



**BASİT DOĞRUSAL REGRESYON MODELİNDE DEĞİŞEN VARYANS
DURUMU İÇİN KULLANILAN TESTLERİN PERFORMANSLARININ
İNCELENMESİ**

Betül POLATDEMİR

**YÜKSEK LİSANS TEZİ
İSTATİSTİK ANA BİLİM DALI**

**GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

OCAK 2024

ETİK BEYAN

Gazi Üniversitesi Fen Bilimleri Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmasında;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
 - Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
 - Tez çalışmasında yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
 - Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
 - Bu tezde sunduğum çalışmanın özgün olduğunu,
- bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

Betül POLATDEMİR

22/01/2024

BASİT DOĞRUSAL REGRESYON MODELİNDE DEĞİŞEN VARYANS DURUMU İÇİN KULLANILAN TESTLERİN PERFORMANSLARININ İNCELENMESİ

(Yüksek Lisans Tezi)

Betül POLATDEMİR

GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Ocak 2024

ÖZET

Basit doğrusal regresyon modelinin varsayımlarından biri, hata teriminin sabit bir varyansa sahip olmasıdır. Bu varsayım karşılanmadığında, modelin tahmin gücü ve güvenilirliği olumsuz etkilenebilir. Bu yüzden değişen varyansın olup olmadığı kontrol edilmelidir. Değişen varyansın kontrolü için grafiksel yöntemler kullanılmakla beraber birçok test geliştirilmiştir. Fakat literatürde hangi durumda hangi testin daha iyi olduğu hakkında çalışmalar oldukça kısıtlıdır. Bu tez çalışmasında literatürde var olan Breusch-Pagan (BP), Spearman Sıra Korelasyonu (SSK), Glejser (G), Golfeld-Quandt (GQ), Park (P), White (W), Harvey-Godfrey (HG), medyan değerine göre Modifiye Edilmiş Levene (MLM), ortalamaya göre Modifiye Edilmiş Levene (MLA) testleri ele alınmış ve simülasyon yöntemi ile testlerin deneysel 1.tip hata oranları ve testin gücü bakımından karşılaştırılması yapılmıştır. Simülasyon çalışmasında, farklı sayıda örnek çapı ve farklı heterojenlik durumları dikkate alınmıştır. Sonuçlara bakıldığında; BP, G, P, W ve HG testlerinin deneysel I.tip hata oranlarının 0,05'e yakın çıktığı, güç değerleri incelendiğinde ise SSK ve G testleri, heterojenlik arttıkça bu testlerin yanı sıra BP testinin diğer testlere göre daha yüksek güç değerine sahip olduğu gözlenmiştir.

Bilim Kodu : 20513
Anahtar Kelimeler : Değişen varyans, regresyon modeli, simülasyon, deneysel I. Tip hata, testin gücü
Sayfa Adedi : 71
Danışman : Doç. Dr. Esra GÖKPİNAR

INVESTIGATION OF THE PERFORMANCE OF TESTS USED FOR
HETEROSCEDASTICITY CONDITION IN THE SIMPLE LINEAR REGRESSION
MODEL

(M. Sc. Thesis)

Betül POLATDEMİR

GAZİ UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

January 2024

ABSTRACT

One of the assumptions of the simple linear regression model is that the error term has a constant variance. If this assumption is not met, the predictive power and reliability of the model may be negatively affected. Therefore, it should be checked whether there is heteroscedasticity. Although graphical methods are used to control heteroscedasticity, many tests have also been developed. However, there are very limited studies in the literature about which test is better in which situation. In this thesis study, Breusch-Pagan (BP), Spearman Rank Correlation (SSK), Glejser (G), Golfeld–Quandt (GQ), Park (P), White (W), Harvey-Godfrey (HG), Modified Levene (MLM) according to the median, Modified Levene (MLA) tests according to the mean were discussed and the tests were compared with the simulation method in terms of experimental type 1 error rates and the power of the test. In the simulation study, different numbers of sample sizes and different heterogeneity situations were taken into account. Looking at the results, it was observed that the estimated type I error rates of the BP, G, P, W and HG tests were close to nominal level 0,05. When the power values were analysed, it was observed that SSK and G tests, as the heterogeneity increased, these tests as well as BP test had higher power values than the other tests.

Science Code : 20513
Key Words : Heteroscedasticity, regression model, simulation, estimated I. Tip error, power of test
Page Number : 71
Supervisor : Assoc. Prof. Dr. Esra GÖKPINAR

TEŐEKKÜR

Bu alıőmanın gerekleőtirilmesinde, kıymetli bilgi ve birikimi ile bana yol gsteren deęerli danıőman hocam Do. Dr. Esra GKPINAR'a, eęitim hayatım boyunca zerimde emeęi olan deęerli hocalarıma, en byk motivasyon kaynaęım canım babama, sabırları iin aileme, maddi ve manevi desteęini esirgemeyen dostlarım Hatice ACAT ve Nur BAYRAM'a, ok uzaklardan akademik tecrbeleriyle her daim yanımda olan Halime TOSUN'a ve manevi olarak desteklerini grdęim tm arkadaőlarıma sonsuz teőekkrlerimi sunarım.



İÇİNDEKİLER

	Sayfa
ÖZET	iv
ABSTRACT.....	v
TEŞEKKÜR.....	vi
İÇİNDEKİLER	vii
ÇİZELGELERİN LİSTESİ.....	ix
ŞEKİLLERİN LİSTESİ.....	xi
SİMGELER VE KISALTMALAR.....	xiii
1. GİRİŞ.....	1
2. BASİT DOĞRUSAL REGRESYON MODELİ	7
2.1. Model Varsayımları	7
2.2. Regresyon Katsayılarının Tahmini	8
2.3. Hipotez Testleri	9
2.3.1. Model parametreleriyle ilgili hipotez testleri	9
2.3.2. Model anlamlılığı ile ilgili hipotez testleri.....	10
2.3.3. Belirleme katsayısı	12
3. DEĞİŞEN VARYANS DURUMU	13
3.1. Varyans Neden Değişir?	14
3.2. Değişen Varyansın Sonuçları	15
3.3. Değişen Varyans Durumunun Tespiti	15
3.3.1. Breusch-Pagan testi	17
3.3.2. Spearman sıra korelasyonu testi.....	18
3.3.3. Glejser testi.....	19
3.3.4. Golfeld-Quandt testi.....	19
3.3.5. Park testi.....	21

	Sayfa
3.3.6. White testi	22
3.3.7. Harvey–Godfrey testi	23
3.3.8. Modifiye edilmiş Levene testi.....	23
3.4. Uygulama.....	24
4. SİMÜLASYON ÇALIŞMASI.....	43
5. SONUÇ.....	65
KAYNAKLAR	67
ÖZGEÇMİŞ	71

ÇİZELGELERİN LİSTESİ

Çizelge	Sayfa
Çizelge 1.1. İstatistiksel yazılım paketlerinde bulunan değişen varyans testleri	4
Çizelge 2.1. ANOVA tablosu.....	11
Çizelge 3.1. 30 ailenin tüketim harcaması (Y) ve gelir verileri (X).....	24
Çizelge 3.2. Regresyon modeli istatistikleri.....	26
Çizelge 3.3. ANOVA tablo sonuçları.....	26
Çizelge 3.4. Regresyon katsayıları analizi	26
Çizelge 3.5. Artık terimlerin değerleri	27
Çizelge 3.6. Park testi hesaplamaları.....	29
Çizelge 3.7. Park testi regresyon modeli istatistikleri	30
Çizelge 3.8. Park testi ANOVA tablo sonuçları.....	30
Çizelge 3.9. Park testi için regresyon katsayıları analizi.....	31
Çizelge 3.10. Golfeld-Quandt testi grup 1 verileri	31
Çizelge 3.11. Golfeld-Quandt testi grup 2 verileri	32
Çizelge 3.12. Golfeld-Quandt testi ile grup 1 regresyon modeli istatistikleri.....	32
Çizelge 3.13. Golfeld-Quandt testi ile grup 1 ANOVA sonuçları	32
Çizelge 3.14. Golfeld-Quandt testi ile grup 1 regresyon katsayıları analizi	32
Çizelge 3.15. Golfeld-Quandt testi ile grup 2 regresyon modeli istatistikleri.....	33
Çizelge 3.16. Golfeld-Quandt ile testi grup 2 ANOVA sonuçları	33
Çizelge 3.17. Golfeld-Quandt testi ile grup 2 regresyon katsayıları analizi	33
Çizelge 3.18. Spearman Sıra Korelasyonu testi hesaplamaları	34
Çizelge 3.19. Glejser testi ile regresyon modeli istatistikleri.....	36
Çizelge 3.20. Glejser testi ile ANOVA sonuçları	36
Çizelge 3.21. Glejser testi ile regresyon katsayıları analizi.....	36
Çizelge 3.22. Breusch-Pagan testi ile regresyon modeli istatistikleri	37
Çizelge 3.23. Breusch-Pagan testi ile ANOVA sonuçları.....	37

Çizelge	Sayfa
Çizelge 3.24. Breusch-Pagan testi ile regresyon katsayıları analizi.....	37
Çizelge 3.25. White testi ile regresyon modeli istatistikleri.....	38
Çizelge 3.26. White testi ile regresyon katsayıları analizi	38
Çizelge 3.27. Harvey–Godfrey testi ile regresyon modeli istatistikleri	39
Çizelge 3.28. Harvey–Godfrey testi ile regresyon katsayıları analizi.....	39
Çizelge 4.1. X değişkeni Log-normal dağılımdan geldiğinde ve n=10 için testlerin deneysel I.tip hata oranları ve güç değerleri	45
Çizelge 4.2. X değişkeni Log-normal dağılımdan geldiğinde ve n=20 için testlerin deneysel I.tip hata oranları ve güç değerleri	47
Çizelge 4.3. X değişkeni Log-normal dağılımdan geldiğinde ve n=30 için testlerin deneysel I.tip hata oranları ve güç değerleri	49
Çizelge 4.4. X değişkeni Log-normal dağılımdan geldiğinde ve n=50 için testlerin deneysel I.tip hata oranları ve güç değerleri	51
Çizelge 4.5. X değişkeni tekdüze dağılımdan geldiğinde ve n=10 için testlerin deneysel I.tip hata oranları ve güç değerleri	53
Çizelge 4.6. X değişkeni tekdüze dağılımdan geldiğinde ve n=20 için testlerin deneysel I.tip hata oranları ve güç değerleri	55
Çizelge 4.7. X değişkeni tekdüze dağılımdan geldiğinde ve n=30 için testlerin deneysel I.tip hata oranları ve güç değerleri	57
Çizelge 4.8. X değişkeni tekdüze dağılımdan geldiğinde ve n=50 için testlerin deneysel I.tip hata oranları ve güç değerleri	59
Çizelge 4.9. Hata terimleri t3 dağılımına sahip olduğunda testlerin deneysel 1.tip hata oranları.....	61
Çizelge 4.10. Hata terimleri t5 dağılımına sahip olduğunda testlerin deneysel 1.tip hata oranları.....	62

ŞEKİLLERİN LİSTESİ

Şekil	Sayfa
Şekil 3.1. Sabit varyans durumu.....	13
Şekil 3.2. Değişen varyans durumu.....	13
Şekil 3.3. Artık değerlerin Y^* 'e göre örüntüleri	16
Şekil 3.4. Y 'nin X 'e bağlı serpm diyagramı.....	26
Şekil 3.5. Y_i 'nin artık değerlerine (e_i) bağlı serpm diyagramı.....	28
Şekil 3.6. X_i 'nin artık değerlerine (e_i) bağlı serpm diyagramı	28
Şekil 4.1. X değişkeni Log-normal dağılımdan geldiğinde ve $n=10$ için testlerin toplamsal modelde deneysel I.tip hata oranları ve güç değerleri.....	44
Şekil 4.2. X değişkeni Log-normal dağılımdan geldiğinde ve $n=10$ için testlerin çarpımsal modelde deneysel I.tip hata oranları ve güç değerleri.....	44
Şekil 4.3. X değişkeni Log-normal dağılımdan geldiğinde ve $n=20$ için testlerin toplamsal modelde deneysel I.tip hata oranları ve güç değerleri.....	46
Şekil 4.4. X değişkeni Log-normal dağılımdan geldiğinde ve $n=20$ için testlerin çarpımsal modelde deneysel I.tip hata oranları ve güç değerleri.....	46
Şekil 4.5. X değişkeni Log-normal dağılımdan geldiğinde ve $n=30$ için testlerin toplamsal modelde deneysel I.tip hata oranları ve güç değerleri.....	48
Şekil 4.6. X değişkeni Log-normal dağılımdan geldiğinde ve $n=30$ için testlerin çarpımsal modelde deneysel I.tip hata oranları ve güç değerleri.....	48
Şekil 4.7. X değişkeni Log-normal dağılımdan geldiğinde ve $n=50$ için testlerin toplamsal modelde deneysel I.tip hata oranları ve güç değerleri.....	50
Şekil 4.8. X değişkeni Log-normal dağılımdan geldiğinde ve $n=50$ için testlerin çarpımsal modelde deneysel I.tip hata oranları ve güç değerleri.....	50
Şekil 4.9. X değişkeni tekdüze dağılımdan geldiğinde ve $n=10$ için testlerin çarpımsal modelde deneysel I.tip hata oranları ve güç değerleri.....	52
Şekil 4.10. X değişkeni tekdüze dağılımdan geldiğinde ve $n=10$ için testlerin toplamsal modelde deneysel I.tip hata oranları ve güç değerleri.....	52
Şekil 4.11. X değişkeni tekdüze dağılımdan geldiğinde ve $n=20$ için testlerin toplamsal modelde deneysel I.tip hata oranları ve güç değerleri.....	54
Şekil 4.12. X değişkeni tekdüze dağılımdan geldiğinde ve $n=20$ için testlerin çarpımsal modelde deneysel I.tip hata oranları ve güç değerleri.....	54

Şekil	Sayfa
Şekil 4.13. X değişkeni tekdüze dağılımdan geldiğinde ve $n=30$ için testlerin çarpımsal modelde deneysel I.tip hata oranları ve güç değerleri.....	56
Şekil 4.14. X değişkeni tekdüze dağılımdan geldiğinde ve $n=30$ için testlerin toplamsal modelde deneysel I.tip hata oranları ve güç değerleri.....	56
Şekil 4.15. X değişkeni tekdüze dağılımdan geldiğinde ve $n=50$ için testlerin toplamsal modelde deneysel I.tip hata oranları ve güç değerleri.....	58
Şekil 4.16. X değişkeni tekdüze dağılımdan geldiğinde ve $n=50$ için testlerin çarpımsal modelde deneysel I.tip hata oranları ve güç değerleri.....	58
Şekil 4.17. Hata terimleri t_3 dağılımına sahip olduğunda testlerin deneysel 1.tip hata oranları	60
Şekil 4.18. Hata terimleri t_5 dağılımına sahip olduğunda testlerin deneysel 1.tip hata oranları	62

SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Simgeler

Açıklamalar

\hat{Y}_i	i. Bağımlı değişkenin tahmini
σ^2	Yığın varyansı
X_i	i. Bağımsız değişken
H_0	Yokluk hipotezi
H_1	Alternatif hipotez
R^2	Belirleme katayısı
Y_i	i. Bağımlı değişken
e_i	Artık terimi
β_0	Regresyon katsayısı (kesim noktası)
β_1	Regrasyon katsayısı (eğim)
ε_i	Hata terimi
α	Anlamlılık düzeyi
m	Parametre sayısı
n	Örnek çapı

Kısaltmalar

Açıklamalar

AKO	Artık kareler ortalaması
AKT	Artık kareler toplamı
ANOVA	Varyans Analizi
BP	Breusch-Pagan testi
EÇO	En çok olabilirlik
EKK	En küçük kareler
G	Glejser testi
GQ	Golfeld-Quandt testi
HG	Harvey-Godfrey testi

Kısaltmalar**Açıklamalar**

LN	Lognormal dağılım
MLA	ortalamaya göre Modifiye Edilmiş Levene testi
MLM	medyan değerine göre Modifiye Edilmiş Levene testi
N	Normal Dağılım
P	Park testi
RKO	Regresyon kareler ortalaması
RKT	Regresyon kareler toplamı
SSK	Spearman Sıra Korelasyonu testi
TKT	Toplam kareler toplamı
U	Tekdüze dağılım
W	White testi
V	Varyans

1. GİRİŞ

Günümüzde istatistiksel yöntemler ekonomi, sağlık, sosyal bilimler ve mühendislik gibi çeşitli alanlarda kullanılmaktadır. İstatistiksel yöntemlerden biri olan regresyon analizinde bağımlı değişken ile bağımsız değişken arasındaki ilişkiyi analiz edip tahmin etmek, bağımsız değişkenin bağımlı değişkeni ne kadar etkilediği, bağımlı değişkenin bağımsız değişken tarafından ne kadar açıklandığının belirlenmesi amaçlanmaktadır. Bağımlı değişken tahmin etmeye çalıştığımız, bağımsız değişken ise tahmini etkileyecek olan değişkendir. Örneğin bir bitkinin büyüme hızı bağımlı değişken ise gübre miktarı bağımsız değişkendir.

Regresyon modelleri, değişkenler arasındaki karmaşık ilişkileri çözerek değişkenlerin etkilerini anlamamıza ve geleceği tahmin etmemize yardımcı olur. Basit doğrusal regresyon modeli ise bir bağımsız değişkenin bir bağımlı değişken üzerindeki etkisini ölçmek için kullanılan en temel modeldir. Regresyon modelinin belirli varsayımları vardır. Bu varsayımlar modelin doğru ve güvenilir sonuçlar üretmesi için önemlidir. Bu varsayımlarından biri, hata teriminin sabit bir varyansa sahip olmasıdır. Bu varsayım karşılanmadığında, modelin tahmin gücünü ve güvenilirliği olumsuz etkilenebilir. Bu yüzden değişen varyansın olup olmadığının kontrol edilmesi ve buna göre hareket edilmesi önemlidir. Değişen varyansın olup olmadığını belirlemek için birçok test geliştirilmiştir. Fakat literatürde hangi durumda hangi testin daha iyi olduğunu ortaya koyan çalışmalar oldukça kısıtlıdır.

Goldfeld ve Quandt (1965) yaptıkları çalışmada parametrik ve parametrik olmayan iki test önermişlerdir. Literatürde yaygın olarak kullanılan Goldfeld-Quandt testi gözlemleri kriterlere göre iki gruba ayırarak bu grupların arasındaki varyansların farklı olup olmadığını ele almıştır.

Glejser (1969) regresyon analizinde değişen varyans durumunun olup olmadığını belirlemek amacıyla bir test önermiştir. Bu test, regresyon modelinden elde edilen artık terimlerinin mutlak değerlerini ile yeni bir yardımcı regresyon modeli oluşturarak elde edilen katsayının istatistik bakımdan anlamlılığını test etmiştir.

Breusch ve Pagan (1979) regresyon analizinde deęişen varyans durumunun olup olmadığını belirlemek amacıyla bir test önermiştir. Bu test, regresyon modelinin artık terimlerinin varyansının bağımsız deęişkenlerle ilişkili olup olmadığını deęerlendirir. Eęer hata terimlerinin varyansı bağımsız deęişkenlere baęlı olarak deęişiyorsa, bu durumun deęişen varyansa işaret edebileceğini söylemiştir.

White (1980) regresyon analizinde deęişen varyans durumunun olup olmadığını belirlemek amacıyla bir test önermiştir. Bu test, regresyon modelinin artık terimlerinin varyansının bağımsız deęişkenlere baęlı olarak deęişip deęişmediğini deęerlendirmiştir.

Harvey (1976) yaptığı çalışmada hata terimlerinin belirli bir tür deęişen varyans sergiledięi bir regresyon modeli dikkate alıp maksimum olabilirlik tahmin yöntemleri geliştirmiş ve iki aşamalı tahmin prosedürüyle karşılaştırmıştır. Deęişen varyans için bir olabilirlik oranı testi önermiştir.

Harvey ve Phillips (1974) yaptıkları çalışmada genel doğrusal modelde deęişen varyansa karşı tam bir parametrik test önermiştir. Önerilen test, kolayca hesaplanan özyinelemeli artıklara dayanmaktadır ve yeni testin esneklięinin ve hesaplama basitlięinin onu pratik bir prosedür olarak çekici kıldığını göstermişlerdir.

Ramsey (1969), Szroeter (1978), Harrison ve McCabe (1979), Evans ve King (1985) tarafından farklı testler önerilmiştir.

Ali ve Giaccotto (1984) yaptıkları çalışmada deęişen varyans için optimum parametrik olmayan testler önererek literatürde güç ve sağlamlık özellikleri açısından tanıtılan testlerle birlikte incelemişlerdir.

Griffiths ve Surekha (1986) yaptıkları çalışmada Szroeter (1978)'in asimptotik testini Goldfeld – Quandt (GQ) testi, Breusch-Pagan (1978) tarafından önerilen Lagrange çarpanı (BP) ve Ramsey (BAMSET) testleri ile karşılaştırmıştır. Monte Carlo deneyini $(\beta_0, \beta_1)=(10,1)$ olacak şekilde ayarlamıştır. İki farklı örneklem büyüklüğü (20,50) ile bağımsız deęişken X'ler için iki farklı dağılımdan (tekdüze ve log-normal) örnek seçmiştir. Bu testlerin güçlerini bulmak için çarpımsal ve toplamsal varyans heterojenlięini arttırarak %5 anlamlılık düzeyinde 5000 tekrarla reddedilme oranlarını hesaplayarak tahmin

etmişlerdir. Gözlemleri artan varyanslara göre sıralamak mümkün olduğunda, Szroeter'in testinin diğer üçünden daha güçlü olduğunu, gözlemler sıralanmadığında Szroeter'in testi, BAMSET ve GQ testinin performanslarının düştüğünü ve BP testinin gözlemlerin sıralı olup olmamasına bağlı olmadığını göstermişlerdir.

Pagan ve Pak (1993) yaptıkları çalışmada değişen varyans için yapılan çalışmaların, yöntemlerin bir incelemesini sunmayı amaçlamış ve testlerin performansını anlamak için yapılması gereken çok fazla çalışma olduğunu göstermişlerdir. Ortaya çıkan sorunlardan bazıları sonlu örnek performansı ile, diğeri ise bunları oluştururken yapılan yardımcı varsayımlardan sapmalara karşı dayanıklı olacak şekilde tasarlanmış testlerin zayıf performansı ile ilgilidir. Simülasyon çalışmalarının kısıtlı olduğunu ve tüm testlerin örnekleme göre davranışının kapsamlı bir şekilde çalışmasını önermişlerdir.

Lyon ve Tsai (1996) yaptıkları çalışmada literatürde var olan değişen varyans testlerini farklı dağılımlardan seçilen X 'ler ile testlerin güç performanslarını incelemişlerdir.

Yüce (2004) yaptığı çalışmada doğrusal ve doğrusal olmayan regresyon modelleri için değişen varyans durumunu incelemiş ve değişen varyansın tespiti yapıldığında düzeltici yöntemleri iki ayrı veri üzerinden incelemiştir.

Carapeto ve Holt (2003) regresyon modellerinde değişen varyans için parametrik ve parametrik olmayan, doğrusal ve doğrusal olmayan regresyon modellerine uygulanabilen Goldfeld-Quandt testine dayalı yeni bir test sunmuştur. Testin gücünü 5000 tekrar ve $n=50$ olarak yaygın olarak kullanılan değişen varyans testleriyle karşılaştırmış hem toplamsal hem çarpımsal değişen varyansı tespit etmede güçlü olduğunu göstermişlerdir.

Çelik (2012) yaptığı çalışmada kelebek dağılan artıklar için değişen varyanslılığın dengelenmesinde kullanılan yöntemlerin sınanmasını ve alternatif yöntemleri araştırmış önerilen yöntemlerin geçerliliğini hem uygulama verisinde hem de farklı örneklem büyüklüğündeki yapay veriler ile incelemiştir.

Cook ve Weisberg (1983) yaptıkları çalışmada değişen varyans durumunun teşhisi için grafiksel ve grafiksel olmayan yöntemleri derleyip tanıtmış ve bu yöntemlerin nasıl iyileştirilebileceği üzerinde çalışmışlardır.

Gülyüz (1992) yaptığı çalışmada değişen varyans durumunun belirlenmesini ve sonuçlarını incelemiştir. Literatürde var olan testleri tesadüfi olarak seçilen 60 ailenin geliri ve zorunlu harcamaları arasında regresyon modelini kurup test sonuçlarını değerlendirmiştir. Yaptığı uygulama sonucunda Bartlett testi sabit varyansın olduğu sonucunu vermiş, diğer testler ise değişen varyansın olduğunu göstermiştir. Gülyüz (1992), bu çalışmasına göre Bartlett testinin güvenilir bir test olmadığını ifade etmiştir.

Çamurlu ve Erilli (2018) yaptıkları çalışmada Kantil Regresyon yöntemi ile değişen varyansın belirlenmesine imkân sağlandığından bahsetmiştir.

Uyanto (2019) yaptığı çalışmada, Breusch-Pagan, Glejser, Goldfeld-Quandt, Harvey-Godfrey, Harrison-McCabe, Park, White testlerini farklı örneklem büyüklüğü (15, 30, 60, 90, 120, 150) ve farklı heterojenlik durumları altında deneysel 1.tip hata oranları ve güçleri bakımından incelemiştir.

Astivia ve Zumbo (2019) yaptıkları çalışmada psikoloji ve sosyal bilimler alanında kullanılan regresyon analizlerinde değişen varyans tespiti ve düzeltilmesi için R ve SPSS programlarının adım adım nasıl kullanıldığını göstermişlerdir.

Kullanılan bazı değişen varyans testlerinin, EViews, IBM SPSS, MATLAB, Minitab, R, SAS, SHAZAM, S-PLUS ve Stata gibi istatistiksel yazılımlarda kullanılıp kullanılmadığı Çizelge 1.1'de gösterilmiştir (Uyanto, 2019). Bu çizelgeye göre, değişen varyansın olup olmadığını test edebilmek için günümüzde en yaygın kullanılan SPSS programında bile bu testlerden hiçbirine rastlanılmadığı gözlenmiştir.

Çizelge 1.1. İstatistiksel yazılım paketlerinde bulunan değişen varyans testleri

İstatistiksel Yazılım	Breusch-Pagan	Glejser	Goldfeld-Quandt	Harvey-Godfrey	Park	White
Eviews	√	√	-	√	-	√
IBM SPSS	-	-	-	-	-	-
MATLAB	-	-	-	-	-	-
Minitab	-	-	-	-	-	-
R	√	-	√	√	-	√
SAS	√	-	-	-	-	√
SHAZAM	√	√	-	√	-	√
S-PLUS	√	√	√	-	-	√
Stata	√	-	√	-	-	√

Literatürü incelediğimizde regresyon modelinde, değişen varyansın olup olmadığını belirlemek için birçok test geliştirilmiştir. Fakat literatürde hangi durumda hangi testin daha iyi olduğunu ortaya koyan çalışmalar oldukça kısıtlıdır. Ayrıca yapılan simülasyon çalışmalarının da oldukça kısıtlı durumlar altında çalıştığı görülmektedir. Bu amaçla, bu tez çalışmasında literatürde en çok kullanılan testler, küçük, orta ve büyük olmak üzere farklı örnek çapları altında, farklı heterojenlik durumlarında, bağımsız değişkenin farklı dağılımları altında performansları incelenerek karşılaştırılmıştır. Ayrıca bu testlerin farklı dağılımlar altında nasıl performans göstereceğini görebilmek için, hata terimlerinin normal dağıldığı durumun yanında, normal dağılmadığı durumlar altında da incelenerek karşılaştırılmıştır. Bu bağlamda regresyon modelinde değişen varyansın tespiti ile ilgili literatüre önemli bir katkı sağlanacağı düşünülmektedir.

Buna göre tez çalışması aşağıdaki gibi oluşturulmuştur. 2. Bölümde Basit doğrusal regresyon modeli ve varsayımları tanıtılmıştır. 3. Bölümde değişen varyans, değişen varyansın nedenleri ve sonuçları ve değişen varyans için literatürde var olan testler tanıtılmış olup gerçek veriler üzerinden uygulama yapılmıştır. 4. Bölümde literatürde var olan testlerin simülasyon yoluyla karşılaştırılması yapılmıştır.



2. BASİT DOĞRUSAL REGRESYON MODELİ

Regresyon analizi değişkenler arasındaki matematiksel ilişkiyi modellemek ve incelemek amacıyla kullanılan bir istatistiksel yöntemdir. Regresyonda değişkenler arasındaki bağıntının biçimi için model ifadesi kullanılır. Regresyon analizinde değişkenler bağımlı ve bağımsız değişken olarak adlandırılır. Bağımlı değişken bağımsız değişkene bağlı olarak değişen bir değişkendir. Bağımsız değişken ise bağımlı değişkeni etkileyen, değiştiren değişken olarak tanımlanmaktadır. Regresyon analizinde bağımlı değişken Y, bağımsız değişken X ile gösterilir. Örneğin; öğrencilerin ders notu Y ile ders çalışma saatleri ise X ile gösterilir.

Basit doğrusal regresyon modeli;

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad , \quad i=1, \dots, n \quad (2.1)$$

Bu modelde β_1 regresyon doğrusunun eğimi, β_0 regresyon doğrusunun Y ekseninin kesim noktasını, ε_i da hata terimini ifade eder (Gamgam ve Altunkaynak, 2015: 11).

2.1. Model Varsayımları

1. Hata terimleri ortalaması 0 ve varyansı σ^2 olan bir rastgele değişkendir. Yani,

$$\varepsilon_i \sim (0, \sigma^2) \text{ dir.}$$

2. Hata terimleri birbirleriyle ilişkisizdir. Bu varsayım hata terimleri arasında otokorelasyon olmaması veya $Kov(\varepsilon_i, \varepsilon_j) = 0$, $i \neq j=1, \dots, n$ şeklinde ifade edilir.

3. Hata terimlerinin varyansları aynıdır. $V(\varepsilon_i) = \sigma^2$ olmalıdır.

4. Hata terimlerinin dağılımları normaldir. Bu varsayım $\varepsilon_i \sim N(0, \sigma^2)$ biçiminde ifade edilebilir. EKK yöntemi ile parametre tahmini için bu varsayıma gerek yoktur. Bu varsayım istatistiksel çıkarımlar (hipotez testleri ve güven aralıkları) için gereklidir.

5. $Kov(\varepsilon_i, X_i) = 0$ olmalıdır. Bu varsayım X'in rastgele olduğu tasarımlar için gereklidir (Gamgam ve Altunkaynak, 2015: 25).

2.2. Regresyon Katsayılarının Tahmini

Basit doğrusal regresyon modellerinin parametrelerinin tahmini için kullanılan en yaygın yöntemler En Küçük Kareler (EKK) Yöntemi ve En Çok Olabilirlik (EÇO) Yöntemidir.

Basit doğrusal regresyon modelinin örnek verisi için modeli aşağıdaki gibidir.

$$Y_i = b_0 + b_1X_i + e_i \quad (2.2)$$

veya

$$\hat{Y}_i = b_0 + b_1X_i \quad (2.3)$$

EKK yöntemi artık terim (e_i) ile ilgili iki koşulun sağlanmasına dayanır. Buna göre $\hat{Y}_i = b_0 + b_1X_i$ doğrusu Eş. 2.4 ve Eş. 2.5'de verildiği gibi artıklar toplamını sıfır ve artıkların kareler toplamını en küçük yapmalıdır.

$$\sum_{i=1}^n e_i = 0. \quad (2.4)$$

ve

$$\sum_{i=1}^n e_i^2 = \min. \quad (2.5)$$

e_i yerine $Y_i - \hat{Y}_i$ konulursa ve gerekli işlemler yapıldıktan sonra b_0 ve b_1 için aşağıdaki eşitlikler elde edilir;

$$b_0 = \bar{Y} - b_1\bar{X} \quad (2.6)$$

$$b_1 = \frac{\sum X_i Y_i - n\bar{X}\bar{Y}}{\sum X_i^2 - n\bar{X}^2} \quad (2.7)$$

b_1 katsayısı regresyon doğrusunun eğimi olup bağımsız değişkendeki 1 birim artışın bağımlı değişkende yapacağı değişikliğin yönünü ve miktarını verir.

b_0 katsayısı ise regresyon doğrusunun Y eksenini kestiği noktadır. Yani bağımsız değişken değeri 0 olan birimin bağımlı değişken bakımından alması beklenen değeri ifade eder.

Hata terimlerinin dağılımları normal yani $\varepsilon_i \sim N(0, \sigma^2)$ ise model parametrelerinin tahmini için En Çok Olabilirlik (EÇO) Yöntemi de kullanılabilir.

Hata terimleri normal dağılıma sahipse olabilirlik fonksiyonu ile gerekli işlemler yapıldıktan sonra aşağıdaki eşitlikler elde edilir;

$$b_0 = \bar{Y} - b_1 \bar{X} \quad (2.8)$$

$$b_1 = \frac{\sum X_i Y_i - n \bar{X} \bar{Y}}{\sum X_i^2 - n \bar{X}^2} \quad (2.9)$$

Bu sonuçlardan anlaşılacağı üzere b_0 ve b_1 parametrelerinin EKK ve en çok olabilirlik tahmin edicileri aynıdır. Bu tahmin ediciler sapmasız ve etkin tahmin edicilerdir (Gamgam ve Altunkaynak, 2015: 18, 20).

Bu çalışmada, bu varsayımlardan varyans homojenliği varsayımı sağlanmadığı durum altında değişen varyans durumu incelenecektir.

2.3. Hipotez Testleri

Bu bölümde basit doğrusal regresyon modeli için hipotez testleri, model parametreleri ve model anlamlılığı için test işlemleri verilmiştir.

2.3.1. Model parametreleriyle ilgili hipotez testleri

Basit doğrusal regresyon modelinde β_0 ve β_1 parametreleriyle ilgili hipotez testleri bu parametrelerin tahmin edicilerinin beklenen değeri, varyansı ve dağılımına bağlıdır. Buna göre n gözlem sayısı ve m modeldeki parametre sayısı olmak üzere,

EKK varsayımı nedeniyle $\varepsilon_i \sim N(0, \sigma^2)$ olduğundan $Y_i \sim N(\beta_0 + \beta_1 X_i, \sigma^2)$ dir.

β_1 katsayısının tahmin edicisinin dağılımı $b_1 \sim N(\beta_1, V(b_1))$ şeklindedir. Burada,

$$V(b_1) = \frac{\sigma^2}{\sum (X_i - \bar{X})^2} \text{ olarak tanımlanır.}$$

Ayrıca β_0 katsayısının tahmin edicisinin dağılımı ise $b_0 \sim N(\beta_0, V(b_0))$ şeklindedir.

$$\text{Burada } V(b_0) = s^2 \left(\frac{\sum(X_i)^2}{n \sum(X_i - \bar{X})^2} \right)$$

$$s^2 = \frac{AKT}{n-m} = \frac{\sum(Y_i - \hat{Y}_i)^2}{n-m} \text{ olarak tanımlanır.}$$

β_0 ile ilgili yokluk ve alternatif hipotez

$$H_0: \beta_0 = 0 \quad H_1: \beta_0 \neq 0$$

olmak üzere, test istatistiği

$$t_h = \frac{b_0}{\sqrt{V(b_0)}}$$

olacaktır. Burada t_h , $n-m$ serbestlik dereceli t dağılımına sahip olup $t_h > t_{n,m;\alpha/2}$

ise H_0 reddedilir ve β_0 parametresi anlamlıdır.

β_1 ile ilgili yokluk ve alternatif hipotez

$$H_0: \beta_1 = 0 \quad H_1: \beta_1 \neq 0$$

olmak üzere, test istatistiği

$$t_h = \frac{b_1}{\sqrt{V(b_1)}}$$

olacaktır. Burada t_h , $n-m$ serbestlik dereceli t dağılımına sahip olup $t_h > t_{n,m;\alpha/2}$

ise H_0 reddedilir ve β_1 parametresi anlamlıdır (Gamgam ve Altunkaynak, 2015: 33, 35).

2.3.2. Model anlamlılığı ile ilgili hipotez testleri

Regresyon modelinin anlamlılığı için yokluk ve alternatif hipotez aşağıdaki gibi kurulur.

$H_0: \beta_1 = 0$ (Kurulan model anlamsızdır)

$H_1: \beta_1 \neq 0$ (Kurulan model anlamlıdır)

Modelin anlamlılık testi için varyans analizi yaklaşımı da kullanılabilir. Varyans analizi bağımlı değişkendeki toplam değişimin parçalanmasına dayanır. Buna göre,

$$(Y_i - \bar{Y}) = (\hat{Y}_i - \bar{Y}) + (Y_i - \hat{Y}_i)$$

yazılabilir. Yukarıdaki ifadenin her iki tarafının karesi alınıp n gözlem için toplanacak olursa aşağıdaki ifade elde edilir.

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

Bu ifade,

$$TKT=RKT+AKT$$

şeklinde yazılabilir. Burada TKT= Toplam varyans, RKT= Regresyon modeli tarafından açıklanan varyans ve AKT= Açıklanmayan varyansı gösterir. Modelin anlamlılığının testi için kurulan varyans analiz tablosu aşağıdaki gibidir.

Çizelge 2.1. ANOVA tablosu

<i>Değişim kaynağı</i>	<i>Serbestlik derecesi</i>	<i>Kareler Toplamları</i>	<i>Kareler Ortalaması</i>	<i>F test istatistiği</i>
Regresyon	m-1	RKT	RKO	F=RKO/AKO
Artık	n-m	AKT	AKO	
Toplam	n-1	TKT		

ANOVA tablosundan bulunan F istatistiği $F_{m-1, n-m}$ dağılımına sahip olup $F_h > F_{m-1, n-m, \alpha}$ ise H_0 hipotezi reddedilir ve model anlamlıdır. Basit doğrusal regresyon analizinde $m=2$ olup $F_{m-1, n-m, \alpha}$ istatistiği, β_1 parametresinin hipotez testinde bulunan $(t_{n, m; \alpha/2})^2$ ile aynıdır (Aydın, 2014: 35).

2.3.3. Belirleme katsayısı

Modelin genel yeterliliğini deęerlendirmek için belirleme katsayısı dikkate alınabilir. Belirleme katsayısı yani R^2 bağımsız deęişkenin bağımlı deęişkeni açıklama miktarını verir ve aşığıdaki gibi tanımlanır.

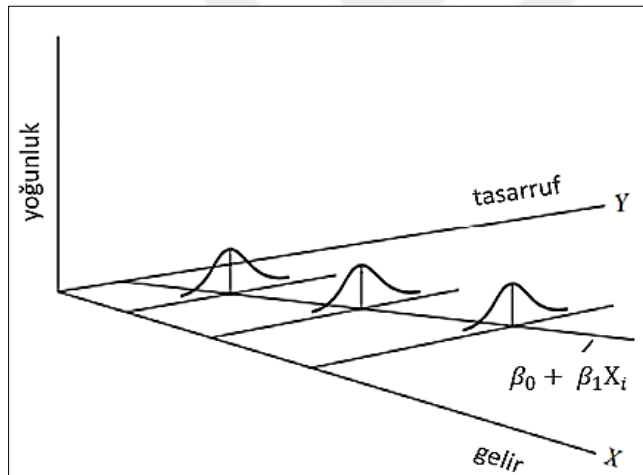
$$R^2 = \frac{RKT}{TKT} \text{ veya } R^2 = 1 - \frac{AKT}{TKT}$$

$0 \leq R^2 \leq 1$ olup, eęer R^2 deęeri 1'e yakınsa kurulan model yada açıklayıcı deęişkenler Y'deki deęişimin büyük kısmını açıklar. Eęer R^2 deęeri 0'a yakınsa kurulan modelin yada açıklayıcı deęişkenlerin Y'deki deęişimi açıklamadığını yani modelin veriye uyum göstermedięi anlaşılır (Gamgam ve Altunkaynak, 2015: 23).

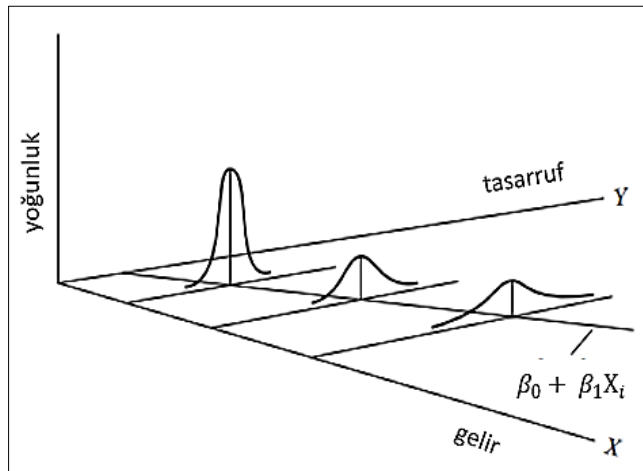
3. DEĞİŞEN VARYANS DURUMU

Klasik doğrusal regresyon modelinin varsayımlarından biri hata terimlerinin varyanslarının aynı ve sabit olduğudur. Değişen varyans durumu ise hata terimlerinin varyanslarının birbirinden farklı olduğu durumda ortaya çıkar. Hata terimlerinin her X_i değerine karşılık aldığı değerlerin ortalamasının sıfır olduğu varsayımı kabul edilmektedir. Ayrıca hata teriminin dağılımı her X_i için aynıdır ve varyansları sabit kabul edilmektedir. Buna karşılık X arttıkça Y_i 'nin varyansı artıyor ise Y_i 'nin aynı varyansa sahip olmadığı ve değişen varyansın var olduğu söylenebilir (Montgomery, Peck ve Vining, 2013: 121).

Aşağıdaki şekillerden Şekil 3.1 hata teriminin varyansın sabit olduğunu, Şekil 3.2 ise hata teriminin değişen varyansa sahip olduğunu göstermektedir.



Şekil 3.1. Sabit varyans durumu



Şekil 3.2. Değişen varyans durumu

Hata teriminin deęişen varyansa sahip olduğunu aynı zamanda aőağıdaki gibi göstermek mümkündür.

$$E(\varepsilon_i^2) = \sigma_i^2 \quad (3.1)$$

Sabit varyansla deęişen varyans arasındaki farkı açıkça belirtmek için örneęin, tasarruf ve gelir deęişkenleri arasında bir regresyon modeli kurulduęunu farz edelim. Bu durumda Y tasarruf, X de geliri göstermek üzere $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$, şeklinde model kurulur. Şekil 3.1 ve Şekil 3.2 den de görüldüğü gibi gelir arttıkça tasarrufların da ortalama olarak arttığı görülmektedir. Ancak Şekil 3.1 de tasarrufların varyansı, gelirin bütün düzeylerinde aynı kalırken, Şekil 3.2 de gelire birlikte artar. Şekil 3.2’de yüksek gelirli ailelerin ortalama olarak, düşük gelirliilerden daha çok tasarruf ettikleri ama tasarruflarının deęişkenliğinin daha fazla olduęu anlaşılmaktadır (Gujarati, 2012: 366).

3.1. Varyans Neden Deęişir?

1. Hatasını öğrenen modellerde, tıpkı insanların öğrenmesi gibi, davranış hataları zamanla azalır ya da hata sayıları daha tutarlı hale gelir. Bu durumda σ_i^2 'nin küçülmesi beklenir. Örneęin belli bir zaman diliminde yapılan yazım hatalarıyla yazma saatlerini ilişkilendiren bir örnek ele alalım. Yazma saatlerinin sayısı arttıkça hem yazım hataları hem de bunların varyansı azalmaktadır.
2. Deęişen varyans, aykırı deęerlerin bir sonucu olarak da ortaya çıkabilir. Böyle durumlarda sabit varyans varsayımını korumak güçleşecektir. Böyle bir gözlemin atılması modelin analiz sonuçlarını önemli ölçüde deęiştirebilir.
3. Klasik doğrusal regresyon modelinin varsayımlarından biri olan doğru model kurma varsayımının çiğnenmesidir. Model kurma hataları, bazı önemli deęişkenlerin modelden dışlanması, deęişen varyansın oluşmasına sebep olabilir. Bir malın talep fonksiyonunda, tamamlayıcı ya da rakip malların fiyatlarını modele koymazsak regresyondan elde edilen artıklar açıkça hata varyansının sabit olmayacağını gösterebilir.
4. Modele eklenen deęişkenlerin birinde veya birkaçındaki çarpıklık deęişen varyansa sebep olabilir. Gelir, servet, eğitim gibi iktisadi deęişkenler bunun örnekleridir.
5. Deęişen varyans, yanlış veri dönüştürmesinden ve yanlış fonksiyon kalıbından kaynaklanabilir. Deęişen varyans sorununa zaman serilerinden çok, kesit verilerinde rastlanabilir (Gujarati, 2012: 366, 378).

3.2. Değişen Varyansın Sonuçları

Değişen varyansın sonuçları aşağıdaki gibi özetlenebilir:

1. Değişen varyans, EKK tahmin edicilerinin yansızlık ve tutarlılık özelliklerini bozmaz. Ama bu tahmin ediciler artık en küçük varyanslı ya da etkin olmaktan çıkar.
2. Değişen varyans varken, EKK tahmin edicilerinin varyansları, bilindik EKK formülleriyle bulunamaz. Ancak alışıldık EKK formülleri kullanılırsa, bunlara dayanan t ile F testleri yanıltıcı sonuçlar verir.

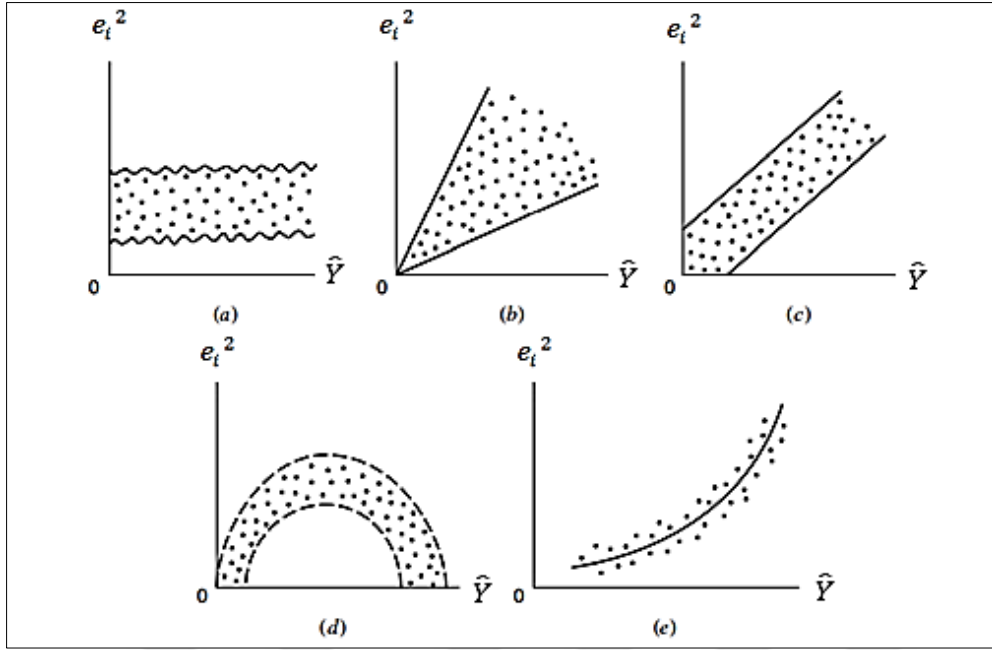
Değişen varyansın olup olmadığını anlamak için kesin kurallar yoktur. Ancak seçilmiş X 'lere karşılık gelen bütün Y 'lerin yığını biliyorsak σ_i^2 bilinebilir. Araştırmalarda belli bir X değerine karşılık tek bir Y gözlem değeri bulunur. Tek bir Y gözleminden de σ_i^2 'yi bilmek olanaksızdır (Gujarati, 2012: 376).

3.3. Değişen Varyans Durumunun Tespiti

Değişen varyansın var olup olmadığını anlamak için biçimsel ve biçimsel olmayan yöntemler geliştirilmiştir. Bu yöntemlerin çoğu gözlemlenebilen artık terimlerinin incelenmesine dayanır.

Biçimsel olmayan yöntemler

Değişen varyansın var olup olmadığını anlamak için kullanılan biçimsel olmayan yöntemler grafik yöntemleridir. Bu yöntemler en sık kullanılan yöntemlerden biridir. Bu yöntemde, değişen varyansın niteliğine dair bir bilgi yoksa uygulamada regresyon çözümlemesi sabit varyans varsayımı altında yapılır. Daha sonra artık teriminin kareleri bulunarak (e_i^2), bu değerlerin bağımlı değişkeninin tahmin değerleri (\hat{Y}) ile düzenli bir yapı sergileyip sergilemediğine bakılır ve sergilenen yapıya göre değişen varyansın durumu hakkında bir çıkarım yapılır. Artık değerlerinin karesi ile \hat{Y} 'nın grafiksel olarak farklı bazı durumları Şekil 3.3'de verilmiştir.



Şekil 3.3. Artık değerlerin \hat{Y} 'e göre örüntüleri

Şekil 3.3'te e_i^2 ile \hat{Y}_i arasındaki ilişkiye baktığımızda;

(a) şeklinde iki değişken arasında düzenli örüntü görülmez, bu da elde edilen veride değişen varyans bulunmayabileceği izlenimi verir. (b), (c), (d), (e) şekilleri ise belirli örüntüler sergiler. (c) doğrusal bir ilişki öngörürken (d) ile (e) ikinci dereceden bir ilişkiyi gösterir.

Fakat biçimsel olmayan yani grafiksel yöntemlerde yukarıdaki şekillerde ifade edildiği gibi belirgin bir durum ortaya çıkmazsa araştırmacı için değişen varyansın olup olmadığının belirlenmesi oldukça zorlaşmaktadır. Bu durumda biçimsel yöntemlerin kullanılması değişen varyansın olup olmadığının belirlenmesinde araştırmacıya daha kolaylık sağlamaktadır (Gujarati, 2012: 377).

Biçimsel yöntemler

Biçimsel yöntemler, değişen varyansın olup olmadığının belirlenmesinde test tekniklerine dayalı olup bu bölümde Breusch-Pagan testi (BP), Spearman korelasyonu testi (SSK), Glejser testi (G), Goldfeld-Quandt testi (GQ), Park testi (P), Harvey-Godfrey testi (HG), White testi (W), meydan değerine göre Modifiye Edilmiş Levene testi (MLM) ve ortalamaya göre Modifiye Edilmiş Levene testi (MLA) tanıtılmıştır.

3.3.1. Breusch-Pagan testi

Breusch ve Pagan (1979), doğrusal bir regresyon modelinde değişen varyans için Lagrange çarpanı testini geliştirmiştir. Bu yöntemle, bir regresyon modelindeki hataların varyansının bağımsız değişkenlerin değerlerine bağlı olup olmadığını test eder.

Breusch-Pagan testini göstermek için basit doğrusal regresyon modelini ele alalım:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

Breusch ve Pagan (1979), hata teriminin varyansını,

$V(\varepsilon_i) = \sigma_i^2$, aşağıdaki gibi tanımlar.

$$\sigma_i^2 = \alpha_1 + \alpha_2 Z_i$$

Yani σ_i^2 , Z 'nin doğrusal bir fonksiyonu olup, burada X değişkeni Z şeklinde ifade edilir. Eğer $\alpha_2 = 0$ ise, $\sigma_i^2 = \alpha_1$ olur, bu da hata teriminin varyansının bir sabite eşit olduğunu gösterir. Öyleyse σ_i^2 'nin sabit olup olmadığını anlamak için $\alpha_2 = 0$ hipotezi test edilir. Bu durumda yokluk hipotezi

$$H_0: \alpha_2 = 0$$

olur. Buna göre test adımları aşağıdaki gibidir:

1. $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ modeli EKK yöntemi ile tahmin edilir ve

e_1, e_2, \dots, e_n artık değerleri bulunur.

2. σ^2 'nin en çok olabilirlik tahmin edicisi $\tilde{\sigma}^2 = \sum \frac{e_i^2}{n}$ elde edilir.

3. $p_i = \frac{e_i^2}{\tilde{\sigma}^2}$ olmak üzere, Z ile regresyon modeli aşağıdaki gibi oluşturulur.

$$p_i = \alpha_1 + \alpha_2 Z_i + v_i$$

Burada v_i modelin hata terimidir.

4. 3.adımdaki regresyon modelinin regresyon kareler toplamı (RKT) bulunarak Q_{hsp} değeri aşağıdaki gibi elde edilir.

$$Q_{hsp} = \frac{1}{2}(RKT)$$

Q_{hsp} , $(m-1)$ serbestlik dereceli ki-kare dağılımına sahiptir. Burada m, modeldeki parametre sayısıdır. Buna göre $Q_{hsp} > \chi_{m-1, \alpha}^2$ ise H_0 hipotezi reddedilir. Yani değişen varyans vardır sonucuna ulaşılır.

3.3.2. Spearman sıra korelasyonu testi

Spearman (1961), iki değişken arasındaki ilişkinin ölçümü için aşağıdaki sıra korelasyon katsayısını önermiştir. Bu katsayı sabit varyans varsayımının geçerli olup olmadığının belirlenmesi için de kullanılabilir. Bu durumda H_0 hipotezi sıra korelasyon katsayısının sıfır olduğu yani değişen varyansın olmadığını, H_1 hipotezinin ise sıra korelasyon katsayısının anlamlı olduğunu yani değişen varyansın olduğunu ifade etmektedir. Buna göre sıra korelasyon katsayısı ρ_s ile ifade edildiğinde hipotezler aşağıdaki gibi kurulur. (Güriş ve Çağlayan, 2018: 552, 553).

$$H_0 : \rho_s = 0$$

$$H_1 : \rho_s \neq 0$$

Sıra korelasyon katsayısı istatistiği r_s aşağıdaki gibi hesaplanır.

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2-1)}$$

Burada d_i , artık terimin mutlak değerinin sıra sayısı $R(|e_i|)$ ile X değişkeninin sıra sayısı $R(X_i)$ arasındaki farkı ifade etmektedir, yani $d_i = R(|e_i|) - R(X_i)$. Bu ilişki katsayısı aşağıda izlenen adımlarla değişen varyansı bulmada kullanılır. Buna göre:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

olmak üzere, test adımları aşağıdaki gibidir:

1. Modelin artık terimleri olan e_i değerleri bulunur.
2. Hem $|e_i|$, hem de X_i 'nin sıra sayıları elde edilerek aradaki fark yani d_i bulunur ve r_s istatistiği hesaplanır.
3. Yokluk hipotezinin testi için, t istatistiği aşağıdaki gibi hesaplanır.

$$t = \frac{r_s \sqrt{n-2}}{\sqrt{1-r_s^2}}$$

Burada t istatistiği $(n-2)$ serbestlik dereceli t dağılımına sahiptir. Eğer hesaplanan t değeri $t_{(n-2),\alpha/2}$ değerinden büyükse H_0 hipotezi reddedilir. Yani değişen varyans vardır sonucuna ulaşılır.

3.3.3. Glejser testi

Glejser (1969), Breusch–Pagan testine benzer bir test geliştirmiştir. İlk adımda regresyon denklemi bulunur. Daha sonra elde edilen artık terimlerinin mutlak değeri, yani $|e_i|$ ile X değişkeni ile ifade edilen Z_i arasında regresyon modeli kurularak aşağıdaki gibi regresyon modeli oluşturulur.

$$|e_i| = \alpha_1 + \alpha_2 Z_i + v_i$$

Burada v_i modelin hata terimidir. Bu durumda H_0 hipotezi aşağıdaki gibi olur.

$$H_0: \alpha_2 = 0$$

Regresyon analizinden tahmin edilen t istatistiği $t_{(n-2),\alpha/2}$ değerinden büyükse H_0 hipotezi reddedilir. Yani değişen varyans vardır sonucuna ulaşılır.

3.3.4. Goldfeld-Quandt testi

Goldfeld ve Quandt (1965) artıkların varyansının tüm gözlem boyunca aynı ise (yani homojen ise) o zaman örneğin bir kısmının varyansının diğer kısmının varyansı ile aynı

olmalıdır fikrine dayalı bir yöntem geliştirmiştir. Buna göre yokluk hipotezi aşağıdaki gibi ifade edilir.

$$H_0 : \sigma_1^2 = \sigma_2^2$$

σ_i^2 'nin X_i ile aşağıdaki gibi ilişkili olduğu varsayılır.

$$\sigma_i^2 = \sigma^2 X_i^2$$

Burada σ^2 sabittir ve X_i değerleri büyüdükçe σ_i^2 'nin de büyüyeceği anlamına gelir. Eğer böyle olursa modelde değişen varyans bulunması çok olasıdır.

Bu yöntemin test adımları aşağıdaki gibidir:

1. Gözlemler en küçük X_i değerinden başlayarak sıralanır.
2. Bir c sayısı belirleyip sıralanan verilerin ortasından c sayısı kadar gözlemi çıkartıp kalan $(n-c)$ gözlem ikiye bölünür.
3. İlk $(n-c)/2$ ile son $(n-c)/2$ gözlemleri için ayrı ayrı 2 tane regresyon modeli kurulup her model için artık kareler toplamları (AKT) bulunur. Buna göre ilk grubun artık kareler toplamı yani AKT_1 ve son grubun artık kareler toplamı yani AKT_2 olmak üzere Goldfeld-Quandt test istatistiği,

$$\lambda = \frac{\frac{AKT_2}{sd}}{\frac{AKT_1}{sd}}$$

şeklinde hesaplanır. Burada $sd = \frac{n-c}{2} - m$ olup, m modeldeki parametre sayısıdır. λ test istatistiği sd serbestlik dereceli F dağılımına sahiptir. Buna göre $\lambda > F_{sd, sd, \alpha}$ ise H_0 hipotezi reddedilir. Yani değişen varyans vardır.

Gözlem sayısı farklıysa, serbestlik dereceleri buna göre değiştirilmelidir. Goldfeld ve Quandt (1965), $n = 30$ için $c = 8$ ve $n = 60$ için $c = 16$ önermektedir. Ancak Goldfeld ve Quandt 1970 yılından sonraki yaptığı çalışmalarında $n = 30$ için $c = 4$ kullanmışlardır. Ayrıca Goldfeld – Quandt testinin başarısı yalnızca (atılacak ortadaki gözlem sayısı) c 'nin

değerine değil, aynı zamanda gözlemleri sıralamada kullanılacak doğru X değişkeninin seçimine de bağlıdır (Gujarati, 2012: 383).

3.3.5. Park testi

Park (1966), σ_i^2 'nin açıklayıcı değişken X_i 'nin bir fonksiyonu olduğunu ileri sürerek alternatif bir Lagrange çarpanı testi geliştirmiştir. Buna göre hatanın varyansının formu aşağıdaki gibidir:

$$\sigma_i^2 = \sigma^2 X_i^\beta e^{v_i}$$

ya da

$$\ln \sigma_i^2 = \ln \sigma^2 + \beta \ln X_i + v_i$$

Burada v_i hata terimidir.

Park (1966), genellikle σ_i^2 bilinmediğinden, onun yerine e_i^2 'yi kullanıp aşağıdaki modeli tanımlamıştır.

$$\ln e_i^2 = \ln \sigma^2 + \beta \ln X_i + v_i \quad (3.2)$$

Burada ilgilenilen yokluk hipotezi aşağıdaki gibidir.

$$H_0 : \beta = 0$$

Eğer β istatistik bakımından anlamlı çıkarsa, değişen varyans olduğu söylenebilir. Anlamsız çıkarsa, sabit varyans varsayımı kabul edilir.

Test adımları şu şekildedir:

1. Değişen varyans olup olmadığına bakılmaksızın regresyon katsayılarının EKK tahmin edicileri bulunur.
2. Elde edilen tahmin denkleminde yola çıkarak e_i bulunur ve E.ş 3.2 'de ifade edildiği gibi regresyon modeli kurularak β katsayısının anlamlı olup olmadığına bakılır.

3.3.6. White testi

Gözlemleri, değişen varyansa yol açtığı düşünülen X değişkenine göre yeniden sıraya koyan Goldfeld-Quandt testinden ya da normallik varsayımına karşı duyarlı olan Breusch-Pagan-Godfrey testinden farklı olarak, White'nin önerdiği değişen varyans testi normallik varsayımına dayanmadığı gibi uygulaması da kolaydır. Buradaki temel düşüncüyü açıklamak için aşağıdaki regresyon modelini ele alalım:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

Buna göre test adımları aşağıdaki gibidir:

1. Regresyon modelinden faydalanılarak e_i artık değerleri bulunur.
2. e_i^2 ile X değişkeni arasında aşağıdaki gibi regresyon modeli kurulur.

$$e_i^2 = \alpha_1 + \alpha_2 X_i + \alpha_3 X_i^2 + v_i \quad (3.3)$$

Yani ilk regresyon modelinin artıklarının karesi ile, açıklayıcı değişken olan X değişkeni ve onun karesine göre regresyon modeli kurulur. Çoklu doğrusal regresyon modelinde açıklayıcı değişkenlerin daha yüksek kuvvetleri de işe katılarak regresyon modeli kurulabilir. Eşitlik (3.3)'de ifade edilen modelin belirleme katsayısı olan R^2 değeri bulunur.

Değişen varyans yoktur hipotezini test etmek için modelden hesaplanan $n.R^2_{hsp}$ test istatistiği kullanılır.

$$n.R^2_{hsp} \sim \chi^2_{m-1}$$

Burada m modeldeki parametre sayısıdır. Bulunan $n.R^2_{hsp}$ değeri, $\chi^2_{sd,\alpha}$ değerinden büyükse H_0 hipotezi reddedilir. Yani değişen varyans vardır.

White testine ilişkin dikkat edilmesi gereken, bir modelde çok sayıda açıklayıcı değişken varsa, bunları, bunların karelerini (ya da daha yüksek üslerini), çapraz çarpımlarını hesaba katmak serbestlik derecesini artırır.

3.3.7. Harvey–Godfrey testi

Harvey (1976) ve Godfrey (1978) hata teriminin varyansının regresyon denklemindeki açıklayıcı değişkenlerin üstel bir fonksiyonu olduğunu varsayarak bir test yöntemi geliştirmişlerdir. Buna göre, ilk adımda regresyon denklemi bulunur daha sonra elde edilen artık terimleri ile aşağıdaki regresyon modeli kurulur.

$$\ln e_i^2 = \alpha_1 + \alpha_2 X_i + v_i \quad (3.4)$$

Değişen varyans yoktur hipotezini test etmek için modelden hesaplanan $n.R^2_{hsp}$ test istatistiği kullanılır.

Burada $n.R^2_{hsp} \sim \chi^2_{m-1}$ olup m modeldeki parametre sayısıdır. Bulunan $n.R^2_{hsp}$ değeri, $\chi^2_{m-1,\alpha}$ değerinden büyükse H_0 hipotezi reddedilir. Yani değişen varyans vardır.

3.3.8. Modifiye edilmiş Levene testi

Modifiye edilmiş Levene testi artık terimlerinin normal dağılıma uymasını gerektirmemektedir. Artık terimleri X'in küçükten büyüğe sıralamasına göre iki gruba ayrılır. Birinci grubun yığına ait hata varyansı σ_1^2 ve 2. grubun yığına ait hata varyansı σ_2^2 olmak üzere, H_0 hipotezi aşağıdaki gibi tanımlanır.

$$H_0 : \sigma_1^2 = \sigma_2^2 \text{ (değişen varyans yoktur)}$$

H_0 hipotezinin testi için, her grubun artık terimlerinin medyanları bulunur. Levene testi iki örneklem t testine dayanır. n_1 ve n_2 her iki gruptaki veri sayısını göstermek üzere $n = n_1 + n_2$ 'dir. Her bir grubun medyanı sırasıyla \tilde{e}_1 ve \tilde{e}_2 olmak üzere, her bir grubun artık terimlerinin grup medyan değerinden mutlak sapmaları

$$d_{i1} = |e_{i1} - \tilde{e}_1| \quad d_{i2} = |e_{i2} - \tilde{e}_2|$$

elde edilir. İki-örneklem t-test istatistiği ise,

$$t_L^* = \frac{\bar{d}_1 - \bar{d}_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

şeklindedir. Burada \bar{d}_1 ve \bar{d}_2 , oluşturulan d_{i1} ve d_{i2} gruplarının sırasıyla örneklem ortalamalarıdır. Ortak varyans s^2 ise,

$$s^2 = \frac{\sum(d_{i1} - \bar{d}_1)^2 + \sum(d_{i2} - \bar{d}_2)^2}{n - 2}$$

şeklindedir. Levene test istatistiği t_L^* , asimptotik olarak n-2 serbestlik dereceli t dağılımını sahiptir, buna göre $|t_L^*| > t_{n-2, \alpha/2}$ ise H_0 hipotezi reddedilir (Neter, Kutner, Nachtsheim ve Wasserman, 1996: 116).

Ayrıca tez çalışmamızda, her bir grubun ortalaması sırasıyla \bar{e}_1 ve \bar{e}_2 olmak üzere, her bir grubun artık terimlerinin grup ortalaması değerinden mutlak sapmaları alınarak Modifiye Edilmiş Levene (MLA) testi tarafımızdan elde edilmiştir. Simülasyon çalışmasına bu testin sonuçları eklenmiş ve diğer testlere göre performansı incelenmiştir.

3.4. Uygulama

Bu bölümde regresyon modelinde değişen varyans durumunun olup olmadığının tespitinde kullanılan testler bir uygulama üzerinde gösterilmiştir.

Gelir değerlerinin tüketim harcamalarını nasıl etkilediği ile ilgili bir araştırma yapılması planlanmaktadır. Bu amaçla, 30 ailenin tüketim harcamaları ve gelir değerleri aşağıdaki gibi elde edilmiştir. Buna göre, bağımlı değişken Y tüketim harcaması değişkeni ve bağımsız değişken X gelir değişkenidir. Ekonomide gelir arttıkça tüketim de artacağı varsayıldığından, Y'nin X'e bağlı basit doğrusal regresyon modeli kurulur (Gujarati, 2012: 384).

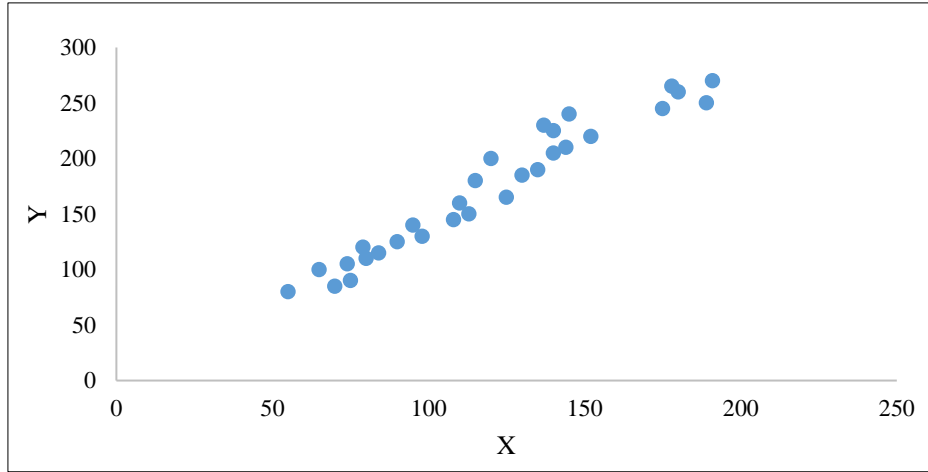
Çizelge 3.1. 30 ailenin tüketim harcaması (Y) ve gelir verileri (X)

Y	X
55	80
65	100

Çizelge 3.1. (devam) 30 ailenin tüketim harcaması (Y) ve gelir verileri (X)

Y	X
70	85
80	110
79	120
84	115
98	130
95	140
90	125
75	90
74	105
110	160
113	150
125	165
108	145
115	180
140	225
120	200
145	240
130	185
152	220
144	210
175	245
180	260
135	190
140	205
178	265
191	270
137	230
189	250

Y'nin X'e bağlı serpilme diyagramı aşağıdaki gibidir.



Şekil 3.4. Y'nin X'e bağlı serpmeye diyagramı

Şekil 3.4'de görüldüğü gibi Y ile X arasında doğrusal bir ilişki olduğu görülmekte olup Y ile X arasında doğrusal regresyon modeli kurulabilir. Buna göre yapılan analiz sonuçları aşağıdaki gibi elde edilmiştir.

Çizelge 3.2. Regresyon modeli istatistikleri

<i>Regresyon Model İstatistikleri</i>	
R^2	0,9466
Standart Hata ($\hat{\sigma}$)	9,183
n	30

Çizelge 3.3. ANOVA tablo sonuçları

ANOVA					
<i>Değişim kaynağı</i>	<i>Serbestlik derecesi</i>	<i>Kareler Toplamları</i>	<i>Kareler Ortalaması</i>	<i>F test istatistiği</i>	<i>p-değeri</i>
Regresyon	1	41886,71342	41886,71	496,71828	0,0000
Artık	28	2361,15325	84,3269		
Toplam	29	44247,86667			

Çizelge 3.4. Regresyon katsayıları analizi

	<i>Katsayılar tahmini</i>	<i>Katsayılar tahmininin Standart Hatası</i>	<i>t test istatistiği</i>	<i>P-değeri</i>
β_0	9,29030744	5,231385657	1,775879	0,0866232
β_1	0,637784558	0,028616659	22,28718	0,000

$$\hat{Y}_i = 9,2903 + 0,6378X_i$$

Doğrusal regresyon modeli varsayımlarından biri olan değişen varyans durumunun olup olmadığının incelenmesi için biçimsel olmayan ve biçimsel yöntemler sırasıyla aşağıdaki gibi hesaplanmıştır.

Biçimsel olmayan yöntemler için artık terimlerin hesaplanması aşağıdaki gibidir;

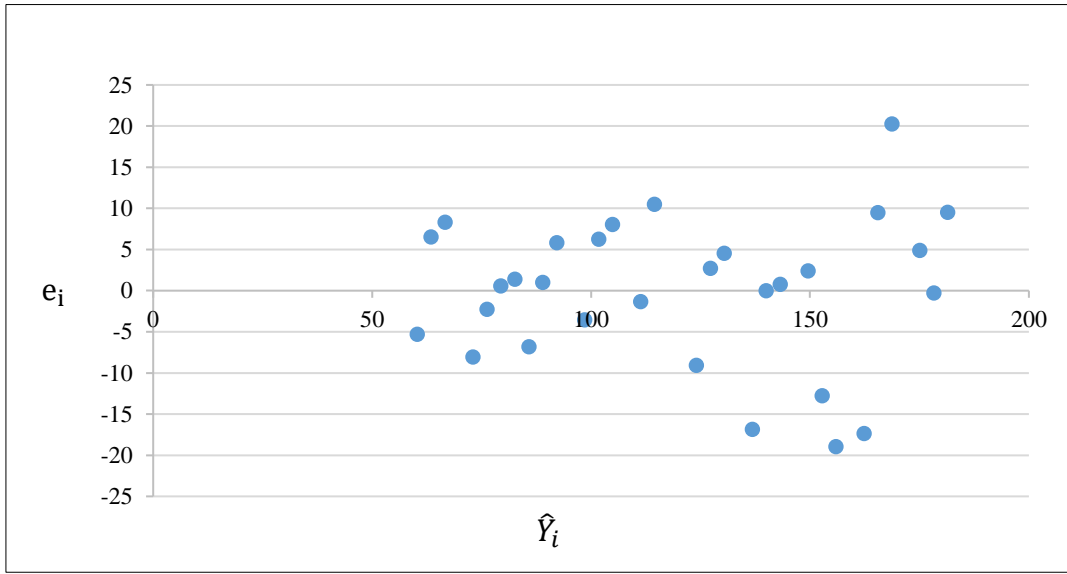
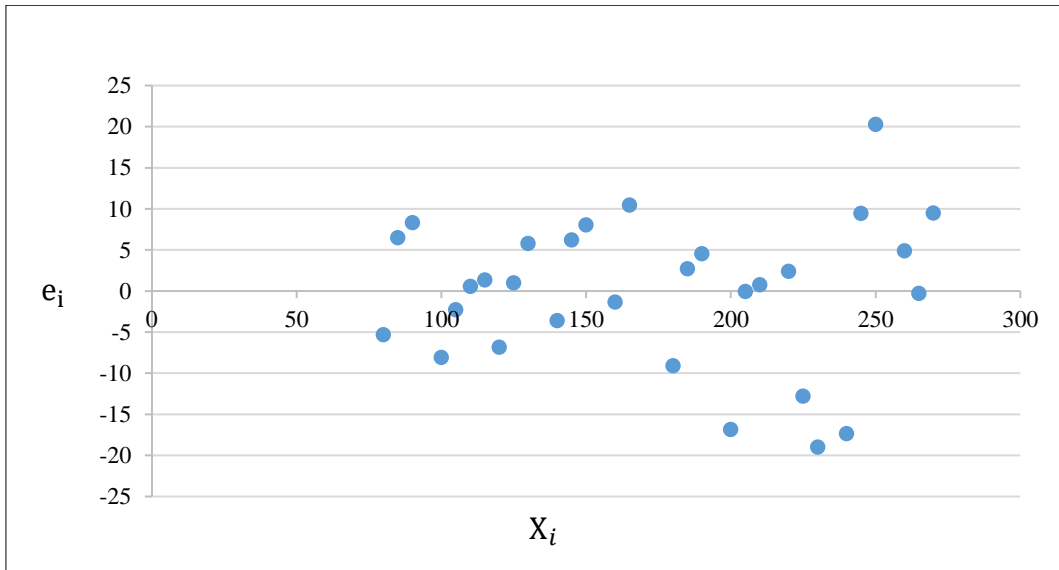
Çizelge 3.5. Artık terimlerin değerleri

\hat{Y}_i	$e_i = Y_i - \hat{Y}_i$
60,313	-5,3131
73,069	-8,0688
63,502	6,4980
79,447	0,5534
85,824	-6,8245
82,636	1,3645
92,202	5,7977
98,580	-3,5801
89,013	0,9866
66,691	8,3091
76,258	-2,2577
111,336	-1,3358
104,958	8,0420
114,525	10,4752
101,769	6,2309
124,092	-9,0915
152,792	-12,7918
136,847	-16,8472
162,359	-17,3586
127,280	2,7195
149,603	2,3971
143,225	0,7749
165,548	9,4525
175,114	4,8857
130,469	4,5306
140,036	-0,0361

Çizelge 3.5. (devam) Artık terimlerin değerleri

\hat{Y}_i	$e_i = Y_i - \hat{Y}_i$
178,303	-0,3032
181,492	9,5079
155,981	-18,9808
168,736	20,2636

Değişen varyans durumunun olup olmadığı grafiksel olarak aşağıda gösterilmiştir.

Şekil 3.5. \hat{Y}_i 'nin artık değerlerine (e_i) bağlı serpmeye diyagramıŞekil 3.6. X_i 'nin artık değerlerine (e_i) bağlı serpmeye diyagramı

Veri içerisinde değişen varyans olup olmadığını anlamak için ilk olarak artıkların serpmeye diyagramına bakabiliriz. X_i ve \hat{Y}_i ile e_i 'lerin ayrı ayrı grafiklerini çizdiğimizde dağılımın belli bir örüntüye sahip olduğunu yani değişen varyans durumunun olduğunu görebiliriz. Fakat bilindiği gibi grafiksel yöntemler subjektif yöntemler olup daha kesin sonuca varmak için biçimsel yöntemleri uygulamak daha doğru olacaktır. Buna göre aşağıda biçimsel yöntemler sırasıyla verilmiştir.

Park testinin uygulaması

Değişen varyansın olup olmadığını testi için Park testinde regresyon analizinden sonra e_i 'ler bulunur sonra $\ln e_i^2 = \ln \sigma^2 + \beta \ln X_i + v_i$ denklemindeki gibi regresyon analizi yapılarak β katsayısının anlamlı olup olmadığına bakılır. Buna göre yokluk ve alternatif hipotezler aşağıdaki gibidir.

$$H_0 : \beta = 0 \quad (\beta \text{ katsayısı anlamsızdır}) \quad (3.5)$$

$$H_1 : \beta \neq 0 \quad (\beta \text{ katsayısı anlamlıdır})$$

Park testine göre Çizelge 3.5'te verilen artıkların karelerinin logaritması alınarak bağımlı değişken, ve X değişkeninin de logaritması alınarak $\ln e_i^2 = \ln \sigma^2 + \beta \ln X_i + v_i$ modeli için regresyon analizi yapılır. Bu model için değişkenler aşağıda verilen tablodaki gibi elde edilir.

Çizelge 3.6. Park testi hesaplamaları

Y_i	X_i	\hat{Y}_i	$e_i = Y_i - \hat{Y}_i$	e_i^2	$\ln e_i^2 = Y_i^*$	$\ln X_i = X_i^*$
55	80	60,313	-5,3131	28,2287	3,3403	4,3820
65	100	73,069	-8,0688	65,1049	4,1760	4,6052
70	85	63,502	6,4980	42,2241	3,7430	4,4427
80	110	79,447	0,5534	0,3062	-1,1834	4,7005
79	120	85,824	-6,8245	46,5732	3,8410	4,7875
84	115	82,636	1,3645	1,8618	0,6215	4,7449
98	130	92,202	5,7977	33,6133	3,5149	4,8675
95	140	98,580	-3,5801	12,8174	2,5508	4,9416
90	125	89,013	0,9866	0,9734	-0,0269	4,8283
75	90	66,691	8,3091	69,0408	4,2347	4,4998
74	105	76,258	-2,2577	5,0971	1,6287	4,6540

Çizelge 3.6. (devam) Park testi hesaplamaları

Y_i	X_i	\hat{Y}_i	$e_i = Y_i - \hat{Y}_i$	e_i^2	$\ln e_i^2 = Y_i^*$	$\ln X_i = X_i^*$
110	160	111,336	-1,3358	1,7845	0,5791	5,0752
113	150	104,958	8,0420	64,6739	4,1694	5,0106
125	165	114,525	10,4752	109,7307	4,6980	5,1059
108	145	101,769	6,2309	38,8245	3,6591	4,9767
115	180	124,092	-9,0915	82,6559	4,4147	5,1930
140	225	152,792	-12,7918	163,6310	5,0976	5,4161
120	200	136,847	-16,8472	283,8288	5,6484	5,2983
145	240	162,359	-17,3586	301,3210	5,7082	5,4806
130	185	127,280	2,7195	7,3959	2,0009	5,2204
152	220	149,603	2,3971	5,7460	1,7485	5,3936
144	210	143,225	0,7749	0,6005	-0,5100	5,3471
175	245	165,548	9,4525	89,3493	4,4926	5,5013
180	260	175,114	4,8857	23,8701	3,1726	5,5607
135	190	130,469	4,5306	20,5266	3,0217	5,2470
140	205	140,036	-0,0361	0,0013	-6,6406	5,3230
178	265	178,303	-0,3032	0,0919	-2,3866	5,5797
191	270	181,492	9,5079	90,3994	4,5042	5,5984
137	230	155,981	-18,9808	360,2691	5,8869	5,4381
189	250	168,736	20,2636	410,6116	6,0176	5,5215

Y_i^* 'ın X_i^* değişkeni ile regresyon modeli için tablo sonuçları aşağıdaki gibidir.

Çizelge 3.7. Park testi regresyon modeli istatistikleri

<i>Regresyon Model İstatistikleri</i>	
R^2	0,0019
Standart Hata($\hat{\sigma}$)	2,843
n	30

Çizelge 3.8. Park testi ANOVA tablo sonuçları

ANOVA					
<i>Değişim kaynağı</i>	<i>Serbestlik derecesi</i>	<i>Kareler Toplamları</i>	<i>Kareler Ortalaması</i>	<i>F testi</i>	<i>p</i>
Regresyon	1	0,4489	0,4489	0,05553	0,8154
Artık	28	226,31	8,0826		
Toplam	29	226,76			

Çizelge 3.9. Park testi için regresyon katsayıları analizi

	<i>Katsayılar tahmini</i>	<i>Katsayılar tahmininin Standart Hatası</i>	<i>t</i>	<i>p-değeri</i>
β_0	1,0140	7,2749	0,1393	0,8901
β_1	0,3359	1,4252	0,2356	0,8154

Park testine göre eğer β istatistik bakımından anlamlı çıkarsa, değişen varyansın var olduğu söylenebilir, anlamsız çıkarsa sabit varyans varsayımı kabul edilir.

Tablo sonuçlarından elde edildiğine göre p değeri= 0,8154 > $\alpha/2 = 0,025$ Eş. 3.5’de verilen H_0 hipotezi reddedilemediği için β katsayısı istatistiksel olarak anlamsız olduğu sonucuna varılır. Buna göre değişen varyansın olduğu söylenemez.

Golfeld-Quandt testinin uygulaması

Golfeld-Quandt testi aşağıdaki gibi uygulanır. Buna göre değişen varyansın olup olmadığının testi için yokluk ve alternatif hipotez

$$H_0 : \sigma_1^2 = \sigma_2^2 \text{ (değişen varyans yoktur)} \quad (3.6)$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2 \text{ (değişen varyans vardır)} \quad \text{şeklinde kurulur.}$$

Test adımlarında anlatıldığı gibi, önce X verileri küçükten büyüğe sıralanır ve ortadaki 4 gözlem veriden çıkarılarak, kalan veri iki gruba ayrılır.

Çizelge 3.10. Golfeld-Quandt testi grup 1 verileri

(Grup 1) Y	(Grup 1) X
55	80
70	85
75	90
65	100
74	105
80	110
84	115
79	120
90	125
98	130
95	140
108	145
113	150

Çizelge 3.11. Golfeld-Quandt testi grup 2 verileri

(Grup 2) Y	(Grup 2) X
135	190
120	200
140	205
144	210
152	220
140	225
137	230
145	240
175	245
189	250
180	260
178	265
191	270

Buna göre 2 ayrı grup için regresyon analizi yapılır.

Grup 1 için regresyon analizi

Çizelge 3.12. Golfeld-Quandt testi ile grup 1 regresyon modeli istatistikleri

<i>Regresyon Model İstatistikleri</i>	
R ²	0,889
Standart Hata($\hat{\sigma}$)	5,855
n	13

Çizelge 3.13. Golfeld-Quandt testi ile grup 1 ANOVA sonuçları

ANOVA					
<i>Değişim kaynağı</i>	<i>Serbestlik derecesi</i>	<i>Kareler Toplamları</i>	<i>Kareler Ortalaması</i>	<i>F test istatistiği</i>	<i>p-değeri</i>
Regresyon	1	3010	3010	87,79	0,000
Artık	11	377,1	34,3		
Toplam	12	3387,2			

Çizelge 3.14. Golfeld-Quandt testi ile grup 1 regresyon katsayıları analizi

	<i>Katsayılar tahmini</i>	<i>Katsayılar tahmininin Standart Hatası</i>	<i>t test istatistiği</i>	<i>P-değeri</i>
β_0	3,401	8,7	0,39	0,7
β_1	0,697	0,07	9,37	0,000

Tablo sonuçlarına göre model tahmin denklemi aşağıdaki gibi elde edilir.

$$\hat{Y}_{1i} = 3,401 + 0,697X_{1i}$$

Grup 2 için regresyon analizi

Çizelge 3.15. Golfeld-Quandt testi ile grup 2 regresyon modeli istatistikleri

<i>Regresyon Model İstatistikleri</i>	
R ²	0,768
Standart Hata($\hat{\sigma}$)	11,82
n	13

Çizelge 3.16. Golfeld-Quandt ile testi grup 2 ANOVA sonuçları

ANOVA					
<i>Değişim kaynağı</i>	<i>Serbestlik derecesi</i>	<i>Kareler Toplamları</i>	<i>Kareler Ortalaması</i>	<i>F test istatistiği</i>	<i>p-değeri</i>
Regresyon	1	5088,9	5088,9	36,4	0,000
Artık	11	1536,8	139,7		
Toplam	12	6625,7			

Çizelge 3.17. Golfeld-Quandt testi ile grup 2 regresyon katsayıları analizi

	<i>Katsayılar tahmini</i>	<i>Katsayılar tahmininin Standart Hatası</i>	<i>t test istatistiği</i>	<i>P-değeri</i>
β_0	-28,02	30,64	-0,91	0,38
β_1	0,794	0,13	6,03	0,000

Tablo sonuçlarına göre model tahmin denklemi aşağıdaki gibi elde edilir.

$$\hat{Y}_{2i} = -28,02 + 0,794X_{2i}$$

Buna göre bu iki gruptan elde edilen artık kareler toplamı ile F test istatistiği λ aşağıdaki gibi hesaplanır.

$$\lambda = \frac{\frac{AKT_2}{sd}}{\frac{AKT_1}{sd}} = \frac{139,7}{34,3} = 4,073$$

F 'in tablo değeri $F_{11,11,0.05} = 2,82$ olarak bulunur. Buna göre $\lambda = 4,073 > F_{11,11,0.05} = 2,82$ olduğundan (3.6) eşitliğinde verilen H_0 hipotezi reddedilir. Yani değişen varyans olduğu söylenebilir.

Spearman sıra korelasyonu testinin uygulaması

Değişen varyansın olup olmadığının testi için yokluk ve alternatif hipotezler aşağıdaki gibidir.

$H_0: \rho_S = 0$ (değişen varyans yoktur)

$H_1: \rho_S \neq 0$ (değişen varyans vardır)

Spearman Sıra Korelasyonu testi için hesaplamalar Çizelge 3.18'de elde edilmiştir.

Çizelge 3.18. Spearman Sıra Korelasyonu testi hesaplamaları

Y_i	X_i	\hat{Y}_i	e_i	$ e_i $	$R(e_i)$	$R(X_i)$	$d = R(e_i) - R(X_i)$
55	80	60,313	-5,3131	5,313072	14	1	13
65	100	73,069	-8,0688	8,068763	20	4	16
70	85	63,502	6,498	6,498005	17	2	15
80	110	79,447	0,5534	0,553391	3	6	-3
79	120	85,824	-6,8245	6,824454	18	8	10
84	115	82,636	1,3645	1,364468	7	7	0
98	130	92,202	5,7977	5,7977	15	10	5
95	140	98,58	-3,5801	3,580146	11	11	0
90	125	89,013	0,9866	0,986623	5	9	-4
75	90	66,691	8,3091	8,309082	21	3	18
74	105	76,258	-2,2577	2,257686	8	5	3
110	160	111,336	-1,3358	1,335837	6	14	-8
113	150	104,958	8,042	8,042009	19	13	6
125	165	114,525	10,4752	10,47524	25	15	10
108	145	101,769	6,2309	6,230932	16	12	4
115	180	124,092	-9,0915	9,091528	22	16	6
140	225	152,792	-12,7918	12,79183	26	23	3
120	200	136,847	-16,8472	16,84722	27	19	8
145	240	162,359	-17,3586	17,3586	28	25	3
130	185	127,28	2,7195	2,719549	10	17	-7
152	220	149,603	2,3971	2,39709	9	22	-13
144	210	143,225	0,7749	0,774935	4	21	-17

Çizelge 3.18. (devam) Spearman Sıra Korelasyonu testi hesaplamaları

Y_i	X_i	\hat{Y}_i	e_i	$ e_i $	$R(e_i)$	$R(X_i)$	$d = R(e_i) - R(X_i)$
180	260	175,114	4,8857	4,885708	13	28	-15
135	190	130,469	4,5306	4,530627	12	18	-6
140	205	140,036	-0,0361	0,036142	1	20	-19
178	265	178,303	-0,3032	0,303215	2	29	-27
191	270	181,492	9,5079	9,507862	24	30	-6
137	230	155,981	-18,9808	18,98076	29	24	5
189	250	168,736	20,2636	20,26355	30	27	3

Tablodan elde edilen sonuçlara göre Spearman Sıra Korelasyonu istatistiği r_s ,

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

$$r_s = 1 - \frac{6 * 3404}{30(30^2 - 1)} = 0,2427$$

şeklinde hesaplanır. Buna göre test istatistiği;

$$t = \frac{r_s \sqrt{n-2}}{\sqrt{1-r_s^2}} = \frac{0,2427 \sqrt{30-2}}{\sqrt{1-0,2427^2}} = \frac{1,2842}{0,9701} = 1,32$$

olur. Hesaplanan t değeri ile tablodaki $t_{28,0.025}$ değeri karşılaştırılır. Hesaplanan t değeri tablo değerinden büyük ise H_0 hipotezi reddedilir.

$t = 1,32 < t_{28,0.025} = 2,048$ olduğundan H_0 hipotezi kabul edilir. Buna göre değişen varyans yoktur.

Glejser testinin uygulaması

Değişen varyans olup olmadığının tespiti için Glejser testi aşağıdaki gibi uygulanır.

İlk olarak $|e_i| = \beta_0 + \beta_1 X_i + v_i$ modeli oluşturulur ve regresyon analizi yapılır. buna göre aşağıda verilen tablo sonuçları elde edilir.

Çizelge 3.19. Glejser testi ile regresyon modeli istatistikleri

<i>Regresyon Model İstatistikleri</i>	
R^2	0,118
Standart Hata($\hat{\sigma}$)	5,475
n	30

Çizelge 3.20. Glejser testi ile ANOVA sonuçları

ANOVA					
<i>Değişim kaynağı</i>	<i>Serbestlik derecesi</i>	<i>Kareler Toplamları</i>	<i>Kareler Ortalaması</i>	<i>F test istatistiği</i>	<i>p- değeri</i>
Regresyon	1	113,14	113,14	3,77	0,062
Artık	28	839,25	29,97		
Toplam	29	952,4			

Çizelge 3.21. Glejser testi ile regresyon katsayıları analizi

	<i>Katsayılar tahmini</i>	<i>Katsayılar tahmininin Standart Hatası</i>	<i>t test istatistiği</i>	<i>P-değeri</i>
β_0	1,112	3,118	0,356	0,723
β_1	0,033	0,017	1,94	0,062

$$H_0 : \beta_1 = 0 \quad (\beta \text{ katsayısı anlamsızdır}) \quad (3.7)$$

$$H_1 : \beta_1 \neq 0 \quad (\beta \text{ katsayısı anlamlıdır})$$

Glejser testine göre eğer β_1 istatistik bakımından anlamlı çıkarsa, değişen varyansın var olduğu, anlamsız çıkarsa sabit varyans varsayımı kabul edilir.

Çizelge 3.21 sonuçlarından görüldüğü gibi p değeri= 0,062 > $\alpha/2 = 0,025$ olup Eş. 3.7'de verilen H_0 hipotezi reddedilemediği için β_1 katsayısı istatistiksel olarak anlamsız olduğu sonucuna varılır. Buna göre değişen varyansın olduğu söylenemez.

Breusch-Pagan testinin uygulaması

Değişen varyans olup olmadığının tespiti için Breusch-Pagan testi aşağıdaki gibi uygulanır.

$$p_i = \frac{e_i^2}{\hat{\sigma}^2} \quad \text{olmak üzere;}$$

$$p_i = \alpha_1 + \alpha_2 Z_i + v_i$$

modeli kurularak

$$H_0 : \alpha_2 = 0 \quad (\alpha_2 \text{ katsayısı anlamsızdır})$$

$$H_1 : \alpha_2 \neq 0 \quad (\alpha_2 \text{ katsayısı anlamlıdır})$$

hipotezi test edilir. Buna göre kurulan regresyon modelinin analiz sonuçları aşağıdaki gibi elde edilir.

Çizelge 3.22. Breusch-Pagan testi ile regresyon modeli istatistikleri

<i>Regresyon Model İstatistikleri</i>	
R ²	0,175
Standart Hata($\hat{\sigma}$)	1,321
n	30

Çizelge 3.23. Breusch-Pagan testi ile ANOVA sonuçları

ANOVA					
<i>Değişim kaynağı</i>	<i>Serbestlik derecesi</i>	<i>Kareler Toplamları</i>	<i>Kareler Ortalaması</i>	<i>F test istatistiği</i>	<i>p-değeri</i>
Regresyon	1	10,42	10,42	5,96	0,021
Artık	28	48,90	1,74		
Toplam	29	59,33			

Çizelge 3.24. Breusch-Pagan testi ile regresyon katsayıları analizi

	<i>Katsayılar tahmini</i>	<i>Katsayılar tahmininin Standart Hatası</i>	<i>t test istatistiği</i>	<i>P-değeri</i>
β_0	-0,7426	0,752	-0,986	0,33
β_1	0,01	0,004	2,44	0,021

Çizelge 3.22'den RKT=10.42 olup , $Q_{hsp} = \frac{1}{2}(RKT)$

$Q_{hsp} = \frac{1}{2}(10,428021) = 5,214$ elde edilir.

$Q_{hsp} > \chi_{1,0.05}^2 = 3,8414$ olduğundan H_0 hipotezi reddedilir. Yani değişen varyans vardır.

White testinin uygulaması

Değişen varyans olup olmadığının tespiti için White testi aşağıdaki gibi uygulanır. Regresyon analizinden sonra e_i 'ler bulunur sonra e_i^2 ler ile X değişkeni arasında regresyon analizi yapılır. Model sonucunda çıkan R^2 ile n çarpılıp test istatistiği bulunur. Buna göre tablo sonuçları aşağıdaki gibi elde edilir.

Çizelge 3.25. White testi ile regresyon modeli istatistikleri

<i>Regresyon Model İstatistikleri</i>	
R^2	0,0828
Standart Hata($\hat{\sigma}$)	121
n	30

Çizelge 3.26. White testi ile regresyon katsayıları analizi

	<i>Katsayılar tahmini</i>	<i>Katsayılar tahmininin Standart Hatası</i>	<i>t test istatistiği</i>	<i>P-değeri</i>
β_0	-7,1074	219,02	-0,032451	0,97435
β_1	0,49617	2,7053	0,18341	0,85585
β_2	0,000262	0,007659	0,0342	0,97296

Tablo sonuçlarına göre kurulan model tahmin denklemi aşağıdaki gibi elde edilir.

$$e_i^2 = -7,1074 + 0,496X_i + 0,000262X_i^2$$

Test istatistiği aşağıdaki gibi hesaplanır.

$$n \cdot R^2_{hsp} = 2.4845$$

$$n \cdot R^2_{hsp} = 2.4845 < \chi^2_{2,0.05} = 5.9915 \text{ olduğu için değişen varyans yoktur sonucuna ulaşılır.}$$

Harvey–Godfrey testinin uygulaması

Harvey - Godfrey testinde regresyon analizinden sonra e_i 'ler bulunur. Daha sonra $\ln e_i^2$ ler ile X değişkeni ile regresyon analizi yapılır

$$\ln e_i^2 = \alpha_1 + \alpha_2 X_i + v_i$$

ve model sonucunda çıkan R^2 ile n çarpılıp test istatistiği bulunur. Buna göre yokluk hipotezi aşağıdaki gibi olmak üzere regresyon modelinden elde edilen analiz saşunçları aşağıdaki çizelgeler de verilmiştir.

$$H_0 : \alpha_2 = 0 \quad (\alpha_2 \text{ katsayısı anlamsızdır})$$

$$H_1 : \alpha_2 \neq 0 \quad (\alpha_2 \text{ katsayısı anlamlıdır})$$

Çizelge 3.27. Harvey–Godfrey testi ile regresyon modeli istatistikleri

<i>Regresyon Model İstatistikleri</i>	
R^2	0,004
Standart Hata($\hat{\sigma}$)	2,84
n	30

Çizelge 3.28. Harvey–Godfrey testi ile regresyon katsayıları analizi

	<i>Katsayılar tahmini</i>	<i>Katsayılar tahmininin Standart Hatası</i>	<i>t test istatistiği</i>	<i>P-değeri</i>
β_0	2,203	1,617	1,362	0,184
β_1	0,003	0,008	0,339	0,736

$$\ln e_i^2 = 2,203 + 0,003X_i$$

$$n \cdot R^2_{hsp} = 0,123 \quad \chi^2_{1,0.05} = 3,8414$$

Bulunan $n \cdot R^2_{hsp}$ değeri, $\chi^2_{1,0.05}$ değerinden küçük olduğu için değişen varyans yoktur sonucuna ulaşılır.

Medyan değerine göre modifiye edilmiş Levene testinin uygulaması

Değişen varyansın olup olmadığının testi için yokluk ve alternatif hipotezler aşağıdaki gibidir.

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

Modifiye edilmiş Levene testinde artık terimleri X'in sıralamasına göre iki gruba ayrılır. Her bir grubun artık değerlerinin medyanları, $\tilde{e}_1 = -0,0580$ ve $\tilde{e}_2 = 3,7510$ olarak elde edilir. Her bir grup için d_{i1} ve d_{i2} değerleri elde edilir.

$$d_{i1} = |e_{i1} - \tilde{e}_1| \quad d_{i2} = |e_{i2} - \tilde{e}_2|$$

Buna göre her grup ortalaması sırasıyla $\bar{d}_1 = 4,6445$, $\bar{d}_2 = 8,2484$ şeklinde hesaplanır.

$s^2 = 33.7962$ olmak üzere test istatistiği ise

$$t_L^* = \frac{\bar{d}_1 - \bar{d}_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$t_L^* = 1,6978$$

olarak elde edilir. $|t_L^*| = 1,6978 < t_{n-2, \alpha/2} = 2,0484$ olduğundan H_0 hipotezi kabul edilir.

Değişen varyans yoktur.

Ortalamaya göre modifiye edilmiş Levene testinin uygulaması

Değişen varyansın olup olmadığının testi için yokluk ve alternatif hipotezler aşağıdaki gibidir.

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

Modifiye edilmiş Levene testinde artık terimleri X'in sıralamasına göre iki gruba ayrılır. Her bir grubun artık değerlerinin ortalaması, $\bar{e}_1 = 0$ ve $\bar{e}_2 = 0$ olarak elde edilir. Her bir grup için d_{i1} ve d_{i2} değerleri elde edilir.

$$d_{i1} = |e_{i1} - \bar{e}_1| \quad d_{i2} = |e_{i2} - \bar{e}_2|$$

Buna göre her grup ortalaması sırasıyla $\bar{d}_1 = 4,6484$, $\bar{d}_2 = 8,7766$ şeklinde hesaplanır.

$s^2 = 21,4205$ olmak üzere test istatistiği ise

$$t_L^* = \frac{\bar{d}_1 - \bar{d}_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$t_L^* = 2,4428$$

olarak elde edilir. $|t_L^*| = 2,4428 > t_{n-2, \alpha/2} = 2,0484$ olduğundan H_0 hipotezi reddedilir.

Değişen varyans vardır.



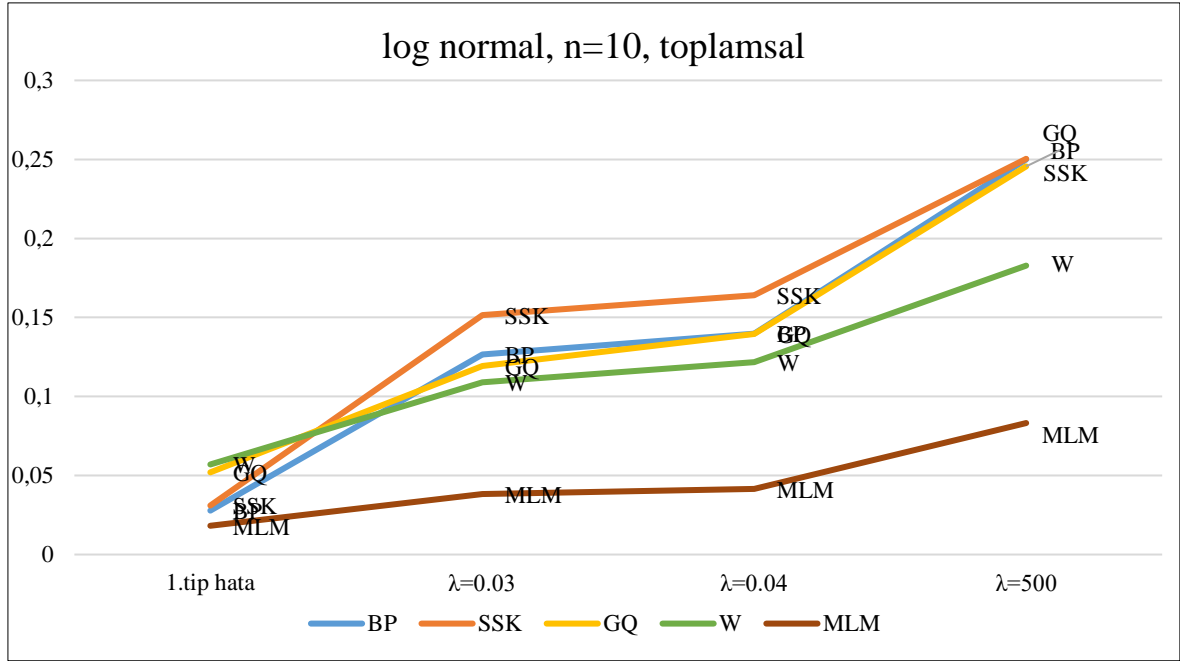
4. SİMÜLASYON ÇALIŞMASI

Bu bölümde 2. bölümde tanıtılan basit doğrusal regresyon modelinde değişen varyansın var olup olmadığını test etmek için kullanılan testlerin Breusch-Pagan (BP), Spearman Sıra Korelasyonu (SSK), Glejser (G), Golfeld-Quandt (GQ), Park (P), White (W), Harvey-Godfrey (HG), medyan değerine göre Modifiye Edilmiş Levene (MLM), ortalamaya göre Modifiye Edilmiş Levene (MLA) testleri ele alınmış ve performansı araştırılmıştır. Bu amaçla simülasyon yöntemi ile testlerin deneysel 1.tip hata oranları ve testin gücü bakımından karşılaştırılması yapılmıştır. Simülasyon çalışmasında, farklı sayıda örnek çapı ve farklı heterojenlik durumları dikkate alınmıştır.

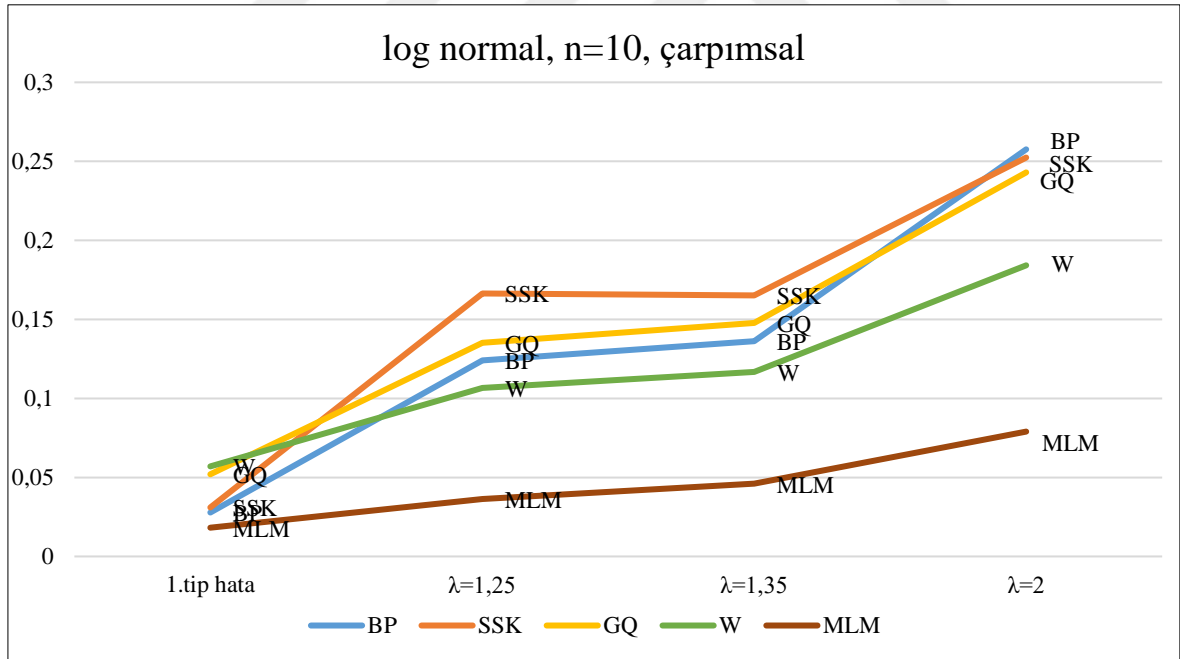
Bu amaçla testlerin deneysel 1. tip hata oranları ve güç değerleri simülasyon yoluyla karşılaştırılmıştır. Çalışmada Griffiths ve Surekha (1986)'da kullanılan simülasyon çalışması referans alınmıştır. Doğrusal regresyon modelini oluşturmak için bağımsız değişken X, iki dağılımdan oluşturulmuştur. Buna göre tekdüze dağılım U[20,100] ve log-normal dağılım LN(3.8,0.4) olacak şekilde üretilmiştir. Örneklem büyüklüğü n =10, 20, 30, 50 ve (β_0, β_1) parametreleri (10,1) olacak şekilde regresyon modeli kurulmuştur. Hata terimleri ortalaması 0 olan normal dağılımdan üretilmiştir. Testlerin deneysel 1. tip hata oranlarını elde etmek için hata teriminin varyansı $\sigma^2=2$ olarak alınmıştır. Testin gücü için hata terimlerinin varyansını elde etmek amacıyla çarpımsal model $V(\varepsilon_i)=\sigma_i^2 = x_i^\lambda$ ve toplamsal model $V(\varepsilon_i)=\sigma_i^2=(1 + \lambda x_i)^2$ şeklinde kurulmuştur. Simülasyon çalışmasında çarpımsal model için $\lambda = 1,25, 1,35$ ve 2, toplamsal model için $\lambda = 0,03, 0,04$ ve 500 olarak alınmıştır. Simülasyon çalışmasında %5 anlamlılık düzeyinde ($\alpha = 0.05$) testlerin deneysel 1. tip hata oranlarını ve güç değerlerini elde etmek için yapılan 5000 tekrar ile reddedilme ortalamaları hesaplanarak yapılmıştır. Ayrıca doğrusal regresyon modelinin varsayımı doğrultusunda X değişkeni simülasyon çalışmasında sabit tutulmuştur. Çalışmada Matlab (2023) programı kullanılmıştır.

Testlerin deneysel 1. Tip hata oranları %6 'yı aştığı zaman testlerin güç değerlendirilmesi anlamlı olmayacaktır. Bu amaçla tabloları daha kolay yorumlamak için testlerin deneysel 1. Tip hata oranları %6 'yı aşanların güç değerleri ** ile gösterilmiştir.

Simülasyon çalışmasında testlerin deneysel 1.tip hataları ve güç değerlerini yorumlamak için ilk olarak grafikler daha sonra tablolar verilmiştir.



Şekil 4.1. X değişkeni Log-normal dağılımdan geldiğinde ve n=10 için testlerin toplamsal modelde deneysel I.tip hata oranları ve güç değerleri



Şekil 4.2. X değişkeni Log-normal dağılımdan geldiğinde ve n=10 için testlerin çarpımsal modelde deneysel I.tip hata oranları ve güç değerleri

Şekil 4.1 ve Şekil 4.2'de toplamsal ve çarpımsal modelde X değişkeni Log-normal dağılımdan geldiğinde n=10 için testlerin deneysel I.tip hata oranları ve güç değerleri grafiksel olarak gösterilmiştir.

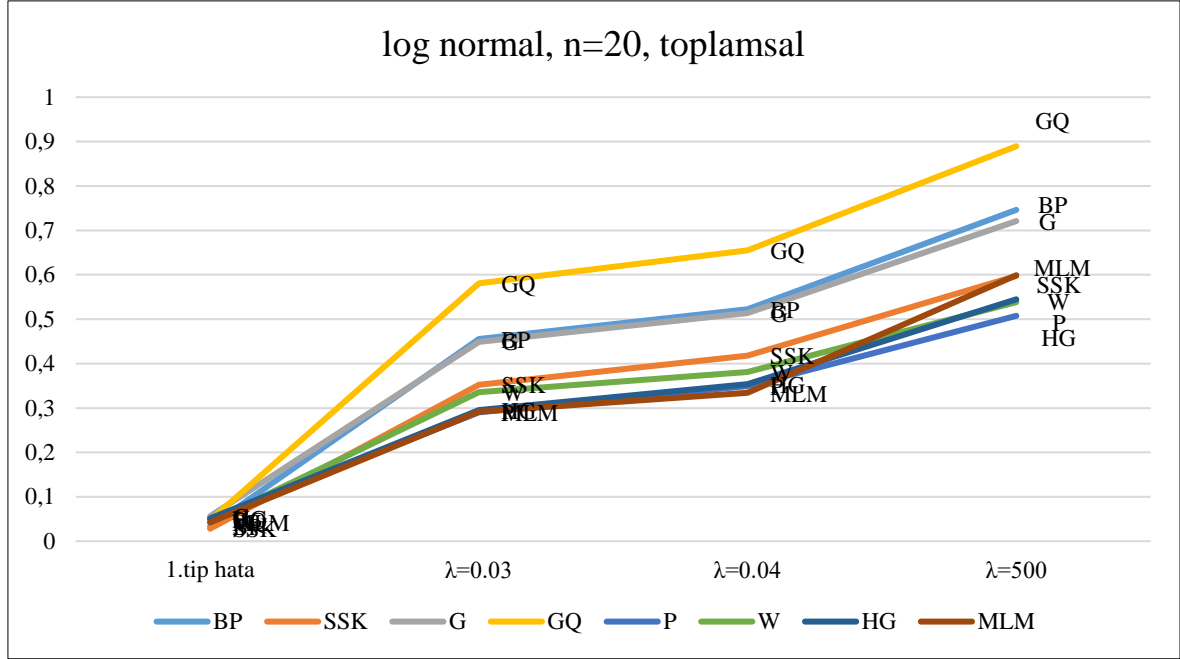
Çizelge 4.1. X değişkeni Log-normal dağılımdan geldiğinde ve n=10 için testlerin deneysel I.tip hata oranları ve güç değerleri

log normal, n=10							
Test	1,tip hata	<i>çarpımsal</i>			<i>toplamsal</i>		
		$\lambda=1,25$	$\lambda=1,35$	$\lambda=2$	$\lambda=0,03$	$\lambda=0,04$	$\lambda=500$
BP	0,0280	0,1240	0,1360	0,2580	0,1270	0,1390	0,2500
SSK	0,0310	0,1664	0,1652	0,2524	0,1516	0,1640	0,2500
G	0,0720	**	**	**	**	**	**
GQ	0,0520	0,1352	0,1478	0,2430	0,1192	0,1396	0,2460
P	0,0640	**	**	**	**	**	**
W	0,0570	0,1066	0,1168	0,1842	0,1090	0,1220	0,1830
HG	0,0690	**	**	**	**	**	**
MLM	0,0180	0,0364	0,046	0,079	0,0382	0,0420	0,0830
MLA	0,1330	**	**	**	**	**	**

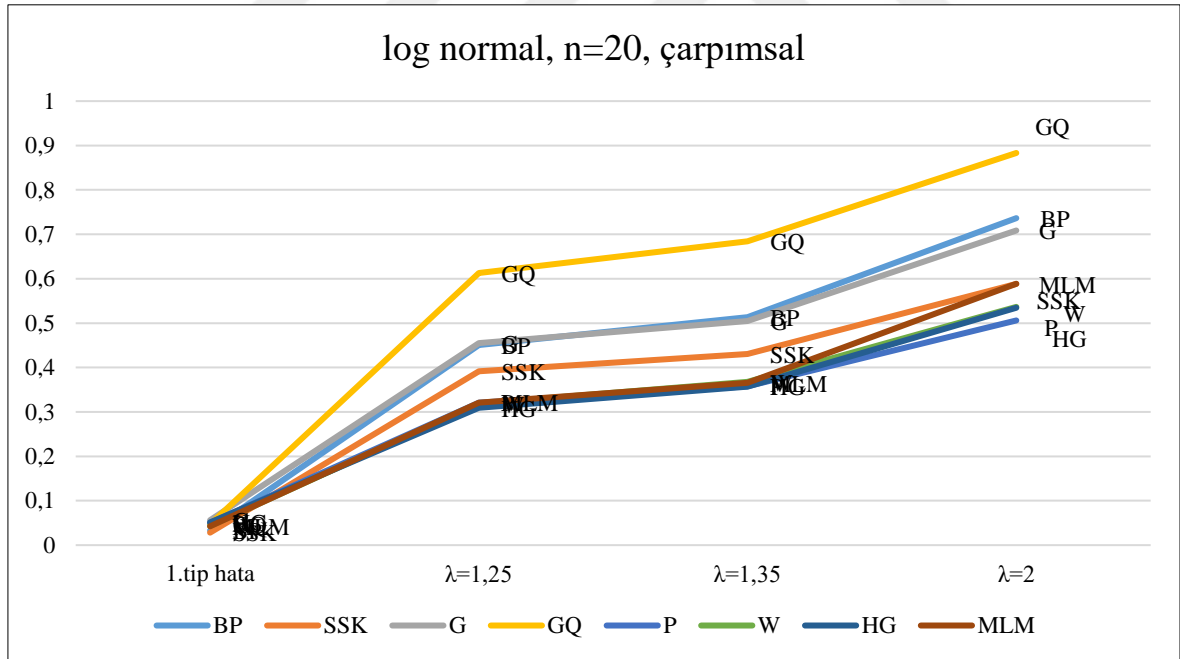
X'ler log-normal dağılımından seçildiğinde örnek çapı n= 10 iken, Çizelge 4.1'deki testlerin deneysel 1. tip hata oranları incelendiğinde G, HG ve P testinin %5 i biraz geçtiği, MLA testinin ise oldukça geçtiği görülmektedir. BP, SSK ve MLM testleri ise %5 değerinden oldukça küçük çıkmıştır. GQ ve W testlerinin ise %5 değerine oldukça yakın sonuç verdiği gözlemlenmiştir.

Çarpımsal model için testin güç değerleri incelendiğinde, SSK testinin diğer testlerden daha yüksek güç değerlerine sahip olduğu görülmektedir. Ayrıca BP testinin deneysel 1. tip hata oranının %5den oldukça küçük çıkıp testin güç değeri SSK testinden sonra iyi bir güç performansına sahip olduğu görülmektedir. SSK, BP, GQ testinin ise heterojenlik arttıkça güç değerlerinin arttığı görülmektedir.

Toplamsal model için testin güç değerleri incelendiğinde SSK testinin diğer testlerden daha yüksek güç değerlerine sahip olduğu görülmektedir. Ayrıca SSK testinin deneysel 1. tip hata oranının %5den oldukça küçük çıkıp testin güç değerinin de en iyi güç performansına sahip olduğu görülmektedir. BP ve GQ testinin ise heterojenlik arttıkça güç değerlerinin arttığı görülmektedir.



Şekil 4.3. X değişkeni Log-normal dağılımdan geldiğinde ve n=20 için testlerin toplamsal modelde deneysel I.tip hata oranları ve güç değerleri



Şekil 4.4. X değişkeni Log-normal dağılımdan geldiğinde ve n=20 için testlerin çarpımsal modelde deneysel I.tip hata oranları ve güç değerleri

Şekil 4.3 ve Şekil 4.4'de toplamsal ve çarpımsal modelde X değişkeni Log-normal dağılımdan geldiğinde n=20 için testlerin deneysel I.tip hata oranları ve güç değerleri grafiksel olarak gösterilmiştir.

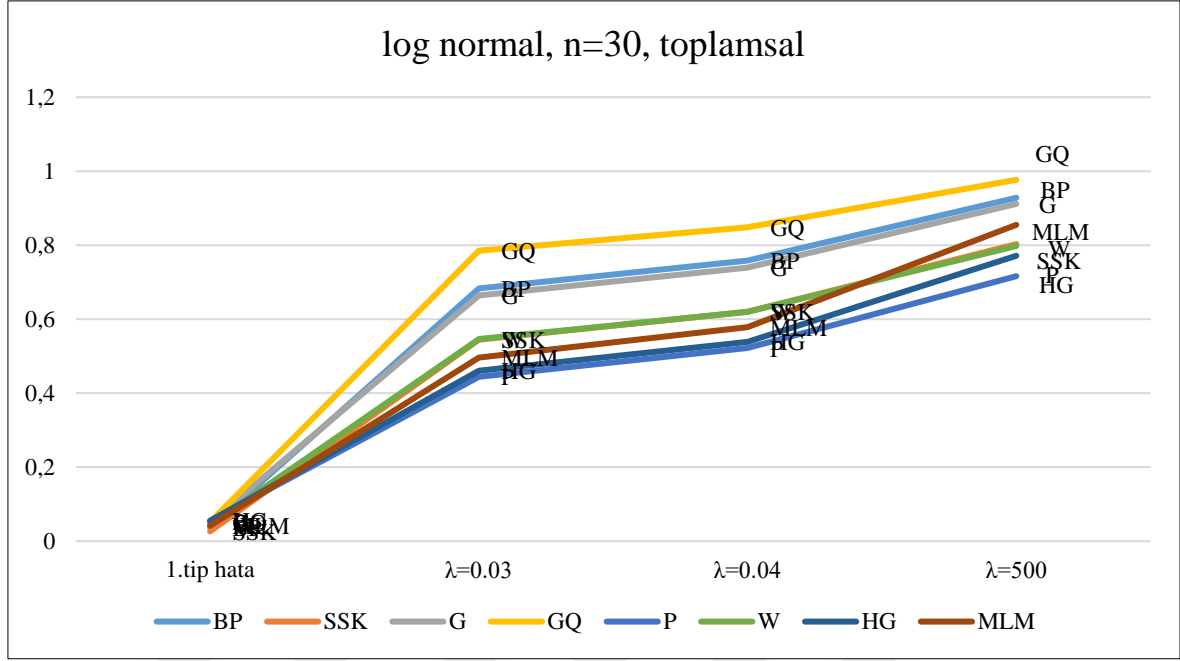
Çizelge 4.2. X değişkeni Log-normal dağılımdan geldiğinde ve n=20 için testlerin deneysel I.tip hata oranları ve güç değerleri

log normal, n=20							
Test	1,tip hata	<i>çarpımsal</i>			<i>toplamsal</i>		
		$\lambda=1,25$	$\lambda=1,35$	$\lambda=2$	$\lambda=0,03$	$\lambda=0,04$	$\lambda=500$
BP	0,0332	0,4508	0,5134	0,7364	0,4552	0,5222	0,7462
SSK	0,0282	0,3916	0,4308	0,5884	0,3524	0,4180	0,5974
G	0,0562	0,4552	0,5050	0,7086	0,4492	0,5138	0,7210
GQ	0,0470	0,6132	0,6846	0,8832	0,5806	0,6550	0,8896
P	0,0502	0,3212	0,3580	0,5060	0,2902	0,3492	0,5076
W	0,0432	0,3186	0,3680	0,5366	0,3354	0,3808	0,5384
HG	0,0512	0,3090	0,3572	0,5346	0,2956	0,3540	0,5448
MLM	0,0424	0,3210	0,3658	0,5886	0,2914	0,3340	0,5990
MLA	0,0696	**	**	**	**	**	**

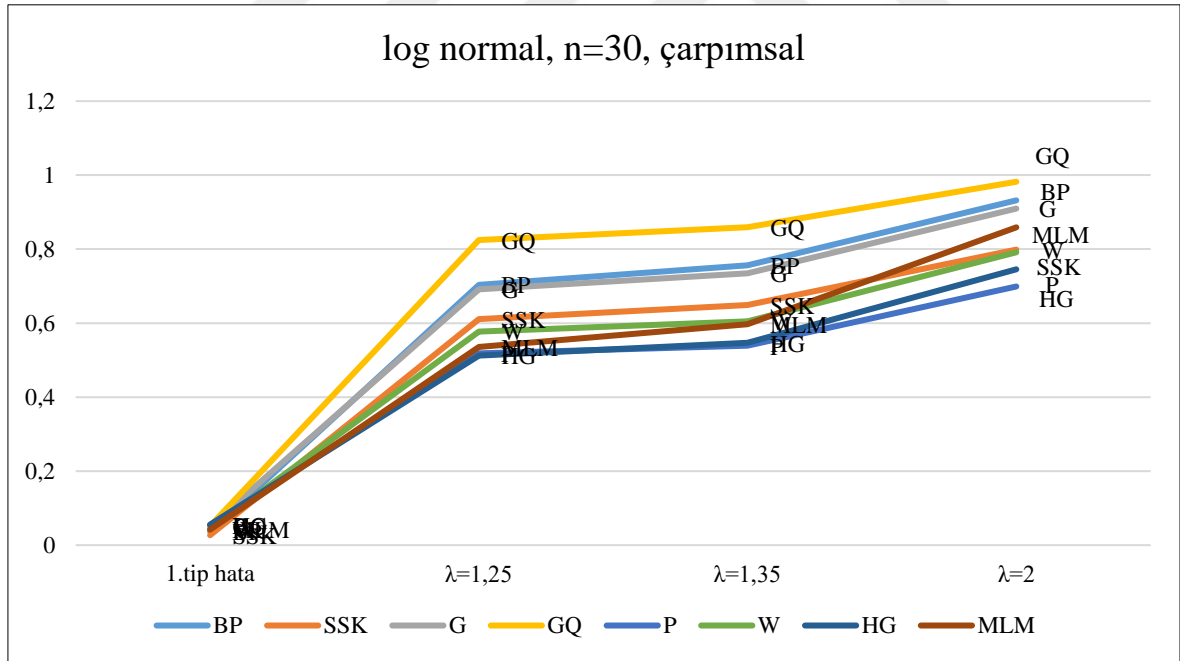
Örnek çapı n= 20 iken, testlerin deneysel 1. Tip hata oranları Çizelge 4.2’den incelendiğinde MLA testinin deneysel 1.tip hata oranının %5’i oldukça geçtiği görülmektedir. BP ve SSK testinin ise deneysel 1.tip hata oranlarının %5 değerinden oldukça küçük çıktığı görülmektedir. G, GQ, P, W, HG ve MLM testlerinin diğer testlere göre %5 değerine daha yakın sonuç verdiği gözlemlenmiştir.

Testlerin güç değerleri incelendiğinde çarpımsal model için, GQ testinin diğer testlerden daha yüksek güç değerlerine sahip olduğu görülmektedir. Ayrıca BP testinin deneysel 1. Tip hata oranının %5den oldukça küçük olduğu, testin güç değeri ise $\lambda=1,25$ durumu hariç GQ testinden sonra en yüksek güç performansına sahip olduğu görülmektedir. Genel olarak heterojenlik arttıkça testlerin güç değerlerinin arttığı da görülmektedir. Ayrıca beklenildiği gibi örnek çapı da arttıkça testlerin güç değerlerinde Çizelge 4.1’e göre artış görülmüştür.

Toplamsal model için testin güç değerleri incelendiğinde çarpımsal modeldekine benzer sonuçların elde edildiği görülmektedir.



Şekil 4.5. X değişkeni Log-normal dağılımdan geldiğinde ve n=30 için testlerin toplamsal modelde deneysel I.tip hata oranları ve güç değerleri



Şekil 4.6. X değişkeni Log-normal dağılımdan geldiğinde ve n=30 için testlerin çarpımsal modelde deneysel I.tip hata oranları ve güç değerleri

Şekil 4.5 ve Şekil 4.6'da toplamsal ve çarpımsal modelde X değişkeni Log-normal dağılımdan geldiğinde n=30 için testlerin deneysel I.tip hata oranları ve güç değerleri grafiksel olarak gösterilmiştir.

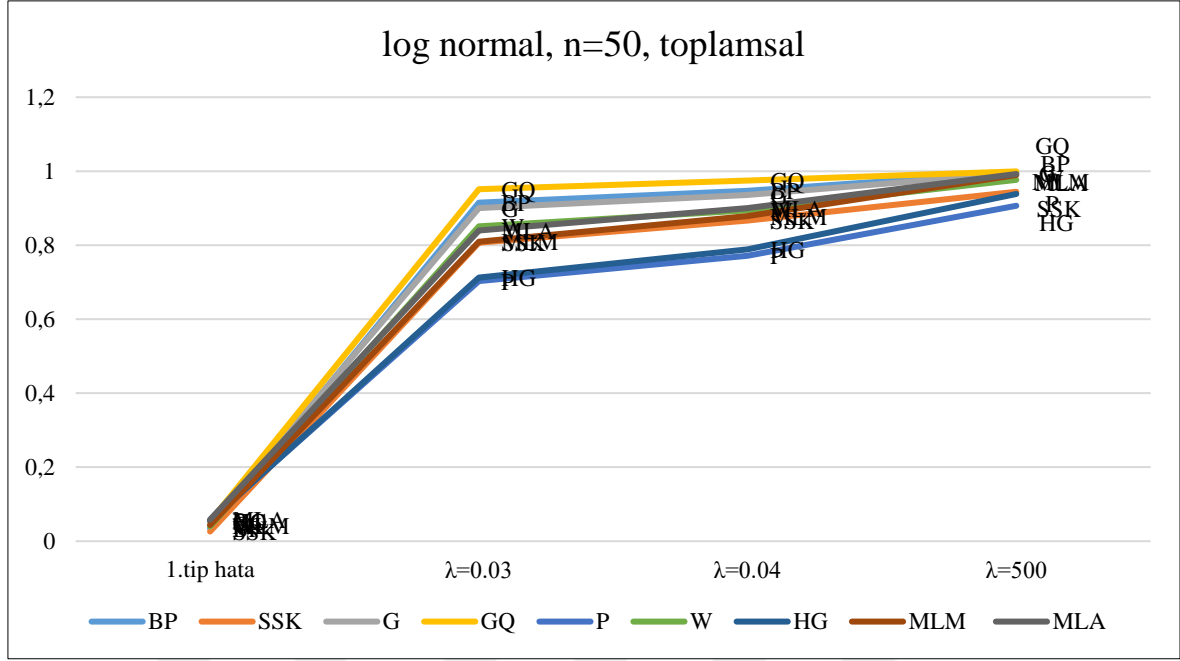
Çizelge 4.3. X değişkeni Log-normal dağılımdan geldiğinde ve n=30 için testlerin deneysel I.tip hata oranları ve güç değerleri

log normal, n=30							
Test	1,tip hata	<i>çarpımsal</i>			<i>toplamsal</i>		
		$\lambda=1,25$	$\lambda=1,35$	$\lambda=2$	$\lambda=0,03$	$\lambda=0,04$	$\lambda=500$
BP	0,0370	0,7038	0,7564	0,9318	0,6828	0,7584	0,9282
SSK	0,0268	0,6110	0,6492	0,7986	0,5446	0,6202	0,8028
G	0,0510	0,6912	0,7344	0,9096	0,6642	0,7392	0,9116
GQ	0,0514	0,8248	0,8596	0,9820	0,7854	0,8482	0,9766
P	0,0546	0,5180	0,5388	0,6990	0,4446	0,5220	0,7162
W	0,0438	0,5774	0,6044	0,7914	0,5468	0,6204	0,7978
HG	0,0552	0,5126	0,5466	0,7454	0,4610	0,5390	0,7714
MLM	0,0420	0,5356	0,5974	0,8588	0,4960	0,5782	0,8550
MLA	0,0698	**	**	**	**	**	**

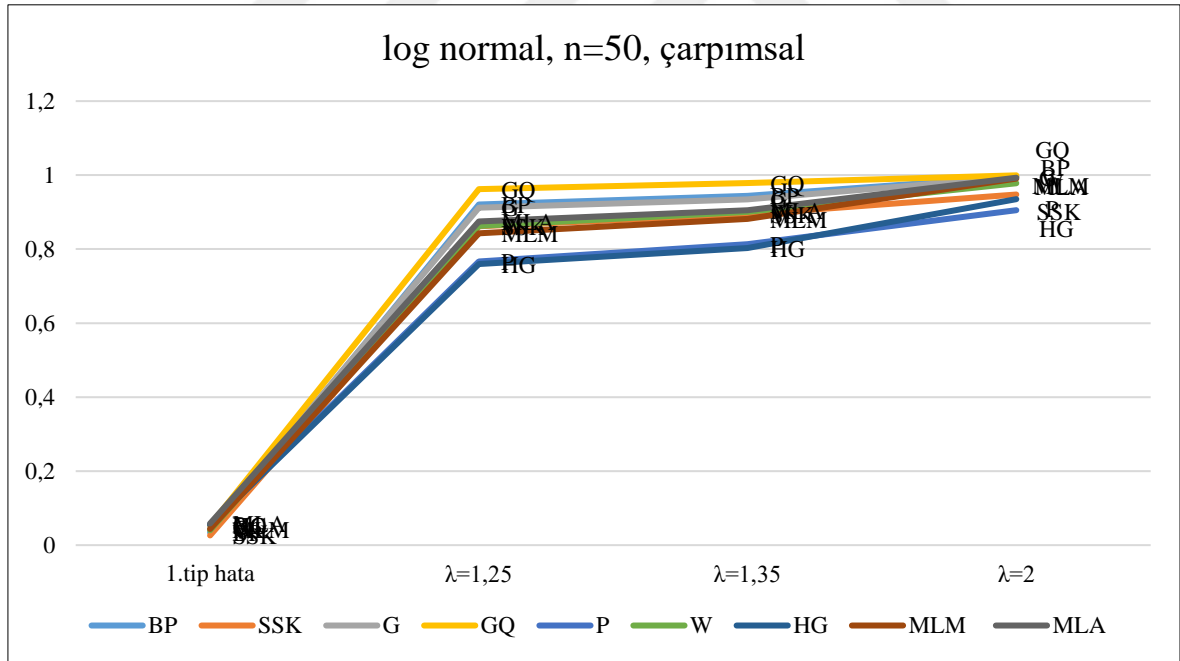
Örnek çapı n= 30 iken Çizelge 4.3'den görüldüğü gibi MLA testinin deneysel 1.tip hata oranının %5'i aştığı %6 civarında değer aldığı görülmektedir. SSK ve BP testi ise %5 değerinden oldukça küçük çıkmıştır. G, GQ, P, W, HG, MLM testlerinin ise %5 değerine yakın sonuç verdiği gözlemlenmiştir.

Testin güç değerleri incelendiğinde, çarpımsal model için GQ testinin diğer testlerden oldukça yüksek güç değerlerine sahip olduğu görülmektedir. Daha sonra BP ve ardından G testleri ise diğer testlere göre daha yüksek güç değerlerine sahip olduğu görülmüştür. Ayrıca BP testinin deneysel 1. Tip hata oranının %5den oldukça küçük çıkıp testin güç değeri ise GQ testinden sonra en yüksek güç performansına sahip olduğu görülmektedir. Genel olarak Çizelge 4.3'den görüldüğü gibi heterojenlik arttıkça testlerin güç değerlerinin arttığı da görülmektedir.

Toplamsal model için testin güç değerleri incelendiğinde çarpımsal modeldekine benzer sonuçların elde edildiği görülmektedir.



Şekil 4.7. X değişkeni Log-normal dağılımdan geldiğinde ve n=50 için testlerin toplamsal modelde deneysel I.tip hata oranları ve güç değerleri



Şekil 4.8. X değişkeni Log-normal dağılımdan geldiğinde ve n=50 için testlerin çarpımsal modelde deneysel I.tip hata oranları ve güç değerleri

Şekil 4.7 ve Şekil 4.8'de toplamsal ve çarpımsal modelde X değişkeni Log-normal dağılımdan geldiğinde n=50 için testlerin deneysel I.tip hata oranları ve güç değerleri grafiksel olarak gösterilmiştir.

Çizelge 4.4. X değişkeni Log-normal dağılımdan geldiğinde ve n=50 için testlerin deneysel 1.tip hata oranları ve güç değerleri

log normal, n=50							
Test	1,tip hata	<i>çarpımsal</i>			<i>toplamsal</i>		
		$\lambda=1,25$	$\lambda=1,35$	$\lambda=2$	$\lambda=0,03$	$\lambda=0,04$	$\lambda=500$
BP	0,0342	0,9208	0,9444	0,9962	0,9148	0,9472	0,9972
SSK	0,0262	0,8614	0,8938	0,9476	0,8066	0,8666	0,9444
G	0,0452	0,9120	0,9348	0,9954	0,9006	0,9358	0,9952
GQ	0,0534	0,9618	0,9790	0,9994	0,9510	0,9746	0,9994
P	0,0568	0,7670	0,8132	0,9052	0,7030	0,7718	0,9066
W	0,0402	0,8624	0,9004	0,9780	0,8512	0,8942	0,9766
HG	0,0546	0,7600	0,8028	0,9350	0,7128	0,7886	0,9384
MLM	0,0436	0,8424	0,8814	0,9906	0,8094	0,8786	0,9896
MLA	0,0584	0,8746	0,9052	0,9930	0,8400	0,9004	0,9924

Örnek çapı n= 50 iken Çizelge 4.4’de verilen testlerin deneysel 1. Tip hata oranları incelendiğinde SSK ve BP testi %5 değerinden oldukça küçük çıkmıştır. G, GQ, P, W, HG, MLM, MLA testlerinin ise %5 değerine yakın sonuç verdikleri gözlemlenmiştir.

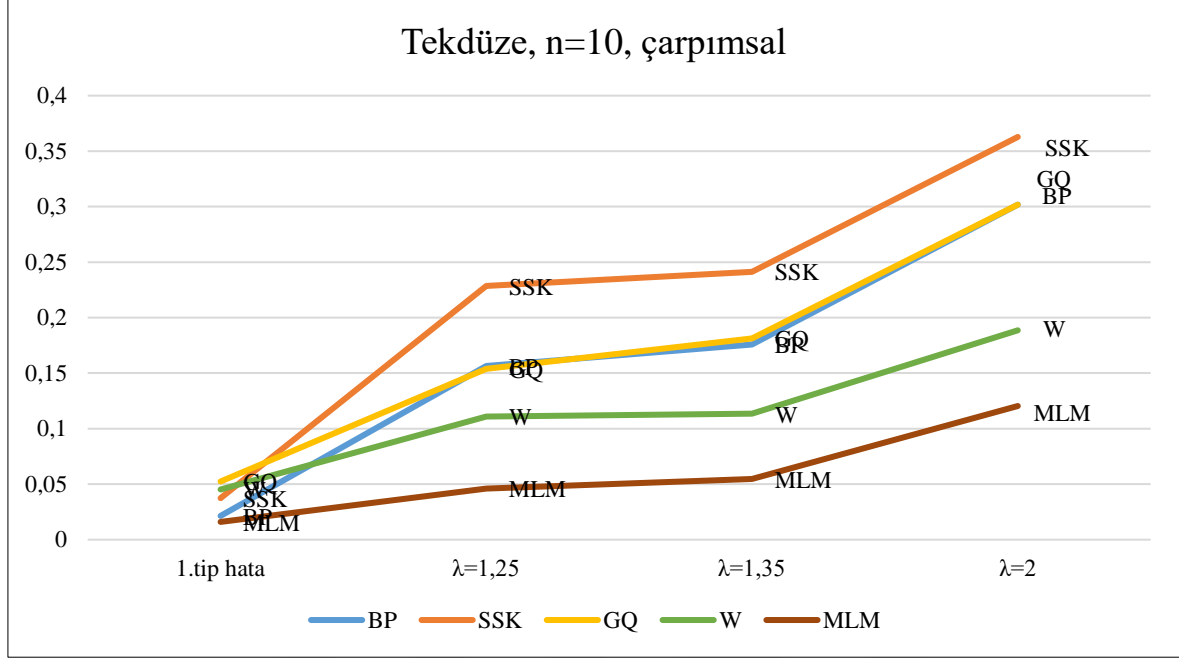
Çarpımsal model için testlerin güç değerleri incelendiğinde GQ testinin diğer testlerden daha yüksek güç değerlerine sahip olduğu görülmektedir. Heterojenlik arttıkça GQ testinin yanı sıra, BP, SSK, G, W, MLM, MLA testlerinin güç değerlerinin birbirine oldukça yaklaştığı görülmektedir. Ayrıca SSK testinin deneysel 1. tip hata oranının %5den oldukça küçük çıkıp testin güç değerinin ise iyi bir güç performansına sahip olduğu görülmektedir.

Toplamsal model için testin güç değerleri incelendiğinde çarpımsal modeldekine benzer sonuçların elde edildiği görülmektedir.

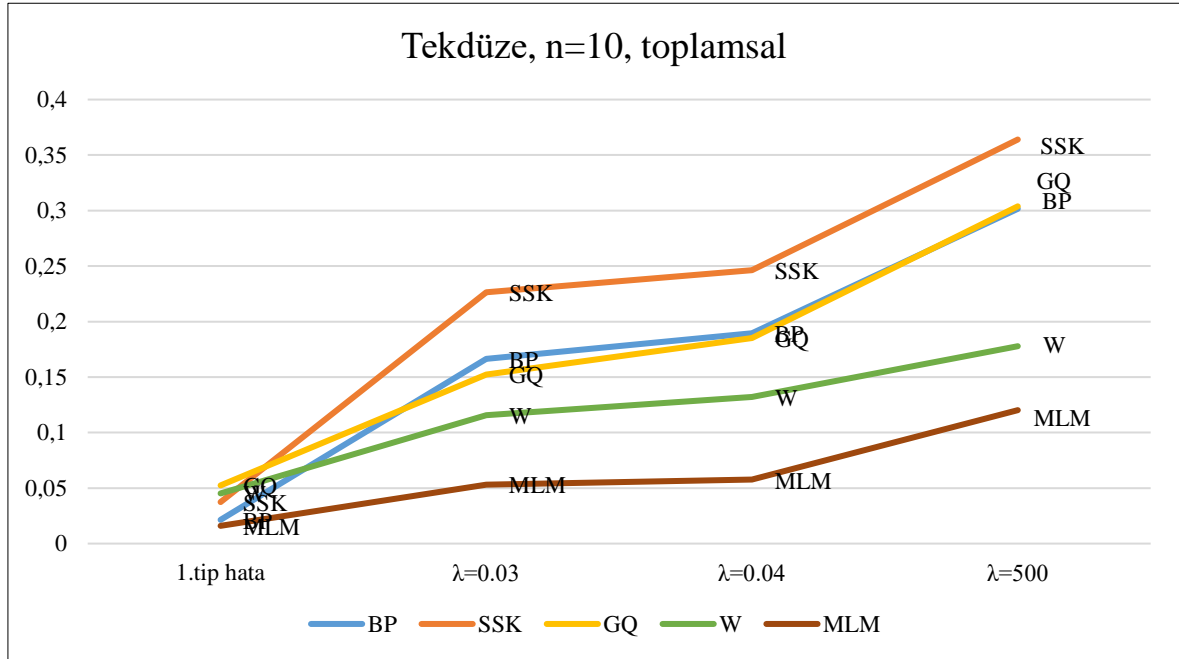
Genel olarak X’ler log-normal dağılımından seçildiğinde testlerin deneysel 1. tip hata oranları ve güç değerleri değerlendirildiğinde, BP, SSK, GQ, W, MLM testlerinin tüm örnek çaplarında %5 e yakın sonuçlar verdiği görülmektedir. Küçük örnek çaplarında BP, SSK, GQ, W, MLM testleri, orta örnek çaplarında BP, SSK, G, GQ, P, W, HG, MLM testleri ve büyük örnek çaplarında incelenen testlerin daha iyi sonuçlar verdiği gözlemlenmiştir.

Testlerin güç değerleri incelendiğinde küçük örnek çaplarında SSK testi, orta örnek çaplarında GQ testi ve büyük örnek çaplarında GQ ve BP testlerinin diğer testlere göre daha iyi performans gösterdiği söylenebilir. BP ve SSK testleri deneysel 1. Tip hatasının %5den küçük olup testin gücü değerinin yüksek çıkması bu testlerin de oldukça iyi olduğunu

göstermiştir. Ayrıca genel olarak heterojenlik ve örnek çapı arttıkça testlerin güç performanslarının da arttığı gözlenmektedir.



Şekil 4.9. X değişkeni tekdüze dağılımdan geldiğinde ve n=10 için testlerin çarpımsal modelde deneysel I.tip hata oranları ve güç değerleri



Şekil 4.10. X değişkeni tekdüze dağılımdan geldiğinde ve n=10 için testlerin toplamsal modelde deneysel I.tip hata oranları ve güç değerleri

Şekil 4.9 ve Şekil 4.10'da toplamsal ve çarpımsal modelde X değişkeni tekdüze dağılımdan geldiğinde n=10 için testlerin deneysel I.tip hata oranları ve güç değerleri grafiksel olarak gösterilmiştir.

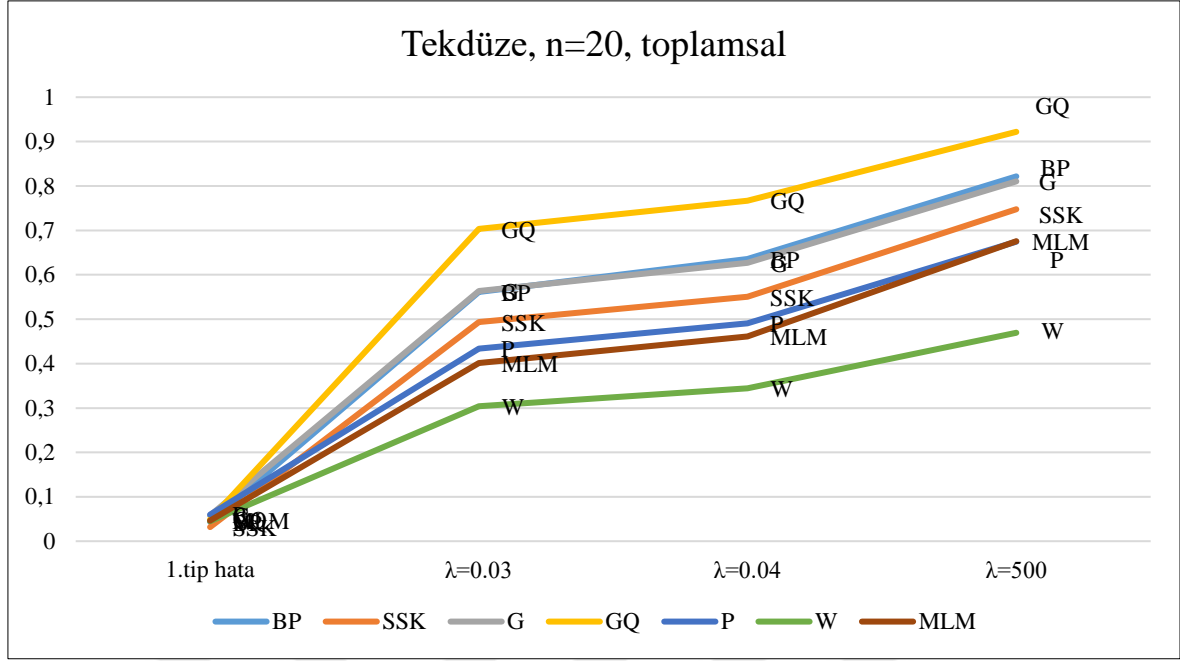
Çizelge 4.5. X değişkeni tekdüze dağılımdan geldiğinde ve n=10 için testlerin deneysel I.tip hata oranları ve güç değerleri

Tekdüze, n=10							
Test	1.tip hata	<i>çarpımsal</i>			<i>toplamsal</i>		
		$\lambda=1,25$	$\lambda=1,35$	$\lambda=2$	$\lambda=0,03$	$\lambda=0,04$	$\lambda=500$
BP	0,0214	0,1566	0,1758	0,3018	0,1664	0,1896	0,3016
SSK	0,0374	0,2286	0,2414	0,3628	0,2264	0,2462	0,3640
G	0,0628	**	**	**	**	**	**
GQ	0,0524	0,1538	0,1812	0,3018	0,1522	0,1852	0,3038
P	0,0658	**	**	**	**	**	**
W	0,0452	0,1108	0,1134	0,1886	0,1156	0,1322	0,1778
HG	0,0678	**	**	**	**	**	**
MLM	0,0160	0,0460	0,0546	0,1204	0,0532	0,0576	0,1202
MLA	0,1156	**	**	**	**	**	**

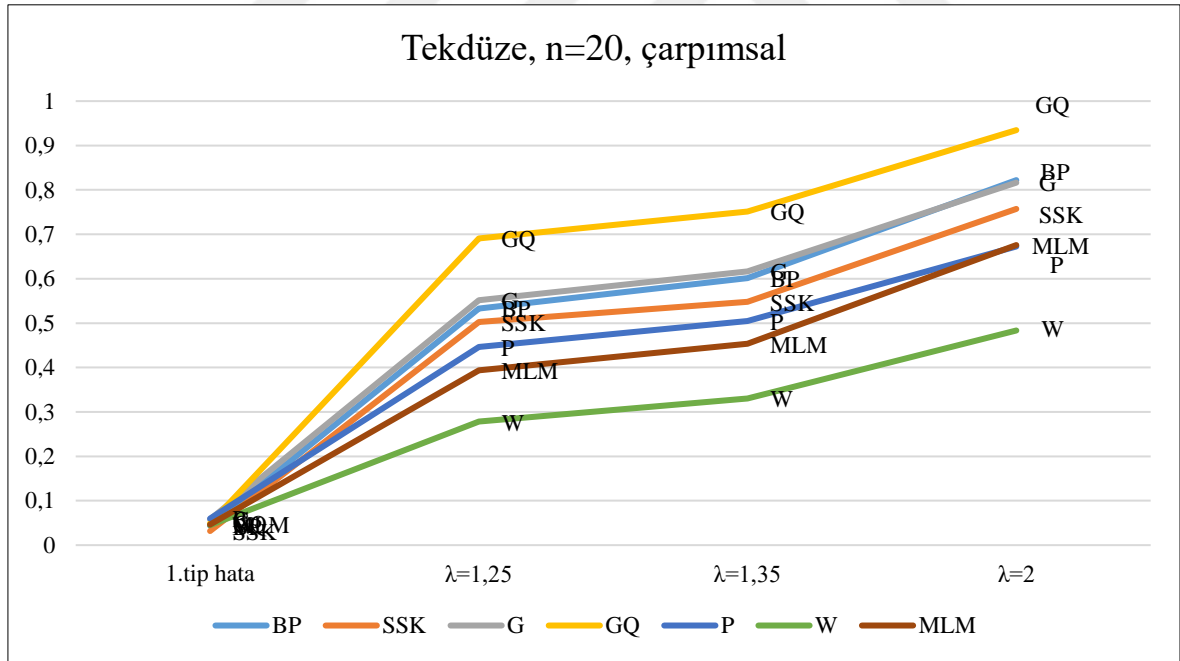
X'ler tekdüze dağılımından seçildiğinde örnek çapı n= 10 iken, testlerin deneysel 1. tip hata oranları incelendiğinde G, P, HG testlerinin deneysel 1.tip hata oranlarının %5 i biraz geçtiği, MLA testinin ise oldukça geçtiği görülmektedir. BP ve MLM testleri ise %5 değerinden oldukça küçük çıkmıştır. SSK, GQ, W testlerinin ise %5 değerine oldukça yakın sonuçlar verdiği gözlemlenmiştir.

Çarpımsal model için testin güç değerleri incelendiğinde SSK testinin diğer testlerden daha yüksek güç değerlerine sahip olduğu görülmektedir. Ayrıca BP testinin deneysel 1. tip hata oranının %5den oldukça küçük çıkıp testin güç değerinin ise SSK testinden sonra iyi bir güç performansına sahip olduğu görülmektedir. Örnek çapı küçükken genel olarak SSK daha sonra BP ve GQ testlerinin güç değerlerinin diğer testlere göre daha iyi çıktığı gözlenmiştir. Genel olarak heterojenlik arttıkça testlerin güç değerlerinin arttığı da görülmektedir.

Toplamsal model için testin güç değerleri incelendiğinde çarpımsal modeldekine benzer sonuçların elde edildiği görülmektedir.



Şekil 4.11. X değişkeni tekdüze dağılımdan geldiğinde ve n=20 için testlerin toplamsal modelde deneysel I.tip hata oranları ve güç değerleri



Şekil 4.12. X değişkeni tekdüze dağılımdan geldiğinde ve n=20 için testlerin çarpımsal modelde deneysel I.tip hata oranları ve güç değerleri

Şekil 4.11 ve Şekil 4.12'da toplamsal ve çarpımsal modelde X değişkeni tekdüze dağılımdan geldiğinde n=20 için testlerin deneysel I.tip hata oranları ve güç değerleri grafiksel olarak gösterilmiştir.

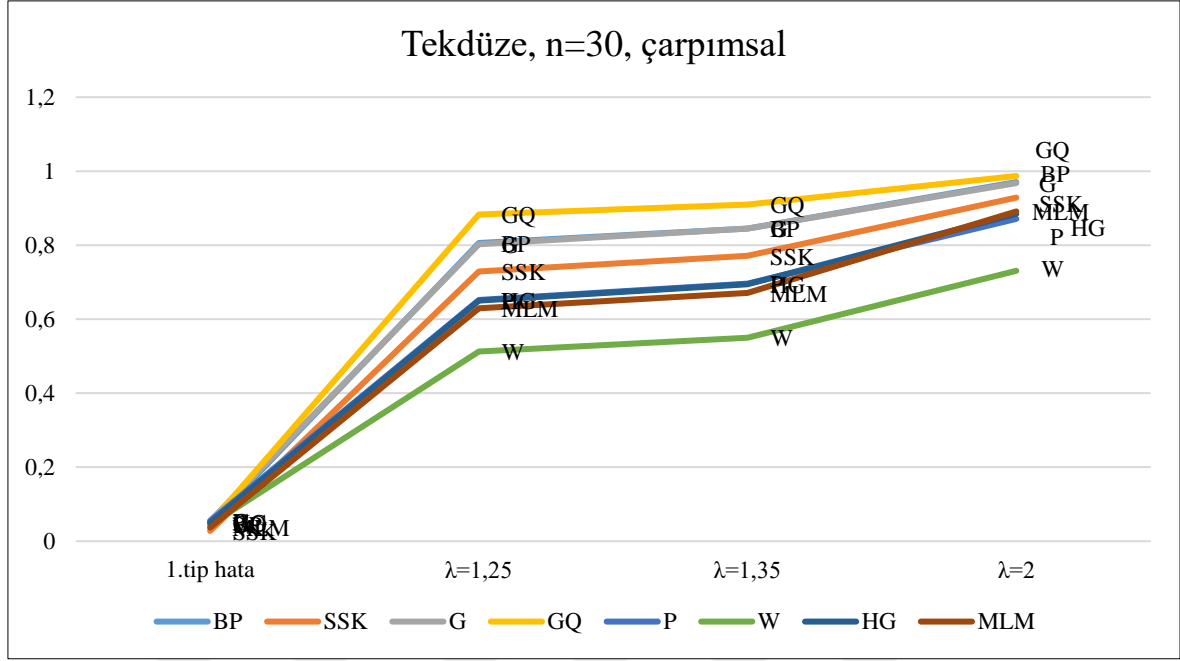
Çizelge 4.6. X değişkeni tekdüze dağılımdan geldiğinde ve n=20 için testlerin deneysel I.tip hata oranları ve güç değerleri

Tekdüze, n=20							
Test	1.tip hata	<i>çarpımsal</i>			<i>toplamsal</i>		
		$\lambda=1,25$	$\lambda=1,35$	$\lambda=2$	$\lambda=0,03$	$\lambda=0,04$	$\lambda=500$
BP	0,0436	0,5332	0,6014	0,8220	0,5616	0,6356	0,8216
SSK	0,0318	0,5026	0,5482	0,7572	0,4934	0,5502	0,7476
G	0,0580	0,5516	0,6166	0,8164	0,5638	0,6268	0,8104
GQ	0,0490	0,6904	0,7512	0,9346	0,7032	0,7670	0,9220
P	0,0598	0,4460	0,5050	0,6726	0,4334	0,4910	0,6746
W	0,0450	0,2780	0,3300	0,4834	0,3038	0,3446	0,4694
HG	0,0608	**	**	**	**	**	**
MLM	0,0464	0,3934	0,4534	0,6758	0,4010	0,4608	0,6756
MLA	0,0740	**	**	**	**	**	**

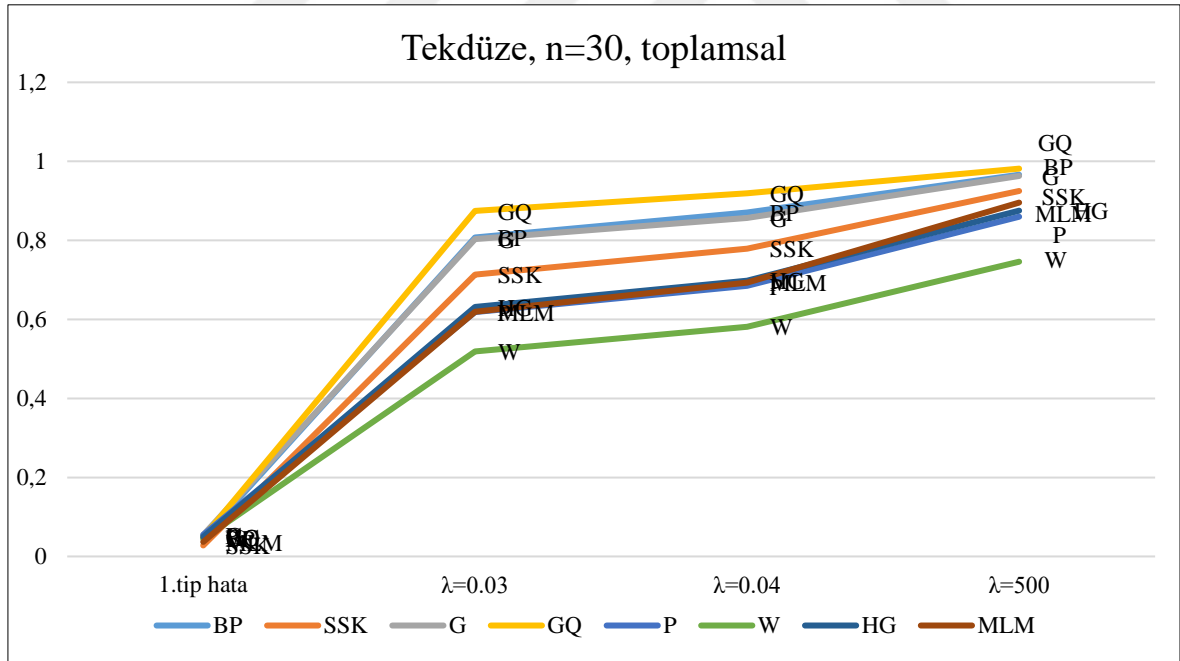
X'ler tekdüze dağılımından seçildiğinde örnek çapı n= 20 iken, testlerin deneysel 1. tip hata oranları incelendiğinde HG, MLA testlerinin %5 i aştığı görülmektedir. BP, SSK, G, GQ, P, W, MLM testleri %5 değerine oldukça yakın sonuç verdiği gözlemlenmiştir.

Çarpımsal model için testin güç değerleri incelendiğinde GQ testinin diğer testlerden daha yüksek güç değerlerine sahip olduğu görülmektedir. Bu testten sonrada G ve BP testlerinin yüksek güç değerlerine sahip olduğu gözlenmiştir. Genel olarak örnek çapı ve heterojenlik arttıkça testlerin güç değerlerinin de arttığı görülmektedir.

Toplamsal model için testin güç değerleri incelendiğinde çarpımsal modeldekine benzer sonuçların elde edildiği görülmektedir.



Şekil 4.13. X değişkeni tekdüze dağılımdan geldiğinde ve n=30 için testlerin çarpımsal modelde deneysel I.tip hata oranları ve güç değerleri



Şekil 4.14. X değişkeni tekdüze dağılımdan geldiğinde ve n=30 için testlerin toplamsal modelde deneysel I.tip hata oranları ve güç değerleri

Şekil 4.13 ve Şekil 4.14'da toplamsal ve çarpımsal modelde X değişkeni tekdüze dağılımdan geldiğinde n=30 için testlerin deneysel I.tip hata oranları ve güç değerleri grafiksel olarak gösterilmiştir.

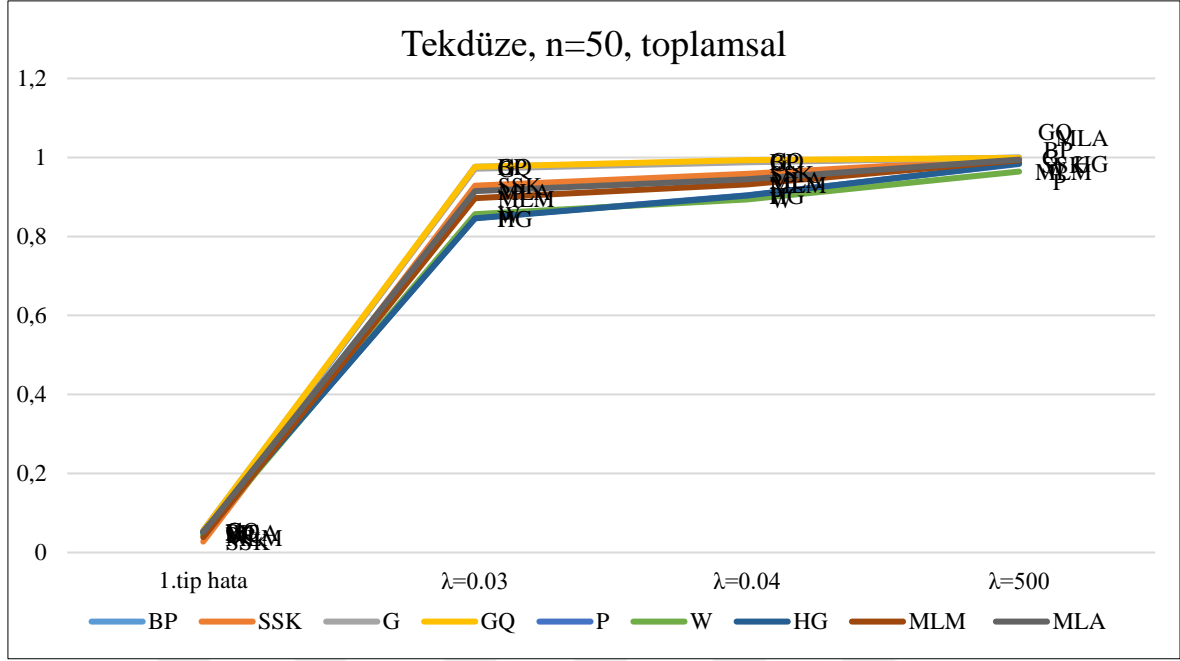
Çizelge 4.7. X değişkeni tekdüze dağılımdan geldiğinde ve n=30 için testlerin deneysel I.tip hata oranları ve güç değerleri

Tekdüze, n=30							
Test	1.tip hata	<i>çarpımsal</i>			<i>toplamsal</i>		
		$\lambda=1,25$	$\lambda=1,35$	$\lambda=2$	$\lambda=0,03$	$\lambda=0,04$	$\lambda=500$
BP	0,0470	0,8062	0,8452	0,9706	0,8074	0,8704	0,9662
SSK	0,0278	0,7294	0,7710	0,9286	0,7136	0,7790	0,9250
G	0,0544	0,8026	0,8450	0,9678	0,8034	0,8570	0,9630
GQ	0,0482	0,8834	0,9102	0,9870	0,8746	0,9194	0,9816
P	0,0548	0,6518	0,6960	0,8712	0,6186	0,6852	0,8598
W	0,0444	0,5130	0,5500	0,7308	0,5188	0,5816	0,7460
HG	0,0498	0,6506	0,6940	0,8846	0,6312	0,6982	0,8752
MLM	0,0368	0,6294	0,6708	0,8912	0,6194	0,6932	0,8956
MLA	0,0620	**	**	**	**	**	**

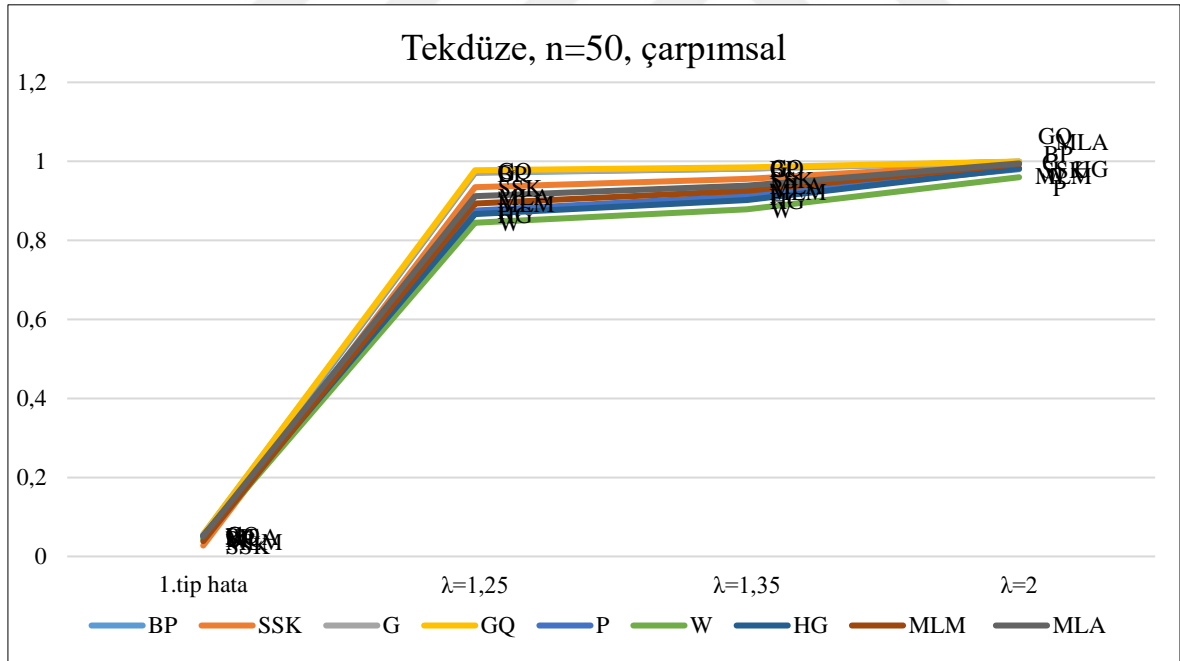
X'ler tekdüze dağılımından seçildiğinde örnek çapı n= 30 iken, testlerin deneysel 1. tip hata oranları incelendiğinde MLA testinin %5 i aştığı görülmektedir. SSK testi ise %5 değerinden oldukça küçük çıkmıştır. BP, G, GQ, P, W, HG, MLM testlerinin ise %5 değerine oldukça yakın sonuçlar verdiği gözlemlenmiştir.

Çarpımsal model için testin güç değerleri incelendiğinde GQ testinin diğer testlerden daha yüksek güç değerlerine sahip olduğu görülmektedir. Bu testten sonrada G ve BP testlerinin yüksek güç değerlerine sahip olduğu gözlenmiştir. Ayrıca SSK testinin deneysel 1. tip hata oranının %5den oldukça küçük çıkıp testin güç değeri ise iyi bir güç performansına sahip olduğu görülmektedir.

Toplamsal model için testin güç değerleri incelendiğinde çarpımsal modeldekine benzer sonuçların elde edildiği görülmektedir.



Şekil 4.15. X değişkeni tekdüze dağılımdan geldiğinde ve n=50 için testlerin toplamsal modelde deneysel I.tip hata oranları ve güç değerleri



Şekil 4.16. X değişkeni tekdüze dağılımdan geldiğinde ve n=50 için testlerin çarpımsal modelde deneysel I.tip hata oranları ve güç değerleri

Şekil 4.15 ve Şekil 4.16'da toplamsal ve çarpımsal modelde X değişkeni tekdüze dağılımdan geldiğinde n=50 için testlerin deneysel I.tip hata oranları ve güç değerleri grafiksel olarak gösterilmiştir.

Çizelge 4.8. X değişkeni tekdüze dağılımdan geldiğinde ve n=50 için testlerin deneysel I.tip hata oranları ve güç değerleri

Tekdüze, n=50							
Test	1.tip hata	<i>çarpımsal</i>			<i>toplamsal</i>		
		$\lambda=1,25$	$\lambda=1,35$	$\lambda=2$	$\lambda=0,03$	$\lambda=0,04$	$\lambda=500$
BP	0,0472	0,9708	0,9814	0,9992	0,9762	0,9894	0,9994
SSK	0,0276	0,9342	0,9556	0,9950	0,9282	0,9586	0,9974
G	0,0558	0,9728	0,9818	0,9992	0,9714	0,9874	0,9998
GQ	0,0556	0,9776	0,9850	0,9996	0,9766	0,9930	0,9988
P	0,0528	0,8752	0,9128	0,9802	0,8468	0,9042	0,9834
W	0,0442	0,8448	0,8784	0,9598	0,8562	0,8936	0,9642
HG	0,0520	0,8664	0,9024	0,9826	0,8460	0,9038	0,9848
MLM	0,0388	0,8934	0,9254	0,9932	0,8966	0,9316	0,9928
MLA	0,0512	0,9122	0,9388	0,9948	0,9146	0,9450	0,9946

X'ler tekdüze dağılımından seçildiğinde örnek çapı n= 50 iken, testlerin deneysel 1. tip hata oranları incelendiğinde SSK testi %5'den oldukça küçük çıkmıştır. BP, G, GQ, P, W, HG, MLM ve MLA testlerinin ise %5 değerine oldukça yakın sonuçlar verdiği gözlemlenmiştir.

Çarpımsal model için testin güç değerleri incelendiğinde GQ testinin diğer testlerden daha yüksek güç değerlerine sahip olduğu görülmektedir. Daha sonra G ve BP testlerinin güçleri diğer testlerden daha yüksek çıkmıştır. Ayrıca SSK testinin deneysel 1. tip hata oranının %5den oldukça küçük çıkıp testin güç değeri ise iyi bir güç performansına sahip olduğu görülmektedir. Heterojenlik arttıkça incelenen tüm testlerin güç değerlerinin arttığı arttığı gözlemlenmiştir.

Toplamsal model için testin güç değerleri incelendiğinde GQ ve G testleri diğer testlerden daha yüksek güç değerlerine sahip olduğu görülmektedir. Heterojenlik arttıkça BP, SSK, GQ, P, W, HG, MLM ve MLA testleri çok iyi bir güç performansı sergilediği görülmüştür.

Genel olarak X'ler tekdüze dağılımından seçildiğinde deneysel 1. tip hata oranları incelendiğinde BP, SSK, GQ, W ve MLM testlerinin tüm örnek çaplarında %5'e yakın sonuçlar verdikleri görülmektedir.

Testlerin güç değerleri incelendiğinde, küçük örnek çapında SSK testi, orta ve büyük örnek çaplarında ise GQ testinin daha yüksek güç değerlerine sahip olduğu söylenebilir.

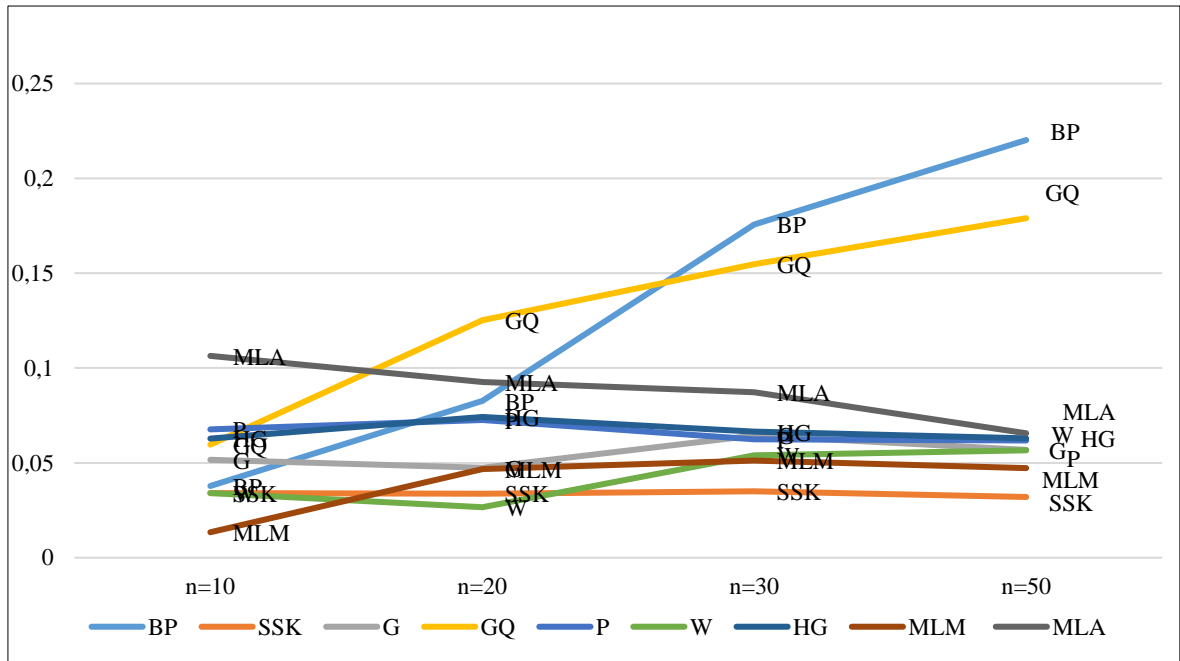
Genel olarak heterojenlik durumlarına göre testler karşılaştırıldığında çarpımsal ve toplamsal modellerde testlerin benzer sonuçlar verdiği gözlenmiştir.

X'lerin Log-normal dağılımdan geldiği durumla tekdüze dağılımından geldiği durumla karşılaştırıldığında da yine benzer durumların ortaya çıktığı gözlenmektedir.

Genel olarak örnek çapı ve heterojenlik arttıkça testlerin güç performanslarında da artış olduğu görülmektedir.

Tüm durumlar göz önüne alındığında GQ testinin diğer testlere göre hem deneysel 1.tip hata oranı bakımından hem de güç bakımından daha iyi performans gösterdiği söylenebilir.

Ayrıca hata terimleri normal dağılmadığında testlerin performansı da incelenmiştir. Buna göre hata teriminin dağılımı 3 ve 5 serbest dereceli student-t dağılımından üretilmiş ve buna göre testlerin deneysel 1.tip hata oranları elde edilmiştir. Bu sonuçlar şekil ve tablo olarak aşağıdaki gibi verilmiştir. Buna göre X'ler log-normal dağılımından ve hata terimleri t_3 dağılımından seçildiğinde testlerin deneysel 1. tip hata oranları Şekil 4.17 ve Çizelge 4.9'da verilmiştir.



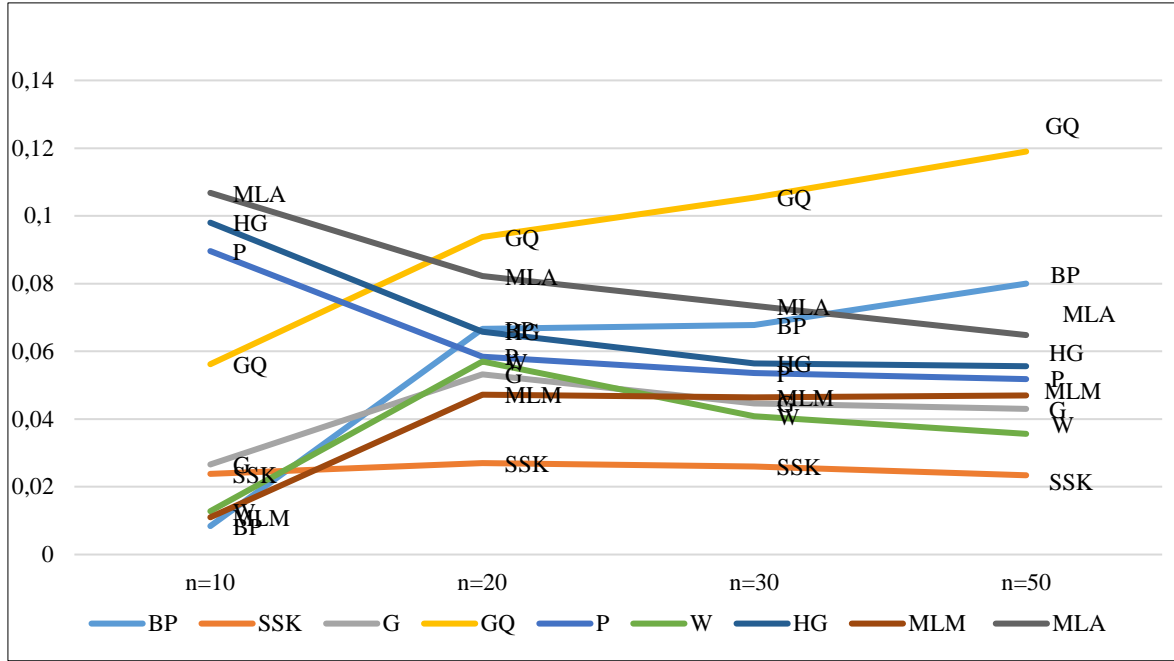
Şekil 4.17. Hata terimleri t_3 dağılımına sahip olduğunda testlerin deneysel 1.tip hata oranları

Çizelge 4.9. Hata terimleri t_3 dağılımına sahip olduğunda testlerin deneysel 1.tip hata oranları

Test	n=10	n=20	n=30	n=50
BP	0,0378	0,0826	0,1756	0,2202
SSK	0,0340	0,0338	0,0350	0,0320
G	0,0516	0,0474	0,0644	0,0568
GQ	0,0596	0,1252	0,1548	0,1790
P	0,0676	0,0726	0,0624	0,0618
W	0,0340	0,0266	0,0540	0,0566
HG	0,0628	0,0742	0,0664	0,0630
MLM	0,0134	0,0468	0,0512	0,0472
MLA	0,1064	0,0926	0,0872	0,0656

X 'ler log-normal dağılımından, hata terimleri t_3 dağılımından seçildiğinde testlerin deneysel 1. tip hata oranları incelendiğinde küçük örnek çaplarında SSK, BP, MLM ve W testlerinin 1.tip hata oranları %5 değerinden oldukça küçük, MLA, HG ve P testlerinin 1.tip hata oranlarının ise %5 büyük çıktığı görülmektedir. G, GQ testlerinin 1.tip hata oranları ise %5 değerine oldukça yakın sonuç verdiği gözlemlenmiştir. Orta örnek çaplarında SSK ve W %5 değerinden oldukça küçük ve GQ, P, HG, MLA testleri %5 değerinden oldukça büyük çıkmıştır. G, MLM testleri %5 değerine oldukça yakın sonuç verdiği gözlemlenmiştir. Büyük örnek çaplarında SSK testi %5 değerinden oldukça küçük, BP, GQ, P, HG, MLA testleri ise %5 değerinden oldukça büyük çıkmıştır. G, W ve MLM testlerinin %5 değerine oldukça yakın sonuç verdiği gözlemlenmiştir.

X 'ler log-normal dağılımından, hata terimleri t_5 dağılımından seçildiğinde testlerin deneysel 1. tip hata oranları Şekil 4.18 ve Çizelge 4.10'da verilmiştir.



Şekil 4.18. Hata terimleri t_5 dağılımına sahip olduğunda testlerin deneysel 1.tip hata oranları

Çizelge 4.10. Hata terimleri t_5 dağılımına sahip olduğunda testlerin deneysel 1.tip hata oranları

Test	n=10	n=20	n=30	n=50
BP	0,0084	0,0666	0,0678	0,0800
SSK	0,0238	0,0270	0,0260	0,0234
G	0,0266	0,0532	0,0446	0,0430
GQ	0,0562	0,0938	0,1054	0,1190
P	0,0896	0,0584	0,0536	0,0518
W	0,0128	0,0570	0,0408	0,0356
HG	0,0980	0,0658	0,0564	0,0556
MLM	0,0110	0,0472	0,0464	0,0470
MLA	0,1068	0,0822	0,0734	0,0648

X 'ler log-normal dağılımından, hata terimleri t_5 dağılımından seçildiğinde testlerin deneysel 1. tip hata oranları incelendiğinde küçük örnek çaplarında SSK, BP, G, W, MLM testlerinin %5 değerinden oldukça küçük ve MLA, HG ve P testleri %5 değerinden oldukça büyük çıkmıştır. GQ, testi %5 değerine oldukça yakın sonuç verdiği gözlemlenmiştir.

Orta örnek çaplarında SSK %5 değerinden oldukça küçük ve BP, GQ, MLA testleri %5 değerinden oldukça büyük çıkmıştır. G, P, W, HG ve MLM testleri %5 değerine oldukça yakın sonuç verdiği gözlemlenmiştir.

Büyük örnek çaplarında SSK testi %5 değerinden oldukça küçük, BP, GQ, MLA testleri %5 değerinden oldukça büyük çıkmıştır. G, P, W, HG, MLM testleri %5 değerine oldukça yakın sonuç verdiği gözlemlenmiştir.





5. SONUÇ

Bu çalışmada, basit doğrusal regresyon modellerinde değişen varyans durumunun tespiti için kullanılan testler üzerinde odaklanılmıştır. Değişen varyans durumları, modelin hata terimlerinin gözlemler arasında sistemli bir şekilde değiştiği anlamına gelir. Bu durum, modelin tahmin gücünü etkileyebilir ve güven aralıkları ile hipotez testleri sonuçlarını yanıltabilir. Değişen varyansın belirlenmesi için biçimsel olmayan ve biçimsel yöntemler kullanılmaktadır. Bu çalışmada literatürde var olan biçimsel yöntemlerden Breusch-Pagan (BP), Spearman Sıra Korelasyonu (SSK), Glejser (G), Goldfeld-Quandt (GQ), Park (P), White (W), Harvey-Godfrey (HG), medyan değerine göre Modifiye Edilmiş Levene (MLM), ortalamaya göre Modifiye Edilmiş Levene (MLA) testleri ele alınmıştır. Yapılan simülasyon çalışmasına göre özellikle, Spearman Sıra Korelasyonu testi, Breusch-Pagan testi ve Goldfeld-Quandt testi, değişen varyans durumlarını belirlemede diğer testlere göre daha iyi performans göstermiştir.

Sonuç olarak, basit doğrusal regresyon modellerinde değişen varyans durumunun doğru bir şekilde belirlenmesi, modelin güvenilirliğini artırabilir ve istatistiksel çıkarımları daha sağlam hale getirebilir. Bu çalışma, regresyon analizi alanında önemli bir konu olan değişen varyans durumlarına odaklanarak, gelecekteki araştırmalara katkıda bulunmayı amaçlamaktadır.



KAYNAKLAR

- Ali, M. M., Giaccotto, C. (1984). A study of several new and existing tests for heteroscedasticity in the general linear model. *Journal of Econometrics*, 26(3), 355-373.
- Astivia, O. L. O., Zumbo, B. D. (2019). Heteroskedasticity in Multiple Regression Analysis: What it is, How to Detect it and How to Solve it with Applications in R and SPSS. *Practical Assessment, Research, and Evaluation*, 24(1), 1.
- Aydın, D. (2014). *Uygulamalı regresyon analizi: kavramlar ve R hesaplamaları*. Ankara: Nobel Akademik Yayıncılık, 35.
- Breusch, T. S., Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the Econometric Society*, 1287-1294.
- Carapeto, M., Holt, W. (2003). Testing for heteroscedasticity in regression models. *Journal of Applied Statistics*, 30(1), 13-20.
- Cook, R. D., Weisberg, S. (1983). Diagnostics for heteroscedasticity in regression. *Biometrika*, 70(1), 1-10.
- Çelik, R. (2012). *Regresyon çözümlemesinde kelebek dağılan artıklar için değişen varyanslılık sorunu ve düzeltilmesi*. Doktora Tezi, Mimar Sinan Güzel Sanatlar Üniversitesi Fen Bilimleri Enstitüsü, İstanbul, 41-50.
- Erilli, N., Çamurlu, S. (2018). Kantil regresyon analizinde bootstrap tahmini. *Erciyes Üniversitesi Fen Bilimleri Enstitüsü Fen Bilimleri Dergisi*, 35(2), 16-25.
- Evans, M. A., King, M. L. (1985). A point optimal test for heteroscedastic disturbances. *Journal of Econometrics*, 27(2), 163-178.
- Gamgam, H., Altunkaynak, B. (2015), *Regresyon Analizi*, Ankara: Seçkin Yayıncılık, 11-54.
- Glejser, H. (1969). A new test for heteroskedasticity. *Journal of the American Statistical Association*, 64(325), 316-323.
- Godfrey, L. G. (1978). Testing for multiplicative heteroskedasticity. *Journal of Econometrics*, 8(2), 227-236.
- Goldfeld, S. M., Quandt, R. E. (1965). Some tests for homoscedasticity. *Journal of the American statistical Association*, 60(310), 539-547.
- Griffiths, W. E., Surekha, K. (1986). A Monte Carlo evaluation of the power of some tests for heteroscedasticity. *Journal of Econometrics*, 31(2), 219-231.
- Gujarati, D. N. (2012), *Temel Ekonometri*, (Çev. Şenesen, Ü. ve Şenesen, G. G.) İstanbul: Literatür Yayıncılık, 376-389.

- Güleryüz, B. (1992). *Değişen Varyans ve Bir Uygulama*. Yüksek Lisans Tezi, Marmara Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul, 63-81.
- Güriş, S. ve Çağlayan, E. (2018), *Ekonometri Temel Kavramlar*, İstanbul: Der Yayınları, 541-571.
- Harrison, M. J., McCabe, B. P. (1979). A test for heteroscedasticity based on ordinary least squares residuals. *Journal of the American Statistical Association*, 74(366a), 494-499.
- Harvey, A. C. (1976). Estimating regression models with multiplicative heteroscedasticity. *Econometrica: Journal of the Econometric Society*, 461-465.
- Harvey, A. C., Phillips, G. D. A. (1974). A comparison of the power of some tests for heteroskedasticity in the general linear model. *Journal of Econometrics*, 2(4), 307-316.
- Levene, H. (1960). Robust tests for equality of variances. *Contributions to Probability and Statistics*, 278-292.
- Lyon, J. D., Tsai, C. L. (1996). A comparison of tests for heteroscedasticity. *Journal of the Royal Statistical Society Series D: The Statistician*, 45(3), 337-349.
- Montgomery, D. C., Peck, E. A., ve Vining, G. G. (2013). *Doğrusal regresyon analizine giriş*. (çev. Prof. Dr. M. A. Erar) Ankara: Nobel Akademik Yayıncılık, 121.
- Neter, J., Kutner, M. H., Nachtsheim, C. J., Wasserman, W. (1996). *Applied Linear Statistical Models*. New York: McGraw Hill, 116.
- Pagan, A.R., Pak, Y., (1993). Testing for heteroskedasticity. *Handbook of Statistics 11: Econometrics*. 489–518.
- Park, R. E. (1966). Estimation with heteroscedastic error terms. *Econometrica (pre-1986)*, 34(4), 888.
- Ramsey, J. B. (1969). Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 31(2), 350-371.
- Spearman, C. (1987). The proof and measurement of association between two things. *The American Journal of Psychology*, 100(3/4), 441-471.
- Szroeter, J. (1978). A class of parametric tests for heteroscedasticity in linear econometric models. *Econometrica: Journal of the Econometric Society*, 1311-1327.
- Uyanto, S. S. (2019). Monte Carlo power comparison of seven most commonly used heteroscedasticity tests. *Communications in Statistics-Simulation and Computation*, 51(4), 2065-2082.
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society*, 817-838.

Yüce, M. (2004). *Doğrusal olmayan regresyonda değişen varyans ve ilgili uygulamalar*. Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul, 63-81.







Gazili olmak ayrıcalıktır