağımlı ve bağımsız değişkenlerin arasındaki ilişkiyi analiz etmektedir. Burada, doğru yöntemin uygulanabilmesi verinin yapısına bağlıdır ve ancak bu şekilde en doğru sonuçlar alınabilmektedir. Bağımlı değişkenin kesikli yapıda olması doğrusal modeller için olmazsa olmaz normallik varsayımını bozmakta ve bu modellerin işlevselliğini ortadan kaldırmaktadır. Bağımlı değişken kategorik bir yapıda olduğunda en küçük kareler (EKK) tekniği yetersiz kalmaktadır. Çünkü varyanslar artık minimum olmayacaktır. Bu sebeple, bu tekniğe alternatif olarak lojistik regresyon modelleri mevcuttur. Burada da amaç aslında diğer tekniklerle aynıdır. Regresyonun temel mantığında olduğu gibi bağımlı ve bağımsız değişkenler arasındaki ilişkiyi minimum değişken sayısı ile ifade eden yüksek uyuma sahip bir model kurmak temel amaçtır (Atasoy, 2001). Doğrusal regresyon analizinde bağımlı değişkenin değeri, lojistik regresyon analizinde ise bağımlı değişkenin alabileceği değerlerden birinin gerçekleşme olasılığı tahmin edilmektedir (Kleinbaum, 1994).

Eğitim bilimleri, psikoloji, tıp, biyoloji, ekonomi gibi disiplinlerde yapılan araştırmalarda veriler genellikle kesiklidir. Örneğin, bir kişinin öğrenme güçlüğü çekip çekmemesi, akciğer kanseri bir hastanın sigara kullanıp kullanmaması, ergenlerin riskli davranışlara eğilimleri olup olmaması, bir tümörün kanser olup olmadığı, bir hastalığın tedaviye cevap verip vermediği gibi ifade edilebilir (Kılıç, 2000; Peng, Lee ve Ingersoll, 2002; Arabacı, 2002; Field, 2005; Mertler ve Vannatta, 2016).

$$\pi(x) = P(Y = 1/X = x) = \frac{e^{(\beta_0 + \beta_1 x)}}{1 + e^{(\beta_0 + \beta_1 x)}} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}} \quad (6)$$

Eşitlik 6 doğrusal olmayan bir lojistik regresyon fonksiyonudur. Lojistik regresyon fonksiyonları lojit dönüşüme uğradıklarında doğrusallaşır. Lojit dönüşüm bir olayın oddsunun logaritmasının alınması işlemidir. Bir olayın oddsu, olayın gerçekleşme olasılığının olayın gerçekleşmeme olasılığına oranıdır. Burada, $\pi(x)$ 0 ile 1 arasında değerler alan bir olasılıktır. Yani Eşitlik 6 X 'in değeri bilindiğinde Y 'nin 1 olması olasılığını verir.

$$\frac{\pi(x)}{1 - \pi(x)} = e^{(\beta_0 + \beta_1 x)} \quad (7)$$

Lojit dönüşüm yapılırken ilk adım olayın odssunun alınmasıdır. Eşitlik 7'de π olayının odssunun nasıl hesaplandığı görülmektedir. Odds alınarak 0 ile 1 arasında değer alan $\pi(x)$ artık 0 ile ∞ arasında değer alacaktır.

$$\text{lojit}_{\pi(x)} = g(x) = \ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \ln(e^{(\beta_0 + \beta_1 x)}) = \beta_0 + \beta_1 x \quad (8)$$

Eşitlik 8'de ikinci adım olarak odds değerinin logaritmasının alınması görülmektedir. Bu adımda lojit dönüşüm tamamlanmış olur ve lojitler $-\infty$ ile $+\infty$ arasında tanımlı hale gelir. Eşitlik 8'de görülüyor ki lojit dönüşüm parametrelerin doğrusal bir fonksiyonunu oluşturmuştur. Lojistik regresyonun tek varsayımı lojit ile bağımsız değişkenler arasındaki ilişkinin doğrusal olmasıdır. Doğrusal regresyon modellerinde β_1 , bağımsız değişkendeki 1 birim değişimin bağımlı değişkende ne kadarlık bir değişime neden olduğunu gösterirken, lojistik modelde bağımsız değişkendeki 1 birim değişimin lojitte ne kadar değişime neden olduğu gösterilir. Doğrusal regresyonda artıklar 0 ortalama ve sabit varyansla normal dağılırken, lojistik regresyonda artıklar sıfır ortalama $\pi(1 - \pi)$ varyansla binom dağılıma uygundur (Hosmer ve Lemeshow, 1989).

2.3.1. Regresyon Katsayılarının Tahmini

2.3.1.1. En Çok Olabilirlik Tahmin Yöntemi

Lojistik regresyon analizinde regresyon katsayılarının tahmini genellikle en çok olabilirlik yöntemi kullanılarak yapılmaktadır. Yöntemin kullanılabilmesi için öncelikle olabilirlik fonksiyonu oluşturulmalıdır. Olabilirlik fonksiyonu gözlenen verinin olasılığını bilinmeyen parametrelerin bir fonksiyonu olarak ifade eder. Yöntemin ana amacı gözlenen veri setini elde etme olasılığını en büyük yapacak şekilde bilinmeyen parametreler için değerler üretmektir. Olabilirlik fonksiyonunu en büyük yapan değerler bilinmeyen parametrelerin en çok olabilirlik tahmin edicileridir. Bir başka deyişle bu yöntemle bir olayın olma olasılığını en büyük yapmak amaçlanmaktadır (Alpar, 2013).

Örneklem büyüklüğü sonsuza giderken en çok olabilirlik tahmin edicileri tutarlı, yeterli ve asimptotik normaldir. Yöntem küçük örneklem için de çalışmaktadır ancak, değişken sayısının 10 katı kadar veri olması istenen durumdur (Alpar, 2013).

2.3.1.2. Bayes Yöntemi

Bayes teoremi örnek bilgileri toplandıktan sonra parametreler hakkında olasılıksal yorumlar yapmayı sağlayan bir teoremdir. Araştırmacı verileri elde ettikten sonra parametrelerin koşullu dağılımı sonsal dağılım olarak değerlendirilir. Sonsallar örnek bilgisi ve önsel bilgilerin birleşimiyle ortaya çıkar. Parametrelerin önsel bilgilerinden yola çıkılarak belirlenen önsel dağılımlar modelleme sürecine katılırlar. Bayesci yaklaşımda önsel bilgilerden önsel dağılımlara sonsal dağılımlardan da sonsal bilgilere uzanan bir akış söz konusudur. Bu yaklaşımla örneklemden elde edilen olabilirlik fonksiyonu ile belirlenen önsel dağılım birleştirilerek çıkarsamalar yapmak mümkündür. Günümüzde Bayesci yaklaşım karar vermede oldukça önemli bir yere sahiptir.

$$p(\theta/y) \propto L(\theta)\pi(\theta) \quad (9)$$

Parametreler için Eşitlik 9 ile ifade edilen sonsal dağılımdan sonsal bilginin elde edilmesi ve elde edilen bilginin anlamlı bir şekilde kullanılması gerekmektedir. Bunun için de Eşitlik 9'da görüldüğü gibi önsel dağılıma ve örnek bilgisine ihtiyaç duyulduğu da açıkça görülmektedir. Önsel dağılımın seçimi bu anlamda Bayesci yaklaşım için oldukça önemlidir. Önsellerin seçimi araştırmacının tecrübelerinin yanında belirli teorik kurallara bağlı olarak seçilir. Son yıllarda gelişen teknoloji ile en karmaşık problemlerde dahi Bayesci yaklaşımın kullanımı ortaya konan yeni yöntemlerle mümkün kılınmıştır (Kumru, 2003).

2.4. Güvenilirlik ve Geçerlilik

Herhangi bir örtük özellik ölçülmek istendiğinde doğrudan ölçme yapılamayacağından dolayı ilgili özelliği ölçmek üzere bir test geliştirilmesi ya da geliştirilmiş mevcut hazır ölçeklerin kullanılması gerekmektedir. Konuyla ilgili karar verme süreçlerinde testin uygulandığı bireylerin verdikleri tepkilerden yola çıkılır.

Bireylerin verdikleri tepkilerin doğru ölçülmesi ancak doğru ölçüm yapan bir ölçek ile mümkündür. Bir ölçme aracının doğru ölçüm yapması ise güvenilir ve geçerli bir ölçek olması ile doğrudan ilişkilidir (Ercan ve Kan, 2004).

2.4.1. Güvenilirlik

Güvenilirlik kavramı tutarlılık ve duyarlılık kavramlarının toplamı olarak ifade edilebilir. Güvenilirlik, geliştirilen bir ölçeğin mevcut tüm maddelerinin birbirleriyle tutarlılıklarını ortaya koyar. Güvenilir bir ölçekle yapılan ölçümler tekrarlanabilir olmalı ve tekrarlardan elde edilen sonuçlar birbirleriyle tutarlı olmalıdır. Elde edilen sonuçlar zamana, mekana, maddeye ve puanlayıcıya göre farklılık göstermemelidir. Duyarlılık ise ölçüm sonuçlarının hassaslığını ifade eder yani duyarlılık, ölçeğin araştırılan özelliği ne kadar hassas ölçtüğü güvenilirliğin bir parçasıdır. Ölçümlerde hassasiyet arttıkça ölçeğin güvenilirliği de artacaktır. Bu duruma bir bireyin ağırlığının ölçümünde klasik tartı kullanılmasıyla dijital tartı kullanılması örnek verilebilir. Dijital tartı daha hassas ölçümler yapabildiği için klasik tartıya göre daha güvenilir bir ölçme aracıdır. Buradan hareketle, bir ölçekte duyarlılık arttıkça güvenilirliğin artacağı söylenebilir. Güvenilirliğin artması demek gerçeğe daha yakın ölçüm sonuçlarının elde edilmesi anlamına gelir. Güvenilirlik ölçeğin ölçüm hatalarından ne denli uzak olduğu ile ilişkili bir kavramdır. Ölçek, ölçümü hatalardan ne kadar uzak yapabilirse o kadar güvenilirdir (Ercan ve Kan, 2004).

Güvenilirlik katsayısının hesaplanmasında tekrar eden ölçümlerden alınan sonuçların birbirleriyle ilişkisi oldukça önemlidir. Güvenilir bir ölçekten, bir uygulamada başarılı olanların tekrar eden ölçümlerde de başarılı olmaları beklenir. Yani, ölçümler arasındaki korelasyon güvenilirliği işaret etmektedir.

$$r_{xy} = \frac{n \sum XY - (\sum X^2)(\sum Y^2)}{\sqrt{[n \sum X^2 - (\sum X)^2][n \sum Y^2 - (\sum Y)^2]}} \quad (10)$$

Eşitlik 10 ile güvenilirliğin aynı ölçümün tekrar eden iki gözlemi arasındaki korelasyona bağlı olduğu belirtilmektedir. Burada, n , kişi sayısı olmak üzere $\sum XY$, tekrar eden gözlemlere ilişkin ölçümlerin çarpımlarının toplamı, $\sum X^2$, ilk gözlemden elde edilen ölçümlerin karelerinin toplamı, $(\sum X)^2$, ilk gözlemden elde edilen ölçümlerin toplamının

karesi, $\sum Y^2$, ikinci gözlemden elde edilen ölçümlerin karelerinin toplamı ve $(\sum Y)^2$, ikinci gözlemden elde edilen ölçümlerin toplamının karesi olarak ifade edilir.

2.4.1.1. Test- Tekrar Test Yöntemi

Uygulanan testin aynı kişilere, belli zamanlarda ve benzer koşullar sağlanarak uygulanmasına dayalı bir yöntemdir. Burada tekrar eden ölçümler arasındaki zaman önem kazanmaktadır. İki ölçüm arasındaki süre uzadıkça alınan ölçüm sonuçlarının arasındaki korelasyon düşebilmektedir. Testin tekrar uygulanmasında geçen zaman ölçülen özelliğin miktarında değişim yaratmayacak kadar tutulmalıdır. Tekrarlı ölçümlerden alınan sonuçların korelasyonları pozitif ve 1'e ne kadar yakınsa ölçek o kadar güvenilirlerdir. Bu yöntem daha çok zekâ, yetenek, kişilik, tutum testlerinde kullanılmaktadır (Gömleksiz ve Erkan, 2016; Arıkan vd., 2017).

2.4.1.2. Paralel Formlar Yöntemi

Bu yöntemde, ölçme aracı hazırlanırken eş değerde olan yani aynı örtük özelliği ölçen, aynı aritmetik ortalama ve standart sapmaya sahip maddelerle bir ölçek daha hazırlanır. Geliştirilen bu iki ölçek farklı zamanlarda aynı gruba uygulanır. Elde edilen ölçüm sonuçlarının korelasyon katsayısı hesaplanır ve yüksek olması halinde iki ölçümün tutarlı olduğu sonucuna varılır. İki test aynı zamanda uygulanmışsa hesaplanan güvenilirlik katsayısına eşdeğerlik katsayısı, farklı zamanlarda yapılmışsa kararlılık katsayısı denir (Gömleksiz ve Erkan, 2016; Arıkan vd., 2017).

2.4.1.3. Test Yarılama Yöntemi

Bu yöntem testin bir kez uygulanması ile güvenilirlik katsayısının tahmin edildiği bir yöntemdir. Test, rastgele iki eşit yarıya bölünür. Ayrılan bu testlerden elde edilen toplam puanlar arasındaki ilişki incelenerek iki testten elde edilen ölçümlerin tutarlılığı saptanmış olur. Elde edilen bu katsayı tüm test için genellenerek güvenilirlik katsayısına ulaşılır. Bu genelleme ise alt testlerin varyanslarının homojen olup olmamasına göre değişkenlik göstermektedir.

$$r_{xx} = \frac{2r_{12}}{1 + r_{12}} \quad (11)$$

Eşitlik 11 ile varyansların homojen olduğu durumda toplam test güvenilirliği için formül verilmiştir. Burada, r_{12} , alt testler arasındaki korelasyonu ifade eder.

$$r_{xx} = \frac{4r_{12}S_1S_2}{S_x^2} \quad (12)$$

Eşitlik 12 ile varyansların homojen olmadığı durumlar için test güvenilirliğinin hesaplanması gösterilmektedir. Burada, r_{xx} , tüm testin güvenilirliği, S_1 , 1. alt testin standart sapması, S_2 , 2. alt testin standart sapması ve S_x^2 , toplam testin varyansını ifade eder.

Bu yöntemde dikkat çekilmesi gereken nokta iki yarı arasındaki korelasyonun tüm testin korelasyonu için bir alt sınır oluşturmasıdır. Eğer, r_{12} değeri güvenilir olarak hesaplandıysa tüm testin güvenilirliğinin hesaplanmasına gerek yoktur (Ebel, 1972; Gömleksiz ve Erkan, 2016).

2.4.1.4. İç Tutarlılık Katsayısı

Bu yöntemde testi tekrar etmek, eş değer formlar oluşturmak ya da testi yarılamak yerine test maddelerinin birbirleriyle uyumu incelenir. Testi oluşturan maddelerin birbiriyle uyumu iç tutarlılığı ifade eder. Testin kendi içinde tutarlı bir yapıya sahip olması güvenilir bir test olduğu anlamına gelir prensibi ile çalışmaktadır. Bu yöntemle testin mümkün tüm yarıları arasındaki güvenilirlik değerlerinin ortalaması hesaplanır (Gömleksiz ve Erkan, 2016). Ölçeğin 1 kez uygulandığı durumlarda kullanıma uygundur. Maddelere verilen tepkiler ile toplam skorun ilişkisine yönelik bir katsayıdır. Genellikle tek boyutlu yapıların ölçüldüğü ölçme araçlarının güvenilirlikleri hesaplanırken kullanılır (Büyüköztürk vd., 2016).

İç tutarlılık yöntemleri yanıtların ikili şekilde verilebildiği ölçek tiplerinde ve daha çok bilişsel testlerde kullanılır. Madde güçlüklerinin farklı olduğu durumlarda Kuder Richardson-20 (KR_{20}) yöntemi kullanılır. Bu hesaplama madde güçlükleri de dahil edilir. KR_{20} değerinin yüksek çıkması kullanılan ölçekteki maddelerin ortak bir yapıyı

ölçtüklere anlamı taşımaktadır. Başka bir deyişle; KR_{20} değeri büyükse ölçek tek boyutludur.

$$KR_{20} = \frac{K}{K-1} \left[1 - \frac{\sum pq}{S_x^2} \right] \quad (13)$$

Eşitlik 13 ile KR_{20} formülü verilmiştir. Burada, pq , madde varyansını, K , madde sayısını ve S_x^2 , test puanlarının varyansını ifade etmektedir. $KR_{20} > 0.70$ olduğunda ölçeğin iç tutarlılığının yüksek olduğu kabul edilir.

Kuder Richardson-21 (KR_{21}) ise test maddelerinin güçlüklerinin eşit olduğu kabulü altında hesaplamalar yapar. Bu yöntemde madde istatistikleri değil test istatistikleri kullanılır.

$$KR_{21} = \frac{K}{K-1} \left[1 - \frac{K\bar{X} - \bar{X}^2}{KS_x^2} \right] \quad (14)$$

Eşitlik 14 ile KR_{21} formülü verilmiştir. \bar{X} , test puanlarının ortalamasını ifade eder. Eşitlik 14'en de görüldüğü gibi formülde sadece test istatistiği kullanılmıştır. $KR_{21} > 0.70$ değerlerini alması tutarlılığın göstergesi olarak kabul edilmektedir (Gömleksiz ve Erkan, 2016; Arıkan vd., 2017).

2.4.1.5. Cronbach α Katsayısı

Tek uygulamaya dayanan bu yöntem KR_{20} ve KR_{21} yöntemlerinin genel hali olarak değerlendirilebilir. Bu yöntem ölçekteki maddelerin homojen bir yapıyı ölçmek üzere bir araya gelip gelmediklerini araştırmaktadır. Katsayının yüksek çıkması durumunda test maddelerinin tutarlı olduğu ve araştırma konusu olan örtük özelliği ölçtüğü yorumu yapılabilir. Güvenilirliği en iyi yansıtan katsayıdır. Cronbach α katsayısı için yapılan açıklamalarda katsayının negatif olamayacağı ve kullanım alanının likert tipli ölçeklere uygun olduğu belirtilir. Fakat Bademci'ye (2006a, 2006b) göre, Cronbach α katsayısı 0-1 kodlanan ikili maddelerde kullanılabilir ve negatif değerler alabilmektedir.

$$\alpha = \frac{K}{K-1} \left[1 - \frac{\sum S_i^2}{S_x^2} \right] \quad (15)$$

Eşitlik 15 ile Cronbach α katsayısının hesaplanması verilmiştir. Burada, K , madde sayısını, S_i^2 , her bir madde puanının varyansını ve S_x^2 , test puanlarının varyansını ifade eder.

Tablo 3. Cronbach α için referans değerleri

Referans Değeri	Güvenilirlik Değerlendirmesi
$0 \leq \alpha \leq 0.40$	Güvenilir Değil
$0.40 \leq \alpha \leq 0.60$	Düşük Güvenilir
$0.60 \leq \alpha \leq 0.80$	Güvenilir
$0.80 \leq \alpha \leq 1$	Yüksek Güvenilir

Tablo 3 ile Eşitlik 15 yardımıyla elde edilen değerlerin yorumlanması için Cronbach α referans değerleri verilmiştir. Tablo 3 ile ölçeğin güvenilirliği ile ilgili yargıya varılır. Güvenilirlik değerinin düşük olması durumunda güvenilirliğin arttırılmasına yönelik adımlar uygulanabilir (Gömleksiz ve Erkan, 2016; Arıkan vd., 2017).

2.4.2. Geçerlilik

Cronbach'a (1984) göre geçerlilik, kullanılan ölçeğin araştırılmak istenen örtük özelliği ölçüp ölçmediğini belirlemek için gereken kanıtların toplandığı bir süreçtir. Geçerlilik ile ölçeğin uygulanmasından sonra alınan ölçüm sonuçlarının farklılığının ne kadarının bireysel farklılıklardan kaynaklandığı önem kazanmaktadır. Bu uygulanan ölçek için istenen bir durumdur ancak, uygulamada bunun tamamen gerçekleşmesi pek mümkün değildir. Ölçme sonuçlarına ölçülmek istenen özelliğin dışında başka özelliklerin karışması çoğu zaman önlenemez. Ancak, ölçümler arasındaki farklarda bireysel farklılıklardan kaynaklanan değişkenlik arttıkça ölçeğin geçerliliği yükselecektir. Daha açık bir şekilde geçerli bir test ölçülmek istenen yapıyı diğer yapıların etkisini elemine ederek ölçebilen testtir (Turgut ve Baykul, 2012). Geçerliliğin düşük olması güvenilirliğin düşük olması ile ilişkilendirilebilir. Güvenilirlik geçerlilik için gerek koşuldur ancak yeter koşul değildir.

Geçerliliğin bir tür kanıt toplama işi olduğunu kabul ederek kanıtları toplamak için çeşitli yollara başvurmak geçerlilik türlerini ortaya çıkarmıştır (Turgut ve Baykul, 2012).

2.4.2.1. Kapsam Geçerliliği

Baykul'a (2015) göre kapsam geçerliliği, testin ölçmeyi amaçladığı yapıyı temsil edebilme kapasitesidir. Bir ölçeğin kapsam geçerliliğine sahip olabilmesi için ölçek ölçülecek özelliğin her yönünü araştırarak maddelerden oluşmalıdır. Örneğin; bir ders ile ilgili sınıf içi uygulamada geliştirilen test ilgili dersin tüm kazanımlarını ölçecek maddelerden oluşmalıdır. Bu tür geçerliliğin tespiti ölçme aracı ile kapsamı arasındaki bağın gücüne dayanır. Bu bağ uzman görüşü alınarak, belirtke tablosu kullanılarak ya da istatistiksel olarak hesaplanabilir. Belirtke tablosu, hedeflenen davranışlar ve program içeriklerinin bulunduğu bir çizelgedir. İstatistiksel olarak kapsam geçerliliğini araştırırken önceden güvenilirliği ve geçerliliği kanıtlanmış aynı kapsamı ölçen bir test ile korelasyon hesaplanarak kapsam geçerliliği ifade edilir (Gömleksiz ve Erkan, 2016).

2.4.2.2. Ölçüt Geçerliliği

Ölçüt geçerliliğinin belirlenmesinde kullanılan aracın önceden geçerliliği kanıtlanmış bir ölçekle karşılaştırılması yoluna gidilir. Eğer iki testten elde edilen puanlar arasındaki ilişki yüksekse ölçüt geçerliliği sağlanmıştır denir. Eldeki ölçek ile geçmişte uygulanmış ve geçerliliği kanıtlanmış ya da mevcut ölçekle eşzamanlı hazırlanmış ve geçerliliği kanıtlanmış ölçek performans açısından karşılaştırıldığında bu tip geçerliliğe uyum geçerliliği denir. Ölçüt geçerliliğinin bir diğer çeşidi yordama geçerliliğidir. Burada eldeki ölçeğin gelecekteki başarıyı ölçme gücü önem kazanır. Örneğin; Lise Geçiş Sınavı (LGS)'na hazırlanan bir öğrencinin sınavdan önce girdiği denemelerde aldığı puan ile LGS'de aldığı puan benzer ise deneme sınavının yordama geçerliliği sağlanmıştır (Arıkan vd., 2017).

2.4.2.3. Yapı Geçerliliği

Geliştirilen test ölçülmek istenen özelliği ölçmede ne kadar başarılı sorusunun cevabı yapı geçerliliği ile açıklanabilir. Özellikle duyuşsal alanda yapılan çalışmalarda yapı geçerliliğinden söz etmek mümkündür. Test, ölçülmek istenen psikolojik yapıyı ölçmeye hizmet edecek maddelerden oluşuyorsa o aracın yapı geçerliliği yüksek olacaktır. Yapı geçerliliği için ideal test ile uygulanan test arasındaki farkın göstergesi demek yanlış olmayacaktır. Yapı geçerliliğini belirlenmesinde faktör analizi ölçekteki maddelerin aynı yapıyı ölçüp ölçmediklerini belirlemek için, gruplar arası ayırıştırma yöntemini, testin, ölçtüğü yapı yönünden birbirinden farklı olduğu bilinen iki gruptan farklı sonuçların alınıp alınmayacağına gözlemlenebilmesi için, istatistiksel yöntemler, test puanları ile aynı yapıyı ölçen, geçerliği yüksek olduğu bilinen bir başka testten elde edilen puanlar arasındaki korelasyonun belirlenmesi için ve testi alan katılımcılarla görüşülmesi, yanıt verilen her bir maddenin kişiye ölçülmek istenen yapıyı düşündürüp düşündürmediğinin belirlenmesi için kullanılır (Gömleksiz ve Erkan, 2016; Arıkan vd., 2017).

2.5. Klasik Test Kuramı

Ölçme teorisinin tarihi incelenirse iki temel kuramın etkilerini görmemek imkansızdır. Bu kuramlardan ilki Klasik Test Kuramıdır (KTK). KTK, ilk olarak 1905 yılında psikolog Charles Spearman tarafından temelleri atılan bir kuramdır. Spearman, aynı yıl ilk kez kuramın matematiksel modelini ortaya koymuştur. Ardından Yule (1911) ve Kelley (1916) Spearman'ın ortaya koyduğu matematiksel modeli geliştirmek üzere çalışmalar yapmışlardır. Binet ve Simon (1916), kurama sistematik bir yaklaşım getirmişlerdir. Kuder ve Richardson (1937) KR_{20} ve KR_{21} formüllerini ortaya koymuşlardır. Lord ve Novick (1968), yöntemi oldukça geliştirerek yöntemin en yaygın kullanıma ulaşmasını sağlamışlardır. Ölçme adımları olan test geliştirme, uygulama, değerlendirme gibi tüm adımlarda KTK'nın kullanıldığı görülmektedir. Ayrıca; başarı, yetenek, tutum gibi birçok örtük özelliğin ölçülmesinde de KTK uygulamalarına literatürde sıkça rastlanmaktadır (Hambleton, Swaminathan ve Rogers 1991).

$$X_i = T_i + E_i \quad (16)$$

Eşitlik 16 ile KTK'nın matematiksel kuruluşu verilmektedir. X_i , gözlenen skor; T_i , gerçek skor ve E_i , ölçmeye karışan rastgele hatadır. KTK'nın matematiksel modeli incelendiğinde Spearman'ın öne sürdüğü bu kuramda gerçek skorun gözlenen skor ve ölçmeye karışan rastgele hataların yardımıyla tahmin edilmek üzere kurulu olduğu görülmektedir. Kuramın temelinde gözlenen test puanı, gerçek puan ve rastgele hata bileşenleri yatmaktadır. Hatasız ölçme olmaz kabulü ile yapılan ölçülerde ölçülmek istenen gerçek puan tam olarak ölçülemez. Gözlenen puanlar yoluyla tahmin edilir. Ölçmeye karışan hata ne kadar az ise yapılan tahminler gerçek değere o kadar yakın olur (Baykul, 2015).

Eşitlik 16'da matematiksel modelden de anlaşılacağı üzere gerçek puanların tahmini yapılırken madde istatistikleri kullanılmamaktadır. Ölçmenin standart hatası kişiden kişiye değişmez, tüm cevaplayıcılar için eşittir. Doğrusal bir model olan KTK, maddeyi değil testi baz alır.

KTK modelleri de tüm modeller gibi bazı varsayımlar altında çalışmaktadır. Gerçek puan sayılabilir sonsuz kere tekrarlanmış ölçümlerden elde edilen gözlenen puanların beklenen değerine eşittir. Gerçek puan ile hata arasında korelasyon sıfırdır. Herhangi ardışık iki dizi ölçmeye ait hata puanları arasında ilişki yoktur. Belirtilen varsayımların sağlanması durumunda KTK modellerini kullanmak mümkündür (Crocker ve Algina, 2008). KTK, doğrusal bir model olduğu için diğer kuramlara göre hesaplamalar açısından daha kolay olmasından dolayı kullanım alanı oldukça geniştir.

Zamanla yapılan çalışmalarda görülmüştür ki KTK bazı durumlarda ölçümler için yeterli ölçüde hassas ve objektif sonuçlar verememektedir. Bunun nedeni KTK üzerindeki sınırlılıklardır. Klasik test kuramında test ve madde istatistikleri madde güçlük düzeyi (p) ve madde ayırt edicilik indeksidir (r_{ix}). Bu istatistikler hesaplanarak madde ve teste ait özellikler hakkında bilgi elde edilmektedir (Osterlind, 1989). Kuramda bahsedilen madde istatistikleri ölçüm yapılan gruptan elde edilir ve teste ait özellikler de madde istatistiklerinden yola çıkılarak elde edilir. Bu sınırlılık madde ve test özelliklerinin birbirinden ayrılamaması sorununa neden olur. Eğer bir testin güçlük düzeyi yüksek ise kişinin yetenek seviyesi düşük, eğer testin güçlüğü düşük ise kişinin yetenek seviyesi yüksek olarak tahmin edilebilir ve bu her zaman gerçeği yansıtmayabilir. Ayrıca, cevaplayıcıların yetenek seviyeleri de test maddelerinin güçlük düzeyini belirlemektedir çünkü KTK'da maddelerin ayırt ediciliği ve güçlüğü belli bir grup için geçerlidir. Madde istatistikleri cevaplayıcıdan bağımsız hesaplanamaz. KTK'da parametre değişmezliği

özelliği yoktur. Parametre değişmezliği, güçlük, ayırıcılık gibi madde parametrelerinin yetenek dağılımından bağımsız olması ve benzer şekilde, bireylerin yetenek dağılımının da madde parametrelerinden bağımsız olmasıdır (Price, 2017). Bu durumda KTK'da çalışırken grup değiştiğinde madde parametreleri de değişecektir. Bu nedenle farklı testleri alan kişileri KTK'ya göre karşılaştırmak oldukça güçtür. Test puanları farklı büyüklüklerde hata içerebilir. Aynı sayıda fakat farklı güçlükteki test sorularını yanıtlayan iki kişinin aynı yetenekte olduğu ileri sürülemez. Belirtilen tüm bu olumsuzluklara rağmen KTK, hem kendisi kullanılarak yapılan çalışmalar vesilesiyle hem de sınırlılıklarından doğan çalışmalarla ortaya çıkan alternatif kuramlara yardım etmesi sebebiyle ölçme teorisine büyük katkıda bulunmuştur.

2.6. Madde Tepki Kuramı

KTK'nın belirtilmiş sınırlılıkları yapılan ölçümlerin hassaslığına ve objektifliğine gölge düşürmüştür. Özellikle, geniş kitlelere uygulanan, sonuçlarına göre kişiler hakkında karar verilen ölçümlerde objektiflik ve hassasiyetin yeterli miktarda olması bu ölçümlerin değerlendirilmesinde kabul edilemez bir durumdur. Sınırlılıkları ortadan kaldıran yeni ve daha güçlü bir kurama ihtiyaç işte bu noktada doğmuştur. Bu ihtiyaçtan hareketle yapılan çalışmalar sonucu Madde Tepki Kuramı (MTK) literatüre kazandırılmıştır. Birçok alanda yapılan ölçümlerin ana amacı ölçülen bireylerin ilgilenilen örtük özelliğe ne derece sahip olduklarının belirlenmesidir. MTK, bu anlamda işini başarıyla yapan bir kuramdır. Ölçülen özellik açısından bireyin yeteneğini en iyi yansıtacak şekilde ölçme aracının düzenlenmesine olanak sağlar (Hambleton ve Swaminathan, 1985). MTK modelleri, bir ölçme aracı vasıtasıyla ölçülmek istenen örtük özellik ile maddeye verilen cevap arasındaki ilişkiyi gösterir. MTK, model temelli ölçümlere dayanmaktadır (Embretson ve Reise, 2000). MTK sayesinde farklı test formlarından alınan puanlar aynı metrik üzerinde konumlanır ve böylece her cevaplayıcı kendi bireyselleştirilmiş test formuna sahip olabilir. Bireylerin maddelere verdikleri yanıtlar maddelerin ve kişilerin özellikleri ile ilgili bilgiler sunmaktadır. MTK ile farklı yetenek düzeylerindeki cevaplayıcıların maddelere verdikleri yanıtların matematik modellemesi yapılır. Bu anlamda MTK, ölçme teorisine istatistiksel bir yaklaşım sağlar ifadesi kullanılabilir (De Gruijter ve Van der Kamp, 2008).

Spearman'ın KTK'nın temellerini atmasından 30 yıl sonra kuramın sınırlılıklarını giderebileceği iddiasıyla MTK'nın ilk çalışmaları başlamıştır. Tucker (1946), MTK'nın

yapıtışı olarak kabul edilen Madde Karakteristik Eğrisi (MKE) terimini literatüre kazandırmıştır. Lord (1952), MTK üzerinde çalışmalarına ağırlık vererek Normal Ogive modeli geliştirmiştir. Birnbaum (1957), Normal Ogive modelin yerine Lojistik Modeli geliştirmiştir. Rasch (1960), kişi ve madde parametrelerinin ayrı ayrı tahmin edilebileceği; ama benzer metrik sistem üzerinde karşılaştırılabileceği bir istatistik modele ihtiyaç olduğunu ortaya koymuştur ve kendi adıyla anılan modeli geliştirerek bu alana büyük bir katkı yapmıştır. Lord ve Novick (1968), mental test skorlarının istatistiksel açıklamalarını yapmışlardır. Birnbaum (1968), MTK modellerinin tanımlarını yapmış ve matematiksel modellerini inşa etmiştir. Berberoğlu (1988), Rasch model ve KTK modelini öğrenci seçme sınavlarında kullanılan testlerdeki objektiflik açısından karşılaştırmış ve Rasch modelle yapılan tahminlerin objektifliği sağladığını, güvenilirlik ve geçerliliklerinin de KTK'ya göre daha yüksek olduğunu ortaya koymuştur. Gücüm (1990), kriter dayanaklı testler için madde seçiminde iki farklı yöntem olarak Rasch modeli ve Yargıcı Kararları Yaklaşımını karşılaştırmıştır ve iki yöntem arasında kriter dayanaklı testler için madde seçiminde önemli bir fark olmadığını belirtmiştir. Berberoğlu ve Çalikoğlu (1992), madde verilerini değerlendirmek için Rasch ve üç parametrelili (3PL) modellerin uygunluğunu karşılaştırmışlar ve elde edilen sonuçlara göre modellerden birinin diğerine önemli bir üstünlüğü olmadığını ortaya koymuşlardır. İki model için madde güçlük parametreleri karşılaştırılabilir ölçüde bulmuşlardır. Ayrıca 3PL modelde küçük örneklem ve düşük yetenek düzeyinde kestirilen madde parametrelerinin daha yanlı olduğunu gözlemlemişlerdir. Engelhard (1992), yazma yeteneğinin değerlendirilmesi amacıyla eyalet çapında yapılan bir sınav için çok yüzeyli Rasch model kullanmıştır ve bu tip sınavlarda karşılaşılan ölçme sorunlarını ortadan kaldırmak için çok yüzeyli Rasch modeli önermiştir. Gelbal (1994), KTK ve Rasch modelden elde edilen madde ve kişi parametrelerinin ortak noktalarını ortaya koymuştur. Analizler sonucunda madde ve kişi parametrelerinin arasında yüksek korelasyonlar olduğunu belirtmiştir. Macmillan (1995), KTK, Genellenebilirlik Kuramı (G-Kuramı) ve Çok Değişkenlik Kaynaklı Rasch Modelin (ÇDKRM) büyük ölçekli performans değerlendirme durumlarında puanlayıcıların katılımı, tutarlılığı ve uyumu açısından karşılaştırmış, sonuç olarak, üç yaklaşımın her durumda aynı sonuçlar üretmediğini ve hangi yaklaşımın veya yaklaşımların kullanılacağına amaca göre seçilmesinin önemli olduğu vurgulamıştır. Güler (2008), matematik başarısının ölçülmesinde KTK, G-Kuramı ve Rasch Model karşılaştırması yapmış ve ölçme araçlarının güvenilir sonuçlar verdiği, puanlayıcıların birbiriyle uyumlu puanlamalar

yaptığı belirlenmiştir. Ayrıca, güvenilirlik belirlenirken en az iki kuramdan yararlanılması gerektiğini tespit etmiştir. Sebok (2010), Northern British Columbia Üniversitesi'nin, Eğitimde Danışmanlık Yüksek Lisans Programına öğrenci seçme sürecini ÇDKRM ve G-Kuramı kullanarak değerlendirmiştir ve iki yöntemin de puanlayıcılar arası uyumu değerlendirmede uygun olduğunu ancak iki yöntemin de avantajları ve dezavantajları olduğunu söylemektedir. Çelen ve Aybek (2013), KTK ve MTK'ya dayalı yöntemlerle sınıf içi ders başarısını ölçmek için geliştirilen bir ölçekten iki kuramdan hesaplanan öğrenci başarıları karşılaştırmışlardır. Bu karşılaştırmaya göre, testteki tüm maddeler her iki kurama göre puanlandığında, elde edilen puanlar arasındaki ilişki pozitif, oldukça yüksek ve istatistiksel olarak anlamlı bulunmuştur. Jin ve Wang (2014), cevaplayıcıların kişisel ya da testten kaynaklı yaşadıkları performans düşüklüklerini göz önüne alarak cevaplayıcıları birden fazla örtük sınıfa bölmüşler ve her bir sınıfa performansı tanımlayan parametre ekleyerek yeni karma MTK modeller ortaya koymuşlardır. Bonifay ve Cai (2017), MTK modellerinin karmaşıklığına sadece parametre sayısı üzerinden değil işlevsel biçimleri açısından bir bakış açısı getirerek açıklayıcı faktör, bifaktör, Deterministic Input Noisy And Gate (DINA) ve Deterministic Input Noisy Or Gate DINO modelleri ile 3PL modeli uyum açısından karşılaştırmışlardır. Yapılan çalışmalar sonucunda model veri uyumu aranırken sadece uyum kriterlerine değil modellerin işlevsel formunun da incelenmesi gerektiği görüşünü ortaya koymuşlardır. Groothuis-Oudshoorn, Van den Heuvel ve Krabbe (2018), MTK tabanlı bir sağlık ölçüm modeli geliştirmişlerdir. Bu model aslında bir Rasch model örneği olup hastanın sağlık durumunu kendisine sunulmuş seçeneklerden seçmesi mantığı ile çalışır. Hastalardan alınan verilere göre yapılan analizlerde modelin çalışmaya uyum gösterdiğini ortaya koymuşlardır. Böylelikle MTK tabanlı modellerin tıp alanında da kullanımının uygun olduğunu göstermişlerdir. Ranyard, McNair, Nicolini ve Duxbury (2020), kişisel finansla ilgili bilgileri kullanabilme konulu MTK tabanlı bir finansal okuryazarlık testi oluşturmuşlardır. Ölçeğin kullanılabilirliğini test etmişler ve geçerlilik, güvenilirliğini kontrol etmişlerdir. Böylelikle finansal refahın temel göstergelerini tahmin eden MTK tabanlı bir model önermişlerdir. Bu çalışma, MTK modellerinin ekonomi alanında kullanılabilirliğine bir örnek olarak gösterilebilir.

2.6.1. Madde Tepki Kuramı Varsayımları

MTK'nın üç temel varsayımı vardır. Bunlardan ilki, tek boyutluluktur. Tek boyutluluk varsayımı, ölçeğin sadece tek bir örtük özelliği ölçtüğünün göstergesidir (Hambleton vd., 1991; Reckase, 2009; DeMars, 2010). Tek boyutluluk varsayımının sağlanması, seçilmiş modelde her bir cevaplayıcı için tek bir yetenek parametresi değerinin var olması demektir. Yeteneği haricinde cevaplayıcının yanıtını etkileyen tek faktör rastgele hatadan kaynaklanmaktadır. Tek boyutluluğun sağlanmadığı durumlarda MTK modelleri ile yapılan madde ve kişi parametre tahminleri yanlış olacaktır. Burada ölçek geliştiriciler için dikkat edilmesi gereken nokta, ölçekte bazen iki farklı örtük özellik yapısı mevcut olsa bile matematiksel olarak ölçeğin yine de tek boyutlu olabileceğidir. Bu, ölçekteki tüm maddelerin her iki örtük özelliği de aynı oranda ölçmesi ile ya da cevaplayıcıların farklılıkları sadece tek bir örtük özellik üzerinde toplanıyorsa mümkündür. Bu durumda, yine tek bir yetenek parametresi hesaplanacaktır (Reckase vd., 1988; DeMars, 2010). Tek boyutluluk için literatürde pek çok yöntem mevcuttur. Drasgow ve Lissak (1983), tek boyutluluk varsayımının gerçek veriler üzerinden sağlanmasının çok zor olduğunu belirtmişler ve bu durumun çözümü için ölçeğin, tek boyutluluk varsayımından sapmasının parametre tahminlerini etkileyip etkilemeyecek düzeyde olduğunu hesaplayan bir prosedür öne sürmüşlerdir. Hattie (1974), tek boyutlulukla ilgili 87 indeks geliştirmiş ve bu indekslerin karşılaştırmalarını yapmıştır. Karşılaştırılan indekslerin genellikle birbirleriyle ilişkili olduğunu ortaya koymuştur. Tate (2003), ise en çok kullanılan 9 yöntemi birbirleriyle karşılaştırmış ve karşılaştırılan yöntemler makul ölçüde iyi performans göstermiştir.

İkinci varsayım yerel bağımsızlıktır. Bu varsayım, ölçekteki her bir maddeye verilen yanıtların istatistiksel olarak birbirinden bağımsız olduğunu ifade eder (De Gruijter ve Van der Kamp, 2008; Reckase, 2009; DeMars, 2010). Bu, örtük özelliğin etkisi koşullandırıldıktan sonra iki madde arasında korelasyon olmaması gerektiği anlamına gelir. Maddeler yalnızca testin ölçtüğü örtük özellik aracılığıyla ilişkilendirilmelidir (Lord ve Novick, 1968). Örtük özelliğin katkısı kaldırıldıktan sonra maddeler arasında anlamlı korelasyonlar varsa, maddeler yerel bağımlıdır ya da hesaba katılmayan başka bir boyut vardır (Lee, 2004). Maddelerin bütün olarak ilişkili olması beklenen bir durumdur burada bahsedilen bağımsızlık, yetenek parametresinin koşullu dağılımında maddelerin birbirleriyle ilişkilerinin olmamasıdır. Eğer madde çiftleri yerel bağımsız değil ise o

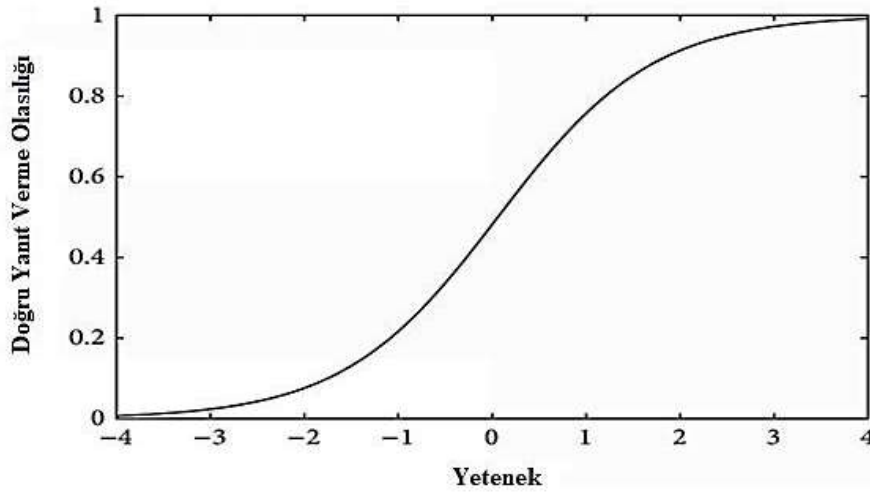
ölçekte başka bir boyut olduğu düşünülmelidir. Bağımlılığı diğer boyutun sağladığı düşünülür. Yen (1984), yerel bağımsızlığın ölçülmesi için Q_3 testini önermiştir. Yen (1993), istatistiksel anlamlılık yerine yerel bağımsızlığı ölçmede korelasyonların etkili olduğunu ortaya koymuştur. Madde çiftleri arasındaki 0.20'den yüksek korelasyonlar için yerel bağımsızlık varsayımının tehlikeye gireceğini ifade etmiştir. Chen ve Thissen (1997), yerel bağımsızlık için verilmiş 4 farklı yöntemi karşılaştırmışlardır. Madde çiftlerinden herhangi birinin arasında yüksek korelasyon gözlenirse bu maddelerin testten çıkartılması ya da maddeleri birleştirme yöntemleri kullanılarak yerel bağımsızlık varsayımının sağlanması mümkün olmaktadır (DeMars, 2010).

Varsayımların sonucusu ise normallik varsayımdır. Bu varsayım da ölçülen özelliğin normal dağılım göstermesi varsayımdır. KTK'nın bazı sınırlılıklarının olduğunun ortaya konulması bunun akabinde MTK modellerinin literatüre kazandırılmasıyla lojistik regresyon denklemleri ölçme teorisinde kullanılmaya başlanmıştır. Lojistik regresyon bilindiği üzere normallik varsayımının sağlanmadığı durumlarda kullanılmaktadır. Bu durumda MTK ile yapılan çalışmalarda normallik varsayımı önemini kaybetmiştir (Birnbaum, 1968).

2.6.2. Madde Karakteristik Eğrisi

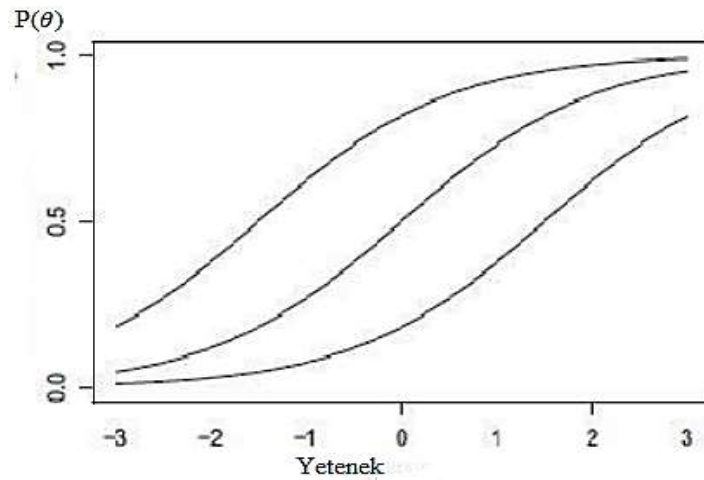
MTK modellerinin temelinde gözlenen skor X ile örtük özellik θ arasındaki ilişki yatmaktadır. Her bir yetenek düzeyinde, bu yetenek düzeyindeki bir katılımcının maddeyi doğru yanıtlamasına ilişkin belirli bir olasılık yer alır ($P(\theta)$). Düşük yetenek düzeyindeki katılımcılar için bu olasılık değerleri düşük, yüksek yetenek düzeyindeki katılımcılar için ise olasılık değerleri büyüktür. $P(\theta)$ değerlerinin yetenek düzeyinin bir fonksiyonu olarak çizilmesi halinde S şeklinde bir eğri elde edilir. Bu eğri, altta yatan özelliğin seviyesi arttığında, maddeye verilen doğru yanıtın nasıl arttığını gösterir. Düşük yetenek düzeylerinde bu olasılık sıfıra yakinken, yüksek yetenek düzeylerine doğru gidildikçe bu olasılık 1'e yaklaşmaktadır. Monoton artan bir fonksiyon olarak tanımlanan bu eğri madde üzerinde kişinin performansı ile madde arasındaki ilişkiyi tanımlar. MTK'da bu eğri Madde Karakteristik Eğrisidir (MKE). MKE, MTK'nın temel taşıdır. MTK üzerindeki pek çok yapı bu eğri üzerine kuruludur. Testteki her bir maddenin ayrı ayrı MKE'leri hesaplanır. Madde karakteristik eğrilerinin toplanması ölçeğin tümünü belirler, böylece madde karakteristik eğrisi doğru yanıtlarının olasılıklarının toplanması, Test Karakteristik

Eğrisini (TKE) oluşturur. TKE, altta yatan örtük bir değişkenin fonksiyonu gibi, yanıtlanan ölçek maddelerinin beklenen sayısını verir. TKE için, olasılıklar madde sayısına bölüldüğünde, TKE ortalama olasılık ya da beklenen doğru oranı verir (Baker, 2001; Crocker ve Algina, 2008; De Gruijter ve Van der Kamp, 2008; DeMars, 2010).



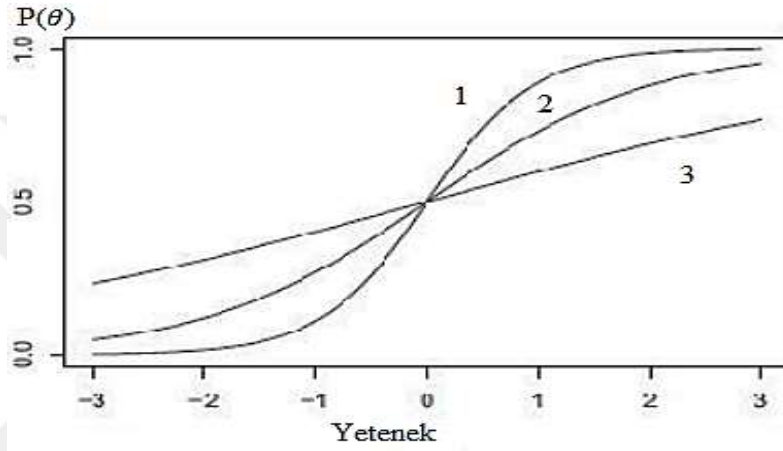
Şekil 2. Madde karakteristik eğrisi

Şekil 2’de bir örneği verilen MKE’yi tanımlamak için güçlük indeksi ve ayırt edicilik indeksi kullanılır. Güçlük, ilgili maddenin yetenek ölçeğinde nerede durduğunu gösterir (DeMars, 2010).



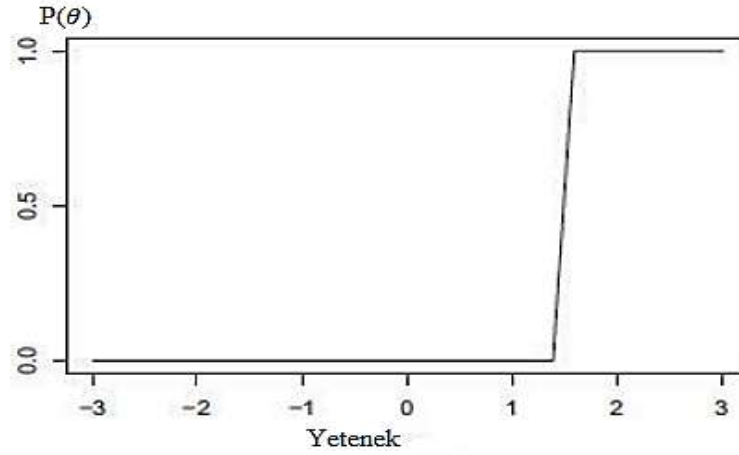
Şekil 3. Ayırt edicilikleri aynı 3 maddeye ait MKE

Şekil 3'te ayırt edicilikleri aynı güçlük düzeyleri farklı 3 maddenin MKE'si görülmektedir. İlk eğriye ait madde diğerlerine göre daha kolay bir maddedir. Bu, düşük yetenek seviyelerinde diğer eğrilere göre daha yüksek bir olasılıkla doğru yanıt verilmiş olmasından anlaşılmaktadır. En alttaki eğri diğer maddelere göre zor bir maddenin eğrisidir. MKE'ye bakıldığında düşük yetenek düzeylerinde olasılık düşüktür. Hatta yetenek seviyesinin en yüksek noktalarında dahi doğru yanıt verme olasılığı diğer eğrilere göre daha düşüktür (DeMars, 2010).



Şekil 4. Güçlükleri aynı 3 maddeye ait MKE

Ayırt edicilik ise MKE'nin orta noktasındaki dikliği ifade eder. Şekil 4'te güçlük düzeyleri aynı ancak ayırt edicilik indeksleri farklı 3 maddeye ait MKE'ler görülmektedir. Bir MKE ne kadar dikse ayırt ediciliği o kadar yüksektir. 1 numaralı eğri bu şekilde ifade edilebilir. Diğer eğrilerle kıyaslandığında 1 numaralı eğride yetenek seviyesi yükseldikçe ilgili maddeye doğru yanıt verme olasılığının daha hızlı bir şekilde yükseldiği görülmektedir. 3 numaralı eğri ise 3 maddenin içinde ayırt ediciliği en düşük maddeyi temsil etmektedir. Nispeten daha düz eğriler ise ayırt ediciliğin daha düşük olduğuna işaret eder (DeMars, 2010).



Şekil 5. Mükemmel ayırt ediciliğe sahip maddeye ait MKE

S şeklindeki MKE, yaygın kullanılmasına rağmen, MKE'lerin tamamı bu formda olmak zorunda değildir. MKE, basamak fonksiyonu şeklinde de olabilmektedir. Şekil 5'de gösterilen eğride minimum örtük özellik θ ile gösterilir ve buna sahip denekler doğru yanıt veremezler. Bu değere eşit ve üstünde değere sahip olanların maddeyi doğru yanıtlama olasılığı 1'dir. Bu tipte bir MKE'de ayırt ediciliğin orta noktası $\theta = 1.5$ noktasıdır. Bu tip bir madde bu yetenek seviyesinin altında ve üstündeki kişileri mükemmel bir şekilde ayırt edebilmektedir ancak, bu noktanın altındaki ya da üstündeki kişileri birbirlerinden ayırt edemez. Bu tip eğriler S tipteki eğrilere nazaran daha az kullanılır. Çünkü gerçek veri, çoğunlukla S tipine uyum sağlamaktadır (Hambleton vd., 1991; Baker, 2001; Crocker ve Algina, 2008).

Bu tiplerin tamamından farklı olarak MKE düz bir yapıya da sahip olabilir. Bu takdirde, MKE'si çizilmiş ilgili madde için bilenle bilmeyeni ayırt etmekten oldukça uzaktır yorumu yapılabilir. Düşük ve yüksek yetenek düzeylerindeki bireylerin ilgili maddeyi doğru yanıtlama olasılığı neredeyse aynıdır. Bu bilgiler ışığında MKE'lerin güçlük ve ayırt edicilik indeksleri ile ifade edilebilecekleri ve MKE'lerin şeklinin böylelikle belirlenebileceği anlaşılmaktadır (Baker, 2001).

2.6.3. Madde Tepki Kuramı Modelleri

2.6.3.1. 1 Parametrelili Lojistik Model

Bu model, bir madde parametresi barındırdığı için 1 Parametrelili Lojistik Model (1 PLM) olarak adlandırılır. Maddeler sadece güçlük parametresi bakımından farklılaşır çünkü modelde ayırt edicilik parametresi tüm maddeler için eşit kabul edilir. Ölçekteki her bir maddenin ayırt ediciliklerinden ortalama bir a indeksi hesaplanır ve ayırt edicilik bu değere sabitlenir. Ayırt edicilik parametresi her madde için eşit olduğundan hiçbir maddenin MKE'si birbirini kesmez. 1 PL modellerde şans ile maddeyi doğru yanıtlama olasılığı olarak ifade edilen c şans parametresinin 0 olduğu kabul edilir (Hambleton ve Swaminathan, 1985; Embretson ve Reise, 2000).

$$P(X_{ij} = 1/\theta_j, b_i) = \frac{\exp(\theta_j - b_i)}{1 + \exp(\theta_j - b_i)} \quad (17)$$

Eşitlik 17'de 1 PLM'nin matematiksel kuruluşu görülmektedir. b_i , i .maddenin güçlük indeksi, θ_j , j . kişinin yetenek düzeyi olmak üzere Eşitlik 17, bu bilgiler ışığında ilgili maddeye doğru yanıt verme olasılığını hesaplar (Birnbaum, 1968).

2.6.3.2. 2 Parametrelili Lojistik Model

Birnbaum (1968), 1 PL modeli geliştirerek 2 PLM'yi literatüre kazandırmıştır. 2 PLM, 1 PLM'den farklı olarak modele madde güçlük parametresinin yanında olarak madde ayırt edicilik indeksi (a) dahil olmuştur. 1 PLM'de olduğu gibi şans parametresi bu modelde de 0 olarak kabul edilir.

$$P_i(X_{ij} = 1/\theta_j, b_i, a_i) = \frac{\exp[a_i(\theta - b_i)]}{1 + \exp[a_i(\theta - b_i)]} \quad (18)$$

Eşitlik 18'de 2 PLM'nin matematiksel kuruluşu görülmektedir. a_i , i . maddenin ayırt edicilik parametresi, b_i , i .maddenin güçlük indeksi ve θ_j , j . kişinin yetenek düzeyini ifade

eder. Eşitlik 18 tüm bu bilinenler ile ilgili maddeye doğru yanıt verilmesi olasılığını hesaplamaktadır (Birnbbaum, 1968).

2.6.3.3. 3 Parametrelili Lojistik Model

Birnbbaum (1968), 2 PLM'yi maddeyi şansla yanıtlanma olasılığı parametresini (c) içine alacak şekilde genişletmiştir. Ancak, bu genişletme işlemi sırasında model lojistik fonksiyondan gelen birtakım kullanışlı özelliklerini kaybetmiştir. Yine de 3 PLM uygulamada sıkça kullanılmaktadır.

$$P_i(\theta) = c_i + (1 - c_i) \frac{\exp[a_i(\theta - b_i)]}{1 + \exp[a_i(\theta - b_i)]} \quad (19)$$

Eşitlik 19 ile 3 PLM'nin matematiksel kuruluşu görülmektedir. a_i , i . maddenin ayırt edicilik parametresi, b_i , i . maddenin güçlük indeksi ve θ_j , j . kişinin yetenek düzeyini ve c_i , şans parametresidir.

c_i parametresi, şans ile maddeyi doğru yanıtlanma olasılığıdır dolayısıyla, kişilerin yetenek seviyelerinin bir fonksiyonu olarak değişmez. Bundan dolayı en düşük ve en yüksek yetenek düzeyindeki katılımcıların maddeyi şansla doğru yanıtlanma olasılıkları aynıdır. Bu parametreye aynı zamanda alt asimptot da denilmektedir. c_i parametresi teorik olarak 0 ile 1 arasında değişen değerler alabilir. Ancak uygulamada 0.35'in üzerindeki c_i değerlerinin kabul edilebilir olmadığı düşünülür. Bu nedenle c_i 0 ile 0.35 arasında değerler alır. Şans parametresinin kullanılması ile güçlük parametresinin tanımında bir miktar değişim olacaktır. 1 PLM ve 2PLM'de güçlük parametresi, maddeyi doğru yanıtlanma olasılığının 0.5 olduğu yetenek düzeyidir. Ancak 3 PLM'de, MKE'nin en küçük değeri artık 0 değil c_i değeridir. Bu sebeple, güçlük parametresinin yetenek ölçeği üzerindeki konumunun belirlenmesi gerekmektedir.

$$\begin{aligned} P(\theta) &= c_i + (1 - c_i)(0,5) \\ &= \frac{1 + c_i}{2} \end{aligned} \quad (20)$$

Eşitlik 20 ile verilen olasılık değeri c_i ile 1 arasındaki orta noktanın değeridir. c_i burada maddenin doğru yanıtlanmasına ilişkin en küçük olasılık değeridir. Bu şekilde de 3 PLM için güçlük parametresinin konumu belirlenmektedir. Yani, 3 PLM’de güçlük parametresi c_i ile 1’in ortalaması değerinde konumlanmaktadır. Ayırt edicilik indeksi, 3 PLM’de, 1 PLM ve 2 PLM’ye benzer şekilde madde karakteristik eğrisinin $\theta = b$ noktasındaki eğimine orantılı olarak tanımlanır. 3 PLM’de MKE’nin $\theta = b$ noktasındaki eğimi $\alpha(1-c_i)/4$ biçiminde hesaplanır (Birnbaum, 1968).

2.6.3.4. 4 Parametrelili Lojistik Model

Barton ve Lord (1981), 3 PLM’ye d_i parametresini ekleyerek 4 PLM’yi geliştirmişlerdir. Bu modelde d_i parametresi ile yüksek yetenek düzeyindeki yanıtlayıcıların kolay bir maddeyi yanıtlamada hata yapabilme olasılıkları dikkate alınmaktadır. d_i parametresine literatürde üst asimptot ya da dikkatsizlik parametresi de denilmektedir.

$$P_i(\theta) = c_i + (d_i - c_i) \frac{\exp[a_i(\theta - b_i)]}{1 + \exp[a_i(\theta - b_i)]} \quad (21)$$

Eşitlik 21 ile 4 PLM’nin matematiksel kuruluşu verilmiştir. a_i , i . maddenin ayırt edicilik parametresi, b_i , i . maddenin güçlük indeksi ve θ_j , j . kişinin yetenek düzeyini, c_i , şans parametresini ve d_i , dikkatsizlik parametresidir. 4 PLM, her maddenin farklı üst asimptot değeri alabilmesine olanak sağlamaktadır. Ancak, modelin ilk ortaya çıkarıldığında tüm maddeler için ortak bir üst asimptot değeri belirlenmiştir bu durum d_i parametresinin yetenek kestirimi konusunda a_i , b_i ve c_i parametrelerine ek olarak bir katkı sağlamadığı sonucuna ulaşılmasına neden olmuştur (Erdemir, 2015). Bundan dolayı araştırmacılar tarafından 4 PLM’ye ihtiyaç olmadığı düşünülmüş ve 2000’li yıllara kadar modelle ilgili pek çalışma yapılmamıştır. Ayrıca, modelin bu yıllara kadar tercih edilmemesinin bir diğer nedeni de üst asimptot değerinin serbestçe tahmin edilmemesidir. Barton ve Lord 4 PLM ile ilgili yaptıkları çalışmalarda üst asimptot değerini 1, 0.99, 0.98 gibi değerler alarak tahmin yapmışlardır (Erdemir, 2015). Ancak, 2000’li yıllardan günümüze kadar 4 PLM ile ilgili Reise ve Waller (2003), Rupp (2003), Rulison ve Loken

(2009), Loken ve Rulison (2010), Magis (2013) gibi isimlerin çalışmaları literatürde mevcuttur (Barton ve Lord, 1981).

2.6.3.5. Aşamalı Tepki Modeli

Doğrudan gözlemlenemeyen örtük özelliklerden herhangi birini ölçmeyi amaçlayan ölçeklerdeki maddeler her zaman ikili olmayabilir. Bazı durumlarda ölçekler çoklu puanlanan maddeler içerirler. Çoklu puanlanan maddeler ikiden fazla yanıt kategorisine sahiptir. Bu maddeler cevaplayıcılarla ilgili daha fazla bilgi sağlamaktadır. Böyle puanlanan maddeler için MTK bünyesinde çoklu puanlanan MTK modelleri geliştirilmiştir. Bunlardan ilki Samejima (1969) tarafından geliştirilmiş olan Aşamalı Tepki Modelidir (ATM). Model, sıralı seçenekli yanıtları modellemek için geliştirilmiştir (Reise ve Yu, 1990). ATM, cevaplayıcının yetenek düzeyi ile belli bir seçeneği işaretlemesi arasında ilişki kurmaktadır (Embretson ve Reise, 2000). Modelde tüm madde seçeneklerinin eşit olması zorunluluğu yoktur. ATM'nin yapısında seçenekleri birbirlerinden ayıran sınır değerleri ve seçenek sayısının 1 eksiği kadar eşik değerleri vardır (Koch, 1983; Ostini ve Nering, 2006).

$$P_{ix} = \frac{\exp[a_i(\theta - \beta_{ij})]}{1 + \exp[a_i(\theta - \beta_{ij})]} \quad (22)$$

Eşitlik 22'de bir ATM'nin matematiksel kuruluşu görülmektedir. a_i , i . maddenin ayırt edicilik indeksi, m_i , eşik parametresi sayısı, β_{ij} , j . seçenek için eşik parametresi, k_i , seçenek sayısı ve θ , yetenek parametresidir. Ayrıca, $m_i = k_i - 1$ biçiminde hesaplanır (Embretson ve Reise, 2000; LaHuis vd., 2011).

ATM literatürde dolaylı MTK modeli olarak da bilinmektedir. Bunun sebebi hesaplanan olasılık değerinin doğrudan Eşitlik 22 ile elde edilememesidir. Her bir seçenek içi ayrı hesaplamalar Eşitlik 23'teki gibi yapılabilir (Embretson ve Reise, 2000; Köse, 2015).

$$P_{ix}^*(\theta) = P_{ix}(\theta) - P_{i(x+1)}(\theta) \quad (23)$$

Samejima'nın 1969'da ilk kez ATM'yi ortaya koymasından sonra literatürde ATM ile yapılmış pek çok çalışmaya rastlamak mümkündür. Koch (1983), farklı bir tutum ölçeğinde ATM kullanmış ve likert ölçekli ATM'lerin tutum ölçümü yapmak için başarıyla kullanılabilirliğini göstermiştir. De Ayala vd. (1989), ATM'leri ilk kez performansın ölçülmesine yönelik kullanmışlardır. Reise ve Yu (1990), ilk kez yapay veriler üzerinde ATM ile çalışmışlardır. Uttaro ve Lehman (1999), Yaşam Kalitesi ölçeğinde ATM kullanılabilirliğini göstermişlerdir. Madera (2003), öğrenci tutumlarının değerlendirilmesi üzerine geliştirilmiş bir ölçekte ATM ile çalışmıştır ve KTK ile ATM ile tahmin edilen parametreler arasında yüksek bir ilişki bulmuştur. Matteucci ve Stracqualursi (2006), öğrencilerin yazma testinden elde ettiği puanları ATM ile değerlendirmiştir. Sukirno ve Siengthai (2010), ATM'lerin kullanılabilirliğini meslek doyum ölçeği üzerinde desteklemiştir. Mielenz vd. (2010), özyeterlilik ölçeğinde ATM ile incelemiş fakat, ölçeğin uç yetenek düzeyindeki katılımcıların bilgi fonksiyonlarını ile ilgili hatalı tahminler elde edildiğini ortaya koymuştur.

2.6.3.6. Rasch Model

Rasch (1960), seçilmiş maddeler üzerinde, bireylerin ilgilen örtük özellik bakımından sergiledikleri davranışlar ile ilgili yaklaşık ölçümler elde etmek için kendi adını verdiği Rasch modeli geliştirmiştir. Rasch, ortaya çıkardığı bu modelle cevaplayıcının ölçek içerisindeki herhangi bir maddedeki herhangi bir seçeneği seçme olasılığının yetenek düzeyi ve madde güçlüğü farkının bir fonksiyonu olduğunu söyler. Rasch model, eğitim bilimleri, psikoloji, tıp, biyoloji gibi pek çok disiplinde kullanım alanı olan bir modeldir.

Rasch model matematiksel olarak 1 PLM ile aynı gibi görünse de ayrı olarak geliştirilmiştir ve 1 PLM gibi 2 PLM'nin özel bir durumu olmayan bir modeldir. Aslına bakılırsa Rasch model, odds veya log-odds terimleriyle tanımlanmıştır. Rasch modelde MKE'ler birbirini kesmez, paralel ilerler (Baker, 2001; De Gruijter ve Van der Kamp, 2008).

$$P(\theta) = P(x = 1/\theta, b_i) = \frac{e^{(\theta-b_i)}}{1 + e^{(\theta-b_i)}} \quad (24)$$

Eşitlik 24 ile Rasch modelin matematiksel kuruluşu verilmektedir. θ , kişinin yetenek düzeyi ve b_i , madde güçlük indeksidir. $P(\theta)$ ile θ yetenek düzeyindeki cevaplayıcının b_i güçlük düzeyindeki maddeye doğru yanıt verme olasılığı ifade edilmektedir. $(\theta - b_i)$ farklarını oluşturmak, j kişinin i maddesinin güçlüğüne karşı yeteneği ölçütünde nasıl bir savaş vereceğini göstermektedir. Her iki parametre de $(-\infty, +\infty)$ aralığında değerler almaktadır dolayısıyla $(\theta - b_i)$ farkı da yine aynı aralıkta değerler alacaktır. Eşitlik 18'de görüldüğü gibi aslında hesaplanan bir olasılık değeridir ve 0 ile 1 arasında değerler almaktadır. Bununla başa çıkmak için, $(\theta - b_i)$ farkı e tabanına taşınır $e^{(\theta - b_i)}$ olur.

Rasch model, parametre değişmezliğini sağlayan en basit modeldir, az sayıda parametre içerir, parametrelerin basit yorumu vardır ve tahmin edilmesi kolaydır, birçok disiplinde kullanılan ölçeklerin barındırdığı madde tiplerinin analizine uyum sağlamaktadır. Rasch model çalışma mantığı açısından diğer MTK modellerinden ayrılmaktadır. Diğer MTK modelleri çalışırken model veri uyumunun en yüksek olduğu MTK modeli seçilmesi esastır. Rasch yaklaşımında ise veriye en uygun Rasch model tipi aranmaktadır. Yani bu durum her veri tipine uygun bir Rasch ailesi üyesi olduğu anlamına gelmektedir (Rasch, 1966).

Ölçme teorisindeki en önemli sorunlardan biri nitelik bildiren gözlemlerin sayısal olarak ifade edilmesi sorunudur (Rasch, 1960). Rasch bu durumu elemine etmek için nitel gözlemleri nicel parametrelerle göstermede kullanılabilecek bir model olan Rasch modeli önermiştir. Rasch (1966), önermiş olduğu model kullanıldığında, ölçmenin yapıldığı gruplardan bağımsız olarak ölçme yapılabileceğini, tahmin edilen parametrelerin gerçeğe daha yakın olacağını ifade etmektedir. Böylelikle, doğrudan gözlem yapılamayan durumlarda da fiziksel ölçümün yapılabildiği durumlar kadar objektif, hassas ve gerçeğe yakın karşılaştırmalar yapılabilmektedir (Rasch, 1960).

2.6.3.7. Kısmi Kredi Modeli

Kısmi kredi modeli (KKM), Rasch modellerin özel bir durumudur. İncelenen veri seti çok sonuçlu maddelerden oluşuyorsa bu durumda model KKM adını alır (Yüksel, 2012). Bazı ölçeklerde madde yanıtlarını ikili kodlamak yerine cevabın doğruluk derecesi ya da uygunluk derecesi önem kazanmaktadır. Böyle madde tepkileri KKM ile modellenmektedir (Masters, 1982). KKM ile araştırmacılar, madde yanıtlarını sadece

dođru ya da yanlış olarak kodlamak yerine, dođru yanıt kavramını görece daha geniş bir alana yayarlar (Yüksel, 2012). KKM, çoktan seçmeli maddelerden esinlenerek türetilmiştir. KKM' de çoktan seçmeli maddelerde katılımcı cevabı yanlış vermiş olsa da verdiği cevap bazı şeyleri bildiğini gösteren bir cevapsa kısmi bir kredi verilmesi söz konusudur. Katılımcının yanlış cevabının altındaki kısmi dođruluk derecesi maddeden maddeye farklılık gösterebilir (Şencan, 2005). KKM, bünyesinde güçlü çeldiricileri barındıran çoktan seçmeli sorular için tasarlanmıştır. Bahsedilen kısmi krediler bu güçlü çeldiriciler arasında dağılır. Her maddenin kendi güçlü çeldiricileri ve bu çeldiricilerin kendi kredilendirilme ölçekleri vardır. Her madde kendi yapısı nasıl bir kredilendirme gerektiriyorsa o şekilde kredilendirilmelidir (Masters ve Wright, 1982). KKM, her maddenin kendi derecelendirme sistemine sahip olduđu maddeleri içeren ölçekler için kullanılabilir. İstatistiksel olarak, bir maddenin genel derecelendirmeden çıkarılarak kendi kısmi kredilendirilme ölçeğinin tanımlamasına izin verilmesi, bazı ek parametrelerin tahmin edilmesini gerekliliğini de beraberinde getirmektedir.

KKM, iki veya daha fazla sıralı kategorili yanıtların analizinde kullanılan tek boyutlu bir modeldir. KKM, bu anlamda ATM'den farklılık göstermektedir. Ayrıca, KKM, ATM'den farklı olarak ayırt edicilik indeksi tüm maddelerde eşit olarak kabul edilmektedir. Rasch ailesinin bir üyesi olduğundan KKM'de de kişi ve madde parametrelerinin ayrı ayrı tahmin edilir ve aynı metrik üzerinde değerlendirilirler.

KKM'nin literatürde birçok ölçüm problemi için gerçekleştirilen uygulamaları mevcuttur. Wright ve Masters (1982), yeni doğanlarla ilgili yapılan çalışmalarda ilk kez KKM kullanmışlardır. Masters ve Evans (1986), eleştirel düşünmenin ölçütleri çalışmasında KKM kullanmışlardır. Adams ve Rowe (1988), ikinci dil yeterliliğinin derecelendirilmesinde KKM kullanmışlardır. Julian ve Wright (1988), KKM'yi ilk kez simülasyonla kullanmışlardır. Koch ve Dodd (1989), bilgisayar uyarlamalı testlerde KKM kullanmışlardır. Doig, Mellor, Piper ve Masters (1994), sosyal eğitimde kavramsal anlama ölçüleri isimli çalışmasında KKM kullanılmıştır.

Bir madde ikiden fazla kategorili ise ilgili maddeden 1 puan almak artan yetenek düzeyi ile artmaz, belli bir eşik değerini aşan kişinin o maddeden daha yüksek bir puan alması olasılığı daha büyüktür. Bu durum, $0 < 1 < 2 < \dots < m_i$ sıralı kategori özelliğinden gelmektedir.

$$\frac{P_{ijx}}{P_{ijx-1} + P_{ijx}} = \frac{\exp(\theta_j - \delta_{ix})}{1 + \exp(\theta_j - \delta_{ix})} \quad x = 1, 2, \dots, m_i \quad (25)$$

Eşitlik 25'te KKM'nin matematiksel kuruluşu verilmiştir. P_{ijx} , i maddesinden j kişinin x puan alması olasılığı, P_{ijx-1} , i maddesinden j kişinin $x - 1$ puan alması olasılığı, θ_j , j kişinin yetenek düzeyi ve δ_{ix} , i maddesinden $x - 1$ yerine x puan alması olasılığını ifade eden madde parametresidir. Eşitlik 25 ile verilen model, j kişinin ilgili maddeden 1 puan almasını değil, 1 puan alması olasılığının 1 olmasını hesaplar. Ayrıca, model bir çift komşu kategori için koşullandırma yapmaktadır. Bu sayede $x - 1$ ve x kategorilerinin karşılaştırılmasını yapar. İşte bu karşılaştırma KKM'nin temel taşı oluşturur.

2.6.3.8. Çok Yüzeyle Rasch Model

Amerikalı istatistikçi Linacre (1989), Rasch model ailesine yeni bir üye kazandırmıştır. Bu üye diğerlerinden farklı olarak yetenek ve madde parametrelerinin yanında puanlayıcı katılığı ve puanlama anahtarı gibi bir takım farklı değişkenlik kaynaklarını da bünyesinde barındırır (Lynch ve McNamara, 1998). Tüm bunlar dikkate alınarak oluşturulan bu model Çok Yüzeyle Rasch Model (ÇYRM) adını almıştır. ÇYRM, açık uçlu sorular için uygun bir seçenektir. Bu modelde puanlayıcılar ile ilgili özellikler cevaplayıcının alacağı puanın değişmesinde etkin rol oynar (Mulqueen vd., 2000).

Açık uçlu maddeler eğitim bilimlerinde sıklıkla kullanılmaktadır. Böyle maddelerden oluşan ölçeklerde en önemli nokta puanlayıcılar tarafından yansız puanlama yapılmasıdır. Açık uçlu maddelerin cevaplayıcılara kendilerini ifade etmede özgür bir alan sağlaması bu tip maddeleri diğer madde türlerinden farklı kılmaktadır (Tan ve Erdoğan, 2004; Turgut ve Baykul, 2012). Açık uçlu maddeler bu yönüyle cevaplayıcılara özgün fikirler üretme, fikirleri değerlendirme, neden-sonuç ilişkileri kurma, karşılaştırma yapma gibi olanaklar sağlar ve böylelikle cevaplayıcıların bu üst düzey becerilerinin ölçülmesi mümkün olur (Tan ve Erdoğan, 2004). Açık uçlu soruların bir diğer önemli özelliği ise şans ile maddeyi doğru yanıtlama olasılığının bu tip maddelerde mümkün olmamasıdır. Bu durum ölçme hatasını azaltarak daha doğru tahminler yapılmasına olanak sağlar (Turgut ve Baykul, 2012). Açık uçlu maddeler kısmi puanlamaya uygundur. Cevaplayıcı maddenin tamamını doğru yanıtlamamış olsa dahi eğer verdiği yanıtın içinde doğru noktalar varsa buradan

belirli bir puan alabilmektedir (Bağcan Büyükturan ve Çıkrıkçı Demirtaşlı, 2012). Bir testin en sancılı süreci şüphesiz doğru maddeleri nihai teste seçebilmektir. Bu anlamda açık uçlu sorular içeren ölçeklerin hazırlanması nispeten daha kolay olmaktadır (Başol, 2013). Tüm bu olumlu özelliklerinin yanında açık uçlu soruların en büyük handikapı puanlayıcının değerlendirmede yansız olamamasıdır (Atılğan vd., 2009). Puanlayıcı, maddelerin zorluklarına ve cevaplayıcının özelliklerine bakmadan değerlendirme yapmalıdır (Lunz vd., 1990).

$$\log\left(\frac{P_{nik}}{P_{nik-1}}\right) = B_n - D_i - C_j - F_k \quad (26)$$

ÇYRM'in matematiksel kuruluşu Eşitlik 26 ile verilmiştir. P_{nik} , n kişinin i maddesinde gösterdiği performansın j puanlayıcısı tarafından k kategorisinde puanlanma olasılığı, P_{nik-1} , n kişinin i maddesinde gösterdiği performansın j puanlayıcısı tarafından $k - 1$ kategorisinde puanlanma olasılığı, B_n , n kişinin yetenek düzeyi, D_i , i maddesinin güçlük düzeyi, C_j , j puanlayıcısının katılık düzeyi ve F_k , $(k - 1)$. kategoriden k . kategoriye geçiş için güçlük düzeyidir. Eşitlik 26 ile modelde değişkenlik kaynaklarının cevaplayıcı, madde ve puanlayıcılar olduğu ve her bir değişkenlik kaynağının parametrelerle temsil edildiği görülmektedir. ÇYRM, birden fazla yüzeyden oluşmaktadır. Burada Rasch modelde olduğu gibi her bir yüzey diğer yüzeylerden bağımsız olmalı ve aynı metrik üzerinde değerlendirilebilmelidir (Schumacker, 1999). ÇYRM, kişilerin ve puanlayıcıların durumlarını bir arada gösteren bir kalibrasyon haritası sunar. Burada, gözlenen değerlerin lojistik dönüşümleri bağımlı değişken yüzeyler ise bağımsız değişken olarak kabul edilmektedir (Behizadeh ve Engelhard, 2014). Değişken haritası, gerçek aralıklı bir ölçektir (Kondo-Brown, 2002). Bu durum, tüm yüzeylerin ve bu yüzeylerdeki elemanların eşzamanlı olarak karşılaştırmalarının yapılmasını mümkün kılar (Güler, 2010).

2.7. Uyum Kriterleri

MTK modellerinin kullanılabilirlik şartlarından biri de model veri uyumudur. Verinin yapısı ile uygun olmayan bir modelle çalışılması, araştırmacının temel amacı olan parametre tahminlerinin objektiflik ve hassaslık özelliklerinin kaybedilmesi anlamını taşır.

MTK modellerinin hangisinin arařtırmacının elindeki veriye uygun modeller olup olmadığının belirlenmesi amacıyla çeřitli model uyumu teknikleri geliřtirilmiřtir.

2.7.1. Sapma Bilgi Kriteri

Spiegelhalter vd. (2002), Bayesci yaklařımla yapılan model seřimlerinde uyum kriterlerinden biri olan Sapma Bilgi Kriterini (Deviance Information Criteria) (DIC) literatüre kazandırmıřlardır. DIC'nin hesaplanmasındaki kolaylık ve uygulanabilirlięi bu kriterin kullanımını arttırmıřtır. Kriterin alıřma mantığı, modeldeki etkili parametre sayısını verilere uygun řekilde dengelemek üzerine kurulmuřtur. DIC, olabilirlik fonksiyonunun logaritması ve sapmanın sonsal daęılımına baęlı olarak geliřtirilmiř bir kriterdir. Model karřılařtırmaları yapılırken DIC deęerinin en dūřuk olduęu model tercih edilir. Bayesci yaklařımla yapılan model seřiminde bilgi verici olmayan nseller kullanıldığında da DIC doęru sonular vermektedir. rnek byklę fazla olduęunda DIC asimptotik bir yaklařımdır. DIC istatistiksel modellemede olduka sık kullanılan bir uyum kriteridir. DIC hesaplamalarında modelin etkili parametre sayısını (p_D) belirlemek olduka nemlidir. p_D 'yi belirlemek iin kullanılan iki hesaplama yntemi mevcuttur.

$$p_D = \bar{D} - D(\bar{\theta}) \quad (27)$$

Eřitlik 27'de Spiegelhalter vd., (2002) tarafından nerilen ilk yntem grlmektedir.

$$\bar{\theta} = \frac{1}{n} \sum_{t=1}^n \theta^{(t)} \quad (28)$$

$\bar{\theta}$, θ 'nın ortalamasıdır ve Eřitlik 28'de grldę gibi hesaplanır.

$$\bar{D} = \frac{1}{n} \sum_{t=1}^n -2\log(p(y/\theta)) \quad (29)$$

\bar{D} , Eřitlik 29'da grldę gibi hesaplanır. \bar{D} tahmini, $D(\theta)$ 'nin Markov Zinciri Monte Carlo (MCMC) simlasyon yntemi ile elde edilmiř deęerlerinden rnek ortalamasının alınmasıyla hesaplanır (Berg vd., 2004).

Gelman vd. (2004), etkili parametre sayısını elde etmek için ikinci hesaplama yöntemini ortaya koymuştur.

$$p_D = p_V = \frac{1}{2}V[D(\bar{\theta})] \quad (30)$$

Eşitlik 30 ile Gelman vd. (2004), tarafından ortaya konan ikinci yöntemin hesaplanması görülmektedir. Etkili parametre sayısını doğru hesaplamak oldukça önemlidir. Bunun sebebi etkili parametre sayısı arttıkça model-veri uyumunun artması ve buna bağlı olarak da sapmaların azalmasıdır. Etkili parametre sayısı p_D , θ 'nın simülasyon ile elde edilmiş değerlerinin örnek ortalamasından elde edilebilir (Berg vd., 2004).

$$DIC = p_D + \bar{D} \quad (31)$$

$$DIC = D(\bar{\theta}) + 2p_D \quad (32)$$

Buradan hareketle DIC değeri Eşitlik 31 ya da Eşitlik 32 yardımıyla hesaplanabilir. DIC'de gözlemlenmiş veriler sonsal dağılımı belirlemek ve model tahmini yapmakta kullanılmaktadır. DIC, aşırı uyum gösteren modelleri seçme eğilimindedir. Bu problemin giderilmesi için Ando (2007), ilk kez Bayesci öngörü kriterini (BPIC) önermiştir.

DIC, örnek büyüklüğü fazla olduğunda Akaike Bilgi Kriteri gibi asimptotik bir yaklaşım halini alır. DIC, Akaike Bilgi Kriterinin Bayesci yaklaşım ile hesaplanmış halidir. DIC, Akaike'den farklı olarak Bayesci bir yaklaşım olduğundan önsel bilgileri dikkate almaktadır (Spiegelhalter vd., 2002).

2.7.2. Akaike Bilgi Kriteri

İstatistiksel çalışmalarda en önemli noktalardan biri doğru modeli seçmektir. Yapılacak tahminlerin doğruluğu modeli doğru seçmekten geçmektedir. Bozdoğan (1987), model belirlemeyi veriyi anlamak olarak değerlendirmiştir. Akaike (1974), modelin doğru seçimi konusunda kendi adıyla anılan Akaike Bilgi Kriterini (Akaike Information Criteria) (AIC) ortaya koymuştur. AIC, tahmin hatasını minimum yapmak için gerçek modeli daha düşük boyutlu bir modele yaklaştırmak mantığı üzerine kurulmuştur. AIC, modellerin kalitesini karşılaştırmak için de kullanılabilir. Ancak, AIC, belli sayıda model arasında

seçtiği model için bu en iyisidir diyemez. Tüm modellerin kötü olduğu bir karşılaştırmada AIC yine bir model seçecektir. AIC'nin seçtiği bu model için kötünün iyisi denilebilir. Model seçilirken AIC değeri en düşük olan model en iyi model olarak değerlendirilir. Psikometri, ekonomi, tıp, mühendislik gibi pek çok disiplinde model seçimi yapılırken AIC değer kullanılmaktadır. AIC model oluşturulurken kaybolan bilgi miktarını belirleme mantığıyla çalışır yani kaybolan bilgi miktarını tahmin ederek model seçimine katkıda bulunur. Süreçte ne kadar az bilgi kaybolursa model o kadar iyi olacaktır.

$$AIC = 2k - 2 \ln(\hat{L}) \quad (33)$$

AIC, Eşitlik 33 yardımıyla hesaplanır. k , modelde tahmin edilmiş parametre sayısı, \hat{L} ise olabilirlik fonksiyonunun maksimum değeridir. AIC değeri minimum olan model en iyi model olarak belirlenir.

Sugiura (1978), küçük örneklem için ikinci bir AIC hesaplama yöntemi geliştirmiştir. Sugiura'ya göre; $n/k \leq 40$ ise AIC yerine AIC_c biçiminde ifade edilen bu kriter kullanılmalıdır.

$$AIC_c = -2 \ln(\hat{L}) + 2k + \frac{2k(k+1)}{(n-k-1)} \quad (34)$$

Eşitlik 34 ile AIC_c kriterinin hesaplama yöntemi verilmiştir. n , etkili örneklem sayısı, k , parametre sayısıdır ve L , olabilirlik fonksiyonudur. Örneklemin boyutu arttıkça AIC_c değeri AIC'ye yaklaşır (Burnham ve Anderson, 2002).

2.7.3. Bayesci Bilgi Kriteri

Schwarz (1978), model seçiminde ilk kez Bayesci Bilgi Kriteri (BIC) yöntemini ortaya koymuştur. BIC, bir model kümesi arasında en iyi modeli seçmek için kullanılan bir yöntemdir. Yöntem olabilirlik fonksiyonuna dayanmaktadır. Model uyumunu yakalamak için ek parametre yöntemine gidilebilir ancak bu durum aşırı uyum sonucunu doğurabilir. Bu sorunu gidermek için BIC ve AIC yöntemlerinde modele parametre sayılarını sınırlandıracak bir terim eklenmektedir.

$$BIC = -2 \log(L) + k \log(n) \quad (35)$$

Eşitlik 35 ile BIC değerinin hesaplanması verilmektedir. Eşitliğin sağ tarafındaki 2. terim AIC ile BIC değerlerini birbirinden ayırmaktadır. Bu terim örnek büyüklüğü ile ilişkili bir terimdir. BIC, isminden de anlaşıldığı gibi Bayesci model seçiminde kullanılan bir yöntemdir.

Bağımlı değişkendeki açıklanamayan bir değişim ve bağımsız değişkenlerin sayısı BIC değerini arttırır. Buradan, düşük BIC değerinin hem daha az açıklayıcı değişken hem de daha iyi uyum anlamına geldiği söylenebilir. BIC değeri önsel dağılımdan bağımsızdır yani, herhangi bir önsel dağılım ile kullanılabilir (Burnham ve Anderson, 2004). BIC değeri karmaşık modeldeki parametre sayısını ifade eder. BIC, model seçiminde tutarlıdır. En iyi model karşılaştırmanın yapıldığı model kümesinin içinde ise BIC o modeli bulmayı garanti eder (Vrieze, 2012).

Literatürde örtük özellik modellerinin seçiminde de bu uyum kriterlerinin kullanıldığı görülmektedir. Lubke ve Neale (2006), AIC ve BIC yöntemlerinin sürekli ve kategorik örtük özellik modellerinin seçimindeki etkinliğini ortaya koymuşlardır. Lubke ve Neale (2006), ortalamalar, gözlemlenen değişkenler ve oranların farklı olduğu durumlarda AIC ve BIC değerlerinin doğru modeli seçmede ne kadar etkin olduğunu göstermişlerdir. Bunun dışında Henson vd., (2007), Nylund vd., (2007), Yang ve Yang, (2007), gibi birçok araştırmacı da bu uyum kriterlerinin çalışma prensipleri, üstünlükleri ve eksiklikleri üzerine çalışmalar yaparak literatüre katkıda bulunmuşlardır (Vrieze, 2012).

2.7.4. LogLikelihood Testi

LogLikelihood (LL) testi, örneklemin yeterince büyük olması varsayımının doğruluğu altında belli bir sayıdaki model kümesinden en iyi modeli seçmek için kullanılan bir yöntemdir. LL, tek başına bir uyum indeksi olarak kullanılamaz. Bunun nedeni, bu değer örnekleme büyüklüğüne bağlı olmasıdır. LL değeri örnekleme genişliklerinden etkilenir. LL değeri farklı modellerden oluşan bir model kümesinde en iyi uyumu yakalamış modeli bulmada yardımcı olur yani modelleri karşılaştırmak için kullanılabilir. LL değerinin maksimum olduğu model karşılaştırmadan sıyrılarak en iyi model olarak seçilir.

$$LL = -2 * \ln \frac{1. \text{ modelin olabilirliđi}}{2. \text{ modelin olabilirliđi}} \quad (36)$$

Eşitlik 36 ile LL değeri hesaplanır. İki model arasındaki anlamlı farklılık, anlamlılık düzeyine veya iki model arasındaki farklı parametre sayısına göre test edilebilir. Eğer LL kritik değerden büyükse modellerdeki fark istatistiksel olarak anlamlıdır. LL değeri iki modeldeki tahmin edilecek parametre sayısının farkı kadar serbestlik derecesine sahip χ^2 dağılımına sahiptir.

2.8. Madde Yanlılığı

Madde yanlılığı, aynı yetenek düzeyinde olmalarına rağmen farklı alt gruplarda konumlanan katılımcıların ölçekteki herhangi bir maddeye doğru yanıt verme olasılıklarının farklı olması olarak tanımlanmaktadır. Yanlı bir ölçekte kişiler hakkında yapılan değerlendirmelerin hatalı olması kaçınılmazdır. Bunun sebebi, alt gruplardan herhangi birinin lehine ya da aleyhine çalışan bir boyut karışmış olmasıdır. Bu durumda eşitlik ilkesi bozulacaktır yani madde her katılımcı için aynı şekilde çalışmayacaktır. Bu durumda ölçmenin olmazsa olmazlarından tahminlerin objektiflik ve hassaslıkları bozulacaktır (Doğan ve Öğretmen, 2005).

Yanlılığın tanımı yapılırken verilen yanıtlardaki farklılıklardan bahsedilmektedir. Bu farklılıklar alt grupların yetenek düzeylerinden kaynaklanıyorsa bu istenen bir durumdur. Çünkü böyle olduğu takdirde, ölçek geçerliliği yüksek demektir ve yapılan tahminlerin hassas ve objektif olacağı anlamına gelir (Osterlind, 1983; Adams ve Rowe, 1988; Mellenberg, 1983; Hambleton vd., 1991).

Ölçekteki maddelerin yanlı olup olmadıkları ilgili maddenin Değişen Madde Fonksiyonu (DMF) içerip içermediği ile ilişkilidir. Burada en önemli nokta bir maddenin tek başına DMF içermesi o maddenin yanlı olması için gerek şart ancak yeter şart değildir. DMF analizi yapılırken odak nokta ölçekteki her bir maddenin aynı kitleden elde edilmiş farklı alt gruplara nasıl davrandığını belirlemektir.

DMF analizi yöntemleri ölçme teorisinin iki kuramına (KTK ve MTK) dayanan yöntemler ile mümkündür. Çoğu teknik bu iki kuramdan birinin uzantısı olarak geliştirilmiştir. Bu nedenle de bu teknikler kuramlardan gelen avantaj ve dezavantajları da bünyelerinde barındırır. Literatüre bakıldığında DMF analizlerinde MTK'nın uzantısı olan

yöntemler KTK'ya dayanan yöntemlerden daha güçlü olduğu öne sürülmektedir. MTK'nın parametre değişmezliği ilkesi sayesinde yani parametrelerin bir gruptan diğerine değişmeyen değerler alması özelliğinden dolayı DMF analizleri bu kurama dayalı yöntemlerle yapıldığında daha iyi sonuçlar vermektedir. Madde parametrelerinin gruptan gruba değişmemesi bu parametrelerden oluşturulan MKE'lerin de benzer olması anlamına gelir. DMF analizi yapılırken farklı gruplardan elde edilen MKE'ler karşılaştırılır (Devine ve Raju, 1982; Rodney ve Drasgow, 1990). Yetenekler eşit olduğu zaman iki grubun MKE'lerinde farklılık gözleniyorsa madde DMF içerir biçiminde yorumlanır. MKE'ler ile maddeler hakkında birçok bilgiye erişmek mümkündür. Madde yanlılığı da bu bilgilerden bir tanesidir.

DMF belirlenirken alt gruplardan ve maddelerin bu alt gruplara nasıl davrandıklarından bahsedilir. Bir madde katılımcılara belirli özelliklerinden dolayı avantaj ya da dezavantaj sağlayabilmektedir. Maddenin dezavantajlı davrandığı düşünülen gruba odak grup, maddenin avantajlı davrandığı düşünülen gruba referans grup denilmektedir (Akın Arıkan vd., 2016). Bu alt gruplardaki katılımcıların madde performansları karşılaştırılmakta ve bu karşılaştırma sonucunda hesaplanan katsayıların durumuna göre maddenin DMF içerip içermediği eğer içeriyorsa ne düzeyde ve hangi grup lehine ya da aleyhine içerdiği belirlenir. DMF içerdiği belirlenen maddeler için uzman görüşü alınarak maddenin yanlı olup olmadığına karar verilir

DMF belirlemek amacıyla KTK'ya ve MTK'ya dayalı çok sayıda istatistiksel yöntem geliştirilmiştir. Bu yöntemlerden bazıları: Mantel-Hanszel, Lojistik Regresyon, Lord's χ^2 İstatistiği, Raju'nun Alan ölçümleri ve Olabilirlik Oranı Testi olarak sıralanabilir (Benito ve Ara, 2000; Camilli ve Shepard, 1994; Hambleton vd., 1991; Oshima ve Morris, 2008).

2.8.1. Breslow-Day İstatistiği

Breslow-Day, belirli bir çerçevede DMF'yi saptamak için Breslow-Day istatistiğini hesaplar. Bu istatistik KTK'ya dayalı bir DMF yöntemi olduğundan toplam test skorlarını baz alarak hesaplama yapar. Bu yöntemde her bir alt gruba dahil olan katılımcıların tepkileri ikili kodlanır.

Tablo 4. Odak ve referans grup için çapraz tablo

Grup	Cevaplar		Toplam
	1	0	
Odak	a_k	b_k	n_{Rk}
Referans	c_k	d_k	n_{Fk}
Toplam	n_{1k}	n_{0k}	N_k

Her bir katılımcının yetenek parametresi toplam test puanı kullanılarak hesaplanır ve test puanlarının her bir tabakası k ile gösterilir. Tablo 4 ile toplam k tabakada N_k katılımcı olduğunda doğru ve yanlış cevaplar ile katılımcı sayılarını içeren 2×2 'lik bir çapraz tablo gösterilmektedir.

$$\psi_k = \frac{a_k * d_k}{b_k * c_k} \quad (37)$$

Eşitlik 37 ile k . tabakadaki odak ve referans grupların başarıları k . tabakada bulunan bu iki grubun doğru cevaplarının oranı olarak değerlendirilir.

Eğer her bir tabakadaki gözlem sayısı yeterince büyükse a_k 'nin dağılımı normal dağılıma yaklaşır. a_k 'nin asimptotik beklenen değeri A_k olmak üzere Breslow-Day istatistiği a_k ve A_k arasındaki sapmalar göz önüne alınarak geliştirilmiştir (Breslow ve Day, 1980).

$$BD = \frac{[\sum_{k=1}^K X_k (a_k - A_k)]^2}{\sum_{k=1}^K X_k^2 V(a_k) - \frac{[\sum_{k=1}^K X_k V(a_k)]^2}{\sum_{k=1}^K V(a_k)}} \quad (38)$$

Eşitlik 38 ile Breslow-Day istatistiğinin hesaplanması verilmiştir. X_k , tabaka değişkeninin k . seviyesinin değeri, A_k , a_k 'nin asimptotik beklenen değeri ve $V(a_k)$, a_k 'nin asimptotik varyansıdır. Breslow-Day istatistiği yaklaşık olarak $k - 1$ serbestlik dereceli χ^2 dağılımına sahiptir. Bu test her bir k tabakasındaki olabilirlik oranlarının eşit olduğu yokluk hipotezinin doğruluğu varsayımı altında hesaplanır.

DMF'nin varlığına karar vermek için geliştirilmiş bazı yöntemlerin aksine Breslow-Day yöntemi her bir tabaka için büyük örneklem gerektirir. Bu durum da yöntemin

kullanışsız olmasına sebep olur. Ayrıca, bu testte saptanan yanlılık geçerliliğin bir göstergesi olarak kullanılmaz (Breslow ve Day, 1980).

2.8.2. Lojistik Regresyon Yöntemi

Lojistik regresyon yöntemi hem iki kategorili hem de çok kategorili maddelerin DMF analizinde oldukça etkili bir yöntemdir (French ve Miller, 1996; Hambleton vd., 1991; Zumbo, 1999). Yöntemin yorumlanmasının kolay oluşu ve etki büyüklüğünün hesaplanması sayesinde DMF belirlemede önemli bir yöntem olarak literatüre kazandırılmıştır (Zumbo, 1999). Lojistik Regresyon yönteminde değişkenler modele belli bir sırayla alınır. Modele eklenen ilk değişken toplam puan değişkenidir. Ardından sırayla hangi açıdan DMF bakıldığına ilişkin değişken ve son olarak toplam puan ile grup etkileşimi değişkeni modele eklenir. Üç basamaktan modelde her bir basamak için bir χ^2 değeri hesaplanır (Gelin ve Zumbo, 2003). χ^2 değerlerinin yanısıra etki büyüklükleri de hesaplanmaktadır. Etki büyüklükleri hesaplanırken R^2 değerleri kullanılır ve bu değerlere bakılarak maddenin DMF durumu yorumlanır. Etki büyüklükleri son basamağa ait R^2 değerinden ilk basamağa ait R^2 değerinin çıkarılmasıyla elde edilir (Atalay Kabasakal ve Kelecioğlu, 2012).

Yöntem ilgili maddenin DMF içermediğini belirten yokluk hipotezinin doğruluğu altında test yapar. Hesaplamadan elde edilen χ^2 ve p değerlerine göre ilgili maddenin DMF içerip içermediği yorumlanabilir.

$$P(u_{ij} = 1/\theta_{ij}) = \frac{e^{(\beta_{0j} + \beta_{1j}\theta_{1j})}}{1 + e^{(\beta_{0j} + \beta_{1j}\theta_{1j})}}, \quad i = \overline{1, n_j}, j = 1, 2 \quad (39)$$

Eşitlik 39'da bir lojistik regresyon modeli görülmektedir. u , maddeye verilen yanıt, θ , yetenek parametresi, β_0 , güçlük parametresi, β_1 , ayırt edicilik parametresidir. Eşitlik 39'da verilmiş lojistik regresyon modeli farklı maddeler için farklı denklemler şeklinde oluşturulur ve her madde için DMF analizi bu şekilde yapılır. Eğer odak ve referans grup için lojistik regresyon eğrileri paralel ise DMF yoktur olarak yorumlanır (Zumbo, 1999).

2.8.3. Lord'un χ^2 İstatistiği

Lord'un χ^2 İstatistiği yönteminde bir maddenin DMF analizi yapılırken odak ve referans gruptan elde edilen madde parametreleri kıyaslanır. Yöntem, bu mantık üzerine kurulmuştur. DMF'nin varlığı ve ölçüsü bu parametrelerin özdeş değerlerde olup olmadıkları ile ilgilidir. Yöntem, iki grup arasında madde parametreleri açısından fark yoktur yokluk hipotezinin doğruluğu altında hesaplamalar yapılmaktadır. Rasch model tek bir madde parametresi barındırdığından hipotezler $H_0: b_R = b_F$ ve $H_1: b_R \neq b_F$ biçiminde kurulmaktadır.

$$z = \frac{b_R - b_F}{SE(b_R - b_F)} \quad (40)$$

Eşitlik 40 ile test istatistiğinin hesaplanması verilmektedir. $z^2 \sim \chi_1^2$ olup serbestlik derecesi tahmin edilen parametre sayısı olarak belirlenir.

$$\chi_i^2 = v_i' \Sigma^{-1} v_i \quad (41)$$

Eşitlik 41'de verilen Σ^{-1} , varyans kovaryans matrisi, v_i , parametreler arası fark vektörüdür.

$$v_i' = [b_{iR} - b_{iF}] \quad (42)$$

Eşitlik 42 ile parametreler arası fark vektörü gösterilmektedir. Bu hesaplamalarda fark vektörünün kullanılmasının sebebi, hesaplamaların bir elips düzleminde yapılmasıdır. Elipste x ekseninin genişliği y eksenindeki boyu ile aynı değildir. Dolayısıyla elips üzerindeki noktalar farklı değerlendirilmelidir. Lord'un χ^2 istatistiği yöntemini kullanabilmek için olabilirlik fonksiyonunun parametre uzayında normal dağılması ve olabilirlikten türetilen Σ 'nın tüm tahminlerinin mevcut olması gerekmektedir (Lord, 1980; McLaughlin ve Drasgow, 1987).

Yapılan hesaplamalar sonrasında elde edilen χ^2 değerleri ya da p değerleri değerlendirilerek ilgili madde ya da maddelerin DMF içerip içermediğine dair yorumda bulunmak mümkün olmaktadır. Ayrıca Lord'un χ^2 yöntemi ile etki büyüklükleri de

hesaplanmaktadır. DMF varlığı için yapılan arařtırmalarda bu etki büyüklükleri de kullanılabilir. DMF varlığı için yapılan arařtırmalarda bu etki büyüklükleri de kullanılabilir.

2.8.4. Raju'nun Alan Ölçümleri

DMF arařtırmalarında odak ve referans gruplarına ait MKE'ler arasında kalan işaretli ve işaretli alanların sıfırdan önemli ölçüde farklı olup olmadığını belirlemek önemlidir. Raju (1988), 1PLM, 2PLM ve 3PLM için iki MKE arasında kalan işaretli ve işaretli alanların hesaplanması için formüller ortaya koymuştur. Ayrıca, Raju, c parametrelerinin eşit olmaması durumunda iki MKE eğrisi arasında kalan alanın sonsuz olduğunu da göstermiştir. Raju'nun literatüre kazandırdığı formüller madde parametrelerinin fonksiyonlarıdır. Uygulama yapılırken sadece parametre tahminleri mevcut olduğundan eğrinin altında kalan alanlar örneklem değişimlerinden etkilenmektedir. Örneklem nedeniyle ne kadar bir değişimden söz edildiği hakkında kesin bir şey söylemek mümkün değildir. İki MKE'nin sadece örneklem büyüklüğü sebebiyle mi yoksa altta yatan başka sebeplerden mi farklılık gösterdiğinin de belirlenmesi mümkün olmamaktadır. Bu durum iki MKE arasında kalan alanın DMF belirlenmesinde kullanılması durumunda yöntemin dezavantajı olarak değerlendirilir.

$F_1 = F_1(\theta) = P_1$ ve $F_2 = F_2(\theta) = P_2$ olmak üzere;

$$P_1 = P_1(\theta) = \frac{\exp[(\theta - b_1)]}{1 + \exp[(\theta - b_1)]} \quad (43)$$

$$P_2 = P_2(\theta) = \frac{\exp[(\theta - b_2)]}{1 + \exp[(\theta - b_2)]} \quad (44)$$

Eşitlik 43 ve Eşitlik 44 yardımıyla DMF analizinde iki MKE arasında kalan işaretli ve işaretli alanlar elde edilmiştir.

$$SA_{kl} = \int_{-\infty}^{\infty} (\hat{F}_1 - \hat{F}_2) d\theta \quad (45)$$

$$UA_{kl} = \int_{-\infty}^{\infty} |\hat{F}_1 - \hat{F}_2| d\theta \quad (46)$$

Eşitlik 45 ve Eşitlik 46 ile iki MKE arasında kalan işaretli ve işaretsiz alanların hesaplanması verilmiştir. k , hangi MTK modelinin kullanıldığını belirtir. l ise alanların özelliklerini gösterir.

$$l = \begin{cases} 0, & a_i' \text{ler eşit ya da değil, işaretli} \\ 1, & a_i' \text{ler eşit, işaretsiz} \\ 2, & a_i' \text{ler eşit değil, işaretsiz} \end{cases}$$

2.8.5. Olabilirlik Oranı

Olabilirlik oranı yöntemi ile DMF belirleme de MTK'ya dayalı yöntemlerden birisidir. Olabilirlik oranı yönteminde, odak ve referans grupların madde parametreleri arasında anlamlı bir fark yoktur yokluk hipotezi test edilir. Yöntemin temeli, sınırlandırılmış ve genişletilmiş iki modelin kurulması ve bu modellerden elde edilen olabilirlik oranlarının karşılaştırılması üzerine kuruludur. Sınırlandırılmış modelde madde parametreleri her iki grup için de aynıdır. Genişletilmiş modelde ise ilgili maddenin parametreleri odak ve referans grup için farklıdır. Ölçekteki her bir madde için aynı adımlar tekrarlanarak DMF sorgulanır.

$$\text{Olabilirlik Oranı (OO)} = \frac{L_S}{L_G} \quad (47)$$

Ölçme aracına dahil olan tüm maddeler için olabilirlik oranları Eşitlik 47'de görüldüğü gibi hesaplandıktan sonra iki modelin elde edilmiş olabilirliklerinin logaritması alınarak bir G^2 değeri elde edilir (Atalay Kabasakal v.d., 2012).

$$G^2 = -2LL_S - (-2LL_G) \quad (48)$$

Eşitlik 48'den elde edilen G^2 istatistiği tüm parametrelerin odak ve referans grup arasında eşit olup olmadığının eş zamanlı olarak değerlendirilmesini sağlar. Hesaplanan G^2 değeri serbestlik derecesi kullanılan MTK modelinin parametre sayısı olan χ^2 dağılımının

kritik değeri ile karşılaştırılarak hipotez testi gerçekleştirilir. Anlamlı olan χ^2 değerleri DMF varlığını göstermektedir (Atalay Kabasakal v.d., 2012).

2.9. Etkili Örneklem Büyüklüğü

Örneklem büyüklüğü, kitleye ait birtakım karakteristikleri tahmin etmek için yapılan araştırmalarda tahminin hassas ve objektif olabilmesi için hesaplanmaktadır. Hambleton vd. (1991), MTK'da madde ve yetenek parametrelerinin tahmininin gerçeğe yakın ve küçük bir hata ile yapılabilmesi için büyük örneklemelere ihtiyaç olduğunu savunmuştur. Ölçme teorisinde yapılan çalışmalarda en önemli hedef bu parametrelerin tahminindeki başarı olduğundan örneklem büyüklüğü bu anlamda MTK ve KTK için kritik bir öneme sahiptir (Orlando, 2004; Şahin, 2012). Ayrıca, MTK'nın en önemli özelliklerinden biri olan parametre değişmezliğinin sağlanabilmesi için belli bir örneklem büyüklüğüne ulaşılmış olması gerekmektedir.

Doğan (2002), MTK'da ve KTK'da farklı örneklemelerden tahmin edilmiş madde parametrelerinin örneklem değişimlerinden etkilendiğini ortaya koymuştur. Köse (2010), MTK'da örneklem büyüklüğünün artmasıyla model veri uyumunun arttığı sonucuna varmıştır.

b parametresi, tahmin edilmesi nispeten daha kolay olan bir parametredir. Diğer parametrelere kıyasla daha küçük örneklemeler ile de hassas tahminler yapılabilir (DeMars, 2010). Ancak, örneklem büyüklüğü arttıkça tahminin standart hatasında azalma olduğu gözlemlenecektir. θ 'nın doğru tahmin edilebilmesi için büyük örneklem ve test uzunluğunun fazla olması gerekmektedir. Test uzunluğu arttıkça θ parametresinin tahmini gerçek değerlere yaklaşır. a parametresinin tahmini hangi modelle çalışıldığına göre farklılık gösterebilir. c parametresinin tahmininde ise daha büyük örneklemeler tahminin kalitesini arttıracaktır (DeMars, 2010).

$$N_{e\ddot{o}b} = \frac{N}{\sum_{t=-\infty}^{\infty} \rho_t} = \frac{N}{1 + 2 \sum_{t=1}^{\infty} \rho_t} \quad (49)$$

Eşitlik 49 ile etkili örneklem büyüklüğünün hesaplanması görülmektedir. N , Markov Zinciri Monte Carlo (MCMC) yöntemi ile hedef sonsal dağılımdan elde edilmiş örnek sayısı, ρ_t , MCMC ile seçilmiş örnekler arasındaki otokorelasyonlardır.

2.10. Test Uzunluęu

Bir testin ana amacı hakkında doğrudan gözlem yapılamayan bir örtük özellięi ölçmektir. İşte bu bahsedilen örtük özellikle ilgili testi alan katılımcılardan yeterince bilgi alabilmek için ve parametre tahminlerini en az hatayla yapabilmek için yeterli test uzunluęunun sağlanması oldukça önemlidir. İki ölçek karşılaştırıldığında nispeten daha uzun testlerin kısa testlere göre güvenilir olduklarını söylemek mümkündür. Bunun yanında, gereksiz uzun bir test ölçek güvenilirliğini arařtırmacıyı yanlış sürükleyecek düzeyde şişirebilmektedir. Lord (1980), parametre tahminlerinin objektiflik ve hassasiyetini örneklem büyüklüęünün yanında test uzunluęunun da etkilediğini ortaya koymuřtur. Swaminathan ve Gifford (1983), 1000 kiřinin girdięi çoktan seçmeli maddelerden oluřan bir sınavda madde sayısının 15'in altına düşmesi halinde parametre tahminlerinin hassaslık ve objektiflikten uzaklařtığını ortaya koymuřlardır. Hambleton ve Cook (1983), ise madde ve kiři parametrelerinin tahminlerinin kalitesi için en az 200 kiřinin girdięi bir sınavda 20'den az maddenin olmaması gerektiğini öne sürmüřlerdir.

3. BULGULAR VE İRDELEME

Bu çalışmada, analizler iki veri seti üzerinden yürütülmüştür. Bunlardan ilki; tarafımızca geliştirilen İstatistiğe Giriş Okuryazarlık testidir. Ölçeğin kullanılabilirliğinin belirlenmesinin ardından Karadeniz Teknik Üniversitesi ve Ondokuz Mayıs Üniversitesi bünyesinde İstatistik dersi alan öğrencilere uygulanmıştır. Diğer veri seti ise Karadeniz Teknik Üniversitesi tarafından geliştirilen ve uygulanan 2019 yılı Yabancı Uyruklu Öğrenci Sınavıdır. İki veri seti için belirli özellikler üzerinden DMF analizleri yapılmış ve kullanılan yöntemler karşılaştırılmıştır.

3.1. İstatistiğe Giriş Okuryazarlık Testi

Karadeniz Teknik Üniversitesi (KTÜ) ve Ondokuz Mayıs Üniversitesi (OMÜ) bünyesinde, İstatistiğe Giriş ya da İstatistik I düzeyinde ders almış olan bölümlere, İstatistiğe Giriş Okuryazarlık testi uygulanmıştır. Test, KTÜ İstatistik ve Bilgisayar Bilimleri Bölümü ve OMÜ İstatistik Bölümü, İstatistiğe Giriş ve İstatistik I derslerinin kazanımlarına uygun olarak hazırlanmıştır. Ölçek, 25 maddeden oluşmaktadır ve her bir maddede 4 seçenek mevcuttur. Öncelikle, KTÜ ve OMÜ’de İstatistik dersi alan bölümler belirlenmiştir. Belirlenen bölümlerde ilgili dersin kaçınıcı yarıyılıda verildiğine bakılarak ölçeği alacak öğrenci profili oluşturulmuştur. Uygulama sonunda, hazırlanmış ölçek KTÜ ve OMÜ’den 1455 öğrenciye ulaştırılmıştır. Bu öğrencilerden 1044 tanesi KTÜ, 411 tanesi ise OMÜ öğrencisidir. Ölçeğin uygulandığı öğrencilerin 730’u kız, 725’i ise erkektir.

Öğrencilerin cevaplarından elde edilen veri seti, 0-1 biçiminde ikili hale getirilmiştir. Burada, doğru yanıt verilmiş maddeler 1, yanlış yanıt verilmiş maddeler 0 ile kodlanmıştır. Verilmiş bu değerler ile her bir öğrencinin ham puanı hesaplanmıştır. Tüm analizler R yazılımı kullanılarak yapılmıştır.

Tablo 5. Toplam puana ilişkin betimleyici istatistikler

Ortalama	Std. Sapma	Maksimum	Minimum	Değişim Aralığı
11.0769	3.1745	22	2	20

Tablo 5'te 0-1 biçimine dönüştürülmüş veriden hesaplanan ham puanlardan elde edilen tanımlayıcı istatistikler verilmiştir.

MTK modellerini kullanabilmek için sağlanması gereken varsayımlar olduğundan öncelikle bu varsayımların sağlanıp sağlanmadığını kontrol etmek adına gerekli analizler yapılmıştır. Sağlatılması gereken ilk varsayım tek boyutluluk varsayımdır. Tek boyutluluk varsayımı için veri setine tek boyutluluk testi uygulanmıştır.

Tablo 6. Verilerin tek boyutluluk ölçüsü

Tek boyutluluk kriteri	Maddeler içi korelasyon değeri
0.9084	0.9225

Tablo 6'da tek boyutluluk testi sonucunda elde edilen sonuçlar görülmektedir. Tek boyutluluğun varlığından söz edebilmek için kriterin 1'e oldukça yakın çıkması beklenmektedir. Bu kriter 1'e ne kadar yakınsa tek boyutluluğun kanıtının o kadar güçlü olduğunu söylemek mümkündür (Drasgow ve Lissak, 1983). Tablo 6'dan ilgili kriterin 0,9084 olarak hesaplandığı görülmektedir. Buna bağlı olarak, ölçme aracının tek boyutlu olduğu ve tek bir örtük özelliği ölçtüğünü söylenir. Ayrıca, ölçme aracı kapsamındaki maddelerin aynı amaca hizmet edip etmediklerinin bir ölçüsü de maddeler içi korelasyon değeri ile yorumlanabilmektedir. Maddeler içi korelasyon değeri ne kadar yüksekse maddelerin o kadar istenen örtük özelliğe hizmet ettiği söylenebilmektedir. Dolayısıyla, Tablo 6'dan elde edilen 0,9225 değeri ile de maddelerin aynı amaca hizmet ettiklerini söylemek mümkün olmaktadır.

MTK için ikinci varsayım yerel bağımsızlık varsayımdır. Bu varsayım için maddeler arası ilişkinin tek bir örtük özellik ile açıklanması yapılabilir. Bu anlamda yerel bağımsızlık ile tek boyutluluk varsayımları birbirleriyle ilişkilidir. Literatüre bakıldığında Lord (1980), Hambleton ve Swaminathan (1985), Embretson ve Reise (2000), gibi araştırmacıların tek boyutluluk varsayımının sağlandığı durumlarda yerel bağımsızlık varsayımının da sağlandığını belirttiklerini görmek mümkündür. Bu durumda Tablo 6'dan görüldüğü üzere tek boyutluluk sağlandığı için yerel bağımsızlık varsayımının da sağlanmaktadır.

Normallik varsayımı, ölçme teorisinin ilk çalışmalarında Normal Ogive model ya da KTK modeli ile yapılan çalışmalarda önemli bir varsayım olmasına rağmen ilerleyen

yıllarda KTK'nın bazı sınırlılıklarından dolayı MTK'nın ortaya atılması ile lojistik regresyon modellerine geçiş yapıldığından bu varsayım önemini kaybetmiştir. Çoğu MTK modelinde normallik varsayımı aranmamaktadır.

Veri setinin hangi MTK modeline uygun olduğunu belirlemek oldukça önemlidir. Aksi halde, veriye uygun olmayan modellerle çalışmak yapılacak tahminlerin doğruluğunu, objektifliğini ve hassasiyetini olumsuz yönde etkileyecek, dolayısıyla araştırmacıları hataya sürükleyecektir. Bu durumdan kaçınmak amacıyla, uyum kriterleri yardımıyla MTK modellerinin karşılaştırılması yapılmalı, tüm modeller arasından en uygun modelin seçilmesi gerekmektedir.

Tablo 7. Model seçimi için uyum kriterleri

	Log.lik	DIC	AIC	BIC
Rasch-1PLM	-17098.65	34197.30	34297.30	345561.40
2PLM	-17883.22	35766.44	35818.44	359955.77
3PLM	-17099.11	34198.23	34300.23	345569.61

Tablo 7'den görüldüğü üzere İstatistik Okuryazarlık testi verisi için MTK modelleri açısından yapılan karşılaştırmada, uyum kriterlerinin değerlerine göre veriden en fazla desteği alan model Rasch modeldir.

Madde analizine başlamadan önce veri setinin aykırı değer analizi yapılmıştır. Tek değişkenli ve çok değişkenli aykırı değer olup olmadığı sırasıyla z puanlarına ve Mahalanobis uzaklıklarına bakılarak değerlendirilmiştir. Elde edilen sonuçlara göre, tek değişkenli 5 aykırı değer tespit edilmiştir ve tespit edilen aykırı değerler veri setinden atılmıştır. Çok değişkenli aykırı değere ise rastlanmamıştır. Ölçekten atılması gereken madde olup olmadığına ise Total Corrected-Item Correlation değerlerine bakılarak karar verilmiştir. Corrected Item-Total Correlation değeri 0.3'ün altında olan maddeler ölçekten çıkarılmalıdır (Cristobal vd., 2007). Elde edilen değerlerin hiçbiri 0.3'ün altında olmadığından ölçekten hiçbir madde çıkartılmamıştır.

Tablo 8. Verilen cevapların oranı ve logit değerler

Maddeler	0	1	logit
S1	0.3219	0.6781	0.7452
S2	0.5440	0.4560	-0.1765
S3	0.7531	0.2469	-1.1152
S4	0.7531	0.2469	-1.1152
S5	0.7531	0.2469	-1.1152
S6	0.3211	0.6788	0.7483
S7	0.8149	0.1850	-1.4827
S8	0.4814	0.5185	0.0743
S9	0.5804	0.4195	-0.3246
S10	0.5687	0.4312	-0.2768
S11	0.4944	0.5055	0.0220
S12	0.4085	0.5914	0.3700
S13	0.4085	0.5914	0.3700
S14	0.5557	0.4442	-0.2237
S15	0.5440	0.4559	-0.1765
S16	0.6182	0.3817	-0.4823
S17	0.6430	0.3569	-0.5886
S18	0.6182	0.3817	-0.4823
S19	0.6182	0.3817	-0.4823
S20	0.6182	0.3817	-0.4823
S21	0.3094	0.9605	0.8024
S22	0.2599	0.7400	1.0461
S23	0.3590	0.6409	0.5796
S24	0.2599	0.7400	1.0461
S25	0.3590	0.6409	0.5796

Tablo 8’de İstatistiğe Giriş Okuryazarlık ölçeğindeki maddelere verilmiş her bir cevabın yüzde oranı ve logit değerleri görülmektedir. Rasch analizinde ham puanlar kullanılmaz. İncelenen değişkeni ham puanlar ile yorumlamak araştırmacıyı yanıltıcı götürecektir. Rasch analizinde aynı ham puana sahip cevaplayıcılar eğer farklı güçlükteki maddelere doğru yanıt vermişlerse metrik üzerinde aynı konuma sahip olmazlar. Dolayısıyla logit değerler ile hesaplama yapmak araştırmacının daha hassas ölçümler yapmasına olanak sağlamaktadır. Logit dönüşümde madde puanlarının kişi puanlarıyla aynı yönde olması için madde puanlarının yönü tersine çevrilir. Bu durum, Rasch analiziyle madde ve kişi puanlarını aynı metrikte karşılaştırma olanağı sağlamaktadır (Elhan ve Atakurt, 2005).

Tablo 9. Her bir madde ve tüm test için güvenilirlik katsayıları

Maddeler	Cronbach α
S1	0.9262
S2	0.9258
S3	0.9224
S4	0.9224
S5	0.9224
S6	0.9224
S7	0.9244
S8	0.9215
S9	0.9213
S10	0.9242
S11	0.9212
S12	0.9202
S13	0.9201
S14	0.9215
S15	0.9208
S16	0.9213
S17	0.9230
S18	0.9213
S19	0.9213
S20	0.9213
S21	0.9246
S22	0.9252
S23	0.9254
S24	0.9252
S25	0.9254
Tüm test	0.9257

Tablo 9’da her bir madde için güvenilirlik katsayıları verilmektedir. Hesaplanan katsayılar incelendiğinde, ölçekteki tüm maddelerin güvenilirliklerinin çok iyi düzeyde olduğunu söylemek mümkündür.

Tablo 10. Madde istatistiklerinin tahmini

Maddeler	Madde güçlük indeksi (b)	Ayırt edicilik indeksi (a)	Düşük asimptot (c)	Yüksek asimptot (d)
S1	-1.178	1	0	1
S2	0.260	1	0	1
S3	1.734	1	0	1
S4	1.734	1	0	1
S5	1.734	1	0	1
S6	-1.183	1	0	1
S7	2.290	1	0	1
S8	-0.135	1	0	1
S9	0.495	1	0	1
S10	0.419	1	0	1
S11	-0.053	1	0	1
S12	-0.598	1	0	1
S13	-0.598	1	0	1
S14	0.335	1	0	1
S15	0.260	1	0	1
S16	0.744	1	0	1
S17	0.912	1	0	1
S18	0.744	1	0	1
S19	0.744	1	0	1
S20	0.744	1	0	1
S21	-1.265	1	0	1
S22	-1.634	1	0	1
S23	-0.923	1	0	1
S24	-1.634	1	0	1
S25	-0.923	1	0	1

Tablo 7 yardımıyla en iyi uyumun Rasch model ile yakalandığı görülmektedir. Rasch modelde madde parametresi açısından sadece güçlük parametresi mevcuttur. Tablo 10'da alt ve üst gruplar yöntemi ile her bir madde için güçlük indekslerinin tahminleri verilmiştir. Seçilen model Rasch model olduğundan ayırıcılık indeksi tüm maddeler için 1'e eşit kabul edilir. Ayrıca düşük asimptot 0 yüksek asimptot 1 değerini almaktadır.

3.1.1. İstatistiğe Giriş Okuryazarlık Testi İçin Cinsiyet Açısından Değişen Madde Fonksiyonu Analizi

Bir ölçme aracı içerisindeki maddelerin yanlı olup olmadıklarını belirlemek için o ölçme aracı içindeki maddelerin değişen madde fonksiyonu içerip içermediklerini incelemek gerekmektedir. Literatürde bu amaca hizmet eden birçok değişen madde fonksiyonu belirleme yöntemi vardır. Analizlerin tamamında anlamlılık düzeyi 0.05 olarak alınmıştır.

Tablo 11. Cinsiyet açısından BD İstatistiği yöntemi ile DMF analizi

Maddeler	Breslow-Day İstatistiği	Serbestlik derecesi	p değeri
S1	18.7397	17	0.3437
S2	11.4325	14	0.6518
S3	16.1332	16	0.4437
S4	23.8173	15	0.0683.
S5	27.4841	14	0.0166*
S6	6.9011	15	0.9603
S7	16.0004	14	0.3134
S8	10.9170	16	0.8146
S9	15.0965	14	0.3716
S10	10.6164	15	0.7793
S11	15.563	14	0.3407
S12	18.7140	14	0.1762
S13	19.4398	13	0.1101
S14	11.0026	14	0.6858
S15	9.4116	13	0.7412
S16	19.0949	14	0.1614
S17	10.4586	14	0.7279
S18	19.3334	14	0.1526
S19	23.4920	15	0.0742.
S20	19.1444	16	0.2612
S21	16.0298	14	0.3116
S22	18.9548	16	0.2710
S23	10.0447	15	0.8169
S24	15.0495	11	0.1802
S25	18.7008	15	0.2276

Tablo 11’de BD yöntemine göre DMF analizinin sonuçları görülmektedir. Odak grup kadınlar olmak üzere; bu analizde cinsiyet açısından DMF araştırılmış ve p değerlerinden yola çıkılarak 5. maddede DMF’ye rastlanmıştır. İlgili maddenin her iki cinsiyete adil davranmadığı kadınlar için dezavantajlı olduğu saptanmıştır.

Tablo 12. Cinsiyet açısından lojistik regresyon yöntemi ile DMF analizi

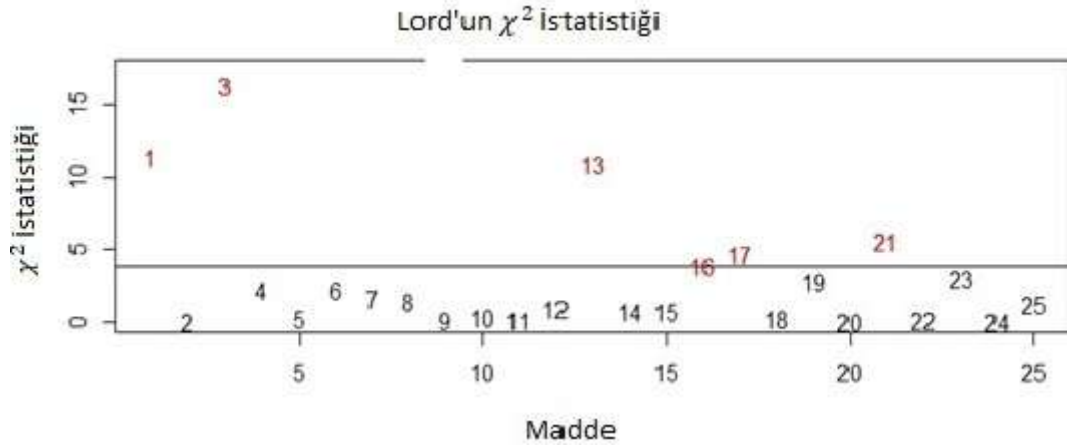
Maddeler	Lojistik regresyon DMF istatistiği	p değeri
S1	14.5553	0.0007***
S2	0.0197	0.9902
S3	20.6136	0.0000***
S4	2.8167	0.2445
S5	1.4206	0.4915
S6	2.5089	0.2852
S7	1.9291	0.3812
S8	5.8367	0.0540.
S9	2.5636	0.2775
S10	1.4710	0.4793
S11	3.0479	0.2178
S12	4.9766	0.0831.
S13	15.1848	0.0005***
S14	2.9019	0.2343
S15	0.9683	0.6162
S16	6.7451	0.0343*
S17	6.5220	0.0384*
S18	2.1156	0.3472
S19	3.0047	0.2226
S20	0.0266	0.9868
S21	7.1878	0.0275*
S22	0.5395	0.7636
S23	6.0915	0.0476*
S24	0.1353	0.9346
S25	1.6097	0.4472

Tablo 12’de LR yöntemi ile cinsiyet açısından DMF analizi yapılmıştır. Elde edilen sonuçlara göre, odak grup kadınlar olmak üzere; 1, 3, 13, 16, 17, 21 ve 23 numaralı maddelerin kadınlar için dezavantajlı olduğu sonucuna ulaşılmıştır.

Tablo 13. Cinsiyet açısından Lord'un χ^2 İstatistiği yöntemi ile DMF analizi

Maddeler	Lord'un χ^2 istatistiği	p değeri
S1	11.3776	0.0007 ***
S2	0.0387	0.8440
S3	16.3964	0.0001 ***
S4	2.1825	0.1396
S5	0.2378	0.6258
S6	2.1687	0.1408
S7	1.5565	0.2122
S8	1.4386	0.2304
S9	0.0634	0.8011
S10	0.2881	0.5914
S11	0.1359	0.7124
S12	0.9282	0.3353
S13	10.9324	0.0009 ***
S14	0.6600	0.4165
S15	0.7454	0.3879
S16	3.8650	0.0493 *
S17	4.7100	0.0300 *
S18	0.2539	0.6143
S19	2.7618	0.0965.
S20	0.0235	0.8782
S21	5.5940	0.0180 *
S22	0.0696	0.7920
S23	2.9901	0.0838.
S24	0.0436	0.8346
S25	1.1714	0.2791

Tablo 13'te Lord'un χ^2 İstatistiği yöntemi yardımıyla ölçme aracında DMF içeren madde ya da maddelerin olup olmadığına dair analiz sonuçları görülmektedir. Odak grup olarak kadınların seçildiği bu analizde elde edilen sonuçlara göre 1, 3, 13, 16, 17 ve 21 numaralı maddelerin kadınların dezavantajlı olmasına neden olan maddeler oldukları belirlenmiştir.



Şekil 6. Lord'un χ^2 istatistiği ile belirlenen cinsiyet için DMF'li maddeler

Şekil 6'da eşik değeri 3.8415'in üzerindeki maddeler yani kırmızı ile gösterilmiş maddeler yanlı maddelerdir.

Tablo 14. Cinsiyet açısından Raju'nun alan ölçümleri yöntemi ile DMF analizi

Maddeler	Raju İstatistiği	p değeri
S1	-3.3731	0.0007 ***
S2	0.1967	0.8440
S3	-4.0492	0.0001 ***
S4	-1.4773	0.1396
S5	-0.4876	0.6258
S6	1.4727	0.1408
S7	1.2476	0.2122
S8	-1.1994	0.2304
S9	-0.2519	0.8011
S10	0.5368	0.5914
S11	-0.3686	0.7124
S12	-0.9634	0.3353
S13	3.3064	0.0009 ***
S14	0.8124	0.4165
S15	-0.8633	0.3879
S16	1.9660	0.0493 *
S17	2.1702	0.0300 *
S18	-0.5039	0.6143
S19	1.6619	0.0965.
S20	0.1532	0.8782
S21	2.3652	0.0180 *
S22	0.2638	0.7920
S23	-1.7292	0.0838.
S24	0.2089	0.8346
S25	-1.0823	0.2791

Tablo 14’te cinsiyete göre Raju’nun Alan Ölçümleri yöntemi ile DMF analizinin sonuçları görülmektedir. Elde edilen sonuçlara göre, odak grup kadınlar olmak üzere; 1, 3, 13, 16, 17 ve 21 kodlu maddelerde DMF varlığı belirlenmiştir. Bu maddeler kadın ve erkeklere eşit davranmamakta kadınları erkeklere göre dezavantajlı duruma sokmaktadırlar.



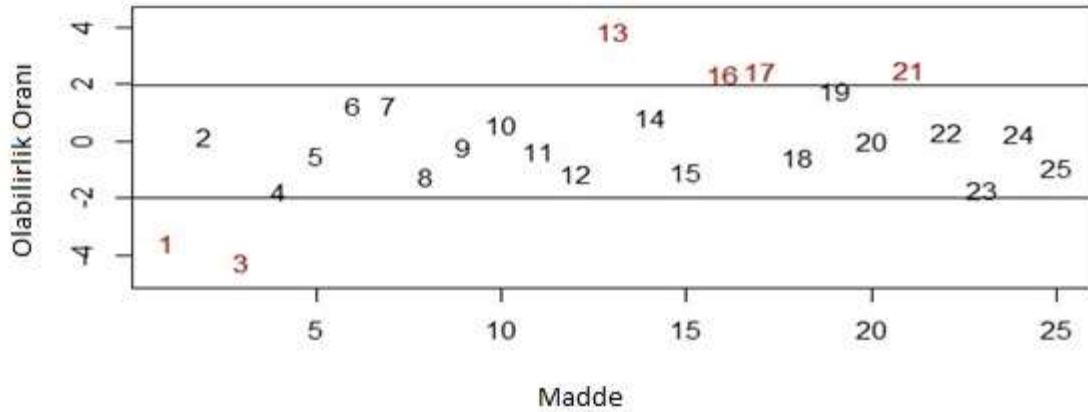
Şekil 7. Cinsiyet açısından Raju'nun alan ölçümleri yöntemi

Şekil 7’de \bar{F} 1.96 eşik değerini aşan maddelerde DMF varlığı saptanmıştır. İlgili maddeler kırmızı ile gösterilmiştir.

Tablo 15. Cinsiyet açısından olabilirlik oranı yöntemi ile DMF analizi

Maddeler	Olabilirlik oranı	p değeri
S1	-3.5423	0.0004 ***
S2	0.1920	0.8477
S3	-4.2615	0.0000 ***
S4	-1.7216	0.0851.
S5	-0.5124	0.6084
S6	1.2753	0.2022
S7	1.2797	0.2006
S8	-1.2338	0.2173
S9	-0.1981	0.8430
S10	0.5780	0.5633
S11	-0.3438	0.7310
S12	-1.1140	0.2653
S13	3.8510	0.0001 ***
S14	0.8365	0.4029
S15	-1.0478	0.2947
S16	2.3378	0.0194 *
S17	2.4690	0.0135 *
S18	-0.5449	0.5858
S19	1.7979	0.0722.
S20	0.0039	0.9969
S21	2.4986	0.0125 *
S22	0.3384	0.7351
S23	-1.7133	0.0867.
S24	0.2963	0.7670
S25	-0.9224	0.3563

Tablo 15'te cinsiyete göre olabilirlik oranı testi ile DMF analizinin sonuçları görülmektedir. Elde edilen sonuçlara göre, odak grup kadınlar olmak üzere; 1, 3, 13, 16,17 ve 21 kodlu maddelerde DMF varlığı belirlenmiştir. DMF içerdiği saptanan maddeler kadınları dezavantajlı konuma getirmektedir.



Şekil 8. Cinsiyet açısından olabilirlik oranı yöntemi

Şekil 8’de \bar{F} 1.96 eşik değerini aşan maddelerde DMF gözlenmiştir. DMF saptanan maddeler kırmızı ile gösterilmektedir.

3.1.2. İstatistiğe Giriş Okuryazarlık Testi İçin Üniversiteler Açısından Değişen Madde Fonksiyonu Analizi

İstatistiğe Giriş Okuryazarlık Testi için bir diğer DMF çalışması öğrencilerin bağlı oldukları üniversiteler açısından ölçekteki maddelerin DMF içerip içermediğine yönelik yapılmıştır. Bu analizde odak grup KTÜ olarak seçilmiştir.

Tablo 16. Üniversiteler açısından BD İstatistiği yöntemi ile DMF analizi

Maddeler	Breslow-Day İstatistiği	Serbestlik derecesi	p değeri
S1	25.3432	17	0.0873.
S2	15.9068	14	0.3191
S3	21.6788	15	0.1165
S4	11.4520	16	0.7808
S5	19.3444	14	0.1522
S6	28.0716	16	0.0310*
S7	6.4208	15	0.9718
S8	9.7868	16	0.8775
S9	11.5897	14	0.6392
S10	11.3600	15	0.7267
S11	10.9862	15	0.7536
S12	15.6203	15	0.4077
S13	10.4175	13	0.6595
S14	13.5033	14	0.4873
S15	11.8929	13	0.5364
S16	13.2522	14	0.5068
S17	5.7468	15	0.9837
S18	11.2782	15	0.7326
S19	11.9458	15	0.6831
S20	11.5423	16	0.7748
S21	8.9036	14	0.8372
S22	14.9783	16	0.5262
S23	9.1826	15	0.8678
S24	11.0999	11	0.4349
S25	29.5702	16	0.0204*

Tablo 16’da maddelerin üniversiteler açısından DMF içerip içermediği BD testi ile analiz edilmiştir. Bu analizde odak grup KTÜ öğrencileri olarak seçilmiş ve 6. ve 25. maddelerde DMF saptanmıştır. Bu maddeler OMÜ öğrencilerine karşı KTÜ öğrencilerini dezavantajlı durumu düşürmektedir.

Tablo 17. Üniversiteler açısından lojistik regresyon yöntemi ile DMF analizi

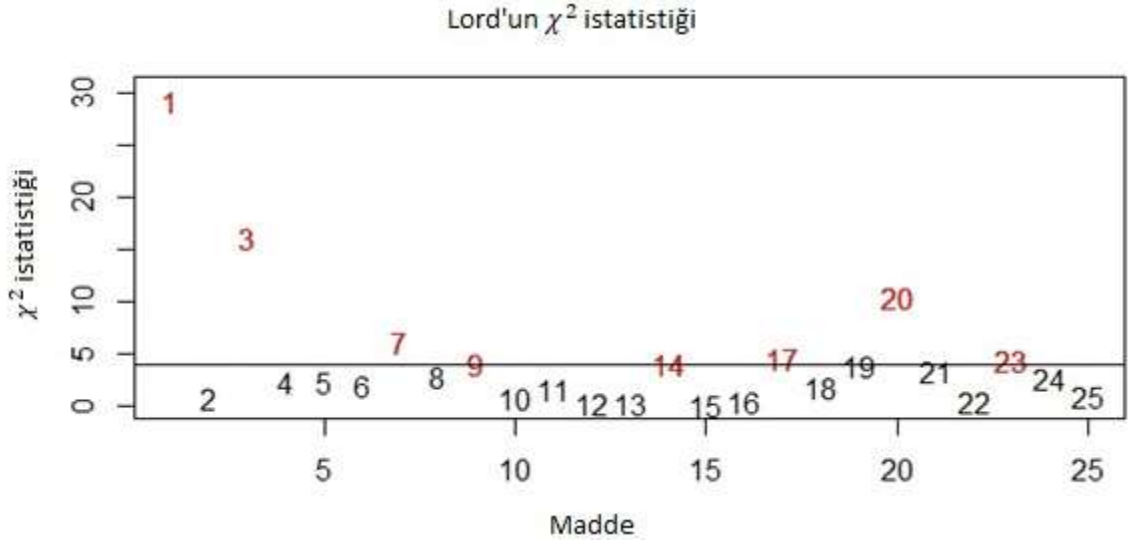
Maddeler	Lojistik regresyon DMF istatistiği	p değeri
S1	4.2213	0.0000***
S2	1.6904	0.4295
S3	18.4197	0.0001***
S4	1.9663	0.3741
S5	3.3721	0.1853
S6	1.6363	0.4412
S7	8.4681	0.0145*
S8	3.3906	0.1835
S9	7.0586	0.0293*
S10	0.4732	0.7893
S11	4.2266	0.1208
S12	1.3802	0.5015
S13	0.8281	0.6610
S14	7.6568	0.0217 *
S15	2.8338	0.2425
S16	1.0724	10.5850
S17	4.8120	0.0902.
S18	1.7856	0.4095
S19	6.2718	0.0435 *
S20	12.9245	0.0016 **
S21	4.2932	0.1169
S22	2.0861	0.3524
S23	5.9815	0.0503.
S24	2.5657	0.2773
S25	4.3619	0.1129

Tablo 17’de LR yöntemi ile üniversiteler açısından DMF analizi yapılmıştır. Elde edilen sonuçlara göre, 1, 3, 7, 9, 14, 19 ve 20 numaralı maddeler odak grup KTÜ öğrencileri için adil çalışmamakta, bu öğrencileri dezavantajlı durumu düşürmektedir.

Tablo 18. Üniversiteler açısından Lord'un χ^2 istatistiği yöntemi ile DMF analizi

Maddeler	Lord'un χ^2 istatistiği	p değeri
S1	29.2768	0.0000***
S2	0.6760	0.4110
S3	16.0900	0.0001***
S4	2.3046	0.1290
S5	2.2214	0.1361
S6	1.9811	0.1593
S7	6.0672	0.0138*
S8	2.7372	0.0980.
S9	3.9154	0.0478*
S10	0.7838	0.3760
S11	1.5389	0.2148
S12	0.1368	0.7115
S13	0.2485	0.6181
S14	3.9906	0.0458*
S15	0.0930	0.7604
S16	0.3399	0.5599
S17	4.4882	0.0341*
S18	1.7530	0.1855
S19	3.8349	0.0502.
S20	10.3638	0.0013**
S21	3.3625	0.0667.
S22	0.2997	0.5841
S23	4.3442	0.0371*
S24	2.6516	0.1034
S25	0.8531	0.3557

Tablo 18'de Lord'un χ^2 İstatistiği yöntemi yardımıyla ölçme aracında DMF içeren madde ya da maddelerin olup olmadığına dair analiz sonuçları görülmektedir. Burada üniversitelere göre DMF analizi yapılmıştır ve odak grup KTÜ olarak belirlenmiştir. Elde edilen sonuçlara göre 1, 3, 7, 9, 14, 17, 20 ve 23 numaralı maddelerin DMF içerdikleri sonucuna ulaşılmıştır. Bu maddeler odak grup KTÜ öğrencilerine adil davranmamaktadır.



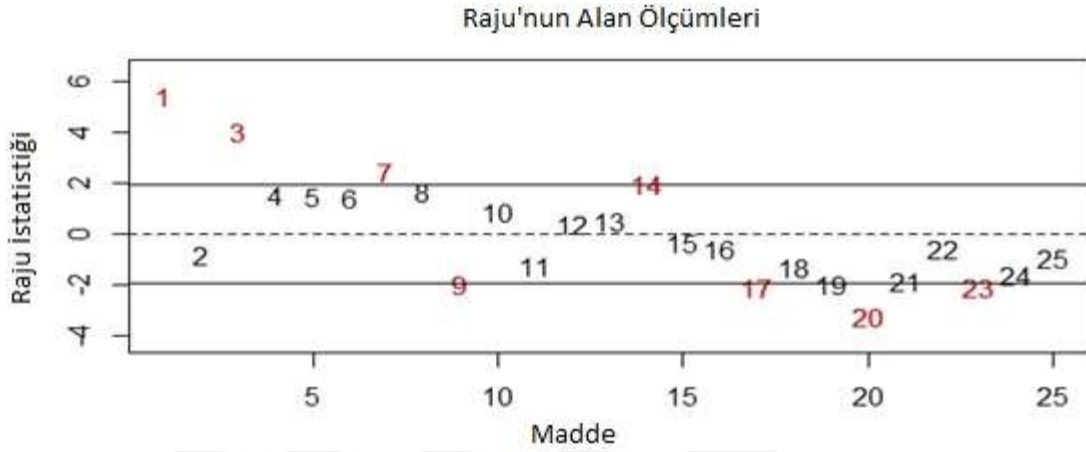
Şekil 9. Lord'un χ^2 istatistiği ile belirlenen üniversiteler için DMF'li maddeler

Şekil 9'da eşik değeri 3.8415'in üzerindeki maddeler yani kırmızı ile gösterilmiş maddeler DMF içeren maddelerdir.

Tablo 19. Üniversiteler açısından Raju'nun alan ölçümleri yöntemi ile DMF analizi

Maddeler	Raju İstatistiği	p değeri
S1	5.4108	0.0000 ***
S2	-0.8222	0.4110
S3	4.0112	0.0001 ***
S4	1.5181	0.1290
S5	1.4904	0.1361
S6	1.4075	0.1593
S7	2.4632	0.0138 *
S8	1.6545	0.0980.
S9	-1.9787	0.0478 *
S10	0.8853	0.3760
S11	-1.2405	0.2148
S12	0.3698	0.7115
S13	0.4985	0.6181
S14	1.9977	0.0458 *
S15	-0.3050	0.7604
S16	-0.5830	0.5599
S17	-2.1185	0.0341 *
S18	-1.3240	0.1855
S19	-1.9583	0.0502.
S20	-3.2193	0.0013 **
S21	-1.8337	0.0667.
S22	-0.5475	0.5841
S23	-2.0843	0.0371 *
S24	-1.6284	0.1034
S25	-0.9236	0.3557

Tablo 19’da üniversitelere göre Raju’nun Alan Ölçümleri ile elde edilen DMF analizinin sonuçları görülmektedir. Elde edilen sonuçlara göre, 1, 3, 7, 9, 14, 17, 20 ve 23 kodlu maddelerde DMF belirlenmiştir. İlgili maddeler KTÜ öğrencilerini Omü öğrencilerine kıyasla dezavantajlı konuma sokmaktadır.



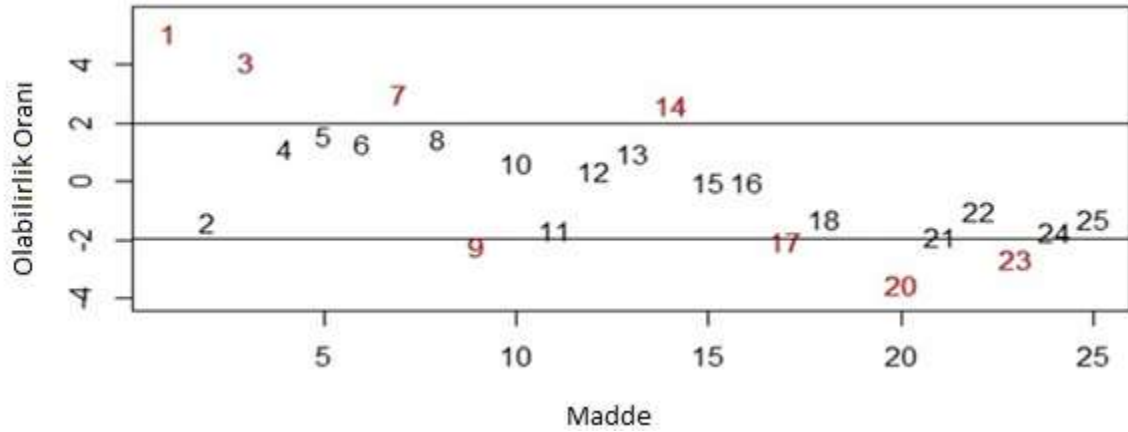
Şekil 10. Üniversiteler açısından Raju yöntemi ile belirlenen DMF içeren maddeler

Şekil 10’da $\bar{x} \pm 1.96$ eşik değerini aşan maddelerin DMF içerdikleri gösterilmiştir. DMF saptanan maddeler kırmızı ile gösterilmiştir.

Tablo 20. Üniversiteler açısından olabilirlik oranı yöntemi ile DMF analizi

Maddeler	Olabilirlik oranı	p değeri
S1	5.0882	0.0000***
S2	-1.4063	0.1596
S3	4.0648	0.0000***
S4	1.1477	0.2511
S5	1.5613	0.1185
S6	1.2801	0.2005
S7	2.9789	0.0029**
S8	1.4730	0.1408
S9	-2.2421	0.0250*
S10	0.6259	0.5314
S11	-1.6823	0.0925.
S12	0.3856	0.6998
S13	0.9684	0.3329
S14	2.6042	0.0092 **
S15	-0.0088	0.9930
S16	-0.0326	0.9740
S17	-2.0766	0.0378*
S18	-1.2665	0.2053
S19	-2.3979	0.1165
S20	-3.5321	0.0004***
S21	-1.8714	0.0613.
S22	-1.0053	0.3147
S23	-2.6479	0.0081**
S24	-1.7210	0.0853.
S25	-1.2946	0.1955

Tablo 20’de üniversiteler baz alınarak Olabilirlik Oranı yöntemiyle yapılan DMF analizlerinden elde edilen sonuçlara göre 1, 3, 7, 9, 14, 17, 20 ve 23 numaralı maddelerin KTÜ öğrencilerine adil davranmamaktadır.



Şekil 11. Üniversiteler için olabilirlik oranı yöntemiyle bulunan DMF’li maddeler

Şekil 11’de \bar{F} 1.96 eşik değerini aşan maddelerin DMF’li oldukları gösterilmiştir. DMF saptanan maddeler kırmızı ile gösterilmiştir.

3.2. Yabancı Uyruklu Öğrenci Sınavı

Yabancı Uyruklu Öğrenci Sınavı (YÖS), yabancı uyruklu katılımcıların Türkiye’de bir üniversitede lisans ya da ön lisans eğitimi alabilmeleri için girmeleri gereken bir sınavdır. YÖS’e çift vatandaşlığı bulunanlar, yabancı uyruklu olanlar ve Türk vatandaşı olup orta öğretim eğitiminin tamamını yurt dışında tamamlamış olanlar katılabilmektedir. YÖS ile adaylar, devlet ve vakıf üniversitelerinin özel sınavla alınan bölümleri hariç tüm bölümlerine yerleşebilirler. Ancak, Güzel Sanatlar, Müzik bölümleri ile Askeri ve Polis Akademilerine girişte istenen özel şartların sağlanması ya da özel yetenek sınavlarından başarılı olunması gerekmektedir. Adayların sınava girebilmeleri için Türk liselerine denk bir okuldan mezun ya da son sınıf öğrencisi olmaları öncelikli şarttır. Yabancı uyruklu öğrenci kontenjanı 11.04.2012 tarih ve 28261 sayılı Resmi Gazetede yayımlanan 6287 sayılı kanunun 14/f maddesine göre Devlet Üniversitelerinde %35, Vakıf Üniversitelerinde %50’yi geçmeyecek şekilde belirlenmiştir.

YÖS, 2011 yılına kadar Öğrenci Seçme ve Yerleştirme Merkezi (ÖSYM) tarafından her yılın haziran ayında gerçekleştirilirken 21.01.2010 yılında alınan kurul kararı doğrultusunda 2010–2011 akademik yılından itibaren YÖS, ÖSYM bünyesinden çıkartılarak her üniversiteye kendi sınavını geliştirip uygulaması için yetki verilmiştir. Ancak, bazı üniversiteler diğer üniversitelerin gerçekleştirdiği sınavlardan alınan puanlar ile öğrenci kabulü yapmaktadır.

Bu çalışmanın ikinci veri seti KTÜ tarafından geliştirilen ve uygulanan KTÜYÖS-2019 sınavıdır. KTÜYÖS-2019, adayların soyut düşünme güçlerini ölçmeyi amaçlayan maddeler ile matematik becerilerini ölçmeyi amaçlayan maddelerden oluşmaktadır. Sınavdan alınan puanların iki yıl geçerliliği vardır. KTÜYÖS-2019, Türkçe, Arapça, İngilizce ve Farsça dillerinde yapılmaktadır. Ölçek, 80 maddeden oluşmaktadır ve her bir madde 5 seçeneklidir. Yanıt sisteminde yanlış dođruyu götürmemektedir.

Tablo 21. 2018-2019 KTÜ yabancı uyruklu öğrenci sayıları (İlk 5 ülke)

Ülkeler	Erkek Öğrenci Sayısı	Kadın Öğrenci Sayısı	Toplam
Azerbaycan	154	35	189
İran	107	56	163
Suriye	97	38	135
Afganistan	93	17	110
Türkmenistan	46	16	62
Toplam	918	324	1242

Tablo 21’de Yükseköğretim Bilgi Yönetim Sisteminden alınan bilgilere göre 2018-2019 Eğitim-Öğretim yılında KTÜ’de eğitim hayatına başlamış ve halen devam etmekte olan öğrenci sayıları ve en fazla katılımı gösteren 5 ülke verilmiştir.

Tablo 22. 2019 KTÜYÖS'e giren ülkeler ve cinsiyete göre öğrenci sayıları

Ülkeler	Erkek Öğrenci Sayısı	Kadın Öğrenci Sayısı	Toplam
İran	966	1129	2054
Suriye	293	190	483
Azerbaycan	196	103	299
Türkiye	118	141	259
Bulgaristan	112	123	235
Almanya	87	102	189
Irak	69	54	123
Afganistan	58	42	100
Toplam	2187	2110	4297

Tablo 22’de KTÜYÖS-2019’a girmiş ilk sekiz ülke ve adayların cinsiyetlerine göre frekansları verilmektedir. Tablo 22’den elde edilen bilgilere göre; KTÜYÖS-2019’a toplam 2054 aday ile en çok İran’dan katılım sağlanmıştır. Türkiye ise sıralamada 4. sıradadır. Toplamda ise 2187 erkek aday, 2110 kadın aday ile KTUYOS-2019’a toplamda 4297 aday girmiştir.

Tablo 23. KTÜYÖS-2019 verisi için tek boyutluluk kriteri

Tek boyutluluk kriteri	Maddeler içi korelasyon değeri
0.9968	0.9325

Tablo 23 ile MTK varsayımlarından ilki olan tek boyutluluk varsayımı araştırılmıştır. Elde edilen tek boyutluluk kriterinin 1'e oldukça yakın olması beklenen bir durumdur (Drasgow ve Lissak, 1983). Tablo 23'te verilmiş tek boyutluluk kriteri değeri ölçeğe tek boyutlu demek için oldukça yeterlidir.

Bir diğer varsayım ise yerel bağımsızlık varsayımdır. Literatürde, tek boyutluluk varsayımı sağlandığında buna bağlı olarak yerel bağımsızlık varsayımının da sağlandığı bilgisi mevcuttur (Lord, 1980; Hambleton ve Swaminathan, 1985; Embretson ve Reise, 2000).

Son varsayım ise ölçülen örtük özelliğin normal dağılım göstermesi varsayımdır. MTK modelleri aslında birer lojistik regresyon modeli olduğundan normallik varsayımı önemini kaybetmiş bir varsayımdır.

Tablo 24. Model-Veri uyum kriterleri

	Log.lik	DIC	AIC	BIC
Rasch-1PLM	-185.534.6	364720.1	365040	366059
2PLM	-185529.8	371059.6	371222	371737
3PLM	-182360	371069.2	371229	371738

Tablo 24'te MTK modelleri için uyum kriterleri ile veriden en yüksek desteği alan modeli belirlemek için yapılan karşılaştırma görülmektedir. Uyum kriterlerinden elde edilen değerlere bakıldığında KTÜYÖS verisi için en uygun modelin Rasch model olduğu görülmektedir.

Madde analizinden önce, ölçekten atılması gereken madde olup olmadığının tespiti için Item Corrected-Total Correlation değerlerine bakılmıştır. Corrected Item-Total Correlation değeri 0.3'ün altında olan maddeler ölçekten çıkarılmalıdır (Cristobal vd., 2007). Elde edilen değerlerin hiçbiri 0.3'ün altında olmadığından ölçekten atılması gereken madde yoktur sonucuna varılmıştır. Buna ek olarak, tek değişkenli ve çok değişkenli aykırı değer analizi yapılmış ve veri setinden çıkarılması gereken aday olup olmadığı incelenmiştir. Tek değişkenli aykırı değer için z puanları, çok değişkenli aykırı değer için Mahalanobis uzaklıkları hesaplanmıştır. Elde edilen sonuçlara göre; 2 tane tek değişkenli 23 tane çok değişkenli aykırı değer tespit edilmiştir. Tespit edilen aykırı değerler veriden çıkarılarak madde analizi yapılmıştır.

Tablo 25. Verilen cevapların oranı ve logit değerler

Maddeler	0	1	logit	Maddeler	0	1	logit
S1	0.3002	0.6997	0.8463	S41	0.2957	0.7042	0.8674
S2	0.2583	0.7416	1.0547	S42	0.2415	0.7584	1.1441
S3	0.3141	0.6858	0.7806	S43	0.2033	0.7966	1.3651
S4	0.3288	0.6711	0.7134	S44	0.2601	0.7398	1.0450
S5	0.2008	0.7991	1.3810	S45	0.2185	0.7814	1.2742
S6	0.2448	0.7551	1.1264	S46	0.1270	0.8729	1.9271
S7	0.3728	0.6271	0.5201	S47	0.3974	0.6025	0.4159
S8	0.2946	0.7053	0.8730	S48	0.3053	0.6946	0.8220
S9	0.2974	0.7025	0.8596	S49	0.2373	0.7626	1.1671
S10	0.3309	0.6690	0.7039	S50	0.1659	0.8340	1.6147
S11	0.2697	0.7302	0.9960	S51	0.2059	0.7940	1.3494
S12	0.2597	0.7402	1.0474	S52	0.3218	0.6781	0.7452
S13	0.3292	0.6707	0.7113	S53	0.2995	0.7004	0.8496
S14	0.2706	0.7293	0.9913	S54	0.2655	0.7344	1.0174
S15	0.2555	0.7444	0.0693	S55	0.1696	0.8303	1.5880
S16	0.2485	0.7514	1.1063	S56	0.3116	0.6883	0.7925
S17	0.3646	0.6353	0.5551	S57	0.2473	0.7526	1.1126
S18	0.3092	0.6907	0.8034	S58	0.1994	0.8005	1.3897
S19	0.3432	0.6567	0.6487	S59	0.1978	0.8021	1.4000
S20	0.3055	0.6944	0.8209	S60	0.1505	0.8494	1.7301
S21	0.6087	0.3912	-0.4422	S61	0.2141	0.7858	1.3003
S22	0.5415	0.4584	-0.1665	S62	0.2173	0.7826	1.2811
S23	0.7295	0.2704	-0.9924	S63	0.3481	0.6518	0.6271
S24	0.3700	0.6299	0.5321	S64	0.1575	0.8424	1.6765
S25	0.6099	0.3900	-0.4471	S65	0.2331	0.7668	1.1904
S26	0.5994	0.4005	-0.4033	S66	0.4191	0.5808	0.3263
S27	0.4810	0.5189	0.0759	S67	0.3823	0.6176	0.4795
S28	0.7319	0.2680	-1.0043	S68	0.1461	0.8538	1.7651
S29	0.6432	0.3567	-0.5894	S69	0.3958	0.6041	0.4227
S30	0.6716	0.3283	-0.7155	S70	0.5306	0.4693	-0.1225
S31	0.6074	0.3925	-0.4363	S71	0.4963	0.5036	0.0144
S32	0.7365	0.2634	-1.0281	S72	0.5322	0.4677	-0.1291
S33	0.8128	0.1871	-1.4689	S73	0.2285	0.7714	1.2166
S34	0.7498	0.2501	-1.0976	S74	0.3658	0.6341	0.5501
S35	0.5496	0.4503	-0.1994	S75	0.5431	0.4568	-0.1731
S36	0.6781	0.3218	-0.7452	S76	0.2820	0.7179	0.9342
S37	0.5219	0.4780	-0.0880	S77	0.4693	0.5306	0.1225
S38	0.6318	0.3681	-0.5401	S78	0.4410	0.5589	0.2370
S39	0.6674	0.3325	-0.6966	S79	0.3970	0.6029	0.4178
S40	0.5403	0.4596	-0.1618	S80	0.6118	0.3881	-0.4549

Rasch analizinde hesaplamalar ham puanlar kullanılarak yapılmaz. Bunun nedeni ölçek içerisinde her maddenin aynı ağırlığa sahip olmadığı kabulüdür. Güçlük indeksleri hesaplandığında da bu kabulün doğruluğu ortaya çıkar. Bu sebeple, maddelere lojit dönüşüm uygulanarak maddelerin her birine güçlükleri oranında ağırlıklar verilir. Tablo 25 ile maddelerin lojit değerleri görülmektedir.

Tablo 26. Güvenilirlik katsayıları

Maddeler	Cronbach α	Maddeler	Cronbach α
S1	0.9306	S41	0.9312
S2	0.9301	S42	0.9308
S3	0.9301	S43	0.9306
S4	0.9300	S44	0.9308
S5	0.9300	S45	0.9306
S6	0.9304	S46	0.9308
S7	0.9303	S47	0.9308
S8	0.9303	S48	0.9307
S9	0.9303	S49	0.9312
S10	0.9303	S50	0.9309
S11	0.9302	S51	0.9311
S12	0.9298	S52	0.9308
S13	0.9302	S53	0.9311
S14	0.9298	S54	0.9307
S15	0.9298	S55	0.9307
S16	0.9300	S56	0.9310
S17	0.9302	S57	0.9315
S18	0.9302	S58	0.9309
S19	0.9303	S59	0.9306
S20	0.9301	S60	0.9307
S21	0.9305	S61	0.9308
S22	0.9307	S62	0.9308
S23	0.9311	S63	0.9320
S24	0.9298	S64	0.9307
S25	0.9305	S65	0.9305
S26	0.9307	S66	0.9320
S27	0.9302	S67	0.9318
S28	0.9311	S68	0.9306
S29	0.9306	S69	0.9315
S30	0.9310	S70	0.9318
S31	0.9308	S71	0.9316
S32	0.9311	S72	0.9315
S33	0.9314	S73	0.9310
S34	0.9316	S74	0.9311
S35	0.9301	S75	0.9317
S36	0.9310	S76	0.9311
S37	0.9298	S77	0.9319
S38	0.9306	S78	0.9304
S39	0.9311	S79	0.9311
S40	0.9301	S80	0.9316
		Tüm Test	0.9315

Ölçmenin doğruluğu açısından ölçeğin güvenilir olması çok önemlidir. Tablo 26’da KTÜYÖS ölçeği için her bir maddenin güvenilirlik katsayısı görülmektedir. Tüm ölçek için 0.9315 değeri bulunmuştur. Elde edilen değerlerden hareketle ölçeğin tamamının ve her bir maddenin güvenilir olduğunu söylemek mümkündür.

Tablo 27. Madde istatistiklerinin tahmini

Maddeler	Madde güçlük indeksi (b)	Ayırt edicilik indeksi (a)	Düşük asimptot (c)	Yüksek asimptot (d)
S1	-1.0634	1	0	1
S2	-1.3204	1	0	1
S3	-0.9819	1	0	1
S4	-0.8981	1	0	1
S5	-1.7169	1	0	1
S6	-1.4082	1	0	1
S7	-0.656	1	0	1
S8	-1.0965	1	0	1
S9	-1.0799	1	0	1
S10	-0.8863	1	0	1
S11	-1.2483	1	0	1
S12	-1.3115	1	0	1
S13	-0.8955	1	0	1
S14	-1.2425	1	0	1
S15	-1.3384	1	0	1
S16	-1.3837	1	0	1
S17	-0.7000	1	0	1
S18	-1.0102	1	0	1
S19	-0.8173	1	0	1
S20	-1.0319	1	0	1
S21	0.5600	1	0	1
S22	0.2122	1	0	1
S23	1.2436	1	0	1
S24	-0.6711	1	0	1
S25	0.5662	1	0	1
S26	0.5111	1	0	1
S27	-0.0949	1	0	1
S28	1.2580	1	0	1
S29	0.7445	1	0	1
S30	0.9018	1	0	1
S31	0.5527	1	0	1
S32	1.2872	1	0	1
S33	1.8185	1	0	1
S34	1.3720	1	0	1
S35	0.2539	1	0	1
S36	0.9387	1	0	1
S37	0.1127	1	0	1
S38	0.6828	1	0	1
S39	0.8782	1	0	1
S40	0.2063	1	0	1
S41	-1.0896	1	0	1
S42	-1.4298	1	0	1
S43	-1.6979	1	0	1
S44	-1.3085	1	0	1
S45	-1.588	1	0	1

Tablo 27'nin devamı

S46	-2.3628	1	0	1
S47	-0.5249	1	0	1
S48	-1.0333	1	0	1
S49	-1.4578	1	0	1
S50	-1.9959	1	0	1
S51	-1.6789	1	0	1
S52	-0.9378	1	0	1
S53	-1.0675	1	0	1
S54	-1.2746	1	0	1
S55	-1.9642	1	0	1
S56	-0.9967	1	0	1
S57	-1.3914	1	0	1
S58	-1.7275	1	0	1
S59	-1.7398	1	0	1
S60	-2.1321	1	0	1
S61	-1.6196	1	0	1
S62	-1.5963	1	0	1
S63	-0.7903	1	0	1
S64	-2.0693	1	0	1
S65	-1.4862	1	0	1
S66	-0.4115	1	0	1
S67	-0.6050	1	0	1
S68	-2.1735	1	0	1
S69	-0.5334	1	0	1
S70	0.1566	1	0	1
S71	-0.0169	1	0	1
S72	0.1649	1	0	1
S73	-1.5181	1	0	1
S74	-0.6936	1	0	1
S75	0.2205	1	0	1
S76	-1.1722	1	0	1
S77	-0.1538	1	0	1
S78	-0.2989	1	0	1
S79	-0.5273	1	0	1
S80	0.5759	1	0	1

Tablo 27'de alt ve üst gruplar yöntemi ile hesaplanmış madde istatistikleri verilmiştir. Tablo 24'te seçilen modelin Rasch model olduğu bilgisinden yola çıkarak güçlük indeksi tahminleri yapılmıştır. Diğer parametrelerin değerleri Rasch modelin özelliklerinden gelmektedir.

KTÜYÖS, farklı milletlerden kadın ve erkek öğrencilerin KTÜ'de öğrenim görme şansı yakalayabilmeleri için gerçekleştirilen bir sınavdır. Öğrencilerin farklı milletlerden olması, cinsiyet ayrımları ve yaşadıkları coğrafi konumlar akla bu ölçek içerisinde uyruğa, cinsiyete ya da bölgelere göre yanlı maddelerin var olup olmadığı sorusunu getirmektedir.

3.2.1. KTÜYÖS-2019 Cinsiyet Açısından Değişen Madde Fonksiyonu Analizi

Cinsiyet için yapılan DMF araştırmasında çalışmalar, en geniş katılım gösteren 5 ülke üzerinden yürütülmüştür. Bu ülkeler; Türkiye, İran, Azerbaycan, Almanya ve Hollanda olarak belirlenmiştir. Bu analizde odak grup olarak kadınlar seçilmiştir.

3.2.1.1. Türkiye-İran Karşılaştırması

İlk karşılaştırma Türkiye ile İran arasında yapılmıştır. Bu karşılaştırmada cinsiyet açısından DMF içeren madde ya da maddelerin varlığı araştırılmıştır. Yapılan karşılaştırmalarda odak grup olarak kadınlar seçilmiştir. DMF içerdikleri saptanan maddeler kadınları erkeklere karşı dezavantajlı duruma düşüren maddelerdir.

Tablo 28. Türkiye-İran cinsiyete göre BD istatistiği sonuçları

Maddeler	Breslow-Day İstatistiği	Serbestlik derecesi	p değeri
S9	78.9439	54	0.0151
S11	76.9805	51	0.0108
S45	70.0310	50	0.0322
S57	82.6148	55	0.0094
S59	69.7797	50	0.0337
S64	72.5743	49	0.0160
S69	73.7160	55	0.0468
S72	81.6639	55	0.0113

Tablo 28’de Türkiye ve İran için odak grup kadın olduğunda BD İstatistiği yöntemi ile DMF analizinin sonuçları görülmektedir. Elde edilen sonuçlara göre, 9, 11, 45, 57, 59, 64, 69 ve 72 numaralı maddeler odak grup kadınlar için adil çalışmamaktadır.

Tablo 29. Türkiye-İran cinsiyete göre lojistik regresyon yöntemi sonuçları

Maddeler	Lojistik Regresyon DMF İstatistiği	p değeri
S9	15.1392	0.0005
S21	8.8916	0.0117
S23	11.4457	0.0033
S24	6.9057	0.0317
S47	6.0016	0.0497
S48	10.8882	0.0043
S62	11.8923	0.0026
S65	12.0133	0.0025
S70	8.5338	0.0140
S76	8.8364	0.0121

Tablo 29’da Türkiye ve İran karşılaştırmasında odak grup kadın olduğunda Lojistik Regresyon yöntemi ile DMF analizinin sonuçları görülmektedir. Elde edilen sonuçlara göre, 9, 21, 23, 24, 47, 48, 62, 65, 70 ve 76 numaralı maddeler kadınlara adil davranmamaktadır.

Tablo 30. Türkiye-İran cinsiyete göre Lord’un χ^2 istatistiği yöntemi

Maddeler	Lord’un χ^2 İstatistiği	p değeri
S9	4.2164	0.0400
S23	6.6375	0.0100
S47	4.5811	0.0323
S48	6.2136	0.0127
S61	4.6109	0.0318
S62	9.6910	0.0019
S65	9.9895	0.0016
S70	5.1021	0.0239
S76	4.5562	0.0328
S80	4.2856	0.0384

Tablo 30’da odak grup kadınlar kabul edilerek Lord’un χ^2 İstatistiği yöntemi ile yapılan DMF analizinin sonuçları görülmektedir. Elde edilen sonuçlara göre; 9, 23, 47, 48, 61, 62, 65, 70, 76 ve 80 numaralı maddelerde DMF’ye rastlanmıştır. Bu maddeler kadınları erkekler karşısında dezavantajlı duruma sokmaktadır.

Tablo 31. Türkiye-İran cinsiyete göre Raju'nun alan ölçümleri yöntemi

Maddeler	Raju'nun Alan Ölçümleri	p değeri
S9	-2.0534	0.0400
S23	-2.5763	0.0100
S47	2.1403	0.0323
S48	2.4927	0.0127
S61	2.1473	0.0318
S62	3.1130	0.0019
S65	3.1606	0.0016
S70	2.2588	0.0239
S76	-2.1345	0.0328
S80	-2.0702	0.0384

Tablo 31'de cinsiyete göre DMF analizi yapılmış ve Raju'nun Alan Ölçümleri yöntemi kullanılmıştır. Yapılan analizde odak grup olarak kadınlar belirlenmiş ve Tablo 31'de verilmiş maddelerin DMF içerdikleri saptanmıştır. İlgili maddeler odak grup kadınları dezavantajlı durumu düşürmektedir.

Tablo 32. Türkiye-İran cinsiyete göre olabilirlik oranı yöntemi

Maddeler	Olabilirlik Oranı	p değeri
S9	-2.3631	0.0181
S23	-2.7785	0.0055
S47	2.2553	0.0241
S48	2.6086	0.0091
S61	2.5703	0.0102
S62	3.1287	0.0018
S65	3.5054	0.0005
S70	2.7883	0.0053
S76	-1.9474	0.0215
S80	-1.9028	0.0371

Tablo 32'de Olabilirlik Oranı yöntemi ile cinsiyet açısından Türkiye-İran için DMF analizinin sonuçları görülmektedir. Analiz için belirlenen odak grup kadınlardır. Tablo 32'de verilen maddeler kadınları erkeklerin karşısında dezavantajlı duruma getirmektedir.

3.2.1.2. Türkiye-Azerbaycan Karşılaştırması

Cinsiyet açısından DMF araştırmasında ikinci karşılaştırma Türkiye ile Azerbaycan arasında yapılmıştır. Karşılaştırmalarda odak grup olarak kadınlar seçilmiştir. DMF içerdiği saptanan maddeler kadınlar için dezavantajlı oldukları düşünülen maddelerdir.

Tablo 33. Türkiye-Azerbaycan cinsiyete göre BD yöntemi sonuçları

Maddeler	Breslow-Day İstatistiği	Serbestlik derecesi	p değeri
S1	74.1107	46	0.0054
S3	73.0990	41	0.0015
S11	60.4124	42	0.0326
S32	61.6937	40	0.0154
S34	70.8539	44	0.0063
S41	77.4388	47	0.0034
S58	72.9027	37	0.0004
S59	55.7887	40	0.0497
S65	68.8631	39	0.0022

Türkiye -Azerbaycan karşılaştırmasında BD yöntemi ile yapılan DMF araştırmasında odak grup kadınlar olmak üzere Tablo 33'te gösterilen maddelerin DMF içerdikleri saptanmıştır. İlgili maddeler kadınlar için dezavantajlıdır.

Tablo 34. Türkiye-Azerbaycan cinsiyete göre lojistik regresyon yöntemi

Maddeler	Lojistik Regresyon	p değeri
S5	6.4642	0.0395
S10	8.3288	0.0155
S13	11.5115	0.0032
S16	6.7331	0.0345
S24	15.7891	0.0004
S48	6.8257	0.0329
S49	0.0241	0.0241
S52	13.7588	0.0010

Tablo 34'te Türkiye-Azerbaycan için LR yöntemi ile yapılmış cinsiyet açısından DMF analizinin sonuçları görülmektedir. Tablo 34'teki ilgili maddeler odak grup kadınlar olarak belirlendiğinde DMF içeren maddeler olarak tespit edilmiştir.

Tablo 35. Türkiye-Azerbaycan cinsiyete göre Lord'un χ^2 istatistiği yöntemi

Maddeler	Lord'un χ^2 İstatistiği	p değeri
S13	3.9776	0.0461
S16	4.6153	0.0317
S24	9.2684	0.0023
S41	3.9659	0.0464
S48	5.6696	0.0173
S49	4.8578	0.0275
S52	6.6364	0.0100

Tablo 35'te belirlenmiş maddeler Lord'un χ^2 İstatistiği yöntemine göre saptanmış cinsiyet açısından DMF analizinde DMF içerdiği belirlenen maddelerdir. Burada, odak grup kadınlar olarak belirlenmiştir yani, ilgili maddeler kadınların erkeklere göre dezavantajlı olmalarına neden olmaktadır.

Tablo 36. Türkiye-Azerbaycan cinsiyete göre Raju'nun alan ölçümleri yöntemi

Maddeler	Raju'nun Alan Ölçümleri	p değeri
S13	-1.9637	0.0496
S16	-2.1195	0.0340
S24	-3.0132	0.0026
S41	2.0211	0.0433
S48	2.4108	0.0159
S49	2.2331	0.0255
S52	2.6058	0.0092

Tablo 36'da Türkiye-Azerbaycan için cinsiyet açısından Raju'nun Alan Ölçümleri Yöntemi ile DMF analizi yapılmış ve gösterilen maddelerin odak grup kadınlara adil davranmadığı saptanmıştır.

Tablo 37. Türkiye-Azerbaycan cinsiyete göre olabilirlik oranı yöntemi

Maddeler	Olabilirlik Oranı	p değeri
S13	-2.3800	0.0173
S16	2.4826	0.0130
S24	-3.6269	0.0003
S41	1.8165	0.0293
S48	2.1897	0.0285
S49	1.9251	0.0442
S52	2.7460	0.0060

Tablo 37’de Türkiye-Azerbaycan için Olabilirlik Oranı yöntemi ile DMF analizi sonuçları verilmiştir. Analiz cinsiyet açısından yapılmış ve odak grup olarak kadınlar olarak belirlenmiştir. Tablo 37’de verilmiş ilgili maddeler kadınları dezavantajlı duruma düşürmektedir.

3.2.1.3. Almanya- Suriye Karşılaştırılması

Almanya ve Suriye arasında cinsiyet açısından yapılan DMF analizinde BD, LR, Lord’un χ^2 İstatistiği, Raju’nun Alan Ölçümleri ve Olabilirlik Oranı yöntemleri kullanılmıştır. DMF içerdikleri belirlenen maddeler odak grup kadınların erkeklere göre bu maddelerde daha dezavantajlı olmalarına neden olmaktadır.

Tablo 38. Almanya-Suriye cinsiyete göre BD yöntemi

Maddeler	Breslow-Day İstatistiği	Serbestlik derecesi	p değeri
S20	56.8452	39	0.0323
S27	62.0759	43	0.0299
S31	68.6067	47	0.0215
S34	73.8684	45	0.0043
S43	75.2491	42	0.0012
S50	72.2234	42	0.0026
S55	64.1753	41	0.0118
S59	59.1548	41	0.0330
S66	68.7618	51	0.0492
S70	75.9722	51	0.0132

Tablo 38’de Almanya-Suriye için cinsiyete göre BD yöntemiyle yapılan DMF analizi sonuçları görülmektedir. DMF analizi yapılırken odak grup olarak kadınlar olarak belirlenmiştir. İlgili maddeler kadınlara erkeklerle eşit bir şekilde davranmamaktadır.

Tablo 39. Almanya-Suriye cinsiyete göre lojistik regresyon yöntemi

Maddeler	Lojistik Regresyon	p değeri
S14	15.9305	0.0003
S16	6.9854	0.0304
S21	11.0055	0.0041
S60	8.0951	0.0175
S70	8.6507	0.0132

Tablo 39’da Almanya-Suriye için cinsiyet açısından LR yöntemi ile DMF analizi sonuçları görülmektedir. Odak grup kadınlar olarak belirlendiğinde Tablo 39’da gösterilen ilgili maddelerin DMF içerdikleri saptanmıştır.

Tablo 40. Almanya-Suriye cinsiyete göre Lord’un χ^2 istatistiği

Maddeler	Lord’un χ^2 İstatistiği	p değeri
S14	7.1695	0.0074
S21	7.8722	0.0050
S31	4.8204	0.0281
S60	6.2616	0.0123
S70	8.2496	0.0041

Tablo 40’ta Almanya-Suriye için cinsiyet açısından Lord’un χ^2 İstatistiği ile yapılmış DMF analizinin sonuçları verilmektedir. Burada, odak grup kadınlar olarak belirlenmiş olup Tablo 39 ile verilen ilgili maddeler DMF içermektedir.

Tablo 41. Almanya-Suriye cinsiyete göre Raju’nun alan ölçümleri

Maddeler	Raju’nun Alan Ölçümleri	p değeri
S14	-2.7018	0.0069
S21	-2.8312	0.0046
S31	-2.2211	0.0281
S60	2.4817	0.0131
S70	2.8463	0.0044

Almanya-Suriye için cinsiyet açısından Raju’nun Alan Ölçümleri ile DMF analizi yapılmıştır. Odak grup olarak kadınlar belirlenmiş ve Tablo 41’de verilen ilgili maddeler kadınlar için dezavantajlıdır.

Tablo 42. Almanya-Suriye cinsiyete göre olabilirlik oranı

Maddeler	Olabilirlik Oranı	p değeri
S14	-3.2275	0.0012
S21	-2.7152	0.0066
S31	-2.4390	0.0147
S60	3.0293	0.0025
S70	2.5762	0.0100

Almanya-Suriye için uyruk açısından Olabilirlik Oranı yöntemiyle yapılan DMF analizinin sonuçları Tablo 42’de görülmektedir. Analizde, odak grup kadınlar olarak belirlenmiş ve Tablo 42’de verilmiş ilgili maddeler kadınlar için dezavantajlıdır.

3.2.1.4. Hollanda-İran Karşılaştırması

Cinsiyet açısından yapılan son karşılaştırma Hollanda ve İran arasında yapılmıştır. Odak grubun kadınlar olarak belirlendiği analizlerde DMF içerdikleri saptanan maddelerin kadınlara adil davranmadıkları belirlenmiştir.

Tablo 43. Hollanda-İran cinsiyete göre BD yöntemi

Maddeler	Breslow-Day İstatistiği	Serbestlik derecesi	p değeri
S9	68.7505	51	0.0493
S11	72.7439	48	0.0121
S37	68.7657	50	0.0403
S57	75.4252	53	0.0232
S64	66.1758	47	0.0339
S69	76.9051	54	0.0220

Hollanda-İran için BD yönteminin cinsiyet açısından sonuçları Tablo 43’te görülmektedir. Odak grup kadınlar olmak üzere, Tablo 43’te verilmiş ilgili maddeler kadınları erkeklere göre dezavantajlı yapmaktadır.

Tablo 44. Hollanda-İran cinsiyete göre lojistik regresyon yöntemi

Madde	Lojistik Regresyon	p değeri
S9	14.8318	0.0006
S10	6.2452	0.0440
S21	10.4350	0.0054
S23	6.4532	0.0397
S57	6.1980	0.0451
S62	10.4100	0.0055
S65	11.4576	0.0033
S70	8.0713	0.0177
S76	6.2597	0.0437

Tablo 44 ile Hollanda-İran karşılaştırmasında odak grup kadınlar iken LR yöntemi ile belirlenmiş DMF içeren maddeler gösterilmektedir.

Tablo 45. Hollanda-İran cinsiyete göre Lord'un χ^2 istatistiği

Maddeler	Lord'un χ^2 İstatistiği	p değeri
S10	5.2004	0.0226
S23	5.3575	0.0206
S62	8.3952	0.0038
S65	9.4026	0.0022
S70	4.3177	0.0377

Hollanda-İran için cinsiyet açısından Lord'un χ^2 İstatistiği yöntemi ile yapılan DMF analizinin sonucu Tablo 45 ile gösterilmiştir. Analizde odak grup kadınlar olarak belirlenmiş ve Tablo 45'teki ilgili maddelerin DMF içerdiği saptanmıştır.

Tablo 46. Hollanda-İran cinsiyete göre Raju'nun alan ölçümleri

Maddeler	Raju'nun Alan Ölçümleri	p değeri
S10	-2.1998	0.0278
S23	-2.2339	0.0255
S62	2.9677	0.0030
S65	3.1371	0.0017
S70	2.1646	0.0304

Hollanda-İran için Raju'nun Alan Ölçümleri yöntemi ile cinsiyet açısından yapılan DMF analizinin sonuçları Tablo 46'da görülmektedir. Analizde odak grup kadınlar olarak belirlenmiş ve elde edilen sonuçlar Tablo 46'da verilmiştir.

Tablo 47. Hollanda-İran cinsiyete göre olabilirlik oranı

Maddeler	Olabilirlik Oranı	p değeri
S10	-2.1450	0.0320
S23	-2.3480	0.0189
S62	2.8350	0.0046
S65	3.3028	0.0010
S70	2.6262	0.0086

Tablo 47 ile Hollanda-İran için cinsiyete göre Olabilirlik Oranı yöntemi ile DMF analizi sonuçları görülmektedir. Analizde, odak grup kadınlar olarak belirlenmiş ve Tablo 47’de verilen maddelerin DMF içerdikleri belirlenmiştir.

3.2.2. KTÜYÖS-2019 Uyrak Açısından Değişen Madde Fonksiyonu Analizi

KTÜYÖS-2019 için yapılan DMF analizinin ikinci ayağı, seçilmiş 5 ülkenin sosyokültürel farkları üzerine düşünülerek gerçekleştirilmiştir. Bu karşılaştırmalar ile farklı kültürlerden gelmenin KTÜYÖS-2019 maddeleri üzerinde avantaj ya da dezavantaj sağlayıp sağlamadığı incelenmiştir.

3.2.2.1. Türkiye-İran Karşılaştırması

Odak grup İran uyruklu vatandaşlar olmak üzere KTK ve MTK’ya dayalı yöntemlerle DMF çalışması yapılmıştır.

Tablo 48. Türkiye-İran uyruk için BD istatistiği sonuçları

Maddeler	Breslow-Day İstatistiği	Serbestlik derecesi	p değeri
S14	62.7395	45	0.0412
S19	68.8629	48	0.0258
S28	68.3756	48	0.0282
S33	67.0552	45	0.0181
S40	79.3754	49	0.0039
S48	84.0046	50	0.0018
S50	80.6286	48	0.0022
S52	99.3313	49	0.0000
S54	97.0923	51	0.0001
S58	66.9671	49	0.0449
S77	74.1140	52	0.0237

Tablo 48’de Türkiye ve İran için odak grup İran vatandaşları olduğunda BD İstatistiği yöntemi ile DMF analizinin sonuçları görülmektedir. Tablo 48’den elde edilen sonuçlara göre, 14, 19, 28, 33, 40, 48, 50, 52, 54, 58 ve 77 numaralı maddeler İran vatandaşlarının dezavantajlı olmalarına neden olur.

Tablo 49. Türkiye-İran uyruk için lojistik regresyon yöntemi

Maddeler	Lojistik Regresyon DMF İstatistiği	p değeri
S1	6.9431	0.0311
S2	10.1805	0.0062
S5	12.5025	0.0019
S6	5.9939	0.0499
S14	9.2168	0.0100
S18	6.0069	0.0496
S20	8.1847	0.0167
S21	7.1286	0.0283
S31	17.0047	0.0002
S35	28.5132	0.0000
S37	7.7118	0.0212
S39	6.9827	0.0305
S40	27.8307	0.0000
S44	7.4500	0.0241
S49	6.7120	0.0349
S51	11.7790	0.0028
S52	8.3469	0.0154
S58	14.8890	0.0006
S61	13.5567	0.0011
S62	11.0216	0.0040
S67	6.0950	0.0475
S70	10.6426	0.0049
S72	13.1039	0.0014
S76	7.4862	0.0237
S77	15.8955	0.0004
S79	9.7219	0.0077

Tablo 49’da Türkiye ve İran için odak grup İran vatandaşları olduğunda LR yöntemi ile DMF analizinin sonuçları görülmektedir. Yapılan analiz sonucu elde edilen sonuçlara göre; 1, 2, 5, 6, 14, 18, 20, 21, 31, 35, 37, 39, 40, 44, 49, 51, 52, 58, 61, 62, 67, 70, 72, 76, 77 ve 79 numaralı maddelerde DMF saptanmıştır.

Tablo 50. Türkiye-İran uyruk için Lord'un χ^2 istatistiği yöntemi

Maddeler	Lord'un χ^2 İstatistiği	p değeri
S2	8.1509	0.0043
S3	5.6852	0.0171
S5	21.4843	0.0000
S6	7.7094	0.0055
S11	6.6423	0.0100
S12	14.4557	0.0001
S13	6.4144	0.0113
S14	15.3909	0.0001
S15	9.7618	0.0018
S16	7.2903	0.0069
S17	5.8132	0.0159
S18	8.5922	0.0034
S20	12.0772	0.0005
S21	4.4288	0.0353
S24	7.7591	0.0053
S25	4.8026	0.0284
S35	22.5602	0.0000
S37	10.9988	0.0009
S40	4.6769	0.0306
S44	4.2486	0.0393
S51	6.4341	0.0112
S52	4.2400	0.0395
S57	6.6099	0.0101
S58	11.2672	0.0008
S62	6.2591	0.0124
S63	9.7496	0.0018
S67	15.5807	0.0001
S69	13.4814	0.0002
S70	14.1306	0.0002
S71	7.7607	0.0053
S72	19.2541	0.0000
S73	3.9250	0.0476
S76	5.5701	0.0183
S77	20.2406	0.0000
S79	11.2692	0.0008
S80	7.9933	0.0047

Tablo 50'de uyuğu göre Türkiye-İran karşılaştırması Lord'un χ^2 İstatistiği yöntemi ile yapılmıştır. Burada, odak grup İran vatandaşlarıdır. Elde edilen sonuçlara göre, Tablo 50'de gösterilen maddelerin DMF içerdikleri saptanmıştır.

Tablo 51. Türkiye-İran uyruk için Raju'nun alan ölçümleri yöntemi

Maddeler	Raju'nun Alan Ölçümleri	p değeri
S2	-2.8550	0.0043
S3	-2.3844	0.0171
S5	-4.6351	0.0000
S6	-2.7766	0.0055
S11	2.5773	0.0100
S12	-3.8021	0.0001
S13	-2.5327	0.0113
S14	-3.9231	0.0001
S15	-3.1244	0.0018
S16	-2.7001	0.0069
S17	-2.4111	0.0159
S18	-2.9312	0.0034
S20	3.4752	0.0005
S21	-2.1045	0.0353
S24	-2.7855	0.0053
S25	-2.1915	0.0284
S35	-4.7498	0.0000
S37	-3.3164	0.0009
S40	-2.1626	0.0306
S44	2.0612	0.0393
S51	2.5366	0.0112
S52	2.0591	0.0395
S57	2.5710	0.0101
S58	3.3567	0.0008
S62	2.5018	0.0124
S63	3.1224	0.0018
S67	3.9472	0.0001
S69	3.6717	0.0002
S70	3.7591	0.0002
S71	2.7858	0.0053
S72	4.3879	0.0000
S73	1.9812	0.0476
S76	2.3601	0.0183
S77	4.4990	0.0000
S79	3.3570	0.0008
S80	2.8272	0.0047

Türkiye-İran arasında uyruğa göre DMF analizi yapılmış ve DMF içerdiği belirlenen maddeler Tablo 51'de verilmiştir. Analiz yapılırken Raju'nun Alan Ölçümleri yöntemi kullanılmış ve odak grup olarak İran vatandaşları seçilmiştir. Tablo 51'de gösterilen ilgili maddeler İran vatandaşları için dezavantajlı maddelerdir.

Tablo 52. Türkiye-İran uyruk için olabilirlik oranı yöntemi

Maddeler	Olabilirlik Oranı	p değeri
S2	-2.1393	0.0324
S3	-1.6816	0.0126
S5	3.6183	0.0003
S6	2.6324	0.0085
S11	1.6438	0.0002
S12	-2.2840	0.0224
S13	-1.6941	0.0202
S14	-2.8862	0.0039
S15	-1.2848	0.0089
S16	-1.5226	0.0127
S17	-1.7519	0.0298
S18	-2.0162	0.0438
S20	-2.8435	0.0045
S21	-2.1418	0.0322
S24	-1.2674	0.0205
S25	-2.0319	0.0422
S35	4.6296	0.0000
S37	-2.5207	0.0117
S40	2.0288	0.0425
S44	2.2503	0.0244
S51	2.1516	0.0314
S52	2.2372	0.0253
S57	2.5710	0.0101
S58	3.3567	0.0008
S62	2.9030	0.0037
S63	1.1525	0.2491
S67	2.0733	0.0381
S69	2.1897	0.0285
S70	2.0419	0.0412
S71	1.0979	0.2722
S72	2.9807	0.0029
S73	1.6580	1.6580
S76	1.5121	0.0130
S77	2.5628	0.0104
S79	2.9379	0.0033
S80	1.6187	0.0105

Tablo 52’de Olabilirlik Oranı yöntemi ile uyruk açısından DMF analizi yapılmıştır. Elde edilen sonuçlara göre, odak grup İran vatandaşları olmak üzere, Tablo 52’de gösterilen ilgili maddelerde DMF saptanmıştır.

3.2.2.2. Türkiye-Azerbaycan Karşılaştırması

Türkiye-Azerbaycan karşılaştırmasında odak grup olarak Azerbaycan belirlenmiştir. Buna göre KTK'ya ve MTK'ya dayalı yöntemlerle elde edilen sonuçlar karşılaştırılmıştır.

Tablo 53. Türkiye-Azerbaycan uyruk için BD yöntemi sonuçları

Maddeler	Breslow-Day İstatistiği	Serbestlik derecesi	p değeri
S1	58.8900	42	0.0435
S19	58.1482	39	0.0249
S26	73.0520	42	0.0021
S32	54.3694	38	0.0414
S33	47.7118	27	0.0083
S40	62.7987	42	0.0204
S42	56.2265	38	0.0287
S51	58.1027	40	0.0320
S52	66.7812	43	0.0116
S53	61.4508	43	0.0336
S54	69.2892	40	0.0028
S58	54.8158	34	0.0133
S68	0.0138	25	0.0138
S77	65.9285	46	0.0285

Tablo 53'te Türkiye-Azerbaycan için uyruk açısından BD yönteminden elde edilmiş, DMF içeren maddeler görülmektedir. Odak grup Azerbaycan vatandaşları olarak belirlenmiştir. Sonuçlara göre, ilgili maddeler Azerbaycan vatandaşlarını Türk vatandaşlarının karşısında dezavantajlı konuma sokmaktadır.

Tablo 54. Türkiye-Azerbaycan uyruk için lojistik regresyon yöntemi

Maddeler	Lojistik Regresyon	p değeri
S5	6.6647	0.0357
S29	6.8717	0.0322
S31	6.1862	0.0454
S35	13.3185	0.0013
S40	11.1518	0.0038
S51	10.6206	0.0049
S62	8.7879	0.0124
S67	6.4214	0.0403
S72	9.2053	0.0100

Tablo 54'te LR yöntemi ile uyruk açısından yapılmış DMF analizinin sonuçları görülmektedir. Elde edilen sonuçlara göre, Tablo 54'te verilmiş ilgili maddeler odak grup Azerbaycan için DMF içermektedir.

Tablo 55. Türkiye-Azerbaycan uyruk için Lord'un χ^2 istatistiği yöntemi

Maddeler	Lord'un χ^2 İstatistiği	p değeri
S2	6.1453	0.0132
S5	7.2854	0.0070
S15	4.9601	0.0259
S17	5.0444	0.0247
S35	8.5245	0.0035
S37	8.3844	0.0038
S51	4.3487	0.0370
S58	4.1762	0.0410
S62	6.9326	0.0085
S63	6.1527	0.0131
S66	15.5643	0.0001
S67	13.0030	0.0003
S69	4.5000	0.0339
S70	9.4838	0.0021
S72	10.2416	0.0014
S77	5.1847	0.0228
S80	6.5483	0.0105

Tablo 55'te Türkiye-Azerbaycan için uyruk açısından Lord'un χ^2 İstatistiği yöntemi ile DMF analizinin sonuçları verilmiştir. Odak grup Azerbaycan vatandaşları olarak belirlenmiş ve ilgili maddelerin DMF içerdikleri saptanmıştır.

Tablo 56. Türkiye-Azerbaycan uyruk için Raju'nun alan ölçümleri yöntemi

Maddeler	Raju'nun Alan Ölçümleri	p değeri
S2	-2.2749	0.0229
S5	-2.5073	0.0122
S15	-2.0235	0.0430
S17	-2.0243	-0.0429
S35	-2.6985	0.0070
S37	-2.6699	0.0076
S51	2.2812	0.0225
S58	2.2280	0.0259
S62	2.8293	0.0047
S63	2.7043	0.0068
S66	4.1787	0.0000
S67	3.8324	0.0001
S69	2.3459	0.0190
S70	3.3135	0.0009
S72	3.4343	0.0006
S77	2.5077	0.0122
S80	2.7891	0.0053

Tablo 56'da Türkiye-Azerbaycan için uyruk açısından DMF analizi Raju'nun Alan Ölçümleri yöntemi ile yapılmıştır. Odak grup Azerbaycan vatandaşları olmak üzere ilgili maddelerin DMF içerdikleri tespit edilmiştir.

Tablo 57. Türkiye-Azerbaycan uyruk için olabilirlik oranı

Maddeler	Olabilirlik Oranı	p değeri
S2	-2.2707	0.0232
S5	-2.3064	0.0211
S15	-2.7386	0.0046
S17	-1.1480	0.0251
S35	-1.8899	0.0288
S37	-1.7916	0.0332
S51	-2.2075	0.0273
S58	1.3800	0.0167
S62	2.7406	0.0061
S63	-2.2480	0.0246
S66	3.3248	0.0074
S67	3.9056	0.0036
S69	2.8389	0.0040
S70	2.4829	0.0138
S72	2.1233	0.0337
S77	2.0565	0.0095
S80	2.8812	0.0378

Tablo 57’de Türkiye-Azerbaycan için odak grup Azerbaycan vatandaşları olmak üzere Olabilirlik Oranı yöntemi ile DMF analizi yapılmış ve ilgili maddelerin Azerbaycan vatandaşları için DMF içeren maddeler olarak belirlenmiştir.

3.2.2.3. Almanya- Suriye Karşılaştırması

DMF analizinde Almanya-Suriye karşılaştırması yapılırken odak grup olarak Suriye belirlenmiştir.

Tablo 58. Almanya-Suriye uyruk için BD yöntemi

Maddeler	Breslow-Day İstatistiği	Serbestlik derecesi	p değeri
S13	42.5769	28	0.0382
S14	34.5026	22	0.0436
S21	40.2938	27	0.0481
S24	39.6015	25	0.0320
S31	40.6695	25	0.0249
S35	45.2102	26	0.0112
S39	42.4320	27	0.0298
S57	43.3382	29	0.0424
S60	37.8830	23	0.0262

Almanya-Suriye için uyruk açısından BD yöntemi ile DMF analizi yapılmıştır. Burada, odak grup Suriye vatandaşları olarak belirlenmiş ve Tablo 58’de saptanmış ilgili maddeler Suriye vatandaşları için DMF içermektedir.

Tablo 59. Almanya-Suriye uyruk için lojistik regresyon yöntemi

Maddeler	Lojistik Regresyon	p değeri
S39	7.2739	0.0263
S51	6.1948	0.0452
S59	6.2563	0.0438

Tablo 59’da Almanya-Suriye için uyruk açısından LR yöntemi ile DMF analizinin sonuçları görülmektedir. Burada, odak grup Suriye vatandaşları olarak belirlenmiş ve

Tablo 59’da verilmiş ilgili maddelerin Suriye vatandaşları için dezavantaj oluşturan maddelerdir.

Tablo 60. Almanya-Suriye uyruk için Lord’un χ^2 İstatistiği

Maddeler	Lord’un χ^2 İstatistiği	p değeri
S9	4.9688	0.0258
S66	3.9994	0.0455

Tablo 60’ta Almanya-Suriye için uyruk açısından Lord’un χ^2 İstatistiği ile yapılmış DMF analizinin sonuçları verilmektedir. Odak grup Suriye vatandaşları olarak belirlenmiş olup Tablo 60’ta verilmiş ilgili maddeler Suriye vatandaşları için DMF içermektedir.

Tablo 61. Almanya-Suriye Uyuğa göre Raju’nun alan ölçümleri

Maddeler	Raju’nun Alan Ölçümleri	p değeri
S9	2.2313	0.0257
S66	-1.9977	0.0458

Tablo 61, Almanya-Suriye için uyruk açısından Raju’nun Alan Ölçümleri yöntemiyle yapılmış DMF analizinin sonuçlarını içermektedir. Analizde, odak grup Suriye vatandaşları olarak belirlenmiş ve ilgili maddelerin Suriye vatandaşları için DMF içermektedir.

Tablo 62. Almanya-Suriye uyruk için olabilirlik oranı yöntemi

Maddeler	Olabilirlik Oranı	p değeri
S9	2.0126	0.0442
S66	-1.3357	0.0181

Tablo 62’de uyuğa göre Olabilirlik Oranı yöntemi ile Almanya ve Suriye vatandaşları için yapılan DMF analizi sonuçları görülmektedir. Analizde odak grup ilgili maddelerde Suriye vatandaşları için DMF saptanmıştır.

3.2.2.4. Hollanda-İran Karşılaştırması

Hollanda-İran karşılaştırması yapılırken KTK ve MTK bazlı yöntemlerin karşılaştırmaları da yapılmıştır. Burada odak grup olarak İran seçilmiştir.

Tablo 63. Hollanda-İran uyruk için BD yöntemi

Maddeler	Breslow-Day İstatistiği	Serbestlik derecesi	p değeri
S10	52.9091	35	0.0266
S18	59.1080	32	0.0025
S21	67.2347	21	0.0008
S24	57.3162	33	0.0054
S48	58.3323	35	0.0079
S54	48.6201	34	0.0498
S74	66.3815	36	0.0015

Tablo 63'te uyruğa göre yapılan Hollanda-İran kıyaslamasının sonuçları görülmektedir. Odak grup olarak İran vatandaşları olarak belirlendiğinde ilgili maddelerde DMF saptanmıştır.

Tablo 64. Hollanda-İran uyruk için lojistik regresyon yöntemi

Maddeler	Lojistik Regresyon	p değeri
S2	7.1509	0.0280
S3	9.1344	0.0104
S5	7.4086	0.0246
S10	10.8257	0.0045
S11	15.1720	0.0005
S16	7.7253	0.0210
S18	6.6849	0.0354
S22	10.0101	0.0067
S28	8.8674	0.0119
S32	10.2555	0.0059
S44	6.2032	0.0450
S47	9.0600	0.0108
S51	8.9055	0.0116
S54	11.7921	0.0028
S66	6.1359	0.0465
S68	7.8882	0.0194
S77	6.3218	0.0424

Hollanda-İran için uyruk açısından LR yöntemi ile yapılan DMF analizinin sonucu Tablo 64'te görülmektedir. Analiz, odak grup İran vatandaşları seçilerek yapılmıştır ve Tablo 64'te verilen maddeler DMF içeren maddeler olarak saptanmıştır.

Tablo 65. Hollanda-İran uyruk için Lord'un χ^2 istatistiği yöntemi

Maddeler	Lord'un χ^2 İstatistiği	p değeri
S2	7.3981	0.0065
S3	6.9820	0.0082
S4	6.0831	0.0136
S5	11.3943	0.0007
S10	5.1795	0.0229
S11	10.6203	0.0011
S14	6.8740	0.0087
S15	9.2781	0.0023
S16	7.8656	0.0050
S41	6.0470	0.0139
S44	4.7045	0.0301
S47	3.8567	0.0495
S51	6.4151	0.0113
S63	4.9524	0.0261
S69	7.3411	0.0067
S70	4.7076	0.0300
S75	6.2103	0.0127
S80	6.0052	0.0143

Hollanda-İran için Lord'un χ^2 İstatistiği yöntemi ile uyruk açısından yapılan DMF analizinin sonuçları Tablo 65'te görülmektedir. Analiz için belirlenen odak grup İran vatandaşlarıdır ve ilgili maddeler bu vatandaşlar için dezavantajlı olarak saptanmıştır.

Tablo 66. Hollanda-İran uyruk için Raju'nun alan ölçümleri

Maddeler	Raju'nun Alan Ölçümleri	p değeri
S2	-3.1047	0.0019
S3	-3.0279	0.0025
S4	-2.8521	0.0043
S5	-3.7518	0.0002
S10	-2.6609	0.0450
S11	-3.6443	0.0078
S14	-3.0055	0.0003
S15	-3.4304	0.0265
S16	-3.1880	0.0027
S41	1.8417	0.0006
S44	1.8417	0.0255
S47	1.5889	0.0112
S51	2.2522	0.0243
S63	1.8583	0.0331
S69	2.3422	0.0192
S70	1.7822	0.0074
S75	2.1043	0.0354
S80	2.0661	0.0388

Hollanda-İran için uyruk açısından Raju'nun Alan Ölçümleri yöntemi ile yapılan DMF çalışmasında DMF içerdiği belirlenen maddeler Tablo 66 ile verilmiştir. Analizde, odak grup İran vatandaşları olarak belirlenmiş olup ilgili maddelerin İran vatandaşları için DMF içerdiği belirlenmiştir.

Tablo 67. Hollanda-İran uyruğa göre olabilirlik oranı

Maddeler	Olabilirlik Oranı	p değeri
S2	-3.1047	0.0160
S3	-3.0279	0.0427
S4	- 1.8994	0.0375
S5	-2.6125	0.0090
S10	-2.2068	0.0273
S11	- 2.8372	0.0046
S14	- 2.0119	0.0442
S15	- 2.4058	0.0161
S16	- 2.0307	0.0423
S41	1.9145	0.0256
S44	2.4881	0.0128
S47	2.1740	0.0297
S51	2.0715	0.0383
S63	1.9868	0.0032
S69	2.3127	0.0207
S70	1.7346	0.0028
S75	1.5051	0.0132
S80	2.1761	0.0295

Tablo 67 ile Hollanda-İran kıyaslamasında Olabilirlik Oranı yöntemi ile uyruk açısından yapılan DMF analizinin sonuçları verilmiştir. Analizde, odak grup İran vatandaşları olarak belirlenmiş ve ilgili maddelerde İran vatandaşları için DMF saptanmıştır.

3.2.3. KTÜYÖS-2019 Bölgeler Açısından Değişen Madde Fonksiyonu Analizi

KTÜYÖS-2019'a ilişkin DMF araştırmasında cinsiyet ve uyruk açısından yapılan DMF analizlerinde hem katılım oranına hem de kültürel anlamda farklılık göstermesi durumuna göre beş ülke seçilmiş ve analizleri yapılmıştır. Ardından benzer coğrafyalarda olmaları sebebiyle benzer kültürel yapı gösteren ülkeler bölgelere ayrılmış, benzer sorular bölgeler arası farklar için de analiz edilmiştir. Bölgeler Batı Asya, Avrupa, Afrika, Güney Asya, Amerika, Orta Doğu ve Orta Asya olarak bölünmüştür. Ülkeler bölgelere ayrılırken benzer sistem ve kültürleri bir arada tutmak amaçlanmış ancak coğrafi bölge faktörü de atlanmamıştır. Farklı kültürlerde yetiştirilerek ve en önemlisi de farklı eğitim sistemlerinden geçirilerek sınava girmeye hak kazanan öğrencilerin madde tepkilerine ilişkin bölgeler arası sosyokültürel bir DMF söz konusu olup olmadığı araştırılmıştır. DMF analizleri için beş farklı bölge karşılaştırılmıştır. Kültürel açıdan yapılan bu karşılaştırmalarda seçilen bölgelerin sosyokültürel açıdan birbirlerinden farklılık göstermeleri önem teşkil etmektedir. Analizlere başlamadan önce her bir ikili karşılaştırma için etkili örneklem büyüklüğünün sağlanıp sağlanmadığı da ayrıca incelenmiştir. Etkili örneklem büyüklüğüne ilişkin analizlerden alınan sonuçlara göre seçilmiş beş karşılaştırmadan iki tanesi (Amerika-Afrika ve Orta Asya-Afrika) etkili örneklem büyüklüğünü sağlamıyorken üç tanesi (Avrupa-Orta Doğu, Batı Asya-Avrupa ve Güney Asya-Avrupa) etkili örneklem büyüklüğünü sağlamıştır.

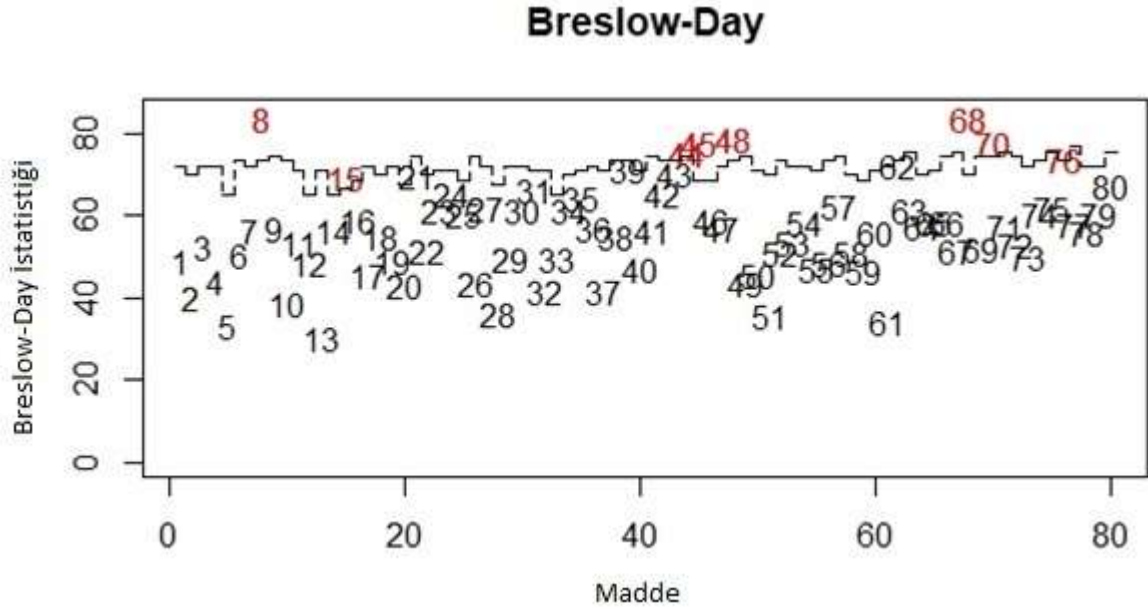
3.2.3.1. Güney Asya-Avrupa Karşılaştırması

Güney Asya-Avrupa karşılaştırmasında odak grup Güney Asya olarak belirlenmiştir. Etkili örneklem büyüklüğünü sağlayan bu karşılaştırmada KTK ve MTK bazlı yöntemler ile DMF analizleri yapılmıştır.

Tablo 68. Güney Asya ve Avrupa için BD yöntemi

Maddeler	Breslow-Day İstatistiği	Serbestlik derecesi	p değeri
S8	83.7441	55	0.0075
S15	69.3357	49	0.0294
S44	75.0017	56	0.0459
S45	77.5850	51	0.0096
S48	78.6665	55	0.0198
S68	83.6985	52	0.0035
S70	77.8536	56	0.0283
S76	73.8081	55	0.0461

Tablo 68 ile Güney Asya ile Avrupa karşılaştırması için BD yönteminden alınan sonuçlar görülmektedir. Burada, odak grup Güney Asya olmak üzere 8, 15, 44, 45, 48, 68, 70 ve 76 numaralı maddelerde DMF olduğu belirlenmiştir.



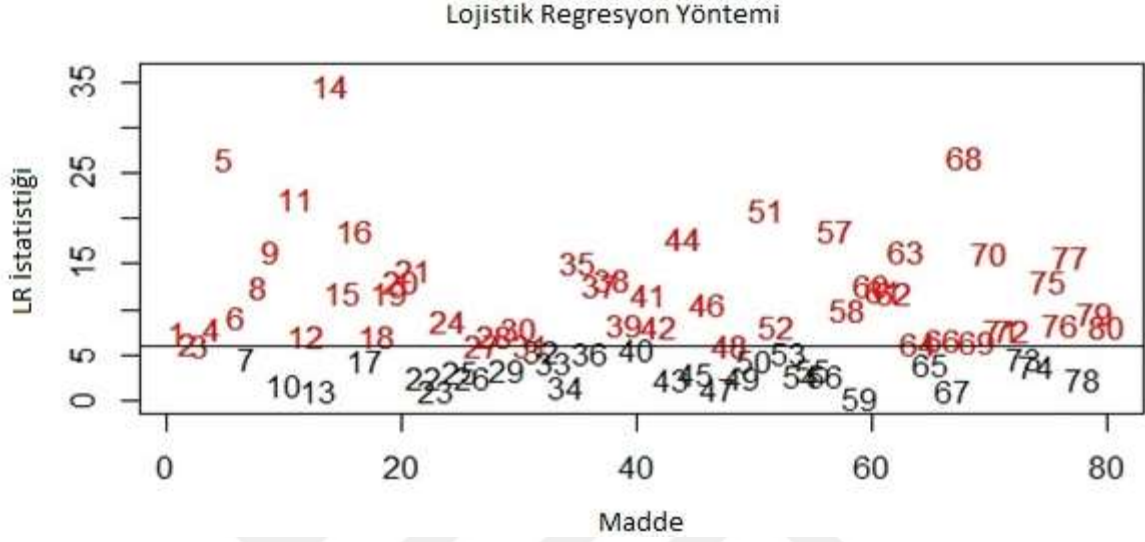
Şekil 12. Güney Asya ve Avrupa için Breslow-Day yöntemi

Şekil 12’de Tablo 68’de verilmiş DMF içerdiği belirlenen maddelerin eşik değerinin üzerinde kırmızı ile işaretli olduğu görülmektedir.

Tablo 69. Güney Asya ve Avrupa için lojistik regresyon yöntemi

Maddeler	Lojistik Regresyon	p değeri
S1	7.5891	0.0225
S2	6.2362	0.0442
S3	6.1393	0.0464
S4	7.9587	0.0187
S5	26.4971	0.0000
S8	12.4347	0.0020
S9	16.5276	0.0003
S11	22.2632	0.0000
S12	7.0432	0.0296
S14	34.6043	0.0000
S15	11.9354	0.0026
S16	18.6729	0.0001
S18	7.1389	0.0282
S19	11.8353	0.0027
S20	13.0805	0.0014
S21	14.3006	0.0008
S24	8.7814	0.0124
S27	6.1506	0.0462
S28	0.0284	0.0284
S30	7.9357	0.0189
S31	6.0150	0.0494
S35	15.2912	0.0005
S37	12.7674	0.0017
S38	13.2654	0.0013
S39	8.3120	0.0157
S41	11.7011	0.0029
S42	8.2043	0.0165
S44	17.9035	0.0001
S46	10.6350	0.0049
S48	5.9976	0.0498
S51	20.9727	0.0000
S52	8.2328	0.0163
S57	18.6336	0.0001
S58	10.1006	0.0064
S60	12.7622	0.0017
S61	12.0029	0.0025
S62	11.9478	0.0025
S63	16.4256	0.0003
S64	6.3474	0.0418
S66	6.7816	0.0337
S68	26.8107	0.0000
S69	6.5426	0.0380
S70	16.2244	0.0003
S71	7.7868	0.0204
S72	7.7342	0.0209
S75	13.0934	0.0014
S76	8.3003	0.0158
S77	15.7672	0.0004
S79	9.6106	0.0082
S80	8.0997	0.0174

Tablo 69’da Güney Asya ve Avrupa için yapılan DMF çalışması için LR yönteminin sonuçları görülmektedir. Odak grup Güney Asya olmak üzere ilgili maddelerde DMF mevcuttur.



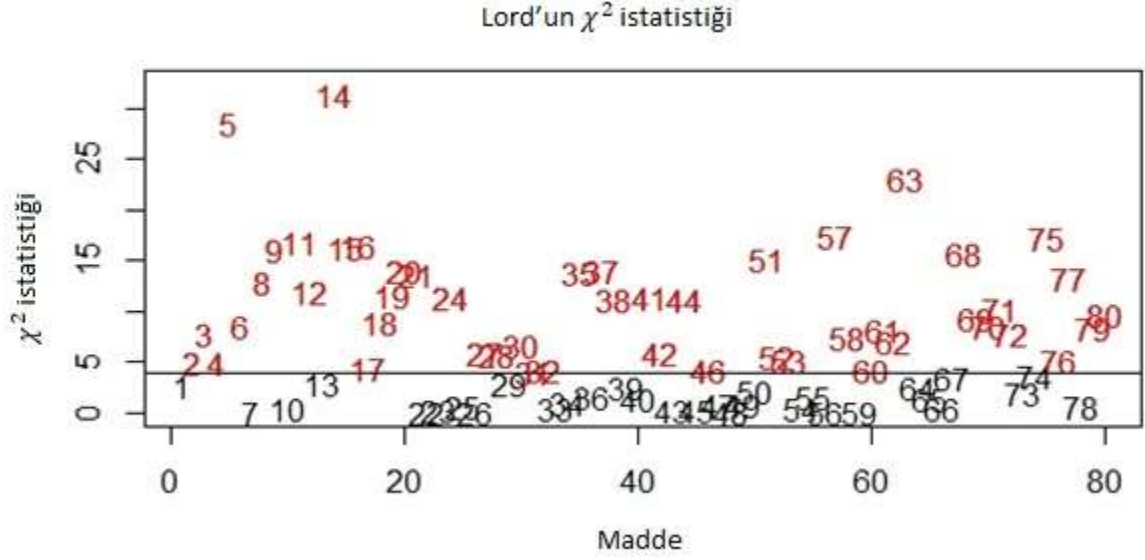
Şekil 13. Güney Asya ve Avrupa için lojistik regresyon yöntemi

Şekil 13’te Tablo 69’da mevcut olan Lojistik Regresyon yöntemine göre DMF içerdiği saptanan maddeler kırmızı ile belirtilmiştir.

Tablo 70. Güney Asya ve Avrupa için Lord'un χ^2 istatistiği

Maddeler	Lord'un χ^2 İstatistiği	p değeri
S2	4.9425	0.0262
S3	7.7146	0.0055
S4	5.0418	0.0247
S5	28.5580	0.0000
S6	8.5225	0.0035
S8	12.8618	0.0003
S9	16.1777	0.0001
S11	16.8337	0.0000
S12	11.8495	0.0006
S14	31.4273	0.0000
S15	16.3091	0.0001
S16	16.3899	0.0001
S17	4.4692	0.0345
S18	9.0098	0.0027
S19	11.4776	0.0007
S20	14.1083	0.0002
S21	13.5699	0.0002
S24	11.4467	0.0007
S27	5.9233	0.0149
S28	5.6498	0.0175
S30	6.6232	0.0101
S31	4.0980	0.0429
S35	13.8648	0.0002
S37	13.9528	0.0002
S38	11.1055	0.0009
S41	11.3019	0.0008
S42	5.9135	0.0150
S44	11.1624	0.0008
S51	15.2455	0.0001
S52	5.4818	0.0192
S53	5.1627	0.0231
S57	17.3525	0.0000
S58	7.3353	0.0068
S60	4.1884	0.0407
S61	8.1070	0.0044
S62	7.0416	0.0080
S63	23.0460	0.0000
S68	15.7895	0.0001
S69	9.3831	0.0022
S70	8.5245	0.0035
S71	10.2578	0.0014
S72	7.8571	0.0051
S75	17.2209	0.0000
S76	5.1747	0.0229
S77	13.2189	0.0003
S79	8.3589	0.0038
S80	9.6584	0.0019

Tablo 70’te Güney Asya ve Avrupa için yapılan DMF çalışmasında Lord’un χ^2 İstatistiği yönteminin sonuçları görülmektedir. Odak grup Güney Asya olmak üzere tabloda belirlenen maddelerin DMF içerdiği saptanmıştır.



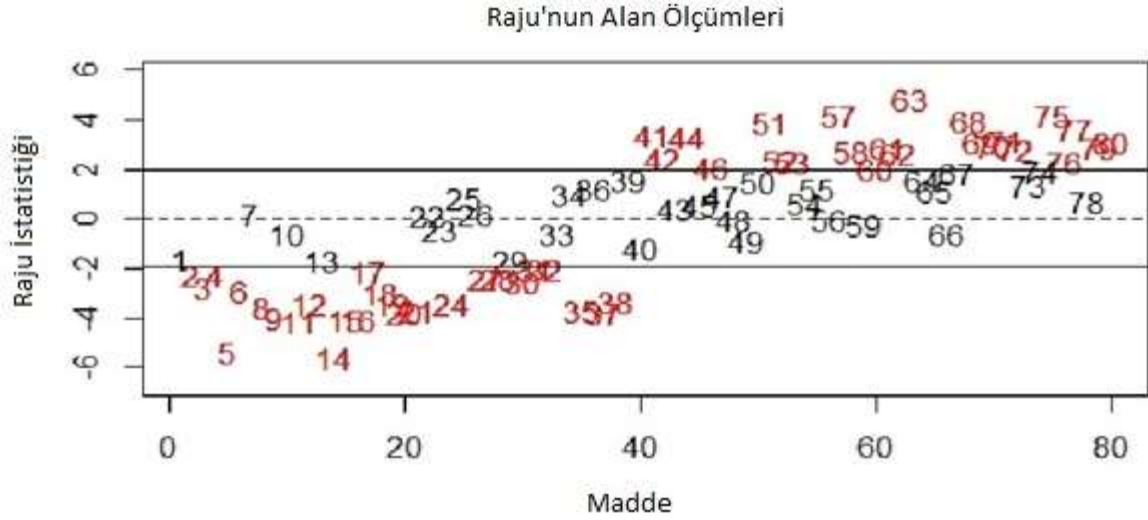
Şekil 14. Güney Asya ve Avrupa için Lord’un χ^2 istatistiği

Tablo 70’te DMF içerdiği belirlenen maddeler Şekil 14’te kırmızı ile gösterilmiştir. Eşik değerinin üzerindeki maddelerin Güney Asya’dan katılan öğrencilerin aleyhine çalıştıkları görülmektedir.

Tablo 71. Güney Asya ve Avrupa için Raju'nun alan ölçümleri

Maddeler	Raju'nun Alan Ölçümleri	p değeri
S2	-2.2232	0.0262
S3	-2.7775	0.0055
S4	-2.2454	0.0247
S5	-5.3440	0.0000
S6	-2.2232	0.0262
S8	-3.5863	0.0003
S9	-4.0221	0.0001
S11	-4.1029	0.0000
S12	-3.4423	0.0006
S14	-5.6060	0.0000
S15	-4.0385	0.0001
S16	-4.0484	0.0001
S17	-2.1141	0.0345
S18	-3.0016	0.0027
S19	-3.3879	0.0007
S20	-3.7561	0.0002
S21	-3.6837	0.0002
S24	-3.3833	0.0007
S27	-2.4338	0.0149
S28	-2.3769	0.0175
S30	-2.5736	0.0101
S31	-2.0244	0.0429
S35	-3.7235	0.0002
S37	-3.7353	0.0002
S38	-3.3325	0.0009
S41	3.3618	0.0008
S42	2.4318	0.0150
S44	3.3410	0.0008
S51	3.9045	0.0001
S52	2.3413	0.0192
S53	2.2722	0.0231
S57	4.1656	0.0000
S58	2.7084	0.0068
S60	2.0466	0.0407
S61	2.8473	0.0044
S62	2.6536	0.0080
S63	4.8006	0.0000
S68	3.9736	0.0001
S69	3.0632	0.0022
S70	2.9197	0.0035
S71	3.2028	0.0014
S72	2.8031	0.0051
S75	4.1498	0.0000
S76	2.2748	0.0229
S77	3.6358	0.0003
S79	2.8912	0.0038
S80	3.1078	0.0019

Tablo 71 ile Güney Asya ve Avrupa için yapılan DMF çalışmasında Raju'nun Alan Ölçümleri yönteminin sonuçları görülmektedir. Odak grup Güney Asya olmak üzere tabloda verilmiş maddelerin DMF içerdiği saptanmıştır.



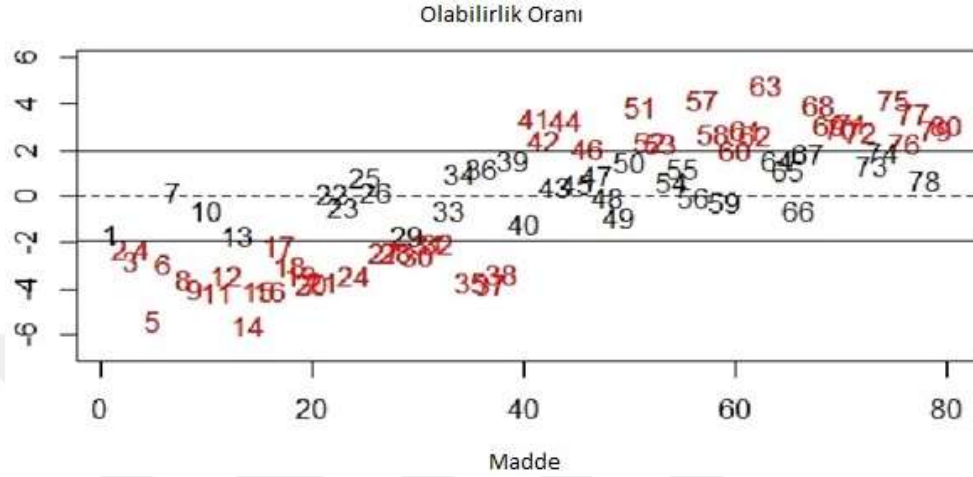
Şekil 15. Güney Asya ve Avrupa için Raju'nun alan ölçümleri

Şekil 15'te Tablo 71'te DMF içerdiği saptanan maddelerin grafik üzerindeki konumları görülmektedir. Analizler sonrası DMF içerdiği belirlenen maddeler kırmızı ile gösterilmektedir.

Tablo 72. Güney Asya ve Avrupa için olabilirlik oranı yöntemi

Maddeler	Olabilirlik Oranı	p değeri
S2	-2.8275	0.0476
S3	-2.3469	0.0189
S4	-2.6295	0.0103
S5	-4.8624	0.0000
S6	-2.7115	0.0067
S8	-3.5487	0.0004
S9	-3.8392	0.0001
S11	-3.7906	0.0002
S12	-2.7909	0.0053
S14	-5.2510	0.0000
S15	-3.3684	0.0008
S16	-3.4477	0.0006
S17	-2.9114	0.0359
S18	-2.2764	0.0228
S19	-3.5130	0.0004
S20	-3.6579	0.0003
S21	-3.5069	0.0005
S24	-3.0117	0.0026
S27	-2.1403	0.0323
S28	-2.5649	0.0103
S30	-2.8838	0.0039
S31	-2.3140	0.0207
S35	-3.7620	0.0002
S37	-3.7389	0.0002
S38	-3.3470	0.0008
S41	2.8926	0.0038
S42	3.0710	0.0021
S44	3.2949	0.0010
S51	3.8684	0.0001
S52	2.4488	0.0143
S53	2.7531	0.0296
S57	3.3618	0.0008
S58	2.8255	0.0047
S60	2.4616	0.0138
S61	3.6484	0.0003
S62	3.2977	0.0010
S63	3.5418	0.0004
S68	4.6246	0.0000
S69	2.4480	0.0144
S70	1.9925	0.0463
S71	2.4925	0.0127
S72	2.2159	0.0267
S75	3.1203	0.0018
S76	2.1653	0.0304
S77	2.2426	0.0249
S79	2.5442	0.0110
S80	2.5611	0.0104

Tablo 72’de gösterilen maddeler Güney Asya ve Avrupa arasındaki DMF çalışmasında Olabilirlik Oranı yöntemine göre odak grup Güney Asya olduğunda DMF içerdiği belirlenen maddelerdir.



Şekil 16. Güney Asya ve Avrupa için olabilirlik oranı yöntemi

Şekil 16 ile Tablo 72’de Olabilirlik Oranı yöntemine göre DMF içerdiği belirlenmiş maddelerin grafik üzerindeki konumları kırmızı ile işaretlenmiş olarak gösterilmektedir.

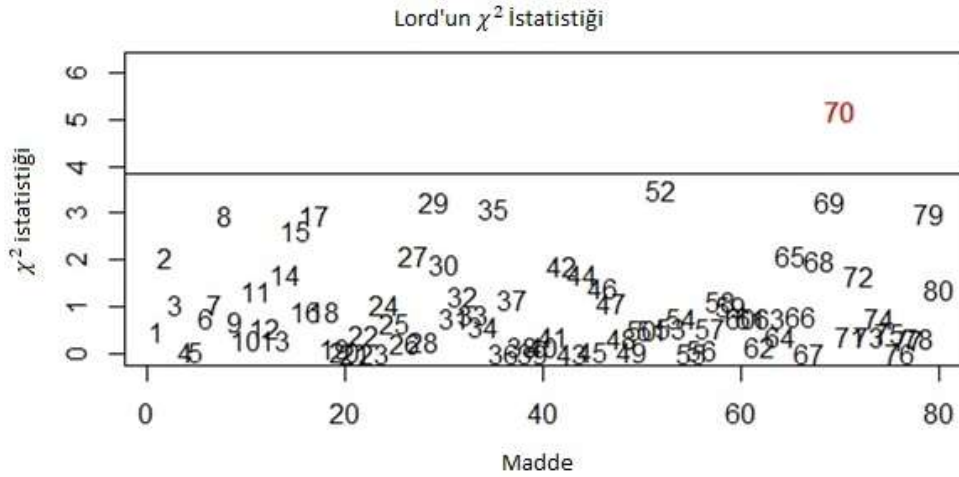
3.2.3.2. Amerika-Afrika Karşılaştırması

Etkili örneklem büyüklüğünü sağlamayan karşılaştırmalardan bir tanesi Amerika ve Afrika karşılaştırmasıdır. Burada, Amerika’dan 18 katılım Afrika’dan ise 40 katılım mevcuttur. 80 maddelik bir ölçek için 58 kişilik bir örneğin yapılacak analizler için yeterli olması mümkün değildir. Test uzunluğu da bu kadar az bir örnek için fazla uzundur. Bu durumda KTK’ya dayalı yöntemler çalışmamış, uyum hatası vermiştir. MTK’ya dayalı yöntemlerde ise beklediğimiz etkili örneklem büyüklüğü ve yeterli test uzunluğu olması halinde yöntemlerin çalışma prensiplerinden farklı sonuçlar elde edilmiştir. Sonuç veren yöntemler için odak grup Afrika olarak belirlenmiştir.

Tablo 73. Amerika ve Afrika için Lord'un χ^2 istatistiği

Maddeler	Lord'un χ^2 İstatistiği	p değeri
S70	5.1785	0.0229

Tablo 73 ile Amerika ve Afrika karşılaştırması için odak grup Afrika olmak üzere Lord'un χ^2 İstatistiği yöntemine göre DMF içerdiği belirlenen sadece 70 nolu maddedir.

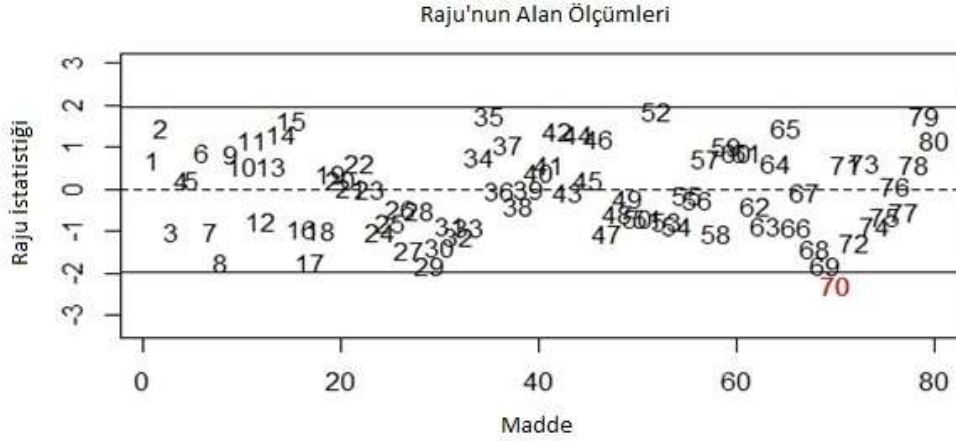
Şekil 17. Amerika ve Afrika için Lord'un χ^2 istatistiği

Şekil 17 ile DMF içerdiği belirlenen tek maddenin 70 nolu madde olduğu görülmektedir. Madde grafik üzerinde kırmızı ile işaretlenmiştir.

Tablo 74. Amerika ve Afrika için Raju'nun Alan Ölçümleri

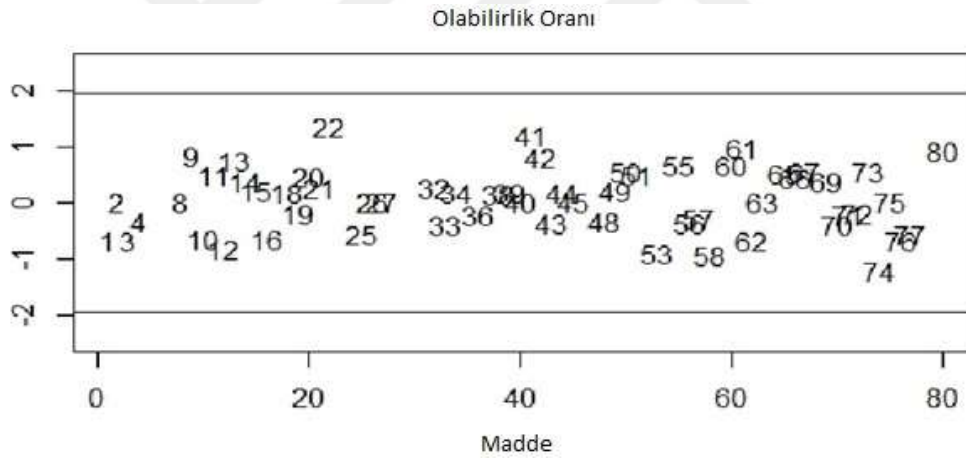
Maddeler	Lord'un χ^2 İstatistiği	p değeri
S70	-2.2756	0.0229

Tablo 74'te Raju'nun Alan Ölçümleri için odak grup Afrika olmak üzere 70 nolu maddenin DMF içerdiği saptanmıştır.



Şekil 18. Amerika ve Afrika için Raju'nun alan ölçümleri

Şekil 18 'de DMF içerdiği belirlenen maddenin konumu gösterilmektedir. Odak grup Afrika iken 70 nolu madde Afrika'dan katılan öğrencilerin aleyhine çalışmaktadır.



Şekil 19. Amerika ve Afrika için olabilirlik oranı

Şekil 19'a göre Amerika ve Afrika için yapılan DMF analizinde Olabilirlik Oranı yönteminin sonucuna göre hiçbir DMF içeren maddeye rastlanmamıştır. Ancak bu durum beklenenin aksi bir durumdur. Eğer bu karşılaştırma için etkili örneklem büyüklüğü sağlansaydı Lord ve Raju yöntemlerinden elde edilen sonuçlar ile Olabilirlik Oranı yöntemi aynı sonuçları veriyor olacaktı. Bu yöntemleri bakımından birbirinden ayıran özellik etkili örneklem büyüklüğü ve test uzunluğudur.

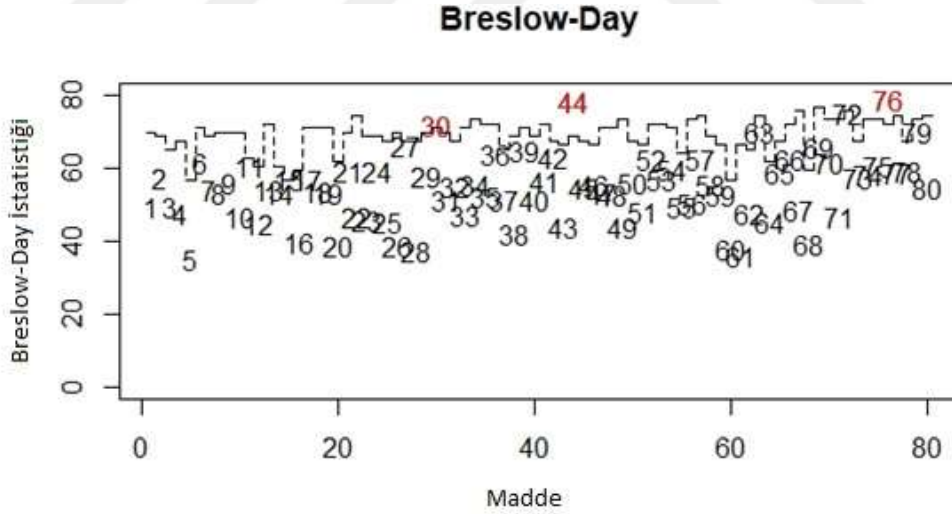
3.2.3.3. Avrupa-Orta Doğu Karşılaştırması

Avrupa-Orta Doğu karşılaştırmasında odak grup Orta Doğu olarak belirlenmiştir. Etkili örneklem büyüklüğünün sağlandığı bu karşılaştırmada analizler, seçilen DMF belirleme yöntemleri kullanılarak yapılmıştır.

Tablo 75. Avrupa ve Orta Doğu için BD yöntemi

Maddeler	Breslow-Day İstatistiği	Serbestlik derecesi	p değeri
S30	71.9010	53	0.043
S44	78.4431	51	0.0081
S76	78.9936	54	0.0149

Tablo 75'te Avrupa ve Orta Doğu için yapılan DMF araştırmasında BD yönteminin sonuçları görülmektedir. Odak grup Orta Doğu olmak üzere, 30, 44 ve 76 kodlu maddelerin Orta Doğu'dan katılan öğrencileri dezavantajlı durumu sokmaktadır.



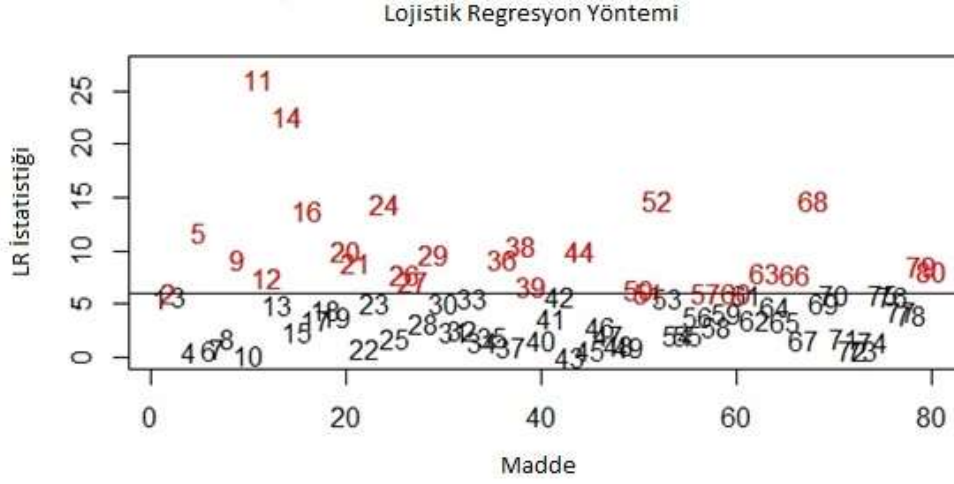
Şekil 20. Avrupa ve Orta Doğu için Breslow-Day yöntemi

Şekil 20'de Tablo 75'ten elde edilmiş sonuçlara göre DMF içerdiği belirlenen maddelerin grafik üzerindeki konumları görülmektedir. Tablo 75'e göre DMF içeren maddeler grafikte kırmızı ile gösterilmiştir.

Tablo 76. Avrupa ve Orta Doğu için lojistik regresyon yöntemi

Maddeler	Lojistik Regresyon	p değeri
S2	6.1275	0.0467
S5	11.7340	0.0028
S9	9.2342	0.0099
S11	26.1931	0.0000
S12	7.4878	0.0237
S14	22.6389	0.0000
S16	13.7462	0.0010
S20	9.9984	0.0067
S21	8.8320	0.0121
S24	14.4012	0.0007
S26	7.8650	0.0196
S27	7.2077	0.0272
S29	9.6991	0.0078
S36	9.2755	0.0097
S38	10.5433	0.0051
S39	6.7076	0.0350
S44	9.9777	0.0068
S50	6.4052	0.0407
S51	6.1254	0.0468
S52	14.7139	0.0006
S57	6.1203	0.0469
S60	6.0759	0.0479
S63	7.9905	0.0184
S66	7.8641	0.0196
S68	14.7154	0.0006
S79	8.6625	0.0132
S80	8.1382	0.0171

Tablo 76’da Avrupa ve Orta Doğu için yapılan DMF araştırmasında LR yöntemine göre DMF içerdiği saptanan maddeler görülmektedir. Odak grup Orta Doğu olmak üzere Tablo 76’da gösterilen maddeler Orta Doğu’dan katılan öğrenciler için DMF içermektedir.



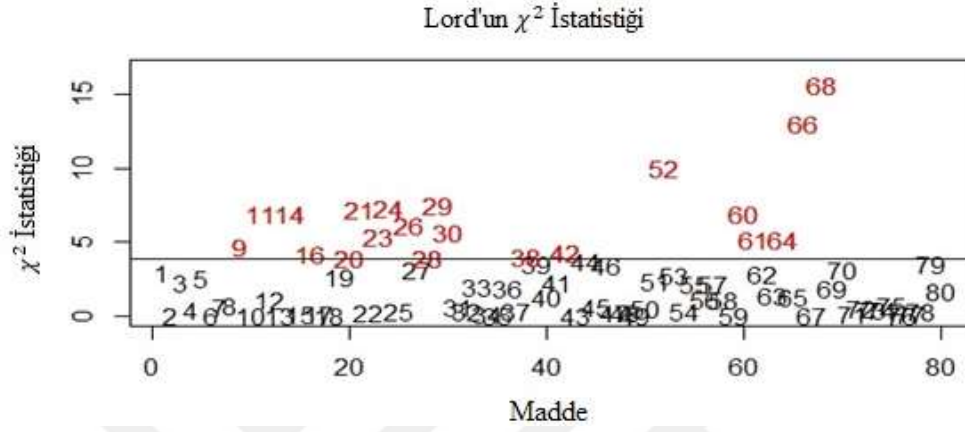
Şekil 21. Avrupa ve Orta Doğu için lojistik regresyon yöntemi

Şekil 21 ile Tablo 76'dan DMF içerdiği saptanan maddelerin grafik konumları görülmektedir. Odak grup Orta Doğu olmak üzere DMF içerdiği belirlenen maddeler kırmızı ile gösterilmektedir.

Tablo 77. Avrupa ve Orta Doğu için Lord χ^2 istatistiği

Maddeler	Lord'un χ^2 İstatistiği	p değeri
S9	4.7281	0.0297
S11	6.9427	0.0084
S14	6.8989	0.0086
S16	4.2414	0.0394
S20	3.9206	0.0477
S21	7.2466	0.0071
S23	5.3353	0.0209
S24	7.3359	0.0068
S26	6.1386	0.0132
S28	0.0490	0.0490
S29	7.5018	0.0062
S30	5.6191	0.0178
S38	3.9681	0.0464
S42	4.2542	0.0392
S52	10.0525	0.0015
S60	6.8841	0.0087
S61	5.1386	0.0234
S64	5.1611	0.0231
S66	13.0298	0.0003
S68	15.6898	0.0001

Tablo 77’de Avrupa ve Orta Doğu için yapılan DMF araştırmasında Lord’un χ^2 İstatistiği yöntemine göre DMF içerdiği saptanan maddeler görülmektedir. Odak grup Orta Doğu olmak üzere Tablo 77 ile gösterilen maddeler Orta Doğu’dan katılan öğrenciler için DMF içermektedir.



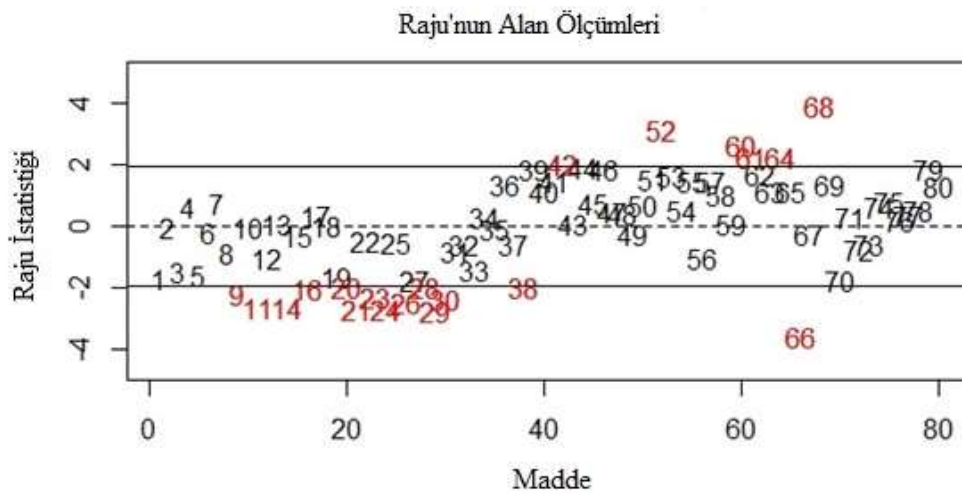
Şekil 22. Avrupa ve Orta Doğu için Lord'un χ^2 istatistiği

Şekil 22’de Tablo 77 ile DMF içerdiği saptanan maddeler grafik üzerinde kırmızı ile gösterilmiştir.

Tablo 78. Avrupa ve Orta Doğu için Raju'nun alan ölçümleri

Maddeler	Raju'nun Alan Ölçümleri	p değeri
S9	-2.1744	0.0297
S11	-2.6349	0.0084
S14	-2.6266	0.0086
S16	-2.0595	0.0394
S20	-1.9801	0.0477
S21	-2.6919	0.0071
S23	-2.3098	0.0209
S24	-2.7085	0.0068
S26	-2.4776	0.0132
S28	-1.9683	0.0490
S29	-2.7389	0.0062
S30	-2.3705	0.0178
S38	-1.9920	0.0464
S42	2.0626	0.0392
S52	3.1706	0.0015
S60	2.6238	0.0087
S61	2.2668	0.0234
S64	2.2718	0.0231
S66	-3.6097	0.0003
S68	3.9610	0.0001

Tablo 78'de Avrupa ve Orta Doğu için yapılan DMF karşılaştırmasında Raju'nun Alan Ölçümleri yöntemi ile elde edilen DMF'li maddeler görülmektedir. Odak grup Orta Doğu olmak üzere Tablo 78 ile verilen maddeler Orta Doğu'dan katılan öğrenciler için DMF içermektedir.



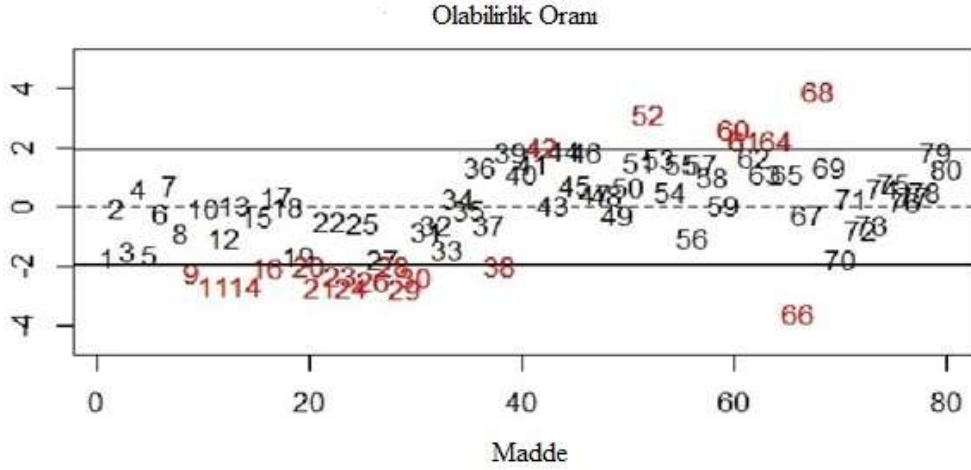
Şekil 23. Avrupa ve Orta Doğu için Raju'nun alan ölçümleri

Şekil 23 ile Tablo 78’de DMF’li olarak belirlenen maddelerin grafik üzerindeki konumları gösterilmektedir. DMF içerdiği saptanan maddeler kırmızı ile işaretlenmiştir.

Tablo 79. Avrupa ve Orta Doğu için olabilirlik oranı

Maddeler	Olabilirlik Oranı	p değeri
S9	390.0000	0.0037
S11	409.0000	0.0009
S14	400.0000	0.0000
S16	428.0000	0.0005
S20	394.0000	0.0015
S21	198.0000	0.0115
S23	191.0000	0.0499
S24	352.0000	0.0002
S26	233.0000	0.0032
S28	285.0000	0.0233
S29	193.0000	0.0031
S30	169.0000	0.044
S38	222.0000	0.0147
S42	485.0000	0.0201
S52	440.0000	0.0004
S60	540.0000	0.0465
S61	508.0000	0.0218
S64	536.0000	0.0194
S66	335.0000	0.0077
S68	557.0000	0.0002

Tablo 79 ile Avrupa ve Orta Doğu için yapılan DMF çalışmasının Olabilirlik Oranı yönteminden elde edilen sonuçları görülmektedir. Odak grup Orta Doğu olmak üzere belirlenen maddeler buradan katılım gösteren adayları dezavantajlı duruma düşürmektedir.



Şekil 24. Avrupa ve Orta Doğu için olabilirlik oranı yöntemi

Şekil 24'te Tablo 79'da DMF içerdiği belirlenen maddeler grafik üzerinde kırmızı ile gösterilmiştir. Eşik değerinin altındaki ve üstündeki maddeler Orta Doğu'dan katılan adayların aleyhine çalışmaktadır.

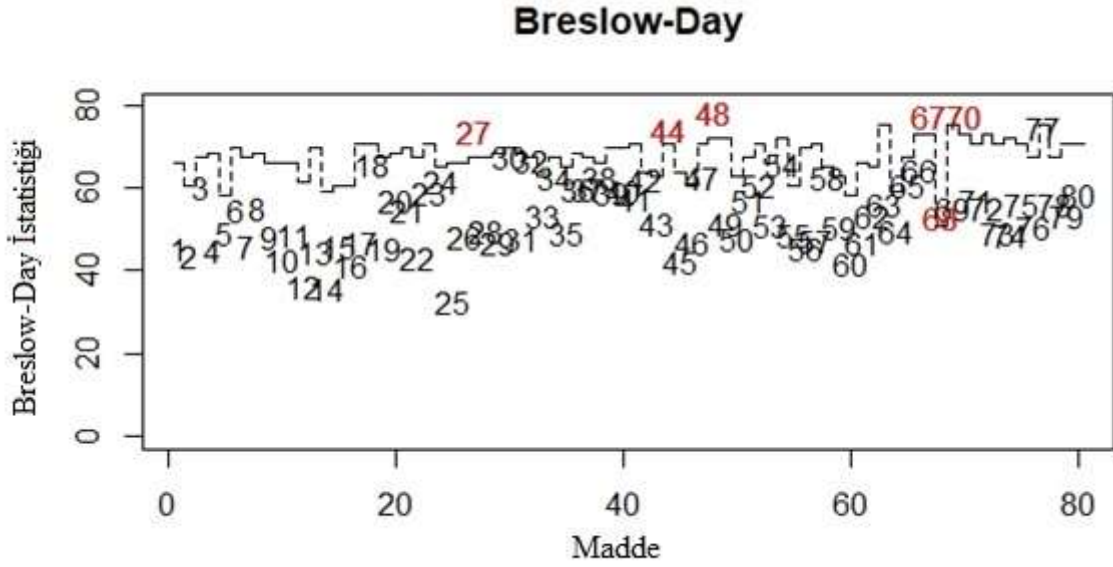
3.2.3.4. Batı Asya-Avrupa Karşılaştırması

Batı Asya ve Avrupa için yapılan karşılaştırmalarda odak grup Batı Asya olarak belirlenmiştir. Etkili örneklem büyüklüğünün sağlandığı bir karşılaştırma olup seçilen yöntemlerden her biri ile DMF sonuçları alınıp karşılaştırılmıştır.

Tablo 80. Batı Asya ve Avrupa için BD yöntemi

Maddeler	Breslow-Day İstatistiği	Serbestlik derecesi	p değeri
S27	74.0248	50	0.0153
S44	74.5205	53	0.0273
S48	78.5969	54	0.0161
S67	77.4989	55	0.0245
S68	52.9204	36	0.0342
S70	77.5930	55	0.0241

Tablo 80'de Batı Asya ve Avrupa için yapılan DMF çalışmasının BD yöntemiyle alınan sonuçları görülmektedir. Verilmiş maddeler, odak grup Batı Asya olmak üzere buradan katılan adayları dezavantajlı konuma düşürmektedir.



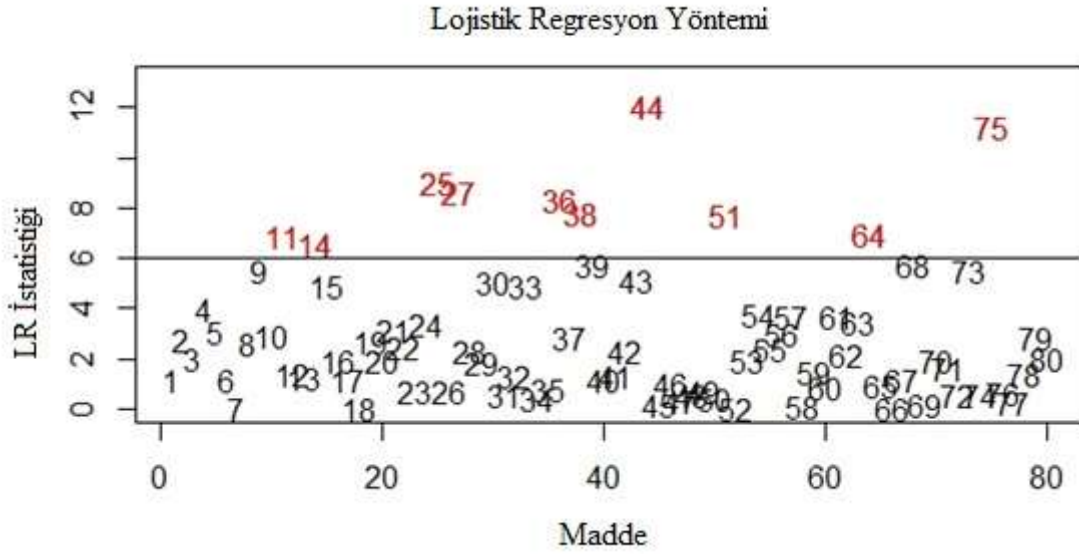
Şekil 25. Batı Asya ve Avrupa için Breslow-Day

Şekil 25'te Tablo 80'de DMF içerdiği belirlenen maddeler grafik üzerinde gösterilmiştir. Batı Asya'dan katılan adayların aleyhine çalışan maddeler kırmızı ile gösterilmektedir.

Tablo 81. Batı Asya ve Avrupa için lojistik regresyon yöntemi

Maddeler	Lojistik Regresyon	p değeri
S11	6.8589	0.0324
S14	6.5978	0.0369
S25	9.0153	0.0110
S27	8.6777	0.0131
S36	8.3325	0.0155
S38	7.8313	0.0199
S44	12.0974	0.0024
S51	7.7335	0.0209
S64	6.9272	0.0313
S75	11.2359	0.0036

Tablo 81'de Batı Asya ve Avrupa için yapılan DMF çalışmasında Lojistik Regresyon yönteminin sonuçları görülmektedir. Odak grup Batı Asya olmak üzere ilgili maddeler buradan katılan adaylar için DMF içermektedir.



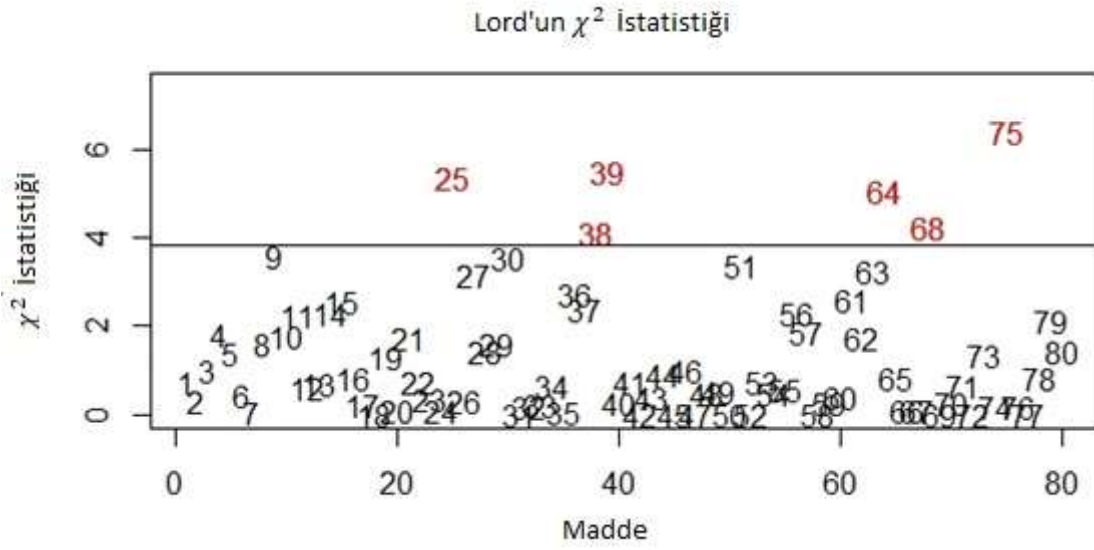
Şekil 26. Batı Asya ve Avrupa için lojistik regresyon yöntemi

Şekil 26’da DMF içerdiği belirlenen Tablo 81 ile verilmiş maddeler grafik üzerinde kırmızı ile gösterilmektedir.

Tablo 82. Batı Asya ve Avrupa için Lord’un χ^2 istatistiği

Maddeler	Lord’un χ^2 İstatistiği	p değeri
S25	5.3412	0.0208
S38	4.1260	0.0422
S39	5.4960	0.0191
S64	5.0512	0.0246
S68	4.2317	0.0397
S75	6.4067	0.0114

Tablo 82’de Batı Asya ve Avrupa için yapılan DMF çalışmasında Lord’un χ^2 İstatistiği yöntemine göre belirlenmiş maddeler görülmektedir. İlgili maddeler Batı Asya’dan katılan adaylar için DMF içermektedir.



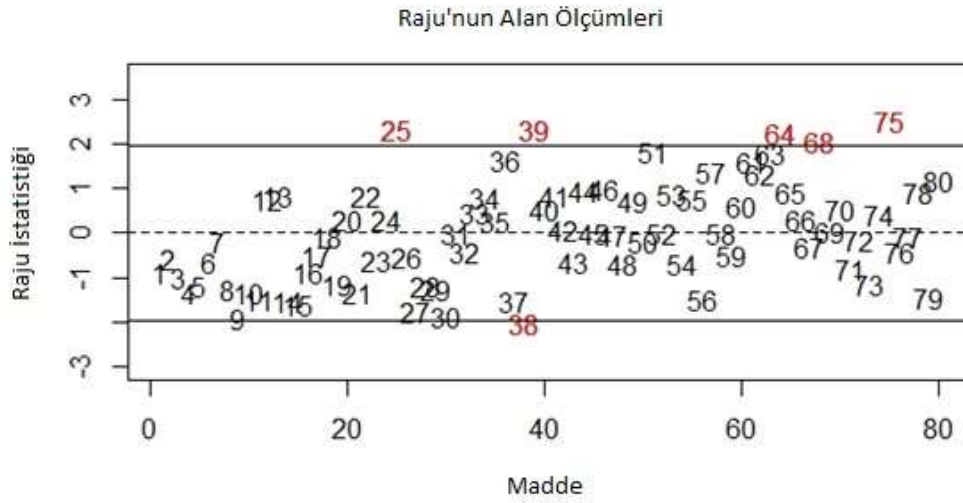
Şekil 27. Batı Asya ve Avrupa için Lord'un χ^2 istatistiği

Şekil 27 ile Tablo 82'de DMF içerdiği saptanan maddeler kırmızı ile gösterilmiştir. Eşik değerinin üzerindeki maddeler Batı Asya'dan katılan adayların aleyhine çalışmaktadırlar.

Tablo 83. Batı Asya ve Avrupa için Raju'nun Alan Ölçümleri

Maddeler	Raju'nun Alan Ölçümleri	p değeri
S25	2.3111	0.0208
S38	-2.0313	0.0422
S39	2.3444	0.0191
S64	2.2475	0.0246
S68	2.0571	0.0397
S75	2.5311	0.0114

Tablo 83'te Batı Asya ve Avrupa için yapılan DMF çalışmasında Raju'nun Alan Ölçümleri yönteminden elde edilen sonuçlar görülmektedir. Verilen maddeler, odak grup Batı Asya'dan katılan adaylar için DMF içermektedir.



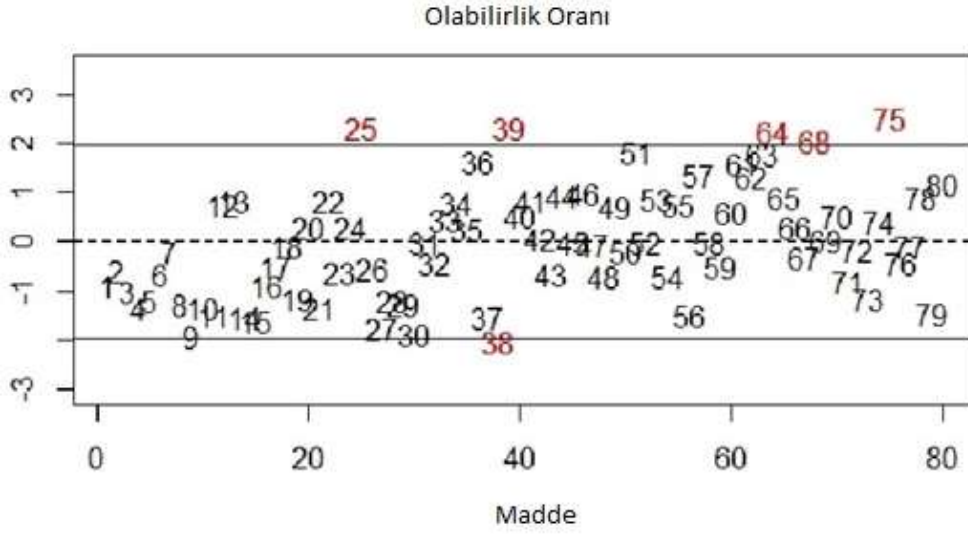
Şekil 28. Batı Asya ve Avrupa için Raju'nun alan ölçümleri

Şekil 28 ile Tablo 83'te verilmiş DMF içeren maddeler grafikte gösterilmiştir. Eşik değerinin altında ve üstünde olan DMF içeren maddeler kırmızı ile işaretlenmiştir.

Tablo 84. Batı Asya ve Avrupa için olabilirlik oranı

Maddeler	Olabilirlik Oranı	p değeri
S25	2.5833	0.0098
S38	-2.0243	0.0429
S39	2.6244	0.0087
S64	1.9749	0.0483
S68	2.6051	0.0092
S75	2.5371	0.0112

Tablo 84 ile Batı Asya ve Avrupa için yapılan DMF çalışmasında Olabilirlik Oranı yönteminden elde edilen sonuçlar görülmektedir. Verilen maddeler odak grup olan Batı Asya'dan katılan adaylar için DMF içermektedir.



Şekil 29. Batı Asya ve Avrupa için olabilirlik oranı

Şekil 29'da Tablo 84'te verilen DMF içerdiği saptanan maddeler grafik üzerinde verilmiştir. Eşik değerinin altında ve üstünde bulunan maddeler DMF içermektedir ve kırmızı ile işaretlenmiştir.

3.2.3.5. Orta Asya-Afrika Karşılaştırması

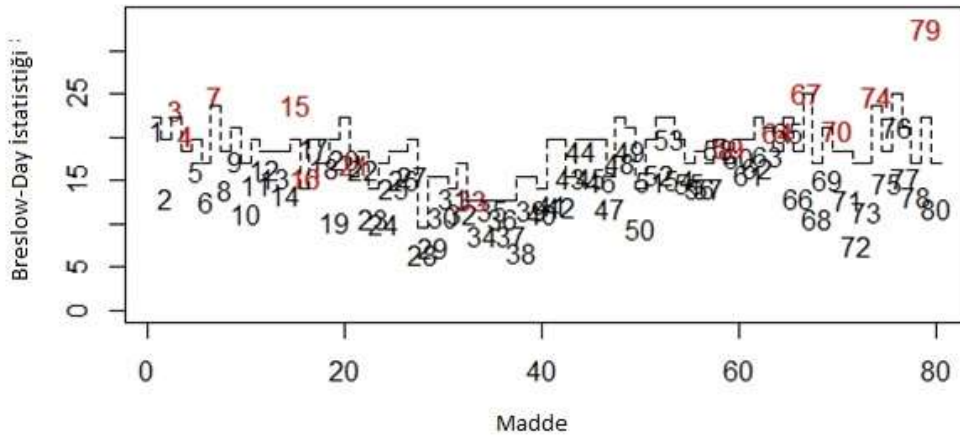
Orta Asya ve Afrika karşılaştırması etkili örneklem büyüklüğünün sağlanamadığı ikinci karşılaştırmadır. Amerika-Afrika karşılaştırmasına kıyasla örnek hacmi daha geniş olduğundan dolayı bu karşılaştırma için KTK'ya dayalı yöntemler çalışmıştır. Ancak, etkili örneklem büyüklüğü sağlanamadığı için ve buna bağlı olarak ideal test uzunluğu yakalanamamış olmasından dolayı alınan sonuçlar beklenenden uzakta kalmıştır. Karşılaştırmada odak grup Afrika olarak belirlenmiştir.

Tablo 85. Orta Asya ve Afrika için BD yöntemi

Maddeler	Breslow-Day İstatistiği	Serbestlik derecesi	p değeri
S3	23.1841	13	0.0395
S4	20.1229	10	0.0281
S7	24.7641	14	0.0370
S15	23.6402	11	0.0143
S16	15.3103	7	0.0322
S21	16.7928	8	0.0323
S33	12.7750	6	0.0468
S59	18.8757	9	0.0263
S64	20.6728	10	0.0235
S67	25.1789	15	0.0476
S70	20.8135	10	0.0224
S74	24.6305	14	0.0384
S79	32.6062	13	0.0020

Tablo 85'te Orta Asya ve Afrika için yapılan DMF çalışmasında BD yönteminin sonuçları görülmektedir. Odak grup Afrika olmak üzere ilgili maddelerin DMF içerdikleri belirlenmiştir.

Breslow-Day



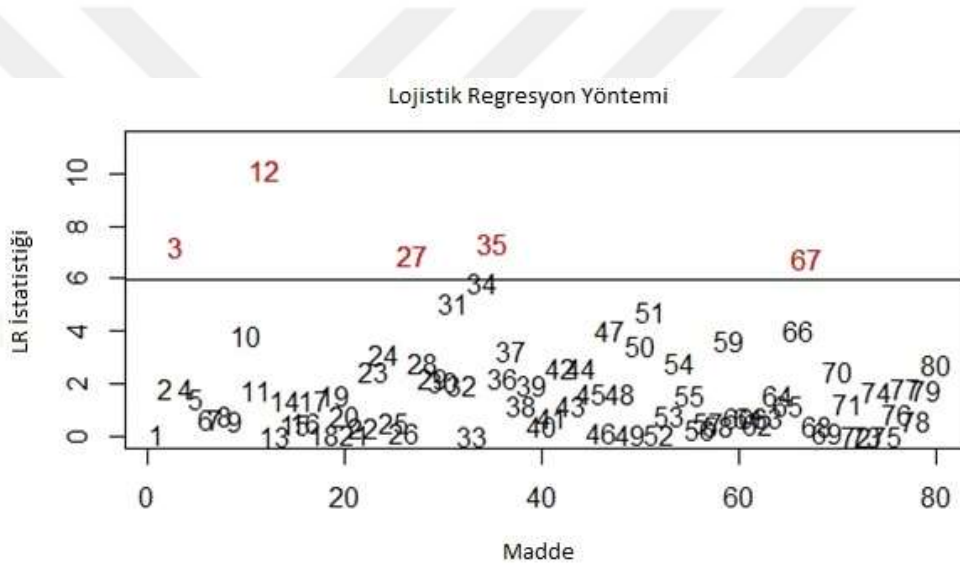
Şekil 30. Orta Asya ve Afrika için Breslow-Day yöntemi

Şekil 30'da Tablo 85 ile DMF içerdiği saptanan maddelerin grafik üzerindeki konumları görülmektedir. DMF içerdiği belirlenen maddeler kırmızı ile gösterilen maddelerdir.

Tablo 86. Orta Asya ve Afrika için lojistik regresyon yöntemi

Maddeler	Lojistik Regresyon	p değeri
S3	7.2589	0.0265
S12	10.1534	0.0062
S27	6.9297	0.0313
S35	7.3379	0.0255
S67	6.7741	0.0338

Tablo 86’da Orta Asya ve Afrika için yapılmış DMF çalışmasında LR yönteminin sonuçları görülmektedir. Odak grup Afrika olmak üzere ilgili maddeler buradan katılan adaylar için DMF içermektedir.



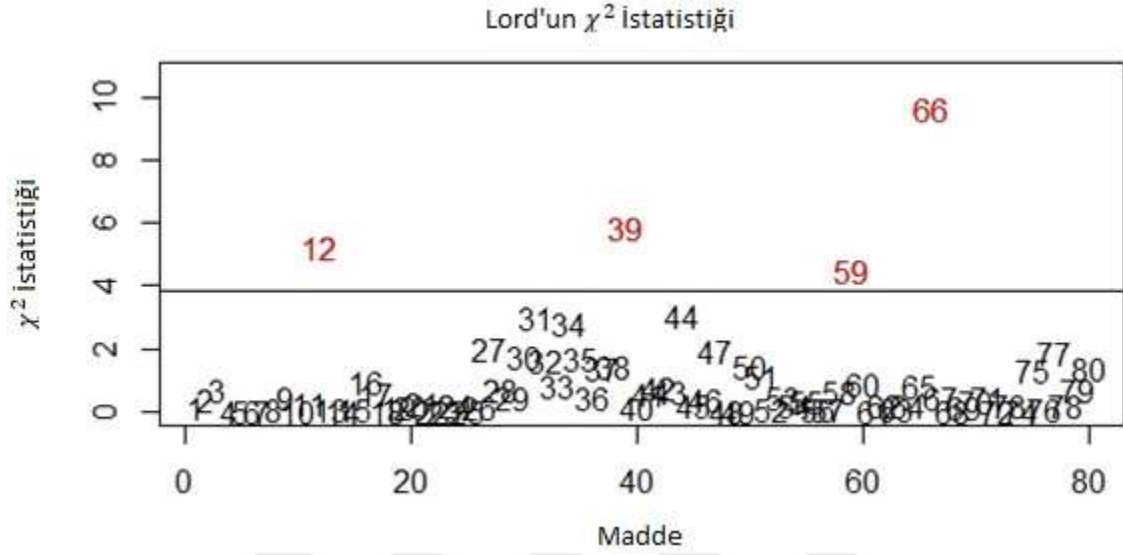
Şekil 31. Orta Asya ve Afrika için lojistik regresyon yöntemi

Şekil 31’de Tablo 86 ile DMF içerdiği belirlenmiş maddeler grafik üzerinde gösterilmiştir. Eşik değerinin üzerinde yer alan DMF’li maddeler kırmızı ile gösterilmektedir.

Tablo 87. Orta Asya ve Afrika için Lord χ^2 istatistiği

Maddeler	Lord’un χ^2 İstatistiği	p değeri
S12	5.2616	0.0218
S39	5.8589	0.0155
S59	4.5089	0.0337
S66	9.6822	0.0019

Tablo 87 ile Orta Asya ve Afrika için yapılan DMF çalışmasında Lord'un χ^2 İstatistiği yönteminin sonuçları görülmektedir. İlgili maddeler odak grup olan Afrika'dan katılan adayları dezavantajlı duruma düşürmektedir.



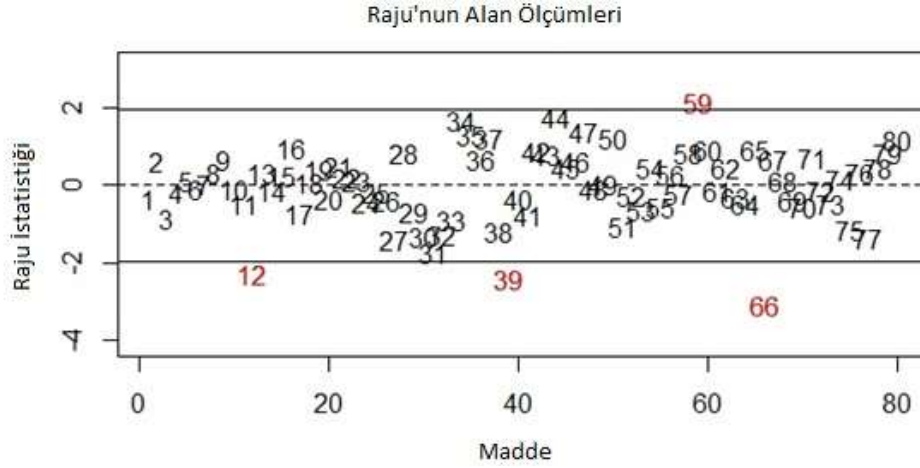
Şekil 32. Orta Asya ve Afrika için Lord'un χ^2 istatistiği

Şekil 32'de Tablo 87 ile belirlenen DMF'li maddelerin grafik konumları verilmiştir. Eşik değerinin üzerindeki DMF'li maddeler kırmızı ile gösterilmektedir.

Tablo 88. Orta Asya ve Afrika için Raju'nun alan ölçümleri

Maddeler	Raju'nun Alan Ölçümleri	p değeri
S12	-2.2938	0.0218
S39	-2.4205	0.0155
S59	2.1234	0.0337
S66	-3.1116	0.0019

Tablo 88 ile Orta Asya ve Afrika için yapılan DMF çalışmasında Raju'nun Alan Ölçümleri yönteminden elde edilen sonuçlar görülmektedir. İlgili maddeler odak grup olan Afrika'dan katılan adaylar için DMF içermektedir.



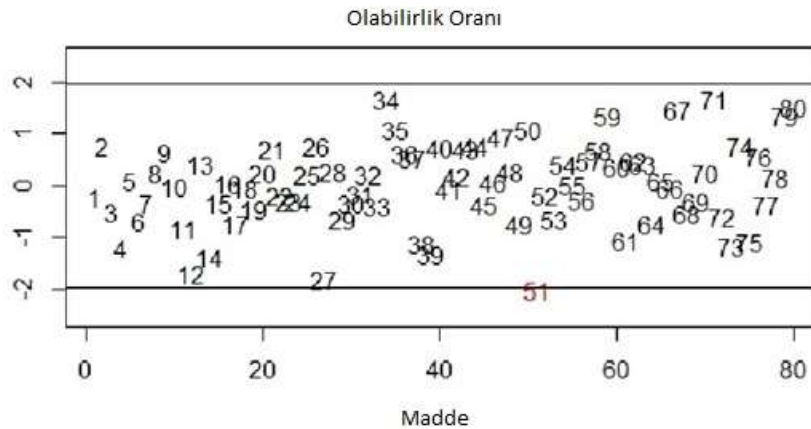
Şekil 33. Orta Asya ve Afrika için Raju'nun alan ölçümleri

Şekil 33 ile Tablo 88'de DMF içerdiği belirlenmiş maddelerin grafik üzerindeki konumları gösterilmektedir. Eşik değerinin altında ve üstünde kalan maddeler DMF'li maddelerdir ve kırmızı ile gösterilmektedir.

Tablo 89. Orta Asya ve Afrika için olabilirlik oranı

Maddeler	Olabilirlik Oranı	p değeri
S51	-2.0161	0.0438

Tablo 89'da Orta Asya ve Afrika için yapılan DMF çalışmasında Olabilirlik Oranı yönteminin sonuçları görülmektedir. Buna göre, 51 kodlu madde odak grup olan Afrika için DMF içermektedir.



Şekil 34. Orta Asya ve Afrika için olabilirlik oranı

Şekil 34'te Tablo 89'da gösterilen DMF içerdiği saptanmış 51 nolu maddenin grafik üzerindeki konumu gösterilmektedir. Eşik değerinin altında kalarak 51 nolu madde DMF'li tek madde olmuştur.

Burada, dikkat edilmesi gereken nokta Lord'un χ^2 İstatistiği ve Raju'nun Alan Ölçümleri yöntemlerinden aynı sonuç alınmasına rağmen Olabilirlik Oranı yönteminin farklı sonuç vermesidir. Bunun nedeni, etkili örneklem büyüklüğünün sağlanamamasıdır. Amerika-Afrika örneğinde olduğu gibi burada da etkili örneklem büyüklüğü sağlanamamış, yöntemler farklı sonuçlar vermiştir. Fakat, iki karşılaştırmayı birbirinden ayıran nokta, Amerika-Afrika örneğinde KTK'ya bağlı yöntemlerin de çalışmamasıdır. Bunun sebebi bu karşılaştırmada örnek sayısının KTK'ya dayalı yöntemlerin de çalışamayacağı kadar küçük olmasıdır. Orta Asya-Afrika karşılaştırmasında ise örnek sayısı KTK yöntemlerinin çalışması için yeterlidir ancak etkili örnek büyüklüğü seviyesinin altındadır. Dolayısıyla, her iki karşılaştırmada MTK'ya dayalı yöntemlerin farklı sonuçlar verdiği görülmektedir. Buradan da MTK yöntemlerinin aynı sonuçları vermesi için etkili örneklem büyüklüğü ve yeterli test uzunluğunun önemi ortaya çıkmaktadır.

4. SONUÇLAR

Geçmişten günümüze birçok disipline ait geliştirilmiş ve uygulanmış testler ile oldukça önemli yargılara varılmış ve insan hayatını etkileyen kararlar verilmiştir. Burada en önemli nokta, bu kararların doğruluğu, hassaslığı ve objektifliğini sağlamaktır. İnsan hayatını bu denli etkileyen bir prosedürün yanlış kararlar vermesi kesinlikle istenmeyen bir durumdur. Herhangi bir disipline ait ölçülmek istenen bir örtük özellik için hazırlanan ölçme aracında bulunan maddelerin ölçeceği alan tüm katılımcılara eşit davranması gerekmektedir. Katılımcıların buldukları bazı özellikler onlara aldıkları testte avantaj ya da dezavantaj sağlamamalıdır. Testin son halinde mevcut tüm maddeler katılımcıların oluşturduğu tüm alt gruplara adil davranmalıdır. Test geliştirme sürecinde en önemli ve belki de en zor süreç bu özellikteki maddeleri oluşturmaktır. Ölçekteki maddelerin yanlılık araştırmasını yapmak için DMF içerip içermedikleri araştırılmalıdır. DMF belirlemek için literatürde KTK ve MTK'ya dayalı pek çok yöntemle rastlamak mümkündür. Bu çalışmada, KTK'ya bağlı yöntemlerden Breslow-Day İstatistiği ve Lojistik Regresyon yöntemi kullanılmış, MTK'ya bağlı yöntemlerden ise Lord'un χ^2 İstatistiği, Raju'nun Alan Ölçümleri ve Olabilirlik Oranı yöntemleri kullanılmıştır.

Bu çalışmada, DMF analizleri iki farklı veri seti üzerinden yapılmıştır. Bunlardan ilki; tarafımızca geliştirilmiş ve uygulanmış İstatistiğe Giriş Okuryazarlık testidir. 25 maddeden oluşan bu testte her bir madde için 4 seçenek mevcuttur. Analizlerin yapılabilmesi için katılımcılardan alınan tepkiler ikili şekilde kodlanmıştır. Bu test, KTÜ ve OMÜ'deki İstatistik dersi almış öğrencilere uygulanmıştır. Toplamda 1455 öğrencinin katılım gösterdiği testte bu öğrencilerden 730'u kız, 725'i erkek, 1044'ü KTÜ öğrencisi ve 411'i OMÜ öğrencisidir. Testin kullanılabilirliği güvenilirlik ve geçerlilik çalışmalarıyla ortaya konulmuştur. Buna göre; tüm test için güvenilirlik 0.9257 olarak hesaplanmıştır. Buradan, hazırlanan testin yüksek güvenilirlikte olduğu sonucuna varılmıştır. Uyum kriterleri ile model karşılaştırması yapılmış verilerden en fazla desteği MTK ailesinden Rasch model almıştır. MTK ile çalışabilmek için varsayımlarının sağlanması gerekliliğinden veriye tek boyutluluk testi yapılmış ve tek boyutluluk kriteri 0.9084 olarak elde edilmiştir. Böylelikle veri setinin tek boyutlu olduğu sonucuna varılmıştır. Buradan hareketle, yerel bağımsızlık varsayımının da sağlandığı gözlenmiştir. Analizlere başlamadan tek ve çok değişkenli aykırı değer analizi yapılmıştır. Tek değişkenli aykırı

değer analizi için z puanları, çok değişkenli aykırı değer analizi için Mahalanobis uzaklıkları hesaplanmış ve çok değişkenli aykırı değere rastlanmazken 5 adet tek değişkenli aykırı değer bulunmuş bu katılımcılar analizin dışında bırakılmıştır. Bunun ardından, Total Corrected-Item Correlation analizi ile ölçekten atılması gereken madde olup olmadığı araştırılmış ve ölçekten atılması gereken madde olmadığı sonucuna ulaşılmıştır. Tüm bu bilgiler ışığında madde analizleri başlamış seçilen model Rasch model olduğundan ve madde parametresi olarak sadece güçlük parametresi barındırdığından Tablo 10' da tüm maddeler için parametre tahminleri elde edilmiştir.

Tablo 90. Cinsiyet için alınan sonuçların karşılaştırılması

BD	LR	LORD	RAJU	MTK-OO
5	1, 3, 13, 16, 17, 21, 23	1, 3, 13, 16, 17, 21	1, 3, 13, 16, 17, 21	1, 3, 13, 16, 17, 21

Tablo 90 ile İstatistiğe Giriş Okuryazarlık testinde cinsiyet açısından yapılan DMF çalışmasının karşılaştırmalı sonuçları görülmektedir. Test için etkili örneklem büyüklüğü ve yeterli test uzunluğu sağlanmış olup alınan sonuçlar buna göre değerlendirilmiştir. Burada odak grup kadınlar olarak belirlenmiştir. DMF içerdiği saptanan maddeler, kadınların erkeklere göre bu maddeler üzerinde dezavantajlı olmalarına neden olmaktadır. Tablo 90' da görüldüğü gibi BD ve LR yöntemleri KTK'ya dayalı yöntemler olmalarına rağmen birbirlerinden çok farklı ve tutarsız sonuçlar vermiştir. Bunun yanında, etkili örneklem büyüklüğü ve yeterli test uzunluğu sağlandığında MTK'ya dayalı tüm yöntemlerin birebir aynı sonuçlar verdiği saptanmıştır.

Tablo 91. Üniversiteler için alınan sonuçların karşılaştırılması

BD	LR	LORD	RAJU	MTK-OO
6, 25	1, 3, 7, 9, 14, 19, 20	1, 3, 7, 9, 14, 17, 20, 23	1, 3, 7, 9, 14, 17, 20, 23	1, 3, 7, 9, 14, 17, 20, 23

Tablo 91 ile İstatistiğe Giriş Okuryazarlık testinde üniversiteler açısından yapılan DMF çalışmasının karşılaştırmalı sonuçları görülmektedir. Test için etkili örneklem büyüklüğü ve yeterli test uzunluğu sağlanmış olup alınan sonuçlar buna göre değerlendirilmiştir. Bu analiz için odak grup olarak KTÜ öğrencileri seçilmiştir. Tablo 91

ile gösterilen maddeler KTÜ'den katılan öğrencileri OMÜ öğrencilerine göre dezavantajlı konuma düşürmektedir. KTK yöntemlerinden BD ve LR yöntemleri yine birbirlerinden aykırı ve tutarsız sonuçlar verirken istenen özelliklerin sağlanması durumunda MTK yöntemleri birebir aynı sonuçları vermişlerdir.

İkinci veri seti KTÜ'nün hazırlayıp uyguladığı 2019 yılı YÖS'dür. Ölçek 80 maddeden oluşmaktadır. Her bir madde 5 seçenekli olup yanlışın doğruyu götürmediği bir sistemde değerlendirme yapılmaktadır. 2019 yılında 4082 kişi KTÜ'de öğrenim hakkı kazanabilmek için bu sınava başvurmuştur. Bunlardan birinciliği 2054 katılımcı ile İran alırken, Türkiye 259 katılımcı ile 4. sıradadır. 2019 yılında bu sınava giren katılımcıların tepkilerinden yola çıkılarak maddelerin DMF araştırması yapılmıştır. Ancak, analizlere başlamadan önce MTK varsayımları incelenmiştir ve bunun için öncelikle tek boyutluluk testi yapılmıştır. Hesaplamalar sonucu tek boyutluluk kriteri 0.9968 olarak elde edilmiştir. Bunun bir sonucu olarak ölçeğin tek boyutlu olduğu sonucuna varılmış ve tüm MTK varsayımlarının sağlandığı gözlemlenmiştir. Ardından hangi MTK modeli ile çalışılacağına karar vermek üzere uyum kriterleri ile model karşılaştırılması yapılmış en iyi modelin Rasch model olduğu sonucuna varılmıştır. Maddelerin her biri ve tüm test için güvenilirlik çalışması yapılmış ve tüm testin güvenilirliği 0.9315 olarak hesaplanmıştır. Buradan, ölçeğin yüksek güvenilir bir test olduğu sonucu çıkmaktadır. Aykırı değer olup olmadığı araştırılmış tek değişkenli aykırı değer analizi için z puanları, çok değişkenli aykırı değer analizi için Mahalanobis uzaklıkları hesaplanmış ve 2 tek değişkenli, 23 çok değişkenli aykırı değer bulunup analizden bu katılımcılar çıkartılmıştır. Ölçekten atılması gereken madde olup olmadığının saptanması için Item Corrected-Total Correlation değerleri incelenmiş ve ölçekten atılması gereken madde olmadığı sonucuna ulaşılmıştır. Tüm bu bilgiler ışığında DMF analizleri KTK ve MTK bazlı yöntemlerle yapılmış ve sonuçlar karşılaştırılıp yorumlanmıştır. DMF analizleri yapılırken cinsiyet açısından, sosyokültürel açıdan ve bölgeler açısından sonuçlar alınmıştır. Sosyokültürel açıdan DMF analizi yapılırken en fazla katılım gösteren 5 ülke seçilmiş bu ülkelerden bazı karşılaştırmalar yapılmıştır. Seçilen ülkeler; Türkiye, Azerbaycan, Almanya, İran, Hollanda ve Suriye'dir. İlgili ülkeler hem katılımcı sayısı dikkate alınarak hem de sosyokültürel açıdan birbirlerinden farklılık göstermeleri sebebiyle seçilmiştir. Bölgeler açısından yapılan DMF analizinde mantık hem yakın kültürlerdeki ülkeleri bir araya getirmek hem de coğrafi konumları dikkate almaktır. Bu bağlamda seçilen bölgeler; Amerika, Afrika, Orta Asya, Avrupa, Orta Doğu, Batı Asya, Güney Asya'dır. Afrika'dan katılan ülkeler, Cezayir,

Kamerun, Çad, Cibuti, Fas, Moritanya, Somali, Sudan, Tunus ve Uganda'dır. Orta Asya'dan katılan ülkeler, Kazakistan, Kırgızistan, Moğolistan, Tacikistan, Türkmenistan ve Özbekistan'dır. Avrupa'dan katılımcı ülkeler, Birleşik Krallık, Ukrayna, İsviçre, İsveç, İspanya, Sırbistan, Romanya, Norveç, Hollanda, Moldova, Makedonya, Kosova, İtalya, Yunanistan, Almanya, Fransa, Danimarka, Bulgaristan, Belçika ve Avusturya'dır. Orta Doğu katılımcıları; Irak, İsrail, Ürdün, Lübnan, Libya, Filistin, Suudi Arabistan, Suriye ve Yemen'dir. Güney Asya'dan katılan ülkeler, Hindistan, Afganistan, İran ve Myanmar'dır. Batı Asya'dan katılan ülkeler, Azerbaycan, Gürcistan, Rusya ve Türkiye'dir. Amerika bir bütün olarak ele alınmıştır.

Tablo 92. Türkiye-İran cinsiyet açısından karşılaştırma

BD	LR	LORD	RAJU	MTK-OO
9, 11, 45, 57, 59, 64, 69, 72	9, 21, 23, 24, 47, 48, 62, 65, 70, 76	9, 23, 47, 48, 61, 62, 65, 70, 76, 80	9, 23, 47, 48, 61, 62, 65, 70, 76, 80	9, 23, 47, 48, 61, 62, 65, 70, 76, 80

Tablo 92'de Türkiye ile İran arasında cinsiyet açısından DMF içeren madde olup olmadığına ilişkin analiz sonuçları görülmektedir. Bu karşılaştırma için etkili örneklem büyüklüğü ve yeterli test uzunluğu sağlanmıştır. Buna göre; KTK'ya dayalı yöntemler incelendiğinde farklı maddelerde DMF saptanırken, MTK bazlı yöntemlerin tamamında birebir aynı sonuçlar alınmıştır.

Tablo 93. Türkiye-Azerbaycan cinsiyet açısından karşılaştırma

BD	LR	LORD	RAJU	MTK-OO
1, 3, 11, 32, 34, 41, 58, 59, 65	5, 10, 13, 16, 24, 48, 49, 52	13, 16, 24, 41, 48, 49, 52	13, 16, 24, 41, 48, 49, 52	13, 16, 24, 41, 48, 49, 52

Tablo 93 ile cinsiyet açısından yapılan DMF araştırmasında Türkiye ile Azerbaycan karşılaştırılmış ve DMF saptanan maddeler belirlenmiştir. Bu karşılaştırma etkili örnek büyüklüğü ve yeterli test uzunluğunun sağlandığı bir karşılaştırmadır. Bu özellikleri sağlayan karşılaştırma sonuçları ile KTK tabanlı yöntemlerin birbirinden tutarsız sonuçlar verdiği, bunun aksine MTK tabanlı yöntemlerin birebir örtüşen sonuçlar verdiği söylenebilir.

Tablo 94. Almanya-Suriye cinsiyet açısından karşılaştırma

BD	LR	LORD	RAJU	MTK-OO
20, 27, 31, 34, 43, 50, 55, 59, 66, 70	14, 16, 21, 60, 70	14, 21, 31, 60, 70	14, 21, 31, 60, 70	14, 21, 31, 60, 70

Tablo 94'te Almanya ile Suriye arasında yapılan cinsiyet açısından karşılaştırma ile yöntemlere göre yapılan DMF analizi sonuçları görülmektedir. Bu karşılaştırmada etkili örnek büyüklüğü ve yeterli test uzunluğu özellikleri sağlanmıştır. Bu özelliklerin sağlanması ile MTK yöntemlerinin tutarlı sonuçlar vermesinin yanında KTK yöntemlerinin farklı maddeleri DMF'li olarak işaretlediği görülmektedir.

Tablo 95. Hollanda-İran cinsiyet açısından karşılaştırma

BD	LR	LORD	RAJU	MTK-OO
9, 11, 37, 57, 64, 69	9, 10, 21, 23, 57, 62, 65, 70, 76	10, 23, 62, 65, 70	10, 23, 62, 65, 70	10, 23, 62, 65, 70

Tablo 95 ile Hollanda ve İran arasında cinsiyet karşılaştırması sonucu saptanan DMF'li maddeler görülmektedir. Etkili örnek büyüklüğü ve yeterli test uzunluğunun sağlandığı bu karşılaştırmadan MTK yöntemlerinin sonuçlarının tutarlı olduğu görülmektedir.

Tablo 96. Türkiye-İran Uyrak açısından karşılaştırma

BD	LR	LORD	RAJU	MTK-OO
14, 19, 28, 33, 40, 48, 50, 52, 54, 58, 77	1, 2, 5, 6, 14, 18, 20, 21, 31, 35, 37, 39, 40, 44, 49, 51, 52, 58, 61, 62, 67, 70, 72, 76, 77, 79	2, 3, 5, 6, 11, 12, 13, 14, 15, 16, 17, 18, 20, 21, 24, 25, 35, 37, 40, 44, 51, 52, 57, 58, 62, 63, 67, 69, 70, 71, 72, 73, 76, 77, 79, 80	2, 3, 5, 6, 11, 12, 13, 14, 15, 16, 17, 18, 20, 21, 24, 25, 35, 37, 40, 44, 51, 52, 57, 58, 62, 63, 67, 69, 70, 71, 72, 73, 76, 77, 79, 80	2, 3, 5, 6, 11, 12, 13, 14, 15, 16, 17, 18, 20, 21, 24, 25, 35, 37, 40, 44, 51, 52, 57, 58, 62, 63, 67, 69, 70, 71, 72, 73, 76, 77, 79, 80

Tablo 96 ile Türkiye İran arasında uyrak açısından DMF çalışmasının sonuçları verilmiştir. Bu karşılaştırma da etkili örnek büyüklüğüne ulaşmış ve yeterli test uzunluğunun sağlanmış olduğu bir karşılaştırmadır. Buna dayanarak sadece MTK bazı yöntemlerin tutarlı sonuçlar verdiği görülmektedir.

Tablo 97. Türkiye-Azerbaycan uyruk açısından karşılaştırma

BD	LR	LORD	RAJU	MTK-OO
1, 19, 26, 32, 33, 40, 42, 51, 52, 53, 54, 58, 68, 77	5, 29, 31, 35, 40, 51, 62, 67, 72	2, 5, 15, 17, 35, 37, 51, 58, 62, 63, 66, 67, 69, 70, 72, 77, 80	2, 5, 15, 17, 35, 37, 51, 58, 62, 63, 66, 67, 69, 70, 72, 77, 80	2, 5, 15, 17, 35, 37, 51, 58, 62, 63, 66, 67, 69, 70, 72, 77, 80

Tablo 97 ile uyruk açısından yapılan DMF araştırmasında Türkiye ile Azerbaycan karşılaştırılmış DMF içerdiği saptanan maddeler gösterilmiştir. Karşılaştırma, etkili örnek büyüklüğü ve yeterli test uzunluğu bakımından istenen özelliklere sahiptir. Sağlanan özellikler ile MTK'ya bağlı yöntemlerin KTK'ya bağlı yöntemlerin aksine tutarlı sonuçlar verdiği görülmektedir.

Tablo 98. Almanya-Suriye uyruk açısından karşılaştırma

BD	LR	LORD	RAJU	MTK-OO
13, 14, 21, 24, 31, 35, 39, 57, 60	39, 51, 59	9, 66	9, 66	9, 66

Tablo 98'de uyruk açısından Almanya ile Suriye karşılaştırmasından elde edilen sonuçlar görülmektedir. Örnek boyutunun ve test uzunluğunun yeterli olduğu bu karşılaştırmada MTK bazlı yöntemlerin birebir aynı sonuçları verdiği görülmektedir.

Tablo 99. Hollanda-İran uyruk açısından karşılaştırma

BD	LR	LORD	RAJU	MTK-OO
10, 18, 21, 24, 48, 54, 74	2, 3, 5, 10, 11, 16, 18, 22, 28, 32, 44, 47, 51, 54, 66, 68, 77	2, 3, 4, 5, 10, 11, 14, 15, 16, 41, 44, 47, 51, 63, 69, 70, 75, 80	2, 3, 4, 5, 10, 11, 14, 15, 16, 41, 44, 47, 51, 63, 69, 70, 75, 80	2, 3, 4, 5, 10, 11, 14, 15, 16, 41, 44, 47, 51, 63, 69, 70, 75, 80

Tablo 99'da uyruk açısından yapılan son karşılaştırma Hollanda ve İran için DMF'li maddeler gösterilmektedir. Örnek büyüklüğü ve test uzunluğunu n yeterli olduğu belirlenen bu karşılaştırmada MTK bazlı sonuçlar KTK'nın aksine birbirleri ile tutarlıdır.

Tablo 100. Güney Asya-Avrupa karşılaştırması

BD	LR	LORD	RAJU	MTK-OO
8, 15, 44, 45, 48, 68, 70, 76	1, 2, 3, 4, 5, 8, 9, 11, 12, 14, 15, 15, 16, 18, 19, 20, 21, 24, 27, 28, 30, 31, 35, 37, 38, 39, 41, 42, 44, 46, 48, 51, 52, 57, 58, 60, 61, 62, 63, 64, 66, 68, 69, 70, 71, 72, 75, 76, 77, 79, 80	2, 3, 4, 5, 6, 8, 9, 11, 12, 14, 15, 16, 17, 18, 19, 20, 21, 24, 27, 28, 30, 31, 35, 37, 38, 41, 42, 44, 51, 52, 53, 57, 58, 60, 61, 62, 63, 68, 69, 70, 71, 72, 75, 76, 77, 79, 80	2, 3, 4, 5, 6, 8, 9, 11, 12, 14, 15, 16, 17, 18, 19, 20, 21, 24, 27, 28, 30, 31, 35, 37, 38, 41, 42, 44, 51, 52, 53, 57, 58, 60, 61, 62, 63, 68, 69, 70, 71, 72, 75, 76, 77, 79, 80	2, 3, 4, 5, 6, 8, 9, 11, 12, 14, 15, 16, 17, 18, 19, 20, 21, 24, 27, 28, 30, 31, 35, 37, 38, 41, 42, 44, 51, 52, 53, 57, 58, 60, 61, 62, 63, 68, 69, 70, 71, 72, 75, 76, 77, 79, 80

Tablo 100 ile bölgeler açısından yapılan karşılaştırmalardan ilki Güney Asya-Avrupa karşılaştırması için sonuçlar verilmiştir. Örnek ve test uzunluğunun yeterli olduğu bu karşılaştırmada BD ile LR yöntemlerinin tutarsız sonuçlar verdiği bunun aksine Lord, Raju ve Olabilirlik Oranı yöntemlerinin birebir aynı sonuçlar verdiği görülmektedir.

Tablo 101. Amerika-Afrika karşılaştırması

BD	LR	LORD	RAJU	MTK-OO
Hesaplanamadı	Hesaplanamadı	70	70	DMF içeren madde yok

Tablo 101 ile etkili örneklem büyüklüğü özelliğini sağlamamış Amerika ve Afrika karşılaştırması için DMF analizinin sonuçları görülmektedir. Örnek büyüklüğü oldukça küçük olduğundan KTK'ya dayalı yöntemler sonuç vermemiştir. Daha güçlü oldukları bilinen MTK yöntemleri ise Lord ve Raju aynı maddeyi işaret ederken Olabilirlik Oranı yöntemi DMF içeren madde olmadığı sonucunu vermiştir. Tüm durumlarda Lord ve Raju aynı sonucu verecektir (Hasançebi vd., 2019). Burada en önemli nokta Olabilirlik Oranı yönteminin Lord ve Raju ile tutarlı olabilmesi için etkili örnek büyüklüğü ve yeterli test uzunluğu özelliklerinin sağlanmış olmasıdır.

Tablo 102. Avrupa-Orta Doğu karşılaştırması

BD	LR	LORD	RAJU	MTK-OO
30, 44, 76	2, 5, 9, 11, 12, 14, 16, 20, 21, 24, 26, 27, 29, 36, 38, 39, 44, 50, 51, 52, 57, 60, 63, 66, 68, 79, 80	5, 11, 14, 16, 20, 21, 23, 24, 26, 28, 29, 30, 38, 42, 52, 60, 61, 64, 66, 68	5, 11, 14, 16, 20, 21, 23, 24, 26, 28, 29, 30, 38, 42, 52, 60, 61, 64, 66, 68	5, 11, 14, 16, 20, 21, 23, 24, 26, 28, 29, 30, 38, 42, 52, 60, 61, 64, 66, 68

Tablo 102 ile Avrupa Orta Doğu karşılaştırmasının sonuçları görülmektedir. Beklendiği gibi istenen özellikleri sağlayan bir karşılaştırma olması sebebiyle Olabilirlik Oranı da Lord ve Raju ile aynı sonuçları vermişken BD ve LR yöntemleri beklendiği gibi tutarsız sonuçlar vermiştir.

Tablo 103. Batı Asya- Avrupa karşılaştırması

BD	LR	LORD	RAJU	MTK-OO
27, 44, 48, 67, 68, 70	11, 14, 25, 27, 36, 38, 44, 51, 64, 75	25, 38, 39, 64, 68, 75	25, 38, 39, 64, 68, 75	25, 38, 39, 64, 68, 75

Tablo 103 ile Batı Asya ve Avrupa karşılaştırılmış istenen özelliklerin sağlanmasıyla tüm MTK yöntemlerinin aynı sonuçları verdiği görülmüştür.

Tablo 104. Orta Asya-Afrika karşılaştırması

BD	LR	LORD	RAJU	MTK-OO
3, 4, 15, 16, 21, 33, 59, 64, 67, 70, 74, 79	3, 12, 27, 35, 67	12, 39, 59, 66	12, 39, 59, 66	51

Tablo 104 ile Orta Asya ve Afrika karşılaştırmasının sonuçları verilmiştir. Bu karşılaştırma etkili örnek büyüklüğünün sağlanmadığı bir karşılaştırmadır. Amerika ve Afrika karşılaştırmasından farkı ise KTK bazlı yöntemlerin bu karşılaştırmada çalışmış olmasıdır. Bunun sebebi, Amerika-Afrika karşılaştırmasına göre örnek büyüklüğünün daha fazla olmasıdır. KTK bazlı yöntemler çalışmış ancak yine birbirleriyle örtüşmeyen sonuçlar vermiştir. Bunun yanında, Lord ve Raju yine aynı sonuçları vermelerine rağmen etkili örnek büyüklüğü sağlanmadığından Olabilirlik Oranı yöntemi diğer MTK yöntemlerinden farklı sonuç vermiştir.

Bu çalışmada, bir ölçek için madde yanlılığının önemi vurgulanmıştır. Uygulamada hatasız ölçüğün mümkün olmadığı ancak hataların minimum olması gerektiği belirtilmiştir. Maddede ölçülmesi beklenen örtük özellik açısından bir kesime avantaj ya da dezavantaj sağlama durumu özellikle de geniş kapsamlı testler için en önemli sorundur. Bu sorunların belirlenmesinde ve giderilmesinde DMF çalışmalarının önemi yadsınamayacak ölçüde büyüktür. Ayrıca, ölçekteki maddelerde DMF araştırması yapılırken sonuçları etkileyen faktörlerin mevcudiyeti ihmal edilemeyecek kadar önemlidir. Bunun sebebi, bu özelliklerin

doğrudan DMF sonuçlarına etki ettiğinin gözlenmesidir. Çalışmada, KTK ve MTK'ya dayalı seçilen yöntemlerden sonuçlar alınmış, alınan sonuçlar birbirleriyle karşılaştırılmıştır. Bu karşılaştırmalar ile yöntemlerin güçlerinin yanı sıra etkili örnek büyüklüğü ve yeterli test uzunluğunun sağlanmasının önemi de görülmektedir. MTK bazlı yöntemlerin analizlerde tutarlı sonuçlar vermeleri itibariyle daha güçlü yöntemler olduklarını söylemek mümkündür. Ancak, güçleri ölçeğin madde sayısına ve ölçeği alan katılımcı sayısına bağlıdır. Yeterli madde sayısına sahip bir ölçeğin örtük özelliği ölçmedeki gücü daha fazladır. Ayrıca, ölçek yeterli sayıda katılımcıya uygulandığında kişilerden alınan tepkiler ile elde edilen bilgilerin doğruluğu artacaktır. Bu özellikler sağlandığında yöntemlerin elde ettikleri güç, yapılan analizlerden ortaya çıkan sonuçların doğruluğu ile orantılıdır. Yöntemlerin karşılaştırmaları sonucu Tablo 94'ten Tablo 104'e kadar görülmektedir. Buradan hareketle, KTK yöntemlerinin örneklem büyüklüğü ve test uzunluğunun sağlandığı durumlarda dahi birbirlerinden tutarsız sonuçlar verdiğini bunun aksine, MTK yöntemlerinin örnek büyüklüğü ve test uzunluğu özelliklerinin sağlanması durumunda birebir aynı ve tutarlı sonuçlar verdikleri görülmektedir. Örnek büyüklüğü ya da test uzunluğu özelliklerinden bir veya ikisinin birden bozulması durumunda ise Lord ve Raju yöntemleri her analizde aynı sonuçlar vermesine rağmen Olabilirlik Oranı yöntemi aykırı sonuçlar vermiştir. Yapılan bu analizlerin sağlanması olması anlamında karşılaştırmalarda DMF saptanan maddeler ölçekten atılarak tekrar DMF analizleri yapılmıştır. Alınan sonuçlara bu maddelerin çıkarılmasıyla ölçeklerin çoğu tamamen temizlenmiştir. Bazı karşılaştırmalarda ise DMF saptanan maddeler çıkartıldıktan sonra da farklı birkaç maddede DMF belirlendiği gözlemlenmiştir. Bu durumda ise belirlenen bu DMF'lerin etki büyüklükleri araştırılmıştır. Etki büyüklüklerine göre DMF ihmal edilebilir düzey, orta düzey ve yüksek düzey DMF olarak kategorilendirilebilir. Saptanan tüm maddeler bu kategorilerden ihmal edilebilir düzeyde çıkmıştır. Dolayısıyla, asıl DMF içeren maddelerin ölçekten çıkarılmasıyla ölçekler temizlenmiştir. Bu durum yapılan analizlerin doğruluğunu kanıtlar niteliktedir.

5. ÖNERİLER

Sunulan tezde, DMF belirlenmesi için kullanılan KTK ve MTK temelli toplam beş yöntem kullanılmıştır. Bu yöntemlerden ikisi KTK temelli BD ve LR iken, diğerleri MTK temelli Lord'un χ^2 İstatistiği, Raju'nun Alan Ölçümleri ve MTK Olabilirlik Oranı yöntemleridir.

Çalışmada, İstatistiğe Giriş Okuryazarlık Testi ve YÖS-2019 için en iyi modelin Rasch model olduğu belirlenmiş ve ele alınan yöntemler ile DMF araştırması yapılmıştır. Elde edilen sonuçlar ışığında MTK bazlı yöntemlerin gücü ortaya konmuştur. Tepki vektöründen en yüksek desteği KTK modelinin aldığı durumda yöntemlerin vereceği sonuçlar araştırılmalıdır. Yöntemlerin gücünün model bazlı olup olmadığı bu anlamda ortaya konabilir. Bireyler hakkında önemli yargılara varılan ÖSS, Kamu Personeli Seçme Sınavı (KPSS), Yabancı Dil bilgisi Seviye Tespit Sınavı (YDS), Akademik Personel Lisansüstü Eğitim Sınavı (ALES), Lise Geçiş Sınavı (LGS) gibi geniş çaplı sınavlar için tezde yapılmış çalışmaların devamlılığı sağlanmalıdır. Sadece eğitim bilimleri alanındaki ölçekler için değil, mental testler, yeterlilik ölçekleri, tutum ölçekleri gibi tüm ölçme araçları için DMF çalışmaları için yürütülmelidir. DMF saptanan maddeler için neler yapılabileceği araştırılmalıdır. Ayrıca, kullanılan her bir yöntemin gücü ve I. tip hata araştırmaları da yapılabilir. Yapılan çalışmada kullanılan yöntemler dışındaki yöntemlerin de sonuçları alınarak karşılaştırmalar genişletilebilir. Farklı yöntemlerle de benzer sonuçların alınıp alınmayacağı incelenebilir. Test uzunluğu ve yeterli örneklem büyüklüğü özelliklerinin sağlanmadığı durumlarda diğer yöntemlerin nasıl cevap verecekleri araştırılabilir. Güvenilirlik ve geçerliliği düşük olan testlerde DMF durumu incelenebilir. İstatistiğe Giriş Okuryazarlık Testi farklı üniversitelerde uygulanıp alınan sonuçlar karşılaştırılabilir.

6. KAYNAKLAR

- Adams, R.J., 1988. Applying the PCM to Educational Diagnosis, Applied Measurement in Education, 1, 4, 347-362.
- Adams, R. J. ve Rowe, K. J., 1988. Item Bias, Keeves, J.P. (Ed.), Pergamon Pres, Oxford, pp.392-398.
- Akaike, H., 1974. A New Look at the Statistical Model Identification, IEEE Transactions on Automatic Control, 19, 6, 716-723.
- Allalouf, A., Hambleton, R.K. ve Sireci, S.G., 1999. Identifying the Causes of DIF in Translated Verbal Items, Journal of Educational Measurement, 36, 3, 185-198.
- Alpar, C. R., 2013. Uygulamalı Çok Değişkenli İstatistiksel Yöntemler, Detay Yayıncılık, Ankara, 886s.
- Altun, A., 2014. Akademik Personel ve Lisansüstü Eğitimi Giriş Sınavı'ndaki (ALES) Maddenin Yerel Bağımsızlığının Nedenlerine İlişkin Sınava Katılanların Görüşleri, New World Science Academy, 9,2, 19-31.
- Ando, T., 2007. Bayesian Predictive Information Criterion for the Evaluation of Hierarchical Bayesian and Empirical Bayes models, Biometrika, 94, 2, 443-458.
- Arabacı, Ö., 2002. Lojistik Regresyon Analizi ve Bir Uygulama Denemesi, Yüksek Lisans Tezi, Uludağ Üniversitesi, Sosyal Bilimler Enstitüsü, Bursa.
- Akın Arıkan, Ç., 2015. Değişen Madde Fonksiyonu Belirlemede MTK-Olabilirlik Oranı, Ordinal Lojistik Regresyon ve Poly-Sıbtest Yöntemlerinin Karşılaştırılmaları. E-International Journal of Educational Research, 6, 1, 1-16.
- Akın Arıkan, Ç., Uğurlu, S. ve Atar, B., 2016. MIMIC, SIBTEST, Lojistik Regresyon ve Mantel-Haenszel Yöntemleriyle Gerçekleştirilen DMF ve Yanlılık Çalışması, Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 31, 1, 34-52.
- Arıkan, S., Çelen, Ü., Demirtaşlı, R.H., Gülleroğlu, H.D., Gültekin, S., Kilmen, S. ve Bilican Demir, S., 2017. Eğitimde Ölçme ve Değerlendirme, Çıkrıkçı Demirtaşlı, N. (Ed.), Anı Yayıncılık, 4, Ankara, 400s.
- Atalay Kabasakal, K. ve Kelecioğlu, H., 2012. Evaluation of Attitude Items in PISA 2006 Student Questionnaire in Terms of Differential Item Functioning, Ankara University Journal of Faculty of Educational Sciences (JFES), 45, 2, 77-96.

- Atalay Kabasakal, K., Gök, B., Kelecioğlu, H. ve Arsan, N., 2012. Değişen Madde Fonksiyonunun Belirlenmesinde Kullanılan Farklı Yöntemlerin Karşılaştırılması: Bir Simülasyon Çalışması, Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 43, 43, 270-281.
- Atasoy, D., 2001. Lojistik Regresyon Analizinin İncelenmesi ve Bir Uygulaması, Yüksek Lisans Tezi, Cumhuriyet Üniversitesi, Sosyal Bilimler Enstitüsü, Sivas.
- Atılğan, H., Kan, A. ve Doğan, N. 2009, Eğitimde ölçme ve değerlendirme, Anı Yayıncılık, 3, Ankara, 462s.
- Bademci, V., 2006a. Tartışmayı Sonlandırmak: Cronbach'ın Alfa Katsayısı, İki Değerli [0,1] Ölçümlenmiş Maddeler ile Kullanılabilir, Atatürk Üniversitesi Kazım Karabekir Eğitim Fakültesi Dergisi, 13, 0, 438-446.
- Bademci, V., 2006b. Güvenilirliği Doğru Anlamak ve Bazı Klişeleri Yıkarak: Bilinenin Aksine Cronbach'ın Alfa Katsayısı Negatif ve "-1"den Küçük Olabilir, İnönü Üniversitesi Eğitim Fakültesi Dergisi, 7, 12, 3-26.
- Bağcan Büyükturan, E. ve Çıkrıkçı Demirtaşlı, N., 2012. Çoktan Seçmeli Testler ile Yapılandırılmış Gridlerin Psikometrik Özellikleri Bakımından Karşılaştırılması, Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi, 45,1, 395-415.
- Bakan Kalaycıoğlu, D. ve Kelecioğlu, H., 2011. Öğrenci Seçme Sınavının Madde Yanlılığı Açısından İncelenmesi, Eğitim ve Bilim, 36, 161, 3-13.
- Baker, F.B., 2001. The Basics of Item Response Theory, Eric, Washington DC, 168s.
- Barton, M.A. ve Lord, F.M., An Upper Asymptote for the Three-Parameter Logistic Item Response Model, Educational Testing Service, New Jersey, 1981.
- Başman, M. ve Kutlu, Ö., 2020. Identification of Differential Item Functioning on Mathematics Achievement According to the Interactions of Gender and Affective Characteristics by Rasch Tree Method, International Journal of Progressive Education, 16, 2, 205-217.
- Başol, G., 2013. Eğitimde Ölçme ve Değerlendirme, Pegem Yayıncılık, 1, Ankara, 312.
- Baştürk, R., 2010. Bilimsel Araştırma Ödevlerinin Rasch Ölçme Modeli ile Değerlendirilmesi, Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 1, 1 51-57.
- Baykul, Y., 2015. Eğitimde ve Psikolojide Ölçme: Klasik Test Teorisi ve Uygulaması, Pegem Akademi, 3, Ankara, 425s.
- Behizadeh, N. ve Engelhard, G., 2014. Development and Validation of A Scale to Measure Perceived Authenticity in Writing, Assessing Writing, 21, 18-36.

- Benito, J.G. ve Ara, M.J.N., 2000. A Comparison of χ^2 , RFA and IRT Based Procedures in the Detection of DIF, Quality ve Quantity, 34, 1, 17-31.
- Berberoğlu, G., 1988. Seçme Amacıyla Kullanılan Testlerde Rasch Modelinin Katkıları, Yayınlanmamış Doktora Tezi, Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.
- Berberoğlu, G. ve Çalikoğlu, G., 1992. The Construction of A Turkish Computer Attitude Scale, Studies in Educational Evaluation, 24, 2, 841-845.
- Berg, A., Meyer, R. ve Yu, J., 2004. Deviance Information Criterion for Comparing Stochastic Volatility Models, Journal of Business & Economics Statistics, 22, 1, 107-120.
- Binet, A. ve Simon, T., 1916. The Development of Intelligence in Children: (The Binet Simon Scale), Williams & Wilkins Co., 1, Baltimore, 366s.
- Birnbaum, A., 1957. Efficient Design and Use of Tests of Ability for Various Decision-Making Problems (Series Report No. 58-16. Project No. 7755-23), Randolph Air Force Base, TX: USAF School of Aviation Medicine.
- Birnbaum, A., 1968. Some Latent Trait Models and Their Use in Inferring An Examinee's Ability, F.M. Lord ve M.R. Novick (Ed.), Addison-Wesley, Reading, pp.397-479.
- Bonifay, W. ve Cai, L., 2017. On the Complexity of Item Response Theory Models, Multivariate Behavioral Research, 52, 4, 465-484.
- Bozdoğan, H., 1987. Model Selection and Akaike's Information Criterion (AIC): The General Theory and Its Analytics Extensions, Psychometrika, 52, 3, 345-370.
- Burnham, K.P. ve Anderson, D.R., 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, Springer Nature, UK, 2, 485s.
- Burnham, K.P. ve Anderson, D.R., 2004. Multimodel Inference: Understanding AC and BIC in Model Selection, Social Methods & Research, 33, 2, 261-304.
- Büyüköztürk, Ş., Kılıç Çakmak, E., Akgün, Ö. E., Karadeniz, Ş. ve Demirel, F., 2016. Bilimsel Araştırma Yöntemleri, Pegem Akademi, 21, Ankara, 358s.
- Breslow, N. E. ve Day, N.E., 1980. Statistical Methods in Cancer Research: Vol. 1-The Analysis of Case-Control Studies, IARC Scientific Publications,1, Lyon, 346s.
- Camilli, G. ve Shepard, L.A., 1994. Methods for Identifying Biased Test Items, Sage, 4, İngiltere, 181s.
- Chen, W.H. ve Thissen, D., 1997. Local Dependence Indexes for Item Pairs Using Item Response Theory, Journal of Educational and Behavioral Statistics, 22, 3, 265-289.

- Cristobal, E., Flavian, C. ve Guinaliu, M., 2007. Perceived E-Service Quality: Measurement Validation and Effects on Consumer Satisfaction and Website Loyalty, Managing Service Quality: An International Journal, 17, 3, 317-340.
- Crocker, L. ve Algina, J., 2008. Introduction to Classical and Modern Test Theory, Cengage Learning Publication, 3, Ohio, 527s.
- Cronbach, L.J., 1984. Essentials of Psychological Testing, Harper & Row, 4, New York, 630s.
- Çelen, Ü. ve Aybek, E.C., 2013. Öğrenci Başarısının Öğretmen Yapımı Bir Testle Klasik Test Kuramı ve Madde Tepki Kuramı Yöntemleriyle Elde Edilen Puanlara Göre Karşılaştırılması, Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 4, 2, 64-75.
- Dancer, L. S., Anderson, A. J. ve Derlin, R. L., 1994. Use of Log-Linear Models for Assessing Differential Item Functioning in A Measure of Psychological Functioning, Journal of Consulting and Clinical Psychology, 62, 4, 710-717.
- De Ayala, R.J., Dodd, B.G. ve Koch, W.R., 1989. A comparison of the Graded Response and Partial Credit Models for Assessing Writing Ability, Annual Meeting of the National Council on Measurement in Education, Mart, San Francisco.
- De Gruijter, D.N.M. ve van der Kamp, L.J.T., 2008. Statistical Test Theory for Behavioral Sciences, Psychometrika, 73, 4, 791-792.
- DeMars, C., 2010. Item Response Theory, Oxford University Press, 1, New York, 130s.
- Devine, P. ve Raju, N., 1982. Extent of Overlap Among Four Item Bias Methods, Educational and Psychological Measurement, 42, 4, 1049-1066.
- Doğan, N. ve Öğretmen, T., 2005. Test ve Madde Yanlılığı, Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi, 5, 1, 89-103.
- Doğan, N., 2002. Klasik Test Kuramı ve Örtük Özellikler Kuramının Örneklem Büyüklükleri Bağlamında Karşılaştırılması, Yayınlanmamış Doktora Tezi, Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.
- Doig, B.A., Mellor, S., Piper, K. ve Maters, G.N., 1994. Conceptual Understanding in Social Education, ACER, 1, Melbourne, 167s.
- Drasgow, F. ve Lissak, R., 1983. Modified Parallel Analysis: A Procedure for Examining the Latent Dimensionality of Dichotomously Scored Item Responses, Journal of Applied Psychology, 68, 3, 363-373.
- Ebel, R. L., 1972. Essentials of Educational Measurement, Prentice-Hall, 1, New Delhi 363s.

- Eells, K., Davis, A., Havighurst, R. J., Herrick, V. E. ve Tyler, R. W., 1951. Intelligence and Cultural Differences: A Study of Cultural Learning and Problem-Solving, University of Chicago Press, 1, Chicago, 388s.
- Elhan, A.H. ve Atakurt, Y., 2005. Ölçeklerin Değerlendirilmesinde Niçin Rasch Analizi Kullanılmalıdır?, Ankara Üniversitesi Tıp Fakültesi Mecmuası, 58, 1, 47-50.
- Embretson, S.E. ve Reise, S.P., 2000. Item Response Theory for Psychologists, Psychology Press, 1, Los Angeles, 384s.
- Engelhard, G., 1992. The Measurement of Writing Ability with a Many-Faceted Rasch Model, Applied Measurement in Education, 5, 3,171-191.
- Ercan, İ. ve Kan, İ., 2004. Ölçeklerde Güvenirlik ve Geçerlik, Uludağ Üniversitesi Tıp Fakültesi Dergisi, 30, 3, 211-216.
- Erdemir, A., 2015. Bir, İki, Üç ve Dört Parametrelili Lojistik Madde Tepki Kuramı Modellerinin Karşılaştırılması, Yayınlanmamış Yüksek Lisans Tezi, Gazi Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.
- Field, A., 2005. Discovering Statistics Using Spss, Sage Publication, 2, İngiltere, 779s.
- French, A. W. ve Miller, T. R., 1996. Logistic Regression and Its Use in Detecting Differential Item Functioning in Polytomous Items, Journal of Educational Measurement, 33, 3, 315-332.
- Gelbal, S., 1994. P Madde Güçlük İndeksi ile Rasch Modelinin b Parametresi ve Bunlara Dayalı Yetenek Ölçüleri Üzerine Bir Karşılaştırma, Doktora Tezi, Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Gelin, M.N. ve Zumbo, B.D., 2003. Differential Item Functioning Results May Change Depending on How An Item is Scored: An İllustration with the Center for Epidemiologic Studies Depression Scale, Educational and Psychological Measurement, 63, 1, 65-74.
- Gelman, A. ve Carlin, J., B., Stern, H. S. ve Rubin, D.B., 2004. Bayesian Data Analysis, Chapman & Hall/CRC, 2, Londra, 666s.
- Gök, B., Kelecioğlu, H. ve Doğan, N., 2010. Değişen Madde Fonksiyonu Belirlemede Mantel-Haenszel ve Lojistik Regresyon Tekniklerinin Karşılaştırılması, Eğitim ve Bilim, 35, 156, 3-16.
- Gömlüksiz, M. ve Erkan, S., 2016. Eğitimde Ölçme ve Değerlendirme, 4, Nobel, Ankara, 338s.
- Green, D. R., 1974. Racial and Ethnic Bias in Achievement Tests and What to Do About It, CA: CTB/McGraw-Hill,1, Monterey.

- Groothuis Oudshoorn, C.G., Van den Heuvel, E. ve Krabbe, P.F.M, 2018. A Preference-Based Item Response Theory Model to Measure Health: Concepts and Mathematics of the Multi-Attribute Preference Response Model, BMC Medical Research Methodology, 18, 1, 1-13.
- Gücüm, B., 1990. Kriter Dayanaklı Testler için Madde Seçiminde İki Farklı Yöntem Olarak Rasch Modeli ve Yargıcı Kararlarının Karşılaştırılması, Doktora Tezi, Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Güler, N. ve Gelbal, S., 2010. Klasik Test Kuramı ve Çok Değişkenlik Kaynaklı Rasch Model Üzerine Bir Çalışma, Eurasian Journal of Educational Research, 38, 10 108-125.
- Güler, N., 2008. Klasik Test Kuramı, Genellenebilirlik Kuramı ve Rasch Model Üzerine Bir Araştırma, Doktora Tezi, Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.
- Hambleton, R. ve Swaminathan, H., 1985. Item Response Theory, Springer Science & Business Media, 1, USA, 310s.
- Hambleton, R. K. ve Cook, L. L., 1983. The Robustness of Item Response Models and the Effects of Test Length and Sample Size on the Precision of Ability Estimates, D. Weiss (Ed.), Academic Press, New York, pp.31-49.
- Hambleton, R., Swaminathan, H. ve Rogers, H.J., 1991. Fundamentals of Item Response Theory, Sage Publications, 2, Londra, 184s.
- Hasançebi, B., Terzi, Y. ve Küçük, Z. 2019. Olasılık Okuryazarlık Testinin Madde Tepki Kuramına Dayalı Yöntemler ile Değişen Madde Fonksiyonu Açısından İncelenmesi, Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi, 19, 3, 643-652.
- Hattie, J.A., 1974. An Emprical Study of Various Indices for Determining Unidimensionality, Multivariate Behavioral Research, 19, 1, 49-78.
- Henson, J.M., Reise, S.P. ve Kim, K.H., 2007. Detecting Mixtures from Structural Model Differences Using Latent Variable Mixture Modelling: A Comparison of Relative Model Fit Statistics, Structural Equation Modelling A Multidisciplinary Journal, 14, 2, 202-226.
- Hosmer, D.W. ve Lemeshow, S., 1989. Applied Logistic Regression, A Wiley-Interscience Publication, 1, New York, 383s.
- Jin, K.Y. ve Wang, W.C., 2014. Item Response Theory Models for Performance Decline During Testing, Journal of Educational Measurement, 51, 2, 178-200.
- Julian, E.R. ve Wright, B.D., 1988. Using Computerized Patient Simulations to Measure the Clinical Competence of Physicians, Applied Measurement in Education, 4, 1, 299-318.

- Kelley, T. L., 1916. A Simplified Method of Using Scaled Data for Purposes of Testing, School and Society, 4, 34, 71-75.
- Kılıç, D., 2000. Lojistik Regresyon Analizi ve Pazarlama Araştırmalarında Bir Uygulama, Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.
- Kleinbaum, D.G., 1994. Logistic Regression A Self-Learning Text, Springer, 1, New York, 251s.
- Koch, W.R., 1983. Likert Scaling Using the Graded Response Latent Trait Model, Applied Psychological Measurement, 7, 1, 15-32.
- Koch, W.R. ve Dodd, B.G., 1989. An Investigation of Procedures for Computerized Adaptive Testing Using Partial Credit Scoring, Applied Measurement in Education, 2, 4, 335-357.
- Kondo-Brown, K., 2002. A FACETS Analysis of Rater Bias in Measuring Japanese Second Language Writing Performance, Language Testing, 19, 1, 3-31.
- Köse, A., 2010. Madde Tepki Kuramına Dayalı Tek Boyutlu ve Çok Boyutlu Modellerin Test Uzunluğu ve Örneklem Büyüklüğü Açısından Karşılaştırılması, Yayımlanmamış Doktora Tezi, Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Köse, A., 2015. Aşamalı Tepki Modeli ve Klasik Test Kuramı Altında Elde Edilen Test ve Madde Parametrelerinin Karşılaştırılması, Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi, 15, 2, 184-197.
- Kuder, G. F. ve Richardson, M. W., 1937. The Theory of Estimation of Test Reliability, Psychometrika, 2, 3, 151-160.
- Kumru, Ö., 2003. Markov Zinciri Monte Carlo Yöntemleri, Yüksek Lisans Tezi, Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.
- LaHuis, D.M., Clark, P. ve O'Brien, E., 2011. An Examination of Item Response Theory Item Fit Indices for the Graded Response Model, Organizational Research Methods, 14, 1, 10-23.
- Lee, Y., 2004. Examining Passage-Related Local Item Dependence (LID) and Measurement Construct Using Q3 Statistics in an EFL Reading Comprehension Test, Language Testing, 21, 1, 74-100.
- Linacre, J.M., 1989. Many-facet Rasch Measurement, MESA Press, 1, Chicago, 149s.
- Loken, E. ve Rulison, K.L., 2010. Estimation of A Four-Parameter Item Response Theory Model, British Journal of Mathematical and Statistical Psychology, 63, 3, 509-525.

- Lord, F.M., 1952. A Theory of Test Scores (Psychometric Monograph No. 7), Psychometric Corporation, 1, Richmond, 93s.
- Lord, F.M. ve Novick, M.R., 1968. Statistical Theories of Mental Test Scores, Addison-Wesley Publishing Company, 1, USA, 585s.
- Lord, F.M., 1980. Applications of Item Response Theory to Practical Testing Problems, Routledge, 1, New York, 288s.
- Lubke, G.H. ve Neale, M.C., 2006. Distinguishing Between Latent Classes and Continuous Factors: Resolution by Maximum Likelihood, Multivariate Behavioral Research, 41, 4, 499-532.
- Lunz, M.E., Wright, B.D. ve Linacre, J.M., 1990. Measuring the Impact of Judge Severity on Examination Scores. Applied Measurement in Education, 3, 4, 331-345.
- Lynch, B.K. ve McNamara, T., 1998. Using G-Theory and Many-Facet Rasch Measurement in the Development of Performance Assessments of the ESL Speaking Skills of Immigrants, Language Testing, 15, 2, 158-180.
- MacDonald, P. ve Paunonen, S.V., 2002. A Monte Carlo Comparison of Item and Person Statistics Based on Item Response Theory Versus Classical Test Theory, Educational and Psychological Measurement, 62, 6, 921-943.
- Macmillan, D.D., 1995. A Comparison of Approaches: Classical Generalizability and Multifaceted Rasch, Doktora Tezi, Alberta University, Kanada.
- Madera, E.K., 2003. Application of the Graded Response Model to the assessment of Student Attitudes, Yayınlanmamış Doktora Tezi, University of Toronto, Toronto.
- Magis, D., 2013. A Note on the Item Information Function of the Four-Parameter Logistic Model, Applied Psychological Measurement, 37, 4, 304-315.
- Masters, G.A., 1982. Rasch Model for Partial Credit Scoring, Psychometrika, 47, 2, 149-174.
- Masters, G.N. ve Wright, B.D., 1982. Defining A Fear-of-Crime Variable: A Comparison of Two Rasch Models, Education Research and Perspectives, 9, 1, 18-32.
- Masters, G.N. ve Evans, J., 1986. Banking Non-Dichotomously Scored Items, Applied Psychological Measurement, 10,4, 355-367.
- Matteucci, M. ve Stracqualursi, L., 2006. Student Assessment via Graded Response Model, Statistica, 66, 4, 435-447.
- McLaughlin, M. E. ve Drasgow, F., 1987. Lord's Chi-Square Test of Item Bias with Estimated and with Known Person Parameters, Applied Psychological Measurement, 11, 2, 161-173.

- Mellenberg, G.J., 1983. Conditional Item Bias Methods, S. H. Irvine ve W. J. Barry (Ed.), Plenum Pres., New York, pp. 123-144.
- Mertler, C.A. ve Vannatta, R.A., 2016. Advanced and Multivariate Statistical Methods: Practical Application and Interpretation. Routledge, 6, California, 390s.
- Mielenz, T.J., Edwards, M.C. ve Callahan, L.F., 2010. Item Response Theory Analysis of Two Questionnaire Measures of Arthritis-Related Self Efficacy Beliefs from Community Based US Samples, *Arthritis*, 2010, 1, 1-8.
- Mulqueen C., Baker D. ve Dismukes R.K., 2000. Using Multifacet Rasch Analysis to Examine the Effectiveness of Rater Training, The 15th Annual Meeting of the Society for Industrial/Organizational Psychology (SIOP), April, Los Angeles.
- Nylund, K. L., Asparouhov, T., ve Muthen, B. O., 2007. Deciding On The Number Of Classes In Latent Class Analysis In Latent Class Analysis And Growth Mixture Modeling: A Monte Carlo Simulation Study, Structural Equation Modeling, 14, 4, 535-569.
- Orlando, M., 2004. Critical Issues to Address When Applying Item Response Theory Models, Conference on Improving Health Outcomes Assessment, Haziran, USA.
- Queensoap, M. ve Orluwene, G.W., 2019. Use of Mantel Haenszel Differential Item Functioning in Detecting Item Bias in Chemistry Achievement Test in Four Ethnic Groups in Nigeria, International Journal of Current Research, 11, 3, 2665-2670.
- Oshima, T.C. ve Morris, S.B., 2008. Raju's Differential Functioning of Items and Tests (DFIT): An NCME Instructional Module on, Educational Measurement Issues and Practice, 27, 3, 43-50.
- Osterlind, S. J., 1983. Test Item Bias, Sage, 1, Newbury, 85s.
- Osterlind, S.J., 1989. Constructing Test Items, Kluwer Academic Publishers, 1, Berlin, 343s.
- Ostini, R. ve Nering, M.L., 2006. Polytomous Item Response Theory Models, Sage Publications, California, 120s.
- Özgüven, İ.E., 1994. Psikolojik Testler, Yeni Doğu Matbaası, 1, Ankara, 454s.
- Özmen, D.T., 2014. PISA 2009 Okuma Testi Maddelerinin Yanlılığı Üzerine Bir Çalışma, Eğitim Bilimleri ve Uygulama Dergisi, 13, 26, 147-165
- Peng, C.Y., Lee, K.L. ve Ingersoll, G.M., 2002. An Introduction to Logistic Regression Analysis and Reporting, The Journal of Educational Research, 96, 1, 3-14.
- Price, L. R., 2017. Psychometric Methods: Theory and Practice, Guilford Press, 2, New York, 552s.

- Raju, N.S., 1988. The Area Between Two Item Characteristic Curves, Psychometrika, 53, 4, 495-502.
- Raju, M.S., Drasgow, F. ve Slinde, J.A. 1993. An Empirical Comparison of the Area Methods, Lord's χ^2 Test and the Mantel-Haenszel Technique for Assessing Differential Item Functioning, Educational and Psychological Measurement, 53, 2, 301-314.
- Ranyard, R., McNair, S., Nicolini, G. ve Duxbury, D., 2020. An Item Response Theory Approach to Constructing and Evaluating Brief and in Depth Financial Literacy Scales, The Journal of Consumer Affairs, 51, 3, 1121-1156.
- Rasch, G., 1960. Probabilistic Models for Some Intelligence and Attainment Test, Paedagogiske Institut, 1, Danimarka, 184s.
- Rasch, G., 1966. An Item Which Takes Individual Differences Into Account, British Journal of Mathematical and Statistical Psychology, 19, 1, 49-57.
- Reckase, M.D., 2009. Multidimensional Item Response Theory, Springer, 1, New York, 339s.
- Reckase, M.D., Ackerman, T.A. ve Carlson, J.E., 1988. Building A Unidimensional Test Using Multidimensional Items, Journal of Educational Measurement, 25, 13, 193-203.
- Reise, S.P. ve Yu, J., 1990. Parameter Recovery in the Graded Response Model Using MULTILOG, Journal of Educational Measurement, 27, 2, 133-144.
- Reise, S. P. ve Waller, N. G., 2003. How Many IRT Parameters Does It Take to Model Psychopathology Items?, Psychological Methods, 8, 2, 164-184.
- Rodney, G.L. ve Drasgow, F., 1990. Evaluation of Two Methods for Estimating Item Response Theory Parameters When Assessing Differential Item Functioning, Journal of Applied Psychology, 75, 2, 164-174.
- Rulison, K. L. ve Loken, E. 2009. I've Fallen and I Can't Get Up: Can High-Ability Students Recover From Early Mistakes in CAT?, Applied Psychological Measurement, 33, 2, 83-101.
- Rupp, A. A., 2003. Item Response Modeling with BILOG-MG and MULTILOG for Windows, International Journal of Testing, 3, 4, 365-384.
- Samejima, F., 1969. Estimation of Latent Ability Using Response Pattern of Graded Scores, Psychometrika, 34, 1, 1-97.
- Scheuneman, J., 1979. A Method of Assessing Bias in Test Items, Journal of Educational Measurement, 16, 3, 143-152.

- Schumacker, R.E., 1999. Many-facet Rasch Analysis with Crossed, Nested, and Mixed Designs, Journal of Outcome Measurement, 3, 4, 323-338.
- Schwarz, G., 1978. Estimating the Dimension of a Model, Ann. Statist., 6, 2, 461 – 464.
- Sebok, S., 2010. Counselling Application Selection Process Using Rasch Analysis and Generalizability Theory, Yüksek Lisans Tezi, University of Northern British Columbia, USA.
- Shepard, L., Camilli, G. ve Averill, M., 1981. Comparison of Procedures for Detecting test-Item Bias with Both Internal and External Ability Criteria, Journal of Educational and Behavioral Statistics, 6, 4, 317-375.
- Seong, Tae-Je ve Subkoviak, M.J., 1987. A Comparative Study Of Recently Proposed İtem Bias Detection Methods, Annual Meeting of the American Educational Research Association, April, Toronto.
- Sırgancı, G. ve Çakan, M., 2020. Sıralı Lojistik Regresyon ve Poly-Sıbtest Yöntemleri ile Değişen Madde Fonksiyonunun Belirlenmesi, Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi, 20, 1, 705-717.
- Sugiura, N., 1978. Further Analysis of the Data by Akaike's Information Criterion and the Finite Corrections, Communications in Statistics, Theory & Methods, 7, 1, 13-26.
- Sukirno, D. ve Siengthai, S., 2010. The Comparison of Graded Response Model and Classical Test Theory in Human Resource Research: A Model Fitness Test, Research and Practice in Human Resource Management, 18, 2, 77-90.
- Sünbül, Ö. ve Ömür Sünbül, S., 2016. Değişen madde fonksiyonunun belirlenmesinde kullanılan yöntemlerde I. Tip hata ve güç çalışması, Elementary Education Online, 15, 3, 882-897.
- Spiegelhalter, D.J., Best, N., Carlin, B.P. ve van der Linde, A., 2002. Bayesian Measures of Model Complexity and Fit (with discussion), Journal of the Royal Statistical Society Series B (Statistical Methodology), 64, 4, 1-34.
- Swaminathan, H. ve Gifford, J., 1983. Estimation of Parameters in the Three-Parameter Latent Trait Model, D. J. Weiss ve R. D. Bock (Ed.), Academic Press, New York, pp. 13–30.
- Şahin, A., 2012. Madde Tepki Kuramında Test Uzunluğu ve Örneklem Büyüklüğünün Model-Veri Uyumu, Madde Parametreleri ve Standart Hata Değerlerine Etkisinin İncelenmesi, Doktora Tezi, Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Şencan, H., 2005. Sosyal ve Davranışsal Ölçümlerde Güvenilirlik ve Geçerlilik, Seçkin Yayıncılık, 1, Ankara, 865s.

- Tan, Ş. ve Erdoğan, A., 2004. Öğretimi Planlama ve Değerlendirme, Pegem Yayıncılık, 5, Ankara, 307s.
- Tate, R., 2003. A Comparison of Selected Empirical Methods for Assessing the Structure of Responses to Test Items, Applied Psychological Measurement, 27, 3, 159-203.
- Tekin, H., 1984. Eğitimde Ölçme ve Değerlendirme, Has-soy Matbaacılık, 1, Ankara, 328s.
- Tucker, L. R., 1946. Maximum Validity of A Test with Equivalent Items, Psychometrika, 11, 1, 1-13.
- Turgut, M.F. ve Baykul, Y., 2012. Eğitimde Ölçme ve Değerlendirme, Pegem Akademi, 4, Ankara, 419s.
- Türkan, A., 2014. Seviye Belirleme Sınavının Rasch Modeline Göre Cinsiyet Değişkeni Açısından Yanlılığının İncelenmesi, Yüksek Lisans Tezi, Gaziantep Üniversitesi, Eğitim Bilimleri Enstitüsü, Gaziantep.
- Ural, A. ve Kılıç, İ., 2005. Bilimsel Araştırma Süreci ve Spss ile Veri Analizi, Detay Yayıncılık, 1, Ankara, 288s.
- Uttaro, T. ve Lehman, A., 1999. Graded Response Modeling of the Quality of Life Interview, Evaluation and Program Planning, 22, 1999, 41-52
- van der Linden, W.J. ve Hambleton, R.K., 1997. Handbook of Modern Item Response Theory, Springer, 1, New York, 512s.
- Vrieze, S.I., 2012. Model Selection and Psychological Theory: A Discussion of the Differences Between the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), Psychological Methods, 17, 2, 228-243.
- Wright, B.D. ve Masters, G.N., 1982. Rating Scale Analysis, Mesa Press, 1, Chicago, 223s.
- Wright, B.D. ve Stone, M.H., 1979. Best Test Design, Mesa Press, 1, 245s.
- Yang, C. ve Yang, C., 2007. Separating Latent Classes by Information Criteria, Journal of Classification, 24, 2, 183-203.
- Yen, W.M., 1984. Effects of Local Item Dependence on the Fit and Equating Performance of the Three-Parameter Logistic Model, Applied Psychological Measurement, 8, 2, 125-145.
- Yen, W.M. 1993. Scaling Performance Assessments: Strategies for Managing Local Item Dependence, Journal of Educational Measurement, 30, 3, 187-213.
- Yule, G.U., 1911. An Introduction to the Theory of Statistics, Charles Griffin & Company Limited, 1, Londra, 600s.

Yüksel, S., 2012. Ölçeklerde Saptanan Madde İşlev Farklılığının Karma Rasch Modelleri ile İncelenmesi, Doktora Tezi, Ankara Üniversitesi, Sağlık Bilimleri Enstitüsü, Ankara.

Zumbo, B. D., 1999. A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as A Unitary Framework For Binary and Likert-Type (Ordinal) Item Scores, Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.



7. EKLER


İstatistiğe Giriş Okuryazarlık Testi

Bölüm:

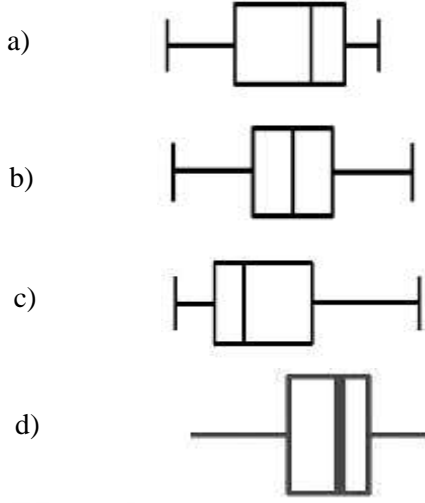
Cinsiyet:

İSTATİSTİĞE GİRİŞ OKURYAZARLIK TESTİ

- Aşağıdakilerden hangisi sürekli değişkenlere örnek olarak verilebilir?
a) Bir sınıfın mevcudu b) Bir ilin nüfusu c) Gelir düzeyi d) Çocuk sayısı
- Kitleyi temsil etmek için örneklemden hesaplanan değerlere ne ad verilir?
a) Örneklem b) Parametre c) İstatistik d) Anakütle
- Aşağıdakilerden hangisi nicel veriye örnek olarak verilebilir?
a) Cinsiyet b) Eğitim durumu c) Göz rengi d) Boy
- Herhangi bir veri kümesindeki elemanların ortalamadan sapmalarının karelerinin toplamının veri sayısına bölümü aşağıdaki kavramlardan hangisi ile ifade edilebilir?
a) Standart hata b) Değişim katsayısı c) Standart sapma d) Varyans
- N tane nesnenin tümünün veya bir kısmının tekrar olmaksızın sıralanması işlemine ne ad verilir?
a) Cebir b) Olasılık c) Permütasyon d) Kombinasyon
- Aşağıdakilerden hangisi kesikli rastgele değişkene örnektir?
a) Kişinin ağırlığı
b) Bir sorunun çözülme süresi
c) Madeni paranın 3 kere atılması deneyinde yazı gelme sayısı
d) Bir arsanın fiyatı
- Bir deney sonucundaki tüm mümkün sonuçlar kümesine ne denir?
a) Olay b) Örnek uzay c) Örneklem d) Olasılık
- İki olaydan birinin gerçekleşmesi diğerinin gerçekleşmesini etkilemiyorsa bu tür olaylar için aşağıdakilerden hangisi söylenebilir?
a) Bağımlı olay b) Rastgele deney c) Bağımsız olay d) Ayrık olay
- Kitle veya örnek birimlerinden belirli değişkenler için çeşitli ölçme düzeylerine göre alınan ölçüm sonuçlarının oluşturduğu küme aşağıdakilerden hangisidir?
a) Parametre b) Denek c) Veri d) Olasılık
- Aşağıdakilerden hangisi nitel veriler için çizilebilecek grafiklerdendir?
a) Yığılımlı sıklık b) Dağılım poligonu c) Dal ve yaprak gösterimi d) Daire dilimi grafiği
- Aşağıdakilerden hangisi nicel veriler için kullanılacak bir konum ölçüsüdür?
a) Ortalama b) Standart sapma c) Dağılım genişliği d) Varyans

12. Ortalamalar için yazılan eşitliklerden hangisi doğrudur?
 a) $G.O. = A.O. \leq H.O.$ b) $A.O. \leq G.O. \leq H.O.$ c) $A.O. \leq H.O. = G.O.$ d) $(A.O. \geq G.O. \geq H.O.)$
13. A kentinden B kentine 4, B kentinden C kentine 5 farklı yol bulunmaktadır. Buna göre, A kentinden C kentine, B kentine uğramak koşulu ile kaç farklı yoldan gidilebilir?
 a) 16 b) 18 c) 20 d) 25
14. $A = \{1,2,3,4,5\}$ kümesinin elemanları kullanılarak 4 basamaklı kaç doğal sayı yazılabilir?
 a) 120 b) 255 c) 400 d) 625
15. $A = \{1,2,3,4,5,6,7\}$ kümesinin elemanlarını kullanarak rakamları farklı 3 basamaklı kaç farklı çift sayı yazılabilir?
 a) 70 b) 84 c) 90 d) 98
16. $A = \{1,2,3,4,5,6\}$ kümesinin elemanları ile rakamları farklı 3 basamaklı 300'den büyük kaç doğal sayı yazılabilir?
 a) 36 b) 45 c) 60 d) 80
17. 5 elemanlı bir kümenin en az 2 elemanlı alt kümelerinin sayısı kaçtır?
 a) 20 b) 26 c) 30 d) 45
18. 4 erkek ve 5 kız öğrenci arasından en az biri erkek olmak üzere 3 kişilik kaç farklı grup kurulabilir?
 a) 44 b) 54 c) 64 d) 74
19. 
 Şekildeki 7 noktadan kaç doğru geçer?
 a) 14 b) 22 c) 24 d) 30
20. Aşağıdakilerden hangisi eşit aralıklı ölçme düzeyinde ölçülen bir değişkendir?
 a) Santigrat cinsinden yıllık ortalama hava sıcaklığı b) Eğitim durumu
 b) Cinsiyet d) Tedavide kullanılan ilaç türü
21. Bir işyerinde çalışan büyük bir grubun aylık gelir ortalaması 1000 TL ve standart sapması 100 TL'dir. Bu grupta geliri 1200 TL olan bir çalışanın Z değeri aşağıdakilerden hangisidir?
 a) -2 b) 2 c) -1 d) 1
22. Bir normal dağılım eğrisinin sağa çarpık olması durumunda aşağıdakilerden hangisinin doğru olduğu söylenebilir?
 a) $Mod = Medyan = \bar{X}$ b) $Mod < Medyan < \bar{X}$
 c) $Mod > Medyan > \bar{X}$ d) $Mod = Medyan > \bar{X}$
23. Bir standart normal dağılımın aşağıdaki özelliklerden hangisini sağlaması beklenir?
 a) $\mu = 0, \sigma = 1, \text{Çarpıklık} = 0, \text{Basıklık} = 3$
 b) $\mu = 0, \sigma = 1, \text{Çarpıklık} = 3, \text{Basıklık} = 3$
 c) $\mu = 1, \sigma = 1, \text{Çarpıklık} = 0, \text{Basıklık} = 3$
 d) $\mu = 0, \sigma = 1, \text{Çarpıklık} = 3, \text{Basıklık} = 3$

24. Aşağıdaki grafiklerden hangisi simetrik bir dağılımı gösterir?



25.

Z	0	0,01	0,02	0,03
1,0	0,3413	0,3438	0,3461	0,3485
1,1	0,3643	0,3665	0,3686	0,3708
1,2	0,3849	0,3869	0,3888	0,3907
1,3	0,4032	0,4049	0,4066	0,4082
1,4	0,4192	0,4207	0,4222	0,4236

Yukarıda Z tablosunun bir kesiti verilmiştir. Bu tabloya göre $P(Z \leq -1.32)$ değeri aşağıdakilerden hangisidir?

- a) 0,0934 b) 0,0926 c) 0,0912 d) 0,0955

ÖZGEÇMİŞ

İlköğretimi Trabzon Prof. İhsan Koz İlköğretim Okulu ve Samsun Gülsüm-Sami Kefeli İlköğretim Okullarında tamamladıktan sonra Lise öğrenimini Samsun Milli Piyango Anadolu Lisesi'nde bitirmiştir. Ardından 2006 yılında KTÜ Fen Fakültesi İstatistik ve Bilgisayar Bilimleri Bölümüne girmiştir. 2010 yılında bu bölümden mezun olduktan hemen sonra KTÜ Fen Bilimleri Enstitüsü İstatistik ve Bilgisayar Bilimleri Anabilim Dalı Yüksek Lisans programına girmeye hak kazandı. 2013 yılında bu programdan “Özel Bariyerli Yarı-Markov Rastgele Yürüyüş Süreci Üzerine Bir Çalışma” başlıklı tezi savunarak mezun oldu ve aynı yıl KTÜ Fen Bilimleri Enstitüsü İstatistik ve Bilgisayar Bilimleri Anabilim Dalı doktora programına başladı.

Hasançebi, B., Terzi, Y. ve Küçük, Z., 2019. Olasılık Okuryazarlık Testinin Madde Tepki Kuramına Dayalı Yöntemler ile Değişen Madde Fonksiyonu Açısından İncelenmesi, Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi, 19, 3, 643-652.

Hasançebi, B., Terzi, Y. ve Küçük, Z., 2020. Madde Güçlük İndeksi ve Madde Ayırt Edicilik İndeksine Dayalı Çeldirici Analizi, Gümüşhane Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 10, 1, 224-240.

Hasançebi, B., Terzi, Y. ve Küçük, Z., 2020. İki Parametrel Madde Tepki Kuramı Modelinde Parametre Değişmezliğinin İncelenmesi, Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 24, 2, 438-444.