

**VARYANSLARIN EŐİTSİZLİĐİ DURUMUNDA
REGRESYON PARAMETRELERİNİN
TEST EDİLMESİ
ÜZERİNE BİR ÇALIŐMA**

Filiz YILMAZ

**YÜKSEK LİSANS TEZİ
İSTATİSTİK**

**GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**KASIM 2007
ANKARA**

**VARYANSLARIN EŐİTSİZLİĐİ DURUMUNDA
REGRESYON PARAMETRELERİNİN
TEST EDİLMESİ
ÜZERİNE BİR ÇALIŐMA**

Filiz YILMAZ

**YÜKSEK LİSANS TEZİ
İSTATİSTİK**

**GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**KASIM 2007
ANKARA**

Filiz YILMAZ tarafından hazırlanan VARYANSLARIN EŞİTSİZLİĞİ DURUMUNDA REGRESYON PARAMETRELERİNİN TEST EDİLMESİ ÜZERİNE BİR ÇALIŞMA adlı bu tezin Yüksek Lisans tezi olarak uygun olduğunu onaylarım.

Prof. Dr. Müslim EKNİ
Tez Danışmanı, İstatistik Anabilim Dalı

Bu çalışma, jürimiz tarafından oy birliği ile İstatistik Anabilim Dalında Yüksek Lisans tezi olarak kabul edilmiştir.

Prof. Dr. Fikri ÖZTÜRK
İstatistik A.D, Ankara Üniversitesi

Prof. Dr. Müslim EKNİ
İstatistik A.D, Gazi Üniversitesi

Prof. Dr. Hamza GAMGAM
İstatistik A.D, Gazi Üniversitesi

Tarih: 02/11/2007

Bu tez ile G.Ü. Fen Bilimleri Enstitüsü Yönetim Kurulu Yüksek Lisans derecesini onamıştır.

Prof. Dr. Nermin ERTAN
Fen Bilimleri Enstitüsü Müdürü

TEZ BİLDİRİMİ

Tez içindeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edilerek sunulduğunu, ayrıca tez yazım kurallarına uygun olarak hazırlanan bu çalışmada orijinal olmayan her türlü kaynağa eksiksiz atıf yapıldığını bildiririm.

Filiz YILMAZ

**VARYANSLARIN EŞİTSİZLİĞİ DURUMUNDA REGRESYON
PARAMETRELERİNİN TEST EDİLMESİ
ÜZERİNE BİR ÇALIŞMA
(Yüksek Lisans Tezi)**

Filiz YILMAZ

**GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
Kasım 2007**

ÖZET

Bu çalışmada, bilinmeyen ve belki de eşit olmayan normal hata varyanslarına sahip bir genel doğrusal model üzerinde hipotez testi yapmak için tek örneğe dayalı olarak geliştirilmiş ve kısmi olarak regresyon parametrelerinin doğrusal fonksiyonlarının bir alt kümesinin elde edilmesinde kullanılan Wen ve ark.nın (2006) tarafından önerilen yöntem tanıtılmaktadır. Bu yöntem, varyans analizine de uygulanabilir. Önerilen yöntem ve regresyon parametrelerinin tahmin edicileri ikinci bölümde tanıtılmıştır. Üçüncü bölümde regresyon parametrelerinin tamamı ya da bir alt kümesi ile ilgili hipotez testleri, varyans analizine uygulanması ve genel doğrusal hipotezlere uzantısı anlatılmıştır. Dördüncü bölümde ise, ilk olarak simülasyon programı kullanılarak türetilmiş veri ile önerilen istatistik $\tilde{F}$ ve bilinen F testi sonuçları tartışılmış, genel doğrusal hipotezlere ilişkin bir örnek verilmiş ve son olarak da $\tilde{F}$ ve bilinen F 'in güçleri bir simülasyon çalışması ile karşılaştırılmıştır.

Bilim Kodu : 205.1.066
Anahtar Kelimeler : Regresyon analizi, homojen olmayan varyans, tek örneğe dayalı yöntem, hipotez testi, varyans analizi
Sayfa Adedi : 76
Tez Yöneticisi : Prof. Dr. Müslim EKNİ

**A STUDY ON TESTING REGRESSION PARAMETERS UNDER
HETEROSCEDASTICITY
(M.Sc. Thesis)**

Filiz YILMAZ

**GAZİ UNIVERSITY
INSTITUTE OF SCIENCE AND TECHNOLOGY**

November 2007

ABSTRACT

In this study, on a general linear model with unknown and possibly unequal normal error variances, One-sample sampling procedure offered by Wen, Chen and Chen (2006) is described to handle the hypothesis testing on all, partial or a subset of linear functions of regression parameters. The procedure can be applied to analysis of variance under various designs of experiments with unequal variances. The offered procedure and the estimators of regression parameters are described in second chapter. In third chapter hypothesis testing on all slope parameters or a subset of regression parameters, its application to analysis of variance and an extension to general linear hypothesis are studied. In fourth chapter, offered test statistic $\tilde{F}$ and usual F test results are compared with data derived from simulation program, an application to general linear hypothesis is given and finally the power of $\tilde{F}$ and the power of usual F are compared with a simulation study.

Science Code : 205.1.066
Key Words : Regression analysis, heteroscedasticity, one sample procedure, hypothesis testing, ANOVA
Page Number : 76
Adviser : Prof.Dr. Müslim EKNİ

TEŞEKKÜR

Çalışmalarım boyunca değerli yardım ve katkılarıyla beni yönlendiren hocam sayın Prof. Dr. Müslim EKNİ'ye, yine hiçbir yardım ve katkısını benden esirgemeyen hocam sayın Araş. Gör. Fikri GÖKPINAR'a, ayrıca öğrencilik hayatım süresince her zaman bana destek olan sevgili aileme teşekkürü bir borç bilirim.

İÇİNDEKİLER

	Sayfa
ÖZET.....	iv
ABSTRACT.....	v
TEŞEKKÜR.....	vi
İÇİNDEKİLER.....	vii
ÇİZELGELERİN LİSTESİ.....	ix
ŞEKİLLERİN LİSTESİ.....	x
SİMGELER VE KISALTMALAR.....	xi
1. GİRİŞ.....	1
2. TEK ÖRNEĞE DAYALI YÖNTEM VE TAHMİNLER.....	5
3. HİPOTEZ TESTLERİ.....	15
3.1. Regresyon Parametrelerinin Tamamının Testi.....	15
3.2. Regresyon Parametrelerinin Bir Alt Kümesinin Testi.....	17
3.3. İki Açıklayıcı Değişkenli Çoklu Regresyon.....	22
3.4. Tek Yönlü ANOVA' ya İlişkin Model.....	25
3.5. Kısmi Test ve İki Yönlü ANOVA Arasındaki İlişki.....	27
3.6. Genel Doğrusal Hipotezler.....	36
4. UYGULAMA.....	41
4.1. Tek Yönlü ANOVA için Uygulama.....	41
4.2. Genel Doğrusal Hipotezler için Sayısal Örnek.....	46
4.3. Simülasyon Çalışması.....	51
5. SONUÇ VE ÖNERİLER.....	57
KAYNAKLAR.....	59

EKLER.....	61
EK-1 U_i ve V_i ağırlıklarının sağlaması gereken koşulların gösterilmesi.....	62
EK-2 U istatistiği.....	65
EK-3 Kukla Değişkenler.....	67
EK-4 Tek Yönlü ANOVA için yapılan uygulamada kullanılan veri.....	73
EK-5 Simülasyon programı.....	75
ÖZGEÇMİŞ.....	76

ÇİZELGELERİN LİSTESİ

Çizelge	Sayfa
Çizelge 3.1. Regresyon ve ANOVA parametreleri arasındaki ilişki I.....	29
Çizelge 3.2. Regresyon ve ANOVA parametreleri arasındaki ilişki II.....	34
Çizelge 4.1. Örnek hacimleri I	42
Çizelge 4.2. $\bar{Y}_i$ ve S_i^2 değerleri.....	42
Çizelge 4.3. U_i ve V_i değerleri.....	42
Çizelge 4.4. Ağırlıklı örnek ortalamaları I.....	43
Çizelge 4.5. Belirlenen parametreler.....	47
Çizelge 4.6. Domuzların ağırlıklarındaki artış miktarı.....	48
Çizelge 4.7. Örnek hacimleri II.....	49
Çizelge 4.8. $\tilde{n}$, U_i ve V_i değerleri.....	50
Çizelge 4.9. Ağırlıklı örnek ortalamaları II.....	50
Çizelge 4.10 $\tilde{F}$ ve bilinen F istatistikleri için çeşitli n , varyans ve G_f değerlerine göre 2000 tekrarda, gerçekte yanlış olan H_0 hipotezi red oranları.....	53

ŞEKİLLERİN LİSTESİ

Şekil	Sayfa
Şekil.4.1. Eş.17.1'deki $\tilde{F}$ test istatistiğinin histogramı.....	44
Şekil.4.2. Eş.13'deki $\tilde{F}_R$ test istatistiğinin histogramı.....	46
Şekil.4.3. Eş.21.1'deki $\tilde{F}$ test istatistiğinin histogramı.....	51

SİMGELER VE KISALTMALAR

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

Simgeler	Açıklama
α	Anlamlılık düzeyi
$\tilde{F}_R$	İndirgenmiş $\tilde{F}$
k	Örnek sayısı
n_i	i 'nci örnek hacmi
$\tilde{n}$	Harmonik örnek ortalaması
r	Rank
U_i, V_i	Ağırlık katsayıları
ν	Serbestlik derecesi
$\tilde{Y}_i$	i 'nci ağırlıklı örnek ortalaması
θ	Lagrange çarpım vektörü
Kısaltmalar	Açıklama
ANOVA	Varyans Analizi
EKK	En Küçük Kareler
HCCME	Ayrı Varyanslı-tutarlı Kovaryans Matris
	Tahmin Edicisi
Gf	Gruplar Arası Fark
m.o.y.f.	Marjinal Olasılık Yoğunluk Fonksiyonu
o.y.f.	Olasılık Yoğunluk Fonksiyonu
o.o.y.f.	Ortak Olasılık Yoğunluk Fonksiyonu

1.GİRİŞ

Bu çalışmada, bilinmeyen ve belki de eşit olmayan normal hata varyanslarına sahip bir genel doğrusal model üzerinde nokta tahmini ve hipotez testi içeren istatistiksel problemlerin çözümü için tek örneğe dayalı bir yöntem geliştirmek ya da kısmi olarak regresyon parametrelerinin doğrusal fonksiyonlarının bir alt kümesini elde etmek için Wen ve ark. (2006) tarafından önerilen yöntem tanıtılmaktadır [1].

Yöntem; bağımsız değişkenler vektörü X_i ' de n_i büyüklüğündeki her bir örneği iki kısma ayırmaktadır. Birinci kısım ilk tahmin için $n_i - 1$ gözlemden oluşur. Diğeri ise en son tahminde tümsel kullanım için, kalan son gözlemden oluşur. Katsayıları ilk kısmın örnek bilgisine bağlı olan, her veri noktasındaki tüm örnek gözlemleri doğrusal bir kombinasyon olarak ifade edilebilir. Bu oluşturulan doğrusal kombinasyon, bilinmeyen ve eşit olmayan varyanslı bir genel doğrusal regresyon modeli altında parametrelerle ilgili yokluk hipotezlerini test ederken, test istatistiklerine ve parametre tahminlerine temel teşkil etmesi amacıyla kullanılır. Önerilen yönteminin avantajı, yokluk hipotezleri altında testlerin dağılımlarının belirli ve bilinmeyen varyanslardan tamamıyla bağımsız olmasıdır. Kritik değerler ve p değerleri çeşitli testler için bir bilgisayar programıyla elde edilebilir. Bu yaklaşım; regresyon modellerinde parametre tahminlerine, parametrelerin bir alt kümesi için yapılan teste ve nicel ya da nitel değişkenlerin ya da bunların kombinasyonlarının açıklayıcı değişken olarak kullanıldığı model seçimi için kısmi bir teste uygulanabilir. Ayrıca, tek yönlü varyans analizine, etkileşimli ya da etkileşimsiz iki yönlü varyans analizine ve üç ya da daha fazla yönlü modellere tamamen rasgeleleştirilmiş tasarım, rasgele blok düzeni, faktöriyel düzen, Latin kare düzenleri altında uygulanabilir. Bu yöntem istatistiksel veri analizi çalışmalarında, hata varyanslarının heterojen olması durumuyla karşılaşıldığında bir alternatif sağlayacaktır.

Regresyon modellerinde varyansların eşitsizliği problemine literatürde bakıldığında, eşit olmayan hata varyanslarının tahminlerine yönelik yapılmış araştırmalarda önerilen çözümlerin varolduğu görülür. 1975'te Horn genel doğrusal modelde hata

varyanslarının tahmin problemini ele almış ve hata varyansları için bir hemen hemen yansız tahmin edici önermiştir. Horn' un pozitif tanımlı hemen hemen yansız tahmin edicisi daha önceden çalışılmış MINQUE' nin (minimum norm quadratic unbiased estimator(s)) geliştirilmiştir. Horn ayrıca ekonomi ile ilgili çalışmalar yapmış ve ortalama hata kareyi azaltmıştır; ancak doğru hata varyanslarını test etmek için bir test yöntemi önermemiştir [2]. 1980' de White genel doğrusal modeldeki hata terimlerinin bilinmeyen varyans kovaryans matrisini test etmek amacıyla ayrı varyanslı-tutarlı kovaryans matris tahmin edicisi (HCCME; heteroskedasticity-consistent covariance matrix estimator) önermiştir. Asimptotik teori yoluyla doğrusal bir hipotezi test etmek için bir test istatistiği ortaya koymak amacıyla HCCME ve regresyon parametrelerinin en küçük kareler (EKK) tahmin edicilerini karesel bir formülde ikame etmiştir. HCCME' nin dezavantajı sonlu örnekte güvenilir olması ve küçük örneğe dayalı dağılımı olmamasıdır [3]. Cribari-Neto (2004), White' ın yöntemine dayalı olarak kovaryans matrisi tahmin edicisinin sonlu örnek davranışı üzerindeki büyük etkisini de hesaba katan bir HCCME önermiştir. Geliştirilmiş kovaryans matris tahmin edicisini HC4 tahmin edicisi olarak adlandırmış ve bunu Monte Carlo metodunu kullanarak t testlerine uygulamıştır. Bununla birlikte doğrusal hipotezler üzerinde bir test gerçekleştirirken, verdiği örnek istatistiği limitel dağılıma ya da koruyan halka yöntemleri (bootstrapping) temeline dayanmıştır. Tahmin edicinin küçük örneğe dayalı dağılımı verilmemiştir [4].

Wen ve arkadaşlarının sunduğu tek örneğe dayalı yöntem, genel doğrusal regresyon modelindeki hata varyanslarının eşitsizliği durumunda önerilmektedir. Bu amaçla, bir yöntem tanımlanmakta, regresyon parametrelerinin yansız tahmin edicileri sağlanmakta ve doğrusal hipotezler için test istatistikleri sunulmaktadır. Ayrıca, sıfır hipotezlerinin doğruluğu altında önerilen istatistiksel yöntemle küçük hacimli örnekler için de geçerli olabilen test istatistikleri de bulunmaktadır.

Klasik genel doğrusal regresyon modeli

$$Y_{ij} = \beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip} + e_{ij} \quad ; \quad e_{ij} \sim N(0, \sigma_i^2)$$

veya

$$Y_{ij} = P'_i \beta + e_{ij}, \quad i = 1, \dots, k, \quad j = 1, \dots, n_i \quad (1)$$

şeklinde gösterilmektedir [1]. Burada, Y_{ij} bağımlı değişken, $X_{i1}, \dots, X_{ip}$ 'ler açıklayıcı değişkenler ve $P'_i = (1, X_{i1}, \dots, X_{ip})$, i 'nci X veri noktasındaki açıklayıcı değişkenlerinin bilinen satır vektörü, $\beta = (\beta_0, \beta_1, \dots, \beta_p)'$ $p+1 \leq k$ 'ya bağlı olarak regresyon parametrelerinin bilinmeyen sütun vektörü ve $e_{ij} \sim N(0, \sigma_i^2)$ şeklinde dağılan bağımsız rasgele değişkenlerdir. Burada, σ_i^2 'ler bilinmeyen ve belki de eşit olmayan yığın varyanslarıdır. Ayrıca her X -veri noktası i 'de n_i tane $Y_{i1}, \dots, Y_{in_i}$ bağımlı değişken değeri gözlemlenir. Bir çok deneysel durumda e_{ij} 'lerin eşit varyansa sahip olduğunu belirten varsayım sağlanmaz. Araştırmalar, hata varyanslarının eşitsizliğinin, sonuçlarla ilişkili olasılıklar üzerinde ciddi bir etkiye sahip olabileceğini göstermiştir [5]. Bu gibi eksiklikleri ortadan kaldırmak için 1978' de Bishop, Stein-tipi iki aşamalı örnekleme yöntemi önermiştir. Bu yöntemde test istatistiklerinin dağılımları hata varyanslarından tamamıyla bağımsızdır. Testin gücü ve aralık uzunluğu istenilen düzeyde kontrol edilebilir; ancak büyük örnek tahminleri hariç kesin bir dağılımları yoktur. İki aşamalı yöntem, ikinci aşamada ilave gözlemler gerektiren bir yöntemdir [6]. Sonuç olarak; proje sonlandırma, bütçe sınırlaması, ilave örneklerin bulunamazlığı ve başka maliyet faktörlerinden dolayı pratikte veri analizi problemi için kullanışlı olmayabilir. Tek örneğe dayalı yöntem ise gelecekteki ilave gözlemleri içermez. Bu da istatistiklerin dağılımlarının bilinmeyen varyanstan tamamıyla bağımsız olmasını sağlar [1]. Bishop, 1976'da varyans analizi probleminde tek aşamalı örneği iki kısma ayırma fikrinden kısaca bahsetmiştir. Chen ve Lam 1989' da bu yöntemi hata varyanslarının eşitsizliği durumunda en büyük ortalama üzerindeki aralık tahminleri üzerinde çalışmış; ancak karmaşıklığından dolayı regresyon problemine genişletememiştir [7].

Bu çalışmada, bilinmeyen ve belki de eşit olmayan normal hata varyanslarına sahip bir genel doğrusal model üzerinde hipotez testi yapmak için tek örneğe dayalı geliştirilmiş bir yöntem ve regresyon parametrelerinin doğrusal fonksiyonlarının bir alt kümesini elde etmek için Wen ve ark.nın (2006) önerisi tanıtılmıştır [1]. Temel

yöntem ‘Tek Örneğe Dayalı Yöntem’ ve regresyon parametrelerinin tahminleri ikinci bölümde verilmiştir. Üçüncü bölümde regresyon parametrelerinin tamamı ve bir alt kümesi ile ilgili hipotez testi problemi ve varyans analizine uygulanması detaylı olarak anlatılırken genel doğrusal hipotezlere uzantısı verilmiştir. Dördüncü bölümde ise Tek Örneğe Dayalı yöntemin simülasyon programı kullanılarak türetilmiş veri ile uygulamaları yapılmıştır.

2. TEK ÖRNEĞE DAYALI YÖNTEM VE TAHMİNLER

Eş.1 altında, tahmin ve hipotez testi için tek örneğe dayalı yöntem aşağıdaki gibi tanımlanmaktadır.

Her X veri noktası P'_i ($i = 1, \dots, k$) 'den n_i (≥ 3) gözlemden oluşan bir rasgele örnek alınarak ve ilk $n_i - 1$ gözlem kullanılarak Y_{ij} ($j = 1, \dots, n_i - 1$) belirlenebilir. S_i^2, σ_i^2 'nin bilinen yansız tahmin edicisidir.

$i = 1, \dots, k$ için

$$\bar{Y}_i = \frac{\sum_{j=1}^{n_i-1} Y_{ij}}{n_i - 1} \quad \text{ve} \quad S_i^2 = \frac{\sum_{j=1}^{n_i-1} (Y_{ij} - \bar{Y}_i)^2}{n_i - 2}$$

'dir. Gözlemlerin ağırlıkları

$$\begin{aligned} U_i &= \frac{1}{n_i} + \frac{1}{n_i} \sqrt{\frac{1}{n_i - 1} [S_{[k]}^2 / S_i^2 - 1]}, \\ V_i &= \frac{1}{n_i} - \frac{1}{n_i} \sqrt{(n_i - 1) [S_{[k]}^2 / S_i^2 - 1]} \end{aligned} \quad (2)$$

olsun. U_i ve V_i katsayıları aşağıdaki koşulları sağlar [EK-1],

$$\begin{aligned} \text{i.} \quad & (n_i - 1)U_i + V_i = 1 \\ \text{ii.} \quad & (n_i - 1)U_i^2 + V_i^2 = S_{[k]}^2 / (n_i S_i^2) \end{aligned} \quad (3)$$

$S_{[k]}^2, S_1^2, \dots, S_k^2$ 'lerin içinde en büyük S_i^2 'dir.

İlk kısma ait $n_i - 1$ tane Y_{ij} ($j = 1, \dots, n_i - 1$) gözlemi ve son gözlem olan Y_{in_i} 'yi de

içeren ağırlıklı örnek ortalaması;

$$\tilde{Y}_{i.} = \sum_{j=1}^{n_i} W_{ij} Y_{ij}, \quad i = 1, \dots, k \quad (4)$$

'dir. Burada,

$$W_{ij} = \begin{cases} U_i & , 1 \leq j \leq n_i - 1, \\ V_i & , j = n_i \end{cases}$$

'dir [1]. Eş.2'de verilen U_i ve V_i katsayıları Chen ve Lam tarafından k -yığın aralık tahmin probleminde de kullanılmıştır [7]. Eş.4'te, n_i genişliğindeki i 'nci örnek iki kısma ayrılmaktadır. İlk kısım $n_i - 1$ gözlemden ve diğeri de kalan gözlemden oluşur. Tek aşamalı örnekleme yöntemi, Chen ve Chen tarafından örnek hacimlerinin eşit olduğu durum için n , n_0 ve $n - n_0$ olmak üzere iki kısma ayrılmıştır. $n_i = n$ ($i = 1, \dots, k$) olan yöntemde U_i ve V_i katsayıları ile ağırlıklı örnek ortalamasına dayalı yöntem kullanılmıştır. Örnek hacimlerinin eşit olarak kullanıldığı tek aşamalı örnekleme yönteminde kullanılan test istatistiği için Monte Carlo simülasyonu ile kritik değerleri gösteren tabloyu elde etmişlerdir [8]. Chen ve Lam n_0 ve $n - n_0$ olarak iki kısma ayrılan tek aşamalı örnekleme için, $n_0 = n - 1$ olarak seçildiğinde W_{ij} ağırlıklarının $1/n$ ' den mutlak farklarının toplamının minimum olduğunu göstermiştir. $1/n$ rasgele bir örnekteki bir gözlemin olağan ağırlığıdır.

$$\sum_{j=1}^n |W_{ij} - 1/n|.$$

Böylece örneğin ilk kısmı $n - 1$ ve ikinci kısmı da 1 gözlemden oluşur [7]. Bu durum örnek hacimleri eşit olmadığı durum için de geçerlidir. $n_{0i} = n_i - 1$ alındığında örnek varyansı ilk tahminde en büyük serbestlik derecesine sahip olur. Chen ve Lam ayrıca

U_i ve V_i ile ağırlıklandırılmış örnek ortalaması $\tilde{Y}_i$ 'nin, $1/n_i$ 'den büyük U_i ve $1/n_i$ 'den küçük V_i ile S_i^2 'ye bağlı olduğunu göstermiştir. Ağırlık V_i 'nin negatif bir değer olma ihtimali ödenmesi gereken bir bedeldir [1].

Ağırlıklı örnek ortalaması $\tilde{Y}_i$, $E(Y_{ij})$ için yansızdır. Gerçekten; $\tilde{Y}_i$ 'nin beklenen değeri alındığında,

$$E(\tilde{Y}_i) = E\left(\sum_{j=1}^{n_i} W_{ij} Y_{ij}\right)$$

$$E(\tilde{Y}_i) = \sum_{j=1}^{n_i} W_{ij} E(Y_{ij}) \text{ olur. Eş.1'den yararlanarak}$$

$$E(\tilde{Y}_i) = \sum_{j=1}^{n_i} W_{ij} E(P_i' \beta + e_{ij})$$

yazılabilir.

$$E(\tilde{Y}_i) = \sum_{j=1}^{n_i} W_{ij} (P_i' \beta + E(e_{ij}))$$

$$e_{ij} \sim N(0, \sigma_i^2) \text{ ve } \sum_{j=1}^{n_i} W_{ij} = (n_i - 1)U_i + V_i \text{ olduğundan}$$

$$E(\tilde{Y}_i) = ((n_i - 1)U_i + V_i) P_i' \beta \text{ ve Eş.3,i'den}$$

$$E(\tilde{Y}_i) = P_i' \beta$$

elde edilir.

$P_i' \beta$ ve Eş.3'te verilen koşullar, Eş.2'deki U_i ve V_i katsayılarının çözümünü sağlar.

Hata varyansları σ_i^2 'lerin homojen olması durumunda U_i ve V_i katsayıları $1/n_i$ 'ye yaklaşması nedeniyle, ağırlıklı örnek ortalaması da bilinen yansız örnek ortalamasına yaklaşır. Ağırlıklı örnek ortalaması $\tilde{Y}_i$, $P_i' \beta$ için yansız ve aynı zamanda tutarlıdır.

Chen, tutarlılığın varlığını $n_i \rightarrow \infty$ ve $S_i^2 \rightarrow \sigma_i^2$, $S_{[k]}^2 \rightarrow \sigma_{[k]}^2$ iken

$\text{var}(\tilde{Y}_{i.}) = \sum_{j=1}^{n_i} W_{ij} \sigma_i^2 = (S_{[k]}^2 / n_i)(\sigma_i^2 / S_i^2)$ ‘nin sıfıra yaklaştığını göstererek sağlamıştır

[9]. $\sigma_{[k]}^2$, en büyük σ_i^2 ’yi ifade etmektedir.

$\text{var}(\tilde{Y}_{i.}) = \sum_{j=1}^{n_i} W_{ij} \sigma_i^2 = (S_{[k]}^2 / n_i)(\sigma_i^2 / S_i^2)$ eşitliği gösterilebilir.

$\text{var}(\tilde{Y}_{i.}) = \sum_{j=1}^{n_i} W_{ij}^2 \text{var}(Y_{ij})$ ‘dir. Eş.1’den yararlanılarak

$\text{var}(\tilde{Y}_{i.}) = \sum_{j=1}^{n_i} W_{ij}^2 \text{var}(\mathbf{P}_i' \boldsymbol{\beta} + e_{ij})$ yazılabilir. Buradan

$\text{var}(\tilde{Y}_{i.}) = \sum_{j=1}^{n_i} W_{ij}^2 \text{var}(e_{ij})$

elde edilir. $e_{ij} \sim N(0, \sigma_i^2)$ ve $\sum_{j=1}^{n_i} W_{ij}^2 = (n_i - 1)U_i^2 + V_i^2$ olduğundan

$\text{var}(\tilde{Y}_{i.}) = ((n_i - 1)U_i^2 + V_i^2) \sigma_i^2$ ve Eş.3,ii’ deki eşitlik yardımı ile

$\text{var}(\tilde{Y}_{i.}) = (S_{[k]}^2 / (n_i S_i^2)) \sigma_i^2$

elde edilir. Gerekli düzenlemelerden sonra

$\text{var}(\tilde{Y}_{i.}) = (S_{[k]}^2 / n_i) (\sigma_i^2 / S_i^2)$

bulunur. Bu durumda

$\lim_{n_i \rightarrow \infty} \text{Var}(\tilde{Y}_{i.}) \approx \lim_{n_i \rightarrow \infty} \sum_{j=1}^{n_i} W_{ij} \sigma_i^2 \approx \lim_{n_i \rightarrow \infty} (S_{[k]}^2 / n_i) (\sigma_i^2 / S_i^2) \rightarrow 0$

‘dır.

Teoride ilk kısımda hangi $n_i - 1$ gözlemin kullanıldığı önemli değildir [1].

S_i^2 'lerin ($i = 1, \dots, k$) bilinmesi durumunda, $\tilde{Y}_i \sim N(P_i' \beta, \sigma_i^2 \sum_{j=1}^{n_i} W_{ij}^2)$ 'dir. T_i gibi bir rasgele değişken

$$T_i = \frac{\tilde{Y}_i - P_i' \beta}{S_i \sqrt{\sum_{j=1}^{n_i} W_{ij}^2}} = \frac{\tilde{Y}_i - P_i' \beta}{S_{[k]} / \sqrt{n_i}} \quad (5)$$

olarak tanımlanmaktadır [1]. $n_i \rightarrow \infty$ iken $S_i^2 \rightarrow \sigma_i^2$ olduğundan $\text{var}(\tilde{Y}_i) = (S_{[k]}^2 / n_i)(\sigma_i^2 / S_i^2)$ ifadesinde σ_i^2 / S_i^2 oranı “1”e yaklaşır. Bu durumda $\sqrt{\text{var}(\tilde{Y}_i)} = S_{[k]} / \sqrt{n_i}$ olur. Eş.5'teki paydalar, $S_{[k]} / \sqrt{n_i}$ ve $S_i \sqrt{\sum_{j=1}^{n_i} W_{ij}^2}$, Eş.3,ii gereği özdeşdir. Bu özdeşlik rasgele değişken T_i 'nin sahip olduğu olasılık dağılımı üzerinde bilinmeyen varyans σ_i^2 'nin etkisini yok etmektedir. T_i 'lerin dağılımlarının türetilmesi için Chen ve Chen' nin önerdiği, tek aşamalı ANOVA problemindeki yöntem kullanılabilir. Bu yöntemde, $S_i^2 = s_i^2 (i = 1, \dots, k)$ olmak üzere $T_i \sim N(0, \sigma_i^2 / s_i^2)$ 'dir. $h_i(s_i^2)$, S_i^2 'nin yoğunluk fonksiyonudur. S_i^2 , $v_i = n_i - 2$ serbestlik dereceli $(\sigma_i^2 / v_i) \chi_{(v_i)}^2$ dağılımına sahiptir [8].

$T_i \sim N(0, \sigma_i^2 / S_i^2 = s_i^2)$ iken T_i 'nin m.o.y.f.,

$$g_{(v_i)}(t_i) = \int_0^\infty N(0, \sigma_i^2 / s_i^2) h_i(s_i^2) ds_i^2$$

$$g_{(v_i)}(t_i) = \int_0^\infty \frac{s_i}{\sqrt{2\pi\sigma_i}} \exp\left\{-\frac{s_i^2 t_i^2}{2\sigma_i^2}\right\} h_i(s_i^2) ds_i^2$$

$$g_{(v_i)}(t_i) = \frac{\Gamma((v_i + 1)/2)}{\Gamma(v_i/2) \sqrt{\pi v_i}} \left(1 + \frac{t_i^2}{v_i}\right)^{-(v_i + 1)/2}$$

olarak tanımlanmaktadır. T_i 'nin m.o.y.f., Student t dağılımının o.y.f ile tamamen aynıdır. Böylece 0 ortalama ve (σ_i^2 / s_i^2) varyans ile koşullu normal dağılıma sahip olarak türetilen T_i istatistiklerinin her biri, Eş.3'te verilen koşulların sağlandığı bilgisi altında ve örnek varyansları S_i^2 'lerin bilinmesi durumunda, koşulsuz ve $v_i = n_i - 2$ serbestlik derecesi ile bağımsız Student t dağılımına sahiptir ve t_{vi} şeklinde gösterilir [1]. Bu yaklaşım ilk olarak Stein tarafından ele alınmıştır [10].

o.y.f.'ları sırasıyla $h_1(s_1^2), \dots, h_k(s_k^2)$ olan $S_1^2, \dots, S_k^2$ ' ler bağımsızdır. $T_1, \dots, T_k$ ' ların m.o.y.f.'ları $T_1, \dots, T_k$ ' ların o.o.y.f. integralinin alınmasıyla

$$\begin{aligned} g(t_1, \dots, t_k) &= \int_0^\infty \dots \int_0^\infty \prod_{i=1}^k [N(0, \sigma_i^2 / s_i^2) h_i(s_i^2)] \prod_{i=1}^k ds_i^2 \\ &= \prod_{i=1}^k \int_0^\infty N(0, \sigma_i^2 / s_i^2) h_i(s_i^2) ds_i^2 \\ &= \prod_{i=1}^k g_{vi}(t_i) \end{aligned}$$

elde edilmektedir [1]. Burada görüldüğü üzere, $T_1, \dots, T_k$ ' ların o.o.y.f.'ları m.o.y.f.'nın çarpımına eşittir. Sonuç olarak, $T_1, \dots, T_k$ ' ların sırasıyla $v_1, \dots, v_k$ serbestlik dereceli bağımsız $t_{v1}, \dots, t_{vk}$ dağılımına sahip oldukları görülmektedir. Eş.5'te verilen

$$\begin{aligned} T_i &= \frac{\tilde{Y}_i - \mathbf{P}_i' \boldsymbol{\beta}}{S_{[k]} / \sqrt{n_i}}, \\ \tilde{Y}_i &= \mathbf{P}_i' \boldsymbol{\beta} + S_{[k]} T_i / \sqrt{n_i} \end{aligned} \tag{5.1}$$

olarak yeniden yazılabilir. Burada $S_{[k]} T_i / \sqrt{n_i}$ ifadesi yerine ε_i kullanılırsa,

$$\tilde{Y}_i = \mathbf{P}_i' \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \dots, k \tag{5.2}$$

olur.

Ayrıca Eş.5.1, bilinen bir doğrusal modelin matris formu olarak da yazılabilir:

$$\tilde{Y} = X\beta + \varepsilon \quad (5.3)$$

Burada,

$$\tilde{Y} = \begin{bmatrix} \tilde{Y}_1 \\ \tilde{Y}_2 \\ \vdots \\ \tilde{Y}_k \end{bmatrix}, \quad X = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1p} \\ 1 & X_{21} & \cdots & X_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & X_{k1} & \cdots & X_{kp} \end{bmatrix} = \begin{bmatrix} P'_1 \\ P'_2 \\ \vdots \\ P'_k \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_k \end{bmatrix},$$

$\tilde{Y} = (\tilde{Y}_1, \dots, \tilde{Y}_k)'$, $k \times 1$ boyutlu bir sütun vektör, $X = (P'_1, \dots, P'_k)$, $k \times k$ boyutlu ($k = p+1$ 'den) bir veri matrisi, $\varepsilon = (\varepsilon_1, \dots, \varepsilon_k)'$, $k \times 1$ boyutlu bir sütun vektör, $\beta = (\beta_0, \beta_1, \dots, \beta_p)'$ ve $P'_i = (1, X_{i1}, \dots, X_{ip})$, Eş.1'de tanımlandığı gibidir.

Eş.5.2'den de görüldüğü gibi, $S_1^2, \dots, S_k^2$ bilinmesi durumunda $\varepsilon_i \sim (0, \frac{S_{[k]}^2}{n_i} \cdot \frac{\sigma_i^2}{S_i^2})$;

$i = 1, \dots, k$ ilişkisiz rasgele değişkenlerdir. Gerçekten de;

$\varepsilon_i = S_{[k]} T_i / \sqrt{n_i}$ eşitliğinden $S_1^2, \dots, S_k^2$ bilindiği durumda,

$E(\varepsilon_i | S_{[k]}) = E(S_{[k]} T_i / \sqrt{n_i})$ yazılabilir. Buradan,

$E(\varepsilon_i | S_{[k]}) = S_{[k]} E(T_i) / \sqrt{n_i}$

olur. $T_i \sim N(0, \sigma_i^2 / S_i^2)$ olduğundan

$E(\varepsilon_i | S_{[k]}) = 0$

bulunur.

ε_i 'nin varyansı,

$\text{var}(\varepsilon_i | S_{[k]}) = \text{var}(S_{[k]} T_i / \sqrt{n_i})$ veya

$$\text{var}(\varepsilon_i | S_{[k]}) = S_{[k]}^2 \text{var}(T_i) / n_i$$

yazılabilir. $T_i \sim N(0, \sigma_i^2 / S_i^2)$ olduğunda

$$\text{var}(\varepsilon_i | S_{[k]}) = \frac{S_{[k]}^2}{n_i} \cdot \frac{\sigma_i^2}{S_i^2}$$

elde edilir.

ε_i ve ε_j ($i \neq j$) arasındaki kovaryans,

$$\text{kov}(\varepsilon_i, \varepsilon_j | S_{[k]}) = E\{(\varepsilon_i - E(\varepsilon_i))(\varepsilon_j - E(\varepsilon_j))\} \text{ 'dır.}$$

$$\varepsilon_i = S_{[k]} T_i / \sqrt{n_i} \quad \text{ve} \quad \varepsilon_j = S_{[k]} T_j / \sqrt{n_j} \quad \text{eşitliklerinden,} \quad E(\varepsilon_i | S_{[k]}) = 0 \quad \text{ve}$$

$E(\varepsilon_j | S_{[k]}) = 0$ olduğu görülmektedir. Bu durumda

$$\text{kov}(\varepsilon_i, \varepsilon_j | S_{[k]}) = E\{(\varepsilon_i - 0)(\varepsilon_j - 0)\}$$

olur. Burada ε_i yerine $S_{[k]} T_i / \sqrt{n_i}$, ε_j yerine de $S_{[k]} T_j / \sqrt{n_j}$ konulursa,

$$\text{kov}(\varepsilon_i, \varepsilon_j | S_{[k]}) = \frac{S_{[k]}}{\sqrt{n_i}} \cdot \frac{S_{[k]}}{\sqrt{n_j}} E(T_i T_j)$$

bulunur. T_i ve T_j bağımsız rasgele değişkenler olduğundan

$$E(T_i T_j) = E(T_i) E(T_j) \text{ 'dir.}$$

$T_i \sim N(0, \sigma_i^2 / S_i^2)$ ve $T_j \sim N(0, \sigma_j^2 / S_j^2)$ iken

$E(T_i T_j) = 0$ olur. Bu durumda

$$\text{kov}(\varepsilon_i, \varepsilon_j | S_{[k]}) = 0$$

elde edilir.

Sonuç olarak,

- i. $E\{E(\varepsilon_i | S_{[k]})\} = 0$,
- ii. $E\{\text{cov}(\varepsilon_i, \varepsilon_j | S_{[k]})\} = 0, \quad i \neq j$
- iii. $\tau_i^2 = E\{\text{var}(\varepsilon_i | S_{[k]})\} = E\left\{\frac{S_{[k]}^2}{n_i} \cdot \frac{\sigma_i^2}{S_i^2}\right\} \approx E\left\{\frac{S_{[k]}^2}{\tilde{n}} \cdot \frac{\sigma_i^2}{S_i^2}\right\}$

(6)

yazılabilir. Eş.6,iii'de $\tilde{n}$, n_i ' ler büyük olduğunda, n_i 'lerin harmonik ortalaması

$$\tilde{n} = [(\sum_{i=1}^k 1/n_i)/k]^{-1} \text{ 'yi ifade etmektedir [1].}$$

τ_i^2 , bilinmeyen varyansların bir fonksiyonudur. $\sigma_{[k]}^2$, σ_i^2 'lerin $(i = 1, \dots, k)$ içinde en büyük olanıdır. $n_i \rightarrow \infty$ iken $P(S_i^2 \rightarrow \sigma_i^2) \rightarrow 1$ ve $P(S_{[k]}^2 \rightarrow \sigma_{[k]}^2) \rightarrow 1$ olan U istatistikleridir [9, EK-2]. Sonuç olarak, $E\{\text{var}(\varepsilon_i | S_{[k]})\} / E\{\sigma_{[k]}^2 / \tilde{n}\}$ oranı hemen hemen 1 ' e yaklaşır.

β ' nın bilinen EKK tahmin edicisi bir nokta tahmin yaklaşımı olarak düşünülebilir. Bu durumda

$$\hat{\beta} = (X'X)^{-1} X' \tilde{Y} \quad (7)$$

'dır [1].

Eş.7'de verilen $\hat{\beta}$, Eş.5.2'de β yerine konursa Eş.1 için aşağıdaki kestirim denklemini elde edilir.

$$\hat{Y}_i = P_i' \hat{\beta}, \quad i = 1, \dots, k \quad (7.1)$$

Eş.5.3'teki $\tilde{Y}$, Eş.7'de yerine yazılırsa

$$\begin{aligned} \hat{\beta} &= (X'X)^{-1} X'(X\beta + \varepsilon) \\ \hat{\beta} &= (X'X)^{-1} X'X\beta + (X'X)^{-1} X'\varepsilon \\ \hat{\beta} &= \beta + (X'X)^{-1} X'\varepsilon \end{aligned} \quad (7.2)$$

elde edilir. Buradan

$$E(\hat{\beta}) = E(\beta + (X'X)^{-1} X'\varepsilon)$$

$E(\hat{\beta}) = \beta + (X'X)^{-1}X'E(\varepsilon) \quad ; \quad \varepsilon = (\varepsilon_1, \dots, \varepsilon_k)'$, $\varepsilon_i \sim (0, \frac{S_{[k]}^2}{n_i} \cdot \frac{\sigma_i^2}{S_i^2})$ olduğundan

$E(\hat{\beta}) = \beta$ bulunur.

Yani $\hat{\beta}$, β 'nin yansız tahmin edicisidir. Varyans kovaryans matrisi aşağıdaki gibidir.

$$\Sigma_{\hat{\beta}} = (X'X)^{-1}X'DX(X'X)^{-1}.$$

Burada, $D = \{\tau_i^2\}$, k boyutlu köşegen matristir. $n_1, \dots, n_k$ büyük olduğunda (ya da eşit olarak $\tilde{n}$ büyük olduğunda), Eş.6,iii ve $n_i \rightarrow \infty$ iken $P(S_i^2 \rightarrow \sigma_i^2) \rightarrow 1$ ve $P(S_{[k]}^2 \rightarrow \sigma_{[k]}^2) \rightarrow 1$ olan U istatistiklerinden yararlanarak,

$$\tau_i^2 \approx E\left\{\frac{\sigma_{[k]}^2}{\tilde{n}}\right\} = \tau^2$$

yazılabilir. Asimptotik olarak Gauss Markov teoremi uygulanırsa

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

$$\Sigma_{\hat{\beta}} = (X'X)^{-1}X'\tau^2X(X'X)^{-1}$$

$$\Sigma_{\hat{\beta}} = \tau^2(X'X)^{-1}X'X(X'X)^{-1}$$

$$\Sigma_{\hat{\beta}} \approx \tau^2(X'X)^{-1} \quad (7.3)$$

bulunur. ε_i 'nin varyansı asimptotik olarak bilinmeyen değer τ^2 'ye eşittir. τ^2 , tüm varyansların $(\sigma_1^2, \dots, \sigma_k^2)$ bir fonksiyonudur [1].

Sonuçta; β 'nın Eş.7'de verilen tahmini, yansız , güçlü olarak tutarlı ve asimptotik olarak minimum varyans kovaryans matrisine sahiptir [1, 11, 12].

3. HİPOTEZ TESTLERİ

3.1. Regresyon Parametrelerinin Tamamının Testi

Eş.1’de verilen klasik genel doğrusal regresyon modelinde, açıklayıcı değişkenlerin tamamının anlamlı olmadığı sıfır hipotezi en az bir tanesinin anlamlı olduğunu ifade eden alternatif hipoteze karşı test edilir.

$$H_0 : \beta_1 = \dots = \beta_p = 0$$

$$H_1 : \text{en az bir } \beta_j \text{ sıfırdan farklıdır, } j = 1, \dots, p .$$

Burada açıklayıcı değişkenlerin tamamının Eş.1’de verilen tam modele anlamlı bir katkı sağlayıp sağlamadıklarını test etmek için genel bir test düşünülmektedir.

$i = 1, \dots, k$ iken Eş.4’ten elde edilen ağırlıklı örnek ortalamaları, Eş.5.2’de verilen tek örneğe dayalı doğrusal regresyon modeli altında,

$$\tilde{Y}_{..} = \sum_{i=1}^k \tilde{Y}_i / k \text{ 'dır. } H_0 : \beta_1 = \dots = \beta_p = 0 \text{ hipotezi için}$$

$$\tilde{F} = \sum_{i=1}^k \left(\frac{\hat{Y}_i - \tilde{Y}_{..}}{S_{[k]} / \sqrt{n_i}} \right)^2 \quad (8)$$

test istatistiğidir [1]. Kestirim denklemi Eş.7.1 ve Eş.5.2’deki model, Eş.8’de yerine konulursa $\tilde{F}$ yeniden yazılabilir.

$$\tilde{F} = \sum_{i=1}^k \left(\frac{\mathbf{P}'_i \hat{\beta} - \mathbf{P}'_i \beta - (\tilde{Y}_{..} - \sum_{j=1}^k \mathbf{P}'_j \beta / k) + \mathbf{P}'_i \beta - \sum_{j=1}^k \mathbf{P}'_j \beta / k}{S_{[k]} / \sqrt{n_i}} \right)^2$$

$$\bar{\mathbf{P}}' = \sum_{j=1}^k \mathbf{P}'_j / k \text{ alınır,}$$

$$\tilde{F} = \sum_{i=1}^k \left(\frac{\mathbf{P}'_i(\hat{\beta} - \beta) - \sum_{j=1}^k (\tilde{Y}_{j.} - \mathbf{P}'_j \beta) / k + (\mathbf{P}'_i - \bar{\mathbf{P}}') \beta}{S_{[k]} / \sqrt{n_i}} \right)^2$$

$$\tilde{F} = \sum_{i=1}^k \left(\frac{\mathbf{P}'_i(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\varepsilon} - \sum_{j=1}^k \boldsymbol{\varepsilon}_j / k + (\mathbf{P}'_i - \bar{\mathbf{P}}') \beta}{S_{[k]} / \sqrt{n_i}} \right)^2$$

yazılabilir. $\{b_{ij}\} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'$ alındığında

$$\tilde{F} = \sum_{i=1}^k \left(\frac{\sum_{j=1}^k b_{ij} \boldsymbol{\varepsilon}_j - \sum_{j=1}^k \boldsymbol{\varepsilon}_j / k + (\mathbf{P}'_i - \bar{\mathbf{P}}') \beta}{S_{[k]} / \sqrt{n_i}} \right)^2$$

$$\tilde{F} = \sum_{i=1}^k \left(\sqrt{n_i} \left(\sum_{j=1}^k b_{ij} \frac{T_j}{\sqrt{n_j}} - \frac{1}{k} \sum_{j=1}^k \frac{T_j}{\sqrt{n_j}} \right) + \frac{(\mathbf{P}'_i - \bar{\mathbf{P}}') \beta}{S_{[k]} / \sqrt{n_i}} \right)^2 \quad \text{ve}$$

üzerinde gerekli düzenlemeler sonucunda

$$\tilde{F} = \sum_{i=1}^k \left(\sqrt{n_i} \sum_{j=1}^k \left(b_{ij} - \frac{1}{k} \right) \frac{T_j}{\sqrt{n_j}} + \frac{(\mathbf{P}'_i - \bar{\mathbf{P}}') \beta}{S_{[k]} / \sqrt{n_i}} \right)^2 \quad \text{olur.}$$

$H_0 : \beta_1 = \dots = \beta_p = 0$ doğruluğu altında

$$\tilde{F} = \sum_{i=1}^k \left(\sqrt{n_i} \sum_{j=1}^k \left(b_{ij} - \frac{1}{k} \right) \frac{T_j}{\sqrt{n_j}} \right)^2 \quad (9)$$

bulunur. $\tilde{F}$ istatistiği, $v_1, \dots, v_k$ serbestlik dereceli bağımsız, Student t dağılımına sahip $T_1, \dots, T_k$ 'ların doğrusal kombinasyonlarının bir karesel fonksiyonudur. Test istatistiğinin p değeri, belirlenen bir α 'dan küçük ise $H_0 : \beta_1 = \dots = \beta_p = 0$ hipotezi reddedilir. p değeri, bir simülasyon programı yardımıyla hesaplanabilir [1].

3.2. Regresyon Parametrelerinin Bir Alt Kümesinin Testi

Genel doğrusal regresyon modelinde, parametrelerin bazı alt kümelerinin anlamlılıklarının test edilmesi de oldukça önemlidir. Bu durum bizi çoklu regresyon analizinde model seçme yöntemlerinden değişken eleme (backward selection), değişken ekleme (forward selection) veya değişken ekleme ve eleme yöntemlerine (stepwise selection) götürür [13].

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_q \\ \text{---} \\ \beta_{q+1} \\ \vdots \\ \beta_p \end{bmatrix}$$

olarak iki alt kümeye ayrılan parametrelerin kısmi yokluk hipotezi,

$$H_0^* : \beta_{q+1} = \dots = \beta_p = 0, \quad q < p \quad (10)$$

'dir. Bu hipotez, modelde yer alan $X_{q+1}, \dots, X_p$ açıklayıcı değişkenlerinin Eş.1'deki modele katkılarının anlamlı olmadığını ifade eder. Eş.1'de verilen tüm açıklayıcı değişkenleri içeren tam modele ilişkin indirgenmiş model

$$e_{ij} \sim N(0, \sigma_i^2); \quad i = 1, \dots, k, \quad j = 1, \dots, n_i \text{ iken}$$

$$Y_{ij} = \beta_0 + \beta_1 X_{i1} + \dots + \beta_q X_{iq} + e_{ij} \quad (11)$$

veya

$$Y_{ij} = Q_i' \beta_R + e_{ij}, \quad (11.1)$$

şeklinde ifade edilebilir [1]. Bağımlı değişken Y_{ij} 'ler Eş.1'de verildiği gibidir.

$Q'_i = (1, X_{i1}, \dots, X_{iq})$, açıklayıcı değişkenlerin $1 \times (q+1)$ boyutlu satır vektörüdür.

$\beta_R = (\beta_0, \beta_1, \dots, \beta_q)$ ise $(q+1) \times 1$ boyutlu, bilinmeyen indirgenmiş regresyon parametrelerinin sütun vektörüdür. Tanımlanan tek örneğe dayalı yöntem, Eş.5.2'ye benzer biçimde kullanılarak,

$$\tilde{Y}_i = Q'_i \beta_R + \varepsilon_i^*, \quad i = 1, \dots, k \quad (11.2)$$

olarak yazılabilir. Burada

$$\varepsilon_i^* = T_i^* S_{[k]} / \sqrt{n_i} \quad \text{ve}$$

$$T_i^* = \frac{\tilde{Y}_i - Q'_i \beta_R}{S_{[k]} / \sqrt{n_i}}, \quad i = 1, \dots, k \quad (11.3)$$

'dır. T_i^* 'lar ($i = 1, \dots, k$) $v_1, \dots, v_k$ serbestlik dereceli $t_{v_1}, \dots, t_{v_k}$ bağımsız Student t dağılımına sahiptir.

EKK ile β_R 'nin tahmini Eş.7'ye benzer olarak elde edilebilir. Burada

$$\tilde{Y} = X_1 \beta_R + \varepsilon^*$$

'dir. $\hat{\beta}_R$,

$$\begin{aligned} \hat{\beta}_R &= (X_1' X_1)^{-1} X_1' \tilde{Y} \\ \hat{\beta}_R &= (X_1' X_1)^{-1} X_1' (X_1 \beta_R + \varepsilon^*) \\ \hat{\beta}_R &= \beta_R + (X_1' X_1)^{-1} X_1' \varepsilon^* \end{aligned}$$

olarak elde edilir [1]. $X_1 = (Q'_1, \dots, Q'_k)$, $k \times (q+1)$ boyutlu indirgenmiş veri matrisidir. $\varepsilon^* = (\varepsilon_1^*, \dots, \varepsilon_k^*)$, $k \times 1$ boyutlu bir vektör ve $S_1^2, \dots, S_k^2$ bilinmesi

durumunda $\varepsilon_i^* \sim \left(0, \frac{S_{[k]}^2}{n_i} \cdot \frac{\sigma_i^2}{S_i^2}\right)$; $i = 1, \dots, k$ ilişkisiz rasgele değişkenlerdir. $\tilde{Y}_{i.}$, Eş.4

kullanılarak bulunan ağırlıklı örnek ortalamasıdır. Eş.11.1'de verilen indirgenmiş modelin kestirim denklemi,

$$\hat{Y}_{i.}^* = Q_i' \hat{\beta}_R, \quad i = 1, \dots, k \quad (11.4)$$

'dir.

Eş.10'da verilen $H_0^* : \beta_{q+1} = \dots = \beta_p = 0$ için bir test istatistiği, aşağıdaki gibi düşünülebilir.

$$\tilde{F}_R = \sum_{i=1}^k \left(\frac{\hat{Y}_{i.} - \hat{Y}_{i.}^*}{S_{[k]} / \sqrt{n_i}} \right)^2 \quad (12)$$

$\tilde{F}_R$, kareler toplamı olarak yorumlanabilir [1]. Eş.1'de verilen tüm açıklayıcı değişkenleri içeren modelin kestirimi $\hat{Y}_{i.} = P_i' \hat{\beta}$ ile Eş.11.1'de verilen indirgenmiş modelin kestirimi $\hat{Y}_{i.}^* = Q_i' \hat{\beta}_R$ kullanılarak test istatistiği $\tilde{F}_R$, aşağıdaki gibi yeniden yazılabilir.

$$\tilde{F}_R = \sum_{i=1}^k \left(\frac{P_i' \hat{\beta} - Q_i' \hat{\beta}_R}{S_{[k]} / \sqrt{n_i}} \right)^2$$

Eşitliğin sağ tarafına $P_i' \beta$ ve $Q_i' \beta_R$ eklenip çıkarılırsa,

$$\tilde{F}_R = \sum_{i=1}^k \left(\frac{P_i'(\hat{\beta} - \beta) - Q_i'(\hat{\beta}_R - \beta_R) + P_i' \beta - Q_i' \beta_R}{S_{[k]} / \sqrt{n_i}} \right)^2$$

elde edilir.

$\hat{\beta} = \beta + (X'X)^{-1} \varepsilon$ ve $\hat{\beta}_R = \beta_R + (X_1'X_1)^{-1} X_1' \varepsilon^*$ idi. Buradan, $\hat{\beta} - \beta = (X'X)^{-1} \varepsilon$ ve $\hat{\beta}_R - \beta_R = (X_1'X_1)^{-1} X_1' \varepsilon^*$ alınırsa

$$\tilde{F}_R = \sum_{i=1}^k \left(\frac{P_i'(X'X)^{-1} X' \varepsilon - Q_i'(X_1'X_1)^{-1} X_1' \varepsilon^* + P_i' \beta - Q_i' \beta_R}{S_{[k]} / \sqrt{n_i}} \right)^2$$

elde edilir. $R_i' = P_i' \beta - Q_i' \beta_R$ (yani $R_i' = (X_{i,q+1}, \dots, X_{ip})$) , $\beta^* = (\beta_{q+1}, \dots, \beta_p)'$,

$T_s = (T_1 / \sqrt{n_1}, \dots, T_k / \sqrt{n_k})'$ ve $T_s^* = (T_1^* / \sqrt{n_1}, \dots, T_k^* / \sqrt{n_k})'$ alınarak ,

$$\tilde{F}_R = \sum_{i=1}^k \left(\frac{P_i'(X'X)^{-1} X' T_s - Q_i'(X_1'X_1)^{-1} X_1' T_s^*}{1 / \sqrt{n_i}} + \frac{R_i' \beta^*}{S_{[k]} / \sqrt{n_i}} \right)^2$$

yazılabilir.

T_i ve T_i^* arasındaki ilişki, $T_i = \frac{\tilde{Y}_i - P_i' \beta}{S_{[k]} / \sqrt{n_i}}$ ve $T_i^* = \frac{\tilde{Y}_i - Q_i' \beta_R}{S_{[k]} / \sqrt{n_i}}$ ifadelerinden

yararlanılarak aşağıdaki gibi gösterilebilir.

$$T_i = \frac{\tilde{Y}_i - P_i' \beta}{S_{[k]} / \sqrt{n_i}}, \quad i = 1, \dots, k \text{ eşitliğinden}$$

$$T_i = T_i^* - \frac{(P_i' \beta - Q_i' \beta_R)}{S_{[k]} / \sqrt{n_i}} \text{ ve } R_i' = P_i' \beta - Q_i' \beta_R \text{ olduğundan}$$

$$T_i = T_i^* - \frac{R_i' \beta^*}{S_{[k]} / \sqrt{n_i}}$$

bulunur. $\beta^* = (\beta_{q+1}, \dots, \beta_p)'$ iken $H_0^*: \beta^* = 0$ hipotezinin doğruluğu altında,

yukarıda verilen T_i ve T_i^* ilişkisini gösteren eşitlikte, sağ tarafın ikinci terimi sıfır

olacağından, $T_i = T_i^*$ 'dir.

Test istatistiği $\tilde{F}_R$ de aşağıdaki gibidir.

$$\tilde{F}_R = \sum_{i=1}^k \left(\frac{\mathbf{P}'_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{T}_s - \mathbf{Q}'_i(\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1\mathbf{T}_s^*}{1/\sqrt{n_i}} \right)^2$$

Başka bir ifadeyle

$$\tilde{F}_R = \sum_{i=1}^k \left(\sqrt{n_i} (\mathbf{P}'_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{Q}'_i(\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1) \mathbf{T}_s \right)^2$$

şeklinde gösterilebilir. $\mathbf{X} = (\mathbf{P}'_1, \dots, \mathbf{P}'_k)$ ve $\mathbf{X}_1 = (\mathbf{Q}'_1, \dots, \mathbf{Q}'_k)$ olduğundan $\{b_{ij}\} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ ve $\{q_{ij}\} = \mathbf{X}_1(\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1$ 'dir. Bu durumda test istatistiği,

$$\tilde{F}_R = \sum_{i=1}^k \left(\sqrt{n_i} \sum_{j=1}^k (b_{ij} - q_{ij}) \frac{T_j}{\sqrt{n_j}} \right)^2 \quad (13)$$

'dir [1]. Test istatistiği $\tilde{F}_R$ 'nin p değeri, belirlenen bir α 'dan küçük ise $H_0^* : \beta^* = 0$ yokluk hipotezi reddedilir. $\tilde{F}_R$ 'nin dağılımı, sırasıyla $v_1, \dots, v_k$ serbestlik dereceli, $t_{v_1}, \dots, t_{v_k}$ bağımsız t dağılımlarının doğrusal kombinasyonlarının karesel bir fonksiyonudur.

Eş.10'da verilen $H_0^* : \beta_{q+1} = \dots = \beta_p = 0$ hipotezinde $q = 0$ olması durumunda, $H_0^* = H_0 : \beta_1 = \dots = \beta_p = 0$ hipotezi olur. $\tilde{Y}_{i.} = \beta_0 + \varepsilon_i^*$ ve $\tilde{Y}_{i.}^* = \sum_{i=1}^k \tilde{Y}_{i.} / k = \tilde{Y}_{..}$ 'dir. Tüm ij 'ler ($(i = 1, \dots, k) \vee (j = 1, \dots, n_i)$) için $q_{ij} = 1/k$ olur. $\tilde{F}_R$ ise, parametrelerin tamamının test edilmesi için önerilen Eş.9'da verilen $\tilde{F}$ istatistiği olur. Kısmi test, $(p+1)$ tane regresyon parametresinin herhangi bir altkümesine uygulanabilir.

3.3. İki Açıklayıcı Değişkenli Çoklu Regresyon

İki açıklayıcı değişkenli çoklu regresyon modeli, Eş.1’de verilen genel doğrusal regresyon modelinin özel bir durumudur [1].

$$Y_{ij} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + e_{ij} , \quad i = 1, \dots, k, \quad j = 1, \dots, n_i \quad (14)$$

‘dır. Burada, β_0, β_1 ve β_2 bilinmeyen parametrelerdir. X_i ’ler açıklayıcı değişkenler ve $e_{ij} \sim N(0, \sigma_i^2)$ ’dir. İki açıklayıcı değişkenli çoklu regresyon modeli, tek örneğe dayalı yöntemle dayalı olarak Eş.5.2’deki gibi ifade edilirse;

$$\begin{bmatrix} \tilde{Y}_1 \\ \tilde{Y}_2 \\ \vdots \\ \tilde{Y}_k \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} \\ 1 & X_{21} & X_{22} \\ \vdots & \vdots & \vdots \\ 1 & X_{k1} & X_{k2} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + S_{[k]} \begin{bmatrix} T_1 / \sqrt{n_1} \\ T_2 / \sqrt{n_2} \\ \vdots \\ T_k / \sqrt{n_k} \end{bmatrix}$$

’dir. Modelin matris formu Eş.5.3’e benzerdir. β_0, β_1 ve β_2 ’nin EKK tahmin edicileri;

$$\hat{\beta}_1 = \frac{(\sum \tilde{y}_i x_{i1})(\sum x_{i2}^2) - (\sum \tilde{y}_i x_{i2})(\sum x_{i1} x_{i2})}{(\sum x_{i1}^2)(\sum x_{i2}^2) - (\sum x_{i1} x_{i2})},$$

$$\hat{\beta}_2 = \frac{(\sum \tilde{y}_i x_{i2})(\sum x_{i1}^2) - (\sum \tilde{y}_i x_{i1})(\sum x_{i1} x_{i2})}{(\sum x_{i1}^2)(\sum x_{i2}^2) - (\sum x_{i1} x_{i2})}$$

ve

$$\hat{\beta}_0 = \tilde{Y}_{..} - \hat{\beta}_1 \bar{X}_1 - \hat{\beta}_2 \bar{X}_2$$

‘dır. Burada, $x_{i1} = (X_{i1} - \bar{X}_1)$, $x_{i2} = (X_{i2} - \bar{X}_2)$, $\tilde{y}_i = (\tilde{Y}_i - \tilde{Y}_{..})$, $\bar{X}_1 = \sum X_{i1} / k$, $\bar{X}_2 = \sum X_{i2} / k$ ve $\tilde{Y}_{..} = \sum \tilde{Y}_i / k$ alınmıştır [1].

İlk olarak $H_0 : \beta_1 = \beta_2 = 0$ ile ilgilenilirse, Eş.8’de verilen

$$\tilde{F} = \sum_{i=1}^k \left(\frac{\hat{Y}_{i.} - \tilde{Y}_{..}}{S_{[k]} / \sqrt{n_i}} \right)^2$$

$H_0 : \beta_1 = \beta_2 = 0$ ’ın doğruluğu altında Eş.9’daki

$$\tilde{F} = \sum_{i=1}^k \left(\sqrt{n_i} \sum_{j=1}^k (b_{ij} - \frac{1}{k}) \frac{T_j}{\sqrt{n_j}} \right)^2$$

test istatistiği olur.

Burada, $\{b_{ij}\} = X(X'X)^{-1}X'$ ve

$$X = \begin{bmatrix} 1 & X_{11} & X_{12} \\ 1 & X_{21} & X_{22} \\ . & . & . \\ . & . & . \\ 1 & X_{k1} & X_{k2} \end{bmatrix}$$

‘dir. $H_0 : \beta_1 = \beta_2 = 0$, Eş.9’daki test istatistiği $\tilde{F}$ ’nın p değeri, α anlamlılık düzeyinden küçük ise reddedilir.

İkinci olarak, kısmi hipotez $H_0^* : \beta_2 = 0$ ya da $H_0^* : \beta_1 = 0$ ele alınabilir. $H_0^* : \beta_2 = 0$ hipotezinin doğruluğu altında indirgenmiş model,

$$Y_{ij} = \beta_0 + \beta_1 X_{i1} + e_{ij} \quad (15)$$

veya başka bir ifadeyle

$$Y_{ij} = Q_i' \beta_R + e_{ij}, \quad i = 1, \dots, k, \quad j = 1, \dots, n_i \quad (15.1)$$

şeklinde gösterilir. Burada $Q'_i = (1, X_{i1})$, $\beta'_R = (\beta_0, \beta_1)$ 'dir. Tek örneğe dayalı yöntemi temel alarak, Eş.15.1'de verilen indirgenmiş model,

$$\begin{bmatrix} \tilde{Y}_1 \\ \tilde{Y}_2 \\ \cdot \\ \cdot \\ \tilde{Y}_k \end{bmatrix} = \begin{bmatrix} 1 & X_{11} \\ 1 & X_{21} \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & X_{k1} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + S_{[k]} \begin{bmatrix} T_1^* / \sqrt{n_1} \\ T_2^* / \sqrt{n_2} \\ \cdot \\ \cdot \\ T_k^* / \sqrt{n_k} \end{bmatrix}$$

şeklinde ya da basit doğrusal regresyon

$$\tilde{Y}_i = X_i \beta_R + \varepsilon_i^*, \quad i = 1, \dots, k$$

olarak ifade edilebilir. Burada $\tilde{Y}_i$ ve ε_i^* , Eş.11.2 ve Eş.11.3'te tanımlandığı gibidir.

$$X_1 = \begin{bmatrix} 1 & X_{11} \\ 1 & X_{21} \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & X_{k1} \end{bmatrix} \text{ 'dir.}$$

$H_0^* : \beta_2 = 0$ doğruluğu altında test istatistiği Eş.13'te verilen

$$\tilde{F}_R = \sum_{i=1}^k \left(\sqrt{n_i} \sum_{j=1}^k (b_{ij} - q_{ij}) \frac{T_j}{\sqrt{n_j}} \right)^2$$

'dir. Eş.14'te tam model için, $\{b_{ij}\} = X(X'X)^{-1}X'$ ve Eş.15.1'de indirgenmiş model için, $\{q_{ij}\} = X_1(X_1'X_1)^{-1}X_1'$ 'dir. $H_0^* : \beta_2 = 0$ ancak Eş.13'te verilen test istatistiği

$\tilde{F}_R$ 'nın p değeri, α anlamlılık düzeyinden küçük ise reddedilir.

3.4. Tek Yönlü ANOVA' ya İlişkin Model

Tek yönlü varyans analizinin, tamamen rasgele tasarım altında Eş.1'de verilen genel doğrusal regresyon modelinin özel bir durumu olduğu gösterilebilir [1].

Rasgeleleştirme ilkesi, gözlem birimine uygulanacak işlemlerin ve sırasının rasgele yapılmasını ya da işlemlerin uygulanması için seçilecek birimlerin ve uygulanacak işlemlerin sırasının rasgele olmasını vurgulamaktadır. Bazı tasarımlarda, rasgeleleştirme ilkesine tam olarak uymak mümkün olmamaktadır, yani rasgeleleştirme üzerinde bazı kısıtlamalar söz konusudur. Kısıtlama olmadığında tasarıma tamamen rasgele tasarım (completely randomized design) denir [14].

Tek yönlü model,

$$Y_{ij} = \mu_i + e_{ij}, \quad i = 1, \dots, k, \quad j = 1, \dots, n_i \quad (16)$$

veya

$$Y_{ij} = \bar{\mu} + \alpha_i + e_{ij}, \quad i = 1, \dots, k, \quad j = 1, \dots, n_i \quad (16.1)$$

olarak düşünülebilir. Burada $\bar{\mu} = \sum_{i=1}^k \mu_i / k$, $\alpha_i = \mu_i - \bar{\mu}$ 'dir. $e_{ij} \sim N(0, \sigma_i^2)$ olan rasgele hatalardır [1].

Eş.16.1'de verilen tek yönlü varyans analizi modeli, Eş.1'de verilen genel doğrusal regresyon modeli gibi

$$Y_{ij} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_{k-1} X_{i,k-1} + e_{ij}$$

olarak yeniden yazılabilir. ANOVA çalışmalarında, işlemlere ait ortalamalar arasında yapılmak istenen herhangi bir kıyaslama β 'lar kullanılarak ifade edilebilir. Bağımlı değişken Y_{ij} 'nin, sayısal olmayan bağımsız değişkenler ile ilişkili olduğu deneysel bir durum için, model yazılması durumunda kukla değişkenler kullanılır. Kukla değişkenler 0 ya da 1 değerini alır.

Modeldeki kukla değişkenler,

Y , 2'inci işlemde ise $X_1 = 1$, değil ise $X_1 = 0$,

Y , 3'üncü işlemde ise $X_2 = 1$, değil ise $X_2 = 0$,

$\vdots$

Y , k 'ıncı işlemde ise $X_{k-1} = 1$, değilse ise $X_{k-1} = 0$

olarak tanımlanabilir [15, EK-3].

$Y_{ij} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_{k-1} X_{i,k-1} + e_{ij}$ modeli için, $\beta_0 = \mu_1$, $\beta_1 = \mu_2 - \mu_1$, $\beta_2 = \mu_3 - \mu_1, \dots, \beta_{k-1} = \mu_k - \mu_1$ 'dir. Tasarım matrisi X , $k \times k$ boyutlu bir karesel matris olarak aşağıdaki gibi verilebilir.

$$X = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ 1 & 1 & 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & 1 & 0 & \dots & 0 & 0 \\ \vdots & & & & & & \vdots \\ 1 & 0 & 0 & 0 & \dots & 0 & 1 \end{bmatrix}$$

Ayrıca, $X(X'X)^{-1}X' = I$, katsayıları, $i = j$ için $b_{ij} = 1$ ve $i \neq j$ için ise $b_{ij} = 0$ olan $k \times k$ boyutlu bir özdeşlik matrisidir.

Eş.8'de verilen $\tilde{F} = \sum_{i=1}^k \left(\frac{\hat{Y}_{i.} - \tilde{Y}_{..}}{S_{[k]} / \sqrt{n_i}} \right)^2$, den

$$\frac{\hat{Y}_{i.} - \tilde{Y}_{..}}{S_{[k]} / \sqrt{n_i}} = \frac{\tilde{Y}_{i.} - \mu_i}{S_{[k]} / \sqrt{n_i}} - \frac{\tilde{Y}_{..} - \bar{\mu}}{S_{[k]} / \sqrt{n_i}} + \frac{\mu_i - \bar{\mu}}{S_{[k]} / \sqrt{n_i}}$$

olarak yeniden yazılabilir. Burada,

$$\frac{\tilde{Y}_{..} - \bar{\mu}}{S_{[k]} / \sqrt{n_i}} = \frac{(1/k) \sum_{j=1}^k (\tilde{Y}_{j.} - \mu_j)}{S_{[k]} / \sqrt{n_i}} = \frac{\sqrt{n_i}}{k} \sum_{j=1}^k \frac{T_j}{\sqrt{n_j}}$$

şeklinde de ifade edilebilir. Böylece

$$\tilde{F} = \sum_{i=1}^k \left(T_i - \frac{\sqrt{n_i}}{k} \sum \frac{T_j}{\sqrt{n_j}} + \frac{\mu_i - \bar{\mu}}{S_{[k]} / \sqrt{n_i}} \right)^2 \quad (17)$$

olur [1]. $H_0 : \beta_1 = \dots = \beta_k = 0$ 'ın doğruluğu altında ($H_0 : \beta_1 = \dots = \beta_k = 0$ ile $H_0 : \mu_1 = \dots = \mu_k$ aynı anlamdadır) test istatistiği,

$$\tilde{F} = \sum_{i=1}^k \left(T_i - \frac{\sqrt{n_i}}{k} \sum_{j=1}^j \frac{T_j}{\sqrt{n_j}} \right)^2 \quad (17.1)$$

'dir [1].

Örnek hacimleri birbirine eşit olduğu durum için tek yönlü ANOVA' ya ilişkin, $\tilde{F}$ istatistiğinin $\tilde{F}_\alpha$ değerleri, Chen ve Chen tarafından hesaplanmıştır [8]. Örnek hacimlerinin eşit olmadığı durumda ise $\tilde{F}$ istatistiğinin p değeri bir bilgisayar programı ile elde edilebilir.

3.5. Kısmi Test ve İki Yönlü ANOVA Arasındaki İlişki

Eş.1'de verilen genel doğrusal regresyon modeli için kısmi test tekniği, tamamen rasgele tasarım ya da rasgele blok tasarımı altında iki yönlü ANOVA modeline uygulanabilir. Etkileşimsiz model,

$$Y_{ijl} = \mu + \alpha_i + \tau_j + e_{ijl}, \quad i = 1, \dots, I, \quad j = 1, \dots, J, \quad l = 1, \dots, n_{ij}, \quad (18)$$

veya

$$Y_{ijl} = \mu_{ij} + e_{ijl}, \quad i = 1, \dots, I, \quad j = 1, \dots, J, \quad l = 1, \dots, n_{ij}, \quad (18.1)$$

şeklinde ifade edilir [1]. Burada, $e_{ijl} \sim N(0, \sigma_{ij}^2)$ olduğu varsayılmaktadır. μ genel ortalama, $\mu_{ij} = \mu + \alpha_i + \tau_j$ 'dir. α_i , A faktörünün i 'nci düzeyinin etkisini ve τ_j de B faktörünün j 'nci düzeyinin etkisini göstermektedir. Ayrıca,

$$\sum_i \alpha_i = 0 \quad \text{ve} \quad \sum_j \tau_j = 0$$

olduğu varsayılmaktadır.

Üç düzeyi olan ($I = 3$) bir A faktörü ile iki düzeyi olan ($J = 2$) bir B faktörünün bulunduğu bir durum düşünülebilir. Kukla değişkenler ,

Y , A faktörünün 2'inci düzeyinde ise $X_1 = 1$, değilse $X_1 = 0$,

Y , A faktörünün 3'üncü düzeyinde ise $X_2 = 1$, değilse $X_2 = 0$,

Y , B faktörünün 2'inci düzeyinde ise $X_3 = 1$, değilse $X_3 = 0$

olarak tanımlanabilir. Eş.18.1'de verilen iki yönlü ANOVA modeli, Eş.1'deki genel doğrusal regresyon modelinin özel bir durumu olarak X_1, X_2 ve X_3 kukla değişkenleri ile

$$Y_{ij} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + e_{ij}, \quad i = 1, \dots, k, \quad j = 1, \dots, n_i \quad (18.2)$$

şeklinde yeniden yazılabilir. $e_{ij} \sim N(0, \sigma_i^2)$ 'dir. Bu durumda, $k = 2 \times 3 = 6$ ve kestirici sayısı ise $p = 3 + 2 - 2 = 3$ 'dir.

Eş.18.2'de verilen regresyon ve Eş.18.1'de verilen ANOVA parametreleri arasındaki ilişki bir çizelge üzerinde aşağıdaki gibi gösterilebilir.

Çizelge 3.1. Regresyon ve ANOVA parametreleri arasındaki ilişki I

		B Faktörü	
		Düzy	
A Faktörü		1	2
	1	$\beta_0(\mu_{11})$	$\beta_0 + \beta_3(\mu_{12})$
	2	$\beta_0 + \beta_1(\mu_{21})$	$\beta_0 + \beta_1 + \beta_3(\mu_{22})$
	3	$\beta_0 + \beta_2(\mu_{31})$	$\beta_0 + \beta_2 + \beta_3(\mu_{32})$

Bölüm 2’de tanımlanan tek örneğe dayalı yöntemi uygulayarak Eş.5.2’ye benzer bir regresyon modeli,

$$\tilde{Y}_{i.} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i, \quad i = 1, \dots, 6$$

‘dır[1]. $S_1^2, \dots, S_k^2$ biliniyorken $\varepsilon_i \sim (0, \frac{S_{[k]}^2}{n_i} \cdot \frac{\sigma_i^2}{S_i^2})$; $i = 1, \dots, k$ ‘dir. Genel doğrusal

regresyon modelinin tasarım matrisi X ve $X'X$ ‘in tersi sırasıyla

$$X = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix} \quad \text{ve} \quad (X'X)^{-1} = \begin{bmatrix} -\frac{2}{3} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{3} \\ -\frac{1}{2} & -1 & -\frac{1}{2} & 0 \\ -\frac{1}{2} & -\frac{1}{2} & -1 & 0 \\ -\frac{1}{3} & -0 & -0 & \frac{2}{3} \end{bmatrix}, \quad \text{dır.}$$

B faktörünün etkisi için, kısmi regresyon hipotezi $H_0^2 : \beta_3 = 0$ kurulur. Bu hipotez B faktörü için iki yönlü ANOVA modeli Eş.18 için kurulabilecek $H_0^2 : \tau_1 = \tau_2 = 0$ ile aynı anlama gelmektedir. H_0^2 doğruluğu altında indirgenmiş regresyon modeli

$$\tilde{Y}_{i.} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i, \quad i = 1, \dots, 6$$

‘dir. İndirgenmiş tasarım matrisi X_1 ve $X_1'X_1$ ‘in tersi de sırasıyla

$$X_1 = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \quad \text{ve} \quad (X_1'X_1)^{-1} = \begin{bmatrix} -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & -1 & -\frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & -1 \end{bmatrix}, \text{dır.}$$

Eş.13'teki istatistiğin yokluk hipotezi altında dağılımının bulunması için, tam model $\tilde{Y}_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$ için $\{b_{ij}\} = X(X'X)^{-1}X'$ matrisi ile indirgenmiş model $\tilde{Y}_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$ için $\{q_{ij}\} = X_1(X_1'X_1)^{-1}X_1'$ matrisi aşağıdaki gibi elde edilmiştir.

$$\{b_{ij}\} = \begin{bmatrix} -\frac{2}{3} & -\frac{1}{6} & -\frac{1}{6} & -\frac{1}{3} & -\frac{1}{6} & -\frac{1}{6} \\ -\frac{1}{6} & \frac{2}{3} & -\frac{1}{6} & -\frac{1}{6} & -\frac{1}{3} & -\frac{1}{6} \\ -\frac{1}{6} & -\frac{1}{6} & \frac{2}{3} & -\frac{1}{6} & -\frac{1}{6} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{6} & -\frac{1}{6} & -\frac{2}{3} & -\frac{1}{6} & -\frac{1}{6} \\ -\frac{1}{6} & \frac{1}{3} & -\frac{1}{6} & -\frac{1}{6} & -\frac{2}{3} & -\frac{1}{6} \\ -\frac{1}{6} & -\frac{1}{6} & -\frac{1}{6} & -\frac{1}{3} & -\frac{1}{6} & -\frac{2}{3} \end{bmatrix}, \quad \{q_{ij}\} = \begin{bmatrix} \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} \end{bmatrix}.$$

Eş.13'te, tam model için $\{b_{ij}\}$ matrisi ve indirgenmiş model için $\{q_{ij}\}$ matrisi yerine konursa, H_0^2 doğruluğu altında test istatistiği,

$$\tilde{F}_R^{(2)} = \sum_{i=1}^6 \left(\sqrt{n_i} \sum_{j=1}^6 (b_{ij} - q_{ij}) \frac{T_j}{\sqrt{n_j}} \right)^2$$

veya örneğe göre basitleştirilerek

$$\tilde{F}_R^{(2)} = \sum_{i=1}^3 \sum_{j=1}^2 \left(\frac{\sqrt{n_{ij}}}{3} \sum_{m=1}^3 \frac{T_{mj}}{\sqrt{n_{mj}}} - \frac{\sqrt{n_{ij}}}{6} \sum_{m=1}^3 \sum_{l=1}^2 \frac{T_{ml}}{\sqrt{n_{ml}}} \right)^2$$

şeklinde olur [1]. T_{mj} 'ler ($m=1,2,3, j=1,2$), serbestlik derecesi $n_{ml} - 2$ olan student t dağılımına sahiptir. Burada, $T_{11}=T_1, T_{21}=T_2, T_{31}=T_3, T_{12}=T_4, T_{22}=T_5, T_{32}=T_6$ ve $n_{11}=n_1, n_{21}=n_2, n_{31}=n_3, n_{12}=n_4, n_{22}=n_5, n_{32}=n_6$ 'dır.

A faktörünün etkisinin test'i için benzer olarak, regresyon modeli için kısmi yokluk hipotezi $H_0^1: \beta_1 = \beta_2 = 0$ 'dır. Bu hipotez, iki yönlü ANOVA modeli Eş.18'de kurulabilecek $H_0^1: \alpha_1 = \alpha_2 = \alpha_3 = 0$ hipotezinin test edilmesi ile aynı anlamdadır.

H_0^1 'ın doğruluğu altında indirgenmiş regresyon modeli,

$$\tilde{Y}_{i.} = \beta_0 + \beta_3 X_{i3} + \varepsilon_i, \quad i=1,...,6$$

olur. Tasarım matrisi X_1^* ,

$$X_1^* = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \quad \text{ve} \quad (X_1^{*'} X_1^*)^{-1} = \begin{bmatrix} -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{2}{3} \end{bmatrix} \quad \text{olmak üzere}$$

$$\{q_{ij}^*\} = X_1^* (X_1^{*'} X_1^*)^{-1} X_1^{*'} \text{ matrisi,}$$

$$q_{ij}^* = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$$

‘dir. H_0^1 ’ın doğruluğu altında, tam model $\tilde{Y}_{i.} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$ için $\{b_{ij}\}$ matrisi ve indirgenmiş model $\tilde{Y}_{i.} = \beta_0 + \beta_3 X_{i3} + \varepsilon_i$ için $\{q_{ij}^*\}$ matrisi kullanılarak Eş.13’e yerleştirilirse test istatistiği,

$$\tilde{F}_R^{(1)} = \sum_{i=1}^6 \left(\sqrt{n_i} \sum_{j=1}^6 (b_{ij} - q_{ij}^*) \frac{T_j}{\sqrt{n_j}} \right)^2$$

veya örneğe göre basitleştirilerek

$$F_R^{(1)} = \sum_{i=1}^3 \sum_{j=1}^2 \left(\frac{\sqrt{n_{ij}}}{2} \sum_{l=1}^2 \frac{T_{il}}{\sqrt{n_{il}}} - \frac{\sqrt{n_{ij}}}{6} \sum_{m=1}^3 \sum_{l=1}^2 \frac{T_{ml}}{\sqrt{n_{ml}}} \right)^2$$

şeklinde yazılabilir [1].

Düzey sayısı 3 olarak alınan A faktörü ile düzey sayısı 2 olarak alınan B faktörü, düzey sayıları sırasıyla I ve J olmak üzere genelleştirildiğinde, A faktörü için $(I-1)$ ve B faktörü için de $(J-1)$ sayıda kukla değişkene ihtiyaç duyulacağı söylenebilir. Bu durumda kestirici sayısı, $p = I + J - 2$ ’dir. Tasarım matrisi X , $k = I \times J$ satır ve $(I + J - 2) + 1$ sütunlu, elemanları 0 ve 1’ler olan bir matristir.

Örnek olarak verilen $\tilde{Y}_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$ regresyon modeline ait, $I = 3$ ve $J = 2$, tasarım matrisinin boyutu 6×4 'tür.

Varyans analizi, genel bir doğrusal regresyon modeli kullanılarak kısmi test yoluyla gerçekleştirilebilir. $\tilde{F}_R^2$ test'i, B faktörü, rasgele blok tasarımında blok olduğu durumda, blok etkisini test eden $H_0^2 : \tau_1 = \tau_2 = 0$ 'ın testinde, aynı durumu regresyon formunda ifade eden $H_0^2 : \beta_3 = 0$ 'ın test edilmesinde ve iki faktörlü faktöriyel tasarımında da aynıdır.

Tamamen rasgele tasarım altında, etkileşimlerin de bulunduğu iki yönlü ANOVA modeli,

$$y_{ijl} = \mu + \alpha_i + \tau_j + (\alpha\tau)_{ij} + e_{ijl}, \quad i = 1,2,3, \quad j = 1,2, \quad l = 1, \dots, n_{ij} \quad (19)$$

'dir. $e_{ijl} \sim N(0, \sigma_{ij}^2)$ 'dir. Aynı durumu ifade eden, tek örneğe dayalı yöntem ile verilen regresyon modeli ise

$$\tilde{Y}_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i1} X_{i3} + \beta_5 X_{i2} X_{i3} + \varepsilon_i, \quad i = 1, \dots, 6 \quad (19.1)$$

'dir [1]. Burada, kukla değişkenler X_1, X_2 ve X_3 aynı şekilde tanımlanır. $S_1^2, \dots, S_k^2$ biliniyorken $\varepsilon_i \sim (0, \frac{S_{[k]}^2}{n_i} \cdot \frac{\sigma_i^2}{S_i^2})$; $i = 1, \dots, k$ şeklindedir. Etkileşim terimleri, $\beta_4 X_{i1} X_{i3}$ ve $\beta_5 X_{i2} X_{i3}$ 'dir [15].

Eş.19'da verilen ANOVA modeli için amaç, tüm ij 'ler için iki faktör arasındaki etkileşimleri yani $H_0^3 : (\alpha\tau)_{ij} = 0$ 'ı test etmektir. Amaç, Eş.19.1'de verilen regresyon modelinde, $H_0^3 : \beta_4 = \beta_5 = 0$ 'ın test edilmesidir. H_0^3 'ın doğruluğu altında indirgenmiş model,

$$\tilde{Y}_{i.} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i \quad (19.2)$$

‘dır.

Çizelge 3.2. Regresyon ve ANOVA parametreleri arasındaki ilişki II

		B Faktörü	
		Düze	
		1	2
A Faktörü	1	$\beta_0(\mu_{11})$	$\beta_0 + \beta_3(\mu_{12})$
	2	$\beta_0 + \beta_1(\mu_{21})$	$\beta_0 + \beta_1 + \beta_3 + \beta_4(\mu_{22})$
	3	$\beta_0 + \beta_2(\mu_{31})$	$\beta_0 + \beta_2 + \beta_3 + \beta_5(\mu_{32})$

H_0^3 ’ın doğruluğu altında, Eş.19.1’de verilen tam model için ve Eş.19.2’de verilen indirgenmiş model için tasarım matrisleri sırasıyla,

$$X = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \end{bmatrix} \quad \text{ve} \quad X_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}, \text{dir.}$$

Tam model için $\{b_{ij}\} = X(X'X)^{-1}X' = I$ olan bir özdeşlik matrisidir.

H_0^3 ’ın doğruluğu altında, indirgenmiş model için $\{q_{ij}\} = X_1(X_1'X_1)^{-1}X_1'$,

$$q_{ij} = \begin{bmatrix} \frac{2}{3} & -\frac{1}{6} & -\frac{1}{6} & -\frac{1}{3} & -\frac{1}{6} & -\frac{1}{6} \\ \frac{1}{6} & \frac{2}{3} & -\frac{1}{6} & -\frac{1}{6} & -\frac{1}{6} & -\frac{1}{6} \\ -\frac{1}{6} & -\frac{1}{6} & \frac{2}{3} & -\frac{1}{6} & -\frac{1}{6} & -\frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & -\frac{1}{6} & \frac{2}{3} & -\frac{1}{6} & -\frac{1}{6} \\ -\frac{1}{6} & -\frac{1}{6} & -\frac{1}{6} & -\frac{1}{6} & \frac{2}{3} & -\frac{1}{6} \\ -\frac{1}{6} & -\frac{1}{6} & -\frac{1}{6} & -\frac{1}{6} & -\frac{1}{6} & \frac{2}{3} \end{bmatrix}, \text{dir.}$$

Eş.19.1 için $\{b_{ij}\}$ matrisi ve Eş.19.2 için $\{q_{ij}\}$ matrisi Eş.13'te yerine konulursa, H_0^3 'ün doğruluğu altında test istatistiği,

$$\tilde{F}_R^{(3)} = \sum_{i=1}^6 \left(\sqrt{n_i} \sum_{j=1}^6 (b_{ij} - q_{ij}) \frac{T_j}{\sqrt{n_j}} \right)^2$$

veya daha basit bir ifadeye sadeleştirilir ise,

$$\tilde{F}_R^{(3)} = \sum_{i=1}^3 \sum_{j=1}^2 \left(T_{ij} - \frac{\sqrt{n_{ij}}}{2} \sum_{l=1}^2 \frac{T_{il}}{\sqrt{n_{il}}} - \frac{\sqrt{n_{ij}}}{3} \sum \frac{T_{mj}}{\sqrt{n_{mj}}} + \frac{\sqrt{n_{ij}}}{6} \sum_{m=1}^3 \sum_{l=1}^2 \frac{T_{ml}}{\sqrt{n_{ml}}} \right)^2$$

olur [1]. T_{ml} , T_j ve n_{ml} , n_j ilişkisi daha önce verildiği gibidir. Sonuç, etkileşim terimlerinin varlığında, A faktörünün I ve B faktörünün J düzeyli olduğu model için geçerlidir. Örnek hacimleri eşit olduğunda, yani $n_{ij} = n$ iken, $H_0^3 : (\alpha\tau)_{ij} = 0$ 'ın testi için kritik değer tabloları Chen ve Chen tarafından verilmiştir [8]. Örnek hacimleri, n_{ij} 'ler birbirine eşit olmadığında, H_0^3 'ün testi için p değeri bir bilgisayar programı ile elde edilebilir.

Kısmi test tekniği Latin kare düzenine, tekrarlanmış ölçümler tasarımına ve diğer tasarımlara da uygulanabilir.

3.6. Genel Doğrusal Hipotezler

Genel doğrusal model

$$i. \quad E(y) = X\beta$$

olarak düşünülün. Burada X , verilen $n \times p$ boyutlu bir matris, β , $p \times 1$ boyutlu kısıtsız bilinmeyen parametre vektörü ve y , $n \times 1$ boyutlu gözlemlenebilir rasgele değişkenler vektörüdür. Tahmin edilmek istenen $s \times 1$ boyutlu

$$ii. \quad A\beta$$

vektörüdür. Burada A , verilen $s \times p$ boyutlu bir matristir.

$s \times n$ boyutlu ve $E(By) = A\beta$ eşitliğini de sağlayan bir B matrisi, $A = BX$ eşitliğini de sağlıyorsa veya başka bir ifadeyle

$$iii. \quad r\begin{pmatrix} X \\ A \end{pmatrix} = r(X)$$

olduğunda $A\beta$ tahmin edilebilir parametrik fonksiyondur. Burada, $r(.)$ rank anlamında kullanılmıştır.

Ayrıca, yokluk hipotezi $A\beta = 0$,

$$iv. \quad r(XT) = r(X) - r(A)$$

olduğu durumda tamamiyle test edilebilirdir. T , A 'nın yokluk uzayını kapsamaktadır [16].

m tane doğrusal, bağımsız tahmin edilebilir fonksiyon

$\psi_1, \dots, \psi_m$ ve

$$\psi_i = \sum_{j=0}^p \alpha_{ij} \beta_j, \quad i = 1, \dots, m$$

olsun [1]. Burada, $(\alpha_{i0}, \alpha_{i1}, \dots, \alpha_{ip})$, $m \times (p+1)$ boyutlu $K = \{\alpha_{ij}\}$ matrisinin i 'inci satır vektörüdür. K , tam satır ranklıdır. Sütun vektörü,

$$\psi = K\beta$$

'dır. $\psi = (\psi_1, \dots, \psi_m)'$ 'dır. Eş.7'de verilen, β 'nın tahmini $\hat{\beta}$ kullanılarak, ψ 'nın EKK tahmini,

$$\hat{\psi} = K\hat{\beta}$$

'dir. $\hat{\beta}$ yerine eşiti olan $(X'X)^{-1}X'\tilde{Y}$ konulursa

$$\hat{\psi} = K(X'X)^{-1}X'\tilde{Y}$$

yazılabilir ve $C = K(X'X)^{-1}X'$ alınırsa

$$\hat{\psi} = C\tilde{Y}$$

olur. $\hat{\psi} = (\hat{\psi}_1, \dots, \hat{\psi}_m)$ ve

$$\hat{\psi}_i = \sum_{j=1}^k c_{ij} \tilde{Y}_j, \quad i = 1, \dots, m$$

'dir [1]. $(c_{i1}, \dots, c_{ik})$, $m \times k$ boyutlu $C = \{c_{ij}\}$ matrisinin i 'inci satır vektörüdür.

Genel doğrusal hipotez,

$$H_0 : K\beta = h \quad (20)$$

olarak düşünülebilir. Burada, h , m boyutlu bir vektördür. $\hat{\beta}^*$, kısıtlı EKK tahmini olduğunda, $(\tilde{Y} - X\beta)'(\tilde{Y} - X\beta)$ kare toplamı, $K\beta = h$ kısıtı altında en küçük olur.

Lagrange çarpım metodu kullanarak, Eş.20'de yer alan $H_0 : K\beta = h$ 'ın doğruluğu altında, $\tilde{Y} = X\beta + \varepsilon$ modeli için, β 'nın kısıtlı EKK tahmini olan $\hat{\beta}^*$ elde edilebilir.

Gerçekten de, Lagrange çarpanlarının bir vektörü olan $2\theta'$ ile, elemanları $\hat{\beta}^*$ ve θ' ya bağlı olan aşağıdaki ifade en küçük yapılabilir [17].

$$(\tilde{Y} - X\hat{\beta}^*)'(\tilde{Y} - X\hat{\beta}^*) + 2\theta'(K\hat{\beta}^* - h)$$

sırasıyla $\hat{\beta}^*$ ve θ' ya göre türevler alındığında,

$$I. \quad X'X\hat{\beta}^* + K'\theta = X'\tilde{Y}$$

$$II. \quad K\hat{\beta}^* = h$$

eşitlikleri elde edilir. Gerekli düzenlemelerden sonra,

$$\hat{\beta}^* = \hat{\beta} - (X'X)^{-1}K'[K(X'X)^{-1}K']^{-1}(K\hat{\beta} - h)$$

bulunur. Eş.1'de verilen tam modelin kestirim denklemi $\hat{Y} = X\hat{\beta}$ ve Eş.20'deki

$H_0 : K\beta = h$ kısıtı altında indirgenmiş model için kestirim denklemi $\hat{Y}^* = X\hat{\beta}^*$ ile

$$\hat{Y} - \hat{Y}^* = X\hat{\beta} - X\hat{\beta}^*$$

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

yazılabilir. $\hat{\beta}^* = \hat{\beta} - (X'X)^{-1}K'[K(X'X)^{-1}K']^{-1}(K\hat{\beta} - h)$ eşitliğinden yararlanarak,

$$\hat{Y} - \hat{Y}^* = X (X'X)^{-1} K' [K(X'X)^{-1} K']^{-1} (K\hat{\beta} - h)$$

elde edilir. Böylece Eş.20'de yer alan hipotezden dolayı, kareler toplamı,

$$Q = (\hat{Y} - \hat{Y}^*)' (\hat{Y} - \hat{Y}^*)$$

gibi karesel bir form ile verilebilir. $\hat{Y}$ yerine $X\hat{\beta}$ ve $\hat{Y}^*$ yerine de $X\hat{\beta}^*$ konularak yeniden yazıldığında

$$Q = (X\hat{\beta} - X\hat{\beta}^*)' (X\hat{\beta} - X\hat{\beta}^*)$$

olur. Burada $\hat{\beta} - \hat{\beta}^* = (X'X)^{-1}K'[K(X'X)^{-1}K']^{-1}(K\hat{\beta} - h)$ yerine yazılırsa

$$Q = (K\hat{\beta} - h)' [K(X'X)^{-1}K']^{-1} (K\hat{\beta} - h)$$

elde edilir. $H_0 : K\beta = h$ için genel bir test istatistiği,

$$\tilde{F}_h = \frac{\tilde{n}Q}{S_{[k]}^2} \quad (21)$$

olarak önerilmektedir [1]. $H_0 : K\beta = h$ 'ın doğruluğu altında, Eş.21'deki test istatistiği $\tilde{F}_h = \tilde{n}Q / S_{[k]}^2$,

$$Q = (K\hat{\beta} - h)' [K(X'X)^{-1}K']^{-1} (K\hat{\beta} - h)$$

eşitliğinde $h = K\beta$ yerine yazıldığında

$$\tilde{F}_h = \tilde{n} (K\hat{\beta} - K\beta)' [K(X'X)^{-1}K']^{-1} (K\hat{\beta} - K\beta) / S_{[k]}^2$$

olur. Burada Eş.7'deki $\hat{\beta} = (X'X)^{-1}X'\tilde{Y}$ ve Eş.7.2'deki $\hat{\beta} = \beta + (X'X)^{-1}X'\varepsilon$ kullanılarak,

$$\tilde{F}_h = \tilde{n} \varepsilon' X(X'X)^{-1} K' [K(X'X)^{-1} K']^{-1} K(X'X)^{-1} X' \varepsilon / S_{[k]}^2$$

yazılabilir. $\varepsilon = S_{[k]} T_s$ ve $T_s = (T_1 / \sqrt{n_1}, \dots, T_k / \sqrt{n_k})'$ olmak üzere

$$\tilde{F}_h = \tilde{n} T_s' X(X'X)^{-1} K' [K(X'X)^{-1} K']^{-1} K(X'X)^{-1} X' T_s$$

'dir. Burada $C = K(X'X)^{-1} X'$ yerine yazılırsa,

$$\tilde{F}_h = \tilde{n} T_s' C' [K(X'X)^{-1} K']^{-1} C T_s \quad (21.1)$$

'dır [1].

Verilen bir X , hipotez matrisi K ve vektör h için, $\tilde{F}_h$ 'ın $H_0 : K\beta = h$ hipotezinin doğruluğu altında bulunan dağılımı kullanılarak bulunacak p değeri, belirlenen bir α anlamlılık düzeyinden küçük ise $H_0 : K\beta = h$ reddedilir.

4. UYGULAMA

4.1. Tek Yönlü ANOVA için Uygulama

Wen ve ark. (2006) tarafından önerilen yöntemin uygulaması için $Y_1 \sim N(94, 4)$, $Y_2 \sim N(96, 3)$, $Y_3 \sim N(97, 1)$, $Y_4 \sim N(95, 2)$ ve $Y_5 \sim N(98, 0,5)$ yığınlarının her birinden sırasıyla 18, 45, 60, 30 ve 23 hacimli örnekler seçilmiştir [EK-4]. Bu veri, ortalamaların eşitliği hipotezini test etmek amacıyla kullanılacaktır.

Eş.16.1’de verilen tek yönlü ANOVA modeli altında,

yığın ortalamaları μ_i olmak üzere;

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu$$

hipotezi test edilmek istenmektedir. Bu amaçla ikinci bölümde anlatılan Tek Örneğe Dayalı yöntem uygulanabilir.

ANOVA modeli kukla değişkenler tanımlanarak

$$\tilde{Y}_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \varepsilon_i$$

gibi bir regresyon modeline dönüştürülebilir. Burada, $\beta_0 = \mu_1$, $\beta_1 = \mu_2 - \mu_1$, $\beta_2 = \mu_3 - \mu_1$, $\beta_3 = \mu_4 - \mu_1$ ve $\beta_4 = \mu_5 - \mu_1$ ’dir. X_1, X_2, X_3, X_4 kukla değişkenleri Bölüm 3.3’teki gibi tanımlanır. Regresyon modeli için tasarım matrisi,

$$X = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

'dır. ANOVA modeli için kurulan $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu$ 'nün test edilmesi ile regresyon modeli için kurulan $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ hipotezinin test edilmesi eş anlamlıdır.

Çizelge 4.1. Örnek hacimleri I

	Örnek1	Örnek2	Örnek3	Örnek4	Örnek5
$n_i - 1$	17	44	59	29	22

olmak üzere, ilk olarak $\bar{Y}_i$ ve S_i^2 istatistikleri

$$\bar{Y}_i = \frac{\sum_{j=1}^{n_i-1} Y_{ij}}{n_i - 1} \quad \text{ve} \quad S_i^2 = \frac{\sum_{j=1}^{n_i-1} (Y_{ij} - \bar{Y}_i)^2}{n_i - 2}, \text{den}$$

Çizelge 4.2. $\bar{Y}_i$ ve S_i^2 değerleri

	Örnek1	Örnek2	Örnek3	Örnek4	Örnek5
$\bar{Y}_i$	93,9535	96,2155	97,2592	94,7431	96,7455
S_i^2	4,10015	2,95930	1,37854	3,25619	0,75705

olarak bulunmuştur.

S_i^2 değerlerinin en büyüğü $S_{[k]}^2$ istatistiğinin değeri $S_{[k]}^2 = 4,10015$ 'tir.

Çizelge 4.3. U_i ve V_i değerleri

	Örnek1	Örnek2	Örnek3	Örnek4	Örnek5
U_i	0,055556	0,024302	0,019715	0,036485	0,062957
V_i	0,05556	-0,06930	-0,16321	-0,05805	-0,38506

elde edilerek

$$W_{ij} = \begin{cases} U_i & , 1 \leq j \leq n_i - 1, \\ V_i & , j = n_i \end{cases}$$

kullanılarak Eş.4'te tanımlanan ağırlıklı örnek ortalaması $\tilde{Y}_i$ 'ler, aşağıdaki gibi hesaplanmıştır.

Çizelge 4.4. Ağırlıklı örnek ortalamaları I

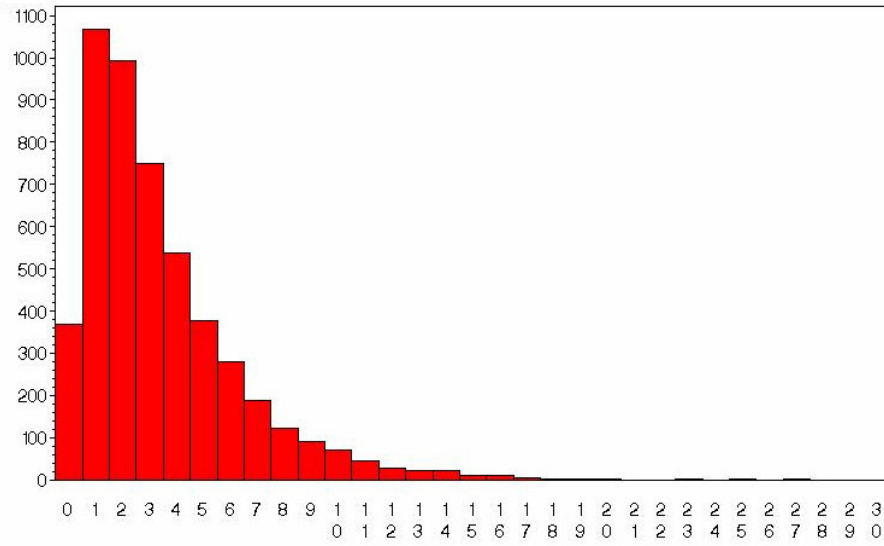
	Örnek1	Örnek2	Örnek3	Örnek4	Örnek5
$\tilde{Y}_i$	94,1300	96,2131	97,2476	94,8292	96,1160

$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ hipotezinin testinde, Eş.17.1'de verilen $\tilde{F}$ 'in olasılık dağılımı, yokluk hipotezinin doğruluğu altında yapılan 2000 deneme ile oluşturulmuştur. Üretilen veri için istatistiğin hesaplanan değeri 55,0305 ile istatistiğin p değeri, oluşturulan dağılımdan 0,0001 olarak bulunmuştur. 2000 tekrar yapılmasının nedeni, test istatistiğinin p değerinin ve $\tilde{F}$ 'in frekans dağılımının istikrarlı, durağan olmasıdır. %1 anlamlılık düzeyinde, p değeri 0,0001 olduğundan $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ (ya da aynı anlamdaki $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$) reddedilir. Yani ortalamalar arasında fark vardır.

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ hipotezi, bilinen F testi kullanılarak test edildiğinde, (4, 171) serbestlik derecesiyle F istatistiği 24,008 değerini almıştır. Bilinen F testi, %1 anlamlılık düzeyinde Tek Örneğe Dayalı yöntem ile aynı sonuca ulaşılır ve ortalamaların eşitliği reddedilir; ancak tek, iki ya da çok yönlü ANOVA modelleri için test yöntemleri, normallik, bağımsızlık ve varyansların homojenliği varsayımlarına dayanır. F testine ilişkin sağlamlık çalışmaları, normallik varsayımının sağlanmamasının, ortalamalar ile ilgili sonuçlar üzerinde çok küçük bir etkisi olduğunu göstermiştir. Bununla birlikte, özellikle örnek hacimlerinin eşit

olmaması durumunda, bağımsızlığın ya da homojen varyans varsayımlarının sağlanmamasının, ortalamalar ile ilgili sonuçlar üzerinde ciddi etkiye sahip olduğu görülmüştür [5]. Burada, varyansların homojenliğine ilişkin varsayım sağlanmadığından F test'i ile ilgili sonuç, sağlıklı bir sonuç değildir.

2000 tekrara dayalı olarak elde edilen Eş.17.1'deki $\tilde{F}$ istatistiğinin frekans dağılımının histogramı Şekil.4.1.'deki gibidir.



Şekil.4.1. Eş.17.1'deki $\tilde{F}$ test istatistiğinin histogramı

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu$ hipotezi reddedildikten sonra, ortalamalar arasında ikili karşılaştırmalar da yapılabilir. 1'inci ve 5'inci ortalamaların eşit olup olmadığı ile ilgilenilirse, Eş.16.1'deki ANOVA modeli altında,

$$H_0^* : \mu_1 = \mu_5$$

hipotezi test edilir. Regresyon yöntemiyle, H_0^* 'ın doğruluğu altında model, $\beta_4 = 0$

($\beta_4 = \mu_5 - \mu_1$) olmak üzere

$$\tilde{Y}_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i^*$$

şeklinde indirgenmiş modele dönüştürülebilir.

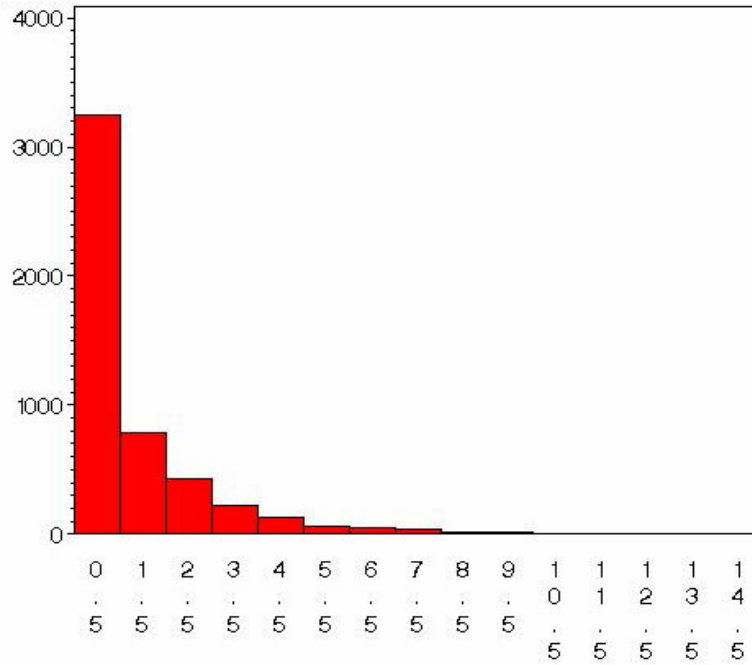
İndirgenmiş tasarım matrisi,

$$X = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

‘dir. $H_0^* : \beta_4 = 0$ kısmi testinde (ya da aynı anlama gelen $H_0^* : \mu_1 = \mu_5$), Eş.13’te verilen $\tilde{F}_R$ ’nin yokluk hipotezinin doğruluğu altında 2000 tekrar sonucu dağılımı oluşturulmuştur. Üretilen veri için, istatistiğin değeri 9,8606 ile oluşturulan dağılımdan istatistiğin p değeri 0,0034 olarak bulunmuştur. Bu durumda, %1 anlamlılık düzeyinde $H_0^* : \beta_4 = 0$ (veya $H_0^* : \mu_1 = \mu_5$) hipotezi reddedilir.

$H_0^* : \mu_1 = \mu_5$, bilinen F testi kullanılarak test edildiğinde, (1, 39) serbestlik derecesiyle F istatistiği 30,37 değerini alır. p değeri ise 0,0001’dir. Bilinen F testi, %1 anlamlılık düzeyinde Tek Örneğe Dayalı yöntem ile aynı sonuca ulaşır. Yani, birinci ve beşinci ortalamaların eşit olduğunu iddia eden yokluk hipotezi reddedilir; ancak varyansların homojenliği varsayımının sağlanmamasından dolayı bilinen F testinin kullanılması önerilmemektedir.

2000 tekrar sonucu elde edilen Eş.13’teki $\tilde{F}_R$ istatistiğinin frekans dağılımının histogramı Şekil.4.2.’de verilmiştir.



Şekil.4.2. Eş.13'teki $\tilde{F}_R$ test istatistiğinin histogramı

4.2. Genel Doğrusal Hipotezler için Sayısal Örnek

Wen ve ark. önerdikleri tek örneğe dayalı örnekleme yöntemini, genel doğrusal bir hipotezi test etmek amacıyla, bir veri seti üzerinde uygulamışlardır [18]. Bu çalışmada kullanılan veri Çizelge 4.5.'te verilmiştir. Deneyde, faktör A düşük, orta ve yüksek olmak üzere üç farklı protein kaynağını ve B faktörü de düşük ve yüksek olmak üzere iki farklı kalori düzeyini ifade etmektedir. Domuzları beslemek amacıyla verilen farklı düzeylerdeki kalori ve protein ile, hayvanların ağırlıklarındaki artış (lbs cinsinden, $1\text{kg} \cong 2,20\text{ lbs}$) altı aylık bir süreçte kaydedilmiştir. İki yönlü ANOVA modeli aşağıdaki gibi kurulmuştur.

$$Y_{ijl} = \mu_{ij} + e_{ijl}, \quad i = 1, \dots, I; \quad j = 1, \dots, J; \quad l = 1, \dots, n_{ij}; \quad e_{ij} \sim N(0, \sigma_{ij}^2).$$

Kullanılan veri, farklı protein kaynaklarını içeren A faktörü ve farklı iki kalori düzeyi için B faktörünün simülasyon çalışmasında bazı özel μ_{ij}, σ_{ij} ve n_{ij} değerleri için aşağıdaki gibi genelleştirilmiştir.

Çizelge 4.5. Belirlenen parametreler

	Düze	B Faktörü	
		düşük	yüksek
A Faktörü	düşük	$\mu_{11} = 90, \sigma_{11} = 1, n_{11} = 10$	$\mu_{12} = 105, \sigma_{12} = 12, n_{12} = 16$
	orta	$\mu_{21} = 100, \sigma_{12} = 5, n_{21} = 15$	$\mu_{22} = 120, \sigma_{22} = 20, n_{22} = 25$
	yüksek	$\mu_{31} = 100, \sigma_{13} = 8, n_{31} = 18$	$\mu_{32} = 130, \sigma_{32} = 25, n_{32} = 20$

Kukla değişkenler Bölüm 3.5'te tanımlandığı gibidir. Kukla değişkenlerin tanımlanmasından sonra, iki yönlü ANOVA modeli genel doğrusal regresyon modeli şeklinde yeniden yazılabilir.

$$Y_{ij} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + e_{ij}, \quad i = 1, \dots, k, \quad j = 1, \dots, n_i.$$

Bu durumda $k = 3 \times 2 = 6$ satır, $p = 3 + 2 - 2 = 3$ kestirici, $n_{11} = n_1, n_{21} = n_2, n_{31} = n_3, n_{12} = n_4, n_{22} = n_5, n_{32} = n_6$ 'dır. Regresyon ve ANOVA parametreleri arasındaki ilişki Çizelge 3.5.1.'de verildiği gibidir. Tek örneğe dayalı yöntemin önerdiği regresyon modeli, $S_1^2, \dots, S_k^2$ biliniyorken $\varepsilon_i \sim (0, \frac{S_{[k]}^2}{n_i} \cdot \frac{\sigma_i^2}{S_i^2})$; $i = 1, \dots, k$ olmak üzere,

$$\tilde{Y}_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i, \quad i = 1, \dots, 6,$$

$$X = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}, \quad (X'X)^{-1} = \begin{bmatrix} -\frac{2}{3} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{3} \\ -\frac{1}{2} & -1 & -\frac{1}{2} & 0 \\ -\frac{1}{2} & -\frac{1}{2} & -1 & 0 \\ -\frac{1}{3} & 0 & 0 & \frac{2}{3} \end{bmatrix}, \text{ dir.}$$

Test edilmek istenen doğrusal hipotez iki yönlü ANOVA modeli altında

$$H_0 = \begin{cases} \mu_{21} = \frac{\mu_{11} + \mu_{31}}{2} \\ \frac{\mu_{11} + \mu_{21} + \mu_{31}}{3} = \frac{\mu_{12} + \mu_{22} + \mu_{32}}{3} - 15 \end{cases}$$

olsun.

Çizelge 4.6. Domuzların ağırlıklarındaki artış miktarı (lbs cinsinden)

Düzey	Düşük		Yüksek	
Düşük	90,282	89,450	93,627	102,579
	90,073	91,851	91,416	106,161
	90,252		91,653	105,365
	90,379		101,102	115,268
	89,115		96,902	92,985
	89,629		102,084	84,909
	89,976		114,389	96,865
	91,111		115,694	106,997
Orta	103,201	107,160	127,452	141,728
	102,159	94,245	88,787	96,200
	106,156		75,826	124,776
	96,056		145,686	139,196
	101,278		114,189	99,681
	93,331		124,493	140,713
	90,085		140,500	133,513
	101,698		153,055	91,931
	106,591		126,039	97,107
	99,220		129,734	113,135
	101,219		149,506	131,948
	93,229		107,551	144,173
	102,759		176,063	
	128,229	109,628	127,001	104,611
	114,671	109,598	151,370	87,044
	108,770	115,240	168,075	143,657
Yüksek	121,901	114,976	109,568	146,720
	101,877	121,090	100,157	105,988
	102,199	112,109	93,417	133,503
	108,523	100,747	113,738	135,534
	102,225	107,431	128,303	152,604
	110,596		130,824	100,935
	110,326		139,196	104,752

Çizelge 4.6.’dan da görüldüğü üzere altı farklı durum söz konusudur. A faktörünün ve B faktörünün birinci düzeyinin yer aldığı veri grubuna “Örnek1”, A faktörünün

ikinci, B faktörünün birinci düzeyinin yer aldığı veri grubuna “Örnek2” vb. denilebilir.

Çizelge 4.7. Örnek hacimleri II

	Örnek1	Örnek2	Örnek3	Örnek4	Örnek5	Örnek6
n_i	10	15	18	16	25	20

‘dir.

Yapılan normallik testi sonucunda p değeri 0,28 değerini almıştır. %1 anlamlılık düzeyinde bu örnek normal dağılımdan gelmiştir hipotezi reddedilememiştir. Varyansların eşitsizliği durumunda önerilen yöntem ile test edilmek istenen

$$H_0 = \begin{cases} \mu_{21} = \frac{\mu_{11} + \mu_{31}}{2} \\ \frac{\mu_{11} + \mu_{21} + \mu_{31}}{3} = \frac{\mu_{12} + \mu_{22} + \mu_{32}}{3} - 15 \end{cases}$$

hipotezi test edilebilir. $H_0 : K\beta = h$ için, hipotez matrisi, h vektörü sırasıyla

$$K = \begin{bmatrix} 0 & 1 & -0,5 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} \text{ ve } h = \begin{bmatrix} 0 \\ 15 \end{bmatrix}, \text{dir.}$$

Harmonik örnek büyüklüğü, $\tilde{n} = [(\sum_{i=1}^k 1/n_i)/k]^{-1}$ ’den hesaplanmıştır.

Çizelge 4.8. $\tilde{n}$, U_i ve V_i değerleri

$\tilde{n}$						
16.0119	Örnek1	Örnek2	Örnek3	Örnek4	Örnek5	Örnek6
U_i	1.46019	0.14506	0.09577	0.10021	0.04000	0.05363
V_i	-12.1417	-1.0308	-0.6280	-0.5031	0.0400	-0.0190

ve ağırlıklı örnek ortalamaları

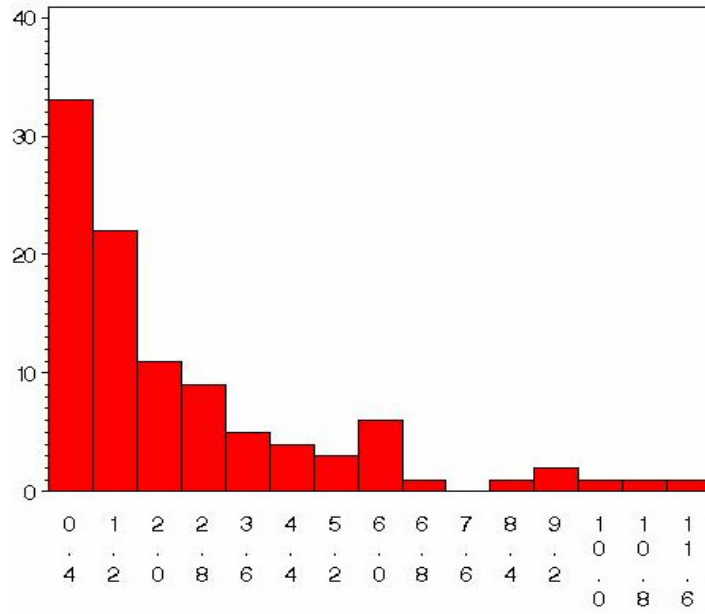
Çizelge 4.9. Ağırlıklı örnek ortalamaları II

	Örnek1	Örnek2	Örnek3	Örnek4	Örnek5	Örnek6
$\tilde{Y}_i$	67.916	106.430	113.788	97.582	124.519	125.237

olarak bulunmuştur.

$H_0 : K\beta = h$ hipotezinin doğruluğu altında, Eş.21.1’de verilen $\tilde{F}$ ’nin dağılımını oluşturmak üzere 1000 deneme yapılmıştır. Çizelge 4.2.1’deki belirlenen parametreler ile üretilmiş veri seti için test istatistiğinin p değeri 0.37 bulunmuştur.

Bu durumda %1 anlamlılık düzeyinde $H_0 : K\beta = h$ hipotezinin reddedilememiştir.



Şekil.4.3. Eş.21.1’deki $\tilde{F}$ test istatistiğinin histogramı

4.3 Simülasyon Çalışması

Çalışmanın bu bölümünde, bir bilgisayar paket programı kullanılarak, varyansların eşitsizliği durumunda kullanılmak üzere, tek örneğe dayalı yöntemin ANOVA için önerdiği Eş.17.1’de verilen test istatistiği $\tilde{F}$ ve bilinen F testinin, güçleri bakımından karşılaştırılması yapılacaktır [EK-5].

Yokluk hipotezi,

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$

‘dir. Eş.16.1’de verilen ANOVA modeli kukla değişkenler tanımlanarak

$$\tilde{Y}_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i$$

olarak ifade edilir. Tasarım matrisi,

$$X = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \text{dir.}$$

Bu simülasyon çalışmasında örnekler normal dağılımdan türetilerek seçilmiştir.

Örnek sayısı (k) 4 ve grup ortalamaları (G_f) arasındaki farklar için 1, 4 ve 10 değerleri alınmıştır. Örnek hacimleri sırasıyla (n_1, n_2, n_3, n_4) olmak üzere birinci durum için küçük ve eşit (4,4,4,4), ikinci durum için küçük ve birbirinden farklı (4,6,8,10), üçüncü durum için büyük ve eşit (10,10,10,10), dördüncü durum için büyük ve birbirinden biraz farklı (10,15,20,25) ve son durum için büyük ve birbirinden çok farklı (10,30,50,70) değerler olarak kullanılmıştır.

Yığın varyansları, homojen (1,1,1,1), az heterojen (1,2,3,4) ve daha heterojen (1,5,9,13) olarak seçilmiştir.

Varyansların eşitsizliği durumunda önerilen tek örneğe dayalı yöntemin tek yönlü varyans analizi için önerdiği $\tilde{F}$ istatistiğinin dağılımını oluşturmak üzere $\mu_0 = 10$ alınarak 2000 deneme yapılmıştır.

H_0 hipotezinin yanlış olduğu biliniyorken, her iki test için de örnek seçme işlemi 2000 defa tekrarlanarak, $\alpha = 0,05$ iken, güçleri yani gerçekte yanlış olan H_0 hipotezi red oranları simülasyon programı ile elde edilmiştir. Çizelge 4.10'da verilmiştir.

Çizelge 4.10. $\tilde{F}$ ve bilinen F istatistikleri için çeşitli n , varyans ve Gf değerlerine göre 2000 tekrarda, gerçekte yanlış olan H_0 hipotezi red oranları [$\alpha = 0,05$]

n_1, n_2, n_3, n_4	$\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2$	Gf	$\tilde{F}$	F
4,4,4,4	1,1,1,1	1	0,4230	0,9110
		4	0,9986	0,9993
		10	0,9999	0,9999
	1,2,3,4	1	0,2495	0,2155
		4	0,8750	0,9960
		10	0,9995	0,9998
	1,5,9,13	1	0,1095	0,1090
		4	0,5025	0,3300
		10	0,9575	0,9495
4,6,8,10	1,1,1,1	1	0,7185	0,9970
		4	0,9999	0,9999
		10	0,9999	0,9999
	1,2,3,4	1	0,4230	0,1945
		4	0,9995	0,9990
		10	0,9999	0,9999
	1,5,9,13	1	0,2205	0,0415
		4	0,8765	0,3330
		10	0,9999	0,9995
10,10,10,10	1,1,1,1	1	0,9050	0,9999
		4	0,9999	0,9999
		10	0,9999	0,9999
	1,2,3,4	1	0,5250	0,5065
		4	0,9999	0,9994
		10	0,9999	0,9999
	1,5,9,13	1	0,2535	0,1170
		4	0,9450	0,7630
		10	0,9999	0,9999

Çizelge 4.10. (Devam). $\tilde{F}$ ve bilinen F istatistikleri için çeşitli n , varyans ve Gf değerlerine göre 2000 tekrarda, gerçekte yanlış olan H_0 hipotezi red oranları [$\alpha = 0,05$]

n_1, n_2, n_3, n_4	$\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2$	Gf	$\tilde{F}$	F
10,15,20,25	1,1,1,1	1	0,9870	0,9999
		4	0,9999	0,9999
		10	0,9999	0,9999
	1,2,3,4	1	0,8320	0,6095
		4	0,9999	0,9990
		10	0,9999	0,9999
	1,5,9,13	1	0,3605	0,0610
		4	0,9995	0,8550
		10	0,9999	0,9999
10,30,50,70	1,1,1,1	1	0,9905	0,9999
		4	0,9999	0,9999
		10	0,9999	0,9999
	1,2,3,4	1	0,9840	0,8690
		4	0,9999	0,9999
		10	0,9999	0,9999
	1,5,9,13	1	0,5870	0,0505
		4	0,9999	0,9820
		10	0,9999	0,9999

Çizelge 4.10 incelendiğinde aşağıdaki bulgular elde edilir.

Her iki test için de, örnek hacimleri arttığında, gerçekte yanlış olan H_0 hipotezini red oranlarının yani güç değerlerinin arttığı görülmektedir.

Varyansların homojen olduğu (1,1,1,1) durumunda, $\alpha = 0,05$ için, örnek hacimleri küçük ve grup ortalamaları arasındaki fark da küçük iken bilinen F testi, $\tilde{F}$ 'ya göre daha güçlüdür. Örneğin, örnek hacmi (4,4,4,4) ve Gf = 1 için bilinen F testinin (0,9110) ile $\tilde{F}$ 'den (0,4230) ile daha güçlü olduğu görülmüştür. Ancak bu durum, örnek hacimleri arttıkça ortadan kalkmış ve Gf= 4 ve Gf= 10 için her iki testin gücü

de birbirine eşit çıkmıştır. Örneğin, örnek hacmi (10,15,20,25), $Gf=4$ ve $Gf=10$ iken her iki testin gücü de 0,9999'dur.

Varyansların heterojenliğinin az olduğu (1,2,3,4) durumunda, $\alpha = 0,05$ için, örnek hacmi ve Gf küçük iken, $\tilde{F}$ 'nin bilinen F testine göre daha güçlü olduğu görülmüştür. Örneğin, örnek hacmi (4,4,4,4) ve $Gf=1$ iken $\tilde{F}$ 'nin gücü (0,2495) ile F 'in gücünden (0,2155) ile daha büyüktür. Aynı şekilde, örnek hacmi (4,6,8,10) ve $Gf=1$ iken $\tilde{F}$ 'nin gücü (0,4230) ile bilinen F 'in gücünden (0,1945) ile çok daha büyüktür. Ancak örnek hacmi ve Gf arttıkça, her iki testin de güç kazandığı görülmektedir. Örneğin, örnek hacmi (10,30,50,70), $Gf=4$ ve $Gf=10$ olmak üzere, her iki testin gücünün de 0,9999 olduğu sonucuna ulaşılmıştır. Varyansların eşitsizliği durumunda, örnek hacmi küçük ve grupların ortalamaları arasındaki fark da küçük olduğunda, $\tilde{F}$, bilinen F istatistiğine göre daha hassas sonuçlar vermiştir.

Varyansların heterojenliğinin fazla olduğu (1,5,9,13) durumunda, $\alpha = 0,05$ için, örnek hacmi büyük ya da küçük olsun, gruplar arası fark küçük olduğunda ($Gf=1$), $\tilde{F}$, bilinen F istatistiğine göre daha güçlüdür. Örneğin, örnek hacmi (4,6,8,10) ve (10,15,20,25), $Gf=1$ iken, $\tilde{F}$ (0,2205, 0,3605) ile bilinen F 'den (0,0415, 0,0610) ile daha güçlüdür. Örnek hacmi ve Gf arttıkça her iki testin de güç kazandığı ancak, $\tilde{F}$ 'nin gücünün $Gf=4$ olduğu her durumda daha büyük olduğu görülmektedir. Her iki testin gücü, $Gf=10$ iken örnek hacmi arttığında eşit bulunmuştur.

Varyansların eşitsizliği durumunda, örnek hacmi küçük ve orta değerler aldığı ve grupların ortalamaları arasındaki fark küçük ya da çok büyük olmadığında, $\tilde{F}$, bilinen F istatistiğine göre daha güçlüdür. Örnek hacmi arttığında, Gf büyük olmasına karşın her iki test de güç bakımından birbirine eşittir.

Bilinen F istatistiği, varyansların homojenliği varsayımının sağlandığı her noktada $\tilde{F}$ 'ya göre daha güçlüdür; ancak varyansların homojenliği bozulduğunda güç kaybetmektedir. $\tilde{F}$ da varyansların homojenliği bozulduğunda güç kaybetmektedir; ancak bu kayıp bilinen F testindekinden daha küçüktür. Örneğin, örnek hacmi

(4,6,8,10), $G_f=1$ iken ve varyanslar homojenken (1,1,1,1), test güçleri $\tilde{F}$ için 0,7185, F için 0,9970'tir. Varyanslar (1,2,3,4) için ise sırasıyla 0,4230 ve 0,1945'tir. Bu durumda yaklaşık olarak, bilinen F %81 oranında güç kaybına uğrarken, $\tilde{F}$, %41 oranında güç kaybetmiştir.

5. SONUÇ VE ÖNERİLER

Bir çok deneysel durumda hata terimlerinin eşit varyansa sahip olduğunu belirten varsayım sağlanmaz.

Bu tez çalışmasında, Wen ve ark. (2006) tarafından önerilen tek örneğe dayalı yöntem tanıtılarak, hata varyanslarının eşitsizliği durumunda, genel doğrusal regresyon modelinde, regresyon parametrelerinin tamamının ya da herhangi bir alt kümesinin test edilmesi üzerinde durularak, ANOVA testlerinin, genel bir doğrusal regresyon modeli kullanılarak kısmi test yoluyla gerçekleştirilebileceği gösterilmiştir.

Önerilen tüm test istatistiklerinin dağılımları, bilinmeyen yığın varyanslarından tamamıyla bağımsız olan, yokluk hipotezleri altında bağımsız Student t dağılımının doğrusal kombinasyonlarının karesel formlarıdır.

Regresyon parametrelerinin bir alt kümesinin test etmek üzere sınanan $H_0 : \beta_{q+1} = \dots = \beta_p = 0$ hipotezi, $q = 0$ olması durumunda, tüm parametrelerin sıfır olduğu bir tümel test olur. Kısmi test, $(p + 1)$ tane regresyon parametresinin herhangi bir altkümesine uygulanabilir.

Varyans analizi, genel doğrusal regresyon modelinin özel bir durumu olarak, kukla değişkenler kullanılarak kısmi test yoluyla gerçekleştirilebilir. ANOVA modelleri için test yöntemleri, normallik, bağımsızlık ve varyansların homojenliği varsayımlarına dayanır. Varyans analizi problemlerinde kullanılan F testi, bu varsayımlar sağlanıyorsa kullanılmalıdır. Çünkü yanlış olan yokluk hipotezini reddetme olasılığı yani testin gücü, diğer testlerden daha büyüktür. Normallik, bağımsızlık ve homojen hata varyanslılığı varsayımlarından birinin bile sağlanmaması durumunda F testinin kullanılması uygun değildir.

Tek örneğe dayalı yöntem, genel doğrusal regresyon modelindeki hata varyanslarının eşitsizliği durumunda önerilmektedir. Uygulamada hata varyanslarının homojenliği

sağlanması zor bir varsayım olduğundan önerilen bu yöntem daha avantajlıdır. Böylece, istatistiksel veri analizi çalışmalarında, hata varyanslarının heterojen olması durumuyla karşılaşıldığında bu yöntem bir alternatif sağlayacaktır.

Yaptığımız çalışmalar ile, tek örneğe dayalı yöntemin önerdiği test istatistiği $\tilde{F}$ 'nın ve bilinen F testinin güçlerinin, yanlış olan sıfır hipotezini red oranlarının, varyansların homojen olması durumunda, örnek hacimleri arttıkça aynı olduğu görülmüştür. Varyansların eşitsizliği durumunda, $\tilde{F}$ 'nın, test edilen ortalamalar arasındaki fark küçük olduğunda bilinen F testine göre daha iyi sonuçlar verdiği bulunmuştur. Varyansların heterojenliği arttığında $\tilde{F}$ 'nın, örnek hacminden bağımsız olarak F testine göre daha çok güç kazandığı sonucuna ulaşılmıştır.

KAYNAKLAR

1. Wen, M. J., Chen, H. J., Chen, S. Y., “ On testing a subset of regression parameters under heteroscedasticity”, *Computational Statistics & Data Analysis*, 1-19 (2006).
2. Horn, S. D., Horn, R. A., Duncan, D. B., “Estimating heteroscedastic variances in linear models”, *J. Amer. Statist. Assoc.*, 70: 380–385 (1975).
3. White, H., “A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity”, *Econometrica*, 48: 817–838 (1980).
4. Cribari-Neto, F., “Asymptotic inference under heteroskedasticity of unknown form”, *Comput. Statist. Data Anal.*, 45: 215–233 (2004).
5. Bishop, T. A., Dudewicz, E. J., “Exact analysis of variance with unequal variances: test procedures and tables”, *Technometrics*, 20: 419-430 (1978).
6. Bishop, T. A., “A Stein two-sample procedure for the general linear model with unequal error variances”, *Comm. Statist. Theory Methods*, A7 (5): 495–507 (1978).
7. Chen, H. J., Lam, K., “Single-stage interval estimation of the largest normal mean under heteroscedasticity”, *Comm. Statist. Theory Methods* , 10: 3703–3718 (1989).
8. Chen, S. Y., Chen, H. J., “Single-stage analysis of variance under heteroscedasticity”, *Comm. Statist. Simulation*, 27 (3): 641–666 (1998).
9. Chen, H.J., “Strong consistency and asymptotic unbiasedness of a natural estimator for a ranked parameter”, *Sankhyā Ser. B Pt.*, 1 (38): 92–94 (1976)
10. Stein, C. M., “A two-sample test for a linear hypothesis whose power is independent of the variance”, *Ann. Math. Statist.*, 16: 243–258 (1945).
11. Myers, R. H., Miton, J. S., “A first course in thr theory of linear statistical models”, *PWS-Kent*, Boston, 83-90, 150-158, 245-253 (1991).
12. Miller, A., “Least-squares computations”, *Subset Selection in Regression* 2nd ed., *A CRC Pres Company*, Boca Raton, London, New York, Washington, 11-17 (2002).
13. Albayrak, A.S. ve arkadaşları, “SPSS Uygulamalı Çok Değişkenli İstatistik Teknikleri”, *Asil Yayın Dağıtım LTD. STİ.*, Ankara, 259-266 (2005).
14. İnternet: Ankara Üni. Web sayfası, “İst306 İstatistik Deney Tasarımı” <http://science.ankara.edu.tr/~ozturk/Dersler/ist306/Giris.doc> (2007).

15. Ott, R.L., “An Introduction Statistical Methods and Data Analysis”, 4th ed., **Wadsworth, Inc.**, Belmont, 897-906 (1993).
16. Alalouf, I.S., Styan, G. P. H., “characterizations of estimability in the general linear model”, *The Annals of Statistics*, 7 (1): 194-200 (1979).
17. Searle, S. R., “ The general hypothesis”, *Linear Models*, **Wiley**, New York, 111-113 (1971).
18. Internet: National Cheng Kung Uni., Technical Reports, “GeneralLinearModels” <http://www.stat.ncku.edu.tw/faculty/mjwen/MjWenTR.htm> (2004).
19. Akdi, Y., “En iyi yansız tahmin ediciler”, Matematiksel İstatistiğe Giriş, **Bıçaklar Kitabevi**, Ankara, 317-322 (2005).

EKLER

EK-1 U_i ve V_i ağırlıklarının sağlaması gereken koşulların gösterilmesi

Eş.3'teki koşullar,

$$\text{i. } (n_i - 1)U_i + V_i = 1$$

$$\text{ii. } (n_i - 1)U_i^2 + V_i^2 = S_{[k]}^2 / (n_i S_i^2)$$

olarak verildi. İlk koşul için,

$$U_i = \frac{1}{n_i} + \frac{1}{n_i} \sqrt{\frac{1}{n_i - 1} [S_{[k]}^2 / S_i^2 - 1]} \text{ ifadesi}$$

$$U_i = \frac{1}{n_i} + \frac{1}{n_i S_i} \sqrt{\frac{S_{[k]}^2 - S_i^2}{n_i - 1}} \text{ olarak yeniden yazılabilir.}$$

Karekök içindeki $\frac{S_{[k]}^2 - S_i^2}{n_i - 1}$, A^2 olsun. Buradan,

$$U_i = \frac{1}{n_i} + \frac{A}{n_i S_i} \text{ yazılabilir ve}$$

$$\text{I. } A = S_i(n_i U_i - 1)$$

bulunur.

$$V_i = \frac{1}{n_i} - \frac{1}{n_i} \sqrt{(n_i - 1) [S_{[k]}^2 / S_i^2 - 1]} \text{ ifadesi}$$

$$V_i = \frac{1}{n_i} - \frac{1}{n_i S_i} \sqrt{(n_i - 1)(S_{[k]}^2 - S_i^2)} \text{ olarak yeniden yazıldığında, karekök içindeki}$$

ifade, $A^2(n_i - 1)$ 'e eşit olur. Buradan,

$$V_i = \frac{1}{n_i} - \frac{A(n_i - 1)}{n_i S_i} \text{ yazılabilir ve}$$

EK-1 (Devam) U_i ve V_i ağırlıklarının sağlaması gereken koşulların gösterilmesi

$$\text{II.} \quad A = \frac{S_i(1 - n_i V_i)}{n_i - 1}$$

olarak bulunur.

I ve II'den

$$S_i(n_i U_i - 1) = \frac{S_i(1 - n_i V_i)}{n_i - 1}$$

yazılabilir. Gerekli işlemlerden sonra,

$$(n_i - 1)U_i + V_i = 1$$

elde edilir. Böylece Eş.3'ün ilk koşulunun geçerliliği görülür.

Diğer koşul için,

$$U_i = \frac{1}{n_i} + \frac{A}{n_i S_i} \text{ olarak yeniden yazılan ifadenin karesi alınarak,}$$

$$U_i^2 = \frac{1}{n_i^2} + \frac{A^2}{n_i^2 S_i^2} + \frac{2A}{n_i^2 S_i}$$

elde edilir. Gerekli işlemlerden sonra,

$$\text{III.} \quad A^2 + 2AS_i = n_i^2 S_i^2 U_i^2 - S_i^2$$

bulunur.

$$V_i = \frac{1}{n_i} - \frac{A(n_i - 1)}{n_i S_i} \text{ olarak tekrar yazılan ifadenin karesi alınarak,}$$

$$V_i^2 = \frac{1}{n_i^2} + \frac{A^2(n_i - 1)^2 - 2A(n_i - 1)S_i}{n_i^2 S_i^2}$$

elde edilir. Gerekli işlemlerden sonra,

$$n_i^2 S_i^2 V_i^2 - S_i^2 = (n_i - 1)(A^2 n_i - (A^2 + 2AS_i))$$

EK-1 (Devam) U_i ve V_i ağırlıklarının sağlaması gereken koşulların gösterilmesi

elde edilir. Bu eşitlikte yer alan A^2 yerine eşiti $(S_{[k]}^2 - S_i^2)/(n_i - 1)$, $A^2 + 2AS_i$ yerine de III'te yer alan eşiti $n_i^2 S_i^2 U_i^2 - S_i^2$ yazılıp gerekli düzenlemeler yapılırsa,

$$(n_i - 1)U_i^2 + V_i^2 = S_{[k]}^2 / (n_i S_i^2)$$

elde edilir. Böylece Eş.3'ün ikinci koşulunun da geçerliliği görülür.

EK-2 U istatistiği

$X_1, \dots, X_n$ dağılım fonksiyonu F (ve olasılık yoğunluk fonksiyonu f) olan bir yığından alınan bir örnek olsun. $\theta(F)$ dağılımın parametresi θ ' nın bir fonksiyonu ve θ 'nın en az bir yansız tahmin edicisi var olsun.

$$\theta(F) = E_F(h(X_1, \dots, X_n))$$

$$\theta(F) = \int \int \dots \int h(x_1, x_2, \dots, x_n) f(x_1) f(x_2) \dots f(x_n) dx_1 dx_2 \dots dx_n$$

olarak yazılabilir. Burada $h(x_1, x_2, \dots, x_n)$ fonksiyonuna çekirdek (öz veya kernel) denir. h fonksiyonu simetrik olarak alınabilir. θ parametresinin bir tahmin edicisi,

$$U_n = U(X_1, X_2, \dots, X_n) = \frac{1}{\binom{n}{m}} \sum_c h(X_{i_1}, X_{i_2}, \dots, X_{i_m})$$

olarak yazılır.

- $\theta(F) = E_F(X) = \mu$ olsun. Çekirdek fonksiyonu $h(x) = x$ alındığında U istatistiği,

$$U_n = \frac{1}{\binom{n}{1}} \sum_i h(X_i) = \frac{1}{n} \sum_i X_i = \bar{X}_n$$

'dir.

- $\theta(F) = [E_F(X)]^2 = \mu^2$ olsun. Çekirdek fonksiyonu $h(x_1, x_2) = x_1 x_2$ seçilirse U istatistiği,

$$U_n = \frac{1}{\binom{n}{2}} \sum_{1 \leq i \leq j \leq n} h(X_i, X_j) = \frac{2}{n(n-1)} \sum_{1 \leq i \leq j \leq n} X_i X_j$$

olur.

- $\theta(F) = \text{Var}_F(X) = \sigma^2$ olsun. Çekirdek fonksiyonu

EK-2 (Devam) U istatistiği

$$h(x_1, x_2) = \frac{x_1^2 + x_2^2 - 2x_1x_2}{2} = \frac{1}{2}(x_1 - x_2)^2 \text{ olarak seçildiğinde } U \text{ istatistiği,}$$

$$U_n = \frac{1}{\binom{n}{2}} \sum_{1 \leq i \leq j \leq n} h(X_i, X_j) = \frac{2}{n(n-1)} \sum_{1 \leq i \leq j \leq n} \frac{1}{2} (X_i - X_j)^2$$

$$U_n = \frac{1}{n(n-1)} \sum_{1 \leq i \leq j \leq n} (X_i - X_j)^2 = \frac{1}{n(n-1)} \sum_{1 \leq i \leq j \leq n} (X_i^2 + X_j^2 - 2X_iX_j)$$

$$U_n = \frac{n}{2n(n-1)} \sum_{i=1}^n X_i^2 + \frac{n}{2n(n-1)} \sum_{j=1}^n X_j^2 - \frac{1}{n(n-1)} \left[\sum_{i=1}^n X_i \right] \left[\sum_{j=1}^n X_j \right]$$

$$U_n = \frac{1}{2(n-1)} \sum_{i=1}^n X_i^2 + \frac{1}{2(n-1)} \sum_{j=1}^n X_j^2 - \frac{1}{n(n-1)} \left[\sum_{i=1}^n X_i \right]^2$$

$$U_n = \frac{1}{(n-1)} \sum_{i=1}^n X_i^2 - \frac{n^2 \bar{X}_n^2}{n(n-1)} = \frac{1}{(n-1)} \sum_{i=1}^n X_i^2 - \frac{n \bar{X}_n^2}{(n-1)}$$

$$U_n = \frac{1}{(n-1)} \left[\sum_{i=1}^n X_i^2 - n \bar{X}_n^2 \right] = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = S_n^2$$

olarak hesaplanır [19].

EK-3 Kukla değişkenler

Regresyon analizi, sayısal değişkenler kullanılarak yapılır. Ancak sosyal araştırmaların bir çoğunda kategorik değişkenler ile çalışılır. Bu durumda çözüm, kukla değişkenleri kullanmaktır. Kukla değişken, özel bir niteliği kodlayan değişkendir. Kukla değişken, her gözlem için 1 ya da 0 değerini alır. Örneğin , erkek için 1 ve kadın için 0 gibi. Kullanılacak kukla değişken sayısı, orijinal değişken sayısının bir eksiğidir $(k - 1)$.

Kukla değişkenler ayrıca ANOVA modelini regresyon modeline dönüştürerek, yığın ortalamalarının eşitliğini test eden hipotezi regresyon parametresi β 'ları kullanarak gerçekleştirmek amacıyla da kullanılır. Bağımlı değişken y 'nin, sayısal olmayan bağımsız değişkenler ile ilişkili olduğu deneysel bir durum için, model yazma problemi düşünülebilir. Örneğin, birbirinden 30 mil uzakta bulunan iki farklı izleme noktasının, bir fırtınada, dakika başına gerçekleşen ortalama yıldırım sayısını kıyaslamak isteyelim. y , bir dakikalık zaman süresince osiloskop'tan (elektriksel işaretlerin ölçülüp değerlendirilmesinde kullanılan alet) ölçülen yıldırım sayısını ifade etsin. İki yığın ortalamasının kıyaslanması için aşağıdaki modeller yazılabilir.

Birinci izleme noktası için; $y = \mu_1 + \varepsilon$

İkinci izleme noktası için; $y = \mu_2 + \varepsilon$.

Bu iki model tek bir model olarak ifade edilebilir.

$$y = \beta_0 + \beta_1 x_1 + \varepsilon$$

burada β_0 ve β_1 bilinmeyen parametreler, ε rasgele hata terimidir. Kukla değişken x_1 aşağıdaki gibi yorumlanır.

Eğer gözlem ikinci izleme noktasında meydana geldiyse $x_1=1$,

Eğer gözlem birinci izleme noktasında meydana geldiyse $x_1=0$.

EK-3 (Devam) Kukla değişkenler

Birinci izleme noktasında meydana gelen gözlemler için modelde x_1 yerine “0” konulursa,

$$y = \beta_0 + \beta_1(0) + \varepsilon = \beta_0 + \varepsilon$$

elde edilir. Burada $\beta_0 = \mu_1$ olduğu görülmektedir.

Aynı şekilde modelde $x_1=1$ yerine konularak, ikinci izleme noktasında meydana gelen gözlemler için de bir eşitlik yazılabilir.

$$y = \beta_0 + \beta_1(1) + \varepsilon = \beta_0 + \beta_1 + \varepsilon$$

$\beta_0 = \mu_1$ olduğundan, burada $\beta_0 + \beta_1$, μ_2 ’e eşit olmalıdır. Böylece, birinci ve ikinci izleme noktalarında meydana gelen gözlemler arasındaki ortalamaların farkı $\beta_1 = \mu_2 - \mu_1$ olur.

$y = \beta_0 + \beta_1 x_1 + \varepsilon$ modelinde, y sayısal olmayan bağımsız değişken izleme noktasına bağlıydı. Bu durum, sayısal olmayan değişkenin ikiden fazla düzeye sahip olduğu duruma genişletilebilir. Bu durumda birden fazla kukla değişken kullanılır. Örneğin, sayısal olmayan değişkenlerin dört düzeyi olduğu bir deneyi düşünelim. Tüm bu düzeyler *işlemler* olarak adlandırılır. Model,

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

olarak ifade edildiğinde kukla değişkenler de sırasıyla

işlem 2 ise $x_1 = 1$, değilse $x_1 = 0$

işlem 3 ise $x_2 = 1$, değilse $x_2 = 0$

işlem 4 ise $x_3 = 1$, değilse $x_3 = 0$

EK-3 (Devam) Kukla değişkenler

'dır. β 'ların yorumlanabilmesi için, beklenen değerlere ilişkin bir tablo hazırlanabilir. $E(\varepsilon) = 0$ varsayımı altında y 'nin beklenen değeri,

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

'dir. İşlem 1'e ait gözlemler için beklenen değer, $x_1 = 0$, $x_2 = 0$ ve $x_3 = 0$ alınarak $E(y) = \beta_0$ bulunur. İşlem 2'ye ait gözlemler için $x_1 = 1$, $x_2 = 0$ ve $x_3 = 0$ alınarak beklenen değer $E(y) = \beta_0 + \beta_1$ elde edilir. Bu şekilde beklenen değerlere ilişkin bir özet tablo verilebilir.

Çizelge 3.1. Beklenen değerler

İşlemler			
1	2	3	4
$E(y) = \beta_0$	$E(y) = \beta_0 + \beta_1$	$E(y) = \beta_0 + \beta_2$	$E(y) = \beta_0 + \beta_3$

İşlem 1'in ortalaması μ_1 , işlem 2'nin ortalaması μ_2 vb. olmak üzere,

$$\mu_1 = \beta_0, \quad \mu_2 = \beta_0 + \beta_1, \quad \mu_3 = \beta_0 + \beta_2 \quad \text{ve} \quad \mu_4 = \beta_0 + \beta_3 \text{ 'tür. Buradan da}$$

$$\beta_0 = \mu_1$$

$$\beta_1 = \mu_2 - \mu_1$$

$$\beta_2 = \mu_3 - \mu_1$$

$$\beta_3 = \mu_4 - \mu_1$$

olduğu görülmektedir. İşlemlere ait ortalamalar arasında yapılmak istenen herhangi bir kıyaslama β ' lar kullanılarak ifade edilebilir. Örneğin, $\mu_4 - \mu_3$, $\beta_3 - \beta_2$ olarak yazılabilir. Bu sonuçlar rasgele blok tasarımı, Latin kare tasarımı ve faktöriyel deneylere da genişletilebilir.

EK-3 (Devam) Kukla değişkenler

Latin kare tasarımına ilişkin bir örnek verelim. Naylon ipliğin döndüğü bir grup makine olsun. Kullanım sürecinde, naylon iplikte meydana gelen kopma, makine operatörünün ipliği yenilemesini gerektirir. Araştırmacılar, 3 farklı makine

operatörü (A, B, C) kullanarak, 3 farklı tipte olan döndürme makinelerinin (I, II, III) çıktılarını karşılaştırmak ile ilgileniyor olsun. Operatör değişikliğine ek olarak, günlük değişkenliği de kontrol edebilmek için 3×3 'lük bir Latin kare tasarımı karar veriliyor. Bu durum için , kukla değişkenler kullanılarak model yazılabilir.

Çizelge 3.2. Latin kare tasarımı

	Günler		
Operatör	1	2	3
A	I	II	III
B	II	III	I
C	III	I	II

Latin kare tasarım modelinde, 3 tane sayısal olmayan bağımsız değişkene bağlı bir y bağımlı değişkeni yer alır. Bağımsız değişkenler; satırlar (operatörler), sütunlar (günler) ve işlemlerdir (makineler). Model,

$$y = \beta_0 + \overbrace{\beta_1 x_1 + \beta_2 x_2}^{\text{operatörler}} + \overbrace{\beta_3 x_3 + \beta_4 x_4}^{\text{günler}} + \overbrace{\beta_5 x_5 + \beta_6 x_6}^{\text{makineler}} + \varepsilon$$

‘dir. Kukla değişkenler,

operatör B ise $x_1 = 1$, değilse $x_1 = 0$

operatör C ise $x_2 = 1$, değilse $x_2 = 0$

2. gün ise $x_3 = 1$, değilse $x_3 = 0$

EK-3 (devam) Kukla değişkenler

3. gün ise $x_4 = 1$, değilse $x_4 = 0$

makine 2 ise $x_5 = 1$, değilse $x_5 = 0$

makine 3 ise $x_6 = 1$, değilse $x_6 = 0$

olarak tanımlanır. Bu model için, $E(\varepsilon) = 0$ iken,

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6$$

‘dır. Kukla değişkenler için değerleri yerine konularak beklenen değerlere ilişkin bir özet tablo verilebilir.

Çizelge 3.3. Latin kare tasarımı için beklenen değerler

Günler			
Operatör	1	2	3
A	I β_0	II $\beta_0 + \beta_3 + \beta_5$	III $\beta_0 + \beta_4 + \beta_6$
B	II $\beta_0 + \beta_1 + \beta_5$	III $\beta_0 + \beta_1 + \beta_3 + \beta_6$	I $\beta_0 + \beta_1 + \beta_4$
C	III $\beta_0 + \beta_2 + \beta_6$	I $\beta_0 + \beta_2 + \beta_3$	II $\beta_0 + \beta_2 + \beta_4 + \beta_5$

β_0 , birinci gün boyunca, operatör A tarafından çalıştırılan birinci makine için ortalama yanıttır. Diğer β ’ları ifade etmek biraz daha zordur. Her makine, tüm operatörler tarafından her gün kullanıldığından, iki makine karşılaştırıldığı zaman, β ’lar satır ve sütunlardan arındırılacaktır. Makine I ve makine II için tablo 3.2’den beklenen değerlerinin ortalamaları alınır

$$\begin{aligned} \text{makine I ortalaması} &= \frac{1}{3} [\beta_0 + (\beta_0 + \beta_2 + \beta_3) + (\beta_0 + \beta_1 + \beta_4)] \\ &= \beta_0 + \frac{1}{3} (\beta_1 + \beta_2 + \beta_3 + \beta_4) \end{aligned}$$

EK-3 (devam) Kukla değişkenler

$$\begin{aligned}
 \text{makine II ortalaması} &= \frac{1}{3}[(\beta_0 + \beta_1 + \beta_5) + (\beta_0 + \beta_3 + \beta_5) + (\beta_0 + \beta_2 + \beta_4 + \beta_5)] \\
 &= \beta_0 + \beta_5 + \frac{1}{3}(\beta_1 + \beta_2 + \beta_3 + \beta_4)
 \end{aligned}$$

İki makine ortalaması arasındaki fark ,

$$\beta_5 = \mu_{II} - \mu_I \text{ 'dir.}$$

benzer olarak makine III ve makine I karşılaştırıldığında,

$$\beta_6 = \mu_{III} - \mu_I \text{ 'dir.}$$

Aynı şekilde,

$$\beta_1 = \mu_B - \mu_A$$

$$\beta_2 = \mu_C - \mu_A$$

$$\beta_3 = \mu_2 - \mu_1$$

$$\beta_4 = \mu_3 - \mu_1$$

olarak ifade edilebilir. Böylelikle β 'lar tanımlanmıştır.

Burada verilen modelde, sayısal olmayan her bir bağımsız değişken için yalnızca ana etki terimleri (yani yalnızca bir x içeren terimler) yer almıştır. Ancak, etkileşim terimleri içeren modeller de yazmak mümkündür [15].

EK-4 Tek yönlü ANOVA için yapılan uygulamada kullanılan veri

Çizelge 4.1. Veri

	Örnek1	Örnek2	Örnek3	Örnek4	Örnek5
1	93,18	99,32	96,79	93,38	96,25
2	94,97	97,30	99,16	94,98	97,02
3	94,49	95,49	98,26	94,55	96,81
4	94,31	96,37	97,48	94,39	96,73
5	91,25	98,34	98,45	91,65	95,42
6	91,43	98,18	99,25	91,82	95,49
7	94,40	93,18	97,17	94,47	96,77
8	98,40	99,24	95,52	98,04	98,49
9	94,67	95,84	96,76	94,71	96,89
10	94,40	97,27	98,38	94,47	96,77
11	94,35	96,41	96,88	94,43	96,75
12	92,67	94,32	95,63	92,93	96,03
13	97,98	97,91	96,89	97,66	98,31
14	93,26	95,79	97,71	93,45	96,28
15	94,13	95,11	95,97	94,23	96,66
16	91,39	93,40	96,84	91,78	95,48
17	91,93	97,75	97,95	92,26	95,71
18	97,13	95,89	97,19	96,90	97,94
19		95,23	97,77	96,40	97,70
20		95,17	98,08	96,12	97,56
21		93,31	94,36	95,08	97,06
22		95,79	97,02	93,44	96,28
23		93,44	97,73	97,83	98,38
24		93,82	97,59	94,56	
25		97,13	98,48	95,02	
26		97,66	95,61	97,55	
27		95,65	97,26	94,66	
28		97,14	97,79	94,47	
29		99,20	97,93	96,32	
30		96,66	99,10	93,26	
31		97,71	96,09		
32		94,43	98,11		
33		94,90	97,72		
34		94,91	97,50		
35		96,18	99,24		
36		96,95	97,29		
37		94,54	96,90		
38		99,29	95,68		
39		95,43	97,84		
40		94,52	93,84		
41		97,22	97,93		
42		96,92	98,02		
43		94,95	98,12		
44		98,22	96,55		
45		96,25	97,19		
46			96,07		

EK-4 (Devam) Tek yönlü ANOVA için yapılan uygulamada kullanılan veri

Çizelge.4.1. (Devam) Veri

	Örnek1	Örnek2	Örnek3	Örnek4	Örnek5
47			95,95		
48			95,36		
49			96,55		
50			96,85		
51			97,62		
52			98,31		
53			98,67		
54			96,91		
55			96,08		
56			98,35		
57			98,57		
58			95,72		
59			98,26		
60			97,33		
$\tilde{Y}_i$	94,1300	96,2131	97,2476	94,8292	96,1160

EK-5 Simülasyon programı

```
%% dagilimi olusturma
clc
```

```

clear
grup=4;% grup sayısını belirtir.
muort=10
mu(1:grup)=muort;
for i=1:grup
    var(i)=1+0*i;%her gruptaki varyansı belirtir.
end
N=2000;%tekrar sayısını belirtir.
n1=4;%secilecekörnek sayısını verir.
    :
X=[1 0 0 0; 1 1 0 0; 1 0 1 0; 1 0 0 1];
for z=1:N
    xk(1:n1)=normrnd(mu(1),var(1),1,n1);
    y1(1:n1)=xk(1:n1);
    :
%% karşılaştırma
gf=0;% gruplar arası farkı belirtir.
mu(2)=muort-0.5*gf;
mu(3)=muort+0.5*gf;
mu(1)=muort-1.5*gf;
mu(4)=muort+1.5*gf;
N=2000;
n(1)=n1;
    :
S(4)=sqrt(SS(4));
Sk=sqrt(SSk);
for i=1:4
    U(i)=1/n(i)+1/n(i)*sqrt((Sk/S(i)-1)/(n(i)-1));
    V(i)=1/n(i)-1/n(i)*sqrt((Sk/S(i)-1)*(n(i)-1));
end
W1(1:n1-1)=U(1);
    :
power_F
power_klasikF
%%

```

ÖZGEÇMİŞ

Kişisel Bilgiler

Soyadı, adı : YILMAZ, Filiz
 Uyuğu : T.C.
 Doğum tarihi ve yeri : 21.07.1983 Ankara

Medeni hali : Bekar
Telefon : 0 (312) 255 81 12
e-mail : flzyilmaz83@hotmail.com.

Eğitim

Derece	Eğitim Birimi	Mezuniyet tarihi
Yüksek lisans	Gazi Üniversitesi /İstatistik Bölümü	2007
Lisans	Gazi Üniversitesi/ İstatistik Bölümü	2005
Lise	Batıkent Lisesi (YDA)	2001

Yabancı Dil

İngilizce, Hollandaca

Hobiler

Seyahat, Yabancı diller, İnternet