



**İKİ PARAMETRELİ ÜSTEL DAĞILMIŞ
ANAKÜTLE ORTALAMALARININ
EŞİTLİĞİNİN TEST EDİLMESİ İÇİN
HESAPLAMALI YAKLAŞIM TESTLERİ**

Doktora Tezi

Mustafa ÇAVUŞ

Eskişehir 2020

**İKİ PARAMETRELİ ÜSTEL DAĞILMIŞ ANAKÜTLE
ORTALAMALARININ EŞİTLİĞİNİN TEST EDİLMESİ İÇİN
HESAPLAMALI YAKLAŞIM TESTLERİ**

Mustafa ÇAVUŞ

DOKTORA TEZİ

**İstatistik Doktora Programı
İstatistik Anabilim Dalı**

Danışman: Prof. Dr. Berna YAZICI

**Eskişehir
Eskişehir Teknik Üniversitesi
Lisansüstü Eğitim Enstitüsü
Kasım 2020**

Bu Tez Çalışması Eskişehir Teknik Üniversitesi BAP Komisyonu tarafından kabul edilen 20DRP047 no.lu proje kapsamında desteklenmiştir.

ÖZET

İKİ PARAMETRELİ ÜSTEL DAĞILMIŞ ANAKÜTLE ORTALAMALARININ EŞİTLİĞİNİN TEST EDİLMESİ İÇİN HESAPLAMALI YAKLAŞIM TESTLERİ

Mustafa ÇAVUŞ

İstatistik Anabilim Dalı

Eskişehir Teknik Üniversitesi, Lisansüstü Eğitim Enstitüsü, Kasım, 2020

Danışman : Prof. Dr. Berna YAZICI

İkiden fazla anakütle ortalamasının eşitliğinin test edilmesi temel istatistik problemlerinden biridir. Normal dağılım ve varyans homojenliği varsayımlarının sağlanması durumunda Klasik F testi problemin çözümü için en güçlü testtir. Uygulamada bahsedilen varsayımların sağlanmadığı durumlarla sıklıkla karşılaşmaktadır. Bu durumda araştırmacılar normallik dönüşümleri ve sağlam tahmincilerden yararlanarak alternatif çözümler geliştirmişlerdir. Fakat bu çözümler her durumda güçlü sonuçlar vermemektedir. Normal dağılım varsayımının sağlanmadığı ve verinin farklı bir parametrik dağılıma benzer bir dağılım gösterdiği durumda problem geniş bir olası çözüm kümesine sahiptir. Literatürde *nuisance* parametre içeren log-normal, ters-normal ve iki parametrelü üstel dağılım için Genelleştirilmiş p-Değeri, Parametrik Bootstrap ve *Fiducial* Yaklaşım yöntemleri kullanılarak testler geliştirilmiştir. Bu çalışmada ölçek parametresinin *nuisance* parametre olması durumunda iki parametrelü üstel dağılımş anakütle ortalamalarının eşitliğinin test edilmesi probleminin çözümü için Hesaplamalı Yaklaşım yöntemine dayalı testler önerilmiştir. Önerilen testlerin performansı literatürde yer alan farklı yöntemlere dayalı olarak geliştirilmiş testler ile karşılaştırılmıştır. Sonuç olarak önerilen testler küçük örneklemelerin yanı sıra diğer örneklem hacimlerinde de daha iyi performans göstermişlerdir. Bunun yanı sıra önerilen testlerin gerçek veri setleri üzerinde uygulamaları yapılmış ve böylelikle uygulamada diğer testlere olan üstünlükleri gösterilmiştir.

Anahtar Sözcükler: İki parametrelü üstel dağılım, hesaplamalı yaklaşım testi, ANOVA, heterojen varyanslılık, normal dağılmama

ABSTRACT

THE COMPUTATIONAL APPROACH TESTS FOR TESTING THE EQUALITY OF POPULATION MEANS UNDER TWO-PARAMETER EXPONENTIAL DISTRIBUTION

Mustafa ÇAVUŞ

Department of Statistics

Eskisehir Technical University, Institute of Graduate Programs, November, 2020

Advisor : Prof. Dr. Berna YAZICI

Testing equality of the several populations' mean is one of the main statistical problems. Classical F test is the most powerful test for the solution of the problem when the assumptions are normality and variance homogeneity. These assumptions may be violated in practice. In this case, the researchers improved alternative solutions using normality transformation techniques and robust estimators. However, these solutions may not give powerful results in all cases. The problem has a large set of possible solutions when the normality assumption is violated. There are some tests have been improved using Generalized p-Value, Parametric Bootstrap and Fiducial Approach methods for log-normal, inverse-normal and two-parameter exponential distribution in the presence of nuisance parameter in the literature. In this study, new tests based on Computational Approach method are proposed to solve the problem of testing the equality of two-parameter exponentially distributed populations' means. The performance of the proposed tests is compared with the alternatives in the literature. As a result, the proposed tests performed better especially in small samples. In addition, the proposed tests have been applied on real data sets, and thus their advantages over the alternatives have been demonstrated.

Keywords: Two-parameter exponential distribution, computational approach test, ANOVA, heteroscedasticity, non-normality

TEŞEKKÜR

Yüksek öğrenimime başladığım günden bugüne kadar olan desteği ve güler yü-
züyle her zaman yanımda olan sevgili danışman hocam *Prof. Dr. Berna YAZICI* ve
akademik çalışmalarım boyunca desteğini her zaman hissettiğim sayın hocam *Doç.*
Dr. Ahmet SEZER'e,

Değerli katkılarından dolayı tez izleme komitesi ve tez savunma jürisinde yer
alan sayın hocalarıma,

Bu günlere gelebilmem için büyük emek harcayan *anneme* ve *babama* en içten
saygılarımı ve teşekkürlerimi sunarım.

Hayatıma anlam katan, sonsuz destekçim ve biricik eşim *Özlem*'e ...

Mustafa ÇAVUŞ

Kasım 2020

ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ

Bu tezin bana ait, özgün bir çalışma olduğunu; çalışmamın hazırlık, veri toplama, analiz ve bilgilerin sunumu olmak üzere tüm aşamalarında bilimsel etik ilke ve kurallara uygun davrandığımı; bu çalışma kapsamında elde edilemeyen tüm veri ve bilgiler için kaynak gösterdiğimi ve bu kaynaklara kaynakçada yer verdiğimi; bu çalışmamın Eskişehir Teknik Üniversitesi tarafından kullanılan "bilimsel intihal tespit proğıyla tarandığını ve hiçbir şekilde "intihal içermediğini" beyan ederim. Herhangi bir zamanda, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun saptanması durumunda, ortaya çıkacak tüm ahlaki ve hukuki sonuçlara razı olduğumu bildiririm.

Mustafa ÇAVUŞ

İÇİNDEKİLER

	<u>Sayfa</u>
BAŞLIK SAYFASI	i
JÜRİ VE ENSTİTÜ ONAYI.....	ii
ÖZET	iii
ABSTRACT.....	iv
TEŞEKKÜR.....	v
ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ.....	vi
İÇİNDEKİLER	vii
TABLolar DİZİNİ	x
ŞEKİLLER DİZİNİ.....	xi
SİMGELER VE KISALTMALAR DİZİNİ	xii
1. GİRİŞ.....	1
2. İKİ PARAMETRELİ ÜSTEL DAĞILMIŞ ANAKÜTLE ORTALA- LAMALARININ EŞİTLİĞİNİN TEST EDİLMESİ İÇİN KUL- LANILAN TESTLER.....	6
2.1 Genelleştirilmiş p-Değeri Yöntemi	6
2.2 Parametrik Bootstrap Yöntemi	7
2.3 Fiducial Yaklaşım Yöntemi	8
2.4 İki Parametrelİ Üstel Dağılım İÇin Kullanılan Testler	8
2.4.1 Genelleştirilmiş p değeri testi	11
2.4.2 Parametrik bootstrap testi	12

2.4.3 Fiducial yaklaşım testi	13
3. İKİ PARAMETRELİ ÜSTEL DAĞILMIŞ ANAKÜTLE ORTALAMALARININ EŞİTLİĞİNİN TEST EDİLMESİ İÇİN ÖNERİLEN TESTLER.....	15
3.1 Hesaplamalı Yaklaşım Yöntemi	15
3.2 Önerilen Hesaplamalı Yaklaşım Testleri.....	16
3.2.1 Hesaplamalı Yaklaşım 1 Testi	19
3.2.2 Hesaplamalı Yaklaşım 2 Testi	20
3.2.3 Hesaplamalı Yaklaşım 3 Testi	21
3.2.4 Hesaplamalı Yaklaşım 4 Testi	22
4. MONTE-CARLO SİMÜLASYON ÇALIŞMALARI	24
4.1 Değerlendirme Kriterleri.....	25
4.1.1 Zhang ve Boos'un düzeltilmiş güç kriteri	27
4.1.2 Lloyd'un asıl güç kriteri	29
4.1.3 Cezalandırılmış güç kriteri	30
4.2 1. Tip Hata Olasılıkları.....	34
4.3 Cezalandırılmış Güç Değerleri.....	39
5. GERÇEK VERİ UYGULAMALARI	56
5.1 doex R - Paketi	56
5.2 Ürün Güvenilirliği Örneği.....	58
5.3 Yaşam Modeli Örnekleri	62
5.3.1 Yaşam modeli örneği 1	62
5.3.2 Yaşam modeli örneği 2	65
5.3.3 Yaşam modeli örneği 3	67
5.4 Ekonomik Gösterge Örneği	70
6. SONUÇ VE TARTIŞMA	74
KAYNAKÇA.....	78

ÖZGEÇMİŞ.....	87
---------------	----



TABLÖLAR DİZİNİ

	<u>Sayfa</u>
Tablo 1. <i>Hipotez testlerinde hata türleri</i>	25
Tablo 2. <i>Testlerin 1.tip hata olasılıkları</i>	26
Tablo 3. <i>Önerilen testlerin $k = 3$ anakütle için 1.tip hata olasılıkları</i> .	35
Tablo 4. <i>Önerilen testlerin $k = 4$ anakütle için 1.tip hata olasılıkları</i> .	36
Tablo 5. <i>Önerilen testlerin $k = 5$ anakütle için 1.tip hata olasılıkları</i> ..	38
Tablo 6. <i>Önerilen testlerin $k = 3$ grup için cezalandırılmış güç değerleri</i>	40
Tablo 7. <i>Önerilen testlerin $k = 3$ grup için cezalandırılmış güç değerleri</i>	43
Tablo 8. <i>Önerilen testlerin $k = 4$ grup için cezalandırılmış güç değerleri</i>	45
Tablo 9. <i>Önerilen testlerin $k = 4$ grup için cezalandırılmış güç değerleri</i>	48
Tablo 10. <i>Önerilen testlerin $k = 5$ grup için cezalandırılmış güç değerleri</i>	50
Tablo 11. <i>Önerilen testlerin $k = 5$ grup için cezalandırılmış güç değerleri</i>	53
Tablo 12. <i>Ürün güvenilirliği örneği verileri</i>	60
Tablo 13. <i>Ürün güvenilirliği örneği üzerinde H_0 hipotezini test etmek için uygulanan testlerin p-değerleri</i>	62
Tablo 14. <i>Yaşam modeli örnek 1 verileri</i>	63
Tablo 15. <i>Yaşam modeli 1 örneği üzerinde H_0 hipotezini test etmek için uygulanan testlerin p-değerleri</i>	64
Tablo 16. <i>Yaşam modeli örnek 2 verileri</i>	65
Tablo 17. <i>Yaşam modeli 2 örneği üzerinde H_0 hipotezini test etmek için uygulanan testlerin p-değerleri</i>	67
Tablo 18. <i>Yaşam modeli örnek 3 verileri</i>	68
Tablo 19. <i>Yaşam modeli 3 örneği üzerinde H_0 hipotezini test etmek için uygulanan testlerin p-değerleri</i>	69
Tablo 20. <i>Ekonomik gösterge örneği verileri</i>	71
Tablo 21. <i>Ekonomik gösterge örneği üzerinde H_0 hipotezini test etmek için uygulanan testlerin p-değerleri</i>	73

ŞEKİLLER DİZİNİ

	<u>Sayfa</u>
Şekil 1. İki parametrelili üstel dağılımın olasılık yoğunluk fonksiyonu	9
Şekil 2. Cezalandırılmış ve asıl güç kriterlerinin testin gücü (0.95) sabit olduğunda aldığı değerler	31
Şekil 3. Cezalandırılmış ve asıl güç kriterlerinin testin gücü (0.80) sabit olduğunda aldığı değerler	32
Şekil 4. Cezalandırılmış ve asıl güç kriterlerinin 1.tip hata olasılığı (0.10) sabit olduğunda aldığı değerler	33
Şekil 5. Cezalandırılmış ve asıl güç kriterlerinin 1.tip hata olasılığı (0.01) sabit olduğunda aldığı değerler	34
Şekil 6. doex paketinin R ortamında kurulması ve yüklenmesi	57
Şekil 7. GPD, PB ve FY testlerinin R ortamında uygulanması	58
Şekil 8. Ürün güvenilirliği örneği verilerinin dağılımı	61
Şekil 9. Yaşam modeli örnek 1 verilerinin dağılımı	64
Şekil 10. Yaşam modeli örnek 2 verilerinin dağılımı	66
Şekil 11. Yaşam modeli örnek 3 verilerinin dağılımı	69
Şekil 12. Ekonomik gösterge örneği verilerinin dağılımı	72

SİMGELER VE KISALTMALAR DİZİNİ

SİMGELER

X	: X vektörü
x	: X vektörünün gözlenen değeri
θ	: ilgilenilen parametre
θ_0	: ilgilenilen parametrenin belirlenen noktadaki değeri
δ	: <i>nuisance</i> parametre
T	: test değişkeni
t	: test değişkeninin gözlenen değeri
ζ	: parametre uzayı
ζ_0	: parametre uzayının gözlenen değeri
U_i	: $(2n - 2)$ serbestlik dereceli ki-kare dağılımdan üretilen rassal örneklem
V_i	: (2) serbestlik dereceli ki-kare dağılımdan üretilen rassal örneklem
R_i	: GPD testi pivot değeri
μ_{Ri}	: GPD testi pivot değerinin beklenen değeri
σ_{Ri}	: GPD testi pivot değerinin varyansı
μ_{Bi}	: PB testi bootstrap örnekleminin beklenen değeri
σ_{Bi}	: PB testi bootstrap örnekleminin varyansı
a	: iki parametrelili üstel dağılımın ölçek parametresi
b	: iki parametrelili üstel dağılımın konum parametresi
μ	: iki parametrelili üstel dağılımın ortalama tahmini
k	: anakütle sayısı
n	: örneklem hacmi
α_i	: 1.tip hata olasılığı
α_0	: nominal seviye
c_α	: kritik değer
c_α^*	: düzeltilmiş kritik değer
β_i	: testin gücü
γ_i	: cezalandırılmış güç

KISALTMALAR

KF	: Klasik F-Testi
GPD	: Genelleştirilmiş P-Değeri
PB	: Parametrik Bootstrap
FY	: Fiducial Yaklaşım
GF	: Genelleştirilmiş F
HY	: Hesaplamalı Yaklaşım
BF	: Brown-Forsythe
MBF	: Modified Brown-Forsythe
ANORE	: Analysis of reciprocals
RML	: Kısıtlandırılmış en çok olabilirlik tahmincisi
ML	: En çok olabilirlik tahmincisi
PML	: Cezalandırılmış en çok olabilirlik tahmincisi
HY1	: Önerilen 1. Hesaplamalı Yaklaşım testi
HY2	: Önerilen 2. Hesaplamalı Yaklaşım testi
HY3	: Önerilen 3. Hesaplamalı Yaklaşım testi
HY4	: Önerilen 4. Hesaplamalı Yaklaşım testi
CvM	: Cramer-von-Mises uyum iyiliği testi

1. GİRİŞ

Klasik tek yönlü varyans analizinde amaç ikiden fazla normal dağılıma uyan anakütlenin ortalamalarının eşitliğini test etmek ve böylece yanıt değişkeni üzerinde ele alınan faktörlerin etkisi olup olmadığına karar vermektir. Tek yönlü varyans analizi başta klinik araştırmalar olmak üzere: tıp, farmakoloji, tarım araştırmaları, kalite kontrol süreci gibi birçok alanda kullanılmaktadır. Klinik araştırmalarda yanıt değişkeni üzerinde bir faktörün çeşitli düzeylerinin etkilerinin farklı olup olmadığı sorusuna cevap aramakta kullanılır. Farmakolojik araştırmalarda geliştirilen ilaç formüllerinin etkisi ve doz düzeylerinin etkisinin araştırılmasında ihtiyaç duyulur. Normal dağılmış anakütlelerden geldiği bilinen ve eşit varyanslı grupların ortalamalarının eşitliğinin test edilmesi, Fisher (1925) tarafından önerilen Klasik F testi (KF) ile yapılır. Ancak gerçek hayatta verilerin normal dağılmaması ve heterojen varyanslılık durumuyla sıklıkla karşılaşmaktadır. Özellikle klinik ve farmakolojik araştırmalarda deneklerin sayısı genellikle azdır. Bu nedenle araştırmacılar küçük örneklem hacimleri ile çok sık çalışmak zorunda kalırlar. Gözlem sayısının az olması gözlem verileri üzerinden elde edilen istatistiklerin dağılımlarının normal olmamasına neden olabilir. Normal dağılım ya da varyans homojenliği varsayımının sağlanmaması KF testinin 1.tip ve 2.tip hata oranının artmasına neden olur. Bahsedilen kısıtlar ile etkilerin farklılığının test edilmesinde doğru yöntemin kullanılması gerekmektedir. Kullanılacak yöntemin varyans homojenliği varsayımının bozulmasından etkilenmemesinin yanı sıra normallikten uzaklaşması karşısında da dayanıklı bir yöntem olması gerekmektedir.

Homojen varyanslılık varsayımı sağlanmadığında KF testi gücünü kaybetmekte ve 1.tip hata oranı nominal seviyenin üzerine çıkmaktadır (Krutchkoff, 1988). Bu durumda araştırmacıların KF testini kullanması yanlış kararlar vermelerine neden olmaktadır. Behrens-Fisher problemi olarak da bilinen bu sorun araştırmacıları yeni yöntemler geliştirme arayışına itmiş ve tam olasılıklı (*exact*), yaklaşık (*approximate*), yeniden örnekleme (*resampling*) yöntemlerine dayalı birçok test geliştirilmiştir. Literatürde yeniden örnekleme yöntemleri, sağladığı avantajlar ve hesaplama teknolojilerindeki gelişmeler ile birlikte Behrens-Fisher ve benzeri *nuisance* parametre içeren problemlerin çözümünde son yıllarda sıklıkla kullanılmıştır. Bu yöntemler-

den ilki Tsui ve Weerahandi (1989) tarafından geliştirilen Genelleştirilmiş p-değeri (GPD) yaklaşımıdır. Bilinmeyen parametreye bağlı olmayan test istatistikleri geliştirmeye imkan sağladığı için son yıllarda popülerliği artmış ve bu yönetime dayalı birçok test geliştirilmiştir. Weerahandi (1995) GPD yöntemine dayalı olarak homojen varyanslı olmayan ikiden fazla anakütle ortalamasının eşitliğini test etmek için Genelleştirilmiş F (GF) testini geliştirmiştir. Gamage ve Weerahandi (1998) GF testinin performansını literatürde yer alan asimptotik alternatifleri olan Welch ve Brown Forsythe testleriyle karşılaştırmış ve daha iyi performans sergilediğini göstermişlerdir. Krishnamoorthy vd. (2007) ise GPD yönteminin *bootstrap* örnekleme dağılımını kullanan Parametrik Bootstrap (PB) adını verdikleri bir versiyonunu geliştirmişlerdir. Bootstrap örneklemlerinden yararlanması nedeniyle küçük örneklem hacimlerinde GPD yöntemine dayalı GF testinden daha iyi bir performans göstermiştir. Pal vd. (2007) çalışmalarında bir tür PB yöntemi olan Hesaplamalı Yaklaşım (HY) yöntemini önermişlerdir. Bu yöntem herhangi bir örnekleme dağılımına ihtiyaç duymadığı için öncesinde belirtilen yöntemlere göre bir avantaj sağlamaktadır. Son olarak ise Li vd. (2011) Fisher'in *fiducial* tahminini kullanarak Fiducial Yaklaşım (FY) yöntemini bu problem üzerinde kullanarak yeni bir test geliştirmişlerdir. GPD, PB, HY ve FY yöntemleri test geliştirmedeki farklı avantajları sebebiyle birçok araştırmacı tarafından farklı problemler üzerinde de kullanılmıştır.

Homojen varyanslılık varsayımının sağlanmaması durumuna bir de normal dağılmama durumu eklendiğinde geliştirilen test istatistiklerinin 1.tip ve 2.tip hata oranı artmaktadır. Bu sorunun çözümü için araştırmacılar homojen olmayan varyans durumunda kullanılan yöntemlere bazı düzeltmeler (*adjustment*) ve iyileştirmeler (*modification*) önermişlerdir. Bu yaklaşım genel olarak üç yöntemi temel almaktadır. Sağlam (*robust*) tahminciler kullanarak test istatistiklerine iyileştirmeler önermek (Oshima ve Algina, 1992; Luh ve Guo, 2005; Cribbie, 2012; Ochuko vd., 2015; 2016; Karagöz, 2016; Karagöz ve Saraçbaşı, 2016; Cavus vd., 2017; Ozdemir vd., 2017), normallik dönüşümleri ile test istatistiklerine düzeltmeler önermek ve sıralı veriler üzerine parametrik olmayan yöntemler geliştirilmektir (Kruskal ve Wallis, 1952; Brunner vd., 1997). Normallik varsayımının sağlanmaması verilerin aykırı değer içermesi ya da bilinen farklı bir parametrik dağılım ailesine uygun dağılmasından kaynaklanabilir. Bu nedenle normallik ve homojen varyanslılık varsayımlarının

sağlanmaması durumuna karşı geliştirilen testler bahsedilen iki durum için de iyi performans göstermeyebilirler. Bu durumda en iyi çözüm bilinen bir parametrik dağılım ailesine dayalı yeni bir test istatistiği geliştirmektir.

Literatürde Behrens-Fisher tipi *nuisance* parametre içeren problemlerin çözümü için GPD, PB, HY ve FY yöntemleri kullanılarak birçok farklı dağılıma uygun test geliştirilmiştir. Li (2009) çalışmasında GPD yöntemini kullanarak ikiden fazla log-normal anakütle ortalamasının sabit varyanslılık durumu sağlanmadığı koşullarda bile güçlü sonuçlar veren bir test geliştirmiştir. Lin ve Wang (2013) GPD yöntemini kullanarak Li (2009)'nin testine bir modifikasyon önermiştir. Gökpınar ve Gökpınar (2015) HY yöntemini kullanarak benzer bir test geliştirmişlerdir. Jafari ve Abdollahnezhad (2017) HY yöntemiyle alternatif bir test daha geliştirmişlerdir. Gökpınar (2017) log-normal anakütle ortalamalarının eşitliğinin testi için önerilen testlerin performanslarını karşılaştırmış ve birbirlerine karşı olan üstünlük ve zayıflıklarını ortaya çıkarmıştır. Üzerinde sık çalışılan diğer bir dağılım ise ters normal dağılımdır. Chhikara ve Folks (1989) tarafından geliştirilen ikiden fazla ters normal dağılmış anakütle ortalamasının eşitliğinin testi probleminin çözümü için kullanılan ANORE (*ANalysis Of REciprovals*) testine alternatif olarak Tian (2006) GPD yöntemine dayalı bir test önermiştir. Bu test anakütle ölçek parametrelerinin bilinmediği ya da eşit olmadığı durumlarda ANORE testine göre daha güçlü sonuçlar vermektedir. Ma ve Tian (2009) aynı problemin çözümü için PB yöntemine dayalı bir test geliştirmişlerdir. Sonrasında ise Gökpınar vd. (2013) HY yöntemine dayalı bir alternatif önermişlerdir. Güven (2018) geliştirilen alternatif testlerin performanslarını karşılaştırmış ve elde ettiği sonuçları değerlendirmiştir. Log-normal ve ters normal dağılımları, bazı durumlarda normal dağılım özelliği taşıyan dağılımlar oldukları için yoğun ilgi çekmişlerdir. Ancak bu dağılımların yanı sıra üzerinde fazla çalışılmamış dağılımların konu edildiği çalışmalar da literatürde mevcuttur. Abdollahnezhad ve Jafari (2015) ise ilk kez güç fonksiyonu (*power function*) dağılımı olarak da bilinen güç yasası (*power law*) dağılımı parametrelerinin hipotez testleri üzerinde HY yöntemi kullanarak çalışmıştır. Niu vd. (2014) güvenilirlik modellerinde sıklıkla kullanılan üstel dağılımın özel bir türü olan Birbaum-Saunders dağılımı gösteren anakütle ortalamalarının eşitliğinin test edilmesi için GPD yöntemine dayalı bir test önermiştir ancak bekledikleri performansı elde edememişlerdir. Guo vd. (2017) ise aynı

problemin çözümü için benzer şekilde GPD yöntemini kullanmışlar ve iyi performans gösteren bir test geliştirebilmişlerdir. Son olarak Malekzadeh ve Jafari (2019) iki parametrelili üstel dağılmış ikinci tip ilerleyen sansürlü verilerden oluşan anakütle ortalamalarının eşitliğinin test edilmesi için GPD, PB ve FY yöntemlerini kullanarak üç test önermişlerdir.

Bu tez çalışmasında *nuisance* parametre içeren iki parametrelili üstel dağılmış anakütle ortalamalarının eşitliğinin test edilmesi ele alınmıştır. İki parametrelili üstel dağılım, bilinen tek parametrelili üstel dağılımdan farklı olarak konum parametresine sahiptir; bu sayede yoğunluk fonksiyonu x-ekseni üzerinde hareket edebilmekte ve daha geniş bir uygulama alanında kullanılmaktadır. Problemde bahsedilen *nuisance* parametre ise bilinmeyen ya da eşit olmayan ölçek parametresidir. Behrens-Fisher tipi bu problemin çözümünün önemini Ghoush ve Razmpour (1984) şu şekilde anlatmaktadır: Dayanıklılık verisi iki parametrelili üstel dağılım gösteren bir ürün ele alındığında, dağılımın ortalaması garanti süresini, ölçek parametresi de hata oranını temsil etmektedir. Böylelikle bu problemin çözümü ile hata oranları bilinmeyen ya da eşit olmayan ürünlerin garanti ya da ortalama dayanıklılık sürelerinin eşitliği güçlü bir şekilde test edilebilecektir. İki parametrelili üstel dağılım üzerinde geliştirilen test istatistikleri incelendiğinde ise, öncelikle Chen (1982) konum parametrelerinin eşitliğinin test edilmesi için *range* istatistiği, Singh (1983) ise olabilirlik oran testi geliştirdiği görülmüştür. Kambo ve Awad (1985) sansürlü veriler için eşit olmayan ölçek parametresi durumunda konum parametrelerinin eşitliğinin test edilmesi için bir test önermiştir. Hsieh (1986) ise alternatif olarak tam olasılıklı bir test geliştirmiştir. Ananda ve Weerahandi (1996) GPD yöntemini kullanarak iki üstel dağılım ortalamasının eşitliğinin test edilmesi için bir test geliştirmiştir. Tiku ve Vaughan (1991) ise iki parametrelili üstel dağılımın konum parametrelerinin karşılaştırılması için bir test geliştirmiş ve bu testi Vaughan ve Tiku (1993) çalışmalarında ikiden fazla anakütle için genişletmişlerdir. Wu (2019) tek aşamalı çoklu karşılaştırma testi geliştirmiştir. Son olarak Malekzadeh ve Jafari (2019) çalışmasıyla iki parametrelili üstel dağılım anakütle ortalamalarının eşit olmayan ölçek parametresi durumunda bile güçlü sonuçlar veren testler önermişlerdir. Ancak bu çalışma ikinci tip ilerleyen sansürlü veriler için gerçekleştirilmiş ve önerilen testler sansürlü olmayan veriler için düzenlenmemiştir. Literatürdeki bu eksikliği gidermek için bu çalışmada öncelikle

Malekzadeh ve Jafari (2019) tarafından geliştirilen yöntemler sansürlü olmayan veri için yeniden düzenlenmiştir. İlgili durum için parametre tahminleri ve dağılımları verilmiş, GPD, PB ve FY yöntemlerine dayalı test önerileri yeniden düzenlenmiştir. Sonrasında ise HY yöntemi kullanılarak problemin çözümü için alternatif bir test geliştirilmiştir. Bahsi geçen yöntemler ile tam veri için düzenlenen ve geliştirilen testlerin performansları Monte-Carlo simülasyon çalışmaları ile araştırılmış ve en iyi performansları sergiledikleri koşullar belirlenmiştir. Bu sayede araştırmacılara en güçlü sonuçları elde edebilecekleri öneriler verilmiştir. Son kısımda ise yaşam çözümlemesi, güvenilirlik analizi ve ekonomik göstergeler içeren gerçek veri seti problemleri üzerinde yapılan uygulamalar ile simülasyon sonuçları desteklenmiştir.

Tezin izleyen bölümleri şu şekilde devam etmektedir: 2.bölümde literatürde ikiden fazla iki parametrelili üstel dağılmış anakütle ortalamasının eşitliğinin test edilmesi probleminin çözümünde kullanılan yöntemler tanıtılmış ve bu yöntemlere dayalı olarak geliştirilen testler verilmiştir. 3.bölümde Hesaplamalı Yaklaşım yöntemi tanıtılmış ve bu yönteme dayalı olarak tez çalışmasında önerilen testler tanıtılmıştır. 4.bölümde önerilen testlerin performansları literatürde yer alan testler ile Monte-Carlo simülasyonları ile karşılaştırılmıştır. Bu bölümde ayrıca simülasyon çalışmalarında testlerin güç kıyaslamalarını yapmak için tez çalışması sürecinde geliştirilen cezalandırılmış güç kriteri verilmiştir. 5.bölümde gerçek veri setleri üzerinde geliştirilen testleri uygulayabilmek için geliştirilen R paketi *doex* tanıtılmış ve bu paket kullanılarak uygulanan testlerden elde edilen sonuçlar karşılaştırmalı olarak verilmiştir. Son bölümde ise tez çalışmasında elde edilen bulgular değerlendirilmiş ve sonuçlar tartışılmıştır.

2. İKİ PARAMETRELİ ÜSTEL DAĞILMIŞ ANAKÜTLE ORTALAMALARININ EŞİTLİĞİNİN TEST EDİLMESİ İÇİN KULLANILAN TESTLER

Bu bölümde öncelikle *nuisance* parametre içeren hipotez testlerinde güçlü test istatistikleri geliştirmek için kullanılan Genelleştirilmiş p-değeri, Parametrik Bootstrap ve Fiducial Yaklaşım yöntemleri tanıtılmıştır. Sonrasında Malekzadeh ve Jafari (2019) tarafından geliştirilen ilerleyen ikinci tip sansürlü verilerde ikiden fazla iki parametrelili üstel dağılmış anakütle ortalamalarının eşitliğini test etmek için kullanılan testler sansürlü olmayan veriler için yeniden düzenlenerek verilmiştir.

2.1. Genelleştirilmiş p-Değeri Yöntemi

Genelleştirilmiş tahmin yöntemi Tsui ve Weerahandi (1989) tarafından *nuisance* parametre olması durumunda bile güçlü sonuçlar veren test istatistikleri geliştirebilmek için önerilen bir yöntemdir. Pivot değer, test değişkeni ve p-değerinin genelleştirilmesine genelleştirilmiş çıkarım denir. Klasik yaklaşımda test istatistiğinin dağılımının uç bölgesi kullanılırken genelleştirilmiş çıkarımda ise test değişkeni için oluşturulan örnek uzayının uç noktası kullanılır. Öncelikle genelleştirilmiş p-değerini hesaplamak için kullanılan genelleştirilmiş red bölgesi ve genelleştirilmiş test değişkeni kavramlarını ele alalım.

Genelleştirilmiş red bölgesi: $T(X; x, \zeta)$ test değişkeninin genelleştirilmiş test değişkeni olduğunu varsayalım.

$$C_X(\zeta) = \{X \in \Omega, T(X; x, \zeta) \geq T(x; x, \zeta)\} \quad (2.1)$$

şeklindeki örneklem uzayının alt kümesine genelleştirilmiş red bölgesi denir.

Genelleştirilmiş test değişkeni: Aşağıda verilen üç özelliği sağlayan $T = T(X; x, \zeta)$ rassal değişkenine genelleştirilmiş test değişkeni denir.

i. $t_g = t(x; x, \zeta)$ bilinmeyen parametreye bağlı değildir.

- ii. $T = T(X; x, \zeta)$ test değişkeninin dağılımı *nuisance* parametreden bağımsızdır.
- iii. $P(T \leq t; \theta)$ ifadesi θ parametresinin monoton fonksiyonudur.

Tsui ve Weerahandi (1989) çalışmasında geleneksel p-değerinin tanımını genelleştirmiş ve böylelikle *nuisance* parametre içeren problemlere çözüm sağlamışlardır. GPD yöntemi, $H_0 : \theta = \theta_0$ nokta tahminine dayalı bir yokluk hipotezinin $H_1 : \theta \neq \theta_0$ alternatif hipotezine karşı testinde parametre uzayı $\zeta = (\theta, \delta)$ ' da *nuisance* parametre δ bulunması durumunda red bölgesinin parametre dağılımlarına uygun olarak üretilen rassal örneklemeler üzerinden belirlenmesine dayanır. Tsui ve Weerahandi (1989) bu yöntemi kullanarak $T(X; x, \zeta)$ test istatistiği için red bölgesinin tanımını genişletmişlerdir. Benzer hipotez testlerinde eğer aşağıda verilen iki koşul sağlanırsa bu yöntemle ilgili test istatistiğinin *nuisance* parametreden bağımsız red bölgesi elde edilebilmektedir:

1. $T(X; x, \zeta_0)$ test istatistiğinin dağılım fonksiyonu ve gözlenen değeri $t = T(x; x, \zeta)$, $\zeta_0 = (\theta_0, \delta)$ olduğunda *nuisance* parametre δ 'dan bağımsız olmalıdır.
2. Tüm θ ve t, x, δ gözlenen değerleri için $P[T(x; x, \zeta) \geq t] \geq P[T(X; x, \zeta_0) \geq t]$ olmalıdır.

Böylelikle genelleştirilmiş p-değeri $p = P[T(X; x, \zeta_0) \geq t]$ olarak ifade edilebilir. Yukarıda verilen koşullardan ilki p-değerinin hesaplanabilirliğini, ikinci koşul ise yansızlığını temin eder. Aynı geleneksel p-değerinde olduğu gibi genelleştirilmiş p-değeri de yokluk hipotezini reddetmek için gerekli kanıtları ölçer.

2.2. Parametrik Bootstrap Yöntemi

Krishnamoorthy vd. (2007) tarafından geliştirilen Parametrik Bootstrap yöntemi, Genelleştirilmiş p-Değeri yönteminden farklı olarak tahmin edilmek istenen modelden yeniden örneklemeler üreterek parametre tahminlerinin örnekleme dağılımı elde etme mantığına dayanmaktadır. Üretilen örneklemeler üzerinden elde edilen parametreler model parametreleriyle yer değiştirilerek örneklem üzerinden parametre tahmini yapar. Test değişkeninde kullanılan *nuisance* parametre yerine parametrenin örnekleme dağılımına uygun rassal örneklemeleri kullanarak güçlü test istatistikleri

elde edilebilir. GPD yöntemi ile geliştirilen benzer test istatistiklerine göre en büyük avantajı örneklem hacminin, karşılaştırılan anakütle sayısının ve hata varyansının arttığı durumlarda 1.tip hata olasılığını daha iyi kontrol edebilmesidir.

2.3. Fiducial Yaklaşım Yöntemi

Li vd. (2011) Fisher'ın *fiducial* tahminini kullanarak hipotez testleri için yeni bir yöntem önermişlerdir. İlgilenilen parametre θ ve *nuisance* parametre δ olduğunda parametre uzayı $\zeta = (\theta, \delta)$ ile ifade edilebilir. Dağılımı bilinen bir U rassal değişkeni $\psi = G(\delta, U)$ fonksiyonu olduğunu varsayalım. Bu eşitliğin δ için tek bir çözümü vardır ve $G^{-1}(\psi, u)$ şeklinde gösterilir. $G^{-1}(\psi, U)$ ve $\theta(G^{-1}(\psi, U))$ fonksiyonlarının dağılımı δ , $\theta(\delta)$ ve gözlenen ψ değerine bağlıdır. FY yöntemi ile p-değeri aşağıda verilen algoritma kullanılarak hesaplanır.

FY yöntemi hesaplama algoritması:

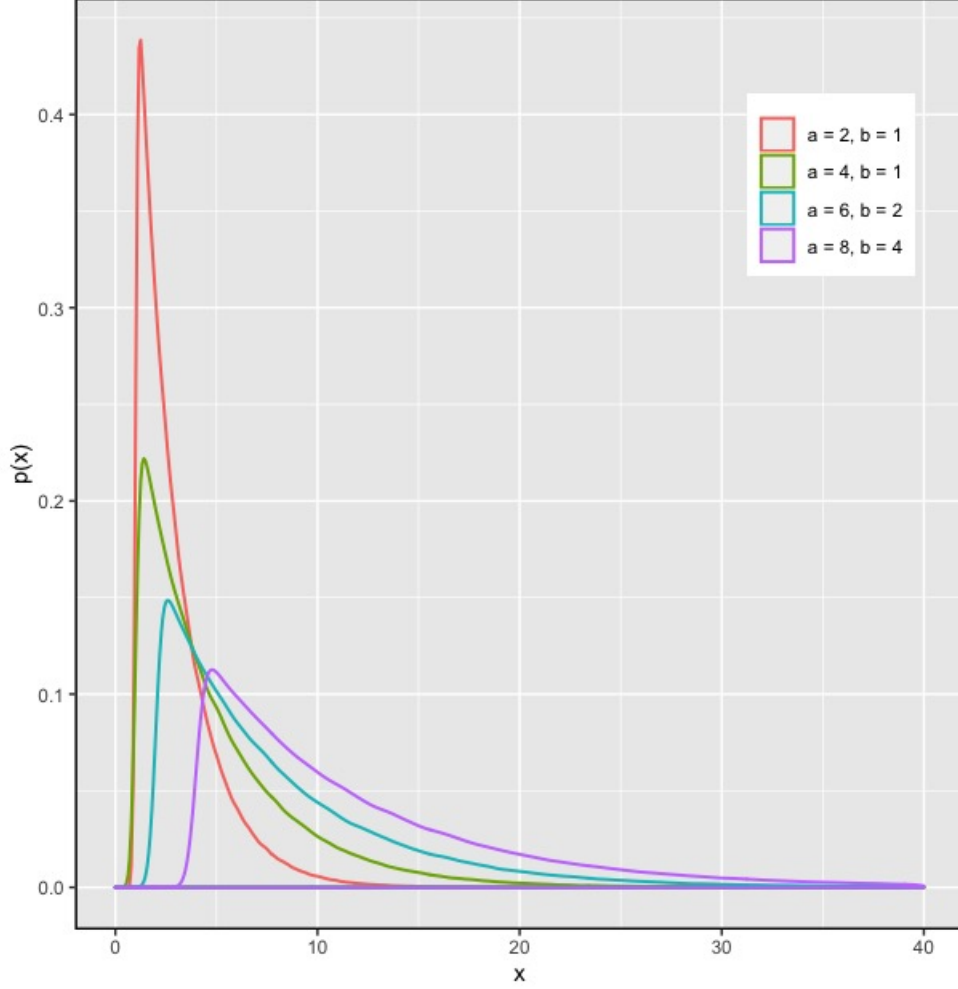
- i. İlgilenilen ve *nuisance* parametrenin tahmincileri $\hat{\theta}$ ve $\hat{\delta}$ hesaplanır.
- ii. Parametrelerin tahmincileri kullanılarak test istatistiğinin gözlenen değeri $\hat{t}^*(x; x, \hat{\zeta})$ hesaplanır.
- iii. Parametrelerin *fiducial* dağılımları elde edilerek test istatistiğinin gözlenen değerleri $\hat{t}^{(m)}(x; x, \hat{\zeta})$, $m = 1, 2, \dots, M$ hesaplanır.
- iv. Bir önceki adımda hesaplanan değerler üzerinden $p = \frac{1}{M} \sum_{m=1}^M I[\hat{t}^{(m)}(x; x, \hat{\zeta}) \geq \hat{t}^*(x; x, \hat{\zeta})]$ değeri hesaplanır.

2.4. İki Parametrelili Üstel Dağılım İçin Kullanılan Testler

Ölçek parametresi a ve konum parametresi b olmak üzere iki parametreden oluşan iki parametrelili üstel dağılımın olasılık yoğunluk fonksiyonu aşağıdaki gibidir:

$$f(x; a, b) = \frac{1}{a} \exp\left(-\frac{x-b}{a}\right), \quad x > b, \quad a > 0 \quad (2.2)$$

bu fonksiyonda görüldüğü gibi gözlem değerleri her zaman konum parametresine eşit ya da daha büyüktür, ölçek parametresi a ise pozitif değerler alır. Pozitif tanımlı olan bu dağılım sağa çarpık bir yapıdadır. Çeşitli parametre değerlerine göre olasılık yoğunluk fonksiyonu grafiği aşağıdaki gibidir.



Şekil 1. İki parametrelili üstel dağılımın olasılık yoğunluk fonksiyonu

Kaydırılmış (*shifted*) üstel dağılım olarak da bilinen iki parametrelili üstel dağılım birçok gerçek hayat problemine çözüm olarak kullanılmaktadır. Bir sistemdeki müşterilerin servis zamanını, yıllık ekstrem yağış miktarlarının modellenmesi, herhangi bir ürünün garanti süresinin hesaplanması ve tedavi sürelerinin modellenmesi gibi birçok durumda kullanılır. Güvenilirlik analizinde ve yaşam testlerinde iki parametrelili üstel dağılımın bilinmeyen ya da eşit olmayan ölçek parametresi, farklı hata oranlarını; ortalama parametresi ise garanti süresini verir (Ghoush ve Razmpour,

1984).

İki parametrelili üstel dağılım parametrelerinin en çok olabilirlik tahmin edicileri aşağıdaki gibidir:

$$\hat{b}_i = X_{i(1)} \quad (2.3)$$

$$\hat{a}_i = (n_i - 1)S_i/n_i \quad (2.4)$$

$$\hat{S}_i = \frac{1}{n_i} \sum_{j=1}^n (X_{(j)} - X_{(1)}) \quad (2.5)$$

$(X_{i(1)}, S_i)$ vektörü (b_i, a_i) için yeterli istatistiktir. Ayrıca S_i ve $X_{i(1)}$ bağımsız rassal değişkenlerdir ve

$$U_i = 2(n_i - 1)S_i/a_i \sim \chi^2_{(2n_i-2)} \quad (2.6)$$

$$V_i = 2n_i(X_{i(1)} - b_i)/a_i \sim \chi^2_{(2)} \quad (2.7)$$

yeterli istatistiklerin dağılımlarından elde edilen bağımsız istatistiklerdir (Viveros ve Balakrishnan, 1994). Bu nedenle, $\mu = a_i + b_i$ parametresinin UMVU tahmincisi aşağıdaki gibi ifade edilebilir.

$$\hat{\mu}_i = X_{i(1)} + \frac{n_i - 1}{n_i} S_i \sim N(b_i, a^2/n_i) \quad (2.8)$$

İki parametrelili üstel dağılmış k anakütlenin ortalamalarının eşitliğinin test edilmesi problemindeki H_0 hipotezi Eşitlik (2.9)'deki gibidir.

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k \quad (2.9)$$

burada $\mu_i = a_i + b_i$ 'dir. Ölçek parametrelerinin eşit olmaması durumunda ikiden fazla iki parametrelili üstel dağılımın konum parametrelerinin karşılaştırılmasına ilişkin Hsieh (1986), ortalamalarının karşılaştırılmasına ilişkin ise Malekzadeh ve Jafari (2019) ikinci tip ilerleyen sansürlü veriler için testler geliştirmişlerdir. Malekzadeh ve Jafari (2019) varyansların eşit olmaması durumunda, normal dağılmış anakütle ortalamalarının eşitliğinin test edilmesi için Cochran (1937) tarafından geliştirilen ve Welch (1951) tarafından Welch testinin geliştirilmesinde kullanılan Cochran test istatistiğinden faydalanmışlardır. Bu istatistik, Klasik F testi istatistiğinin homojen olmayan varyanslılık durumuna karşı bir ağırlık parametresi w_i ile düzeltilmiş halidir. Cochran test istatistiği kullanılarak eşit olmayan ölçek parametresi içeren ikiden fazla iki parametrelili üstel dağılım için Eşitlik (2.10)'deki istatistik oluşturulabilir.

$$T_t = \sum_{i=1}^k \frac{n_i \hat{\mu}_i^2}{S_i^2} - \frac{\left(\sum_{i=1}^k \frac{n_i \hat{\mu}_i}{S_i^2} \right)^2}{\sum_{i=1}^k \frac{n_i}{S_i^2}} \quad (2.10)$$

T_t istatistiği ilerleyen bölümlerde verilen GPD, PB, FY ve HY testlerinin p-değerlerinin hesaplanmasında kullanılır.

2.4.1. Genelleştirilmiş p değeri testi

Genelleştirilmiş p-değeri yaklaşımı Tsui ve Weerahandi (1989) tarafından geliştirilmiştir. İkiden fazla anakütle ortalamasının eşitliğinin test edilmesi için öncelikle Weerahandi (1995) tarafından normal dağılım gösteren anaküteller için Genelleştirilmiş F-testi geliştirilmiştir. Sonrasında log-normal, ters normal ve bazı çarpık dağılımlar için bu yöntem kullanılmıştır. *Nuisance* parametre içeren problemlerin çözümü için kullanılan bu yöntem ile *nuisance* parametreden bağımsız güçlü test istatistikleri geliştirilebilmektedir.

Bu yöntemde öncelikle ilgilenilen dağılımın parametrelerinin yeterli istatistikleri elde edilir. İki parametrelili üstel dağılımın yeterli istatistikleri kullanılarak, (i) *nuisance* parametreden bağımsız ve (ii) gözlenen değeri μ_i *nuisance* parametreden bağımsız olması nedeniyle ortalama parametresi θ_i için genelleştirilmiş pivot değeri R_i elde edilebilir. $(X_{i(1)}, S_i)$ vektörünün gözlenen değerleri cinsinden R_i genelleştirilmiş pivot değerinin beklenen değeri μ_{R_i} ve varyansı $\sigma_{R_i}^2$ aşağıdaki gibi elde edilir.

$$\mu_{Ri} = x_{i(1)} + \frac{(n_i - 1)^2 s_i}{n_i^2 - 2n_i} \quad (2.11)$$

$$\sigma_{Ri}^2 = \frac{(n_i - 1)^4 s_i^2}{n_i^2 (n_i - 2)^2} \left(\frac{1}{n_i - 3} \right) \quad (2.12)$$

Cochran test istatistiği, R_i genelleştirilmiş pivot değerinin beklenen değeri ve varyansı kullanarak Eşitlik (2.13)'deki gibi elde edilebilir.

$$T_{GPD} = \sum_{i=1}^k \frac{\mu_{Ri}^2}{\sigma_{Ri}^2} - \frac{\left(\sum_{i=1}^k \frac{\mu_{Ri}}{\sigma_{Ri}^2} \right)^2}{\sum_{i=1}^k \frac{1}{\sigma_{Ri}^2}} \quad (2.13)$$

eğer $T_{GPD} > T_t$ ise Eşitlik (2.9)'de verilen H_0 hipotezi reddedilir. GPD testinin p-değeri ise 10.000 Monte-Carlo denemesi ile aşağıdaki gibi hesaplanabilir.

$$p_{GPD} = P(T_{GPD} \geq T_t) \quad (2.14)$$

Eğer hesaplanan $p_{GPD} < \alpha_0$ ise H_0 hipotezi reddedilir.

2.4.2. Parametrik bootstrap testi

Parametrik Bootstrap yaklaşımı Krishnamoorthy vd. (2007) tarafından varyansların eşit olmadığı durumda normal dağılmış anakütle ortalamalarının eşitliğinin test edilmesi probleminin çözümünde kullanılmıştır. $V_i \sim \chi_{(2)}^2$ ve $U_i \sim \chi_{(2n_i-2)}^2$ dağılan rassal örneklemeler olsun. Cochran test istatistiğinden hareketle elde edilen PB test istatistiği Eşitlik (2.15)'teki gibi elde edilir.

$$T_{PB} = \sum_{i=1}^k \frac{n_i \mu_{Bi}^2}{S_i^2} - \frac{\left(\sum_{i=1}^k \frac{n_i \mu_{Bi}}{S_i^2} \right)^2}{\sum_{i=1}^k \frac{n_i}{S_i^2}} \quad (2.15)$$

burada $\mu_B = (s_i/2n_i)(V_i + U_i)$ ve $S_{Bi} = s_i U_i / (2n_i - 2)$. PB testinin p-değeri 10.000 Monte-Carlo denemesi ile aşağıdaki gibi hesaplanabilir.

$$p_{PB} = P(T_{PB} \geq T_t) \quad (2.16)$$

Eğer hesaplanan $p_{PB} < \alpha_0$ ise H_0 hipotezi reddedilir.

2.4.3. Fiducial yaklaşım testi

Fiducial Yaklaşımı Li vd. (2011) tarafından varyansların eşit olmadığı durumda normal dağılmış anakütle ortalamalarının eşitliğinin test edilmesi probleminin çözümünde kullanılmıştır. $U_i \sim \chi^2_{(2n_i-2)}$ ve $V_i \sim \chi^2_{(2)}$ dağılan rassal örneklemeler olsun. $X_{i(1)}$ ve S_i fonksiyonlarını rassal örneklemeler cinsinden şu şekilde yeniden düzenleyebiliriz:

$$S_i = \frac{\mu_i U_i}{2(n_i - 1)} \quad (2.17)$$

$$X_{i(1)} = \frac{\mu_i V_i}{2n_i} + a_i \quad (2.18)$$

$(x_{i(1)}, s_i)$ gözlenen değerleri üzerinden parametre tahminleri aşağıdaki gibi elde edilir.

$$a_i = x_{i(1)} - \frac{(n_i - 1)s_i v_i}{n_i u_i} \quad (2.19)$$

$$\mu_i = \frac{2(n_i - 1)s_i}{u_i}. \quad (2.20)$$

Cochran test istatistiğini kullanarak fiducial yaklaşımı ile test istatistiği T_{FY} , Eşitlik (2.21)'daki gibi oluşturulabilir.

$$T_{FY} = \sum_{i=1}^k \frac{f_i}{l_i n_i^2 u_i^2} - \frac{\left(\sum_{i=1}^k \frac{f_i}{l_i s_i n_i u_i} \right)^2}{\sum_{i=1}^k w_i} \quad (2.21)$$

burada $f_i = (n_i - 1)(u_i v_i - 2n_i u_i)$. FY testinin p-değeri 10.000 Monte-Carlo denemesi ile aşağıdaki gibi hesaplanabilir.

$$p_{FY} = P(T_{FY} \geq T_t) \quad (2.22)$$

Eğer hesaplanan $p_{FY} < \alpha_0$ ise H_0 hipotezi reddedilir.



3. İKİ PARAMETRELİ ÜSTEL DAĞILMIŞ ANAKÜTLE ORTALAMALARININ EŞİTLİĞİNİN TEST EDİLMESİ İÇİN ÖNERİLEN TESTLER

GPD, PB ve FY testlerine alternatif olarak bu çalışmada Hesaplamalı Yaklaşım yöntemine dayalı olarak ikiden fazla ($k > 2$) iki parametrelili üstel dağılmış anakütlenin ortalamalarının eşitliğinin testi probleminin çözümü için testler geliştirilecektir. Bu problemde ölçek parametresi *a nuisance* parametre olarak ele alınmıştır. Böylelikle ölçek parametrelerinin bilinmediği ya da eşit olmadığı durumlarda bile güçlü sonuçlar veren testler geliştirilmiş olacaktır.

Bu bölümde PB yönteminin özel bir hali olan Pal vd. (2007) tarafından geliştirilen Hesaplamalı Yaklaşım yöntemi tanıtılmış ve hesaplama algoritması verilmiştir. HY yönteminde kullanılan iki-parametrelili üstel dağılımın ölçek parametresinin kısıtlandırılmış en çok olabilirlik tahmincileri elde edilmiştir. Sonrasında ortalama değerinin hesaplanmasında kullanılan parametrelerin en çok olabilirlik ve cezalandırılmış en çok olabilirlik yöntemleri kullanılarak 4 farklı HY testi varyasyonu oluşturulmuştur. Varyasyonlar oluşturulurken ilk test önerisinde ilgilenilen ortalama teriminin ($\hat{\mu}_i = \hat{a}_i + \hat{b}_i$) en çok olabilirlik tahmincisi kullanılmış, ikinci test önerisinde μ teriminin en çok olabilirlik tahmininin yanı sıra Malekzadeh ve Jafari (2019) çalışmasında kullandıkları konum parametresinin (b_i) yansızlık düzeltmesi kullanılmış, üçüncü test önerisinde μ teriminin tahmininde ölçek parametresinin (a_i) en çok olabilirlik ve konum parametresinin cezalandırılmış en çok olabilirlik tahmincisi ile yansızlık düzeltmesi kullanılmış, dördüncü test önerisinde ise ölçek parametresinin en çok olabilirlik tahmincisi ve konum parametresinin cezalandırılmış en çok olabilirlik tahmincisi kullanılarak geliştirilmiştir. İzleyen bölümlerde önerilen dört test istatistiği tanıtılmıştır.

3.1. Hesaplamalı Yaklaşım Yöntemi

Hesaplamalı Yaklaşım yöntemini kullanarak, bir veri setine ait parametrik modele dayalı herhangi bir örnekleme dağılımına ihtiyaç duyulmadan test istatistiği geliştirilebilmektedir. Gelişen hesaplama teknolojisi sayesinde HY yöntemi son yıllarda sıklıkla farklı çalışmalarda kullanılmıştır. $H_0 : \theta = \theta_0$ nokta tahminine dayalı

bir yokluk hipotezinin $H_1 : \theta \neq \theta_0$ alternatif hipotezine karşı testinde $\zeta = (\theta, \delta)$ parametre uzayında δ *nuisance* parametre bulunması durumunda kullanılan test istatistiği $T = T(X; x, \zeta)$ olsun. HY yöntemi ile p-değeri aşağıda verilen algoritma kullanılarak hesaplanır.

HY yöntemi hesaplama algoritması:

- i. İlgilenilen ve *nuisance* parametrenin en çok olabilirlik tahmincileri $\hat{\theta}_{ML}$ ve $\hat{\delta}_{ML}$ hesaplanır.
- ii. Parametrelerin en çok olabilirlik tahmincileri kullanılarak test istatistiğinin gözlenen değeri $\hat{t}^*(x; x, \hat{\zeta}_{ML})$ hesaplanır.
- iii. H_0 hipotezinin doğruluğu altında ilgilenilen ve *nuisance* parametrenin kısıtlandırılmış (restricted) en çok olabilirlik tahmincileri uzayı $\zeta_{RML} = (\hat{\theta}_{RML}, \hat{\delta}_{RML})$ hesaplanır.
- iv. Parametrelerin kısıtlandırılmış en çok olabilirlik tahmincileri kullanılarak $X_{ij} \sim f(\hat{\theta}_{RML}, \hat{\delta}_{RML})$, $j = 1, 2, \dots, n_i$, $i = 1, 2, \dots, k$, M tane rassal örneklem oluşturulur.
- v. Oluşturulan rassal örneklem üzerinden test istatistiğinin gözlenen değerleri $\hat{t}^{(m)}(x; x, \hat{\zeta})$, $m = 1, 2, \dots, M$ hesaplanır.
- vi. Bir önceki adımda hesaplanan değerler üzerinden $p = \frac{1}{M} \sum_{m=1}^M I[\hat{t}^{(m)}(x; x, \hat{\zeta}_{RML}) \geq \hat{t}^*(x; x, \hat{\zeta}_{ML})]$ değeri hesaplanır.

HY hesaplama algoritması kullanılarak oluşturulan HY testlerinin p-değeri hesaplanabilir. Hesaplanan p-değeri ile belirlenen nominal seviyeye göre ele alınan H_0 hipotezinin reddedilip edilmeyeceğine karar verilir.

3.2. Önerilen Hesaplamalı Yaklaşım Testleri

Bu bölümde bir önceki bölümde tanıtılan HY yöntemine dayalı olarak bilinmeyen ya da eşit olmayan ölçek parametresinin var olması durumunda ikiden fazla iki parametrelili anakütle ortalamasının eşitliğinin test edilmesi için geliştirilen test istatistiği aşağıdaki gibidir.

$$T_{HY} = \sum_{i=1}^k \frac{n_i \mu_i^2}{a_i^2} - \frac{\left(\sum_{i=1}^k \frac{n_i \mu_i}{a_i^2} \right)^2}{\sum_{i=1}^k \frac{n_i}{a_i^2}} \quad (3.1)$$

HY testlerinin hesabında kullanılan iki parametrelili üstel dağılımın kısıtlandırılmış en çok olabilirlik tahmincileri aşağıdaki gibi elde edilmiştir:

$$f(x_i; a, b) = \frac{1}{a} \exp \left(- \frac{x_i - b}{a} \right), \quad x_i > b, \quad a > 0 \quad (3.2)$$

İki parametrelili üstel dağılımın olabilirlik fonksiyonu $L(\cdot)$ aşağıdaki gibidir:

$$\begin{aligned} L(a, b \mid x_1, x_2, \dots, x_n) &= f(x_1 \mid a, b) f(x_2 \mid a, b) \dots f(x_n \mid a, b) \\ &= \prod_{i=1}^n f(x_i \mid a, b) \\ &= \prod_{i=1}^n \frac{1}{a} \exp \left(- \frac{x_i - b}{a} \right) \\ &= \frac{1}{a^n} \exp \left[- \frac{1}{a} \sum_{i=1}^n (x_i - b) \right] \end{aligned} \quad (3.3)$$

burada kısıtlandırılmış en çok olabilirlik tahmincilerinin elde edilmesi için H_0 hipotezinin doğruluğu altında b parametresinin ilgilenilen ve *nuisance* parametre cinsinden $b = \mu - a$ şeklinde ifade edilebilir.

$$L(a, \mu - a \mid x_1, x_2, \dots, x_n) = \frac{1}{a^n} \exp \left[- \frac{1}{a} \sum_{i=1}^n (x_i - \mu + a) \right] \quad (3.4)$$

Olabilirlik fonksiyonunun logaritması alınır:

$$\begin{aligned}
\ln L(a, \mu - a \mid x_1, x_2, \dots, x_n) &= \prod_{i=1}^n \ln f(x_i \mid a, \mu - a) \\
&= \prod_{i=1}^n \ln \left[\frac{1}{a^n} \exp \left(-\frac{1}{a} \sum_{i=1}^n (x_i - \mu + a) \right) \right] \\
&= -\frac{1}{a} \sum_{i=1}^n x_i + \frac{n\mu}{n} - \frac{na}{a} - n \ln a \\
&= -\frac{1}{a} \sum_{i=1}^n x_i + \frac{n\mu}{n} - n - n \ln a
\end{aligned} \tag{3.5}$$

Log-olabilirlik fonksiyonunun ölçek parametresine göre türevi alınarak sıfıra eşitlenir, böylelikle en çok olabilirlik tahmincisi elde edilir.

$$\frac{d \ln L(a, \mu - a)}{da} = \frac{\sum_{i=1}^n (x_i - \mu)}{a^2} - \frac{n}{a} = 0 \tag{3.6}$$

Ölçek parametresinin kısıtlandırılmış en çok olabilirlik tahmincisi aşağıdaki gibi elde edilmiştir.

$$\hat{a}_{RML} = \mu - x_{(1)} \tag{3.7}$$

Ölçek parametresinin kısıtlandırılmış en çok olabilirlik tahmincisi üzerinden ortalama parametresinin kısıtlandırılmış en çok olabilirlik tahmincisi ise aşağıdaki gibi elde edilmiştir.

$$\hat{\mu}_{RML} = \frac{\sum_{i=1}^n x_i - x_{(1)}}{n - 1} + x_{(1)} \tag{3.8}$$

Burada elde edilen kısıtlandırılmış en çok olabilirlik tahmincileri HY testlerinin kritik değerlerinin hesaplanmasında kullanılmıştır. Varyasyonlar oluşturulurken ölçek

ve ortalama için en çok olabilirlik tahmincilerinin yanı sıra Zheng (2013) tarafından iki parametrelili üstel dağılımın ölçek ve konum parametresi için önerilen cezalandırılmış (penalized) en çok olabilirlik tahmincileri kullanılmıştır. Ölçek parametresi için cezalandırılmış en çok olabilirlik tahmincisi $\hat{a}_{i(PML)}$ ile konum parametresi tahmincisi ise $\hat{b}_{i(PML)}$ ile gösterilmiştir:

$$\hat{a}_{i(PML)} = \frac{\sum_{i=1}^n x_i - nx_{(1)}}{n-1} \quad (3.9)$$

$$\hat{b}_{i(PML)} = \frac{nx_{(1)} - \frac{\sum_{i=1}^n x_i}{n}}{n-1} \quad (3.10)$$

HY yöntemi kullanılarak iki parametrelili üstel dağılan anakütle ortalamalarının eşitliğinin test edilebilmesi için dört farklı test varyasyonu önerilmiştir. Varyasyonlar μ_i ortalama tahmininde kullanılan b_i konum parametresinin farklı tahmincileri ile oluşturulmuştur. Bu sayede iki parametrelili üstel dağılımın konum parametresinin en çok olabilirlik tahmincisinin yanlışlığının etkisinin arındırılması hedeflenmiştir. Bilindiği gibi yanlış en çok olabilirlik tahmincilerine alternatif olarak daha düşük hata kareler ortalamasına sahip cezalandırılmış en çok olabilirlik tahmincileri geliştirilmiştir (Zheng, 2013). Aşağıdaki sırasıyla geliştirilen Hesaplamalı Yaklaşım 1 testi (HY1), Hesaplamalı Yaklaşım 2 testi (HY2), Hesaplamalı Yaklaşım 3 testi (HY3), Hesaplamalı Yaklaşım 4 testi (HY4) varyasyonlar için izleyen bölümlerde hesaplama algoritmaları verilmiştir.

3.2.1. Hesaplamalı Yaklaşım 1 Testi

Bu testin hesabında parametrelerin en çok olabilirlik tahmincileri ve ortalama tahmini için $\hat{\mu}_i = \hat{a}_i + x_{(1)}$ kullanılmıştır. HY1 testi hesaplama algoritması aşağıdaki adımlardan oluşmaktadır.

HY1 testi algoritması:

- i. Ölçek parametresinin $\hat{a}_i = \sum_{i=1}^k (x_i - x_{(1)}) / (n_i - 1)$ ve ortalama parametresinin

- en çok olabilirlik tahminlerini $\hat{\mu}_i = \hat{a}_i + x_{(1)}$ hesapla
- ii. İlk adımda hesaplanan değerleri kullanarak test istatistiğinin gözlenen değerini t_{HY}^* hesapla.
- iii. H_0 hipotezinin doğruluğu altında ilgilenilen parametre $\hat{\mu}_{RML}$ ve *nuisance* parametre olan ölçek parametresinin $\hat{a}_{RML}$ kısıtlandırılmış en çok olabilirlik tahmincilerini hesapla.
- iv. Kısıtlandırılmış en çok olabilirlik tahmincilerini kullanarak $X_{ij} \sim \exp 2(\hat{a}_{RML}, \hat{\mu}_{RML} - \hat{a}_{RML})$, $j = 1, 2, \dots, n_i$, $i = 1, 2, \dots, k$, M tane rassal örneklem oluşturulur.
- v. Oluşturulan rassal örneklem üzerinden test istatistiğinin gözlenen değerleri $\hat{t}_{HY}^{(m)}$, $m = 1, 2, \dots, M$ hesaplanır.
- vi. Bir önceki adımda hesaplanan değerler üzerinden $p_{HY1} = \frac{1}{M} \sum_{m=1}^M I(\hat{t}_{HY}^{(m)} \geq \hat{t}_{HY}^*)$ değeri hesaplanır.

HY1 testi algoritması sonucunda elde edilen $p_{HY1} \leq \alpha_0$ ise H_0 hipotezi reddedilir.

3.2.2. Hesaplamalı Yaklaşım 2 Testi

Bu testin hesabında parametrelerin en çok olabilirlik tahmincileri, ortalama tahmini için ise Malekzadeh ve Jafari (2019) kullandığı yansızlık düzeltmesi ile $\hat{\mu}_i = \hat{a}_i + \frac{n_i-1}{n_i} \hat{b}_i$ tahmincisi kullanılmıştır. HY2 testi hesaplama algoritması aşağıdaki adımlardan oluşmaktadır:

HY2 testi algoritması:

- i. Ölçek parametresinin $\hat{a}_i = \sum_{i=1}^k (x_i - x_{(1)}) / (n_i - 1)$ ve ortalama parametresinin en çok olabilirlik tahminleri $\hat{\mu}_i = \hat{a}_i + \frac{n_i-1}{n_i} x_{(1)}$ hesaplanır.
- ii. İlk adımda hesaplanan değerler kullanarak test istatistiğinin gözlenen değerini t_{HY}^* hesaplanır.

- iii. H_0 hipotezinin doğruluğu altında ilgilenilen parametre $\hat{\mu}_{RML}$ ve *nuisance* parametre olan ölçek parametresinin $\hat{a}_{RML}$ kısıtlandırılmış en çok olabilirlik tahmincileri hesaplanır.
- iv. Kısıtlandırılmış en çok olabilirlik tahmincilerini kullanarak $X_{ij} \sim \exp2(\hat{a}_{RML}, \hat{\mu}_{RML} - \hat{a}_{RML})$, $j = 1, 2, \dots, n_i$, $i = 1, 2, \dots, k$, M tane rassal örneklem oluşturulur.
- v. Oluşturulan rassal örneklem üzerinden test istatistiğinin gözlenen değerleri $\hat{t}_{HY}^{(m)}$, $m = 1, 2, \dots, M$ hesaplanır.
- vi. Bir önceki adımda hesaplanan değerler üzerinden $p_{HY2} = \frac{1}{M} \sum_{m=1}^M I(\hat{t}_{HY}^{(m)} \geq \hat{t}_{HY}^*)$ değeri hesaplanır.

HY2 testi algoritması sonucunda elde edilen $p_{HY2} \leq \alpha_0$ ise H_0 hipotezi reddedilir.

3.2.3. Hesaplamalı Yaklaşım 3 Testi

Bu testin hesabında parametrelerin cezalandırılmış en çok olabilirlik tahmincileri ve ortalama tahmini için bir de yansızlık düzeltmesi $\hat{\mu}_i = \hat{a}_i + \frac{n_i-1}{n_i} \hat{b}_{i(PML)}$ kullanılmıştır. HY3 testi hesaplama algoritması aşağıdaki adımlardan oluşmaktadır:

HY3 testi algoritması:

- i. Ölçek parametresinin $\hat{a}_i = \sum_{i=1}^k (x_i - x_{(1)}) / (n_i - 1)$ ve ortalama parametresinin en çok olabilirlik tahminleri $\hat{\mu}_i = \hat{a}_i + \frac{n_i-1}{n_i} n_i (x_{(1)} - \frac{\sum_{i=1}^{n_i} x_i}{n_i}) / (n_i - 1)$ hesaplanır.
- ii. İlk adımda hesaplanan değerler kullanarak test istatistiğinin gözlenen değerini t_{HY}^* hesaplanır.
- iii. H_0 hipotezinin doğruluğu altında ilgilenilen parametre $\hat{\mu}_{RML}$ ve *nuisance* parametre olan ölçek parametresinin $\hat{a}_{RML}$ kısıtlandırılmış en çok olabilirlik tahmincileri hesaplanır.

- iv. Kısıtlandırılmış en çok olabilirlik tahmincilerini kullanarak $X_{ij} \sim \exp 2(\hat{a}_{RML}, \hat{\mu}_{RML} - \hat{a}_{RML})$, $j = 1, 2, \dots, n_i$, $i = 1, 2, \dots, k$, M tane rassal örneklem oluşturulur.
- v. Oluşturulan rassal örneklem üzerinden test istatistiğinin gözlenen değerleri $\hat{t}_{HY}^{(m)}$, $m = 1, 2, \dots, M$ hesaplanır.
- vi. Bir önceki adımda hesaplanan değerler üzerinden $p_{HY3} = \frac{1}{M} \sum_{m=1}^M I(\hat{t}_{HY}^{(m)} \geq \hat{t}_{HY}^*)$ değeri hesaplanır.

HY3 testi algoritması sonucunda elde edilen $p_{HY3} \leq \alpha_0$ ise H_0 hipotezi reddedilir.

3.2.4. Hesaplamalı Yaklaşım 4 Testi

Bu testin hesabında parametrelerin cezalandırılmış en çok olabilirlik tahmincileri ve ortalama tahmini için $\hat{\mu}_i = \hat{a}_i + \hat{b}_{i(PML)}$ kullanılmıştır. HY4 testi hesaplama algoritması aşağıdaki adımlardan oluşmaktadır:

HY4 testi algoritması:

- i. Ölçek parametresinin $\hat{a}_i = \sum_{i=1}^k (x_i - x_{(1)}) / (n_i - 1)$ ve ortalama parametresinin en çok olabilirlik tahminleri $\hat{\mu}_i = \hat{a}_i + n_i(x_{(1)} - \frac{\sum_{i=1}^{n_i} x_i}{n_i}) / (n_i - 1)$ hesaplanır.
- ii. İlk adımda hesaplanan değerler kullanarak test istatistiğinin gözlenen değerini t_{HY}^* hesaplanır.
- iii. H_0 hipotezinin doğruluğu altında ilgilenilen parametre $\hat{\mu}_{RML}$ ve *nuisance* parametre olan ölçek parametresinin $\hat{a}_{RML}$ kısıtlandırılmış en çok olabilirlik tahmincileri hesaplanır.
- iv. Kısıtlandırılmış en çok olabilirlik tahmincilerini kullanarak $X_{ij} \sim \exp 2(\hat{a}_{RML}, \hat{\mu}_{RML} - \hat{a}_{RML})$, $j = 1, 2, \dots, n_i$, $i = 1, 2, \dots, k$, M tane rassal örneklem oluşturulur.
- v. Oluşturulan rassal örneklem üzerinden test istatistiğinin gözlenen değerleri $\hat{t}_{HY}^{(m)}$, $m = 1, 2, \dots, M$ hesaplanır.

vi. Bir önceki adımda hesaplanan değerler üzerinden $p_{HY4} = \frac{1}{M} \sum_{m=1}^M I(\hat{t}_{HY}^{(m)} \geq \hat{t}_{HY}^*)$ değeri hesaplanır.

HY4 testi algoritması sonucunda elde edilen $p_{HY4} \leq \alpha_0$ ise H_0 hipotezi reddedilir. İzleyen bölümde tez çalışmasında geliştirilen Hesaplamalı Yaklaşım testleri ve literatürde yer alan GPD, PB ve FY testlerinin performansları Monte-Carlo simülasyon çalışmaları ile karşılaştırılmıştır.



4. MONTE-CARLO SİMÜLASYON ÇALIŞMALARI

Bu bölümde önerilen test istatistiklerinin performansları 1.tip hata olasılığı ve izleyen bölümde ele alınan sorunlar nedeniyle klasik güç yerine cezalandırılmış güç kriterleri kullanılarak simülasyon çalışmaları ile araştırılmıştır. Öncelikle literatürde yer alan alternatif güç kriterleri ele alınmış ve özellikleri verilmiştir. Verilen kriterlerin eksikleri tartışılmış ve cezalandırılmış güç kriterine neden ihtiyaç duyulduğu açıklanmıştır.

Sonrasında ise sırasıyla $k = 3, 4, 5$ sayıda anakütlenin ortalamalarının eşitliğinin testi için geliştirilen ve literatürde yer alan yöntemlerin 1.tip hata olasılığı ve cezalandırılmış güç değerleri verilmiştir. Simülasyon çalışmasında örneklem hacmi, ölçek parametresi değerleri, tasarım tipi ve etki büyüklükleri konfigürasyon parametreleri olarak belirlenmiştir. Örneklem hacimlerinde sırasıyla küçük, orta ve büyük örneklem hacimleri ele alınmıştır. Tasarım tipi olarak dengeli ve dengeli olmayan (dengesiz) tasarımlar kullanılmıştır.

Dengeli bir tasarım tüm anakütlelerin örneklem hacimlerinin eşit olduğu; dengeli olmayan bir tasarım ise örneklem hacimlerinin eşit olmadığı tasarım olarak tanımlanır. Dengeli olmayan tasarımlar farklı şekillerde ortaya çıkabilir. Bazı durumlarda dengeli tasarıma çok benzer iken, bazı durumlarda ise oldukça dengesiz olabilmektedir. Bu bölümde tüm örneklem hacimleri için dengeli tasarımlar ile iki farklı dengeli olmayan tasarım ele alınmıştır. Bu tasarımlar göreceli olarak dengeli olmayan ve en dengesiz tasarım olarak adlandırılmış; böylelikle tasarım dengesizliğinin artması durumunda testlerin performanslarındaki değişim gözlemlenebilmiştir. Tasarım dengesizliğine ilişkin tüm kıyaslamalar benzer örneklem hacimlerine göre kendi içerisinde değerlendirilmiştir.

Örneklem hacimleri $n = 10$ etrafında değiştiğinde küçük, $n = 30$ etrafında değiştiğinde orta ve $n = 50$ etrafında değiştiğinde ise büyük örneklem olarak ifade edilmiştir. Tasarım tipleri örneklem hacimleri $n_i = (n, n, n)$ olduğunda dengeli, $n_i = (0.8n, n, 1.2n)$ olduğunda dengeli olmayan ve $n_i = (0.5n, n, 1.5n)$ olduğunda en dengesiz tasarım olarak adlandırılmıştır. Ölçek parametresi için ise iki farklı düzey ele alınmış; $a = 2$ düşük, $a = 5$ yüksek ölçek parametresi değeri olarak belirlenmiştir. Etki büyüklüğünü belirlemek için her anakütle sayısı k için en son sırada yer alan

anakütlenin ölçek parametresi a_k değeri farklılaştırılarak 3 farklı etki büyüklüğü belirlenmiştir.

Her bir anakütle sayısı k için hesaplanan 1.tip hata olasılıkları ve düşük - yüksek ölçek parametresi değerleri için ayrı olarak cezalandırılmış güç değerleri tablolar halinde izleyen bölümlerde verilmiştir.

4.1. Değerlendirme Kriterleri

Bu bölümde testin gücüne alternatif olarak geliştirilen kriterler tanıtılmıştır. Literatürde yer alan kriterlerin eksiklikleri verilmiş, bu eksikliklere sahip olmayan ve tez çalışması sürecinde geliştirilen cezalandırılmış güç kriteri tanıtılmıştır. Alternatifleri ile karşılaştırılmış ve avantajları ortaya konmuştur.

Θ_0 ve Θ_1 alt uzayları Θ uzayının ayrık uzayları olmak üzere ve $H_0 : \theta \in \Theta_0$ hipotezine karşı $H_1 : \theta \in \Theta_1$ hipotezinin test edildiği durumu ele alalım. $x = (x_1, x_2, \dots, x_n)$ rassal bir örneklem ve $T(x) = T(x_1, x_2, \dots, x_n)$ rassal örneklem için bir test istatistiği, c_α ise α nominal düzeyi için kritik değer olsun. Eğer $T(x) > c_\alpha$ ise H_0 hipotezi reddedilir, değilse kritik değer yaklaşımı ile reddedilemeyeceğine karar verilir. H_0 hipotezinin testinde verilen karara ve hipotezin tutarlılığına göre dört sonuç ortaya çıkar: bunlardan ikisi olası doğru kararlar, diğer ikisi ise olası yanlış kararlardır. H_0 hipotezinin doğru olmasına rağmen reddedilmesi olasılığına 1. tip hata olasılığı denir ve ayrıca α tipi hata olarak da adlandırılır. 2.tip hata olasılığı ise yanlış olan H_0 hipotezinin reddedilmeme olasılığıdır ve β tipi hata olarak da adlandırılır. Bu durumda testin gücü ise yanlış olan H_1 hipotezinin reddedilmesi olasılığı yani $(1 - \beta)$ olarak tanımlanır.

Tablo 1. *Hipotez testlerinde hata türleri*

		H_0	
		Doğru	Yanlış
Karar	Red	1. tip hata (α)	Doğru karar ($1 - \alpha$)
	Red değil	Doğru karar ($1 - \beta$)	2. tip hata (β)

Birçok araştırmacı test geliştirme üzerine çalışmakta ve geliştirdiği yeni testlerin Monte-Carlo simülasyonlarıyla performansını araştırmaktadır. Bu araştırmalarda 1.tip hata olasılığı ve testin gücü performansın ölçülmesi için sık kullanılan kriterler-

dir. Bir testin gücü, yanlış olan H_0 hipotezinin reddedilme olasılığı ile ölçülür. Ancak doğru olan H_0 hipotezini reddetmeme becerisine de sahip olması gerekir. Bir hipotez testinin performansı araştırılırken ya da karşılaştırılırken bu iki beceri birlikte değerlendirilmeli yani olası iki hata da dikkate alınmalıdır. Parker vd. (2015) çalışmalarında bu duruma vurgu yapmışlardır, çünkü bir test nominal seviyeye yakın bir 1.tip hata olasılığına sahip olsa bile gücü yeterince yüksek olmayabilir. Literatürde 1.tip hata olasılığı ve gücü birlikte değerlendirmeden performans araştırması yapan çalışmalar bulunmaktadır (Karagoz, 2016; Karagoz ve Saracbası, 2016). Blanca vd. (2017) F-testinin performansını etkileyen durumları araştırırken yalnızca 1.tip hata olasılığını kriter olarak kullanmışlar ve testin gücünü yani ilgilenilen durumların testin gücünü de etkilediğini göz ardı etmiştir. Bu durum yazarların yalnızca tek tip hata üzerinde yoğunlaştıklarını göstermektedir. Bu nedenle araştırmacılar bir testin performansını değerlendirirken Neyman-Pearson teoremine uygun olarak nominal seviyeye göre 1.tip hata olasılığını kontrol ederken testin gücünün mümkün olduğunca yüksek olması olarak tanımlanan ideal durumu temel almalıdır. 1.tip hata olasılıklarının eşit olmadığı durumlarda testin gücünü kullanarak karşılaştırma yapılması doğru bir yaklaşım değildir. Bu problemi daha iyi anlamak için aynı H_0 hipotezinin testi için kullanılan A, B, C, D ve E testlerinin kullanıldığı durumu ele alalım. Bu testlere ilişkin aynı durumda hesaplanan 1. tip hata olasılığı ve güç değerleri Tablo 2’de verilmiştir. Buna göre H_0 hipotezinin testi için kullanılması gereken test hangisidir?

Tablo 2. *Testlerin 1.tip hata olasılıkları*

Test	$1 - \beta_i$	α_i
A	0.95	0.06
B	0.90	0.04
C	0.99	0.07
D	0.85	0.05
E	0.90	0.03

Neyman-Pearson teoremine göre bir test 1.tip hata olasılığını kontrol edebilirken en yüksek güce sahipse en iyi performans gösteren test olarak tanımlanır. Tablo 2’de verilen sonuçlara göre, en yüksek güç değerine sahip test C aynı zamanda nominal

seviyeye göre 1.tip hata olasılığını kontrol etmede en başarılı testtir. Ancak D testi 1.tip hata olasılığını en iyi kontrol eden test iken en düşük güç değerine sahiptir. Açıkça görülmektedir ki, eşit ya da benzer olmayan 1.tip hata olasılıkları testlerin güç kıyaslamalarını geçersiz hale getirmektedir. Bu durum testlerin performanslarının kıyaslanmasında sık karşılaşılan bir problemdir.

Literatürde bu problem için farklı çözümler yer almaktadır. Bunlardan biri Zhang ve Boos (1994) hipotez testi için Monte-Carlo simülasyonları ile kritik değeri yeniden hesaplayan düzeltilmiş (*adjusted*) güç kriteridir. Dielman ve Rose (1997a; 1997b) ve Dielman (2005; 2006; 2009) geçerli kıyaslamalar için testlerin boylarının (1.tip hata olasılığı) benzer olması gerektiğini, eğer değilse testin gücünün yerine düzeltilmiş güç kriterinin kullanılması gerektiğini belirtmiştir. Cressie vd. (2003) ise çözüm olarak testin gücü ve 1.tip hata olasılığı arasındaki farkı kullanmış ve farkın ne kadar hızlı bir artış gösterdiğini dikkate almıştır. Bir diğer kriter ise Lloyd (2005)'un ROC eğrileri ile testin gücünün, 1.tip hata olasılığının nominal seviyeye olan uzaklığına göre asıl (*intrinsic*) güç adımı verdiği çözümüdür. Constantini ve Lupi (2013) tahmin edilen ROC eğrisi üzerindeki bir noktanın 1.tip hata olasılığına karşılık gelen güç değerini temsil ettiğini vurgular. Trinquart vd. (2014) farklı 1.tip hata olasılıklarını hesaba katan alternatif düzeltilmiş bir güç kullanmıştır. Heritier vd. (2017) eşit olmayan 1.tip hata olasılığı durumunda testlerinin performanslarının kıyaslanmasının geçerli olmayacağını vurgulayarak testin boyuna göre düzeltilmiş bir güç kriteri kullanmıştır. Görüldüğü gibi bu problem literatürde sıkça ele alınmış ve son yıllarda testin gücüne alternatif olarak geliştirilen kriterler sıklıkla kullanılmıştır.

İzleyen bölümde literatürde yer alan Zhang ve Boos (1994) tarafından önerilen düzeltilmiş güç ve Lloyd (2005) tarafından geliştirilen esas güç kriterleri verilmiştir. Bu kriterlerin yetersiz kaldığı noktaların var olması nedeniyle tez çalışması sürecinde geliştirilen cezalandırılmış güç kriteri tanıtılmıştır.

4.1.1. Zhang ve Boos'un düzeltilmiş güç kriteri

Zhang ve Boos (1994) kritik değere göre yeniden düzenlenen ve böylece 1.tip hata olasılığının sabit tutulabildiği düzeltilmiş (*adjusted*) güç adımı verdikleri bir kriter geliştirmiştir. Nominal düzeyin α_0 olarak belirlendiği ve c_α kritik değerine göre karar verilen bir T test istatistiğini ele alalım. T test istatistiğinin H_0 hipote-

zinin doğruluğu altında belirlenen nominal düzeye göre gözlenen 1.tip hata olasılığı aşağıdaki gibi hesaplanır:

$$\hat{p}_0 = \frac{1}{k} \sum_{i=1}^k I(T_{0i} \geq c_\alpha), \quad i = 1, 2, \dots, k \quad (4.1)$$

burada $I(.)$ bir işaret fonksiyonunu ve $T_{01}, T_{02}, \dots, T_{0k}$ ise k Monte-Carlo adımımda hesaplanan T test istatistiğinin gözlenen değerleridir. Hesaplanan değerlere göre α_0 nominal düzeye için düzeltilmiş kritik değer c_α^* aşağıdaki gibi belirlenir:

$$c_\alpha^* = T_0^{-1}(\alpha_0) \quad (4.2)$$

Düzeltilmiş güç ise bir önceki fonksiyonda yer alan c_α kritik değerinin c_α^* düzeltilmiş kritik değer ile yer değiştirilmesi ile hesaplanır.

$$\hat{p}_1 = \frac{1}{k} \sum_{i=1}^k I(T_{1i} \geq c_\alpha^*), i = 1, 2, \dots, k \quad (4.3)$$

Böylelikle istenilen nominal düzeye göre 1.tip hata olasılığı eşit olacak şekilde testlerin güç değerleri hesaplanabilmektedir. Chang ve Suchindran (1997), Chen ve Wei (2003), Chen (2006) ve Lilienthal vd. (2018) çalışmalarının Monte-Carlo simülasyonlarında bu kriteri kullanmışlardır. Ancak düzeltilmiş güç kriterinin bazı önemli eksiklikleri bulunmaktadır. Bunlardan ilki hesaplama adımları bakımından maliyetli olmasıdır. Özellikle geniş kapsamlı simülasyon çalışmalarında gereken süreyi birkaç katına çıkarabilmektedir. İkincisi ise aynı test istatistiğine sahip ancak farklı serbestlik derecesi olan testlerin performanslarının karşılaştırılmasında işlevsizdir. Bunun nedeni ise testin gücüne düzeltme uygularken yalnızca test istatistiğini göz önüne alması ve serbestlik derecesini göz ardı etmesidir. Bu durumu ortaya koyan en önemli örnek Cavus vd. (2019a) çalışmasında ele alınan Brown-Forsythe (BF) ve Modified Brown-Forsythe (MBF) testi karşılaştırmasıdır. MBF testi Mehrotra (1997) tarafından BF testine yalnızca serbestlik derecesi üzerinde yapılan bir düzenleme önerisidir. Yani iki testin de test istatistikleri aynı ancak serbestlik dereceleri farklı-

dır. Bu durum düzeltilmiş güç kriterinin her iki test içinde aynı koşullar altında aynı değeri döndürmesine neden olmaktadır. Bu durum bahsedildiği gibi kriterin işlevsiz olmasına sebep olmaktadır.

4.1.2. Lloyd'un asıl güç kriteri

Wilk ve Gnanadesikan (1968) tarafından ilk kez kullanılan 1.tip hata olasılığı - güç eğrileri (*size-power curve*) ile testlerin güçlerini karşılaştırmak 1.tip hata olasılığının çok yüksek olduğu durumlarda bile başvurulacak yöntemlerden biridir. Bu eğriler ROC (*Receiver Operating Characteristic*) eğrilerine benzemektedirler. Bu noktadan hareketle Lloyd (2005) testlerin farklı nominal seviyelerde güçlerinin 1.tip hata olasılıklarına karşı değişimleri görselleştirmek için kullanmıştır. Geliştirdiği asıl (*intrinsic*) güç kriterinin bu eğri üzerinde aldığı değerleri ise probit fonksiyonunu kullanarak aşağıdaki gibi hesaplamıştır.

$$L(\alpha_0) = \phi[\phi^{-1}(\beta_i) - \phi^{-1}(\alpha_i) - \phi^{-1}(\alpha_0)] \quad (4.4)$$

burada ϕ probit fonksiyonunu, β_i 2. tip hata olasılığını ve α_i ise 1.tip hata olasılığını göstermektedir. Akıns vd. (2005), Wang ve Qu (2007), Hanson ve Hall (2012), Parker vd. (2015), Trinquart vd. (2014), Heritier vd. (2017) ve Dag, vd. (2018), Lloyd'un önerisini kullanarak testin gücünü 1.tip hata olasılığının nominal seviyeye olan uzaklığına göre yeniden hesapladılar. Constantini ve Lupi (2013) ROC eğrisi üzerindeki herhangi bir noktanın nominal seviyeye göre verilen 1.tip hata olasılığını kullanarak esas gücün tahmin edilebileceğini vurgulamıştır. Jack ve Kokoszka (2017) ise 1.tip hata olasılığının nominal seviyeye çok yakın olduğu durumlarda bu kriteri kullanmışlardır.

Bir önceki bölümde verilen düzeltilmiş güç kriterine benzer olarak asıl güç kriterinin de bazı eksiklikleri bulunmaktadır. Bunlardan ilki Neyman-Pearson teoremini her koşulda dikkate almaması ve 1.tip hata olasılığının nominal seviyede negatif yönde uzaklaşmasına rağmen güç fonksiyonunun pozitif olarak düzeltmeye devam etmesidir. Bu durum araştırmacıların testlerin performansını kıyaslarken yanlış kararlar vermesine neden olabilmektedir. Diğeri ise maksimum güç değerinde iş-

levsiz olmasıdır. Eğer bir testin güç değeri 1 olursa bu kriter 1.tip hata olasılığına bakmaksızın sonsuz değerini döndürmektedir. Bu olumsuz sonuç Cavus vd. (2019a) çalışmalarında ele alınmış ve uygulama çalışması ile ortaya konmuştur.

4.1.3. Cezalandırılmış güç kriteri

Zhang ve Boos (1994) ve Lloyd (2005)'un önerdikleri kriterler bazı durumlarda kullanışlı olmalarına rağmen eksiklikleri bulunmaktadır. Lloyd'un önerisi Neyman-Pearson teoremini her koşulda dikkate almaz ve 1.tip hata olasılığı nominal seviyede negatif yönde uzaklaşmasına rağmen güç fonksiyonuna pozitif oranda düzeltme yapmaya devam eder. Zhang ve Boos'un önerisi ise hesaplama adımlarının yavaşlığı bakımından Monte-Carlo simülasyon çalışmalarında kullanışlı değildir. Ayrıca test istatistikleri üzerinden bir düzeltme yaptığı için aynı test istatistiğine sahip ancak farklı serbestlik dereceli asimptotik testlerin karşılaştırılmasında kullanılamamaktadır. Bu sorunlar nedeniyle tez çalışması sürecinde aşağıda verilen cezalandırılmış güç (*penalized power*) kriteri geliştirilmiştir (Cavus vd., 2019a).

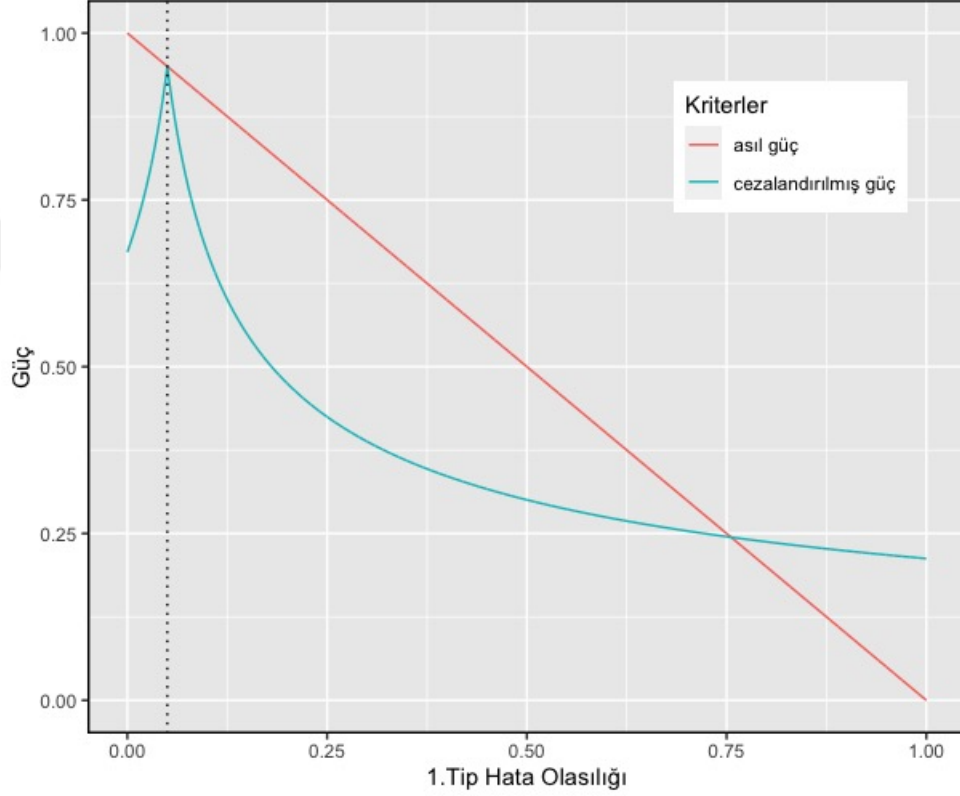
$$\gamma_i = \frac{1 - \beta_i}{\sqrt{1 + |1 - \frac{\alpha_i}{\alpha_0}|}} \quad (4.5)$$

burada β_i 2.tip hata olasılığını, α_i 1. tip hata olasılığını ve α_0 ise nominal seviyeyi göstermektedir. Cezalandırılmış güç, güç fonksiyonunu 1.tip hatanın nominal seviyeden yüzdesel sapmasının karekökü ile cezalandırır. Sapmayı ölçmek için yukarıda verilen formülün paydasında verilen mutlak yüzdesel bir ölçü kullanılır. Böylelikle 1.tip hata olasılıkları birbirinden farklı olan test ya da testlerin güç değerleri Neyman-Pearson teoremine uygun olarak kıyaslanabilir. Ancak testlerin performansları değerlendirilirken cezalandırılmış güç ve 1.tip hata olasılığının birlikte kullanılması önerilir. Cezalandırılmış güç kriterinin özellikleri şu şekildedir:

- i. 0 ve 1 aralığında değerler alır: $\gamma_i \in [0, 1]$
- ii. 1.tip hata olasılığı nominal seviyeye eşit olması durumunda ($\alpha_i = \alpha_0$) testin gücüne eşit olur: $\alpha_i = \alpha_0$, eğer $\gamma_i = 1 - \beta_i$

Görüldüğü gibi cezalandırılmış güç test istatistiği üzerinden değil testin gücü ve 1.tip

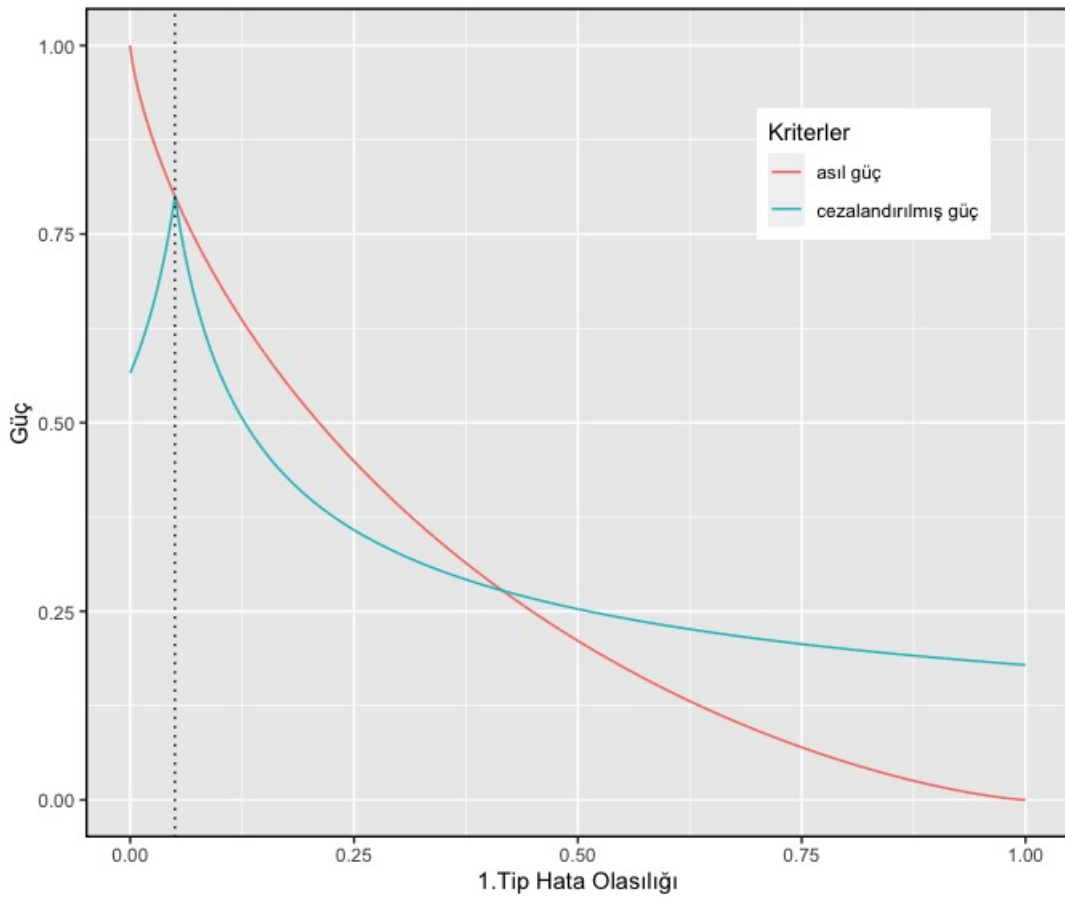
hata olasılığı üzerinden bir düzeltme yapmaktadır. Bu nedenle düzeltilmiş güç kriterinin aynı test istatistiğine sahip yöntemlerin karşılaştırılmasında herhangi bir sorun ile karşılaşmamaktadır. Diğer yandan testin gücünü 1.tip hata olasılığının nominal seviyeye olan uzaklığına göre cezalandırdığı için asıl güç kriterinin bu noktadaki eksikliği ile karşılaşmamaktadır. Bu durumu daha detaylı incelemek için aşağıda verilen şekilleri inceleyelim.



Şekil 2. Cezalandırılmış ve asıl güç kriterlerinin testin gücü (0.95) sabit olduğunda aldığı değerler

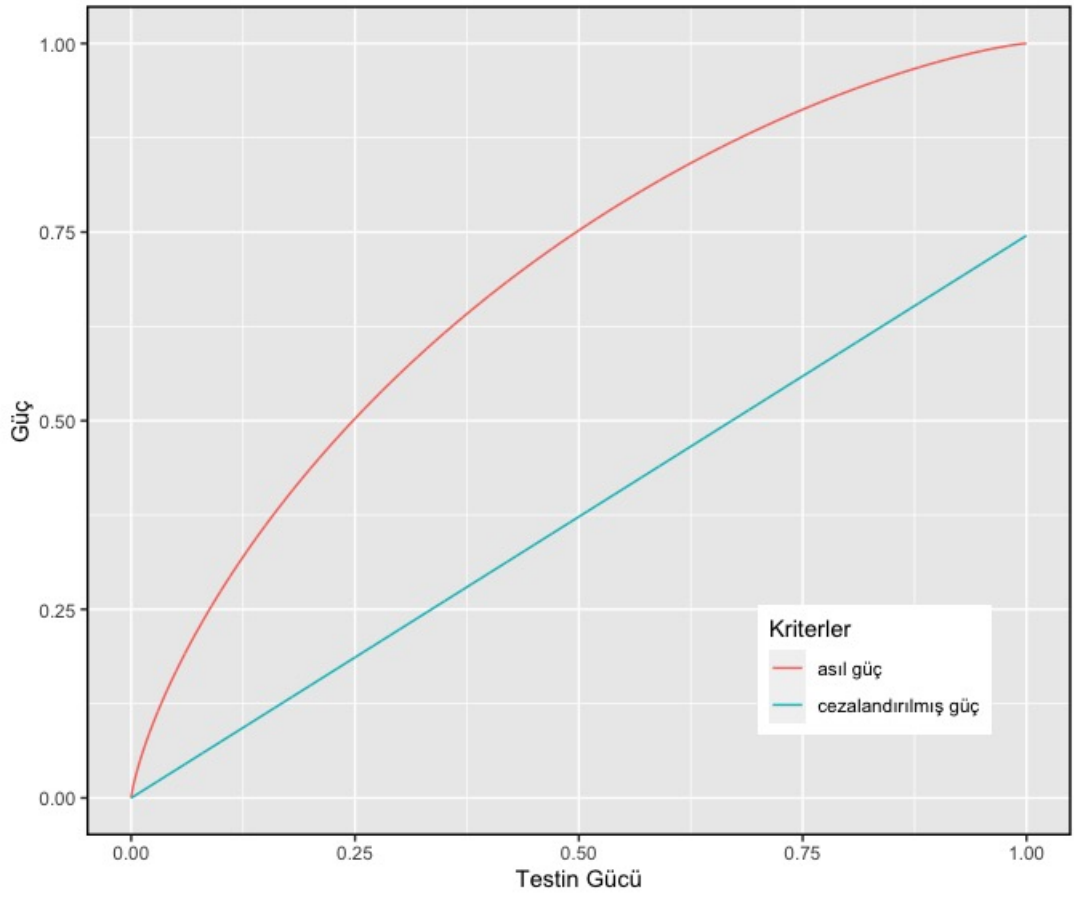
Şekil 2'de testin gücü 0.95 ve nominal seviye 0.05 olduğunda değişen 1.tip hata olasılığı değerlerine göre cezalandırılmış ve asıl güç kriterlerinin aldığı değerler verilmiştir. Cezalandırılmış güç 1.tip hata olasılığının nominal seviyeye yakınlığına göre testin gücünü cezalandırırken, asıl güç kriteri bu durumu dikkate almamaktadır. Cezalandırılmış güç kriterine göre maksimum güç değeri 1.tip hata olasılığının nominal seviyeye eşit olduğu durumda gözlenmekte ve 0.95 değerini almaktadır. Ancak asıl güç kriteri ise 1.tip hata olasılığının sıfıra eşit olduğu durumda maksimum güç değeri gözlenmekte ve 1 değerini almaktadır. Bu durum testin gücünün sabit ve 0.95 alındığı bu grafikte gerçek dışı bir sonuç vermektedir.

Benzer durumu bir de testin gücü 0.80 ve nominal seviye 0.05 olduğunda kriterlerin aldığı değerler üzerinden inceleyelim. Şekil 3’de verilen kriterlerin aldığı değerlere göre kriterlerin testin gücünü cezalandırma yapıları önceki grafiktekine benzerdir. Testin gücünün sabit ve 0.80 olduğu bu grafikte ise asıl güç değeri 1.tip hata olasılığının sıfıra yaklaşması durumunda 1’e yaklaşımaktadır. Bu durum gerçek güç değerinin olduğundan büyük görülmesine neden olmaktadır. İki grafik arasındaki tek farklılık testin gücü 0.95 ve nominal seviye 0.05 olduğu duruma göre asıl güç kriterinin testin gücüne doğrusal bir cezalandırma yapmamasıdır.



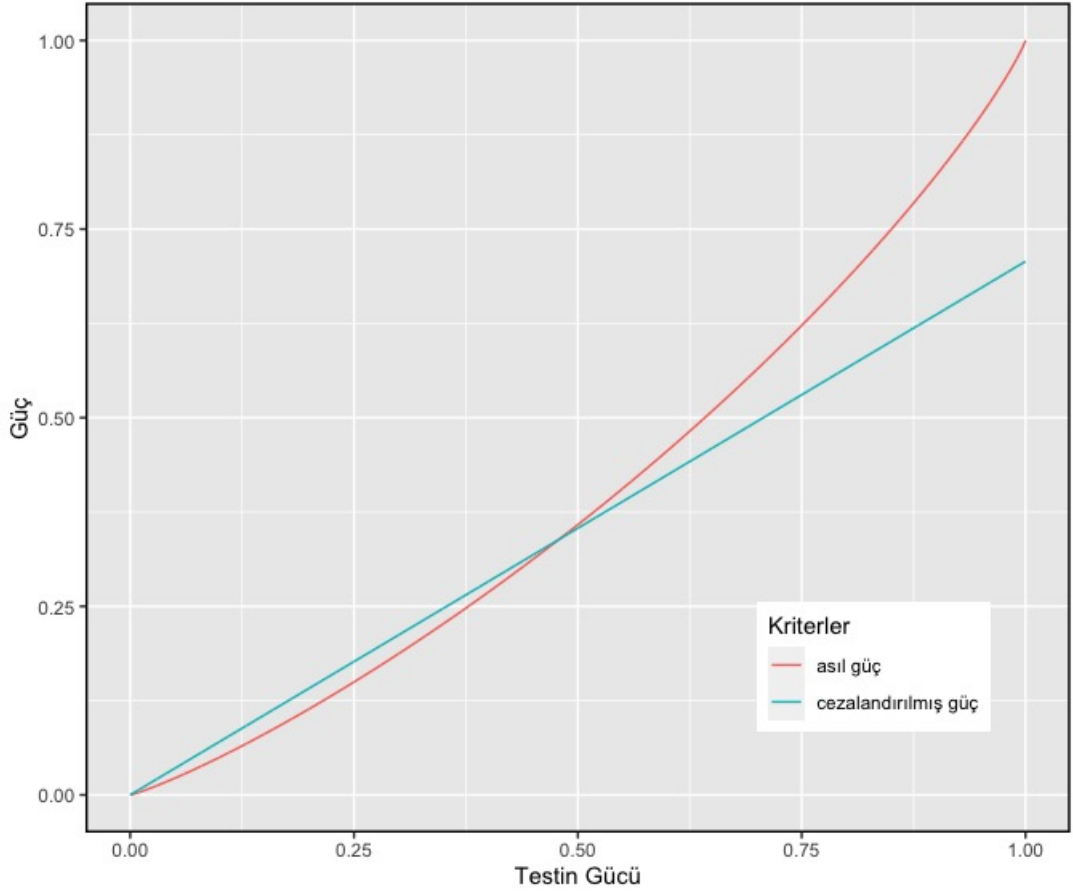
Şekil 3. Cezalandırılmış ve asıl güç kriterlerinin testin gücü (0.80) sabit olduğunda aldığı değerler

Şekil 4’te 1.tip hata olasılığının 0.10 ve nominal seviyenin 0.05 olduğu durumda gerçek güç değerlerine karşı kriterlerin verdiği değerlerin değişimi görülmektedir. Kriterlerin verdiği değerler karşılaştırıldıklarında 1.tip hata olasılığının nominal seviyeden yüksek olması durumunda cezalandırılmış ve asıl güç değerleri birbirine yakındır.



Şekil 4. Cezalandırılmış ve asıl güç kriterlerinin 1.tip hata olasılığı (0.10) sabit olduğunda aldığı değerler

Şekil 5'te 1.tip hata olasılığının 0.01 ve nominal seviyenin 0.05 olduğu durumda gerçek güç değerlerine karşı kriterlerin verdiği değerlerin değişimi görülmektedir. Kriterlerin verdiği değerler karşılaştırıldıklarında 1.tip hata olasılığının nominal seviyeden düşük olması durumunda asıl güç değerleri cezalandırılmış güç değerlerinden oldukça yüksektir.



Şekil 5. Cezalandırılmış ve asıl güç kriterlerinin 1.tip hata olasılığı (0.01) sabit olduğunda aldığı değerler

Sonuç olarak cezalandırılmış güç kriterinde alternatif kriterlerin sahip olduğu eksiklikler gözlemlenmediği için testin gücüne en iyi alternatiftir. Bu nedenle Monte-Carlo simülasyon çalışmasında performansı araştırılan testlerin 1.tip hata olasılıklarının eşit olmaması nedeniyle testin gücünün yerine kullanılacaktır. Bu sayede Neyman-Pearson teoremine uygun geçerli performans kıyaslamaları yapılabilecektir.

4.2. 1. Tip Hata Olasılıkları

Hipotez testleri için geliştirilen test istatistiklerinin gösterdiği performans, testlerin kullanılabilirliği açısından önemlidir. Performans, bir testin 1.tip hata olasılığının nominal seviyeye göre kontrol edebilmesi ve aynı zamanda gücünün mümkün olduğunca yüksek olması ile ölçülür. Bu bölümde çalışma süresince geliştirilen testler ve literatürde yer alan alternatiflerinin nominal seviyeye göre 1.tip hata olasılığını

kontrol edebilme özellikleri araştırılmıştır. İncelemelerde 1.tip hata olasılığı nominal seviyenin altında olan testler tutucu, üzerinde olanlar ise liberal testler olarak adlandırılmaktadır. Simülasyon kurgusunda ölçek parametresi a_i , konum parametresi b_i , örneklem hacmi n_i ve anakütle sayısı k konfigürasyon parametreleri olarak belirlenmiştir. Ayrıca dengeli ve dengeli olmayan tasarımlar oluşturulmuş ve tasarımlarda örneklem hacimleri arasındaki farklar arttırılarak; tasarım dengesizliğinin performans üzerindeki etkisi araştırılmıştır. Nominal seviye $\alpha_0 = 0.05$ ve konum parametreleri $b_i = 1, 1, 1$ olarak alınmıştır. $k = 3$ anakütle ortalamasının eşitliğinin test edilmesi probleminin çözümü için önerilen ve kullanılan testlerin 1.tip hata olasılıkları Tablo 3'te verilmiştir.

Tablo 3. Önerilen testlerin $k = 3$ anakütle için 1.tip hata olasılıkları

n_i	a_i	b_i	GPD	PB	FY	HY1	HY2	HY3	HY4
10, 10, 10	2, 2, 2	1, 1, 1	0.006	0.008	0.001	0.046	0.042	0.012	0.038
	5, 5, 5	1, 1, 1	0.006	0.010	0.001	0.055	0.059	0.017	0.041
8, 10, 12	2, 2, 2	1, 1, 1	0.006	0.012	0.002	0.065	0.072	0.031	0.039
	5, 5, 5	1, 1, 1	0.008	0.013	0.002	0.053	0.056	0.039	0.053
5, 10, 15	2, 2, 2	1, 1, 1	0.014	0.043	0.007	0.053	0.055	0.041	0.054
	5, 5, 5	1, 1, 1	0.012	0.043	0.007	0.055	0.057	0.044	0.043
30, 30, 30	2, 2, 2	1, 1, 1	0.026	0.040	0.020	0.055	0.057	0.049	0.055
	5, 5, 5	1, 1, 1	0.029	0.046	0.022	0.055	0.054	0.048	0.053
24, 30, 36	2, 2, 2	1, 1, 1	0.030	0.045	0.023	0.056	0.054	0.042	0.045
	5, 5, 5	1, 1, 1	0.029	0.045	0.022	0.057	0.056	0.019	0.045
15, 30, 45	2, 2, 2	1, 1, 1	0.033	0.048	0.024	0.055	0.059	0.017	0.041
	5, 5, 5	1, 1, 1	0.032	0.052	0.024	0.065	0.050	0.031	0.039
50, 50, 50	2, 2, 2	1, 1, 1	0.036	0.047	0.031	0.053	0.056	0.039	0.053
	5, 5, 5	1, 1, 1	0.034	0.045	0.030	0.053	0.055	0.041	0.054
40, 50, 60	2, 2, 2	1, 1, 1	0.035	0.049	0.031	0.055	0.057	0.044	0.043
	5, 5, 5	1, 1, 1	0.036	0.049	0.030	0.055	0.057	0.049	0.055
25, 50, 75	2, 2, 2	1, 1, 1	0.039	0.049	0.030	0.055	0.054	0.048	0.053
	5, 5, 5	1, 1, 1	0.040	0.051	0.033	0.051	0.048	0.041	0.041

Tablo 3'te elde edilen sonuçlar incelendiğinde küçük örneklem hacimlerinde HY1, HY2 ve HY4 testleri 1.tip hata olasılığını kontrol edebilirken PB testi orta ve büyük örneklem hacimlerinde benzer performans göstermektedir. GPD ve FY test-

leri ise yalnızca büyük örneklem hacimlerinde 1.tip hata olasılığını tatmin edici bir seviyede kontrol edebilmektedir. Genel olarak testlerin 1.tip hata olasılığını kontrol edebilme becerileri karşılaştırılacak olursa incelenen tüm durumlarda GPD, PB, FY ve HY3 testleri tutucu; HY1, HY2 ve HY4 testleri ise çoğunlukla liberal, bazı durumlarda ise tutucu testlerdir.

Anakütle sayısı $k = 4$ olduğu durumda elde edilen 1.tip hata olasılığı değerleri Tablo 4'te verilmiştir.

Tablo 4. Önerilen testlerin $k = 4$ anakütle için 1.tip hata olasılıkları

n_i	a_i	b_i	GPD	PB	FY	HY1	HY2	HY3	HY4
10, 10, 10, 10	2, 2, 2, 2	1	0.008	0.011	0.001	0.071	0.067	0.020	0.058
	5, 5, 5, 5	1	0.008	0.011	0.001	0.071	0.067	0.020	0.058
6, 8, 10, 14	2, 2, 2, 2	1	0.008	0.018	0.003	0.082	0.079	0.023	0.048
	5, 5, 5, 5	1	0.008	0.018	0.003	0.082	0.079	0.023	0.048
4, 8, 12, 16	2, 2, 2, 2	1	0.009	0.037	0.005	0.088	0.089	0.026	0.042
	5, 5, 5, 5	1	0.009	0.037	0.005	0.088	0.089	0.027	0.042
30, 30, 30, 30	2, 2, 2, 2	1	0.033	0.044	0.027	0.059	0.056	0.040	0.057
	5, 5, 5, 5	1	0.033	0.044	0.027	0.059	0.056	0.040	0.057
18, 24, 30, 42	2, 2, 2, 2	1	0.028	0.039	0.018	0.050	0.052	0.039	0.044
	5, 5, 5, 5	1	0.028	0.039	0.018	0.050	0.052	0.039	0.044
12, 24, 36, 48	2, 2, 2, 2	1	0.031	0.051	0.029	0.054	0.057	0.051	0.049
	5, 5, 5, 5	1	0.031	0.051	0.029	0.054	0.057	0.051	0.049
50, 50, 50, 50	2, 2, 2, 2	1	0.029	0.037	0.026	0.060	0.059	0.051	0.060
	5, 5, 5, 5	1	0.029	0.037	0.026	0.060	0.059	0.051	0.060
30, 40, 60, 70	2, 2, 2, 2	1	0.041	0.055	0.038	0.058	0.057	0.047	0.056
	5, 5, 5, 5	1	0.041	0.055	0.038	0.058	0.057	0.047	0.056
20, 40, 60, 80	2, 2, 2, 2	1	0.044	0.050	0.029	0.052	0.048	0.041	0.042
	5, 5, 5, 5	1	0.044	0.050	0.029	0.052	0.048	0.041	0.042

Tablo 4'te verilen sonuçlar incelendiğinde anakütle sayısının 3'ten 4'e çıkması HY4 testi dışında tüm testlerin performansını olumsuz etkilemiştir. Genel olarak GPD, PB, FY ve HY3 testi tutucu bir yapı sergilerken, HY1 ve HY2 testinin liberal bir yapı sergilediği gözlenmiştir. Küçük örneklem hacimlerinde HY4 testi dışında tüm testlerin 1. tip hata olasılıkları nominal seviyeden uzaktır. Orta örneklem hacimlerinde ise GPD ve FY testinin 1. tip hata olasılığının nominal seviyeden uzak

olduğu görülmektedir. Büyük örneklem hacimlerinde FY testinin, yanı sıra dengeli tasarımlarda ise buna ek olarak GPD ve PB testinin 1. tip hata olasılığı nominal seviyeden uzaktır. Tasarım dengesizliği GPD, PB ve FY testlerinin performansını olumsuz etkilemez iken HY testlerinin 1.tip hata olasılığını arttırıcı bir etkisi olduğu görülmektedir. Bu durum HY testlerinin 1.tip hata olasılığının nominal seviyenin üzerine çıkmasına neden olmaktadır. Dengeli tasarımlarda genel olarak HY testleri daha iyi performans göstermekte, tüm durumlarda HY4 testinin 1.tip hata olasılığının nominal seviyeye çok yakın olduğu görülmektedir. Dengesiz tasarımlarda PB testi yalnızca orta örneklem hacimlerinin göreceli olarak daha az dengesiz tasarımı-
mında ve büyük örneklem hacimlerinin dengesiz tasarımlarında 1.tip hata olasılığı-
nın nominal seviyeye yakın olduğu görülmektedir. Ölçek parametresi a_i değerindeki artışın ise testlerin 1.tip hata olasılığını kontrol etme becerisi üzerinde herhangi bir etkisi olmadığı gözlenmektedir.

Anakütle sayısı $k = 5$ olduğu durumda elde edilen 1.tip hata olasılığı değerleri Tablo 5'te verilmiştir.

Tablo 5. Önerilen testlerin $k = 5$ anakütle için 1.tip hata olasılıkları

n_i	a_i	b_i	GPD	PB	FY	HY1	HY2	HY3	HY4
10, 10, 10, 10, 10	2, 2, 2, 2, 2	1	0.005	0.008	0.002	0.063	0.061	0.016	0.055
	5, 5, 5, 5, 5	1	0.005	0.008	0.002	0.063	0.061	0.016	0.055
6, 8, 10, 12, 14	2, 2, 2, 2, 2	1	0.006	0.019	0.001	0.070	0.073	0.020	0.049
	5, 5, 5, 5, 5	1	0.006	0.019	0.001	0.070	0.073	0.020	0.049
4, 6, 10, 14, 16	2, 2, 2, 2, 2	1	0.017	0.044	0.010	0.093	0.096	0.028	0.040
	5, 5, 5, 5, 5	1	0.017	0.044	0.010	0.093	0.096	0.028	0.040
30, 30, 30, 30, 30	2, 2, 2, 2, 2	1	0.022	0.036	0.012	0.052	0.053	0.037	0.052
	5, 5, 5, 5, 5	1	0.022	0.036	0.012	0.052	0.053	0.037	0.052
18, 24, 30, 36, 42	2, 2, 2, 2, 2	1	0.038	0.055	0.025	0.057	0.058	0.044	0.052
	5, 5, 5, 5, 5	1	0.038	0.055	0.025	0.057	0.058	0.044	0.052
12, 18, 30, 42, 48	2, 2, 2, 2, 2	1	0.040	0.047	0.025	0.050	0.051	0.034	0.033
	5, 5, 5, 5, 5	1	0.040	0.047	0.025	0.050	0.051	0.034	0.033
50, 50, 50, 50, 50	2, 2, 2, 2, 2	1	0.030	0.043	0.024	0.046	0.047	0.038	0.047
	5, 5, 5, 5, 5	1	0.030	0.043	0.024	0.046	0.047	0.038	0.047
30, 40, 50, 60, 70	2, 2, 2, 2, 2	1	0.043	0.066	0.040	0.053	0.050	0.039	0.046
	5, 5, 5, 5, 5	1	0.043	0.066	0.040	0.053	0.050	0.039	0.046
20, 30, 50, 70, 80	2, 2, 2, 2, 2	1	0.042	0.046	0.024	0.043	0.042	0.037	0.036
	5, 5, 5, 5, 5	1	0.042	0.046	0.024	0.043	0.042	0.037	0.036

Tablo 5’te verilen sonuçlar incelendiğinde anakütle sayısındaki artış genel olarak testlerin 1.tip hata olasılıklarını düşürmüştür. HY testlerinin 1.tip hata olasılıkları artmış ve $k = 3$ anakütle sayısında gözlenen durumdakinden daha liberal bir yapı sergilemiştir. GPD, PB ve FY testinin 1.tip hata olasılıkları ise azalış göstererek sifıra yaklaşmıştır. Tablo 4’te verilen sonuçlara benzer şekilde GPD, PB, FY ve HY3 testi tutucu (*conservative*) bir yapı sergilerken, HY1 ve HY2 testinin liberal bir yapı sergilediği gözlenmiştir. Küçük örneklem hacimlerinde HY4 testi ve en dengesiz tasarımda PB testinin 1.tip hata olasılıkları nominal seviyeye yakındır. Orta örneklem hacimlerinde de HY1, HY2, HY4 ve PB testinin 1.tip hata olasılıklarının nominal seviyeye yakın olduğu gözlenmiştir. Büyük örneklem hacimlerinde ise PB, HY1, HY2, HY4 testinin 1.tip hata olasılıkları nominal seviyeye yakındır. Küçük örneklem hacimlerinde tasarımın dengesizleşmesi GPD, PB ve FY testinin performansını olumsuz etkilemezken, HY1, HY2 ve HY4 testinin performansını olumsuz etkilemiştir. Orta örneklem hacimlerinde tasarımın dengesizleşmesi yalnızca HY4

testinin performansını olumsuz etkilemektedir. Büyük örneklerde ise benzer şekilde yalnızca HY4 testinin performansını olumsuz etkilemiştir. Ölçek parametresi a_i değerindeki artış ise testlerin 1.tip hata olasılığını kontrol etme becerisi üzerinde herhangi bir etkisi olmadığı gözlenmektedir.

Tüm sonuçlar incelendiğinde örneklem hacimleri ve tasarım tipinin farklı anakütle sayıları için testlerin 1.tip hata olasılıklarını kontrol etme becerisi üzerinde farklı etkileri olduğu gözlenmiştir. $k = 3$ anakütle sayısı için HY1 ve HY2 testi 1.tip hata olasılığını diğer testlerden daha iyi kontrol ederken, anakütle sayısının artması bu testlerin liberal yapıda olmalarına neden olmaktadır. Artan anakütle sayısına bağlı olarak HY4 testinin 1.tip hata olasılığının nominal seviyeye en yakın olduğu gözlenmiştir. Tasarım dengesizliğinin çoğunlukla GPD, PB ve FY testlerinin performansını olumsuz etkilemezken bazı durumlarda HY testlerinin 1.tip hata olasılığını kontrol edememesine neden olmaktadır. Bu durumda HY testlerinin tasarım dengesizliğine karşı duyarlı olduğu ortaya çıkmıştır. Ölçek parametresinin değerindeki artış ise testlerin performansında çok önemli bir etki yaratmadığı gözlenmiştir. Bu bölümde elde edilen sonuçlara göre hiçbir test tüm durumlarda 1.tip hata olasılığını ideal bir şekilde kontrol edememektedir.

4.3. Cezalandırılmış Güç Değerleri

Bir önceki bölümde bahsedilen problemlerden dolayı bu bölümde testlerin gücünü kıyaslamak için en uygun alternatif olan cezalandırılmış güç kriteri kullanılmıştır. Böylelikle 1.tip hata olasılığına göre düzeltilmiş güç değeri üzerinden geçerli karşılaştırmalar yapılabilmektedir. 1.tip hata olasılığı hesaplamalarında kullanılan konfigürasyon parametrelerine ek olarak bu bölümde etki büyüklüğünü kontrol edebilmek için a_3 parametresi kullanılmıştır. Parametrenin düşük ve yüksek düzeyleri için sırasıyla 3, 4, 5 ve 6, 8, 10 değerleri alınarak testin gücü hesaplanmıştır. Tüm senaryolarda konum parametreleri $b_i = 1, 1, 1$ olarak alınmıştır. Tablo 6'da düşük ($a_i = 2$), Tablo 7'de ise yüksek ölçek parametresi değerleri ($a_i = 5$) ve $k = 3$ için göz önüne alınan durumlarda testlerin cezalandırılmış güç değerleri verilmiştir.

Tablo 6. Önerilen testlerin $k = 3$ grup için cezalandırılmış güç değerleri

n_i	a_i	b_i	GPD	PB	FY	HY1	HY2	HY3	HY4
10, 10, 10	2, 2, 3	1	0.0101	0.0143	0.0021	0.0847	0.0752	0.0226	0.0691
	2, 2, 4	1	0.0254	0.0334	0.0055	0.0991	0.0919	0.0234	0.0808
	2, 2, 5	1	0.0429	0.0581	0.0097	0.1145	0.1012	0.0317	0.0898
8, 10, 12	2, 2, 3	1	0.0245	0.0342	0.0039	0.0839	0.0773	0.0256	0.0635
	2, 2, 4	1	0.0622	0.0779	0.0096	0.1077	0.1095	0.0380	0.0939
	2, 2, 5	1	0.1181	0.1386	0.0177	0.1135	0.1160	0.0334	0.0967
5, 10, 15	2, 2, 3	1	0.0559	0.1119	0.0147	0.0579	0.0733	0.0230	0.0407
	2, 2, 4	1	0.1590	0.2180	0.0265	0.0693	0.0875	0.0213	0.0371
	2, 2, 5	1	0.3010	0.3475	0.0413	0.0658	0.0808	0.0179	0.0281
30, 30, 30	2, 2, 3	1	0.1610	0.2420	0.1253	0.3060	0.2901	0.2272	0.3021
	2, 2, 4	1	0.4997	0.6363	0.4303	0.6721	0.6567	0.5812	0.6731
	2, 2, 5	1	0.7297	0.8488	0.6697	0.9004	0.8759	0.8103	0.8984
24, 30, 36	2, 2, 3	1	0.2444	0.3225	0.1755	0.3040	0.3175	0.2734	0.3166
	2, 2, 4	1	0.6301	0.7455	0.5232	0.7469	0.7418	0.6895	0.7457
	2, 2, 5	1	0.8083	0.9180	0.7411	0.9276	0.9153	0.8773	0.9228
15, 30, 45	2, 2, 3	1	0.3307	0.4043	0.2160	0.3528	0.3737	0.3477	0.3381
	2, 2, 4	1	0.7178	0.8277	0.5868	0.8019	0.8045	0.7852	0.7736
	2, 2, 5	1	0.8447	0.9643	0.7630	0.9296	0.9179	0.9232	0.9160
50, 50, 50	2, 2, 3	1	0.3808	0.4686	0.3422	0.4672	0.4617	0.4604	0.4701
	2, 2, 4	1	0.8119	0.9107	0.7675	0.8953	0.8795	0.9189	0.8963
	2, 2, 5	1	0.8812	0.9700	0.8447	0.9477	0.9310	0.9842	0.9477
40, 50, 60	2, 2, 3	1	0.4722	0.5521	0.4015	0.5091	0.5283	0.5276	0.5332
	2, 2, 4	1	0.8448	0.9522	0.8029	0.9096	0.9228	0.9335	0.9315
	2, 2, 5	1	0.8821	0.9921	0.8549	0.9506	0.9594	0.9776	0.9684
25, 50, 75	2, 2, 3	1	0.5582	0.6071	0.4440	0.5556	0.5927	0.5562	0.5654
	2, 2, 4	1	0.8840	0.9662	0.8103	0.9052	0.9257	0.8951	0.9144
	2, 2, 5	1	0.9063	0.9937	0.8489	0.9440	0.9623	0.9266	0.9525

3 anakütle ortalamasının eşitliğinin testi için yapılan simülasyon çalışmasının sonuçları incelendiğinde aşağıdaki bulgular elde edilmiştir.

- Küçük örneklem hacimlerinden dengeli tasarımda HY testleri diğer testlerden daha güçlüdür. Tasarımın en dengesiz ($n_i = 5, 10, 15$) olduğu düzende HY testlerinin güçlerinde düşüş gözlenmiştir. Aksine tasarımın dengesizliğinin artması durumunda GPD, PB ve HY testlerinin gücü artmaktadır. Dengeli tasarımda

tüm etki büyüklüklerinde HY1 testi en güçlü, dengeli olmayan ($n_i = 8, 10, 12$) tasarımıda düşük ve orta düzey etki büyüklüklerinde en güçlü testler HY1 ve HY2, yüksek etki büyüklüğünde ise PB testi en güçlü testtir. En dengesiz tasarımıda PB testinin en güçlü test olduğu görülmektedir. Yüksek etki büyüklükleri göz önüne alındığında dengeli tasarımıda HY1 ve HY2 testlerinin gücü 0.10 düzeyinin üzerindedir. Dengeli olmayan tasarımıda GPD, PB, HY1 ve HY2 testlerinin gücü 0.10 düzeyinin üzerindedir. En dengesiz tasarımıda ise GPD ve PB testlerinin gücünün 0.30 düzeyinin üzerinde olduğu görülmüştür.

- Orta örneklem hacimlerinden dengeli tasarımıda HY testlerinin gücü diğerlerinden daha yüksektir. Tasarımın dengesizliğinin artması küçük örneklemelerde gözlenen durumun aksine testlerin güçleri üzerinde olumsuz bir etki oluşturmamıştır. Ancak tasarımın dengesizleşmesi GPD, PB ve FY testlerinin performansı diğer testlere göre daha yüksek oranda artmaktadır. Dengeli tasarımıda HY1 ve HY4 testleri en güçlü testlerdir. Dengeli olmayan tasarımıda PB, HY1, HY2 ve HY4 testleri en güçlüdür. En dengesiz tasarımıda düşük etki büyüklüğünde PB ve HY2 testleri, orta düzey etki büyüklüğünde PB, HY1 ve HY2 testleri, yüksek etki büyüklüğünde ise PB testi en güçlüdür. Yüksek etki büyüklükleri göz önüne alındığında, dengeli tasarımıda PB, HY2 ve HY4 testlerinin gücü 0.85 düzeyinin, HY1 testinin ise 0.90 düzeyinin üzerindedir. Dengeli olmayan tasarımıda HY3 testinin gücü 0.85, PB, HY1, HY2 ve HY4 testlerinin gücü 0.90 düzeyinin üzerindedir. En dengesiz tasarımıda ise GPD testinin gücü 0.85, PB ve HY testlerinin gücü 0.90 düzeyinin üzerinde olduğu görülmüştür.
- Büyük örneklem hacimlerinde dengeli tasarımıda GPD ve FY testleri dışındaki tüm testlerin güçleri birbirine yakındır. Tasarımın dengesizleşmesi GPD, PB ve FY testlerinin güçlerini pozitif yönde, HY2, HY2 ve HY4 testlerinin güçleri üzerinde önemli bir etkiye neden olmazken, HY3 testinin gücünü negatif yönde etkilemektedir. Dengeli tasarımıda düşük etki büyüklüğünde PB ve HY testleri, orta düzey ve yüksek etki büyüklüklerinde PB ve HY3 testleri en güçlüdür. Dengeli olmayan tasarımıda düşük ve orta düzey etki büyüklüklerinde PB, HY2, HY3 ve HY4 testleri, yüksek etki büyüklüğünde PB, HY3 ve HY4 testleri

en güçlüdür. En dengesiz tasarımda ise düşük etki büyüklüğünde PB ve HY2 testleri, orta düzey ve yüksek etki büyüklüklerinde PB testi en güçlü testtir. Dengeli tasarımın yüksek etki büyüklüğünde GPD ve FY testlerinin gücü 0.84, diğer testlerinin gücü 0.93 düzeyinin üzerindedir. Dengeli olmayan tasarımda GPD ve FY testlerinin gücü 0.85, diğer testlerinin gücü de 0.95 düzeyinin üzerindedir. En dengesiz tasarımda ise FY testinin gücü 0.84 ve diğer testlerin 0.90 düzeyinin üzerindedir.

Bir sonraki tabloda ise aynı senaryolar bu defa problemde yer alan *nuisance* parametre olarak tanımlanan a_i ölçek parametresinin daha yüksek değer aldığı durumda elde edilen sonuçlar verilmiştir. Artan ölçek parametresine bağlı olarak etki büyüklüğünü kontrol etmek için kullanılan a_3 değeri 6, 8, 10 olarak alınmıştır.

Tablo 7. Önerilen testlerin $k = 3$ grup için cezalandırılmış güç değerleri

n_i	a_i	b_i	GPD	PB	FY	HY1	HY2	HY3	HY4
10, 10, 10	5, 5, 6	1	0.0215	0.0234	0.0164	0.0599	0.0586	0.0118	0.0524
	5, 5, 8	1	0.0298	0.0321	0.0266	0.0880	0.0850	0.0299	0.0810
	5, 5, 10	1	0.0419	0.0436	0.0369	0.0965	0.0935	0.0322	0.0858
8, 10, 12	5, 5, 6	1	0.0356	0.0367	0.0349	0.0620	0.0644	0.0210	0.0525
	5, 5, 8	1	0.0651	0.0642	0.0622	0.0858	0.0893	0.0303	0.0672
	5, 5, 10	1	0.1046	0.1051	0.1023	0.1077	0.1095	0.0380	0.0939
5, 10, 15	5, 5, 6	1	0.0221	0.0631	0.0094	0.0623	0.0770	0.0247	0.0416
	5, 5, 8	1	0.0726	0.1325	0.0171	0.0614	0.0910	0.0170	0.0398
	5, 5, 10	1	0.1577	0.2186	0.0266	0.0693	0.1080	0.0187	0.0444
30, 30, 30	5, 5, 6	1	0.1509	0.1628	0.1329	0.0991	0.0954	0.0779	0.0991
	5, 5, 8	1	0.4060	0.4592	0.3732	0.3769	0.3619	0.3042	0.3759
	5, 5, 10	1	0.5755	0.6553	0.5404	0.6721	0.6567	0.5812	0.6731
24, 30, 36	5, 5, 6	1	0.0758	0.0868	0.0742	0.1166	0.1173	0.1040	0.1212
	5, 5, 8	1	0.2967	0.3371	0.2831	0.4361	0.4443	0.4041	0.4436
	5, 5, 10	1	0.5172	0.5825	0.4921	0.7615	0.7656	0.7088	0.7736
15, 30, 45	5, 5, 6	1	0.0886	0.1306	0.0588	0.1154	0.1236	0.1115	0.1058
	5, 5, 8	1	0.4225	0.4979	0.2934	0.4605	0.4720	0.4365	0.4327
	5, 5, 10	1	0.7141	0.8159	0.5891	0.8019	0.8045	0.7852	0.7736
50, 50, 50	5, 5, 6	1	0.2984	0.3276	0.2862	0.1211	0.1199	0.1059	0.1211
	5, 5, 8	1	0.6130	0.6734	0.5938	0.5988	0.5872	0.5921	0.5997
	5, 5, 10	1	0.6634	0.7277	0.6399	0.8953	0.8795	0.9189	0.8963
40, 50, 60	5, 5, 6	1	0.1163	0.1323	0.1123	0.1411	0.1511	0.1432	0.1505
	5, 5, 8	1	0.4950	0.5549	0.4724	0.6407	0.6591	0.6599	0.6712
	5, 5, 10	1	0.6630	0.7370	0.6277	0.9096	0.9228	0.9335	0.9315
25, 50, 75	5, 5, 6	1	0.1498	0.1752	0.1097	0.1713	0.1814	0.1611	0.1602
	5, 5, 8	1	0.6815	0.7300	0.5755	0.7099	0.7266	0.6785	0.6674
	5, 5, 10	1	0.8951	0.9605	0.8291	0.9486	0.9433	0.8874	0.8828

Ölçek parametresinin yüksek değeri için elde edilen güç değerleri incelendiğinde dengeli tasarımlarda HY1 ve HY2 testleri GPD, PB ve FY testlerine göre daha iyi performans göstermektedir. Ancak ölçek parametresinin değerinin artması testlerin güçlerinin düşmesine neden olmuştur. Örneklem hacimlerinin büyüklüklerine göre sonuçlar incelendiğinde aşağıdaki sonuçlar elde edilmiştir:

- Küçük örneklemeler incelendiğinde HY testleri bu durumdan GPD, PB ve FY

testlerine göre daha az etkinlenmiştir. GPD, PB ve FY testleri dengeli ve dengeli olmayan tasarımda çok düşük düzeyde etkilenmiş, en dengesiz tasarımda ise yaklaşık %50 oranında güç kaybetmiştir. HY testleri ise bunun aksine dengeli tasarımda yaklaşık %10 oranında bir güç kaybı, dengeli olmayan tasarımlarda ise önemli oranda bir kayıp gözlenmemiştir.

- Orta örneklem hacimlerinde de benzer şekilde GPD, PB ve FY testlerindeki güç kaybı HY testlerine göre daha fazladır. HY testlerinde yaklaşık %10 oranında, GPD, PB ve FY testlerinde %15 oranında kayıplar gözlenmiştir. Tasarım tipleri karşılaştırıldığında ise kayıplar dengeli tasarımlarda daha yüksek orandadır. Kayıplar sonrasında dengeli ve dengeli olmayan tasarımda en güçlü testler PB, HY2 ve HY4, en dengesiz tasarımda ise PB testidir.
- Büyük örneklem hacimlerinde kayıplar yalnızca dengeli ve dengeli olmayan tasarımlarda gözlenmiş, en dengesiz tasarımda önemli bir kayıp gözlenmemiştir. Kayıplar küçük ve orta örneklem hacimlerine benzer olarak GPD, PB ve FY testlerinde daha yüksektir. HY testlerinde kayıplar %5 düzeyinde iken GPD, PB ve FY testlerinde bu düzey %20'nin üzerindedir. Kayıplar sonrasında en güçlü testler dengeli tasarımda PB ve HY3 testleri, dengeli olmayan tasarımda HY testleri, en dengesiz tasarımda ise PB ve HY2 testleridir.

Tablo 8'de düşük, Tablo 9' da ise yüksek ölçek parametresi değerleri ve $k = 4$ için göz önüne alınan durumlarda testlerin cezalandırılmış güç değerleri verilmiştir.

Tablo 8. Önerilen testlerin $k = 4$ grup için cezalandırılmış güç değerleri

n_i	a_i	b_i	GPD	PB	FY	HY1	HY2	HY3	HY4
10,10,10,10	2,2,2,3	1	0.0096	0.0135	0.0028	0.0864	0.0847	0.0293	0.0780
	2,2,2,4	1	0.0170	0.0315	0.0036	0.0990	0.0942	0.0372	0.0928
	2,2,2,5	1	0.0229	0.0457	0.0057	0.0982	0.0933	0.0356	0.0854
6,8,12,14	2,2,2,3	1	0.0243	0.0375	0.0043	0.0757	0.0859	0.0250	0.0726
	2,2,2,4	1	0.0678	0.0804	0.0072	0.0765	0.0835	0.0306	0.0775
	2,2,2,5	1	0.1445	0.1249	0.0079	0.0765	0.0788	0.0290	0.0618
4,8,12,16	2,2,2,3	1	0.0474	0.0713	0.0058	0.0678	0.0757	0.0263	0.0418
	2,2,2,4	1	0.1468	0.1220	0.0094	0.0671	0.0772	0.0214	0.0409
	2,2,2,5	1	0.2861	0.1978	0.0152	0.0580	0.0645	0.0214	0.0334
30,30,30,30	2,2,2,3	1	0.1296	0.2041	0.0919	0.1961	0.2013	0.1534	0.2023
	2,2,2,4	1	0.4103	0.5641	0.3435	0.5505	0.5679	0.4546	0.5629
	2,2,2,5	1	0.7006	0.8315	0.6133	0.8046	0.8259	0.7175	0.8204
18,24,36,42	2,2,2,3	1	0.2992	0.3359	0.1655	0.3110	0.3275	0.2626	0.2910
	2,2,2,4	1	0.7183	0.7578	0.5341	0.7800	0.7953	0.6917	0.7427
	2,2,2,5	1	0.8200	0.8845	0.7325	0.9620	0.9512	0.8673	0.9118
12,24,36,48	2,2,2,3	1	0.3414	0.3565	0.1578	0.3349	0.3465	0.3069	0.2990
	2,2,2,4	1	0.7551	0.8169	0.5513	0.7727	0.7783	0.7703	0.7693
	2,2,2,5	1	0.8385	0.9723	0.7846	0.9372	0.9141	0.9565	0.9614
50,50,50,50	2,2,2,3	1	0.3063	0.3813	0.2713	0.3670	0.3719	0.3693	0.3679
	2,2,2,4	1	0.7276	0.8080	0.6941	0.8271	0.8350	0.8832	0.8289
	2,2,2,5	1	0.8350	0.8891	0.8154	0.9083	0.9160	0.9842	0.9074
30,40,60,70	2,2,2,3	1	0.5616	0.5568	0.4221	0.5153	0.5376	0.5488	0.5367
	2,2,2,4	1	0.9058	0.9325	0.8639	0.9062	0.9179	0.9509	0.9251
	2,2,2,5	1	0.9206	0.9525	0.8971	0.9285	0.9366	0.9713	0.9449
20,40,60,80	2,2,2,3	1	0.6425	0.6050	0.4171	0.5834	0.5982	0.5422	0.5376
	2,2,2,4	1	0.9317	0.9720	0.8048	0.9600	0.9629	0.8994	0.9071
	2,2,2,5	1	0.9449	1	0.8383	0.9806	0.9806	0.9206	0.9285

4 anakütle ortalamasının eşitliğinin testi için yapılan simülasyon çalışmasının sonuçları incelendiğinde aşağıdaki bulgular elde edilmiştir.

- Küçük örneklem hacimlerinden dengeli tasarımda HY1 testi diğer testlerden daha güçlüdür. Tasarımın dengesizleşmesi HY testlerin gücünü $k = 3$ anakütle sayısındaki sonuçlara benzer olarak negatif yönde, GPD, PB ve FY testlerinin gücünü pozitif yönde etkilemektedir. Bu etki pozitif yönde daha yüksek şekilde

gözlenmektedir. Ayrıca eşitliği test edilen anakütle sayısının artması testlerin gücü üzerinde negatif yönde bir etki yaratmıştır. Dengeli tasarımlarda HY testlerinde göreceli olarak daha az, dengeli olmayan tasarımlarda ise GPD, PB ve FY testlerinde daha az gözlenmektedir. Dengeli tasarımda en güçlü test HY1 testidir. Dengeli olmayan tasarımda düşük ve orta etki büyüklüklerinde HY2, yüksek etki büyüklüğünde ise PB testi en güçlüdür. En dengesiz tasarımda düşük etki büyüklüğünde HY2 testi, orta düzey etki büyüklüğünde GPD testi, yüksek etki büyüklüğünde GPD testi en güçlüdür. Yüksek etki büyüklükleri göz önüne alındığında dengeli tasarımda HY1 ve HY2 testlerinin gücü 0.10 düzeyindedir. Dengeli olmayan tasarımda GPD ve PB testlerinin gücü 0.10 düzeyinin üzerindedir. En dengesiz tasarımda ise GPD ve PB testlerinin gücü 0.20 düzeyinin üzerindedir.

- Orta örneklem hacimlerinden dengeli tasarımda PB ve HY testlerinin gücü diğerlerinden daha yüksektir. Tasarım dengesizleştikçe testlerin güçleri artmakta, en yüksek artış HY3 testinde gözlenmiştir. Anakütle sayısındaki artışın etkisi dengeli tasarımda negatif, dengeli olmayan tasarımlarda ise pozitif yöndedir. Görüldüğü gibi anakütle sayısındaki artış orta örneklem hacimlerinde farklılaşmaya başlamıştır. Dengeli tasarımda PB, HY1, HY2 ve HY4 testleri en güçlü testlerdir. Dengeli olmayan tasarımda düşük etki büyüklüğünde PB, HY1 ve HY2 testleri, orta düzey ve yüksek etki büyüklüklerinde ise HY1 ve HY2 testleri en güçlüdür. En dengesiz tasarımda düşük etki büyüklüğünde GPD, PB, HY1 ve HY2 testleri, orta düzey etki büyüklüğünde PB testi, yüksek etki büyüklüğünde ise PB ve HY4 testleri en güçlüdür. Yüksek etki büyüklükleri göz önüne alındığında, dengeli tasarımda PB, HY1, HY2 ve HY4 testlerinin gücü 0.80 düzeyinin üzerindedir. Dengeli olmayan tasarımda GPD, PB ve HY3 testlerinin gücü 0.80, HY1, HY2 ve HY4 testlerinin gücü 0.90 düzeyinin üzerindedir. En dengesiz tasarımda ise GPD testinin gücü 0.80, PB ve HY testlerinin gücü 0.90 düzeyinin üzerinde olduğu görülmüştür.
- Büyük örneklem hacimlerinde GPD ve FY testi dışında tüm testlerin güçleri birbirine yakındır. Tasarım dengesizliğinin artması HY3 testi dışında tüm testlerin güçlerinin artmasına neden olmaktadır. HY3 testinin gücündeki düşüş ise

dramatik düzeyde değildir. $k = 3$ durumundaki sonuçlar ile karşılaştırıldığında ise dengeli tasarımlarda bir düşüş, dengeli olmayan tasarımlarda ise herhangi bir değişiklik gözlenmemiştir. Dengeli tasarımda PB ve HY testleri en güçlü testlerdir. Dengeli olmayan tasarımda PB ve HY3 testleri en güçlüdür. En dengesiz tasarımda ise düşük etki büyüklüğünde GPD, orta düzey etki büyüklüğünde PB, HY1 ve HY2 testleri, yüksek etki büyüklüğünde ise PB testi en güçlü testtir. Dengeli tasarımın yüksek etki büyüklüğünde GPD, PB ve HY testlerinin gücü 0.80, HY testlerinin gücü 0.90 düzeyinin üzerindedir. Dengeli olmayan tasarımda GPD, FY, HY1 ve HY2 testlerinin gücü 0.90, PB, HY3 ve HY4 testlerinin gücü de 0.95 düzeyinin üzerindedir. En dengesiz tasarımda ise HY3 ve HY4 testlerinin gücü 0.90, GPD, PB, HY1 ve HY2 testlerinin gücü 0.95 düzeyinin üzerindedir.

Tablo 9. Önerilen testlerin $k = 4$ grup için cezalandırılmış güç değerleri

n_i	a_i	b_i	GPD	PB	FY	HY1	HY2	HY3	HY4
10,10,10,10	5,5,5,6	1	0.0066	0.0090	0.0021	0.0758	0.0709	0.0241	0.0703
	5,5,5,8	1	0.0125	0.0172	0.0028	0.0977	0.0893	0.0342	0.0901
	5,5,5,10	1	0.0170	0.0315	0.0036	0.1077	0.1003	0.0365	0.0990
6,8,12,14	5,5,5,6	1	0.0118	0.0187	0.0029	0.0684	0.0748	0.0236	0.0566
	5,5,5,8	1	0.0317	0.0492	0.0043	0.0780	0.0929	0.0244	0.0687
	5,5,5,10	1	0.0678	0.0804	0.0072	0.0780	0.0863	0.0299	0.0733
4,8,12,16	5,5,5,6	1	0.0163	0.0463	0.0044	0.0663	0.0716	0.0279	0.0383
	5,5,5,8	1	0.0667	0.0820	0.0065	0.0640	0.0753	0.0270	0.0429
	5,5,5,10	1	0.1468	0.1220	0.0094	0.0679	0.0768	0.0270	0.0402
30,30,30,30	5,5,5,6	1	0.0397	0.0737	0.0281	0.0702	0.0718	0.0543	0.0728
	5,5,5,8	1	0.1745	0.2684	0.1299	0.2800	0.2873	0.2090	0.2816
	5,5,5,10	1	0.4103	0.5641	0.3435	0.5601	0.5679	0.4584	0.5679
18,24,36,42	5,5,5,6	1	0.0683	0.0960	0.0351	0.1089	0.1127	0.0886	0.1002
	5,5,5,8	1	0.3983	0.4463	0.2280	0.3951	0.4215	0.3394	0.3846
	5,5,5,10	1	0.7183	0.7578	0.5341	0.7723	0.7877	0.6701	0.7427
12,24,36,48	5,5,5,6	1	0.0783	0.1109	0.0453	0.1049	0.1200	0.0945	0.0942
	5,5,5,8	1	0.4444	0.4624	0.2325	0.4233	0.4384	0.3846	0.3973
	5,5,5,10	1	0.7551	0.8169	0.5513	0.7521	0.7852	0.7351	0.7547
50,50,50,50	5,5,5,6	1	0.0621	0.0873	0.0526	0.1178	0.1167	0.0966	0.1158
	5,5,5,8	1	0.4137	0.5060	0.3748	0.5347	0.5276	0.4707	0.5287
	5,5,5,10	1	0.7276	0.8080	0.6971	0.8971	0.8894	0.8282	0.8991
30,40,60,70	5,5,5,6	1	0.1464	0.1621	0.1069	0.1451	0.1545	0.1346	0.1424
	5,5,5,8	1	0.6867	0.6817	0.5649	0.6776	0.7040	0.6406	0.6678
	5,5,5,10	1	0.9058	0.9325	0.8639	0.9570	0.9703	0.9090	0.9420
20,40,60,80	5,5,5,6	1	0.1446	0.1570	0.0831	0.1602	0.1575	0.1545	0.1471
	5,5,5,8	1	0.7644	0.7310	0.5379	0.7046	0.6790	0.7278	0.7148
	5,5,5,10	1	0.9317	0.9720	0.8048	0.9334	0.8891	0.9674	0.9580

Tablo 9’da verilen yüksek ölçek parametresi durumunda elde edilen güç değerleri incelendiğinde düşük ölçek parametresi senaryosuna benzer olarak dengeli tasarımlarda HY1 ve HY2 testleri GPD, PB ve FY testlerine göre daha iyi performans göstermektedir. Ancak ölçek parametresinin değerinin artması testlerin güçlerinin düşmesine neden olmuştur. Örneklem hacimlerinin büyüklüklerine göre sonuçlar incelendiğinde:

- Küçük örneklemeler incelendiğinde HY testleri bu durumdan GPD, PB ve FY testlerine göre daha az etkinlenmiştir. GPD, PB ve FY testleri dengeli ve dengeli olmayan tasarımda çok düşük düzeyde etkilenmiş, en dengesiz tasarımda ise yaklaşık %50 oranında güç kaybetmiştir. HY testleri ise bunun aksine dengeli tasarımda yaklaşık %10 oranında bir güç kaybı, dengeli olmayan tasarımlarda ise önemli oranda bir kayıp gözlenmemiştir.
- Orta örneklem hacimlerinde GPD, PB ve FY testlerindeki güç kaybı HY testlerine göre daha fazladır. HY testlerinde yaklaşık %20 oranında, GPD, PB ve FY testlerinde ise dengeli tasarımda %30, dengeli olmayan tasarımlarda %10 oranında kayıplar gözlenmiştir. Burada dikkat çekilmesi gereken önemli nokta kayıpların $k = 3$ anakütle sayısına göre daha fazla olduğudur. Özellikle HY testlerin genel bir kayıp artışı, GPD, PB ve FY testlerinde ise sadece dengeli tasarımda önemli bir kayıp artış oranı gözlenmiştir. Tasarım tipleri karşılaştırıldığında ise kayıplar dengeli tasarımlarda daha yüksek orandadır. Kayıplar sonrasında dengeli ve dengeli olmayan tasarımda en güçlü testler PB, HY1, HY2 ve HY4, en dengesiz tasarımda ise PB testidir. Yine bir önceki anakütle sayısında elde edilen sonuçlar ile karşılaştırdığında PB testinin performansında artış gözlenmektedir.
- Büyük örneklem hacimlerinde kayıplar GPD, PB ve FY testlerinde diğer testlere göre daha fazladır. HY testlerinde kayıplar %5 düzeyinde iken GPD, PB ve FY testlerinde bu düzey dengeli tasarımda %10 dengeli olmayan tasarımlarda %5 civarındadır. Kayıplar sonrasında en güçlü testler dengeli tasarımda HY1, HY2 ve HY4, dengeli olmayan tasarımlarda PB, HY3 ve HY4 testleridir.

Tablo 10’da düşük, Tablo 11’de ise yüksek ölçek parametresi değerleri ve $k = 5$ için göz önüne alınan durumlarda testlerin cezalandırılmış güç değerleri verilmiştir.

Tablo 10. Önerilen testlerin $k = 5$ grup için cezalandırılmış güç değerleri

n_i	a_i	b_i	GPD	PB	FY	HY1	HY2	HY3	HY4
10,10,10,10,10	2,2,2,2,3	1	0.0065	0.0170	0.0014	0.0815	0.0170	0.0744	0.0576
	2,2,2,2,4	1	0.0123	0.0280	0.0029	0.0969	0.0216	0.0849	0.0685
	2,2,2,2,5	1	0.0210	0.0383	0.0036	0.1041	0.0208	0.0791	0.0736
6,8,10,12,14	2,2,2,2,3	1	0.0204	0.0322	0.0064	0.0753	0.0213	0.0624	0.0533
	2,2,2,2,4	1	0.0430	0.0605	0.0078	0.0844	0.0190	0.0713	0.0597
	2,2,2,2,5	1	0.0955	0.0951	0.0092	0.0753	0.0134	0.0525	0.0533
4,6,10,14,16	2,2,2,2,3	1	0.0458	0.0756	0.0119	0.0779	0.0283	0.0438	0.0551
	2,2,2,2,4	1	0.1288	0.1266	0.0157	0.0765	0.0250	0.0438	0.0541
	2,2,2,2,5	1	0.2616	0.1587	0.0171	0.0628	0.0200	0.0338	0.0444
30,30,30,30,30	2,2,2,2,3	1	0.0905	0.1564	0.0618	0.1991	0.1363	0.2049	0.1408
	2,2,2,2,4	1	0.3139	0.4552	0.2329	0.5206	0.3751	0.5226	0.3681
	2,2,2,2,5	1	0.5741	0.7239	0.4741	0.7557	0.5960	0.7698	0.5343
18,24,30,36,42	2,2,2,2,3	1	0.3062	0.3061	0.1511	0.2785	0.2504	0.2726	0.1970
	2,2,2,2,4	1	0.7292	0.7323	0.4736	0.7112	0.6699	0.7325	0.5029
	2,2,2,2,5	1	0.8738	0.9153	0.7201	0.8913	0.8825	0.9316	0.6303
12,18,30,42,48	2,2,2,2,3	1	0.3761	0.3186	0.1560	0.3228	0.2193	0.2272	0.2282
	2,2,2,2,4	1	0.8079	0.7848	0.4793	0.7713	0.6040	0.6203	0.5454
	2,2,2,2,5	1	0.9074	0.9480	0.7422	0.9545	0.8121	0.8146	0.6749
50,50,50,50,50	2,2,2,2,3	1	0.2688	0.3578	0.2239	0.3885	0.3269	0.3875	0.2747
	2,2,2,2,4	1	0.7091	0.8308	0.6578	0.8635	0.7705	0.8654	0.6106
	2,2,2,2,5	1	0.8384	0.9300	0.8038	0.9625	0.8881	0.9635	0.6806
30,40,50,60,70	2,2,2,2,3	1	0.5460	0.4630	0.3871	0.5510	0.4817	0.5138	0.3896
	2,2,2,2,4	1	0.9160	0.8408	0.8544	0.9730	0.8800	0.9353	0.6880
	2,2,2,2,5	1	0.9356	0.8686	0.9092	0.9980	0.9035	0.9603	0.7057
20,30,50,70,80	2,2,2,2,3	1	0.6406	0.5485	0.3739	0.5283	0.4802	0.4702	0.3726
	2,2,2,2,4	1	0.9210	0.9372	0.7706	0.9081	0.8695	0.8636	0.6421
	2,2,2,2,5	1	0.9285	0.9613	0.8103	0.9266	0.8891	0.8821	0.6552

5 anakütle ortalamasının eşitliğinin testi için yapılan simülasyon çalışmasının sonuçları incelendiğinde aşağıdaki bulgular elde edilmiştir.

- Küçük örneklem hacimlerinden dengeli tasarımda HY1, HY3 ve HY4 testleri diğer testlerden daha güçlüdür. Tasarımın dengesizleşmesi testlerin gücünü negatif yönde, GPD ve PB testlerinin gücünü pozitif yönde etkilemektedir. Anakütle sayısındaki artış göz önüne alındığında tüm testlerin performansı-

nın olumsuz etkilendiği gözlenmiştir. GPD testinin performansında önemli bir değişiklik olmazken; HY3 testinin performansında artış olduğu görülmüştür. Dengeli tasarımda HY1 testi en güçlü testtir. Dengeli olmayan tasarımda düşük ve orta düzey etki büyüklüğünde HY1, yüksek etki büyüklüğünde GPD ve PB testleri en güçlüdür. En dengesiz tasarımda düşük etki büyüklüğünde PB ve HY1 testleri, orta düzey etki büyüklüğünde GPD ve PB, testleri yüksek etki büyüklüğünde GPD testi en güçlüdür. Yüksek etki büyüklükleri göz önüne alındığında dengeli tasarımda HY1 testinin gücü 0.10 düzeyinin üzerindedir. Dengeli olmayan tasarımda GPD ve PB testlerinin gücü 0.10 düzeyine yakındır. En dengesiz tasarımda ise GPD testinin gücü 0.25 düzeyinin üzerindedir.

- Orta örneklem hacimlerinden dengeli tasarımda HY3 testinin gücü diğerlerinden daha yüksektir. Tasarım dengesizleştikçe testlerin güçleri artmaktadır. Artış PB ve HY1 testlerinde diğerlerine göre daha yüksektir. Anakütle sayısındaki artışın etkisi ise yalnızca GPD testinin orta düzey ve yüksek etki büyüklüklerinde artış şeklinde gözlenmiştir. Ancak diğer testlerin ise güçlerinde azalışa neden olmuştur. Tasarım tiplerine göre sonuçlar irdelendiğinde dengeli tasarımda en güçlü testlerin HY1 ve HY3 olduğu görülmüştür. Dengeli olmayan tasarımda düşük etki büyüklüğünde GPD ve PB testleri, orta düzey etki büyüklüğünde GPD, PB, HY1 ve HY3 testleri, yüksek etki büyüklüğünde ise HY3 testi en güçlü testtir. Yüksek etki büyüklükleri göz önüne alındığında dengeli tasarımda PB, HY1 ve HY3 testlerinin gücü 0.70 düzeyinin üzerindedir. Dengeli olmayan tasarımda GPD, HY1 ve HY2 testlerinin gücü 0.80, PB ve HY3 testlerinin gücü 0.90 düzeyinin üzerindedir. En dengesiz tasarımda ise HY2 ve HY3 testlerinin gücü 0.80, GPD, PB ve HY1 testlerinin gücü 0.90 düzeyinin üzerinde olduğu görülmüştür.
- Büyük örneklem hacimlerinde PB, HY1 ve HY3 testleri dengeli tasarımda diğer testlerden daha güçlüdür. Tasarımın dengesizleşmesi HY1, HY3 ve HY4 testlerinin gücünün düşmesine neden olmaktadır. Bu düşüş HY3 testinde daha fazladır. Anakütle sayısının artması GPD ve HY1 testleri dışında tüm testlerin performansını olumsuz etkilemiştir. GPD testinin gücü üzerinde önemli bir değişikliğe neden olmamışken, HY1 testinin düşük ve orta düzey etki büyük-

lüklerinde performansında artışa neden olmuştur. Dengeli tasarımda HY1 ve HY3 testleri en güçlü testlerdir. Dengeli olmayan tasarımda düşük etki büyüklüğünde GPD, HY1, orta düzey ve yüksek etki büyüklüklerinde HY1 testi en güçlü testtir. En dengesiz tasarımda ise düşük etki büyüklüğünde GPD, orta düzey ve yüksek etki büyüklüklerinde GPD ve PB testleri, yüksek etki büyüklüğünde PB testi en güçlü testtir. Dengeli tasarımın yüksek etki büyüklüğünde GPD, FY ve HY2 testlerinin gücü 0.80, PB, HY1 ve HY3 testlerinin gücü 0.90 düzeyinin üzerindedir. Dengeli olmayan tasarımda PB testinin gücü 0.80, GPD, FY, HY1, HY2 ve HY3 testlerinin gücü 0.90 düzeyinin üzerindedir. En dengesiz tasarımda ise FY, HY2 ve HY3 testlerinin gücü 0.80, GPD, PB ve HY1 testlerinin gücü 0.90 düzeyinin üzerindedir.

Tablo 11. Önerilen testlerin $k = 5$ grup için cezalandırılmış güç değerleri

n_i	a_i	b_i	GPD	PB	FY	HY1	HY2	HY3	HY4
10,10,10,10,10	5,5,5,5,6	1	0.0036	0.0088	0.0014	0.0625	0.0147	0.0553	0.0442
	5,5,5,5,8	1	0.0080	0.0177	0.0021	0.0869	0.0193	0.0839	0.0615
	5,5,5,5,10	1	0.0123	0.0280	0.0029	0.0969	0.0216	0.0849	0.0685
6,8,10,12,14	5,5,5,5,6	1	0.0066	0.0212	0.0036	0.0662	0.0182	0.0574	0.0468
	5,5,5,5,8	1	0.0226	0.0401	0.0064	0.0778	0.0213	0.0634	0.0550
	5,5,5,5,10	1	0.0430	0.0605	0.0078	0.0844	0.0190	0.0713	0.0597
4,6,10,14,16	5,5,5,5,6	1	0.0248	0.0539	0.0089	0.0736	0.0300	0.0402	0.0521
	5,5,5,5,8	1	0.0636	0.0841	0.0127	0.0808	0.0300	0.0466	0.0572
	5,5,5,5,10	1	0.1288	0.1266	0.0157	0.0765	0.0250	0.0438	0.0541
30,30,30,30,30	5,5,5,5,6	1	0.0280	0.0530	0.0204	0.0728	0.0561	0.0765	0.0515
	5,5,5,5,8	1	0.1201	0.2024	0.0087	0.2574	0.1782	0.2657	0.1820
	5,5,5,5,10	1	0.3139	0.4552	0.2329	0.5206	0.3751	0.5226	0.3681
18,24,30,36,42	5,5,5,5,6	1	0.0727	0.1106	0.0425	0.0975	0.0841	0.0951	0.0689
	5,5,5,5,8	1	0.4050	0.4014	0.2107	0.3630	0.3184	0.3716	0.2567
	5,5,5,5,10	1	0.7292	0.7323	0.4736	0.7112	0.6699	0.7325	0.5029
12,18,30,42,48	5,5,5,5,6	1	0.1050	0.1166	0.0449	0.1059	0.0679	0.0708	0.0749
	5,5,5,5,8	1	0.4957	0.4089	0.2009	0.4198	0.2916	0.3032	0.2969
	5,5,5,5,10	1	0.8079	0.7848	0.4793	0.7713	0.6040	0.6203	0.5454
50,50,50,50,50	5,5,5,5,6	1	0.0625	0.0927	0.0527	0.1020	0.0826	0.1010	0.0721
	5,5,5,5,8	1	0.3651	0.4730	0.3188	0.5012	0.4328	0.5090	0.3544
	5,5,5,5,10	1	0.7091	0.8308	0.6578	0.8635	0.7705	0.8654	0.6106
30,40,50,60,70	5,5,5,5,6	1	0.1264	0.1271	0.0895	0.1290	0.1059	0.1174	0.0912
	5,5,5,5,8	1	0.6715	0.5762	0.5121	0.6860	0.6066	0.6563	0.4851
	5,5,5,5,10	1	0.9160	0.8408	0.8544	0.9730	0.8800	0.9353	0.6880
20,30,50,70,80	5,5,5,5,6	1	0.1569	0.1453	0.0819	0.1337	0.1167	0.1131	0.0945
	5,5,5,5,8	1	0.7614	0.6909	0.4826	0.6629	0.6076	0.6028	0.4688
	5,5,5,5,10	1	0.9210	0.9372	0.7706	0.9081	0.8695	0.8636	0.6421

Tablo 11’de verilen yüksek ölçek parametresi durumunda elde edilen güç değerleri incelendiğinde düşük ölçek parametresi senaryosuna benzer olarak dengeli tasarımlarda HY1 ve HY2 testleri GPD, PB ve FY testlerine göre daha iyi performans göstermektedir. Ancak bu fark örneklem hacmi arttıkça düşmektedir. Ayrıca ölçek parametresi değerinin artması testlerin güçlerinin düşmesine neden olmuştur. Örneklem hacimlerinin büyüklüklerine göre sonuçlar incelendiğinde:

- Küçük örneklemeler incelendiğinde HY testleri bu durumdan GPD, PB ve FY testlerine göre daha az etkinlenmiştir. GPD, PB ve FY testleri yaklaşık %50 oranında güç kaybetmiştir. HY testleri ise bunun aksine dengeli tasarımda yaklaşık %10 oranında bir kayıp gözlenmiştir. Kayıplar sonrasında dengeli ve dengeli olmayan tasarımlarda en güçlü testler HY1, HY2 ve HY4, en dengesiz tasarımda ise GPD ve PB testidir.
- Orta örneklem hacimlerinde GPD, PB ve FY testlerindeki güç kaybı HY testlerine göre daha fazladır. HY testlerinde yaklaşık %20 oranında, GPD, PB ve FY testlerinde ise dengeli tasarımda %20, dengeli olmayan tasarımlarda %10 oranında kayıplar gözlenmiştir. Tasarım tipleri karşılaştırıldığında ise kayıplar dengeli tasarımlarda daha yüksek orandadır. Kayıplar sonrasında dengeli tasarımda en güçlü testler HY1 ve HY3, dengeli olmayan tasarımda en güçlü testler GPD, PB, HY1 ve HY3, en dengesiz tasarımda ise GPD, PB ve HY1 testidir. Önceki anakütle sayısında elde edilen sonuçlar ile karşılaştırıldığında burada GPD testinin performansında artış dikkat çekmektedir.
- Büyük örneklem hacimlerinde kayıplar GPD, PB ve FY testlerinde diğer testlere göre dengeli tasarımda daha fazladır. HY testlerinde kayıplar %10 düzeyinde iken GPD, PB ve FY testlerinde bu düzey %15 civarındadır. Kayıplar sonrasında en güçlü testler dengeli tasarımda PB, HY1 ve HY3, dengeli olmayan tasarımlarda GPD, HY1 ve HY3, en dengesiz tasarımda ise GPD, PB ve HY1 testleridir. Görüldüğü gibi anakütle sayısı $k = 5$ olduğunda GPD testinin performansı diğer testlere göre öne çıkmaktadır.

Bu bölümde elde edilen cezalandırılmış güç sonuçları göz önüne alındığında $k = 3$ anakütle sayısı için dengeli ve dengeli olmayan tasarımlarda HY1 testi, en dengesiz tasarımda PB testi en güçlüdür. Örneklem hacmi büyüdükçe dengeli ve dengeli olmayan tasarımda HY4, en dengesiz tasarımda ise HY2 testinin performansı da öne çıkmaktadır. Ölçek parametresinin değeri yükseldiğinde dengeli ve dengeli olmayan tasarımda PB testi gücünü kaybetmektedir. $k = 4$ anakütle sayısı için dengeli tasarımda HY1, HY2 ve HY4 testi, dengeli olmayan tasarımlarda HY1, HY2 ve HY4 testinin yanı sıra GPD ve PB testi en güçlü testlerdir. Örneklem hacmi büyüdükçe HY3 testinin performansında da belirgin artışlar gözlenmekte ve en güçlü testler

grubuna dahil olmaktadır. Ölçek parametresi değeri yükseldiğinde GPD testi gücünü kaybetmekte, PB, HY3 ve HY4 testinin performansı ön plana çıkmaktadır. $k = 5$ anakütle sayısı için dengeli tasarımda HY1 ve HY3 testi, dengeli olmayan tasarımlarda ise bu testlerin yanı sıra GPD ve PB testi en güçlüdür. Örneklem hacmi arttığında ve tasarım dengesizleştğinde HY1 ve HY3 testlerinde performans düşmekte, GPD ve PB testleri performanslarını korumaktadır. Ölçek parametresi değerinin artışı ise dengeli olmayan tasarımlarda GPD ve PB testinin yanı sıra HY3 testinin performansı da ön plana çıkmaktadır. Sonuç olarak HY testleri genel olarak dengeli tasarımlarda, PB testi ise dengeli olmayan tasarımlarda iyi performans sergilemektedir. k anakütle sayısı arttığında GPD testinin dengeli olmayan tasarımlarda performansı ön plana çıkmaktadır. PB ve HY1 testi ölçek parametresi yani *nuisance* parametre değerinin artışından etkilenmemektedir.

5. GERÇEK VERİ UYGULAMALARI

Bu bölümde ikiden fazla iki parametrelili üstel dağılmış anakütlenin ortalamalarının eşitliğinin test edilmesi probleminin çözümü için geliştirilen testler üç farklı gerçek hayat problemi üzerinde uygulanmıştır. Bunlardan ilki ürün güvenilirliği üzerine, ikinci bölümde üç farklı yaşam modeli ve son olarak da ekonomik gösterge problemleri üzerinedir. Uygulamalara geçmeden önce oluşturulan testlerin açık kaynaklı erişimini mümkün kılmak için geliştirilen R paketi *doex* tanıtılmış; içeriğinde yer alan testler ve kullanımları hakkında bilgi verilmiştir.

5.1. *doex* R - Paketi

Son yıllarda oldukça önem kazanan açık kaynaklı yazılımlar tekrar edilebilir bilimin en önemli unsurlarındandır. *R* ve *Python* gibi açık kaynak kodlu yazılımlar geliştiren karmaşık bilimsel yöntemlerin kullanılabilirliğini sağlaması açısından çok faydalıdır. Açık kaynak kodlu yazılımların, geliştirilen paketler sayesinde her geçen gün bilim dünyasındaki yaygın etkisini arttırmaktadır. Hesaplama teknolojisindeki gelişmeler yeniden örnekleme yöntemlerine dayalı istatistiksel tekniklerin geliştirilmesi ve kullanılması sürecini hızlandırmıştır. Ancak bu yöntemlerin karmaşık yapısı ve yoğun istatistiksel bilgi gerektirmesi farklı disiplinlerden araştırmacıların bu yöntemleri kullanabilmesinin önünde önemli bir engel teşkil etmektedir. Bu nedenle tez çalışması boyunca üzerinde çalışılan yöntemler ve literatürdeki alternatifleri *R* platformunda *doex* (*Design Of EXperiment*) adıyla bir paket içerisinde implemente edilmiş ve araştırmacıların kullanımına sunulmuştur (Cavus ve Yazici, 2019b).

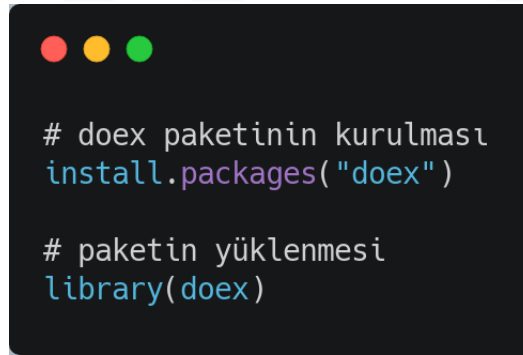
doex temel olarak KF testinin homojen varyanslılık ve/veya normal dağılım varyanslarının ihlal edilmesi durumları için literatürde yer alan ve tez çalışması boyunca geliştirilen testlerden oluşmaktadır. İçerdiği yöntemler aşağıdaki gibi gruplandırılabilir:

1. Homojen varyanslı olmayan normal dağılmış anakütle ortalamalarının eşitliğinin test edilmesi için geliştirilen Approximate F (Asiribo ve Gurland, 1990), Alexander-Govern (Alexander ve Govern, 1994), Alvandi et al. Genelleştirilmiş F (Alvandi vd., 2012), B-square (Ozdemir ve Kurt, 2006), Brown-Forsythe (Brown

ve Forsythe, 1974), Box F (Box, 1954), Cochran F (Cochran, 1937), Fiducial Approach (Li vd., 2011), Genelleştirilmiş F (Weerahandi, 1995), Johansen F (Johansen, 1980), Modified Brown-Forsythe (Mehrotra, 1997), Modified Welch (Hartung vd., 2002), One Stage (Chen ve Chen, 1998), One Stage Range (Chen ve Chen, 2000), Parametric Bootstrap (Krishnamoorthy vd., 2007), Permutation F (Berry ve Mielke, 2002), Scott-Smith (Scott ve Smith, 1971), Welch-Aspin (Aspin, 1948), Welch F (Welch, 1951) testleri,

2. Aykırı değerden kaynaklı normallik varsayımının bozulması durumunda ikiden fazla homojen varyanslı olmayan anakütle ortalamasının karşılaştırılması için geliştirilen Modified Generalized F testi (Cavus vd., 2017),
3. Tez çalışması sürecinde önerilen üstel dağılmış anakütle ortalamalarının eşitliğinin test edilmesi probleminin çözümü için kullanılan GPD, PB, FY ve HY testleri yer almaktadır.

doex paketinin kurulumu ve çağırılması Şekil 6'de gösterilmiştir.



```
# doex paketinin kurulması
install.packages("doex")

# paketin yüklenmesi
library(doex)
```

Şekil 6. *doex* paketinin R ortamında kurulması ve yüklenmesi

doex paketinde yer alan ikiden fazla iki-parametrelili üstel dağılım ölçek parametrelerinin eşit olmaması durumunda anakütle ortalamalarının eşitliğinin test edilmesi için kullanılan GPD, PB ve FY testlerinin R ortamında kullanımı Şekil 7'de gösterildiği gibidir.

```
# GPD testinin uygulanması
gpv_exp(component$lifetime, component$supplier, rept = 10000)

# PB testinin uygulanması
pb_exp(component$lifetime, component$supplier)

# FY testinin uygulanması
fa_exp(component$lifetime, component$supplier)
```

Şekil 7. *GPD, PB ve FY testlerinin R ortamında uygulanması*

Yukarıda kullanılan *R* fonksiyonları pivot veri üzerinde kullanılırlar. Fonksiyonların veri, grup bilgisi ve tekrar sayısı olmak üzere üç girdisi vardır. Veri gözlem değerinden, grup bilgisi her bir gözlem değerinin ait olduğu grup bilgisinden ve tekrar sayısı ise GPD, PB ve FY testlerinde kritik bölgelerin oluşturulması için gereken tekrar sayısıdır. Tekrar sayısı fonksiyonlarda default olarak 10.000 alınmıştır. Değiştirilmek istenmediği durumda girdi olarak gösterilmesine gerek yoktur.

doex paketi sayesinde geliştirilen yöntemler araştırmacılar tarafından herhangi bir kısıtlama olmaksızın ücretsiz bir şekilde CRAN üzerinden indirilerek kullanılabilir. Böylelikle geliştirilen testlerin gerçek hayat problemlerine çözüm olarak kullanılabilmesi sağlanmıştır. Paketin kullanılabilirliğini sonraki bölümlerde ele alınan örneklerdeki uygulamalar ile gösterilmiştir.

5.2. Ürün Güvenilirliği Örneği

Veri seti yerel bir fabrikanın ürettiği elektronik cihazlarda yer alan bir komponente ilişkin bozulma sürelerinden oluşmaktadır (Cavus ve Yazici, 2020c). 2016 yılı Mayıs ayı içerisinde fabrikanın 4 farklı tedarikçisinden temin ettiği komponentlerde tespit edilen arızalar sonucunda tamiri mümkün olmayan parçalar üzerinden yıllık bazda hesaplanmıştır. Burada test edilmesi gereken problem farklı tedarikçilerden temin edilen aynı işi gören komponentin ortalama bozulma süreleri arasında istatistiksel olarak anlamlı bir farkın olup olmadığıdır. Bilindiği gibi elektronik komponentlerin bozulma süresi genellikle sağa çarpık bir dağılım gösterir; en iyi modeli

sağlayan parametrik dağılım ailesi ise üstel dağılımdır. Bu problemin çözümünde normal dağılım varsayımına dayalı olarak oluşturulan herhangi bir testin kullanımı hipotez hakkında yanlış kararlar verilmesine neden olabilir. Bu yüzden tez çalışmasında geliştirilen ve literatürde yer alan iki parametrelili üstel dağılıma dayalı testler problemin çözümü için kullanılacaktır. Dört farklı tedarikçiden temin edilen A, B, C ve D elektronik komponentlerin bozulma süreleri yıllık bazda Tablo 12’de verilmiştir. Tablonun alt kısmında her bir ürün grubuna ilişkin örneklem hacmi ve parametre tahminleri verilmiştir.

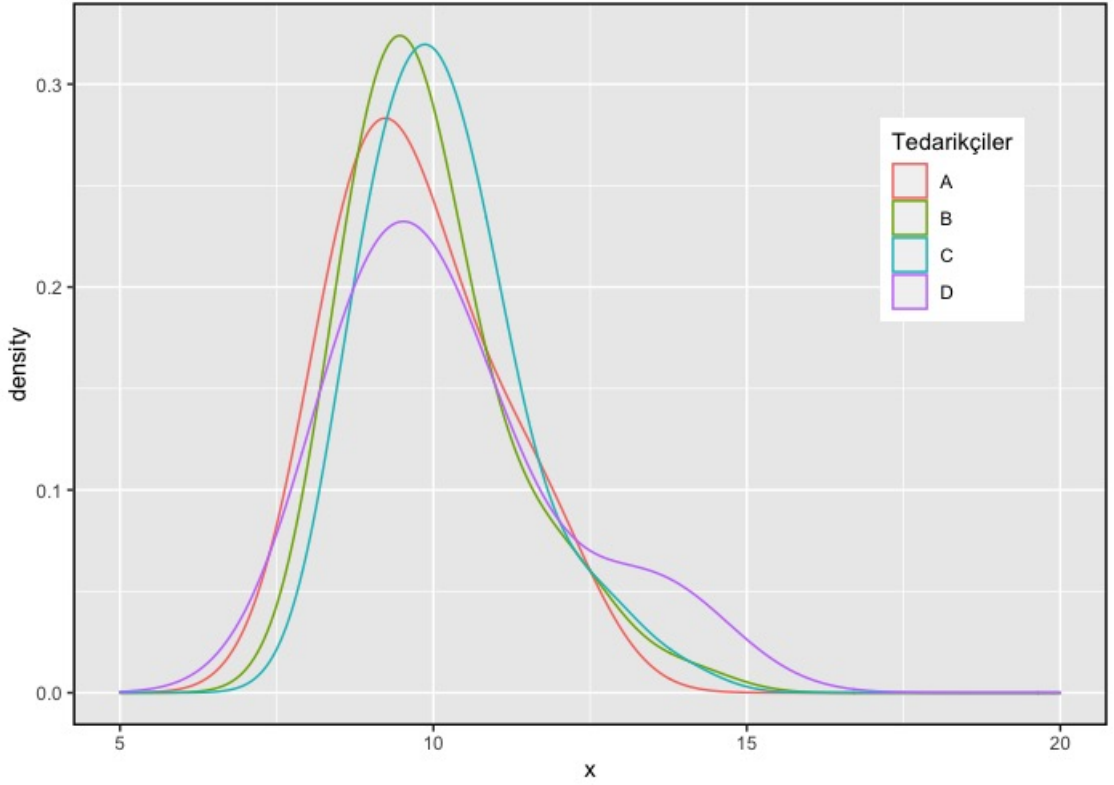


Tablo 12. *Ürün güvenilirliği örneği verileri*

Tedarikçi	A	B	C		D	
	8.38	9.98	9.67	10.90	8.76	8.62
	9.82	9.54	9.54	9.57	11.06	10.01
	11.60	9.46	12.85	9.62	8.57	13.09
	11.21	11.51	12.41	10.63	10.37	9.96
	8.76	14.08	10.92	10.66	8.89	8.80
	9.86	11.94	8.47	10.29	12.85	10.83
	8.85	10.60	9.66	10.29	10.41	9.99
	10.67	9.61	9.01	10.10	8.95	9.02
	12.09	8.81	10.10	9.12	10.24	8.98
	8.51	10.84	9.95	9.98	8.93	9.04
	9.09	9.72	8.73	10.81	11.47	13.94
	9.94	8.40	9.09	12.10	9.58	11.01
	9.80	8.62	9.54	10.84	11.13	
	9.11	9.79	8.51	8.99	9.70	
	8.62	9.09	11.17	12.73	9.74	
		10.73	9.74	10.44	10.60	
		10.45	9.69	9.97	12.68	
		8.88	8.93	8.89	9.14	
		8.86	10.68	8.87	9.80	
		12.01	12.21	11.15	9.03	
		9.72	10.07	9.78	8.41	
		10.03	8.76	9.81	9.62	
		9.96		9.32	9.51	
		8.81		8.43	11.68	
		9.43		10.11	13.93	
		8.67		11.55	9.59	
		9.16		10.03	10.73	
n_i	15	49	54		12	
$\hat{a}_i$	1.47	1.60	1.82		1.80	
$\hat{b}_i$	8.38	8.40	8.41		8.62	
$\hat{\mu}_i$	9.85	10	10.23		10.42	

R’da goftest paketinde yer alan Cramer-von Mises (CvM) uyum iyiliği testi kullanılarak ürün güvenilirliği verisinin iki-parametreliliği üstel dağılım gösterip göster-

mediği test edilmiştir (Faraway vd., 2019). CvM testinin sonucunda p-değeri 0.6786 olarak elde edilmiş ve 0.05 anlamlılık düzeyinde verinin iki-parametreliliği üstel dağılım gösterdiği görülmüştür. Tablo 12 incelendiğinde A tedarikçisinden temin edilen komponentin ortalama bozulma süresi 9.85, B komponentinin 10, C komponentinin 10.23 ve D komponentinin 10.42 yıldır. Veri setinin tasarım tipi incelendiğinde komponent gruplarının ortalama örneklem hacminin 32 ve dengeli olmadığı görülmektedir. Komponent gruplarının dağılımları ise Şekil 8’te verilmiştir.



Şekil 8. Ürün güvenilirliği örneği verilerinin dağılımı

Dağılım grafiğinde görüldüğü gibi komponent gruplarının ortalaması birbirine yakın ve dağılımları sağa çarpıktır. Farklı tedarikçilerden temin edilen komponentlerin ortalama bozulma süreleri eşitliğinin testinde $H_0 : \mu_A = \mu_B = \mu_C = \mu_D$ hipotezinin reddedilebilmesi için gerekli kanıt olup olmadığını ortaya çıkarmak için GPD, PB, FY ve HY testleri uygulanmıştır ve elde edilen p-değerleri Tablo 13’de verilmiştir.

Tablo 13. *Ürün güvenilirliği örneği üzerinde H_0 hipotezini test etmek için uygulanan testlerin p-değerleri*

Test	p-değeri
GPD	0.6770
PB	0.7466
FY	0.7493
HY1	0.7510
HY2	0.7171
HY3	0.6824
HY4	0.7128

Tablo 13’de verilen testlerin p-değerlerine göre $\alpha = 0.05$ anlamlılık düzeyinde H_0 hipotezi reddedilemez. Böylelikle farklı tedarikçilerden temin edilen komponentlerin ortalama bozulma süreleri arasında istatistiksel olarak anlamlı bir fark bulunmadığı sonucuna varılır.

5.3. Yaşam Modeli Örnekleri

Bu bölümde tez çalışmasında üzerinde çalışılan problemin en sık karşılaşıldığı alanlardan biri olan yaşam modellemesi örnekleri ele alınmıştır. Küçük örneklem hacimleri ile sık karşılaşılan yaşam modellemesi örneklerinde güçlü testler ile çalışmak çok önemlidir.

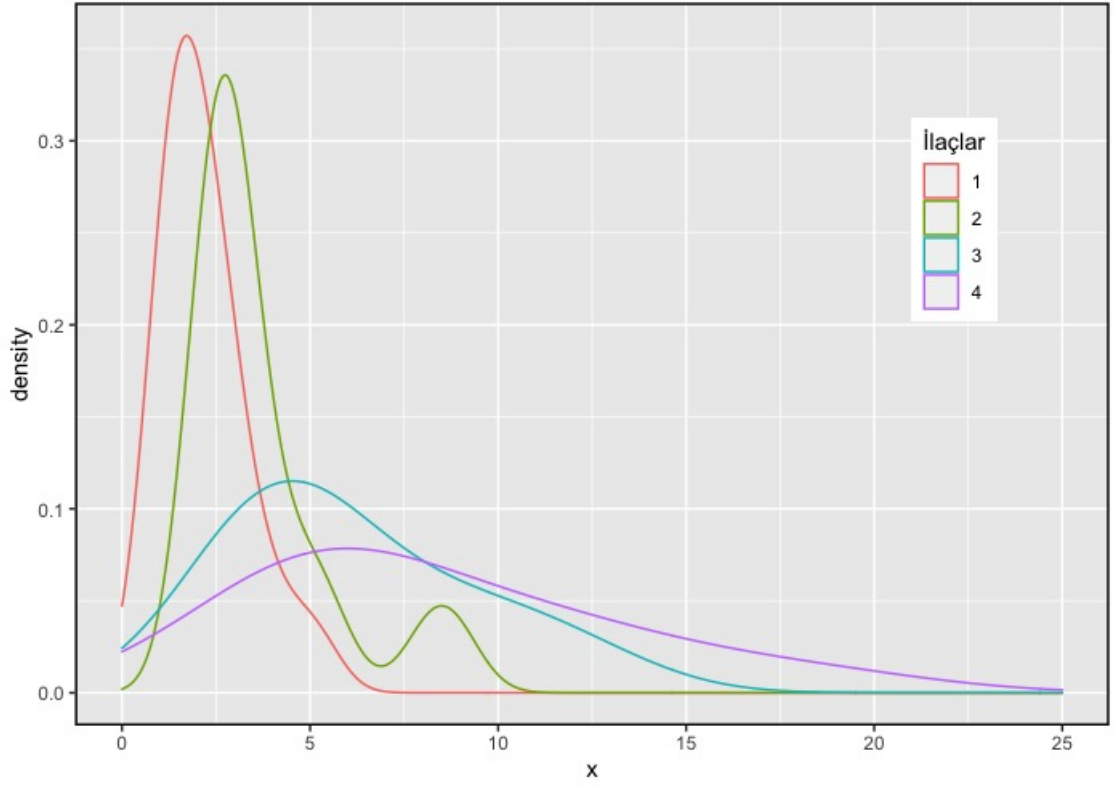
5.3.1. Yaşam modeli örneği 1

Klinik denemelerde ilaçların remisyon sürelerinin karşılaştırılması sık karşılaşılan bir problemdir. Bir önceki örnek uygulamada da bahsedildiği gibi yaşam süresi verileri genellikle normal dağılım göstermezler ve sıklıkla üstel dağılım kullanılarak modellenirler. Bu örnekte Wu ve Wu (2005) çalışmalarında kullandıkları 20 hasta üzerinde lösemi tedavisinde kullanılan 4 farklı ilaca karşı gözlemlenen remisyon sürelerinden oluşan veri seti ele alınmıştır. Tablo 14’te veri seti ve ilgili parametre tahminleri verilmiştir.

Tablo 14. *Yaşam modeli örnek 1 verileri*

İlaç	1	2	3	4
	1.034	2.214	4.158	5.115
	2.344	4.976	4.025	4.498
	1.266	8.154	5.170	4.617
	1.563	2.686	11.909	4.651
	1.169	2.271	4.912	4.533
	4.118	3.139	4.629	4.513
	1.013	2.214	3.955	7.641
	1.509	4.480	6.735	5.971
	1.109	8.847	3.140	12.130
	1.965	2.239	12.446	4.600
	5.136	3.473	8.777	4.914
	1.533	2.761	6.321	17.169
	1.716	2.833	3.256	5.497
	2.778	2.381	8.250	11.332
	2.546	3.548	3.759	18.922
	2.626	2.414	5.205	13.712
	3.413	2.832	3.071	6.309
	1.929	5.551	3.147	10.086
	2.061	3.376	9.773	9.293
	2.951	2.968	10.218	11.787
n_i	20	20	20	20
$\hat{a}_i$	1.238	1.530	3.233	4.070
$\hat{b}_i$	1.013	2.214	3.071	4.498
$\hat{\mu}_i$	2.251	3.744	6.304	8.568

CvM uyum iyiliği testi kullanılarak verinin iki-parametreliliği üstel dağılım gösterip göstermediği test edilmiştir. Testinin sonucunda p-değeri 0.4468 olarak elde edilmiş ve 0.05 anlamlılık düzeyinde verinin iki-parametreliliği üstel dağılım gösterdiği görülmüştür. Tablo 14 incelendiğinde 1. ilacın ortalama remisyon süresi 2.25 gün, 2. ilacın 3.75, 3. ilacın 6.35 ve 4. ilacın 8.56 gündür. Veri seti dengeli bir tasarıma sahiptir. İlaçların remisyon sürelerinin dağılımları ise Şekil 9’te verilmiştir.



Şekil 9. Yaşam modeli örnek 1 verilerinin dağılımı

Dağılım grafiğinde görüldüğü gibi farklı ilaçların tedavisinde hastaların remisyon sürelerinin dağılımları sağa çarpıktır. Burada problem hastalık tedavisi için uygulanan farklı ilaçların remisyon (etki süreleri) arasında anlamlı bir istatistiksel farkın olup olmadığıdır. Çözümü için bahsedilen sebeplerden dolayı iki parametrelili üstel dağılım kullanılarak geliştirilen testler kullanılmıştır ve elde edilen p-değerleri Tablo 15'te verilmiştir.

Tablo 15. Yaşam modeli 1 örneği üzerinde H_0 hipotezini test etmek için uygulanan testlerin p-değerleri

Test	p-değeri
GPD	0.0001*
PB	0.0002*
FY	0.0001*
HY1	0.0001*
HY2	0.0001*
HY3	< 0.0001*
HY4	< 0.0001*

Sonuçlardan görüldüğü gibi tüm testlere göre $\alpha = 0.05$ anlamlılık düzeyinde H_0 hipotezi reddedilir. Böylelikle ilaçların ortalama remisyon süreleri arasında istatistiksel olarak anlamlı bir fark bulunduğu sonucuna varılır.

5.3.2. Yaşam modeli örneği 2

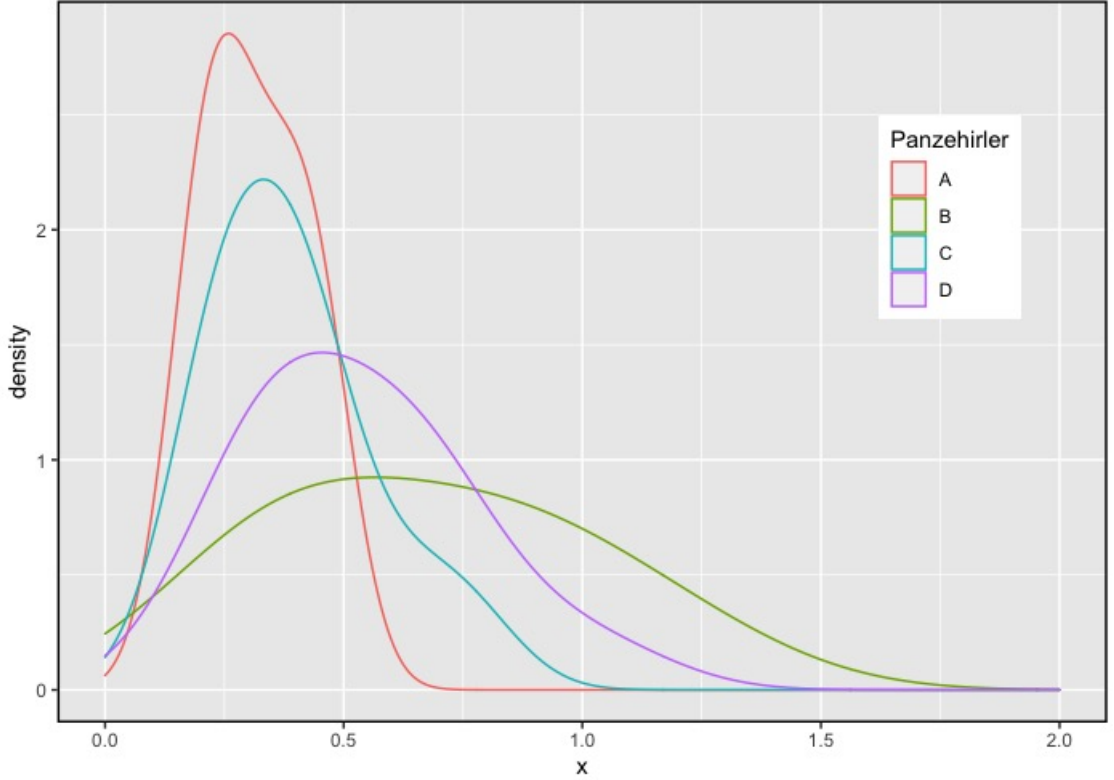
Cavus ve Yazici (2020a) çalışmasında örnek olaydaki problemi GPD, PB ve FY testlerini kullanarak çözmeyi denemişlerdir. Bu örnek olayda bir zehire karşı geliştirilen panzehirlerin etkisi fareler üzerinde denenmiştir. Box ve Cox (1964) çalışmasından alınan veri setinde A, B, C ve D panzehirleri tedavisine karşı farelerin tedaviye cevap verme sürelerinden oluşmaktadır. İlgili veri seti ve gruplara göre parametre tahminleri Tablo 16’te verilmiştir.

Tablo 16. *Yaşam modeli örnek 2 verileri*

Tedaviler	A	B	C	D
	0.31	0.82	0.43	0.45
	0.45	1.10	0.45	0.71
	0.46	0.88	0.63	0.66
	0.43	0.72	0.76	0.62
	0.36	0.92	0.44	0.56
	0.29	0.61	0.35	1.02
	0.40	0.49	0.31	0.71
	0.23	1.24	0.40	0.38
	0.22	0.30	0.23	0.30
	0.21	0.37	0.25	0.36
	0.18	0.38	0.24	0.31
	0.23	0.29	0.22	0.33
n_i	12	12	12	12
$\hat{a}_i$	0.15	0.42	0.19	0.26
$\hat{b}_i$	0.18	0.29	0.22	0.30
$\hat{\mu}_i$	0.33	0.71	0.41	0.56

CvM uyum iyiliği testi kullanılarak verinin iki-parametrelili üstel dağılım gösterip göstermediği test edilmiştir. Testinin sonucunda p-değeri 0.6814 olarak elde edilmiş ve 0.05 anlamlılık düzeyinde verinin iki-parametrelili üstel dağılım gösterdiği

görülmüştür. Tablo 16’te verilen istatistiklere bakıldığında panzehir tedavisine farelerin cevap verme süresi A panzehirinde ortalama 0.33 hafta, B panzehirinde 0.71, C panzehirinde 0.41 ve D panzehirinde 0.56 haftadır. Yanı sıra veri seti küçük hacimli dengeli tasarım örneğidir. Tedaviye cevap verme sürelerinin dağılımları Şekil 10’te verilmiştir.



Şekil 10. Yaşam modeli örnek 2 verilerinin dağılımı

Şekil 10’te görüldüğü üzere panzehirlere göre tedavi süreleri sağa çarpık bir dağılım göstermektedir. Bu örnekte amaç panzehirlerin tedavi süresi üzerinde etkilerinin eşit olup olmadığının testidir. Böylelikle kullanılacak panzehir ya da panzehirler belirlenerek en etkin tedavi sağlanabilecektir. Çözümü için geliştirilen testler kullanılmış ve elde edilen p-değerleri Tablo 17’da verilmiştir.

Tablo 17. *Yaşam modeli 2 örneği üzerinde H_0 hipotezini test etmek için uygulanan testlerin p-değerleri*

Test	p-değeri
GPD	0.0475*
PB	0.0205*
FY	0.0630
HY1	0.0032*
HY2	0.0038*
HY3	0.0077*
HY4	0.0036*

Yukarıdaki tabloda verilen testlerden yalnızca FY testi dışında tüm testlerden elde edilen p-değeri sonuçlarına göre 0.05 anlamlılık düzeyinde H_0 hipotezi reddedilir. Böylelikle panzehirlerin tedavi sürelerinin eşit olmadığı sonucuna varılır.

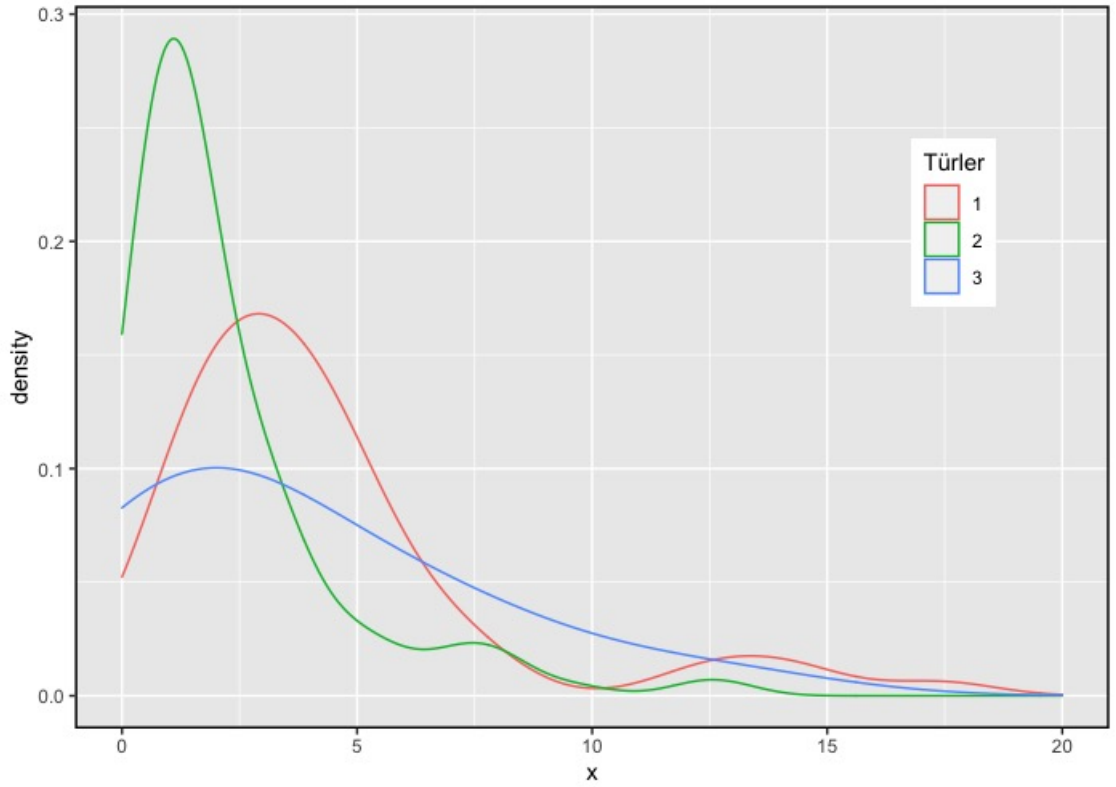
5.3.3. Yaşam modeli örneği 3

Bu örnekte malignant melanoma tanısı konan hastaların taşıdıkları türlere göre tümör boyutları arasında farklılık olup olmadığını test edilmiştir. Veri seti malignant melanoma tanısı konan hastalarda yapılan ölçümlerden oluşmaktadır. 1962-1997 yılları arasında University Hospital of Odense Plastik Cerrahi Kliniği'nde ameliyat ile tümörleri alınan her bir hastanın tümör kalınlıkları (cm) ve türleri kaydedilmiştir. Andersen vd. (1993) çalışmasından alınan veriler ve tümör tiplerine ilişkin parametre tahminleri Tablo 18'de verilmiştir.

Tablo 18. *Yaşam modeli örnek 3 verileri*

Tür	1					2					3
	12.08	3.56	0.81	1.34	0.58	0.81	5.64	3.22	0.81	0.81	6.76
	4.84	3.54	5.16	1.29	2.58	7.73	9.66	1.45	0.32	2.10	0.65
	5.16	0.97	5.80	8.38	0.81	0.97	0.10	0.48	3.22	3.87	2.90
	12.88	4.83	1.62	1.94	3.54	12.88	5.48	0.16	2.74	0.65	3.22
	7.41	1.62	0.97	0.16	0.97	2.58	4.83	0.16	4.84	1.94	0.16
	4.19	6.44	1.94	0.16	1.78	4.09	0.97	1.29	1.62	0.65	12.56
	3.87	14.66	2.26	1.29	1.94	0.64	0.97	1.94	0.65	2.10	8.54
	4.84	2.58	3.22	0.32	1.29	0.97	5.16	0.81	0.65	1.94	1.29
	2.42	3.87	3.54	1.13	3.22	3.22	2.90	0.65	1.29	1.13	1.29
	5.80	3.54	1.94	1.62	1.53	1.62	3.87	7.09	1.62	7.06	4.84
	7.06	1.34	3.06	1.37	1.29	3.87	1.94	0.16	3.54	6.12	0.65
	5.48	2.24	0.81	0.24	1.62	0.32	0.16	1.62	3.22	0.48	1.62
	7.73	3.87	1.76	0.81	1.62	0.32	0.64	1.29	0.65	2.26	6.12
	13.85	3.54	2.26	1.29	0.32	3.22	1.45	0.48	1.03	2.90	1.45
	2.34	17.42	0.81	1.29	1.29	2.58	4.82	0.64	7.09		
	4.19	3.22	2.26	0.97	0.97	1.13	1.29	3.22	1.29		
	4.04	4.51	3.54	1.13	3.06	0.97	7.89	1.94	0.65		
	4.84	2.58		1.29	3.54	1.94	0.81	2.58	1.78		
	0.32	1.62		0.48	1.62	0.65	1.29	2.58	12.24		
	2.58	2.10		2.26	2.58	0.97	0.64	0.81	8.06		
n_i	57					134					14
$\hat{a}_i$	4.06					2.16					3.83
$\hat{b}_i$	0.32					0.10					0.16
$\hat{\mu}_i$	4.38					2.26					3.99

CvM uyum iyiliği testi kullanılarak verinin iki-parametreliliği üstel dağılım gösterip göstermediği test edilmiştir. Testinin sonucunda p-değeri 0.4005 olarak elde edilmiş ve 0.05 anlamlılık düzeyinde verinin iki-parametreliliği üstel dağılım gösterdiği görülmüştür. Tablo 18 incelendiğinde 1.ilacın tedavisi uygulanan hastaların ortalama tümör boyutları 4.38 cm, 2. ilacın 2.26 cm ve 3. ilacın 3.99 cm'dir. Veri setinin tasarım tipi incelendiğinde komponent gruplarının ortalama örneklem hacminin 68 ve dengeli olmadığı görülmektedir. Tümör boyutlarının uygulanan tedavi yöntemlerine göre dağılımları ise Şekil 8'te verilmiştir.



Şekil 11. Yaşam modeli örnek 3 verilerinin dağılımı

Dağılımlar incelendiğinde sağa çarpık oldukları açıkça görülmektedir. Test edilmesi gereken durum ise tedavi türlerine göre ortalama tümör boyutları arasında istatistiksel olarak anlamlı bir fark olup olmadığıdır. Çözümü için geliştirilen testler kullanılmış ve elde edilen p-değerleri Tablo 19’de verilmiştir.

Tablo 19. Yaşam modeli 3 örneği üzerinde H_0 hipotezini test etmek için uygulanan testlerin p-değerleri

Test	p-değeri
GPD	0.0017*
PB	0.0132*
FY	0.0161*
HY1	0.0072*
HY2	0.0079*
HY3	0.0194*
HY4	0.0138*

Yukarıdaki tabloda verilen testlerin p-değeri sonuçlarına göre 0.05 anlamlılık düze-

yinde H_0 hipotezi reddedilir. Tümör tiplerine göre tümör boyutları arasında istatistiksel olarak anlamlı bir fark bulunduğu sonucuna varılır.

5.4. Ekonomik Gösterge Örneği

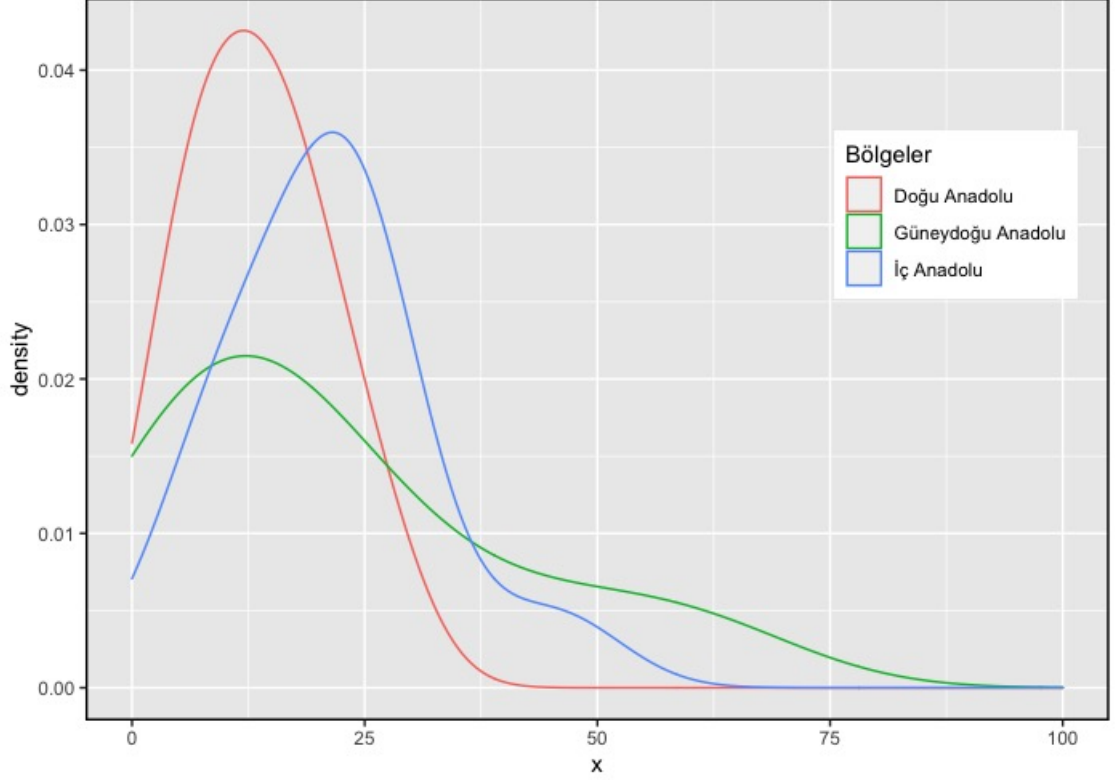
Ekonomik göstergelerden oluşan veriler genellikle üstel ya da Pareto dağılım gösterir. Bu örnekte ise Türkiye’de 2017 yılında illere göre elde edilen tarım gelirleri 1.000.000 TL cinsinden ele alınmış ve Tablo 20’de parametre tahminleriyle birlikte verilmiştir. Burada yanıtlanması gereken soru coğrafi bölgelere göre tarım gelirleri arasında bir farklılık olup olmadığıdır. Cavus ve Yazici (2020b) çalışmalarında normal dağılım göstermeyen ekonomik verilerin normallik dönüşümü ya da parametrik olmayan yöntemler kullanılarak anakütle ortalamalarının eşitliğinin test edilmesi probleminde yanlış kararlar alınmasına neden olacağını belirtmiş ve simülasyon çalışmalarıyla bu durumu desteklemişlerdir.

Tablo 20. *Ekonomik gösterge örneği verileri*

Bölge	Akdeniz	Ege	Marmara	Karadeniz	İç Anad.	D. Anad.	G. Anad.
	66.74	83.72	9.623	9.61	21.70	26.02	23.06
	19.81	48.74	20.58	12.50	45.53	8.51	15.48
	16.06	35.62	21.13	6.81	109.48	15.13	6.33
	57.23	50.29	14.31	2.14	26.20	16.38	60.68
	65.56	73.97	51.52	3.61	7.50	13.92	43.92
	28.45	34.12	32.88	18.03	24.0	7.54	25.08
	25.15	17.65	52.74	9.15	21.92	19.34	9.33
	13.02	12.33	7.09	39.24	12.92	14.70	5.81
			8.77	24.80	13.16	6.62	6.48
			23.28	22.88	28.24	3.64	
			3.10	17.73	22.69	24.76	
				15.63	21.63	17.88	
				22.73	7.90	8.74	
				12.52		5.10	
				15.13			
				7.15			
				4.06			
				5.01			
n_i	8	8	11	18	13	14	9
$\hat{a}_i$	26.83	36.83	21.09	12.36	22.12	10.55	17.98
$\hat{b}_i$	13.02	12.33	3.10	2.14	7.50	3.64	5.81
$\hat{\mu}_i$	39.86	49.21	24.19	14.51	29.62	14.20	23.80

CvM uyum iyiliği testi kullanılarak verinin iki-parametreliliği üstel dağılım gösterip göstermediği test edilmiştir. Testinin sonucunda p-değeri 0.6786 olarak elde edilmiş ve 0.05 anlamlılık düzeyinde verinin iki-parametreliliği üstel dağılım gösterdiği görülmüştür. Tablo 20’de verilen istatistikler incelendiğinde coğrafik bölgelere göre 2017 yılı ortalama tarım gelirleri arasında büyük farklılıklar olduğu görülmektedir. Örnekte özellikle aralarında ciddi bir fark bulunan İç Anadolu, Doğu Anadolu ve Güneydoğu Anadolu bölgeleri ele alınmıştır ve bu üç bölgenin ortalamalarının eşit olup olmadığı test edilecektir. Ortalama gelirleri incelendiğinde ise İç Anadolu’nun yaklaşık 30 milyon, Doğu Anadolu’nun ortalama 14 milyon, Güneydoğu Anadolu’nun ise ortalama 24 milyon TL’dir. Ölçek parametreleri de farklılıklar göstermektedir.

Ele alınan veri setinin tasarımı dengeli olmayan küçük örneklem düzenine benzerdir. Bölgelerin gelir dağılımları ise Şekil 8’te verilmiştir.



Şekil 12. Ekonomik gösterge örneği verilerinin dağılımı

Dağılım grafiğinde birçok gelir verisi dağılımına benzer olarak sağa çarpık bir dağılım yapısındadır. İç Anadolu, Doğu Anadolu ve Güneydoğu Anadolu bölgelerinin tarım gelirleri arasındaki farklılık $H_0 : \mu_{IA} = \mu_{DA} = \mu_{GA}$ hipotezinin testi ile belirlenmiş ve elde edilen p-değerleri Tablo 21’de verilmiştir.

Tablo 21. *Ekonomik gösterge örneği üzerinde H_0 hipotezini test etmek için uygulanan testlerin p-değerleri*

Test	p-değeri
GPD	0.0830
PB	0.0932
FY	0.1487
HY1	0.0267*
HY2	0.0264*
HY3	0.0594
HY4	0.0336*

Sonuçlar incelendiğinde testlerin H_0 hipotezi hakkında verilecek karar konusunda farklılıklar bulunmaktadır. HY1, HY2 ve HY4 testleri sonucunda elde edilen H_0 hipotezinin $\alpha = 0.05$ anlamlılık düzeyinde reddedilmesi için yeterli kanıt olduğunu gösterirken, GPD, PB, FY ve HY3 testlerinden elde edilen sonuçlar ise yeterli kanıt olmadığını göstermektedir. Bu gibi durumlarda bir önceki bölümde verilen Monte-Carlo simülasyon sonuçlarından faydalanılmalıdır. H_0 hipotezinin testinde kullanılan veride yer alan grup sayısı, örneklem hacmi ve tasarım düzeni incelendiğinde karşılaştırılan grup sayısının 3, toplam örneklem hacminin 36 ve tasarımın dengesiz olduğu görülmektedir. Ölçek parametreleri de hesaba katıldığında Tablo 4'te yer alan simülasyon sonuçlarına göre dengeli olmayan küçük örneklem hacminde GPD, PB ve FY testlerinin performansı 1.tip hata olasılığı ve cezalandırılmış güç hesaba katılarak değerlendirildiğinde 1.tip hata olasılığını kontrol edemediği görülmektedir. Bu bilgidен yararlanıldığında bu testler yerine 1.tip hata olasılığını kontrol edebilen HY1, HY2 ve HY4 testlerinden elde edilen sonuçlar izlenerek H_0 hipotezinin $\alpha = 0.05$ anlamlılık düzeyinde reddedilmesi için yeterli kanıt olduğu ve reddedilmesi gerektiği sonucuna varılır.

Görüldüğü gibi iki parametrelili üstel dağılmış anakütle ortalamalarının eşitliğinin test edilmesi çok farklı alanlarda kullanılan geniş bir uygulama alanına sahiptir. Geliştirilen testler sayesinde güvenilirlik analizinde, biyoistatistik uygulamalarında ve ekonomik analizlerde kullanılabilir. Geleneksel yöntemlerin kısıtlayıcı varsayımlarından etkilenmeyen HY testleri ile literatürde yer alan testlerin araştırmacıları yanlış kararlar almasına neden olmasının önüne geçilebilmektedir.

6. SONUÇ VE TARTIŞMA

Varyans homojenliği ve normal dağılım varsayımları KF testinin kullanım alanını kısıtlamaktadır. Varsayım bozulmalarının neden olduğu 1.tip hata olasılığı artışı ve testin gücündeki azalış araştırmacıları alternatif çözüm yolları ve yöntemleri geliştirmeye mecbur kılmıştır. Genelleştirilmiş Behrens-Fisher problemi olarak adlandırılan ve varsayımlardan varyans homojenliğinin sağlanmaması durumunda literatürde geliştirilen tam olasılıklı, yaklaşık ve yeniden örnekleme yöntemlerine dayalı birçok test bulunmaktadır. Yapılan araştırmalarda bu yöntemlerin üstünlük ve zayıflıkları detaylı olarak birçok kez incelenmiştir. Ancak normallik varsayımının bozulması durumu ortaya çıkış nedenine bağlı olarak daha karmaşık bir problemdir ve geniş bir çözüm alanı gerektirmesi çözülmeyi bekleyen bir problem olmasında neden olmaktadır. Ortaya çıkış nedenleri ise aykırı değer veya farklı bir parametrik dağılım ailesine uygun bir dağılım gözlenmesi olarak sınıflandırılabilir.

Aykırı değerden kaynaklanan normallik varsayımının bozulması durumunda homojen varyanslı olmayan anakütle ortalamalarının eşitliğinin test edilmesi için sağlam tahminciler kullanılarak yapılan düzeltmeler bulunmaktadır. Cavus vd. (2017) sağlam tahmincilerden M-tahmincilerini kullanarak GF testine alternatif olarak MGF testini önermişlerdir ve bu testin normal dağılım varsayımının yalnızca aykırı değerden kaynaklı bozulması durumunda güçlü sonuçlar elde etmişlerdir. Cavus vd. (2018) çalışmalarında MGF testinin gerçek veri seti üzerinde uygulamasını yaparak literatürdeki testlerden daha güçlü sonuçlar verdiğini göstermiştir. Ancak bu çalışmalar ve bu konuda yapılan benzer çalışmalarda (Karagoz, 2016; Karagoz ve Saracbası, 2016) göstermektedir ki, sağlam tahminciler kullanılarak önerilen düzeltilmiş testler çarpıklığın artış hızına bağlı olarak belirli bir noktadan sonra performansları istenilen seviyenin uzağında kalmaktadır. Bu durumda çözüm, normal dağılım dışında ilgili problemde gözlenen verinin yapısına bağlı olarak en uygun parametrik dağılım ailesi üzerinde bir test geliştirilmesidir.

GPD, PB ve FY yöntemlerine dayalı olarak normal dağılım dışındaki parametrik dağılımlar için geliştirilen testler Behrens-Fisher tipi problemlerin çözümünde sıklıkla kullanılmıştır. Son yıllarda artan hesaplama teknolojisi ile birlikte benzer problemlerin çözümünde HY yöntemi de sıklıkla tercih edilmektedir. Bunun ne-

deni diğer yöntemlerde öncelikle örnekleme dağılımı bilgisine ihtiyaç duyulurken bu yöntemde ihtiyaç duymadan çözümler geliştirilebilmesidir. Normal, log-normal, ters normal gibi farklı dağılımlar için yapılan hipotez testlerinde *nuisance* parametrenin var olması durumunda HY yaklaşımına dayalı birçok test geliştirilmiştir. Chang vd. (2011) gamma dağılımı için, Gokpinar vd. (2013) ters normal dağılım için, Abdollahnezhad ve Jafari (2015) power-law, Gokpinar ve Gokpinar (2017) ve Abdollahnezhad ve Aghadoust (2018) log-normal dağılım için HY yöntemine dayalı testler geliştirmişlerdir. HY yönteminin örnekleme dağılımına ihtiyaç duymaması bu denli sık kullanılmasının en önemli nedenidir. Ancak görüldüğü gibi henüz üzerinde çalışılmayan parametrik dağılım aileleri bulunmaktadır ve bunlardan biri de birçok farklı uygulama alanı olan iki parametrelili üstel dağılımdır.

Bilinmeyen ya da eşit olmayan ölçek parametresi içeren iki parametrelili üstel dağılmış anakütle ortalamalarının eşitliğinin test edilmesi probleminin çözümü için GPD, PB ve FY yöntemlerini kullanarak Malekzadeh ve Jafari (2019) literatürde yapılan ilk ve tek çalışmadır, ancak yalnızca ikinci tip ilerleyen sansürlü veri yapısı ele alınmıştır. Bu tez çalışmasında ise amaçlanan sansürlü olmayan veriler için benzer problemin çözümünde kullanılacak güçlü testler oluşturmaktır. Bu çalışmada öncelikle bahsedilen çalışmadan yola çıkarak sansürlü olmayan veriler için geliştirilen bu testler yeniden düzenlenmiştir. Ayrıca küçük örneklem hacimlerinde daha iyi sonuçlar veren HY yöntemine dayalı yeni çözümler önerilmiştir. Çözümler oluşturulurken ilk test önerisinde ilgilenilen ortalama teriminin en çok olabilirlik tahmincisi kullanılmış, ikinci test önerisinde ortalama teriminin en çok olabilirlik tahmininin yanı sıra Malekzadeh ve Jafari (2019) çalışmasında kullandıkları konum parametresinin yansızlık düzeltmesi kullanılmış, üçüncü test önerisinde ortalama teriminin tahmininde ölçek parametresinin en çok olabilirlik ve konum parametresinin cezalandırılmış en çok olabilirlik tahmincisi ile yansızlık düzeltmesi kullanılmış, dördüncü test önerisinde ise ölçek parametresinin en çok olabilirlik tahmincisi ve konum parametresinin cezalandırılmış en çok olabilirlik tahmincisi kullanılarak geliştirilmiştir. Farklı çözümler önerilmesinin sebebi iki parametrelili üstel dağılımın konum parametresinin en çok olabilirlik tahmincisinin yanlış olmasıdır. Elde edilen yeni testlerin performansları sansürlü olmayan veriler üzerinde Monte-Carlo simülasyon çalışmaları ile araştırılmıştır.

Monte-Carlo simülasyon çalışmalarında ise karşılaştırılan testlerin 1.tip hata olasılıklarının eşit ya da benzer olmadığı görülmüştür. Bu durumda testlerin güçlerini karşılaştırmak doğru bir yaklaşım değildir. Bunun nedeni liberal bir yapıda olan yani H_0 hipotezini reddetme eğiliminde olan bir testin hem 1.tip hata olasılığı hem de gücünün yüksek değerler olması doğaldır. Tam aksi durumda testin tutucu bir yapıda olması da aksi durumun gerçekleşmesi demektir. Literatürde bu sorunu çözmek için Zhang ve Boos (1994) tarafından düzeltilmiş güç ve Lloyd (2005) tarafından ise asıl güç kriterleri testin gücüne alternatif olarak önerilmiştir. Bu öneriler temel olarak testin gücünü 1.tip hata olasılığının nominal seviyenin etrafında aldığı değerlere göre düzenlemektir. Ancak bu önerilerin bazı eksiklikleri bulunmaktadır ve 3.bölümde detaylı olarak bahsedilmiştir. Bu nedenle tez çalışmasında kullanmak üzere alternatiflerinin eksikliklerini taşımayan cezalandırılmış güç kriteri geliştirilmiştir (Cavus vd., 2019a). Böylelikle 1.tip hata olasılığı farklı olan testlerin güç karşılaştırmaları sağlıklı bir şekilde yapılabilmektedir.

4. Bölüm’de verilen Monte-Carlo simülasyon sonuçlarına göre açıkça görülmektedir ki, HY yöntemine dayalı testler 1.tip hata olasılıkları bakımından diğer yöntemler kullanılarak oluşturulan testlere göre özellikle küçük örneklem hacimlerinde çok iyi performans göstermişlerdir. Önerilen HY testleri kendi içlerinde karşılaştırıldıklarında ise HY1 testi alternatiflerine göre daha dengeli bir performans sergilemiştir. Örneklem hacmi değişiminden HY3 ve HY4 testine göre çok daha az etkilenmiş, HY2’ye benzer bir performans gösterse de nominal seviyeye daha yakın bir seyir izlemiştir. Dengeli tasarımlarda 1.tip hata olasılığını kontrol etmedeki becerisine benzer olarak HY yöntemine dayalı testler diğer yöntemlere göre daha iyi sonuçlar vermektedir. PB testi $k = 3$ anakütle ortalamasının eşitliğinin test edilmesinde dengeli olmayan tasarımlarda HY yöntemine dayalı testlere göre daha güçlü iken, *nuisance* parametrenin aldığı değer arttığında aradaki fark azalmaktadır. $k = 4$ anakütle sayısı için dengeli tasarımda HY1, HY2 ve HY4 testi, dengeli olmayan tasarımlarda HY1, HY2 ve HY4 testinin yanı sıra GPD ve PB testi en güçlü testlerdir. Örneklem hacmi büyüdükçe HY3 testinin performansında da belirgin artışlar gözlenmekte ve en güçlü testler grubuna dahil olmaktadır. Ölçek parametresi değeri yükseldiğinde GPD testi gücünü kaybetmekte, PB, HY3 ve HY4 testinin performansı ön plana çıkmaktadır. $k = 5$ anakütle sayısı için dengeli tasarımda HY1 ve

HY3 testi, dengeli olmayan tasarımlarda ise bu testlerin yanı sıra GPD ve PB testi en güçlüdür. Örneklem hacmi arttığında ve tasarım dengesizleştğinde HY1 ve HY3 testlerinde performans düşmekte, GPD ve PB testleri performanslarını korumaktadır. Ölçek parametresi değerinin artışı ise dengeli olmayan tasarımlarda GPD ve PB testinin yanı sıra HY3 testinin performansı da ön plana çıkmaktadır. Sonuç olarak HY testleri genel olarak dengeli tasarımlarda, PB testi ise dengeli olmayan tasarımlarda iyi performans sergilemektedir. k anakütle sayısı arttığında GPD testinin dengeli olmayan tasarımlarda performansı ön plana çıkmaktadır.

Simülasyon sonuçları 5. Bölüm’de yer alan gerçek hayat problemleri üzerinde yapılan uygulamalarda desteklenmiştir. Ürün güvenilirliği, yaşam modellemesi ve ekonomik gösterge örnekleri üzerinde çalışmada ele alınan testler uygulanmıştır. Örneklerde testler H_0 hipotezinin reddine ilişkin aynı kararı vermelerine rağmen, yalnızca son örnekte farklı sonuçlar vermişlerdir. Bu noktada simülasyon çalışmasındaki sonuçlardan faydalanarak hipotez testine ilişkin uygulanması gereken en uygun yöntemlerin HY1 ve HY2 testleri olduğu görülmüş ve verdikleri sonuçlar kullanılmıştır.

Özetle bu çalışmada gerçek hayat problemlerinde sıklıkla karşılaşılan normal dağılım varsayımının bozulması durumunda *nuisance* parametre içeren bir problemin çözümü üzerinde çalışılmıştır. Normal dağılmama durumunda çarpık dağılım gösteren verilerin modellenmesinde kullanılan iki parametrelili üstel dağılım ele alınmıştır. İki den fazla anakütle ortalamasının eşit olmayan ölçek parametresi durumunda eşitliğin test edilmesi için Malekzadeh ve Jafari (2019) çalışmasında verilen GPD, PB ve FY testleri sansürlü olmayan veriler için yeniden düzenlenmiş ve HY yöntemine dayalı alternatif testler geliştirilmiştir. Tüm bu testlerin performansları geliştirilen cezalandırılmış güç kriterinden faydalanarak Monte-Carlo simülasyonları ile araştırılmış ve çeşitli senaryolar altında hangi testin kullanılması gerektiğine dair sonuçlar verilmiştir. Son bölümde ise gerçek hayat problemlerinden örnekler verilmiş ve bu örnekler üzerinde testlerin uygulamaları yapılmıştır. Hesaplama yapısı gereği karmaşık olan bu testlerin kullanılabilirliğini sağlamak için R üzerinde *doex* adında bir paket geliştirilmiştir (Cavus ve Yazici, 2019b). Pakete eklenen hazır fonksiyonlar ile tez çalışmasında elde edilen bulgulardan faydalanılmasının arttırılması hedeflenmiş ve böylece çalışmanın yaygın etkisi arttırılmıştır.

KAYNAKÇA

- Abdollahnezhad, K., and Aghadoust, S. (2018) Computational approach test in one-way fixed effects ANOVA models of log-normal samples. *Journal of Statistical Modelling: Theory and Applications*, 1 (1): 1–11.
- Abdollahnezhad, K., and Jafari, A.A. (2015) Inferences on the parameters of power law distribution. *ProbStat Forum*, 8: 130–139.
- Akins, R. B., Tolson, H. and Cole, B. R. (2005) Stability of response characteristics of a delphi panel: Application of bootstrap data expansion. *BMC Medical Research Methodology*, 5 (37): 1–12.
- Alexander, R.A. and Govern, D.M. (1994) A new and simpler approximation for ANOVA under variance heterogeneity. *Journal of Educational Statistics*, 19(2), 91-101.
- Alvandi, S.M., Jafari, A.A. and Mardani-Fard, H.A. (2012) One-way ANOVA with unequal variances. *Communications in Statistics: Theory and Methods*, 41, 4200-4221.
- Ananda, M.M.A. and Weerahandi, S. (1996) Testing the difference of two exponential means using generalized p-values. *Communications in Statistics - Simulation and Computation*, 25, 521– 532.
- Andersen, P.K., Borgan, O., Gill, R.D. and Keiding, N. (1993) *Statistical Models Based on Counting Processes*. Springer-Verlag.
- Asiribo, O. and Gurland, J. (1990) Coping with variance heterogeneity. *Communications in Statistics: Theory and Methods*, 19 (11), 4029-4048.
- Aspin, A.A. (1948) An examination and further development of a formula arising in the problem of comparing two means. *Biometrika*, 35, 88-96.
- Berry, K.J. and Mielke, P.W. (2002) The Fisher-Pitman permutation test: an attractive alternative to the F test. *Psychological Reports*, 90, 495-502.
- Blanca, M.J., Alarcon, R., Arnau, J., Bono, R., and Bendeyan, R. (2017) Non-normal data: is ANOVA still a valid option. *Psicothema*, 29 (4), 552-557.

- Box, G.E.P. (1954) Some theorems on quadratic forms applied in the study of analysis of variance problems. *Annals of Mathematical Statistics*, 25, 290-302.
- Box, G.E.P. and Cox, D.R. (1964) An analysis of transformation. *Journal of the Royal Statistical Society*, 26, 211-252.
- Brown, M.B. and Forsythe, A.B. (1974) The small sample behavior of some statistics which test the equality of several means. *Technometrics*, 16, 129–132.
- Brunner, E., Dette, H., and Munk, A. (1997) Box-type approximations in nonparametric factorial designs. *Journal of the American Statistical Association*, 92, 1494–1502.
- Cavus, M., Yazici, B. and Sezer, A. (2017) Modified tests for comparison of group means under heteroscedasticity and non-normality caused by outliers. *Hacet-tepe Journal of Mathematics and Statistics*, 46 (3), 493-510.
- Cavus, M., Yazici, B. and Sezer, A. (2018) Analyzing regional export data by the modified generalized F-test. *International Journal of Economic and Administrative Studies*, 541–552.
- Cavus, M., Yazici, B. and Sezer, A. (2019a) Penalized power approach to compare the power of the tests when Type I error probabilities are different. *Communications in Statistics - Simulation and Computation*.
doi: 10.1080/03610918.2019.1588310.
- Cavus, M. and Yazici, B. (2019b) doex: The One-Way Heteroscedastic ANOVA Tests. R package version 1.2. <https://CRAN.R-project.org/package=doex>.
- Cavus, M. and Yazici, B. (2020a) Usage of doex R-package for ANOVA type problems in lifetime modelling. *International Artificial Intelligence in Health Congress*, Izmir, Turkey.
- Cavus, M. and Yazici, B. (2020b) Comparison of Hsieh test and ANOVA for log-transformed on income data. *20th International Symposium on Econometrics, Operations Research and Statistics*, Ankara, Turkey.
- Cavus, M. and Yazici, B. (2020c) One-way non-normal ANOVA in reliability analysis using with doex. *The European R Users Meeting 2020*. Milano, Italy.

- Chang, H. Y., and C. M. Suchindran. 1997. Testing overdispersion in data with censoring using the mixture of exponential families. *Communication in Statistics*, 26 (12):2945–66.
- Chang, C.H., Lin, J.J., and Pal, N. (2011) Testing the equality of several gamma means: a parametric bootstrap method with applications. *Computational Statistics*, 26:55-76.
- Chen, H.J. (1982) A new range statistic for comparisons of several exponential location parameters. *Biometrika*, 69, 257–260.
- Chen, X. 2006. The adjustment of random baseline measurements in treatment effect estimation. *Journal of Statistical Planning and Inference*, 136 (12):4161–75.
- Chen, S.Y. and Chen, H.J. (1998) Single-stage analysis of variance under heteroscedasticity. *Communications in Statistics - Simulation and Computation*, 27 (3), 641-666.
- Chen, S.Y. and Chen, H.J. (2000) A range test for the equality of means when variances are unequal. *American Journal of Mathematical and Management Sciences*, 20, 145-170.
- Chen, X., and L. Wei. 2003. A comparison of recent methods for the analysis of small-sample cross-over studies. *Statistics in Medicine*, 22 (18):2821–33.
- Chhikara, R.S., and Folks, J.L. (1989) *The Inverse Gaussian Distribution: Theory, Methodology, and Applications*. Marcel Dekker, New York.
- Cochran, W.G. (1937) Problems arising in the analysis of a series of similar experiments. *Journal of the Royal Statistical Society*, 4, 102-118.
- Constantini, M. and Lupi, C. (2013) A simple panel-CDF test for unit roots. *Oxford Bulletin of Economics and Statistics*, 75 (2), 276–96.
- Cressie, N., L. Pardo, and M. C. Pardo. 2003. Size and power considerations for testing log-linear models using divergence test statistics. *Statistica Sinica*, 13:555–70.
- Cribbie, R.A. (2012) Effects on non-normality on test statistics for one-way independent groups designs. *British Journal of Mathematical and Statistical*

- Psychology*, 65 (1), 56-73.
- Dag, O., Dolgun, A. and Konar, N. M. (2018) Onewaytests: An R package for one-way tests in independent group designs. *The R Journal*, 10 (1), 175–99.
- Dielman, T. E. 2005. Least absolute value regression: Recent contributions. *Journal of Statistical Computation and Simulation*, 75 (4):263–86.
- Dielman, T. E. 2006. Variance estimates and hypothesis tests in least absolute value regression. *Journal of Statistical Computation and Simulation*, 76 (2):103–14.
- Dielman, T. E. 2009. A note on hypothesis tests after correction for autocorrelation: Solace for the Cochrane-Orcutt method. *Journal of Modern Applied Statistical Methods*, 8 (1):100–109.
- Dielman, T. E., and E. L. Rose. 1997a. A note on hypothesis testing in LAV multiple regression: A small sample comparison. *Computational Statistics and Data Analysis*, 23 (3):381–8.
- Dielman, T. E., and E. L. Rose. 1997b. Estimation and testing in least absolute value regression with serially correlated disturbances. *Annals of Operations Research*, 74:239–57.
- Faraway, J., Marsaglia, G., Marsaglia, J., and Baddeley, A. (2019) goftest: Classical Goodness-of-Fit Tests for Univariate Distributions. R package version 1.2-2. <https://CRAN.R-project.org/package=goftest>.
- Fisher, R.A. (1925) *Statistical methods and scientific inference*. Oxford, England. Oliver and Boyd.
- Gamage, J. and Weerahandi, S. (1998) Size performance of some tests in one-way ANOVA. *Communications in Statistics: Simulation and Computation*, 27 (3), 625-640.
- Ghosh, M. and Razmpour, A. (1984) Estimation of the common location parameter of several exponentials. *Sankhya: The Indian Journal of Statistics, Series A*, 46, 383–394.
- Gokpinar, E.Y. and Gokpinar, F. (2015) Testing equality of variances for several normal populations. *Communications in Statistics- Simulation and Compu-*

- tation, 46 (1), 38-52.
- Gokpinar, E.Y. (2017) Comparison of tests for the equality of several log-normal means. *Gazi University Journal of Science*, 30 (3), 209-222.
- Gokpinar, E.Y. and Gokpinar, F. (2017) Testing the equality of several log- normal means based on a computational approach. *Communications in Statistics - Simulation and Computation*, 46 (3), 1998-2010.
- Gokpinar, E.Y., Polat, E., Gokpinar, F. and Günay, S. (2013) A new computational approach for testing equality of inverse Gauss means under heterogeneity. *Hacettepe Journal of Mathematics and Statistics*, 42 (5), 585-590.
- Guo, X., Wu, H., Li, G. and Li, Q. (2017) Inference for the common mean of several Birnbaum–Saunders populations. *Journal of Applied Statistics*, 44 (5), 941-954.
- Güven, G . (2018). A comparison of five tests for the equality of inverse Gaussian means under heteroscedasticity of scale parameters. *Gazi University Journal of Science*, 31 (2), 628-641.
- Hanson, T. A., and J. R. Hall. 2012. Statistical arbitrage and robust tests for cointegration. Midwest Finance Association 2013 Annual Meeting Paper.
- Hartung, J., Argaç, D. and Makambi, K. (2002) Small sample properties of tests on homogeneity in one-way ANOVA and meta-analysis. *Statistical Papers*, 41, 197-235.
- Heritier, S., Lloyd, C. J. and Lo, S. N. (2017) Accurate p-values for adaptive designs with binary endpoints. *Statistics in Medicine*, 36 (17), 2643–2655.
- Hsieh, H.K. (1986) An exact test for comparing location parameters of k exponential distributions with unequal scaled based on Type II censored data. *Technometrics*, 28, 157–164.
- Jack, A., and Kokoszka, P. (2017) Wavelet semi-parametric inference for long memory in volatility in the presence of a trend. *Journal of Statistical Computation and Simulation*, 87 (8), 1498–1519.
- Jafari, A.A., and Abdollahnezhad, K. (2017) Testing the equality means of seve-

- ral log-normal distributions. *Communications in Statistics - Simulation and Computation*, 46 (3), 2311–2320.
- Johansen, S. (1980) Whe Welch-James approximation to the distribution of the residual sum of squares in a weighted linear regression. *Biometrika*, 67 (1), 58-92.
- Kambo, N.S. and Awad, A.M. (1985) Testing equality of location parameters of k exponential distributions. *Communications in Statistics - Theory and Methods*, 14, 567–583.
- Karagoz, D. (2016). Modified Welch test statistic for ANOVA under Weibull distribution. *Hacettepe Journal of Mathematics and Statistics*, 45, 56-65.
- Karagoz, D. and Saraçbaşı, T. (2016). Robust Brown-Forsythe and robust modified Brown-Forsythe ANOVA tests under heteroscedasticity for contaminated Weibull distribution. *Revista Colombiana de Estadística*, 39 (1), 17-32.
- Krishnamoorthy, K., Lu, F. and Mathew, T. (2007) A parametric bootstrap approach for anova with unequal variances: Fixed and random models. *Computational Statistics and Data Analysis*, 51, 5731-5742.
- Kruskal, W.H., and Wallis, W.A. (1952) Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47 (260), 583-621.
- Krutchkoff, R.G. (1988) One-way fixed effects analysis of variance when the error variances may be unequal. *Journal of Statistical Computation and Simulation*, 30(4), 259-271.
- Li, X. (2009) A generalized α -value approach for comparing the means of several log-normal populations. *Statistics and Probability Letters*, 79 (11), 1404-1408.
- Li, X., Wang, J. and Liang, H. (2011) Comparison of several means: a fiducial based approach. *Computational Statistics and Data Analysis*, 55, 1993-2002.
- Lilienthal, J., R. Fried, and A. Schumann. 2018. Homogeneity testing for skewed and cross-correlated data in regional flood frequency analysis. *Journal of Hydrology* 556:557–71.
- Lin, S.H., and Wang, R.S. (2013) Modified method on the means for several log-

- normal distributions. *Journal of Applied Statistics*, 40 (1), 194-208.
- Lloyd, C. J. (2005) Estimating test power adjusted for size. *Journal of Statistical Computation and Simulation*, 75 (11), 921-933.
- Luh, W.M. and Guo, J.H. (2005) Heteroscedastic test statistics for one-way analysis of variance: the trimmed means and Hall's transformation conjunction. *The Journal of Experimental Education*, 74 (1), 75-100.
- Ma, C.X. and Tian, L. (2009) A parametric bootstrap approach for testing equality of inverse Gaussian means under heterogeneity. *Communications in Statistics-Simulation and Computation*, 38, 1153-1160.
- Malekzadeh, A. and Jafari, A.A. (2019) Inference on the equality means of several two-parameter exponential distributions under progressively Type II censoring. *Communications in Statistics - Simulation and Computation*.
- Mehrotra, D.V. (1997) Improving the Brown-Forsythe solution to the generalized Behrens-Fisher problem, 26 (3), 1139-1145.
- Niu, C. , Guo, X., Xu, W., Zhu, L. (2014) Comparison of several Birbaum-Saunders distributions. *Journal of Statistical Computation and Simulation*, 84, 2721-2733.
- Ochuko, T.K., Abdullah, S., Zain, Z.B. and Yahaya, S.S.S. (2015) The modification and evaluation of the Alexander-Govern test in terms of power. *Modern Applied Sciences*, 9 (13), 1-21.
- Ochuko, T.K., Abdullah, S., Zain, Z.B. and Yahaya, S.S.S. (2016) Graphical representation of the power rates for the winsorized modified Alexander-Govern test. *International Journal of Statistics and Applications*, 6 (2), 58-80.
- Oshima, T.C., Algina, J. (1992) Type I error rates for James's second-order test and Wilcox's H test under heteroscedasticity and non-normality. *British Journal of Mathematical and Statistical Psychology*, 45 (2), 255-263.
- Ozdemir, A.F. and Kurt, S. (2006) One way fixed effect analysis of variance under variance heterogeneity and a solution proposal. *Selçuk Journal of Applied Mathematics*, 7 (2), 81-90.
- Ozdemir, A.F., Wilcox, R. and Yildiztepe, E. (2017) Comparing j independent

- groups with a method based on trimmed means. *Communication in Statistics-Simulation and Computation*, 1-12.
- Pal, N., Lim, W.K. and Ling, C.H. (2007) A computational approach to statistical inferences. *Journal of Applied Probability and Statistics*, 2 (1), 13-35.
- Parker, C., Paparoditis, E. and Palitidis, D. (2015) Tapered block bootstrap for unit root testing. *Journal of Time Series and Econometrics*, 7 (1), 37-67.
- Scott, A. and Smith, T. (1971) Interval estimates for linear combinations of means. *Applied Statistics*, 20, 276-285.
- Singh, N. (1983) The likelihood ratio test for the equality of location parameters of exponential populations based on Type II censored samples. *Technometrics*, 25, 193-195.
- Tian, L. (2006) Testing equality of inverse Gaussian means under heterogeneity based on generalized test variable. *Computational Statistics Data Analysis*, 51, 1156-1162.
- Tiku, M.L. and Vaughan, D.C. (1991) Testing equality of location parameters of two exponential distributions from censored samples. *Communication in Statistics-Theory and Methods*, 20, 929-944.
- Trinquart, L., Ioannidis, Chatellier, J., Ravaud, P. (2014) A test for reporting bias in trial networks: simulation and case studies. *BMC Medical Research Methodology*, 14 (112):112-1. doi:10.1186/1471-2288-14-112.
- Tsui, K. and Weerahandi, S. (1989) Generalized p-values in significance testing of hypotheses in the presence of nuisance parameters. *Journal of the American Statistical Association*, 84, 602-607.
- Vaughan, D.C. and Tiku, M.L. (1993) Testing the equality of location parameters of exponential populations from censored samples. *Communication in Statistics-Theory and Methods*, 22, 2567-2581.
- Viveros, R. and Balakrishnan, N. (1994) Interval estimation of parameters of life from progressively censored data. *Technometrics*, 36, 84-91.
- Wang, L. and Qu, A. (2007) Robust tests in regression models with omnibus alter-

- natives and bounded influence. *Journal of American Statistical Association*, 102 (477), 347–358.
- Weerahandi, S. (1995) ANOVA under unequal error variances. *Biometrics*, 51, 589–599.
- Welch, B.L. (1951) On the comparison of several mean values. *Biometrika*, 38, 330–336.
- Wilk, M. B. and Gnanadesikan, R. (1968) Probability plotting methods for the analysis of data. *Biometrika*, 55 (1), 1–17.
- Wu, S.F. (2019) One stage multiple comparisons with the control for exponential mean lifetimes based on doubly censored samples under heteroscedasticity. *Communications in Statistics Simulation and Computation*.
- Wu, S. F. and Wu, C. C. (2005) Two stage multiple comparisons with the average for exponential location parameters under heteroscedasticity. *Journal of Statistical Planning and Inference*, 134 (2), 392–408.
- Zhang, J. and Boos, D.D. (1994) Adjusted power estimates in monte carlo experiments. *Communications in Statistics-Simulation and Computation*, 23 (1), 165–173.
- Zheng, M. (2013) *Penalized maximum likelihood estimation of two-parameter exponential distributions*. Master Dissertation. The Faculty of Graduate School of the University of Minnesota.

ÖZGEÇMİŞ

Adı-Soyadı : Mustafa ÇAVUŞ
Yabancı Dili : İngilizce
Doğum Yeri ve Yılı : Kütahya - 1989
E-Posta : mustafacavus@eskisehir.edu.tr
ORCID : 0000-0002-6172-5449

Eğitim ve Mesleki Geçmişi:

- 2016 - 2020, Eskişehir Teknik Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı (Doktora Programı)
- 2014 - 2016, Anadolu Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı (Yüksek Lisans Programı)
- 2009 - 2014, Anadolu Üniversitesi, İktisat Fakültesi, Uluslararası İlişkiler Bölümü (Lisans Programı)
- 2007 - 2013, Anadolu Üniversitesi, Fen Fakültesi, İstatistik Bölümü (Lisans Programı)
- 2011 - 2012, Warsaw University of Technology, Faculty of Mathematics and Information Science, Department of Computer Science (Erasmus Learning Mobility Program)

Yayınlar ve Bilimsel/Sanatsal Faaliyetleri:

1. Uluslararası Hakemli Dergilerde Yayınlanan Makaleler

Kan Kilinc, B. and **Cavus, M.** (2015) Flexible logistic models with correlated variables, *Turkish Clinics Biostatistics Journal*, 7 (2), 90-95.

Cavus, M., Yazici, B. and Sezer, A. (2017) Modified tests for comparison of group means under heteroskedasticity and non-normality caused by outlier(s), *Hacettepe Journal of Mathematics and Statistics*, 46 (3), 492-510.

- Kan Kilinc, B. and **Cavus, M.** (2017) İkili yanıt değişkenine sahip modellerin yeterliliklerine ilişkin benzetim çalışması, *Sakarya üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 21 (2), 169-177.
- Cavus, M.**, Yazici, B. and Sezer A. (2018) Analysing regional export data by the modified generalized F-test, *International Journal of Economics and Administrative Studies*, 18. EYI Special Issue, 541-552.
- Cavus, M.**, Ozdemir, O., Sezer A. (2019) İstanbul ve Eskişehir bölgesi minimum maksimum yağış miktarlarının ekstrem değerler dağılımları ailesi ile modellenmesi, *Süleyman Demirel üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 33, 33-39.
- Cavus, M.**, Yazici, B. and Sezer A. (2019) Penalized power approach to compare the power of the tests when Type I error probabilities are different, *Communications in Statistics-Simulation and Computation*.
- Cavus, M.**, Kan Kilinc, B., Yazici, B., Tekeli, S., Gunsoy, G., Gunsoy, B. and Karaduman, C. (2019) Modeling the contribution of distance education to students' preparation for the professions, *Turkish Journal of Distance Education*.

2. Yazılan Uluslararası Kitaplar ve Kitaplardaki Bölümler

- Cavus, M.**, Sezer, A. and Yazici, B. (2015) A simulation study on generalized Pareto mixture model. *Computational Problems in Science and Engineering*, Springer, 342, 249-259.

3. Uluslararası Bilimsel Toplantılarda Sunulan Bildiriler

- Cavus, M.**, Sezer A. and Yazici B. (2014) Generating random samples from generalized Pareto mixture model. *Applied Mathematics, Simulating and Modelling*. Florence, Italy.
- Cavus, M.** and Ozdemir, O. (2014) Performance of the simulation method with Gumbel distribution. *Applied Mathematics, Simulating and Modelling*. Flo-

rence, Italy.

Cavus, M. and Kan Kilinc, B. (2015) Comparison of additive logistic regression models with multicollinearity. *16th International Symposium on Econometrics, Operations Research and Statistics*. Edirne, Turkey.

Cavus, M. and Kan Kilinc, B. (2015) "The influence of multicollinearity in binary logistic regression and additive models. *60th World Statistics Congress*. Rio De Janerio, Brasil.

Sezer A., **Cavus, M.** and Ozdemir O. (2015) Sample size determination by the information based test. *The 71st Deming Conference on Applied Statistics*. New Jersey, USA.

Cavus, M., Yazici, B. and Sezer A. (2015) Comparison of means under non-normality with outliers for unbalanced data. *8th International Conference of the ERCIM WG on Computational and Methodological Statistics*. London, United Kingdom.

Kan Kilinc, B. and **Cavus, M.** (2015) Extracting the major factors affecting the traffic accidents using decision trees. *8th International Conference of the ERCIM WG on Computational and Methodological Statistics*. London, United Kingdom.

Cavus, M., Yazici, B. and Sezer A. (2016) Comparison of two approaches for the generalized F-test. *10th International Statistics Days Conference*. Giresun, Turkey.

Cavus, M. and Ozdemir, O. (2016) Performance of the inverse transformation method for extreme value distributions. *10th International Statistics Days Conference*. Giresun, Turkey.

Cavus, M., Yazici, B. and Sezer A. (2016) Modified test for comparison of group means in the presence of outlier(s). *2nd International Researchers-Statisticians and Young Statisticians Congress*. Ankara, Turkey.

Asfha H.D., Kan Kilinc, B. and **Cavus, M.** (2016) Appraisal of the performance of recursive partitioning, conditional inference tree, and evolutionary learning

- tree: a comparative study. *International Conference on Trends and Perspectives in Linear Statistical Inference*. Istanbul, Turkey.
- Cavus, M.** and Kan Kilinc, B. (2016) Qualitative data analysis using nonlinear principal component analysis. *10th International Statistics Days Conference*. Giresun, Turkey.
- Cavus, M.**, Yazici, B. and Sezer A. (2016) Robust methods to compare means under violation of normality with outliers. *12th German Probability and Statistics Days*. Bochum, Germany.
- Cavus, M.**, Yazici, B. and Sezer A. (2017) Comparison of the pairwise post-hoc comparison tests under heteroskedasticity. *3rd International Researchers-Statisticians and Young Statisticians Congress*. Konya, Turkey.
- Cavus, M.**, Yazici, B. and Sezer A. (2017) Comparison of the average amounts of exports for geographical regions in Turkey with the modified generalized F-test. *18th International Symposium on Econometrics Operations Research and Statistics*. Trabzon, Turkey.
- Tekeli S., Kan Kilinc, B., **Cavus, M.**, Gunsoy, G., Gunsoy, B., Yazici, B. and Karaduman C. (2017) Does distance learning meet the life goals and expectation. *International Conference on Computational and Statistical Methods in Applied Sciences*. Samsun, Turkey.
- Cavus, M.**, Yazici, B. and Sezer A. (2017) A new performance criterion for hypothesis testing. *International Conference on Computational and Statistical Methods in Applied Sciences*. Samsun, Turkey.
- Cavus, M.**, Yazici, B. and Sezer A. (2017) Comparison of the modified generalized F-test with the non-parametric alternatives. *10th International Statistics Congress*. Ankara, Turkey.
- Cavus, M.**, Yazici, B. and Sezer A. (2018) A criterion to evaluate power of the test. *4th International Researchers-Statisticians and Young Statisticians Congress*. Izmir, Turkey.
- Cavus, M.**, Yazici, B. and Sezer A. (2019) Comparison of some normality tests in

the presence of outliers. *DAGStat 2019*. Munich, Germany.

Cavus, M. and Yazici, B. (2019) A simulation study on the unbalanced design properties of the generalized p-value based tests. Kusadasi, Turkey

Cavus, M. and Yazici, B. (2019) A comparison study on the procedures for testing equality of exponential distributed mean lifetimes in realibility analysis. *Analytical Methods in Statistics 2019*. Liberec, Czech Republic

Yazici, B. and **Cavus, M.** (2019) Penalized power properties of computation methods for the generalized F-test. *11.International Statistics Congress*. Bodrum, Turkey

Yazici, B., Marwa, B.G. and **Cavus, M.** (2019) Nonparametric modeling via wavelet smoothing. *5th International Researchers, Statisticians and Young Statisticians Congress 2019*. Kusadasi, Turkey

Cavus, M. and Yazici, B. (2020) Usage of doex R package for ANOVA type problems in life modeling. *International Artificial Intelligence in Health Congress 2020*. Izmir, Turkey

Cavus, M. and Yazici, B. (2020) Comparison of Hsieh test and ANOVA for log-transformed on income data. *20th International Symposium on Econometrics, Operations Research and Statistics*. Ankara, Turkey

Cavus, M. and Yazici, B. (2020) One-way non-normal ANOVA in reliability analysis using with doex. *The European R Users Meeting 2020*. Milano, Italy

4. Ulusal ve Uluslararası Projeler

Araştırmacı, Ekstrem Değerler Dağılımları Ailesindeki Dağılımların Parametrelerinin Belirlenmesi: İstanbul ve Eskişehir Bölgesi Minimum ve Maksimum Yağış Miktarları için Son 50 Yıla Ait Uygulama, Anadolu üniversitesi Bilimsel Araştırma Projesi, 2015 - 2016.

Araştırmacı, Toplamsal Lojistik Regresyon Modellerinde Sınıflama Araştırması, Anadolu üniversitesi Bilimsel Araştırma Projesi, 2015 - 2016.

Araştırmacı, Deney Tasarımında Genelleştirilmiş p-Değeri, Anadolu üniversitesi

Bilimsel Araştırma Projesi, 2015 - 2016.

Araştırmacı, Güvenilirlik Analizi Uygulamaları, Anadolu üniversitesi Bilimsel Araştırma Projesi, 2016 - 2017.

Araştırmacı, Anadolu üniversitesi Açık ve Uzaktan Eğitim Sistemindeki öğrencilerin Yaşam Hedefleri, üniversite Yaşam Beklentileri ve Memnuniyetleri, Anadolu Üniversitesi Bilimsel Araştırma Projesi, 2016 - 2018.

Araştırmacı, Ürün üzerindeki Elektronik Komponentlerin Güvenilirlik Analizleri İçin Stres Faktörlerinin Araştırılması ve Modellenmesi, Sanayi Destekli Bilimsel Araştırma Projesi, Mart 2016 - Aralık 2016.

Araştırmacı, Homojen olmayan varyans durumunda normal dağılmış anakütle ortalamalarının karşılaştırılması için kullanılan testlerin cezalandırılmış güç özellikleri, Eskişehir Teknik üniversitesi Bilimsel Araştırma Destek Projesi, 2019 - 2020.