ÖZET

Yüksek Lisans Tezi

TESADÜFİ BLOKLAR DÜZENİNDE KOVARYANS ANALİZİ VE UYGULAMASI

Tülin CANTAY
Selçuk Üniversitesi
Fen Bilimleri Enstitüsü
İstatistik Ana Bilim Dalı

Danışman: Yrd. Doç. Dr. Aşır GENÇ

2005, 83 Sayfa

Jüri: Yrd. Doç. Dr. Aşır GENÇ
Yrd. Doç. Dr. M. Fedai KAYA
Yrd. Doç. Dr.Abdurrahman TOZLUCA

Kovaryans analizi; varyans analizi ile regresyon analizinin bir kombinasyonu olup, deneysel çalışma sonuçlarını açıklamada büyük avantajlar ve kolaylıklar sağlamaktadır.Model, yarı deneysel çalışmalarda istatistiksel kontrolü yapmak için geliştirilmiş önemli bir yöntemdir.Modelin güçlü varsayımları olması nedeniyle sonuçlarda ki duyarlılık daha fazladır.

Anahtar Kelimeler: Kovaryans, Kovaryans Analizi, Tesadüfi Bloklar

ABSTRACT

MS Thesis

COVARIANCE ANALYSIS ON RANDOM BLOCKS DESIGN and APPLICATION

Tülin CANTAY
Selçuk University
Graduate School of Natural and Applied Sciences
Department of Statistics

Supervisor: Assist. Prof. Dr. Aşır GENÇ

2006, 83 Page

Jury: Assist. Prof. Dr. Aşır GENÇ
Assist. Prof. Dr. M. Fedai KAYA
Assist. Prof.Dr. Abdurrahman TOZLUCA

Covariance Analysis is combination of variance analysis and regression analysis, therefore experimental investigations can be made easly and advantageously. Model is very important process and made for controlling the statistical analysis in the semi experimental investigations. The Model has powerfull hypothesis so that sensitivity of the results is much better.

Key Words: Covariance , Covariance Analysis, Random Blocks

ÖNSÖZ

Tesadüfi bloklar düzeninde kovaryans analizi ve uygulaması adlı tezimin seçiminde ve gerçekleşmesinde yardımını esirgemeyen ve çalışmaya teşvik eden değerli hocam Yrd. Doç. Dr. Aşır GENÇ'e, Araştırma görevlisi Aydın KARAKOCA'ya ve desteklerini hiçbir zaman eksik etmeyen aileme teşekkür ederim.

İÇİNDEKİLER

ÖZET ..... **i**
ABSTRACT ..... **ii**
ÖNSÖZ ..... **iii**
İÇİNDEKİLER ..... **iv**
GİRİŞ **1**

1. REGRESYON ANALİZİ **4**

1.1. Doğrusal Regresyon 4
1.1.1. Regresyon Katsayılarının Kestirimi 6
1.1.2. Açıklanan ve Açıklanamayan Değişim 8
1.1.3. Belirleme Katsayısı 11
1.1.4. Hipotez Testleri ve Güven Aralıkları 11
1.2. Çoklu Doğrusal Regresyon Analizi 17
1.2.1. Çoklu Doğrusal Regresyon Modeli 17
1.2.2. Çoklu Doğrusal Regresyon Modeli Üzerinde Temel Varsayımlar 18
1.2.3. Tahmin Edicilerde Aranan Özellikler 20
1.2.4. Çoklu Doğrusal Regresyon Modellerinde Parametre Tahmini 21
1.2.4.1. En Küçük Kareler Yöntemi 21
1.2.4.2. En Çok Olabilirlik Yöntemi 23
1.2.5. Çoklu Regresyon Analizinde Hipotez Testleri ve Güven Aralıkları 26
1.2.5.1. Regresyon Parametrelerinin Anlamlılık Testi 26
1.2.5.2. Güven Aralıkları 28
1.2.6. Regresyon Analizinde Genel Lineer Hipotezler 30

2. VARYANS ANALİZİ **31**

2.1. Bir Yönlü Varyans Analizi 31
2.1.1. F Değeri 37
2.1.2. Bir Yönlü Varyans Analizinde Hipotez Testi 39
2.1.3. Tek Faktörlü Tesadüfi Bloklar Deney Düzeni İçin Varyans Analizi 41

3. KOVARYANS ANALİZİ 45

3.1. Kovaryans Analizi ile İlgili Varsayımlar 45
3.2. Kovaryans Analizi İçin Doğrusal Model 46
3.2.1. Tahmin 48
3.2.2. Hipotez Testi 51
3.3. Bir Kovariate (Ortak Değişken)'li Tek Yönlü Model 52
3.3.1 Model 52
3.3.2. Tahmin 52
3.3.3. Hipotez Testi 54
3.3.3a Denemeler 54
3.3.3b. Eğim 56
3.3.3c. Eğimlerin Homojenlik Testi 57
3.4. Tesadüfi Bloklar Düzeninde Kovaryans Analizi 61

4. UYGULAMA 67

4.1. Uygulama I 67
4.2. Uygulama II 72

5. SONUÇ 80
KAYNAKLAR 82

GİRİŞ

Kovaryans analizi, iki veya daha fazla grupta, bir bağımlı değişkenin ortalamalarının karşılaştırılması sırasında, söz konusu değişkene etki eden başka sürekli değişkenin etkisinin ortadan kaldırılması veya bu etkinin artırılması maksadıyla kullanılan bir istatistiksel süreçtir.

Kovaryans analizi grup ortalamaları arasındaki farkı ölçerken regresyon analizi ile varyans analizini birleştirir. Kovaryans analizinde varyans analizi sonuçlarında bağımlı değişken ile kodeğişken arasındaki doğrusal ilişki için düzeltme yapılır. Kovaryans analizinin diğer istatistiksel yöntemlerden avantajı, sonuçtaki hata varyansını azaltma ve denekler arasındaki diğer farklılıkları dikkate alarak, grup farklılıklarını ortaya koyabilme kabiliyetidir.

Farklı ölçülebilen değişkenlerden kaynaklanan ve bütün grubu etkileyen bağımlı değişkendeki değişmeyi kontrol altına alarak, hata varyansı azaltılır. Böyle bir değişken varyans analizinde bağımsız veya bağımlı değişken değildir. Ancak değişmeye katkıda bulunur ve gruplar arasındaki farkın büyüklüğünü azaltır. Kovaryans analizinde söz konusu değişkenden kaynaklanan değişim ölçülebilir ve hata varyansından ayırt edilebilir. Hata varyansının azaltılması suretiyle analizin gücü artar.

Kovaryans analizinde bağımsız ve bağımlı değişkenler dışında, araya giren bir veya daha fazla değişkende vardır. Böyle değişkene ‘covariate(kodeğişken)’de denir. Kovaryans analizinde birden fazla bağımsız değişken ve birden fazla covariate(kodeğişken) bulunabilir(Akgül 2005).

Varyans analizinde ölçümle gelen bağımsız değişkenlerin etkisi ortaya konulamamaktadır. Kovaryans analizinde ölçümle gelen ve ortak değişken olarak olan isimlendirilen değişkenin etkisi ortaya konulabilmekte ve böylelikle diğer bağımsız değişkenlerin etkiside daha net olarak ortaya konulabilmektedir.Buna

karşın varyans analizinin kullanımı, kovaryans analizinin kullanımına kıyasla daha yaygındır.

Bu çalışmadaki amaç; tesadüfi bloklar düzeninde kovaryans analizinin temel varsayımları, teorik yapısı anlatıldıktan sonra, tesadüfi bloklar düzeninde kovaryans analizinin varyans analizine kıyasla deney hatasını azalttığını ve çalışmanın güvenilirliğini arttırdığını, verilen iki örnekle ortaya koyabilmektir.

Çalışmanın birinci bölümünde regresyon analizi, ikinci bölümünde varyans analizi ile ilgili temel bilgiler verilmiştir. Üçüncü bölümde kovaryans analizinin temel varsayımları ve teorik yapısından bahsettikten sonra, tek kovariate(ortak değişken)'li kovaryans analizi ile tesadüfi bloklar düzeninde kovaryans analizi hakkında bilgi verilmiştir. Çalışmanın son bölümünde ise konu ile ilgili iki uygulama yapılmıştır. Her iki uygulamada da Selçuk Üniversitesi İstatistik Paket Programı Selçuk Stat ve SPSS 13.0 programları kullanılmıştır.

Birinci uygulamada; Türkiye tarımında önemli yeri olan fındık ürünü üzerinde kovaryans analizi yapılmıştır. Yıllık sıcaklık ortalamasının kabuklu fındığın iç randımanı üzerindeki etkisinin incelenmesi amacıyla böyle bir çalışma yapılmıştır.Bu çalışmada amaç; kabuklu fındığın iç randımanının bölgelere göre farklılık gösterip göstermediğini incelemek eğer gösteriyorsa bu farklılığa etkisi olduğu düşünülen, yıllık sıcaklık ortalaması açısından, tesadüfi bloklar düzeninde bir kovaryans analizi yapmaktır.

İkinci uygulamada; MI (miyokard enfarktüsü) geçirme riskinin bir belirteci olan, damarlarda meydana gelen bozukluğun ifadesi, CFR (Coronary Flow Rezerve) değeri için kovaryans analizi yapılmıştır. Bu uygulamada amaç; normal sigara içen, light sigara içen ve sigara içmeyen gruplarda; light sigara içimi sonrası oluşan CFR değerleri ile başlangıç (bazal) CFR değerleri arasında bir fark olup olmadığını, tek kovariate(ortak değişken)'li kovaryans analizi ile incelemektir.

LİTERATÜR TARAMASI

Fisher, R.A. 1932. Kovaryans analizini ilk kez 1932 yılında tarımsal araştırmalarda uygulamıştır. Sonradan birçok biyolojik araştırmalarda bu teknikten yararlanılmıştır.

Steel, M. ve Torrie, D. 1960. İki bağımsız değişkenin söz konusu olduğu tek faktörlü tesadüfi bloklar deney düzeni üzerine bir çalışma yaptı. Bu çalışmadaki deney 11 soya çeşidi üzerinde dört tekrarlı olarak yapıldı ve bitkilerin olgunlaşmasındaki gecikmenin ve toprak sathına yayılmasının, kansere yakalanma ile bir ilşkisi olup olmadığının araştırılması ve 11 soya çeşidi arasında kansere yakalanma yönünden farklar bulunup bulunmadığının ortaya çıkarılması amaçlandı. Eğer, olgunlaşmadaki gecikmenin veya toprak sathına yayılmanın ya da her ikisinin birden bitkilerin kansere yakalanması yönünden etkileri varsa kovaryans analizi ile çeşitlere ait ortalamaların düzeltilmesi düşünülmüştür.

Leclerc, A. 1962. Tesadüfi parseller deney düzeninde bir sera denemesi uygulamıştır. Üç buğday çeşidine ait bitkilerin kök ağırlıklarını karşılaştırma amacına yönelik bu araştırmada, çimlenmedeki bazı aksaklıklardan dolayı saksılardaki bitki sayıları farklı alınarak, bu farklılıkların bitkilerin kök gelişmesini etkilemesi beklendiği gibi kovaryans analizi vasıtasıyla üç çeşide ait kök ağırlıklarının bitki adedine göre düzeltilmesi yapılmıştır.

Snedecor, T. 1965. Latin kareler deney düzenine kovaryans analizini uygulamıştır. Dört soya çeşidinin verim yönünden karşılştırılmasının amaçladığı bu araştırmada, bitkilerdeki kanser enfeksiyonu yüzdesinin verim üzerinde etki yapacağı düşünülerek, kovaryans analiziyle dört soya çeşidine ait verim ortalamalarının, bitkilerdeki kanser enfeksiyonu yüzdesine göre düzeltilmesi yapılmıştır.

Demirsoy, Ş. 1993. Tarafından, et ve yapağı yönlü yeni bir tip geliştirmek amacıyla yapılan melezleme ile elde edilen Lincoln x Türk Merinosu; Hampshire x Türk Merinosu; Alman Siyah Baş x Türk Merinosu kuzularında büyüme dönemindeki canlı ağırlıklarını etkileyen ırk, cinsiyet, doğum tipi ve ana yaşı faktörlerinin kovaryans analizi ile incelemesini yapmıştır.

1. REGRESYON ANALİZİ

Biri bağımsız değişken, diğeri bağımlı değişken kabul edilen iki değişken arasındaki ilişkinin matematiksel bir fonksiyonla ifade edilebilmesi için yapılan araştırmalar regresyon analizine konu teşkil eder. Değişken sayısı ikiden fazladır. Böyle durumlarda değişkenlerden biri bağımlı, diğerleri bağımsız değişken durumunda olacaktır.

Bağımlı değişken ile bir ya da bir kaç değişken arasında kurulan modeldeki parametreleri tahmin ederek, bağımsız değişkenlerin belirlenen değerleri için bağımlı değişkenin alacağı değeri tahmin etmeye Regresyon Problemi denir(Çömlekçi 1999).

1.1. Doğrusal Regresyon

Burada bir bağımlı değişken ve bir bağımsız değişken varken bu iki değişken arasındaki ilişkiyi belirleyecek modelin nasıl olduğunun ortaya çıkartılması açıklanmaya çalışılacaktır. n tane birimin her birinden bağımsız değişken X ve bağımlı değişken Y değerleri saptanmış olsun. Bu durumda $(X_1,Y_1),(X_2,Y_2)...(X_n,Y_n)$ olmak üzere n tane gözlem değeri olacaktır. Acaba X ve Y değişkenleri arasındaki ilişki nasıldır? Bu ilişkiyi matematiksel eşitlik olarak ifade edilebilir mi? Bu soruların yanıtını verebilmek için (X_iY_i), $i=1,2,...,n$ gözlem çiftlerini koordinat eksenlerine işaretlemek gerekir. Buna regresyonda serpme diyagramının hazırlanması denir. n tane gözlem çiftinin her biri için serpme diyagramında kesişim noktaları bulunduğunda n tane nokta oluşacaktır. Bu noktaların konumuna bakılarak modelin nasıl olduğuna karar verilir. Eğer noktalar bir doğru etrafında toplanıyor ise doğrusal bir model kullanılmalıdır.

Noktalar bir doğru etrafında toplanmışsa yada bu duruma yakınlık gösteriyorsa seçilecek regresyon modeli X bağımsız, Y bağımlı değişken olmak üzere

$$Y=\beta_0 + \beta_1 X + \varepsilon \qquad (1.1)$$

olacaktır. Bu modele "**Basit Doğrusal Regresyon Modeli**" denir.

Eşitlikteki ε, deneysel hata *Y*'deki değişkenliğin doğrusal ilişki ile açıklanamayan kısmı olup *0* ortalama σ^2 varyans ile normal dağılıma uyan rasgele değişkendir ve sabit bir değer olan *X*'in değerine bağlı değildir. β_0 ve β_1 regresyon katsayıları olarak adlandırılır ve tahmin edilmeleri gerekir.

Basit doğrusal regresyon modeli şu varsayımlara dayanır:

1. *Açıklayıcı değişken değerleri sabit sayılardır.*

2. *Hata teriminin ortalaması sıfırdır.*

3. *Hata teriminin varyansı σ^2 dir ve bu varyans X değerlerine göre değişmez*

4. *Hata teriminin değerleri birbirleriyle ilişkili değildir.*

5. *Hata terimi, ortalaması ve varyansı (2) ve (3) nolu varsayımlarda belirlenmiş olan bir normal değişkendir* (Myers 1990).

Kitleden seçilen n birimlik örneklem için doğrusal regresyon denklemi;

$$Y = b_0 + b_1 X_i + e, \qquad (i = 1,2,...,n) \qquad (1.2)$$

biçiminde tanımlanır. Bilinen bir X_i değeri için, Y_i değeri tahmin edilir. Tahmini doğrusal regresyon denklemi ise;

$$\hat{Y}_i = b_0 + b_1 X_i \qquad (i=1,2,...,n) \qquad (1.3)$$

biçimindedir. (1.2) ve (1.3) eşitliklerinde;

Y_i = i ‘inci gözleme ilişkin gerçek Y değeri

$\hat{Y}_i$ = i ‘inci gözleme ilişkin Y ‘nin tahmini değeri

X_i = i ‘inci gözleme ilişkin bağımsız değişkenin alacağı değer

b_0= Regresyon doğrusunun y eksenini kestiği noktayı gösteren kesim noktasıdır. β_1’in tahminidir.

b_1=Regresyon katsayısıdır. Doğrunun eğimini gösterir.Bağımsız değişkendeki bir birimlik değişmenin bağımlı değişkende yapacağı değişikliği gösterir. β_1’in tahminidir.

e_i = i’inci gözlemin hata terimidir. Gözlenen değer ile tahmin değer arasındaki farktır.

Yani $e_i = Y_i - \hat{Y}_i$ dir. Hata terimleri ortalaması sıfır, varyansı σ^2 olan normal dağılıma sahiptir. $e \sim N(0,\sigma^2)$.

1.1.1. Regresyon Katsayılarının Kestirimi

Rassal olarak seçilen n tane gözlem biriminde (X_i,Y_i) $i=1,2,...,n$ değerleri saptanarak serpme diyagramı hazırlanmış olsun. Serpme diyagramında oluşan n tane noktanın bir doğrunun etrafında toplandığını kabul edelim. E.K.K. bu doğrunun belirlenmesini sağlar. $\hat{Y} = b_0 + b_1X$ olarak tanımlanan bu doğru aşağıdaki iki koşulu sağlamalıdır.

1. Her (X_i,Y_i) çiftine karşı gelen nokta ile bu noktanın E.K.K. ile elde edilecek doğru üzerindeki dik izdüşümleri arasındaki farkların toplamı sıfır olmalıdır. Yani, $e_i = Y_i - \hat{Y}_i$ ile gösterilen hataların toplamının sıfır olması istenmektedir.

2. Bu farkların kareleri toplamı minimum olmalıdır.

$Y=b_0 + b_1X$ doğrusu elde edildiğinde;

$$\sum_{i=1}^{n}(Y_i - \hat{Y}_i) = 0 \qquad \text{ve} \qquad \sum_{i=1}^{n}(Y_i - \hat{Y}_i^2) \text{ minumum olmalıdır.}$$

$\hat{Y}_i$ değerini ikinci koşulda yerine koyalım.

$$\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i)^2 = 0$$

$$S = \sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i)^2 = 0$$

olsun.

Yukarıda sözü edilen doğrunun birinci koşulu sağlaması için S olarak tanımlanan terimin b_0 ve b_1 istatistiklerine göre kısmi türevlerinin alınıp sıfıra eşitlenmesi gerekir. İkinci türevlerde pozitif ise ikinci koşul sağlanmış olur.

$$\frac{\partial S}{\partial b_0} = -2\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i)$$

$$\frac{\partial S}{\partial b_1} = -2\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i)X_i$$

Bu iki eşitliği sıfıra eşitleyip gerekli işlemler yapılırsa aşağıdaki eşitlikler elde edilir.

$$\sum_{i=1}^{n} Y_i = nb_0 + b_1 \sum_{i=1}^{n} X_i$$

$$\sum_{i=1}^{n} Y_i X_i = b_0 \sum_{i=1}^{n} X_i + b_1 \sum_{i=1}^{n} X_i^2$$

Bunlara "normal eşitlikler" denir. İki bilinmeyenli bu iki eşitlik çözülürse b_0 ve b_1 istatistikleri

$$b_1 = \frac{\sum(Y_i - \overline{Y})(Xi - \overline{X})}{\sum(X_i - \overline{X})^2} \tag{1.4}$$

$$b_0 = \overline{Y} - b_1 \overline{X} \tag{1.5}$$

elde edilir.

1.1.2. Açıklanan ve Açıklanamayan Değişim

1. Örneklem Ortalaması Etrafındaki Değerlerin Değişimi

Bu değişim $\sum(Y_i - \overline{Y})^2$ ile verilir ve toplam değişim yada genel kareler toplamı olarak bilinir ve aksi belirtilmedikçe burada *YOAKT* olarak ifade edilecektir. Bu değişim, örneklem standart sapması ve varyansı tanımı üzerine oturtulmuştur.

2. Regresyon Doğrusu Etrafındaki Değerlerin Değişimi

Bu değişim $\sum(Y_i - \hat{Y}_i)^2$ ile verilir ve açıklanamayan değişim ya da hata kareler toplamı olarak bilinir. Bu bölümde açıklanamayan değişim, regresyondan ayrılış kareler toplamı *RAKT* olarak ifade edilecektir. Eğer iki değişken arasındaki korelasyon ± 1.00 ise tüm değerler regresyon doğrusu üzerine düşecektir. Diğer bir deyişle, ideal (tam) bir ilişki açıklanamayan değişim olmayacaktır. Fakat, gerçekte korelasyon ideallikten daha azdır. Dolayısıyla değerlerin bir çoğu regresyon doğrusu üzerine düşmeyecektir. Regresyon doğrusunun sapması, değişimi verecektir ki sapmalar küçük olduğu oranda, bağımlı değişkende ki açıklanamayan değişimde küçük olacaktır.

3. Ortalama Etrafındaki Tahmini Değerlerin Değişimi

Bu değişim $\sum(\hat{Y}_i - \overline{Y})^2$ ile verilir ve açıklanan değişim yada regresyon kareler toplamı RKT olarak bilinir, İki bileşenden oluşan değişim, matematiksel olarak eşitlik (1.6) ile tanımlanır.

Toplam Değişim = Açıklanamayan Değişim + Açıklanan Değişim

$$\sum(Y_i - \overline{Y})^2 = \sum(Y_i - \hat{Y}_i)^2 + \sum(\hat{Y}_i - \overline{Y})^2 \qquad (1.6)$$

Bu bileşenlerin açık ifadeleri aşağıdaki gibi verilebilir.

$$YOAKT = \sum(Y_i - \overline{Y})^2 = \sum Y_i^2 - \frac{(\sum Y_i)^2}{n} \qquad (1.7)$$

YOAKT: Toplam değişim ya da genel kareler toplamı

$$RKT = \sum(\hat{Y}_i - \overline{Y})^2 = \frac{\left[\sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n}\right]^2}{\sum X_i^2 - \frac{(\sum X_i)^2}{n}}$$

$$RKT = \frac{(XYOAÇT)^2}{XOAKT}$$

$$RKT = b_1 \, XYOAÇT \qquad (1.8)$$

RKT: Açıklanan değişim ya da regresyon kareler toplamı

Regresyon serbestlik derecesi, (katsayıların sayısı -1)'dir. Basit doğrusal regresyon için serbestlik derecesi,

RSD = 2 – 1 = 1 olacaktır.

Genel serbestlik derecesi,

*YOASD=n-1'*dir. (1.9)

YOASD: Genel serbestlik derecesi

Eşitlik serbestlik dereceleri için düşünüldüğünde, regresyondan ayrılış serbestlik derecesi yada hata kareler toplamının serbestlik derecesi,

RASD = YOASD – RSD = (n-l) -1 = n-2 (1.10)

RASD: Regresyondan ayrılış serbestlik derecesi ya da hata kareler toplamının serbestlik derecesi

RAKT'nin *RASD*'ne bölünmesi ile elde edilen regresyondan ayrılış kareler ortalaması;

$$RAKO = \frac{RAKT}{RASD} \tag{1.11}$$

olarak bulunur. *RAKO* yada hata kareler ortalaması, bir başka deyişle hataların varyansı, σ^2 nin tahmini değeridir.

RAKO: Regresyondan ayrılış kareler ortalaması

RAKT: Açıklanamayan değişim ya da regresyondan ayrılış kareler toplamı

RASD: Regresyondan ayrılış serbestlik derecesi ya da hata kareler toplamının serbestlik derecesi

RKT'nin *RSD*'ye bölünmesi ile regresyon kareler ortalaması elde edilir ve

$$RKO = \frac{RKT}{RSD} \tag{1.12}$$

olarak bulunur.

RKO: Regresyon kareler ortalaması

1.1.3. Belirleme Katsayısı

Açıklanan değişimin toplam değişime oranıdır. Diğer bir deyişle, bağımsız değişkenin, bağımlı değişkende ki değişimin yüzde kaçını açıkladığını gösterir. R^2 notasyonu ile verilen belirleme katsayısı (determinasyon katsayısı) şu şekilde tanımlanır $\hat{Y} = b_0 + b_1 X$ modeli için belirleme katsayısı,

R^2 = *(Açıklanan değişim) / (Toplam değişim)*

$$R^2 = \frac{RKT}{YOAKT} = \frac{\sum(\hat{Y} - \overline{Y})^2}{\sum(Y_i - \overline{Y})^2} \tag{1.13}$$

olarak tanımlanır. Toplam değişimin açıklanamayan kısmı ise,

$$1\text{-}R^2 = \frac{RAKT}{YOAKT}$$

ile verilir ve belirlenememe katsayısı adını alır.

Belirleme katsayısı regresyonda önemli bir kavramdır. R^2 değerinin bire yakın olması her araştırmacının arzuladığı bir durumdur. Bu durumda bağımlı değişkendeki varyansın büyük bir kısmının modeldeki bağımsız değişkenin açıkladığı sonucuna varılır.

1.1.4. Hipotez Testleri ve Güven Aralıkları

Basit doğrusal regresyonda gerek güven aralıklarının bulunmasında gerekse hipotez testlerinde tahmin edicisi bulunan; $Y = \beta_0 + \beta_1 X$ modeli için aşağıdaki varsayımlar yapılmalıdır.

1. Bağımsız değişken olan X değerleri sabit olup hatasız ölçülmüşlerdir.

2. X'in verilen herhangi bir değeri için Y bağımlı değişkeni ortalaması

$\mu = \beta_0 + \beta_1 X$ ve varyansı σ^2 olan normal dağılıma uyar.

3. $\sigma_\varepsilon^2 = \sigma^2$ şeklinde tanımlanan varyans her bir X değeri için aynıdır ve σ^2'nin tahmin edicisi

$$S^2 = \frac{HKT}{n-2}$$

eşitliğiyle gösterilir .

Basit doğrusal regresyon modelinin yeterliliğini belirlemek için hipotez testine ve güven aralıklarının oluşturulmasına gerek vardır. Regresyon modelinin yeterliliğini ve eğimi üzerine kurulacak hipotez testlerinde, hata bileşenleri ε_i 'nin normal dağılıma sahip olduğu varsayımı yapılmalıdır.

Regresyon analizi ile ilgili hipotezler $\beta = 0$ yani regresyon hattı eğiminin sıfır olduğu şekilde formüle edilirler. Böyle bir hipotez, birisi "*F*" diğeri de "*t*" olmak üzere istatistik vasıtasıyla kontrol edilebilirler.

$\hat{Y} = b_0 - b_1 X$ regresyon modelinde b_0 istatistiği β_0 parametresinin ve b_1 istatistiği de β_1 parametresinin tahmin edicisidir. Bu nedenle b_0 istatistiği β_0 parametresine ilişkin hipotezlerin testlerinde ve güven aralıklarının oluşturulmasında kanıt olarak kullanılır.

β_1 parametresine ilişkin hipotezlerin testlerinde ve güven aralıklarının oluşturulmasında kanıt olarak kullanılacak b_1 istatistiğinin örnekleme bölünümü normal bölünüme yeterince yakındır. b_1 istatistiğinin beklenen değeri ve varyansı da bilinirse diğer işlemler daha önceki test ve güven aralıklarından farklı değildir(Apaydın 1994).

$E(b_1)$: b_1 istatistiğinin beklenen değeri

$V(b_1)$: b_1 istatistiğinin varyansı

$s.h(b_1)$: b_1 istatistiğinin standart hatası.

$E(b_1)$ değeri yokluk hipotezinde belirtilen değere eşittir. $V(b_1)$ değeri elde edilsin. (2.4) eşitliğinde yer alan payın parantezi açıldığında;

$$b_1 = \frac{\sum\left[X_i - \overline{X})Y_i - \overline{Y}(X_i - X)\right]}{\sum(X_i - \overline{X})^2} \tag{1.15}$$

$$b_1 = \frac{\sum(X_i - \overline{X})Y_i}{\sum(X_i - \overline{X})^2}$$

olur. (1.15) eşitliğini aşağıdaki gibi de yazılabilir.

$$b_1 = \frac{\left[(X_1 - \overline{X})Y_1 + (X_2 - \overline{X})Y_2 + ... + (X_n - \overline{X})Y_n\right]}{\sum(X_i - \overline{X})^2}$$

$$D = a_1Y_1 + a_2Y_2 + ... + a_nY_n$$

ise D'nin varyansı şu şekilde tanımlanır.

$$V(D) = a_1^2V(Y1) + a_2^2V(Y2) + ... + a_2^nV(Yn)$$

a ve Y bağımsız ise aşağıdaki eşitlikler yazılabilir.

$$V(D) = (a_1^2 + a_2^2 + ... + a_n^2)\sigma^2$$

$$V(D) = (\sum a_i^2)\sigma^2$$

$$V(Y_i) = \sigma^2$$

$$b_1 = \left(\frac{\sum(X_i - \overline{X})}{\sum(X_i - \overline{X})^2}\right)Y_i$$

Yukarıdaki eşitlikteki parantez içindeki terim a olarak düşünülürse $V(b_1)$ ulaşılır.

$$V(b_1) = \left(\frac{\sum(X_i - \overline{X})}{\sum(X_i - X)^2} \right) \sigma^2$$

$$V(b_1) = \frac{\sigma^2}{\sum(X_i - X)^2}$$

σ^2 bilinmediğinde, yerine S^2 istatistiği kullanılır.

$$V(b_1) = \frac{S^2}{\sum(X_i - X)^2} \tag{1.16}$$

$$S^2 = \frac{\sum(Y_i - \hat{Y}_i)^2}{n-2} \tag{1.17}$$

$$S.h(b_1) = \frac{S}{(\sum(X_i - \overline{X})^2)^{1/2}} \tag{1.18}$$

$$\frac{b_1 - E(b_1)}{s.h.(b_1)} \approx t_{n-2}$$

olduğundan hipotez testlerinde ve güven aralıklarının oluşturulmasında *n-2* serbestlik dereceli *t* bölünümünden yararlanılır β_1 parametresine ilişkin hipotezler aşağıdaki gibi olsun;

$H_0 : \beta_1 = \beta_{1.0}$

$H_1 : \beta_1 \neq \beta_{1.0}$ $\qquad -\infty \leq \beta_{1.0} \leq \infty$

Yokluk hipotezinin test edilebilmesi için b_1 istatistiği gereklidir. b_1 istatistiğinin örnekleme bölünümü normal olup parametreleri şu şekildedir;

$E(b_1) = \beta_{1.0}$

$$S.h(b_1) = \frac{S}{(\sum(X_i - \overline{X})^2)^{1/2}}$$

n-2 serbestlik dereceli t bölünümü test istatistiği olduğundan

$$V(b_0) = V(\overline{Y}) + \overline{X}^2 V(b_1)$$

$$V(b_0) = \frac{S^2}{n}.\overline{X}^2.\frac{S^2}{\sum(X_i - \overline{X})^2}$$

olarak yazılabilir. Gerekli kısaltmalar yapılırsa $V(b_0)$ ve $s.h.(b_0)$ değerleri elde edilir.

$$V(b_0) = S^2\left(\frac{\sum X_i^2}{n\sum(X_i - \overline{X})^2}\right)$$

$$s.h.(b_0) = \left(\frac{\sum X_i^2}{n\sum(X_i - \overline{X})^2}\right)^{1/2}.S$$

bu durumda;

$$\frac{b_0 - E(b_0)}{s.h.(b_0)} \approx t_{n-2}$$

olduğundan güven aralıklarının oluşturulmasında ve hipotezlerin testlerinde *n-2* serbestlik dereceli t-bölünümünden yararlanılan yokluk ve karşıt hipotezler,

$H_0 : \beta_0 = \beta_{0.0}$

$H_1 : \beta_0 \neq \beta_{0.0}$ $\qquad -\infty \leq \beta_{1.0} \leq \infty$

iken H_0 hipotezinin test için gerekli işlemler pek farklı değildir.

$E(b_0) = \beta_{0.0}$

$$s.h.(b_0) = S\left(\frac{\sum X_i^{\,2}}{n\sum (X_i - \overline{X})^2}\right)^{1/2}$$

t_{n-2} değeri hesaplanarak karar verilmesi şu şekilde olacaktır.

$$t_{n-2} = \frac{b0 - E(b_0)}{s.h.(b_0)}$$

$t_{n-2} > {}_{1-a/2}\, t_{n-2}$ ya da $t_{n-2} < {}_{-1-a/2} t_{n-2}$ ise H_0 red edilir.

H_1: $\beta_0 > \beta_{0.0}$ ise $t_{n-2} > {}_{1-a/2}\, t_{n-2}$ olması durumunda H_0 red edilir.

H_1: $\beta_0 > \beta_{0.0}$ ise $t_{n-2} > {}_{-1-a/2}\, t_{n-2}$ olduğundan H_0 hipotezi red edilir.

1.2. ÇOKLU DOĞRUSAL REGRESYON ANALİZİ

1.2.1. Çoklu Doğrusal Regresyon Modeli

Regresyon analizi istatistik uygulamalarda yaygın olarak kullanılan araçlardan biridir. Regresyon analizi bağımlı bir değişken ($\underline{Y}$) ile bir veya daha fazla açıklayıcı değişken ($\underline{X}_1, \underline{X}_{2,} \cdots, \underline{X}_n$) arasındaki ilişkiyi açıklama ve değerlendirme ile ilgilidir.

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i \qquad (1.19)$$

Şeklinde en genel haliyle verilen çoklu doğrusal regresyon modeli için matris formunda gösterim ise ,

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon} \qquad (1.20)$$

şeklindedir. (1.20) ile belirtilen gösterim için,

$\underline{Y}$: n x 1 boyutlu rasgele değişkenlerin gözlenebilir vektörü

X : n x p boyutlu açıklayıcı değişkenlerin gözlem değerlerini içeren matris

$\underline{\beta}$: bilinmeyen parametreler vektörü

$\underline{\varepsilon}$: nx1 aynı dağılıma sahip $E(\underline{\varepsilon}) = 0$ ve $Cov(\underline{\varepsilon}) = \sigma^2 I$ koşulunu sağlayan ve gözlenemeyen bağımsız rasgele değişkenlerin vektörüdür.

(1.19) veya (1.20) formunda verilen Çoklu Doğrusal Regresyon (ÇDR) modelleri için bilinmeyen parametrelerin tahmini için En Küçük Kareler (EKK) ve En Çok Olabilirlik (ML) yöntemleri verilecektir. Ancak bu yöntemlerden önce ÇDR modelleri üzerinde geçerli temel varsayımların ve tahmin edicilerin genel özelliklerinin bilinmesi gereklidir.

1.2.2 Çoklu Doğrusal Regresyon Modeli Üzerinde Temel Varsayımlar

Çoklu doğrusal regresyonda elde edilen sonuçların geçerliliği temel varsayımların geçerliliğine bağlıdır. Buna göre Çoklu doğrusal regresyon modellerinde temel varsayımlar aşağıdaki gibidir.

Varsayım 1 : *Hata terimleri birbirinden bağımsızdır.*

Bu varsayıma göre hata teriminin herhangi bir değeri başka bir değerinden bağımsızdır. Yani, herhangi bir ε_i teriminin başka bir ε_j ile ortak varyansları sıfıra eşittir yani $\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0$ ' dir. Bu varsayım sağlanmıyorsa hata terimleri arasında otokorelasyon (ardışık bağımlılık) vardır. Hata terimlerinin gözlenen değerlere karşılık grafiği çizilerek otokorelasyonun varlığı incelenebilir veya Durbin-Watson test istatistiği ile test edilebilir.

Varsayım 2 : *Hata terimlerinin beklenen değeri sıfır (*$\mathrm{E}[\varepsilon_i] = 0$*)' dır.*

Bağımlı değişken için $\varepsilon_i = y_i - \hat{y}_i$ hata terimleri bazen negatif bazen pozitif değer almalarına karşın ortalamaları sıfırdır. Varsayımın sağlanmadığı durumlarda parametre tahminleri gerçek değerlerinden daha büyük veya daha küçük değerler alırlar. Bu durumda yanlı parametre tahminleri elde edilir.

Varsayım 3 : *Hata terimleri sabit varyanslıdır. (*$\mathrm{Var}(\varepsilon_i) = \sigma^2 \mathrm{I}$*). i=1,2,...,n*

Doğrusal regresyon modelinin önemli varsayımlarından biri olan sabit varyanslılık varsayımına göre hata terimi varyansı bağımsız değişkendeki değişmelere bağlı olarak değişmeyip aynı kalır ve

$$\mathrm{Var}(\varepsilon_i) = \mathrm{E}[\varepsilon_i - \mathrm{E}[\varepsilon_i]]^2 = \mathrm{E}[\varepsilon_i^2] = \sigma_\varepsilon^2 \qquad (1.21)$$

şeklinde ifade edilir.

Varsayım 4 : *Hata terimleri normal dağılıma sahiptir.*

Çoklu doğrusal regresyon modellerinde hata terimlerinin normal dağılıma sahip olması kendi sıfır ortalamaları etrafında simetrik bir dağılım göstermeleridir. Hata terimlerinin küçük değerler alma olasılığı yüksek, büyük değerler alması olasılığı küçük olacaktır. Bu varsayımın sağlanması parametre tahminlerinden yararlanarak parametre tahminleri için güven aralıkları oluşturmak ve gerekli hipotezleri test etmeye imkan sağlar.

Varsayım 5 : X *matrisinin sütunları birbirinden bağımsızdır Bağımsız değişkenler arasında çoklu doğrusal bağlantı yoktur.*

Bu varsayıma göre çoklu doğrusal regresyon modellerinde bağımsız değişkenler arasında doğrusal veya doğrusala yakın bir ilişki yoktur. Bu varsayım sağlanmaması durumunda $(X'X)^{-1}$ matrisinin tersi alınamayacağından, parametre tahminleri yapılamayacak, tersinin alınabildiği durumlarda ise parametre tahminlerinin varyansları büyük olacaktır. Bunun sonucu olarak parametre tahminleri tutarsız olacaktır.

Varsayım 6: *Gözlem sayısı(n), açıklayıcı değişken sayısı (k)' ndan büyüktür.*

Çoklu doğrusal regresyon modellerinde parametre tahminlerinin yapılabilmesi için X veri matrisinin tam sütun ranklı olması gerekir bu yüzden oluşturulacak model için gerekli gözlem sayısı , modeldeki bağımsız değişken sayısından fazla olmalıdır.

Varsayım 7 : *Model spesifikasyonu doğrudur.*

Modelde tanımlama hatası yoktur. Fonksiyonel biçim doğru seçilmiştir. Model kurulurken incelenen bağımlı değişkeni açıklayan önemli değişkenlerin modele alındığı, modelin matematiksel kalıbının ve denklem sayısının doğru belirlendiği varsayılmaktadır.

1.2.3 Tahmin Edicilerde Aranan Özellikler

Dağılımını bilinen ancak, parametresi bilinmeyen bir kitleden alınan örneklem yardımıyla kitlenin bilinmeyen parametrelerinin tahmin edilmesi problemine parametre tahmini problemi denilmektedir.

Parametre tahminleri çeşitli yöntemlerle yapılmaktadır. Bu yöntemler arasında en iyi tahminleri elde eden yöntemi seçebilmek için bir tahminin iyiliğini yada başka yöntemle bulunan tahmine göre daha iyi olduğunu gösterecek bazı kriterlere ihtiyaç vardır. Genellikle yapılan tahminin, anakütle parametresinin gerçek değerine yakın olması ve bu gerçek parametre etrafında dar bir alanda değişmesi istenir. Anakütle parametresine yakınlık, tahminlerin örnekteki dağılımlarının ortalaması ve varyansları ile ölçülür. Tahmin edicilerde aranan başlıca özellikler şunlardır.

Yansızlık : Bir tahmin edicinin yanı (sapması, biası), tahmin edicinin beklenen değeri ile tahmin edilen parametrenin anakütle değeri arasındaki farktır. $\underline{\hat{\beta}}$, $\underline{\beta}$ parametresi için bir tahmin edici olmak üzere,

$$Yan = E\left(\underline{\hat{\beta}}\right) - \underline{\beta} \tag{1.22}$$

dır. Beklenen değeri, anakütle parametresine eşit olan tahmincilere yansız tahminciler denir (Tarı 1999).

Asimptotik Yansızlık : $\underline{\hat{\beta}}$, $\underline{\beta}$ parametresi için bir tahmin edici olmak üzere,

$$\lim_{n \to \infty} E\left(\underline{\hat{\beta}}\right) = \underline{\beta} \tag{1.23}$$

şartını sağlıyorsa $\hat{\underline{\beta}}$, $\underline{\beta}$ parametresi için asimptotik yansız bir tahmin edicidir (Tarı 1999).

Tutarlılık : $\hat{\underline{\beta}}$, $\underline{\beta}$ parametresi için asimptotik yansız bir tahmin edici olmak üzere,

$$\lim_{n \to \infty} E\left(\hat{\underline{\beta}}\right) = \underline{\beta}$$

$$\lim_{n \to \infty} Var\left(\hat{\underline{\beta}}\right) = 0 \qquad (1.24)$$

ise $\hat{\underline{\beta}}$ tutarlı bir tahmin edicidir (Tarı 1999).

Asimptotik Etkinlik : Tutarlı tahmin ediciler arasında en küçük asimptotik varyanslı tahmin edici asimptotik etkin tahmin edicidir.

1.2.4 Çoklu Doğrusal Regresyon Modellerinde Parametre Tahmini

ÇDR modellerinde bilinmeyen model parametrelerinin tahmin edilmesi bilinmeyen parametreler vektörü $\underline{\beta}$'nın bir $\hat{\underline{\beta}}$ tahmin edicisi ile tahmin edilmesidir. Bu amaçla EKK ve ML tahmin edicilerinden yaralanılmaktadır.

1.2.4.1 En küçük kareler yöntemi (EKK)

En küçük kareler yönteminde amaç ;

$$\phi\left(\underline{\beta}\right) = \left(\underline{Y} - X\underline{\beta}\right)' \left(\underline{Y} - X\underline{\beta}\right) \qquad (1.25)$$

karesel formunu $\underline{\beta} \in \Re^p$ parametre kümesi üzerinden minimize eden değerleri bulmaktır.

$$\frac{\partial \phi(\underline{\beta})}{\partial \underline{\beta}} = -2\mathrm{X}'\underline{\mathrm{Y}} + 2\mathrm{X}'\mathrm{X}\underline{\beta} \qquad (1.26)$$

olmak üzere (1.25) ifadesini minimize eden $\underline{\hat{\beta}}$ vektörü $\underline{\beta}$' nın EKK tahmin edicisidir. (1.25) ile ifade edilen denklemin sıfıra eşitlenmesiyle ,

$$\underline{\hat{\beta}} = (X'X)^{-1} X'\underline{Y} \qquad (1.27)$$

şeklinde EKK tahmin edicisi elde edilir.

Gauss-Markov Teoremi : Çoklu doğrusal regresyon modelinin varsayımlarının geçerliliği altında EKK tahmin edicileri, sapmasız, doğrusal tahmin ediciler arasında en küçük varyanslı olanıdır (Graybill 1961).

(1.27) ile verilen $\underline{\beta}$ 'nın EKK tahmin edicisi $\underline{\hat{\beta}}$ hipotez modellerinde aynı zamanda En Çok Olabilirlik tahmin edicisi olmakla birlikte Gauss Markov teoremine göre de $\underline{\beta}$ 'nın en iyi lineer yansız tahmin edicisidir. EKK tahmin edicisi

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

için;

$$
\begin{aligned}
E\left[\underline{\hat{\beta}}\right] &= E\left[(X'X)^{-1}X'\underline{Y}\right] \\
&= (X'X)^{-1}X'E\left[\underline{Y}\right] \\
&= (X'X)^{-1}X'E\left[X\underline{\beta}+\underline{\varepsilon}\right] \\
&= (X'X)^{-1}X'X\underline{\beta}+\underbrace{E\left[\underline{\varepsilon}\right]}_{0} \\
&= \underline{\beta}
\end{aligned}
\tag{1.28}
$$

olduğundan $\underline{\hat{\beta}}$, $\underline{\beta}$ nın yansız bir tahmin edicisidir (Graybill 1961).

1.2.4.2 En çok olabilirlik yöntemi

(1.19) veya (1.20) formunda verilen ÇDR modeli için varsayım 1,2 ve 4 sağlanırsa modele hipotez modeli denir ve,

$$
\underline{Y}_{nx1} = X_{nxp}\underline{\beta}_{px1} + \underline{\varepsilon}_{nx1}, \quad \underline{\varepsilon} \sim N\left(\underline{0}, \sigma^2 I\right) \tag{1.29}
$$

ile gösterilir. Hipotez modelinde $\underline{\beta} \in \Re^p$ ve σ^2 parametrelerini tahmin etmek için en çok olabilirlik yöntemi göz önüne alınsın. Olabilirlik fonksiyonu

$$
\ell(\underline{Y};\underline{\beta},\sigma^2) = \frac{1}{(\sqrt{2\pi})^{n/2}(\sigma^2)^{n/2}} e^{\frac{-1}{2\sigma^2}(\underline{Y}-X\underline{\beta})'(\underline{Y}-X\underline{\beta})} \tag{1.30}
$$

dır. Olabilirlik fonksiyonunun logaritması

$$\frac{\partial \ln \ell}{\partial \underline{\beta}} = -\frac{1}{2\sigma^2}(-2\mathrm{X}'\underline{\mathrm{Y}} + 2\mathrm{X}'\mathrm{X}\underline{\beta})$$

$$\frac{\partial \ln \ell}{\partial \sigma^2} = -\frac{\mathrm{n}}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}((\underline{\mathrm{Y}} - \mathrm{X}\underline{\beta})'(\underline{\mathrm{Y}} - \mathrm{X}\underline{\beta})) \qquad (1.31)$$

olmak üzere (1.30) ifadesinde türevlerin sıfıra eşitlenmesiyle

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

ve

$$\hat{\sigma}^2 = \frac{1}{n}(\underline{Y} - X\underline{\hat{\beta}})'(\underline{Y} - X\underline{\hat{\beta}}) \qquad (1.32)$$

elde edilir.

$\underline{\varepsilon} \sim N(\underline{0}, \sigma^2 I)$ ve $\underline{Y} \sim N(X\underline{\beta}, \sigma^2 I)$olduğundan $\underline{Y}$' nin bir lineer dönüşümü olan $\underline{\hat{\beta}}$ da normal dağılımlıdır. Böylece (1.28) ve

$$\begin{aligned}\mathrm{cov}(\underline{\hat{\beta}}) &= \mathrm{cov}((\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'\underline{\mathrm{Y}}) \\ &= (\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'\,\mathrm{cov}(\underline{\mathrm{Y}})((\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}')' \\ &= (\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'\sigma^2\mathrm{I}(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}' \\ &= \sigma^2(\mathrm{X}'\mathrm{X})^{-1}\end{aligned} \qquad (1.33)$$

olmak üzere $\underline{\hat{\beta}} \sim N(\underline{\beta}, \sigma^2(X'X)^{-1})$ dır. $\hat{\sigma}^2$ tahmin edicisi için,

$$E(\hat{\sigma}^2) = E(\frac{1}{n}(\underline{Y} - X\underline{\hat{\beta}})'(\underline{Y} - X\underline{\hat{\beta}}))$$

$$= \mathrm{E}(\frac{1}{n}(\underline{Y} - X(X'X)^{-1}X'\underline{Y})'(\underline{Y} - X(X'X)^{-1}X'\underline{Y})')$$

$$= \frac{1}{\mathrm{n}}E(\underline{Y}'(I - X(X'X)^{-1}X')\underline{Y}) \qquad (1.34)$$

$$= \frac{1}{\mathrm{n}}\{tr[(I - X(X'X)^{-1}X')\sigma^2 I] + (X\underline{\beta})'(I - X(X'X)^{-1}X')(X\underline{\beta})\}$$

$$= \frac{1}{\mathrm{n}}(n-p)\sigma^2$$

olduğundan $\hat{\sigma}^2$ yansız değildir. σ^2 nin yansız bir tahmin edicisi

$$\mathrm{S}^2 = \frac{(\underline{\mathrm{Y}} - \mathrm{X}\underline{\hat{\beta}})'(\underline{\mathrm{Y}} - \mathrm{X}\underline{\hat{\beta}})}{\mathrm{n} - \mathrm{p}}$$

$$= \frac{\underline{\mathrm{Y}}'(\mathrm{I} - \mathrm{X}(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}')\underline{\mathrm{Y}}}{\mathrm{n} - \mathrm{p}} \qquad (1.35)$$

dır. Böylece

$$\frac{\underline{Y}'(I - X(X'X)^{-1}X')\underline{Y}}{\sigma^2} = \frac{(n-p)S^2}{\sigma^2} \sim \chi^2_{(n-p)} \qquad (1.36)$$

elde edilir.

$\underline{\varepsilon} \sim N(\underline{0}, \sigma^2 I)$ olan basit doğrusal hipotez modeli için $\underline{\beta}$ parametresinin en çok olabilirlik tahmin edicisi olmak üzere

$$\underline{\hat{\beta}} \sim N(\underline{\beta}, \sigma^2(X'X)^{-1}) \qquad (1.37)$$

ve

$$\frac{(n-2)S^2}{\sigma^2} \sim \chi^2_{(n-2)} \qquad (1.38)$$

dir (Graybill 1961).

1.2.5 Çoklu Regresyon Analizinde Hipotez Testleri ve Güven Aralıkları

Çoklu Regresyon denklemi elde edildikten sonra çeşitli hipotezler test edilebilir ve güven aralıkları oluşturulabilir. Bu hipotez testlerinden regresyon parametrelerin topluca anlamlılığının test edildiği F testi ve bireysel parametrelerin anlamlılığının testi olan t testi aşağıdaki gibi yapılmaktadır.

1.2.5.1.Regresyon parametrelerinin anlamlılık testi

a) Tümel F testi

Bağımlı değişkenin bağımsız değişkenler tarafından açıklanıp açıklanmadığı, diğer bir deyişle, bağımlı değişkenle bağımsız değişkenler kümesi arasında doğrusal bir ilişki olup olmadığı test edilir. F dağılımı yardımıyla yapılan bu testte ;

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$$

şeklindeki çoklu doğrusal regresyon modeli için ;

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$$

şeklindeki bağımlı değişken $\underline{Y}$ 'nin bağımsız değişkenler $X_1, \underline{X}_2, \ldots, \underline{X}_p$ tarafından açıklanamadığını belirten sıfır hipotezine alternatif olarak,

H_1: En az bir β_i sıfırdan farklıdır.

hipotezi kurulur. Hipotez F test istatistiği yardımıyla test edilir.

$$F = \frac{RKO}{HKO} \tag{1.39}$$

RKO= (Regresyon Kareler Toplamı) / (p-1)

HKO= (Hata Kareler Toplamı) / (n-p)

$$F = \frac{\left(\sum (\hat{y}_i - \bar{y})^2\right)/(p-1)}{\left(\sum (y - \hat{y})^2\right)/(n-p)} \tag{1.40}$$

şeklinde elde edilen F değeri, α anlam seviyesinde *(p-1)* ve *(n-p)* serbestlik dereceli F tablo değeri ile karşılaştırılır. Tablo değerinden büyük F değerleri için H_0 reddedilir.

b) Bireysel parametrelerin testi (*t* testi)

$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i$, $i = 1,2,...n$ şeklinde verilen çoklu doğrusal regresyon modelinde bireysel parametreler için,

$H_0 : \beta_i = 0$

$H_1 : \beta_i \neq 0$

hipotezleri,

$$t = \frac{b_i}{s\sqrt{c_{ii}}} \tag{1.41}$$

test istatistiği yardımıyla test edilir. *(n-p)* serbestlik dereceli tablo değeri ile hesaplanan *t* değeri karşılaştırılır. Tablo değerinden büyük *t* değerine sahip parametreler için sıfır hipotezi reddedilir.

1.2.5.2 Güven aralıkları

Modelde $\underline{\varepsilon} \sim N(\underline{0}, \sigma^2 I)$ olması durumunda yani model hipotez modeli olduğunda parametre tahminleri yanında parametreler için güven aralıkları elde etmek mümkündür. Bilindiği gibi

$$C = (c_{ij}) = (X'X)^{-1} \tag{1.42}$$

olmak üzere,

$$\underline{\hat{\beta}} = (X'X)^{-1} X'\underline{Y} \sim N(\underline{\beta}, \sigma^2 C) \tag{1.43}$$

$$\frac{(n-2)S^2}{\sigma^2} = \frac{\underline{Y}'(\mathrm{I} - X(X'X)^{-1}X')\underline{Y}}{(n-2)} \sim \chi^2_{(n-2)} \tag{1.44}$$

dir. $\underline{Y}$ rasgele vektörünün bir fonksiyonu olan $(X'X)^{-1}X'\underline{Y}$ lineer formu ile $\underline{Y}'(\mathrm{I} - X(X'X)^{-1}X')\underline{Y}$ karesel formu bağımsızdır yani $\underline{\hat{\beta}}$ vektörünün her bir bileşeni olan rasgele değişken , S^2 'den bağımsızdır.

β_0 için güven aralığı:

$$\begin{aligned} &\frac{\hat{\beta}_0 - \beta_0}{\sqrt{\sigma^2 c_{11}}} \sim N(0,1) \\ &\frac{(n-2)S^2}{\sigma^2} \sim \chi^2_{(n-2)} \end{aligned} \tag{1.45}$$

olmak üzere , bağımsızlıktan

$$\left(\frac{\hat{\beta}_0-\beta_0}{\sqrt{\sigma^2 c_{11}}}\right)\Bigg/\sqrt{\left(\frac{(n-2)S^2}{\sigma^2}\right)\Bigg/(n-2)}\sim t_{(n-2)} \tag{1.46}$$

ve

$$\frac{\hat{\beta}_0-\beta_0}{\sqrt{\sigma^2 c_{11}}}\sim t_{(n-2)} \tag{1.47}$$

olmak üzere β_0 için $1-\alpha$ lık güven aralığı

$$\left(\hat{\beta}_0 - s\sqrt{c_{11}}t_{1-\alpha/2;n-2};\hat{\beta}_0 + s\sqrt{c_{11}}t_{1-\alpha/2;n-2}\right) \tag{1.48}$$

dır.

β_1 için güven aralığı:

Benzer düşüncelerle β_1 için $1-\alpha$ lık güven aralığı

$$\left(\hat{\beta}_1 - s\sqrt{c_{22}}t_{1-\alpha/2;n-2};\hat{\beta}_1 + s\sqrt{c_{22}}t_{1-\alpha/2;n-2}\right) \tag{1.49}$$

dir. σ^2 için güven aralığı :

$$\frac{(n-2)S^2}{\sigma^2}\sim \chi^2_{(n-2)} \tag{1.50}$$

olmak üzere σ^2 için $1-\alpha$ lık güven aralığı

$$\left(\frac{(n-2)S^2}{\chi^2_{(n-2);1-\alpha/2}};\frac{(n-2)S^2}{\chi^2_{(n-2);\alpha/2}}\right) \tag{1.51}$$

dir.

1.2.6 Regresyon analizinde genel lineer hipotezler

$\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$, $\varepsilon \sim N(0,\sigma^2 \mathrm{I})$ H : q × p tipinde rank(H)=q , (q ≤ p) olan bir matris ve $\underline{h}$: $q \times 1$ tipinde bir vektör olmak üzere

$$H_0 : H\underline{\beta} = h$$
$$H_1 : H\underline{\beta} \neq h$$

hipotezini α anlam düzeyinde test etmede aşağıdaki istatistik kullanılmaktadır .

$$\mathrm{W}(\underline{\mathrm{Y}}) = \frac{(\mathrm{H}\underline{\hat{\beta}} - \mathrm{h})' \left[\mathrm{H}(\mathrm{XX})^{-1}\mathrm{H}\right]^{-1} (\mathrm{H}\underline{\hat{\beta}} - \mathrm{h}) / \mathrm{q}}{\underline{\mathrm{Y}}(\mathrm{I} - \mathrm{X}(\mathrm{XX})^{-1}\mathrm{X})^{-1} / (\mathrm{n} - \mathrm{p})} \tag{1.52}$$

$$= \frac{(\mathrm{H}\underline{\hat{\beta}} - \mathrm{h})' \left[\mathrm{H}(\mathrm{X}'\mathrm{X})^{-1}\mathrm{H}'\right]^{-1} (\mathrm{H}\underline{\hat{\beta}} - \mathrm{h}) / \mathrm{q}}{\mathrm{SSE}/(\mathrm{n} - \mathrm{p})}$$

istatistiği H_0 hipotezi altında ,

$$W(\underline{Y}) \sim F(q, n-p) \tag{1.53}$$

dağılımlıdır (Graybill 1961).

2.VARYANS ANALİZİ

Varyans analizi (V.A) ölçü ile belirtilen kitlelerde normal dağılıma uyan üç yada daha fazla örneklem (gruplar) arasındaki farklılığın önemli olup olmadığını araştıran ve bu farklılığı meydana getiren sebepleri kontrolde kullanılan istatistiksel bir tekniktir. (V.A.)özellikle gruplar arasındaki farklılıklar miktar olarak belirlenemediğinden çok yararlıdır. Ayrıca örneklemler arasındaki farklılığın ortalamalardan mı yoksa varyanslardan mı geldiğinin açıklık kazanması için öncelikle, örneklemlerin varyansları homojen olan kitlelerden alındığı varsayımı yapılır.

Varyans analizinin ikiden fazla ortalamanın toplu olarak kıyaslanmasında kullanılabilmesi için şu koşulların mevcut olması gerekir:

1. Kıyaslanan değişik diziler, normal bölünme gösteren ana yığınlardan çekilmiş olmalıdır.
2. Dizilerin ayrıldığı ana yığınların varyansları eşit olmalıdır(Çömlekçi 1999).

2.1. Bir Yönlü Varyans Analizi

Bu modelde tek faktör ve bu faktörün çeşitli seviyeleri ya da denemeler vardır. Burada amaç denemelerin bağımlı değişken üzerindeki etkisinin incelenmesidir. Örneklemlerin alındığı kitleler normal dağılımlı ve eşit varyanslı olmak üzere inceleme altındaki değişken veya değişkenler dış etkenlerden etkilenmiyorsa bu durumda ortalamalar varyans analizi ile aksi durumda kovaryans analizi ile karşılaştırılırlar.

Varyans analizi değişkenlik kaynaklarını bileşenlere ayırma ve bu bileşenlere ait varyansları karşılaştırma yöntemidir.

Tek etkene göre bağımsız örneklemlerin ortalamalarını genel kareler toplamının parçalama yöntemiyle karşılaştıran varyans analizine basit veya tek yönlü varyans analizi denir.

Normal dağılıma sahip bir kitleden rasgele çek her biri m_i genişliğindeki k tane grup alınsın veya normal dağılıma sahip k tane kitleden rasgele çekilmiş her biri m_i genişliğinde birer grup alındığında verilerin gruplara dağılımları Tablo 1.'deki gibi özetlenebilir.

Tablo.1 Bir Yönlü Varyans Analizi Deney Düzeninde Gruplar ve Gözlem Değerleri

Gruplar(Denemeler)	Gözlemler	Toplam	Gözlem Sayısı	Ortalama
1	$y_{11}, y_{12}, \ldots, y_{1m_i}$	T_1	m_1	$\overline{Y}_1$
2	$y_{21}, y_{22}, \ldots, y_{2m_i}$	T_2	m_2	$\overline{Y}_2$
⋮	⋮	⋮	⋮	⋮
k	$y_{k1}, y_{k2}, \ldots, y_{km_i}$	T_k	m_k	$\overline{Y}_k$

Kitlelerde veya gruplarda varyansların eşit oldukları varsayımı altında varyans analizi için doğrusal model;

$$X_{ij} = \mu + \tau_i + \varepsilon_{ij} \qquad i = 1,2,\ldots,k \qquad \text{ve} \qquad j = 1,2,\ldots,m_i \tag{2.1}$$

şeklindedir. Bu denklemde,

X_{ij} : Normal dağılımlı kitleye sahip bağımsız gözlemler, başka bir deyişle i'nci grupta j' inci gözlem değeri $X_{ij} \sim N(\mu + \alpha_i, \sigma^2)$ dir.

μ : Tüm kitlelerde X_{ij}'lerin ortalamasıdır. Buna genel ortalama denir.

τ_i : Değişkenin i 'inci işlem veya satırdaki etkisi

m_i : Bir gruptaki veri sayısı

ε_{ij} : i'inci grupta j 'inci gözleme ait hata terimi

anlamındadır. Burada $\tau = \mu_i - \mu$ olarak tanımlanır. *i*'inci deneme ortalamasıdır. Deneme seviyeleri özel seçildiği için değeri sabit olarak düşünülür. Sıfır ortalama ve σ^2 varyans ile normal dağılım gösteren rasgele değişkendir.

$\varepsilon \sim N(0,\sigma^2)$'dir ve $\sum_{i=1}^{k} \tau_i = 0$ ifadesinin doğru olduğu varsayılır. Bu modelde başka bir varsayım, grupların seçildiği k kitlelerin varyanslarının homojen olmasıdır. Bu nedenle grupların homojenlik kontrolünün yapılması gerekir.

Tablo 2.6'da T_i , i 'inci gruptaki gözlemlerin toplamı, m_i, *i* 'inci gruptaki gözlemlerin sayısı ve X_i, i 'inci gruptaki gözlemlerin ortalamasıdır.T tüm gözlemlerin genel toplamıdır ve

$$T = \sum_{i=1}^{k} \sum_{j=1}^{mi} X_{ij} = \sum T_{i.}$$

$$T_{i.} = \sum_{j=1}^{mi} X_{ij} \qquad (i = 1,2,...,k) \tag{2.2}$$

dir.

Gerçekte μ , τ_i ve ε_{ij} değerleri bilinmez. Burada amaç grup etkilerini ve genel ortalamayı tahmin etmek ve *k* kitle ortalaması arasındaki farklılığı test etmek olacaktır. μ' yü tahmin etmek için genel örneklem ortalaması,

$$\overline{X}_{..} = \frac{\sum_{i=1}^{k} \sum_{j=1}^{mi} X_{ij}}{\sum_{i=1}^{k} m_i} = \frac{T_{..}}{n} = \frac{\sum T_{i.}}{n} \tag{2.3}$$

olarak tanımlanır. μ_i 'nin tahmin edicisi ise,

$$\overline{X}_{i.} = \frac{\sum_{j=1}^{mi} X_{ij}}{m_i} = \frac{T_{i.}}{m_i}; \qquad i = 1,2,..,k \tag{2.4}$$

dir.

k grup etkisinin eşit olup olmadığı yoklanmak istenildiğine göre $\Sigma\tau_i = 0$ özelliği kullanılarak yokluk hipotezi,

$$H_0 : \tau_1 = \tau_2 = ... = \tau_k = 0 \tag{2.5}$$

biçimindedir. $\Sigma\tau_i = 0$ olduğundan dolayı μ_i 'lerin (i = 1,2,...,k) μ'ye eşit olduğu görülür. Buradan k kitlenin ortalamasının eşit olduğunu ifade eden yokluk hipotezi (2.5) hipotezine eşdeğer olarak,

$$H_0 : \mu_1 = \mu_1 = ... = \mu_k \tag{2.6}$$

biçiminde kurulabilir.

Kitle ortalamalarının eşit olmadığı biçimindeki karşıt hipotez ise,

$H_1 : \tau_i \neq 0$ bir *i* içindir.

Eğer yokluk hipotezi doğru ise her bir gözlem, genel ortalama ve ε_{ij} rasgele hatanın toplamı olacaktır ($X_{ij} = \mu + \varepsilon_{ij}$).

(2.5) yada (2.6) eşitliği ile verilen hipotezler için yapılan test süresince Varyans Çözümlemesi denilecektir. Çözümleme toplam değişimin parçalanarak incelenmesidir. Bu nedenle çözümlemenin adında varyans sözcüğü kullanılmaktadır. Bu durumda toplam değişimin ölçüsü,

$$\sum_{i=1}^{k}\sum_{j=1}^{mi}(X_{ij}-\overline{X})^2=\sum_{i=1}^{k}\sum_{j=1}^{mi}(X_{ij}-\overline{X}_{i.})^2+\sum_{i=1}^{k}m_i(X_i-\overline{X})^2 \qquad (2.7)$$

biçimindedir.(2.7) eşitliğinde

$\sum_{i=1}^{k}\sum^{mi}(Xi_j-X_{..})$: Genel kareler toplamı *(GnKT).*

$\sum_{i=1}^{k}\sum_{j=1}^{mi}(X_{ij}-\overline{X}_{i.})^2$: Grup içi kareler toplamı *(GİKT).*

$\sum_{i=1}^{k}m_i(X_{i.}-\overline{X}_{..})^2$: Gruplar arası kareler toplamı *(GAKT)'*dır.

Bu tanımlamadan sonra (2.7) eşitliği,

$$GnKT=GİKT+GAKT$$

simge ile de gösterilebilir.

Gruplar arası kareler toplamı genel kareler toplamının faktör tarafından açıklanan kısmını verir. Bu değerin büyük çıkması beklenir. Gruplar içi kareler toplamı genel kareler toplamının faktör tarafından açıklanamayan kısmıdır.

Ortalamalar tamsayı değil ise *GnKT*, *GAKT* ve *GİKT* için aşağıdaki hesaplama formülleri kullanılabilir.

$$GnKT=\sum_{i=1}^{k}\sum_{j=1}^{mi}X_{ij}{}^2-\frac{(T)^2}{n} \qquad (2.8)$$

Burada $(T)^2$ /n düzeltme terimidir ve *DT* ile gösterilir.

$$GAKT=\sum_{i=1}^{k}\frac{(T_{i.})^2}{m_i}-DT \qquad (2.9)$$

$$GIKT = \sum_{i=1}^{k}\sum_{j=1}^{mi} X_{ij}{}^{2} - \sum_{i=1}^{k}\frac{(T_{i.})^{2}}{m_i} \qquad (2.10)$$

Genel ortalama ile grup ortalamaları arasındaki farkların mutlak değerleri büyüdükçe *GAKT*'da büyür. *GİKT* 'nın büyük olması faktörün iyi seçilmiş olduğunun göstergesidir.

Bulunan kareler toplamına karşılık gelen serbestlik derecesi Genel Serbestlik Derecesi olup, *GnSD* şeklinde gösterilir.

GnSD = n-1'dir (2.11)

Gruplar Arası Kareler Toplamına karşılık gelen serbestlik derecesine Gruplar Arası Serbestlik Derecesi denir ve *GASD* şeklinde gösterilir.

GASD=k-1 (2.12)

Gruplar içi Kareler Toplamına karşılık gelen serbestlik derecesine Gruplar İçi Serbestlik Derecesi denir ve *GİSD* şeklinde gösterilir.

$$GİSD = n - k \qquad \text{veya} \quad \sum_{i=1}^{k}(m_i - 1) \qquad (2.13)$$

Bulunan kareler toplamı ve serbestlik dereceleri yardımı ile Gruplar Arası Kareler ortalaması *GAKO* ile gösterilir.

$$GAKO = \frac{GAKT}{GISD}'dir. \qquad (2.14)$$

Gruplar İçi Kareler Ortalaması *GIKO* ile gösterilir ve

$$GIKO = \frac{GIKT}{GISD}\text{'dir.} \qquad (2.15)$$

2.1.1 F Değeri

Normal dağılıma sahip bir kitleden m genişliğinde bir grup çekildiğini kabul edelim. Bu durumda grup varyansı, kitle varyansının kestirim değeri olarak düşünülebilir

Normal dağılıma sahip bir kitleden her biri m genişliğindeki k tane grup çekildiğinde grup varyansları kullanılarak toplanmış varyans hesaplanabilir. Hesaplanan değerler kitle varyansının kestirim değeri olarak kullanılır. Toplanmış varyans S_T^2 gösterilir.

$$S_T^2 = \frac{\sum_{i=1}^{k} X_{ij}^2 - \frac{\left(\sum_{J=1}^{m_i} Xi_j\right)}{m_i}}{\sum_{i=1}^{k}(m_i - 1)} \tag{2.16}$$

dir. Yukarıdaki eşitlikte pay *GİKT* payda *GİSD* olduğundan toplanmış varyansın gruplar içi kareler ortalamasının kendisi olduğu anlaşılır. Böylece toplanmış varyans,

$$S_T^2 = \frac{GIKT}{GISD} = GIKO \tag{2.17}$$

dir. Normal dağılım gösteren bir kitleden her biri m_i genişliğinde ve varyansları eşit k sayıda grup çekilmiş olsun. Grup ortalamaları da kitle ortalamalarının iki yanında, kitle ortalamasından uzaklaştıkça gittikçe uzanan sıklıkta normal dağılım gösterirler. Grup ortalamalarının göstermiş oldukları dağılımın varyansları *S* ile gösterilir ve

$$S_{\overline{X}}{}^2 = \frac{\sum_{i=1}^{k} m_i (X_{i.} - \overline{X})}{k-1} \tag{2.18}$$

dir. Yukarıdaki eşitlikte pay *GAKT* ve payda *GASD*'dir. Böylece;

$$S_{\overline{X}}^2 = \frac{GAKT}{GASD} = GAKO \tag{2.19}$$

dir. Kitle varyansının kestiriminde $S_{\overline{X}}^2$ 'de kullanılabilir. Bu durumda kitle varyansının hesaplanması iki ayrı yolla hesaplanmış olur.

Normal dağılım gösteren bir kitleden her bir m_i genişliğinde varyansları eşit k sayıda grup çekilmiş olsun. Grup ortalamaları arasında farkın önemli bulunmadığı ($\mu_1 . = \mu_{2.} = ... = \mu_{k.}$) durumunda S_T^2 'nin $S_{\overline{X}}^2$ 'ye eşit veya hata sınırları içerisinde birbirinin aynı olması gerekir. Yani bu iki değerin oranlarının 1'e eşit olması beklenir.

Araştırmacı işlemleri gruplara rasgele dağıtır ve grup ortalamalarına bakarak işlemler arasında önemli farkın olup olmadığını anlamak ister. Eğer işlemler arasında önemli fark varsa $S_{\overline{X}}^2$ 'nin S_T^2 'den daha büyük olması gerekir. Bu durumda grup ortalamaları varyansının toplanmış varyansa bölümünden,

$$F = \frac{GAKO}{GIKO} \tag{2.20}$$

elde edilir ve bu değerin l 'den önemli derecede farklı olup olmadığına bakılır.

F tablosunda üst sırada ve sol sütunda l 'den sonsuza kadar serbestlik dereceleri yer almakta ve seçilen her yanılma olasılığı için ayrı bir *F* tablosu verilmiş bulunmaktadır.

F tablosu kullanılırken *GAKO* ile ilgili serbestlik derecesi üst satırdan *GIKO* ile ilgi serbestlik derecesi de sol sütundan bulunur. Bu iki serbestlik derecesinin tablo üzerinde kesiştikleri noktadaki *F* değeri seçilen yanılma olasılığındaki rasgele hatanın üst sınırını verir. Bu değer F_l ile gösterilir.

F_t değerini bulmada kareler ortalamasının büyük alanının serbestlik derecesine üst sıradan ve küçük alanının serbestlik derecesine sol sütundan bakmakta önerilir fakat önemlilik açısından herhangi bir farklılık söz konusu olmamaktadır.

2.1.2 Bir Yönlü Varyans Analizinde Hipotez Testi

Bir yönlü varyans analizinde kitle varyanslarının homojen olduğu koşulu altında gruplar arasındaki farkın önemli olup olmadığını test etmek için izlenen süreç şöyledir:

1. Yokluk hipotezi

H_0 : $\mu_1 = \mu_2 = ... = \mu_k$

H_1 : μ_i ‘ lerden en az biri farklıdır.

2. (2.8), (2.9) ve (2.10) eşitliklerinden GnKT, GAKT ve GİKT hesap1anır.

Gruplardaki gözlem sayıları birbirinden farklı yani

$m_1 = m_2 = ... = m_k$ ise $n = \sum_{i=1}^{k} m_i$ ’dir

3. H_0 hipotezi altında test istatistiği

$$F_H = \frac{GAKO}{GIKO}' dir.$$

4. Yorum yapılır.

$F_H > F_T$ ise H_0 hipotezi red edilir. Aksi durumda H_0 hipotezi kabul edilir. Burada F_T : α yanılma olasılığında ki s_1 ve s_2 serbestlik derecesindeki F tablo değeridir.

Tablo 2. Varyans Analizi Tablosu

Değişim Kaynakları D.K.	**Serbestlik Derecesi S.D.**	**Kareler Toplamı K.T.**	**Kareler Ortalaması K.O.**	**Test *F* İsatistiği**
Genel	GnSD=n-1	GnKT		$F_H = \frac{GAKO}{GIKO}$
Gruplar Arası	GASD=k-1	GAKT	GAKO=$\frac{GAKT}{k-1}$	
Grup İçi(Hata)	GİSD=n-k	GİKT	GIKO=$\frac{GIKT}{n-k}$	

2.1.3 Tek Faktörlü Tesadüfi Bloklar Deney Düzeni İçin Varyans Analizi

Deney düzenlemenin temel amacı deneysel hatayı azaltmaktır. Bir blokta söz konusu olan her denemenin yapılmış olduğu ve tek bir faktörün etkisinin araştırıldığı deney düzenine tek faktörlü tesadüfi bloklar deney düzeni denir.

Deneme sayısı t ve tekrarlama sayısı r olan tek faktörlü tesadüfi bloklar deney düzeni için veri düzeni Tablo 3. deki gibidir.

Denemeler	**Bloklar**			
	Tek.1	**Tek.2**	**…**	**Tek.r**
M_1	Y_{11}	Y_{12}	…	Y_{1r}
M_2	Y_{21}	Y_{22}	…	Y_{2r}
⋮	⋮	⋮	⋮	⋮
M_t	Y_{t1}	Y_{t2}	…	Y_{tr}

Tek etkenli tamamlanmış tesadüfi bloklar deney düzeni için model denklemi;

$$Y_{ij}=\mu+\beta_i+\tau_j+\varepsilon_{ij} \qquad (i = 1,2,...,r)\ (j = 1.2,...,t) \tag{2.21}$$

eşitliği ile elde edilir. Burada μ_j, ve μ kitle ortalamaları olmak üzere;

$\beta_i = \mu_i - \mu$; olup i blok etkisini,

$\tau_j = \mu_j - \mu$; olup j. deneme etkisini göstermektedir. (2.21) eşitliğindeki model denklemi için örnekleme ait genel kareler toplamı

$$GKT=\sum_{j=1}^{t}\sum_{i=1}^{r}(\overline{Y}_{ij}-\overline{Y})^2$$

$$=\sum_{j=1}^{t}\sum_{i=1}^{r}(\overline{Y}_{i}-\overline{Y})^2+\sum_{j=1}^{t}\sum_{i=1}^{r}(\overline{Y}_{j}-\overline{Y})^2+\sum_{j=1}^{t}\sum_{i=1}^{r}(\overline{Y}_{ij}-\overline{Y}_{i}-\overline{Y}_{j}+\overline{Y})^2 \tag{2.22}$$

$$GKT = KT_{Blok} + KT_{Deneme} + HKT$$

şeklinde parçalara ayrılır. Burada,

KT_{Blok} : Bloklar için kareler toplamı

KT_{Deneme} : Denemeler için kareler toplamı

dır.

Bu deney düzeninde varyans analizinin amacı, denemelerin (faktör seviyelerinin) ve blokların bağımlı değişken üzerinde aynı etkiyi yapıp yapmadığı ayrı olarak kontrol etmektir. Bu amaca yönelik olarak söz konusu kontrol işlemi bir hipotez yoklaması ile yapılabilir.

Burada test edilecek hipotezler;

(i) Deneme etkileri için;

$H_0 : \tau_j = 0$ (tüm j 'ler için)

$H_1 : \exists \tau_j \neq 0$

(ii) Blok etkileri için;

$H_0 : \beta_i = 0$ (tüm i'ler için)

$H_1 : \exists \beta_i \neq 0$

şeklinde kurulur. Her iki durumda da H_0 hipotezi doğru iken hem denemelerin hem de blokların bağımlı değişken üzerine aynı etkiyi yapacağı varsayımından hareketle istatistiği ve örnekleme dağılımı elde edilir.

Tablo 4. Tek Faktörlü Tesadüfi Bloklar Deney Düzeni İçin Varyans Analizi Tablosu

Kaynak	**Serbestlik Derecesi Sd**	**Kareler Toplamı KT**	**Kareler ortalaması KO**	**Test F İstatistiği**
Denemeler (τ_j)	$t-1$	KT_{Deneme}	$KO_{Deneme} = \frac{KT_{Deneme}}{t-1}$	$F_{(t-1);(t-1)(r-1)} = \frac{KO_{Deneme}}{HKO}$
Bloklar (β_i)	$R-1$	KT_{Blok}	$KD_{Blok} = \frac{KT_{Blok}}{r-1}$	$F_{(r-1);(t-1)(r-1)} = \frac{KO_{Blok}}{HKO}$
Hata (ε_{ij})	$(t-1)(r-1)$	HKT	$HKO = \frac{HKT}{(t-1)(r-1)}$	
Genel	$Tr-1$	GKT		

Tablo 4.'de verilen kareler toplamları, bu değerlerin eşiti olan aşağıdaki bağıntılar ile hesaplanabilir.

$$KT_{Deneme} = \sum_{j=1}^{t}\sum_{i=1}^{r}(\overline{Y}_j - \overline{Y})^2 = \sum_{j=1}^{t}\frac{Y_j^2}{r} - \frac{Y^2}{rt} \tag{2.23}$$

$$KT_{Blok} = \sum_{j=1}^{t}\sum_{i=1}^{r}(\overline{Y}_i - \overline{Y})^2 = \sum_{j=1}^{r}\frac{Y_i^2}{r} - \frac{Y^2}{rt} \tag{2.24}$$

$$HKT = \sum_{j=1}^{t}\sum_{i=1}^{r}Y_{ij}^{\ 2} - \sum_{j=1}^{t}\frac{Y_j^2}{r} - \sum_{i=1}^{r}\frac{Y_i^2}{t} + \frac{Y^2}{rt} \tag{2.25}$$

Burada;

Y_i *: i.* blok toplamıdır.

Şimdi $P(F \leq F_{(t-1);(t-1)(r-1)} = p$ ihtimalini göz önüne alalım. α anlamlılık düzeyi olmak üzere, eğer $p \geq \alpha$ ise (*i*)'de deneme etkileri için kurulan H_0 hipotezi kabul edilir ve denemelerin (faktör seviyelerinin) bağımlı değişken üzerinde aynı etkiyi yaptığı sonucuna verilir. Eğer $p < \alpha$ ise H_0 reddedilir ve en az bir denemenin diğerlerinden farklı etki yaptığı ortaya çıkar.

$P(F \leq F_{(r-1);(t-1)(r-1)}) = p$ ihtimali göz önüne alınırsa, α anlamlılık düzeyi olmak üzere, eğer $p \geq \alpha$ ise (*ii*)'de blok etkileri için kurulan H_0 hipotezi kabul edilir ve blokların bağımlı değişken üzerinde aynı etkiyi yaptığı sonucuna ulaşılır. Eğer $p < \alpha$ ise H_0 reddedilir ve en az bir blok diğerlerinden farklı etki yapıyor sonucuna varılır.

Ayrıca deneme veya blok etkileri karşılaştırılmak istendiğinde Newman-Keuls en küçük önem genişliği testi kullanılabilir (Çelik 1997).

3. KOVARYANS ANALİZİ

Bağımlı değişken üzerinde bir veya daha fazla etkenin etkisi araştırıldığında, bazen bağımlı değişken ile birlikte değişen başka bir değişken veya değişkenler vardır. Çoğunlukla bu diğer değişkeni (veya değişkenleri) deney süresince değişmez bir düzeyde denetim altında tutmak mümkün olmaz, ancak bu değişken (veya değişkenler) bağımlı değişkenler ile ölçülebilir. Bu değişkene, bağımlı değişken ile birlikte değiştiğinden "birlikte değişen değişken" adı verilir ve genellikle (X) ile gösterilir. Denemelerin bağımlı değişken üzerindeki etkisini ölçmek için öncelikle birlikte değişen değişkenin (veya değişkenlerin) bağımlı değişken üzerinde etkisinin giderildiği ve denemelerin bağımlı değişken üzerindeki kalan miktarının etkisi için çözümlendiği yönteme "Kovaryans Analizi" denir. Kovaryans analizi varyans analizi ile regresyon analizinin bir kombinasyonudur (Akgül 2005).

3.1. Kovaryans Analizi ile İlgili Varsayımlar

Kovaryans analizi için varsayımlar, varyans analizi ile regresyon analizi için verilen varsayımların bir kombinasyonu olarak düşünülebilir. Geçerli bir kovaryans analizi için gerekli varsayımlar şunlardır (Akgül 2005):

(i) Bağımsız değişkenler deneme konularından etkilenmemektedir.

(ii) Blok ve deneme farkları çıkarıldıktan sonra bağımsız değişken ile bağımlı değişken arasındaki regresyon doğrusaldır ve blok ile denemelerden etkilenmemektedir. Başka bir ifadeyle, bağımlı değişken ile bağımsız değişken arasındaki ilişki, blok ve deneme etkileri için düzeltildikten sonra $y = \alpha + \beta x$ formundadır.

(iii) Regresyon katsayısı β, blok ve denemelerden bağımsızdır.

(iv)Denemede, kontrol edilemeyen faktörlerin etkileri yani deneme hataları ε_{ij}'ler ortalaması sıfır ve varyansı σ^2 olan normal dağılım gösterirler (Akgül 2005).

3.2. Kovaryans Analizi İçin Doğrusal Model

En genel haliyle kovaryans analizi modeli;

$$y = Z\alpha + X\beta + \varepsilon \tag{3.1}$$

şeklindedir. Burada;

Z ; 0 ve 1 leri

α ; μ ve α_i , β_i , γ_{ij} gibi faktör ve interaksiyon parametrelerini veya diğer etkileri

X ; Kovariate değerlerini

β ; Kovariatelerin kastsayılarını

içerir. Böylece kovariateler (3.1)'in sağ tarafında bağımsız değişken gibidir (Rencher 2001).

Tek yönlü bir kovariate'li model aşağıdaki gibi ifade edilir.

$$y_{ij} = \mu + \alpha_i + \beta x_{ij} + \varepsilon_{ij} \qquad i=1,2,...,k \qquad j=1,2,...,n \tag{3.2}$$

Burada;

α_i ; muamele etkisi

x_{ij} ; y_{ij} ile aynı örneklemde gözlenen kovariate değerleri

β ; x_{ij} 'nin y_{ij} ile ilişkisinin eğimi

biçimindedir.

Eğer (3.2) bir regresyon modeli olarak göz önüne alınırsa k grup için $\mu + \alpha_i$ regresyon sabitleri olarak düşünülür. kn gözlem (3.2) için ;

$$y = Z\alpha + X\beta + \varepsilon$$

formunda (3.1)'de ki gibi yazılabilir.

Burada;

$$Z=\begin{pmatrix}1 & 1 & 0 & \cdots & 0\\ \vdots & \vdots & \vdots & & \vdots\\ 1 & 1 & 0 & \cdots & 0\\ 1 & 0 & 1 & \cdots & 0\\ \vdots & \vdots & \vdots & & \vdots\\ 1 & 0 & 0 & \cdots & 1\end{pmatrix}, \quad \alpha=\begin{pmatrix}\mu\\ \alpha_1\\ \vdots\\ \alpha_k\end{pmatrix} \quad , \quad \mathrm{X}=x=\begin{pmatrix}x_{11}\\ \vdots\\ x_{1n}\\ x_{2n}\\ \vdots\\ x_{kn}\end{pmatrix} \tag{3.3}$$

ve $\beta=\beta$ 'dır. Bu durumda Z , (3.3a) deki X ile aynıdır.

$$\begin{pmatrix}y_1\\ y_2\\ \vdots\\ y_k\end{pmatrix}=\begin{pmatrix}j & j & 0 & \cdots & 0\\ j & 0 & j & \cdots & 0\\ \vdots & \vdots & \vdots & & \vdots\\ j & 0 & 0 & \cdots & 0\end{pmatrix}\begin{pmatrix}\mu\\ \alpha_1\\ \vdots\\ \alpha_k\end{pmatrix}+\begin{pmatrix}\varepsilon_1\\ \varepsilon_2\\ \vdots\\ \varepsilon_k\end{pmatrix} \tag{3.3a}$$

Dolayısıyla;

$$y=X\beta+\varepsilon$$

dır. Tek yönlü q kovariate'li model;

$$y_{ij}=\mu+\alpha_i+\beta_1 x_{ij1}+\ldots+\beta_q x_{ijq}+\varepsilon_{ij} \qquad i=1,2,\ldots,k \qquad j=1,2,\ldots,n \tag{3.4}$$

Bu durumda (3.3) 'de verilen; Z , α ve $X\beta$ aşağıdaki formdadır.

$$X\beta=\begin{pmatrix}x_{111} & x_{112} & \cdots & x_{11q}\\ x_{121} & x_{122} & \cdots & x_{12q}\\ \vdots & \vdots & & \vdots\\ x_{kn1} & x_{kn2} & \cdots & x_{knq}\end{pmatrix}\begin{pmatrix}\beta_1\\ \beta_2\\ \vdots\\ \beta_q\end{pmatrix} \tag{3.5}$$

olur. Tek kovariateli iki yönlü model ise;

$$y_{ijk}=\mu+\alpha_i+\gamma_i+\delta_{ij}+\beta x_{ijk}+\varepsilon_{ijk}$$

$$i=1,2,\ldots,a \quad j=1,2,\ldots,c \quad k=1,2,\ldots,n \tag{3.6}$$

şeklinde iki yönlü model bir kaç kovariate içeren modele genişletilebilir(Rencher 2001). Bu durumda model parametrelerinin tahmini önem arzeder.

3.2.1. Tahmin

Eşitlik (3.1)' de genel durumda α ve β için tahmin ediciler geliştirilsin.

$y = Z\alpha + X\beta + \varepsilon$

Z'nin aşırı parametreli ANOVA modelindeki tam ranklı durumdan daha az ranklı ve X'in regresyon modelindeki gibi tam ranklı olduğu varsayılsın.Aynı zamanda aşağıdaki koşulların bulunduğu varsayılsın.

$$E(\varepsilon)= 0 \quad \text{ve} \quad \text{cov}(\varepsilon) = \sigma^2 \text{I}$$

Model,

$$\begin{aligned} y &= Z\alpha + X\beta + \varepsilon \\ &= (Z, X)\begin{pmatrix}\alpha \\ \beta\end{pmatrix} + \varepsilon \\ &= U\theta + \varepsilon \end{aligned} \tag{3.7}$$

biçiminde ifade edilebilir. Burada;

$$\text{U} = (Z, X) \quad ; \quad \theta = \begin{pmatrix}\alpha \\ \beta\end{pmatrix}$$

dır. Eşitlik (3.7) için normal denklemler $U'U\hat{\theta} = U'y$ olmak üzere parçalanmış şekli;

$$\begin{pmatrix}Z' \\ X'\end{pmatrix}(Z, X)\begin{pmatrix}\hat{\alpha} \\ \hat{\beta}\end{pmatrix} = \begin{pmatrix}Z' \\ X'\end{pmatrix} y$$

olup,

$$\begin{pmatrix}Z'Z & Z'X \\ X'Z & X'X\end{pmatrix}\begin{pmatrix}\hat{\alpha} \\ \hat{\beta}\end{pmatrix} = \begin{pmatrix}Z'y \\ X'y\end{pmatrix} \tag{3.8}$$

normal denklemlere ulaşılır. Buradan;

$$Z'Z\,\hat{\alpha} + Z'X\hat{\beta} = Z'\,y$$
$$X'Z\,\hat{\alpha} + X'X\hat{\beta} = X'\,y \qquad (3.10)$$

$Z'Z$ nin genelleştirilmiş tersini kullanarak (3.9) daki $\hat{\alpha}$ için çözülürse;

$$\hat{\alpha} = (Z'Z)^{-}Z'y - (Z'Z)^{-}Z'X\hat{\beta}$$
$$= \hat{\alpha}_0 - (Z'Z)^{-}Z'X\hat{\beta} \qquad (3.11)$$

$\hat{\alpha}_0 = (Z'Z)^{-}Z'y$, kovariatesiz $y = Z\alpha + \varepsilon$ modeli için normal denklemlerin çözümüdür.

$\hat{\beta}$ yı çözmek için (4.11) 'i (4.10) da yerine koyulduğunda,

$$X'Z\ [\ (Z'Z)^{-}Z'y - (Z'Z)^{-}Z'X\hat{\beta}\ ] + X'X\hat{\beta} = X'\,y$$

veya

$$X'Z\,(Z'Z)^{-}Z'y + X'\left[I - Z\,(Z'Z)^{-}Z'\right]X\hat{\beta} = X'\,y \qquad (3.12)$$

elde edilir.

$P = Z(Z'Z)^{-}Z'$ (3.13) dönüşümü tanımlanırsa (3.12) 'de;

$$X'(I-P)X\hat{\beta} = X'\,y - X'P\,y = X'(I-P)y$$

haline gelir.

X'lerin elemanları belirgin bir biçimde, Z'de ki 0 ve 1 'lerden ilişkisiz bir bağımsız olduğunu varsayabiliriz. Böylece $X'(I-P)X$ nonsingulerdir ve $\hat{\beta}$ 'nın bir x çözümü;

$$\hat{\beta} = [\ X'(I-P)X\ ]^{-1}X'(I-P)y \qquad (3.14)$$

veya

$$= E_{xx}^{-1}\,e_{xy} \qquad (3.15)$$

dır.

Burada;

$$E_{xx} = X'(I-P)X \quad \text{ve} \quad e_{xy} = X'(I-P)y \tag{3.16}$$

dır. (3.1) ve (3.7) deki kovaryans analizi modeli için SSE'yi SSE_{yx} açılımı yapılırsa;

$$\begin{aligned}
\text{SSE}_{yx} &= y'y - \hat{\theta}'U'y \\
&= y'y - (\hat{\alpha}', \hat{\beta}')\begin{pmatrix} Z'y \\ X'y \end{pmatrix} \\
&= y'y - (\hat{\alpha}'Z'y, -\hat{\beta}'X'y) \\
&= y'y - [\ \hat{\alpha}_0' - \hat{\beta}'X'Z(Z'Z)^{-}\ ]Z'y - \hat{\beta}'X'y \\
&= y'y - \hat{\alpha}_0'Z'y - \hat{\beta}'X'\left[I - Z(Z'Z)^{-}Z'\right]y \\
&= \text{SSE}_y - \hat{\beta}'X'(I-P)y
\end{aligned} \tag{3.17}$$

bulunur. $\hat{\alpha}_0$ (3.11) de; P (3.13) de; tanımlanmış olduğundan yukarıdaki gibi yazılabilir.

$\text{SSE}_y = y'y - \hat{\alpha}_0' Z'y$ ise kovariatesiz ANOVA modeli $y = Z\alpha + \varepsilon$ için hesaplanan SSE ile aynıdır (Rencher 2001). (3.6) 'yı kullanarak (3.7) 'yi yeniden yazılırsa;

$$\text{SSE}_{yx} = e_{yy} - e_{xy}' E_{xx}^{-1} e_{xy} \tag{3.18}$$

elde edilir.

Burada;

$$e_{yy} = SSE_y = y'(I-P)y \tag{3.19}$$

olup,

(3.18) eşitliğinde SSE'deki azalma görülebilmektedir.

$E_{xx} = X'(I-P)X$ nonsingular ve E_{xx} 'in pozitif tanımlı olduğu gösterilebilir (Rencher 2001). Böylece,

$e'_{xy} E_{xx}^{-1} e_{xy} \rangle 0$ ve $SSE_{yx} \langle SSE_y$ 'dir

3.2.2 Hipotez Testi

Hipotez testini yapabilmek için (4.1) deki ε için ortalaması 0, varyansı $\sigma^2 I$ olan normal dağılımdan geldiği varsayılır. $N_n\ (0, \sigma^2 I)$ n;Z veya X için satır sayısı iken, (4.7) modeli kullanılarak α hakkında $H_0 : C\theta = 0$, C= (C_1,0) hipotezi ifade edilsin;

$$H_0 : (C_1, 0)\begin{pmatrix}\alpha\\ \beta\end{pmatrix} = 0 \quad \text{veya} \quad H_0 : C_1\alpha = 0$$

genel hipotez testini kullanabiliriz.

β ile ilgili hipotezler $H_0 : C\theta = 0$ formunda ifade edilebilir.

$$H_0 : C\theta = (0, C_2)\begin{pmatrix}\alpha\\ \beta\end{pmatrix} = 0 \quad \text{veya} \quad H_0 : C_2\beta = 0$$

En basit haliyle $H_0 : \beta = 0$ hipotezi ile kovariatelerin modele ait olmadığı hipotezi test edilmek istenmektedir (3.1).

$H_0 : \beta = 0$ için genel doğrusal hipotez testi yapmak için ; $\text{cov}(\hat{\beta})$, $\hat{\beta}$ (3.14) için $\hat{\beta} = [\ X'(I-P)X\ ]^{-1} X'(I-P)y$ ye ihtiyaç vardır. I-P ıdempotent olduğundan $\text{cov}(\hat{\beta})$ aşağıdaki gibi bulunur (Rencher 2001):

$$\text{cov}(\hat{\beta}) = [\ X'(I-P)X\ ]^{-1} X'(I-P)\ \sigma^2 I (I-P)X [\ X'(I-P)X\ ]^{-1}$$

$$= \sigma^2 \left[\; X'(I-P)X \; \right]^{-1} \tag{3.20}$$

$H_0 : \; \beta = 0$ için

$$\text{SSH} = \hat{\beta}' \, X'(I-P)X \, \hat{\beta} \tag{3.21}$$

dır.

(3.16)' yı kullanarak ;

$$\text{SSH} = \; e'_{xy} \, E_{xx}^{-1} \, e_{xy} \tag{3.22}$$

elde edilir.

SSH , kovariateler yüzünden SSE'de meydana gelen azalmaya denktir (Rencher 2001).

3.3.Bir Kovariate (Ortak Değişken)'li Tek Yönlü Model

3.3.1. Model

Tek yönlü model (3.2)'de tanıtıldı.

$$y_{ij} = \mu + \alpha_i + \beta \, x_{ij} + \varepsilon_{ij} \qquad i=1,2,...,k \quad j=1,2,...,n \tag{3.23}$$

Tüm kn tane gözlem (3.1) formunda yazılabileceğinden (3.3) de verilen Z , α ve β için;

$$\mathbf{y} = \mathbf{Z}\alpha + \mathbf{X}\beta \; + \varepsilon \; = Z\alpha + x\beta + \varepsilon$$

biçimindedir.

3.3.2. Tahmin

Parametre tahmini,

$$\hat{\alpha} = \hat{\alpha}_0 - (Z'Z)^{-} \, Z' \, X\hat{\beta} \; = \hat{\alpha}_0 - (Z'Z)^{-} \, Z' \, x\hat{\beta}$$

$$= \begin{pmatrix} 0 \\ \bar{y}_1 \\ \bar{y}_2 \\ \vdots \\ \bar{y}_k \end{pmatrix} - \begin{pmatrix} 0 \\ \hat{\beta}\,\bar{x}_1 \\ \hat{\beta}\,\bar{x}_2 \\ \vdots \\ \hat{\beta}\,\bar{x}_k \end{pmatrix} = \begin{pmatrix} 0 \\ \bar{y}_1 - \hat{\beta}\,\bar{x}_1 \\ \bar{y}_2 - \hat{\beta}\,\bar{x}_2 \\ \vdots \\ \bar{y}_k - \hat{\beta}\,\bar{x}_k \end{pmatrix} \tag{3.24}$$

$$E_{xx} = e_{xx} = \sum_{i=1}^{k}\sum_{j=1}^{n}(x_{ij} - \bar{x}_i)^2$$

$$e_{xy} = \sum_{ij}(x_{ij} - \bar{x}_i)(y_{ij} - \bar{y}_i) \qquad (3.25)$$

$$e_{yy} = \sum_{ij}(y_{ij} - \bar{y}_i)^2$$

olup, (3.15) yardımıyla β için tahmin edici

$$\hat{\beta} = \frac{e_{xy}}{e_{xx}} = \frac{\sum_{ij}(x_{ij} - \bar{x}_i)(y_{ij} - \bar{y}_i)}{\sum_{ij}(x_{ij} - \bar{x}_i)^2} \qquad (3.26)$$

bulunur. (3.18) , (3.19) ve (3.25) deki sonuç kullanılırsa;

$$SSE_{yx} = e_{yy} - e'_{xy}\, E_{xx}^{-1}\, e_{xy} = e_{yy} - \frac{e_{xy}^2}{e_{xx}}$$

$$= \sum_{ij}(y_{ij} - \bar{y}_i)^2 - \frac{\left[\sum_{ij}(x_{ij} - \bar{x}_i)(y_{ij} - \bar{y}_i)\right]^2}{\sum_{ij}(x_{ij} - \bar{x}_i)^2} \qquad (3.27)$$

hesaplanır. *k(n-1)-1* serbestlik derecesine sahiptir. β 'nın tahmin edilmesinden dolayı SSE_{yx} 'in serbestlik derecesi 1 azalmıştır.

$SSE_y = e_{yy}$ k(n-1) ve e_{xy}^2 / e_{xx} bir serbestlik derecesine sahiptir.

Kovaryans analizi kullanırken SSE_y 'den SSE_{yx} 'e olan azalmanın en azından serbestlik derecesi kaybını karşılamasını beklerler (Rencher 2001).

3.3.3. Hipotez Testi

Hipotez testlerini yaparken (3.23) 'teki ε_{ij}'lerin bağımsız olduklarını ve $N(0,\sigma^2)$ dağılımına sahip oldukları varsayılır. Öncelikle deneme etkilerinin eşitliği test edilmek istensin.

3.3.3a. Denemeler

$H_{01}: \alpha_1 = \alpha_2 = ... = \alpha_k$ test etmek için (kovariate için düzeltilmiş) tam ve indirgenmiş model yaklaşımını kullanacağız.

Tüm model (3.23) ve indirgenmiş model ($\alpha_1 = \alpha_2 = ... = \alpha_k = \alpha$) ise

$$y_{ij} = \mu + \alpha + \beta x_{ij} + \varepsilon_{ij}$$

$$y_{ij} = \mu^* + \beta x_{ij} + \varepsilon_{ij} \qquad i=1,2,...,k \quad j=1,2,...,n \tag{3.28}$$

Bu aslında basit regresyon modeli ile aynıdır.
İndirgenmiş model için SSE (SSE_{rd}) ile gösterilecektir ve aşağıdaki gibidir.

$$SSE_{rd} = \sum_{ij}(y_{ij} - \bar{y}_i)^2 - \frac{\left[\sum_{ij}(x_{ij} - \bar{x}_i)(y_{ij} - \bar{y}_i)\right]^2}{\sum_{ij}(x_{ij} - \bar{x}_i)^2} \tag{3.29}$$

kn-1-1=kn-2 serbestlik derecesine sahiptir.

H_{01}'i test etmek için kareler toplamı ifade edilirse;

$$SS(\alpha \mid \mu, \beta) = SS(\mu, \alpha, \beta) - SS(\mu, \beta)$$

$$SSE_{yx} = y'y - SS(\mu, \alpha, \beta)$$

$$SSE_{rd} = y'y - SS(\mu, \beta)$$

buradan;

$$SS(\alpha \mid \mu, \beta) = SSE_{rd} - SSE_{yx} \tag{3.30}$$

$kn\text{-}2 - [k(n-1)-1] = k\text{-}1$ serbestlik derecesine sahiptir.

$H_{01}: \alpha_1 = \alpha_2 = ... = \alpha_k$ için test istatistiği ise

$$F = \frac{SS(\alpha \mid \mu, \beta)/(k-1)}{SSE_{yx}/[k(n-1)-1]} \tag{3.31}$$

olup $F[(k-1), k(n-1)-1]$ dağılımına sahiptir. (H_{01} doğru iken)

(3.30)'dan

$SSE_{rd} = SS(\alpha \mid \mu, \beta) + SSE_{yx}$ bulunur.

$$SST_{yx} = SS(\alpha \mid \mu, \beta) + SSE_{yx} \tag{3.32}$$

(3.27) deki $SSE_{yx} = e_{yy} - \frac{e_{xy}^2}{e_{xx}}$ eşitlikten;

$$SST_{yx} = t_{yy} - \frac{t_{xy}^2}{t_{xx}} \tag{3.33}$$

$SST_{yx} = SSE_{rd}$

$$t_{yy} = \sum_{ij} (y_{ij} - \bar{y}_i)^2 \quad , \; t_{xy} = [\sum_{ij} (x_{ij} - \bar{x}_i)(y_{ij} - \bar{y}_i)] \quad , \; t_{xx} = \sum_{ij} (x_{ij} - \bar{x}_i)^2 \tag{3.34}$$

elde edilir(Rencher 2001).

$H_{01}: \alpha_1 = \alpha_2 = ... = \alpha_k$ için hesaplanan toplam kareler toplamı aşağıdaki tabloda verilmiştir.

Kaynak	Kovariate (Ortak Değişken)'e Göre Düzeltilmiş Kareler Toplamı	SD
Deneme	$SS\langle\alpha\,|\,\mu,\beta\rangle = SST_{yx} - SSE_{yx}$	*k-1*
Hata	$SSE_{yx} = e_{yy} - \frac{e^2{}_{xy}}{e_{xx}}$	*k(n-1)-1*
Genel	$SST_{yx} = t_{yy} - \frac{t^2{}_{xy}}{t_{xx}}$	*kn-2*

3.3.3b. Eğim

$H_{02} \;:\; \beta = 0$

Kovariateler önemli değildir hipotezi test edilir.

$$SSH = \frac{e^2{}_{xy}}{e_{xx}} \tag{3.35}$$

$H_{02} \;:\; \beta = 0$ hipotezinin kontrolü için kullanılacak F testi aşağıdaki gibi hesaplanır.

$$\mathrm{F_H} = \frac{e^2{}_{xy} / e_{xx}}{SSE_{yx} / \left[k(n-1)-1\right]} \tag{3.36}$$

$F_T\left[1, k(n-1)-1\right]$ tablo değeri ile karşılaştırıldığında $F_H < F_T$ ise H_{02} kabul edilir (Rencher 2001).

3.3.3c Eğimlerin Homojenlik Testi

Eğimlerin homojenliği;

$$H_{O3} : \beta_1 = \beta_2 = ... = \beta_k = 0$$

hipotezinde β_i i. gruptaki y ve x değişkenleri arasındaki regresyon doğrusunun eğimini göstermek üzere H_{03} hipotezi , k regresyon doğrusunun paralel olduğunu ifade etmektedir.bu hipotezi test ederken kullanılacak model;

$$y_{ij} = \mu + \alpha_i + \beta_i x_{ij} + \varepsilon_{ij} \qquad i=1,2,...,k \quad j=1,2,...,n \tag{3.38}$$

kn gözlemli matris formu $y = Z\alpha + X\beta + \varepsilon$ biçimindedir ve

$$X\beta = \begin{pmatrix} x_1 & 0 & \cdots & 0 \\ 0 & x_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & x_k \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} \tag{3.39}$$

olup,

$$\hat{\beta} = E_{xx}^{-1} e_{xy} = \left[\; X'(I-P)X \; \right]^{-1} X'(I-P)y$$

$$I - P = I - Z(Z'Z)^{-} Z'$$

$$= \begin{pmatrix} I - \frac{1}{n}J & 0 & \cdots & 0 \\ 0 & I - \frac{1}{n}J & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & I - \frac{1}{n}J \end{pmatrix}$$

dır. Buradan;

$$E_{xx} = X'(I-P)X$$

$$= \begin{pmatrix} x_1'\left(I-\frac{1}{n}J\right)x_1 & 0 & \cdots & 0 \\ 0 & x_2'\left(I-\frac{1}{n}J\right)x_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & x_k'\left(I-\frac{1}{n}J\right)x_k \end{pmatrix}$$

$$= \begin{pmatrix} \sum_j \left(x_{1j}-\bar{x}_1\right)^2 & 0 & \cdots & 0 \\ 0 & \sum_j \left(x_{2j}-\bar{x}_2\right)^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sum_j \left(x_{kj}-\bar{x}_k\right)^2 \end{pmatrix} \tag{3.40}$$

veya

$$= \begin{pmatrix} e_{xx,1} & 0 & \cdots & 0 \\ 0 & e_{xx,2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & e_{xx,k} \end{pmatrix} \tag{3.41}$$

olup

$$e_{xx,i} = \sum_j \left(x_{ij}-\bar{x}_{i.}\right)^2$$

dır. Burada,

$$e_{xy} = X'(I-P)y$$

$$= \begin{pmatrix} x_1' & 0 & \cdots & 0 \\ 0 & x_2' & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & x_k' \end{pmatrix} \begin{pmatrix} I-\frac{1}{n}J & 0 & \cdots & 0 \\ 0 & I-\frac{1}{n}J & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & I-\frac{1}{n}J \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \\ y_k \end{pmatrix}$$

$$= \begin{pmatrix} x_1'\left(I - \frac{1}{n}J\right)y_1 \\ x_2'\left(I - \frac{1}{n}J\right)y_2 \\ \vdots \\ x_k'\left(I - \frac{1}{n}J\right)y_k \end{pmatrix}$$

$$= \begin{pmatrix} \sum_j \left(x_{1j} - \bar{x}_1\right)\left(y_{1j} - \bar{y}_1\right) \\ \sum_j \left(x_{2j} - \bar{x}_2\right)\left(y_{2j} - \bar{y}_2\right) \\ \vdots \\ \sum_j \left(x_{kj} - \bar{x}_k\right)\left(y_{kj} - \bar{y}_k\right) \end{pmatrix} \tag{3.42}$$

$$= \begin{pmatrix} e_{xy,1} \\ e_{xy,2} \\ \\ e_{xy,k} \end{pmatrix} \tag{3.43}$$

ve

$$e_{xy,i} = \sum_j \left(x_{ij} - \bar{x}_i\right)\left(y_{ij} - \bar{y}_i\right)$$

olup,

$$\hat{\beta} = E_{xx}^{-1} e_{xy} = \begin{pmatrix} e_{xy,1}/e_{xx,1} \\ e_{xy,2}/e_{xx,2} \\ \vdots \\ e_{xy,k}/e_{xx,k} \end{pmatrix}$$

dır. Buradan,

$$SSE(F)_{y.x} = e_{yy} - e_{xy}' E_{xx}^{-1} e_{xy} = e_{yy} - e_{xy}' \hat{\beta}$$

$$= e_{yy} - (e_{xy,1}, e_{xy,2}, \ldots, e_{xy,k}) \begin{pmatrix} e_{xy,1}/e_{xx,1} \\ e_{xy,2}/e_{xx,2} \\ \vdots \\ e_{xy,k}/e_{xx,k} \end{pmatrix}$$

$$SSE(F)_{y.x} = e_{yy} - \sum_{i=1}^{k} \frac{e_{xy,i}^{\ 2}}{e_{xx,i}} \tag{3.44}$$

bulunur. *k(n-1) – k =k(n-2)* serbestlik derecesi hesaplanır.

$H_{O3} : \beta_1 = \beta_2 = \beta = 0$ hipotezi için $SSE(R)_{yx}$ aşağıdaki gibi hesaplanır.

$$SSE(R)_{yx} = e_{yy} - \frac{e_{xy}^{\ 2}}{e_{xx}} \tag{3.45}$$

sd=k(n-1)-1

H_{03} 'ü test etmek için kareler toplamı;

$$SSE(R)_{yx} - SSE(F)_{yx} = \sum_{i=1}^{k} \frac{e_{xy,i}^{\ 2}}{e_{xx,i}} - \frac{e_{xy}^{\ 2}}{e_{xx}} \tag{3.46}$$

ve

sd= k(n-1)-1-k(n-2) =k-1 'dir. Test istatistiği ;

$$F = \frac{\left[\sum_{i_1}^{k} e_{xy,i}^2 / e_{xx,i} - e_{xy}^2 / e_{xx}\right] / (k-1)}{SSE(F)_{y.x} / k(n-2)} \tag{3.47}$$

biçiminde hesaplanır. $F_t\left[(k-1), k(n-2)\right]$ tablo değeri ile karşılaştırılır.
$F_H < F_T$ ise H_{03} kabul edilir (Rencher 2001).

3.4. Tesadüfi Bloklar Düzeninde Kovaryans Analizi

Konu sayısı t ve tekrarlama sayısı r olan tesadüfi bloklar düzeninde,tek faktörlü bir deneme ve her deneme birimi üzerinde iki gözlem yapılmış olsun. Bu gözlem sonuçlarını bağımlı değişken için Y_{ij} ve bağımsız değişken için X_{ij} ile gösterilsin.

Böyle bir deneme için veri düzeni Tablo 5.'deki gibi verilebilir.

Tablo 5. Tek Faktörlü Tesadüfi Bloklar Deney Düzeninde Kovaryans Analizi İçin Veri Düzeni

Denemeler	Bloklar			
	Tek. 1	Tek.2	...	Tek r
M_1	X_{11} Y_{11}	X_{12} Y_{12}	...	X_{1r} Y_{1r}
M_2	X_{21} Y_{21}	X_{22} Y_{22}	...	X_{2r} Y_{2r}
⋮	⋮	⋮	⋮	⋮
M_t	X_{t1} Y_{t1}	X_{t2} Y_{t2}	...	X_{tr} Y_{tr}

Bu veri düzeni için *X* ve *Y*'ye ait Genel Kareler Toplamı ve Çarpımlar Toplamı aşağıdaki bağlantılarla hesaplanabilir (Düzgüneş 1963).

(i) Tüm veriler üzerinden *X* ve *Y*' ye ait genel kareler ve çarpımlar toplamı ;

$$\sum X^2 = \sum_{j=1}^{t}\sum_{i=1}^{n} X_{ij}{}^2 - \frac{X^2}{rt} \qquad (4.1)$$

$$\sum XY = \sum_{j=1}^{t}\sum_{i=1}^{n} X_{ij}Y_{ij} - \frac{(X)(Y)}{rt} \tag{4.2}$$

$$\sum Y^2 = \sum_{j=1}^{t}\sum_{i=1}^{n} Y_{ij}{}^2 - \frac{Y^2}{rt} \tag{4.3}$$

(ii) Denemeler için kareler ve çarpımlar toplamı ;

$$T_{XX} = \frac{\sum_{j=1}^{t} X_j{}^2}{r} - \frac{(X)^2}{rt} \tag{4.4}$$

$$T_{XY} = \frac{\sum_{j=1}^{t} X_j Y_j}{r} - \frac{(X)(Y)}{rt} \tag{4.5}$$

$$T_{XY} = \frac{\sum_{j=1}^{t} Y_j{}^2}{r} - \frac{(Y)^2}{rt} \tag{4.6}$$

(iii) Bloklar için kareler ve çarpımlar toplamı ;

$$R_{XX} = \frac{\sum_{i=1}^{n} X_i{}^2}{t} - \frac{(X)^2}{rt} \tag{4.7}$$

$$R_{XY} = \frac{\sum_{i=1}^{n} X_i Y_i}{t} - \frac{(X)(Y)}{rt} \tag{4.8}$$

$$R_{YY} = \frac{\sum_{i=1}^{n} Y_i^2}{t} - \frac{(Y)^2}{rt} \tag{4.9}$$

(iv) Hata için kareler ve çarpmalar toplamı;

$$E_{XX} = \sum X^2 - R_{XX} - T_{XX} \tag{5.0}$$

$$E_{XY} = \sum XY - R_{XY} - T_{XY} \tag{5.1}$$

$$E_{YY} = \sum Y^2 - R_{YY} - T_{YY} \tag{5.2}$$

(v) Denemelerle hataların toplamı ile oluşan 'Deneme+Hata' için kareler ve çarpımlar toplamı da;

$$S_{XX} = T_{XX} + E_{XX} \tag{5.3}$$

$$S_{XY} = T_{XY} + E_{XY} \tag{5.4}$$

$$S_{YY} = T_{YY} + E_{YY} \tag{5.5}$$

Bu bilgileri kullanarak bağımlı değişkene ait kareler toplamı, bağımsız değişken üzerindeki regresyon için düzeltilebilir. Buna göre regresyon katsayısı;

$$b = \frac{E_{XY}}{E_{XX}} \tag{5.6}$$

hata varyansı;

$$\hat{\sigma}_{YX}{}^2 = S_{YX}^2 = \frac{E_{YY} - \frac{E^2{}_{XY}}{E_{XX}}}{f_e} \tag{5.7}$$

Burada f_e , serbestlik derecesidir.

Tablo 5.'deki veri düzeni için kovaryans analizi tablosu Tablo 6. ile verildi.

Tablo 6. Tesadüfi Bloklar Deney Düzeninde Kovaryans Analizi

Kaynak	sd	Çarpım Toplamları			Sd	Düzeltilmiş $\sum Y^2$	KO
		XX	XY	YY			
Genel	$rt-1$	$\sum X^2$	$\sum XY$	$\sum Y^2$			
Bloklar	$r-1$	R_{XX}	R_{XY}	R_{YY}			
Denemeler	$t-1$	T_{XX}	T_{XY}	T_{YY}			
Hata	$(r-1)(t-1)$	E_{XX}	E_{XY}	E_{YY}	$(r-1)(t-1)$-1		S_{YX}^2
Denemeler + Hata	$r(t-1)$	S_{XX}	S_{XY}	S_{YY}	$r(t-1)$-1	$S_{YY} - \frac{S_{XY}{}^2}{S_{XX}} = B$	
Düzeltilmiş Denemeler					$t-1$	$B - A = C$	$\frac{C}{t-1}$

Bu formüller vasıtasıyla hesaplanacak kareler ve çarpımlar toplamları, kovaryans analizi için olduğu kadar *X* ve *Y* nin varyans analizleri içinde gerekli değerleri sağlamaktadır.

"Düzeltilmemiş bağımlı değişken için deneme ortalamaları arasında fark yoktur" şeklindeki hipotezin kontrolü için test istatistiği;

$$F = \frac{T_{YY}/t-1}{E_{XX}/(r-1)(t-1)} \tag{5.8}$$

şeklinde tanımlanan bir $F_{[\,t-1;(r-1)(t-1)\,]}$ istatistiğidir. Benzer şekilde "bağımsız değişken için deneme ortalamaları arasında fark yoktur" bir başka ifadeyle "bağımsız değişkenler deneme konularından etkilenmemektedir" şeklindeki hipotezin kontrolü için test istatistiği de;

$$F = \frac{T_{XX}/t-1}{E_{XX}/(r-1)(t-1)} \tag{5.9}$$

formülü ile verilen bir $F_{[\,t-1;(r-1)(t-1)\,]}$ istatistiğidir. Bunun önemli çıkması genellikle bağımsız değişkenlerin denemelerden etkilendiğini gösterdiğinden, bu gibi durumlarda bağımlı değişken ortalamalarının ilgili bağımsız değişkenlere göre bir düzeltmeye tabii tutulması deneme etkilerinin bir kısmını ortadan kaldırmaktadır. Bu nedenle kovaryans analizinden önce böyle bir istatistik kontrolünün yapılmasında yarar vardır (Düzgüneş 1963).

Düzeltilmiş deneme ortalamaları arasında fark olmadığına dair bir hipotezin kontrolü için her şeyden önce bağımsız değişkenlere göre düzeltilmiş bağımlı değişkenlere ait "Hata" ve "Denemeler+Hata" kareler toplamlarının hesaplanması gerekir. Çünkü Tablo 6.'dan da görüleceği gibi düzeltilmiş deneme ortalamaları için kareler toplamı bu ikisi arasındaki farktan bulunmaktadır. Bunun *t-1* serbestlik derecesine bölünmesi ise düzeltilmiş deneme ortalamaları için kareler ortalamasını vermektedir. Kareler ortalamasının hata varyansı S_{YX}^2'e bölünmesinden de söz konusu hipotezin kontrolü için gerekli F istatistiği elde edilmektedir. Yani;

$$F = \frac{Düzeltilmiş\ \ deneme\ ortalamaları}{S^2_{YX}} \qquad (6.0)$$

şeklinde bulunur ve (r-1)(t-1)-1 ile t-1 serbestlik derecelerine göre kontrol edilir.

F istatistiğinin değeri anlamlı bulunduğu taktirde, bu durum bağımlı değişkene ait deneme ortalamaları arasında gerçekten fark olduğu, başka bir ifadeyle, denemelerden dolayı varyansın önemli olduğunu gösterir. Eğer düzeltilmiş deneme ortalamaları değil de, düzeltilmemiş ortalamaları arasındaki farklar önemli bulunacak olursa, düzeltilmemiş ortalamalar arasındaki farka denemelerin değil sadece bağımsız değişkenlerin etki yaptığı söylenir.

Deneme Ortalamalarının Düzeltilmesi

Düzeltilmiş deneme ortamaları, bütün bağımsız değişken ortalamaları birbirine eşit olduğunda elde edilecek bağımlı değişken ortalamalarının tahmini ölçüleridir.

Deneme ortalamalarının düzeltilmesi,

$$\overline{Y}_j' = \overline{Y}_j - b(\overline{X}_j - \overline{X}) \qquad (j=1,2,...,t) \qquad (6.1)$$

eşitliği kullanılarak yapılmaktadır.

Burada;

$\overline{Y}_j'$: *j.* denemeye ait düzeltilmiş bağımlı değişken ortalaması

$\overline{Y}_j$: *j.* denemeye ait düzeltilmemiş bağımlı değişken ortalaması

b : Bağımlı değişken ile bağımsız değişken arasındaki doğrusal regresyon katsayısı

$\overline{X}_j$: *j.* Denemeye ait bağımsız değişken ortalaması

$\overline{X}$: Bağımsız değişkenin genel ortalaması

dır.

4. UYGULAMA

4.1 Uygulama I

Tesadüfi Bloklar Düzeninde Kovaryans Analizi için Uygulama

Araştırma materyali olarak Tablo 8.' deki veriler kullanılmıştır.

Bu tabloda;

Y : Kabuklu fındığın iç randımanı (%kg)

X : Yıllık sıcaklık ortalaması (^0C)

Kabuklu fındığın iç randımanına ait veriler Giresun Fiskobirlik Genel Müdürlüğü'nden, yıllık sıcaklık ortalaması oranlarına ait veriler ise Ankara Meteoroloji Genel Müdürlüğü'nden alınmıştır (Yurtsever 1984).

Tablo.8, 14 bloktan ve 4 ayrı denemeden oluşmaktadır. Fiskobirlik Genel Müdürlüğü, fındık alımlarını gerçekleştirdiği kooperatifleri dört bölgeye ayırmıştır.Tablo.8' deki verilerde, bu 4 bölgenin dikkate alındığı denemelerden oluşmaktadır. Aynı zamanda bu tablodaki Y, X değişkenlerine ait veriler elde edilirken, her bir bölge için o bölgede yer alan illerin toplam değerlerinin ortalaması alınmıştır.

Tablo 8. Kovaryans Analizi İçin Veri Tablosu

Denemeler	Doğu Bölgesi		Giresun Bölgesi		Ordu Bölgesi		Batı Bölgesi	
Bloklar	Y	X	Y	X	Y	X	Y	X
B1	43,03	13,3	47,64	14,2	49,05	13,7	47,64	12,5
B2	46,18	12,9	50,03	13,9	50,38	13,6	47,49	12,4
B3	45,49	13,6	48,9	14,2	50,2	14	46,33	12,6
B4	45,3	12,8	46,53	13,4	50,11	13,4	48,1	12,2
B5	46,57	13,1	50,56	13,9	49,75	14	48,68	12,5
B6	40,67	13,4	49,03	14,2	48,28	14,4	48,8	12,7
B7	44,56	13,2	49,18	14,2	49,17	14,1	48,53	12,6
B8	46,19	13,4	50,7	14,1	49,08	14	48,23	12,2
B9	45,36	12,2	48,68	13,3	48,48	13,2	48,63	11,6
B10	47,37	12,5	52,05	13,4	50,95	13,3	46,02	12,1
B11	44,85	13,9	46,1	14,7	41,6	14,1	37,46	13,6
B12	45,46	13,9	51,84	14,6	50,35	14,6	50,09	13
B13	44,81	14	48,95	14,4	48,47	14,3	47,53	12,7
B14	45,2	13,2	50,34	13,9	49,24	13,8	47,47	12,1

Bu tabloda bloklar, fındık üretimini gerçekleştiği alanları ifade etmektedir.

Araştırmada öncelikle, kabuklu fındığın iç randımanı açısından bölgeler arasında farklılık olup olmadığı varyans analizi ile incelenmiştir..Elde edilen sonuçlar (Tablo 9)'da verilmiştir.

Tablo 9.

Dependent Variable: y

Kaynak	Kareler toplami	Serbestlik derecesi	Kareler ortalamasi	F	p degeri
Duzeltilmis model	157,907[a]	3	52,636	10,613	,000
Sabit	127080,939	1	127080,939	25623,459	,000
bolge	157,907	3	52,636	10,613	,000
hata	257,897	52	4,960		
Toplam	127496,743	56			
duzeltilmis toplam	415,804	55			

a. Belirleme Katsayisi= ,380 (Duzeltilmis Belirleme Katsayisi = ,344)

Kabuklu fındığın iç randımanı açısından bölgeler arasında %5 anlam seviyesinde istatistiksel olarak anlamlı bir fark bulunmuştur. Farklılığın hangi bölgeler arasında bulunduğunu belirlemek amacıyla çoklu karşılaştırma testi yapılıp,sonuçlar Tablo.10'da verilmiştir.

Tablo 10.

Çoklu Karsilastirmalar

Dependent Variable: y

LSD

(I) bolge	(J) bolge	Ortalamala r Farki	Std. Hata	P	%95 Guven Araligi	
					Alt sinir	Ust sinir
1,00	2,00	-4,2493*	,84173	,000	-5,9383	-2,5602
	3,00	-3,8621*	,84173	,000	-5,5512	-2,1731
	4,00	-2,1402*	,84173	,014	-3,8293	-,4512
2,00	1,00	4,2493*	,84173	,000	2,5602	5,9383
	3,00	,3871	,84173	,647	-1,3019	2,0762
	4,00	2,1091*	,84173	,015	,4200	3,7981
3,00	1,00	3,8621*	,84173	,000	2,1731	5,5512
	2,00	-,3871	,84173	,647	-2,0762	1,3019
	4,00	1,7219*	,84173	,046	,0329	3,4110
4,00	1,00	2,1402*	,84173	,014	,4512	3,8293
	2,00	-2,1091*	,84173	,015	-3,7981	-,4200
	3,00	-1,7219*	,84173	,046	-3,4110	-,0329

*. anlamli farkliliklar %5 seviyesinde

Kabuklu fındığın iç randımanı açısından bakıldığında, Giresun bölgesi ve Ordu bölgesi hariç diğer tüm bölgeler arasında %5 önem seviyesinde istatistiksel olarak anlamlı bir farklılık bulunmuştur.

Aynı zamanda bölgeler arası faklılığa sebebep olduğu düşünülen, yıllık sıcaklık ortalamasınında bölgelere göre farklılık gösterip göstermediği varyans analizi ile incelenerek elde edilen sonuçlar Tablo.11'de verilmiştir.

Tablo 11.

Dependent Variable: sicaklik

Kaynak	Kareler Toplami	Serbestlik Derecesi	Kareler Ortalamasi	F	P
Düzeltilmis Model	20,972[a]	3	6,991	32,663	,000
Sabit	10074,129	1	10074,129	47069,930	,000
bolge	20,972	3	6,991	32,663	,000
Hata	11,129	52	,214		
Toplam	10106,230	56			
Düzeltilmis Toplam	32,101	55			

a. Belirleme katsayisi = ,653 (Düzeltilmis belirleme katsayisi = ,633)

Yılık sıcaklık ortalaması açısından bakıldığında bölgeler arasında istatistiksel olarak anlamlı farklılık bulunmuştur.Bu farklılığı yaratan bölgelei tespit etmek amacıyla çoklu karşılaştırma yapılarak, elde edilen sonuçlar Tablo.12'de verilmiştir.

Tablo 12.

Çoklu Karsilastirmalar

Dependent Variable: sicaklik

LSD

(I) bolge	(J) bolge	Ortalamalar Farki(I-J)	Std. Hata	P	%95 Güven araligi	
					Alt sinir	Üst Sinir
1,00	2,00	-,7857*	,17486	,000	-1,1366	-,4348
	3,00	-,6500*	,17486	,000	-1,0009	-,2991
	4,00	,7571*	,17486	,000	,4063	1,1080
2,00	1,00	,7857*	,17486	,000	,4348	1,1366
	3,00	,1357	,17486	,441	-,2152	,4866
	4,00	1,5429*	,17486	,000	1,1920	1,8937
3,00	1,00	,6500*	,17486	,000	,2991	1,0009
	2,00	-,1357	,17486	,441	-,4866	,2152
	4,00	1,4071*	,17486	,000	1,0563	1,7580
4,00	1,00	-,7571*	,17486	,000	-1,1080	-,4063
	2,00	-1,5429*	,17486	,000	-1,8937	-1,1920
	3,00	-1,4071*	,17486	,000	-1,7580	-1,0563

*. Anlamli farkliliklar %5 anlam seviyesinde

Yıllık sıcaklık ortalaması açısından, Giresun ve Batı bölgesi hariç diğer tüm bölgelerde istatistiksel olarak anlamlı bir fark bulunmuştur.

Kabuklu fındığın iç randımanının, bölgelere göre farklılık göstermesine etki ettiği düşünülen yıllık sıcaklık ortalaması dikkate alınarak kovaryans analizi yapıldığında elde edilen sonuçlar Tablo.13'de verilmiştir.

Tablo 13.

Dependent Variable: y

kaynak	Kareler toplami	Serbestlik Derecesi	Kareler Ortalamasi	F	P degeri
Duzeltilmis model	201,892[a]	7	28,842	6,472	,000
Sabit	269,964	1	269,964	60,578	,000
bolge	15,167	3	5,056	1,134	,345
sicaklik	25,449	1	25,449	5,710	,021
bolge * sicaklik	16,836	3	5,612	1,259	,299
Hata	213,912	48	4,456		
Toplam	127496,743	56			
Duzeltilmis toplam	415,804	55			

a. Belirleme Katsayisi = ,486 (Duzeltilmis belirleme Katsayisi = ,411)

Kabuklu fındığın iç randımanının, bölgelere göre farklılık göstermesinde yıllık sıcaklık ortalaması istatistiksel olarak anlamlı bir farklılık yaratmaktadır.

Tablo 9.'da kabuklu fındığın iç randımanı açısından bölgeler arasında farklılık olup olmadığı varyans analizi ile incelendiğinde, elde edilen hata kareler toplamı 257,897 iken; tesadüfi bloklar düzeninde kovaryans analizi uygulaması ile elde edilen hata karaler toplamı Tablo 13.'de görüldüğü gibi 213,912 bulunmuştur. Tesadüfi bloklar düzeninde kovaryans analizi uygulaması ile deney hatası azalmıştır.Böylece analizin güvenilirliği artmıştır.

4.2. Uygulama II

Bir Kovariate (Ortak Değişken)'li Kovaryans Analizi için Uygulama

Bu çalışmada, Konya Başkent Üni. Kardiyoloji servisinde, normal sigara içen, light sigara içen ve sigara içmeyen, üç grup, toplam 45 hasta üzerinden elde edilen, MI geçirme riskinin bir belirteci olan başlangıç CFR değerlerinin ve light sigara içimi sonrası oluşan CFR değerlerinin, Kovaryans analizi ile çözümlemesi yapılmıştır.

VERİLER

Normal sigara içen		Sigara içmeyen		Light sigara içen	
x1	y1	x2	y2	x3	y3
29	31	31	33	19	27
37	24	17	18	24	27
22	24	20	24	24	24
29	24	21	35	18	18
24	22	40	45	19	30
26	20	26	30	34	36
21	31	20	24	28	37
21	25	21	23	29	30
24	26	24	26	24	28
39	24	23	24	30	43
22	24	19	20	21	28
22	24	20	35	18	27
24	24	24	28	18	24
24	28	34	35	20	22
24	27	25	27	24	28

X = Hastaların başlangıçtaki CFR değerleri

Y = Light sigara içiminden sonra hastalarda oluşan yeni CFR değerleri

TEK YÖNLÜ VARYANS ANALİZİ SONUÇLARI BAŞLANGIÇ CFR DEĞERLERİ İÇİN

TANIMLAYICI ISTATISTIKLER

X

	N	ORTALAMA	STD SAPMA	STD HATA	ORTALAMA ICIN %95 GUVEN ARALIGI		Minimum	Maximum
					ALT SINIR	UST SINIR		
1,00	15	25,8667	5,51448	,42383	22,8128	28,9205	21,00	39,00
2,00	15	24,3333	6,26403	,61737	20,8644	27,8022	17,00	40,00
3,00	15	23,3333	5,02375	,29713	20,5513	26,1154	18,00	34,00
toplam	45	24,5111	5,59473	,83401	22,8303	26,1920	17,00	40,00

VARYANS ANALIZI TABLOSU

X

	KARELER TOPLAMI	SD	KARELER ORTALAMASI	F	P
GRUPLAR ARASI	48,844	2	24,422	,772	,468
GRUP ICI	1328,400	42	31,629		
GENEL	1377,244	44			

Başlangıç CFR değerleri açısından üç grup ortalaması arasında %5 anlam seviyesinde istatistiksel bir fark olmadığı görülmektedir (p=0,468>0,05).

TEK YÖNLÜ VARYANS ANALİZİ LİGHT SİGARA İÇİMİNDEN SONRAKİ CFR DEĞERLERİ İÇİN

TANIMLAYICI ISTATISTIKLER

Y

	N	ORTALAMA	STD SAPMA	STD HATA	ORTALAMA ICIN %95 GÜVEN ARALIGI ALT SINIR	ORTALAMA ICIN %95 GÜVEN ARALIGI ÜST SINIR	Minimum	Maximum
1,00	15	25,2000	3,00476	,77583	23,536	26,8640	20,00	31,00
2,00	15	28,4667	7,08990	1,83060	24,540	32,3929	18,00	45,00
3,00	15	28,6000	6,23126	1,60890	25,149	32,0508	18,00	43,00
TOPLAM	45	27,4222	5,80943	,86602	25,677	29,1676	18,00	45,00

VARYANS ANALIZI TABLOSU

Y

	KARELER TOPLAMI	SD	KARELER ORTALAMASI	F	P
GRUPLAR ARASI	111,244	2	55,622	1,701	,195
GRUP ICI	1373,733	42	32,708		
GENEL	1484,978	44			

Light sigara içiminden sonraki CFR değerleri açısından üç grup ortalaması arasında %5 anlam seviyesinde istatistiksel bir fark olmadığı görülmektedir (p=0,195>0,05).

Modelimiz;

$y_{ij} = \mu + \alpha_i + \beta x_{ij} + \varepsilon_{ij}$ i= 1,2,3 j= 1,2,...,15

μ : y_{ij} 'lerin ortalamalarıdır.

α_i : Deneme etkisi

x_{ij} : y_{ij} 'ye bağlı kovariate gözlem değeri

β : x'in y'ye göre eğimi

Öncelikle düzeltilmiş kareler toplamı hesabı yapılarak aşağıdaki sonuçlar hesaplanır;

$$e_{xx} = \sum_{i=1}^{k} \sum_{j=1}^{n} (x_{ij} - \bar{x}_i)^2$$

$$= 1328,4$$

$$e_{xy} = \sum_{i=1}^{k} \sum_{j=1}^{n} (x_{ij} - \bar{x}_i)(y_{ij} - \bar{y}_i)$$

$$= 790,0667$$

$$e_{yy} = \sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij} - \bar{y}_i)^2$$

$$= 1373,733$$

Bu hesaplamalardan sonra modeldeki kovariatelerin etkisi tahmin edilir.

$$\hat{\beta} = \frac{e_{xy}}{e_{xx}}$$
$$= 0,59$$

Düzeltilmiş xy hata kareler toplamı hesaplanır.

$$SSE_{yx} = e_{yy} - \frac{e^2{}_{xy}}{e_{xx}}$$ *sd= k(n-1)-1* *sd*=41

$= 903,84$

Adım I:

H_{01} : $\alpha_1 = \alpha_2 = \alpha_3$

H_{11} : En az bir α diğerlerinden farklıdır.

Her bir grup için ; Kovariate göre düzeltilmiş deneme ortalamalarının eşit olup olmadığı hipotezi test edilir.

$$SSE_{rd} = \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_i)^2 - \frac{\left[\sum_{i=1}^{k}\sum_{j=1}^{n}(x_{ij} - \bar{x}_i)(y_{ij} - \bar{y}_i)\right]^2}{\sum_{i=1}^{k}\sum_{j=1}^{n}(x_{ij} - \bar{x}_i)^2}$$

sd= kn-1-1 =kn-2

$SSE_{rd} = 520257,8$ sd= 40

$$SS\langle\alpha|\mu,\beta\rangle = SSE_{rd} - SSE_{yx}$$

$$F_H = \frac{SS\langle\alpha|\mu,\beta\rangle/(k-1)}{SSE_{yx}/[k(n-1)-1]}$$

$F_H = 11779,46$

$F_T = F[(k-1)/k(n-1)-1]$

$F_T\ (3,41) = 2,84$

$F_H > F_T$ olduğu için H_{01} hipotezi red edilir.Yani; deneme ortalamaları eşit değildir.

SSE_{rd} ,genel kareler yoplamının bir fonksiyonu olduğu için düzeltme yapıldıktan sonra SSE_{rd} 'yi SST_{yx} olarak gösterebiliriz.

$$SST_{yx} = SS\langle\alpha|\mu,\beta\rangle + SSE_{yx}$$

$$SST_{yx} = t_{yy} - \frac{t^2_{xy}}{t_{xx}}$$

Tek kovariate olan tek yönlü modelde $H_{01} : \alpha_1 = \alpha_2 = ... = \alpha_k$ hipotezinin testi için hazırlanan kovaryans analizi tablosu aşağıdaki gibidir.

Kaynak	Kovariate (Ortak değişken)'e Göre Düzeltilmiş Kareler Toplamı	SD
Deneme	$SS\langle\alpha\|\mu,\beta\rangle = SST_{yx} - SSE_{yx}$	*k-1*
Hata	$SSE_{yx} = e_{yy} - \frac{e^2_{xy}}{e_{xx}}$	*k(n-1)-1*
Genel	$SST_{yx} = t_{yy} - \frac{t^2_{xy}}{t_{xx}}$	*kn-2*

Kaynak	Kovariate (Ortak değişken)'e Göre Düzeltilmiş Kareler Toplamı	SD
Deneme	$SS\langle\alpha\|\mu,\beta\rangle = 519354$	2
Hata	$SSE_{yx} = 903{,}84$	41
Genel	$SST_{yx} = 520257{,}8$	43

ADIM 2:

$H_{02} : \beta = 0$
$H_{12} : \beta \neq 0$

Kovariateler önemli değildir hipotezi test edilir.

$$SSH = \frac{e^2{}_{xy}}{e_{xx}}$$
$$= 469{,}89$$

$$F_H = \frac{e^2{}_{xy} / e_{xx}}{SSE_{yx} / [k(n-1)-1]}$$

$$=21{,}32$$

$F_T = F(1,41) = 4{,}0$

$F_H > F_T$ olduğu için H_{02} hipotezi red edilir.Yani; Covariateler önemlidir.

ADIM 3 :

Eğimlerin homojenliği;

$H_{O3} : \beta_1 = \beta_2 = \beta = 0$

H_{13} : Eğimler homojen değildir.

hipotezinde β_i i. gruptaki y ve x değişkenleri arasındaki regresyon doğrusunun eğimini göstermek üzere H_{03} hipotezi 3 regresyon doğrusunun paralel olduğunu ifade etmektedir.bu hipotezi test ederken;

$$SSE(F)_{y.x} = e_{yy} - \sum_{i=1}^{k} \frac{e_{xy,i}{}^2}{e_{xx,i}}$$

$$e_{xx,i} = \sum_j (x_{ij} - x_{i.})^2 \quad \text{ve} \quad e_{xy,i} = \sum_j (x_{ij} - x_{i.})(y_{ij} - y_{i.})$$

$e_{xx,1} = 1460,34$ $\quad e_{xy,1} = 585,4$

$e_{xx,2} = 1378,72$ $\quad e_{xy,2} = 712,75$

$e_{xx,3} = 1440,02$ $\quad e_{xy,3} = 658,69$

$$SSE(F)_{yx} = e_{yy} - \sum e_{xx,i} \rightarrow SSE(F)_{yx} = 469,3$$

$$SSE(R)_{yx} = e_{yy} - \frac{e_{xy}^{\ 2}}{e_{xx}} \rightarrow SSE(R)_{yx} = 1373,73 - 469,89$$
$$= 903,84$$

H_{03}'ü test etmek için kareler toplamı;

$$SSE(R)_{yx} - SSE(F)_{yx} = 903,84 - 469,3$$
$$= 434,54$$

olup, F değeri;

$$F = \frac{\frac{434,54}{2}}{\frac{469,3}{39}} = \frac{217,27}{12,03} = 18,06$$ olup $F_{[2,39]} = 3,23$ ile karşılaştırılırsa $F_h > F_t$ olduğundan en az bir grubun eğiminin diğerlerinden farklı olduğu görülür.

Light sigara içimi , normal sigara –light sigara ve sigara içmeyen gruplarındaki CFR değerlerini etkilemektedir.

5. SONUÇ

Kovaryans analizi; regresyon analizi ile varyans analizinin bir kombinasyonu olup, istatistik bilim dalının önemli yöntemlerinden birirdir.

Kovaryans analizi deneysel çalışma sonuçlarını açıklamada büyük avantajlar ve kolaylıklar sağlamaktadır. Yöntem deneysel çalışmalardaki deney hatasını azaltmak için geliştirilmiştir. Modelin güçlü varsayımları olması nedeniyle sonuçlardaki duyarlılık daha fazladır.

Çalışmamızdan elde edilen sonuçları şu şekilde sıralayabiliriz:

- Tesadüfi bloklar düzeninde kovaryans analizi uygulaması, varyans analizine kıyasla deney hatasını azaltmaktadır. Bunun nedeni; bağımlı değişken ile birlikte değişen ortak değişkenin etkisinin kontrol altına alınmasıdır. Bu sayade analizin güvenilirliği artmaktadır.

- Birinci uygulamada; kabuklu fındığın iç randımanına ait veriler Giresun Fiskobirlik Genel Müdürlüğü'nden, yıllık sıcaklık ortalaması oranlarına ait veriler ise Ankara Meteoroloji Genel Müdürlüğü'nden alınmıştır (Yurtsever 1994).

Bu uygulamada öncelikle, kabuklu fındığın iç randımanı açısından bölgeler arasında farklılık olup olmadığı varyans analizi ile incelenmiştir. Aynı zamanda bu farklılığa sebep olduğu düşünülen yıllık sıcaklık ortalamasının bölgelere göre farklılık gösterip göstermediği varyans analizi ile kontrol edilmiştir. Daha sonra kabuklu fındığın iç randımanının, bölgelere göre farklılık göstermesine etki ettiği düşünülen yıllık sıcaklık ortalaması dikkate alınarak kovaryans analizi yapılmıştır. Yıllık sıcaklık ortalamasının yarattığı farklılık istatistiksel olarak anlamlı bulunmuştur.

- İkinci uygulamada; Konya Başkent Üni. Kardiyoloji servisinde, normal sigara içen, light sigara içen ve sigara içmeyen, üç grup, toplam 45 hasta üzerinden elde edilen, MI geçirme riskinin bir belirteci olan başlangıç CFR(Coronary Flow

Rezerve) değerlerinin ve light sigara içimi sonrası oluşan CFR değerlerinin, kovaryans analizi ile çözümlemesi yapılmıştır. Bunun için öncelikle regresyon analizi uygulanıp, daha sonrada düzeltilmiş değerler üzerinden varyans analizi uygulanmıştır. Böylelikle bağımlı değişken ile ortak değişken arasındaki doğrusal ilşki için bir düzeltme yapılmış oldu ve hata varyansı düştü. Kullanılan F istatistiğinde; Light sigara içimi , normal sigara –light sigara ve sigara içmeyen gruplarındaki CFR değerlerini etkilediği sonucuna varılmıştır.

KAYNAKLAR

1- Akgül, A., 2005. Tıbbi Araştırmalarda İstatistiksel Analiz Teknikleri, Emek Ofset, Ankara.

2- Alpar, R., 1997. Uygulamalı Çok Değişkenli İsatistiksel Yöntemlere Giriş-1, Bağırgam, Ankara.

3- Çelik, C., 1997. İstatistiksel Uygulamalarda Kovaryans Analizi ve Hipertansif, Sağlıklı Gebe Kadınlarda Yaş, Ortalama Arteryel Basınç Değişkenlerinin Değerlendirilmesi, Yüksek Lisans Tezi, D.Ü. Fen Bilimleri Enstitüsü, Diyarbakır.

4- Çömlekçi, N., 1994. Temel İstatistik, Bilim Teknik Yayınevi, İstanbul.

5- Demirsoyi, Ş., 1993. Kuzularda Büyümeyi Etkileyen Çevresel Faktörlerin Kovaryans Analizi ile İncelenmesi, Yüksek Lisans Tezi, A.Ü. Sağlık Bilimleri Enstitüsü, Ankara.

6- Düzgüneş, O., 1963. Bilimsel Araştırmalarda İstatistik Prensipleri ve Metotları, Ege Üniversitesi Matbaası, İzmir.

7- Graybill, F.A., 1961. An Introduction to Linear Statistical Models, McGraw-Hill, New York.

8- Muluk, Z., Toktamış, Ö.,Kurt, S., 1994. Deney Düzenlemede İstatistiksel Yöntemler, Ege Üniversitesi Basımevi, İzmir.

9- Myers, R.H., 1990. Classical and Modern Regression With Application, Duxbury Pres, California.

10- Neter, J.,Kutneri, M. H., Nachtsheim, C. J., Wasserman, W., 1996. Applied Linear Statistical Models, WCB McGraw-Hill Pres, Boston.

11- Özdamar, K., 1999. Paket Programları ile İstatistiksel Veri Analizi, Kaan Kitabevi, Eskişehir.

12- Rawlings, J.O., Pantula,S.G., Dickey, D.A., 1998. Applied Regression Analysis, Springer Werlag, New York.

13- Rencher, A. 2001. Lineer Models in Statistics, John Wıley & Sons, New York.

14- Tarı, R., 1999. Ekonometri, Alfa Yayınları, İstanbul.

15- Yurtsever, N., 1974. İstatistik Metotları III, Topraksu Genel Müdürlüğü Yayınları, Ankara.