REGRESYON ANALİZİNDE GENELLEŞTİRİLMİŞ P DEĞERİ

Seray MANKIR

Yüksek Lisans Tezi

İstatistik Anabilim Dalı

Ağustos – 2014

Bu tez çalışması Anadolu Üniversitesi Bilimsel Araştırma Projeleri Komisyonu Başkanlığı tarafından desteklenmiştir. Proje No: 1302F034

ANADOLU ÜNİVERSİTESİ

JÜRİ VE ENSTİTÜ ONAYI

Seray Mankır'ın "**Regresyon Analizinde Genelleştirilmiş p Değeri**" başlıklı **İstatistik** Anabilim Dalındaki, Yüksek Lisans Tezi 22.07.2014 tarihinde, aşağıdaki jüri tarafından Anadolu Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin ilgili maddeleri uyarınca değerlendirilerek kabul edilmiştir.

	Adı-Soyadı	**İmza**
Üye (Tez Danışmanı)	**: Prof. Dr. Berna YAZICI**	
Üye	**: Doç. Dr. Özlem ALPU**	
Üye	**: Yrd. Doç. Dr. Ahmet SEZER**	

Anadolu Üniversitesi Fen Bilimleri Enstitüsü Yönetim Kurulu'nun .... tarih ve ……………….. sayılı kararıyla onaylanmıştır.

Enstitü Müdürü

ÖZET

Yüksek Lisans Tezi

REGRESYON ANALİZİNDE GENELLEŞTİRİLMİŞ P DEĞERİ

Seray MANKIR

Anadolu Üniversitesi

Fen Bilimleri Enstitüsü

İstatistik Anabilim Dalı

Danışman: Prof. Dr. Berna YAZICI

2014, 46 sayfa

Regresyon analizinde, iki regresyon modelinin ve katsayılarının karşılaştırılmasında, değişkenlere ait dağılım bilinmediği durumlarda genelleştirilmiş p değeri kullanılabilir. Genelleştirilmiş p değeri, yalnız yaklaşık çözümler sağlayan klasik p değerinin genelleştirilmiş halidir. İlgilenilen parametreyi açıklarken bunun yanı sıra ilgilenilmeyen (nuisance parametre) ama problemi çözmek için açıklanması gereken parametrelerin olması durumunda da genelleştirilmiş p değerinden yararlanılması kesin çözümü sağlar. Genelleştirilmiş p değeri, gözlem sayısı küçük örneklerde daha iyi ve doğru sonuçlar verir. Bu tezde, uygulamada yaygın olarak kullanılan klasik p değerleri yerine farklı varsayımların sağlanmadığı durumlarda alternatif olarak kullanılabilecek genelleştirilmiş p değeri, teorik olarak araştırılmış, uygun simülasyon programı yazılmış ve regresyon analizinde bir uygulaması verilmiştir.

Anahtar Kelimeler: Genelleştirilmiş p Değeri, Regresyonda Genelleştirilmiş p Değeri, Varsayım Bozulumunda Regresyon Analizi, Regresyon Katsayılarının Karşılaştırılması

ABSTRACT

Master of ScienceThesis

GENERALIZED P VALUES IN REGRESSION ANALYSIS

Seray MANKIR

Anadolu University

Graduate School of Sciences

Statistics Program

Supervisor: Prof. Dr. Berna YAZICI

2014, 46 pages

In regression analysis, in case of comparing two regression models and coefficients where the distribution of variables is not known, generalized p values may be used. The generalized p value is an extended version of the classical p value which provides only approximate solutions. Use of approximate methods, generalized p value, have beter results, performance with small samples. In this thesis, the generalized p value – which may be used alternatively when different assumptions aren’t fulfilled - is researched theoretically; the suitable simulation program is written and an application in regression analysis is given.

Keywords: Generalized p Values, Generalized p Values in Regression, Regression Analysis in Case of Assumption Violation, Comparison of Regression Coefficients

TEŞEKKÜR

Bu araştırmanın gerçekleştirilmesinde birçok kişinin katkısı olmuştur. Öncelikle araştırmamın başından itibaren akademik, maddi ve manevi desteğiyle beni yüreklendiren, yaptığım çalışmalara olan inancı ile beni destekleyen, verdiği dönütler ile araştırmama ve öğrenme sürecime büyük katkılar sağlayan, en zor anlamlarımda yanımda olduğunu bana hissettiren, her zaman sabır ve anlayış gösteren değerli hocam, tez danışmanım Prof. Dr. Berna YAZICI'ya en içten duygularımla teşekkür ederim.

Araştırmam sırasında olumlu eleştirileri ile beni yönlendiren Dr.Samaradasa Weerahandi'ye teşekkürlerimi sunarım.

İstatistik alt yapımın oluşmasında emeği geçen Anadolu Üniversitesi İstatistik Bölümü öğretim üyelerine teşekkürlerimi sunarım.

Yaşamımın her anında yanımda olan, beni sabırla ve sevgiyle yetiştiren, sevgilerini, desteklerini ve başarıma olan inançlarını her zaman hissettiren, akademik gelişimim için maddi ve manevi desteklerini hiç esirgemeyen canım annem Pervin MANKIR'a, canım babam Ahmet MANKIR'a ve canım ablam Sinem MANKIR ÖZTAN'a sonsuz teşekkür ederim.

Bu süreci benimle yaşayan, sabrı, ilgisi ve desteğiyle hep yanımda olan, inancıyla ve sevgisiyle bana güç veren çok sevgili arkadaşım Aydoğan KAHVECİOĞLU'na ve desteklerini her zaman yanımda hissettiğim, isimlerini burada yazamadığım çok sevgili arkadaşlarım ve dostlarıma çok teşekkür ederim.

İÇİNDEKİLER

TABLOLAR DİZİNİ

1. GİRİŞ

İki veya daha fazla regresyon modelinin karşılaştırılması istatistiksel uygulamalarda standart bir sorundur. Regresyon modellerinin karşılaştırılması, regresyon modelinin yapısındaki değişikliğin, önemli bir değişkenin parametresini değiştirip değiştirmeyeceğini test etmek istendiğinde sık sık ortaya çıkar. İki veya daha fazla regresyonun karşılaştırılmasına ilişkin sonuçlarıyla ilgilenen okuyucular Kullback ve Rosenblatt(1957) ve Koschat ve Weerahandi'dir(1992). İki modelin ayrı ayrı parametreleri tahmin edilirken her bir regresyon modeli için yeterince veri olduğu kabul edilir. Chow(1960) iki ayrı regresyon modeli için yeterli veri olmadığı durumlarda bir çözüm sunar. Ancak, birleştirilmiş ve tek regresyonu uygulamak için yeterli veri vardır. Birçok uygulamada eşit varyans varsayımını sağlamak zordur. Bu yüzden de varyans eşitliği varsayımı sağlanmadığı durumlar için yapılan bazı çalışmalar vardır. Toyoda(1974) bu varsayımın sonuçları üzerinde çalışarak, değişen varyans durumunda yapılan çalışmalara öncülük etmiştir. Weerahandi(1987) genelleştirilmiş p değerlerine dayanan daha güçlü testler sunmuştur. Ancak bu varsayımların sağlanmadığı durumda regresyon analizinde genelleştirilmiş p değeri ile ilgili az sayıda çalışma vardır. Varyans homojenliği varsayımı gerçekleşmediğinde iki regresyon modelinin karşılaştırılması genelleştirilmiş p değeri ile yapılabilir. Bu tez kapsamında uygulamada yaygın olarak kullanılan klasik p değerleri yerine farklı varsayımların bozulduğu durumlarda alternatif olarak kullanılabilecek genelleştirilmiş p değeri konusu teorik olarak araştırılacaktır. Özellikle genelleştirilmiş p değeri kullanarak homojen varyans varsayımının bozulması durumunda regresyon analizi ve ortak parametreler içeren problemlerin tam çözümü sağlanarak, beraberinde de regresyon modellerinin homojenliğinin testi incelenecektir. İstatistiksel veri analizi konusunda yaygın kullanıma sahip regresyon analizi için karşılaşılan bu varsayım bozulumu sorunlarında izlenecek yollardan biri olan genelleştirilmiş p değeri ele alınacaktır. Geliştirilen ve var olan teorik yapı için simülasyon programları geliştirilecek, farklı birim sayıları, farklı değişken yapıları ve farklı varyanslılık durumları için simülasyon programları yazılacak ve literatüre bu simülasyon programları kazandırılacaktır.

2. REGRESYON ANALİZİ

Regresyon analizi, bağımlı değişkeni ile bir veya daha fazla sayıda bağımsız değişkenleri arasındaki ilgiyi sayısal hale dönüştürmede kullanılan istatistiksel analizdir. Ayrıca, bilinen bulgulardan, bilinmeyen gelecekteki olaylarla ilgili tahminler yapılmasına izin verir. Regresyon, bağımlı ve bağımsız değişken(ler) arasındaki ilişkiyi ve doğrusal eğri kavramını kullanarak, bir tahmin eşitliği geliştirir. Bağımsız değişken olarak bir değişken kullanılırsa basit regresyon, bağımsız değişkenler olarak iki veya daha fazla değişken kullanılırsa çoklu regresyon analizinden söz etmek mümkündür. Amaç her bağımsız değişkenin bağımlı değişkendeki toplam değişime olan katkısının saptanması ve dolayısıyla bağımsız değişkenlerin doğrusal kombinasyonunun değerinden hareketle bağımlı değişken değerinin tahmin edilmesidir (Field,2000). Değişkenler arasındaki ilişkiyi açıklamak için kullanılan matematiksel model regresyon modeli olarak adlandırılır (Birkes&Dodge,1993). Aşağıda basit doğrusal regresyon analizi anlatılmıştır.

2.1.Basit Doğrusal Regresyon Modeli

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad\qquad i = 1, \dots, n$$

şeklinde bir bağımlı ve bir de bağımsız değişken içeren bir modeldir.

Y; Bağımlı (açıklanan) değişken

X; Bağımsız (açıklayıcı) değişken

$\boldsymbol{\beta_0}$; Sabit olup, X=0 olduğunda Y’nin aldığı değerdir.

$\boldsymbol{\beta_1}$; Regresyon katsayısı olup, X’in kendi birim cinsinden 1 birim değişmesine karşılık, Y’de kendi birimi cinsinden meydana gelecek ortalama değişme miktarını ifade eder, eğimdir.

$\boldsymbol{\varepsilon}$**;** Hata terimi.

2.1.1. Parametrelerin (Katsayıların) Tahmini

Bir regresyon modeli oluşturulurken genelde en küçük kareler ve en büyük olabilirlik (maximum likelihood) teknikleri olarak bilinen iki yaklaşımdan birisi kullanılır. Eğer hata teriminin normal dağılım göstermesi şeklinde bir varsayım varsa en büyük olabilirlik, hata teriminin dağılışı ile ilgili herhangi bir varsayım söz konusu değilse en küçük kareler (EKK) tekniği kullanılarak parametreler tahmin edilir. Tabi EKK yönteminin de uygulanabilmesi için çeşitli varsayımların sağlanması gerekir ve tahmin modeli öyle oluşturulur. Eğer kurulan regresyon modeli veriye uygun değilse alınan sonuçlar da yanıltıcı olacaktır (Wilcox 1997). Basit doğrusal regresyon analizinde bulunacak olan regresyon modellerinin tahmin amaçlı kullanılabilmesi için; hata terimlerinin ($\varepsilon_i = Y_i - \hat{Y}_i$) rassal olup normal dağılım göstermesi, hataların beklenen değerinin 0 ve varyanslarının da sabit olup σ^2'e eşit olması, hataların birbirinden bağımsız olması ($cov(\varepsilon_i, \varepsilon_j) = 0$), hata terimleri ile bağımlı değişken arasında korelasyonun olmaması gibi bazı varsayımların sağlanması gerekmektedir. Bu varsayımlardan birisinin sağlanamaması durumunda EKK tahminleri kararlı ve küçük varyanslı olma özelliğini kaybederek yanlı, tutarsız veya etkisiz olacaktır.

2.1.2. Basit Doğrusal Regresyon Modelinde Parametrelerin EKK Yöntemiyle Tahmin Edilmesi

Anakütle regresyon modelinde yer alan β_0 ve β_1 parametrelerinin örneklemden elde edilen tahminleri $\hat{\beta}_0$ ve $\hat{\beta}_1$ olarak yazıldığında, tek değişkenli regresyon modeli

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i \qquad\qquad i = 1, 2, \dots, n$$

biçimindedir. Modelde yer alan $\hat{\beta}_0$ ve $\hat{\beta}_1$ terimlerinin değerlerini bulmak için kullanılan EKK yönteminin temelini, toplam sapmaların karelerinin toplamını en küçük yapacak değerlerin bulunması oluşturmaktadır.

$$\hat{\varepsilon}_i = Y_i - \hat{Y}_i$$

eşitlikte verilen ifade ile hesaplanan hata terimleri pozitif, negatif ve sıfır değerine sahip olurken bu farkların toplamı

$$\sum_{i=1}^{n} \hat{\varepsilon}_i = \sum_{i=1}^{n} (Y_i - \hat{Y}_i) = 0$$

olur. EKK yöntemi, β_0 ve β_1 parametrelerinin tahminleri olan $\hat{\beta}_0$ ve $\hat{\beta}_1$'nın farkını en küçük yapacak biçimde aşağıdaki gibi belirler

$$min \sum_{i=1}^{n} \hat{\varepsilon}_i^2 = min \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2.$$

Burada regresyon katsayılarının EKK tahminlerini elde edebilmek için (2.1) eşitliğinin en küçük değerine L denilirse ve bu eşitlikte $\hat{\beta}_0$ ve $\hat{\beta}_1$'e göre kısmi türevler alınıp sıfıra eşitlendiğinde I. ve II. normal denklemleri elde edilir.

$$min \sum_{i=1}^{n} (Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i))^2 = L \qquad (2.1)$$

$$\sum_{i=1}^{n} Y_i = \hat{\beta}_0 n + \hat{\beta}_1 \sum_{i=1}^{n} X_i \qquad (I)$$

$$\sum_{i=1}^{n} X_i Y_i = \hat{\beta}_0 \sum_{i=1}^{n} X_i + \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 \qquad (II)$$

Bu denklemler üzerinden gerekli çözümler yapıldığında β_0 ve β_1 parametrelerinin tahminleri olan $\hat{\beta}_0$ ve $\hat{\beta}_1$

$$\hat{\beta}_1 = \frac{n[\sum_{i=1}^{n} X_i Y] - (\sum_{i=1}^{n} X_i)(\sum_{i=1}^{n} Y_i)}{n(\sum_{i=1}^{n} X_i^2) - (\sum_{i=1}^{n} X_i)^2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

$$\hat{\beta}_0 = \frac{\sum_{i=1}^{n} Y_i + \hat{\beta}_1 \sum_{i=1}^{n} X_i}{n} = \bar{Y} - \hat{\beta}_1 \bar{X}$$

eşitlikleri elde edilir. Dolayısıyla, yukarıda EKK yöntemiyle tahmin edilen $\hat{\beta}_0$ ve $\hat{\beta}_1$ değerleri yerlerine yazılarak regresyon modeli oluşturulur.

2.2. Çoklu Doğrusal Regresyon Modeli

Bir bağımlı değişken ve birden fazla bağımsız değişkenin yer aldığı regresyon modelleridir. k açıklayıcı değişkenden oluşan regresyon modelini her gözlem i için aşağıdaki gibi yazabiliriz:

$$y_i = \hat{\beta}_1 + \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} + \cdots + \hat{\beta}_k x_{ki} + \varepsilon_i \qquad i = 1, \dots, n$$

Burada y_i bağımlı değişkenin *i*nci gözleme ilişkin değerini, x_{ij}, $j = 1, \dots, k$, *i*nci açıklayıcı değişkenin *j*nci gözlem değerini, ε_i ise rassal hata terimini ifade etmektedir. k bilinmeyenli n denklemden oluşan bu sistemi aşağıdaki gibi yazabiliriz:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{21} & x_{31} & \cdots & x_{k1} \\ 1 & x_{22} & x_{32} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{2n} & x_{3n} & \cdots & x_{kn} \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Böylece modeli aşağıdaki gibi ifade edebiliriz:

$$\underbrace{Y}_{n\times 1} = \underbrace{X}_{n\times k} \underbrace{\hat{\beta}}_{k\times 1} + \underbrace{\varepsilon}_{n\times 1}$$

Çoklu regresyon modellerinde de hata terimlerinin beklenen değerinin 0 ve varyanslarının da sabit olup σ^2 'e eşit olması, hata terimleri arasında ardışık bağımlılık olmaması ($cov(\varepsilon_i, \varepsilon_j) = 0$), hata terimlerinin açıklayıcı değişkenlerden bağımsız olması, açıklayıcı değişkenler arasında tam doğrusal ilişki bulunmaması gibi bazı varsayımların sağlanması gerekmektedir.

2.2.1. Çoklu Doğrusal Regresyon Modelinde Parametrelerin EKK Yöntemiyle Tahmin Edilmesi

Bilinmeyen parametre vektörü β 'nın EKK tahmin edicisi hata kareler toplamı minimum yapılarak bulunur.

$\varepsilon = Y - X\hat{\beta}$ olmak üzere hata kareler toplamı;

$$\sum_{i=1}^{n} \varepsilon_i^2 = \varepsilon' \varepsilon$$

$$= (Y - X\hat{\beta})'(Y - X\hat{\beta}) = (Y' - \hat{\beta}'X')(Y - X\hat{\beta})$$

$$= Y'Y - Y'X\hat{\beta} - \hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta}$$

olarak elde edilir. $\beta'X'Y$ için $(\beta'X'Y)' = Y'X\beta$'dir. Buna göre hata kareler toplamı;

$$\varepsilon'\varepsilon = Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta}$$

Normal denklemleri;

$$\frac{\partial(\varepsilon'\varepsilon)}{\partial\beta} = -2X'Y + 2X'X\beta = 0$$

$$X'X\beta = X'Y$$

elde edilir. Buradan β'nın EKK tahmin edicisi

$$\hat{\beta} = (X'X)^{-1}X'Y$$

olarak elde edilir. Hata terimlerinin birbirleriyle ilişkisiz ve varyansları eşit varsayımı altında $\hat{\beta}$ 'nın varyans-kovaryans matrisi aşağıda verilmiştir:

$$V(\hat{\beta}) = E[\left(\hat{\beta} - E(\hat{\beta})\right)\left(\hat{\beta} - E(\hat{\beta})\right)^T]$$

$$= E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T]$$

$$= E[((X^TX)^{-1}X^Tu)((X^TX)^{-1}X^Tu)^T]$$

$$= E[(X^TX)^{-1}X^Tuu^TX(X^TX)^{-1}]$$

$$= (X^TX)^{-1}X^T\underbrace{E(uu^T)}_{\sigma^2 I_n}X(X^TX)^{-1}$$

$$= \sigma^2(X^TX)^{-1}\underbrace{X^TI_nX}_{X^TX}(X^TX)^{-1}$$

$$= \sigma^2\underbrace{(X^TX)^{-1}X^TX}_{I_k}(X^TX)^{-1}$$

$$= \sigma^2(X^TX)^{-1}$$

Bu formülde hata varyansı σ^2 bilinmediğinden verilerden hareketle tahmin edilmesi gerekir. σ^2'nin yansız bir tahmin edicisi aşağıdaki gibi bulunabilir:

$$s^2 = \frac{1}{n-k}\varepsilon'\varepsilon$$

$$s^2 = \frac{1}{n-k}\left(Y - X\hat{\beta}\right)^{'}\left(Y - X\hat{\beta}\right)$$

Öyleyse EKK tahmin edicilerinin varyans-kovaryans matrisi tahmin edicisi

$$\widehat{V(\hat{\beta})} = s^2(X^{'}X)^{-1}$$

şeklinde yazılabilir. Bu matris $k \times k$ boyutlu, simetrik, kare ve pozitif tanımlıdır. Köşegen elemanları EKK tahmin edicilerinin varyanslarını, köşegen dışı elemanlar ise kovaryansları ifade eder.

2.3. Parametrenin ve Modelin Anlamlılığı

Regresyon modelindeki bağımlı ve bağımsız değişkenler arasında ilişkinin varlığı, eğer bir ilişki var ise bu ilişkinin gücü, değişkenler arasındaki ilişkinin türü, bağımsız değişkenlerin bağımlı değişkeni ne kadar etkilediği ve kurulan modelinin anlamlılığı çoklu regresyon analizi ile cevaplanabilir.

$$H_0 : \hat{\beta}_1 = \hat{\beta}_2 = \cdots = \hat{\beta}_k = 0$$

$$H_1 : En\ az\ biri\ diğerlerinden\ farklıdır(\hat{\beta}_j \neq 0).$$

Hipotezlerinden, H_0 hipotezini reddetmek; bağımsız değişkenlerden en az biri modele anlamlı bir şekilde katkıda bulunuyor demektir.

Regresyon analizi tablosu oluşturulması için gerekli tanımlamalar:

$$\underbrace{\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}}(Y_i - \bar{Y})^2}_{\boldsymbol{GKT}} = \underbrace{\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}}(Y_i - \hat{Y}_i)^2}_{\boldsymbol{AKT}} + \underbrace{\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}}(\hat{Y}_i - \bar{Y})^2}_{\boldsymbol{RAKT}}$$

GKT = Genel Kareler Toplamı = $\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}(Y_i^2 - 2Y_i\bar{Y} + \bar{Y}^2) =$ $i=1nYi2-nY2=Y'Y-(i=1nYi)2n$

AKT = Artık Kareler Toplamı = $Y^{'}Y - \hat{\beta}'X'Y$

RAKT = Regresyonla Açıklanan Kareler Toplamı = $\hat{\beta}^{'}X^{'}Y - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n}$

olmak üzere;

Tablo 2.1. Regresyon Analizi İçin Varyans Analizi Tablosu

Değişim Kaynağı	Kareler Toplamı (SS)	Sd (Df)	Kareler Ort. (MS)	F_h
RAKT (Regression)	$\hat{\beta}'X'Y - \frac{(\sum_{i=1}^{n} Y_i)^2}{n}$	k	$\frac{\hat{\beta}'X'Y - \frac{(\sum_{i=1}^{n} Y}{n}}{k}$	$\frac{\frac{RAKT}{k}}{\frac{AKT}{n-(k+1)}}$
AKT (Residual)	$Y'Y - \hat{\beta}'X'Y$	n-(k+1)	$\frac{Y'Y - \hat{\beta}'X'Y}{n-(k+1)}$	
GKT (Total)	$Y'Y - \frac{(\sum_{i=1}^{n} Y_i)^2}{n}$	n-1		

H_0 hipotezi altında $F_{hesaplanan} \sim F_{k,n-(k+1),\alpha}$ dağılımına sahiptir.

$F_{hesaplanan}$ değeri, $F_{k,n-(k+1),\alpha}$ tablo değerinden küçük ise H_0 hipotezi kabul edilir. Başka bir deyişle, regresyon modelini oluşturan katsayılardan hiç biri modele katkı sağlamıyor, α anlamlılık düzeyinde istatistiksel olarak anlamlı olmadıkları söylenir.

$F_{hesaplanan}$ değeri, $F_{k,n-(k+1),\alpha}$ tablo değerinden büyük ise H_0 hipotezi reddedilir. Başka bir deyişle, regresyon modelini oluşturan katsayılardan en az birinin α anlamlılık düzeyinde istatistiksel olarak anlamlı olduğu söylenir. Bu katsayılardan en az biri regresyon modeline katkı sağlamaktadır.

3. GENELLEŞTİRİLMİŞ P DEĞERİ

Genelleştirilmiş test değişkeni kavramı ve genelleştirilmiş p değeri ilk olarak Tsui ve Weerahandi (1989) tarafından bazı istatistiksel testlerde kullanılmak için önerilmiştir. Bu yöntem birçok istatistiksel analizde kullanılmaktadır. Özellikle bilgisayarların hızlanması ile birlikte son yıllarda daha yaygın olarak kullanılan yöntemde test istatistiğinin dağılımı teorik olarak elde edilemediğinden iteratif işlemler ile p değeri elde edilir.

3.1. Varyansları Eşit İki Regresyon Modelinin Karşılaştırılması

İki regresyon modelinde parametrelerin bazıları farklıyken diğer parametreler ortak olabilir. Bu yüzden aşağıdaki yapıda iki doğrusal regresyon modeli ele alınsın

$$y_j = U_j c + V_j d_j + W_j t_j + \varepsilon_j, \qquad j = 1,2 \tag{3.1}$$

Hata terimlerinin normal dağıldığı varsayılsın;

$$\varepsilon_i \sim N\left(0, \sigma_j^2 I_j\right), \qquad i = 1,2,\dots,n$$

I_j , $j = 1,2,$ n_i boyutlu birim matris iken $\sigma_1^2 = \sigma_2^2$ olduğu varsayılır. y_1 ve y_2 vektörlerinin uzunlukları n_1 ve n_2 ve bağımlı değişkenlerden oluşur. Örnek veri için açıklayıcı değişkenleri U_j, V_j ve W_j matrisleri temsil eder. U_j ve W_j matrislerine karşılık gelen değişkenler her iki örneklem için benzerken, V_1 ve V_2 matrislerine karşılık gelen değişkenler ise birbirinden farklıdır. Örneğin, bir örneklem için yanıt değişkenini etkileyen açıklayıcı değişken bilinirken, diğer örneklem için bilinmeyebilir, ya da bir örneklem için yeterli sayıda açıklayıcı değişken toplanırken diğer örneklem için toplanmayabilir. c her iki örneklem için bilinen parametre vektörü ve uzunluğu p_c dir. d_1 ve d_2 vektörlerinin de uzunlukları sırasıyla p_{d_1} ve p_{d_2} dir. Bunlar iki örneklem için de farklı olması beklenen parametreleri de içerir. Başka bir deyişle, farklı dağılımlardan olması beklenen V_1 ve V_2 matrislerini oluşturan değişkenlerin katsayılarıdır. Her ikisinin de uzunlukları p_t ile gösterilen t_1 ve t_2 vektörlerinin eşitliği test edilmek istenirse:

$H_0 : t_1 = t_2$ hipotezine karşılık $H_1 : t_1 \neq t_2$ hipotezi kurulur. H_0 hipotezi, farklı iki regresyon modeline ait katsayıların eşit olduğunu öne sürerken, H_1 hipotezi ise bu katsayıların birbirinden farklı olduğunu öne sürer.

Her iki örnekleme de ait olan yanıt değişkenleri sıralı bir şekilde vektör haline getirilir $y' = (y_1', y_2')$. Değişkenlere ait matrisler ise;

$$X_{12} = \begin{bmatrix} U_1 & 0 & V_1 & 0 & W_1 \\ 0 & U_2 & 0 & V_2 & W_2 \end{bmatrix} \quad (n_1 + n_2) \times \left(2p_c + p_{d_1} + p_{d_2} + p_t\right) \tag{3.2}$$

ve

$$X_{1,2} = \begin{bmatrix} U_1 & 0 & V_1 & 0 & W_1 & 0 \\ 0 & U_2 & 0 & V_2 & 0 & W_2 \end{bmatrix} \quad (n_1 + n_2) \times \left(2p_c + p_{d_1} + p_{d_2} + 2p_t\right) \tag{3.3}$$

şeklindedir. 1,2 indeksli notasyon $t_1 \neq t_2$ durumu için kullanılırken, 12 indeksli notasyon ise $t_1 = t_2$ durumunu ifade etmek için kullanılır, ayrıca 1 ve 2 indeksleri de sadece 1. ve 2. örneklemlerin regresyon değişkenlerini tanımlamak için kullanılmaktadır. Matrislere ait artık kareler toplamları da sırasıyla $s_{12}^2,\ s_{1,2}^2$'dir. Eğer iki regresyon modelinin ortak parametreleri yoksa artık kareler toplamı; ayrı ayrı modellere ait artık kareler toplamlarının toplanmasıyla elde edilir $s_{1,2}^2 = s_1^2 + s_2^2$.

$H_0 : t_1 = t_2$ hipotezi altında $\qquad F = \frac{(s_{12}^2 - s_{1,2}^2)/p_t}{s_{1,2}^2/l} \sim F_{p_t,l}$

$l = (n_1 + n_2) - \left(2p_c + p_{d_1} + p_{d_2} + 2p_t\right)$.

Bu istatistiğin gözlemlenen anlamlılık düzeyi

$$p = 1 - H_{p_t,l}\left[\frac{(s_{12}^2 - s_{1,2}^2)/p_t}{s_{1,2}^2/l}\right]$$

ile verilirken $H_{k,l}$, k ve l serbestlik dereceli F dağılımının birikimli dağılım fonksiyonudur. Bilinen α düzeyi ile karşılaştırılır $p < \alpha$ ise H_0 hipotezi reddedilir.

Başka bir deyişle, farklı regresyon modellerine ait katsayılar α anlamlılık düzeyinde istatistiksel olarak birbirinden farklıdır.

3.2. Ortak Parametreler Olmadan İki Regresyon Modelinin Karşılaştırılması

İki regresyon modelini karşılaştırırken bilinen parametre olmadığı durum söz konusu ise, diğer bir deyişle modelde c = 0 olduğu varsayılsın. Genelleştirilmiş p değerlerine bağlı yansız testleri vermek için, $\tilde{y}_j = a_j^{-1} y_j$, $\tilde{V}_j = a_j^{-1} V_j$ ve $\widetilde{W}_j = a_j^{-1} W_j$ şeklinde tanımlansın. Tildalı matrislerin oluşturulmasında $a_1^2 = s_1^2/R$ ve $a_2^2 = s_2^2/(1-R)$ normalleştirme faktörleri kullanılır. R belirlenen parametrelere sahip beta dağılımından elde edilen rassal değişkendir. Matrisler ise;

$$\tilde{X}_{12} = \begin{bmatrix} \mathrm{U}_1 & 0 & \tilde{V}_1 & 0 & \widetilde{W}_1 \\ 0 & \mathrm{U}_2 & 0 & \tilde{V}_2 & \widetilde{W}_2 \end{bmatrix} \qquad (n_1 + n_2) \times (2p_c + p_{d_1} + p_{d_2} + p_t), \qquad (3.4)$$

ve

$$\tilde{X}_{1,2} = \begin{bmatrix} U_1 & 0 & \tilde{V}_1 & 0 & \widetilde{W}_1 & 0 \\ 0 & U_2 & 0 & \tilde{V}_2 & 0 & \widetilde{W}_2 \end{bmatrix} \quad (n_1 + n_2) \times \left(2p_c + p_{d_1} + p_{d_2} + 2p_t\right) \quad (3.5)$$

şeklinde ifade edilir.

$\tilde{S}_{1,2}^2(a_1^2, a_2^2)$ ve $\tilde{S}_{12}^2(a_1^2, a_2^2)$ notasyonları ise sırasıyla $\tilde{X}_{1,2}$ ve $\tilde{X}_{12}$ nin hata kareler toplamlarıdır. $\tilde{S}_T^2$, rassal değişkenlere ait kareler toplamı iken $\tilde{s}_T^2$ ise gözlemlenmiş değişkenlere karşılık gelen kareler toplamıdır. Herhangi bir pozitif c değeri için;

$$\tilde{S}_T^2(\mathrm{c}a_1^2, ca_2^2) = c^{-1} \tilde{S}_T^2\,(a_1^2, a_2^2). \qquad (3.6)$$

S_1^2 ve S_2^2 ise iki regresyon modelinden ayrı ayrı olarak hesaplanmış hata kareler toplamlarıdır. Bu tanımlamalarla birlikte, hipotezi test etmek için yararlanılan Weerahandi(2005)'e göre p değeri

$$p = 1 - E_R\left\{H_{k,l}\left[\frac{l}{k}\left(\tilde{s}_{12}^2\left(\frac{s_1^2}{R}\,,\,\frac{s_2^2}{(1-R)}\right) - 1\right)\right]\right\}, \qquad (3.7)$$

$\tilde{s}^2_{12}$, $a_1^2 = s_1^2/R$ ve $a_2^2 = s_2^2/(1-R)$ aracılığıyla kurulmuş normalleştirilmiş artık kareler toplamıdır. Ayrıca $k = p_t$, $l = n_1 + n_1 - (2p_c + p_{d_1} + p_{d_2} + 2p_t)$ olmak üzere $H_{k.l}$, k ve l serbestlik dereceli F dağılımının birikimli dağılım fonksiyonudur ve beklenen değeri parametreleri aşağıdaki gibi verilen beta dağılımdan elde edilir (Weerahandi,2005)

$$R \sim Beta\left(\frac{n_1 - 2p_t - p_{d1}}{2}, \frac{n_2 - 2p_t - p_{d2}}{2}\right).$$

Gözlenen anlamlılık değeri, p (3.7) eşitliği yardımıyla verilerin nasıl daha iyi ölçülmesini sağlayarak farklı iki regresyon modeline ait katsayıların karşılaştırılmasını ifade eden $H_0: t_1 = t_2$ hipotezini sınar. (3.7) eşitliğindeki beklenen değer sayısal iterasyon işlemler aracılığıyla hesaplanabilir. Eğer $p < \alpha$ ise H_0 hipotezi reddedilir. Ayrıca R gibi bazı yazılım programları kullanılarak p'nin tam değeri hesaplanabilir.

3.2.1. Test Değişkeninin Türetilmesi ve p Değerinin Elde Edilmesi

Dikkat edilirse

$$S_1^2/\sigma_1^2 + S_2^2/\sigma_2^2 = \tilde{S}^2_{1,2}(\sigma_1^2, \sigma_2^2)$$

eşitliği yazılabilir.

$$R = \frac{S_1^2/\sigma_1^2}{\tilde{S}^2_{1,2}(\sigma_1^2,\sigma_2^2)} \sim Beta\left(\frac{n_1 - 2p_t - p_{d1}}{2}, \frac{n_2 - 2p_t - p_{d2}}{2}\right)$$

muhtemel test değişkenini ele alalım

$$T = \frac{\tilde{S}^2_{12}(\sigma_1^2,\sigma_2^2)}{\tilde{s}^2_{12}(s_1^2\sigma_1^2/S_1^2, s_2^2\sigma_2^2/S_2^2)} \ .$$

Beta rassal değişkenine dayanarak, T şu şekilde yeniden yazılabilir

$$T = \left(\frac{\tilde{S}^2_{12}(\sigma_1^2,\sigma_2^2)}{\tilde{S}^2_{1,2}(\sigma_1^2,\sigma_2^2)}\right) \frac{1}{(\tilde{s}^2_{12}(s_1^2/R, s_2^2/(1-R)))} \ .$$

Burada ikinci faktör bilinmeyen parametrelerden hiç birine bağlı değil ve her iki faktör de simetrik ve idempotent olduğu için istatistiksel olarak birbirlerinden bağımsızdırlar. Böylece T bir test değişkenidir ve yansız testlerin tanımlanmasında yararlanılabilir. Test değişkeninin gözlenmiş değeri $t_{obs} = 1$ 'dir. Dolayısıyla p değeri elde edilebilir:

$$p = Pr(T \geq 1)$$

$$= Pr\left(\frac{\tilde{S}_{12}^2}{\tilde{S}_{1,2}^2}(\sigma_1^2, \sigma_2^2) \geq \tilde{s}_{12}^2(s_1^2/R, s_2^2/(1-R))\right)$$

$$= Pr\left(\frac{\tilde{S}_{12}^2 - \tilde{S}_{1,2}^2}{\tilde{S}_{1,2}^2}(\sigma_1^2, \sigma_2^2) \geq \tilde{s}_{12}^2(s_1^2/R,\ s_2^2/(1-R)) - 1\right)$$

$$= 1 - E_R\left\{H_{k,l}\left(\frac{l}{k}\left(\tilde{s}_{12}^2\left(\frac{s_1^2}{R}\ ,\ \frac{s_2^2}{(1-R)}\right) - 1\right)\right)\right\},$$

olduğu iddia edilir (Weerahandi,2005).

3.3. Varyansların Eşit Olmadığı Durumda İki Genel Modelin Karşılaştırılması

İki regresyon modeli aşağıdaki yapıda, varyansları eşit olmadığı durumda ($\sigma_1^2 \neq \sigma_2^2$) ve uygulamak için yeterli gözlem değerleri olduğu varsayılsın.

$n_j > p_c + p_{d_i} + p_t, \quad j = 1,2.$

$$y_j = U_j c + V_j d_j + W_j t_j + \varepsilon_j, \quad j = 1,2$$

$$\tilde{X}_{12} = \begin{bmatrix} \tilde{U}_1 & 0 & \tilde{V}_1 & 0 & \tilde{W}_1 \\ 0 & \tilde{U}_2 & 0 & \tilde{V}_2 & \tilde{W}_2 \end{bmatrix}, \qquad (n_1 + n_2) \times (2p_c + p_{d_1} + p_{d_2} + p_t)$$

ve

$$\tilde{X}_{1,2} = \begin{bmatrix} \tilde{U}_1 & 0 & \tilde{V}_1 & 0 & \tilde{W}_1 & 0 \\ 0 & \tilde{U}_2 & 0 & \tilde{V}_2 & 0 & \tilde{W}_2 \end{bmatrix}, \qquad (n_1 + n_2) \times (2p_c + p_{d_1} + p_{d_2} + 2p_t)$$

matrislerinden elde edilen hata kareler toplamları sırasıyla $\tilde{S}_{12}^2(a_1^2, a_2^2)$ ve $\tilde{S}_{1,2}^2(a_1^2, a_2^2)$ şeklinde gösterilsin. Daha sonra, farklı regresyon modellerine ait katsayıların karşılaştırılması $H_0: t_1 = t_2$ için yapılan testler test değişkenine dayanarak elde edilir:

$$T = \frac{(\tilde{S}_{12}^2 - \tilde{S}_{1,2}^2)(\sigma_1^2, \sigma_2^2)}{(\tilde{s}_{12}^2 - \tilde{s}_{1,2}^2)(s_1^2\sigma_1^2/S_1^2, s_2^2\sigma_2^2/S_2^2)} \tag{3.8}$$

bu test değişkeni ile verilen p değeri

$$p = 1 - E_{R_1,R_2}\left\{H_{k,l}\left[\frac{l}{k} R_2 \tilde{s}^2 \left(\frac{s_1^2}{R_1}, \frac{s_2^2}{(1-R_1)}\right)\right]\right\} \tag{3.9}$$

şeklindedir. $\tilde{s}^2 = \tilde{s}_{12}^2 - \tilde{s}_{1,2}^2$, normalleştirilmiş artık kareler toplamları farkını ifade eder. Her biri s_1^2/R_1 ve $s_2^2/(1-R_1)$ normalleştirme faktörü kullanılarak normalleştirilir. Ayrıca k= p_t , $l = n_1 + n_1 - \left(p_c + p_{d_1} + p_{d_2} + 2p_t\right)$ olmak üzere $H_{k.l}$, k ve l serbestlik dereceli F dağılımının birikimli dağılım fonksiyonudur ve beklenen değeri bağımsız beta rassal değişkenine göre alınır.

$$R_1 \sim Beta\left(\frac{n_1 - p_c - p_t - p_{d_1}}{2}, \frac{n_2 - p_c - p_t - p_{d_2}}{2}\right),$$

$$R_2 \sim Beta\left(\frac{n_1 + n_2 - 2p_c - 2p_t - p_{d_1} - p_{d_2}}{2}, \frac{p_c}{2}\right).$$

Eşitlik (3.9) ile verilen gözlemlenmiş anlamlılık değeri p, $H_0: t_1 = t_2$ hipotezinin doğru olup olmadığını belirtir. α düzeyinde, eğer $p < \alpha$ ise H_0 hipotezi reddedilir.

3.3.1. Test Değişkeninin Türetilmesi ve p Değerinin Elde Edilmesi

R_1 ve R_2 'nin tanımı daha önceki kısımda anlatılana benzer olarak yapılır.

$$R_1 = \frac{S_1^2/\sigma_1^2}{S_1^2/\sigma_1^2 + S_2^2/\sigma_2^2}, \; R_2 = \frac{S_1^2/\sigma_1^2 + S_2^2/\sigma_2^2}{\tilde{S}_{1,2}^2(\sigma_1^2 + \sigma_2^2)}$$

$\tilde{S}_j^2 = S_j^2/\sigma_j^2$, j=1,2 ve $\tilde{S}_{12}^2$, $\tilde{S}_{1,2}^2$,$\tilde{S}_1^2 + \tilde{S}_2^2$ ve $\tilde{S}_1^2$ kareler toplamları ifade etsin.

$\tilde{S}_{12}^2 - \tilde{S}_{1,2}^2$, $\tilde{S}_{1,2}^2 - (\tilde{S}_1^2 + \tilde{S}_2^2)$, $(\tilde{S}_1^2 + \tilde{S}_2^2) - \tilde{S}_1^2$ ve $\tilde{S}_1^2$ matrisleri simetrik ve idempotent oldukları için birbirlerinden bağımsız ve ki-kare dağılım gösterirler. Bunların merkezi olmayan parametreleri sırasıyla $\mu' \tilde{P}_{12}\mu, 0, 0, 0$ şeklinde ifade edilir. Sonuçtan ve bağımsız ki-kare dağılımın oranından da görüyoruz ki

$\tilde{S}_{12}^2 - \tilde{S}_{1,2}^2$,$\tilde{S}_{1,2}^2 - (\tilde{S}_1^2 + \tilde{S}_2^2)$,$(\tilde{S}_1^2 + \tilde{S}_1^2)$ ve R_1 birbirinden tamamen bağımsızdır. Sonuç olarak,

$$R_1 = \frac{\tilde{S}_1^2}{\tilde{S}_1^2 + \tilde{S}_2^2} \sim Beta\left(\frac{n_1 - p_c - p_t - p_{d1}}{2}, \frac{n_2 - p_c - p_t - p_{d2}}{2}\right),$$

$$R_2 = \frac{\tilde{S}_1^2 + \tilde{S}_2^2}{\tilde{S}_{1,2}^2} \sim Beta\left(\frac{n_1 + n_2 - 2p_c - 2p_t - p_{d1} - p_{d2}}{2}, \frac{p_c}{2}\right),$$

$$\text{ve} \qquad F = \frac{\tilde{S}_{12}^2 - \tilde{S}_{1,2}^2/k}{\tilde{S}_{1,2}^2/l} \sim F_{k,l}(\mu' \tilde{P}_{12}\mu),$$

$k = p_c,\ l = n_1 + n_2 - p_c - p_{d_1} - p_{d_2} - 2p_t$ ve $\mu' = (c', d_1', d_2', t')\tilde{X}_{12}'$, birbirinden bağımsızdır. Son olarak $S_1^2/\sigma_1^2 = R_1 R_2 \tilde{S}_{1,2}^2(\sigma_1^2, \sigma_2^2)$ ve

$S_2^2/\sigma_2^2 = (1 - R_1)\, R_2 \tilde{S}_{1,2}^2(\sigma_1^2, \sigma_2^2)$ olduğu belirtilirse, T test değişkeni (3.8) formülü kullanılarak ve (3.6) tanımından da yararlanılarak

$$T = \left(\frac{\tilde{S}_{12}^2 - \tilde{S}_{1,2}^2}{\tilde{S}_{1,2}^2}(\sigma_1^2, \sigma_2^2)\right) R_2^{-1} [\tilde{s}^2(s_1^2/R_1, s_2^2/(1 - R_1))]^{-1}$$

şeklinde yazılabilir.

T'nin bu açılımında sadece ilk faktör H_0 hipotezinin test edilmesine bağlıdır ve bu terim merkezi olmayan F dağılıma sahiptir. Bu durum aynı zamanda T'nin yansız test olduğunun kanıtıdır. Sıfır hipotezi doğru olduğunda, merkezi olmayan F dağılımı, merkezi F dağılımına dönüşür böylece T'ye bağlı olasılıklar bilinmeyen parametre serbestliğidir. Çünkü test istatistiğinin gözlenmiş değeri 1'dir.

4. UYGULAMA

R istatistiksel hesaplama ve grafikleri için bilgisayar programı olup aynı zamanda programlama dilidir. Açık kaynak kodlu olan R, istatistiksel yazılım geliştirme ve veri analizi alanında kullanılmaktadır. Regresyon modellerinin karşılaştırılması sırasında varsayım bozulumu durumları için hazır veri seti üzerinde (Weerahandi,2005) uygulama yapılacak ve R programlama dilinde uygulanışı gösterilecektir.

A ve B ilaçları, hastalara kalp atım hızını düşürücü etkisini karşılaştırmak için uygulanıyor. Bir grup hastaya (n=8) her bir ilacın üç dozundan biri rassal olarak seçilerek veriliyor ve bir saat sonra kalp atım hızlarındaki düşüş (Y) kaydediliyor. Kaydedilen veriler Tablo 4.1.'de yer almaktadır.

Tablo 4.1. Yaş, Doz ve Kalp Atım Hızındaki Düşüş

	A İlacı							
Yaş	28	32	35	41	45	52	56	58
Doz	1.5	2.0	1.5	1.0	2.0	1.5	1.0	2.0
Y	8	16	14	11	21	18	20	29
	B İlacı							
Yaş	36	38	41	44	44	48	50	51
Doz	2.0	1.5	1.5	1.0	2.0	1.0	1.0	1.5
Y	16	17	14	8	22	14	7	23

Her bir ilaç için regresyon modeli aşağıdaki gibi yazılabilir:

$$Y = \beta_1 + \beta_2 Doz + \beta_3 Yaş + \varepsilon$$

Bu regresyon modellerine ilişkin varyansları eşit kabul ettiğimizde Bölüm 3.1.'de yer alan eşitlikten yararlanmak için regresyon modellerine ait artık kareler toplamları sırasıyla $s_1^2 = 15.45$ ve $s_2^2 = 73.17$ olarak bulunur. Ortak parametre olmadan bu iki regresyon modelinin artık kareler toplamı ise $s_{1,2}^2 = s_1^2 + s_2^2 = 88.62$ olur.

$$H_0 \ : \ \beta_{A2} = \beta_{B2}$$

Hipotezi altında parametreleri ve artık kareler toplamı (3.2) matris formundan hesaplanırsa;

$$\hat{\beta} = (X'X)^{-1}X'Y$$

$$X_{12} = \begin{bmatrix} 1 & 0 & 1.5 & 28 & 0 \\ 1 & 0 & 2 & 32 & 0 \\ 1 & 0 & 1.5 & 35 & 0 \\ 1 & 0 & 1 & 41 & 0 \\ 1 & 0 & 2 & 45 & 0 \\ 1 & 0 & 1.5 & 52 & 0 \\ 1 & 0 & 1 & 56 & 0 \\ 1 & 0 & 2 & 58 & 0 \\ 0 & 1 & 2 & 0 & 36 \\ 0 & 1 & 1.5 & 0 & 38 \\ 0 & 1 & 1.5 & 0 & 41 \\ 0 & 1 & 1 & 0 & 44 \\ 0 & 1 & 2 & 0 & 44 \\ 0 & 1 & 1 & 0 & 48 \\ 0 & 1 & 1 & 0 & 50 \\ 0 & 1 & 1.5 & 0 & 51 \end{bmatrix}_{16\times 5} \qquad Y = \begin{bmatrix} 8 \\ 16 \\ 14 \\ 11 \\ 21 \\ 18 \\ 20 \\ 29 \\ 16 \\ 17 \\ 14 \\ 8 \\ 22 \\ 14 \\ 7 \\ 23 \end{bmatrix}_{16\times 1} \qquad X'Y = \begin{bmatrix} 137 \\ 121 \\ 409 \\ 6360 \\ 5311 \end{bmatrix}_{5\times 1}$$

$$(X'X)^{-1} = \begin{bmatrix} 3.6373 & 2.7532 & -0.8285 & -0.0511 & -0.0355 \\ 2.7532 & 14.7478 & -1.6260 & -0.0049 & -0.2792 \\ -0.8285 & -1.6260 & 0.4893 & 0.0014 & 0.0209 \\ -0.0511 & -0.0049 & 0.0014 & 0.0011 & 6.32E-05 \\ -0.0355 & -0.2792 & 0.0209 & 6.32E-05 & 0.0056 \end{bmatrix}_{5\times 5}$$

$$\hat{\beta} = (X'X)^{-1}X'Y = \begin{bmatrix} -21.17 \\ -17.44 \\ 10.62 \\ 0.50 \\ 0.393 \end{bmatrix}_{5\times 1}$$

$$AKT = Y'Y - \hat{\beta}'X'Y = 4706 - 4603.049 = 102.951 = s_{12}^2$$

İlaç etkilerinin eşit varsayımından dolayı artık kareler toplamı artmıştır. Bu yüzden H_0 hipotezini test etmek için gözlenen değerlerden F istatistiği hesaplanabilir:

$$F = \frac{(s_{12}^2 - s_{1,2}^2)/p_t}{s_{1,2}^2/(n-k)} = \frac{14.33/1}{88.62/10} = 1.617$$

p değeri ise $p = 1 - H_{1.10}(1.617) = 0.233$ olur. $p = 0.233 > \alpha = 0.05$ olduğundan H_0 hipotezi kabul edilir yani iki ilacın etkileri istatistiksel olarak önemli derecede farklı değildir. Uygulamanın R programlama dilinde yazılmış kodları EK-1'de verilmiştir.

Örnek veri Bölüm 3.1'de incelenirken iki regresyon modelinin varyanslarının eşitliği varsayımı altında iki ilacın etkilerinin eşit $\mathrm{H_0} : \beta_{A2} = \beta_{B2}$ hipotezi test edildi. Şimdi ise Bölüm 3.2'den çıkan sonuçları uygulayarak ve varsayımı göz önüne almayarak hipotez yeniden test edilsin. Modellere ait artık kareler toplamları sırasıyla $s_1^2 = 15.45$ ve $s_2^2 = 73.17$ 'di. Böylece (3.7) eşitliği yardımıyla gözlemlenen anlamlılık değeri

$$p = 1 - E_R\left\{H_{1,10}\left[10\left(\tilde{s}_{12}^2\left(\frac{15.45}{R}\ ,\ \frac{73.17}{(1-R)}\right) - 1\right)\right]\right\},$$

$$= 0.295$$

R yazılım programı kullanılarak bulunur ve beklenen değer beta rassal değişkenine göre alınır

$$R \sim Beta\,(2.5, 2.5)\ .$$

Sonuç olarak, hata varyanslarının eşitliği varsayımı olmadan da önceki bölümden elde edilen aynı sonuca ulaşılır, $p = 0.295$ değeri α değerinden büyük olduğu için H_0 hipotezi kabul edilir, diğer bir deyişle, iki ilacın etkileri arasındaki farklılık istatistiksel olarak anlamlı değildir. Uygulamanın R programlama dilinde yazılmış kodları EK-2'de verilmiştir.

Bölüm 3.1 ve Bölüm 3.2'nin uygulanması için kullanılan örnek veri Bölüm 3.3'ün uygulanması için de kullanılsın. Bölüm 3.2'de regresyon denklemlerinin karşılaştırılması sırasında hata varyanslarının eşitlik varsayımı olmadan, A ve B ilaçlarının etkilerinin eşitliği test edilmişti, fakat yaş etkisinin iki regresyon modeli için farklı olabileceği düşünülmüştü. Şimdi bu değişkenin etkisinin her iki regresyon modeli içinde benzer olması varsayımı test edilsin. Sıfır hipotezini test etmek için uygun p değeri aşağıdaki formül yardımıyla hesaplanarak

$$p = 1 - E_{R_1,R_2}\left\{H_{1,11}\left[11R_2\tilde{s}^2\left(\frac{15,45}{R_1}, \frac{73,17}{(1-R_1)}\right)\right]\right\}$$

$p = 0.253$ olarak bulunur. Beklenen değer beta rassal değişkenine göre alınır

$$R_1 \sim Beta(2.5, 2.5) \text{ ve } R_2 \sim Beta(5.0, 0.5)$$

Bu p değeri, A ve B ilaçlarının etkilerinin eşitliğini veren sıfır hipotezini reddetmez. Böylece; benzer yaş etkisi varsayımı, daha önce bulunan sonucu etkilemediği sonucuna varılır. Uygulamanın R programlama dilinde yazılmış kodları EK-3'te verilmiştir.

5. SİMÜLASYON ÇALIŞMASI

İki farklı regresyon modellerine ait değişkenlerin dağılımları bilinmediği veya varyansların eşitliği varsayımı sağlanmadığı durumlarda regresyon modellerinin karşılaştırılmasında genelleştirilmiş p değeri yöntemi kullanılabilir. Bu bölümde, 3.Bölümde yer alan durumlarla karşılaşıldığında R programlama dili kullanılarak genelleştirilmiş p değerleri hesaplanmış ve regresyon modellerinin karşılaştırılması yapılmıştır. Karşılaştırmayı yaparken değişkenlere ait dağılımları önceden belirleyerek hesaplanan p değerleri tablolarda gösterilmiştir.

Simülasyonlar yapılırken regresyon modellerine ait değişkenler parametreleri önceden belirlenmiş dağılımlardan 8 birimlik örnekler şeklinde seçilmiştir. Seçilen bu değişkenler (3.4) matrisi uygulanarak 500 tekrar sonucu genelleştirilmiş p değerleri elde edilmiştir ve simülasyon çalıştırılarak tablolar oluşturulmuştur.

$$y_j = U_j c + V_j d_j + W_j t_j + \varepsilon_j, \quad j = 1,2$$

şeklinde iki regresyon modeli verilsin, bu iki modeli karşılaştırırken birinci modele ait V_1 değişkeni Normal(0,2.5), W_1 değişkeni ise Gama(2,2) dağılımdan seçilirken, diğer modele ait V_2 değişkeni Weibul(1,1.5), W_2 değişkeni ise Gama(2,2) dağılımından seçilsin. Sekiz birimlik örneklerden oluşan modelleri benzer parametreleri olmadığını kabul ederek karşılaştırırken, R programlama dilinde 500 tekrarlı simülasyonla genelleştirilmiş p değeri hesaplanmıştır. Tablo 5.1.'de simülasyon programı ile elde edilen genelleştirilmiş p değerleri gösterilmiştir. Bu örneğe ait R programlama dilinde yazılmış kodlar EK-4'de verilmiştir.

Tablo 5.1. Birim Sayısı 8, V_1 Değişkeni N(0,2.5), V_2 Değişkeni Weibull(1,1.5), W_1 ve W_2 Değişkenleri ise Gama(2,2) Dağılımlarından Oluşan, 500 Tekrar Sonucu Genelleştirilmiş p Değerleri

0.8563	0.9484	0.8493	0.9247	0.9790
0.5150	0.3005	0.6902	0.1937	0.4306
0.5561	0.9829	0.7998	0.7484	0.3430
0.7751	0.4706	0.9430	0.3357	0.9444
0.6093	0.5218	0.1002	0.1016	0.2897
0.0837	0.7206	0.3605	0.4529	0.4019
0.7548	0.4253	0.5839	0.6500	0.7650
0.9367	0.5483	0.7634	0.3068	0.1265
0.6241	0.5555	0.3008	0.4745	0.6148
0.0863	0.9463	0.3758	0.1375	0.9144
0.6310	0.8814	0.2994	0.4900	0.9682
0.0368	**0.0186**	0.8868	0.7659	0.1246
0.1015	0.6196	0.6317	0.2656	0.5202
0.7911	0.4672	0.3707	0.6429	0.3557
0.5453	0.6247	0.6601	0.9529	0.8344
0.8907	0.3795	0.6434	0.7684	0.8592
0.4738	0.7820	0.6650	0.8782	0.0842
0.2432	0.4913	0.8121	0.9177	0.7930
0.9265	0.2534	0.4391	0.3391	0.7036
0.1568	0.6859	0.6100	0.5789	0.2109

20×5

Regresyon modellerinin karşılaştırılması amacıyla genelleştirilmiş p değerine bakılarak ortak parametre olmadığı varsayımı altında farklı dağılımlardan oluşan değişkenler ile simülasyonlar yapılmıştır. H_0 hipotezinin doğruluğu altında 500 tekrar yapılarak 0.05 anlamlılık düzeyinde genelleştirilmiş p değerleri elde edilmiştir. 500 tekrarlı simülasyonun çalıştırılması sonucuyla elde edilen genelleştirilmiş p değerlerinin bulunduğu Tablo 5.1.'e bakıldığında sadece 2 tane değerin anlamlılık düzeyi 0.05'den küçük veya yakın olduğu gözlenmiştir. Diğer kalan 98 tane genelleştirilmiş p değerleri ise H_0 hipotezini kabul etmek için yeterlidir. Başka bir deyişle, Tablo 5.1.'i oluşturan regresyon modelleri arasında istatistiksel olarak anlamlı bir fark yoktur.

Tablo 5.2. Birim Sayısı 8, V_1 Değişkeni N(0,2.5), V_2 Değişkeni Weibull(1,1.5), W_1 ve W_2 Değişkenleri ise LogNormal(0,0.1) Dağılımlarından Oluşan, 500 Tekrar Sonucu Genelleştirilmiş p Değerleri

0.9362	0.1137	0.4018	0.0815	0.4526
0.9788	0.5223	0.1757	0.6772	0.7006
0.6610	0.5654	0.3545	0.6040	0.2361
0.6558	0.6768	0.7088	0.2635	0.6668
0.8827	0.9508	0.6986	0.8871	0.4842
0.5180	0.1423	0.6714	0.5441	**0.0341**
0.4457	0.3547	0.8740	0.3472	0.9961
0.6285	0.4653	0.1546	0.8764	0.3515
0.7102	0.9879	0.2539	0.2377	0.4868
0.7233	0.9113	0.3790	0.4037	0.4941
0.3965	0.6479	0.1398	0.6233	0.4138
0.9184	0.8661	0.4723	0.9529	0.3455
0.8813	0.8850	**0.0365**	0.9624	0.3428
0.4692	0.5769	0.1047	0.7413	0.0533
0.3938	0.6935	0.0971	0.9011	0.9527
0.2436	0.6093	0.8608	0.9542	0.3202
0.4707	0.3753	0.6043	0.0607	0.9203
0.4722	0.3093	0.9973	0.2476	0.5791
0.8546	0.0943	0.2412	0.4964	0.7725
0.0434	0.0518	0.8386	0.2167	**0.0090**

20×5

İki regresyon modeli karşılaştırılırken birinci modele ait V_1 değişkeni Normal(0,2.5), W_1 değişkeni ise LogNormal(0,0.1) dağılımlarından seçilirken, ikinci modele ait V_2 değişkeni Weibul(1,1.5), W_2 değişkeni ise LogNormal(0,0.1) dağılımından seçilsin. Sekiz birimlik örneklerden oluşan modelleri ortak parametreleri olmadığını kabul ederek karşılaştırırken, R programlama dilinde 500 tekrarlı simülasyonla genelleştirilmiş p değeri hesaplanmıştır. H_0 hipotezinin doğruluğu altında 0.05 anlamlılık düzeyinde simülasyon sonucu elde edilen genelleştirilmiş p değerlerinin bulunduğu Tablo 5.2.'e bakıldığında sadece 4 tane değerin anlamlılık düzeyi 0.05'den küçük veya yakın olduğu gözlenmiştir. Diğer kalan 96 tane genelleştirilmiş p değerleri ise H_0 hipotezini kabul etmek için yeterlidir. Başka bir deyişle, Tablo 5.2.'yi oluşturan regresyon modelleri arasında istatistiksel olarak anlamlı bir fark yoktur.

Tablo 5.3. Birim Sayısı 8, V_1 ve V_2 Değişkenleri N(0,2.5), W_1 ve W_2 Değişkenleri ise Weibull(1,5) Dağılımlarından Oluşan, 500 Tekrar Sonucu Genelleştirilmiş p Değerleri

0.6217	0.7250	0.4959	0.5243	0.1189
0.9128	0.9938	0.1674	0.2454	0.2014
0.6798	0.8106	0.1436	0.4663	0.9415
0.1453	0.0625	0.4723	0.2617	0.4192
0.9952	0.1209	0.1470	**0.0263**	0.1608
0.7877	0.9724	0.9954	0.9726	0.1370
0.7446	0.5069	0.3016	0.0636	0.0269
0.4272	0.9473	0.1694	0.3590	0.8369
0.1910	0.9378	0.7051	0.9600	0.7789
0.4443	0.9121	0.4454	0.2128	0.1108
0.5227	0.2430	0.5157	0.6084	0.2811
0.5078	0.6633	0.5282	0.4679	0.4514
0.5361	0.8192	0.9784	0.6186	0.4386
0.1779	0.8345	0.3745	0.5696	0.5707
0.4576	0.8397	0.6498	0.9870	0.1932
0.2740	**0.0138**	0.9244	0.5699	0.4586
0.8816	0.1509	0.6289	0.5774	0.2924
0.4902	0.5377	0.5176	0.1295	0.7977
0.8861	0.5714	0.2620	0.4893	0.2338
0.6535	0.0537	0.7014	0.9541	0.0620

20×5

İki regresyon modeli karşılaştırılırken birinci modele ait V_1 değişkeni Normal(0,2.5), W_1 değişkeni ise Weibul(1,5) dağılımlarından seçilirken, ikinci modele ait V_2 değişkeni Normal(0,2.5), W_2 değişkeni ise Weibul(1,5) dağılımından seçilsin. Sekiz birimlik örneklerden oluşan modelleri ortak parametreleri olmadığını kabul ederek karşılaştırırken, R programlama dilinde 500 tekrarlı simülasyonla genelleştirilmiş p değeri hesaplanmıştır. H_0 hipotezinin doğruluğu altında 0.05 anlamlılık düzeyinde simülasyon sonucu elde edilen genelleştirilmiş p değerlerinin bulunduğu Tablo 5.3.'e bakıldığında sadece 2 tane değerin anlamlılık düzeyi 0.05'den küçük veya yakın olduğu gözlenmiştir. Diğer kalan 98 tane genelleştirilmiş p değerleri ise H_0 hipotezini kabul etmek için yeterlidir. Başka bir deyişle, Tablo 5.3.'ü oluşturan regresyon modelleri arasında istatistiksel olarak anlamlı bir fark yoktur.

Tablo 5.4. **Birim Sayısı 8, V_1 ve V_2 Değişkenleri N(0,2.5), W_1 ve W_2 Değişkenleri ise Weibull(1,1.5) Dağılımlarından Oluşan, 500 Tekrar Sonucu Genelleştirilmiş p Değerleri**

0.3512	0.4068	0.6428	0.3124	0.3141
0.3949	0.8744	0.6421	0.7002	0.9103
0.3362	0.9166	0.4513	0.0915	0.1327
0.5499	0.8194	0.5271	0.4409	0.3994
0.2090	0.8087	0.0719	0.8366	0.3245
0.9079	0.1719	0.8942	0.8396	0.4447
0.3955	**0.0230**	0.8035	0.7218	0.3529
0.6096	0.2341	**0.0476**	0.1583	0.6691
0.3106	0.4332	0.4400	0.3399	0.9829
0.1565	0.5114	0.5387	0.8165	0.6498
0.7256	0.7241	0.4689	0.9818	0.2077
0.4568	0.4757	0.7523	0.2572	0.1254
0.2509	0.5401	0.6033	0.3051	0.5876
0.7442	0.0838	0.4401	0.5442	0.5946
0.2545	0.5058	0.4806	0.9364	0.7090
0.4349	0.3434	0.0603	0.1136	0.6262
0.2507	0.1810	0.8795	0.2688	0.2816
0.6462	0.8162	0.9644	0.7121	0.4897
0.5854	0.3057	0.1393	0.6102	0.9334
0.8522	0.2739	0.8232	0.1635	0.1710

20×5

İki regresyon modeli karşılaştırılırken birinci modele ait V_1 değişkeni Normal(0,2.5), W_1 değişkeni ise Weibul(1,1.5) dağılımlarından seçilirken, ikinci modele ait V_2 değişkeni Normal(0,2.5), W_2 değişkeni ise Weibul(1,1.5) dağılımından seçilsin. Sekiz birimlik örneklerden oluşan modelleri ortak parametreleri olmadığını kabul ederek karşılaştırırken, R programlama dilinde 500 tekrarlı simülasyonla genelleştirilmiş p değeri hesaplanmıştır. H_0 hipotezinin doğruluğu altında 0.05 anlamlılık düzeyinde simülasyon sonucu elde edilen genelleştirilmiş p değerlerinin bulunduğu Tablo 5.4.'e bakıldığında sadece 2 tane değerin anlamlılık düzeyi 0.05'den küçük veya yakın olduğu gözlenmiştir. Diğer kalan 98 tane genelleştirilmiş p değerleri ise H_0 hipotezini kabul etmek için yeterlidir. Başka bir deyişle, Tablo 5.4.'ü oluşturan regresyon modelleri arasında istatistiksel olarak anlamlı bir fark yoktur.

Tablo 5.5. Birim Sayısı 8, V_1 ve V_2 Değişkenleri N(0,2.5), W_1 ve W_2 Değişkenleri ise Gama(9,0.5) Dağılımlarından Oluşan, 500 Tekrar Sonucu Genelleştirilmiş p Değerleri

0.5855	0.8378	0.1099	0.5307	0.9569
0.9666	0.1453	0.3514	0.4958	0.5149
0.5183	0.6842	0.4408	0.6349	0.7684
0.3429	0.2207	0.4353	0.1939	0.5287
0.7931	0.1277	0.3449	0.1648	0.5794
0.2190	0.5311	0.0818	0.3802	0.6908
0.6637	0.5502	0.6817	0.8831	0.6387
0.9748	0.3979	0.5839	0.1657	0.1899
0.4438	0.4581	0.9608	0.6348	0.2271
0.6384	0.2654	0.2339	0.3753	0.6532
0.5094	0.5385	0.5070	0.1233	0.5852
0.7218	0.2046	0.5413	0.7129	0.1634
0.2456	0.3525	0.7753	0.9282	0.4027
0.3627	0.8557	0.6640	**0.0402**	**0.0460**
0.8611	0.1410	0.5164	0.3807	0.6450
0.2513	0.7480	0.9542	0.1612	0.9121
0.3221	0.4979	0.1229	0.4538	0.7499
0.4753	0.5583	0.2122	0.1015	0.7901
0.9552	0.5857	0.2674	0.7621	0.8563
0.3709	0.2110	0.1607	0.4079	0.9283

20×5

İki regresyon modeli karşılaştırılırken birinci modele ait V_1 değişkeni Normal(0,2.5), W_1 değişkeni ise Gama(9,0.5) dağılımlarından seçilirken, ikinci modele ait V_2 değişkeni Normal(0,2.5), W_2 değişkeni ise Gama(9,0.5) dağılımından seçilsin. Sekiz birimlik örneklerden oluşan modelleri ortak parametreleri olmadığını kabul ederek karşılaştırırken, R programlama dilinde 500 tekrarlı simülasyonla genelleştirilmiş p değeri hesaplanmıştır. H_0 hipotezinin doğruluğu altında 0.05 anlamlılık düzeyinde simülasyon sonucu elde edilen genelleştirilmiş p değerlerinin bulunduğu Tablo 5.5.'e bakıldığında sadece 2 tane değerin anlamlılık düzeyi 0.05'den küçük veya yakın olduğu gözlenmiştir. Diğer kalan 98 tane genelleştirilmiş p değerleri ise H_0 hipotezini kabul etmek için yeterlidir. Başka bir deyişle, Tablo 5.5.'i oluşturan regresyon modelleri arasında istatistiksel olarak anlamlı bir fark yoktur.

Tablo 5.6. Birim Sayısı 8, V_1 ve V_2 Değişkenleri N(0,2.5), W_1 ve W_2 Değişkenleri ise N(0,1.5) Dağılımlarından Oluşan, 500 Tekrar Sonucu Genelleştirilmiş p Değerleri

0.7144	0.4042	0.3474	0.0939	0.4310
0.7029	0.8466	0.4890	0.3035	0.3434
0.9775	0.4884	0.8283	0.3206	0.8089
0.3307	0.9637	0.1609	0.6701	0.7486
0.1152	0.5376	0.1823	0.8855	0.1950
0.5541	0.5426	0.5396	0.2448	0.8057
0.6196	0.8169	0.5398	0.2000	0.5284
0.8464	0.8729	0.9160	0.5322	0.9041
0.2203	0.5274	0.4056	0.2245	0.1429
0.7612	0.1846	0.6980	0.9872	0.2707
0.0999	0.2148	0.6434	0.4861	0.3516
0.1089	0.4228	0.6092	0.6121	0.9929
0.5894	0.3305	0.4306	0.6196	0.3853
0.8848	0.8869	0.9655	0.3194	0.5123
0.5783	0.4645	0.4691	0.2263	0.1355
0.9787	0.8525	0.3137	0.6805	0.3482
0.1803	0.3286	**0.0193**	0.1684	0.2289
0.2799	**0.0358**	0.7603	0.6865	0.5068
0.1269	0.9823	0.0733	0.1658	0.2556
0.7844	0.7881	0.7044	0.1782	0.6754

20×5

İki regresyon modeli karşılaştırılırken birinci modele ait V_1 değişkeni Normal(0,2.5), W_1 değişkeni ise Normal(0,1.5) dağılımlarından seçilirken, ikinci modele ait V_2 değişkeni Normal(0,2.5), W_2 değişkeni ise Normal(0,1.5) dağılımından seçilsin. Sekiz birimlik örneklerden oluşan modelleri ortak parametreleri olmadığını kabul ederek karşılaştırırken, R programlama dilinde 500 tekrarlı simülasyonla genelleştirilmiş p değeri hesaplanmıştır. H_0 hipotezinin doğruluğu altında 0.05 anlamlılık düzeyinde simülasyon sonucu elde edilen genelleştirilmiş p değerlerinin bulunduğu Tablo 5.6.'ya bakıldığında sadece 2 tane değerin anlamlılık düzeyi 0.05'den küçük veya yakın olduğu gözlenmiştir. Diğer kalan 98 tane genelleştirilmiş p değerleri ise H_0 hipotezini kabul etmek için yeterlidir. Başka bir deyişle, Tablo 5.6.'yı oluşturan regresyon modelleri arasında istatistiksel olarak anlamlı bir fark yoktur.

Tablo 5.7. Birim Sayısı 8, V_1 ve V_2 Değişkenleri N(0,2.5), W_1 ve W_2 Değişkenleri ise Beta(2,5) Dağılımlarından Oluşan, 500 Tekrar Sonucu Genelleştirilmiş p Değerleri

0.7647	0.2294	0.4774	0.6559	0.1831
0.3859	0.5398	0.5782	0.0875	**0.0036**
0.5428	0.9446	0.6963	0.3710	0.7371
0.2821	0.1270	0.2907	0.1758	0.6182
0.0259	0.8258	0.3023	0.3094	0.7815
0.6635	0.5293	0.6056	0.8606	0.1788
0.7144	0.1175	0.7791	0.3032	0.2564
0.6042	0.9275	0.1011	0.2218	0.5285
0.5250	0.4740	0.8946	0.4273	0.9915
0.5939	0.3219	0.3291	0.9279	0.0991
0.0824	0.1691	0.3842	0.6526	0.0760
0.4942	0.5708	0.8274	0.4259	0.4830
0.5639	0.1013	0.2615	0.2057	0.4270
0.0360	0.9762	0.1682	0.6039	0.9098
0.0294	0.9829	0.2853	0.7891	0.6095
0.0227	0.8653	0.4410	0.8892	0.2905
0.3747	0.7692	0.9182	0.6190	0.2188
0.2960	0.2053	0.6266	0.8022	**0.0195**
0.9466	0.8868	0.3586	0.9576	0.4579
0.5569	0.6886	0.5242	0.5913	0.1344

20×5

İki regresyon modeli karşılaştırılırken birinci modele ait V_1 değişkeni Normal(0,2.5), W_1 değişkeni ise Beta(2,5) dağılımlarından seçilirken, ikinci modele ait V_2 değişkeni Normal(0,2.5), W_2 değişkeni ise Beta(2,5) dağılımından seçilsin. Sekiz birimlik örneklerden oluşan modelleri ortak parametreleri olmadığını kabul ederek karşılaştırırken, R programlama dilinde 500 tekrarlı simülasyonla genelleştirilmiş p değeri hesaplanmıştır. H_0 hipotezinin doğruluğu altında 0.05 anlamlılık düzeyinde simülasyon sonucu elde edilen genelleştirilmiş p değerlerinin bulunduğu Tablo 5.7.'ye bakıldığında sadece 6 tane değerin anlamlılık düzeyi 0.05'den küçük veya yakın olduğu gözlenmiştir. Diğer kalan 94 tane genelleştirilmiş p değerleri ise H_0 hipotezini kabul etmek için yeterlidir. Başka bir deyişle, Tablo 5.7.'yi oluşturan regresyon modelleri arasında istatistiksel olarak anlamlı bir fark yoktur.

Tablo 5.8. Birim Sayısı 8, V_1 Değişkeni N(0,2.5), V_2 Değişkeni Weibull(1,1.5), W_1 ve W_2 Değişkenleri ise Beta(2,2) Dağılımlarından Oluşan, 500 Tekrar Sonucu Genelleştirilmiş p Değerleri

0.3088	0.1688	0.4836	0.3246	0.6730
0.8030	0.8865	0.2113	0.9180	0.2523
0.6734	0.9567	**0.0090**	0.3926	0.6730
0.8706	**0.0247**	0.9520	0.3855	0.6697
0.5490	0.1234	0.4766	0.7312	0.2899
0.3759	**0.0352**	0.2472	0.1529	0.7313
0.3194	0.3018	0.9921	0.4040	0.7465
0.6333	0.9137	0.0745	0.1196	0.9515
0.1996	0.1807	0.9561	0.1709	0.3779
0.1111	0.6573	0.2440	0.3582	0.7871
0.0995	0.1768	0.3879	0.8219	0.4293
0.5751	0.8116	0.7328	0.0669	0.1508
0.8853	0.6844	0.2898	0.7790	0.5142
0.7120	0.7411	0.2855	0.0570	0.5537
0.9080	0.4808	0.7872	0.3622	0.2621
0.1599	0.7139	0.1942	0.1776	0.3233
0.1779	0.8808	0.5922	0.9805	0.5338
0.6661	0.6441	0.7668	**0.0208**	0.2195
0.6169	0.4242	0.2207	0.9321	0.3071
0.3967	0.1150	0.0547	0.6493	0.6001

20×5

İki regresyon modeli karşılaştırılırken birinci modele ait V_1 değişkeni Normal(0,2.5), W_1 değişkeni ise Beta(2,2) dağılımlarından seçilirken, ikinci modele ait V_2 değişkeni Weibull(1,1.5), W_2 değişkeni ise Beta(2,2) dağılımından seçilsin. Sekiz birimlik örneklerden oluşan modelleri ortak parametreleri olmadığını kabul ederek karşılaştırılırken, R programlama dilinde 500 tekrarlı simülasyonla genelleştirilmiş p değeri hesaplanmıştır. H_0 hipotezinin doğruluğu altında 0.05 anlamlılık düzeyinde simülasyon sonucu elde edilen genelleştirilmiş p değerlerinin bulunduğu Tablo 5.8.'e bakıldığında sadece 4 tane değerin anlamlılık düzeyi 0.05'den küçük veya yakın olduğu gözlenmiştir. Diğer kalan 96 tane genelleştirilmiş p değerleri ise H_0 hipotezini kabul etmek için yeterlidir. Başka bir deyişle, Tablo 5.8.'i oluşturan regresyon modelleri arasında istatistiksel olarak anlamlı bir fark yoktur.

Tablo 5.9. Birim Sayısı 8, V_1 ve V_2 Değişkenleri N(0,2.5), W_1 ve W_2 Değişkenleri ise Üstel(1) Dağılımlarından Oluşan, 500 Tekrar Sonucu Genelleştirilmiş p Değerleri

0.0312	0.1752	0.9895	0.2456	0.5567
0.3816	0.7777	0.8786	0.2010	0.7445
0.1470	0.2289	0.6977	0.6394	**0.0142**
0.4415	0.6643	0.4451	0.3674	0.4909
0.4846	0.8392	0.8159	0.7370	0.9067
0.2840	0.2756	0.6219	0.1963	0.3563
0.7185	0.3122	0.2076	0.3912	0.9618
0.3664	0.2568	0.5426	0.2167	0.4882
0.7003	0.6745	0.5698	0.1824	0.2776
0.4132	0.8040	0.8485	0.7014	0.8031
0.0706	0.5946	0.5978	0.3057	0.8828
0.3182	0.7729	0.0707	0.9713	0.7306
0.1891	0.8907	0.3892	0.4287	0.2596
0.7126	0.4972	0.4712	0.0800	0.5583
0.2885	0.8672	0.6890	0.6169	0.4625
0.2154	0.3104	0.1089	0.4412	0.0656
0.8605	0.7012	0.2929	0.9901	0.3235
0.0196	0.8068	0.5333	0.1678	0.3579
0.7418	0.1376	0.4568	0.7255	0.2360
0.6470	0.2862	0.4244	0.3090	0.8297

20×5

İki regresyon modeli karşılaştırılırken birinci modele ait V_1 değişkeni Normal(0,2.5), W_1 değişkeni ise Üstel(1) dağılımlarından seçilirken, ikinci modele ait V_2 değişkeni Normal(0,2.5), W_2 değişkeni ise Üstel(1) dağılımından seçilsin. Sekiz birimlik örneklerden oluşan modelleri ortak parametreleri olmadığını kabul ederek karşılaştırılırken, R programlama dilinde 500 tekrarlı simülasyonla genelleştirilmiş p değeri hesaplanmıştır. H_0 hipotezinin doğruluğu altında 0.05 anlamlılık düzeyinde simülasyon sonucu elde edilen genelleştirilmiş p değerlerinin bulunduğu Tablo 5.9.'a bakıldığında sadece 3 tane değerin anlamlılık düzeyi 0.05'den küçük veya yakın olduğu gözlenmiştir. Diğer kalan 97 tane genelleştirilmiş p değerleri ise H_0 hipotezini kabul etmek için yeterlidir. Başka bir deyişle, Tablo 5.9.'u oluşturan regresyon modelleri arasında istatistiksel olarak anlamlı bir fark yoktur.

Tablo 5.10. Birim Sayısı 8, V_1 ve V_2 Değişkenleri N(0,2.5), W_1 ve W_2 Değişkenleri ise LogNormal(0,0.5) Dağılımlarından Oluşan, 500 Tekrar Sonucu Genelleştirilmiş p Değerleri

0.5818	0.6466	0.6702	0.3156	0.5557
0.7957	0.2681	0.9108	0.7020	0.6359
0.3405	0.4216	0.0843	0.9315	0.2650
0.4489	0.5973	0.8398	0.6777	0.9149
0.0009	0.4925	0.1849	0.7125	0.1120
0.6266	0.9452	0.5578	0.6981	0.8531
0.3834	**0.0251**	0.9707	0.1309	0.2926
0.3229	0.4835	0.8296	0.7896	0.8302
0.1101	0.1341	0.2919	0.5237	0.8029
0.5165	0.9147	0.2086	0.5528	0.4495
0.9122	0.8753	0.1960	0.9210	0.5844
0.5286	0.7260	0.8119	0.8963	0.4759
0.5513	0.2548	0.5428	0.5646	0.7802
0.6080	0.2848	0.6517	**0.0314**	0.3134
0.8779	0.1929	0.9674	0.2933	0.8108
0.9900	0.8899	0.5843	0.1300	0.5943
0.7374	0.2554	0.1827	0.7747	0.3935
0.3165	0.1160	0.8157	0.9968	0.6902
0.5260	0.2311	0.3044	0.8428	0.6688
0.9115	0.5316	0.2297	0.7374	0.7429

20×5

İki regresyon modeli karşılaştırılırken birinci modele ait V_1 değişkeni Normal(0,2.5), W_1 değişkeni ise LogNormal(0,0.5) dağılımlarından seçilirken, ikinci modele ait V_2 değişkeni Normal(0,2.5), W_2 değişkeni ise LogNormal(0,0.5) dağılımından seçilsin. Sekiz birimlik örneklerden oluşan modelleri ortak parametreleri olmadığını kabul ederek karşılaştırılırken, R programlama dilinde 500 tekrarlı simülasyonla genelleştirilmiş p değeri hesaplanmıştır. H_0 hipotezinin doğruluğu altında 0.05 anlamlılık düzeyinde simülasyon sonucu elde edilen genelleştirilmiş p değerlerinin bulunduğu Tablo 5.10.'a bakıldığında sadece 3 tane değerin anlamlılık düzeyi 0.05'den küçük veya yakın olduğu gözlenmiştir. Diğer kalan 97 tane genelleştirilmiş p değerleri ise H_0 hipotezini kabul etmek için yeterlidir. Başka bir deyişle, Tablo 5.10.'u oluşturan regresyon modelleri arasında istatistiksel olarak anlamlı bir fark yoktur.

Tablo 5.11. V_1 Değişkeni Gama(2,2), V_2 Değişkeni Beta(2,5), W_1 Değişkeni Gama(2,2) ve W_2 Değişkeni ise Beta(2,5) Dağılımlarından Oluşan, 10000 Tekrar Sonucu Testin Gücü

0.6500	0.7083	0.5832	0.8457	0.8565
0.9693	0.8446	0.9105	0.4455	0.9759
0.8386	0.9040	0.9690	0.9178	0.9780
0.8052	0.8184	0.9995	0.8123	0.9943

İki regresyon modeli karşılaştırılırken birinci modele ait V_1 değişkeni ve W_1 değişkeni Gama(2,2) dağılımlarından seçilirken, ikinci modele ait V_2 değişkeni ve W_2 değişkeni ise Beta(2,5) dağılımından seçilsin. Sekiz birimlik örneklerden oluşan modelleri ortak parametreleri olmadığını kabul ederek karşılaştırırken, R programlama dilinde 10000 tekrarlı simülasyonla testin gücü hesaplanmıştır. Regresyon katsayıları karşılaştırılan değişkenler farklı dağılımlardan seçilerek; yanlış olduğu bilinen katsayıların eşitliğini ifade eden H_0 hipotezi test edilmiştir. Güç, bir hipotez testinin isabetliliği için önemli bir kriterdir ve her zaman maksimize edilmek istenir. Genellikle testin gücünün 0.80'den yüksek olması yeterlidir. Tablo 5.11.'e bakıldığında testin gücünün neredeyse 1'e yaklaştığı değerler görülmektedir. Diğer bir deyişle, farklı dağılımlardan oluştuğu bilinen değişkenlerin katsayılarının eşitliğini ifade eden H_0 hipotezi reddedilerek yüksek test gücü hesaplanmıştır.

Tablo 5.12. V_1 Değişkeni N(0,2.5), V_2 Değişkeni N(0,2.5), W_1 Değişkeni Gama(2,2) ve W_2 Değişkeni ise Beta(2,5) Dağılımlarından Oluşan, 10000 Tekrar Sonucu Testin Gücü

0.9189	0.6563	0.6748	0.9250	0.7575
0.9072	0.6347	0.9899	0.8721	0.8583
0.7738	0.9952	0.7216	0.2221	0.9728
0.9853	0.9763	0.9857	0.9962	0.5986

İki regresyon modeli karşılaştırılırken birinci modele ait V_1 değişkeni N(0,2.5), W_1 değişkeni ise Gama(2,2) dağılımlarından seçilirken, ikinci modele ait V_2 değişkeni N(0,2.5), W_2 değişkeni ise Beta(2,5) dağılımından seçilsin. Sekiz birimlik örneklerden oluşan modelleri ortak parametreleri olmadığını kabul ederek karşılaştırırken, R programlama dilinde 10000 tekrarlı simülasyonla testin gücü hesaplanmıştır. Tablo 5.12.'ye bakıldığında testin gücünün neredeyse 1'e yaklaştığı değerler görülmektedir. Diğer bir deyişle, farklı dağılımlardan oluştuğu bilinen değişkenlerin katsayılarının eşitliğini ifade eden H_0 hipotezi reddedilerek yüksek test gücü hesaplanmıştır.

Tablo 5.13. **V_1 Değişkeni N(0,2.5), V_2 Değişkeni N(0,2.5), W_1 Değişkeni N(0,2.5) ve W_2 Değişkeni ise Üstel(1) Dağılımlarından Oluşan, 10000 Tekrar Sonucu Testin Gücü**

0.8893	0.9072	0.9918	0.7996	0.6790
0.8109	0.9642	0.8933	0.8768	0.9044
0.7845	0.9320	0.7899	0.8911	0.8431
0.8583	0.9747	0.6548	0.8289	0.7838

İki regresyon modeli karşılaştırılırken birinci modele ait V_1 değişkeni ve W_1 değişkeni N(0,2.5) dağılımlarından seçilirken, ikinci modele ait V_2 değişkeni N(0,2.5), W_2 değişkeni ise Üstel(8,1) dağılımından seçilsin. Sekiz birimlik örneklerden oluşan modelleri ortak parametreleri olmadığını kabul ederek karşılaştırırken, R programlama dilinde 10000 tekrarlı simülasyonla testin gücü hesaplanmıştır. Tablo 5.13.'e bakıldığında testin gücünün neredeyse 1'e yaklaştığı değerler görülmektedir. Diğer bir deyişle, farklı dağılımlardan oluştuğu bilinen değişkenlerin katsayılarının eşitliğini ifade eden H_0 hipotezi reddedilerek yüksek test gücü hesaplanmıştır.

Tablo 5.14. **V_1 Değişkeni N(0,2.5), V_2 Değişkeni N(0,2.5), W_1 Değişkeni Gama(9,0.5) ve W_2 Değişkeni ise Beta(2,5) Dağılımlarından Oluşan, 10000 Tekrar Sonucu Testin Gücü**

0.9407	0.1399	0.2676	0.8793	0.8912
0.8706	0.5634	0.7607	0.8623	0.7650
0.9680	0.5994	0.9321	0.9054	0.4329
0.9736	0.9125	0.9388	0.6617	0.5377

İki regresyon modeli karşılaştırılırken birinci modele ait V_1 değişkeni N(0,2.5), W_1 değişkeni Gama(9,0.5) dağılımlarından seçilirken, ikinci modele ait V_2 değişkeni N(0,2.5), W_2 değişkeni ise Beta(2,5) dağılımından seçilsin. Sekiz birimlik örneklerden oluşan modelleri ortak parametreleri olmadığını kabul ederek karşılaştırırken, R programlama dilinde 10000 tekrarlı simülasyonla testin gücü hesaplanmıştır. Tablo 5.14.'e bakıldığında testin gücünün neredeyse 1'e yaklaştığı değerler görülmektedir. Diğer bir deyişle, farklı dağılımlardan oluştuğu bilinen değişkenlerin katsayılarının eşitliğini ifade eden H_0 hipotezi reddedilerek yüksek test gücü hesaplanmıştır.

Tablo 5.15. **V_1 Değişkeni N(0,2.5), V_2 Değişkeni N(0,2.5), W_1 Değişkeni Weibull(1,1.5) ve W_2 Değişkeni ise Gama(2,2) Dağılımlarından Oluşan, 10000 Tekrar Sonucu Testin Gücü**

0.7929	0.8054	0.6123	0.7937	0.9376
0.5514	0.9809	0.9995	0.9004	0.8985
0.9792	0.9221	0.9453	0.9549	0.9809
0.9868	0.8056	0.9310	0.8749	0.3980

İki regresyon modeli karşılaştırılırken birinci modele ait V_1 değişkeni N(0,2.5), W_1 değişkeni Weibull(1,1.5) dağılımlarından seçilirken, ikinci modele ait V_2 değişkeni N(0,2.5), W_2 değişkeni ise Gama(2,2) dağılımından seçilsin. Sekiz birimlik örneklerden oluşan modelleri ortak parametreleri olmadığını kabul ederek karşılaştırırken, R programlama dilinde 10000 tekrarlı simülasyonla testin gücü hesaplanmıştır. Tablo 5.15.'e bakıldığında testin gücünün neredeyse 1'e yaklaştığı değerler görülmektedir. Diğer bir deyişle, farklı dağılımlardan oluştuğu bilinen değişkenlerin katsayılarının eşitliğini ifade eden H_0 hipotezi reddedilerek yüksek test gücü hesaplanmıştır.

Tablo 5.16. **V_1 Değişkeni Weibull(1,1.5), V_2 Değişkeni N(0,2.5), W_1 Değişkeni Weibull(1,1.5) ve W_2 Değişkeni ise LogNormal(0,0.1) Dağılımlarından Oluşan, 10000 Tekrar Sonucu Testin Gücü**

0.9070	0.7870	0.4139	0.9331	0.9970
0.6701	0.5949	0.9013	0.9853	0.8756
0.8626	0.9289	0.9584	0.8579	0.8301
0.9889	0.7253	0.8152	0.6447	0.6564

İki regresyon modeli karşılaştırılırken birinci modele ait V_1 değişkeni Weibull(1,1.5), W_1 değişkeni Weibull(1,1.5) dağılımlarından seçilirken, ikinci modele ait V_2 değişkeni N(0,2.5), W_2 değişkeni ise LogNormal(0,0.1) dağılımından seçilsin. Sekiz birimlik örneklerden oluşan modelleri ortak parametreleri olmadığını kabul ederek karşılaştırırken, R programlama dilinde 10000 tekrarlı simülasyonla testin gücü hesaplanmıştır. Tablo 5.16.'ya bakıldığında testin gücünün neredeyse 1'e yaklaştığı değerler görülmektedir. Diğer bir deyişle, farklı dağılımlardan oluştuğu bilinen değişkenlerin katsayılarının eşitliğini ifade eden H_0 hipotezi reddedilerek yüksek test gücü hesaplanmıştır.

Tablo 5.17. **V_1 Değişkeni Weibull(1,1.5), V_2 Değişkeni Beta(2,5), W_1 Değişkeni Beta(2,5) ve W_2 Değişkeni ise Weibull(1,1.5) Dağılımlarından Oluşan, 10000 Tekrar Sonucu Testin Gücü**

0.8295	0.9199	0.9842	0.9319	0.8218
0.8618	0.9264	0.9952	0.8975	0.8853
0.8713	0.2097	0.8812	0.6971	0.8114
0.6815	0.6108	0.6756	0.8790	0.9478

İki regresyon modeli karşılaştırılırken birinci modele ait V_1 değişkeni Weibull(1,1.5), W_1 değişkeni Beta(2,5) dağılımlarından seçilirken, ikinci modele ait V_2 değişkeni Beta(2,5), W_2 değişkeni ise Weibull(1,1.5) dağılımından seçilsin. Sekiz birimlik örneklerden oluşan modelleri ortak parametreleri olmadığını kabul ederek karşılaştırırken, R programlama dilinde 10000 tekrarlı simülasyonla testin gücü hesaplanmıştır. Tablo 5.17.'ye bakıldığında testin gücünün neredeyse 1'e yaklaştığı değerler görülmektedir. Diğer bir deyişle, farklı dağılımlardan oluştuğu bilinen değişkenlerin katsayılarının eşitliğini ifade eden H_0 hipotezi reddedilerek yüksek test gücü hesaplanmıştır.

Tablo 5.18. **V_1 Değişkeni Weibull(1,1.5), V_2 Değişkeni N(0,2.5), W_1 Değişkeni Beta(2,5) ve W_2 Değişkeni ise LogNormal(0,0.1) Dağılımlarından Oluşan, 10000 Tekrar Sonucu Testin Gücü**

0.6846	0.7996	0.9785	0.8267	0.9495
0.8636	0.8920	0.7395	0.3734	0.9218
0.6090	0.9281	0.4373	0.9998	0.8896
0.7810	0.9931	0.7361	0.9510	0.4724

İki regresyon modeli karşılaştırılırken birinci modele ait V_1 değişkeni Weibull(1,1.5), W_1 değişkeni Beta(2,5) dağılımlarından seçilirken, ikinci modele ait V_2 değişkeni N(0,2.5), W_2 değişkeni ise LogNormal(0,0.1) dağılımından seçilsin. Sekiz birimlik örneklerden oluşan modelleri ortak parametreleri olmadığını kabul ederek karşılaştırırken, R programlama dilinde 10000 tekrarlı simülasyonla testin gücü hesaplanmıştır. Tablo 5.18.'e bakıldığında testin gücünün neredeyse 1'e yaklaştığı değerler görülmektedir. Diğer bir deyişle, farklı dağılımlardan oluştuğu bilinen değişkenlerin katsayılarının eşitliğini ifade eden H_0 hipotezi reddedilerek yüksek test gücü hesaplanmıştır.

Tablo 5.19. **V_1 Değişkeni N(0,2.5), V_2 Değişkeni N(0,2.5), W_1 Değişkeni LogNormal(0,0.1) ve W_2 Değişkeni ise Üstel(1) Dağılımlarından Oluşan, 10000 Tekrar Sonucu Testin Gücü**

0.9381	0.9652	0.9397	0.7884	0.7546
0.9980	0.5642	0.4939	0.8917	0.2404
0.8482	0.9485	0.9265	0.8655	0.8255
0.9892	0.8700	0.7266	0.7380	0.7076

İki regresyon modeli karşılaştırılırken birinci modele ait V_1 değişkeni N(0,2.5), W_1 değişkeni LogNormal(0,0.1) dağılımlarından seçilirken, ikinci modele ait V_2 değişkeni N(0,2.5), W_2 değişkeni ise Üstel(1) dağılımından seçilsin. Sekiz birimlik örneklerden oluşan modelleri ortak parametreleri olmadığını kabul ederek karşılaştırırken, R programlama dilinde 10000 tekrarlı simülasyonla testin gücü hesaplanmıştır. Tablo 5.19.'a bakıldığında testin gücünün neredeyse 1'e yaklaştığı değerler görülmektedir. Diğer bir deyişle, farklı dağılımlardan oluştuğu bilinen değişkenlerin katsayılarının eşitliğini ifade eden H_0 hipotezi reddedilerek yüksek test gücü hesaplanmıştır.

Tablo 5.20. **V_1 Değişkeni N(0,2.5), V_2 Değişkeni N(0,2.5), W_1 Değişkeni Gama(2,2) ve W_2 Değişkeni ise Beta(2,2) Dağılımlarından Oluşan, 10000 Tekrar Sonucu Testin Gücü**

0.5617	0.8571	0.8864	0.9543	0.3763
0.6107	0.9672	0.8431	0.4464	0.6643
0.6308	0.7722	0.2250	0.7034	0.9790
0.8914	0.6282	0.9635	0.6885	0.8251

İki regresyon modeli karşılaştırılırken birinci modele ait V_1 değişkeni N(0,2.5), W_1 değişkeni Gama(2,2) dağılımlarından seçilirken, ikinci modele ait V_2 değişkeni N(0,2.5), W_2 değişkeni ise Beta(2,2) dağılımından seçilsin. Sekiz birimlik örneklerden oluşan modelleri ortak parametreleri olmadığını kabul ederek karşılaştırırken, R programlama dilinde 10000 tekrarlı simülasyonla testin gücü hesaplanmıştır. Tablo 5.20.'ye bakıldığında testin gücünün neredeyse 1'e yaklaştığı değerler görülmektedir. Diğer bir deyişle, farklı dağılımlardan oluştuğu bilinen değişkenlerin katsayılarının eşitliğini ifade eden H_0 hipotezi reddedilerek yüksek test gücü hesaplanmıştır.

Tablolarda verilen test güçleri arasından W_1 değişkenin Gama(2,2) dağılımından ve W_2 değişkeninin Beta(2,5) dağılımından seçilerek hesaplanan test gücünün çoğunluğu daha yüksek çıkarak diğer test güçleri arasında en iyi sonucu vermiştir.

6. SONUÇ

Regresyon analizinde yaygın olarak kullanılan klasik p değerleri farklı varsayımların bozulduğu durumlarda alternatif olarak geliştirilmiş genelleştirilmiş p değerlerinin kullanılabileceği görülmüştür. İki farklı regresyon modelinin karşılaştırılması sırasında oluşabilecek değişen varyans, benzer parametre olmama ve denklemleri oluşturan değişkenlerin dağılımının bilinmeyecek kadar az sayıda olması durumlarında genelleştirilmiş p değerinin doğru sonuç verdiği gözlenmiştir. Karşılaştırmaları yapmak için kullanılan genelleştirilmiş p değeri, R programlama dilinde 500 tekrar sonucu hesaplanmıştır. Simülasyonda, varsayımların sağlanmadığı durumlarda değişkenlerin dağılımlarını önceden belirleyerek, regresyon modelleri arasında fark yoktur hipotezi test edilmiştir. Simülasyon sonucu oluşturulan genelleştirilmiş p değerleri, $\alpha = 0.05$ anlamlılık düzeyinden büyük çıkarak H_0 hipotezi kabul edilmiştir. Ayrıca yeterli büyüklükte test gücü de hesaplanmıştır. Böylelikle hem doğru olan H_0 hipotezinin kabul edilmesi hem de elde edilen yüksek test gücü sonucunda genelleştirilmiş p değerinin kullanımının uygun olduğu görülmüştür. Sonuç olarak iki regresyon modelinin karşılaştırılması sırasında kullanılan genelleştirilmiş p değeri, eşit varyans, değişen varyans ve benzer parametre olmama durumlarında aynı sonucu vermiş, varsayımların bozulmasından etkilenmemiştir.

KAYNAKLAR

Birkes D, Dodge Y, (1993). Alternative Methods of Regression. John Wiley Sons. New York, pp. 80–140.

Field, Andy (2000). Discovering Statistics, Sage Publications.

Tsui, K.-W.,Weerahandi, S., (1989).” Generalized p-values in significance testing of hypotheses in the presence of nuisance parameters”. ***Journal of the American Statistical Association***,84, 602-607.

Weerahandi, S.(2005). *Exact Statistical Methods for Data Analysis. Springer*, New York.

Weerahandi, S. (1987), “Testing Regression Equality With Unequal Variances,” ***Econometrica***, 55, 1211-1215.

Wilcox RR, (1997). Introduction to Robust estimation and Hypothesis Testing. Academic Press. San Diego.

Yiğit, E. , Gamgam, H. (2011). “Homojen Olmayan Varyans Varsayımı Altında Ortalamaların Eşitliği İçin Bazı Test İstatistikleri ve Karşılaştırmaları”. ***Anadolu University Journal of Science And Technology –B, Theoretical Sciences***, 1:57-71.

EK-1 Varyanslar Eşit Olduğu Durumda Regresyon Modellerinin Karşılaştırılması

```
ones=rep(1,1,8)

zeros=rep(0,1,8)

v1=c(28,32,35,41,45,52,56,58)

w1=c(1.5,2,1.5,1,2,1.5,1,2)

y1=c(8,16,14,11,21,18,20,29)

v2=c(36,38,41,44,44,48,50,51)

w2=c(2,1.5,1.5,1,2,1,1,1.5)

y2=c(16,17,14,8,22,14,7,23)

sample1=c(ones,w1,v1)

matrix1=matrix(sample1,ncol=3)

betahat1=solve(t(matrix1)%*%matrix1)%*%t(matrix1)%*%y1

betahat1

yhat1=matrix1%*%betahat1

ss1=sum((y1-yhat1)^2)

ss1

sample2=c(ones,w2,v2)

matrix2=matrix(sample2,ncol=3)

betahat2=solve(t(matrix2)%*%matrix2)%*%t(matrix2)%*%y2

betahat2

yhat2=matrix2%*%betahat2
```

```
ss2=sum((y2-yhat2)^2)

ss2

ss1comma2=ss1+ss2

sample12=c(ones,zeros,zeros,ones,v1,zeros,zeros,v2,w1,w2)

matrix=matrix(sample12,ncol=5)

y=c(y1,y2)

beta.hat<- solve(t(matrix) %*% matrix) %*% t(matrix) %*% y

y.hat=matrix%*%beta.hat

ss12=sum((y-y.hat)^2)

F=((ss12-ss1comma2)/1)/(ss1comma2/10)

d=pf(F,1,10)

p=1-d

p
```

EK-2 Ortak Parametre Olmadığı Durumda Regresyon Modellerinin Karşılaştırılması

```
sample1<-c(1,1,1,1,1,1,1,1,28,32,35,41,45,52,56,58,1.5,2,1.5,1,2,1.5,1,2)

matrix1<-matrix(sample1,ncol=3)

 matrix1

y1<-c(8,16,14,11,21,18,20,29)

beta.hat1 <- solve(t(matrix1) %*% matrix1) %*% t(matrix1) %*% y1

 beta.hat1

y.hat1 <- matrix1 %*% beta.hat1

ss1<-sum((y1 - y.hat1)^2)

ss1

sample2<-c(1,1,1,1,1,1,1,1,36,38,41,44,44,48,50,51,2,1.5,1.5,1,2,1,1,1.5)

matrix2<-matrix(sample2,ncol=3)

matrix2

y2<-c(16,17,14,8,22,14,7,23)

 beta.hat2 <- solve(t(matrix2) %*% matrix2) %*% t(matrix2) %*% y2

beta.hat2

y.hat2 <- matrix2 %*% beta.hat2

ss2<-sum((y2 - y.hat2)^2)

ss2

d=rep(0,1,500)

sstilde=rep(0,1,500)
```

```
for(i in 1:500){

R=rbeta(1,2.5,2.5)

alpha1=ss1/R

alpha2=ss2/(1-R)

w1tilde=(1/sqrt(alpha1)) *c(1.5,2,1.5,1,2,1.5,1,2)

w2tilde=(1/sqrt(alpha2))*c(2,1.5,1.5,1,2,1,1,1.5)

wtilde=c(w1tilde,w2tilde)

v1tilde=(1/sqrt(alpha1))*c(28,32,35,41,45,52,56,58)

v2tilde=(1/sqrt(alpha2))*c(36,38,41,44,44,48,50,51)

sampletilde=c(1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,v1tilde,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,v2tilde,wtilde)

matrixtilde=matrix(sampletilde,ncol=5)

y1tilde=(1/sqrt(alpha1))*c(8,16,14,11,21,18,20,29)

y2tilde=(1/sqrt(alpha2))*c(16,17,14,8,22,14,7,23)

ytilde=c(y1tilde,y2tilde)

beta.hattilde<- solve(t(matrixtilde) %*% matrixtilde) %*% t(matrixtilde) %*%
ytilde

y.hattilde=matrixtilde%*%beta.hattilde

sstilde[i]=sum((ytilde-y.hattilde)^2)

d[i]=pf((10*(sstilde[i]-1)),1,10)

p=1-mean(d)

p
```

EK-3 Varyanslar Eşit Olmadığı Durumda Regresyon Modellerinin Karşılaştırılması

```
ones=rep(1,1,8)

zeros=rep(0,1,8)

sample1<-c(1,1,1,1,1,1,1,1,1.5,2,1.5,1,2,1.5,1,2,28,32,35,41,45,52,56,58)

matrix1<-matrix(sample1,ncol=3)

matrix1

y1<-c(8,16,14,11,21,18,20,29)

beta.hat1 <- solve(t(matrix1) %*% matrix1) %*% t(matrix1) %*% y1

beta.hat1

y.hat1 <- matrix1 %*% beta.hat1

ss1<-sum((y1 - y.hat1)^2)

ss1

sample2<-c(1,1,1,1,1,1,1,1,2,1.5,1.5,1,2,1,1,1.5,36,38,41,44,44,48,50,51)

matrix2<-matrix(sample2,ncol=3)

matrix2

y2<-c(16,17,14,8,22,14,7,23)

beta.hat2 <- solve(t(matrix2) %*% matrix2) %*% t(matrix2) %*% y2

beta.hat2

y.hat2 <- matrix2 %*% beta.hat2

ss2<-sum((y2 - y.hat2)^2)

ss2
```

```
d=rep(0,1,500)

sstilde=rep(0,1,500)

for(i in 1:500){

R1=rbeta(1,2.5,2.5)

R2=rbeta(1,5.0,0.5)

alpha1=ss1/R1

alpha2=ss2/(1-R1)

u1tilde=(1/sqrt(alpha1)) *c(1,1,1,1,1,1,1,1)

u2tilde=(1/sqrt(alpha2)) *c(1,1,1,1,1,1,1,1)

w1tilde=(1/sqrt(alpha1)) *c(1.5,2,1.5,1,2,1.5,1,2)

w2tilde=(1/sqrt(alpha2))*c(2,1.5,1.5,1,2,1,1,1.5)

wtilde=c(w1tilde,w2tilde)

v1tilde=(1/sqrt(alpha1))*c(28,32,35,41,45,52,56,58)

v2tilde=(1/sqrt(alpha2))*c(36,38,41,44,44,48,50,51)

y1tilde=(1/sqrt(alpha1))*c(8,16,14,11,21,18,20,29)

y2tilde=(1/sqrt(alpha2))*c(16,17,14,8,22,14,7,23)

ytilde=c(y1tilde,y2tilde)

sample1comma2tilde=c(u1tilde,zeros,zeros,u2tilde,v1tilde,zeros,zeros,v2tilde,w1t
ilde, zeros,zeros,w2tilde)

matrix1comma2tilde=matrix(sample1comma2tilde,ncol=6)

beta.hat1comma2tilde<- solve(t(matrix1comma2tilde) %*% matrix1comma2tilde)
%*% t(matrix1comma2tilde) %*% ytilde

y.hat1comma2tilde=matrix1comma2tilde%*%beta.hat1comma2tilde
```

ANADOLU ÜNİVERSİTESİ

```
ss1comma2tilde=sum((ytilde-y.hat1comma2tilde)^2)

sample12tilde=c(u1tilde,zeros,zeros,u2tilde,v1tilde,zeros,zeros,v2tilde,w1tilde,w2
tilde)

matrix12tilde=matrix(sample12tilde,ncol=5)

beta.hat12tilde<- solve(t(matrix12tilde) %*% matrix12tilde) %*%
t(matrix12tilde) %*% ytilde

y.hat12tilde=matrix12tilde%*%beta.hat12tilde

ss12tilde=sum((ytilde-y.hat12tilde)^2)

sstilde=ss12tilde-ss1comma2tilde

d[i]=pf((11*R2*sstilde),1,11)

}

p=1-mean(d)

p
```

EK-4 Ortak Parametreler Olmadığı Durumda Simülasyon Çalışması

```
ones=rep(1,1,8)

zeros=rep(0,1,8)

v1=rnorm(8,0,2.5)

w1= rgamma(8,2,2)

y1=rnorm(8,0,2.5)

v2= rweibull(8,1,1.5)

w2= rgamma(8,2,2)

y2= rweibull(8,1,1.5)

sample1=c(ones,v1,w1)

matrix1=matrix(sample1,ncol=3)

betahat1=solve(t(matrix1)%*%matrix1)%*%t(matrix1)%*%y1

betahat1

yhat1=matrix1%*%betahat1

ss1=sum((y1-yhat1)^2)

ss1

sample2=c(ones,v2,w2)

matrix2=matrix(sample2,ncol=3)

betahat2=solve(t(matrix2)%*%matrix2)%*%t(matrix2)%*%y2

betahat2

yhat2=matrix2%*%betahat2

ss2=sum((y2-yhat2)^2)
```

```
ss2
d=rep(0,1,500)
for(i in 1:500){
R=rbeta(1,2.5,2.5)
alpha1=ss1/R
alpha2=ss2/(1-R)
v1tilde=(1/sqrt(alpha1))*v1
w1tilde=(1/sqrt(alpha1))*w1
y1tilde=(1/sqrt(alpha1))*y1
v2tilde=(1/sqrt(alpha2))*v2
w2tilde=(1/sqrt(alpha2))*w2
y2tilde=(1/sqrt(alpha2))*y2
ytilde=c(y1tilde,y2tilde)
sampletilde=c(ones,zeros,zeros,ones,v1tilde,zeros,zeros,v2tilde,w1tilde,w2tilde)
matrixtilde=matrix(sampletilde,ncol=5)
betahattilde=solve(t(matrixtilde)%*%matrixtilde)%*%t(matrixtilde)%*%ytilde
yhattilde=matrixtilde%*%betahattilde
sstilde=sum((ytilde-yhattilde)^2)
d[i]=pf((10*(sstilde-1)),1,10)
}
p=1-mean(d)
p
```