

T.C.
MERSİN ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ
BİYOİSTATİSTİK VE TIBBİ BİLİŞİM
ANABİLİM DALI

**BİRDEN ÇOK DEĞERLENDİRİCİNİN BİRDEN
ÇOK TANI TESTİNİN BULUNDUĞU
ÇALIŞMALARDA DİAGNOSTİK PERFORMANS
ANALİZİ**

Gülhan OREKİCİ TEMEL
DOKTORA TEZİ

DANIŞMAN
Doç. Dr. E. Arzu KANIK

Tez No:16

MERSİN 2011

T.C.
MERSİN ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ
BİYOİSTATİSTİK VE TIBBİ BİLİŞİM
ANABİLİM DALI

**BİRDEN ÇOK DEĞERLENDİRİCİNİN BİRDEN
ÇOK TANI TESTİNİN BULUNDUĞU
ÇALIŞMALARDA DİAGNOSTİK PERFORMANS
ANALİZİ**

Gülhan OREKİCİ TEMEL
DOKTORA TEZİ

DANIŞMAN
Doç. Dr. E. Arzu KANIK

Bu tez, Mersin Üniversitesi Bilimsel Araştırma Projeleri Birimi tarafından BAP-SBE BTB (GOT) 2009-9 DR nolu proje olarak desteklenmiştir.

Tez No: 16

MERSİN 2011

T.C.
MERSİN ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ
BİYOİSTATİSTİK VE TIBBİ BİLİŞİM
ANABİLİM DALI

**BİRDEN ÇOK DEĞERLENDİRİCİNİN BİRDEN
ÇOK TANI TESTİNİN BULUNDUĞU
ÇALIŞMALARDA DİAGNOSTİK PERFORMANS
ANALİZİ**

Gülhan OREKİCİ TEMEL
DOKTORA TEZİ

DANIŞMAN
Doç. Dr. E. Arzu KANIK

Bu tez, Mersin Üniversitesi Bilimsel Araştırma Projeleri Birimi tarafından BAP-SBE BTB (GOT) 2009-9 DR nolu proje olarak desteklenmiştir.

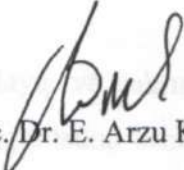
Tez No: 16

MERSİN 2011

Mersin Üniversitesi Sağlık Bilimleri Enstitüsü

Biyoistatistik ve Tıbbi Bilişim Anabilim dalı Doktora programı çerçevesinde yürütülmüş olan "Birden Çok Değerlendiricinin Birden Çok Tanı Testinin Bulunduğu Çalışmalarda Diagnostik Performans Analizi" başlıklı çalışma, Jürimiz tarafından doktora tezi olarak kabul edilmiştir.

Tez Savunma Tarihi: 19.01.2011


Doç. Dr. E. Arzu KANIK

Mersin Üniversitesi Tıp Fakültesi

Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı

Jüri Başkanı



Prof. Dr. Nazan ALPARSLAN

Çukurova Üniversitesi Tıp Fakültesi

Biyoistatistik ve Tıbbi Bilişim AD.

Jüri Üyesi



Prof. Dr. Yusuf ÇELİK

Dicle Üniversitesi Tıp Fakültesi

Biyoistatistik ve Tıbbi Bilişim AD.

Jüri Üyesi



Doç. Dr. Bahar TAŞDELEN

Mersin Üniversitesi Tıp Fakültesi

Biyoistatistik ve Tıbbi Bilişim AD.

Jüri Üyesi



Yrd. Doç. Dr. Gülçin YAPICI

Mersin Üniversitesi Tıp Fakültesi

Halk Sağlığı AD.

Jüri Üyesi

Bu tez, Enstitü Yönetim Kurulunun 10/02/2011 tarih ve 2011/51 sayılı kararı ile kabul edilmiştir.

Prof. Dr. Ülkü ÇÖMELEKOĞLU



TEŐEKKÜR

Tez alıřmam boyunca deęerli katkılarını her ařamada veren ve bu alıřmayı ynlendiren deęerli hocam ve Anabilim Dalı Bařkanım Do. Dr. E.Arzu Kanık'a en iten saygılarımı ve teőekkrlerimi sunarım. Ayrıca doktora eęitimim boyunca deęerli bilgilerinden yararlandıęım hocam Do. Dr. Bahar Tařdelen'e teőekkr ederim.

alıřmamda bilgisayar programlama konusunda yardımını sunan Yrd. Do. Dr. İrem Ersz'e teőekkr ederim.

Tez dnemim boyunca anlayıř ve sabrından dolayı sevgili eřim Devrim ve oęlum ınar'a, yardımlarını ve hořgrsn benden esirgemeyen ve ęrenim hayatımı borlu olduęum aileme sonsuz teőekkr ederim.

Glhan OREKİCİ TEMEL

Ocak 2011

İÇİNDEKİLER

KABUL VE ONAY	i
TEŞEKKÜR.....	ii
İÇİNDEKİLER	iii
ŞEKİLLER DİZİNİ	vi
ÇİZELGELER DİZİNİ	viii
ÖZET	xi
ABSTRACT	x
1.GİRİŞ.....	1
2. GENEL BİLGİLER.....	4
2.1. Çok Testli Çok Değerlendiricili Diagnostik Performans Analizi.....	4
2.1.1. Çok Testli Çok Değerlendiricili (ÇTÇD) Diagnostik Doğruluk Analizinin Tarihsel Gelişimi	4
2.1.2.Çok Testli Çok Değerlendiricili (ÇTÇD) ROC Analizlerinin Amaçları.....	5
2.1.3. Çok Testli Çok Değerlendiricili (ÇTÇD) Deneme Tasarımları	6
2.1.4. Çok Testli Çok Değerlendiricili Doğruluk Analizleri İçin Örnek Genişliği Hesaplaması.....	9
2.2. Çok Testli Çok Değerlendiricili (ÇTÇD) ROC Analizlerinde Kullanılan Yöntemler	11
2.2.1. Dorfman-Berbaum-Metz (DBM) Metodu/ Sözde Değerlerle ANOVA.....	12
2.2.1.1. DBM Metodu Teorisi.....	12
2.1.2. DBM Metodunun Avantajları.....	15
2.2.2. Obuchowski ve Rockette Metodunun (Düzeltilmiş F Metodu) Teorisi	15
2.2.3.Çok Değişkenli WMW istatistikleri	18
2.2.4. Hiyerarşik Ordinal Regresyon ROC Metodu	19

2.2.5. Bootstrap of Components-of-Variance “BWC Method”	19
2.3. Çok Testli Çok Değerlendiricili (ÇTÇD) ROC Analizlerinde Kullanılan Yöntemlerin Karşılaştırılması	20
3. GEREÇ ve YÖNTEM.....	21
3.1. Değerlendiriciler Arası Çok Yüksek ve Çok Düşük Uyum Etkisi İçin Yapılan Simülasyon Çalışması.....	22
3.2. Tanı Testleri Arası Çok Güçlü ve Çok Zayıf Korelasyonun Etkisi İçin Yapılan Simülasyon Çalışması.....	23
4. BULGULAR	24
4.1. Değerlendiriciler Arasındaki Yüksek ve Düşük Uyumun Diagnostik Doğruluk Üzerine Etkisi	24
4.1.1. Değerlendiriciler Arasındaki Yüksek ve Düşük Uyumun Etkisinin Tanı Testlerinin Doğrulukları Üzerine Etkisi	24
4.1.2. Değerlendiriciler Arasındaki Yüksek ve Düşük Uyumun Etkisinin Değerlendiricilerin Doğrulukları Üzerine Etkisi.....	36
4.2. Tanı Testleri Arasındaki Güçlü ve Zayıf Korelasyonun Diagnostik Doğruluk Üzerine Etkisi	47
4.2.1. Tanı Testleri Arasındaki Güçlü ve Zayıf Korelasyonun Etkisinin Tanı Testlerinin Doğrulukları Üzerine Etkisi	47
4.2.2. Tanı Testleri Arasındaki Güçlü ve Zayıf Korelasyonun Etkisinin Değerlendiricilerin Doğrulukları Üzerine Etkisi	58
5. TARTIŞMA	69
5.1. Değerlendiriciler Arasındaki Yüksek ve Düşük Uyumun Diagnostik Doğruluklar Üzerindeki Etkisinin Neden Olduğu Tip I Hata Yapma Olasılıkları	69
5.1.1. Tanı Testlerinin Diagnostik Doğrulukları Arasında Gerçekte Farklılık Olmadığında Farklılık Var Olarak Bulunma Olasılıkları	69
5.1.2. Değerlendiricilerin Diagnostik Doğrulukları Arasında Gerçekte Farklılık Olmadığında Farklılık Var Olarak Bulunma Olasılıkları.....	73
5.2. Tanı Testleri Arasındaki Güçlü ve Zayıf Korelasyonun Diagnostik Doğruluklar Üzerindeki Etkisinin Neden Olduğu Tip I Hata Yapma Olasılıkları	77
5.2.1. Tanı Testlerinin Diagnostik Doğrulukları Arasında Gerçekte Farklılık Olmadığında Farklılık Var Olarak Bulunma Olasılıkları	77

5.2.2. Deęerlendiricilerin Diagnostik Doğrulukları Arasında Gerçekte Farklılık Olmadığında Farklılık Var Olarak Bulunma Olasılıkları.....	78
6.SONUÇLAR ve ÖNERİLER.....	80
KAYNAKLAR.....	83
EKLER.....	87
EK I: Deęerlendiriciler Arasında Düşük /Yüksek Uyum İçin Matlab Paket Programı Kodları	87
EK II: Tanı Testleri Arasında Güçlü/Zayıf Korelasyon İçin Matlab Paket Programı Kodları	92
EK III: Deęerlendiriciler Arasında Uyum Çok Yüksek Olduęu Durumdaki Varyans Analiz Tablosu Simülasyon Sonuçları.....	96
EK IV: Deęerlendiriciler Arasında Uyum Çok Düşük Olduęu Durumda Varyans Analiz Tablosu Simülasyon Sonuçları	106
EK V: Tanı Testleri Arasında Korelasyon Çok Güçlü Olduęu Durumda Varyans Analiz Tablosu Simülasyon Sonuçları	116
EK VI: Tanı Testleri Arasında Korelasyon Çok Zayıf Olduęu Durumdaki Varyans Analiz Tablosu Simülasyon Sonuçları.....	126
ÖZGEÇMİŞ.....	136

ŞEKİLLER DİZİNİ

- Şekil 4.1.** Değerlendiriciler arasında uyum çok yüksekken, 2,5,7 tanı testi için, tanı testlerinin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I Hata yapma olasılıklarının gösterim grafiği 32
- Şekil 4.2.** Değerlendiriciler arasında uyum çok yüksekken, 2,5,7 değerlendirici için tanı testlerinin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I Hata yapma olasılıklarının gösterim grafiği 33
- Şekil 4.3.** Değerlendiriciler arasında uyum çok düşükken, 2,5,7 tanı testi için, tanı testlerinin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I Hata yapma olasılıklarının gösterim grafiği 34
- Şekil 4.4.** Değerlendiriciler arasında uyum çok düşükken, 2,5,7 değerlendirici için tanı testlerinin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I Hata yapma olasılıklarının gösterim grafiği 35
- Şekil 4.5.** Değerlendiriciler arasında uyum çok yüksekken, 2,5,7 tanı testi için değerlendiricilerin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I Hata yapma olasılıklarının gösterim grafiği 43
- Şekil 4.6.** Değerlendiriciler arasında uyum çok yüksekken, 2,5,7 değerlendirici için değerlendiricilerin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I Hata yapma olasılıklarının gösterim grafiği 44
- Şekil 4.7.** Değerlendiriciler arasında uyum çok düşükken, 2,5,7 tanı testi için değerlendiricilerin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I Hata yapma olasılıklarının gösterim grafiği 45
- Şekil 4.8.** Değerlendiriciler arasında uyum çok düşükken, 2,5,7 değerlendirici için tanı testlerinin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I Hata yapma olasılıklarının gösterim grafiği 46
- Şekil 4.9.** Tanı testleri arasında korelasyon çok güçlü olduğunda, 2,5,7 tanı testi için, tanı testlerinin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I Hata yapma olasılığı gösterim grafiği 54
- Şekil 4.10.** Tanı testleri arasında korelasyon çok güçlü olduğunda, 2,5,7 değerlendirici için, tanı testlerinin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I Hata yapma olasılığı gösterim grafiği 55
- Şekil 4.11.** Tanı testleri arasında korelasyon çok zayıf olduğunda, 2,5,7 tanı testi için, tanı testlerinin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I Hata yapma olasılığı gösterim grafiği 56
- Şekil 4.12.** Tanı testleri arasında korelasyon çok zayıf olduğunda, 2,5,7 değerlendirici için, tanı testlerinin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I Hata yapma olasılığı gösterim grafiği 58
- Şekil 4.13.** Tanı testleri arasında korelasyon çok güçlü olduğunda, 2,5,7 tanı testi için, değerlendiricilerin diagnostik doğruluk farklılıklarına ilişkin olabilecek Tip I Hata yapma olasılığı gösterim grafiği 65
- Şekil 4.14.** Tanı testleri arasında korelasyon çok güçlü olduğunda, 2,5,7 değerlendirici için, değerlendiricilerin diagnostik doğruluk farklılıklarına ilişkin olabilecek Tip I Hata yapma olasılığı gösterim grafiği 66
- Şekil 4.15.** Tanı testleri arasında korelasyon çok zayıf olduğunda, 2,5,7 tanı testi için, değerlendiricilerin diagnostik doğruluk farklılıklarına ilişkin olabilecek Tip I Hata yapma olasılığı gösterim grafiği 67

Şekil 4.16. Tanı testleri arasında korelasyon çok güçlü olduğunda 2,5,7 değerlendirici için, değerlendiricilerin diagnostik doğruluk farklılıklarına ilişkin olabilecek Tip I Hata yapma olasılığı gösterim grafiği.....	68
Şekil 5.1. İki değerlendirici arasında çok yüksek ve düşük uyum varken, tanı testlerinin diagnostik doğruluk farklılıklarına ilişkin olabilecek Tip I Hata yapma olasılıklarının grafik gösterimi	70
Şekil 5.2. Beş değerlendirici arasında çok yüksek ve düşük uyum varken, tanı testlerinin diagnostik doğruluk farklılıklarına ilişkin olabilecek Tip I Hata yapma olasılıklarının grafik gösterimi	71
Şekil 5.3. Yedi değerlendirici arasında çok yüksek ve düşük uyum varken, tanı testlerinin diagnostik doğruluk farklılıklarına ilişkin olabilecek Tip I Hata yapma olasılıklarının grafik gösterimi	73
Şekil 5.4. İki değerlendirici arasında çok yüksek ve düşük uyum varken, değerlendiricilerin diagnostik doğruluk farklılıklarına ilişkin olabilecek Tip I Hata yapma olasılıklarının grafik gösterimi	74
Şekil 5.5. Beş değerlendirici arasında çok yüksek ve düşük uyum varken, değerlendiricilerin diagnostik doğruluk farklılıklarına ilişkin olabilecek Tip I Hata yapma olasılıklarının grafik gösterimi	75
Şekil 5.6. Yedi değerlendirici arasında çok yüksek ve düşük uyum varken, değerlendiricilerin diagnostik doğruluk farklılıklarına ilişkin olabilecek Tip I Hata yapma olasılıklarının grafik gösterimi	77
Şekil 5.7. İki-beş ve yedi değerlendirici arasında korelasyon çok güçlü ve zayıf olduğunda, tanı testlerinin diagnostik doğruluk farklılıklarına ilişkin, Tip I Hata yapma olasılıklarının grafik gösterimi	78
Şekil 5.8. İki-beş ve yedi değerlendirici arasında korelasyon çok güçlü ve zayıf olduğunda, değerlendiriciler arasındaki diagnostik doğruluk farklılıklarına ilişkin, Tip I Hata yapma olasılıklarının grafik gösterimi.....	79

ÇİZELGELER DİZİNİ

Çizelge 2.1. Geleneksel Tasarım	7
Çizelge 2.2. Eşleşmemiş değerlendirici, Eşleşmemiş vaka deneme tasarımı	8
Çizelge 2.3. ÇTÇD Doğruluk Analizleri için Faktöriyel Deneme Tasarımı.....	9
Çizelge 2.4. DBM Metodu Örnek Genişliği.....	10
Çizelge 2.5. ÇTÇD ROC Yöntemlerinin Karşılaştırılması.....	20
Çizelge 4.1. İki değerlendirici arasında uyum çok yüksek ve düşükken, tanı testlerinin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I Hata yapma olasılıkları	26
Çizelge 4.2. Beş değerlendirici arasında uyum çok yüksek ve düşükken, tanı testlerinin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I Hata yapma olasılıkları	28
Çizelge 4.3. Yedi değerlendirici arasında uyum çok yüksek ve düşükken, tanı testlerinin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I Hata yapma olasılıkları	30
Çizelge 4.4. İki değerlendirici arasında çok yüksek ve düşük uyum varken, değerlendiricilerin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I Hata yapma olasılıkları	37
Çizelge 4.5. Beş değerlendirici arasında çok yüksek ve düşük uyum varken, değerlendiricilerin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I Hata yapma olasılıkları	39
Çizelge 4.6. Yedi değerlendirici arasında çok yüksek ve düşük uyum varken, değerlendiricilerin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I Hata yapma olasılıkları	41
Çizelge 4.7. İki tanı testi arasında korelasyon çok güçlü ve zayıf olduğunda, tanı testlerinin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I Hata yapma olasılıkları	48
Çizelge 4.8. Beş tanı testi arasında korelasyon çok güçlü ve zayıf olduğunda, tanı testlerinin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I Hata yapma olasılıkları	50
Çizelge 4.9. Yedi tanı testi arasında korelasyon çok güçlü ve zayıf olduğunda, tanı testlerinin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I Hata yapma olasılıkları	52
Çizelge 4.10. İki tanı testi arasında korelasyon çok güçlü ve zayıf olduğunda, değerlendiricilerin diagnostik doğruluk farklılıklarına ilişkin olabilecek Tip I Hata yapma olasılığı.....	59
Çizelge 4.11. Beş tanı testi arasında korelasyon çok güçlü ve zayıf olduğunda, değerlendiricilerin diagnostik doğruluk farklılıklarına ilişkin olabilecek Tip I Hata yapma olasılığı.....	61
Çizelge 4.12. Yedi tanı testi arasında korelasyon çok güçlü ve zayıf olduğunda, değerlendiricilerin diagnostik doğruluk farklılıklarına ilişkin olabilecek Tip I Hata yapma olasılığı.....	63

ÖZET

Birden Çok Değerlendirici Birden Çok Tanı Testinin Bulunduğu Çalışmalarda Diagnostik Performans Analizi

ROC verilerinin analizinde geleneksel metotlar ya tek bir değerlendiricinin bir grup vaka üzerindeki sonuçlarını ya da birden fazla değerlendiricinin yine bir grup vaka üzerindeki sonuçları üzerinden yapılır. Geleneksel ROC analizinde amaç değerlendiricilerin performansını değerlendirmek değildir. Fakat geleneksel ROC analizinden farklı olarak hem tanı testlerinin vaka gruplarını ayırmadaki performansını hem de değerlendiricilerin performansını multi reader multi case MRMC analizleriyle çözümlenmek mümkündür. Bu tür çalışmaların temel amacı ROC eğrileri toplamak değildir. Amaç, değerlendiricilerin ya da değerlendiricilerin ortalamaları arasındaki tanı testinin doğruluğunu ortaya çıkarmaktır. Her değerlendiricinin her vakayı değerlendirdiği tam çapraz deneme planlı ROC verilerinin değerlendirilmesini Dorfman'ın 1992'de yapmış olduğu çalışmasında ilk defa MRMC çalışmaları olarak değinmişlerdir. Çalışmada her defasında bir vakanın dışarıda tutulduğu Jackknife yeniden örnekleme yöntemini kullanarak klasik varyans analiz çalışması yapmıştır. Ve bu amaçla DBM metodunu önermiştir. Bu çalışmada çoklu değerlendirici ve çoklu tanı testinin bulunduğu çalışmalarda değerlendiriciler arasındaki uyumun ve tanı testleri arasındaki uyumun diagnostik doğruluk sonuçlarına etkisine bakılmıştır. Bu nedenle değerlendiriciler arasında uyumun çok yüksek (0.90) veya çok düşük (0.10), tanı testleri arasındaki uyumun çok yüksek (0.90) veya çok düşük (0.10) olduğu durumlar olmak üzere 2 farklı deneme planı oluşturulmuş ve simülasyon çalışması yapılmıştır. Birden çok değerlendirici ve birden çok tanı testinin bulunduğu doğruluk çalışmalarında değerlendiriciler ve tanı testleri arasında uyum ve uyumsuzluğun var olması kaçınılmazdır. Her koşuldaki Tip I Hata artma riskini ortadan kaldırmak için örnek genişliği, tanı testi sayısı ve değerlendirici sayısının dengelenmesi gerekmektedir.

Anahtar Kelimeler: Birden Çok Değerlendirici-Birden Çok Tanı Testi, DBM metot, OR metot

ABSTRACT

Multi-Reader Multi Case Study in Diagnostic Performance Analysis

Diagnostic evaluation for one group observation may be done by one or more reader and the traditional ROC analysis is often used to analyse test results. In traditional ROC analysis, the objective is not to evaluate the performance of estimators. But, it is possible to resolve both the performance of diagnostic tests in classification of observation groups and the performance of readers by multireader multicase (MRMC) analyses unlike traditional ROC analysis. The main objective of these kinds of studies is not to collect ROC curves. The objective is to reveal the accuracy of estimators or the accuracy of diagnostic test between the averages of readers. Assessment of complete cross validation of ROC data with a trial plan, in which every reader assess every observation, was referred as MRMC studies for the first time in 1992 by a study of Dorfman.

In the study, he conducted the classical variance analysis study by using Jackknife resampling method, in which an observation is left out every time, and for this purpose, he recommended DBM method. In this study, the effect of coherence among estimators and among diagnostic tests in studies that comprise multiple estimators and multiple diagnostic tests on the results of diagnostic accuracy was analyzed. For this reason, two different trial plans were generated as situations, in which the coherence among estimators is either high (0.90) or low (0.10) and the correlation between diagnostic tests is either powerful (0.90) or poor (0.10), and then simulation study was conducted. The existence of coherence or incompatibility among estimators and diagnostic tests in the accuracy studies that have multiple readers and diagnostic tests is inevitable. The sample size, the number of diagnostic test and the number of readers should be balanced in order to remove the risks of Type I error increase in all conditions.

Key Words: Multi Reader-Multi Case, DBM method, OR method.

1.GİRİŞ

Hastanın durumu hakkında klinik karar verirken hastalığın varlığını ya da yokluğunu ortaya çıkaran tanı testleri, görüntüleme yöntemleri, laboratuvar bulguları, biyolojik markerlar çoğu zaman tek başlarına karar vermede yeterli değildirler. Gerçeği % 100 e yakın olasılıklarla ortaya koyan test sonuçları gold standart testler çoğu zaman ya hasta için riskli ve zahmetli ya da oldukça pahalıdır. Ayrıca her tanı için gold standart test de bulunamayabilir. Teknolojik gelişmelere bağlı olarak tanı amaçlı kullanılan klasik testlere alternatif olarak çok sayıda yeni testler geliştirilmeye devam edilmektedir. Bu testlerin var olan standart testlerden üstün, zayıf ya da alternatif olduklarının test edilmesi bu amaçla yapılan araştırmaların birçoğunun konusunu oluşturmaktadır. Bu araştırmalarda kullanılacak tanı testinin yapısı, tanı testi sayısı, değerlendirici sayısı, değerlendiriler arası uyum dikkate alınmalı ve gerekli minimum örnek genişlikleri de hesaplanmalıdır.

Birbirinin alternatifi olan testler ile karar verilirken bir gold standartın var olduğu, tanı testleri arasında korelasyonun olmadığı ve tek bir değerlendiricinin var olduğu durumda genellikle ortaya çıkan senaryolar aşağıdaki gibi özetlenebilir.

Birinci durumda standart teste göre kullanımı daha kolay ve/veya daha ekonomik olan bir test, standart teste göre performansı bakımından karşılaştırılır. Diğer bir durum ise birden fazla tanı testi ve/veya kuralı uygulanarak birleştirilir ve elde edilen sonuçlar referans testle karşılaştırılır. Üçüncü bir durum ise çok değişkenli geleneksel istatistik metotları kullanarak tanı testlerinin birleştirilmesi şeklindedir. Bu değerlendirmeleri tanı koyma gücü ve maliyetleri açısından karşılaştırmak gerekir (1-5).

Uygulamada birden fazla değerlendiricinin ve değerlendiriciler arası uyumun düşük ya da yüksek olduğu durumlar söz konusu olabilir. Birden fazla değerlendiricinin var olduğu durumda tanı testlerinin diagnostik performansı değerlendirilirken tanı testinin performansını tek bir değerlendiriciymiş gibi değerlendiriciler için genellemek doğru değildir. Ayrıca da bir ya da birden fazla değerlendirici var olsa bile tanı testlerinin performansları karşılaştırılır iken testler arasında farklı korelasyon yapıları var olabilir.

Bu nedenle tanı testlerinin karşılaştırılması ve birleştirilmesi değerlendirilir iken, değerlendiriciler arası uyumunda birlikte tartışılması gerekmektedir. Değerlendiriciler arası uyum tartışılır iken değerlendirici sayısı, tanı testi sayısı, tanı testinin ölçüm düzeyi ve örnek genişliği önem taşımaktadır (6-10).

Ayrıca tanı testleri tek bir değerlendirici ile yapıldığında değerlendiricinin kişisel kararlarına bağlı olarak sübjektiflik içerebilir. Değerlendiricilerin performansı değerlendiricilerin deneyimleri, eğitimleri ve vakaların mevcut durumlarıyla (yaş, cinsiyet, hastalığın evresi, takip eden hastalıklar, hastalığın ölüm oranı...) ile ilgilidir. Bu nedenle diagnostik testlerin doğruluğu tartışılırken çoklu değerlendiricilerin raporları da modelde ele alınmalıdır. Hem çoklu değerlendirici hem de çoklu tanı testi sonuçlarının ele alındığı doğruluğu ortaya koyan modeller mevcuttur (11,12).

ROC verilerinin analizinde geleneksel metotlar ya tek bir değerlendiricinin bir grup vaka üzerindeki sonuçlarını ya da birden fazla değerlendiricinin yine bir grup vaka üzerindeki sonuçları üzerinden yapılır. Geleneksel ROC analizinde amaç değerlendiricilerin performansını değerlendirmek değildir. Fakat geleneksel ROC analizinden farklı olarak hem tanı testlerinin vaka gruplarını ayırmadaki performansını hem de değerlendiricilerin performansını MRMC (Multi Reader Multi Case) analizleriyle çözümlenmek mümkündür. Bu yöntem bu tezde bundan sonra kısaca ÇTÇD (Çok Testli Çok Değerlendiricili) ROC analizi olarak ifade edilecektir. Bu analiz ROC verilerinin toplanması ile başlar ve bir çeşit faktöriyel deneme tasarımını oluşturur. Bu tür çalışmaların temel amacı ROC eğrileri toplamak değildir. Amaç, değerlendiricilerin ya da değerlendiricilerin ortalamalarından yararlanarak tanı testlerinin ve değerlendiricilerin ortak bir modelle diagnostik performansını ortaya çıkarmaktır.

Her değerlendiricinin her vakayı değerlendirdiği tam çapraz deneme planlı ROC verilerinin değerlendirilmesini Dorfman ve ark.(1992), çalışmalarında ilk defa ÇTÇD olarak isimlendirmişlerdir. Çoklu değerlendiricili ROC çalışmaları analiz edilirken kullanılan farklı istatistik analiz yöntemleri vardır (13).

Metotlardan en sık kullanılanı Dorfman ve ark. (1992) tarafından önerilen DBM yöntemi ve Obuchowski ve Rockette (1995) tarafından önerilen düzeltilmiş F metodudur kısaca OR olarak adlandırılır. DBM sözde değerler için üç yönlü ANOVA

yaparken, OR doğruluk tahminleri için ilişkilendirilmiş hataların iki yönlü ANOVA analizini yapar. ÇTÇD çalışmaları değerlendiriciler örnekleme ve hasta örnekleme olarak iki örneklem içerir. Eğer çalışma sonuçları, çalışma örneklemine benzer hastalara ve çalışma örneklemine benzer değerlendiricilere genellenebilir ise istatistik analizlerde hem hasta hem de değerlendiriciler için rasgele etki modeli kullanılır. Bu nedenle ÇTÇD çalışmalarında vakalar ve değerlendiriciler için rasgele etki modeli kullanılır. Tanı testleri için ise etkilerin faktörler arasında değişmediğini varsayan sabit etki modeli kullanılır (14,15,16).

ÇTÇD deneme planlarında diagnostik doğruluk tartışılırken değerlendiriciler arası uyum/uyumsuzluk durumunun, tanı testleri arasındaki uyum/uyumsuzluk durumunun, tanı testi ve değerlendirici sayısının ve tanı testlerinin kategorisinin birlikte değerlendirilmesi gerekmektedir. Fakat ÇTÇD ROC analizlerinde bu faktörlerin ve kombinasyonlarının diagnostik doğruluk üzerine etkisinin araştırıldığı bir çalışma literatürde mevcut değildir. Bu nedenle ÇTÇD deneme planlı çalışmalarda değerlendiriciler arasında uyumun çok yüksek (0.90) veya çok düşük (0.10), tanı testleri arasındaki korelasyonun çok güçlü (0.90) veya çok zayıf (0.10) olduğu durumlar olmak üzere 2 farklı deneme planı oluşturulmuştur. Her bir deneme planında da vakaların örnek büyüklüğü (15-30 ve 100), tanı testi sayısı (t=2-5-7), tanı testinin ölçüm düzeyi (2-3-5-7-10) ve değerlendirici sayısı (2-5-7) alınarak kombinasyonlar hazırlanmıştır. Ve tüm kombinasyonlarda % 5 teorik Tip I Hata koşulu altında red edilen H_0 hipotezleri sayılarak gerçekleşen Tip I Hatalara bakılmıştır.

2. GENEL BİLGİLER

2.1. Çok Testli Çok Değerlendiricili Diagnostik Performans Analizi

ROC verilerinin analizinde genellikle geleneksel metotlar kullanılır. Bu durumda ya tek bir değerlendiricinin bir grup vaka üzerindeki sonuçları ya da birden fazla değerlendiricinin yine bir grup vaka üzerindeki sonuçları analiz edilir (17). Diagnostik değerlendirme bir ya da birden fazla değerlendirici tarafından yapılmış olabilir. Bu durumda da değerlendirici grubu ya da her bir tanı testi için ortalama eğri altında kalan alanlar hesaplanarak performans değerlendirmesi yapılır (18,19). Ancak klinikte karşılaşılan veri yapısı aslında bu kadar basit değildir. Tanı amaçlı veriler, korelasyonlu veri (correlated data) ya da kümelenmiş veri (clustered data) şeklinde olabilir. Bir vakadan farklı tanı testleri için ölçümler alındığında korelasyonlu veri, bir vakanın farklı bölgelerinden örneğin sağ ve sol gözünden ya da sağ ve sol el bileklerinden ölçümler alındığında kümelenmiş veriler elde edilir. Bu deneme düzenine farklı değerlendiricilerin verileri de eklendiğinde veri analizinde geleneksel ROC analizinden farklı olarak hem tanı testlerinin vaka gruplarını ayırmadaki performansını, hem de değerlendiricilerin performansını ÇTÇD ROC analizleriyle çözümlenmek gerekir (11).

2.1.1. Çok Testli Çok Değerlendiricili (ÇTÇD) Diagnostik Doğruluk Analizinin Tarihsel Gelişimi

Her değerlendiricinin her vakayı değerlendirdiği tam çapraz deneme planlı ROC verilerinin değerlendirilmesini Dorfman ve ark.(1992), çalışmasında ilk defa ÇTÇD çalışmaları olarak değinmişlerdir. Çalışmada her defasında bir vakanın dışarıda tutulduğu Jackknife yeniden örnekleme yöntemini kullanarak klasik varyans analiz çalışması yapmış ve bu amaçla DBM metodunu önermiştir (13).

Daha sonra Obuchowski ve ark. çalışmalarında korelasyonlu yapıdaki ROC verileri için değerlendiriciler ve hastalar için olması gereken minimum örnek genişlikleri üzerinde durmuşlardır (20).

Toledano ve ark. çalışmalarında korelasyonlu ROC verilerinin analizi için Genelleştirilmiş Tahmin Denklemleri (Estimated Equation Metot, GEE) ve Ordinal Regresyonu kombine ederek genel bir çerçeve oluşturmuşlardır (21).

Beiden ve ark. çalışmalarında geleneksel ANOVA tahminlerine alternatif olarak bir bootstrap yöntemi önermiştir. Beiden, Wagner ve Campbell (BWC) modeli ROC değerlerinin varyansı vakalar ve değerlendiriciler, tanı testleri ile birlikte değerlendirici, hasta ve diğer interaksyonlar yüzünden varyans bileşenlerinin toplamı olacağını modellerinde belirtmişlerdir (22).

Obuchowski ve ark. çalışmalarında korelasyonlu yapıdaki ROC verileri için değerlendiriciler ve hastalar için olması gereken minimum örnek genişliklerine ilişkin araştırmacıya kolaylık sağlaması amacı ile tablolar hazırlamıştır (23).

Zhou ve ark. diagnostik doğruluk için kapsamlı bir kitap yazmışlardır (11).

Song ve ark. çalışmasında çok tanılı çok değerlendiricili ROC çalışmalarına yeni bir marjinal yöntem önermişlerdir (24).

Hillis ve ark. çalışmalarında korelasyonlu ROC verilerinde kullanılan DBM ve OR yöntemlerini karşılaştırmışlardır (15). Hillis ve ark. çalışmalarında korelasyonlu ROC verilerinde kullanılan DBM ve OR yöntemlerinin serbestlik derecelerini karşılaştırmış ve her iki yöntem için yeni bir serbestlik dereceleri önermiştir (25).

Klinik uygulamalarda çok testli çok değerlendiricili ROC analizlerinin uygulamasına en çok değerlendiricilerin performanslarını karşılaştırmak amacı ile Radyoloji araştırmalarında rastlanmaktadır (26-29).

2.1.2.Çok Testli Çok Değerlendiricili (ÇTÇD) ROC Analizlerinin Amaçları

Aynı hastadan alınan tanı test sonuçlarının birbiri ile ilişkili olması ve buna ilaveten aynı değerlendiricilerin birden fazla vaka üzerinde karar vermesine karşın, analizlerde bu ilişkilerin göz ardı edilmesi büyük sorunlara yol açar. Bu tür verilerin var olduğu çalışmalarda doğruluk analizleri geleneksel yöntemlerden farklıdır. Bu tür verilerin analizinde, genel doğruluğun unsurları olan, tanı testleri arasındaki varyasyon, değerlendiriciler arası varyasyon, vakalar arası varyasyon ve bunların arasındaki etkileşimin dikkate alınması gerekir. Bu deneme tasarımı araştırmacıya hastanın tanı sonucunun değerlendiriciden değerlendiriciye ve tanı testinden tanı testine değişip

değişmediği konusunda fikir verir. Ayrıca bir çeşit meta analiz çalışması olarak, birden fazla tanı test sonuçlarının veya birden fazla değerlendiricilerin sonuçlarının birleştirilmesine olanak sağlayarak genelleme şansı şansı verir (11).

2.1.3. Çok Testli Çok Değerlendiricili (ÇTÇD) Deneme Tasarımları

Birden fazla tanı testinin ve birden fazla değerlendiricinin bulunduğu çalışmalarda farklı deneme tasarımları mevcuttur. Çalışmalarda amaç iki ya da daha fazla tanı testi sonuçlarını karşılaştırmak ise eşleştirilmiş değerlendirici (a paired-reader) ya da eşleştirilmemiş değerlendirici (unpaired-reader) tasarımları kullanılır. Eşleştirilmiş değerlendirici tasarımında her değerlendirici bütün tanı testi sonuçlarını değerlendirir. Fakat eşleştirilmemiş değerlendirici tasarımında farklı değerlendiriciler farklı tanı testi sonuçlarını yorumlar. Eşleştirilmiş deneme tasarımı eşleştirilmemiş deneme tasarımından daha güçlüdür. Bu durumda daha az vaka ve daha az değerlendirici ile daha doğru istatistiksel kararlar verilebilir.

En genel tasarım Obuchowski ve Rockette 1995 'de önerdiği ya eşleşmiş vaka, eşleşmiş değerlendirici tasarımıdır. Bu tasarım hastalara bütün tanı testlerinin uygulandığı bütün değerlendiricilerin bütün test sonuçlarını değerlendirdiği tasarımıdır. Geleneksel tasarım Çizelge 2.1'de görülmektedir.

Çizelge 2.1: Geleneksel Tasarım

Vaka	Değerlendirici 1		Değerlendirici j		Değerlendirici J	
	Test 1	Test 2	Test 1	Test 2	Test 1	Test 2
1	Y_{111}	Y_{121}	Y_{11j}	Y_{12j}	Y_{11J}	Y_{12J}
2	Y_{211}	Y_{221}	Y_{21j}	Y_{22j}	Y_{21J}	Y_{22J}
.
k	Y_{k11}	Y_{k21}	Y_{k1j}	Y_{k2j}	Y_{k1J}	Y_{k2J}
.
C	Y_{c11}	Y_{c21}	Y_{c1j}	Y_{c2j}	Y_{c1J}	Y_{c2J}

Geleneksel deneme tasarımında toplam c vaka J değerlendirici ve I tanı testi mevcuttur. Burada her bir değerlendirici C*I farklı yorumlama yapar. Burada Y_{k1j} ve Y_{k2j} j. Değerlendirici tarafından k. hasta için 1. ve 2. test sonucunun değerlendirilmesidir. Geleneksel tasarım vaka sayısını en az kullanabildiğinden en yaygın kullanılan tasarımdır.

ÇTÇD çalışmalarında eşleştirilmiş ve eşleştirilmemiş deneme tasarımları hem eşleşmiş hem de eşleşmemiş vaka (paired-and unpaired-patient) tasarımı ile birlikte kullanılabilir. Çizelge 2.2’de eşleşmemiş değerlendirici-eşleşmemiş vaka deneme tasarımı kullandığı tasarım görülmektedir.

Çizelge 2.2: Eşleşmemiş değerlendirici, Eşleşmemiş vaka deneme tasarımı

Vaka	Test 1			Vaka	Test 2		
	D _a	D _b	D _c		D _d	D _f	D _g
1	Y _{1a1}	Y _{1b1}	Y _{1c1}	1'	Y _{1'd2}	Y _{1'f2}	Y _{1'g2}
2	Y _{2a1}	Y _{2b1}	Y _{2c1}	2''	Y _{2'd2}	Y _{2'f2}	Y _{2'g2}
.				.			
K	Y _{ka1}	Y _{kb1}	Y _{kc1}	k'	Y _{k'd2}	Y _{k'f2}	Y _{k'g2}
.							
C	Y _{ca1}	Y _{cb1}	Y _{ca1}	c'	Y _{c'd2}	Y _{c'f2}	Y _{c'g2}

Bu tasarım çok kullanılmayan bir deneme tasarımıdır. Pepe, 1997 de yapmış olduğu bir çalışmada bir çeşit mamaografî çalışması planlamıştır. Bu çalışmanın amacı bir değerlendiricinin doğruluğunu etkileyen başka bir faktörü ortaya çıkarmaktır (30).

Ayrıca eşleşmemiş vaka eşleşmiş değerlendirici, eşleşmiş vaka eşleşmemiş değerlendirici, hybrid deneme tasarımları da mevcuttur (11,31).

ÇTÇD ROC çalışmalarında en genel deneme tasarımı R değerlendiricinin C vakayı K tanı testini birlikte değerlendirdiği eşleşmiş vaka, eşleşmiş değerlendirici tasarımıdır.

Kabul edelim ki, R değerlendirici, C vaka ve K tanı testi var olsun. Bu durumda her değerlendiricinin C*K tane sonucu ve bütün çalışmanın ise R*K*C tane sonucu vardır. ÇTÇD analizlerinin deneme düzenleri bir tür faktöriyel deneme düzenidir. Her değerlendirici bir ya da birden fazla test sonucunu değerlendirir. Çizelge 2.3' de ÇTÇD ROC analizi için deneme tasarımı görülmektedir.

Çizelge 2.3: ÇTÇD Doğruluk Analizleri için Faktöriyel Deneme Tasarımı

Vaka (Hasta/Sağlam)	TANI TESTLERİ															
	1				2							K			
	Değerlendirici				Değerlendirici				...				Değerlendirici			
	R ₁	R ₂	.	R _r	R ₁	R ₂	.	R _r	R ₁	R ₂	...	R _r
1	Y ₁₁₁	Y ₁₂₁	.	Y _{1r1}	Y ₁₁₂	Y ₁₂₂	.	Y _{1r2}	Y _{11k}	Y _{12k}	...	Y _{1rk}
2	Y ₂₁₁	Y ₂₂₁	.	Y _{2r1}	Y ₂₁₂	Y ₂₂₂	.	Y _{2r2}	Y _{21k}	Y _{22k}	...	Y _{2rk}
.
.
c	Y _{C11}	Y _{C21}	.	Y _{Cr1}	Y _{C12}	Y _{C22}	.	Y _{Cr2}	Y _{C1k}	Y _{C2k}	.	Y _{Crk}

2.1.4. Çok Testli Çok Değerlendiricili Doğruluk Analizleri İçin Örnek Genişliği Hesaplaması

Obuchowski, çok değerlendiricili ROC çalışmalarında çalışmanın tasarımına bağlı olarak vaka ve değerlendirici varyasyonu, doğruluklar arası ilişkiden türetilen etki büyüklüğünü dikkate alarak bir matematiksel modele ile örnek büyüklüğü çizelgesi oluşturmuştur.

İki tanı testinin doğrulukları arasında farklılık olup olmadığını ortaya koymak amacı % 80 Güç ve Tip I Hatayı % 5 olarak Çizelge 2.4 'deki örnek büyüklüğü çizelgesi geliştirmiştir (23). Bu çizelge, ayırma gücü 0.75 veya 0.90 olan referans teste karşılık, alternatif tanı testi ile, ROC eğrisi altında kalan alanlar arası farklılığı (0.05,0.10, 0.15) değerlendirirken, değerlendiriciler arası varyasyon dikkate alınarak Çizelge 2.4'de oluşturulmuştur. Ayrıca bu çizelgede vakaların hasta ve sağlamların oranını dengeli (1/1) ve hastalar sağlamların iki katı (2/1) ve dört katı (4/1) olacak şekilde çalışmanın tasarımına bağlı olarak ayırmıştır.

Çizelge 2.4: DBM Metodu Örnek Genişliği

			Dört			Altı			On		
			Değerlendirici			Değerlendirici			Değerlendirici		
			Değerlendiriciler Arası Çeşitlilik/Varyasyon								
			Düşük	Orta	Yüksek	Düşük	Orta	Yüksek	Düşük	Orta	Yüksek
Tam Testinin Ayırma Gücü (ROC=0.75)	ROC ₁ - ROC ₂ =0.05	Oran=1/1	571	-	-	246	3769	-	116	201	-
		Oran=2/1	679	-	-	293	4479	-	138	239	-
		Oran=4/1	983	-	-	424	6488	-	200	345	-
	ROC ₁ - ROC ₂ =0.10	Oran=1/1	133	291	-	60	78	943	29	32	51
		Oran=2/1	159	345	-	71	92	1120	35	38	60
		Oran=4/1	229	500	-	103	133	1622	50	55	87
	ROC ₁ - ROC ₂ =0.15	Oran=1/1	59	77	2975	27	30	46	20	20	20
		Oran=2/1	70	92	3536	32	35	54	30	30	30
		Oran=4/1	101	132	5122	50	51	78	50	50	50
Tam Testinin Ayırma Gücü (ROC=0.90)	ROC ₁ - ROC ₂ =0.05	Oran=1/1	287	-	-	124	1896	-	59	101	-
		Oran=2/1	363	-	-	157	2395	-	74	128	-
		Oran=4/1	548	-	-	236	3618	-	112	193	-
	ROC ₁ - ROC ₂ =0.10	Oran=1/1	67	146	-	31	39	474	20	20	26
		Oran=2/1	85	185	-	38	50	599	30	30	32
		Oran=4/1	128	279	-	58	75	905	50	50	50
	ROC ₁ - ROC ₂ =0.15	Oran=1/1	30	39	1497	20	20	23	20	20	20
		Oran=2/1	38	49	1890	30	30	30	30	30	30
		Oran=4/1	57	74	2856	50	50	50	50	50	50

2.2. Çok Testli Çok Değerlendiricili (ÇTÇD) ROC Analizlerinde Kullanılan Yöntemler

ÇTÇD ROC çalışmalarında deneme tasarımındaki varyasyon kaynakları, farklı varsayımlarla ve farklı yöntemlerle modeli tanımlamışlardır. Değerlendiriciler arasındaki çeşitlilik belirlenirken metotlar arasında modelin kuruluş şekli bakımından farklılıklar vardır. Modellerde interaksiyonlar, korelasyonlar, varyanslar, kovaryatlar farklı yöntemlerle modelde yer alırlar (14,32).

Çoklu değerlendiricili ROC çalışmaları analiz edilir iken birkaç farklı yöntem vardır. Metotlardan en sık kullanılanı Dorfman ve ark. (1992) tarafından önerilen DBM yöntemi ve Obuchowski ve Rockette (1995) OR tarafından önerilen düzeltilmiş F metodudur. DBM sözde değerler için üç yönlü ANOVA yaparken, OR doğruluk tahminleri için ilişkilendirilmiş hataların iki yönlü ANOVA analizini yapar. ANOVA analizlerinde eğer çalışma sonuçları çalışma örneğine benzer hastalara ya da çalışma örneğine benzer değerlendiriciler için genellenebilir ise hem hasta hem de değerlendiriciler için rasgele etkili modeli kullanılır. ÇTÇD çalışmaları değerlendiriciler örnekleme ve hasta örnekleme olarak iki örneklem içerir. Eğer çalışma sonuçları çalışma örneğine benzer hastalara ya da çalışma örneğine benzer değerlendiriciler genellenebilir ise istatistik analizlerde hem hasta hem de değerlendiriciler için rasgele etkili model kullanılır.

Çok testli çok değerlendiricili çalışmada hastalar ve değerlendiricilerin rasgele etkili, tanı testlerinin özel seçilmiş olduğu varsayılır. Bu nedenle varyans analizleri karışık (mixed) etkili model kullanılarak yapılır.

Değerlendiricilerin doğruluğu ve güvenirliliği, hastaların özelliklerinden (yaş, cinsiyet, hastalık aşaması, ölüm oranı) etkilenme durumlarıyla, değerlendiricilerin deneyim ve eğitimleriyle ilgilidir. Bu değişkenler kovaryat olarak adlandırılır. Bazı istatistiksel metotlar kovaryatları içerebilir. Bu modeller hastalar arasındaki farklılığı ve değerlendiriciler arasındaki farklılığı ortaya çıkarır. Bu metotlar aşağıdaki bölümlerde açıklanmıştır.

2.2.1. Dorfman-Berbaum-Metz (DBM) Metodu/ Söзде Değerlerle ANOVA

2.2.1.1. DBM Metodu Teorisi

Çoklu değerlendiricili ROC çalışmalarında en sık kullanılan metot Dorfman-Berbaum-Metz (DBM) metodudur. Dorfman ROC verilerinin çoklu değerlendiricili ile yapıldığı çalışmalarda (mixed effect) karışık etkili ANOVA modelini kullanmaktadır. Modelin testinde amaç “değerlendiricilerin doğrulukları arasında fark yoktur”, “diagnostik testlerin doğruluğu arasında fark yoktur” ve “değerlendiricilerin ortalama doğruluğu bütün diagnostik testler için aynıdır” ifadeli yokluk hipotezlerini test etmektir. Yokluk hipotezinin testinde Quenouille-Tukey Jackknife (pseudo values) söзде değerleri ile ANOVA hesaplaması yapılır. Karışık etkili ANOVA, söзде değerler ile yapılır. Söзде değerler orijinal verinin bir çeşit transformasyonu gibi düşünülür.

t tane tanı şekli ($i=1,2,\dots,t$), r tane değerlendirici ($j=1,2,\dots,r$) ve c tane de vakanın bulunduğu bir veri seti ($k=1,2,\dots,c$) olsun. ROC verileri analiz edilirken her değerlendirici için Maksimum Likelihood (ML) tahmin edicisi kullanılarak, her değerlendirici-tanı kombinasyonu için binormal ROC eğrisi elde edilir. Her değerlendirici-tanı kombinasyonu için ROC eğrisinden bir A_z indeksi elde edilir. ML tahmini j değerlendirici tarafından i tanı değişkeni için bütün vakalar üzerinden yapılır ve A_{ij} ile gösterilir. Sonra 1. vaka silinir ve ML ile A_{ij} nin yeni bir tahmini hesaplanır. Sonra 1. vaka örnekleme yeniden dahil edilir ve 2. vaka silinir ve A_{ij} nin yeni bir tahmini hesaplanır ve böylece bu işlemler bir döngü şeklinde devam eder. Bu durum örneklemedeki her vaka bir kez silininceye kadar devam eder. Bu işlemler, eğri altında kalan alan c kez hesaplanır. Çünkü örnekleme üzerinde c tane vaka vardır.

Bu durumda $A_{ij(k)}$ k . vaka silindiğinde A_{ij} nin ML tahminini gösterir. J değerlendirici ve i tanı için k tane söзде değer A_{ij*k} ile gösterilir ve denklem 2.1 ile hesaplanır.

$$A_{ij*k} = cA_{ij} - (c-1)A_{ij(k)} \quad (2.1)$$

Burada c ve c-1 terimleri k. vaka silinerek oluşturulan alt kümede k. vakanın ağırlığını gösterir. Bu ilişki, bütün vakalar üzerinde A_{ij*k} k. Vakanın ortalama etkisi olarak düşünülebilir. Silinen vakanın hasta ya da sağlam olup olmasına bağlı olarak, negatif ya da pozitif vaka sayısından bir vaka düşülür.

DBM metodu (Dorfman, 1992) karma etkili doğrusal ANOVA modeli olarak kabul edilir. Bu metodun modeli denklem 2.2’de verilmiştir.

$$Y_{ijk} = \mu + \tau_i + R_j + C_k + (\tau R)_{ij} + (\tau C)_{ik} + (RC)_{jk} + (\tau RC)_{ijk} + \varepsilon_{ijk} \quad (2.2)$$

Burada

μ : Populasyon ortalamasını,

τ_i : i tanı testinin sabit etkisini

R_j : j değerlendiricinin tesadüfi etkisini

C_k : k vakanın tesadüfi etkisini

ε_{ijk} : k vaka, j değerlendirici ve i. tanın hata miktarını gösterir.

Parantez ile verilen ifadeler interaksyonu gösterir.

İnteraksiyon terimlerinin (tanı_testi*değerlendirici, tanı_testi*vaka, değerlendirici*vaka ve tanı_testi*vaka*değerlendirici) hepsi rasgele etkilidir.

Bir tanı testinin performansı için F istatistiği denklem 2.3’de verilmiştir.

$$F_{DBM} = \frac{MS(T)_{pseudo}}{MS(T * R)_{pseudo} + MS(T * C)_{pseudo} - MS(T * R * C)_{pseudo}} \quad (2.3)$$

Burada;

MS(T): Tanı etkisinin kareler toplamını

MS(T*R): Tanı-değerlendirici interaksiyon etkisinin kareler toplamını

MS(T*C): Tanı-vaka interaksiyon etkisinin kareler toplamını

MS(T*R*C): Tanı*değerlendirici*vaka interaksiyon etkisinin kareler toplamını göstermektedir.

Eğer F değeri ; $(F_{df_1, dfsatt}); (1-\alpha)$ F tablo değerinden büyükse tanı testlerinin diagnostik doğrulukları arasındaki farklılık anlamlıdır. F istatistiğine karşılık gelen kritik tablo değerlerinin serbestlik dereceleri denklem 2.4 ve denklem 2.5'deki gibi bulunur.

Burada

$$df_1 = (t-1) \quad (2.4)$$

$$dfsatt = \frac{[MS(T * R)_{pseudo} + MS(T * C)_{pseudo} - MS(T * R * C)_{pseudo}]^2}{\frac{MS(T * R)_{pseudo}^2}{(t-1)(r-1)} + \frac{MS(T * C)_{pseudo}^2}{(t-1)(c-1)} + \frac{MS(T * R * C)_{pseudo}^2}{(t-1)(r-1)(c-1)}} \quad (2.5)$$

ÇTÇD ROC analizlerinde performans değeri eğri altında kalan alandır (13,15,33-36).

2.1.2. DBM Metodunun Avantajları

- Jackknife yeniden örnekleme tekniğini kullanır.

- Ortalama ROC eğrisi hesapladığı için, her bir bireyin tek başına eğri altında kalan alanı üzerindeki bireysel etkisini yansıtmaya olanağı sağlar.
- LABMRMC 1.0 ve LABMRMC 2.0 ve DBM ROC yazılımları internetten ücretsiz olarak indirilebilir ve kullanılabilir (37-39).
- İyi bilinen ANOVA modelini kullanır.

2.2.2. Obuchowski ve Rockette Metodunun (Düzeltilmiş F Metodu) Teorisi

Obuchowski ve Rockette tarafından OR metodu olarak kısaltılan ve düzeltilmiş F testi olarak da ifade edilen matematiksel bir model önerilmiştir. Modelde değerlendiriciler sabit etkili, tanı testleri ve vakalar rasgele etkili kullanılarak tahmin yapılır. Buradaki doğruluk değeri, DBM metodundaki gibi sözde değerler kullanılarak değil, gerçek değerler kullanılarak yapılır. DBM metodu aynı vakalardan çoklu değerlendiricilerin her birinin performansının çoklu tekrarlarına izin verir. Fakat düzeltilmiş F testi faktoriyel deneme tasarımı için yalnız bir replikasyonda hesaplanır ve bu deneme tasarımı için model denklem 2.6'daki gibi yazılabilir.

$$\hat{\theta}_{ij;gerçek} = \mu + \tau_i + R_j + (\tau R)_{ij} + \varepsilon_{ij} \quad (2.6)$$

Burada

i tanı testi (i=1,2,...,t) j değerlendirici (j=1,2,...,r) olmak üzere

$\hat{\theta}_{ij;gerçek}$ j.değerlendirici, i. tanı test için eğri altında kalan alanın tahmin değerini,

μ : Populasyon ortalamasını,

τ_i : i tanının sabit etkisini

R_j : j değerlendiricinin rasgele etkisini

$(\tau R)_{ij}$: Değerlendirici-tanı testi interaksiyon teriminin rasgele etkisini

ε_{ij} : j değerlendirici ve i tanının hata terimini gösterir.

R_j ve $(\tau R)_{ij}$ terimlerinin ortalaması sıfır varyansları σ_R^2 ve $\sigma_{\tau R}^2$ olan ortak bağımsız normal dağılım gösterdikleri varsayılır. Eşitlik 2.6 'daki ε_{ij} 'de ortalaması sıfır ve varyansı sabit (σ_ε^2) normal dağıldığı varsayılır.

Ayrıca $\varepsilon_{ij}; R_j$ ve $(\tau R)_{ij}$ da bağımsız olduğu varsayılır. Fakat her tanı için her değerlendiricinin aynı vakayı okuması sonucunda, ε_{ij} 'nin ortak bağımsız olduğu varsayımını kabul etmez. Bunu yerine testler ve değerlendiriciler arasındaki hataların kovaryansları oluşturulur ve denklem 2.7'deki gibi ifade edilir.

$$\text{Cov}(\varepsilon_{ij}, \varepsilon_{i'j'}) = \begin{bmatrix} \text{Cov1} & i \neq i' & j = j' \\ \text{Cov2} & i = i' & j = j' \\ \text{Cov3} & i \neq i' & j \neq j' \end{bmatrix} \quad (2.7)$$

Obuchowski ve Rockette (15,16) bu kovaryansları denklem 2.8'deki gibi sıralamışlardır.

$$\text{Cov1} \geq \text{Cov2} \geq \text{Cov3} \geq 0 \quad (2.8)$$

Bu kovaryans değerlerinden; denklem 2.9 farklı tanı testleri-aynı değerlendirici için oluşturulan Cov1'in, denklem 2.10 aynı tanı testi-aynı değerlendirici için oluşturulan Cov2 'nin, denklem 2.10'da farklı tanı testi-farklı değerlendirici için oluşturulan Cov3 değerinin hesaplanmasını göstermektedir.

$$\text{Cov1} = \text{Cov}(\varepsilon_{ij}, \varepsilon_{i'j'}) = \text{Cov}(\hat{\theta}_{ij}, \hat{\theta}_{i'j'} / R, \tau R) \quad (2.9)$$

$$Cov2 = Cov(\varepsilon_{ij}, \varepsilon_{ij'}) = Cov(\hat{\theta}_{ij}, \hat{\theta}_{ij'} / R, \tau R) \quad (2.10)$$

$$Cov3 = Cov(\varepsilon_{ij}, \varepsilon_{ij'}) = Cov(\hat{\theta}_{ij}, \hat{\theta}_{ij'} / R, \tau R) \quad (2.11)$$

Buradaki kovaryans $Cov(. / R, \tau R)$; şartlı değerlendirici ve tanı-değerlendirici etkisinin kovaryansıdır.

Eğer hatalar bağımsız ise bir tanının etkisini test etmek için ($H_0 : \tau_1 = \tau_2 = \dots = \tau_i$)

denklem 2.12’de kullanılan F istatistiği kullanılır.

$$F = \frac{r \sum_{i=1}^t (\hat{\theta}_i - \hat{\theta}_{..})^2 / (t-1)}{\sum_{i=1}^t \sum_{j=1}^r (\hat{\theta}_{ij} - \hat{\theta}_i - \hat{\theta}_{.j} + \hat{\theta}_{..})^2 / [(t-1)(r-1)]} = \frac{MS(T)_{\hat{\theta}_{ij}; orjinal}}{MS(T * R)_{\hat{\theta}_{ij}; orjinal}} \quad (2.12)$$

Fakat denklem 2.12’deki F istatistiği, denklem 2.6 ‘daki modelde hatalar korelasyonlu olduğu için geçersizdir. Bu nedenle Obuchowski ve Rockette (1995) denklem 2.6’daki modele bir düzeltme yaparak denklem 2.13’deki F istatistiğini oluşturmuşlardır.

$$F = \frac{MS(T)_{\hat{\theta}_{ij}; orjinal}}{MS(T * R)_{\hat{\theta}_{ij}; orjinal} + r(cov2 - cov3)} \quad (2.13)$$

Uygulamada Cov2 ya da Cov3 bilinmez bu nedenle ya pilot bir çalışma verilerinden ya da önceki çalışmalardan tahminleri kullanılır. Bu nedenle F istatistiği denklem 2.14 ‘deki gibi yazılır

$$F = \frac{MS(T)_{\hat{\theta}_{ij};orjinal}}{MS(T * R)_{\hat{\theta}_{ij};orjinal} + r(\hat{c}v_2 - \hat{c}v_3)} \quad (2.14)$$

Karar eğer F değeri ; $(F_{df_1, df_2}; (1 - \alpha))$ tablo değerinden büyükse tanı testlerinin diagnostik doğrulukları arasında farklılık anlamlıdır. F istatistiğine karşılık gelen kritik tablo değerleri için serbestlik dereceleri denklem 2.15 ve denklem 2.16'da verilmiştir. (15,34,40-42).

Burada

$$df_1 = (t - 1) \quad (2.15)$$

$$df_2 = (t - 1) * (r - 1) \quad (2.16)$$

2.2.3.Çok Değişkenli WMW istatistikleri

Song (1997) çoklu değerlendircili çalışmalarda nonparametrik bir yaklaşım önermiştir. Önerilen modelin herhangi bir matematiksel modeli yoktur ve model tamamen non-parametrikdir. Model her vaka için bir güven skoru oluşturularak kurulur. Bu modelde de değerlendiriciler modele, sabit etkili olarak katılırlar. Bu metotta diagnostik doğruluk olarak eğri altında kalan alan kullanılır. Bu yöntem için herhangi bir bilgisayar yazılımı yoktur (41).

2.2.4. Hiyerarşik Ordinal Regresyon ROC Metodu

Ishwaran ve Gatsonis (2000) çoklu değerlendiricili ROC deneme tasarımlarında Bayesci hiyerarşik ordinal regresyon modelini (HROC) geliştirmişlerdir. Radyoloji verilerinde çok düzeyli kümelenmiş veri yapısında kullanmışlardır. HROC modelleri her vaka için bir gizli değişken oluşturur. Ordinal bağımlı değişkenin değerleri çok değişkenli gizli değişkenlerin aralıklarının bölünmesi ile belirlenir. Modelde değerlendiriciler modele rasgele ya da sabit etkili olarak girebilirler. Ayrıca modele kovaryat eklemek mümkündür. HROC modellerinde diagnostik doğruluk ölçümü için eğri altında kalan alan kullanılmaz sadece test sonuçlarını kullanılarak analiz yapılır. (42-44).

2.2.5. Bootstrap of Components-of-Variance “BWC Method”

Beiden, Wagner ve Campbell (2000) çoklu değerlendiricili ROC çalışmaları için DBM metodundaki jackknife yeniden örnekleme yöntemine alternatif olarak bir bootstrap yeniden örnekleme metodu önermişlerdir. Bu metoda göre iki farklı şekilde veriler yeniden örnekleme yapılabilmektedir. Birincisi orjinal veriyi vakalar ve değerlendiriciler üzerinde; ikincisi orjinal veriyi sadece hastalar üzerinde bootstrap yeniden örnekleme yöntemini kullanarak kullanma şeklindedir (22,45,46).

Modelde veriler bootstrap yeniden örnekleme metodu kullanılarak elde edilir ve varyans analizindeki varyasyon kaynakları bu yeni verilerle hesaplanır. Oluşturulan yeni veriler yine eğri altında kalan alan verileridir. Bu veriler ile varyasyon kaynakları hesaplanırken, varyansı vaka ve değerlendiricilerden kaynaklı, her değerlendirici*tanı testi kombinasyonundan kaynaklı ve değerlendirici*vaka kombinasyonlarından kaynaklı varyasyon kaynaklarının toplamı olarak varsayar. Ayrıca varyans tahminlerinde dağılım serbestliği vardır. Değerlendiriciler modele rasgele ya da sabit etkili olarak girebilirler. Model kovaryat içermez. Model için herhangi bir bilgisayar yazılımı yoktur.

2.3. Çok Testli Çok Değerlendiricili (ÇTÇD) ROC Analizlerinde Kullanılan Yöntemlerin Karşılaştırılması

Çok testli çok değerlendiricili ROC analizlerinde kullanılan beş yöntemin model, değerlendiricilerin modele etkisi, kullanılan modelin herhangi bir kovaryatta izin verip vermemesi, modelde kullanılan doğruluk ölçümü ve bir bilgisayar yazılımlarının olup olmaması yönünden Çizelge 2.5’de karşılaştırılmıştır (14).

Çizelge 2.5: ÇTÇD ROC Yöntemlerinin Karşılaştırılması

Özellikler	DBM	OR	Çok değişkenli WMW	BWC	HROC
Model	Sözde Değerler	Doğruluğun özet ölçümü	Her hasta için güven skoru	Doğruluğun özet ölçümü	Her hasta için gizli değişken
Değerlendirici Etkisi	Rasgele ya da sabit etkili	Rasgele ya da sabit etkili olabilir.	Sabit etkilidir.	Rasgele ya da sabit etkili	Rasgele ya da sabit etkili
Kovaryat	Var	Yok	Yok	Yok	Var.
Doğruluk ölçümü	Diagnostik istatistiklerden herhangi biri kullanılır.	Diagnostik istatistiklerden herhangi biri kullanılır.	Eğri Altında Kalan Alan kullanılır.	Diagnostik istatistiklerden herhangi biri kullanılır.	Diagnostik istatistiklerden herhangi biri kullanılır.
Yazılım	MRMC1.0 ³⁸ MRMC 2.0 ³⁸ DBM MRMC 2.0 ³⁹ LABMRCM ³⁷	OBUMRM ⁴⁷ DBM MRMC 2.0 ³⁹	Yok	Yok	Yok

3. GEREÇ ve YÖNTEM

Çoklu değerlendirici ve çoklu tanı testinin bulunduğu çalışmalarda değerlendiriciler arasındaki uyumun ve tanı testleri arasındaki korelasyonun diagnostik doğruluk sonuçlarını (ÇTÇD ROC analiz sonuçlarını) etkilediği düşünülmektedir. Ayrıca tanı testinin sayısı ve tanı testinin ölçüm düzeyi, değerlendirici sayısı, vaka sayısı da doğruluk sonuçlarını etkilemektedir. Bu nedenle ÇTÇD deneme planlı çalışmalarda değerlendiriciler arasında uyumun çok yüksek (0.90) veya çok düşük (0.10), tanı testleri arasındaki korelasyonun çok güçlü (0.90) veya çok zayıf (0.10) olduğu durumlar olmak üzere 2 farklı deneme planı oluşturulmuş ve simülasyon çalışması yapılmıştır. Bu iki farklı simülasyon denemelerinde amaç değerlendiriciler arasındaki uyumun ve tanı testleri arasındaki korelasyonun diagnostik doğruluklarını ne derece etkilediklerini ortaya koymaktır. Her bir kombinasyon için üretilen verilerde gerçekte değerlendiricilerin ortalama doğrulukları bakımından fark yoktur ve bu farksızlık bütün diagnostik testler için aynıdır. Yani değerlendiriciler ve tanı testlerinin doğrulukları arasında farklılık yoktur. Bu nedenle değerlendiricilerin diagnostik doğrulukları için Hipotez takımı I ve tanı testlerinin diagnostik doğrulukları için de Hipotez takımı II aşağıdaki gibi kurulabilir.

Hipotez 1:

H_0 =Değerlendiricilerin ortalama diagnostik doğrulukları bakımından fark yoktur.

H_1 = Değerlendiricilerin ortalama diagnostik doğrulukları bakımından fark vardır.

Üretilmiş olan bütün kombinasyonlarda Hipotez takımı I için Tip I Hata yapma olasılığını, değerlendiricilerin diagnostik doğrulukları arasında gerçekte farklılık yokken var olarak bulunma ihtimali olarak ifade edebiliriz.

Hipotez II:

H_0 =Tanı testlerinin ortalama diagnostik doğrulukları bakımından fark yoktur.

H_1 = Tanı testlerinin ortalama diagnostik doğrulukları bakımından fark vardır.

Bu durumda da üretilmiş olan bütün kombinasyonlarda Hipotez takımı II için Tip I Hata yapma olasılığını, tanı testlerinin diagnostik doğrulukları arasında gerçekte farklılık yokken var olarak bulunma ihtimali olarak ifade edebiliriz.

3.1. Değerlendiriciler Arası Çok Yüksek ve Çok Düşük Uyum Etkisi İçin Yapılan Simülasyon Çalışması

Bu deneme planında amaç değerlendiriciler arasındaki çok yüksek ve çok düşük uyumun tanı testlerinin ve değerlendiricilerin diagnostik doğruluklarına (eğri altında kalan alan) etkisini incelemektir. Bu etki değerlendirici sayısı, tanı testi sayısı ve kategorisi ve de örnek genişliği çerçevesinde değerlendirilecektir. Vakaların örnek büyüklüğü (15,30 ve 100), tanı testi sayısı ($t=2,5,7$), tanı testinin ölçüm düzeyi (2,3,5,7,10) ve değerlendirici sayısı (2,5,7) alınarak kombinasyonlar hazırlanmıştır. Değerlendiriciler arası çok yüksek uyum için iki, beş ve yedi değerlendiricinin her bir kombinasyonu için her bir tanı testlerine verdikleri kararlarda 0.90'lık uyumu yakalamaları sağlanmıştır. Değerlendiriciler arası çok düşük uyum için iki, beş ve yedi değerlendiricinin her bir kombinasyonu için her bir tanı testlerine verdikleri kararlarda 0.10'lık uyumu yakalamaları sağlanmıştır. Her koşulda tanı testlerinin ayırma gücü % 50 olarak alınmıştır. Deneme planlarına ait verilere Matlab 7.0 paket programında 1000 kez üretilmiş ve bu 1000'lik verilerin analizleri genel bilgilerde detayları verilen DBM Metodunun algoritmasını kullanan LABMRMC 1.0 paket programında gerçekleştirilmiştir (48,37) Ek I'de değerlendiriciler arasındaki uyuma ait veri üretim aşaması için oluşturulan sürecin deneme planlarının Matlab kodları verilmiştir. Sonuçların grafik gösterimi SPSS 19.0 istatistik paket programının demo versiyonu kullanılarak yapılmıştır (49).

3.2. Tanı Testleri Arası Çok Güçlü ve Çok Zayıf Korelasyonun Etkisi İçin Yapılan Simülasyon Çalışması

Bu deneme planında amaç tanı testleri arasındaki çok güçlü ve çok zayıf korelasyonun tanı testlerinin, değerlendiricilerin diagnostik doğruluklarına (eđri altında kalan alan) etkisini incelemektir. Bu etki değerlendirici sayısı, tanı testi sayısı ve kategorisi ve de örnek genişliđi çerçevesinde değerlendirilecektir. Vakaların örnek büyüklüđü (15,30 ve 100), tanı testi sayısı ($t=2,5,7$), tanı testinin ölçüm düzeyi (2,3,5,7,10) ve değerlendirici sayısı (2,5,7) alınarak kombinasyonlar hazırlanmıştır. Tanı testleri arası çok güçlü korelasyon için iki, beş ve yedi tanı testi sonucunda her bir kombinasyonu için tanı testi sonuçları arasında 0.90'lık korelasyonu yakalamaları sağlanmıştır. Tanı testleri arası çok zayıf korelasyon için iki, beş ve yedi tanı testi sonucunda her bir kombinasyonu için tanı testi sonuçları arasında 0.10'lık korelasyonu yakalamaları sağlanmıştır. Her koşulda tanı testlerinin ayırma gücü % 50 olarak alınmıştır. Deneme planlarına ait verilerin üretimi her kombinasyon için Matlab 7.0 paket programında 1000 kez üretilmiş ve bu 1000'lik verilerin analizleri DBM Metodunun algoritmasını kullanan LABMRMC 1.0 paket programında gerçekleştirilmiştir (48,37). Ek II'de tanı testleri arasındaki korelasyona veri üretim aşaması için oluşturulan sürecinin deneme planlarının Matlab kodları verilmiştir. Sonuçların grafik gösterimi SPSS 19.0 0 istatistik paket programının demo versiyonu kullanılarak yapılmıştır (49).

4. BULGULAR

4.1.Değerlendiriciler Arasındaki Yüksek ve Düşük Uyumun Diagnostik Doğruluk Üzerine Etkisi

Bu simülasyon denemelerinde amaç değerlendiriciler arasındaki uyumun diagnostik doğruluklarını ne derece etkilediklerini ortaya koymaktır. Her bir kombinasyon için üretilen verilerde gerçekte değerlendiricilerin ortalama doğrulukları bakımından fark yoktur ve bu farksızlık bütün diagnostik testler için aynıdır. Yani değerlendiriciler ve tanı testlerinin doğrulukları arasında farklılık yoktur.

Ek III'de değerlendiriciler arasında uyum çok yüksekken değerlendirici sayısı, tanı testi sayısı ve kategorisi ve farklı örnek büyüklükleri için üretilmiş olan kombinasyonlardaki varyans analiz tablosunun bütününde olabilecek tüm hata kaynakları için Tip I Hata yapma olasılıkları verilmiştir. Ek IV'de ise değerlendiriciler arasında uyum çok düşükken değerlendirici sayısı, tanı testi sayısı ve kategorisi ve farklı örnek büyüklükleri için üretilmiş olan kombinasyonlardaki varyans analiz tablosunun bütününde olabilecek tüm hata kaynakları için Tip I Hata yapma olasılıkları verilmiştir.

4.1.1. Değerlendiriciler Arasındaki Yüksek ve Düşük Uyumun Etkisinin Tanı Testlerinin Doğrulukları Üzerine Etkisi

Değerlendirici sayısı, tanı testi sayısı ve kategorisi, değerlendiriciler arasındaki çok yüksek ve çok düşük uyum durumları için üretilmiş olan kombinasyonlardaki tanı testlerinin diagnostik doğrulukları arasında gerçekte farklılık yokken var olarak bulunma ihtimalini gösteren Tip I Hata yapma olasılıkları iki değerlendirici için Çizelge 4.1, beş değerlendirici için Çizelge 4.2 ve yedi değerlendirici için Çizelge 4.3'de verilmiştir. Grafik gösterimleri ise değerlendiriciler arasında uyum çok yüksekken örnek genişliği sayısı 15,30 ve 100; tanı testi sayısı 2,5 ve 7 için değerlendirici sayılarına ait Tip I Hata yapma olasılığı Şekil 4.1'de gösterilmektedir. Şekil 4.2'de ise değerlendirici sayısı 2-5 ve 7 alınarak tanı testi sayılarına ait gerçekleşen Tip I Hata yapma olasılıkları gösterilmektedir. Değerlendiriciler arasında uyum çok yüksekken örnek genişliği sayısı 15,30 ve 100; tanı testi sayısı 2,5 ve 7 için değerlendirici sayılarına ait Tip I Hata yapma olasılığı Şekil 4.3'de gösterilmektedir. Şekil 4.4'de ise

değerlendirici sayısı 2,5 ve 7 alınarak tanı testi sayılarına ait gerçekleşen Tip I Hata yapma olasılıkları gösterilmektedir.

Çizelge 4.1: İki değerlendirici arasında uyum çok yüksek ve düşükken, tanı testlerinin diagnostik doğruluk farklılıklarına ilişkin,

Olabilecek Tip I Hata yapma olasılıkları

Tanı Testi Sayısı	Örnek Genişliği	İki Değerlendirici Arasında Uyum Çok Yüksek					İki Değerlendirici Arasında Uyum Çok Düşük				
		Tanı Testinin Kategori Sayısı					Tanı Testinin Kategori Sayısı				
		K=2	K=3	K=5	K=7	K=10	K=2	K=3	K=5	K=7	K=10
İki tanı testi	N=15	35.0	4.0	4.2	3.4	3.7	11.1	3.3	2.7	3.9	2.8
	N=30	50.6	3.5	3.2	2.6	2.9	16.5	2.9	2.4	3.2	3.6
	N=100	52.4	3.3	4.1	3.0	2.8	18.1	4.6	2.9	4.3	2.3
Beş tanı testi	N=15	55.0	3.2	3.5	2.2	3.0	8.2	3.4	1.9	2.0	2.6
	N=30	56.1	3.3	2.8	2.5	3.9	5.0	2.4	3.4	2.4	2.2
	N=100	55.9	3.5	3.6	3.3	2.7	3.8	2.5	2.2	2.7	2.6
Yedi tanı testi	N=15	51.2	2.9	2.7	3.0	2.2	7.9	2.3	2.3	1.5	2.3
	N=30	55.4	2.7	3.4	2.8	3.0	2.8	2.9	2.8	3.4	2.4
	N=100	59.5	2.6	3.7	3.1	2.6	4.4	1.9	2.7	3.6	2.7

Çizelge 4.1 incelendiğinde İki değerlendirici arasında çok güçlü bir uyum varken tanı testi sayısı ve örnek genişliğinden etkilenmeksizin tanı testinin kategorisi 3,5,7 ve 10 olduğu durumda tanı testlerinin diagnostik doğrulukları arasında gerçekte farklılık yokken var olarak bulunma ihtimalini gösteren Tip I Hata yapma olasılığı %5'in altındadır. Fakat tanı testi binary yapıda olduğu durumda 2 tanı testinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 35.0, 30 için % 50.6 ve 100 için % 52.4'dür. Beş tanı testinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 55.0, 30 için % 56.1 ve 100 için % 55.9 dur. Yedi tanı testinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 51.2, 30 için % 55.4 ve 100 için % 59.5'dir

Buna karşın iki değerlendirici arasında uyum yokken yine kategori sayısı 3,5,7 ve 10 olduğu durumda Tip I Hata yapma olasılığı örnek genişliği ve tanı testi sayısından etkilenmeksizin % 5'in altındadır. Fakat tanı testi binary yapıda olduğu durumda 2 tanı testinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 11.1, 30 için % 16.5 ve 100 için % 18.1'dür. Beş tanı testinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 8.2, 30 için % 5.0 ve 100 için % 3.8 dir. Yedi tanı testinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 7.9, 30 için % 2.8 ve 100 için % 4.4'dir.

Çizelge 4.2: Beş değerlendirici arasında uyum çok yüksek ve düşükken, tanı testlerinin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I Hata yapma olasılıkları

Tanı Testi Sayısı	Örnek Genişliği	Beş Değerlendirici Arasında Uyum Çok Yüksek					Beş Değerlendirici Arasında Uyum Çok Düşük				
		Tanı Testinin Kategori Sayısı					Tanı Testinin Kategori Sayısı				
		K=2	K=3	K=5	K=7	K=10	K=2	K=3	K=5	K=7	K=10
İki tanı testi	N=15	20.6	2.3	2.3	3.4	3.7	2.4	3.6	3.6	3.3	2.3
	N=30	30.4	1.9	0.9	3.4	2.8	1.0	2.8	2.3	3.1	2.3
	N=100	29.1	3.8	0.2	0.2	2.9	0.0	3.3	4.5	2.9	3.8
Beş tanı testi	N=15	18.6	4.0	2.5	2.2	2.3	0.0	2.7	2.0	2.1	2.5
	N=30	10.8	3.2	3.3	3.0	2.6	0.0	3.7	2.1	2.3	3.1
	N=100	-	-	-	-	-	-	-	-	-	-
Yedi tanı testi	N=15	20.4	2.5	2.9	2.6	3.5	0.0	2.7	2.2	2.4	1.9
	N=30	13.4	2.6	3.1	3.0	2.8	0.0	2.7	1.7	2.6	1.8
	N=100	-	-	-	-	-	-	-	-	-	-

-Paket programın kapasitesi yeterli gelmediği için hesaplanamamaktadır.

Çizelge 4.2 incelendiğinde beş değerlendirici arasında çok güçlü bir uyum varken tanı testi sayısı ve örnek genişliğinden etkilenmeksizin tanı testinin kategorisi 3,5,7 ve 10 olduğu durumda Tip I Hata yapma olasılığı %5'in altındadır. Fakat tanı testlerinin binary yapıda olduğu durumda 2 tanı testinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 20.6, 30 için % 30.4 ve 100 için % 29.1'dir. Beş tanı testinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 18.6, 30 için % 10.8 dir. Yedi tanı testinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 20.4, 30 için % 13.4'dür.

Buna karşın beş değerlendirici arasında uyum yokken tanı testi sayısı, kategorisi ve örnek genişliğinden etkilenmeksizin Tip I Hata yapma olasılığı % 5'in altında yer almaktadır.

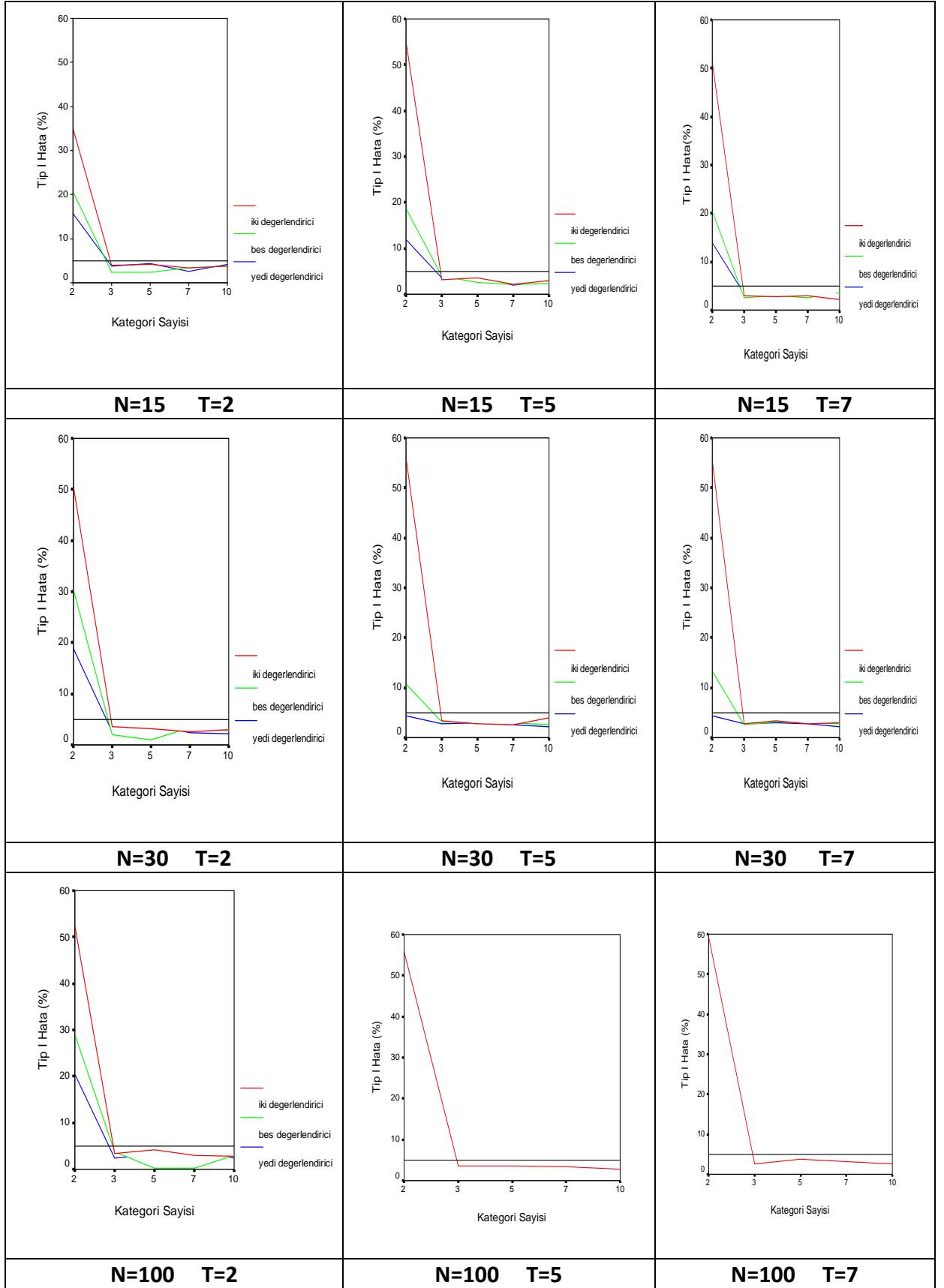
Çizelge 4.3: Yedi değerlendirici arasında uyum çok yüksek ve düşükken, tanı testlerinin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I Hata yapma olasılıkları

Tanı Testi Sayısı	Örnek Genişliği	Yedi Değerlendirici Arasında Uyum Çok Yüksek					Yedi Değerlendirici Arasında Uyum Çok Düşük				
		Tanı Testinin Kategori Sayısı					Tanı Testinin Kategori Sayısı				
		K=2	K=3	K=5	K=7	K=10	K=2	K=3	K=5	K=7	K=10
İki tanı testi	N=15	15.6	3.8	4.3	2.6	4.1	0.7	2.1	2.6	2.6	2.7
	N=30	18.9	2.5	4.2	2.4	2.1	0.2	2.9	2.2	2.5	3.2
	N=100	20.4	2.3	3.2	3.9	2.4	0.0	2.8	2.4	1.8	2.0
Beş tanı testi	N=15	12.0	3.5	3.5	2.0	3.0	0.0	2.9	2.0	2.3	2.2
	N=30	4.3	2.7	3.0	2.5	2.2	0.0	2.8	3.6	2.1	2.0
	N=100	-	-	-	-	-	-	-	-	-	-
Yedi tanı testi	N=15	14.0	3.4	3.2	3.5	2.4	0.0	3.1	2.7	2.7	2.8
	N=30	4.4	2.7	3.0	2.7	2.1	0.0	3.5	2.0	3.1	1.6
	N=100	-	-	-	-	-	-	-	-	-	-

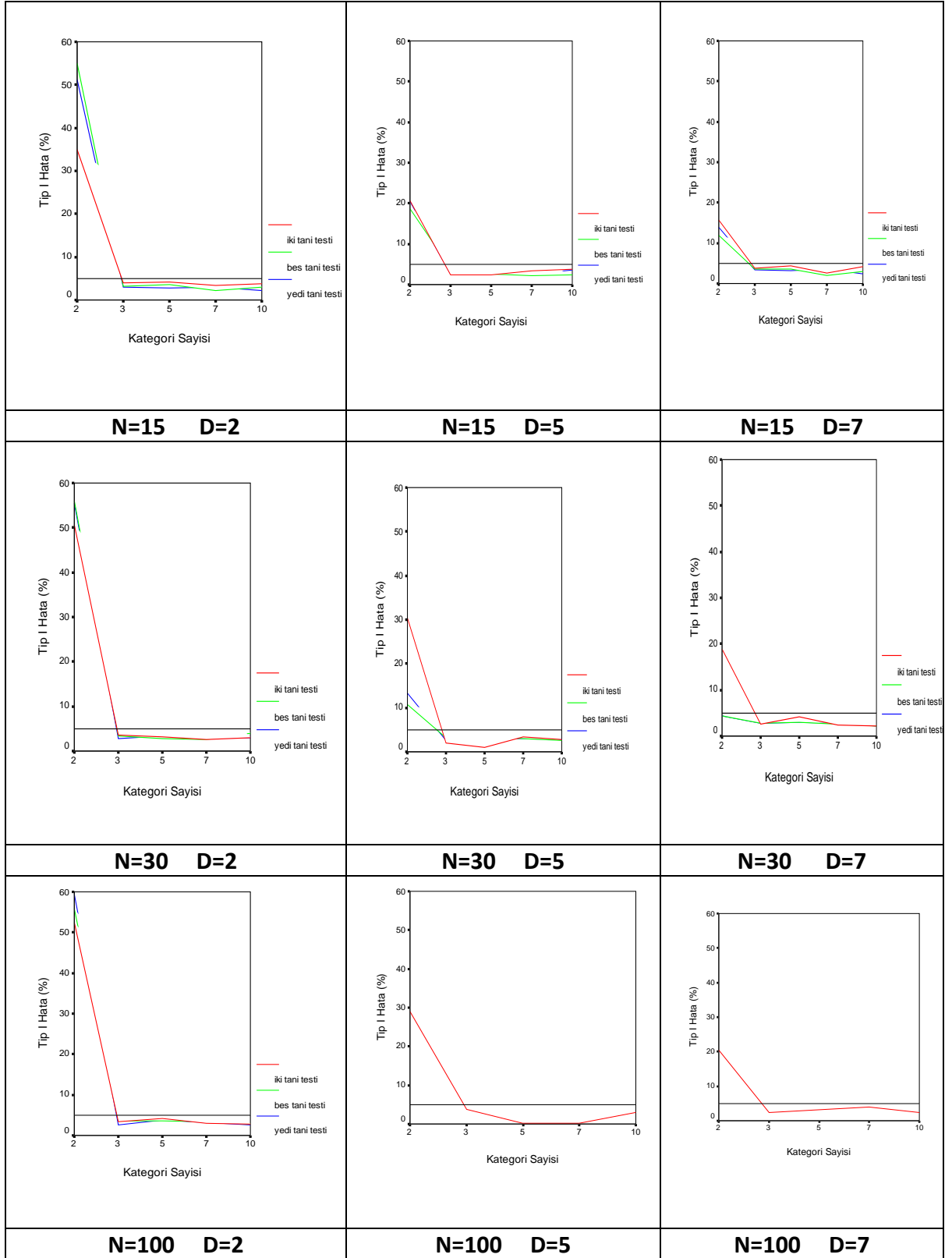
-Paket programın kapasitesi yeterli gelmediği için hesaplanamamaktadır.

Çizelge 4.3 incelendiğinde yedi değerlendirici arasında çok güçlü bir uyum varken tanı testi sayısı ve örnek genişliğinden etkilenmeksizin tanı testinin kategorisi 3,5,7 ve 10 olduğu durumda Tip I Hata yapma olasılığı %5'in altındadır. Fakat tanı testi binary yapıda olduğu durumda 2 tanı testinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için %15.6, 30 için %18.9 ve 100 için %20.4'dir. Beş tanı testinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 12.0 ve örnek genişliği 30 için % 4.3 dir. Yedi tanı testinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 14.0, ve örnek genişliği 30 için % 4.4'dür.

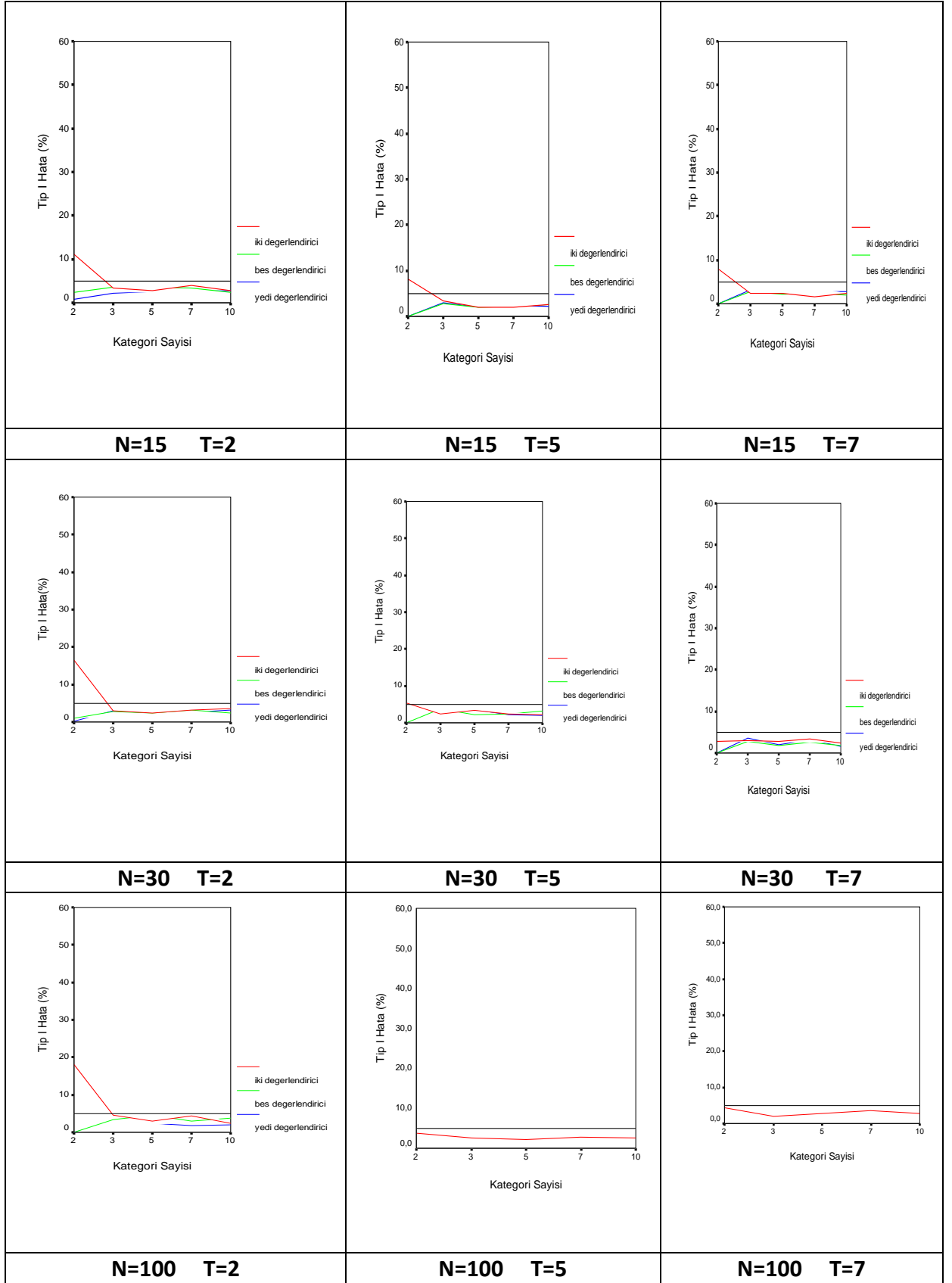
Buna karşın yedi değerlendirici arasında uyum yokken tanı testi sayısı, kategorisi ve örnek genişliğinden etkilenmeksizin Tip I Hata yapma olasılığı 5'in altında yer almaktadır.



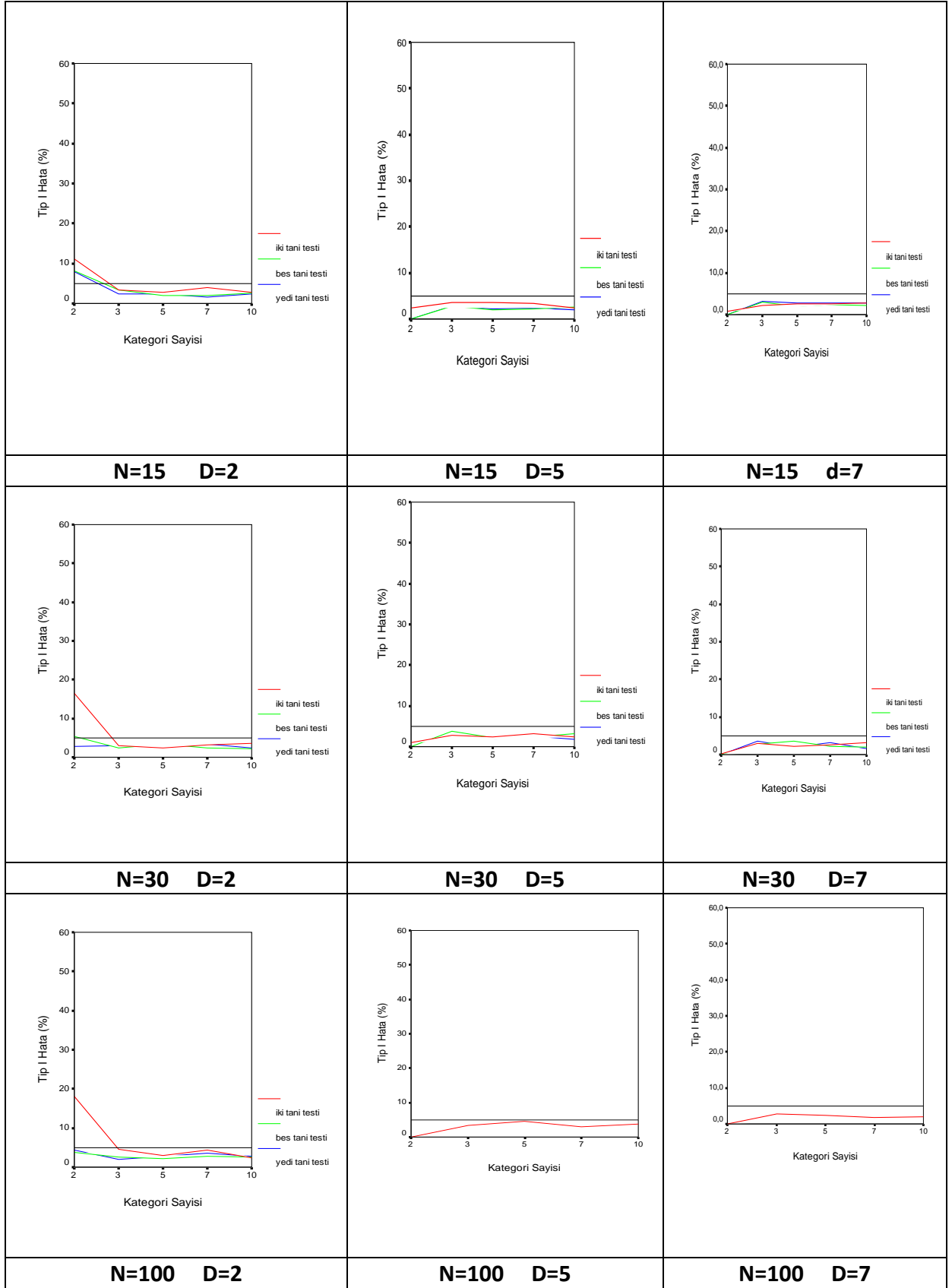
Şekil 4.1: Değerlendiriciler arasında uyum çok yüksekken, 2,5,7 tanı testi için, tanı testlerinin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I Hata yapma olasılıkları için gösterim grafiği



Şekil 4.2: Değerlendiriciler arasında uyum çok yüksekken, 2-5-7 değerlendirici için, tanı testlerinin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I Hata yapma olasılıkları için gösterim grafiği



Şekil 4.3: Değerlendiriciler arasında uyum çok düşükken, 2,5,7 tanı testi için, tanı testlerinin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I Hata yapma olasılıkları için gösterim grafiği



Şekil 4.4: Değerlendiriciler arasında uyum çok düşükken, 2,5,7 değerlendirici için, tanı testlerinin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I Hata yapma olasılıkları için gösterim grafiği

4.1.2. Deęerlendiriciler Arasındaki Yüksek ve Düşük Uyumun Etkisinin Deęerlendiricilerin Doğrulukları Üzerine Etkisi

Deęerlendirici sayısı, tanı testi sayısı ve kategorisi, deęerlendiriciler arasındaki çok yüksek ve çok düşük uyum durumları için üretilmiş olan kombinasyonlardaki deęerlendiricilerin diagnostik doğrulukları arasında gerçekte farklılık yokken farklılık var olarak bulunma ihtimalini gösteren Tip I Hata yapma olasılıkları iki deęerlendirici için Çizelge 4.4, beş deęerlendirici için Çizelge 4.5 ve yedi deęerlendirici için Çizelge 4.6'da verilmiştir. Grafik gösterimleri ise deęerlendiriciler ve tanı testi sayısı ve kategorisine karşılık, her bir örnek genişliği için Şekil 4.5-Şekil 4.8'de gösterilmektedir.

Çizelge 4.4: İki değerlendirici arasında çok yüksek ve düşük uyum varken, değerlendiricilerin diagnostik doğruluk farklılıklarına ilişkin olabilecek Tip I Hata yapma olasılıkları

Tanı Testi Sayısı	Örnek Genişliği	İki Değerlendirici Arasında Uyum Çok Yüksek					İki Değerlendirici Arasında Uyum Çok Düşük				
		Tanı Testinin Kategori Sayısı					Tanı Testinin Kategori Sayısı				
		K=2	K=3	K=5	K=7	K=10	K=2	K=3	K=5	K=7	K=10
İki tanı testi	N=15	14.0	1.4	0.2	0.4	0.3	39.3	4.5	4.5	4.2	3.4
	N=30	16.8	4.2	4.1	2.8	2.7	56.7	4.2	3.6	3.3	4.2
	N=100	16.8	4.2	5.0	4.6	5.0	64.5	5.6	4.2	1.6	4.6
Beş tanı testi	N=15	16.8	4.2	4.7	3.8	1.7	43.8	4.6	5.0	4.1	4.0
	N=30	36.5	4.7	4.8	5.1	5.0	58.9	4.8	4.0	4.7	5.0
	N=100	39.6	4.7	5.3	4.6	5.2	81.4	5.4	5.8	3.2	5.2
Yedi tanı testi	N=15	19.6	2.7	2.9	2.1	2.5	39.6	3.8	3.9	3.7	4.4
	N=30	36.0	3.4	5.2	4.5	3.2	54.9	4.3	5.2	4.6	4.6
	N=100	37.3	4.3	4.7	5.2	4.8	81.2	5.6	4.3	4.9	5.0

Çizelge 4.4 incelendiğinde iki değerlendirici arasında çok güçlü bir uyum varken tanı testi sayısı ve örnek genişliğinden etkilenmeksizin tanı testinin kategorisi 3,5,7 ve 10 olduğu durumda değerlendiriciler diagnostik doğrulukları arasında gerçekte farklılık yokken var olarak bulunma ihtimalini gösteren Tip I Hata yapma olasılığı %5'in altındadır. Fakat tanı testinin binary yapıda olduğu durumda 2 tanı testinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 14.0, 30 için % 16.8 ve 100 için % 16.8'dir. Beş tanı testinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 16.8, 30 için % 36.5 ve 100 için % 39.6'dır. Yedi tanı testinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 19.6, 30 için % 36.0 ve 100 için % 37.3'dür.

Buna karşın iki değerlendirici arasında uyum yokken yine kategori sayısı 3,5,7 ve 10 olduğu durumda Tip I Hata yapma olasılığı örnek genişliği ve tanı testi sayısından etkilenmeksizin % 5'in altındadır. Fakat tanı testi binary yapıda olduğu durumda 2 tanı testinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 39.3, 30 için % 56.7 ve 100 için % 64.5'dir. Beş tanı testinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 43.8, 30 için % 58.9 ve 100 için % 81.4' dür. Yedi tanı testinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 39.6, 30 için % 54.9 ve 100 için % 81.2'dir.

Çizelge 4.5: Beş değerlendirici arasında çok yüksek ve düşük uyum varken, değerlendiricilerin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I Hata yapma olasılıkları

Tanı Testi Sayısı	Örnek Genişliği	Beş Değerlendirici Arasında Uyum Çok Yüksek					Beş Değerlendirici Arasında Uyum Çok Düşük				
		Tanı Testinin Kategori Sayısı					Tanı Testinin Kategori Sayısı				
		K=2	K=3	K=5	K=7	K=10	K=2	K=3	K=5	K=7	K=10
İki tanı testi	N=15	35.1	2.4	1.8	2.0	1.6	75.7	4.2	3.4	2.7	2.8
	N=30	50.9	4.3	3.7	4.2	3.7	91.3	4.8	6.0	4.8	5.0
	N=100	52.9	5.1	5.2	4.6	5.9	100	4.6	6.1	4.9	5.1
Beş tanı testi	N=15	52.2	4.2	3.4	3.8	3.7	79.6	5.4	3.7	3.2	4.6
	N=30	83.4	6.1	4.4	4.7	3.8	93.0	5.3	4.9	3.2	4.9
	N=100	-	-	-	-	-	-	-	-	-	-
Yedi tanı testi	N=15	51.2	4.1	2.2	4.4	2.5	78.7	4.3	3.8	3.6	2.2
	N=30	79.1	5.8	5.4	5.0	4.8	92.9	5.6	4.2	4.0	4.9
	N=100	-	-	-	-	-	-	-	-	-	-

-Paket programın kapasitesi yeterli gelmediği için hesaplanamamaktadır.

Çizelge 4.5 incelendiğinde beş değerlendirici arasında çok güçlü bir uyum varken tanı testi sayısı ve örnek genişliğinden etkilenmeksizin tanı testinin kategorisi 3,5,7 ve 10 olduğu durumda değerlendiriciler arasında gerçekte farklılık yokken var olarak bulunma ihtimalini gösteren Tip I Hata yapma olasılığı % 6'nın altındadır. Fakat tanı testinin binary yapıda olduğu durumda 2 tanı testinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 35.1, 30 için % 50.9 ve 100 için % 52.9'dur. Beş tanı testinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 52.2, 30 için % 83.4'dür. Yedi tanı testinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 51.2, 30 için % 79.1'dir.

Buna karşın beş değerlendirici arasında uyum yokken yine kategori sayısı 3,5,7 ve 10 olduğu durumda Tip I Hata yapma olasılığı örnek genişliği ve tanı testi sayısından etkilenmeksizin % 6'nın altındadır. Fakat tanı testi binary yapıda olduğu durumda 2 tanı testinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 75.7, 30 için % 91.3 ve 100 için % 100'dür. Beş tanı testinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 79.6, 30 için % 93.0'dür. Yedi tanı testinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 78.7, 30 için % 92.9'dur.

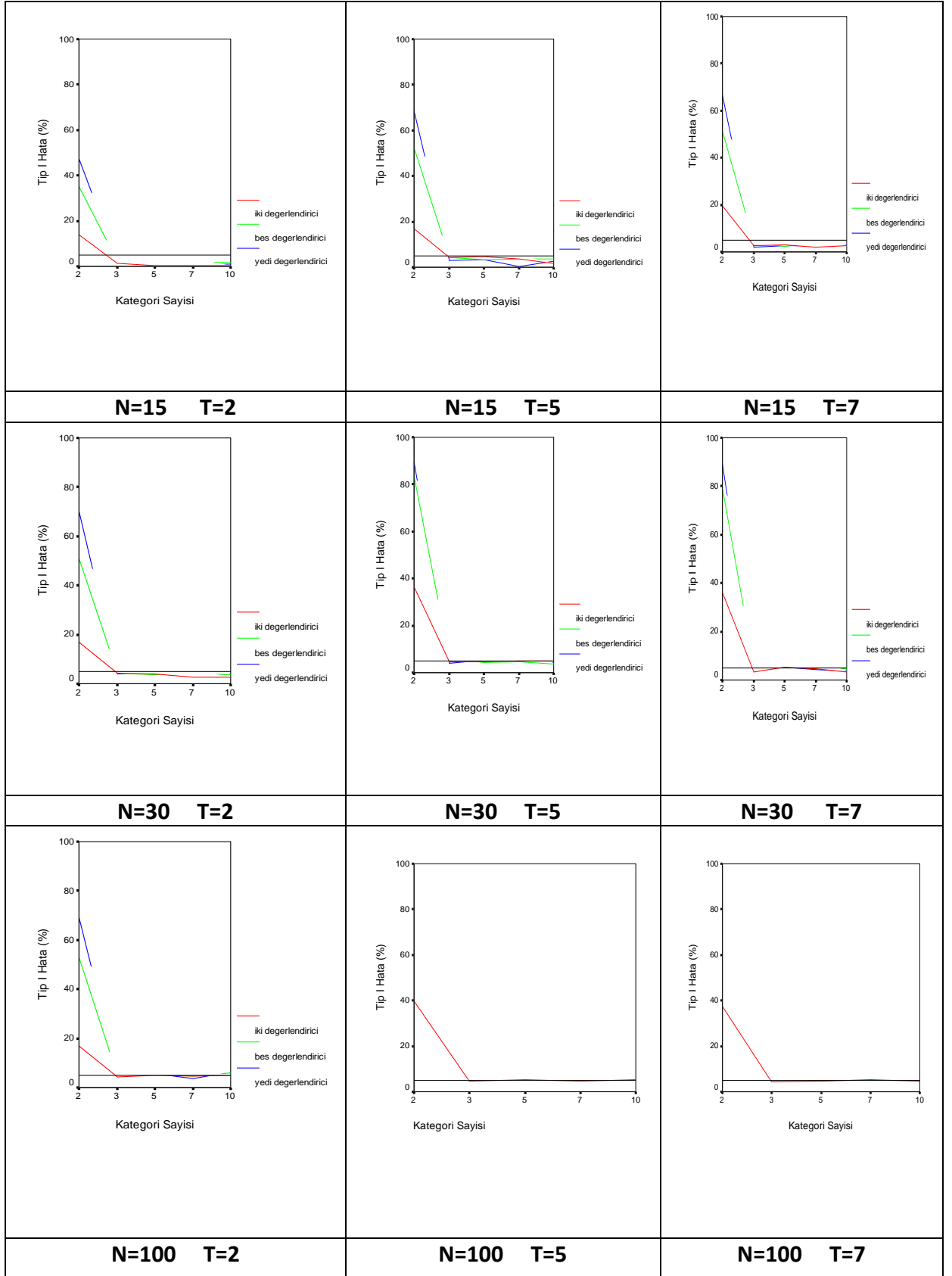
Çizelge 4.6: Yedi değerlendirici arasında çok yüksek ve düşük uyum varken, değerlendiricilerin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I Hata yapma olasılıkları

Tanı Testi Sayısı	Örnek Genişliği	Yedi Değerlendirici Arasında Uyum Çok Yüksek					Yedi Değerlendirici Arasında Uyum Çok Düşük				
		Tanı Testinin Kategori Sayısı					Tanı Testinin Kategori Sayısı				
		K=2	K=3	K=5	K=7	K=10	K=2	K=3	K=5	K=7	K=10
İki tanı testi	N=15	47.3	2.3	1.7	0.8	1.1	87.5	3.7	3.8	2.7	3.6
	N=30	70.2	4.0	4.9	3.8	4.0	96.9	4.6	3.6	4.2	4.9
	N=100	69.2	6.5	5.7	3.7	6.3	100	4.3	5.8	5.4	4.7
Beş tanı testi	N=15	68.6	3.0	3.2	0.4	2.7	89.5	5.5	3.1	2.2	2.2
	N=30	89.4	4.1	5.3	4.7	4.3	98.5	3.9	4.6	4.7	4.6
	N=100	-	-	-	-	-	-	-	-	-	-
Yedi tanı testi	N=15	66.4	2.1	2.5	2.6	3.1	88.0	2.9	2.6	3.8	2.1
	N=30	89.2	4.4	5.4	4.2	3.5	98.0	5.0	4.7	3.2	3.0
	N=100	-	-	-	-	-	-	-	-	-	-

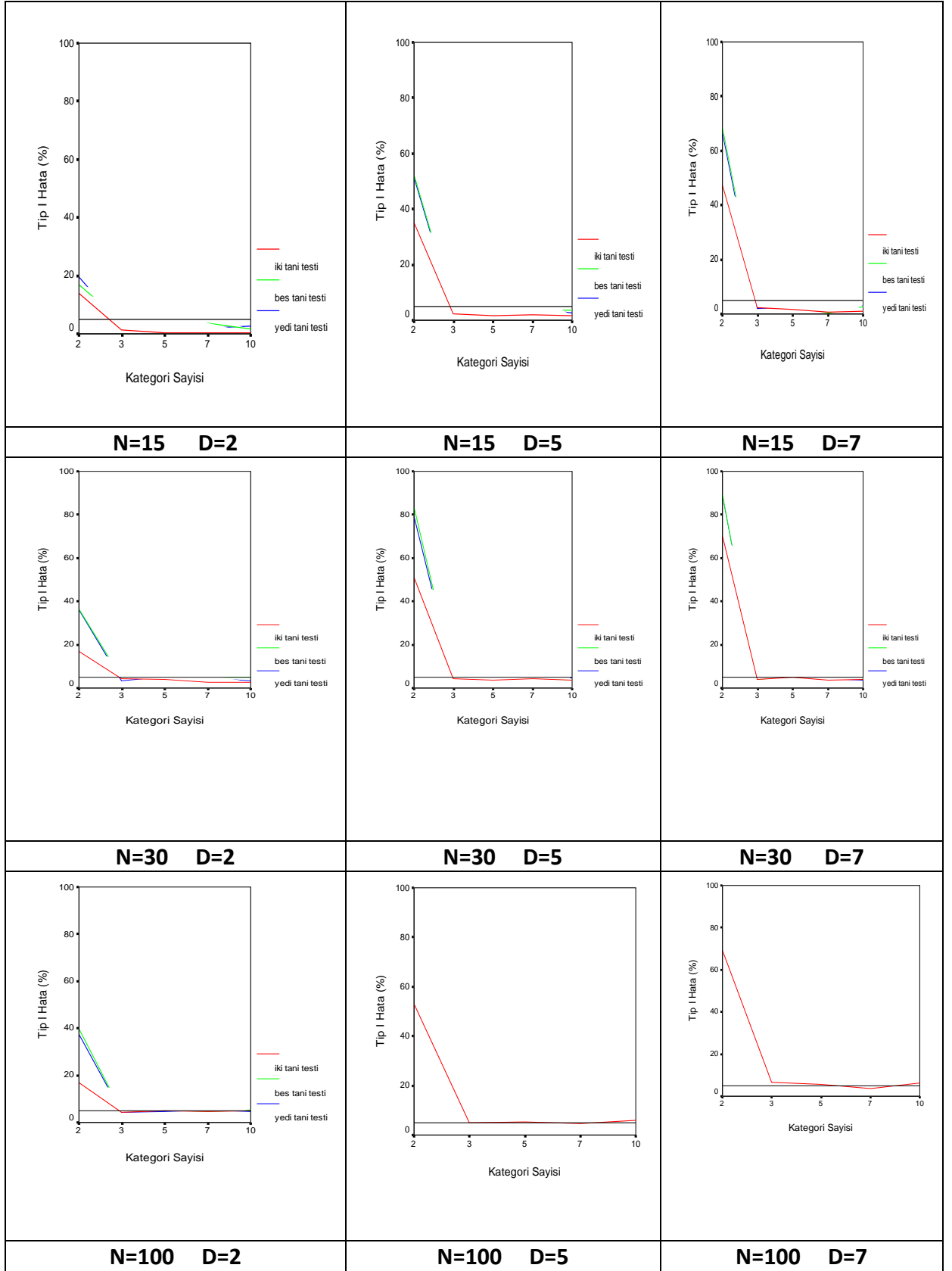
-Paket programın kapasitesi yeterli gelmediği için hesaplanamamaktadır.

Çizelge 4.6 incelendiğinde yedi değerlendirici arasında çok güçlü bir uyum varken tanı testi sayısı ve örnek genişliğinden etkilenmeksizin tanı testinin kategorisi 3,5,7 ve 10 olduğu durumda değerlendiriciler arasında gerçekte farklılık yokken var olarak bulunma ihtimalini gösteren Tip I Hata yapma olasılığı % 6'nın altındadır. Fakat tanı testi binary yapıda olduğu durumda 2 tanı testinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 47.3, 30 için % 70.2 ve 100 için % 69.2'dir. Beş tanı testinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 68.6, 30 için % 89.4'dür. Yedi tanı testinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 66.4, 30 için % 89.2'dir.

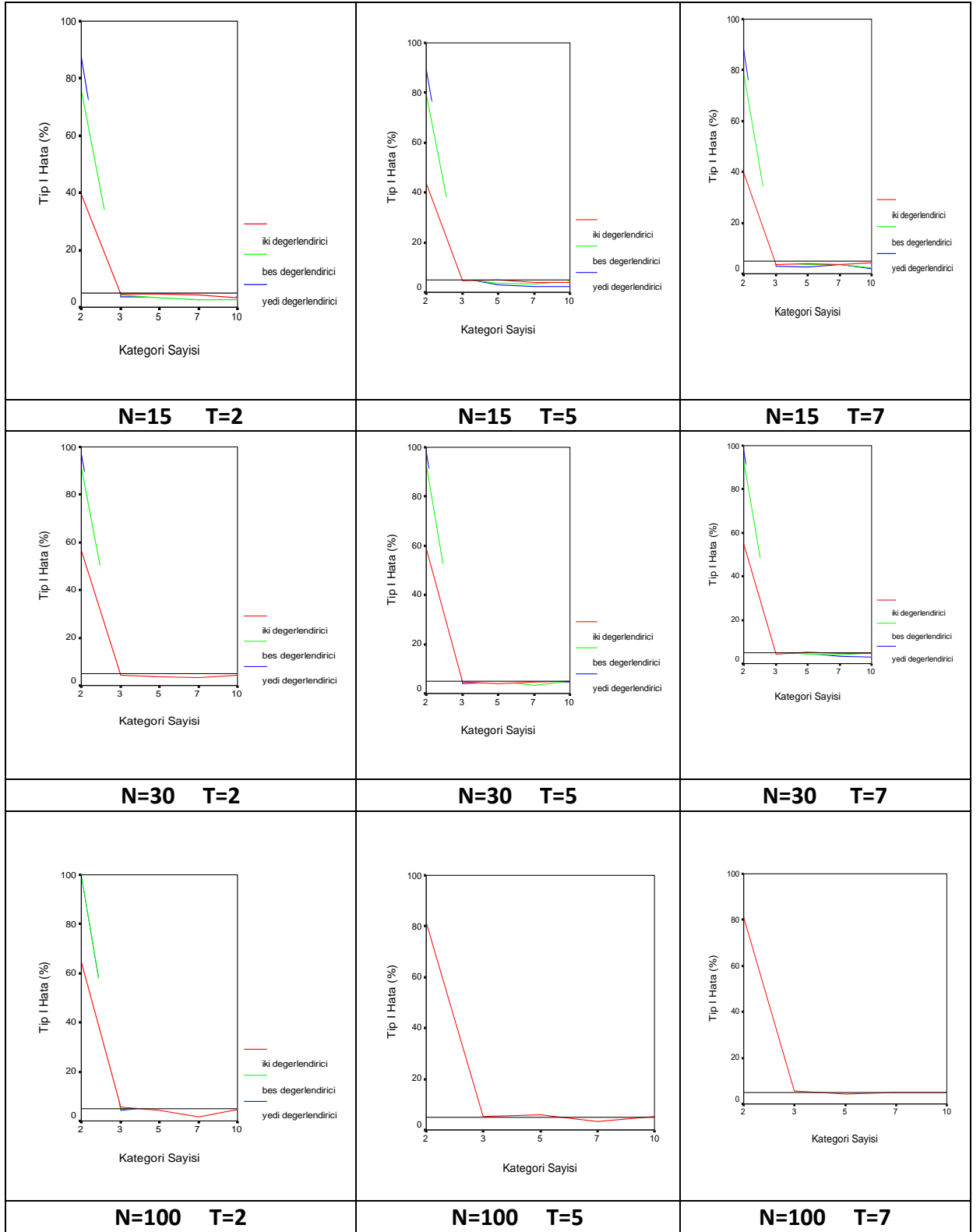
Buna karşın yedi değerlendirici arasında uyum yokken yine kategori sayısı 3,5,7 ve 10 olduğu durumda Tip I Hata yapma olasılığı örnek genişliği ve tanı testi sayısından etkilenmeksizin % 5'in altındadır. Fakat tanı testi binary yapıda olduğu durumda 2 tanı testinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 87.5, 30 için % 96.9 ve 100 için % 100'dür. Beş tanı testinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 89.5, 30 için % 98.5' dir. Yedi tanı testinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 88.0, 30 için % 98.0'dir.



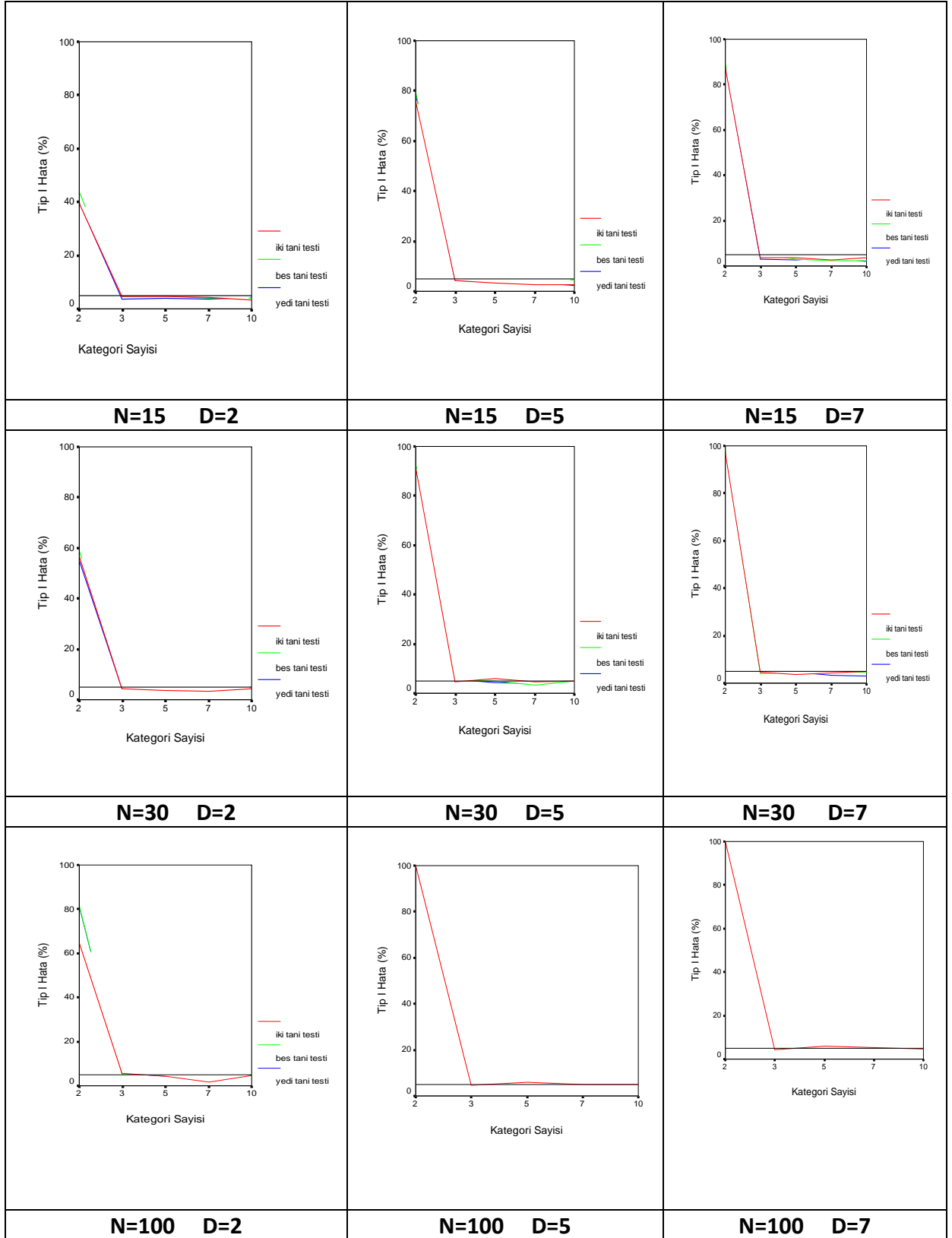
Şekil 4.5: Değerlendiriciler arasında uyum çok yüksekken, 2-5-7 tanı testi için değerlendiricilerin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I Hata yapma olasılıklarının gösterim grafiği



Şekil 4.6: Değerlendiriciler arasında uyum çok yüksekken, 2,5,7 değerlendirici için değerlendiricilerin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I Hata yapma olasılıkları gösterim grafiği



Şekil 4.7: Değerlendiriciler arasında uyum çok düşükken, 2,5,7 tanı testii için değerlendiricilerin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I Hata yapma olasılıkları gösterim grafiği



Şekil 4.8: Değerlendiriciler arasında uyum çok düşükken, 2,5,7 değerlendirici için değerlendiricilerin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I Hata yapma olasılıkları gösterim grafiği

4.2.Tanı Testleri Arasındaki Güçlü ve Zayıf Korelasyonun Diagnostik Doğruluk Üzerine Etkisi

Bu simülasyon denemelerinde amaç tanı testleri arasındaki korelasyonun diagnostik doğrulukları ne derece etkilediklerini ortaya koymaktır. Her bir kombinasyon için üretilen verilerde gerçekte tanı testlerinin ortalama doğrulukları bakımından fark yoktur ve yine bu farksızlık bütün değerlendiriciler için aynıdır. Yani değerlendiricilerin ve tanı testlerinin doğrulukları arasında yine farklılık yoktur.

Ek V’de ise tanı testleri arasında korelasyon çok güçlü olduğunda; değerlendirici sayısı, tanı testi sayısı ve kategorisi ve farklı örnek büyüklükleri için üretilmiş olan kombinasyonlardaki varyans analiz tablosunun bütününde olabilecek tüm hata kaynakları için Tip I Hata yapma olasılıkları verilmiştir.

Ek VI’da ise tanı testleri arasında korelasyon çok zayıf olduğunda; değerlendirici sayısı, tanı testi sayısı ve kategorisi ve farklı örnek büyüklükleri için üretilmiş olan kombinasyonlardaki varyans analiz tablosunun bütününde olabilecek tüm hata kaynakları için Tip I Hata yapma olasılıkları verilmiştir.

4.2.1. Tanı Testleri Arasındaki Güçlü ve Zayıf Korelasyonun Etkisinin Tanı Testlerinin Doğrulukları Üzerine Etkisi

Değerlendirici sayısı, tanı testi sayısı ve kategorisi, tanı testleri arasında korelasyonun çok yüksek ve çok zayıf olma ihtimalleri için üretilmiş olan kombinasyonlardaki tanı testlerinin diagnostik doğrulukları arasında gerçekte farklılık yokken var olarak bulunma ihtimalini gösteren Tip I Hata yapma olasılıkları iki tanı testi için Çizelge 4.7, beş tanı testi için Çizelge 4.8 ve yedi tanı testi için Çizelge 4.9’da verilmiştir. Grafik gösterimleri ise değerlendiriciler ve tanı testi sayısı ve kategorisine karşılık, her bir örnek genişliği için Şekil 4.9-Şekil 4.12’de gösterilmektedir.

Çizelge 4.7: İki tanı testi arasındaki korelasyon çok güçlü ve zayıf olduğunda tanı testlerinin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I Hata yapma olasılıkları

Değerlendirici Sayısı	Örnek Genişliği	İki Tanı Testi Arasındaki Korelasyon Çok Güçlü					İki Tanı Test Arasındaki Korelasyon Çok Zayıf				
		Tanı Testinin Kategori Sayısı					Tanı Testinin Kategori Sayısı				
		K=2	K=3	K=5	K=7	K=10	K=2	K=3	K=5	K=7	K=10
İki değerlendirici	N=15	0.2	0.0	0.0	0.0	0.0	0.0	2.2	1.7	2.2	1.4
	N=30	0.0	0.9	0.5	1.5	0.6	0.0	2.1	2.7	2.3	2.6
	N=100	0.0	1.3	1.9	2.2	2.1	0.1	3.3	3.7	3.0	3.4
Beş değerlendirici	N=15	0.0	0.0	0.0	0.0	0.0	0.0	1.6	2.0	1.4	0.2
	N=30	0.0	0.8	0.5	0.6	0.6	0.0	2.2	1.4	1.5	1.6
	N=100	0.0	1.9	1.7	1.7	1.4	0.0	2.7	2.8	1.9	2.2
Yedi değerlendirici	N=15	0.0	0.0	0.0	0.0	0.0	0.0	2.2	1.8	1.4	1.6
	N=30	0.0	0.6	0.4	1.0	0.4	0.0	2.8	2.0	2.8	2.8
	N=100	0.0	1.3	1.6	1.7	0.9	0.0	3.2	2.8	3.3	2.1

Çizelge 4.7 incelendiğine iki tanı testi arasında çok güçlü ve çok zayıf korelasyon varken değerlendirici sayısı ve örnek genişliği ve tanı testinin kategorisinden etkilenmeksizin tanı testlerinin diagnostik doğrulukları arasında gerçekte farklılık yokken var olarak bulunma ihtimalini gösteren Tip I Hata yapma olasılıkları daima %5'in altındadır.

Çizelge 4.8: Beş tanı testi arasındaki korelasyon çok güçlü ve zayıf olduğunda tanı testlerinin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I Hata yapma olasılıkları

Değerlendirici Sayısı	Örnek Genişliği	Beş Tanı Testi Arasında Korelasyon Çok Güçlü					Beş Tanı Test Arasında Korelasyon Çok Zayıf				
		Tanı Testinin Kategori Sayısı					Tanı Testinin Kategori Sayısı				
		K=2	K=3	K=5	K=7	K=10	K=2	K=3	K=5	K=7	K=10
İki değerlendirici	N=15	0.1	0.0	0.2	0.2	0	0.0	1.4	1.4	1.5	1.5
	N=30	0.0	1.3	1.1	1.4	0.9	0.0	2.8	2.3	2.3	2.1
	N=100	0.0	1.7	1.6	2.4	2.9	0.0	3.5	1.9	0.2	3.0
Beş değerlendirici	N=15	0.0	0.0	0.0	0.0	0.0	0.0	1.6	1.4	1.3	1.5
	N=30	0.0	1.1	0.7	0.4	0.8	0.0	1.4	1.9	2.4	1.6
	N=100	-	-	-	-	-	-	-	-	-	-
Yedi değerlendirici	N=15	0.0	1	0.0	0.0	0.0	0.0	1.4	1.7	1.9	1.6
	N=30	0.0	0.6	1.8	1.5	1.4	0.0	2.3	2.3	2	1.4
	N=100	-	-	-	-	-	-	-	-	-	-

-Paket programın kapasitesi yeterli gelmediği için hesaplanamamaktadır.

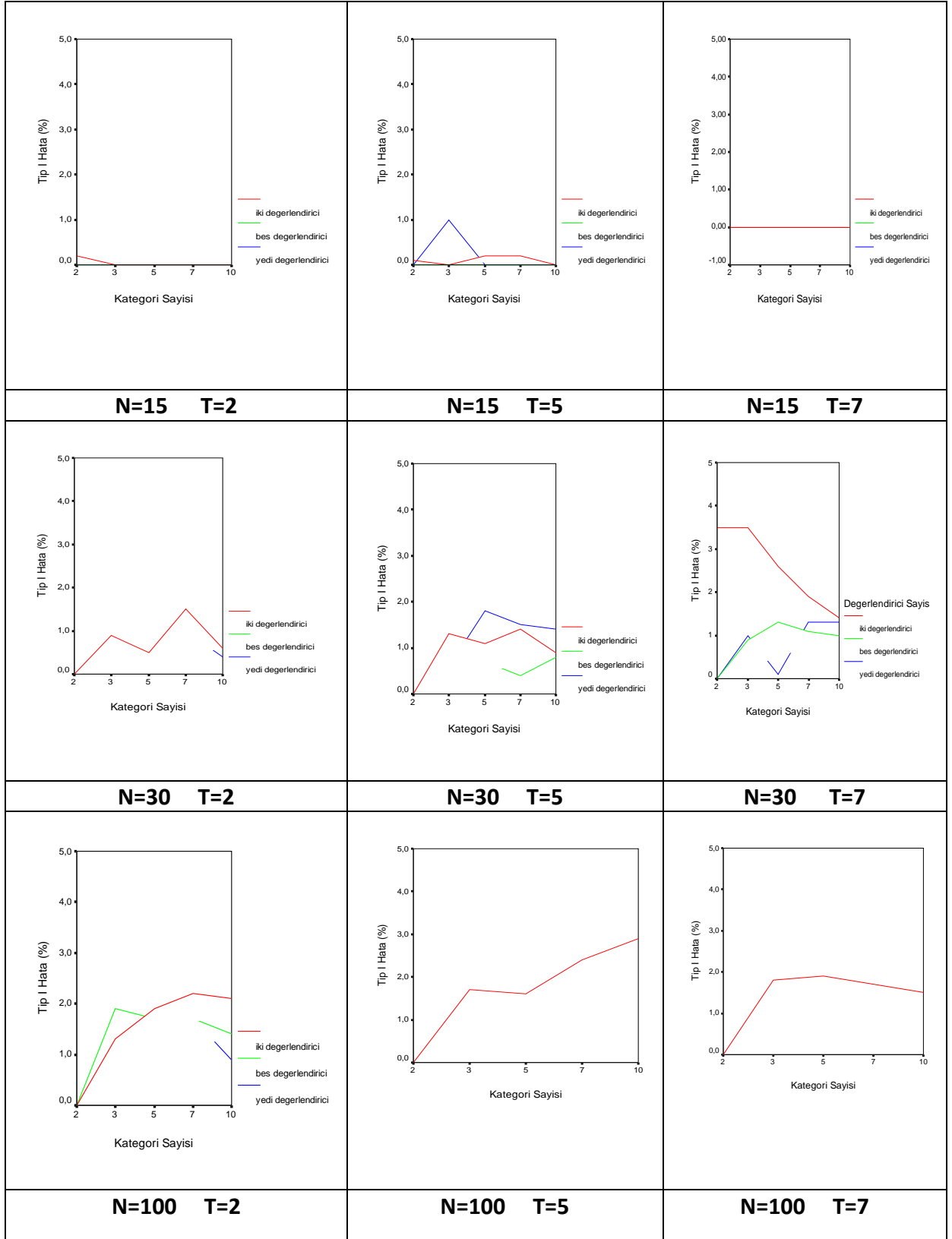
Çizelge 4.8 incelendiğinde beş tanı testi arasında çok güçlü ve çok zayıf korelasyon varken değerlendirici sayısı ve örnek genişliği ve tanı testinin kategorisinden etkilenmeksizin tanı testlerinin doğrulukları arasında gerçekte farklılık yokken var olarak bulunma ihtimalini gösteren Tip I Hata yapma olasılıkları daima % 5'in altındadır.

Çizelge 4.9: Yedi tanı testi arasındaki korelasyon çok güçlü ve zayıf olduğunda tanı testlerinin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I Hata yapma olasılıkları

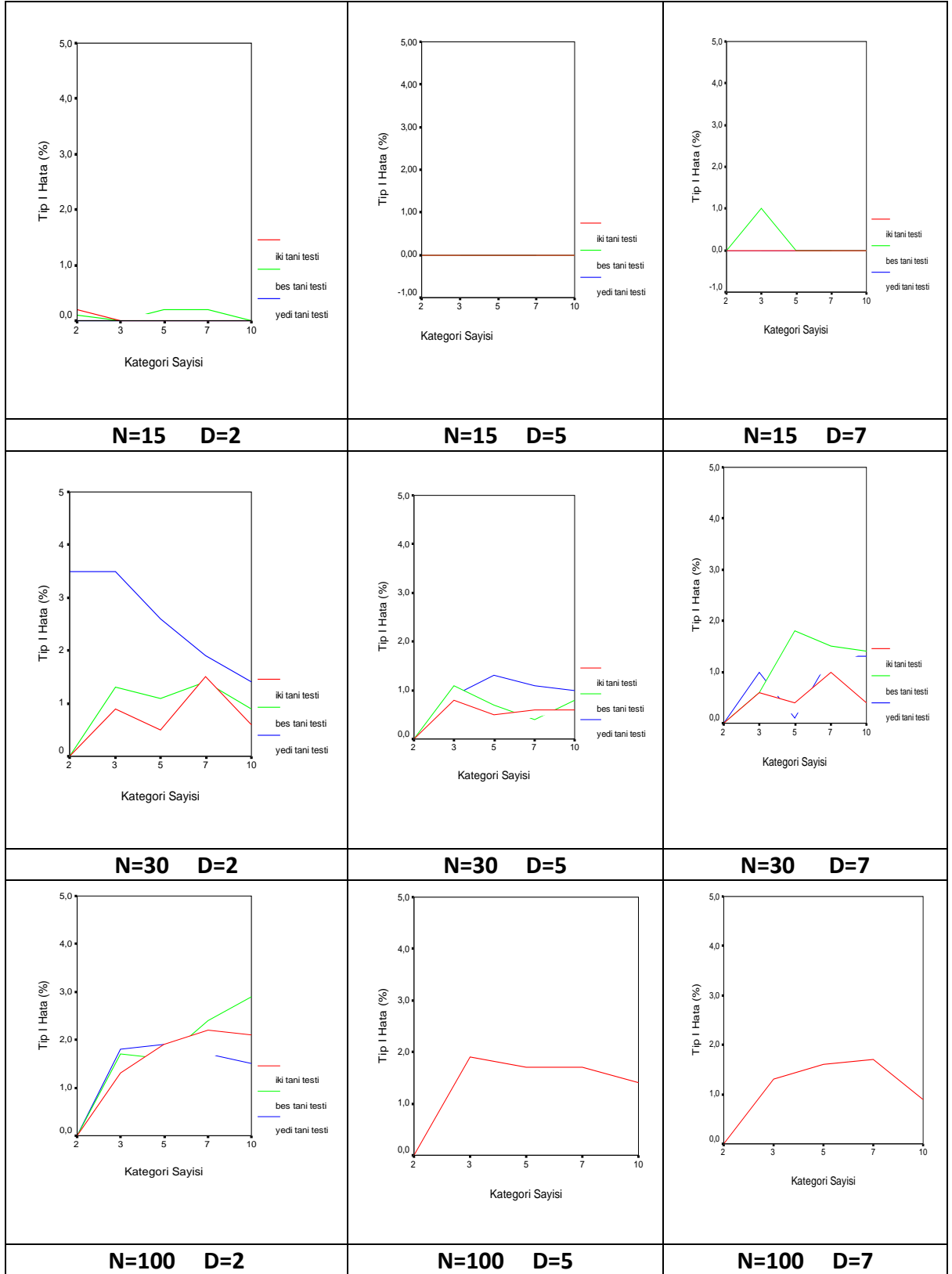
Değerlendirici Sayısı	Örnek Genişliği	Yedi Tanı Testi Arasında Korelasyon Çok Güçlü					Yedi Tanı Test Arasında Korelasyon Çok Zayıf				
		Tanı Testinin Kategori Sayısı					Tanı Testinin Kategori Sayısı				
		K=2	K=3	K=5	K=7	K=10	K=2	K=3	K=5	K=7	K=10
İki değerlendirici	N=15	0.0	0.0	0.0	0.0	0.0	0.0	1.4	1.3	1.3	1.2
	N=30	3.5	3.5	2.6	1.9	1.4	0.0	2.6	2.2	2.1	1.2
	N=100	0.0	1.8	1.9	1.7	1.5	0.0	3.4	2.8	2.4	2.1
Beş değerlendirici	N=15	0.0	0.0	0.0	0.0	0.0	0.0	1.8	1.4	1.1	1.2
	N=30	0.0	0.9	1.3	1.1	1.0	0.0	3.1	2.5	2.6	1.4
	N=100	-	-	-	-	-	-	-	-	-	-
Yedi değerlendirici	N=15	0.0	0.0	0.0	0.0	0.0	0.0	1.8	1.8	1.7	1.6
	N=30	0.0	1.0	0.1	1.3	1.3	0.0	2.3	3.4	1.9	2.0
	N=100	-	-	-	-	-	-	-	-	-	-

-Paket programın kapasitesi yeterli gelmediği için hesaplanamamaktadır.

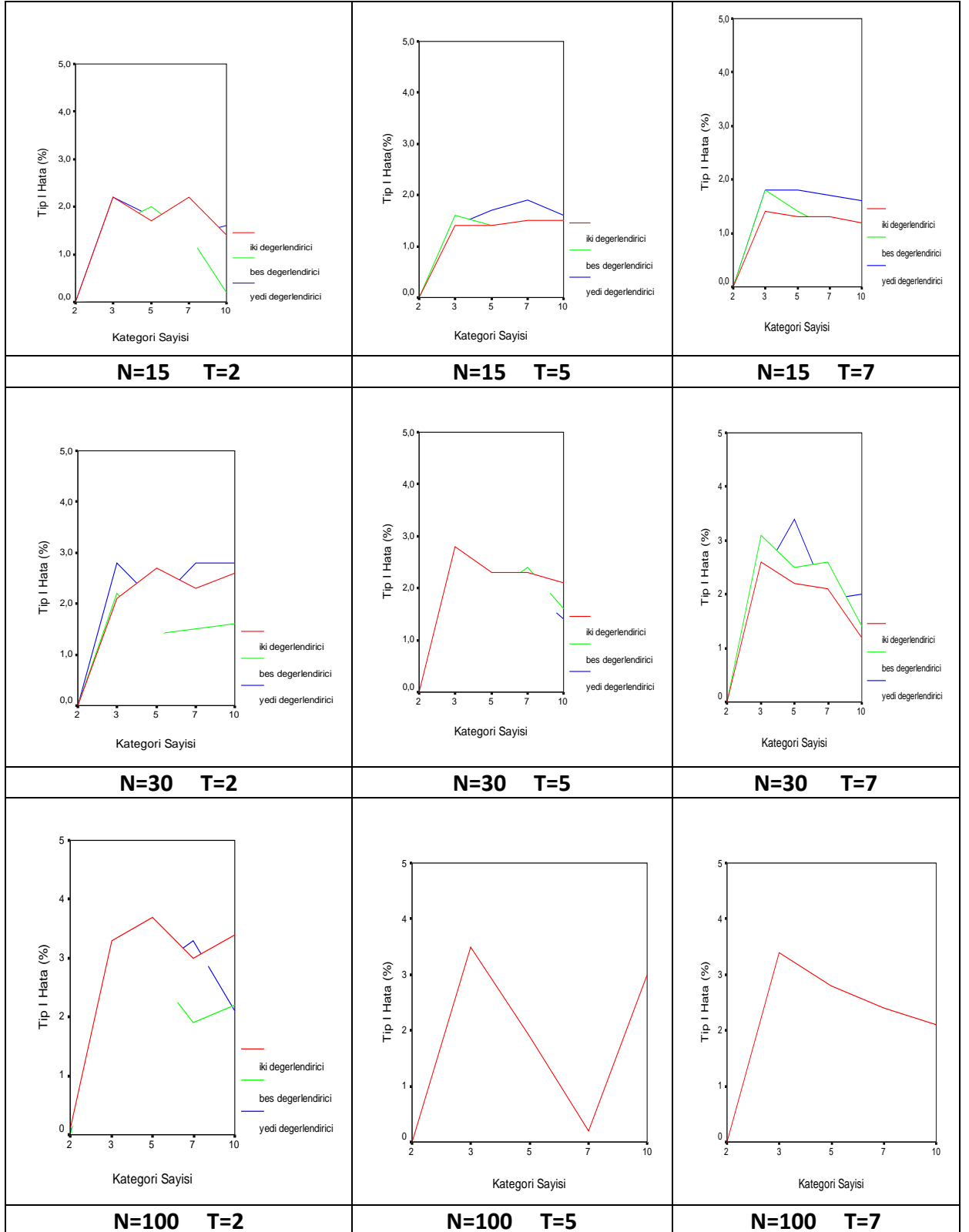
Çizelge 4.9 incelendiğinde yedi tanı testi arasında çok güçlü ve çok zayıf korelasyon varken değerlendirici sayısı ve örnek genişliği ve tanı testinin kategorisinden etkilenmeksizin tanı testlerinin doğrulukları arasında gerçekte farklılık yokken var olarak bulunma ihtimalini gösteren Tip I Hata yapma olasılıkları daima %5'in altındadır.



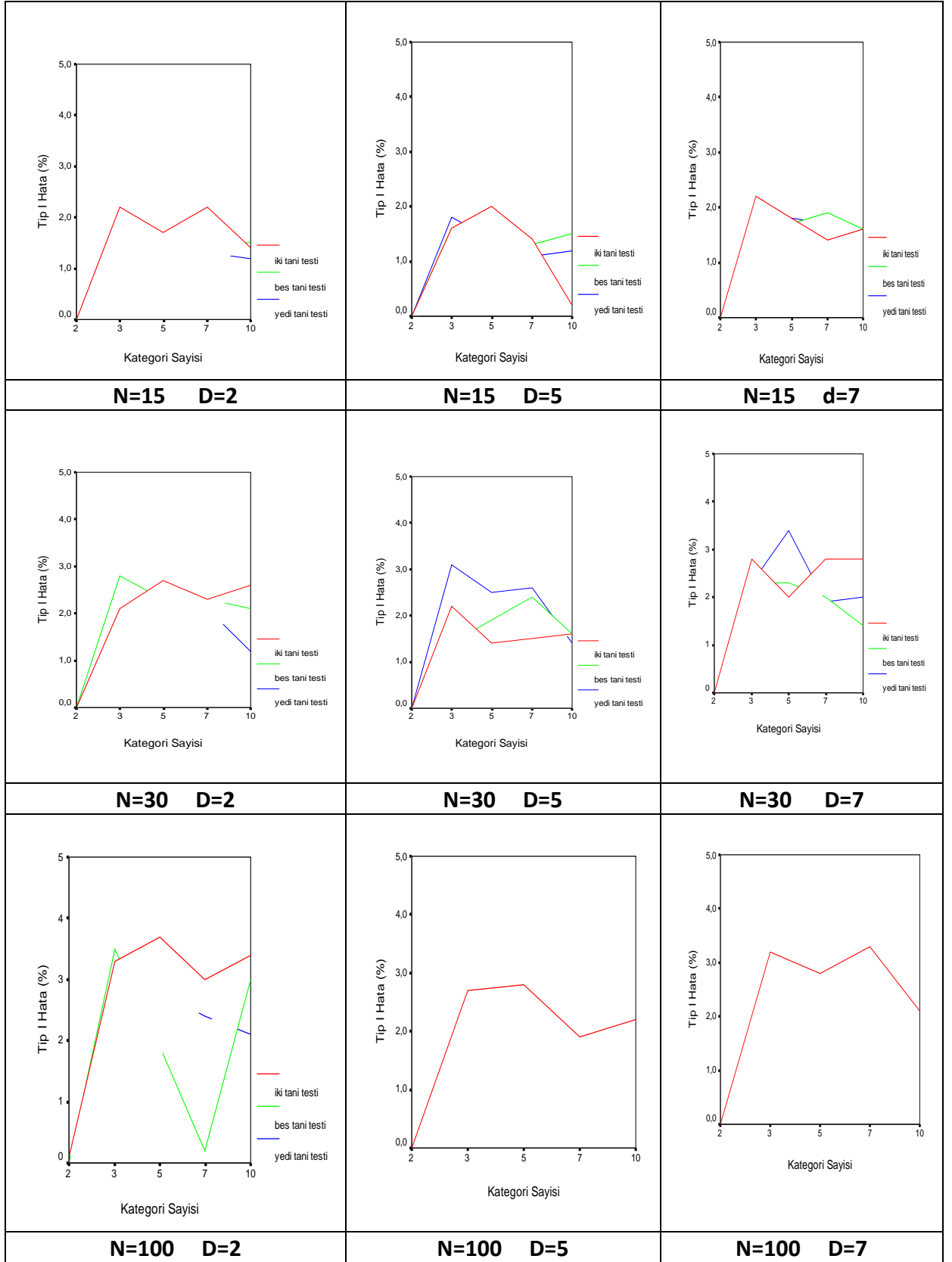
Şekil 4.9: Tanı testleri arasında korelasyon çok güçlü olduğunda, 2-5-7 tanı testi için, tanı testlerinin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I hata yapma olasılığı gösterim grafiği



Şekil 4.10: Tani testleri arasında korelasyon çok güçlü olduğunda, 2-5-7 değerlendirici için tani testlerinin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I hata yapma olasılığı gösterim grafiği



Şekil 4.11: Tanı testleri arasında korelasyon çok zayıf olduğunda, 2-5-7 tanı testi için, tanı testlerinin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I hata yapma olasılığı gösterim grafiği



Şekil 4.12: Tani testleri arasında korelasyon çok zayıf olduğunda, 2-5-7 değerlendirici için tani testlerinin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I hata yapma olasılığı gösterim grafiği

4.2.2. Tanı Testleri Arasındaki Güçlü ve Zayıf Korelasyonun Etkisinin Değerlendiricilerin Doğrulukları Üzerine Etkisi

Değerlendirici sayısı, tanı testi sayısı ve kategorisi, tanı testleri arasındaki korelasyonun çok yüksek ve çok zayıf olma ihtimalleri için üretilmiş olan kombinasyonlardaki değerlendiricilerin diagnostik doğrulukları arasında gerçekte farklılık yokken, farklılık var olarak bulunma ihtimalini gösteren Tip I Hata yapma olasılıkları iki tanı testi için Çizelge 4.10, beş tanı testi için Çizelge 4.11 ve yedi tanı testi için Çizelge 4.12’de verilmiştir. Grafik gösterimleri ise değerlendiriciler ve tanı testi sayısı ve kategorisine karşılık, her bir örnek genişliği için Şekil 4.13-Şekil 4.16’da gösterilmektedir.

Çizelge 4.10: İki tanı testi arasında korelasyon çok güçlü ve çok zayıf olduğunda, değerlendiricilerin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I Hata yapma olasılıkları

Değerlendirici Sayısı	Örnek Genişliği	İki Tanı Testi Arasında Korelasyon Çok Güçlü					İki Tanı Testi Arasında Korelasyon Çok Zayıf				
		Tanı Testinin Kategori Sayısı					Tanı Testinin Kategori Sayısı				
		K=2	K=3	K=5	K=7	K=10	K=2	K=3	K=5	K=7	K=10
İki değerlendirici	N=15	40.9	5.9	5.5	4.5	3.9	41.6	5.5	3.7	3.3	4.2
	N=30	60.8	4.8	3.3	4.6	3.5	60.4	4.8	3.5	3.1	3.3
	N=100	65.4	5.2	5.3	4.3	4.1	63.2	4.5	5.0	4.4	5.5
Beş değerlendirici	N=15	78.7	4.7	3.6	3.1	4.2	70.1	3.9	3.5	3.4	2.7
	N=30	89.6	4.5	3.7	5.4	3.1	89.9	5.2	3.3	3.2	3.1
	N=100	97.3	5.1	5.7	5.2	5.2	96.9	5.0	4.7	5.0	4.4
Yedi değerlendirici	N=15	86.0	4.6	3.7	2.9	3.5	86.6	4.7	3.0	2.6	3.1
	N=30	94.4	5.1	3.4	4.8	2.3	95.6	4.3	3.2	3.2	2.8
	N=100	99.5	5.3	4.3	4.9	4.9	99.2	4.8	4.3	4.7	5.2

-Paket programın kapasitesi yeterli gelmediği için hesaplanamamaktadır.

Çizelge 4.10 incelendiğinde iki tanı testi arasında korelasyon çok güçlü olduğunda değerlendirici sayısı ve örnek genişliğinden etkilenmeksizin tanı testinin kategorisi 3-5-7 ve 10 olduğu durumda Tip I Hata yapma olasılığı % 6'nın altındadır. Fakat tanı testlerinin kategorisi binary yapıda olduğu durumda 2 değerlendiricinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 40.9, 30 için % 60.8 ve 100 için % 65.4'dür. Yedi değerlendiricinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 78.7, 30 için % 89.6 ve 100 için % 97.3 dür. Yedi tanı testinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 86.0, 30 için % 99.4 ve 100 için % 99.5'dir.

Bununla birlikte iki tanı testi arasında korelasyon çok zayıf yine kategori sayısı 3-5-7 ve 10 olduğu durumda Tip I Hata yapma olasılığı örnek genişliği ve değerlendirici sayısından etkilenmeksizin % 6'nın altındadır. Fakat tanı testi binary yapıda olduğu durumda 2 değerlendiricinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 41.6, 30 için % 60.4 ve 100 için % 63.2'dir. Beş değerlendiricinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 70.1, 30 için % 89.9 ve 100 için % 96.9 dur. Yedi değerlendiricinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 86.6, 30 için % 95.6 ve 100 için % 99.2'dir.

Çizelge 4.11: Beş tanı testi arasında korelasyon çok güçlü ve zayıf olduğunda, değerlendiricilerin diagnostik doğruluk farklılıklarına ilişkin olabilecek Tip I Hata yapma olasılıkları

Değerlendirici Sayısı	Örnek Genişliği	Beş Tanı Testi Arasında Korelasyon Çok Güçlü					Beş Tanı Test Arasında Korelasyon Çok Zayıf				
		Tanı Testinin Kategori Sayısı					Tanı Testinin Kategori Sayısı				
		K=2	K=3	K=5	K=7	K=10	K=2	K=3	K=5	K=7	K=10
İki değerlendirici	N=15	39.1	5.6	4.5	3.9	3.2	41.8	5.4	2.7	3.2	3.9
	N=30	65.4	5.5	4.6	4.3	4.6	63.2	4.9	4.5	5.8	4
	N=100	60.5	5.8	5.2	5.5	5.8	62.7	5.6	4.8	4.6	4.2
Beş değerlendirici	N=15	75.4	4.9	3.1	4.3	2.3	76.6	3.4	3.9	3.6	3.3
	N=30	90.6	5.1	4.4	3.5	4.5	94.0	4.5	4.9	3.9	4
	N=100	-	-	-	-	-	-	-	-	-	-
Yedi değerlendirici	N=15	84.9	5.3	4.2	4	3.4	89.9	4.6	3.1	2.4	3.0
	N=30	94.2	4.5	3.8	3.2	4.0	98.4	4.4	4.2	4.4	4.5
	N=100	-	-	-	-	-	-	-	-	-	-

-Paket programın kapasitesi yeterli gelmediği için hesaplanamamaktadır.

Çizelge 4.11 incelendiğinde beş tanı testi arasında korelasyon çok güçlü olduğunda, değerlendirici sayısı ve örnek genişliğinden etkilenmeksizin tanı testinin kategorisi 3,5,7 ve 10 olduğu durumda Tip I Hata yapma olasılığı % 6'nın altındadır. Fakat tanı testi binary yapıda ve 2 değerlendiricinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 39.1, 30 için % 65.4 ve 100 için % 60.5'dir. Beş değerlendiricinin var olduğu durumda olasılıklar örnek genişliği 15 için % 75.4, 30 için % 90.6 dir. Yedi değerlendiricinin var olduğu durumda da örnek genişliği 15 için % 84.9 ve 30 için % 94.2'dür.

Bununla birlikte beş tanı testi arasında korelasyon çok zayıf, tanı testi binary yapıda ve 2 değerlendiricinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 41.8, 30 için % 63.2 ve 100 için % 62.7'dir. Beş değerlendiricinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 76.6, 30 için % 94.0' dir. Yedi değerlendiricinin var olduğu durumda da bu olasılık örnek genişliği 15 için % 89.9 ve 30 için % 98.4'dür.

Çizelge 4.12: Yedi tanı testi arasında korelasyon çok güçlü ve zayıf olduğunda, değerlendiricilerin diagnostik doğruluk farklılıklarına ilişkin

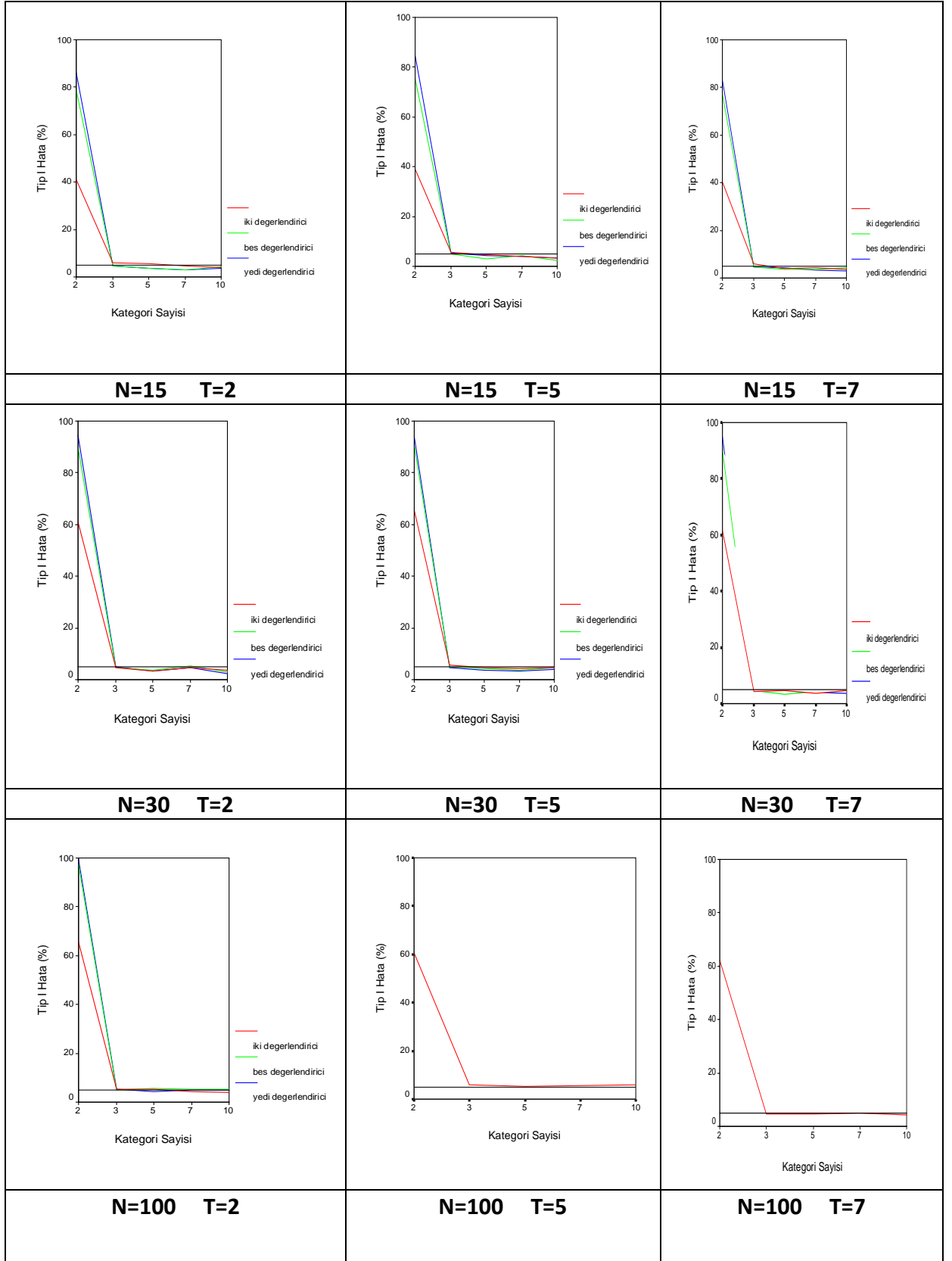
olabilecek Tip I Hata yapma olasılıkları

Değerlendirici Sayısı	Örnek Genişliği	Yedi Tanı Testi Arasında Korelasyon Çok Güçlü					Yedi Tanı Testi Arasında Korelasyon Çok Zayıf				
		Tanı Testinin Kategori Sayısı					Tanı Testinin Kategori Sayısı				
		K=2	K=3	K=5	K=7	K=10	K=2	K=3	K=5	K=7	K=10
İki değerlendirici	N=15	40.4	5.9	3.9	4.3	3.6	46.6	3.7	3.7	3.6	3.6
	N=30	61.7	4.4	4.6	3.5	4.6	60.3	4.5	3.9	5.0	5.9
	N=100	62.3	4.5	4.6	4.9	4.2	61.1	4.3	5.4	4.9	4.2
Beş değerlendirici	N=15	77.0	4.6	3.7	3.7	4.4	83.0	3.5	4.0	3.3	3.2
	N=30	90.0	4.8	3.3	4.3	4.5	89.4	5.6	4.6	3.9	4.5
	N=100	-	-	-	-	-	-	-	-	-	-
Yedi değerlendirici	N=15	83.0	4.8	4.4	3.2	2.9	75.4	4.2	3.0	2.3	3.0
	N=30	94.9	6.1	5.2	3.9	3.6	96.6	6.3	5.5	4.6	2.6
	N=100	-	-	-	-	-	-	-	-	-	-

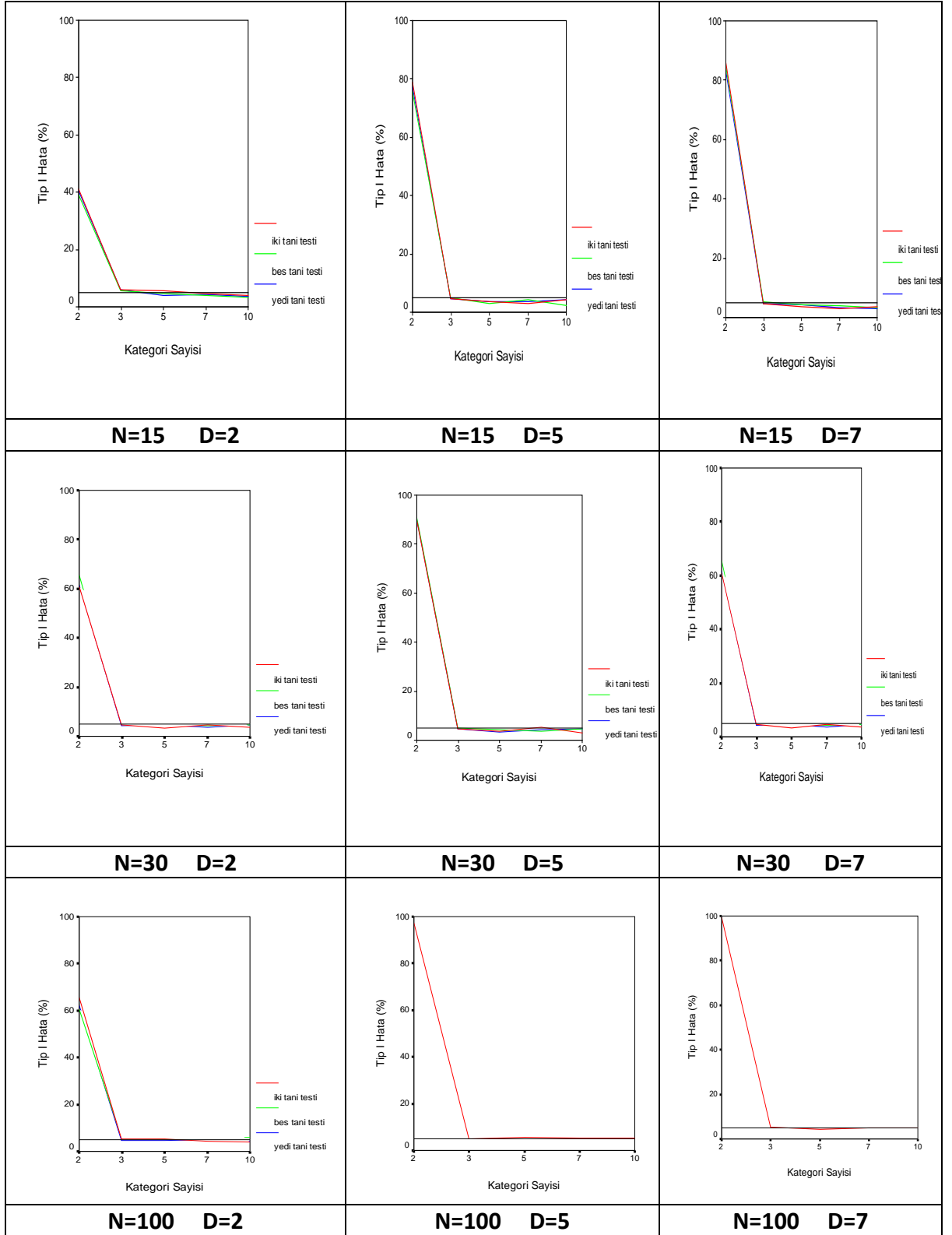
-Paket programın kapasitesi yeterli gelmediği için hesaplanamamaktadır.

Çizelge 4.12 incelendiğinde yedi tanı testi arasında çok güçlü bir korelasyon var olduğunda, değerlendirici sayısı ve örnek genişliğinden etkilenmeksizin tanı testinin kategorisi 3,5,7 ve 10 olduğu durumda Tip I Hata yapma olasılığı %5'in altındadır. Fakat tanı testi binary yapıda ve 2 değerlendiricinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 40.4, 30 için % 61.7 ve 100 için % 62.3'dür. Beş değerlendiricinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 77.0, 30 için % 90.0 dir. Yedi değerlendiricinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 83.0 ve 30 için % 94.9'dur.

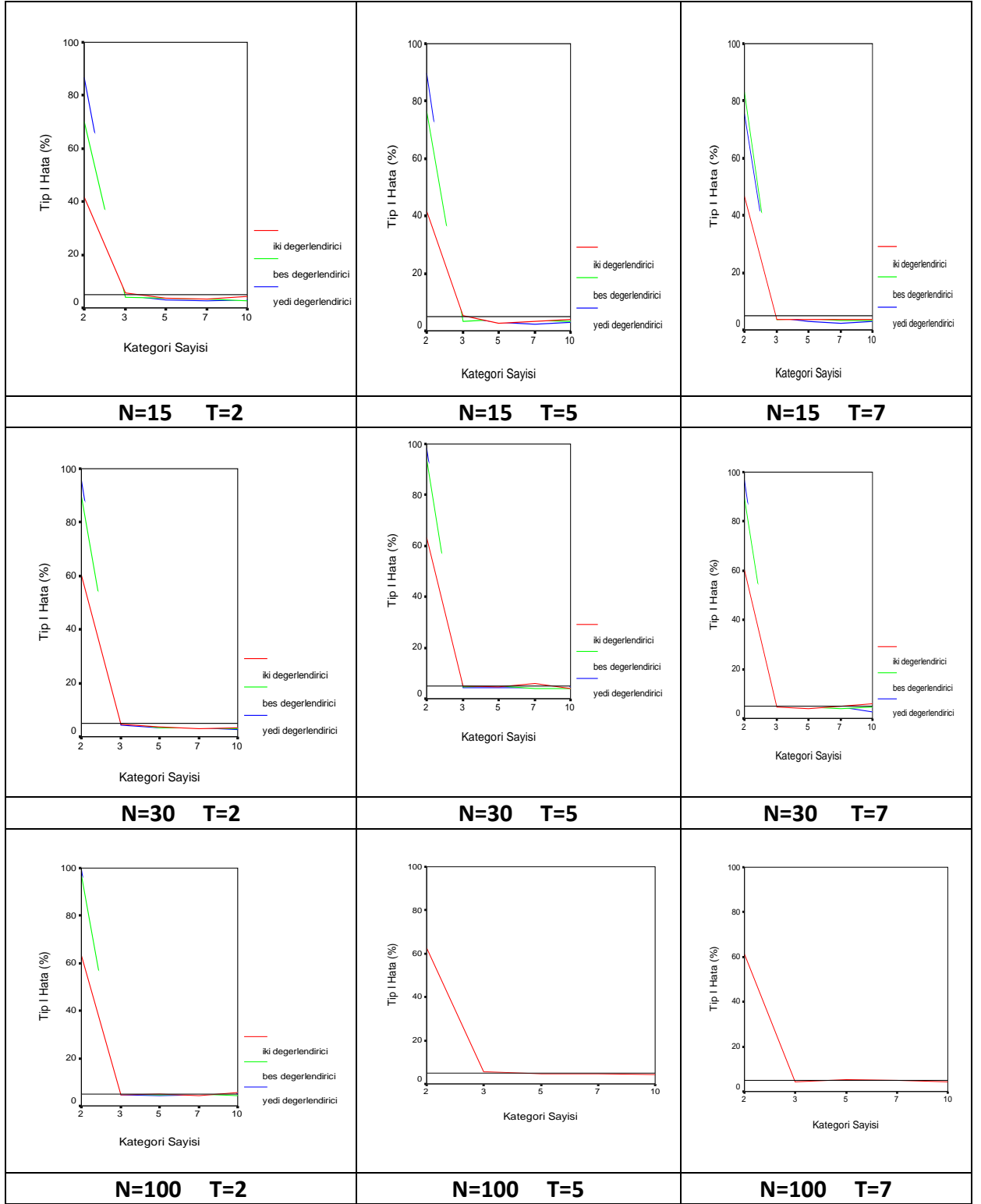
Bununla birlikte beş tanı testi arasında korelasyon çok zayıf, tanı testleri binary yapıda ve 2 değerlendiricinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 46.6, 30 için % 60.3 ve 100 için % 61.1'dir. Beş değerlendiricinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 83.0, 30 için % 89.4 dür. Yedi değerlendiricinin var olduğu durumda hata yapma olasılığı örnek genişliği 15 için % 75.4 ve 30 için % 96.6'dır.



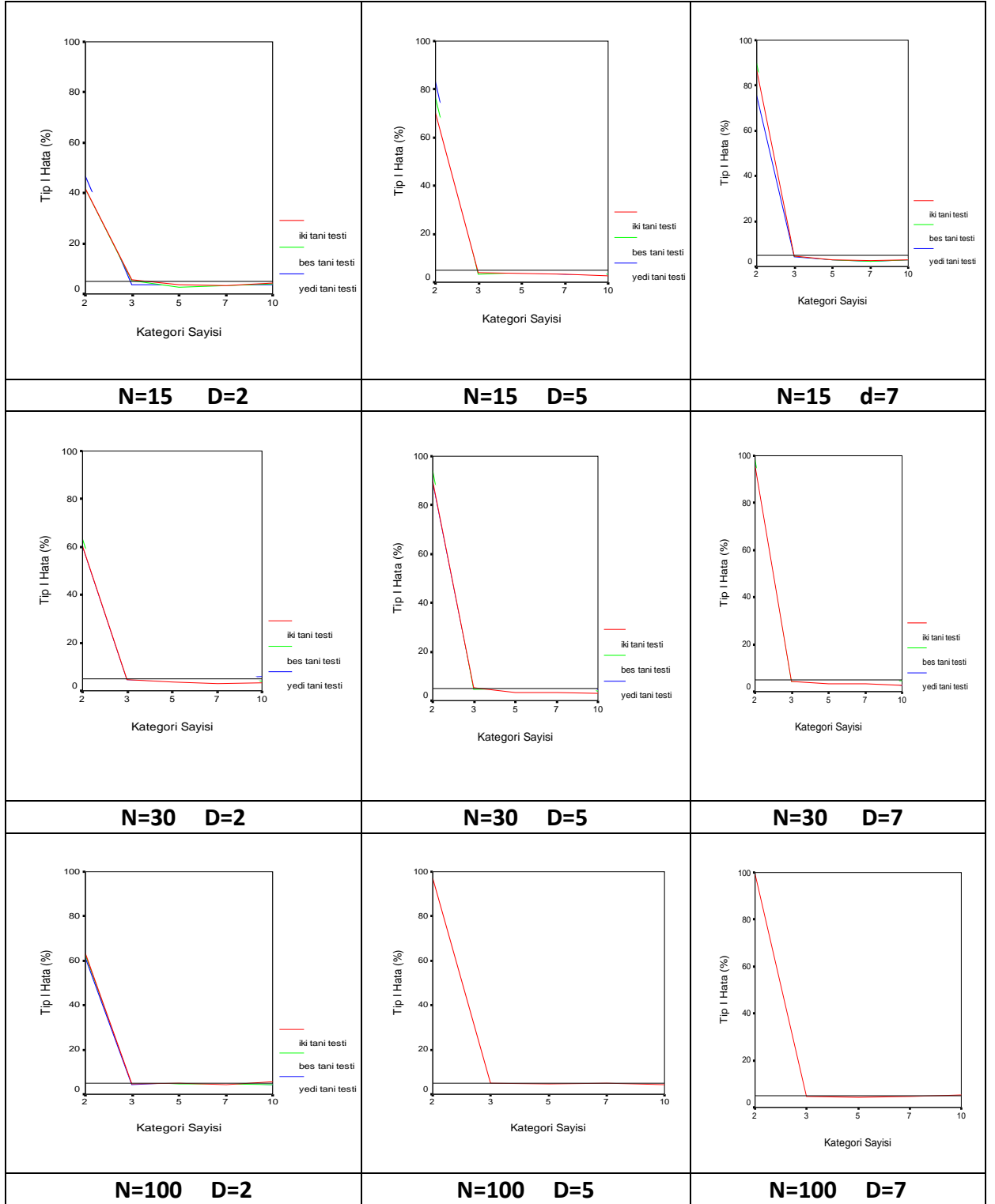
Şekil 4.13: Tanı testleri arasında korelasyon çok güçlü olduğunda, 2-5-7 tanı testi için değerlendiricilerin diagnostik doğruluk farklılıklarına ilişkin olabilecek Tip I Hata yapma olasılığı gösterim grafiği



Şekil 4.14: Tanı testleri arasında korelasyon çok güçlü olduğunda, 2-5-7 değerlendirici için değerlendiricilerin diagnostik doğruluk farklılıklarına ilişkin olabilecek Tip I Hata yapma olasılığı gösterim grafiği



Şekil 4.15: Tanı testleri arasında korelasyon çok zayıf olduğunda, 2-5-7 tanı testi için değerlendiricilerin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I Hata yapma olasılığı gösterim grafiği



Şekil 416: Tani testleri arasında korelasyon çok zayıf olduğunda, değerlendirici sayısı 2-5-7, değerlendiricilerin diagnostik doğruluk farklılıklarına ilişkin olabilecek Tip I Hata yapma olasılığı gösterim grafiği

5. TARTIŞMA

Bu çalışmada, değerlendirici sayısı ve değerlendiriciler arası uyumun, tanı testi sayısı ve tanı testleri arası korelasyonun ve tanı testinin kategori sayısının farklı vaka sayılarındaki deneme planlarında değerlendiriciler ve tanı testleri için diagnostik doğruluk yani ROC eğrileri altında kalan alanlar arası farka olan etkilerine bakılmıştır. Bu tez çalışmasında tartışma bölümü daha etkin ve anlaşılır olabilmesi açısından Gereç ve Yöntem de belirtilen iki farklı simülasyon denemeleri için ayrı ayrı yapılandırılmış ve bulgulardan elde veriler tartışma amacıyla grafikler yardımıyla birleştirilmiş ve bu bölümde sunulmuştur.

5.1. Değerlendiriciler Arasındaki Yüksek ve Düşük Uyumun Diagnostik Doğruluklar Üzerindeki Etkisinin Neden Olduğu Tip I Hata Yapma Olasılıkları

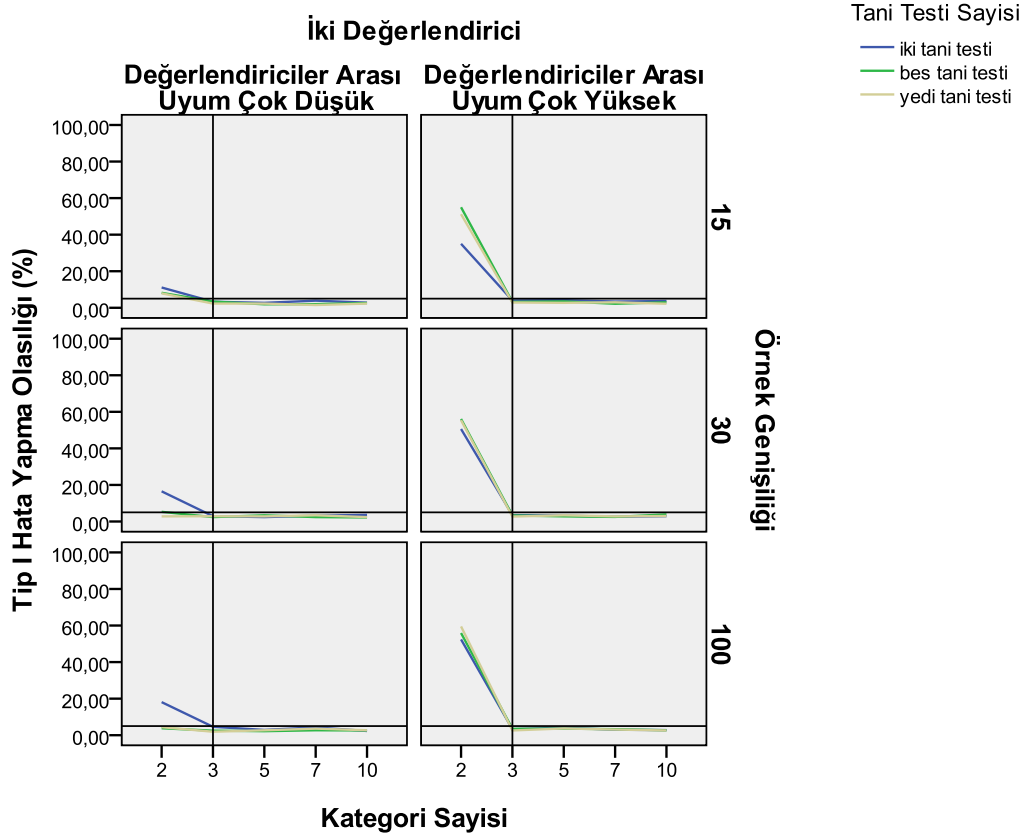
5.1.1. Tanı Testlerinin Diagnostik Doğrulukları Arasında Gerçekte Farklılık Olmadığında Farklılık Var Olarak Bulunma Olasılıkları

İki değerlendirici arasında çok yüksek bir uyumun olduğu durumlarda tanı testi sayısı ve örnek genişliğinden etkilenmeksizin tanı testinin kategorisi 3,5,7 ve 10 olduğu durumda tanı testlerinin diagnostik doğrulukları arasında gerçekte farklılık yokken fark var bulunma ihtimalini gösteren Tip I Hata yapma olasılıkları %5'in altında olduğu saptanmıştır. Fakat tanı testlerinin binary olduğu durumlarda tanı testinin sayısından bağımsız olarak Tip I Hata olasılıklarının % 5 'in çok üzerinde olduğu gözlenmiştir. Ayrıca tanı testinin sayısı ve örnek genişliği artıkça da hata yapma olasılıkları artmaktadır.

Buna karşın iki değerlendirici arasında uyum çok düşükken yine kategori sayısı 3,5,7 ve 10 olduğu durumda Tip I Hata yapma olasılığı örnek genişliği ve tanı testi sayısından etkilenmeksizin % 5'in altındadır. Klinik olarak bu tablonun kullanıldığı ortamlara örnek olarak ultrason bulgularından Meme CA tanısı koymada kullanılan

BRADS kriterleri gösterilebilir. 6 tanı testinden oluşan bu kriterin başarısı için en az 2 değerlendirici tarafından uygulanması önerilmektedir (10).

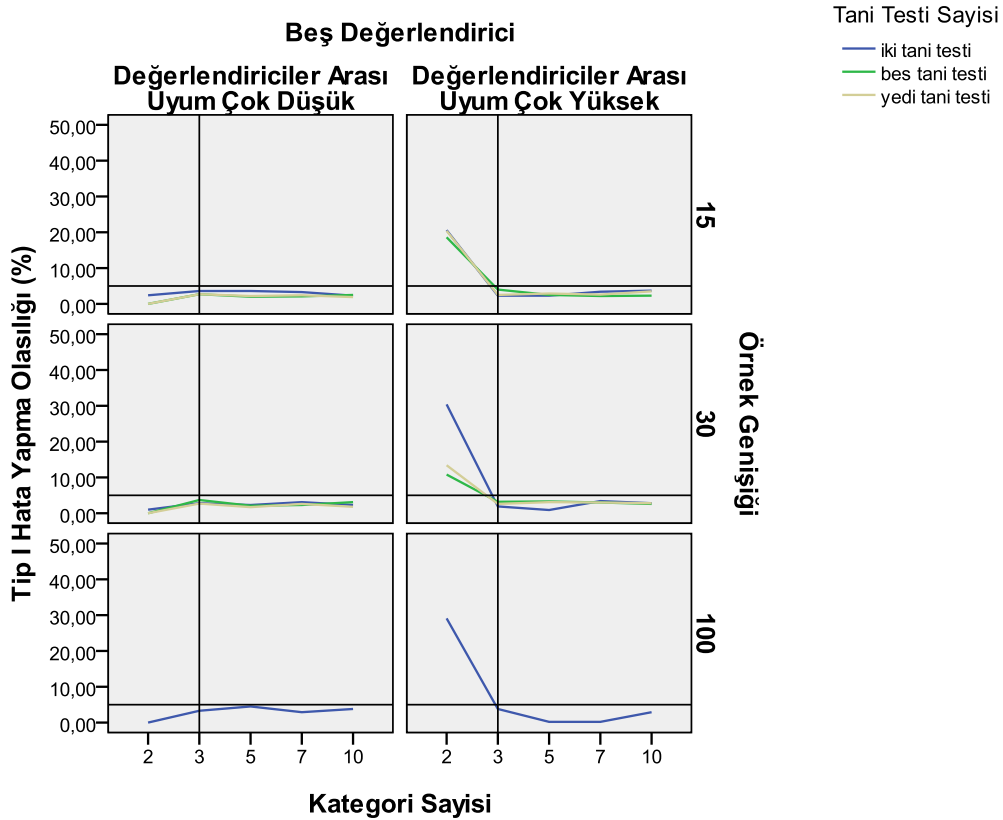
Değerlendiriciler arasında uyum çok düşükken, tanı testi binary yapıda olduğu durumda hata yapma olasılığı tanı testi sayısı ve örnek genişliğinden etkilenmektedir. Tanı testi sayısı 2 olduğu durumda örnek genişliği artıkça hata yapma olasılığı da artmaktadır. Buna karşın tanı testi sayısı 5 ve 7 olduğu durumda örnek genişliliğini 15’den 30 ve 100 çıkarma hata yapma düzeyini % 8’lerden % 5’in altına düşürmektedir. İki değerlendirici arasında uyum çok düşük ve çok yüksekken, tanı testi sayısı, örnek genişliği, tanı testinin kategorilerine göre tanı testlerinin diagnostik doğrulukları arasında farklılık yokken var bulunma olasılığını gösteren, Tip I Hata yapma olasılıkları Şekil 5.1’de sunulmuştur.



Şekil 5.1: İki değerlendirici arasında çok yüksek ve düşük uyum varken, tanı testlerinin diagnostik doğruluk farklılıklarına ilişkin olabilecek Tip I Hata yapma olasılıklarının grafik gösterimi

Beş değerlendirici arasında çok yüksek bir uyum varken tanı testi sayısı ve örnek genişliğinden etkilenmeksizin tanı testinin kategorisi 3,5,7 ve 10 olduğu durumda Tip I Hata yapma olasılığı % 5'in altındadır. Fakat tanı testi binary yapıda olduğu durumlarda tanı testinin sayısı ve örnek genişliği için üretilen tüm kombinasyonlarda Tip I Hata yapma olasılığı yine % 5'in üzerinde yer almaktadır. Fakat tanı testi sayısı ve örnek genişliğinin birlikte artışı ile Tip I hatalar % 10' a kadar çekilebilmektedir.

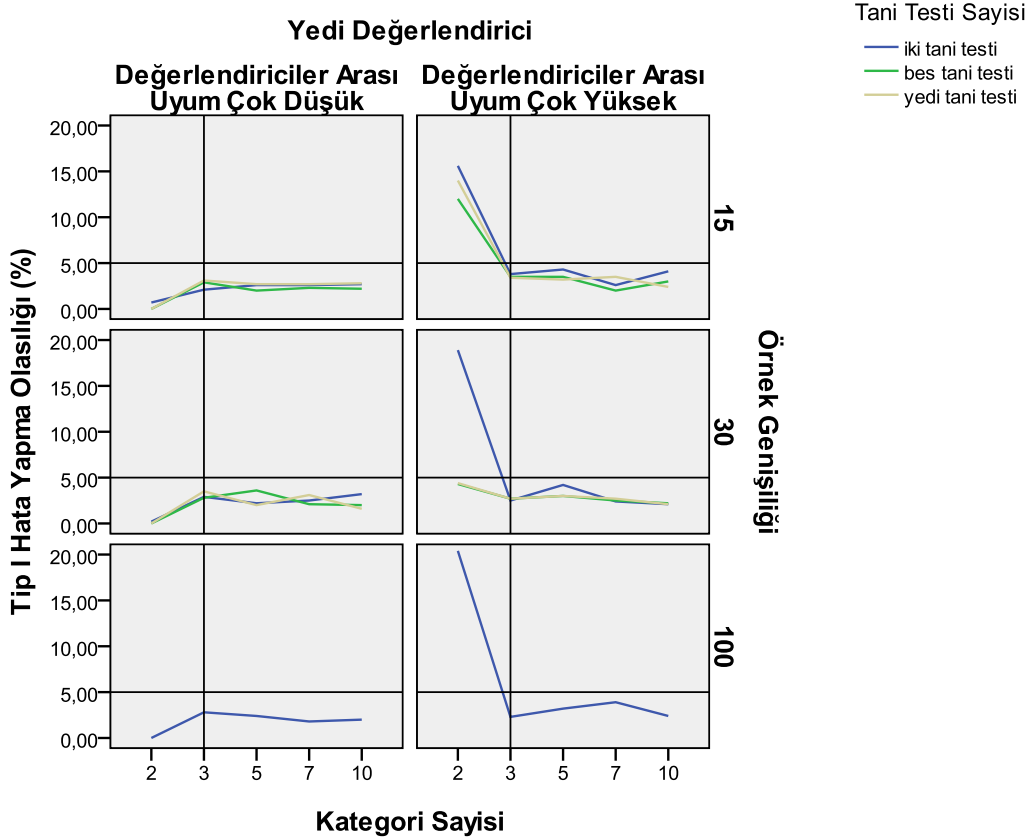
Buna karşın beş değerlendirici arasında düşük uyum olduğunda tanı testi sayısı, kategorisi ve örnek genişliğinden etkilenmeksizin Tip I Hata yapma olasılığı % 5'in altında elde edilmektedir. Beş değerlendirici arasında uyum çok düşük ve çok yüksek olduğunda tanı testi sayısı, örnek genişliği, tanı testinin kategorilerine göre tanı testlerinin diagnostik doğrulukları arasında farklılık yokken var bulunma olasılığını gösteren Tip I Hata yapma olasılıkları Şekil 5.2'de sunulmuştur.



Şekil 5.2: Beş değerlendirici arasında çok yüksek ve düşük uyum varken, tanı testlerinin diagnostik doğruluk farklılıklarına ilişkin olabilecek, Tip I Hata yapma olasılıklarının grafik gösterimi

Yedi deęerlendirici arasında ok yksek bir uyum varken tanı testi sayısı ve rnek geniřlięinden etkilenmeksizin tanı testinin kategorisi 3,5,7 ve 10 olduęu durumda Tip I Hata yapma olasılıęı % 5'in altındadır. Fakat tanı testinin kategorisi binary yapıda olduęu durumlarda tanı testinin sayısı ve rnek geniřlięi ne olursa olsun Tip I Hata yapma olasılıęı yine % 5'in zerinde yer almaktadır. Fakat yedi deęerlendirici arasında uyum ok yksek olduęunda tanı testinin kategorisi binary yapıda olduęu durumda hata yapma olasılıęı tanı testi sayısı ve rnek geniřlięinden etkilenmektedir. Tanı testi sayısı 2'den 5 ve 7'ye ıkarıldıęında rnek geniřlięi de 15'den 30'a ıkarılırsa Tip I Hata yapma olasılıęı % 5'in altına dřmektedir.

Buna karřın yedi deęerlendirici arasında uyum ok dřkken tanı testi sayısı, kategorisi ve rnek geniřlięinden etkilenmeksizin Tip I Hata yapma olasılıęı % 5'in altında yer almaktadır. İki beř ve yedi deęerlendirici arasında uyum ok yksek ve dřkse rnek geniřlięinden ve deęerlendirici sayısından etkilenmeksizin tanı testinin kategorisi 3,5,7 ve 10 olduęu durumda Tip I Hata yapma olasılıęı % 5'in altındadır. Fakat aynı kořullarda tanı testi binary yapıda olduęunda Tip I Hata yapma olasılıęı % 20 kadar ıkmaktadır. Yedi deęerlendirici arasında uyum ok dřk yksekse tanı testi sayısı, rnek geniřlięi, tanı testinin kategorilerine gre, tanı testlerinin diagnostik doęrulukları arasında farklılık yokken var bulunma olasılıęını gsteren Tip I Hata yapma olasılıęları Őekil 5.3'de sunulmuřtur.



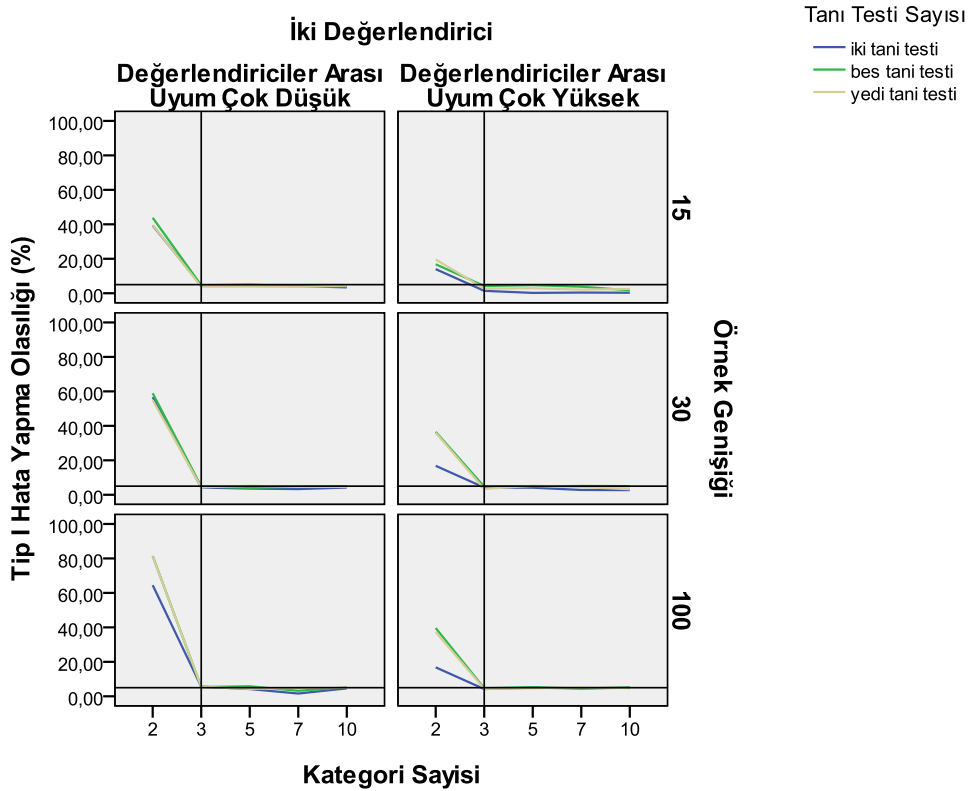
Şekil 5.3: Yedi değerlendirici arasında çok yüksek ve düşük uyum varken, tanı testlerinin diagnostik doğruluk farklılıklarına ilişkin olabilecek, Tip I Hata yapma olasılıklarının grafik gösterimi.

5.1.2. Değerlendiricilerin Diagnostik Doğrulukları Arasında Gerçekte Farklılık Olmadığında Farklılık Var Olarak Bulunma Olasılıkları

İki değerlendirici arasında çok yüksek bir uyum varsa tanı testi sayısı ve örnek genişliğinden etkilenmeksizin tanı testinin kategorisi 3,5,7 ve 10 olduğu durumda değerlendiricilerin diagnostik doğrulukları arasında gerçekte farklılık olmadığında var olarak bulunma olasılığını gösteren Tip I Hata % 5'in altındadır. Fakat tanı testlerinin binary yapıda olduğu durumda tanı testinin sayısı ne olursa olsun gerçekleşen Tip I Hatalar % 5 'in çok üzerindedir. Ayrıca tanı testinin sayısı ve örnek genişliği arttıkça da bu olasılıklar artmaktadır.

Buna karşın iki değerlendirici arasında uyum çok düşükse yine kategori sayısı 3,5,7 ve 10 olduğu durumda Tip I Hata yapma olasılıkları, örnek genişliği ve tanı testi sayısından etkilenmeksizin % 5'in altında olmasına karşın tanı testinin binary olduğu

bazı durumlarda % 5'in altına düşmektedir. Değerlendiriciler arasında uyum çok düşükse ve tanı testi binary yapıdaysa Tip I hata olasılığı tanı testi sayısı ve örnek genişliğinden etkilenmektedir ve tanı testinin sayısı ve örnek genişliği arttıkça da hata yapma olasılığı artmaktadır. İki değerlendirici arasında uyum çok düşük ve çok yüksek olduğunda tanı testi sayısı, örnek genişliği, tanı testinin kategorilerine göre değerlendiricilerin diagnostik doğrulukları arasında gerçekte fark yokken var olarak bulunma olasılığını gösteren, Tip I Hata yapma olasılıkları Şekil 5.4'de sunulmuştur.

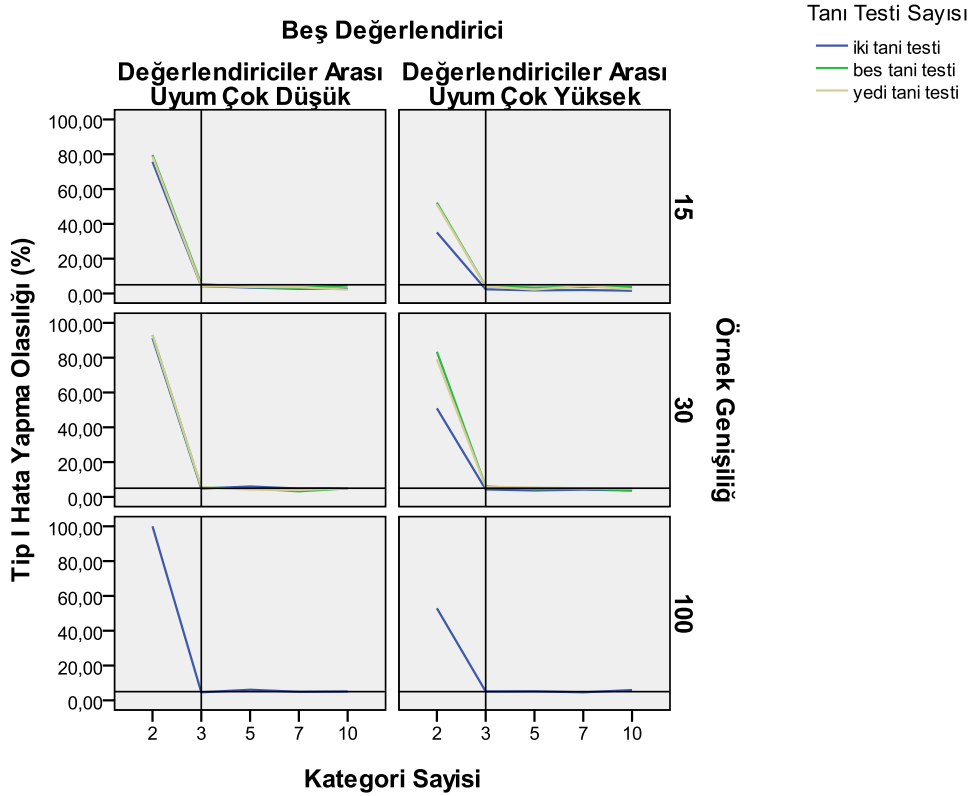


Şekil 5.4: İki değerlendirici arasında çok yüksek ve düşük uyum varken değerlendiricilerin diagnostik farklılıklarına ilişkin olabilecek, Tip I Hata yapma olasılıklarının grafik gösterimi.

Beş değerlendirici arasında çok yüksek bir uyum varsa tanı testi sayısı ve örnek genişliğinden etkilenmeksizin tanı testinin kategorisi 3,5,7 ve 10 olduğu durumda Tip I Hata yapma olasılığı % 6'nın altındadır. Fakat tanı testinin binary yapıda olduğu durumlarda tanı testinin sayısı ve örnek genişliği ne olursa olsun Tip I Hata yapma

olasılığı yine % 5'in üzerinde yer almaktadır. Tanı testinin sayısı ve örnek genişliği artıkça da hata yapma olasılığı artmaktadır.

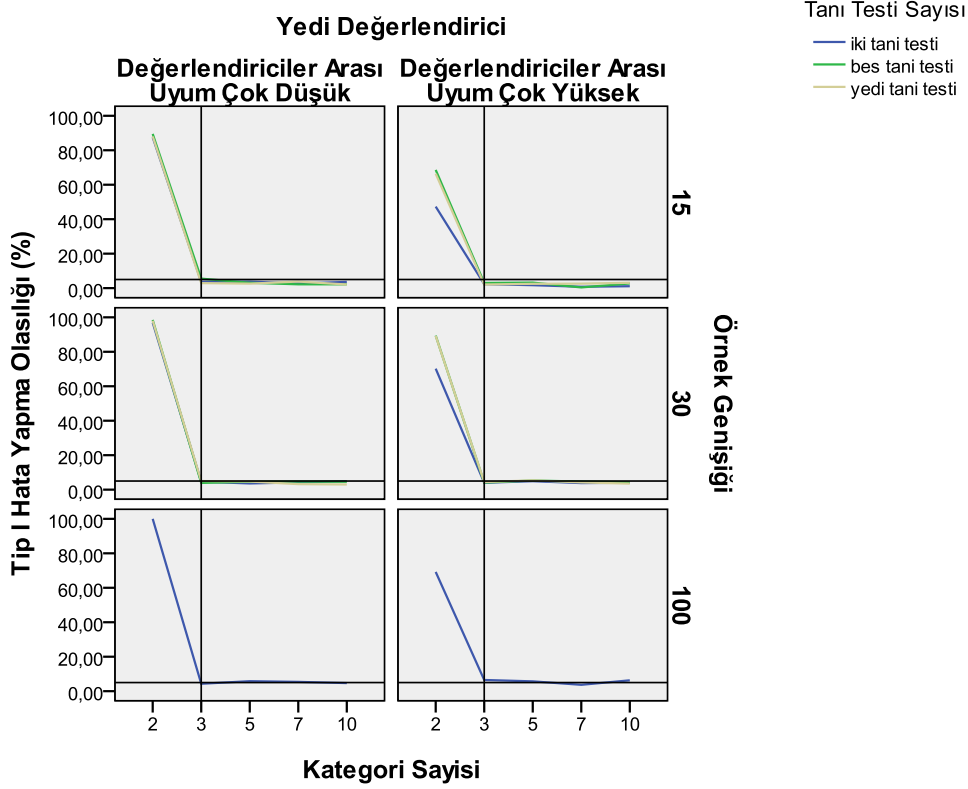
Buna karşın beş değerlendirici varsa ve aralarındaki uyum düşükse tanı testi sayısı ve örnek genişliğinden etkilenmeksizin tanı testinin kategorisi 3,5,7 ve 10 olduğu durumda Tip I Hata yapma olasılığı % 6'nın altındadır. Fakat tanı testinin kategorisi binary yapıda olduğu durumlarda tanı testinin sayısı ve örnek genişliği ne olursa olsun Tip I Hata yapma olasılığı yine % 5'in üzerinde hatta bazı durumlarda % 100' e kadar ulaşmaktadır. Ve tanı testinin sayısı ve örnek genişliği artıkça da hata yapma olasılığı artmaktadır. Beş değerlendirici arasında uyum çok düşük ve çok yüksek iken tanı testi sayısı, örnek genişliği, tanı testinin kategorilerine göre Tip I Hata yapma olasılıkları Şekil 5.5'de sunulmuştur.



Şekil 5.5: Beş değerlendirici arasında çok yüksek ve düşük uyum varken, değerlendiricilerin diagnostik doğruluk farklılıklarına ilişkin olabilecek, Tip I Hata yapma olasılıklarının grafik gösterimi.

Yedi deęerlendirici varsa ve aralarında çok yüksek bir uyum varsa beş deęerlendirici olduęu duruma benzer bir tablo karřımıza çıkmakta ve tanı testi sayısı ve örnek geniřlięinden etkilenmeksizin tanı testinin kategorisi 3-5-7 ve 10 olduęu durumda Tip I Hata yapma olasılıęı % 6'nın altında gerekleřmektedir. Fakat tanı testinin kategorisi binary yapıda olduęu durumlarda tanı testinin sayısı ve örnek geniřlięi ne olursa olsun Tip I Hata yapma olasılıęı yine % 5'in üzerinde hatta bazı durumlarda % 89.4' lere çıkmaktadır. Ve tanı testinin sayısı ve örnek geniřlięi artıka da hata yapma olasılıęı artmaktadır.

Buna karřın yedi deęerlendirici arasında uyum dūřuk iken tanı testi sayısı ve örnek geniřlięinden etkilenmeksizin tanı testinin kategorisi 3-5-7 ve 10 olduęu durumda Tip I Hata yapma olasılıęı % 5'in altındadır. Fakat tanı testinin kategorisi binary yapıda olduęu durumlarda tanı testinin sayısı ve örnek geniřlięi ne olursa olsun Tip I Hata yapma olasılıęı yine % 5'in üzerinde hatta bazı durumlarda % 100' e kadar ulařmaktadır. Ve tanı testinin sayısı ve örnek geniřlięi artıka da hata yapma olasılıęı artmaktadır. Yedi deęerlendirici arasında uyum çok dūřuk ve çok yüksek iken tanı testi sayısı, örnek geniřlięi, tanı testinin kategorilerine gre bu Tip I Hata yapma olasılıęları Őekil 5.6'de sunulmuřtur.



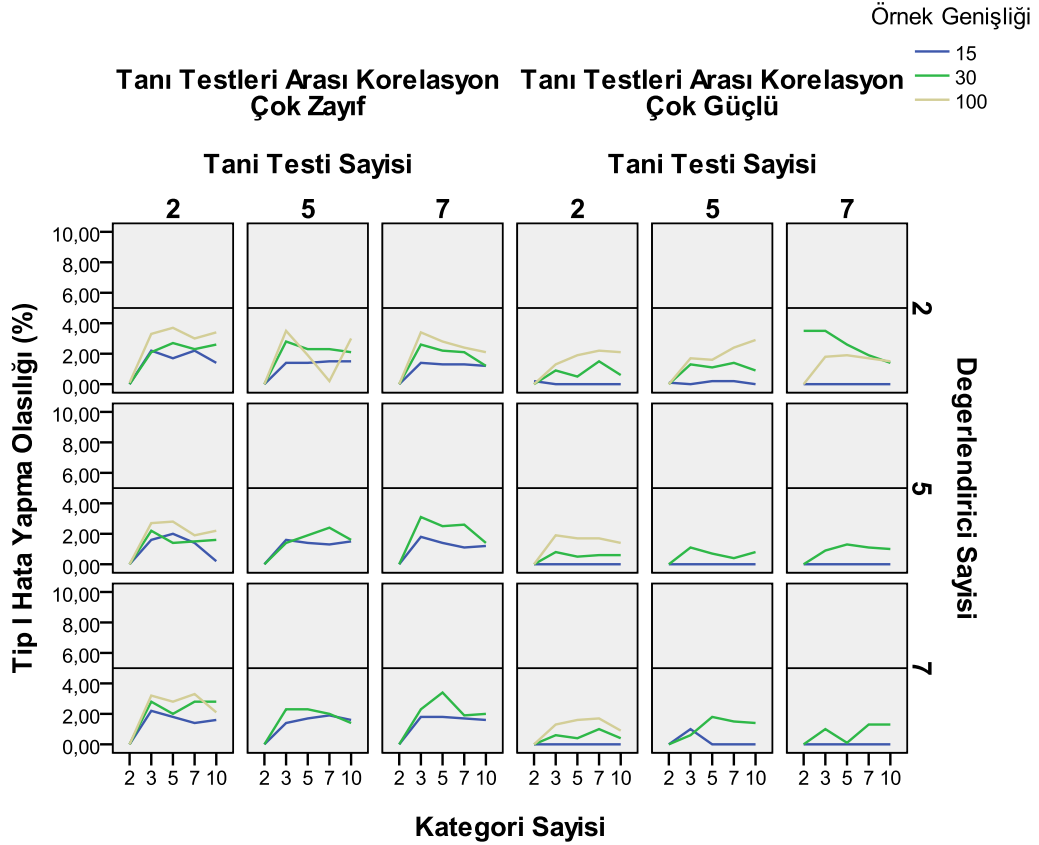
Şekil 5.6: Yedi değerlendirici arasında çok yüksek ve düşük uyum varken, değerlendiricilerin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I Hata yapma olasılıklarının grafik gösterimi.

5.2. Tanı Testleri Arasındaki Güçlü ve Zayıf Korelasyonun Diagnostik Doğruluklar Üzerindeki Etkisinin Neden Olduğu Tip I Hata Yapma Olasılıkları

5.2.1. Tanı Testlerinin Diagnostik Doğrulukları Arasında Gerçekte Farklılık Olmadığında Var Olarak Bulunma Olasılıkları

Tanı testleri arasında çok güçlü ve zayıf korelasyon olduğu durumlarda tanı testi sayısı, değerlendirici sayısı ve kategorisi ve hatta örnek genişliğinden etkilenmeksizin tanı testlerinin diagnostik doğrulukları arasında gerçekte farklılık yokken var olarak bulunma olasılığını gösteren Tip I Hatalar % 5'in altında elde edilmiştir. Bu bulgular DBM metodunda söz edilen ve bu metotla tanı testleri arası korelasyonun tanı testlerinin performanslarının klasik yöntemlerle karşılaştırılmasında yaratacağı olumsuz etkileri giderilebileceği savını desteklemektedir. Tanı testleri arasında korelasyonun çok

zayıf ve çok güçlü olduğu durumlar için değerlendirici sayısı, örnek genişliği, tanı testinin kategorilerine göre tanı testlerinin diagnostik doğrulukları arasında gerçekte fark yokken var olarak bulunma ihtimalini gösteren, Tip I Hata yapma olasılıkları Şekil 5.7’de sunulmuştur.

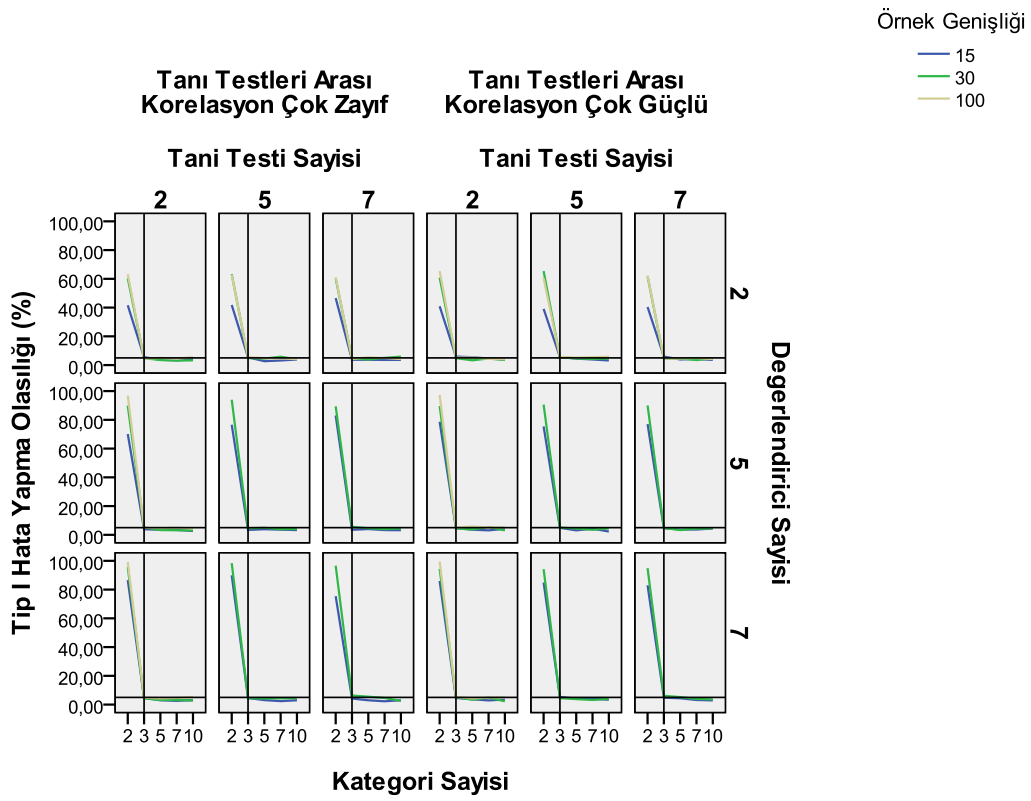


Şekil 5.7: İki-beş ve yedi tanı testi arasında korelasyon çok güçlü ve çok zayıf iken, tanı testlerinin diagnostik doğruluk farklılıklarına ilişkin, olabilecek Tip I Hata yapma olasılıklarının grafik gösterimi.

5.2.2. Değerlendiricilerin Diagnostik Doğrulukları Arasında Gerçekte Farklılık Olmadığında Farklılık Var Olarak Bulunma Olasılıkları

İki, beş ve yedi tanı testi arasında çok güçlü ve zayıf korelasyon varken değerlendirici sayısı ve örnek genişliğinden etkilenmeksizin tanı testinin kategorisi 3,5,7 ve 10 olduğu durumda değerlendiricilerin diagnostik doğrulukları arasında gerçekte farklılık olmadığında farklılık var olarak bulunma olasılıklarını gösteren Tip I Hata değerleri % 6'nın altındadır. Fakat binary tanı testleri için değerlendirici sayısı ne olursa

olsun gerçekleşen Tip I Hatalar % 5 'in çok üzerindedir ve değerlendirici sayısı ve örnek genişliği artıkça bu değerlerin arttığı gözlenmektedir. Değerlendirici sayısı ve örnek genişliği artıkça bu hata yapma olasılığı % 90'ların üzerine çıkmaktadır. Bu kombinasyonlardaki değerlendiricilerin diagnostik doğrulukları arasında gerçekte fark yokken fark var olarak bulunma olasılığını gösteren Tip I Hata yapma olasılıkları Şekil 5.8' de sunulmuştur.



Şekil 5.8: İki, beş ve yedi tanı testi arasında çok güçlü ve çok zayıf korelasyon varken, değerlendiriciler arasındaki diagnostik doğruluk farklılığına ilişkin, Tip I Hata yapma olasılıklarının grafik gösterimi.

6.SONUÇLAR ve ÖNERİLER

Birden çok tanı testi ve birden çok değerlendiricinin bulunduğu diagnostik doğruluk çalışmalarında değerlendiriciler arası uyum ve tanı testleri arasındaki korelasyon varlığının diagnostik doğruluk üzerinde oldukça önemli bir etkiye sahip olduğu bu tez çalışması sonuçlarından ortaya çıkmaktadır. Değerlendiriciler arası uyum klinikte veri güvenilirliğini ifade etmektedir. Fakat güvenilirlik ancak geçerlilik var olduğu durumda önem ifade eder. Klinikte değerlendiriciler arası uyumun, tutarlılığın yani güvenilirliğin yüksek olması beklenen hatta istenen bir durum olmasına karşın tanı testleri arasındaki korelasyonun yüksek olması tanı testinin gücünü azaltacağından istenen bir durum değildir. Tanı testlerinin kararlarının bağımsız olması kararların birleştirilebilirliği açısından önemlidir.

İki ve beş değerlendirici arasında yüksek uyum varken yani güvenilirlik söz konusu olduğunda, tanı testinin binary yapıda olması, tanı testlerinin diagnostik doğrulukları üzerinde olumsuz etkiye neden olmakta ve tanı testlerinin geçerliliğini bozmaktadır. Değerlendirici sayısı iki değerlendiriciden sırasıyla beş ve yedi değerlendiriciye çıkarıldıkça ve tanı testi sayısı ve örnek genişliği artırıldıkça bu sorun azalmaktadır. Değerlendirici sayısı, tanı testi sayısı ve örnek genişliğinin tüm kombinasyonlarında tanı testinin binary yapıda değilse değerlendiriciler arası yüksek uyumun getirdiği olumsuz etki çözümlenmektedir.

Değerlendiriciler arasında çok düşük uyum olması sadece iki değerlendirici ve iki tanı testinin var olduğu deneme planlarında sorun yaratmaktadır. Değerlendiriciler arası düşük uyumun neden olduğu olumsuz etki ise tanı testi sayısını en az 5'e ve gruplardaki vaka sayısını en az 30'a çıkarmakla giderilebilmektedir. Değerlendirici sayısı 2'den fazla olduğunda, tanı testi sayısı, örnek genişliği ve tanı testlerinin kategorisi için simüle edilen tüm kombinasyonlarda tanı testlerinin diagnostik doğrulukları üzerinde bir sorun yaratmamaktadır.

Değerlendiriciler arasındaki çok yüksek ya da çok düşük uyum, tanı testlerinin binary yapıda olduğu tüm deneme planlarında değerlendiricilerin diagnostik doğruluklarına ilişkin Tip I hatalar üzerinde olumsuz etkilidir. Tanı testi binary yapıdan

farklılaştığında değerlendiriciler arası çok yüksek ya da çok düşük uyum diagnostik doğruluk üzerinde etkili olmamaktadır.

Dorfman ve arkadaşlarının (1992) ortaya attığı DBM yöntemi Jackknife yeniden örnekleme yöntemi ile tanı testleri arasındaki korelasyonun tanı testlerinin diagnostik doğruluk sonuçlarını etkilemediğini iddia etmişlerdir. Bu çalışmada yapılan simülasyon denemeleri de bu sonucu doğrulamaktadır. Tanı testleri arasında çok güçlü korelasyonun var olduğu deneme planlarında değerlendirici sayısı, tanı testi sayısı, tanı testinin kategorisi ve örnek genişliği için üretilen tüm durumlarda bu korelasyonun tanı testlerinin diagnostik doğrulukları üzerine olumsuz bir etkisi gözlenmemiştir. Tanı testleri arasında çok güçlü ve zayıf korelasyonun var olduğu deneme planlarında, tanı testi sayısı, değerlendirici sayısı ve örnek genişliğine ait tüm durumlarda tanı testinin binary yapıda olması değerlendiricilerin diagnostik doğrulukları üzerinde olumsuz etkiye neden olmaktadır. Tanı testinin kategorisi en az 3 olduğunda bu sorun gözlenmemektedir.

Birden çok değerlendirici ve birden çok tanı testinin bulunduğu doğruluk çalışmalarında değerlendiriciler ve tanı testleri arasında uyum ve uyumsuzluğun olması kaçınılmazdır. Her koşuldaki Tip I Hata artma riskini ortadan kaldırmak için örnek genişliği, tanı testi sayısı ve değerlendirici sayısının dengelenmesi gerekmektedir.

Dorfman ve ark. (1992) çalışmasında tanı testlerinin kategori düzeyi ile ilgili herhangi bir kısıtlama bulunmamaktadır. Gallas ve ark. (2007) çalışmasında da tanı testinin kategorisi binary yapı için çoklu tanıli çok değerlendiricili deneme planını kullanarak simülasyon denemeleri yapmışlardır (50). DBM metotunu kullanmalarına karşın Tip I hatalar üzerine bir sonuç yayınlamamışlardır. Bu tez çalışmasında kurgulanan tüm deneme planları ve kombinasyonlarda elde edilen simülasyon sonuçları tanı testinin binary yapıda olmasının Tip I Hatanın beklenen %5 değerinin oldukça üstünde değerler almasına neden olduğunu göstermektedir. Bu çalışma bulgularına dayanarak binary tanı testleri için elde edilen Tip I hataların %5 seviyesini koruyabilmesi için örneklem genişliğinin en az 100 ve aynı zamanda değerlendirici sayısının ve kullanılan tanı testi sayısının en az 5 olması gerektiği söylenilebilir.

Tezin kısıtlılıkları; bu tez çalışmasında değerlendiriciler arasındaki uyumun yüksek ve düşük olması koşulunda; farklı değerlendirici sayısı, tanı testi sayısı ve kategorisi, gruplardaki vaka sayısı kombinasyonu için toplam 135 farklı deneme planı yaratılmıştır. Ayrıca tanı testleri arasındaki korelasyonun yüksek ve zayıf olması koşulunda da aynı kombinasyonlar için toplam 135 farklı deneme planı yaratılmıştır. Toplam kombinasyon sayısı 270 olup her bir kombinasyon 1000 kez simüle edilmiş tek bir kombinasyonun simülasyonlarının tamamlanması yaklaşık 2.5 saat sürmüştür. Her kombinasyonda deneme planları oluşturulurken örnek genişlikleri dengeli tutulmuştur. Paket programın getirdiği sıkıntılardan kaynaklı olarak tanı testi sayısı ve değerlendirici sayısının en az 5'e çıktığı deneme planlarında, gruplardaki vaka sayısı 100 olduğunda varyans analiz tablosu oluşturulamamıştır. Programın getirdiği bir başka sınırlılık ise tanı testlerinin ayırt ediciliğinin(eğri altında kalan alan) yüksek olduğu denemelerde değerlendiriciler arasında uyum ve tanı testleri arasında korelasyon güçlü ise program varyans analiz tablosunu oluşturamamaktadır. Sonuç olarak LABMRMC 1.0 paket programı güçlü uyum, yüksek korelasyon ve güçlü ayırt edicilik koşullarında sorunlar yaratmaktadır.

KAYNAKLAR

1. **Pepe MS, Cai T, Longton G.** Combining Predictors for Classification Using the Area under the Receiver Operating Characteristic Curve. *Biometrics*, 2006; 62: 221-229.
2. **Harasymiw J, Bean P.** The Combined Use of the Early Detection of Alcohol Consumption (EDAC) Test and Carbohydrate-Deficient Transferrin to Identify Heavy Drinking Behaviour in Males. *Alcohol&Alcoholism*, 2001; 36 (4): 349-353.
3. **Su JQ, Liu SJ,** Linear Combinations of Multiple Diagnostic Markers. *Journal of the American Statistical Association*, 1993; 88 (424): 1350-1355.
4. **Liu B, Metz CC, Jiang Y.** Effect of Correlation on Combining Diagnostic Information from Two Images of the Same Patient. *Medical Physics*, 2005; 32 (11): 3329-3338.
5. **Pepe MS, Thompson ML.** Combining Diagnostic Test Results to Increase Accuracy, *Biostatistics*, 2000; 1 (2): 12-140
6. **Shoukri MM,** Measures of Interobserver Agreement. Washington: Chapman&Hall/CRC, **2004.**
7. **Sadatsafavi M, Najafzadeh M, Lynd L, Marra C,** Reliability Studies of Diagnostic Tests are not Using Enough Observers for Robust Estimation of Interobserver Agreement: A Simulation Study. *Journal of Clinical Epidemiology*, 2008; 61: 722-727.
8. **Cohen JA,** A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 1990; 20(1): 37-46.
9. **Kanık A, Orekici Temel G, Ersöz İ.** Fleiss Kappa ve Krippendorff Alpha Uyum Katsayılarının Örneklem Genişliği, Değerlendirici Sayısı ve Kullanılan Ölçeğin Kategori Sayısından Etkilenme Durumları Üzerine Bir Benzetim Çalışması. *Türkiye Klinikleri J Biostat*, 2010; 2(2): 74-81.
10. **Duce MN, Apaydın FD, Akkurt ÖE, Polat A, Kanık A, Orekici Temel G.** Inter and Intraobserver Agreement in BI-RADS US: Are They Good Enough?. Insights into Imaging Book Of Abstracts, March, Vienna, Austria, **2010.**
11. **Zhou X, Obuchowski N, McClish D.,** Statistical Methods in Diagnostic Medicine, Wiley, **2002.**
12. **Obuchowski NA.** Fundamentals of Clinical Research for Radiologists. *American Roentgen Ray Society*, 2005; 184: 364-372.
13. **Dorfman D, Berbaum K, Metz C.** Receiver operating Characteristic Rating Analysis: Generalization to the Population of Readers and Patients with Jackknife Method, *Investigate Radiology*, 1992; 27:723-731
14. **Obuchowski NA, Beiden SV., Berbaum SK, Hills S.L, Ishwaran H, Song HH, Wagner RF.** Multireader Multicase Receiver Operating Characteristic Analysis: An Empirical Comparison of Five Methods. *Academic Radiology*, 2004; 11:980-995.
15. **Hillis SL, Obuchowski NA, Schartz KM, Berbaum KS.** A Comparison of the Dorfman-Berbaum-Metz and Obuchowski-Rockette Methods for Receiver Operating Characteristic (ROC) Data. *Statistics in Medicine*, 2005; 24: 1579-1607.

16. Obuchowski NA. Fundamentals of Clinical Research for Radiologists. American Roentgen Ray Society, 2005; 184: 364-372.
17. **Hanley JA, McNeil BJ.** The Meaning and Use of The Area Under a Receiver Operating Characteristic (ROC) Curve. *Radiology*, 1982; 143: 29-36.
18. **Beam CA.** Analysis of Clustered Data in Receiver Operating Curve Studies. *Statistical Methods in Medical Research*, 1998; 7 (4): 324-336.
19. http://www.lerner.ccf.org/qhs/software/roc_analysis.php
20. **Obuchowski N.** Multi-Reader Multi Modality ROC Studies: Hypothesis Testing and Sample Size Estimation Using an ANOVA Approach with Dependent Observations., With rejoinder. *Academic Radiology*, 1995; 2:522-529.
21. **Toledano A, Gatsonis C.** Regression Analysis of Correlated Receiver Operating Characteristic Data. *Academic Radiology.*, 1995;2:30-36.
22. **Beiden S, Wagner, Campbell G.** Components-of-Variance Models and Multiple-Bootstrap Experiments: An Alternative Method for Random-effects, Receiver Operating Characteristic Analysis. *Academic Radiology*, 2000;7:341-349
23. **Obuchowski N.** Sample Size Tables for Receiver Operating Characteristic Studies. *American Roentgen Ray Society*, 2000; 175: 603-608.
24. **Song X.**, A marginal model approach for analysis of multi-reader multi-test receiver operating characteristic (ROC) data, *Biostatistics*, 2005; 6:303-312
25. **Hillis S.** A Comparison of Denominator Degrees of Freedom Methods for Multiple Observer ROC Analysis. *Statistics in Medicine*, 2007; 26(3): 596-619
26. **Usami H, Ikeda M, Ishigaki T, Fukushima H, Shimamoto K.** The Influence of Liquid Crystal (LCD) Monitors on Observer Performance for the Detection of Nodular Lesions on Chest Radiographs. *Eur Radiol*, 2006; 16: 726-732.
27. **Park CM, Lee JH, Goo JM, Han DH, Kim JH, Lim KY, Kim SH, Kang JJ, Kim KG, Lee CH, Chun EJ, Im JG.** Comparison of Observer Performance on Soft-Copy Reading of Digital Chest Radiographs: High Resolution Liquid-Crystal Display Monitors Versus Cathode-Ray Tube Monitors. *European Journal of Radiology*, 2008; 66:13-18.
28. **Cha JH, Moon KW, Cho N, Lee EH, Park JS, Jang MJ.** LCD Versus CRT Monitors for Digital Mammography: A Comparison of Observer Performance for the Detection of Clustered Microcalcifications and Masses. *Acta Radiologica*, 2009; 10:1104-1108.
29. **Sung JM, Chung MJ, Lee KS, Choe BK.** The Influence of Liquid Crystal Display Monitors on Observer Performance for the Detection of Interstitial Lung Markings on Both Storage Phosphor and Flat-Panel-Detector Chest Radiography. *European Journal of Radiology*, 2010; 74: 275-279.
30. **Pepe MS, Urban N, Rutter C, Longton G,** Design of a Study to Improve Accuracy in Reading Mammograms. *J. Clin. Epidemiol*, 1997;50: 1327-1338
31. **Obuchowski NA.,** Multireader Receiver Operating Characteristic Studies: A Comparisons of study Designs. *Acad Radiol*, 1995; 2: 709-716.
32. **Obuchowski NA.** New Methodological Tools for Multiple-Reader ROC Studies. *Radiology*, 2007; 243: 10-12.

33. **Dorfman DD, Berbaum KS, Lenth RV, Chen YF, Donaghy BA.** Monte Carlo Validation of a Multireader method for Receiver Operating Characteristic Discrete Rating Data: Factorial Experimental Design. *Academic Radiology*, 1998;5: 591-602.
34. **Obuchowski NA, Rockette HE.** Hypothesis Testing of the Diagnostic Accuracy for Multiple Diagnostic tests: An Approach with Dependent Observations. *Commun Stat Simul Computation*, 1995; 24: 285-308.
35. **Hillis SL, Berbaum, KS.** Power estimation for the Dorfman-Berbaum-Metz method. *Academic Radiology*, 2004; 11: 1260-1273.
36. **Hillis SL.** Monte Carlo validation of the Dorfman-Berbaum-Metz Method Using Normalized Pseudovalues and Less Data-Based Model Simplification. *Academic Radiology*, 2005;12: 1534-1541.
37. LABMRMC. Available at: http://xray.bsd.uchicago.edu/krl/KRL_ROC/software_index.htm
38. **Schartz KM, Hillis SL, Berbaum KS, Dorfman DD: MRMC2.0.** Available at: <http://perception.radiology.uiowa.edu>.
39. DBM MRMC. Available at:<http://perception.radiology.uiowa.edu/Software/ReceiverOperatingCharacteristicROC/DBMMRMC/tabid/116/Default.aspx>
40. **Hillis LS, Berbaum KS., Metz CE.** Recent Developments in the Dorfman-Berbaum-Metz Procedure for Multireader ROC Study Analysis. *Academic Radiology*, 2008; 15; (5): 647-661.
41. **Song HH.** Analysis of Correlated ROC Areas in Diagnostic Testing. *Biometrics*, 1997; 53: 370-382.
42. **Ishwaran H, Gatsonis CA,** A General Class of Hierarchical Ordinal Regression Models with Applications to Correlated ROC Analysis. *Can J Stat*, 2000;28:731-750
43. **Wang F, Gatsonis C.A.** Hierarchical Models for ROC Curve Summary Measures: Design and Analysis of Multi-Reader, Multi-Modality Studies of Medical Tests. *Statistics in Medicine*, 2008; 27: 243-256.
44. **Toledano AY.** Three Methods for Analysing Correlated ROC Curves: A Comparison in Real Data Sets from Multi-Reader, Multi-Case Studies with a Factorial Design. *Statistics in Medicine*, 2003; 22: 2919-2933.
45. **Beiden S., Wagner R, Campbell G, Metz CE, Jiang Y.** Components-of-Variance Models for Random Effects ROC Analysis: The Case of Unequal Variance Structures Across Modalities. *Academic Radiology*, 2001; 8: 605-615.
46. **Beiden S, Wagner R, Campbell G, Chan HP.** Analysis of Uncertainties in Estimates of Components of Variance in Multivariate ROC Analysis. *Academic Radiology*, 2001; 8: 616-622.
47. OBUMRM. Available at: <http://www.bio.ri.ccf.org/OBUMRM/OBUMRM.html>
48. MATLAB The Language of Technical Computing Version 7.6.0.324 (R2008a).
49. IBM SPSS Statistics. Version 19.0 Trial Software, Chicago: SPSS Inc., 2010.

50. **Gallas BD., Pennello GA, Myres KJ.** Multireader Multicase Variance Analysis for Binary Data. *Virtual Journal for Biomedical Optics*, 2007;24(12):70-80.

EKLER

EK I: Değerlendiriciler Arasında Düşük /Yüksek Uyum İçin Matlab Paket Programı Kodları

```
function dataSimRunExeAuto2()

%----Similasyon icin parametreler aliniyor----%

Deg = input('Degerlendirici sayisini giriniz: ');
HasG = input('Hasta grubu sayisini giriniz: ');
KonG = input('Kontrol grubu sayisini giriniz: ');
Ted = input('Tedavi sayisini giriniz: ');

disp('Hasta grubu icin');
basH = input('Degerlendirme icin baslangic degerini giriniz: ');
bitH = input('Degerlendirme icin bitis degerini giriniz: ');

disp('Kontrol grubu icin');
basK = input('Degerlendirme icin baslangic degerini giriniz: ');
bitK = input('Degerlendirme icin bitis degerini giriniz: ');

disp('Uyum oranini giriniz');
Uyum = input('(Rasgele dagilim icin oran sifir(0) girilmelidir): ');
Uyum = Uyum / 100;

SimN = input('Similasyonun tekrar sayisini giriniz: ');
if SimN>9999
    warning('Similasyon sayisi 9999 den fazla olamaz');
    finish
end

sayac1=0;
sayac2=0;
sayac3=0;
sayac4=0;

%----Veri Uretimi Basliyor----%
for k=1:SimN

    fid = fopen('a_temp.txt','w');

    fid2 = fopen('veriSimAll.txt','a');
    if k>1
        fprintf(fid2,'\r\n');
    end

    fprintf(fid,'LABMRC version of SAMPLE_IN\r\n');
    fprintf(fid,'Değerlendirici %d\r\n'. 1);

    fprintf(fid2,'LABMRC version of SAMPLE_IN\r\n');
    fprintf(fid2,'Değerlendirici %d\r\n'. 1);

    for i = 1:Ted
        fprintf(fid,'"Treat %d"'. i);
```

```

fprintf(fid2.'Treat %d'. i);
if i ~= Ted
    fprintf(fid.'\r\t');
    fprintf(fid2.'\r\t');
else
    fprintf(fid.'\r\n');
    fprintf(fid2.'\r\n');
end
end
for j = 1:Ted
    fprintf(fid.'S');
    fprintf(fid2.'S');
    if j ~= Ted
        fprintf(fid.'\r\t');
        fprintf(fid2.'\r\t');
    else
        fprintf(fid.'\r\n');
        fprintf(fid2.'\r\n');
    end
end
end

ilk_degH = randint(Ted.HasG.[basH.bitH]);
ilk_degK = randint(Ted.KonG.[basK.bitK]);

yaz_ilkDegH = ilk_degH';
yaz_ilkDegK = ilk_degK';

for s = 1:HasG
    fprintf(fid.' %d'. yaz_ilkDegH(s:));
    fprintf(fid.'\r\n');
    fprintf(fid2.' %d'. yaz_ilkDegH(s:));
    fprintf(fid2.'\r\n');
end

fprintf(fid.'*\r\n');
fprintf(fid2.'*\r\n');

for t = 1:KonG
    fprintf(fid.' %d'. yaz_ilkDegK(t:));
    fprintf(fid.'\r\n');
    fprintf(fid2.' %d'. yaz_ilkDegK(t:));
    fprintf(fid2.'\r\n');
end

fprintf(fid.'*\r\n');
fprintf(fid2.'*\r\n');

UyumsuzH = HasG - round(HasG * Uyum);
UyumsuzK = KonG - round(KonG * Uyum);

fprintf(fid.'Değerlendirici %d\r\n'. 2);
fprintf(fid2.'Değerlendirici %d\r\n'. 2);

for i = 2:Deg

    degH = ilk_degH;
    degK = ilk_degK;

```

```

indexH = randint(Ted. UyumsuzH. [1.HasG]);
indexK = randint(Ted. UyumsuzK. [1.KonG]);

for s = 1:Ted
    for t = 1:UyumsuzH
        degH(s.indexH(s.t)) = randint(1.1.[basH.bitH]);
    end
end

for s = 1:Ted
    for t = 1:UyumsuzK
        degK(s.indexK(s.t)) = randint(1.1.[basK.bitK]);
    end
end

yaz_degH = degH';
yaz_degK = degK';

for s = 1:HasG
    fprintf(fid. ' %d'. yaz_degH(s.));
    fprintf(fid. '\r\n');
    fprintf(fid2. ' %d'. yaz_degH(s.));
    fprintf(fid2. '\r\n');
end

fprintf(fid. '*\r\n');
fprintf(fid2. '*\r\n');

for t = 1:KonG
    fprintf(fid. ' %d'. yaz_degK(t.));
    fprintf(fid. '\r\n');
    fprintf(fid2. ' %d'. yaz_degK(t.));
    fprintf(fid2. '\r\n');
end

fprintf(fid. '*\r\n');
fprintf(fid2. '*\r\n');

if(i~=Deg)
    fprintf(fid. 'Değerlendirici %d\r\n'. i+1);
    fprintf(fid2. 'Değerlendirici %d\r\n'. i+1);
else
    fprintf(fid. '#');
    fprintf(fid2. '#');
end

end

fclose(fid);
fclose(fid2);

%-----
%Exe uygulaması basılıyor...

robot = java.awt.Robot;

```

```

!LABMRMC.exe &
robot.keyPress(java.awt.event.KeyEvent.VK_Y);
pause(.1)
robot.keyPress(java.awt.event.KeyEvent.VK_ENTER);
pause(.3)

robot.keyPress(java.awt.event.KeyEvent.VK_TAB);
pause(.1)
robot.keyPress(java.awt.event.KeyEvent.VK_T);
robot.keyPress(java.awt.event.KeyEvent.VK_ENTER);
for i=1:6
robot.keyPress(java.awt.event.KeyEvent.VK_TAB);
pause(.1)
end
robot.keyPress(java.awt.event.KeyEvent.VK_A);
pause(.1)
robot.keyPress(java.awt.event.KeyEvent.VK_ENTER);
pause(.3)

robot.keyPress(java.awt.event.KeyEvent.VK_Y);
pause(.1)
robot.keyPress(java.awt.event.KeyEvent.VK_ENTER);
pause(.3)
robot.keyPress(java.awt.event.KeyEvent.VK_S);
robot.keyPress(java.awt.event.KeyEvent.VK_O);
robot.keyPress(java.awt.event.KeyEvent.VK_N);
robot.keyPress(java.awt.event.KeyEvent.VK_U);
robot.keyPress(java.awt.event.KeyEvent.VK_C);

sayac1 = sayac1+1;

if sayac1==10
    sayac2 = sayac2+1;
    sayac1 = 0;
end
if sayac2==10
    sayac3 = sayac3+1;
    sayac2 = 0;
end
if sayac3==10
    sayac4 = sayac4+1;
    sayac3 = 0;
end

eval(['robot.keyPress(java.awt.event.KeyEvent.VK_' .num2str(sayac4) .')'
])
eval(['robot.keyPress(java.awt.event.KeyEvent.VK_' .num2str(sayac3) .')'
])
eval(['robot.keyPress(java.awt.event.KeyEvent.VK_' .num2str(sayac2) .')'
])
eval(['robot.keyPress(java.awt.event.KeyEvent.VK_' .num2str(sayac1) .')'
])

pause(.1)

robot.keyPress(java.awt.event.KeyEvent.VK_ENTER);
robot.keyRelease(java.awt.event.KeyEvent.VK_ENTER);

```

```
pause(2)

robot.keyPress(java.awt.event.KeyEvent.VK_Q);
pause(.1)
robot.keyPress(java.awt.event.KeyEvent.VK_ENTER);
pause(.3)

robot.keyPress(java.awt.event.KeyEvent.VK_ALT);
robot.keyPress(java.awt.event.KeyEvent.VK_F4);

robot.keyRelease(java.awt.event.KeyEvent.VK_ALT);
robot.keyRelease(java.awt.event.KeyEvent.VK_F4);

pause(.1)
robot.keyPress(java.awt.event.KeyEvent.VK_TAB);
pause(.1)
robot.keyPress(java.awt.event.KeyEvent.VK_ENTER);
pause(.1)
%-----
clear i lett robot;

end
```


EK II: Tanı Testleri Arasında Güçlü/Zayıf Korelasyon İçin Matlab Paket Programı Kodları

```
function dataSimRunExeAuto()

%----Similasyon icin parametreler aliniyor----%

Deg = input('Degerlendirici sayisini giriniz: ');
HasG = input('Hasta grubu sayisini giriniz: ');
KonG = input('Kontrol grubu sayisini giriniz: ');
Ted = input('Tedavi sayisini giriniz: ');

disp('Hasta grubu icin');
basH = input('Degerlendirme icin baslangic degerini giriniz: ');
bitH = input('Degerlendirme icin bitis degerini giriniz: ');

disp('Kontrol grubu icin');
basK = input('Degerlendirme icin baslangic degerini giriniz: ');
bitK = input('Degerlendirme icin bitis degerini giriniz: ');

disp('Uyum oranini giriniz');
Uyum = input('(Rasgele dagilim icin oran sifir(0) girilmelidir): ');
Uyum = Uyum / 100;

SimN = input('Similasyonun tekrar sayisini giriniz: ');
if SimN>9999
    warning('Similasyon sayisi 9999 den fazla olamaz');
    finish
end

sayac1=0;
sayac2=0;
sayac3=0;
sayac4=0;

%----Veri Uretimi Basliyor----%
for k=1:SimN

    fid = fopen('a_temp.txt','w');

    fid2 = fopen('veriSimAll.txt','a');
    if k>1
        fprintf(fid2,'\r\n');
    end

    fprintf(fid,'LABMRC version of SAMPLE_IN\r\n');
    fprintf(fid,'Değerlendirici %d\r\n', 1);

    fprintf(fid2,'LABMRC version of SAMPLE_IN\r\n');
    fprintf(fid2,'Değerlendirici %d\r\n', 1);

    for i = 1:Ted
        fprintf(fid,'"Treat %d"', i);
        fprintf(fid2,'"Treat %d"', i);
        if i ~= Ted
```

```

        fprintf(fid. '\r\t');
        fprintf(fid2. '\r\t');
    else
        fprintf(fid. '\r\n');
        fprintf(fid2. '\r\n');
    end
end
for j = 1:Ted
    fprintf(fid. 'S');
    fprintf(fid2. 'S');
    if j ~= Ted
        fprintf(fid. '\r\t');
        fprintf(fid2. '\r\t');
    else
        fprintf(fid. '\r\n');
        fprintf(fid2. '\r\n');
    end
end
end

for i = 2:Deg+1

    ilk_satH = randint(1.HasG.[basH.bitH]);
    UyumsuzH = HasG - round(HasG * Uyum);
    uyumsuzMatH = randint(Ted-1.UyumsuzH.[basH.bitH]);
    uyumMatH = [];

    ilk_satK = randint(1.KonG.[basK.bitK]);
    UyumsuzK = KonG - round(KonG * Uyum);
    uyumsuzMatK = randint(Ted-1.UyumsuzK.[basK.bitK]);
    uyumMatK = [];

    for j=1:Ted-1
        uyumMatH = [uyumMatH; ilk_satH(:,(UyumsuzH+1):HasG)];
        uyumMatK = [uyumMatK; ilk_satK(:,(UyumsuzK+1):KonG)];
    end

    ekMatH = [uyumsuzMatH uyumMatH];
    hasMat = [ilk_satH; ekMatH]';

    ekMatK = [uyumsuzMatK uyumMatK];
    konMat = [ilk_satK; ekMatK]';

    for s = 1:HasG
        fprintf(fid. ' %d'. hasMat(s,:));
        fprintf(fid. '\r\n');
        fprintf(fid2. ' %d'. hasMat(s,:));
        fprintf(fid2. '\r\n');
    end
    fprintf(fid. '*\r\n');
    fprintf(fid2. '*\r\n');
    for t = 1:KonG
        fprintf(fid. ' %d'. konMat(t,:));
        fprintf(fid. '\r\n');
        fprintf(fid2. ' %d'. konMat(t,:));
        fprintf(fid2. '\r\n');
    end
end

```

```

fprintf(fid. '*\r\n');
fprintf(fid2. '*\r\n');
if(i~=Deg+1)
    fprintf(fid. 'Değerlendirici %d\r\n'. i);
    fprintf(fid2. 'Değerlendirici %d\r\n'. i);
else
    fprintf(fid. '#');
    fprintf(fid2. '#');
end

end

fclose(fid);
fclose(fid2);

%-----
%Exe uygulaması başlıyor...

robot = java.awt.Robot;
!LABMRMC.exe &
robot.keyPress( java.awt.event.KeyEvent.VK_Y );
pause(.1)
robot.keyPress( java.awt.event.KeyEvent.VK_ENTER );
pause(.3)

robot.keyPress( java.awt.event.KeyEvent.VK_TAB );
pause(.1)
robot.keyPress( java.awt.event.KeyEvent.VK_T );
robot.keyPress( java.awt.event.KeyEvent.VK_ENTER );
for i=1:6
robot.keyPress( java.awt.event.KeyEvent.VK_TAB );
pause(.1)
end
robot.keyPress( java.awt.event.KeyEvent.VK_A );
pause(.1)
robot.keyPress( java.awt.event.KeyEvent.VK_ENTER );
pause(.3)

robot.keyPress( java.awt.event.KeyEvent.VK_Y );
pause(.1)
robot.keyPress( java.awt.event.KeyEvent.VK_ENTER );
pause(.3)
robot.keyPress( java.awt.event.KeyEvent.VK_S );
robot.keyPress( java.awt.event.KeyEvent.VK_O );
robot.keyPress( java.awt.event.KeyEvent.VK_N );
robot.keyPress( java.awt.event.KeyEvent.VK_U );
robot.keyPress( java.awt.event.KeyEvent.VK_C );

sayac1 = sayac1+1;

if sayac1==10
    sayac2 = sayac2+1;
    sayac1 = 0;
end
if sayac2==10
    sayac3 = sayac3+1;
    sayac2 = 0;
end
end

```

```

if sayac3==10
    sayac4 = sayac4+1;
    sayac3 = 0;
end

eval(['robot.keyPress(java.awt.event.KeyEvent.VK_' .num2str(sayac4).' ')']
)
eval(['robot.keyPress(java.awt.event.KeyEvent.VK_' .num2str(sayac3).' ')']
)
eval(['robot.keyPress(java.awt.event.KeyEvent.VK_' .num2str(sayac2).' ')']
)
eval(['robot.keyPress(java.awt.event.KeyEvent.VK_' .num2str(sayac1).' ')']
)

pause(.1)

robot.keyPress(java.awt.event.KeyEvent.VK_ENTER);
robot.keyRelease(java.awt.event.KeyEvent.VK_ENTER);

pause(2)

robot.keyPress(java.awt.event.KeyEvent.VK_Q);
pause(.1)
robot.keyPress(java.awt.event.KeyEvent.VK_ENTER);
pause(.3)

robot.keyPress(java.awt.event.KeyEvent.VK_ALT);
robot.keyPress(java.awt.event.KeyEvent.VK_F4);

robot.keyRelease(java.awt.event.KeyEvent.VK_ALT);
robot.keyRelease(java.awt.event.KeyEvent.VK_F4);

pause(.1)
robot.keyPress(java.awt.event.KeyEvent.VK_TAB);
pause(.1)
robot.keyPress(java.awt.event.KeyEvent.VK_ENTER);
pause(.1)
%-----
clear i lett robot;

end

```

EK III: Değerlendiriciler Arasında Uyum Çok Yüksek Olduğu Durumdaki Varyans Analiz Tablosu Simülasyon Sonuçları

N=15	T=2 ROC=0.50 (D=2 ve Aralarındaki Uyum=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	35	4.0	4.2	3.4	3.7
Vaka	98.7	99.7	99.3	100	100
Değerlendirici	14	1.4	0.2	0.4	0.3
Vaka×Tanı Testi	98.2	99.7	99.3	100	100
Değerlendirici×Tanı Testi	11.2	0.5	0.2	0	0.3
Değerlendirici×Vaka	23	0.4	0.1	0	0
N=30	T=2 ROC=0.50 (D=2 ve Aralarındaki Uyum=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	50.6	3.5	3.2	2.6	2.9
Vaka	100	100	100	100	100
Değerlendirici	16.8	4.2	4.1	2.8	2.7
Vaka×Tanı Testi	99.6	100	100	100	100
Değerlendirici×Tanı Testi	16.7	4.3	4.8	4.0	4.0
Değerlendirici×Vaka	0	0	0	0	0
N=100	T=2 ROC=0.50 (D=2 ve Aralarındaki Uyum=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	52.4	3.3	4.1	3.0	2.8
Vaka	99.3	99.7	100	100	100
Değerlendirici	16.8	4.2	5.0	4.6	5.0
Vaka×Tanı Testi	99.4	99.7	100	100	100

Değerlendirici×Tanı Testi	16.7	5.1	5.0	6.8	4.5
Değerlendirici×Vaka	0	0	0	0	0
N=15	T=5 ROC=0.50 (D=2 ve Aralarındaki Uyum=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	55	3.2	3.5	2.2	3
Vaka	100	94.4	100	100	100
Değerlendirici	16.8	4.2	4.7	3.8	1.7
Vaka×Tanı Testi	100	94.4	100	100	100
Değerlendirici×Tanı Testi	25.3	0.3	2.4	3.6	0
Değerlendirici×Vaka	0	0	0	0	0
N=30	T=5 ROC=0.50 (D=2 ve Aralarındaki Uyum=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	56.1	3.3	2.8	2.5	3.9
Vaka	100	100	100	100	100
Değerlendirici	36.5	4.7	4.8	5.1	5.0
Vaka×Tanı Testi	100	100	100	100	100
Değerlendirici×Tanı Testi	38.5	3.3	3.0	1.2	3.9
Değerlendirici×Vaka	0	0	0	0	0
N=100	T=5 ROC=0.50 (D=2 ve Aralarındaki Uyum=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	55.9	3.5	3.6	3.3	2.7
Vaka	99.7	99.9	99.9	100	100
Değerlendirici	39.6	4.7	5.3	4.6	5.2

Vaka×Tanı Testi	100	99.9	99.9	100	100
Değerlendirici×Tanı Testi	41.7	4.7	4.6	3.0	5.7
Değerlendirici×Vaka	0	0	0	0	0
N=15	T=7 ROC=0.50 (D=2 ve Aralarındaki Uyum=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	51.2	2.9	2.7	3	2.2
Vaka	100	99.6	100	99.9	100
Değerlendirici	19.6	2.7	2.9	2.1	2.5
Vaka×Tanı Testi	100	99.6	100	99.9	100
Değerlendirici×Tanı Testi	27.1	0.3	0	0.1	2
Değerlendirici×Vaka	0	0	0	0	0
N=30	T=7 ROC=0.50 (D=2 ve Aralarındaki Uyum=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	55.4	2.7	3.4	2.8	3.0
Vaka	99.9	100	100	99.9	100
Değerlendirici	36	3.4	5.2	4.5	3.2
Vaka×Tanı Testi	99.9	100	100	99.9	100
Değerlendirici×Tanı Testi	38.5	3.8	3.3	2.4	2.3
Değerlendirici×Vaka	0	0	0	0	0
N=100	T=7 ROC=0.50 (D=2 ve Aralarındaki Uyum=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	59.5	2.6	3.7	3.1	2.6
Vaka	100	96.9	100	100	100

Değerlendirici	37.3	4.3	4.7	5.2	4.8
Vaka×Tanı Testi	100	96.9	100	100	100
Değerlendirici×Tanı Testi	39.0	5.3	4.1	4.8	5.1
Değerlendirici×Vaka	1	0	0	1	0
N=15	T=2 ROC=0.50 (D=5 ve Aralarındaki Uyum=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	20.6	2.3	2.3	3.4	3.7
Vaka	100	100	100	100	100
Değerlendirici	35.1	2.4	1.8	2.0	1.6
Vaka×Tanı Testi	100	100	100	100	100
Değerlendirici×Tanı Testi	35.9	2.8	2.1	1.2	1.8
Değerlendirici×Vaka	0	0	0	0	0
N=30	T=2 ROC=0.50 (D=5 ve Aralarındaki Uyum=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	30.4	1.9	0.9	3.4	2.8
Vaka	100	100	100	100	100
Değerlendirici	50.9	4.3	3.7	4.2	3.7
Vaka×Tanı Testi	100	100	100	100	100
Değerlendirici×Tanı Testi	51.1	4.9	6.5	4.1	3.1
Değerlendirici×Vaka	0	0	0	0	0
N=100	T=2 ROC=0.50 (D=5 ve Aralarındaki Uyum=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	29.1	3.8	0.2	0.2	2.9

Vaka	100	100	100	100	100
Değerlendirici	52.9	5.1	5.2	4.6	5.9
Vaka×Tanı Testi	100	100	100	100	100
Değerlendirici×Tanı Testi	52.6	5.5	5.5	6.3	6.0
Değerlendirici×Vaka	0	0	0	0	0
N=15	T=5 ROC=0.50 (D=5 ve Aralarındaki Uyum=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	18.6	4.0	2.5	2.2	2.3
Vaka	100	99.5	99.9	100	100
Değerlendirici	52.2	4.2	3.4	3.8	3.7
Vaka×Tanı Testi	100	99.5	99.9	100	100
Değerlendirici×Tanı Testi	63.9	0.7	1.2	1.0	0.6
Değerlendirici×Vaka	0	0	0	0	0
N=30	T=5 ROC=0.50 (D=5 ve Aralarındaki Uyum=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	10.8	3.2	3.3	3.0	2.6
Vaka	100	100	100	100	100
Değerlendirici	83.4	6.1	4.4	4.7	3.8
Vaka×Tanı Testi	100	100	100	100	100
Değerlendirici×Tanı Testi	88.1	4.7	3.7	3.4	3.2
Değerlendirici×Vaka	0	0	0	0	0
N=100	T=5 ROC=0.50 (D=5 ve Aralarındaki Uyum=0.90)				
	K=2	K=3	K=5	K=7	K=10

Tanı Testi	Programın Kapasitesi Yeterli Değil				
Vaka					
Değerlendirici					
Vaka×Tanı Testi					
Değerlendirici×Tanı Testi					
Değerlendirici×Vaka					
N=15	T=7 ROC=0.50 (D=5 ve Aralarındaki Uyum=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	20.4	2.5	2.9	2.6	3.5
Vaka	100	100	100	99.8	99.9
Değerlendirici	51.2	4.1	2.2	4.4	2.5
Vaka×Tanı Testi	100	100	100	99.8	99.9
Değerlendirici×Tanı Testi	61.1	1.9	1.1	1.1	0.3
Değerlendirici×Vaka	0	0	0	0	0
N=30	T=7 ROC=0.50 (D=5 ve Aralarındaki Uyum=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	13.4	2.6	3.1	3	2.8
Vaka	100	100	100	100	100
Değerlendirici	79.1	5.8	5.4	5.0	4.8
Vaka×Tanı Testi	100	100	100	100	100
Değerlendirici×Tanı Testi	84.3	4.4	4.1	3.1	3.1
Değerlendirici×Vaka	0	0	0	0	0
N=100	T=7 ROC=0.50 (D=5 ve Aralarındaki Uyum=0.90)				

	K=2	K=3	K=5	K=7	K=10
Tanı Testi	Programın Kapasitesi Yeterli Değil				
Vaka					
Değerlendirici					
Vaka×Tanı Testi					
Değerlendirici×Tanı Testi					
Değerlendirici×Vaka					
N=15	T=2 ROC=0.50 (D=7 ve Aralarındaki Uyum=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	15.6	3.8	4.3	2.6	4.1
Vaka	99.8	99.9	99.2	45.8	100
Değerlendirici	47.3	2.3	1.7	0.8	1.1
Vaka×Tanı Testi	99.8	99.9	99.2	45.8	100
Değerlendirici×Tanı Testi	46.9	2.9	1.9	0.9	1.7
Değerlendirici×Vaka	0	0	0	0	0
N=30	T=2 ROC=0.50 (D=7 ve Aralarındaki Uyum=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	18.9	2.5	4.2	2.4	2.1
Vaka	100	100	100	100	100
Değerlendirici	70.2	4.0	4.9	3.8	4
Vaka×Tanı Testi	100	100	100	100	100
Değerlendirici×Tanı Testi	70.4	4.8	3.8	3.4	3.6
Değerlendirici×Vaka	0	0	0	0	0

N=100	T=2 ROC=0.50 (D=7 ve Aralarındaki Uyum=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	20.4	2.3	3.2	3.9	2.4
Vaka	100	99.8	100	100	100
Değerlendirici	69.2	6.5	5.7	3.7	6.3
Vaka×Tanı Testi	100	99.8	100	100	100
Değerlendirici×Tanı Testi	69.3	5.7	4.5	5.8	6.6
Değerlendirici×Vaka	0	0	0	0	0
N=15	T=5 ROC=0.50 (D=7 ve Aralarındaki Uyum=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	12.0	3.5	3.5	2.0	3.0
Vaka	100	100	100	100	100
Değerlendirici	68.6	3.0	3.2	0.4	2.7
Vaka×Tanı Testi	100	100	100	100	100
Değerlendirici×Tanı Testi	77.2	1.4	0.5	0	0.6
Değerlendirici×Vaka	0.0	0	0	0	0
N=30	T=5 ROC=0.50 (D=7 ve Aralarındaki Uyum=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	4.3	2.7	3.0	2.5	2.2
Vaka	100	100	100	100	100
Değerlendirici	89.4	4.1	5.3	4.7	4.3
Vaka×Tanı Testi	100	100	100	100	100
Değerlendirici×Tanı Testi	94.5	3.0	2.9	3.4	2.9

Değerlendirici×Vaka	0	0	0	0	0
N=100	T=5 ROC=0.50 (D=7 ve Aralarındaki Uyum=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	Programın Kapasitesi Yeterli Değil				
Vaka					
Değerlendirici					
Vaka×Tanı Testi					
Değerlendirici×Tanı Testi					
Değerlendirici×Vaka					
N=15	T=7 ROC=0.50 (D=7 ve Aralarındaki Uyum=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	14	3.4	3.2	3.5	2.4
Vaka	100	100	100	100	99.9
Değerlendirici	66.4	2.1	2.5	2.6	3.1
Vaka×Tanı Testi	100	100	100	100	99.9
Değerlendirici×Tanı Testi	74.1	1.7	0.4	0.3	0.4
Değerlendirici×Vaka	0	0	0	0	0
N=30	T=7 ROC=0.50 (D=7 ve Aralarındaki Uyum=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	4.4	2.7	3.0	2.7	2.1
Vaka	100	100	100	100	89.6
Değerlendirici	89.2	4.4	5.4	4.2	3.5
Vaka×Tanı Testi	100	100	100	100	89.6
Değerlendirici×Tanı	94.8	42	3.1	2.8	1.4

Testi					
Değerlendirici×Vaka	0	0	0	0	0
N=100	T=7 ROC=0.50 (D=7 ve Aralarındaki Uyum=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	Programın Kapasitesi Yeterli Değil				
Vaka					
Değerlendirici					
Vaka×Tanı Testi					
Değerlendirici×Tanı Testi					
Değerlendirici×Vaka					

EK IV: Değerlendiriciler Arasında Uyum Çok Düşük Olduğu Durumda Varyans Analiz Tablosu Simülasyon Sonuçları

N=15	T=2 ROC=0.50 (D=2 ve Aralarındaki Uyum=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	11.1	3.3	2.7	3.9	2.8
Vaka	75.7	67	70.8	70.8	69.5
Değerlendirici	39.3	4.5	4.5	4.2	3.4
Vaka×Tanı Testi	60.1	67	68.8	69.4	71.3
Değerlendirici×Tanı Testi	38.1	3.2	3.6	3.5	3.2
Değerlendirici×Vaka	0.3	0	0	0	0
N=30	T=2 ROC=0.50 (D=2 ve Aralarındaki Uyum=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	16.5	2.9	2.4	3.2	3.6
Vaka	83.2	92.3	93.4	93	89.9
Değerlendirici	56.7	4.2	3.6	3.3	4.2
Vaka×Tanı Testi	77.3	93.2	92.3	91.7	91.5
Değerlendirici×Tanı Testi	54.1	4.1	4.5	3.8	5.3
Değerlendirici×Vaka	0	0	0	0	0
N=100	T=2 ROC=0.50 (D=2 ve Aralarındaki Uyum=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	18.1	4.6	2.9	4.3	2.3
Vaka	92.3	100	100	100	100
Değerlendirici	64.5	5.6	4.2	1.6	4.6
Vaka×Tanı Testi	90.9	100	100	100	100

Değerlendirici×Tanı Testi	64.5	6.3	5.1	6.5	5.9
Değerlendirici×Vaka	0	0	0	0	0
N=15	T=5 ROC=0.50 (D=2 ve Aralarındaki Uyum=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	8.2	3.4	1.9	2.0	2.6
Vaka	94.9	79.0	82.9	82.3	84.1
Değerlendirici	43.8	4.6	5.0	4.1	4.0
Vaka×Tanı Testi	93.1	98.7	99.4	99.3	99.5
Değerlendirici×Tanı Testi	76.6	4.5	3.2	3.0	2.6
Değerlendirici×Vaka	0	0	0	0	0
N=30	T=5 ROC=0.50 (D=2 ve Aralarındaki Uyum=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	5.3	2.4	3.4	2.4	2.2
Vaka	91.2	96.7	97.2	96.9	96.7
Değerlendirici	58.9	4.8	4.0	4.7	5.0
Vaka×Tanı Testi	95.4	100	99.9	100	100
Değerlendirici×Tanı Testi	86.7	3.7	4.8	4.3	3.8
Değerlendirici×Vaka	0	0	0	0	0
N=100	T=5 ROC=0.50 (D=2 ve Aralarındaki Uyum=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	3.8	2.5	2.2	2.7	2.6
Vaka	89.6	99.6	100	100	100
Değerlendirici	81.4	5.4	5.8	3.2	5.2

Vaka×Tanı Testi	96.8	99.6	100	100	100
Değerlendirici×Tanı Testi	96	4.6	3.5	3.9	4.7
Değerlendirici×Vaka	0	0	0	0	0
N=15	T=7 ROC=0.50 (D=2 ve Aralarındaki Uyum=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	7.9	2.3	2.3	1.5	2.3
Vaka	92.7	79.6	84.3	83.9	83.2
Değerlendirici	39.6	3.8	3.9	3.7	4.4
Vaka×Tanı Testi	92.1	99	99.6	99.3	99
Değerlendirici×Tanı Testi	75.7	6.0	3.2	3.5	2.8
Değerlendirici×Vaka	0	0	0	0	0
N=30	T=7 ROC=0.50 (D=2 ve Aralarındaki Uyum=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	2.8	2.9	2.8	3.4	2.4
Vaka	90.8	95.6	97.1	97.5	96.8
Değerlendirici	54.9	4.3	5.2	4.6	4.6
Vaka×Tanı Testi	94.3	99.8	100	100	100
Değerlendirici×Tanı Testi	88	3.6	3.8	3.7	3.4
Değerlendirici×Vaka	0	0	0	0	0
N=100	T=7 ROC=0.50 (D=2 ve Aralarındaki Uyum=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	4.4	1.9	2.7	3.6	2.7
Vaka	90.9	98	99.8	100	99.9

Değerlendirici	81.2	5.6	4.3	4.9	5.0
Vaka×Tanı Testi	97.3	98	99.8	100	99.9
Değerlendirici×Tanı Testi	95.3	5.2	4.6	4.4	4.3
Değerlendirici×Vaka	0	0	0	0	0
N=15	T=2 ROC=0.50 (D=5 ve Aralarındaki Uyum=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	2.4	3.6	3.6	3.3	2.3
Vaka	96.8	93.2	94.8	94.4	94.3
Değerlendirici	75.7	4.2	3.4	2.7	2.8
Vaka×Tanı Testi	82.2	92.4	94.5	95.4	94.6
Değerlendirici×Tanı Testi	77.9	4.2	2.8	3.7	4.1
Değerlendirici×Vaka	0	0	0	0	0
N=30	T=2 ROC=0.50 (D=5 ve Aralarındaki Uyum=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	1.0	2.8	2.3	3.1	2.3
Vaka	96.6	98.0	99.8	99.5	99.8
Değerlendirici	91.3	4.8	6.0	4.8	5.0
Vaka×Tanı Testi	88.6	97.9	99.7	99.4	99.8
Değerlendirici×Tanı Testi	89.6	5.9	4.9	3.7	2.5
Değerlendirici×Vaka	0	0	0	0	0
N=100	T=2 ROC=0.50 (D=5 ve Aralarındaki Uyum=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0	3.3	4.5	2.9	3.8

Vaka	94.8	100	100	100	100
Değerlendirici	100	4.6	6.1	4.9	5.1
Vaka×Tanı Testi	97.8	100	100	100	100
Değerlendirici×Tanı Testi	100	6.5	4.3	5.2	2.9
Değerlendirici×Vaka	0	0	0	0	0
N=15	T=5 ROC=0.50 (D=5 ve Aralarındaki Uyum=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0	2.7	2.0	2.1	2.5
Vaka	100	94.6	95.5	96	97.1
Değerlendirici	79.6	5.4	3.7	3.2	4.6
Vaka×Tanı Testi	99.8	100	100	100	100
Değerlendirici×Tanı Testi	98.7	4.4	2.8	2.7	2.3
Değerlendirici×Vaka	0	0	0	0	0
N=30	T=5 ROC=0.50 (D=5 ve Aralarındaki Uyum=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0	3.7	2.1	2.3	3.1
Vaka	99.8	99.9	99.8	99.9	100
Değerlendirici	93	5.3	4.9	3.2	4.9
Vaka×Tanı Testi	100	100	100	100	100
Değerlendirici×Tanı Testi	99.9	4.0	4.6	3.2	3.4
Değerlendirici×Vaka	0	0	0	0	0
N=100	T=5 ROC=0.50 (D=5 ve Aralarındaki Uyum=0.10)				
	K=2	K=3	K=5	K=7	K=10

Tanı Testi	Programın Kapasitesi Yeterli Değil				
Vaka					
Değerlendirici					
Vaka×Tanı Testi					
Değerlendirici×Tanı Testi					
Değerlendirici×Vaka					
N=15	T=7 ROC=0.50 (D=5 ve Aralarındaki Uyum=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0	2.7	2.2	2.4	1.9
Vaka	100	95.2	96.9	96.7	96.3
Değerlendirici	78.7	4.3	3.8	3.6	2.2
Vaka×Tanı Testi	99.7	100	100	100	100
Değerlendirici×Tanı Testi	99	4.2	2.9	2.5	1.7
Değerlendirici×Vaka	0	0	0	0	0
N=30	T=7 ROC=0.50 (D=5 ve Aralarındaki Uyum=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0	2.7	1.7	2.6	1.8
Vaka	99.5	100	99.9	100	99.9
Değerlendirici	92.9	5.6	4.2	4	4.9
Vaka×Tanı Testi	100	100	100	100	100
Değerlendirici×Tanı Testi	99.9	5.1	4.2	3.3	3.3
Değerlendirici×Vaka	0	0	0	0	0
N=100	T=7 ROC=0.50 (D=5 ve Aralarındaki Uyum=0.10)				

	K=2	K=3	K=5	K=7	K=10
Tanı Testi	Programın Kapasitesi Yeterli Değil				
Vaka					
Değerlendirici					
Vaka×Tanı Testi					
Değerlendirici×Tanı Testi					
Değerlendirici×Vaka					
N=15	T=2 ROC=0.50 (D=7 ve Aralarındaki Uyum=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0.7	2.1	2.6	2.6	2.7
Vaka	99.8	85.9	99.3	97.4	97.3
Değerlendirici	87.5	3.7	3.8	2.7	3.6
Vaka×Tanı Testi	90.9	86.3	98.4	97.6	98.2
Değerlendirici×Tanı Testi	86.5	4.4	3.3	3.4	2.7
Değerlendirici×Vaka	0	0	0	0	0
N=30	T=2 ROC=0.50 (D=7 ve Aralarındaki Uyum=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0.2	2.9	2.2	2.5	3.2
Vaka	99.5	99.8	100	100	100
Değerlendirici	96.9	4.6	3.6	4.2	4.9
Vaka×Tanı Testi	96.4	99.8	100	100	100
Değerlendirici×Tanı Testi	97.2	3.0	5.3	3.8	4.5
Değerlendirici×Vaka	0	0	0	0	0

N=100	T=2 ROC=0.50 (D=7 ve Aralarındaki Uyum=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0	2.8	2.4	1.8	2.0
Vaka	99.6	99.2	100	100	100
Değerlendirici	100	4.3	5.8	5.4	4.7
Vaka×Tanı Testi	99.6	99.2	100	100	100
Değerlendirici×Tanı Testi	99.6	5.1	3.6	3.4	4.7
Değerlendirici×Vaka	0	0	0	0	0
N=15	T=5 ROC=0.50 (D=7 ve Aralarındaki Uyum=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0	2.9	2.0	2.3	2.2
Vaka	100	98.1	98.2	98.6	99.0
Değerlendirici	89.5	5.5	3.1	2.2	2.2
Vaka×Tanı Testi	100	100	100	100	100
Değerlendirici×Tanı Testi	99.9	3.9	2.4	2.3	1.5
Değerlendirici×Vaka	0	0	0	0	0
N=30	T=5 ROC=0.50 (D=7 ve Aralarındaki Uyum=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0	2.8	3.6	2.1	2
Vaka	100	99.9	100	99.8	93.9
Değerlendirici	98.54	3.9	4.6	4.7	4.6
Vaka×Tanı Testi	100	99.9	100	99.8	93.9
Değerlendirici×Tanı Testi	100	3.1	3.5	3.2	2.7

Değerlendirici×Vaka	0	0	0	0	0
N=100	T=5 ROC=0.50 (D=7 ve Aralarındaki Uyum=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	Programın Kapasitesi Yeterli Değil				
Vaka					
Değerlendirici					
Vaka×Tanı Testi					
Değerlendirici×Tanı Testi					
Değerlendirici×Vaka					
N=15	T=7 ROC=0.50 (D=7 ve Aralarındaki Uyum=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0	3.1	2.7	2.7	2.8
Vaka	99.8	97.2	99	98.6	99.4
Değerlendirici	88	2.9	2.6	3.8	2.1
Vaka×Tanı Testi	99.8	99.8	100	100	100
Değerlendirici×Tanı Testi	99.6	4.4	2.3	2.7	1.4
Değerlendirici×Vaka	0	0	0	0	0
N=30	T=7 ROC=0.50 (D=7 ve Aralarındaki Uyum=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0	3.5	2.0	3.1	1.6
Vaka	100	99.2	100	100	74.5
Değerlendirici	98	5.0	4.7	3.2	3.0
Vaka×Tanı Testi	100	99.2	100	100	74.5
Değerlendirici×Tanı	100	5.0	3.7	3.4	2.3

Testi					
Değerlendirici×Vaka	0	0	0	0	0
N=100	T=7 ROC=0.50 (D=7 ve Aralarındaki Uyum=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	Programın Kapasitesi Yeterli Değil				
Vaka					
Değerlendirici					
Vaka×Tanı Testi					
Değerlendirici×Tanı Testi					
Değerlendirici×Vaka					

EK V: Tanı testleri Arasında Korelasyon Çok Güçlü Olduğu Durumda Varyans Analiz Tablosu Simülasyon Sonuçları

N=15	D=2 ROC=0.50 (T=2 ve Aralarındaki Korelasyon=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0.2	0	0	0	0
Vaka	16.1	4.8	4.4	5.2	5.9
Değerlendirici	40.9	5.9	5.5	4.5	3.9
Vaka×Tanı Testi	20.2	26.4	32.7	30.2	33.5
Değerlendirici×Tanı Testi	3.0	0	0	0	0
Değerlendirici×Vaka	54.6	30.7	28.1	30.4	26.4
N=30	D=2 ROC=0.50 (T=2 ve Aralarındaki Korelasyon=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0	0.9	0.5	1.5	0.6
Vaka	17.1	5.9	5.8	6.2	4
Değerlendirici	60.8	4.8	3.3	4.6	3.5
Vaka×Tanı Testi	32.3	31.9	29.8	29.6	30.3
Değerlendirici×Tanı Testi	0.1	3.2	3	2.5	1.7
Değerlendirici×Vaka	48.5	31.8	30	34.7	30.4
N=100	D=2 ROC=0.50 (T=2 ve Aralarındaki Korelasyon=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0	1.3	1.9	2.2	2.1
Vaka	14.9	5.5	5.1	5.0	4.5
Değerlendirici	65.4	5.2	5.3	4.3	4.1
Vaka×Tanı Testi	41.2	32.1	31.8	32.4	30.8

Değerlendirici×Tanı Testi	0	5.2	5.0	4.4	4.6
Değerlendirici×Vaka	48.4	32.2	31.3	32.0	33.1
N=15	D=5 ROC=0.50 (T=2 ve Aralarındaki Korelasyon=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0	0	0	0	0
Vaka	31.7	5.4	4.8	4.3	5.9
Değerlendirici	78.7	4.7	3.6	3.1	4.2
Vaka×Tanı Testi	25.5	24.6	24.9	26.0	23.0
Değerlendirici×Tanı Testi	3.0	0	0	0	0
Değerlendirici×Vaka	59.1	34.8	33.1	37.1	38.2
N=30	D=5 ROC=0.50 (T=2 ve Aralarındaki Korelasyon=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0	0.8	0.5	0.6	0.6
Vaka	30.9	5.6	6.6	4.9	6.2
Değerlendirici	89.6	4.5	3.7	5.4	3.1
Vaka×Tanı Testi	29.3	24.5	26.6	25.8	25.4
Değerlendirici×Tanı Testi	0.5	3.8	2.7	1.26	3.6
Değerlendirici×Vaka	52.9	31.4	33.3	33.1	34
N=100	D=5 ROC=0.50 (T=2 ve Aralarındaki Korelasyon=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0	1.9	1.7	1.7	1.4
Vaka	30.2	6.3	5.9	6.3	5.5
Değerlendirici	97.3	5.1	5.7	5.2	5.2

Vaka×Tanı Testi	28	27.1	27.8	28.1	30.0
Değerlendirici×Tanı Testi	0	4.2	5	5.2	4.0
Değerlendirici×Vaka	53.4	29.9	30.5	31.7	31

N=15	D=7 ROC=0.50 (T=2 ve Aralarındaki Korelasyon=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0	0	0	0	0
Vaka	41.6	5.4	5.8	6.0	6.0
Değerlendirici	86.0	4.6	3.7	2.9	3.5
Vaka×Tanı Testi	26.2	22.4	26.5	25.1	23.0
Değerlendirici×Tanı Testi	41	0.3	0.2	0	0
Değerlendirici×Vaka	55.4	37.8	35.5	39.2	38.1
N=30	D=7 ROC=0.50 (T=2 ve Aralarındaki Korelasyon=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0	0.6	0.4	1.0	0.39
Vaka	34.9	5.2	5.8	6.2	5.5
Değerlendirici	94.4	5.1	3.4	4.8	2.3
Vaka×Tanı Testi	28.1	26.1	23.4	27.2	26.3
Değerlendirici×Tanı Testi	10	3.7	2.1	20.1	2.6
Değerlendirici×Vaka	53	31.6	33.3	31.5	32.7
N=100	D=7 ROC=0.50 (T=2 ve Aralarındaki Korelasyon=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0	1.3	1.6	1.7	0.9

Vaka	29	6.3	6.2	5	5.7
Değerlendirici	99.5	5.3	4.3	4.9	4.9
Vaka×Tanı Testi	26.5	25.5	25.8	29.7	26.3
Değerlendirici×Tanı Testi	0	4.7	4.4	4.2	4.1
Değerlendirici×Vaka	55.2	30.2	31.8	29.5	32.1
N=15	D=2 ROC=0.50 (T=5 ve Aralarındaki Korelasyon=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0.1	0	0.2	0.2	0
Vaka	24.4	11.9	11.9	14.2	12.7
Değerlendirici	39.1	5.6	4.5	3.9	3.2
33.1	32.5	31.7	31.6	32.2	33.7
Değerlendirici×Tanı Testi	2.6	4	0	0.2	0.2
Değerlendirici×Vaka	50.9	27.3	29.5	29.6	30.0
N=30	D=2 ROC=0.50 (T=5 ve Aralarındaki Korelasyon=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0	1.3	1.1	1.4	0.9
Vaka	20.4	12.8	10.9	12	12.3
Değerlendirici	65.4	5.5	4.6	4.3	4.6
33.1	32.3	31.7	32.8	31.7	31.7
Değerlendirici×Tanı Testi	0.4	2.6	1.9	2.1	1.5
Değerlendirici×Vaka	48.8	29.1	29.1	31.6	30.2
N=100	D=2 ROC=0.50 (T=5 ve Aralarındaki Korelasyon=0.90)				
	K=2	K=3	K=5	K=7	K=10

Tanı Testi	0	1.7	1.6	2.4	2.9
Vaka	19.1	11.5	10.8	11.6	18.5
Değerlendirici	60.5	5.8	5.2	5.5	5.8
Vaka×Tanı Testi	28.4	30.3	30.5	30.6	40.5
Değerlendirici×Tanı Testi	0	3.3	3.5	3.8	6.4
Değerlendirici×Vaka	51.5	32.2	31	29.5	43.6
N=15	D=5 ROC=0.50 (T=5 ve Aralarındaki Korelasyon=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0	0	0	0	0
Vaka	35.1	6.8	6.9	9.3	7.1
Değerlendirici	75.4	4.9	3.1	4.3	2.3
Vaka×Tanı Testi	29.7	23.4	29.2	29.2	29.3
Değerlendirici×Tanı Testi	6.3	0.6	0.4	0.1	0
Değerlendirici×Vaka	50.8	32.3	35.3	36.0	40.9
N=30	D=5 ROC=0.50 (T=5 ve Aralarındaki Korelasyon=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0	1.1	0.7	0.4	0.8
Vaka	28.1	7.2	7.1	8.3	8.8
Değerlendirici	90.6	5.1	4.4	3.5	4.5
Vaka×Tanı Testi	29.8	28.2	29.3	32.3	29.2
Değerlendirici×Tanı Testi	1.1	2.5	2.1	1.7	1.6
Değerlendirici×Vaka	27	30.9	31.4	30.8	32.3
N=100	D=5 ROC=0.50 (T=5 ve Aralarındaki Korelasyon=0.90)				

	K=2	K=3	K=5	K=7	K=10
Tanı Testi	Programın Kapasitesi Yeterli Değil				
Vaka					
Değerlendirici					
Vaka×Tanı Testi					
Değerlendirici×Tanı Testi					
Değerlendirici×Vaka					
N=15	D=7 ROC=0.50 (T=5 ve Aralarındaki Korelasyon=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0	1	0	0	0
Vaka	43.9	7.0	6.3	6.4	6.4
Değerlendirici	84.9	5.3	4.2	4	3.4
Vaka×Tanı Testi	29.6	24.0	27.4	27.5	26.7
Değerlendirici×Tanı Testi	7.5	0.7	0	0	0.2
Değerlendirici×Vaka	49.4	35.9	39.6	38.6	38.8
N=30	D=7 ROC=0.50 (T=5 ve Aralarındaki Korelasyon=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0	0.6	1.8	1.5	1.4
Vaka	32.3	6.9	6.4	6.8	7.2
Değerlendirici	94.2	4.5	3.8	3.2	4.0
Vaka×Tanı Testi	28.6	27.8	28.6	27.9	28.4
Değerlendirici×Tanı Testi	1.8	2.7	2.7	1.4	1.2
Değerlendirici×Vaka	50.9	32.3	33	34.8	33.1

N=100	D=7 ROC=0.50 (T=5 ve Aralarındaki Korelasyon=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	Programın Kapasitesi Yeterli Değil				
Vaka					
Değerlendirici					
Vaka×Tanı Testi					
Değerlendirici×Tanı Testi					
Değerlendirici×Vaka					
N=15	D=2 ROC=0.50 (T=7 ve Aralarındaki Korelasyon=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0	0	0	0	0
Vaka	24.0	11.3	14.2	13.6	13.9
Değerlendirici	40.4	5.9	3.9	4.3	3.6
Vaka×Tanı Testi	33.7	30.7	31.2	29.8	31.7
Değerlendirici×Tanı Testi	1.8	0.1	0.1	0	0
Değerlendirici×Vaka	50	29.9	27.8	31.9	31.8
N=30	D=2 ROC=0.50 (T=7 ve Aralarındaki Korelasyon=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	3.5	3.5	2.6	1.9	1.4
Vaka	45.7	34.7	28.9	19.6	10.4
Değerlendirici	61.7	4.4	4.6	3.5	4.6
Vaka×Tanı Testi	74.1	67.1	58.9	42.5	26.7
Değerlendirici×Tanı Testi	9.4	9.0	7.1	5.1	2.6

Değerlendirici×Vaka	79.3	68.7	58.5	43.2	26.4
N=100	D=2 ROC=0.50 (T=7 ve Aralarındaki Korelasyon=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0	1.8	1.9	1.7	1.5
Vaka	22.8	12.3	13.4	11.6	13.1
Değerlendirici	62.3	4.5	4.6	4.9	4.2
Vaka×Tanı Testi	4.8	29.2	32.2	28.3	30.9
Değerlendirici×Tanı Testi	0	3.7	4.1	2.9	3.7
Değerlendirici×Vaka	51.2	31.0	32.1	31.7	32.7
N=15	D=5 ROC=0.50 (T=7 ve Aralarındaki Korelasyon=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0	0	0	0	0
Vaka	36.2	7.8	6.9	6.3	8.3
Değerlendirici	77	4.6	3.7	3.7	4.4
Vaka×Tanı Testi	30.3	27.7	26.4	28.1	24.8
Değerlendirici×Tanı Testi	5.5	0.3	0	0	0.1
Değerlendirici×Vaka	50.5	35.5	36.3	37.2	37.3
N=30	D=5 ROC=0.50 (T=7 ve Aralarındaki Korelasyon=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0	0.9	1.3	1.1	1.0
Vaka	27.9	7.1	7.8	7.2	5.8
Değerlendirici	90.0	4.8	3.3	4.3	4.5
Vaka×Tanı Testi	27.9	25.5	26.5	26	25.8

Değerlendirici×Tanı Testi	0.4	2.8	2.5	2.4	1.6
Değerlendirici×Vaka	51.8	29.8	31.5	29.9	33.8
N=100	D=5 ROC=0.50 (T=7 ve Aralarındaki Korelasyon=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	Programın Kapasitesi Yeterli Değil				
Vaka					
Değerlendirici					
Vaka×Tanı Testi					
Değerlendirici×Tanı Testi					
Değerlendirici×Vaka					
N=15	D=7 ROC=0.50 (T=7 ve Aralarındaki Korelasyon=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0	0	0	0	0
Vaka	41.1	7.8	7.0	8.3	7.6
Değerlendirici	83.0	4.8	4.4	3.2	2.9
Vaka×Tanı Testi	27.5	24.5	29.7	27.4	27.7
Değerlendirici×Tanı Testi	8.4	0.9	0.2	0	0.2
Değerlendirici×Vaka	49.6	35.9	39.7	42.3	40.6
N=30	D=7 ROC=0.50 (T=7 ve Aralarındaki Korelasyon=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0	1.0	0.1	1.3	1.3
Vaka	37.2	8.17	6.9	6.2	5.8
Değerlendirici	94.9	6.1	5.2	3.9	3.6

Vaka×Tanı Testi	26.5	26.7	26.7	27.3	27.7
Değerlendirici×Tanı Testi	1.4	2.3	2.0	2.3	2.4
Değerlendirici×Vaka	52.8	30.1	31.9	32.2	32
N=100	D=7 ROC=0.50 (T=7 ve Aralarındaki Korelasyon=0.90)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	Programın Kapasitesi Yeterli Değil				
Vaka					
Değerlendirici					
Vaka×Tanı Testi					
Değerlendirici×Tanı Testi					
Değerlendirici×Vaka					

EK VI: Tanı Testleri Arasında Korelasyon Çok Zayıf Olduğu Durumdaki Varyans Analiz Tablosu Simülasyon Sonuçları

N=15	D=2 ROC=0.50 (T=2 ve Aralarındaki Korelasyon=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0	2.2	1.7	2.2	1.4
Vaka	18.4	7.2	4.8	5.1	5.8
Değerlendirici	41.6	5.5	3.7	3.3	4.2
Vaka×Tanı Testi	4.3	5.8	6.0	5.8	5.7
Değerlendirici×Tanı Testi	0.2	5.2	5.0	3.9	4.0
Değerlendirici×Vaka	19.1	1.9	1.0	1.5	1.3
N=30	D=2 ROC=0.50 (T=2 ve Aralarındaki Korelasyon=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0	2.1	2.7	2.3	2.6
Vaka	18	5.3	6.2	5.2	5.4
Değerlendirici	60.4	4.8	3.5	3.1	3.3
Vaka×Tanı Testi	4.6	5.7	5.7	6.4	6.0
Değerlendirici×Tanı Testi	0.5	6.3	5.2	4.2	5.1
Değerlendirici×Vaka	31	1.4	2.6	2.5	2.7
N=100	D=2 ROC=0.50 (T=2 ve Aralarındaki Korelasyon=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0.1	3.3	3.7	3.0	3.4
Vaka	14.9	4.5	4.0	4.8	4.3
Değerlendirici	63.2	4.5	5.0	4.4	5.5
Vaka×Tanı Testi	4.8	6.7	5.4	6.7	5.8

Değerlendirici×Tanı Testi	0.1	5.6	3.4	4.8	3.7
Değerlendirici×Vaka	36	7.4	10.2	8	10
N=15	D=5 ROC=0.50 (T=2 ve Aralarındaki Korelasyon=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0	1.6	2.0	1.4	0.2
Vaka	37.5	6.0	2.8	5.4	4.6
Değerlendirici	70.1	3.9	3.5	3.4	2.7
Vaka×Tanı Testi	4.6	6.6	5.7	5.1	5.1
Değerlendirici×Tanı Testi	0.2	4.4	3.6	3.7	3.5
Değerlendirici×Vaka	41.6	6.5	8.7	9.1	8.6
N=30	D=5 ROC=0.50 (T=2 ve Aralarındaki Korelasyon=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0	2.2	1.4	1.5	1.6
Vaka	33.8	5.3	5.4	5.8	6.5
Değerlendirici	89.9	5.2	3.3	3.2	3.1
Vaka×Tanı Testi	5.2	6.2	4.8	4.6	4.4
Değerlendirici×Tanı Testi	0.2	4.4	5.4	4.6	4.5
Değerlendirici×Vaka	48.5	9.1	11.3	12.7	10.6
N=100	D=5 ROC=0.50 (T=2 ve Aralarındaki Korelasyon=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0	2.7	2.8	1.9	2.2
Vaka	27.2	5.5	5.1	5.2	5.1
Değerlendirici	96.9	5.0	4.7	5.0	4.4

Vaka×Tanı Testi	5.1	5.8	7.2	6.5	6.0
Değerlendirici×Tanı Testi	0	4.5	4.0	4.7	4.3
Değerlendirici×Vaka	51.2	26.5	25.7	27.1	26.2
N=15	D=7 ROC=0.50 (T=2 ve Aralarındaki Korelasyon=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0	2.2	1.8	1.4	1.6
Vaka	51.2	6.5	5.3	4.3	5.8
Değerlendirici	86.6	4.7	3.0	2.6	3.1
Vaka×Tanı Testi	6.1	7.5	5.7	5.4	6.6
Değerlendirici×Tanı Testi	0.3	3.6	2.6	2.0	1.9
Değerlendirici×Vaka	50.1	11.8	11.4	13.3	12.6
N=30	D=7 ROC=0.50 (T=2 ve Aralarındaki Korelasyon=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0	2.8	2.0	2.8	2.8
Vaka	35.8	6.6	5.5	5.8	6.1
Değerlendirici	95.6	4.3	3.2	3.2	2.8
Vaka×Tanı Testi	5.5	5.8	6.0	6.2	5.9
Değerlendirici×Tanı Testi	0	4.5	4.7	4.8	3.8
Değerlendirici×Vaka	50.1	13.6	16.3	15.8	15.9
N=100	D=7 ROC=0.50 (T=2 ve Aralarındaki Korelasyon=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0	3.2	2.8	3.3	2.1
Vaka	29.5	5.5	5.3	5.3	4.5

Değerlendirici	99.2	4.8	4.3	4.7	5.2
Vaka×Tanı Testi	4.0	5.7	5.2	5.4	5.4
Değerlendirici×Tanı Testi	0	4.5	4.4	4.1	4.5
Değerlendirici×Vaka	52.3	27.1	30.4	29.7	28.7
N=15	D=2 ROC=0.50 (T=5 ve Aralarındaki Korelasyon=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0	1.4	1.4	1.5	1.5
Vaka	27.2	17.4	18.9	16.5	13.7
Değerlendirici	41.8	5.4	2.7	3.2	3.9
Vaka×Tanı Testi	3.7	6.0	6.1	6.0	5.5
Değerlendirici×Tanı Testi	0	4.1	3.8	3.4	3.4
Değerlendirici×Vaka	33.2	15.4	16.9	15.0	14.1
N=30	D=2 ROC=0.50 (T=5 ve Aralarındaki Korelasyon=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0.0	2.8	2.3	2.3	2.1
Vaka	26.6	18.1	19.1	17.9	17.9
Değerlendirici	63.2	4.9	4.5	5.8	4
Vaka×Tanı Testi	4.7	4.9	5.7	7	6.4
Değerlendirici×Tanı Testi	0.1	4.4	3.6	4.2	3.9
Değerlendirici×Vaka	40.5	16.3	16.8	18.6	17.7
N=100	D=2 ROC=0.50 (T=5 ve Aralarındaki Korelasyon=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0	3.5	1.9	0.2	3.0

Vaka	24.1	15.4	16.8	14.7	15.9
Değerlendirici	62.7	5.6	4.8	4.6	4.2
Vaka×Tanı Testi	4.6	5.8	5.2	6.3	5.8
Değerlendirici×Tanı Testi	0.0	3.8	4.6	3.9	3.9
Değerlendirici×Vaka	48.1	29.9	31.9	27.3	29.2
N=15	D=5 ROC=0.50 (T=5 ve Aralarındaki Korelasyon=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0	1.6	1.4	1.3	1.5
Vaka	43.9	12.3	10.6	12.3	12.7
Değerlendirici	76.6	3.4	3.9	3.6	3.3
Vaka×Tanı Testi	6.0	5.8	6.9	6.8	6.8
Değerlendirici×Tanı Testi	0	3.8	1.9	2.3	2.2
Değerlendirici×Vaka	43.2	23.8	28.3	33.4	35.3
N=30	D=5 ROC=0.50 (T=5 ve Aralarındaki Korelasyon=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0.0	1.4	1.9	2.4	1.6
Vaka	40.8	12.0	13.5	11.8	11.4
Değerlendirici	94	4.5	4.9	3.9	4
Vaka×Tanı Testi	5.5	5.5	6.2	7	7.3
Değerlendirici×Tanı Testi	0.0	5.5	4.2	3.6	3.1
Değerlendirici×Vaka	51.7	26.5	31.6	31.2	31.4
N=100	D=5 ROC=0.50 (T=5 ve Aralarındaki Korelasyon=0.10)				
	K=2	K=3	K=5	K=7	K=10

Tanı Testi	Programın Kapasitesi Yeterli Değil				
Vaka					
Değerlendirici					
Vaka×Tanı Testi					
Değerlendirici×Tanı Testi					
Değerlendirici×Vaka					
N=15	D=7 ROC=0.50 (T=5 ve Aralarındaki Korelasyon=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0	1.4	1.7	1.9	1.6
Vaka	60.2	9.8	13	11.1	10.4
Değerlendirici	89.9	4.6	3.1	2.4	3.0
Vaka×Tanı Testi	7.6	7.0	8.3	6.5	5.9
Değerlendirici×Tanı Testi	0.1	4.3	2.0	1.7	1.8
Değerlendirici×Vaka	49.2	30.4	31.6	37.7	34
N=30	D=7 ROC=0.50 (T=5 ve Aralarındaki Korelasyon=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0.0	2.3	2.3	2	1.4
Vaka	42.9	10.7	13.2	13.3	11
Değerlendirici	98.4	4.4	4.2	4.4	4.5
Vaka×Tanı Testi	6.3	5.6	6.2	5.8	5.8
Değerlendirici×Tanı Testi	0.0	4.2	4.4	3.7	3.8
Değerlendirici×Vaka	50.0	29.6	32.2	30.6	31.1
N=100	D=7 ROC=0.50 (T=5 ve Aralarındaki Korelasyon=0.10)				

	K=2	K=3	K=5	K=7	K=10
Tanı Testi	Programın Kapasitesi Yeterli Değil				
Vaka					
Değerlendirici					
Vaka×Tanı Testi					
Değerlendirici×Tanı Testi					
Değerlendirici×Vaka					
N=15	D=2 ROC=0.50 (T=7 ve Aralarındaki Korelasyon=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0.0	1.4	1.3	1.3	1.2
Vaka	30.5	14.1	18.1	17.4	16.6
Değerlendirici	46.6	3.7	3.7	3.6	3.6
Vaka×Tanı Testi	6.0	5.3	6.4	6.5	4.3
Değerlendirici×Tanı Testi	3.0	3.2	3.1	3.3	3.1
Değerlendirici×Vaka	37.6	13.8	14.6	16.9	12.4
N=30	D=2 ROC=0.50 (T=7 ve Aralarındaki Korelasyon=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0.0	2.6	2.2	2.1	1.2
Vaka	30.7	17.8	17.3	19.5	17.5
Değerlendirici	60.3	4.5	3.9	5	5.9
Vaka×Tanı Testi	5.4	5.6	6.7	5.7	6.2
Değerlendirici×Tanı Testi	0.4	3.8	4.9	3.9	4
Değerlendirici×Vaka	39.7	18.5	18.2	18.9	17.8

N=100	D=2 ROC=0.50 (T=7 ve Aralarındaki Korelasyon=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0.0	3.4	2.8	2.4	2.1
Vaka	22.7	19.0	17.9	17.2	16.4
Değerlendirici	61.1	4.3	5.4	4.9	4.2
Vaka×Tanı Testi	5.0	4.9	5.4	4.7	5.7
Değerlendirici×Tanı Testi	0.0	5.7	4.8	4.9	4.6
Değerlendirici×Vaka	47.1	28.2	31.3	29.4	27.2
N=15	D=5 ROC=0.50 (T=7 ve Aralarındaki Korelasyon=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0	1.8	1.4	1.1	1.2
Vaka	50.5	12.5	13.7	12.5	12.1
Değerlendirici	83.0	3.5	4.0	3.3	3.2
Vaka×Tanı Testi	6.1	5.2	5.7	6.4	5.4
Değerlendirici×Tanı Testi	0.1	4.6	3.0	1.6	3.1
Değerlendirici×Vaka	50.6	29.8	31	36.2	30.9
N=30	D=5 ROC=0.50 (T=7 ve Aralarındaki Korelasyon=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0.0	3.1	2.5	2.6	1.4
Vaka	33.7	15.1	11.6	13.4	11
Değerlendirici	89.4	5.6	4.6	3.9	4.5
Vaka×Tanı Testi	28.9	5.2	5.7	5.1	5.8
Değerlendirici×Tanı Testi	1.4	5.3	3.5	4	3.8

Değerlendirici×Vaka	53.9	27.3	29.5	29.3	31.1
N=100	D=5 ROC=0.50 (T=7 ve Aralarındaki Korelasyon=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	Programın Kapasitesi Yeterli Değil				
Vaka					
Değerlendirici					
Vaka×Tanı Testi					
Değerlendirici×Tanı Testi					
Değerlendirici×Vaka					
N=15	D=7 ROC=0.50 (T=7 ve Aralarındaki Korelasyon=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0	1.8	1.8	1.7	1.6
Vaka	53.1	10.4	12.4	11.0	10.7
Değerlendirici	75.4	4.2	3.0	2.3	3.0
Vaka×Tanı Testi	5.8	6.9	6.3	4.5	7.0
Değerlendirici×Tanı Testi	0	3.2	3.4	1.9	1.9
Değerlendirici×Vaka	41.3	32.0	32.1	36.3	31.1
N=30	D=7 ROC=0.50 (T=7 ve Aralarındaki Korelasyon=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	0.0	2.3	3.4	1.9	2.0
Vaka	39.7	11.6	13.9	14.1	7.9
Değerlendirici	96.6	6.3	5.5	4.6	2.6
Vaka×Tanı Testi	5.9	5.6	5.6	6.3	5.6
Değerlendirici×Tanı	0.0	5.2	3.3	3.3	3.3

Testi					
Değerlendirici×Vaka	47.9	28.5	30.1	32.3	32
N=100	D=7 ROC=0.50 (T=7 ve Aralarındaki Korelasyon=0.10)				
	K=2	K=3	K=5	K=7	K=10
Tanı Testi	Programın Kapasitesi Yeterli Değil				
Vaka					
Değerlendirici					
Vaka×Tanı Testi					
Değerlendirici×Tanı Testi					
Değerlendirici×Vaka					

ÖZGEÇMİŞ

24 Haziran 1976'da Tarsus'ta doğdu. 1998 yılında 19 Mayıs Üniversitesi Fen Edebiyat Fakültesi İstatistik Bölümünden mezun oldu. 1998 yılında Mersin Üniversitesi İ.İ.B.F. 'de Öğretim Görevlisi olarak göreve başladı. 2000 yılında Mersin Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalında Yüksek Lisans öğrencisi oldu. 2004 yılında Yüksek Lisansdan mezun oldu. 2006 yılında aynı Anabilim Dalında Doktora eğitimime başladı. 2006 yılından itibaren de Mersin Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalında Öğretim Görevlisi olarak çalışmaktadır.