



## Değişen Madde Fonksiyonunun Belirlenmesinde Kullanılan Yöntemlerde I. Tip Hata ve Güç Çalışması<sup>1</sup>

### Type I Error Rates and Power Study of Several Differential Item Functioning Determination Methods

Önder Sünbül, Yrd. Doç. Dr., Mersin Üniversitesi, [ondersunbul@mersin.edu.tr](mailto:ondersunbul@mersin.edu.tr)  
Seçil Ömür Sünbül, Yrd. Doç. Dr., Mersin Üniversitesi, [secilomur@mersin.edu.tr](mailto:secilomur@mersin.edu.tr)

**ÖZ.** Bu çalışmada, Değişen Madde Fonksiyonunun belirlenmesinde kullanılan çeşitli yöntemlerin belirli koşullardaki I. Tip hata ve güç oranlarının belirlenmesi amaçlanmıştır. I. Tip hata ve güç oranlarının belirlenmesi simülasyon çalışmasıyla yapılmıştır. Klasik Test Kuramı'na dayalı değişen madde fonksiyonunun belirlenmesinde kullanılan yöntemlerde, örneklem büyüklüğünün, madde sayısının, referans grubundaki örneklem büyüklüğünün toplam örneklem büyüklüğüne oranlarının değişimlenmesinin I. Tip hata oranları ortalamalarını çok fazla değiştirmedeği görülmektedir. Madde Tepki Kuramı'na dayalı değişen madde fonksiyonunun belirlenmesinde kullanılan yöntemlerde ise, örneklem büyüklüğünün artmasıyla I. Tip hata oranlarının ortalamasının azaldığı, madde sayısının artmasıyla I. Tip hata oranlarının ortalamasının değişmediği ve referans ve odak gruplardaki örneklem büyüklüğü oranlarının eşit olduğu koşulda I. Tip hata oranları ortalamasının daha düşük olduğu görülmektedir. Güç oranlarının belirlenmesi için yapılan çalışmada, hem Klasik Test Kuramı'na hem de Madde Tepki Kuramı'na dayalı değişen madde fonksiyonu belirlenmesinde kullanılan yöntemlerde, örneklem büyüklüğünün ve değişen madde fonksiyonu miktarının artmasıyla güç oranlarının ortalamasının da arttığı görülmektedir. Tüm yöntemlerde, madde sayısının değişimlenmesinin güç oranlarının ortalamalarını çok fazla değiştirmedeği, ayrıca, değişen madde fonksiyonlu madde sayısının oranının artırılmasının ve referans ve odak gruplardaki örneklem büyüklüğü oranlarının farklı olması güç oranlarının ortalamasını azalttığı görülmektedir.

**Anahtar Kelimeler:** Değişen Madde Fonksiyonu, I. Tip Hata Oranı, Güç Çalışması

**ABSTRACT.** In this study Type I Error rates and power study of several differential item functioning determination methods were investigated for several conditions. For this purpose simulation studies were conducted for methods which belong to Classical Test Theory (CTT) and Item Response Theory (IRT). The results for Type I Error rates of differential item functioning determination methods under the framework of CTT show that none of manipulation criteria have significant effect on average Type I Error rates. For IRT based methods it has been observed that the increase of sample size causes the decrease of average Type I error rates while increasing the number of items has no significant effect on average Type I Error rates, and that the equal ratio of reference and focal sample sizes cause lower average Type I error rates. The results for power study for both differential item functioning determination methods of CTT and IRT show that the increase of sample size and the magnitude of differential item functioning cause the increase of average power rates. No significant effect of the number of items on average power rates could be observed.

**Keywords:** Differential item functioning, Type 1 error rates, Power study

#### SUMMARY

**Purpose and significance:** During the formal education process many decisions are made for students by using the results of educational tests. These tests must hold two important psychometric facilities like as reliability and validity. However, several treats occur during testing process which jeopardize reliability and validity. One of those treats is biased test items. Biased items generate biased test results and biased tests results generate base for biased decisions for test takers. This sequence ends with the violation of validity. Before labeling an item as biased, differential item functioning (DIF) investigation must be conducted. DIF occurs when two groups with equal abilities

<sup>1</sup> Bu çalışma IV.Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi'nde sözlü bildiri olarak sunulmuştur.

generate a different probability of answering an item correctly. There are several DIF detection methods in the literature which belong to Classical Test Theory and Item Response Theory. Since reliability of these indexes will depend several conditions, reliability of these DIF methods must be investigated for several conditions in term of Type I error rate and power rate.

To evaluate the precision of decisions which were made by using these DIF indexes is an important issue. There are several studies in literature which focus on Type I error rates and power of DIF detection indexes. This study aims to contribute to literature by checking the Type I error and power rates of Manthel- Haenszel, Logistic Regression, Lord's Chi Square and Raju Non-Signed Area DIF detection methods.

**Methods:** To investigate Type I error rates of several DIF determination methods, non-DIF data was generated according to 2PL IRT model for several conditions with 100 replications. After the data generation, average rate of false positive DIF items has been determined and graphed. To investigate power of several DIF determination methods, data which includes DIF items was generated according to 2PL IRT model for several conditions with 100 replication. After the data generation, average rate of true positive DIF items has been determined and graphed. Both of Type I Error and Power data were generated by using codes which were written in R programming language. All the DIF analyses were conducted by using "difR" R package.

**Results:** The results for Type I Error rates of DIF determination methods of CTT show that none of the manipulation criteria have any significant effect on average Type I Error rates. For IRT methods it has been observed that the increase of sample size causes the decrease of average Type I error rates while increasing the number of items has no significant effect on average Type I Error rates, and that the equal ratio of reference and focal sample sizes cause lower average Type I error rates. The results for power study for both of the frameworks of CTT and IRT differential item functioning determination methods show that the increase of sample size and the magnitude of differential item functioning cause the increase of average power rates. No significant effect of the number of items on average power rates could be observed.

**Discussion and Conclusions:** The results showed that the increase of the ratio of the items which have differential item functioning decreased the average power rates. Another cause of the decrease of average power rate has been found out to be the inequality of sample sizes of reference and focal group for every method.

---

## GİRİŞ

Bireylere çeşitli amaçlar doğrultusunda testler ve ölçekler uygulanmakta ve bireylerin bu araçlardan aldıkları puanlara bakılarak onlar hakkında önemli kararlar verilmektedir. Bu kararların yerinde ve isabetli olabilmesi için bireylere uygulanan ölçme araçlarının taşıması gereken bir takım özelliklere sahip olması gerekmektedir. Bu özellikler, güvenilirlik ve geçerliktir. Bir başka ifadeyle, bir ölçme aracı ne kadar güvenilir ve geçerli ise, o ölçme aracından elde edilecek sonuçlar o kadar gerçeğe yakın ve objektif olacaktır. Ölçmek istenilen özellik hakkında doğru bilgi elde edebilmek ve elde edilen sonuçlara dayanarak doğru kararlar verebilmek için gerçek ve gözlenen puanların farklılaşmasının, bir başka deyişle hatanın, mümkün olduğunca azolması gerekmektedir. Ölçme sonuçlarına karışan hatalardan bir tanesi de sistematik hatadır ve sistematik hata geçerliği tehdit eden bir unsurdur.

Yanlılık, sistematik hata kapsamında yer alır. Yanlılık, farklı alt gruplardaki bireylerin test puanlarının, buldukları gruba bağlı olarak sistematik hata içermesidir (Camilli ve Shepard, 1994). Bireylere uygulanan sınavın, sınava girenler içinde yer alan hiçbir alt gruba avantaj sağlamaması, bir başka ifadeyle, sınavın yanlı olmaması beklenir.

Madde yanlılığının belirlenebilmesi için öncelikle, testin tek boyutlu olup olmadığının ve sonra maddelerin Değişen Madde Fonksiyonu (DMF) içerip içermediğinin belirlenmesi gerekmektedir. DMF, aynı yetenek düzeyinde, fakat farklı gruplarda yer alan bireylerin maddeyi doğru cevaplandırma olasılıklarının farklılaşması olarak tanımlanır (Zumbo, 1999; Hambleton,

Swaminathan ve Rogers, 1991). Başka bir ifadeyle, bir madde DMF içeriyorsa, o madde, aynı yetenek düzeyinde ancak farklı alt gruplarda yer alan bireylerden herhangi bir grubuna avantaj sağladığını göstermektedir. Bu açıdan bakıldığında ülkemizde yapılan çeşitli sınavlarda değişen madde fonksiyonunu ya da madde yanlılığını inceleyen birçok çalışmanın yer aldığı görülmektedir (Öğretmen ve Doğan, 2004; Acar, 2008; Anıl ve Güzeller, 2010; Anıl, Güzeller, Çokluk ve Şekercioğlu, 2010; Kelecioğlu, Atalay ve Öztürk, 2010; Doğan ve Sevindik, 2011; Çepni, 2011; Karakaya ve Kutlu, 2012, Kan, Sünbül ve Ömür, 2013).

DMF'nin ortaya çıkmasının iki nedeni olabilir. Bunlar alt gruplar arasındaki gerçek farklılığı gösteren madde etkisi ve madde yanlılığıdır (Camilli ve Shepard, 1994). Madde etkisi, farklı gruplarda yer alan bireylerin bir maddeyi doğru cevaplama olasılıklarının, söz konusu o madde ile ölçülmek istenilen özellik bütünde bir farklılaşma olduğunda görülür (Zumbo,1999).Bir başka ifadeyle, madde etkisi madde üzerinde farklı grupların farklı performans göstermeleridir ve bu farklılık maddenin yanlı olduğu anlamını taşımaz, bu farklılık gruplar arasındaki gerçek yetenek farkıdır. Ancak, maddeden kaynaklı olarak, gruplarda farklı performanslar gözleniyorsa bu durumda istatistiksel yanlıktan söz edilebilir. İstatistiksel yanlılık, bazı parametrelerin kestirilmesinde, gruplara göre farklılık olmasını, bir başka ifadeyle, bir veya daha çok sayıdaki parametreye dair kestirimin olması gerektiğinden çok düşük ya da yüksek olması halini ifade eder (Camilli ve Shepard, 1994).

DMF'yi belirlemek amacıyla Klasik Test Kuramı'na dayalı olarak; varyans analizi, Dönüştürülmüş Madde Güçlüğü (TID), Mantel-Haenszel (MH), Lojistik Regresyon (LR) ve Madde Tepki Kuramı'na dayalı olarak, olabilirlik oran testi, Lord'un  $\chi^2$ , Raju'nun alan ölçüleri gibi çok sayıda istatistiksel yöntem geliştirilmiştir (Hambleton, Swaminathan ve Rogers, 1991; Camilli ve Shepard,1994; Osterlind,1983).

Bu araştırmada DMF'nin belirlenmesinde, MH, LR, Lord'un  $\chi^2$  ve Raju'nun alan ölçüsü yöntemleriyle sınırlandırılmıştır ve DMF belirlemede kullanılan yöntemlerin I. Tip hata ve güç oranlarının çeşitli koşullar altında nasıl değiştiği belirlenmeye çalışılmıştır. Aşağıda araştırmada kullanılacak olan yöntemler kısaca açıklanmıştır:

*Mantel Haenszel (MH)*: Bu yöntem  $\chi^2$  tekniğine dayanmaktadır. Bu teknikte, gözlenen puanlar birkaç kategoriye ayrılır. Gruplar her kategori için, maddeyi doğru cevaplama oranları açısından karşılaştırılır. Grupların kategorilerde verdiği cevaplar oranından  $\chi^2$  değeri elde edilir. Hesaplanan  $\chi^2$  değeri, ilgili serbestlik derecesi ve alfa düzeyindeki  $\chi^2$  dağılımındaki tablo değeri ile karşılaştırılarak manidarlık testi yapılmaktadır (Holland ve Wainer, 1993; Camilli ve Shepard, 1994). MH istatistiğinin daha kolay yorumlanabilmesi için logaritmik dönüşüm yapılmaktadır. Ayrıca, MH büyüklüğüne göre DMF'nin düzeyi hakkında da yorum yapılabilmektedir. Bu araştırma kapsamında, ETS (Educational Test Service) tarafından da kabul gören ve B (orta:  $1 \leq \Delta MH \leq 1.5$ ) ve C (yüksek:  $\Delta MH > 1.5$ ) düzeyinde yer alan maddeler DMF'li olarak belirlenmiştir.

*Lojistik Regresyon (LR)*: Bu yöntemle DMF belirlenirken, maksimum olabilirlik yöntemleri ile tahmin edilen olasılık fonksiyonuna başvurulur (Osterlind,1983). Bu yöntemde, bağımsız değişken olarak grup aidiyeti ve toplam test puanı, bağımlı değişken olarak da madde puanı (1,0) alınmaktadır. Bu yöntem, kurulan modeller sayesinde maddeye verilen cevaplar ve toplam puan üzerinden DMF'nin sınımlanmasını sağlar. Bu yöntemde  $\chi^2$  değeri elde edilebildiği gibi  $R^2$  değeri de hesaplanır.  $\Delta R^2$  istatistiği, DMF miktarının yorumlanmasında kullanılan bir etki büyüklüğü belirtir. Bu araştırmada, Zumbo ve Thomas (1996) tarafından belirlenen sınır değerlerinde B (orta:  $0.13 \leq \Delta R^2 < 0.26$ ) ve C (yüksek:  $\Delta R^2 \geq 0.26$ ) düzeyinde yer alan maddeler DMF'li olarak belirlenmiştir.

*Lord'un  $\chi^2$* : Bu yöntem, Lord (1982) tarafından önerilmiş olup  $\chi^2$  tekniğini temel almaktadır ve uygun MTK modelleri kullanılarak hem tek biçimli hem de tek biçimli olmayan DMF belirlenmesinde kullanılmaktadır (Camilli ve Shepard, 1994). Burada her iki grup için de kestirilen a ve b parametreleri farkları hesaplanmaktadır. Madde parametreleri arasındaki farkların varyans-kovaryans matrisleri elde edilmektedir. Daha sonra ise, kestirilen parametre sayısındaki serbestlik derecesine sahip  $\chi^2$  dağılımının elde edilmektedir (Camilli ve Shepard, 1994). Etki büyüklüğü  $\Delta MH$

etki büyüklüğüyle benzerdir ve etki büyüklüğün belirlenmesinde ETS'nin delta ölçeği sınıflaması kullanılmıştır (Holland ve Thayer, 1985) ve bu çalışmada B ve C düzeyinde olan maddeler DMF'li olarak belirlenmiştir.

*Raju'nun İşaretsiz Alan Ölçüsü:* Bu yöntemde, her iki grup için kestirilen madde karakteristik eğrileri arasında kalan alan hesaplanmaktadır. Bu alanın matematiksel olarak hesaplanmasını sağlayan integral formülünü geliştirilmiştir (Camilli ve Shepard, 1994). Her iki grup için kestirilen madde karakteristik eğrileri arasında kalan alanın olmaması maddede DMF'nin bulunmadığını göstermektedir. Etki büyüklüğünün hesaplanmasında ETS'in belirlemiş olduğu delta ölçeği dikkate alınmıştır ve bu çalışmada B ve C düzeyinde yer alan maddeler DMF'li olarak belirlenmiştir.

Bu çalışmada, DMF'nin belirlenmesi amaçlanmış ve DMF belirlemede kullanılan yöntemlerin çeşitli koşullardaki I. Tip hata ve güç oranları belirlenmeye çalışılmıştır. *I. Tip hata*, madde gerçekte DMF içermediği halde, madde DMF içeriyor şeklinde kararın verilmesidir. DMF belirlemede *güç* ise, maddenin DMF içermesi gerekirken, DMF içeriyor şeklinde kararın verilmesidir. Maddelerin DMF'li olup olmadıkları belirlenirken istenilen, güç oranlarının yüksek, I. Tip hata oranlarının ise düşük olmasıdır. Tablo 1'de I. Tip hata ve DMF belirleme gücü gösterilmektedir.

**Tablo 1. I. Tip Hata ve DMF Belirleme Gücü**

		Hipotez Testi Sonucunda Verilen Karar	
		H <sub>0</sub> reddedildi "DMF var"	H <sub>0</sub> kabul edildi "DMF yok"
Gerçek	H <sub>0</sub> doğru "DMF yok"	I. Tip hata	Doğru karar
	H <sub>0</sub> yanlış "DMF var"	Doğru karar (DMF Belirleme Gücü)	II. Tip hata

İlgili literatür incelendiğinde, DMF belirlemede kullanılan yöntemlerin I. Tip hata ve güç oranlarının belirlenmesine ilişkin çeşitli araştırmaların yapıldığı görülmektedir. Yapılan araştırmalarda, çeşitli koşullar manipüle edilerek I. Tip hata ve güç oranlarının nasıl değiştiği araştırılmıştır. Örneğin, Uttaro and Millsap (1994), MH yöntemiyle DMF belirlenirken, farklı test uzunluklarında (20-40), I. Tip hata oranlarının da farklılaştığını göstermiş ve I. Tip hata oranlarının, 20 maddelik testte 40 maddelik testten daha fazla olduğunu belirtmiştir. Roussos ve Stout (1996), yaptıkları simülasyon çalışmasında DMF'yi belirlemek için MH ve SIBTEST yöntemlerini kullanmış, örneklem büyüklüklerini ve yetenek dağılımlarını manipüle etmiş ve bu yöntemlerdeki 1. Tip hatayı araştırmışlardır. Yetenek dağılımı benzer ve farklı örneklem büyüklüklerinde MH ve SIBTEST yöntemlerinin 1. Tip hata oranlarının düşük olduğu ancak, yetenek dağılımları farklılaştığında her iki yöntemin de 1. Tip hata oranlarının arttığı, bununla birlikte SIBTEST yönteminin hata oranının MH yönteminden daha az olduğunu bulmuşlardır. Finch (2005), DMF'nin belirlenmesinde MIMIC modeli kullanarak simülasyon çalışması yapmış, madde sayısı, örneklem büyüklüğü, yetenek ortalamaları farklılıkları, DMF'li madde sayısı gibi faktörleri manipüle ederek, MH, SIBTEST ve MTK-OO yöntemlerini karşılaştırmıştır. Çalışma sonucunda, MIMIC modelde 50 maddenin DMF'nin belirlenmesinde etkili olduğu, testte %20'den fazla DMF'li madde olduğunda tüm yöntemlerde 1. Tip hata oranının şiştiği ve güç oranının düştüğü, MH yönteminin hem iki hem de üç parametrelili modelde diğer yöntemlere göre daha iyi performans gösterdiği, SIBTEST yönteminde iki parametrelili modelin üç parametrelili modelden daha iyi performans gösterdiği sonucuna ulaşmıştır. Scott ve diğerleri (2009), örneklem büyüklüğünü, madde sayısını ve anlamlılık düzeylerini değişimleyerek I. Tip hata ve güç çalışması yapmışlar ve madde sayısının değişimlenmesinin etkisinin çok az olduğu, tek biçimli olmayan DMF'yi belirlemek ya da 1. Tip hata oranını düşürmek için büyük örneklem büyüklüklerinin gerekli olduğu ve DMF belirlemede LR yönteminin performansının iyi olduğu sonucuna ulaşmışlardır. Atar ve Kamata (2011), çoklu puanlanan maddelerde, örneklem büyüklüğü, DMF miktarı, örneklem büyüklüğü oranları gibi faktörleri değişimleyerek, I. Tip hata ve güç oranlarını incelemişlerdir.

Yaptıkları çalışma sonucunda, genel olarak tüm simülasyon koşullarında, olabilirlik oran testi ve logistik regresyon yöntemlerinin I. Tip hatayı kontrol etmede iyi olduğu ve olabilirlik oran testinin orta ve büyük örneklemelerden elde edilen güç oranlarının yüksek olduğu sonucuna ulaşmışlardır. Cuevas ve Cervantes (2012), logistik regresyon yöntemini kullanarak, örneklem büyüklüğünü, örneklem büyüklüğü oranlarını, test uzunluğunu ve DMF miktarını değişimleyerek, bu faktörlerin ortak etkilerini de göz önünde bulundurarak I. Tip hata ve II. Tip hata çalışması yapmışlardır. Kabasakal, Arslan, Gök ve Kelecioğlu (2014), yaptıkları simülasyon çalışmasında, örneklem büyüklüğü oranını, yetenek dağılımı farklılıklarını, test uzunluğunu, DMF içeren madde oranını ve model türünü değişimleyerek, MH, SIBTEST ve Madde Tepki Kuramı Olabilirlik Oranı (MTK-00) yöntemlerinin performanslarını karşılaştırmışlardır. Bu çalışma sonucunda, tek biçimli DMF'nin belirlenmesinde SIBTEST yönteminin en yüksek hataya ve incelenen tüm koşullarda MH yönteminin en yüksek güce sahip olduğunu bulmuşlardır. Keklik (2014) örneklem büyüklüğünü, yetenek dağılımını ve örneklem büyüklüğü oranlarını değişimleyerek, DMF belirlemede kullanılan MH ve LR yöntemlerinin I. Tip hata oranlarını ve güç oranlarını karşılaştırmıştır. Çalışma sonunda yetenek dağılımları farklılaştığında MH ve LR tekniklerinde şişirilmiş I. Tip hataların ortaya çıktığını, grup yetenek dağılımları eşitken MH ve LR I. Tip hata oranlarının örneklem büyüklüğüne göre dalgalanma gösterdiğini ve örneklem büyüklüğü oranları farklı olduğunda, odak ve referans grupların örneklem büyüklükleri arttıkça MH ve LR testlerinin istatistiksel güçlerinin de düştüğünü bulmuştur.

Yukarıdaki çalışmalar incelendiğinde, çeşitli koşullar manipüle edilerek I. Tip hata ve güç oranlarının nasıl değiştiğinin araştırıldığı görülmektedir. Ancak dikkat edilirse benzer yöntemlerin kullanıldığı görülmektedir. Bu çalışma kapsamında ise, çeşitli koşullar ve bu koşulların düzeyleri değişimlenerek DMF'nin belirlenmesinde kullanılan hem Klasik Test Kuramı'na hem de Madde Tepki Kuramı'na dayalı yöntemlerin, hem temel etkilerinin hem de ortak etkilerinin I. Tip hata ve güç oranları belirlenmeye çalışılmıştır. Bu açıdan bakıldığında, gerek kullanılan yöntemler, gerek kullanılan koşullar ve onların düzeyleri farklılaştırıldığından, ayrıca faktörlerin temel etkileriyle birlikte ortak etkilerinin de ele alınmasından ötürü özgün bir çalışma olacağı ve alana katkı getireceği düşünülmektedir.

Araştırmada aşağıda belirtilen sorulara cevap aranmaya çalışılmıştır.

1. Çeşitli koşullar altında DMF belirleme yöntemlerinin 1. Tip Hata oranları nasıldır?
  - a) Koşulların temel etkileri için nasıldır?
  - b) Koşulların ortak etkileri için nasıldır?
2. Çeşitli koşullar altında DMF belirleme yöntemlerinin güç oranları nasıldır?
  - a) Koşulların temel etkileri için nasıldır?
  - b) Koşulların ortak etkileri için nasıldır?

#### **Araştırmanın Amacı**

Bu çalışmanın amacı, değişen madde fonksiyonunun belirlenmesinde kullanılan Klasik Test Kuramı'na dayalı Mantel-Haenszel (MH), Lojistik Regresyon (LR) yöntemleri ve Madde Tepki Kuramına dayalı, Lord'un  $\chi^2$  ve Raju'nun işaretli alan ölçüleri yöntemlerinin çeşitli koşullar altındaki (örneklem büyüklüğü, madde sayısı, referans/toplam gruptaki örneklem büyüklüğü oranı, DMF'li madde oranı) I. Tip hata ve güç oranlarının belirlenmesidir.

## **YÖNTEM**

#### **Araştırmanın Türü**

Bu çalışmada değişen madde fonksiyonunun belirlenmesinde kullanılan bazı yöntemlerin I. Tip hata ve güç oranları çeşitli değişkenler altında incelenmiştir. Yöntemlerin performansı değişkenlerle ilişkilendirilmeye çalışılmıştır. Bu çalışma, mutlak bir nedensellik atfetme amacı gütmendiğinden ilişkisel bir çalışma olarak düşünülebilir. İlişkisel araştırmalarda, nedensellik dışı ilişkiler aranabilmektedir (Erkuş, 2009).

## Veri Üretimi

Bu çalışmada, değişen madde fonksiyonunun belirlenmesinde kullanılan çeşitli yöntemlerin belirli koşullardaki I. Tip hata ve güç oranlarının belirlenmesi simülasyon çalışmasıyla yapılmıştır. I. Tip hata ve güç oranların belirlenmesi için ayrı ayrı veri üretilmiştir.

Çeşitli DMF belirleme yöntemlerinin çeşitli koşullarda gösterdikleri 1. Tip hata oranlarının belirlenmesi amacıyla, hem referans hem de odak grup için örneklem büyüklüğü (N1=500, N2=1000, N3=1500, N4=2000 ve N5=3000), referans grubundaki örneklem büyüklüğünün bütün gruptaki örneklem büyüklüğüne oranı (RT1=0.50, RT2=0.75) ve madde sayısı (MS1=40, MS2=60) değiştirilerek veri üretilmiştir. Veri üretimi iki parametrelili lojistik modele dayalı olarak yapılmıştır. Çalışma kapsamında, madde parametresi olarak kullanılan a parametresi ortalaması 0.8, standart sapması 0.04 olan normal dağılımdan elde edilirken, b parametresi ise minimum değeri -2, maksimum değeri +2 olan uniform dağılımdan seçkisiz çekilerek belirlenmiştir. Bireylerin yetenek dağılımına ilişkin değerleri ise ortalaması 0, standart sapması 1 olan normal dağılımdan elde edilmiştir. Referans ve odak grup için ayrı ayrı veri üretilmekle beraber üretilen veriler için madde parametreleri ortaktır.

Çeşitli DMF belirleme yöntemlerinin çeşitli koşullardaki güç oranlarının belirlenmesi için hem referans hem de odak grup için örneklem büyüklüğü (500, 1000, 1500, 2000 ve 3000), referans gruptaki örneklem büyüklüğünün bütün gruptaki örneklem büyüklüğüne oranı (0.50, 0.75), madde sayısı (40, 60), değişen madde fonksiyonu içeren madde oranı (%5, %10) ve DMF düzeyi (b1=+0.50, b2=+0.75) değiştirilerek veri üretilmiştir. Veri üretimi I. Tip hata çalışmasında olduğu gibi iki parametrelili lojistik modele dayalı olarak yapılmıştır. Referans ve odak grup için a parametresi ortalaması 0.8 standart sapması 0.04 olan normal dağılımdan elde edilmiştir. b parametresi ise referans grup için minimum değeri -2 maksimum değeri +2 olan uniform dağılımdan seçkisiz çekilerek belirlenmiştir. Odak grup verilerinin üretiminde kullanılacak olan b parametreleri, referans grubun b parametreleri üzerinde manipülasyon yapılarak elde edilmiştir. DMF içermesi istenen madde sayısı kadar b parametresine DMF miktarı olan 0.5 veya 0.75 eklenmiştir. Bireylerin yetenek dağılımına ilişkin değerler ise hem referans grup için hem de odak grup için ortalaması 0 standart sapması 1 olan normal dağılımdan elde edilmiştir. Böylelikle bireylerin yetenek dağılımı farklılaşmazken istenen sayıda maddenin referans ve odak grupları için istenen güçlük düzeylerinde farklılaşma meydana getirilerek DMF içeren madde oluşturulmuştur. Referans ve odak grup için ayrı ayrı veri üretilmekle beraber üretilen veriler aynı matriste birleştirilmiştir.

Hem I. Tip hata çalışmasında hem de güç çalışmasında değişimleme ölçütlerinin düzeylerinin çaprazlanması ile oluşan her bir durum için 100 tekrarlama yapılmıştır. Veri üretimi R.3.0.1 programı kullanılarak yazılan kod ile gerçekleştirilmiştir.

## Verilerin Analizi

Değişen madde fonksiyonunun (DMF) belirlenmesinde, Klasik Test Kuramı'na dayalı Mantel-Haenszel (MH), Lojistik Regresyon (LR) yöntemleri ve Madde Tepki Kuramına dayalı, Lord'un  $\chi^2$  ve Raju'nun işaretli alan ölçüleri yöntemleri kullanılmıştır. Madde Tepki Kuramı için yapılan analizlerde, "ltm" motoru kullanılarak iki parametrelili lojistik modele göre kestirim yapılmıştır. DMF belirlemeye yönelik analizler R.3.0.1 programı ve "difR" paketi kullanılarak gerçekleştirilmiştir. difR, Magis, Beland ve Raiche (2015) tarafından yazılmış olan değişen madde fonksiyonu belirleme yöntemlerine yönelik indekleri barındıran bir R paketidir.

I. Tip hata ve güç çalışması için elde edilen DMF analiz sonuçları ayrı ayrı matrisler halinde depolanmış ve değerlendirilmiştir.

I. Tip hata çalışmasında her bir koşul için gerçekleştirilen 100 tekrarlama sonucunda, DMF içermemesi gerekirken DMF içeriyor şeklinde analiz çıktısı elde edilen maddelerin oranı belirlenip raporlaştırılmıştır. Güç çalışması için ise, DMF içermesi gerekirken, DMF içeriyor şeklinde analiz çıktısı elde edilen maddelerin oranı belirlenip raporlaştırılmıştır.

## BULGULAR

Bu bölümde araştırma soruları doğrultusunda elde edilen bulgular yer almaktadır.

### I. Tip Hata Oranları İçin Elde Edilen Bulgular

Tüm koşul düzeyleri için I. Tip hata oranlarından elde edilen sonuçlar Tablo 2'de gösterilmektedir. Bu sonuçlar incelendiğinde, MH yönteminden elde edilen sonuçların minimum değerinin 0.046, maksimum değerinin ise 0.056, LR yönteminden elde edilen sonuçların minimum değerinin 0.045 maksimum değerinin 0.058, Lord'un  $\chi^2$  yönteminden elde edilen sonuçların minimum değerinin 0.034 maksimum değerinin 0.264 ve Raju'nun işaretli alan yönteminden elde edilen sonuçların minimum değerinin 0.047 maksimum değerinin 0.230 olduğu görülmüştür.

En düşük I. Tip hata oranı ortalamasının (0.034), örneklem büyüklüğü 2000, madde sayısı 60, referans/toplam örneklem büyüklüğü oranı 0.5 (referans ve odak gruptaki birey sayıları birbirine eşit) olduğu koşulda ve Lord'un  $\chi^2$  yöntemiyle elde edilmiştir. En yüksek I. Tip hata oranı ortalaması (0.264) ise, örneklem büyüklüğü 500, madde sayısı 60, referans/toplam örneklem büyüklüğü oranı 0.75 (referans ve odak gruptaki birey sayıları farklı) olduğu koşulda ve Lord'un  $\chi^2$  yöntemiyle elde edilmiştir. Bu durum bize Madde Tepki Kuramı'na dayalı yöntemlerin kullanılabilirliği için öncelikle örneklem büyüklüğünün önemli olduğunu göstermektedir.

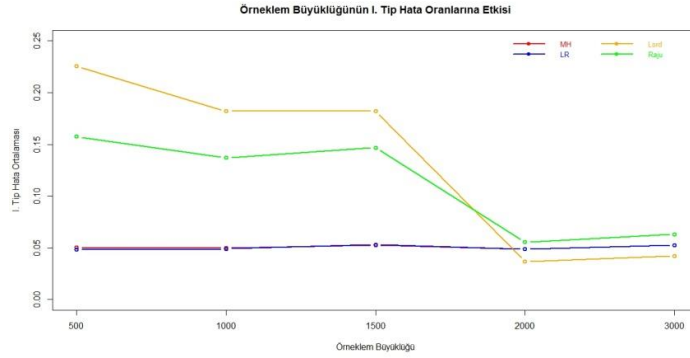
**Tablo 2.** Tüm Koşul Düzeyleri İçin I. Tip Hata Oranları

Örneklem Büyüküğü	Madde Sayısı	Örneklem Büyüküğü Oranı		Lojistik Regresyon		
		R/T	MH	Lord $\chi^2$	Raju	
500	40	0.50	0,053	0,050	0,188	0,082
		0.75	0,048	0,050	0,260	0,218
	60	0.50	0,046	0,045	0,190	0,100
		0.75	0,054	0,048	0,264	0,230
1000	40	0.50	0,051	0,049	0,175	0,119
		0.75	0,048	0,050	0,185	0,145
	60	0.50	0,051	0,050	0,172	0,126
		0.75	0,049	0,047	0,196	0,158
1500	40	0.50	0,049	0,050	0,172	0,124
		0.75	0,053	0,049	0,194	0,160
	60	0.50	0,054	0,055	0,179	0,142
		0.75	0,055	0,058	0,184	0,161
2000	40	0.50	0,048	0,054	0,037	0,048
		0.75	0,048	0,045	0,038	0,065
	60	0.50	0,048	0,048	0,034	0,047
		0.75	0,051	0,048	0,038	0,062
3000	40	0.50	0,050	0,052	0,038	0,053
		0.75	0,056	0,055	0,047	0,067
	60	0.50	0,053	0,055	0,044	0,063
		0.75	0,052	0,048	0,039	0,069

Faktörlerin temel etkileri ve ortak etkilerine ilişkin grafikler ve açıklamalar aşağıda sunulmuştur.

### Örneklem Büyüklüğünün I. Tip Hata Oranlarına Etkisi

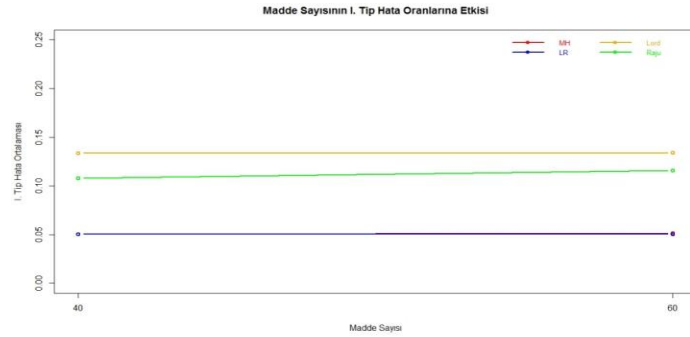
Örneklem büyüklüğünün I. Tip Hata Oranlarına etkisi Şekil 1'de gösterilmektedir. Şekil 1 incelendiğinde, MH ve LR yöntemlerinde örneklem büyüklüğü arttıkça, 1. Tip hata oranlarının ortalama değerlerinin çok fazla değişmediği, ancak Lord'un  $\chi^2$  ve Raju'nun işaretli alan yöntemlerinde örneklem büyüklüğü arttıkça 1. Tip hata oranlarının ortalama değerlerinin azaldığı görülmektedir. Minimum örneklem büyüklüğünde (500) I. Tip hata oranı en düşük MH ve LR yöntemlerinden elde edilirken, örneklem büyüklüğü 2000 ve daha fazlasına ulaşıldığında özellikle Lord'un  $\chi^2$  yönteminde en düşük I. Tip hata oranının elde edildiği göze çarpmaktadır.



Şekil 1. Örneklem Büyüklüğünün I. Tip Hata Oranlarına Etkisi

### Madde Sayısının I. Tip Hata Oranlarına Etkisi

Madde sayısının I. Tip hata oranlarına etkisi Şekil 2'de gösterilmektedir. Şekil 2 incelendiğinde, tüm yöntemlerde madde sayısı arttırıldığında 1. Tip hata oranlarının ortalama değerlerinin çok fazla değişmediği görülmektedir. Ancak hem 40 hem de 60 madde koşullarında, Lord'un  $\chi^2$  ve Raju'nun işaretli alan yöntemlerinden elde edilen ortalama değerlerinin, MH ve LR yöntemlerinden elde edilen ortalama değerlerine göre daha yüksek olduğu görülmektedir.



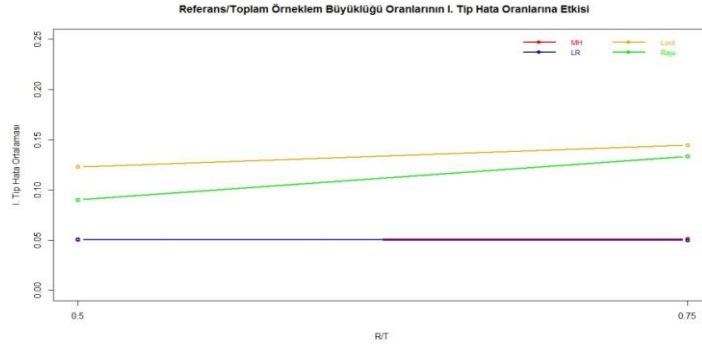
Şekil 2. Madde Sayısının I. Tip Hata Oranlarına Etkisi

### Referans / Toplam Örneklem Büyüklüğü Oranlarının I. Tip Hata Oranlarına Etkisi

Referans grubunun toplam örneklem büyüklüğüne oranlarının I. Tip hata oranlarına etkisi Şekil 3'te gösterilmektedir. Şekil 3 incelendiğinde, referans/toplam örneklem büyüklüğü oranlarının değişimlenmesi sonucunda MH ve LR yöntemlerinden elde edilen ortalama değerlerinin çok fazla değişmediği görülmektedir. Referans grubunun tüm gruptaki örneklem büyüklüğüne oranı hem 0.5 hem de 0.75 olduğu durumlarda, en yüksek I. Tip hata ortalama değerinin Lord'un  $\chi^2$  yönteminden elde edildiği görülmektedir. Ayrıca, referans grubunun tüm gruptaki örneklem büyüklüğüne oranı 0.75 olduğu koşulda Lord'un  $\chi^2$  ve Raju'nun işaretli alan yöntemlerinden elde edilen ortalama



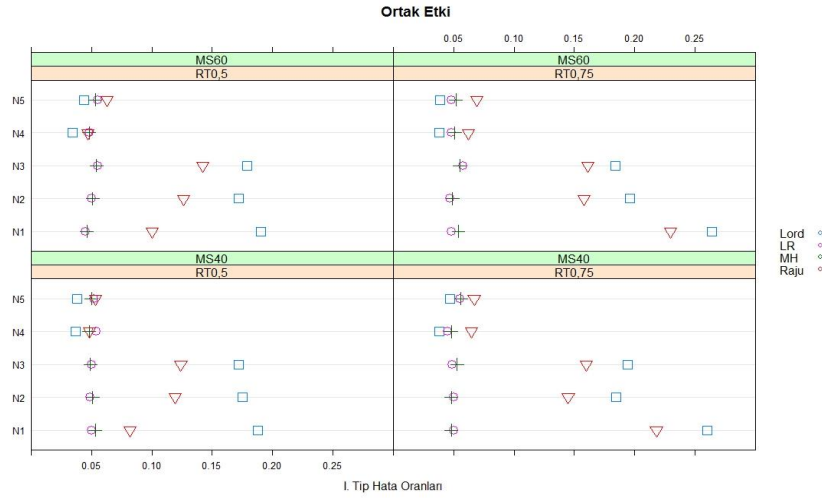
değerlerinin 0.5 olduğu koşuldan elde edilen ortalama değerlerinden daha yüksek olduğu görülmektedir.



Şekil 3. Referans / Toplam Örneklem Büyüklüğü Oranlarının I. Tip Hata Oranlarına Etkisi

### Tüm Koşulların Ortak Etkilerinin I. Tip Hata Oranlarına Etkisi

Tüm koşulların ortak etkilerinin I. Tip hata oranlarına etkisi Şekil 4'te gösterilmektedir.



Şekil 4. Tüm Koşulların Ortak Etkilerinin I. Tip Hata Oranlarına Etkisi

Şekil 4 incelendiğinde, örneklem büyüklüğü arttıkça Lord'un  $\chi^2$  ve Raju'nun işaretli alan yöntemlerinden elde edilen 1. Tip hata oranlarının ortalama değerlerinin azaldığı, MH ve LR yöntemlerinde örneklem büyüklüğü arttıkça (N1'den N5'e gidildikçe), 1. Tip hata oranlarının ortalama değerlerinin çok fazla değişmediği görülmektedir. Bu durumda Madde Tepki Kuramı'na dayanan yöntemler için örneklem büyüklüğünün önemli iken, Klasik Test Kuramı'na dayalı yöntemler için bu çalışmada kullanılan minimum örneklem büyüklüğünün yeterli olduğu görülmektedir. Madde sayısı artırıldığında, tüm yöntemlerden elde edilen I. Tip hata oranları ortalamalarını çok değişmediği görülmektedir. Referans/toplam örneklem büyüklüğü oranlarının değişimlenmesi MH ve LR yöntemlerinden elde edilen ortalama değerlerinin çok fazla değiştiği görülmektedir. Referans/toplam örneklem büyüklüğü oranlarının 0.75 olduğu durumda, Lord'un  $\chi^2$  ve Raju'nun işaretli alan yöntemlerinden elde edilen I. Tip hata ortalama değerlerinin, 0.5 olduğu durumdan daha yüksek olduğu görülmektedir. Şekil 4 toplu olarak değerlendirildiğinde, örneklem büyüklüğünün düşük (500), madde sayısının 60 ve R/T oranının 0.75 olduğu (referans ve odak gruptaki birey sayılarının farklı) durumlarda Madde Tepki Kuramı'na dayalı yöntemlerden elde edilen hata oranı ortalamalarının yüksek olduğu görülmektedir.

## Güç Oranları İçin Elde Edilen Bulgular

Tüm koşul düzeyleri için güç oranlarından elde edilen sonuçlar Tablo 3'te gösterilmektedir.

**Tablo 3. Tüm Koşul Düzeyleri İçin Güç Oranları**

Örneklem Büyüklüğü	Madde Sayısı	DMF'li Madde Oranı	Örneklem Büyüklüğü Oranı (R/T)	b Farkı	Güç Oranları			
					MH	Lojistik Regresyon	Lord $\chi^2$	Raju
500	40	% 5	0.50	0.50	0.430	0.375	0.165	0.140
			0.75	0.75	0.760	0.725	0.490	0.395
			0.50	0.50	0.415	0.340	0.290	0.310
		% 10	0.75	0.75	0.655	0.635	0.430	0.425
			0.50	0.50	0.365	0.290	0.142	0.128
			0.75	0.75	0.735	0.662	0.458	0.362
	60	% 5	0.50	0.50	0.480	0.437	0.263	0.197
			0.75	0.75	0.727	0.663	0.477	0.377
			0.50	0.50	0.320	0.260	0.233	0.250
		% 10	0.75	0.75	0.643	0.593	0.410	0.403
			0.50	0.50	0.417	0.347	0.213	0.170
			0.75	0.75	0.753	0.693	0.442	0.350
1000	40	% 5	0.50	0.50	0.775	0.720	0.570	0.405
			0.75	0.75	0.950	0.945	0.865	0.705
			0.50	0.50	0.585	0.510	0.400	0.350
		% 10	0.75	0.75	0.905	0.875	0.750	0.635
			0.50	0.50	0.702	0.600	0.462	0.375
			0.75	0.75	0.930	0.905	0.795	0.672
	60	% 5	0.50	0.50	0.570	0.480	0.370	0.338
			0.75	0.75	0.902	0.828	0.702	0.580
			0.50	0.50	0.713	0.637	0.513	0.467
		% 10	0.75	0.75	0.957	0.933	0.847	0.700
			0.50	0.50	0.563	0.500	0.403	0.380
			0.75	0.75	0.913	0.883	0.747	0.637
1500	40	% 5	0.50	0.50	0.653	0.595	0.428	0.342
			0.75	0.75	0.958	0.932	0.832	0.730
			0.50	0.50	0.543	0.457	0.332	0.325
		% 10	0.75	0.75	0.857	0.802	0.673	0.565
			0.50	0.50	0.885	0.845	0.735	0.605
			0.75	0.75	1.000	0.990	0.980	0.850
	60	% 5	0.50	0.50	0.785	0.680	0.555	0.450
			0.75	0.75	0.985	0.960	0.895	0.780
			0.50	0.50	0.855	0.785	0.682	0.555
		% 10	0.75	0.75	0.998	0.990	0.955	0.852
			0.50	0.50	0.752	0.645	0.542	0.425
			0.75	0.75	0.960	0.930	0.838	0.682
2000	40	% 5	0.50	0.50	0.877	0.817	0.743	0.643
			0.75	0.75	0.990	0.980	0.967	0.843
			0.50	0.50	0.817	0.713	0.597	0.523
		% 10	0.75	0.75	0.980	0.963	0.907	0.780
			0.50	0.50	0.825	0.773	0.683	0.558
			0.75	0.75	0.990	0.983	0.962	0.885
	60	% 5	0.50	0.50	0.742	0.658	0.515	0.488
			0.75	0.75	0.970	0.942	0.868	0.752
			0.50	0.50	0.945	0.910	0.920	0.840
		% 10	0.75	0.75	1.000	0.995	0.990	0.970
			0.50	0.50	0.895	0.815	0.805	0.735
			0.75	0.75	0.995	0.990	0.980	0.970
2000	40	% 5	0.50	0.50	0.925	0.895	0.872	0.832
			0.75	0.75	1.000	0.998	0.982	0.970
			0.50	0.50	0.842	0.782	0.815	0.780
		% 10	0.75	0.75	0.998	0.998	0.965	0.958
			0.50	0.50	0.973	0.937	0.913	0.890
			0.75	0.75	1.000	0.997	0.987	0.983
	60	% 5	0.50	0.50	0.857	0.753	0.770	0.787
			0.75	0.75	0.997	0.993	0.970	0.977
			0.50	0.50	0.925	0.885	0.883	0.853
		% 10	0.75	0.75	1.000	1.000	0.990	0.983
			0.50	0.50	0.880	0.802	0.823	0.790
			0.75	0.75	0.997	0.995	0.967	0.945

**Tablo 3. Tüm Koşul Düzeyleri İçin Güç Oranları Devamı**

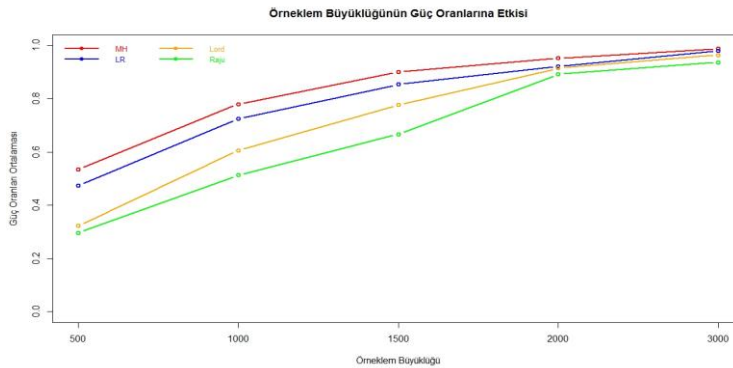
3000	40	% 5	0.50	0.50	0.995	0.995	0.965	0.930
			0.75	0.75	1.000	1.000	0.995	0.995
			0.75	0.50	0.975	0.945	0.920	0.895
		% 10	0.75	0.75	1.000	1.000	1.000	0.995
			0.50	0.50	0.990	0.982	0.962	0.912
			0.75	0.75	1.000	1.000	0.992	0.985
	60	% 5	0.50	0.50	0.990	0.990	0.967	0.957
			0.75	0.75	1.000	1.000	0.993	0.987
			0.75	0.50	0.980	0.960	0.960	0.927
		% 10	0.75	0.75	0.997	0.997	0.997	0.997
			0.50	0.50	0.983	0.970	0.937	0.923
			0.75	0.75	1.000	1.000	0.995	0.987
		0.75	0.50	0.945	0.907	0.858	0.730	
		0.75	0.75	0.998	0.995	0.987	0.922	

Bu sonuçlar incelendiğinde, MH yönteminden elde edilen sonuçların minimum değerinin 0.303, maksimum değerinin ise 1.000, LR yönteminden elde edilen sonuçların minimum değerinin 0.230 maksimum değerinin 1.000, Lord'un  $\chi^2$  yönteminden elde edilen sonuçların minimum değerinin 0.142 maksimum değerinin 1.000 ve Raju'nun işaretli alan yönteminden elde edilen sonuçların minimum değerinin 0.128 maksimum değerinin 0.997 olduğu görülmüştür.

En düşük güç oranı ortalamasının (0.128), örneklem büyüklüğü 500, madde sayısı 40, referans/toplam örneklem büyüklüğü oranı 0.5 (referans ve odak gruptaki birey sayıları birbirine eşit), DMF'li madde oranı %10 ve b farkı 0.50 olduğu koşulda ve Raju'nun işaretli alanyöntemiyle elde edildiği görülmektedir. Bu durumda MTK yöntemlerinde örneklem büyüklüğünün etkisi bir kez daha göze çarpmakla birlikte KTK yöntemleri için de örneklem büyüklüğünün etkisi olduğu görülmektedir. Ayrıca tablo genel olarak incelendiğinde, DMF miktarının da tüm yöntemlerin güç oranlarını etkilediği görülmektedir.

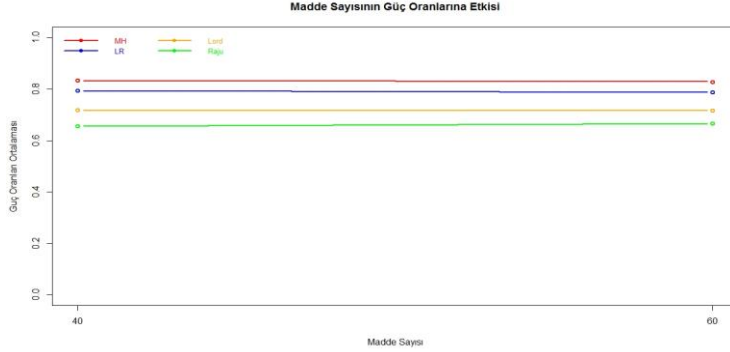
### Örneklem Büyüklüğünün Güç Oranlarına Etkisi

Örneklem büyüklüğünün güç oranlarına etkisi Şekil 5'te gösterilmektedir. Şekil 5 çeşitli örneklem büyüklükleri açısından incelendiğinde, tüm yöntemlerde örneklem büyüklüğü arttıkça elde edilen ortalama güç oranlarının da arttığı görülmektedir. 2000 örneklem büyüklüğünden sonra hem MTK'ya hem de KTK'ya bağlı yöntemlerden elde edilen ortalama güç oranlarının çok yaklaşık sonuçlar verdiği görülmektedir. Tüm örneklem büyüklüklerinde, en yüksek güç oranı ortalamasının MH yöntemiyle elde edildiği görülmektedir. Ayrıca Lord'un  $\chi^2$  ve Raju'nun işaretli alan yöntemlerinden elde edilen güç oranları ortalamasının yüksek örneklem büyüklüklerinde maksimum olduğu görülmektedir.

**Şekil 5. Örneklem Büyüklüğünün Güç Oranlarına Etkisi**

### Madde Sayısının Güç Oranlarına Etkisi

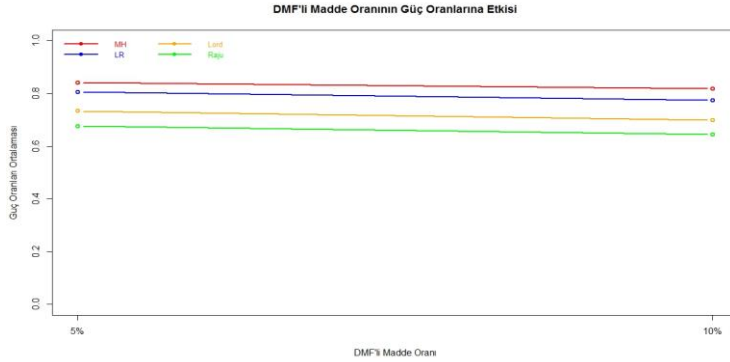
Madde sayısının güç oranlarına etkisi Şekil 6'da gösterilmektedir. Şekil 6 incelendiğinde, tüm yöntemlerde madde sayısının artırılmasının ortalama güç oranlarını çok fazla değiştirmedeği görülmektedir. Tüm yöntemler içinde, her iki madde sayısı koşulunda, güç oranları ortalaması en yüksek MH, en düşük Raju'nun işaretli alan yönteminden elde edildiği görülmektedir.



Şekil 6. Madde Sayısının Güç Oranlarına Etkisi

### DMF'li Madde Oranının Güç Oranlarına Etkisi

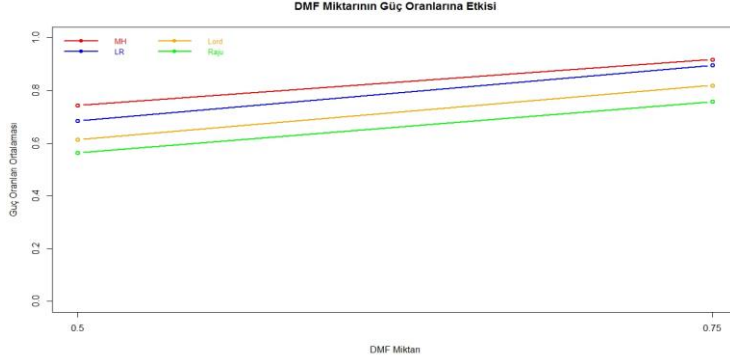
DMF'li madde oranının güç oranlarına etkisi Şekil 7'de gösterilmektedir. Şekil 7 incelendiğinde, DMF'li madde sayısının %5 olduğu koşuldan elde edilen sonuçların %10 olduğu koşuldan elde edilen sonuçlardan biraz daha yüksek olduğu görülmektedir. Başka bir ifadeyle, DMF'li madde sayısının artırılmasının ortalama güç oranlarını az miktarda düşürdüğü söylenebilir.



Şekil 7. DMF'li Madde Oranının Güç Oranlarına Etkisi

### DMF Miktarının Güç Oranlarına Etkisi

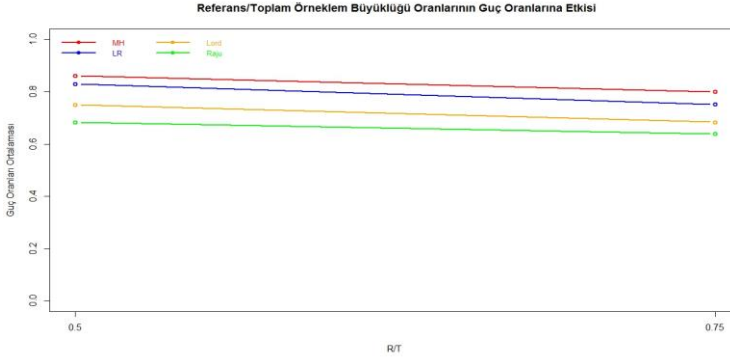
DMF miktarının güç oranlarına etkisi Şekil 8'de gösterilmektedir. Şekil 8 DMF miktarları açısından incelendiğinde, tüm yöntemlerde DMF miktarları arttıkça elde edilen ortalama güç oranlarının da arttığı görülmektedir. Yöntemler içinde farklı DMF miktarlarında güç oranları ortalaması en yüksek MH yönteminin, en düşük Raju'nun işaretli alan ölçüleri yönteminin olduğu görülmektedir.



**Şekil 8.** DMF Miktarının Güç Oranlarına Etkisi

### Referans/Toplam Örneklem Büyüklüğü Oranlarının Güç Oranlarına Etkisi

Referans/toplamörneklem büyüklüğü oranlarının güç oranlarına etkisi Şekil 9'da gösterilmektedir. Şekil 9 incelendiğinde, tüm yöntemlerde, referans/toplam örneklem büyüklüğü oranı 0.5 olan koşuldan elde edilen güç oranı ortalamasının, 0.75 olan koşuldan elde edilen güç oranı ortalamasından daha yüksek olduğu görülmektedir. Başka bir ifadeyle, referans ve odak gruplardaki örneklem büyüklüklerinin eşit olduğu koşuldan elde edilen ortalama güç oranlarının, eşit olmadığı koşuldan elde edilen güç oranlarından daha yüksek olduğu görülmektedir.

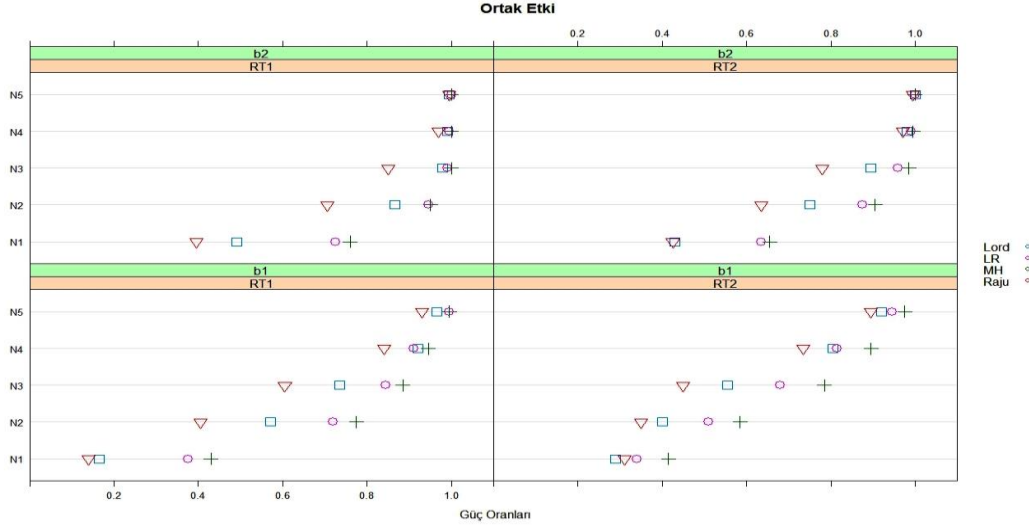


**Şekil 9.** Referans/Toplam Örneklem Büyüklüğü Oranlarının Güç Oranlarına Etkisi

### Çeşitli Koşulların Ortak Etkilerinin Güç Oranlarına Etkisi

Çeşitli koşulların ortak etkilerinin güç oranlarına etkisi Şekil 10'da gösterilmektedir. Çeşitli koşulların ortak etki grafiği oluşturulurken, koşulların temel etkileri göz önünde bulundurulmuştur. Bütün ortak etkileşimlerin birlikte verilmesi mümkün olmadığından farklılaşma yaratan koşulların ortak etki grafiği çizilmiştir.

Şekil 10 incelendiğinde, örneklem büyüklüğü ve DMF miktarı arttıkça tüm yöntemlerden elde edilen güç oranlarının ortalama değerlerinin arttığı görülmektedir. Ayrıca, tüm yöntemlerde, referans ve odak gruplardaki örneklem büyüklüklerinin eşit olduğu koşuldan elde edilen ortalama güç oranlarının, eşit olmadığı koşuldan elde edilen güç oranlarından daha yüksek olduğu görülmektedir.



**Şekil 10.** Tüm Koşulların Ortak Etkilerinin Güç Oranlarına Etkisi

## SONUÇLAR ve TARTIŞMA

Bu çalışmada, DMF belirlenmesinde kullanılan Klasik Test Kuramı'na dayalı MH, LR yöntemleri ve Madde Tepki Kuramına dayalı, Lord'un  $\chi^2$  ve Raju'nun işaretersiz alan ölçüleri yöntemlerinin çeşitli koşullar altındaki I. Tip hata ve güç oranlarının belirlenmesi amaçlanmıştır. I. Tip hata oranlarının belirlenmesi için yapılan çalışmada, MH ve LR yöntemlerinde, örneklem büyüklüğünün değişimlenmesinin I. Tip hata oranları ortalamalarını çok fazla değiştirmedeği ancak Lord'un  $\chi^2$  ve Raju'nun işaretersiz alan yöntemlerinde örneklem büyüklüğü arttıkça I. Tip hata oranlarının ortalama değerlerinin azaldığı görülmektedir. Burada MH ve LR yöntemlerinin düşük örneklemelerde daha iyi sonuçlar verdiği, çalışmada yer alan minimum örneklem büyüklüğünden (500) sonra önemsizleştiği sonucu çıkartılabilir. Ayrıca, MTK'ya bağlı yöntemlerin örneklem büyüklüklerine karşı daha duyarlı olduğu, çeşitli örneklem büyüklüklerinde Raju'nun işaretersiz alan ölçüleri yönteminin I. Tip hata oranının Lord'un  $\chi^2$  yönteminden daha az olduğu görülmektedir. Buradan hareketle, düşük örneklem büyüklüklerinde KTK'ya dayalı yöntemlerin kullanılması, MTK'ya dayalı yöntemlerin ise daha büyük örneklem büyüklüklerinde kullanılması gerektiği sonucu çıkartılabilir. Örneklem büyüklüklerinin yöntemlerin güç oranlarına etkisine bakıldığında ise, hem KTK'ya hem de MTK'ya dayalı yöntemlerde, örneklem büyüklüğünün artmasıyla güç oranlarının ortalamasının da arttığı görülmektedir. Çalışmada kullanılan minimum örneklem büyüklüğünde (500), MH ve LR yöntemlerinin güç oranları ortalamasının daha fazla olduğu, ancak 2000 örneklem büyüklüğünden sonra, tüm yöntemlerden elde edilen güç oranları ortalamasının birbirlerine yaklaştığı görülmektedir. Bu durumda, I. Tip hatanın belirlenmesinde belirtildiği gibi, MTK'ya dayalı yöntemlerle güç oranlarının belirlenmesinde de yüksek örneklem büyüklüklerinin kullanılması gerektiği söylenebilir.

MH ve LR yöntemlerinde, referans/toplam gruptaki örneklem büyüklüğü oranlarının değişimlenmesinin I. Tip hata oranları ortalamalarını çok fazla değiştirmedeği ancak Lord'un  $\chi^2$  ve Raju'nun işaretersiz alan ölçüleri yöntemlerinde referans ve odak gruptaki örneklem büyüklüğü oranlarının eşit olduğu koşulda I. Tip hata oranları ortalamasının daha düşük olduğu görülmektedir. Bu durumda, I. Tip hata oranı açısından, referans ve odak gruptaki birey sayıları farklı olduğunda KTK'ya dayalı yöntemlerin kullanılmasının daha uygun olacağı söylenebilir. Güç oranları açısından bakıldığında ise, tüm yöntemlerde referans ve odak gruptaki örneklem büyüklüğü oranlarının

farklı olması güç oranlarının ortalamasını azalttığı görülmektedir. Bu durumda hem referans hem de odak gruplardaki birey sayılarının eşit olmasının güç oranlarını arttıracığı söylenebilir.

Madde sayısının değişimlenmesinin tüm yöntemlerin I. Tip hata ve güç oranları ortalamalarını çok fazla değiştirmedeği görülmektedir. Farklı madde sayılarında (40 ve 60) I. Tip hata oranları ortalaması en düşük MH ve LR yöntemlerinin, güç oranları ortalaması en yüksek MH yöntemi olduğu görülmektedir. Uttaro and Millsap (1994), MH yönteminin I. Tip hata oranlarının, 20 maddelik testte 40 maddelik testten daha fazla olduğunu belirtmişlerdi. Bu durumda, MH yöntemiyle DMF belirlenirken, diğer yöntemlere göre hem I. Tip hata oranı ortalamasının düşük hem de güç oranı ortalamasının yüksek olmasından dolayı, 40 madde ve daha fazlasında kullanılması önerilebilir.

Tüm yöntemlerde, DMF'li madde sayısının oranının arttırılmasının güç oranlarının ortalamasını azalttığı görülmektedir. Yöntemler içinde farklı DMF miktarlarında güç oranları ortalamasının en yüksek MH yöntemiyle, en düşük Raju'nun işaretli alan ölçüleri yöntemiyle elde edildiği görülmektedir.

Bu çalışma kapsamında tek biçimli olan değişen madde fonksiyonu incelenmiştir. Benzer çalışmalarda ek olarak tek biçimli olmayan değişen madde fonksiyonu da incelenebilir. Çalışma kapsamında bireylerin yetenek düzeyleri sabit tutulmuştur. Referans ve odak gruplardaki bireylerin yetenek dağılımları değişimlenerek benzer çalışmalar yapılabilir. Bu çalışmada madde sayısı 40 ve 60 olarak değişimlenmiştir. Madde sayısının daha düşük olduğu durumlarda 1. Tip hata ve güç oranlarının nasıl değiştiği incelenebilir. Çalışma kapsamında, MH, LR, Lord'un  $\chi^2$ 'si ve Raju'nun işaretli alan yöntemleri kullanılarak DMF belirlenmeye çalışılmıştır. DMF belirlemede kullanılan başka yöntemlerle benzer çalışmalar yapılabilir.

## KAYNAKÇA

- Acar, T. (2008). Maddenin farklı fonksiyonlaşmasını belirlemede kullanılan genelleştirilmiş aşamalı doğrusal modelleme, lojistik regresyon ve olabilirlik oranı tekniklerinin karşılaştırılması. Yayınlanmamış doktora tezi, Hacettepe Üniversitesi, Ankara.
- Anıl, D. ve Güzeller, C. O. (2010). Seviye Belirleme Sınavı Yedinci Sınıf Fen ve Teknoloji Alt Testi ile Diğer Alt Testler Arasındaki İlişkinin Yol Analizi ile İncelenmesi. *AİBÜ Eğitim Fakültesi Dergisi*, 11(1), 1-10
- Anıl, D., Güzeller, C. O., Çokluk, Ö. ve Şekercioğlu, G. (2010). Level Determination Exam (SBS-2008) the Determination of the Validity and Reliability of 7th Grade Mathematics Sub-Test, *Procedia Social and Behavioral Sciences*, 2(2) 5292-5298
- Atar, B. ve Kamata A. (2011). Comparison of IRT Likelihood Ratio Test And LogisticRegression DIF Detection Procedures. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 41, 36-47.
- Cuevas, M. ve Cervantes, V. H. (2012). Differential item functioning detection with logistic regression. *Mathematics and Social Sciences*, 3, 45-59
- Çepni, Z. (2011). Değişen madde fonksiyonlarının SIBTEST, mantel haenzsel, lojistik regresyon ve madde tepki kurami yöntemleriyle incelenmesi. Yayınlanmamış doktora tezi, Hacettepe Üniversitesi, Ankara.
- Camilli, G. ve Shepard, L. A. (1994). Methods for identifying biased test items. London: Sage Publications.
- Doğan, N. ve Sevindik, H. (2011). İlköğretim 6.Sınıflar İçin Uygulanan Seviye Belirleme Sınavının Uygunluk Geçerliği. *Eğitim ve Bilim Dergisi*, 36 (110), 309-319.
- Erkuş, A. (2009). Davranış bilimleri için bilimsel araştırma süreci. Ankara: Seçkin Yayıncılık.
- Finch, W. H. (2005). The MIMIC model as a method for detecting DIF: Comparison with Mantel-Haenzsel, SIBTEST and the IRT likelihood ratio. *Applied Psychological Measurement*, 29 (4), 278-295.
- Hambleton, R K., Swaminathan, H. ve Rogers, H. J. (1991). Fundamentals of item response theory. London: Sage Publication.
- Holland, P.W. ve Thayer, D. T. (1988). Differential item performance and the Mantel Haenzsel procedure. In H.Wainer & H. Braun (Eds), *Test validity* (pp.129-145). Hillsdale, NL: Erlbaum.
- Jodoin, M. G., ve Gierl, M. L. (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 İncelenmesi, *Mersin Üniversitesi Eğitim Fakültesi Dergisi*,9(2), 207-222.
- Karakaya, İ. ve Kutlu, Ö. (2012). Seviye Belirleme Sınavındaki Türkçe Alt Testlerinin Madde Yanlılığının İncelenmesi. *Eğitim ve Bilim Dergisi*, 37(165).
- Kelecioğlu, H., Atalay, K ve Öztürk, N. (2010). Seviye Belirleme Sınavı 7. Sınıf Matematik Alt Testinin Madde-Hedef Uyumunu Açısından İncelenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, Yaz 2010,1(1),37-43
- Magis, D., Beland, S., ve Raiche, G. (2015). difR: Collection of methods to detect dichotomous differential item functioning (DIF). R package version 4.6
- Osterlind, S. (1983). Test item bias. Newbury Park: Sage Publications.
- Öğretmen, T. ve Doğan, N. (2004). OKÖSYS Matematik Alt Testine Ait Maddelerin Yanlılık Analizi. İnönü Üniversitesi Eğitim Fakültesi Dergisi, 5(8)
- Rogers, H. J. ve Swaminathan, H. (1993). A comparison of logistic regression and Mantel Haenszel procedures for detecting differential item functioning. *Applied Psychological Measurement*, 17(2), 105-116.
- Roussos, L. A. ve Stout, W. (1996). Simulation studies of the effects of small sample size and studied item parameters on SIBTEST and Mantel- Haenszel Type I Error performance. *Journal of Educational Measurement*, 33(2), 215-230.
- Scott, N.W., Fayers, P. M., Aaronson, N. K., Bottomley, A., DeGraeff, A., Groenvold, M., Gundy, C., Koller, M., ve Petersen, M. A. (2009). A simulation study provided sample size guidance for differential item functioning (DIF) studies using short scales. *Journal of Clinical Epidemiology*, 62, 288-295.
- Shealy, R. T. ve Stout, W. F. (1993). An item response theory model for test bias and differential test functioning. In P.W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp.197-239). Hillsdale, NJ: Erlbaum.
- Swaminathan, H. ve Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, 27(4), 361-370.
- Uttaro, T. ve Millsap, R. E. (1994). Factors influencing the Mantel-Haenszel procedure in the detection of differential item functioning. *Applied Psychological Measurement*, 18, 15-25.
- Zumbo, D. B. (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 item scores. Ottawa: Directorate of Human Resources Research and Evaluation, Department of National Defense.