

**ÇOKLU GÖSTERGELER ÇOKLU NEDENLER VE  
LOJİSTİK REGRESYON YÖNTEMLERİNİN DEĞİŞEN  
MADDE FONKSİYONUNU BELİRLEME  
PERFORMANSLARI**

**PERFORMANCE OF MULTIPLE INDICATORS MULTIPLE  
CAUSES AND LOGISTIC REGRESSION PROCEDURES IN  
DETECTING DIFFERENTIAL ITEM FUNCTIONING**

**Seçil ARSLAN**

Hacettepe Üniversitesi

Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin

Eğitim Bilimleri Anabilim Dalı, Eğitimde Ölçme ve Değerlendirme Bilim Dalı İçin  
Öngördüğü

Yüksek Lisans Tezi

olarak hazırlanmıştır.

2014

Eđitim Bilimleri Enstitüsü M¼d¼rl¼đ¼'ne,

Seçil ARSLAN'ın hazırladığı "ÇOKLU GÖSTERGELER ÇOKLU NEDENLER VE LOJİSTİK REGRESYON YÖNTEMLERİNİN DEĐİŐEN MADDE FONKSİYONUNU BELİRLEME PERFORMANSLARI" başlıklı bu çalıŐma j¼rimiz tarafından **Eđitim Bilimleri Anabilim Dalı, Eđitimde Ölçme ve Deđerlendirme Bilim Dalı'nda Yüksek Lisans Tezi** olarak kabul edilmiŐtir.

Başkan

Prof. Dr. Selahattin GELBAL

¼ye (DanıŐman)

Doç. Dr. Burcu ATAR

¼ye

Doç. Dr. Őeref TAN

¼ye

Doç. Dr. Nuri DOĐAN

¼ye

Yrd. Doç. Dr. Derya ÇOBANOĐLU AKTAN

ONAY

Bu tez Hacettepe Üniversitesi Lisansüstü Eđitim-Öđretim ve Sınav Yönetmeliđi'nin ilgili maddeleri uyarınca yukarıdaki j¼ri üyeleri tarafından 29 / 05/ 2014 tarihinde uygun gör¼lm¼Ő ve Enstitü Yönetim Kurulunca ...../...../..... tarihinde kabul edilmiŐtir.

Prof. Dr. Berrin AKMAN  
Eđitim Bilimleri Enstitüsü M¼d¼r¼

# ÇOKLU GÖSTERGELER ÇOKLU NEDENLER VE LOJİSTİK REGRESYON YÖNTEMLERİNİN DEĞİŞEN MADDE FONKSİYONUNU BELİRLEME PERFORMANSLARI

**Seçil ARSLAN**

## ÖZ

Bu araştırmada iki kategorili veriler için Çoklu Göstergeler Çoklu Nedenler (ÇGÇN) (Multiple Indicators Multiple Causes, MIMIC) ve Lojistik Regresyon (LR) yöntemlerinin DMF belirleme performansları incelenmiştir. Bu iki yöntemin performansları 1. tip hata ve güç oranlarına göre değerlendirilmiştir. Çalışmada ele alınan koşullar: örneklem büyüklüğü (2000 ve 4000 bireylik), grupların yetenek dağılımları ( $N(0,1)$  ve  $N(-0.5,1)$ ) ve DMF gösteren madde yüzdesidir (%10 ve %20). Referans gruptaki bireylerin yetenek parametreleri ( $N(0,1)$ ), testin uzunluğu (30 madde), DMF gösterdiği belirtilen maddeler için gruplar arasındaki madde güçlük parametrelerindeki değişim (0.6) ve odak ve referans grup oranı (1:1) sabit tutulan koşullardır.

Araştırmada yöntemlerin performansları 8 farklı simülasyon koşuluna göre incelenmiştir. Çalışmada kullanılan DMF belirleme yöntemlerinden ÇGÇN için MPlus 6, LR içinse SAS 9.1.3 programı kullanılmıştır.

Elde edilen bulgular doğrultusunda, ele alınan koşullara göre yöntemlerin performansları 1. tip hata oranları bakımından farklılık göstermiştir. Bazı koşullar için ÇGÇN yönteminin bazı koşullar için LR yönteminin 1. tip hata oranları daha düşüktür; bazı koşullarda bu değerler birbirine çok yakın çıkmıştır. ÇGÇN yönteminin 1. tip hata oranının daha düşük olduğu koşul sayısı, LR yönteminin 1. tip hata oranının daha düşük olduğu koşul sayısından fazladır. Her iki yöntem için de tüm koşullarda 1. tip hata oranları 0.05'in üzerindedir.

Örneklem büyüklüğünün artması ÇGÇN yöntemi için 1. tip hata oranını önemli ölçüde azaltırken, LR yöntemi için 1. tip hata oranını çok fazla etkilememiştir. DMF'li madde yüzdesinin artması ÇGÇN yönteminin 1. tip hata oranını pek fazla etkilemezken, LR yönteminin 1. tip hata oranını arttırmıştır. Odak grubun yetenek dağılımının farklı olması ise her iki yöntemin de 1. tip hata oranlarını çok

etkilememiştir. Sonuç olarak 1. tip hata oranları bakımından bu iki yöntem karşılaştırıldığında ele alınan koşullardan örneklem büyüklüklerindeki değişim ÇGÇN yöntemi için daha etkili olurken, DMF'li madde yüzdelerindeki değişim LR yöntemi için daha etkili olmuştur.

Çalışmanın sonuçları yöntemlerin güç oranlarındaki değişimler bakımından incelendiğinde; her iki modelin de tüm koşullar için güç oranları kabul edilebilir seviyenin (0.70) üzerinde olduğu sonucuna ulaşılmıştır; ancak ÇGÇN yönteminin güç oranının daha yüksek olduğu koşul sayısı, LR yönteminin güç oranının daha yüksek olduğu koşul sayısından fazladır. Örneklem büyüklüğünün artması her iki yöntem için güç oranını artırmıştır. Odak grubun yetenek dağılımının referans grubun yetenek dağılımından farklı olması her iki yöntemin de güç oranını düşürmüştür. Yetenek dağılımı koşullundaki güç oranındaki bu düşüş LR yönteminde daha fazladır. DMF gösteren madde yüzdelerinin artması her iki yöntemin güç oranını az miktarda artırmıştır. Sonuç olarak güç oranlarındaki değişime bakıldığında her iki yöntem için en etkili değişkenin örneklem büyüklüğü olduğu görülmüştür.

**Anahtar sözcükler:** Değişen madde fonksiyonu, ÇGÇN, LR.

**Danışman:** Doç. Dr. Burcu ATAR, Hacettepe Üniversitesi, Eğitim Bilimleri Anabilim Dalı, Eğitimde Ölçme ve Değerlendirme Bilim Dalı

# PERFORMANCE OF MULTIPLE INDICATORS MULTIPLE CAUSES AND LOGISTIC REGRESSION PROCEDURES IN DETECTING DIFFERENTIAL ITEM FUNCTIONING

Seçil ARSLAN

## ABSTRACT

In this thesis, DIF detection performances of MIMIC and LR methods for dichotomous data are investigated. Performances of these two methods are compared with according to their type-I error and power rates. Conditions covered in the study are: sample size (2000 and 4000 respondents), ability distributions of groups ( $N(0,1)$  and  $N(-0.5,1)$ ), and the percentage of items with DIF (10% and 20%). Ability parameters of the respondents in the reference group ( $N(0,1)$ ), exam length (30 items), the variation in item power parameters between groups for the items that contain DIF (0.6), and the ratio of focus group to reference group (1:1) are the conditions that are held constant. In the survey, performances of the methods are investigated with respect to 8 distinct simulation conditions. MPlus 6 and SAS 9.1.3 computer applications are used for MIMIC and LR DIF detection methods, respectively.

In the light of the results obtained, performances of the methods show differences with respect to type-I error rates, according to the covered conditions. In some cases type-I error rate of MIMIC method is less than the one of LR method and vice versa; in some cases these values are too close to each other. Number of conditions where MIMIC method has less type-I error rate is more than the number of conditions where LR method has less type-I error rate. For both methods, type-I error rates are over 0.05 for all conditions.

Increase in the sample size does not affect type-I error rate too much for LR method, while it decreases type-I error rate for MIMIC method dramatically. Increase in the percentage of items with DIF increases type-I error rate for LR method, where as it does not affect type-I error rate for MIMIC method very much. Having the ability distribution of focus group different does not affect type-I error

rates much for both methods. As a result; when the two methods are compared according to their type-I error rates, it can be concluded that the change in sample size is more effective for MIMIC model, while the change in the percentage of items with DIF is more effective for LR method.

When the results of the study are investigated with respect to the variations in the power rates of methods; it has been found out that the power rates of both models are over the acceptable limit (0.70) for all conditions. However, the number of conditions where MIMIC method has higher power rates is greater than the number of conditions where LR method has higher power rates. Increase in sample size increases power rates of both methods. Having different skill distributions for focus and reference groups decreases power rates of both methods. The decrease in power rate in the different ability distribution case is more in LR method. Increase in the percentage of items with DIF increases the power rates of both methods for a little amount. As a result; when the two methods are compared according to their power rates, it can be concluded that the most effective parameter for both methods is the sample size.

**Keywords:** Differential item functioning, MIMIC, LR.

**Advisor:** Assoc. Prof. Burcu ATAR, Hacettepe University, Department of Educational Science, Program in Measurement and Evaluation

## ETİK BEYANNAMESİ

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

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

beyan ederim.



Seçil ARSLAN

Sevgili anneme ve babama...



## TEŞEKKÜR

Bu süreçte içten, sıcak tavırlarıyla ve samimiyetiyle bana destek olan; her soruma cevap bulmamı sağlayan; kendisinden çok şey öğrendiğim; heyecanımı, telaşımı anlayan ve benden daha çok heyecanlanacak kadar özenle tezimin her aşamasında bana yardımcı olan sevgili danışmanım Doç. Dr. Burcu ATAR'a; ilk günden beri bana inanan, beni destekleyen, yardımını hiçbir zaman esirgemeyen ve akademik anlamda çok şey öğrenmemi sağlayan değerli hocam Prof. Dr. Selahattin GELBAL'a; yine akademik olarak onlardan çok şey öğrendiğim ve çok sevdiğim hocalarım Prof. Dr. Hülya Kelecioğlu'na, Doç. Dr. Nuri DOĞAN'a ve Doç. Dr. Duygu ANIL'a; tezime olan katkılarından dolayı Doç. Dr. Şeref TAN'a; bu sürecin neredeyse her aşamasını takip eden ve fikirleriyle bana çok yardımcı olan sevgili hocam Yrd. Doç. Dr. Derya ÇOBANOĞLU AKTAN'a; tezimi özenle okuyarak değerli fikirlerini paylaşan sevgili arkadaşım Dr. Kübra ATALAY KABASAKAL'a; yine tezime katkıları olan, heyecanımı ve sevincimi paylaşan, hayatımda önemli kararlar alırken benden bile çok düşünen, sevgisi ve emeği ile dostluğumuzu çoğaltan canım dostlarım Esin YILMAZ'a ve Eda DEMİRDÜZEN'e; hayatımda karşılaştığım en sıcak, en samimi, en eğlenceli, en yardımsever, en düşünceli insanların bir araya toplanmış olma ihtimalini gerçek yapan; yanlarında olduğum süre içinde bana adeta sevgi dolu bir masal yaşatan uzaktayken bile aslında hep yanımda olduklarını hissettiren canım dostlarım Demet BARAN BULUT'a, Berna AYGÜN'e, Bahar ŞAHİN'e ve Tuğba KARAKAYALI'ya; tanıdığım günden beri çok uzun yıllardır tanıyormuşçasına yakınlık gösteren, mutluluğumu kendi mutlulukları bilen, sürecin telaşını bana unutturan sevgili dostlarım Leman SÜRER ve Can SÜRER'e; her kararımın arkasında duran, bana inanan ve destek veren, her sevincimde benden daha çok sevinen, hayatının neredeyse tamamını benimle dolduran ve bundan mutluluk duyan canım annem Semra ARSLAN'a ve canım ablam Serap GENCAN'a; hayatımda "iyi ki" diye başladığım her cümlemin içinde olan; her sevincime, heyecanıma, derdime ortak olan; her şeye onunla daha pozitif baktığım; kendi işini bir kenara bırakıp bana yardımcı olmaya çalışan; kendimi çok değerli hissetmemi sağlayan; tüm yardımları ve bana verdiği mutluluk için ne kadar teşekkür etsem az olacağını bildiğim Bilal UĞURLU'ya teşekkür ederim.

# İÇİNDEKİLER

ÖZ .....	iii
ABSTRACT .....	v
TEŞEKKÜR .....	ix
İÇİNDEKİLER .....	x
TABLolar DİZİNİ .....	xii
ŞEKİLLER DİZİNİ .....	xiii
SİMGELER VE KISALTMALAR DİZİNİ .....	xiii
1. GİRİŞ .....	1
1.1. Problem Durumu .....	1
1.2. Araştırmanın Amacı ve Önemi .....	2
1.3. Problem Cümlesi .....	4
1.3.1. Alt Problemler .....	4
1.4. Sınırlılıklar .....	4
1.5. Araştırmanın Kuramsal Temeli .....	4
1.5.1. Yanlılık .....	4
1.5.2. Değişen Madde Fonksiyonu (DMF) .....	5
1.5.3. DMF Analizleri İçin Genel Bir Çerçeve .....	6
1.5.3.1. Odak Grupların Sayısı .....	7
1.5.3.2. Metodolojik Yaklaşım .....	7
1.5.3.3. DMF Türü .....	8
1.5.3.4. Madde Temizleme .....	9
1.5.3.5. TB DMF İçin MTK'ya Dayalı Olmayan Yöntemler .....	11
1.5.3.5.1. Lojistik Regresyon Yöntemi .....	12
1.5.3.5.2. ÇGÇN Yöntemi .....	14
1.5.3.6. TBO DMF İçin MTK'ya Dayalı Olmayan Yöntemler .....	15
1.5.3.7. MTK'ya Dayalı DMF Belirleme Yöntemleri .....	16
2. İLGİLİ ARAŞTIRMALAR .....	17
3. YÖNTEM .....	21
3.1. Araştırmanın Yöntemi .....	21
3.2. Çalışma Grubu .....	21
3.2.1. Çalışma Grubunun Özellikleri .....	21
3.2.2. Simülasyon Koşulları .....	21
3.3. Verilerin Türetilmesi .....	22
3.4. Kullanılan Programlar ve Verilerin Analizi .....	25
3.5. 1. Tip Hata ve Gücün Belirlenmesi .....	26
4. BULGULAR VE TARTIŞMA .....	27
4.1. Alt Problem 1'e İlişkin Bulgular ve Yorumlar .....	28
4.2. Alt Problem 2'ye İlişkin Bulgular ve Yorumlar .....	33
5. SONUÇ ve ÖNERİLER .....	38
5.1. Sonuçlar .....	38
5.2. Öneriler .....	39

5.2.1. Arařtırmanın Sonularına Dayalı Öneriler .....	40
5.2.2. Arařtırmacılar İin Öneriler .....	40
KAYNAKA .....	42
EKLER DİZİNİ .....	46
<a href="#">EK-1: SK1'e Ait GN Yöntemi Kod Dosyası .....</a>	
<a href="#">EK-2: SK1'e Ait SAS Kod Dosyası .....</a>	
<a href="#">EK-3: Etik Kurulu İzin Muafiyet Formu .....</a>	
ÖZGEMİŐ .....	55

## TABLolar DİZİNİ

Tablo 1.1: Geleneksel Değişen Madde Fonksiyonu (DMF) Belirleme Yöntemleri.....	11
Tablo 1.2: $\Delta R^2$ Değerlerinin Yorumlanması İçin Önerilen Sınıflama Kategorileri.....	14
Tablo 3.1: Simülasyon Koşulları Planı.....	22
Tablo 3.2: Madde Parametreleri.....	24
Tablo 4.1: Örneklem Büyüklükleri, Yetenek Dağılımları ve DMF'li Madde Yüzdelerine Göre 1. Tip Hata Oranları.....	27
Tablo 4.2: Örneklem Büyüklükleri, Yetenek Dağılımları ve DMF'li Madde Yüzdelerine Göre Güç Oranları.....	33

## ŞEKİLLER DİZİNİ

Şekil 1.1. Tek Biçimli DMF Bulunan Madde Grafiği.....	8
Şekil 1.2. Tek Biçimli Olmayan DMF Bulunan Madde Grafiği.....	9
Şekil 1.3.ÇGÇN Yöntemi Kullanılarak Y <sub>1</sub> Maddesindeki DMF'yi Belirleme.....	14

## SİMGELER VE KISALTMALAR DİZİNİ

**MTK:** Madde Tepki Kuramı

**DFA:** Doğrulayıcı Faktör Analizi

**DMF:** Değişen Madde Fonksiyonu

**TB DMF:** Tek Biçimli Değişen Madde Fonksiyonu

**TBO DMF:** Tek Biçimli Olmayan Değişen Madde Fonksiyonu

**ÇGÇN:** Çoklu Göstergeler Çoklu Nedenler

**MIMIC:** Multiple Indicators Multiple Causes

**LR:** Lojistik Regresyon

**SIBTEST:** Eş Zamanlı Madde Yanlılığı Testi

**1-PL:** Bir-Parametrelili Lojistik

**2-PL:** İki-Parametrelili Lojistik

**3-PL:** Üç-Parametrelili Lojistik

**SK:** Simülasyon Koşulu

# 1. GİRİŞ

Bu bölümde problem durumu, araştırmanın amacı ve önemi, problem cümlesi, alt problemler, sınırlılıklar ve araştırmanın kuramsal temeline yer verilmiştir.

## 1.1. Problem Durumu

Günümüzde modern toplumun tamamlayıcı bir parçası haline gelen testler, bireyler hakkında önemli kararlar alınmasında kullanılmaktadır. Testlerin hayatımızdaki artan etkisi, sosyal ve politik açıdan bir tartışma konusu haline gelmiştir. Bu denli önemli sonuçları olan testlerin geçerli olması vazgeçilmez bir gerekliliktir; çünkü geçerliliği düşük ölçümlere dayanan araştırma sonuçları yanıltıcıdır. Test ve madde yanlılığı ise testin geçerliliğini düşüren durumlardır. Yanlılık ölçme sonuçlarında sistematik hatalara sebep olabilir; bu da seçme ve yerleştirme yaparken sonuçları çarpıtabilir. Tüm bu açılardan bakıldığında testin bütün bireyler için adil olması ve belirli bir gruba karşı yanlı olmaması gerekir (Clauser, Mazor ve Hambleton, 1991).

Testte yer alan maddeler testi cevaplayan hiçbir gruba avantaj ya da dezavantaj sağlamamalıdır. Yansızlık, hem test geliştiricileri hem de test uygulayıcıları için ahlaki bir zorunluluktur ve bu 30 asırdan fazla süredir var olan bir temel ilkedir (Holland ve Wainer, 1993). Bir testin yanlı olmaması için, testte yer alan maddelerin tüm bireyler için benzer yapıyı ölçmeleri ve benzer ölçme özelliklerine sahip olmaları gerekmektedir. Gruplar arasında gerçek yetenek seviyesi farklılıkları olabilir; ancak böyle olup olmadığını geçerli bir şekilde belirlerken maddeler, yetenek seviyelerine göre eşleştirilen bireylerden oluşan grupta benzer şekilde çalışmalıdır. Örneğin; bir konuda aynı yetenek seviyesine sahip bireylerin cinsiyetlerine, etnik kökenlerine ya da diğer grup üyeliklerine bakılmaksızın, o konuyla ilgili bir testte yer alan maddelere doğru cevap verme olasılıkları benzer olmalıdır (Woods, 2009).

Test maddeleri, ölçülmesi amaçlanan yapının yanında ölçülmesi amaçlanmayan yapılar da içerdiğinden yanlı olabilir. Madde, ilgili olduğu temel faktörün dışında ikinci ya da çok sayıda faktörle bağlantı kurabilir. Yapı ile ilişkisi olmayan bu faktörler bireylerin performansını etkileyebilir. Bu durum, test yanlılığı olarak bilinir. Yanlılık analizinde kullanılan istatistiksel bir teknik olan değişen madde fonksiyonu

(DMF), son zamanlardaki arařtırmaların büyük bir kısmına konu olmuřtur (Zumbo, 1999).

Bir madde, cevaplayanların bir grubunda diđer grupta olduđundan farklı bir řekilde işleyebilir. Üzerinde çalıřılan grup odak grup, madde üzerindeki performanslarına bakılarak karşılařtırma yapılan grup ise referans grup olarak adlandırılır (Holland ve Wainer, 1993; French ve Miller, 1996). Test maddeleri odak ve referans gruplar arasında farklı bir řekilde işlese, bu gruplardan elde edilecek olan puanlar karşılařtırılmaz, eđer karşılařtırılırsa bu deđerlendirme adil olmaz. DMF yöntemlerinin sağladıđı bilgiler madde güçlüğü ve ayırt edicilik indislerinde deđerişimlere sebep olan özelliklerin anlaşılmasında yol göstericidir (Holland ve Wainer, 1993). Bu bakımdan DMF analizleri tüm test geliřtiricilerini ve testi alan bireyleri ilgilendirmektedir. Birçok testin kapsam ve yapı geçerliliđini belirlemede DMF analizlerinden yararlanılır (Holland ve Wainer, 1993). Sonuç olarak DMF belirleme yöntemleri test geliřtirme ve deđerlendirme sürecinin tamamlayıcı bir parçası haline gelmiřtir.

## **1.2. Arařtırmanın Amacı ve Önemi:**

Bu çalıřmayla, psikometrik özellikleri güçlü testler oluşturmak ve testlerle gerçekteřtirilen seçme ve yerleřtirme işlemlerinde yansızlıđı sağlamak üzere kullanılan DMF belirleme yöntemlerine yönelik arařtırmalara katkıda bulunmak amaçlanmıřtır. Testlerde DMF gösteren maddelerin belirlenmesi için birçok yöntem geliřtirilmiřtir. İki kategorili puanlanan maddeler için kullanılan DMF belirleme yöntemlerinden bazıları; Mantel-Haenszel yöntemi (Holland ve Thayer, 1988), standartlařtırma yöntemi (Dorans ve Kulick, 1986), lojistik regresyon yöntemi (Swaminathan ve Rogers, 1990), SIBTEST yöntemi (Shealy ve Stout, 1993), madde tepki kuramına (MTK) dayanan ki-kare testi (Lord,1980), madde tepki kuramı olabilirlik oranı (MTK-OO) testi (Thissen ve Steinberg ve Wainer, 1988) ve çoklu göstergeler çoklu nedenler (ÇGÇN) yöntemidir (Finch, 2005; Oort, 1998). ÇGÇN yöntemi, bahsedilen diđer yöntemlere göre oldukça yenidir ve özellikle iki kategorili verilerde ÇGÇN yöntemiyle alan yazında az sayıda çalıřma bulunmaktadır (Finch, 2005). Bu nedenle, ÇGÇN yönteminin DMF belirlemede hangi kořullar altında daha etkili sonuçlar sağlayacađının arařtırılmasının önemli olduđu düşünölmektedir.

ÇGÇN yöntemi, sürekli ya da süreksiz çoklu gruplama değişkenleriyle DMF analizi yapmaya izin verdiğinden, tek bir kategorik değişkenle DMF analizi yapmaya izin veren geleneksel DMF belirleme yöntemlerine göre daha kullanışlıdır. Fleishman, Spector ve Altman (2002) ÇGÇN yöntemi gibi gizil değişken kullanan yaklaşımların, tüm bir sınavın performansının anlaşılmasında DMF'nin etkisini kontrol etmeye yardımcı olma olasılığından söz etmişlerdir. Buna ek olarak, bu yöntemle birlikte, ikiden fazla grup için DMF'nin varlığını denetlemek mümkündür ve çoklu alt yapı değişkenleri (kategorik ve süreklilerin her ikisini de içeren) analize dahil olabilir. Muthen (1988) ÇGÇN yönteminin sadece DMF'nin ölçülmesine izin vermediğini aynı zamanda alt yapı değişkenleri ve örtük özellik arasındaki ilişkinin tam olarak araştırılmasını sağladığını belirtmiştir. ÇGÇN yönteminin DMF araştırmalarına sağlayacağı bu avantajlar düşünüldüğünde ve Türkiye'de eğitim alanında ÇGÇN yönteminin kullanıldığı bir araştırmaya rastlanmadığından ÇGÇN yöntemi ilgi çekici bir araştırma konusu haline gelmiştir.

ÇGÇN yönteminin DMF içeren maddeleri belirlemede nasıl sonuçlar verdiğinin incelenmesi için daha geleneksel bir DMF belirleme yöntemi referans alınarak karşılaştırmalar yapılması amaçlanmıştır. Bu amaçla, çalışmada ÇGÇN yöntemi ile lojistik regresyon (LR) yönteminin performansları çeşitli koşullar altında karşılaştırılmıştır. Bu koşullar; örneklem büyüklükleri, grupların yetenek dağılımları ve testteki DMF gösteren maddelerin yüzdeleridir.

Yanılılık çalışmalarında kullanılmak üzere seçilen DMF belirleme yöntemlerinin kullanılan teste ve testin uygulandığı grupların özelliklerine uygun olması büyük bir önem taşımaktadır. Bu çalışmayla iki kategorili verilerde farklı koşullar ele alınarak ÇGÇN yönteminin incelenen koşullardan hangisi için daha doğru sonuçlar verdiği araştırılmak istenmiştir. Finch (2005) yaptığı çalışmada iki kategorili MTK modellerinden 2-PL modele veya 3-PL modele göre, iki farklı örneklem büyüklüğü (600 birey ve 1000 birey), iki farklı test uzunluğu (20 ve 50 madde), iki farklı DMF gösteren madde yüzdesi (%0 ve %15) koşulları altında üretilen veri setleriyle ÇGÇN yönteminin performansını incelemiştir. Bu çalışmada ele alınan koşullar Finch'in (2005) çalışmasında ele alınan koşullardan test uzunluğu, DMF gösteren madde yüzdesi ve gruplar arası yetenek dağılımı bakımından farklıdır. Çalışmada kullanılan veriler 3-PL MTK modeline göre üretilmiş, test uzunluğu gerçek uygulamalara benzerlik göstermesi için 30 madde olarak alınmıştır. ÇGÇN



yönteminin farklı örneklem büyüklüklerinde benzer çalışıp çalışmadığı önemli bir sorudur (Wang ve Shih, 2010). Bu nedenle çalışmada farklı örneklem büyüklüklerinin karşılaştırılması amaçlanmıştır. Çalışmada ele alınan koşullardan bir diğeri DMF gösteren madde yüzdesi olup, %10 ve %20 olmak üzere iki düzeyde ele alınmıştır.

### **1.3. Problem Cümlesi:**

ÇGÇN ve LR DMF belirleme yöntemlerinin 1. tip hata oranları ve güçleri; örneklem büyüklüğüne, grupların yetenek dağılımlarına ve DMF gösteren madde yüzdesine göre nasıl değişmektedir?

#### **1.3.1. Alt Problemler:**

1) ÇGÇN ve LR DMF belirleme yöntemlerinin 1. tip hata oranları örneklem büyüklüklerine, yetenek dağılımlarına ve DMF gösteren madde yüzdelere göre nasıl değişmektedir?

2) ÇGÇN ve LR DMF belirleme yöntemlerinin güçleri örneklem büyüklüklerine, yetenek dağılımlarına ve DMF gösteren madde yüzdelere göre nasıl değişmektedir?

### **1.4. Sınırlılıklar:**

Sonuçların genellenmesi, bu çalışmada ele alınan iki farklı örneklem büyüklüğü, iki farklı yetenek dağılımı ve iki farklı DMF içeren madde yüzdesi ile sınırlıdır.

Bu benzetim çalışmasında referans grup büyüklüğü ve odak grup büyüklüğü arasındaki oran 1:1 olarak alınmıştır. Ancak, gerçek uygulamalarda bu iki grubun örneklem büyüklükleri oranıyla ilgili daha farklı durumlar söz konusu olabilir. Ayrıca bu çalışmada testte yer alan madde sayısı 30 olarak alınmıştır. Gerçek uygulamalarda daha kısa veya daha uzun testlerle karşılaşmak mümkündür.

### **1.5. Araştırmanın Kuramsal Temeli**

#### **1.5.1. Yanlılık**

Test yanlılığı, bir testin belli bir grubun üyelerini nasıl ölçtüğüyle ilgili yapılan sistematik hata ya da testin geçerliğin düşmesi olarak tanımlanır (Camilli ve Shepard, 1994).

Birçok araştırmacı (Linn ve Drasgow, 1987; Linn, Levine, Hastings ve Wardrop, 1981; ve Shepard, Camilli ve Averill, 1981) madde yanlılığını temelde benzer

şekillerde tanımlamışlardır: Farklı gruplardan eşit yetenek düzeyine sahip bireylerin maddeyi doğru cevaplandırma olasılıkları eşit değilse, bu madde yanlıdır. Shephard, Camilli ve Averill (1981) yanlılığı, bir grubu diğerinden daha çok etkileyen bir tür geçersizlik olarak tanımlamışlardır. Bu tanımlamalardan ilki farklılık hesaplamaya dayanırken, ikincisi ise farklılığın ötesinde, değerlendirme anlamı içerir ve yanlılığın bir grup üzerindeki etkisinden bahseder (Holland ve Wainer, 1993).

Bir grubun test puanlarının diğer gruptan daha düşük olması, ölçülen temel yetenek açısından gruplar arasındaki gerçek farklılıktan ya da testteki yanlılıktan kaynaklanabilir. Aslında yanlılık araştırmalarının amacı, grup farklılıklarının sebebinin gerçek ya da ölçme sürecinden kaynaklı olup olmadığını açığa çıkarmaktır (Camilli ve Shepard, 1994; Hamilton, 1999).

Yanlılık analizlerinin uygun şekilde anlaşılması için madde etkisi ve yanlılık kavramların farkından bahsedilmelidir. Madde etkisi, farklı gruplardan bireylerin bir maddeye doğru şekilde yanıt verme olasılıklarının farklı olduğu durumlarda görülür; çünkü madde tarafından ölçülen yetenek açısından gruplar arasında gerçek farklılıklar vardır. Bireyler genellikle testlerin ölçtüğü özelliklerde farklılık gösterdikleri için madde etkisi sık rastlanan bir durumdur. Örneğin, standartlaştırılmış matematik testlerinin ortalamalarında erkeklerin puanları genellikle kızlarınkinden daha yüksektir (Zumbo,1999). Camilli (2006), farklılığın var olmasının, yanlılığın olduğu anlamına gelmediğini; çünkü var olan farklılığın gerçek yetenek farklılığı olabileceğini belirtmiştir (Wiberg, 2007).

Literatür incelendiğinde, yanlılığın çok erken zamanlarda sosyal ve istatistiksel olmak üzere iki farklı anlamda kullanıldığı görülmektedir. Yanlılığı belirlemek için çeşitli istatistiksel yöntemler geliştirilmiştir ve değişen madde fonksiyonu (DMF) yanlılık analizlerinde bir standart haline gelmiştir (Holland ve Wainer, 1993; Kristjansson, Aylesworth, McDowell ve Zumbo, 2005).

### **1.5.2. Değişen Madde Fonksiyonu (DMF)**

Test yanlılığının değerlendirmesi, testin yapısal geçerliliğini sağlamak açısından önemlidir. DMF değerlendirmesi ise bu sürecin önemli bir parçasıdır. DMF, tüm yetenek seviyeleri kontrol edildikten sonra, farklı gruplardan cevaplayıcıların bir

madde üzerindeki doğru yanıt olma olasılıkları farklılaştığında görülür (Crane, Belle ve Larson, 2004; Mazor, Kanjee ve Clauser, 1995).

Bir madde yanlı değilse, bu maddeye verilen cevaplar sadece maddenin ölçmeye çalıştığı temel özelliklerle ilişkili olacaktır. Eğer madde yanlılığı söz konusuysa, maddeye verilen cevaplar ölçülmek istenilen özelliğin yanı sıra başka birtakım faktörlerle de ilişkili olacaktır. Bu durumda maddede DMF'nin varlığından söz edilir. Bir testte çok sayıda DMF'li madde bulunuyorsa, bu maddeler test puanlarına dayalı sonuçları ve testin yapı geçerliliğini tehdit eder (Crane, Belle ve Larson, 2004).

Bahsedildiği üzere DMF, madde yanlılığı için gerekli olan ancak yeterli olmayan bir durumdur. Buna göre, eğer bir madde için DMF söz konusu değilse, madde yanlılığı görülmez. Ancak, DMF'nin varlığından bahsediliyorsa, bu durum madde yanlılığını açıklamak için yeterli değildir; madde yanlılığını ortaya çıkarmak için ek analizlere (içerik analizi, gözlemsel değerlendirme vb.) başvurulmalıdır (Camilli ve Shepard, 1994; Zumbo, 1999).

Madde yanlılığı ve DMF'nin her ikisini de uygun olmayan madde içeriği ya da yapısından ayırt etmek gerekir. Eğer bir madde herkes açısından olumsuz ise yanlı olarak saptanmayacaktır, basit şekliyle yansızlık tüm grupların eşit şekilde etkilenmiş olması demektir (Zumbo, 1999).

DMF belirleme yöntemleri ilk olarak Cardall ve Coffman (1964), Cleary ve Hilton (1968), Angoff ve Ford (1973), Lord (1976) ve Scheuneman (1979) tarafından öne sürülmüştür. Tarihsel önemine karşın bu yöntemler sonrasında pek kullanılmamıştır. Bunlardan "Geleneksel Metodlar" olarak bahsedilir. DMF belirleme yöntemlerinin daha önceki gözden geçirildiği çalışmalar arasında Ironson ve Subkoviak (1979), Rudner, Getson ve Knight (1980), Shepard, Camilli ve Averill (1981) ve Berk (1982) yer almaktadır. Daha yeni çalışmalar arasında Millsap ve Everson (1993), Camilli ve Shepard (1994), Clauser ve Mazor (1998), Penfield ve Camilli (2007) ve Osterlind ve Everson (2009) bulunmaktadır (Magis, Beland, Teurlinckx ve Boeck, 2010).

### **1.5.3. DMF Analizleri İçin Genel Bir Çerçeve**

Magis, Beland, Teurlinckx ve Boeck (2010) DMF belirleme sürecini 4 temel ögeye dayandırmıştır, bu ögerler şunlardır: odak grupların sayısı, metodolojik yaklaşım

(MTK'ya dayalı, MTK'ya dayalı olmayan), DMF türü (tek biçimli ya da tek biçimli olmayan) ve madde temizlemenin kullanılıp kullanılmadığı.

### **1.5.3.1. Odak Grupların Sayısı**

DMF analizleri genellikle bir referans grup ile bir odak grubun karşılaştırılmasına dayanır. Pratikte birden fazla odak grup ile analizler gerçekleştirilebilir. Örneğin, bu çok çeşitli okulların öğrencilerinin performansları karşılaştırılırken bir çeşit okul referans alınarak yapılan çalışmalarda görülebilir. Diğer bir durumda, gruplardan hiçbiri bir referans grup değildir; ancak karşılaştırmada bir tanesiyle hala ilgileniliyordur. Eğer bir referans grup yoksa, tüm odak grup ve referans grup arasındaki ya da bütün gruplar arasındaki ikili karşılaştırma uygulamaları yapmak ortak bir yaklaşımdır. Ancak çoklu grup karşılaştırmalarının birçok dezavantajı vardır. İlk dezavantajı, anlamlılık seviyesi için kontrol gerektirmesidir. Bu kontrol Bonferroni düzeltmesi kullanarak sağlanabilir. Çoklu karşılaştırmaların ikinci dezavantajı ise, DMF'li maddeleri belirlemedeki güç oranının genellikle bütün grupları eşzamanlı olarak karşılaştıran tek bir analize göre gücünün daha düşük olmasıdır (Magis, Beland, Teurlinckx ve Boeck, 2010).

### **1.5.3.2. Metodolojik Yaklaşım**

DMF belirleme yöntemlerini, MTK'ya dayalı yöntemler ve MTK'ya dayalı olmayan yöntemler olmak üzere iki grupta incelemek mümkündür (Camilli ve Shephard, 1994). MTK'ya dayalı yöntemler arasında olabirlik oranı testleri (Thissen ve Steinberg ve Wainer, 1988), Lord'un ki-kare testi (Lord,1980) ve Raju'nun alan ölçümleri yöntemi (Raju, 1988, 1990) sayılabilir. MTK'ya dayalı olmayan yöntemler arasında ise; Mantel-Haenszel (Mantel ve Haenszel, 1959), lojistik regresyon (Swaminathan ve Rogers, 1990), standartlaştırma yöntemi (Dorans ve Kluick, 1986), SIBTEST (Roussos ve Stout,1996) ve Breslow-Day testi (Breslow ve Day,1980) yer alır.

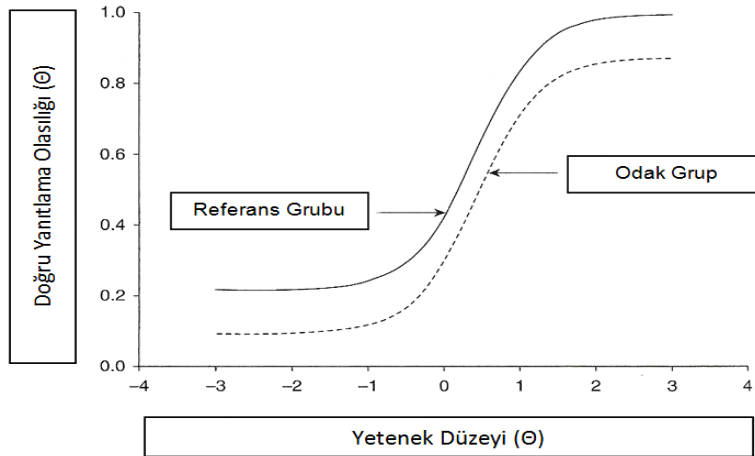
DMF belirleme yöntemlerinin bir başka sınıflaması ise, gözlenen puan ve örtük puanlar yaklaşımıdır. Mantel-Haenszel, standartlaştırma yöntemi, ve lojistik regresyon gözlenen puan yöntemleri arasında yer almaktadır. Olabirlik oranı testleri, SIBTEST, ki kare, maddenin ve testlerin farklı fonksiyonlaşması (DFIT) ve test karakteristik eğrilerinin alanlarının karşılaştırmaları ise örtük puan yöntemleri

arasında yer almaktadır (Potenza ve Dorans, 1995; akt. Atalay, Gök, Kelecioğlu ve Arslan, 2012).

### 1.5.3.3. DMF Türü

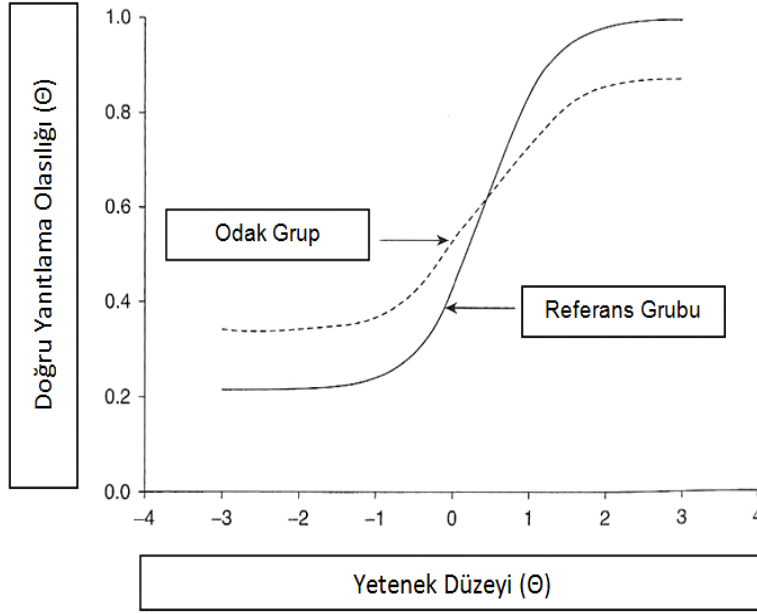
Değişen madde fonksiyonu, tek biçimli (uniform) ve tek biçimli olmayan (non-uniform) değişen madde fonksiyonu olmak üzere iki farklı şekilde görülür (Camili ve Shephard, 1994).

**1. Tek Biçimli (Uniform) DMF:** Aynı yetenek seviyelerindeki iki grubun ilgili maddeyi doğru cevaplama olasılığındaki farklılık sabit ise tek biçimli DMF meydana gelir. Yani bireylerin yetenek seviyeleri değiştikçe maddeyi doğru cevaplama olasılıklarındaki farklılık da aynı düzeyde değişmektedir. Örneğin; alt grup olarak cinsiyet alındığında, erkeklerin maddeyi doğru cevaplama olasılığı farklı yetenek düzeylerinin her noktasında kızlara göre daha yüksek ise, yani düzlemde bir kesişim söz konusu değilse bu durumda tek biçimli DMF'den bahsedilir (Mellenberg, 1982; akt. Bekçi, 2007). Madde performansı doğru cevap olasılığı kullanılarak tanımlandığında tek biçimli DMF, olasılık oranının yetenek dağılımı boyunca sabit olduğu durumlarda ortaya çıkar (Camili ve Shephard, 1994, Penfield, 2003). Şekil 1.1'de tek biçimli DMF bulunan maddenin madde karakteristik eğrisi verilmiştir.



Şekil 1.1. Tek Biçimli DMF Bulunan Madde Grafiği (Osterlind ve Everson, 2009)

**2. Tek Biçimli Olmayan (Nonuniform) DMF:** Gruplar arasında maddeyi doğru yanıtlama olasılığındaki farklılığın, tüm yetenek seviyelerinde tutarlı olmadığı durumlarda tek biçimli olmayan DMF meydana gelir (Camilli ve Shepard, 1994). Yani bireylerin yetenek seviyeleri değiştikçe maddeyi doğru cevaplama olasılıkları da değişmektedir. Bir başka deyişle; yetenek seviyeleri ile grup kategorileri arasında bir kesişim varsa tek biçimli olmayan DMF olduğu söylenebilir (Mellenberg, 1982; akt. Bekçi, 2007). Şekil 1.2'de tek biçimli olmayan DMF bulunan maddenin madde karakteristik eğrisi verilmiştir.



**Şekil 1.2. Tek Biçimli Olmayan DMF Bulunan Madde Grafiği (Osterlind ve Everson, 2009)**

DMF türü MTK modeli ile doğrudan ilişkilidir. Örneğin, 1PL model ayırt edicilik parametresini içermediği için yalnızca TB (tek biçimli) DMF'yi belirlemede kullanılabilir. 2-PL ve 3-PL modeller de bu amaçla kullanılabilir. Bu modeller ayrıca TBO (tek biçimli olmayan) DMF'yi belirlemek için de kullanılabilirler; çünkü ayırt edicilik parametresini içermektedirler (Magis, Beland, Teurlinckx ve Boeck, 2010).

#### 1.5.3.4. Madde Temizleme

Bir testte bir ya da daha fazla sayıda DMF içeren madde bulunması testteki diğer maddeler için DMF araştırmalarının sonuçlarını etkileyebilir. Örneğin; DMF içermeyen maddeler yanlışlıkla DMF'li olarak saptanabilir. Bu durum testin 1. tip

hatasını istenmeyen şekilde arttırır (Clauser, Mazor ve Hambleton, 1993). Özetle, DMF içermediği bilinen maddelerden oluşan öncül madde takımının içinde DMF'li maddelerin belirlenmesi bu durumu tam olarak açıklamaktadır. Bahsedilen sorunun üstesinden gelebilmek için birçok araştırmacı (Candell ve Drasgow, 1988; Clauser, Mazor ve Hambleton, 1993; Fidalgo, Mellenbergh ve Muñiz, 2000; Holland ve Thayer, 1988; Lautenschlager ve Park, 1988; Wang ve Su, 2004; Wang ve Yeh, 2003) DMF'li maddeleri tekrarlı olarak elemeyi önermiştir ve bu işlem, madde temizleme olarak adlandırılmıştır. Madde temizleme işleminin basamakları aşağıdaki şekildedir:

1. Tüm maddeler DMF içerip içermemeleri bakımından tek tek test edilmelidir.
2. 1. basamaktan elde edilen sonuçlara göre DMF içeren maddelerden oluşan bir grup belirlenir.
3. İlk durumda 1. basamaktan hiç DMF içeren madde tespit edilmemişse ya da yapılan analizin sonuçları bir önceki analizin sonuçlarıyla aynı ise 6. basamağa gidilir. Bu durumun aksine, 1. basamakta gerçekleştirilen analizlerden DMF içeren madde ya da maddeler elde edildiyse ve ya yapılan bir önceki DMF analizlerinden farklı bir durum söz konusuysa 4. basamağa geçilir.
4. 2. basamakta oluşturulan gruptaki, DMF içerdiği belirlenen maddelerden DMF içerdiğinden emin olunan maddeler testten çıkarılır ve geriye kalan maddeler DMF içerip içermemeleri bakımından tek tek test edilir.
5. 4. basamakta yapılan analizlerde DMF içeren maddeler belirlenirse 3. basamağa geri dönülür ve aynı işlemler tekrarlanır.
6. İşlemler durdurulur.

Bahsedilen tekrarlama işlemleri için, doğruluğundan emin olunan bir durdurma kuralı yoktur. İşlemleri durdurmak için araştırmacı yapacağı tekrarların sayısını belirlemelidir ve bu sayıya ulaştığında işlemleri durdurmalıdır.

Araştırmacı isterse 2. basamakta işlemlerini durdurabilir ancak bu durumda geri kalan tüm maddeler DMF'siz olarak varsayılacaktır. Bu ise, yanıltıcı sonuçlara sebep olabilir (Clauser, Mazor ve Hambleton, 1993; Magis, Beland, Teurlinckx ve Boeck, 2010; Wang ve Su, 2004).

Tablo 1.1 odak grup sayılarına, metodolojik yaklaşımlara ve DMF türüne göre geleneksel DMF belirleme yöntemlerini listeler. Bu yöntemlerden her biri temizleme yapmadan ya da temizleme işlemi yapılarak kullanılabilir.

**Tablo 1.1: Geleneksel Değişen Madde Fonksiyonu (DMF) Belirleme Yöntemleri**

Yapı	DMF Etkisi	Grup sayısı	
		2	>2
MTK'ya dayanmayan	TB	Mantel-Haenszel Standartlaştırma SIBTEST	İkili karşılaştırmalar Genelleştirilmiş MH
MTK'ya dayanmayan	TBO	Lojistik Regresyon Lojistik Regresyon Breslow-Day TBO.MH TBO.SIBTEST	İkili karşılaştırmalar
MTK'ya dayalı	TB	OO Lord Raju	İkili karşılaştırmalar Genelleştirilmiş Lord
MTK'ya dayalı	TBO	OO Lord Raju	İkili karşılaştırmalar Genelleştirilmiş Lord

Not: TBO.MH, Tek biçimli olmayan DMF için düzeltilmiş Mantel-Haenszel; TBO.SIBTEST, Tek biçimli olmayan DMF için düzeltilmiş SIBTEST; OO, Olabilirlik oranı yöntemi (Magis, Beland, Teurlinckx ve Boeck, 2010).

#### 1.5.3.5. TB DMF İçin MTK'ya Dayalı Olmayan Yöntemler

Birçok geleneksel yöntem, MTK'ya dayalı olmayan yöntemler sınıfına aittir ve TB DMF'yi belirlemek üzere tasarlanmıştır. Mantel-Haenszel (MH), standartlaştırma ve SIBTEST yöntemleri bu gruptadır. Camilli ve Shepard (1994)'ün belirttiği üzere lojistik regresyon (LR) yöntemi, MTK'ya dayalı olan ve MTK'ya dayalı olmayan yöntemler arasında bir köprü görevi görmektedir.

MH (Mantel ve Haenszel, 1959), DMF çalışmalarında oldukça yaygın olarak kullanılan bir yöntemdir. Bu yöntem, toplam test puanlarına (ya da puanların ortalamalarına) bakmak koşuluyla madde cevapları ile grup üyeliği arasında bir ilişki olup olmadığını test etmeyi amaçlar (Magis, Beland, Teurlinckx ve Boeck, 2010). Bu yöntemde grupların puanları eşleştirilir ve belirli bir testteki tüm maddeler hakkında iki cevaplayıcı grubunun performansı karşılaştırılır. İki grup arasında madde performansındaki her bir farklılaşma, değişen madde fonksiyonunu gösterir (Holland ve Wainer, 1993).

Standartlaştırma yöntemi (Dorans ve Kulick, 1986), MH yöntemine yakın bir yaklaşıma dayanmaktadır. Standartlaştırma yönteminde, her grubun doğru cevaplarının oranı ve toplam test puanlarının her bir değeri için karşılaştırma



yapılır. Yani, standartlaştırılmış  $p$  farkı test istatistiğinin sonucudur ve gruplar arası başarı farklarının ağırlıklı bir ortalaması olarak görülebilir.

SIBTEST, standartlaştırma yönteminin genelleştirilmiş bir hali olarak görülebilir. SIBTEST istatistiği standartlaştırma yöntemine göre çok sayıda yapısal avantaja sahiptir. Bu yöntemle, her bir maddeyi ayrı ayrı test etmek yerine madde setinin DMF analizini yapabilir.

Lojistik regresyon yaklaşımında (Swaminathan ve Rogers, 1990), maddeye doğru cevap verme olasılığı toplam test puanlarına, grup üyeliğine ve bu ikisi arasındaki ilişkiye dayalıdır. TB DMF temel grup etkisinin test edilmesi ile belirlenebilirken; TBO DMF etkileşim etkisinin test edilmesiyle belirlenebilir.

Çoklu gruplar için yukarıda belirtilen yöntemlerden hiçbiri (MH, standartlaştırma, SIBTEST ve LR) odak ve referans grupların her biri ya da tüm gruplar arasındaki ikili karşılaştırmalarda kullanılamaz. MTK'ya dayalı olmayan yöntemler arasında yer alan MH, çoklu gruplar için "genelleştirilmiş Mantel-Haenszel yöntemi" olarak Penfield (2001) tarafından ortaya atılmıştır. LR yöntemi de regresyon eşitliklerinde çoklu grup belirleyicileri ile çoklu gruplar için genelleştirilebilir (Millsap ve Everson, 1993).

#### 1.5.3.5.1. Lojistik Regresyon Yöntemi

DMF'yi ortaya çıkarmak için lojistik regresyonun kullanımı, Rogers (1989), Tian, Pang ve Boss (1990), Swaminathan (1993), Rogers ve Swaminathan (1993) tarafından incelenmiştir. Bu yöntem MH tekniğinin genellemesidir; ancak tek biçimli olmayan DMF'yi tanımlaması açısından MH istatistiğine göre daha üstündür.

LR yöntemi, bir maddeyi doğru yanıtlama olasılığının tahmini için aşağıdaki gibidir:

$$P(u = 1|\theta) = \frac{e^{(\beta_0 + \beta_1\theta)}}{1 + e^{(\beta_0 + \beta_1\theta)}} \quad (1)$$

Denklemden  $u$  maddeye verilen yanıt,  $\theta$  bireyin gözlenen yeteneği,  $\beta_0$  kesişim parametresi,  $\beta_1$  eğim parametresidir. Bu verilen, sahip olunan bağımsız değişkenlerden iki kategorili bağımlı bir değişkenin kestirimi için kullanılan standart lojistik regresyon denklemdir. LR modeli, değişen madde fonksiyonunu belirlemede ilgilenilen iki grup için ayrı eşitlikler olarak belirtilebilir:

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

Burada  $u_{ij}$ ,  $j$  grubundaki  $i$ . bireyin maddeye verdiği yanıt,  $\beta_{0j}$  kesişim parametresi,  $\beta_{1j}$   $j$ . grup için eğim parametresi,  $\theta_{ij}$   $j$ . gruptaki  $i$ . bireyin yeteneğidir. Burada DMF için kabul edilmiş tanım şu şekildedir: Eğer bireyler aynı yetenek düzeyindeyse ve farklı grupların maddeyi doğru cevaplama olasılıkları farklıysa bu madde DMF gösterir. Bunu şu şekilde açıklayabiliriz: İki grup için Lojistik Regresyon eğrileri aynıysa yani  $\beta_{01} = \beta_{02}$  ve  $\beta_{11} = \beta_{12}$  ise, DMF'nin varlığından söz edilemez. Ancak  $\beta_{11} = \beta_{12}$  ve  $\beta_{01} \neq \beta_{02}$  olduğunda, lojistik regresyon eğrileri paralel olduğundan TB DMF görüldüğü sonucuna varılabilir. Eğer  $\beta_{01} = \beta_{02}$  ve  $\beta_{11} \neq \beta_{12}$  ise, eğriler paralel olmadığı için TBO DMF'nin varlığından söz edilebilir (Rogers ve Swaminathan, 1990).

Lojistik regresyon yöntemi ile DMF'yi araştırmada ise üç alt model kullanılır:

$$z = \beta_0 + \beta_1 X,$$

$$z = \beta_0 + \beta_1 X + \beta_2 G,$$

$$z = \beta_0 + \beta_1 X + \beta_2 G + \beta_3 GX, \quad (3)$$

Burada  $X$  test puanlarını,  $G$  grubu ve  $GX$  ise grup ve test puanları kesişimini temsil eden değişkendir. DMF'nin varlığını test etmek için modelde  $X$  değişkeninin anlamlı olması modelin geçerliğini,  $G$  değişkeninin modeldeki anlamlılığı tek biçimli DMF'yi,  $GX$  değişkeninin modeldeki anlamlılığı ise tekbiçimli olmayan DMF'yi gösterir (Gierl, Jodoin ve Ackerman, 2000; akt. Bekçi, 2007).

Lojistik regresyon yönteminin üç aşamasında da etki büyüklüğü kestiricisi elde edilir. Etki büyüklüğü  $R^2$  değeri ile hesaplanır ve bu değer iki ve çok kategorili maddeler için kullanılabilir. Çalışılan maddenin TB DMF gösterip göstermediğini test etmek için ikinci alt modelin uyumu ile birinci alt modelin uyumları karşılaştırılır. İki modelin  $R^2$  değerlerinin farkı DMF miktarının yorumlanmasında kullanılan etki büyüklüğünü verir. Çalışılan maddenin TBO DMF gösterip göstermediğini test etmek içinse üçüncü alt modelin uyumu ile ikinci alt modelin uyumları karşılaştırılır. Aynı şekilde, bu iki modelin  $R^2$  değerlerinin farkı DMF miktarının yorumlanmasında kullanılan etki büyüklüğünü verir (Zumbo, 1999). Etki büyüklüğü miktarının,  $\Delta R^2$ , yorumlanması için Zumbo ve Thomas (1996) ile Jodoin ve Gierl (2001) Tablo 1.2'de verilen iki ayrı sınıflamayı önermişlerdir.

**Tablo 1.2:  $\Delta R^2$  Değerlerinin Yorumlanması İçin Önerilen Sınıflama Kategorileri**

Düzye	Zumbo ve Thomas (1996)	Jodoin ve Gierl (2001)	DMF Etki Büyüklüğü
A	$\Delta R^2 < 0.13$	$\Delta R^2 < 0.035$	Yok veya ihmal edilebilir düzeyde
B	$0.13 \leq \Delta R^2 < 0.26$	$0.035 \leq \Delta R^2 < 0.070$	Orta düzeyde
C	$\Delta R^2 \geq 0.26$	$\Delta R^2 \geq 0.070$	Yüksek düzeyde

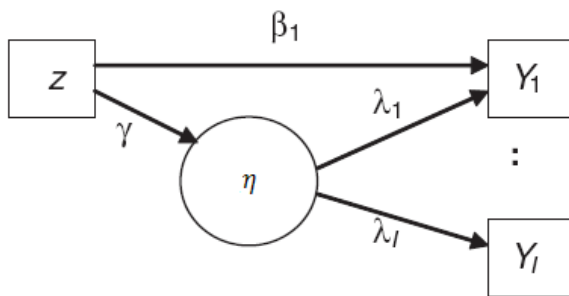
### 1.5.3.5.2. ÇGÇN Yöntemi

Doğrulamalı faktör analizi (DFA), tek biçimli ve tek biçimli olmayan DMF'nin belirlenmesinde kullanılan bir yaklaşımdır. DFA yaklaşımının, DMF'nin belirlenmesinde bir potansiyele sahip olduğu yapılan çalışmalarla desteklenmiştir; çünkü bu model ikincil bir faktör üzerinde grup farklılıklarının kıyaslanmasına izin verir. DFA'nın bu bağlamda kullanımı için değişim indeksleri, çok gruplu DFA ve ÇGÇN yöntemlerinin kullanımını içeren birçok yaklaşım geliştirilmiştir (Finch, 2005).

ÇGÇN yöntemleri pratikte belirli maddelerde TB DMF'nin varlığını kontrol etmek için kullanılmışlardır. Araştırmacılar DMF'nin yönünü ve büyüklüğünü saptamak için bu yaklaşımdan elde edilen model parametre hesaplarının MTK türünde madde güçlüğü değerlerine nasıl dönüştürülebileceğini göstermişlerdir. DMF bağlamında ÇGÇN yöntemi aşağıdaki gibidir:

$$y_i^* = \lambda_i \eta + \beta_i z_k + \varepsilon_i, \quad (4)$$

Yukarıdaki  $y_i^* = i$  değişkeni gizil yanıtı ( $y_i^* > \tau_i$ , gözlenen bir değişken,  $y_i = 1$ ;  $\tau_i$ , madde güçlüğüne ilişkin sınır parametresi),  $\eta$  =gizil özellik,  $\lambda_i = i$  değişkeni için faktör yükü ve  $\varepsilon_i$  =tesadüfi hatadır;  $z_k$  =grup üyeliğini gösteren bir yapay değişkendir;  $\beta_i$  =grup değişkeni ve yanıt arasındaki eğimdir.

**Şekil 1.3. ÇGÇN Yöntem Kullanarak  $Y_1$  Maddesindeki DMF'yi Belirleme (Wang, Shih ve Yang, 2009)**

ÇGÇN, çoklu gruplama değişkenleriyle DMF analizi yapmaya izin veren bir yöntemdir ve Şekil 1.3'te yer alan  $z$  sembolü bahsedilen çoklu gruplama değişkenlerini içeren bir vektör olarak tanımlanmıştır.  $z$  vektörü sürekli ya da süreksiz bir değer olabilir. Böylece ÇGÇN yönteminin, sadece tek bir süreksiz gruplama değişkeniyle DMF analizi yapmaya izin veren geleneksel DMF belirleme yöntemlerine (MH, SIBTEST ya da MTK-OO gibi) göre daha esnek olduğu söylenebilir (Wang, Shih ve Yang,2009).

ÇGÇN ile DMF belirlenmesinin altında yatan temel metod, bir gruplama değişkeni için doğrudan ve dolaylı etkilerin her ikisinin birden değerlendirmesini kapsar. Grup değişkeninin ( $z$ ), gizil özellik ( $\eta$ ) üzerinden madde yanıtlarına olan dolaylı etkisi incelenerek bu gizil değişkenin ortalamasının gruplar boyunca farklı olup olmadığı belirtilir; böylece gizil özellik üzerindeki grup farklılıkları için hesaplama yapılır. Grup değişkeninin ( $z$ ), madde yanıtları ( $Y_i$ ) üzerindeki doğrudan etkisi incelenerek (yani  $\beta_1 \neq 0$ 'dır) gruplar boyunca yanıt olasılıklarında bir farklılık olup olmadığı belirtilir. Bu ilişki, gruplar için gizil özelliğin ortalamasındaki farklılıkların kontrolünden sonra, TB DMF'nin testidir (Finch, 2005).

#### **1.5.3.6. TBO DMF İçin MTK'ya Dayalı Olmayan Yöntemler**

Yukarıda açıklandığı üzere, LR yaklaşımı TBO DMF belirlemek için de kullanılabilen bir yöntemdir; ancak bu tür DMF belirlemede tek yaklaşım değildir. Çok sayıda alternatif bulunmaktadır. Breslow-Day (BD) testi (Breslow ve Day, 1980) madde cevapları ve grup üyeliği arasındaki ilişkinin toplam test puanlarının dağılımı boyunca homojen olup olmadığını belirler. Eğer homojen değilse TBO DMF'nin varlığı söz konusudur. BD istatistiği asimtotik ki-kare dağılımına sahiptir. Bununla birlikte serbestlik dereceleri, toplam test puanlarının sayısı kadardır (Penfield, 2003).

İkinci olarak, birçok araştırmacı (Finch ve French, 2007; Narayanan ve Swaminathan, 1996) TBO DMF'nin belirlenmesi için TB DMF belirleme yönteminin bu yönde değiştirilmesini önermiştir. TBO DMF'yi belirlemek için düzeltilmiş MH (Mazor, Clauser ve Hambleton, 1994) ve düzeltilmiş SIBTEST versiyonları (Li & Stout, 1996) vardır. Bunlar Tablo 1.1'de TBO.MH ve TBO.SIBTEST olarak belirtilmiştir.

Çoklu gruplar ve TBO DMF için, son zamanlardaki Bayesian (Soares, Gonçaves, ve Gamerman, 2009) yaklaşımlarından ayrı olarak, literatürde herhangi bir yöntem gözükmemektedir. Bir olası yaklaşım genelleştirilmiş MH metodunu TBO DMF içeriğine genişletmektir (Magis, Beland, Teurlinckx ve Boeck, 2010). Alternatif şekilde LR yöntemi de birden çok odak grup için kullanılabilir (Millsap ve Everson, 1993).

#### **1.5.3.7. MTK'ya Dayalı DMF Belirleme Yöntemleri**

MTK'ya dayalı yöntemlerden çoğu her iki DMF türü için de kullanılabilir. 1-PL model sadece TB DMF için kullanılabilirken, 2-PL ve 3-PL modeller TB ve TBO DMF için uygundur. MTK'ya dayalı DMF belirleme yöntemlerinden ilki OO (olabilirlik oranı) testidir. Bu yöntem kendi içinde ikiye ayrılır: her iki gruptaki bireyler için aynı olan madde parametreleri ile sıkı bir ilişkisi olan, sınırlandırılmış model ve bireylerin ait oldukları gruplar arasında değişime izin veren, genişletilmiş model. OO testi yöntemi kavramsal olarak lojistik regresyon yöntemine yakındır (Braddy, Meade ve Johnson, 2006).

İkinci yöntem Lord'un ki-kare testidir (Lord, 1980). Bu yöntem, her iki gruptaki eşit madde parametrelerinin yokluk hipotezi testine dayanmaktadır. Kim, Cohen ve Park (1995) Lord'un ki-kare testini birden çok odak grubun yer aldığı bir yöntemle genişletmiştir. Bu genelleştirilmiş Lord'un testi olarak adlandırılmıştır.

Üçüncü yöntem Raju'nun alan ölçümleri yöntemidir (Raju, 1988, 1990). Bu yöntemle DMF'nin belirlenmesinde, madde karakteristik eğrileri dikkate alınır. Referans ve odak grup için madde karakteristik eğrileri arasındaki alan hesaplanır. Aynı parametre değerine sahip maddelerin madde karakteristik eğrilerinin de aynı olması beklenir. Bu yöntemde, bir madde için, alt gruplardan elde edilen madde karakteristik eğrileri çizilir. Eğer madde bu alt gruplara göre farklı şekilde işliyorsa, söz konusu iki eğri arasında bir alan gözlemlenir. Başka bir deyişle, madde karakteristik eğrileri arasında fark bulunduğu anda, burada DMF'nin varlığından söz edilir (Camilli ve Shepard, 1994; akt. Kan, Sünbül ve Ömür, 2013). Ancak, her bir maddeye yönelik her iki grubun şans parametrelerinin eşit olma zorunluluğu bu yöntem için önemli bir sınırlılıktır (Magis, Beland, Teurlinckx ve Boeck, 2010).

## 2. İLGİLİ ARAŞTIRMALAR

Bu bölümde ÇGÇN ve LR yöntemlerinin kullanıldığı analizlerin sonuçlarına yönelik 1. tip hata ve güç oranlarının incelendiği bazı çalışmalara yer verilmiştir.

Finch (2005) yaptığı çalışmada ÇGÇN DFA modelini geleneksel DMF belirleme yöntemlerinden Mantel-Haenszel, SIBTEST ve MTK olabilirlik oranı ile karşılaştırmıştır. DMF belirlemede kullanılan ÇGÇN yönteminin şans parametresi olduğu zaman daha kısa testler için yüksek 1. tip hata oranına sahip olduğu sonucuna ulaşılmıştır. Ancak daha uzun testler ya da şans parametresinin olmadığı koşullarda bu yöntem klasik yöntemlere göre uygulanabilir bir seçenektir. ÇGÇN yöntemi, 50 maddelik sınav koşulu için incelenen yöntemler arasında en düşük 1. tip hata oranına sahiptir. ÇGÇN yönteminin güç oranı ise çalışmada ele alınan diğer yöntemlerin güç oranlarına göre daha yüksektir. Ayrıca diğer üç yöntemin performansı DMF içermeyen maddelerden olumsuz şekilde etkilenirken, ÇGÇN yönteminin 1. tip hatasında sadece az miktarda artış ve gücünde yine az miktarda düşüş gözlenmiştir. Bu sebeple çok sayıda DMF'li madde içeren durumlar için ÇGÇN yöntemi tercih edilebilir. Bunlara ek olarak, DMF içermeyen maddelerin etkili olduğu durumlarda diğer yöntemlerin güç oranlarının odak grubun büyüklüğünden ÇGÇN yöntemine göre daha çok etkilendikleri gözlenmiştir. Aynı zamanda, 3-PL modele sahip daha az sayıda madde içeren sınav koşulları için 1. tip hata oranı oldukça yüksek bulunduğundan, DMF belirlemede ÇGÇN yöntemine yönelik diğer yöntemlere göre olumlu bir kıyaslama yapılamamıştır. 3-PL modelde farklı test uzunlukları kullanılarak ÇGÇN yöntem ve diğer yöntemlerin performanslarının karşılaştırılmasına yönelik daha fazla çalışma yapılması önerilmiştir.

Finch ve French (2007) yaptıkları çalışmada SIBTEST, LR, MTK-OO ve DFA DMF belirleme yöntemlerini kullanmışlardır. Çalışmada bu dört yöntemin, TBO-DMF'yi belirlerkenki güç ve 1. tip hata oranları karşılaştırılmıştır. Örneklem büyüklüğü, yetenek dağılımları, TBO-DMF içeren madde yüzdeleri ve verilerin dayandığı temel modeller çalışmada ele alınan değişkenlerdir. Bu çalışmanın sonuçlarına göre, TBO-DMF içeren maddeleri belirlemede en iyi performansı SIBTEST yöntemi göstermiştir. DFA ve MTK-OO yöntemleri içinse TBO-DMF içeren madde

yüzdesinin %10 olduğu koşulda, bu maddeleri doğru şekilde belirlemenin mümkün olmadığı belirtilmiştir.

Shih ve Wang (2009) yaptıkları araştırmada DMF'siz madde seti ile ÇGÇN yöntemini kullanmışlardır. Çalışmada kullanılan bağımsız değişkenler; madde tepki modeli, örneklem büyüklüğü, test uzunluğu, DMF içeren madde yüzdesi ve DMF içermeyen madde uzunluğudur. DMF içermeyen madde uzunluğu 1, 2, 4 ve 10 maddelik olarak belirlenmiştir. DMF içermeyen maddelerin dışındaki tüm maddeler DMF içerip içermemeleri bakımından test edilmiştir. DMF içermeyen madde uzunluğu 1 adet olduğu koşulda diğer tüm bağımsız değişkenler için bakıldığında 1. tip hata oranı ortalamaları 0.03 ve 0.08 arasındadır. Bu sonuçlar beklenen 0.05 değerine oldukça yakındır yani 1 adet DMF içermeyen madde uzunluğu iyi kontrol edilebilen 1. tip hata oranı sağlamaktadır. Güç oranları ise örneklem büyüklüğü ve madde tepki modelinden oldukça etkilenmiştir. 1 adet DMF içermeyen madde uzunluğu, 2PL model ve 3000 bireylik örneklem büyüklüğü için güç oranı 0.89'a kadar yükselmiştir. 2,4 ve 10 maddelik DMF'siz madde içeren koşullar için de sonuçlar 1 adet DMF'siz madde içeren koşullardakine benzer olarak, beklenen 0.05 değerine yakın çıkmıştır; güç oranları ise testteki ankor uzunluğu arttıkça yükselmiştir ve 4 adet DMF'siz maddenin, özellikle geniş sayıda örneklem için yüksek DMF belirleme gücü sağlamaya yetecek uzunlukta olduğundan bahsedilmiştir. Bu araştırma ile M-SA'nın standart ÇGÇN yöntemine (M-ST) üstün geldiği anlaşılmıştır. M-SA yöntemi çalışılan tüm maddeleri DMF açısından tek bir bilgisayar düzenlemesiyle belirleyebilirken M-ST yöntemi birden fazla düzenlemeleri gerektirdiği belirtilmiştir. Wang ve Shih (2010), saf DMF'siz maddeli ÇGÇN yöntemini çok kategorili maddelerde de DMF değerlendirmesi için kullanmışlardır.

Finch ve French (2011) yaptıkları simülasyon çalışmasında gizil değişken kullanan yöntemlerin performanslarını belirlemeyi amaçlamışlardır. Araştırmada, gizil değişken kullanan yöntemlerden ÇGÇN yöntemi kullanılmıştır. Çalışmada hiyerarşik veya iç içe geçmiş yapıdaki veriler kullanılmıştır. Standart yöntemler ve gizil değişken kullanan yöntemler arasında 1. tip hata ve güç oranları açısından benzerlik görülmüştür.

Kim ve Yoon (2011) yaptıkları araştırmada çok gruplu DFA ve MTK yaklaşımlarını karşılaştırmışlardır. Çalışmada kullanılmak üzere, farklı koşullara uygun veriler

türetilmiştir. Sonuçlara bakıldığında her iki yaklaşımın da DMF'li maddeleri belirlemede yüksek güç oranına sahip oldukları görülmüştür.

MacIntosh ve Hashim (2003), ÇGÇN yönteminden elde edilen model parametre hesaplarının MTK türünde madde güçlüğü ve ayırt ediciliği değerlerine dönüştürülürken yapılan hatayı hesaplamayı amaçlamışlardır. Çalışmada kullanılan veriler gerçek sınav sonuçlarına dayalıdır ve çok kategoriliden iki kategoriliye çevrilerek elde edilmiştir. Araştırmanın sonucuna göre ÇGÇN yöntemine ait parametrelerin dönüştürülmesi ile elde edilen madde güçlüğü ve ayırt ediciliği parametreleri ve MTK yöntemleriyle elde edilen madde güçlüğü ve ayırt ediciliği parametrelerinin birbirine çok yakın olduğu görülmüştür.

Woods (2009) DMF'li maddelerin belirlenmesinde ÇGÇN yönteminin kullanılmasının nasıl sonuçlar getireceği üzerinde durmuştur ve çalışmada çok kategorili veriler kullanmıştır. Araştırmada kullanılan diğer yöntem MTK-OO testidir. Çalışmada, ÇGÇN yöntemi için uygun örneklem büyüklüğünün henüz tam olarak belirlenmediği ifade edilmiş ve bu bakımdan araştırmada küçük odak gruplar için ÇGÇN yönteminin DMF'li maddeleri belirlemedeki performansı incelenmiştir. Araştırmanın sonuçlarına göre, ÇGÇN yönteminin DMF belirleme performansının MTK-OO testi yöntemininkine göre daha doğru sonuçlar verdiği görülmüştür.

Woods, Oltmanns ve Turkheimer (2009) DMF'li maddeleri belirlemede ÇGÇN yapısal eşitlik modelini kullanmışlardır. Çalışma gerçek verilerle yürütülmüştür ve bu veriler çok kategorilidir. Sonuçlar, testte yer alan maddelerin çoğunda DMF olduğunu göstermiştir. Araştırmanın sonuçlarına yönelik olarak ise DMF içerdiği tespit edilen maddelerin testten atılması önerilmiştir.

İlgili araştırmalara bakıldığında, ÇGÇN yönteminin çoğunlukla çok kategorili verilerle yürütülen çalışmalarda kullanıldığı görülmüştür. Bu araştırma ise iki kategorili veriler için, ÇGÇN ve lojistik regresyon yöntemleri kullanılarak yürütülen bir DMF belirleme çalışmasıdır. Çalışmada 8 farklı koşula yönelik yapılan analizler sonucunda ÇGÇN ve LR yöntemlerinin 1. tip hata oranları ve güçleri karşılaştırılmıştır. Karşılaştırılan koşullarda ele alınan değişkenler örneklem büyüklüğü, grupların yetenek dağılımları ve DMF gösteren madde yüzdesidir.



Çalıřmada 2000 ve 4000 bireylik iki farklı örneklem büyüklüğü ele alınmıřtır ve çalıřma 3-PL modele dayalı verilerle yürütölmüřtür.

## 3. YÖNTEM

### 3.1. Araştırmanın Yöntemi

Bu çalışma özü itibariyle bir simülasyon (benzetim) çalışmasıdır. Bu araştırma iki kategorili veriler için, ÇGÇN ve Lojistik Regresyon yöntemleri kullanılarak yürütülen bir DMF belirleme çalışmasıdır.

### 3.2. Çalışma Grubu

#### 3.2.1. Çalışma Grubunun Özellikleri

Bu simülasyon çalışmasındaki veriler R programı kullanılarak 8 farklı koşula göre türetilmiştir. Verilerin türetilmesi işlemi, her bir koşul için 100 kere gerçekleştirilmiştir. Bunun amacı, ne kadar çok sayıda analiz yapılırsa doğru sonuçlara o kadar yakın sonuçlar elde edilebileceğidir.

#### 3.2.2. Simülasyon Koşulları

Bu simülasyon çalışmasında DMF analiz sonuçları, referans grup ve odak grup olmak üzere iki grup için karşılaştırılmıştır. Sabit tutulan koşullar ve değişen koşullar aşağıda açıklanmıştır.

#### **Sabit Koşullar:**

Referans gruptaki bireylerin yetenek düzeylerinin dağılımı, madde sayısı (testin uzunluğu 30) ve DMF büyüklüğü sabit tutulan koşullardandır. Her koşul için referans gruptaki bireylerin yetenek parametreleri standart normal dağılıma uygun şekilde,  $N(0,1)$ , örneklenmiştir. 30 adet, iki kategorili puanlanan madde üretilmiştir. DMF gösterdiği belirtilen maddeler için gruplar arasındaki madde güçlük parametrelerindeki değişim 0.6 olarak sabit bir değer alınmıştır. Odak ve referans grup oranı da (1:1) sabit tutulan bir diğer koşuldur.

#### **Değişen Koşullar:**

**Örneklem büyüklüğü:** Bu çalışmada örneklem büyüklüğü koşulu için 2000 ve 4000 olmak üzere iki düzey ele alınmıştır. 2000 bireylik örneklem büyüklüğü için 1000 bireylik referans ve 1000 bireylik odak grup oluşturulmuştur. 4000 bireylik örneklem büyüklüğü içinse 2000 bireylik referans ve 2000 bireylik odak grup oluşturulmuştur.

**Odak grubun yetenek dağılımı:** Odak grubun yetenek dağılımı koşulu için iki düzey ele alınmıştır. Birinci düzeyde odak gruptaki bireylerin yetenek parametreleri bireylerin yetenek düzeylerinin ortalaması 0, standart sapması 1 olan standart normal dağılım  $N(0,1)$ , ikinci düzeyde yetenek düzeylerinin ortalaması  $-0,5$  ve standart sapması 1 olan normal dağılım  $N(-0,5, 1)$ , gösterecek şekilde üretilmiştir.

**DMF'li madde yüzdesi:** Testteki DMF'li madde yüzdesi koşulu için iki düzey ele alınmıştır. Birinci düzeyde testteki maddelerin %10'u (testteki 30 maddeden 3 madde), ikinci düzeyde testteki maddelerin %20'si (testteki 30 maddeden 6 madde) DMF'li madde olarak belirlenmiştir.

**Tablo 3.1: Simülasyon Koşulları Planı**

Simülasyon Koşul Numarası	Örneklem Büyüklüğü	Yetenek Dağılımı		DMF'li Madde Yüzdesi
		Referans	Odak	
1	2000 (1000 R/1000 O)	(0,1)	(0,1)	%10 (3 madde)
2	4000 (2000 R/2000 O)	(0,1)	(0,1)	%10 (3 madde)
3	2000 (1000 R/1000 O)	(0,1)	(0,1)	%20 (6 madde)
4	4000 (2000 R/2000 O)	(0,1)	(0,1)	%20 (6 madde)
5	2000 (1000 R/1000 O)	(0,1)	(-0,5,1)	%10 (3 madde)
6	4000 (2000 R/2000 O)	(0,1)	(-0,5,1)	%10 (3 madde)
7	2000 (1000 R/1000 O)	(0,1)	(-0,5,1)	%20 (6 madde)
8	4000 (2000 R/2000 O)	(0,1)	(-0,5,1)	%-20 (6 madde)

### 3.3 Verilerin Türetilmesi

İki kategorili (0/1) puanlanmış maddeler için veriler 3-parametrelili lojistik madde tepki kuramı ( 3-PL MTK ) modeli kullanılarak türetilmiştir. 3-parametrelili lojistik (3-PL) madde tepki fonksiyonu (MTF) aşağıdaki eşitlikle gösterilebilir:

$$P(U_i = 1|\theta) = c_i + \frac{1-c_i}{1+e^{-1.7a_i(\theta-b_i)}} \quad (5)$$

Eşitlikte  $\theta$  bireyin yetenek düzeyi,  $c_i$   $i$  maddesi için şans parametresi,  $a_i$   $i$  maddesi için ayırt edicilik parametresi ve  $b_i$   $i$  maddesi için güçlük parametresidir.

3-PL MTF her bir madde için doğru bir cevap üretme olasılığı ve  $\theta$  arasındaki ilişki sağlanarak kestirilebilir. 3-PL MTF, 2-PL MTF'na şans parametresinin ( $c$ ) eklenmiş halidir. 2-PL MTF aşağıdaki şekilde açıklanabilir:

$$P(U_i = 1|\theta) = \frac{1}{1+e^{-1.7a_i(\theta-b_i)}} \quad (6)$$

MTK modelleri gibi DFA modelleri de bir test maddesine verilen cevaplarla bir birey gizil yeteneği arasında bağlantı kurmak için kullanılabilir. MTK ve DFA modelleme yaklaşımları; maddeye verilen cevapla birey gizil yeteneği arasındaki ilişkinin yapısını, madde güçlüklerini ve ayırıcılıklarını kapsayacak şekilde maddelerin kendilerini ve her bir bireyin temel yeteneğini içerecek şekilde bireylerin kendilerini tanımlayarak parametre değerlerini sağlarlar. ÇGÇN DFA modeli kullanılarak elde edilen parametre değerlerinin MTK modellerine ait parametre değerlerine kolayca dönüştürülebildiği gösterilmiştir (MacIntosh ve Hashim, 2003; Muthen, Kao ve Burstein, 1991).

Tek-faktörlü DFA modeli aşağıdaki şekildedir:

$$y_i^* = \lambda_i \eta + \varepsilon_i, \quad (7)$$

Burada  $y_i^* = i$  değişkeni gizil yanıtı ( $y_i^* > \tau_i$ , gözlenen bir değişken,  $y_i = 1$ ;  $\tau_i$ , madde güçlüğüne ilişkin sınır parametresi),  $\eta$ =gizil özellik,  $\lambda_i = i$  değişkeni için faktör yükü ve  $\varepsilon_i$  =tesadüfi hata.

Bu faktör çözümlenmeli model iki kategorili madde yanıtlarından oluşan veriler için McDonald (1967) ve Lord ve Novick (1968) tarafından tanımlanan normal ogive dağılım modeline (ortalaması 0, standart sapması 1 olan dağılım) eşdeğerdir. Muthen, Kao ve Burstein (1991) MTK formülünde yer alan ayırt edicilik parametresinin,  $a$ , ÇGÇN DFA modelinden  $\lambda$ 'nin değerinin kullanılarak elde edilebileceğini kanıtlayarak bu çalışmayı ÇGÇN yöntemi için genişletmişlerdir.

$$a_i = \frac{\lambda_i}{\sqrt{(1-\lambda_i^2)\sqrt{\sigma_{\eta\eta}}}}, \quad (8)$$

Burada  $\sigma_{\eta\eta}$  =gizil özellik varyansıdır.

Sırasıyla, ÇGÇN çalışmasında, MTK güçlük parametresi,  $b$ , hem  $\lambda$  hem de  $\tau$ 'nin kullanımını içerir:

$$b_i = [(\tau_i - \beta_i z_k) \lambda_i^{-1} - \mu_\eta] \sigma_{\eta\eta}^{-\frac{1}{2}}, \quad (9)$$

Burada  $z_k$  =grup belirleyici, 1 odak gruba üyeliği temsil ettiğinde 0 referans grubu temsil eder;  $\beta_i$  =madde yanıtları ve grup arasındaki ilişkinin ölçüsüdür (  $\beta_i$ 'nin anlamlı değeri değişen madde fonksiyonunun varlığını belirtir); ve  $\mu_\eta$  =gizil özellik ortalamasıdır (Finch,2005).

Bu araştırmada kullanılan madde parametreleri Finch'in (2005) çalışmasında kullanılan madde parametrelerinden rastgele seçilerek oluşturulmuş olup seçilen parametreler Tablo 3.2'de gösterilmiştir.

**Tablo 3.2: Madde Parametreleri**

Madde	Referans Grup			Odak Grup DMF Yüzdesi % 10			Odak Grup DMF Yüzdesi % 20		
	$a_i$	$b_i$	$c_i$	$a_i$	$b_i$	$c_i$	$a_i$	$b_i$	$c_i$
1	1.10	-0.70	0.20	1.10	-0.70	0.20	1.10	<b>-0.10*</b>	0.20
2	0.70	-0.60	0.20	0.70	-0.60	0.20	0.70	-0.60	0.20
3	1.40	0.10	0.20	1.40	0.10	0.20	1.40	0.10	0.20
4	0.40	0.80	0.20	0.40	<b>1.40*</b>	0.20	0.40	<b>1.40*</b>	0.20
5	1.40	-0.40	0.20	1.40	-0.40	0.20	1.40	-0.40	0.20
6	1.60	-0.10	0.16	1.60	-0.10	0.16	1.60	-0.10	0.16
7	1.20	0.50	0.20	1.20	0.50	0.20	1.20	0.50	0.20
8	1.20	1.40	0.11	1.20	1.40	0.11	1.20	1.40	0.11
9	1.80	1.40	0.12	1.80	1.40	0.12	1.80	1.40	0.12
10	2.00	1.60	0.16	2.00	1.60	0.16	2.00	1.60	0.16
11	1.00	1.60	0.13	1.00	1.60	0.13	1.00	1.60	0.13
12	1.50	1.70	0.09	1.50	1.70	0.09	1.50	1.70	0.09
13	0.70	-0.50	0.20	0.70	-0.50	0.20	0.70	-0.50	0.20
14	1.20	-0.30	0.20	1.20	-0.30	0.20	1.20	-0.30	0.20
15	0.90	0.20	0.20	0.90	<b>0.80*</b>	0.20	0.90	<b>0.80*</b>	0.20
16	0.70	-0.40	0.20	0.70	-0.40	0.20	0.70	-0.40	0.20
17	1.00	0.70	0.15	1.00	0.70	0.15	1.00	0.70	0.15
18	1.60	1.10	0.12	1.60	1.10	0.12	1.60	<b>1.70*</b>	0.12
19	1.10	2.00	0.06	1.10	2.00	0.06	1.10	2.00	0.06
20	1.10	2.40	0.09	1.10	2.40	0.09	1.10	2.40	0.09
21	1.70	1.30	0.17	1.70	1.30	0.17	1.70	1.30	0.17
22	0.90	1.00	0.15	0.90	1.00	0.15	0.90	1.00	0.15
23	0.50	-0.60	0.20	0.50	-0.60	0.20	0.50	-0.60	0.20
24	1.30	0.40	0.18	1.30	0.40	0.18	1.30	0.40	0.18
25	1.30	1.40	0.06	1.30	1.40	0.06	1.30	1.40	0.06
26	1.10	1.20	0.05	1.10	1.20	0.05	1.10	<b>1.80*</b>	0.05
27	0.90	0.80	0.20	0.90	<b>1.40*</b>	0.20	0.90	<b>1.40*</b>	0.20
28	0.40	-0.40	0.20	0.40	-0.40	0.20	0.40	-0.40	0.20
29	0.80	-0.70	0.20	0.80	-0.70	0.20	0.80	-0.70	0.20
30	1.00	1.10	0.13	1.00	1.10	0.13	1.00	1.10	0.13

DMF gösteren madde sayısının testteki madde sayısının %10'u kadar olduğu simülasyon koşullarında madde 4, madde 15 ve madde 27 DMF gösteren maddeler olarak belirlenmiştir. DMF gösteren madde sayısının testteki madde sayısının %20'si kadar olduğu simülasyon koşullarında madde 4, madde 15 ve madde 27'ye ek olarak madde 1, madde 18 ve madde 26 DMF gösteren maddeler olarak belirlenmiştir. DMF büyüklüğü 0.6 olarak alınmış olup DMF gösteren her bir maddenin madde güçlüğü parametresinde odak grup aleyhinde 0.6 birimlik fark olarak ele alınmıştır.

Tablo 3.1'deki benzetim koşulları ve yukarıda belirtilen DMF'li maddeler dikkate alınarak R programı yardımıyla her bir koşul için 100 tekrar olmak üzere 8 farklı koşula uygun veri setleri oluşturulmuştur.

### **3.4. Kullanılan Programlar ve Verilerin Analizi**

Verilerin DMF analizlerinde lojistik regresyon yöntemi için SAS 9.1.3 (SAS Enstitüsü, 2006) programı, ÇGÇN yöntemi için Mplus 6 (Muthen & Muthen, 2010) programı kullanılmıştır.

Mplus sürekli, iki kategorili, sıralı, sırasız, sayılabilen ya da bunların kombinasyonlarından oluşan veriler için uygun bir analiz programıdır. ÇGÇN yöntemi belirli maddeler için tek biçimli DMF'yi belirlemede kullanılır. Çalışmada iki kategorili puanlanan veriler için tek biçimli DMF belirlemek amacıyla Mplus programından yararlanılmıştır, her bir koşul için türetilen 100 veri setini bir kerede analiz edebilmek için Mplus programı batch dosyaları üzerinden çalıştırılmıştır. Birinci simülasyon koşuluna (SK1'e) ait ÇGÇN yöntemi kod dosyası EK1'de verilmiştir. Maddelere yönelik anlamlılık seviyeleri incelenmiş ve 0,05 alfa değerinden düşük olanlar DMF'li olarak kabul edilmiştir. Analizler sonucunda her bir koşula ait 100'er veri seti için ayrı ayrı 1. tip hatalar ve güçler hesaplanıp ortalamaları alınmıştır. Koşulların ortalama 1. tip hataları ve güçleri karşılaştırılmıştır.

SAS programı varyans analizi (ANOVA), çoklu regresyon, lojistik regresyon, path analizi, çoklu varyans analizi (MANOVA) gibi çeşitli istatistiksel tekniklerin yapılabildiği kod dosyası ile çalışan bir programdır. Çalışmada kullanılan uygun kod dosyası ile yapılan analiz sonucunda her bir koşula yönelik ortalama 1. tip

hata ve güçler elde edilmiştir. Birinci simülasyon koşuluna (SK1'e) ait SAS kod dosyası EK2'de verilmiştir.

### **3.5. 1. Tip Hata ve Gücün Belirlenmesi**

DMF'nin içeriğinde 1. tip hata, DMF göstermeyen bir maddenin yanlışlıkla DMF gösteren madde şeklinde sınıflanması olarak tanımlanmıştır. Güç kavramı ise, gerçekte DMF gösteren bir maddenin DMF gösteriyor olarak doğru şekilde sınıflanmasıdır. DMF araştırmaları için her iki kavram da eşit miktarda önemlidir. (Vaughn ve Wang, 2010).

Araştırmada, 1. tip hata çalışmasında DMF göstermeyen maddeler kullanılmıştır. 2 kategorili puanlanan verilerin incelendiği bu çalışmada simülasyon koşullarından SK1, SK2, SK5 ve SK6 olmak üzere 4 koşul için 27 madde DMF göstermemektedir; simülasyon koşullarından SK3, SK4, SK7 ve SK8 olmak üzere 4 koşul içinse 24 madde DMF göstermemektedir.

Güç çalışmasında DMF gösteren maddeler kullanılmıştır. Simülasyon koşullarından SK1, SK2, SK5 ve SK6 olmak üzere 4 koşul için 3 madde DMF göstermektedir; simülasyon koşullarından SK3, SK4, SK7 ve SK8 olmak üzere 4 koşul içinse 6 madde DMF göstermektedir.

Çalışmada örneklem büyüklüğü, yetenek dağılımı ve DMF'li madde yüzdesinin 1. tip hata ve güç oranları üzerindeki etkileri araştırılmıştır. Çalışmada anlamlılık seviyesi ( $\alpha$  düzeyi) 0,05 olarak kabul edilmiştir. Orta düzeyde tek biçimli ya da tek biçimli olmayan DMF'lerin belirlenmesi için kabul edilebilir güç oranı 0,70 ve üzeridir ( $\geq 0,70$ ) (González-Romá, Hernández ve Gómez-Benito, 2006). 1. tip hata oranı için ise istenen 0,05'den küçük bir değer olmasıdır.

#### 4. BULGULAR VE TARTIŞMA

Bu bölümde iki alt probleme yönelik bulgulara ve yorumlara yer verilmiştir. Çalışmanın içerdiği tüm koşulların kombinasyonları yoluyla uygulanan iki DMF belirleme yöntemine yönelik 1. tip hata oranları Tablo 4.1’de gösterilmiştir. Alt problem 1’e geçmeden önce bulgulara ait genel yorumlara aşağıda yer verilmiştir. Bu çalışma için kabul edilebilir 1. tip hata oranı 0.05’tir.

**Tablo 4.1: Örneklem Büyüklükleri, Yetenek Dağılımları ve DMF’li Madde Yüzdelerine Göre 1. Tip Hata Oranları**

DMF %	Örneklem Büyüklüğü	Yetenek Dağılımı R/O	ÇGÇN	LR
10	2000	(0,1) (0,1)	.121	.069
		(0,1) (-0.5,1)	.120	.068
	4000	(0,1) (0,1)	.065	.087
		(0,1) (-0.5,1)	.090	.097
20	2000	(0,1) (0,1)	.129	.122
		(0,1) (-0.5,1)	.128	.129
	4000	(0,1) (0,1)	.076	.244
		(0,1) (-0.5,1)	.078	.189

*Not.* DMF% DMF’li maddelerin yüzdesini ifade eder. LR= Lojistik Regresyon Yöntemi; ÇGÇN = Çoklu Göstergeler, Çoklu Nedenler Yaklaşımı.

Tablo 4.1 genel olarak incelendiğinde, bazı koşullar için ÇGÇN yönteminin bazı koşullar için ise LR yönteminin 1. tip hata oranları daha düşüktür; kimi koşullarda bu değerler birbirine çok yakın çıkmıştır; ancak, ÇGÇN yönteminin 1. tip hata oranının daha düşük olduğu koşul sayısı, LR yönteminin 1. tip hata oranının daha düşük olduğu koşul sayısından fazladır. Her iki yöntem için de 1. tip hata oranları her koşulda 0.05’in üzerindedir.

ÇGÇN yöntemi için 2000 kişilik örneklemelerde 1. tip hata oranları yüksektir. DMF’li madde yüzdesinin %10, örneklemdeki kişi sayısının 2000 olduğu ancak yetenek dağılımlarının farklı olduğu koşullar için 1. tip hata oranları birbirine yakındır yani yetenek dağılımlarının farklı olması bu koşullar için çok fazla değişikliğe sebep olmamıştır. Benzer incelemeler DMF’li madde yüzdelerinin %20, örneklem sayısının 2000 olduğu koşullar için yapıldığında sonuç benzer çıkmıştır. Bu yöntem için DMF’li madde yüzdelerinin artırılması 2000 kişilik örneklemelerde, aynı yetenek dağılımına sahip veriler için fazla bir değişiklik yaratmamıştır. Örneklem büyüklüğü arttığında ise 1. tip hata oranları daha düşük çıkmıştır. Bu oranlar kabul edilebilir değer olan 0.05’e en yakın değerlerdir. Özellikle çalışmanın genelinde en düşük 1. tip hata oranı 4000 kişilik, %10 DMF’li madde içeren normal dağılıma



sahip benzetim koşulu içindir. 2000 kişilik örneklemin aksine 4000 kişilik örneklem için yetenek dağılımının değişmesi 1. tip hata oranının artmasına sebep olmuştur. Ancak bu artış %10 DMF'li madde içeren koşullar için daha belirgindir. Genel olarak ÇGÇN yöntemi için örneklem sayısı arttıkça 1. tip hata oranının azaldığı görülmektedir.

LR yönteminde ise en düşük 1. tip hata oranı DMF'li madde oranının %10 olduğu koşullar içindir. DMF'li madde yüzdesinin artması ile 1. tip hata oranının arttığı görülmektedir. Yetenek dağılımlarının değişmesi ile gerçekleşen 1. tip hata oranındaki değişiklikler 4000 kişilik örneklemelerde daha fazladır. Özellikle %20 oranında DMF'li madde içeren 4000 kişilik örneklem için odak grubun yetenek seviyesi daha düşük olması durumunda 1. tip hata oranının azaldığı görülmüştür.

#### **4.1. Alt Problem 1'e İlişkin Bulgular ve Yorumlar**

ÇGÇN ve LR DMF belirleme yöntemlerinin 1. tip hata oranları örneklem büyüklüklerine, yetenek dağılımlarına ve DMF gösteren madde yüzdelere göre nasıl değişmektedir?

Birey sayısının 2000 olduğu, referans gruptaki ve odak gruptaki bireylerin yetenek dağılımlarının aynı olduğu ve testteki DMF gösteren madde sayısının 3 olduğu simülasyon koşulunda, ÇGÇN yöntemine göre gerçekleştirilen DMF analizinin 1. tip hata oranı LR yöntemiyle gerçekleştirilen DMF analizinin 1. tip hata oranına göre daha yüksek bulunmuştur. Ancak örneklem büyüklüğü iki katına çıktığında ÇGÇN yönteminin 1. tip hata oranı azalırken LR yönteminin 1. tip hata oranının arttığı gözlenmiştir. ÇGÇN yönteminin 1. tip hata oranı örneklem büyüklüğü arttıkça azalırken, LR yönteminin 1. tip hata oranı örneklem büyüklüğü arttıkça artmaktadır. Ancak LR yönteminde örneklem büyüklüğüne bağlı 1. tip hata oranındaki artış, ÇGÇN yöntemindeki örneklem büyüklüğüne bağlı 1. tip hata oranındaki düşüş kadar belirgin değildir.

Birey sayısının 2000 olduğu, referans gruptaki ve odak gruptaki bireylerin yetenek düzeylerinin standart normal dağılıma sahip olduğu ve testteki DMF gösteren madde sayısının 6 madde olduğu simülasyon koşulunda, her iki yöntemin 1. tip hata oranı birbirine yakın bulunmuştur. Testte DMF gösteren madde yüzdesi arttıkça LR yönteminin 1. tip hata oranının da arttığı görülmektedir.

Referans ve odak grubun yetenek dağılımlarının aynı olduğu, %20 DMF gösteren madde içeren koşullar için örneklem sayısı arttığında ÇGÇN yönteminin 1. tip hata oranı azalırken LR yönteminin 1. tip hata oranı artmıştır. ÇGÇN yönteminin 1. tip hata oranındaki azalma oranı %10 DMF gösteren madde içeren koşullardakiyle benzerlik göstermekteyken LR yönteminin 1. tip hata oranındaki artma %20 DMF gösteren madde içeren koşullarda %10'luk koşullara göre daha fazladır. DMF gösteren madde oranındaki değişim LR yönteminin 1. tip hata oranını oldukça etkilemiştir.

%10 DMF gösteren madde içeren, referans ve odak grubun yetenek dağılımlarının farklı olduğu, birey sayısının 2000 ve 4000 olduğu simülasyon koşulları için ÇGÇN ve LR yöntemlerinin 1. tip hata oranları ile %10 DMF gösteren madde içeren, referans ve odak grubun yetenek dağılımlarının aynı olduğu, birey sayısının 2000 ve 4000 olduğu koşullardaki ÇGÇN ve LR yöntemlerinin 1. tip hata oranları birbirine yakın çıkmıştır.

Referans ve odak grubun yetenek dağılımlarının farklı olduğu, %20 DMF gösteren madde içeren, birey sayısının 2000 olduğu simülasyon koşulu için her iki yöntemin 1. tip hata oranı birbirine yakındır.

%20 DMF gösteren madde içeren, odak grubun yetenek düzeyinin referans gruba göre daha düşük olduğu ve birey sayısının 2000 olduğu simülasyon koşulundaki ÇGÇN yönteminin 1. tip hata oranı ile %20 DMF gösteren madde içeren, referans ve odak gruptaki bireylerin yetenek dağılımlarının aynı olduğu ve birey sayısının 2000 olduğu koşullar için ÇGÇN yönteminin 1. tip hata oranı birbirine yakın çıkmıştır. %20 DMF gösteren madde içeren, birey sayısının 4000 olduğu ve referans ve odak grubun yetenek dağılımlarının farklı olduğu koşuldaki LR yönteminin 1. tip hata oranı; %20 DMF gösteren madde içeren, birey sayısının 4000 olduğu ve referans ve odak grubun yetenek dağılımlarının aynı olduğu koşuldaki LR yönteminin 1. tip hata oranına göre daha düşük çıkmıştır.

Birey sayısının 2000 olduğu, %10 DMF gösteren madde içeren koşullar için referans ve odak grubun yetenek düzeylerinin farklı olması her iki yöntemin 1. tip hata oranlarında önemli miktarda değişikliğe sebep olmamıştır.

Birey sayısının 4000 olduđu ve %10 DMF gösteren madde ieren kořullar iin odak grubun yetenek dzeyinin referans gruba gre daha dřk hale gelmesi her iki yntemin 1. tip hata oranını arttırmıřtır. rnekleme sayısı arttıka yetenek dađılımlarındaki bu farklılık sonucu yntemlerin 1. tip hata oranlarını birbirine yaklařtırmıřtır.

Birey sayısının 2000 olduđu ve DMF madde sayısının testteki maddelerin %20'si kadar olduđu kořullar iin odak grubun yetenek dzeyinin referans gruba gre daha dřk hale gelmesi her iki yntemin 1. tip hata oranlarında nemli miktarda deđiřikliđe sebep olmamıřtır.

Birey sayısının 4000 olduđu ve DMF madde sayısının testteki maddelerin %20'si kadar olduđu kořullar iin odak grubun yetenek dzeyinin referans gruba gre daha dřk hale gelmesi GN ynteminin 1. tip hata oranında nemli miktarda deđiřikliđe sebep olmazken LR ynteminin 1. tip hata oranında nemli miktarda dřř gzlenmiřtir.

Birey sayısının 2000 olduđu ve referans ve odak gruptaki bireylerin yetenek dzeylerinin standart normal dađılıma sahip olduđu kořullarda testteki DMF gsteren maddelerin yzdesi iki katına ıktıđında GN ynteminin 1. tip hata oranı nemli miktarda deđiřiklik gstermezken LR regresyon ynteminin 1. tip hata oranının yaklařık olarak iki katına ıktıđı gzlenmiřtir.

Birey sayısının 2000 olduđu ve referans ve odak gruptaki bireylerin yetenek dađılımlarının farklı olduđu kořullarda testteki DMF gsteren maddelerin yzdesi iki katına ıktıđında GN ynteminin 1. tip hata oranı nemli miktarda deđiřiklik gstermezken LR regresyon ynteminin 1. tip hata oranının yaklařık olarak iki katına ıktıđı gzlenmiřtir.

Birey sayısının 4000 olduđu ve referans ve odak gruptaki bireylerin yetenek dzeylerinin standart normal dađılıma sahip olduđu kořullarda testteki DMF gsteren maddelerin yzdesi iki katına ıktıđında GN ynteminin 1. tip hata oranı nemli miktarda deđiřiklik gstermezken LR regresyon ynteminin 1. tip hata oranında nemli miktarda artıř gzlenmiřtir. Bu artıř birey sayısının 2000 ve referans ve odak gruptaki bireylerin yetenek dzeylerinin standart normal dađılıma sahip olduđu kořul iin DMF gsteren madde yzdesi iki katına ıktıđında LR ynteminin 1. tip hata oranındaki artıř miktarından daha fazladır.

Birey sayısının 4000 olduğu ve referans gruptaki bireylerin yetenek düzeylerinin standart normal dağılıma sahip olduğu, odak gruptaki bireylerin yetenek düzeylerinin ortalaması -0,5 ve standart sapması 1 olan normal dağılıma sahip olduğu koşullarda testteki DMF gösteren maddelerin yüzdesi iki katına çıktığında ÇGÇN yönteminin 1. tip hata oranı önemli miktarda değişiklik göstermezken LR yönteminin 1. tip hata oranının yaklaşık olarak iki katına çıktığı gözlenmiştir.

DMF gösteren madde yüzdesindeki artıştan LR yönteminin 1. tip hata oranındaki değişimin daha fazla olduğu görülmektedir. LR yönteminin 1. tip hata oranındaki en fazla artış, birey sayısının 4000 olduğu ve referans ve odak gruptaki bireylerin yetenek düzeylerinin standart normal dağılıma sahip olduğu koşullarda gerçekleşmiştir.

1. tip hata oranı ÇGÇN yöntemi için LR yönteminde olduğundan farklı bir şekilde değişim göstermektedir. ÇGÇN yöntemi için 1. tip hata oranının kabul edilebilir değer olarak belirlenen 0.05'e en yakın olduğu koşul 4000 kişilik örneklemden oluşan, referans ve odak grubun yetenek dağılımlarının standart normal dağılım gösterdiği, %10 DMF gösteren madde içeren koşuldur. 4000 kişilik örneklem büyüklüğüne sahip koşullar için ÇGÇN yönteminin 1. tip hata oranları LR yönteminin 1. tip hata oranlarından daha düşüktür. Ancak bu oran hiçbir koşulda 0.065'in altına düşmemiştir.

ÇGÇN yönteminin DMF belirlemedeki performansının diğer yöntemlerle karşılaştırıldığı simülasyon çalışmalarında 3-PL modelin kullanıldığı kısa test (20 maddelik test) koşulları için 1.tip hata oranları 0.10'dan büyük 0.20'dense küçük çıkmıştır (Finch, 2005). Finch ilgili çalışmasında 600 ve 1000 bireylik örneklemler kullanmıştır. 3-PL model kullandığında örneklem büyüklüğü arttıkça genel anlamda 1. tip hata oranlarında azalma gözlenirken; 2-PL model kullandığında 1. tip hata oranlarında çoğunlukla artış gözlenmiştir. Bu çalışmada 3-PL model kullanılmış ve testin uzunluğu Finch'in kullandığı iki test uzunluğunun (20 ve 50 madde) neredeyse ortalaması kadar (30 madde) alınmıştır ve analiz sonuçlarında örneklem büyüklüğü arttıkça 1. tip hata oranlarının azaldığı görülmüştür. Bu açıdan iki çalışma birbirini destekler niteliktedir. Finch'in yine aynı çalışmasında 3-PL modelin kullanıldığı koşullarda DMF gösteren madde yüzdesi arttığında, 1.tip hata oranları artmıştır. DMF gösteren madde yüzdesinin artmasıyla 1. tip hata

oranlarının artması bu çalışmada da söz konusudur ancak artış miktarları Finch'in (2005) çalışmasındaki kadar belirgin değildir.

Çalışmanın bulguları LR yöntemiyle gerçekleştirilen analiz sonuçlarına göre incelendiğinde farklı bir değişim göstermektedir. 1. tip hata oranının kabul edilebilir değer olarak belirlenen 0.05'e en yakın olduğu koşullar 2000 bireylik örnekleme sahip, referans ve odak grubun yetenek dağılımlarının standart normal dağılım gösterdiği, %10 DMF gösteren madde içeren koşul ve 2000 bireylik örnekleme sahip, odak grubun yetenek düzeyinin referans gruba göre daha düşük olduğu %10 DMF gösteren madde içeren koşuldur. %10 DMF gösteren madde içeren koşullar için 1. tip hata oranları ÇGÇN yöntemine göre daha düşüktür. Ancak bu oran hiçbir koşulda 0.068'in altına düşmemiştir.

Finch ve French (2007) çalışmalarında; 500, 750, 1000 ve 1500 bireylik olmak üzere dört farklı örneklem büyüklüğü kullanmışlardır. Çalışmada LR ve DFA yönteminin 1. tip hatasının örneklem büyüklüğündeki artıştan önemli miktarda etkilenmediği görülmüştür. Bu çalışmada %10 DMF gösteren madde içeren koşullarda örneklem büyüklükleri arttıkça LR yönteminin 1. tip hata oranında önemli miktarda değişiklik gözlenmezken ÇGÇN yönteminin 1. tip hata oranı azalmıştır. LR yönteminin 1. tip hata oranındaki bu değişiklik bakımından iki çalışma birbirini destekler niteliktedir. Bu araştırmada %20 DMF gösteren madde içeren koşullarda örneklem büyüklüğü arttıkça LR yönteminin 1. tip hata oranının önemli miktarda arttığı, ÇGÇN yönteminin 1. tip hata oranının ise önemli miktarda azaldığı gözlenmiştir. Bu bakımdan iki çalışmanın sonuçları birbirinden farklıdır.

Çalışmanın içerdiği tüm koşulların kombinasyonları yoluyla uygulanan iki DMF belirleme yöntemi için güç oranları Tablo 4.2'de gösterilmiştir. Alt problem 2'ye geçmeden önce bulgulara ait genel yorumlara aşağıda yer verilmiştir. Bu çalışma için kabul edilebilir güç oranı 0.70 ve üzeridir.

**Tablo 4.2: Örneklem Büyüklükleri, Yetenek Dağılımları ve DMF'li Madde Yüzdelerine Göre Güç Oranları**

<i>DMF %</i>	<i>Örneklem Büyüklüğü</i>	<i>Yetenek Dağılımı R/O</i>	<i>ÇGÇN</i>	<i>LR</i>
10	2000	(0,1) (0,1)	.770	.800
		(0,1) (-0.5,1)	.750	.700
	4000	(0,1) (0,1)	.933	.910
		(0,1) (-0.5,1)	.910	.817
20	2000	(0,1) (0,1)	.852	.827
		(0,1) (-0.5,1)	.780	.772
	4000	(0,1) (0,1)	.977	.935
		(0,1) (-0.5,1)	.943	.872

Her iki yöntemin de tüm koşullar için güç oranları kabul edilebilir seviyenin üzerindedir. Örneklem büyüklüğünün 4000 olduğu koşullar için ÇGÇN yönteminin güç oranı oldukça yüksektir. LR yöntemi için ise örneklem büyüklüğünün fazla olduğu ve her iki grubun da yetenek dağılımlarının standart normal dağılıma sahip olduğu koşullarda güç oranı oldukça yüksek gözükmemektedir. Yapılan çalışmalarda 1. tip hata oranının 0.10'dan yüksek olduğu durumlarda güç oranı sonuçları rapor edilmemiştir. Bu tarz, 1. tip hata oranının yüksek olduğu, durumlarda alfanın belirlenen seviyesindeki gücün standart tanımının anlamlı olmadığı belirtilmiştir (Finch, 2005). Bu çalışmada ise koşullara göre yöntemlerin güç oranlarını karşılaştırmak amacıyla tüm güç oranlarına yer verilmiştir. Ancak 1.tip hata oranının 0,10'dan yüksek olduğu durumlar için güç oranları yatay olarak gösterilmiştir.

#### **4.2. Alt Problem 2'ye İlişkin Bulgular ve Yorumlar**

ÇGÇN ve LR DMF belirleme yöntemlerinin güçleri örneklem büyüklüklerine, yetenek dağılımlarına ve DMF gösteren madde yüzdelerine göre nasıl değişmektedir?

Referans gruptaki ve odak gruptaki bireylerin yetenek düzeylerinin standart normal dağılıma sahip olduğu ve DMF madde sayısının testteki maddelerin %10'u kadar olduğu koşullar için örneklem büyüklüğü arttığında her iki yöntemin de güç oranlarının arttığı ve yüksek bir değere ulaştığı görülmektedir.

Referans gruptaki bireylerin yetenek düzeylerinin standart normal dağılıma sahip olduğu, odak gruptaki bireylerin yetenek düzeylerinin ortalaması -0,5 ve standart sapması 1 olan normal dağılıma sahip olduğu ve DMF madde sayısının testteki

maddelerin %10'u kadar olduđu kořullarda örneklem büyüklüğü arttığıında her iki yöntemin de güç oranlarının arttığı ve yüksek bir değere ulařtığı görülmektedir. Birey sayısının 4000 olduđu ve referans gruptaki bireylerin yetenek düzeylerinin standart normal dağılıma sahip olduđu, odak gruptaki bireylerin yetenek düzeylerinin ortalaması -0,5 ve standart sapması 1 olan normal dağılıma sahip olduđu kořul için ÇGÇN yönteminin güç oranı LR yönteminin güç oranına göre daha yüksektir.

Referans gruptaki ve odak gruptaki bireylerin yetenek düzeylerinin standart normal dağılıma sahip olduđu ve DMF madde sayısının testteki maddelerin %20'si kadar olduđu kořullarda örneklem büyüklüğü arttığıında her iki yöntemin de güç oranlarının arttığı ve yüksek bir değere ulařtığı görülmektedir.

Referans gruptaki bireylerin yetenek düzeylerinin standart normal dağılıma sahip olduđu, odak gruptaki bireylerin yetenek düzeylerinin ortalaması -0,5 ve standart sapması 1 olan normal dağılıma sahip olduđu ve DMF madde sayısının testteki maddelerin %20'si kadar olduđu kořullarda örneklem büyüklüğü arttığıında her iki yöntemin de güç oranlarının arttığı ve yüksek bir değere ulařtığı görülmektedir. Birey sayısının 4000 olduđu, referans gruptaki bireylerin yetenek düzeylerinin standart normal dağılıma sahip olduđu, odak gruptaki bireylerin yetenek düzeylerinin ortalaması -0,5 ve standart sapması 1 olan normal dağılıma sahip olduđu ve DMF madde sayısının testteki maddelerin %20'si kadar olduđu kořullarda ÇGÇN yönteminin güç oranı LR yönteminin güç oranına göre daha yüksektir.

Birey sayısının 2000 olduđu ve %10 DMF gösteren madde içeren kořullar için odak grubun yetenek düzeyinin referans gruba göre daha düşük hale gelmesi her iki yöntemin güç oranında azalmaya sebep olmuřtur ancak ÇGÇN yöntemin güç oranı için bu azalma önemsiz miktarda iken LR yönteminin güç oranı için biraz daha belirgindir.

Birey sayısının 4000 olduđu ve %10 DMF gösteren madde içeren kořullar için odak grubun yetenek düzeyinin referans gruba göre daha düşük hale gelmesi her iki yöntemin güç oranında azalmaya sebep olmuřtur ancak ÇGÇN yöntemin güç oranı için bu azalma önemsiz miktarda iken LR yönteminin güç oranı için biraz daha belirgindir.

Birey sayısının 2000 olduđu ve %20 DMF gösteren madde ieren kořullar iin odak grubun yetenek düzeyinin referans gruba gre daha dřk hale gelmesi her iki yntemin g oranında azalmaya sebep olmuřtur ancak bu azalma her iki yntemin g oranı iin nemsiz miktardadır.

Birey sayısının 4000 olduđu ve %20 DMF gösteren madde ieren kořullar iin odak grubun yetenek düzeyinin referans gruba gre daha dřk hale gelmesi her iki yntemin g oranında azalmaya sebep olmuřtur ancak bu azalma her iki yntemin g oranı iin nemsiz miktardadır.

Odak grup iin yetenek dađılımlarının farklı olması LR ynteminin g oranını GN ynteminin g oranına gre daha ok etkilediđi grlmektedir.

Birey sayısının 2000 olduđu ve referans ve odak gruptaki bireylerin yetenek düzeylerinin standart normal dađılıma sahip olduđu kořullarda DMF gösteren madde yzdelerinin artması her iki yntemin g oranlarının az miktarda artmasına sebep olmuřtur.

Birey sayısının 4000 olduđu ve referans ve odak gruptaki bireylerin yetenek düzeylerinin standart normal dađılıma sahip olduđu kořullarda DMF gösteren madde yzdelerinin artması her iki yntemin g oranlarının az miktarda artmasına sebep olmuřtur.

Birey sayısının 2000 olduđu ve referans gruptaki bireylerin yetenek düzeylerinin standart normal dađılıma sahip olduđu, odak gruptaki bireylerin yetenek düzeylerinin ortalaması -0,5 ve standart sapması 1 olan normal dađılıma sahip olduđu kořullarda DMF gösteren madde yzdesinin artması her iki yntemin g oranlarının az miktarda artmasına sebep olmuřtur.

Birey sayısının 4000 olduđu ve referans gruptaki bireylerin yetenek düzeylerinin standart normal dađılıma sahip olduđu, odak gruptaki bireylerin yetenek düzeylerinin ortalaması -0,5 ve standart sapması 1 olan normal dađılıma sahip olduđu kořullarda DMF gösteren madde yzdesinin artması her iki yntemin g oranlarının az miktarda artmasına sebep olmuřtur.

İncelenen kořullar iin DMF gösteren madde yzdelerindeki deđiřim GN ve LR DMF belirleme yntemlerinin g oranlarında nemli bir deđiřikliđe sebep olmamıřtır.



ÇGÇN yöntemi için güç oranları LR yöntemine benzer bir değişim göstermektedir. ÇGÇN yöntemi için güç oranının mükemmel en yakın olduğu koşul 4000 kişilik örneklemden oluşan, referans ve odak grubun yetenek dağılımlarının standart normal dağılım gösterdiği, %20 DMF gösteren madde içeren koşuldur. Tek bir koşul dışında güç oranı LR yöntemininkinden daha büyüktür. Sadece 2000 kişilik örneklemden oluşan, referans ve odak grubun yetenek dağılımlarının standart normal dağılım gösterdiği, %10 DMF gösteren madde içeren koşul için LR yönteminin güç oranı ÇGÇN yönteminin güç oranından biraz daha yüksek çıkmıştır. ÇGÇN yöntemi için en düşük güç oranı 0.770 iken en yüksek güç oranı 0.977'dir.

ÇGÇN yönteminin DMF belirlemedeki performansının diğer yöntemlerle karşılaştırıldığı simülasyon çalışmalarında güç oranının en belirgin değişiklik gösterdiği koşul uzun testler (50 maddelik testler) için, 1000 bireylik örnekleme DMF gösteren madde yüzdesinin %0'dan %15'e çıkarıldığı koşuldur (Finch,2005). Bu çalışmada, Finch'in (2005) çalışmasında bahsedilen koşula en yakın olarak 2000 kişilik örnekleme sahip, referans ve odak grubun yetenek dağılımlarının standart normal dağılıma sahip olduğu koşul için DMF gösteren madde yüzdesinin %10'dan %20'ye çıkarıldığı koşul alınabilir. Ancak bu değişen koşullar için güç oranındaki değişim örneklem büyüklüğünün değişmesi sonucunda gerçekleşen değişim miktarı kadar etkili olmamıştır. Yapılan çalışmalarda diğer taraftan ÇGÇN yönteminin gücünün değişen faktörlere karşı oldukça etkilenmez olduğundan bahsedilmiştir (Finch, 2005). Bu çalışmada ise örneklem büyüklüğündeki değişimin güç oranındaki değişimde oldukça etkili olduğu görülmüştür. İlgili çalışmalarda genel olarak ÇGÇN yönteminin güç oranının SIBTEST ve M-H gibi daha klasik yöntemlerdeki kadar yüksek olduğu ve hatta bazı durumlarda, 2-PL model kullanıldığında, bu yöntemlerinkine göre daha yüksek güç oranına sahip olduğu sonucuna ulaşılmıştır (Finch, 2005). Bu çalışmada da benzer sonuçlar söz konusudur, güç oranı neredeyse tüm koşullar için LR yönteminde olduğundan daha yüksektir. Finch (2005) çalışmasında 600 ve 1000 kişilik örneklemler kullanmıştır. 2-PL ve 3-PL model kullandığında örneklem büyüklüğü arttıkça güç oranlarında artış gözlenmiştir. Bu çalışmada 3-PL model kullanılmış ve testin uzunluğu Finch'in kullandığı iki test uzunluğunun (20 ve 50 madde) neredeyse ortalaması kadar (30 madde) alınmıştır ve analiz sonuçlarında örneklem

büyükliđü arttıkça gücün arttıđı görülmüştür. Bu açıdan iki çalıřma birbirini destekler niteliktedir.

Çalıřmanın bulguları LR yöntemi ile gerçekleştirilen analiz sonuçlarına bakıldıđında benzer bir deđişim göstermektedir. Güç oranının en yüksek deđeri 4000 kişilik örnekleme sahip, referans ve odak grubun yetenek dađılımlarının standart normal dađılım gösterdiđi, %20 DMF gösteren madde içeren koşulda gözlenmiřtir. Güç oranı neredeyse tüm koşullar için ÇGÇN yöntemine göre daha düşüktür. Ancak bu oran hiçbir koşulda 0.70'in altına düşmemiřtir.

Finch ve French'in (2007) çalıřmasında, ele alınan tüm koşullar için LR ve DFA yöntemlerinin güç oranları 0.70'in altında çıkmıřtır. Bahsedilen çalıřmada örneklem büyüklüđü arttıkça LR yönteminin güç oranı artarken, DFA yönteminin güç oranı azalmıřtır. Bu çalıřmada, örneklem büyüklüđü arttıkça ÇGÇN ve LR yöntemlerinin her ikisinin güç oranı artmıřtır. Bu iki çalıřma LR yönteminin örneklem büyüklüđü koşuluna göre güç oranındaki artış bakımından birbirini destekler niteliktedir.

## 5. SONUÇ ve ÖNERİLER

Bu bölümde araştırmadan elde edilen sonuçlara ve bu sonuçlara dayalı olarak sunulabilecek önerilere yer verilmiştir.

### 5.1. Sonuçlar

Bu araştırmada 8 farklı koşula yönelik yapılan analizler sonucunda ÇGÇN ve LR yöntemlerinin 1. tip hata oranları ve güçleri karşılaştırılmıştır. Karşılaştırılan koşullarda ele alınan değişkenler örneklem büyüklüğü, grupların yetenek dağılımları ve DMF gösteren madde yüzdesidir. ÇGÇN ve LR DMF belirleme yöntemlerinden hangisinin ele alınan hangi koşullarda daha iyi performans gösterdiği araştırılmıştır.

Araştırmada 8 koşul karşılaştırıldığında iki yöntemin 1. tip hata ve güç oranlarına yönelik sonuçlara aşağıda yer verilmiştir.

#### 1. tip hata oranlarına yönelik karşılaştırmalarda;

Bulgular genel olarak incelendiğinde DMF belirlemede ÇGÇN yaklaşımının örneklem büyüklüğünün 4000 olduğu koşullarda LR yöntemine göre, LR yöntemininse DMF gösteren madde yüzdesinin %10 olduğu koşullarda ÇGÇN yöntemine göre 1. tip hata oranlarının daha düşük olduğu görülmektedir. Genel olarak ÇGÇN yönteminin 1. tip hata oranının LR yöntemine göre daha düşük olduğu koşul sayısı daha fazla olmakla birlikte her iki yöntem için de 1. tip hata oranları çalışmada ele alınan her koşulda kabul edilebilir alfa düzeyinin ( $\alpha=0.05$ ) üzerindedir.

Örneklem büyüklüğünün artması ÇGÇN yöntemi için 1. tip hata oranını önemli ölçüde azaltırken LR yöntemi için 1. tip hata oranını çok fazla etkilememiştir. DMF'li madde yüzdelерinin artması ÇGÇN yönteminin 1. tip hata oranını pek fazla etkilemezken LR yöntemininkini arttırmıştır. Odak grubun yetenek dağılımının farklı olması ise her iki yöntemin de 1. tip hata oranlarını çok etkilememiştir. Sonuç olarak 1. tip hata oranları bakımından bu iki yöntem karşılaştırıldığında ÇGÇN yöntemi için örneklem büyüklüklerindeki değişim daha etkili olurken LR yöntemi için DMF'li madde yüzdelерindeki değişim daha etkili olmuştur.

## **Güç oranlarına yönelik karşılaştırmalarda;**

Alan yazında DMF belirleme yöntemlerinin performanslarının 1. tip hata oranları ve güç değerleri bakımından karşılaştırıldığı çalışmalarda, 1. tip hata oranlarının 0.10'dan yüksek olduğu koşullarda güç değerleri rapor edilmemiştir (Finch, 2005). Bu tarz durumlarda alfanın belirlenen seviyesindeki gücün standart tanımının anlamlı olmadığı belirtilmiştir (Finch, 2005). Bu çalışmada 1.tip hata oranının 0.10'dan yüksek olduğu koşullar için güç oranları yatay olarak gösterilmiştir ve koşullara göre yöntemlerin güç oranlarını karşılaştırmak amacıyla tüm güç oranlarına yer verilmiştir.

Bulgular genel olarak incelendiğinde, her iki yöntemin de tüm koşullar için güç oranları kabul edilebilir seviyenin (0.70) üzerindedir. Örneklem büyüklüğünün daha fazla olduğu koşullar için ÇGÇN yönteminin güç oranları oldukça yüksektir. LR yöntemi içinse örneklem büyüklüğünün daha fazla olduğu ve her iki grubun da yetenek dağılımlarının standart normal dağılıma sahip olduğu koşullarda güç oranları oldukça yüksek gözükmektedir. Genel olarak ÇGÇN yönteminin güç oranları, birey sayısının 2000 kişi olduğu, DMF gösteren madde yüzdesinin %10 olduğu ve referans ve odak grubun yetenek dağılımlarının standart normal dağılıma sahip olduğu, tek bir koşul dışında LR yöntemine göre her zaman daha yüksektir. Her iki yöntem için de güç oranları her koşulda 0.70'in üzerindedir.

Örneklem büyüklüğünün artması her iki yöntem için güç oranını artırmıştır. Odak grubun yetenek dağılımının referans grubun yetenek dağılımından farklı olması her iki yöntemin güç oranını düşürmüştür. Yetenek dağılımındaki bu değişikliğin düşüş miktarı olarak LR yönteminin güç oranını daha fazla etkilediği görülmektedir. DMF gösteren madde yüzdelerinin artması her iki yöntemin güç oranını az miktarda artırmıştır. Sonuç olarak güç oranlarındaki değişime bakıldığında her iki yöntem için en etkili değişkenin örneklem büyüklüğü olduğu görülmektedir.

## **5.2.Öneriler**

Bu araştırmanın ÇGÇN yöntemiyle gerçekleştirilecek DMF belirleme çalışmalarına referans olacağı ve gerçek sınav sonuçlarının analizlerinde örneklem büyüklüğü ve yetenek dağılımlarına göre uygun DMF belirleme yöntemine karar vermede araştırmacılara kolaylık sağlayacağı düşünülmektedir.

### **5.2.1. Araştırmanın Sonuçlarına Dayalı Öneriler**

Bu çalışma ile testin özellikleri ve testin uygulanacağı grubun özelliklerine uygun DMF belirleme tekniklerinin seçiminde araştırmacılara güvenilir bir kaynak sağlanabilir. Böylece daha güvenilir DMF belirleme teknikleri kullanılarak testlerin daha eşit ve iyi hale gelmesi sağlanabilir.

Bu araştırmadan elde edilen sonuçlara dayanarak DMF analizi çalışmalarında, daha küçük örneklerde, yaklaşık 2000 gibi, LR yönteminin kullanılması önerilebilir. Bu yöntem ile DMF gösteren maddeler belirlenmeye çalışıldıktan sonra bu maddelerle ilgili yanlılık araştırması yapmak için uzman kanısına başvurulabilir. Daha büyük örnekler içinse, yaklaşık 4000 gibi, ÇGÇN yöntemi kullanılabilir. Bu yöntem ile DMF gösteren maddeler belirlendikten sonra diğer yöntemlerde olduğu gibi yanlılık araştırması yapmak için uzman kanısına başvurulabilir.

### **5.2.2. Araştırmacılar İçin Öneriler**

Analizlerde kullanılan Mplus ve SAS programları 2000 ve 4000 kişilik veriler için oldukça hızlı çalışan ve kolay yorumlanabilen sonuçlar üreten, kullanışlı programlardır. Gelecekte yapılacak çalışmalarda araştırmacılar rahatlıkla bu programları kullanabilirler.

Çalışmada iki farklı örneklem büyüklüğü, odak gruba yönelik iki farklı yetenek dağılımı ve iki farklı DMF gösteren madde yüzdesi koşullarının çaprazlanmasıyla toplam 8 farklı koşul ele alınmıştır. Testteki madde sayısı koşullarda sabit tutulmuştur. Gelecek çalışmalarda testteki madde sayısı artırılarak uzun testlerde sonuçların nasıl etkilendiği incelenebilir.

ÇGÇN yönteminin farklı örneklem büyüklüklerinde nasıl performans gösterdiği önemli bir konudur. Çalışmada 2000 ve 4000 bireylik iki farklı örneklem büyüklüğü ele alınmıştır. Ancak 4000 bireylik örneklem büyüklüğünde bile istenilen 1. tip hata oranlarına ulaşamamıştır. Bundan sonraki çalışmalarda daha büyük örneklem büyüklükleri üzerinde çalışılarak ÇGÇN yöntemi için ideal örneklem büyüklüğü araştırılabilir.

Çalışmada referans grup büyüklüğü ve odak grup büyüklüğü arasındaki oran 1:1 olarak alınmıştır. Ancak gerçek sınavlarda bu iki grubun örneklem büyüklükleri oranlarıyla ilgili çok farklı durumlar söz konusu olabilir. Bu nedenle farklı oranlar kullanılarak çalışmalar yapılabilir.

Çalışma 3-PL modele dayalı verilerle yürütülmüştür. 2-PL modele dayalı verilerle benzer çalışmalar yürütülebilir ve bu çalışmalar arasında karşılaştırmalar yapılabilir.

Bu araştırmada iki kategorili puanlanan veriler ile çalışılmıştır. Çok kategorili puanlanan ya da iki kategorili ve çok kategorili puanlanan verilerin birlikte kullanıldığı veriler üzerinde çalışmalar yapılabilir.

Çalışmada benzetim verileri kullanılmıştır. Buna ek olarak gerçek veriler de kullanılıp sonuçlar karşılaştırılabilir.

## KAYNAKÇA

- Atalay, K., Gök, B., Kelecioğlu, H. ve Arslan, N. (2012). Değişen madde fonksiyonunun belirlenmesinde kullanılan farklı yöntemlerin karşılaştırılması: bir simülasyon çalışması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 43, 270- 281.
- Bekçi, B. (2007). *Orta öğretim kurumları öğrenci seçme ve yerleştirme sınavının değişen madde fonksiyonlarının cinsiyete ve okul türüne göre incelenmesi*. Yüksek Lisans Tezi, Hacettepe Üniversitesi, Ankara.
- Braddy, P. W., Meade, A. W. & Johnson, E. C. (2006). Practical implications of using different tests of measurement invariance for polytomous measures. Paper presented at the 21st Annual Conference of the Society for Industrial and Organizational Psychology, Dallas, TX.
- Camili, G., & Shephard, L. A. (1994). *Methods for identifying biased test items*. London: Sage Publications.
- Clauser, B. E., Mazor, K. & Hambleton, R. K. (1991). Influence of the criterion variable on the identification of differential functioning test items using the Mantel-Haenszel statistic. *Applied Psychological Measurement*, 15, 353-359.
- Clauser, B. E., Mazor, K. M., & Hambleton, R. K. (1993). The effects of purification of matching criterion on the identification of DIF using the Mantel–Haenszel procedure. *Applied Measurement in Education*, 6, 269-279.
- Crane, P. K., Belle, G., & Larson, E. B. (2004). Test bias in a cognitive test: differential item functioning in the CASI. *Statistics in Medicine*, 23, 241-256.
- Finch, H. (2005). The MIMIC model as a method for detecting DIF: comparison with Mantel-Haenszel, SIBTEST, and the IRT likelihood ratio. *Applied Psychological Measurement*, 29, 278-295.
- Finch, W. H. & French, B. F. (2007). Detection of crossing differential item functioning: a comparison of four methods. *Educational and Psychological Measurement*, 67, 565-582.
- Finch, W. H. & French, B. F. (2011). Estimation of MIMIC model parameters with multilevel data. *Structural Equation Modeling: A Multidisciplinary Journal*, 18(2), 229-252.
- Fleishman, J. A., Spector, W. D., & Altman, B. M. (2002). Impact of differential item functioning on age and gender differences in functional disability. *Journal of Gerontology: Social Sciences*, 57B(5), S275-S283.
- French, B. F., Maller, S. J. (2007). Iterative Purification and Effect Size Use With Logistic Regression for Differential Item Functioning Detection. *Educational and Psychological Measurement*, 67, 378-379.
- French, A. W., & Miller, T. R. (1996). Logistic regression and its use in detecting differential item functioning in polytomous items. *Journal of Educational Measurement*, 33, 315-332.

- Gonzalez-Roma, V., Hernandez, A. & Gomez-Benito, J. (2006). Power and Type one error of the mean and covariance structure analysis model for detecting differential item functioning in graded response items. *Multivariate Behavioral Research*, 41, 29-53.
- Hamilton, L. S. (1999). Detecting gender-based differential item functioning on a constructed-response science test. *Applied Measurement in Education*, 12(3), 211-235.
- Holland, P. W., & Wainer, H. (1993). *Differential item functioning*. London: Lawrence Erlbaum Associates.
- Jodoin, M. G., & Gierl, M. J. (2001). Evaluating Type I error and power rates using an effect size measure with the logistic regression procedure for DIF detection. *Applied Measurement in Education*, 14, 329-349.
- Kan, A., Sünbül, Ö. ve Ömür, S. (2013). 6.-8. sınıf seviye belirleme sınavları alt testlerinin çeşitli yöntemlere göre değişen madde fonksiyonlarının incelenmesi. *Mersin Üniversitesi Eğitim Fakültesi Dergisi*, 9(2), 207-222.
- Kim, S.-H., Cohen, A. S., & Park, T.-H. (1995). Detection of differential item functioning in multiple groups. *Journal of Educational Measurement*, 32, 261-276.
- Kim, E. S. & Yoon, M. (2011). Testing measurement invariance: a comparison of multiple-group categorical CFA and IRT. *Structural Equation Modeling: A Multidisciplinary Journal*, 18(2), 212-228.
- Kristjansson, E., Aylesworth, R., McDowell, I., & Zumbo, B. (2005). A comparison of four methods for detecting differential item functioning in ordered response items. *Educational and Psychological Measurement*, 65, 935-953.
- Li, H.-H., & Stout, W. (1996). A new procedure for detection of crossing DIF. *Psychometrika*, 61, 647-677.
- MacIntosh, R., & Hashim, S. (2003). Variance estimation for converting MIMIC model parameters to IRT parameters in DIF analysis. *Applied Psychological Measurement*, 27(5), 372-379.
- Magis, D., Beland, S., Teurlinckx, F., & Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods*, 42(3), 847-862.
- Mazor, K. M., Kanjee, A., & Clauser, B. E. (1995). Using logistic regression and the Mantel-Haenszel with multiple ability estimates to detect differential item functioning. *Journal of Educational Measurement*, 32, 131-144.
- Mazor, K. M., Clauser, B. E., & Hambleton, R. K. (1994). Identification of nonuniform differential item functioning using a variation of the Mantel-Haenszel procedure. *Educational & Psychological Measurement*, 54, 284-291.
- Millsap, R. E. & Everson, H. T. (1993). Methodology review: statistical approaches for assessing measurement bias. *Applied Psychological Measurement*, 17, 297-334.



- Muthen, B. O. (1988). Some uses of structural equation modeling in validity studies: Extending IRT to external variables. In H. Wainer & H. Braun (Eds.), *Test validity* (pp. 213-238). Hillsdale, NJ: Lawrence Erlbaum.
- Muthen, B. O., Kao, C. F., & Burstein, L. (1991). Instructionally sensitive psychometrics: Application of a new IRT-based detection technique to mathematics achievement test items. *Journal of Educational Measurement*, 28(1), 1-22.
- Narayanan, P., & Swaminathan, H. (1996). Identification of items that show nonuniform DIF. *Applied Psychological Measurement*, 20, 257-274.
- Osterlind, S. J., & Everson, H. T. (2009). *Differential item functioning* (2nd ed.). Thousand Oaks, CA: Sage.
- Penfield, R. D. (2001). Assessing differential item functioning among multiple groups: a comparison of three Mantel-Haenszel procedures. *Applied Measurement In Education*, 14(3), 235-259.
- Penfield, R. D. (2003). Applying the Breslow–Day test of trend in odds ratio heterogeneity to the analysis of nonuniform DIF. *Alberta Journal of Educational Research*, 49, 231-243.
- Penfield, R. D., & Camilli, G. (2007). Differential item functioning and item bias. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics: Vol. 26. Psychometrics* (pp. 125-167). Amsterdam: Elsevier.
- Rogers, H. J., & Swaminathan, H. (1993). A Comparison of Logistic Regression and Mantel-Haenszel Procedures for Detecting Differential Item Functioning. *Applied Psychological Measurement*, 17, 105-106.
- Shepard, L. A., Camilli, G., & Averill, M. (1981). Comparison of procedures for detecting test item bias with both internal and external ability criteria. *Journal of Educational Statistics*, 6, 317-375.
- Shih, C.-L., & Wang, W.-C. (2009). Differential item functioning detection using the multiple indicators, multiple causes MIMIC method with a pure short anchor. *Applied Psychological Measurement*, 33, 184-199.
- Soares, T. M., Gonçalves, F. B., & Gamerman, D. (2009). An integrated Bayesian model for DIF analysis. *Journal of Educational & Behavioral Statistics*, 34, 348-377.
- Swaminathan, H., & Rogers, H. J. (1990). Detecting Differential Item Functioning Using Logistic Regression Procedures. *Journal of Educational Measurement*, 27, 362-363.
- Vaughn, B. K., & Wang, Q. (2010). DIF trees: Using classification trees to detect differential item functioning. *Educational and Psychological Measurement*, 70, 941-952.
- Wang, W.-C. & Shih, C.-L. (2010). MIMIC methods for assessing differential item functioning in polytomous items. *Applied Psychological Measurement*, 34, 166-180.

- Wang, W.-C., Shih, C.-L. & Yang, C.-C. (2009). The MIMIC method with scale purification for detecting differential item functioning. *Educational and Psychological Measurement, 69*, 713-731.
- Wang, W.-C., & Su, Y.-H. (2004). Effects of average signed area between two item characteristic curves and test purification procedures on the DIF detection via the Mantel–Haenszel method. *Applied Measurement in Education, 17*, 113-144.
- Wiberg, M. (2007). *Measuring and detecting differential item functioning in criterion-referenced licensing test*, [Çevirim-içi:  
[http://www.edusci.umu.se/digitalAssets/59/59534\\_em-no-60.pdf](http://www.edusci.umu.se/digitalAssets/59/59534_em-no-60.pdf)], Erişim Tarihi: 8 Mayıs 2014.
- Woods, C. M. (2009). Evaluation of MIMIC-Model methods for DIF testing with comparison to Two-Group Analysis. *Multivariate Behavioral Research, 44*, 1-27.
- Woods, C. M., Oltmanns, T. F., & Turkheimer, E. (2009). Illustration of MIMIC-model DIF testing with the schedule for nonadaptive and adaptive personality. *J Psychopathol Behav Assess, 31*, 320-330.
- 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.

## **EKLER DİZİNİ**

## EK 1. SK 1'e Ait ÇGÇN Yöntemi Kod Dosyası

```
Ngroups = 2;  
Variable:  
Names are group i1-i31;  
Categorical i1-i31;  
Usevariables are group i1-i31;  
Analysis:  
Estimator = wls;  
Model:  
f1 by i1-i31;  
f1 on group;  
i1 on group;  
i2 on group;  
i3 on group;  
i4 on group;  
i5 on group;  
. . .  
i30 on group;  
i31 on group@0;  
Output: Modindices (all)
```

## EK 2. SK 1'e Ait SAS Kod Dosyası

```
libname lrdif "D:\LR\LR1";
%macro main(rep, num);
data scdr;
infile "D:\Data\SC1\scdr&rep..dat";
input group 1
      item1 3
          item2 5
          item3 7
          item4 9
          item5 11
          item6 13
      item7 15
          item8 17
          item9 19
          item10 21
          item11 23
          item12 25
          item13 27
          item14 29
          item15 31
          item16 33
          item17 35
          item18 37
          item19 39
          item20 41
      item21 43
      item22 45
          item23 47
          item24 49
          item25 51
          item26 53
          item27 55
          item28 57
          item29 59
          item30 61;
total = item1 + item2 + item3 + item4 + item5 + item6 + item7 + item8 + item9 +
item10
      + item11 + item12 + item13 + item14 + item15 + item16 + item17 + item18 +
item19 + item20
      + item21 + item22 + item23 + item24 + item25 + item26 + item27 + item28 +
item29 + item30;
run;
proc means noprint data=scdr;
var total;
output out=tmeans mean=tmean stddev=tsd;
run;
proc means noprint data=scdr;
```

```

var group;
output out=gmeans mean=gmean stddev=gsd;
run;
data scdr;
set scdr;
if _N_=1 then set tmeans;
diff1=total-tmean;
totalz=diff1/tsd;
if _N_=1 then set gmeans;
diff2=group-gmean;
groupz=diff2/gsd;
run;
proc logistic data = scdr descending outest = modela&num noprint;
model item&num = total / scale = none aggregate rsquare;
run;
proc logistic data = scdr descending outest = modelb&num noprint;
model item&num = total group / scale = none aggregate rsquare;
run;
data modela&num (keep = _NAME_ LOGLI);
set modela&num;
LOGLI = -2*_LNLIKE_;
run;
data modelb&num (keep = _NAME_ LOGLII);
set modelb&num;
LOGLII = -2*_LNLIKE_;
run;
data model&num;
merge modela&num modelb&num;
by _NAME_;
run;
data model&num;
set model&num;
GSQUAREU = LOGLI-LOGLII;
run;
data lregdr&rep;
length _NAME_ $6;
set model1 model2 model3 model4 model5 model6 model7 model8 model9
model10
    model11 model12 model13 model14 model15 model16 model17 model18
model19 model20
    model21 model22 model23 model24 model25 model26 model27 model28
model29 model30;
run;
%mend main;
%main(1, 1); run;
%main(1, 2); run;
%main(1, 3); run;
%main(1, 4); run;
%main(1, 5); run;
.

```

```

.
.
%main(1, 30); run;
%main(2, 1); run;
%main(2, 2); run;
%main(2, 3); run;
%main(2, 4); run;
%main(2, 5); run;
.
.
.
%main(100, 30); run;
%macro main(num);
data error100;
set lregdr1 lregdr2 lregdr3 lregdr4 lregdr5 lregdr6 lregdr7 lregdr8 lregdr9 lregdr10
    lregdr11 lregdr12 lregdr13 lregdr14 lregdr15 lregdr16 lregdr17 lregdr18 lregdr19
lregdr20
    lregdr21 lregdr22 lregdr23 lregdr24 lregdr25 lregdr26 lregdr27 lregdr28 lregdr29
lregdr30
    lregdr31 lregdr32 lregdr33 lregdr34 lregdr35 lregdr36 lregdr37 lregdr38 lregdr39
lregdr40
    lregdr41 lregdr42 lregdr43 lregdr44 lregdr45 lregdr46 lregdr47 lregdr48 lregdr49
lregdr50
    lregdr51 lregdr52 lregdr53 lregdr54 lregdr55 lregdr56 lregdr57 lregdr58 lregdr59
lregdr60
    lregdr61 lregdr62 lregdr63 lregdr64 lregdr65 lregdr66 lregdr67 lregdr68 lregdr69
lregdr70
    lregdr71 lregdr72 lregdr73 lregdr74 lregdr75 lregdr76 lregdr77 lregdr78 lregdr79
lregdr80
    lregdr81 lregdr82 lregdr83 lregdr84 lregdr85 lregdr86 lregdr87 lregdr88 lregdr89
lregdr90
    lregdr91 lregdr92 lregdr93 lregdr94 lregdr95 lregdr96 lregdr97 lregdr98 lregdr99
lregdr100;
if _name_="item&num";
run;
data error100;
set error100;
if gsquareu>3.84 then P=2;
run;
proc freq data=error100 noprint;
tables P / out=item&num;
run;
data error100SC1u;
set item1 item2 item3 item5 item6 item7 item8 item9 item10
    item11 item12 item13 item14 item16 item17 item18 item19 item20
    item21 item22 item23 item24 item25 item26 item28 item29 item30;
if P=2;
run;
proc means data=error100SC1u noprint;
var count;

```

```

output out=total100SC1u sum=sum1;
run;
data total100SC1u;
set total100SC1u;
ExpVal = (_FREQ_*100)*0.05;
run;

data lrdif.totale1100SC1u (drop = _TYPE_ _FREQ_);
set total100SC1u;
Percent = (sum1*0.05)/ExpVal;
ALevel = 0.0500000000;
Rep = 100;
SC = 1;
run;
%mend main;
%main(1); run;
%main(2); run;
%main(3); run;
%main(5); run;
.
.
.
%main(30); run;
%macro main(num);
data power100;
set lregdr1 lregdr2 lregdr3 lregdr4 lregdr5 lregdr6 lregdr7 lregdr8 lregdr9 lregdr10
    lregdr11 lregdr12 lregdr13 lregdr14 lregdr15 lregdr16 lregdr17 lregdr18 lregdr19
lregdr20
    lregdr21 lregdr22 lregdr23 lregdr24 lregdr25 lregdr26 lregdr27 lregdr28 lregdr29
lregdr30
    lregdr31 lregdr32 lregdr33 lregdr34 lregdr35 lregdr36 lregdr37 lregdr38 lregdr39
lregdr40
    lregdr41 lregdr42 lregdr43 lregdr44 lregdr45 lregdr46 lregdr47 lregdr48 lregdr49
lregdr50
    lregdr51 lregdr52 lregdr53 lregdr54 lregdr55 lregdr56 lregdr57 lregdr58 lregdr59
lregdr60
    lregdr61 lregdr62 lregdr63 lregdr64 lregdr65 lregdr66 lregdr67 lregdr68 lregdr69
lregdr70
    lregdr71 lregdr72 lregdr73 lregdr74 lregdr75 lregdr76 lregdr77 lregdr78 lregdr79
lregdr80
    lregdr81 lregdr82 lregdr83 lregdr84 lregdr85 lregdr86 lregdr87 lregdr88 lregdr89
lregdr90
    lregdr91 lregdr92 lregdr93 lregdr94 lregdr95 lregdr96 lregdr97 lregdr98 lregdr99
lregdr100;
if _name_="item&num";
run;
data power100;
set power100;
if gsquareu>3.84 then P=2;
run;

```



```

proc freq data=power100 noprint;
tables P / out=item&num;
run;
data power100SC1u;
set item4 item15 item27;
if P=2;
run;
proc means data=power100SC1u noprint;
var count;
output out=total100SC1u sum=sum1;
run;
data total100SC1u;
set total100SC1u;
ExpVal = (_FREQ_*100)*0.95;
run;
data lrdif.totalp1100SC1u (drop = _TYPE_ _FREQ_);
set total100SC1u;
Percent = (sum1*0.95)/ExpVal;
ALevel = 0.9500000000;
Rep = 100;
SC = 1;
run;
data power100SC1u4;
set item4;
if P=2;
run;
proc means data=power100SC1u4 noprint;
var count;
output out=total100SC1u4 sum=sum1;
run;
data total100SC1u4;
set total100SC1u4;
ExpVal = (_FREQ_*100)*0.95;
run;
data lrdif.totalp1100SC1u4 (drop = _TYPE_ _FREQ_);
set total100SC1u4;
Percent = (sum1*0.95)/ExpVal;
ALevel = 0.9500000000;
Rep = 100;
SC = 1;
run;
data power100SC1u15;
set item15;
if P=2;
run;
proc means data=power100SC1u15 noprint;
var count;
output out=total100SC1u15 sum=sum1;
run;
data total100SC1u15;

```

```

set total100SC1u15;
ExpVal = (_FREQ_*100)*0.95;
run;
data lrdif.totalp1100SC1u15 (drop = _TYPE_ _FREQ_);
set total100SC1u15;
Percent = (sum1*0.95)/ExpVal;
ALevel = 0.9500000000;
Rep = 100;
SC = 1;
run;
data power100SC1u27;
set item27;
if P=2;
run;
proc means data=power100SC1u27 noprint;
var count;
output out=total100SC1u27 sum=sum1;
run;
data total100SC1u27;
set total100SC1u27;
ExpVal = (_FREQ_*100)*0.95;
run;
data lrdif.totalp1100SC1u27 (drop = _TYPE_ _FREQ_);
set total100SC1u27;
Percent = (sum1*0.95)/ExpVal;
ALevel = 0.9500000000;
Rep = 100;
SC = 1;
run;
%mend main;
%main(4); run;
%main(15); run;
%main(27); run;

```

### EK 3. Etik Kurulu İzin Muafiyet Formu



**HACETTEPE ÜNİVERSİTESİ  
EĞİTİM BİLİMLERİ ENSTİTÜSÜ  
TEZ ÇALIŞMASI ETİK KURUL İZİN MUAFİYETİ FORMU**

Form: 40

**HACETTEPE ÜNİVERSİTESİ  
EĞİTİM BİLİMLERİ ENSTİTÜSÜ  
EĞİTİM BİLİMLERİ ANABİLİM DALI BAŞKANLIĞI'NA**

Tarih: / /

Tez Başlığı / Konusu: ÇOKLU GÖSTERGELER ÇOKLU NEDENLER VE LOJİSTİK REGRESYON YÖNTEMLERİNİN DEĞİŞEN MADDE FONKSİYONUNU BELİRLEME PERFORMANSLARI

Yukarıda başlığı/konusu gösterilen tez çalışmam:

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

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

Gereğini saygılarımla arz ederim.

Tarih ve İmza

**Adı Soyadı:** Seçil ARSLAN

**Öğrenci No:** N11224121

**Anabilim Dalı:** Eğitim Bilimleri Anabilim Dalı

**Programı:** Eğitimde Ölçme ve Değerlendirme Bilim Dalı

**Statüsü:**  Y.Lisans  Doktora  Bütünleşik Dr.

#### **DANIŞMAN GÖRÜŞÜ VE ONAYI**

Doç. Dr. Burcu ATAR

## ÖZGEÇMİŞ

<b>Adı Soyadı</b>	Seçil ARSLAN
<b>Doğum Yeri</b>	Malatya
<b>Doğum Yılı</b>	1987
<b>Medeni Hali</b>	Bekâr

### Eğitim ve Akademik Durumu

<b>Lisans</b>	Hacettepe Üniversitesi- Fen Bilgisi Öğretmenliği	2010
<b>Yüksek Lisans</b>	Hacettepe Üniversitesi- Eğitimde Ölçme ve Değerlendirme	2012
<b>Yabancı Dil</b>	İngilizce	
<b>İş Deneyimi</b>	Recep Tayyip Erdoğan Üniversitesi- Araştırma Görevlisi	2013
<b>İş Deneyimi</b>	Hacettepe Üniversitesi- Araştırma Görevlisi	2013- Halen