EKONOMETRİ I

EKONOMETRİ PROGRAMI

PROF. DR. NİLGÜN ÇİL

İSTANBUL ÜNİVERSİTESİ AÇIK VE UZAKTAN EĞİTİM FAKÜLTESİ
İSTANBUL ÜNİVERSİTESİ AÇIK VE UZAKTAN EĞİTİM FAKÜLTESİ

EKONOMETRİ PROGRAMI

EKONOMETRİ I

PROF. DR. NİLGÜN ÇİL
Yazar Notu

Elinizdeki bu eser, İstanbul Üniversitesi Açık ve Uzaktan Eğitim Fakültesi’nde okutulmak için hazırlananmış bir ders notu niteliğindedir.
ÖNSÖZ
İÇINDEKİLER

ÖNSÖZ........................................................................................................................................... I
KISALTMALAR .................................................................................................................................. V
YAZAR NOTU .................................................................................................................................... VI

1. EKONOMETRİNİN TANIMI AMACI VE METODOLOJİSİ .......................................................... 1
   1.1. Ekonometri Nedir? Ekonometrinin Gerekliliği ................................................................. 7
   1.2. Ekonometrinin Amacı ve Metodolojisi .............................................................................. 8
   1.3. Ekonometrik Model .......................................................................................................... 11
   1.4. Ekonometride Kullanılan Veriler ..................................................................................... 12
   1.5. Ekonometride Bilgisayarın, Paket Programların Yeri .................................................. 14

2. REGRESYON ANALİZINE GİRİŞ .......................................................................................... 20
   2.1. Basit ve Çok Değişkenli Regresyon Modelleri ................................................................. 26
   2.2. Regresyon Analizi ........................................................................................................... 27
   2.3. Anakütle Regresyon Modeli ........................................................................................... 28
   2.4. Örnek Kütle Regresyon Modeli ....................................................................................... 35
   2.5. Hata Teriminin Kaynakları ............................................................................................. 38

3. KLASİK DOĞRUSAL REGRESYON MODELİNİN TEMEL VARSAYIMLARI .......... 51

4. ANA KÜTLE REGRESYON MODELİNİN TAHMİNİ: EN KÜÇÜK KARELER YÖNTEMI .... 73
   4.1. En Küçük Kareler Yöntemi ............................................................................................ 79
   4.2. En Küçük Kareler Yönteminin Diğer Bir Uygulama Biçimi: Sapmalar Yöntemi ...... 81
   4.3. En Küçük Kareler Yönteminin Özellikleri ................................................................... 90
   4.4. Uygulama ....................................................................................................................... 93

5. EN KÜÇÜK KARELER TAHMİNCİLERİNIN STANDART HATALARI VE REGRESYON DOĞRUSUNUN UYGUNLUĞU ................................................................. 102
   5.1. En Küçük Kareler Tahmincilerinin Varyans ve Kovaryansı ........................................... 108
   5.2. Regresyon Doğrusunun Verilere Uygunluğu ................................................................. 112
      5.2.1. Belirginlik Katsayısı ................................................................................................. 112
      5.2.2. Tahminin Standart Hatası ...................................................................................... 117
      5.2.3. Uyum Katsayısı ve Uygunluk Katsayları ............................................................... 117

6. ELASTİKİYET VE HATA TERİMLERİNİN NORMAL DAĞILIM VARSAYIMI ..... 125
   6.1. Elastikiyet ....................................................................................................................... 131
   6.2. Hata Terimlerinin Normal Dağılımı ............................................................................ 131
7. TAHMİN EDİLEN PARAMETRELER İLE İLGİLİ HİPOTEZ TESTLERİ ve GÜVEN ARALIKLARI ....................................................................................................................... 141
7.1. Aralık Tahmini ve Hipotez Testleri ............................................................................................................................ 147
7.2. $\beta_0$ ve $\beta_1$ için Hipotez Testi ............................................................................................................................ 147
7.3. $\beta_0$ ve $\beta_1$ için Güven Aralığı ............................................................................................................................ 152
7.4. Ana Kütle Hata Teriminin Varyansının Güven Aralığı ........................................................................................................ 153
7.5. Ana Kütle Hata Teriminin Varyansı İçin Ki-kare $\chi^2$ Testi ........................................................................................................ 155
8. EN KÜÇÜK KARELER TAHMİN EDİCİLERİN ÖZELLİKLERİ ....................................................................................................................... 164
8.1. Küçük Örnek Özellikleri ............................................................................................................................ 170
8.1.1. Doğrusallık .................................................................................................................................. 170
8.1.2. Sapmasız (Eğilimsiz, Sistematik Hatasız) ........................................................................................................ 170
8.1.3. En Küçük Varyans ............................................................................................................................ 171
8.1.4. Etkinlik .................................................................................................................................. 172
8.1.5. Doğrusal En İyi Sapmasızlık .................................................................................................................... 172
8.1.6. En Küçük Orta Karesi Kareli Hata ........................................................................................................ 173
8.1.7. Yeterlilik .................................................................................................................................. 173
8.2. Büyük Örnek Özellikleri ............................................................................................................................ 173
8.2.1. Asimptotik Sapmasızlık ........................................................................................................................ 173
8.2.2. Tutarlılık .................................................................................................................................. 174
8.2.3. Asimptotik Etkinlik ........................................................................................................................ 174
9. ORİJİNDEN GEÇEN REGRESYON VE DOĞRUSAL OLMAYAN MODELLER ....................................................................................................................... 180
9.1. Orijinden Geçen Regresyon ............................................................................................................................ 186
9.2. Doğrusal Olmayan Modeller ............................................................................................................................ 189
9.2.1. Parametrelerin Özellikleri ile Doğrusallıkta sapma .......................................................................................... 190
9.2.2. Değişkenlerin Özellikleri ile Doğrusallıkta sapma .................................................................................. 195
10. ÇOK DEĞİŞKENLİ REGRESYON ANALİZİ ....................................................................................................................... 201
10.1. Çok Değişkenli Regresyon Analizi ............................................................................................................................ 207
10.2. Çok Değişkenli Regresyonda Parametre Tahmini ........................................................................................................ 209
10.3. Çok Değişkenli Regresyonda Tahmin Edilen Parametrelerin Varyans ve Kovaryansı ............................................................................................................................ 211
10.4 Çoklu Belirginlik Katsayısı ($R^2$) ............................................................................................................................ 212
10.5. Varyans Analiz Tablosu Yaklaşımı: $F$ Testi ........................................................................................................ 213
11. MODELE YENİ BİR BAĞIMSIZ DEĞİŞKENİN EKLEME GEREKLİLİĞİNİN TESTİ ................................................................. 221
12. ÇOK DEĞİŞKENLİ REGRESYON MODELLERİYLE İLGİLİ BİR UYGULAMA ... 236
13. REGRESYON MODELLERİNİN YAPISAL KARARLILIKLARININ TESTİ .......... 257
   13.1. Yapısal Kararlılığın Chow Testi ile Sınanması ................................................... 263
   13.2. Yapısal Farklılaşmanın Gölge Değişken Yöntemiyle Belirlenmesi .................... 267
14. NORMALLİK VARSAYIMI ................................................................................................................................. 275
   14.1. Jarque-Bera Testi (Normallik Varsayımı) ........................................................................ 283
KAYNAKÇA ......................................................................................................................................................... 293
KISALTMALAR
YAZAR NOTU
1. EKONOMETRİNİN TANIMI AMACI VE METODOLOJİSİ
Bu Bölümde Neler Öğreneceğiz?

1.1. Ekonometri Nedir? Ekonometrinin Gerekliliği

1.2. Ekonometrinin Amacı ve Metodolojisi

1.3. Ekonometrik Model

1.4. Ekonometride Kullanılan Veriler

1.5. Ekonometride Bilgisayarın, Paket Programların Yeri
Bölüm Hakkında İlişki Oluşturan Sorular
Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri
Anahtar Kavramlar
Giriş

Bu bölümde öncelikle ekonometri biliminin gerekliliği üzerinde durularak ekonometrinin kapsamı ve amaçları açıklanacak; iktisat, istatistik ve matematik bilim dallarıyla ilişkisi ele alınacaktır. Ardından ekonometrik bir modelin tanımı yapılarak ekonometride kullanılan veri türleri hakkında bilgiler verilecektir.

Ekonometrinin başlangıç düzeyindeki temel kavram ve analizler ayrıntılı olarak ele alınmadan önce genel olarak ekonometri bilimi ile ilgili bilgiler verilmesinin faydalı olacağını düşünüldüğü bu bölümde öğrenciler ekonometrinin amacı ile ekonometrik model ve veri türleri hakkında önsel olarak bilgi sahibi olmaları amaçlanmıştır.
1.1. Ekonometri Nedir? Ekonometrinin Gerekliliği

İktisat öğrencisinin karşılaştığı ilk temel kavram iktisadi değişkenler arasındaki ilişkilerdir. İktisat ilminin dayandığı iktisat teorisi, iktisadi değişkenler arasındaki ilişkilerin özelliklerini ve iktisadi olayları açıklar. Şüphesiz en iyi teori, bütün değişkenleri ve bütün bağlantıları eksiksiz olarak ele alan ve ekonominin işleyişini bir zaman dilimi içinde (statik) veya zaman akımı çerçevesinde (dinamik) gösterebilen teoridir.


İktisat teorisi, matematik ve istatistikin bir bileşimi olan ekonometri, bu üç bilim dalının her birinden bütünüyle farklıdır. İktisat teorisi, matematik ve istatistikin her biri, bunların bileşimleri iktisadi istatistikler, matematiksel istatistik ve matematiksel iktisat günümüz iktisadi olaylardaki nicel ilişkilerin kavranması için gerektiği olmakla birlikte hiçbir tek başına yeterli değildir.

Ekonometri iktisat teorisinin belirlendiği iktisadi ilişkileri en uygun, en gerçekçi, işlenebilir bir matematiksel bir model içinde ele alarak, değişkenler arasındaki bağlantıyı kuran parametrelerin sayısal değerlerini en doğru tahmin edeceğin yöntemleri araştırır.
Genel olarak ekonometri, iktisadi verileri ve aralarındaki ilişkileri analiz etmek için iktisat teorisi ve istatistik yöntemleri kullanma bilimi ve sanatıdır. Makro iktisat, mikro iktisat, çalışma ekonomisi, finans, pazarlama ve iktisat politikası dahil olmak üzere ekonominin birçok bilim dalında ekonometrik yöntemler kullanılmaktadır.

1.2. Ekonometrinin Amacı ve Metodolojisi

Ekonometrinin amacı iktisat teorisinin belirlediği iktisadi ilişkileri kantitatif bir model çerçevesinde ele alarak değişkenlerin değişme sebeplerini sayisal ölçülerle ifade etmek ve bu değişkenlerin gelecekte alabileceği değerleri tahmin edebilmek ve iktisat teorilerinin doğruluğunu test etmektir.

Genel olarak ekonometri, teorik ekonometri ve uygulamalı ekonometri olarak ikiye ayrılır. Her ikisine de klasik ya da Bayesgil geleneklerle yaklaşılabilir. Ancak ekonometriye giriş düzeyindeki kitaplar genellikle klasik yaklaşım tercih edilirken, bayesgil yaklaşım kullanılmamaktadır. Bu derste de klasik yaklaşım kullanılacaktır.

Teorik ekonometri ile uygulamalı ekonometrinin birbirinden tamamen ayrılmış olan alanlarını belirlemek ve iki disiplinin kesin sınırlarını çizmek zordur. Teorik ekonometri, bir yandan en uygun ekonometrik modellerin geliştirilmesi sorununu incelerken, diğer yandan ekonometrik modeller için en uygun parametre tahmin yöntemlerini inceler. Uygulamalı ekonometri ise, gerçek verilere dayanarak ilgili iktisadi olay için en uygun kabul edilen ekonometrik modelin parametrelerini en iyi biçimde tahmin edip bunun sonuçlarını incelemek gibi konularla ilgilenir.


Ekonomik modellerin ekonometrik analizinde gerekli adımlar aşağıdaki şemada verilmiştir.

**Şekil-1.2:** Ekonometrik analizin adımları
Yukarıda verdğimiz şemayı bir örnekle şu şekilde açıklayabiliriz. İktisat teorisi bir kişiinin tüketiminin, harcanabilir gelirine, servetine, zevk ve alışkanlıklarına bağlı olduğunu öngörür. Bu kesin bir ilişkiye ifade etmektedir. Çünkü tüketimin bütünüyle bu üç etmen tarafından belirlendiği, bunun dışında hiçbir etmenin tüketimi etkilemediği sonucu ortaya çıkmaktadır.

Matematiksel iktisatta tüketimin yukarıdaki soyut iktisadi ilişkisi matematik terimleri ile ifade edilir ve aşağıdaki tüketim denklemi yazılabilir.

\[ C_t = \beta_0 + \beta_1Y_t + \beta_2W_t + \beta_3t \]

Burada,

\[ C_t = \text{tütüketimi} \]

\[ Y_t = \text{harcanabilir geliri} \]

\[ W = \text{serveti} \]

\[ t = \text{zevk ve alışkanlıklar} \]

göstermektedir. \( \beta_0, \beta_1, \beta_2 \) ve \( \beta_3 \) ise matematiksel modelin katsaylarıdır. Bu matematiksel model kesin ilişkileri göstermektedir, çünkü tüketim sadece eşitliğin sağ tarafından iktisadi değişkenler tarafından açıklanmaktadır. Ancak iktisadi gerçekler göstermektedir ki, tüketimi iktisadi krizler, vergi düzenlemeleri gibi başka iktisadi unsurlar da etkileyebilmektedir. Ayrıca insan davranışı, psikolojik ve toplumsal etmenlerin etkisi altındadır ve düzenli değildir. Dolayısıyla tüketim modeli sıraladığımız değişkenlerin de etkisi altında olduğuna göre matematiksel modeldeki kesinlikten uzaklaşacaktır. Model dışındaki bu etmenler dikkate alındığında model aşağıdaki gibi rastlantısal (stokastik) biçimi alır ve ekonometrik model olarak adlandırılır.

\[ C_t = \beta_0 + \beta_1Y_t + \beta_2W_t + \beta_3t + u_t \]

Yukarıdaki denklemede yer alan \( u_t \), hata terimi adını alır. Ekonometrik modelin, matematiksel modelden farklı bünyesinde hata terimini barındırması ve dolayısıyla kesin ilişkileri değil, rastlantısal ilişkileri ifade etmesidir.

Bundan sonraki adım ekonometrik modelde yer alan değişkenlere ilişkin gerçek verilerin sağlanmasıdır. Daha sonra model ekonometrinin kendine özgü yöntemleri ile tahmin edilir ve modelin katsaylarının sayısal değerleri elde edilir.

Modelin tahmininden sonraki aşama, modelin geçerliliğinin test edilmesidir. Tahmin edilen modelin sonuçlarının öncelikle iktisat teorisi ile uyumlu olması gerekmektedir. Şiyle
ki; tüketim modelinde harcanabilir gelirin katsayısı ($\beta_1$) tanımlı gerekli (ileride tekrar üzerinde durulacaktır) marjinal tüketim meyline eşittir. Marjinal tüketim meyli ise 0 ile 1 değerleri arasında yer alıktıdır. Eğer model tahmini sonucunda $\beta_1$, bu sınırlar dışında ise tahmin sonuçları kabul edilemez, bu durumda model yeniden tahmin edilmelidir. Modelin geçerliliği testinden diğer bir kasit, modelin ekonometrik açıdan temel varsayımlara uygunluğudur. Test sonuçlarına göre uygun model herhangi bir hipotezin testinde, iktisat politikası modelinin oluşumunda ve geleceğin tahmininde kullanılabılır.

Yukarıda da görüldüğü üzere, matematiksel terimlerle ifade edilen, genel iktisat teorisiyle iktisadi gerçeklerin nicel ölçümünün birleştirildiği, kendine özgü tahmin yöntemleri olan ekonometrinin üç temel amacı vardır. Bunlar:

1. Ekonometri, iktisat teorilerinin doğruluğunun sınanmasını amaçlar.
2. İktisadi ilişkilerin katsayılara sayısal tahminleri yapar.
3. Parametrelerin sayısal tahminlerinden iktisadi değişkenlerin gelecekteki değerlerini tahmin etmektedir.

Ekonometrinin amacı iktisadi olayları ve iktisadi değişkenler arasındaki ilişkileri kantitatif bir model içinde ele alarak değişkenlerin değişme sebeplerini sayısal ölçülebilecekleri ifade etmek ve bu değişkenlerin gelecekte alabileceği değerleri tahmin edebilmektir. Buna göre ekonometri biliminin nihai hedefi iktisadi olayların kantitatif ölçülerle izahını yapabilmek ve geleceğin yine kantitatif olarak tahminini yapabilmektir.

1.3. Ekonometrik Model

Ekonometrik modellerin en önemli özelliği rastlantsal bir değişken olan hata terimi içermesidir. İktisat teorisi ve matematiksel iktisat, iktisadi değişkenler arasında kesin ilişkiler kurar. Ekonometri, iktisadi ilişkilerin kesin olduğunu kabul etmez ve bu ilişkilerin rassal bileşeni ile ilgilenir.

Aşağıda matematiksel bir model verilmiştir.

$$Y = \alpha + \beta X$$

Burada yer alan $\alpha$ ve $\beta$ modelin parametreleridir. Matematiksel modelde X değişkeninin sabit bir değerine karşın Y değişkeni belli bir değer alır. Örneğin;

$$Y = 2 + 0.5X$$

Matematiksel modelinde X değişkeni 1’ eşit ise, Y değişkeni 2.5 değerini alacaktır, başka bir değer alması beklenmez. Ancak ekonometrik model kesin olmayan ilişkileri gösterdikten X değişkeni sabit bir değer eşit iken Y farklı değerler alabilir. Ekonometrik model aşağıdaki gibidir.
\[ Y_i = \alpha + \beta X_i + u_i \]

Ekonometrik

modelde, modelin sağ tarafından değişken genel olarak bağımsız değişken, sol tarafından değişken ise bağımlı değişken olarak adlandırılır. Ancak aşağıda gösterildiği üzere farklı biçimlerde de adlandırılmaktadır.

<table>
<thead>
<tr>
<th>Y</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bağımlı Değişken</td>
<td>Bağımsız Değişken</td>
</tr>
<tr>
<td>Açıkланan Değişken</td>
<td>Açıklayan Değişken</td>
</tr>
<tr>
<td>İçsel(endogen) Değişken</td>
<td>Dışsal(egzogen) Değişken</td>
</tr>
</tbody>
</table>

Tablo-1.1: Bağımlı ve bağımsız değişken kavramları

Hata terimi, çeşitli iktisadi büyüklükler arasında kesin ilişkiler öngören iktisat teorisi ve matematiksel iktisat tarafından göz ardı edilir. Ekonometride ise iktisadi ilişkilerin rassal bileşenleriyle ilgili varsayımlarda bulunmuş ve yeni yöntemler geliştirilmiştir.

1.4. Ekonometride Kullanılan Veriler

Veriler, iktisadi olaylar ve ilişkiler hakkında bilgi veren ve sayılarla ifade edilebilen bilgi (enformasyon) kaynaklardır. Ekonometrik modeli tahmin edebilmek için verilere ihtiyaç vardır ve ekonometrik modelin başarısı uygun verilerin bulunmasına bağlıdır.

Ekonometrik modelin tahmini için kullanılan veriler aşağıdaki gibidir:

1. **Zaman serisi verileri**, bir değişkenin dönemler itibariyle gözlenen değerler takımıdır. Nicel veya nitel olabilmektedir. 2010 Ocak ayı için İMKB-100 endeksi kapanış fiyatları, 2000-2010 yılları arası Akbank hisse senedine ilişkin yıllık getiri rakamları, üçer aylık GSMH rakamlarının her biri zaman serisi verisine örnek teşkil etmektedir. Örneklerden de anlaşılacağı üzere, zaman serisi verileri yıllık, 6 aylık, 3 aylık, günlük v.b. frekanslarla oluşturulabilmektedir.

Genel olarak zaman serisi verisi, ekonometrik modelde yer alan değişkenlerin zaman boyutu çerçevesindeki değişimlerinin sayisal ifadesidir. Zaman serisi verilerinin ekonometrik modelde gösterimi ile ilgili aşağıdaki basit model açıklayıcı olacaktır.

\[ C_t = \alpha_0 + \alpha_1 Y_t \]

Bu modelde yer alan

\[ C_t = t \] zamanındaki tüketimi,
\( Y_t = t \) zamanındaki geliri,
\[ \alpha_0, \alpha_1 = \text{modelin parametreleridir.} \]
\( t \) zaman dizisi, verileri olarak alınacak değerler tüketim(\( C \)) ve gelirin(\( Y \))
\( t=1,2,...,T \)örnekleme dönemi için sayısal değerlerdir.

2. **Yatay kesit verileri**, belli bir zaman noktasında derlenen verilerden oluşmaktadır ve
tek tek birimlere ilişkin değerler hakkında bilgi verir. Örneğin; İMKB’de işlem gören hisse
senetlerinin 6 Temmuz 2010 tarihindeki kapamış fiyatları, G20 ülkelerinin 2009 yılı büyüme
rakamları, AB ülkelerinin 2008 yılı nüfus artış hızları ve iller itibariyle 2005 yılı GSMH
rakamları.

Yatay kesit verilerin ekonometrik modelde gösterimi aşağıda gibidir.
\[ C_i = \alpha_0 + \alpha_i Y_i \]
Bu modelde yer alan \( i \) tüketicinin belli zaman noktasındaki tüketimi,
\( Y_i = i \) tüketicinin belli zaman noktasındaki geliri
\( \alpha_0, \alpha_1 = \text{modelin parametreleridir.} \)

Yatay kesit verilerinin kullanıldığı modellerin özelliği zaman boyutunun sabit
olmasıdır. Modellen matematiksel biçimi \( i=1,2,...,n \) tüketici için aynıdır.

3. **Panel veriler**, yatay kesit verilerinin çeşitli zamanlardaki değerleridir. İMKB-30’da
yer alan hisse senetlerinin 1990- 2009 yılları arasında getiri rakamları, 1960 -2009 imalat
sanayi sektörler itibariyle verimlilik rakamları, 1950- 2009 AB ülkeleri işsizlik oranları v. b.

Hem zaman hem de kesit verilerinin kullanıldığı basit bir modelde bu durum
aşağıda gibi gösterilir.
\[ C_{it} = \alpha_0 + \alpha_i Y_{it} \]
Yine bu modelde \( i \)ler yatay kesit verisini, \( t \)ler ise zaman serisini göstermektedir.
\[ C_{it} = tzaman noktasında \( i \) tüketicinin belli zaman noktasındaki tüketimi,
Y_{it} = tzaman noktasında \( i \) tüketicinin belli zaman noktasındaki geliri
gösterir. Buna göre panel veri setindeki \( n \) birim sayısı gösterilir iken , zaman
dönemleri sayısı ise \( T \) ile gösterilir. Örneğin 13 hisse senedinin haftalık iş günü(5) itibariyle
kapanış fiyatları var ise \( n=13, \ T=5 \) olmak üzere elimizde bu örnek ile ilgili \( nxT=13x5=65 \) veri vardır.

Kısaca;

- Zaman serisi verileri çoklu zaman dönemlerinde tek bir birim içerir.
- Yatay-kesit verisi çoklu birim içerip tek bir zaman noktasında gözlemlenir.
- Panel veri çoklu birim içerir, her birim için iki veya daha fazla zaman noktasında gözlemlenir.

Ekonometrik modelde kullanılan veriler gelir, tasarruf, tüketim, döviz kuru, faiz oranı gibi **nicel değişken** olabildiği gibi, cinsiyet, eğitim düzeyi (ilköğretim, lise, üniversite) gibi **nitel değişken** de olabilir. Yine bir örnek üzerinde açıkladığımız olursak; bir iş yerinde çalışanların kazançları, işyerindeki deneyimlerine ve eğitim düzeylerine bağlıdır. Kazanç ilişkisini ekonometrik model aşağıdaki gibi yazılabilir.

\[
Y_i = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + u_i \quad i = 1,\ldots,n
\]

modelde;

\[
Y_i = \text{i. çalışanın maaşı}
\]

\[
X_{1i} = \text{i. çalışanın işyeri deneyimi (1 yıl, 5 yıl, 16 yıl gibi)}
\]

\[
X_{2i} = \text{i. çalışanın eğitim düzeyi (lise ve üniversite)}
\]

ifade etmektedir.

\( n \) sayıda kişiye ait verilerin yer alacağı kazanç modelinde kullanılan değişkenlerde maaş ve işçi deneyimine ilişkin veriler nicel olmasına karşın, eğitim düzeyi nitel bir değişkendir. Nitel değişkenler, ekonometrik modele gölge değişken yöntemleri ile modele ilave edilir. (Gölge değişkenler başlığı altında incelenecektir.)

**1.5. Ekonometride Bilgisayarın, Paket Programların Yeri**

Son çeyrek yılda bilgisayar teknolojisindeki gelişim, özellikle PC’lerin yaygın kullanımını ve buna bağlı olarak ekonometrik paket programların yazılımı, ekonometrik çalışmalarla ve uygulamalarla önemli ölçüde katkını olmuştur, ekonometrinin artan ivmesinde önemli yer almıştır. Derslerimiz ilerledüğünde modelin tahmini için kullanılan matematiksel işlemlerin hisse de kolay olmadığı herhaldece görülecektir. Ancak paket program aracılığı ile bu modellerin çözümlü ancak birkaç dakika almaktaadır.

Uygulamada en çok kullanılan programlar arasında E-Wiews, SPSS, MINITAB, MICRO TSP ‘yı sayabiliriz. Bu ders için şu ana kadar, hazır paket programları temin edilemediğinden Excel paket programı kullanılarak ilgili uygulamalar yapılarak. Derse alın öğrencilerin Excel bilgilerini geliştirmeleri dersin uygulamalarının daha anlaşılması olması
açısından önemlidir. İlerleyen aşamalarda her konu ile ilgili Excel yazılımı kullanılarak uygumla yapılacaktır.
Uygulamalar
Uygulama Soruları
Bu Bölümde Ne Öğrendik Özeti
Bölüm Soruları

1. Ekonometri bilimi ile ilgili üç tanımlamız.

2. Ekonometrinin iktisat ve istatistik bilimleri ile ilişkisi hakkında ne söyleyebilirsiniz?

3. Ekonometrinin amaçlarını yazarak yorumlayınız.

4. Ekonometrik model ile matematiksel model arasındaki farkı açıklayın.

5. Ekonometri ile iktisat bilim dalları, değişkenler arasındaki ilişkilere yaklaşımları açısından karşılaştırmınız.

6. Ekonometri için veri ne anlama gelmektedir?

7. Ekonometride kullanılan veri türlerini belirtiniz.

8. Ekonometride kullanılan veri türlerini birer örnek model yazarak açıklayın.

9. Ekonometri, iktisadi ilişkilerin kesinliği ile ilgili nasıl bir varsayım yapmaktadır?

2. REGRESYON ANALİZİNE GİRİŞ
Bu Bölümde Neler Öğreneceğiz?

2.1. Giriş

2.2. Basit ve Çok Değişkenli Regresyon Modelleri

2.3. Regresyon Analizi

2.4. Anakütle Regresyon Modeli

2.5. Örnek Kütle Regresyon Modeli

2.6. Hata Teriminin Kaynakları

EK Excel programı
Bölüm Hakkında İlgı Oluşturan Sorular
Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri
Anahtar Kavramlar


Giriş

Bu bölümde ilk olarak tek denklemli regresyon modelleri ve bu modellerin önemi üzerinde durulacak, basit ve çok değişkenli regresyon modelleri hakkında bilgi verilmiştir. Ardından regresyon analizi ile ilgili temel kavramlar detaylı olarak ele alınmıştır. Bu bağlamda regresyon ve korelasyon analizi arasındaki farklılıklar, regresyon modelinin temel bileşenleri, ana kütle ve örnek kütle kavramları, koşullu olasılık hesaplamaları ve hata terimi açıklanmıştır.

Yukarıda bahsedilen tanımlamaların, regresyon analizinin temel kavramları ve bileşenlerinin anlaşılması açısından, model tahmini aşamasına geçilmeden önce yararlı olacağını düşünülmüştür. Buna bağlı olarak öğrencilerin regresyon çözümlemesi, değişken tanımlamaları ve hata terimleri ile ilgili temel bilgilere sahip olmaları hedeflenmiştir.
2. REGRESYON ANALİZİNE GİRİŞ

Ekonometri konularıyla ilgili literatürde en çok rastlanan ve ilgi gören model türü tek denklemli doğrusal regresyon modelleridir. 1980 sonrası teorik ve uygulamalı ekonometrideki hızlı gelişime ile birlikte iktisadi çerçeveler ve veriler daha kompleks modellerle izah edilebiliyorsa da, tek denklemli modeller hala önemli bir yer tutmaktadır. Bunun sebepleri:

1. Geleneksel iktisat teorisi bir sonucu bir dizi sebebe bağlamakta ve bu tek denklemli bir model çerçevesinde ele alın bir tutuma sahip bulunmaktadır.

2. Gerçekler çok defa basit kalıplara sokulamayacak kadar karmaşık olmasına rağmen, iktisad olaylarda sebeb-sonuç (kozalite) ilişkilerinin doğrusal bir model içinde sunulması önemli ölçüde kolaylık sağlayacaktır.

Tek denklemli regresyon modellerinde yalnız bir bağımlı değişken vardır ve bu bağımlı değişken bir ya da birden çok bağımsız değişkenin doğrusal fonksiyonu olarak belirtilmiştir.


2.1. Basit ve Çok Değişkenli Regresyon Modelleri


İlişkileri mümkün olduğu kadar basitleştirebilmek için, tek bir ilişki üzerinde durulan ve bu ilişkinin de sadece iki değişkeni içerdığı şeklindeki varsayımдан hareketle, en basit durum basit regresyon modeli;

\[ Y_i = \beta_0 + \beta_1 X_i + u_i \quad i = 1,2,\ldots,n \]

ile gösterilirken, ikiden fazla değişkenin yer aldığı çok değişkenli regresyon modeli ise;

\[ Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_k X_{ik} + u_i \quad i = 1,2,\ldots,n \]
Yukarıdaki modellerde de görüldüğü üzere, basit regresyon modelinde bağımlı değişkeni açıklayan sadece bir tane bağımsız değişken \( (X) \) yer alırken, çok değişkenli regresyon modelinde birden fazla bağımsız değişken \( (X_1, X_2, \ldots, X_k) \) vardır. \( i \) ise \( 1 \) den \( n \) e kadar olan gözlemleri ifade etmektedir. Her iki modelde de sadece tek bir bağımlı değişken \( (Y) \) yer almaktadır.

Örneğin, bir malın talebi \( (D) \) en basit ifade ile malın fiyatının \( (P) \) fonksiyonudur. Malın fiyatı artarsa malın talebi düşer, malın fiyatı düşersese de talebi artar. Fiyattan talebe doğru bir nedensellik söz konusu olduğuna göre, talep bağımlı değişken, fiyat ise bağımsız değişendir. Buna göre basit regresyon modeli aşağıdaki gibi kurulacaktır.

\[
D_i = \beta_0 + \beta_1 P_i + u_i, \quad t = 1, 2, \ldots, t
\]

Ancak bir malın talebi, malın fiyatının yanı sıra bu malı talep eden kişilerin gelir düzeyine \( (Y) \), malın tamamlayıcısı malı varsa bu tamamlayıcı malın fiyatına \( (P_T) \) ve nihayet rakip mal varsa rakip malın fiyatına \( (P_R) \) bağlı olarak değişiklik gösterecek. İlgili değişkenler modele dahil edildiğinde, model çok değişkenli regresyon modeli olarak adlandırılacak ve aşağıdaki şekilde gösterilecektir.

\[
D_i = \beta_0 + \beta_1 P_i + \beta_2 Y_i + \beta_3 P_T + \beta_4 P_R + u_i, \quad t = 1, 2, \ldots, t
\]

Yukarıdaki model çok değişkenli regresyon modeli olup; malın talebi dört tane bağımsız değişken ve hata teriminin \( (u_i) \) bir fonksiyonudur.

### 2.2. Regresyon Analizi

Regresyon analizi genel olarak nicel değişkenler arasındaki ilişkilerin incelenmesi olarak tanımlanmaktadır. Regresyon, bağımlı ve bağımsız değişkenler arasındaki bağlantıyı kuran parametrelerin değerinin tahmin imkanını araştırır.

Regresyon analizinde, değişkenler arasındaki ilişki fonksiyonel ya da kesinlikle istatistiksel ilişkilerdir. Değişkenler arasındaki istatistiksel ilişkilerde, genellikle stokastik (tesadüfi-rastlantısal) değişkenler yanı olasılık dağılımı olan değişkenler kullanılır. Fonksiyonel ya da kesinlikle de değişkenler kullanılır ancak bunlar tesadüfi ya da stokastik değil, deterministik değişkenlerdir.

Basit korelasyon analizinin amacı, yukarıda da ifade ettiği üzere iki değişken arasındaki doğrusal ilişkinin yönünü ve derecesini ölçmektedir (Hatırlanacağı üzere korelasyon katsayısı \( r = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2 \sum y_i^2}} \) formülü ile hesaplanmaktadır.) Formülde yer alan unsurlar ise \( x_i = X_i - \bar{X} \) ve \( y_i = Y_i - \bar{Y} \) eşittir. Korelasyon katsayısı \( r \), \( -1 \leq r \leq 1 \) değerleri arasında yer alır. Regresyon analizinde ise bağımlı değişkenin ortalama değeri, bağımsız değişken(ler)in değişmeyen değerlerine dayanılarak tahmin edilmektedir.


Bir regresyon modelinin kurulabilmesi için önsel olarak;

1. Sebep-sonuç ilişkisine göre bağımlı bağımsız değişken ayrımının
2. Bağımsız değişkenlerin ve sayısının
3. Modelin fonksiyonel biçiminin

belirlenmesi gerekir. Ancak sebep-sonuç ilişkisi ekonometride regresyonun dışında olup çoğu kez iktisat kuramı tarafından saptanmaktadır.

2.3. Anakütte Regresyon Modeli

Regresyon çözümlemesi büyük ölçüde, bağımsız değişkenin değeri bilindiği ya da sabit olduğu zaman bağımlı değişkenin ana kütledeki ortalama değeri ile ilgilendir.

Basit regresyon modeli, bağımlı \( Y \) değişkeni ile bağımsız \( X \) değişkeni arasında bağlantılı sağlamaktadır. \( X \)’in her sabit değeri için ana kütlede rastlantısal ilişki gereği bağımlı değişkenin en az iki değerinin bulunması zorunludur. Böylece her \( X_i \) değeri için \( Y_j \) değerleri elde edilecek, bu da bir dağılım oluşturacaktır.

Rastlantısal ilişkiyi açıklayabilmek için belirli bir zaman boyutu içinde, hane halkları yatay kesitinde hipotetik veriler kullanılarak haftalık tüketim harcamaları ile harcanabilir gelir arasında ilişki incelenecektir. Tüketim harcamaları harcanabilir gelirin bir fonksiyonudur. Buna göre, \( X \) haftalık harcanabilir gelir, \( Y \) ise tüketim harcamalarıdır. Öncelikle 60 aileden oluşan ana kütle, gelirleri yaklaşıklı olarak aynı olan ailelerden oluşan 10 ayrı gruba ayrılır ve
alt ana kütle olarak adlandırılan her grup aynı gelir düzeyindeki farklı ailelerin tüketim harcamalarını içerir.

\[ n = 60 \text{ aile} \]

\[ X_i = \text{Haftalık harcanabilir gelir} \]

\[ Y_i = \text{Haftalık tüketim harcamaları} \]

Tablo-2.1: Haftalık harcanabilir gelir ve tüketim harcaması

| X_i | 55 | 65 | 70 | 65 | 70 | 80 | 75 | -88 | 115 | 5 | 6 | 4 | 7 | 6 | 5 | 7 | 6 | 7 |
|-----|----|----|----|----|----|----|----|-----|-----|---|---|---|---|---|---|---|---|---|---|
| Y_i | 80 | 90 | 94 | 103 | 116 | 130 | 144 | -88 | 115 | 7 | 6 | 5 | 6 | 5 | 7 | 6 | 7 | 7 |

Tablo 1’deki her bir sütun yukarıda ifade edildiği üzere alt ana kütle olarak adlandırılmakta ve sabit bir \( X_i \) değerlerine karşılık gelen \( Y_i \) bağımlı değişkeninin koşullu dağılımını vermektedir. Buna göre, yukarıdaki her bir sütun belli bir gelir düzeyine karşılık gelen tüketim harcamalarının koşullu dağılımını vermektedir. Tabelodan da açıkça görüldüğü üzere, belirli bir gelir düzeyine sahip \( (X_i) \) bir gruba ait tüm hane halklarının aynı tüketim harcamasında \( (Y_i) \) bulunması beklenmez. Diğer bir ifade ile bağımsız değişken (gelir) sabit değerine karşılık bağımlı değişkenin (tüketim) birbirinden farklı değerleri vardır. Örneğin, 80 $ eşit gelire sahip 5 aile, 55$, 60$, 65$, 70$ ve 75 $ olmak üzere farklı tüketim harcamalarında bulunmaktadır. Farklı tüketim harcamaları ise bir olasılık dağılım oluşturmaktadır. Bazıları daha fazla, bazıları daha az harcarlarsa da harcama rakamlarının söz konusu gelir düzeyini hedef alan bir değer etrafında toplanacakları beklenebilir. Böylece olasılık dağılımlı her alt ana kütle için beklenen değer (koşullu ortalama) \( E(Y|X = X_i) \) (veya kısaca \( E(Y|X_i) \) ve \( E(Y) \) olarak gösterilir) hesaplanır.

**Beklenen değer** (Koşullu ortalama): \( Y \) rastlantısal değişkeninin çok sayıda veya meydana gelişeacağı değerlerin ortalamasıdır.

Beklenen değerin hesaplanabilmesi için öncelikle \( Y \)'nin her alt ana kütle için koşullu olasılıklarının bulunması gerekir. Her alt ana kütle için \( Y \)'nin koşullu olasılığı \( P(Y|X) \) ile gösterilir.
Haftalık harcanabilir gelir (Y) 80$ birbirinden farklı 5 tane haftalık tüketim harcaması olduğuna göre, bunların her birinin gerçekleşme olasılığı 1/5'dir. Diğer bir ifade ile 80$ geliri olan 5 ailenin rastlantısal olarak seçilen bir ailenin 55$ haftalık tüketim harcaması olan aile olma koşullu olasılığı 1/5 'e eşittir ve aşağıdaki gösterilir:

\[ P(Y=55/X=80)=1/5 \]

Diğer bir alt ana kütle için örnek verecek olursak, haftalık harcanabilir geliri 220$ olan 7 aile vardır. Bu ailelerin tüketimleri birbirlerinden farklıdır ve her birinin gerçekleşme olasılığı 1/7'dir. Buna göre 220$ haftalık geliri olan 7 ailenin rastlantısal olarak seçilen bir ailenin 157$ haftalık tüketim harcaması olan aile olmasının koşullu olasılığı 1/7'ye eşittir.

\[ P(Y=157/X=220)=1/7 \]

Y’nin her bir koşullu olasılık dağılımı için koşullu ortalama (beklenen değer) \( E(Y|X=X_i) \) hesaplanacaktır. Her beklenen değer \( X_i \) ’nin doğrusal bir fonksiyonudur.

X=80 $ iken Y’nin koşullu ortalaması veya diğer bir ifade ile beklenen değeri:

\[
E(Y|X=80) = \frac{55}{5} + \frac{60}{5} + \frac{65}{5} + \frac{70}{5} + \frac{75}{5} = (55 + 60 + 65 + 70 + 75) \frac{1}{5} = 65
\]

\[ E(Y|X=80) = 65 \]

‘e eşittir.

Kısaca aynı 80 $ gelir seviyesindeki 5 ailenin birbirinden farklı tüketim harcamalarının koşullu ortalaması 65 $'a eşittir.

Diğer bir alt ana kütle, X=220 $ iken Y’nin koşullu ortalaması veya beklenen değeri:

\[
E(Y|X=220) = (135 + 137 + 140 + 152 + 157 + 160 + 162) \frac{1}{7} = 149
\]

\[ E(Y|X=220) = 149 \]

‘a eşittir.

Böylece 10 alt ana kütle için 10 tane koşullu ortalama (beklenen değer) hesaplanacaktır. Hesaplanan koşullu ortalamalar Tablo-2.2’de yer almaktadır.

Bu uygulamada alt ana kütülerin birim mevcutları farklıdır. Alt ana kütle birim mevcutlarının aynı olması halinde, ana kütle içinde bütün alt ana kütülerin koşullu olasılıkları aynı olacaktır.
Her alt ana kütlein bir dağılımı olduğuna göre, alt ana kütlenin koşullu varyansları ve dolayısıyla standart sapması hesaplanabilir. Bilindiği üzere varyans ve standart sapma olasılık dağılımının dağılışını diğer bir ifade ile yayılışını ölçer. Alt ana kütle koşullu varyansı ise:

$$\sigma_i^2 = E\left[ (Y - E(Y|X = X_i))^2 \right] = \frac{1}{N} \sum_{i=1}^{N} \left[ Y_i - E(Y|X = X_i) \right]^2$$

ile ifade edilmektedir. Örneğin

X=80 iken Y’nin koşullu varyansı:

$$E\left[ Y_i - E(Y|X = 80)^2 \right] = E\left( Y|X = 80 \right) = 65 \text{ olduğu bilindiğine göre}$$

$$E\left[ Y_i - 65 \right]^2 = \frac{55 - 65)^2 + (60 - 65)^2 + (65 - 65)^2 + (70 - 65)^2 + (75 - 65)^2}{5} = 50$$

’e eşittir

X=100 iken Y’nin koşullu varyansı ise

$$E\left[ Y_i - E(Y|X = 100)^2 \right] =$$

$$E\left[ Y_i - 77 \right]^2 = \frac{(65 - 77)^2 + (70 - 77)^2 + (74 - 77)^2 + (80 - 77)^2 + (85 - 77)^2 + (88 - 77)^2}{6} = 66$$

’ya eşittir.

Hesaplamalara göre alt ana kütle varyanslarının farklı olduğu sonucuna varılmıştır.
Ana kütle regresyon fonksiyonu, \( X \) veriğen Y’nin ana kütle ortalamasını \( X \) ile fonksiyonel ilişkili olduğunu gösterir. Yani X’teki değişmeye karşılık Y’nin ortalama tepkisini gösterir. Böylece, **doğrusal ana kütle regresyon modeli**, 

\[ E(Y \mid X_i) = \beta_0 + \beta_1 X_i \]

şeklinde ifade edilmektedir. Ana kütle regresyon modeli, ana kütle regresyon denklemi olarak da adlandırılmaktadır. Ana kütle regresyon modelinde yer alan \( \beta_0 \) ve \( \beta_1 \) modelin bilinmeyen parametreleridir. \( \beta_0 \) sabit (otonom) parametre, \( \beta_1 \) ise eğim parametresi olarak adlandırılır. \( \beta_0 \) parametresi, ana kütle regresyon doğrusunun koordinat sisteminde Y eksenini kestiği nokta, \( \beta_1 \) ise ana kütle regresyon doğrusunun eğimidir. \( \beta_0 \) \( X \) bağımlı değişkenindeki 1 birimlik değişme gerçekleştiğinde \( E(Y \mid X_i) \) deki değişme miktarını, diğer bir ifade ile marjinal değişmeyi göstermektedir ve aşağıdaki gibi ifade edilmektedir.

\[ \beta_1 = \frac{\Delta E(Y \mid X_i)}{\Delta X} = \frac{dE(Y \mid X_i)}{dX} \]

Doğrusallıktan kasıtlı, Y nin beklenen değerinin \( X \) ve parametrelerin doğrusal bir fonksiyonu olmasıdır. \( \beta_0 \) ve \( \beta_1 \) parametrelerinin üstlerinin 1 olması dolayısıyla model doğrusaldır. Hem parametreler hem de değişkenler açısından doğrusal olan regresyon modelleri birinci mertebeden regresyon modelleri olarak adlandırılır.

\[ E(Y \mid X_i) = \beta_0 + \beta_1 X_i^2 \]

İkinci dereceden doğrusal bir regresyon modelidir. Bağımsız değişkenin üstlerinin azamisi modelin mertebesini verir.

\[ Y_i = \beta_0 + \beta_1 X_i + u_i \]

Ana kütle regresyon modeli \( E(Y \mid X_i) = \beta_0 + \beta_1 X_i \) olduğuna göre, bağımlı değişkenin gözlemlenen değeri \( Y_i \) aşağıdaki gibi yazılabilir.

\[ Y_i = E(Y \mid X_i) + u_i \]

Yukarıdaki eşitliğe göre hata terimi \( (u_i) \), bağımlı değişkenin gözlemlenen değeri ile beklenen değeri arasındaki farka eşittir.

\[ u_i = Y_i - E(Y \mid X_i) \]

Böylece birinci alt ana kütle için hata terimleri sırasıyla

\[ 55-65=-10 \]

\[ 60-65=-5 \]
65-65=0
70-65=5
75-65=10

olarak hesaplanır. $u_i$ artı ve eksi değer alabilen gözlememeyen tesadüfi bir terimdir. Ana kütle hata terimi olarak adlandırılır. Kısaca bağımlı değişkenin değeri iki unsur tarafından belirlenmektedir.

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

1. **Sistematisk kısım** ($\beta_0 + \beta_1 X_i$) alt ana kütle ortalamalarından geçen bir doğru olup ana kütle regresyon doğrusu olarak adlandırılır ve aşağıdaki gibi gösterilir.

$$E(Y | X_i) = \mu_{\beta X_i} = \beta_0 + \beta_1 X_i$$

$X=0$ olduğunda $Y$'nin beklenen değeri ($E(Y | X_i = 0)$) $\beta_0$’a eşittir.

$$E(Y | X_i = 0) = \beta_0$$

ancak $\beta_0$’ı bu açıdan yorumlanması, $X=0$ etrafında veri bulunmasına bağlıdır. Aksi halde bu yorum geçerli değildir.

$\beta_1$, yukarıda da değinildiği üzere $X$ teki bir birimlik değişme karşısında $E(Y | X_i)$ deki değişme miktarını, diğer bir ifade ile marjinal değişmeyi verir. $\beta_1$ aynı zamanda ana kütle regresyon doğrusunun eğimidir.

$$\beta_1 = \frac{d E(Y | X_i)}{d X_i}$$

2. **Stokastik kısım**, hata terimlerini simgeleyen $u_i$, bağımlı değişken ile bağımsız değişken arasında ortalama ilişkiyi gösteren ($\beta_0 + \beta_1 X_i$) kısım arasındaki farka eşittir. Alt örnekler itibariyle dikkate alındığında ise her alt ana kütlinin birimlerinin alt ana kütle ortalamasından sapmalarını göstermektede, eğer $Y_i > E(Y | X_i)$ ise artı, $Y_i < E(Y | X_i)$ ise eksi değer almaktadır.

$u_i$, bağımlı değişken ($Y_i$) ile aynı özellikleri göstermesine rağmen gözlemlenemeyen rastlantısal değişkendir.

$$Y_i = E(Y_i | X_i) + u_i$$
\[ E(Y_i | X_i) = \mu_{Y_i} = \beta_1 + \beta_2 X_i \]

\[ Y_i = \beta_1 + \beta_2 X_i + u_i \]

Hata terimi, rastlantısal bir değişken olduğuna göre bağımlı değişken (Y) gibi hata terimi de olasılık dağılımına sahiptir ve dolayısıyla koşullu ortalamaşı (beklenen değeri) ve varyansı hesaplanabilir. Hata teriminin (\( u_i \)) koşullu ortalamaşı (beklenen değeri) sıfıra eşittir. Şöyle ki;

\[ Y_i = E(Y_i | X_i) + u_i \]

Eşitliğin her iki yanının beklenen değeri alınır.

\[ E(Y_i | X_i) = E[E(Y_i | X_i)] + E(u_i | X_i) \]
\[ = E(Y_i | X_i) + E(u_i | X_i) \]

\[ E[u_i | X_i] = 0 \]

olduğundan

\[ E(u_i | X_i) = 0 \]

Ana kütle regresyon doğrusunun bağımlı değişkenin beklenen değerlerinden geçtiği varsayımı, hata terimlerinin beklenen değerlerinin sıfır olduğu anlamına gelmektedir.

\[ \mathcal{B}(Y | X_i) = \beta_1 + \beta_2 X_i = \mu_{X_i} \]

Şekil-2.1: Ana kütle Regresyon Fonksiyonu

Yukarıdaki şekilde görüldüğü üzere \( n \) sayındaki (yukarıdaki örnekte \( n = 60 \)’a eşittir.) \( Y \) ve \( X \) gözlem çiftleri bir dağılım diyagramı üzerinde gösterilebilir. Geometrik olarak, ana kütle regresyon doğrusu, açıklıayıcı değişkenlerin sabit değerlerine karşılık gelen bağımlı değişkenin koşullu ortalamalarından (beklenen değerlerinden) geçer.
Örnek Kütle Regresyon Modeli

Uygulamada ana kütlenin gözlemlenmesi ve $\beta_0 + \beta_1 X_i$ doğrusunun bilinmemesi nedeniyle, anakütle yerine çoğu kez örnekten hareket edilir. $\beta_0$, $\beta_1$ ve hata teriminin varyansi ($\sigma^2$) bilinmeyen parametrelerdir. Bu parametreler ana kütle yerine örnek gözlemlerine dayanılarak istatistiksel olarak tahmin edilir. Genellikle X’in sabit değerleri için, Y örneklemeye değerleri bulunmaktadır.

Ana kütleden çekilen rastlantısal çekilen Örnek I:

<table>
<thead>
<tr>
<th>Y</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>65</td>
<td>80</td>
</tr>
<tr>
<td>80</td>
<td>100</td>
</tr>
<tr>
<td>79</td>
<td>120</td>
</tr>
<tr>
<td>113</td>
<td>140</td>
</tr>
<tr>
<td>125</td>
<td>160</td>
</tr>
<tr>
<td>115</td>
<td>180</td>
</tr>
<tr>
<td>144</td>
<td>200</td>
</tr>
<tr>
<td>157</td>
<td>220</td>
</tr>
<tr>
<td>155</td>
<td>240</td>
</tr>
<tr>
<td>178</td>
<td>260</td>
</tr>
</tbody>
</table>
Ana kütleden çekilen rastlantısal çekilen örnek II:

<table>
<thead>
<tr>
<th>Y</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>55</td>
<td>80</td>
</tr>
<tr>
<td>74</td>
<td>100</td>
</tr>
<tr>
<td>90</td>
<td>120</td>
</tr>
<tr>
<td>103</td>
<td>140</td>
</tr>
<tr>
<td>107</td>
<td>160</td>
</tr>
<tr>
<td>135</td>
<td>180</td>
</tr>
<tr>
<td>144</td>
<td>200</td>
</tr>
<tr>
<td>160</td>
<td>220</td>
</tr>
<tr>
<td>189</td>
<td>240</td>
</tr>
<tr>
<td>150</td>
<td>260</td>
</tr>
</tbody>
</table>

Şekil-2.2: Örnek regresyon doğruları

Yukarıdaki iki farklı örneklemde de görüldüğü üzere X değeri, tekrarlanan örneklemlerde değişmemekte, sabit kalmaktadır. Ancak Y stokastik (rastlantısal) bir değişken olduğu için X in her bir değerine birden fazla Y değeri karşılık gelmekte ve dolayısıyla yukarıda da görüldüğü üzere yinelenen örneklemlerde Y’nin değeri değişmektedir.

Regresyon analizindeki birinci amaç, örneklem verilerinin kullanılarak tahmin edilen örnek regresyon fonksiyonuna dayanarak (diğer bir ifade ile örnekten tahmin edilen bilgilerle) ana kütle regresyon fonksiyonunu tahmin etmektir.
\[ \hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i \]

\( \hat{Y}_i \) ana kütle regresyon modelinde \( E(Y \mid X) \) nin tahminidir. \( \hat{\beta}_1, \hat{\beta}_2 \) sırasıyla \( \beta_1 \) ve \( \beta_2 \) nin tahmini verir.

\[ Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i \]

\( \hat{u}_i \) kalıntı olarak adlandırılmaktadır, ana kütledeki hata terimini \( u_i \) ’nin tahminidir. Böylece

ana kütle regresyon fonksiyonu

\[ Y_i = \beta_1 + \beta_2 X_i + u_i \]

yerini örnek kütle regresyon fonksiyonu

\[ Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i \]

alacaktır. Burada önemli bir husus, ana kütle yerine örnekten hareket edildiği durumda ana kütle regresyon denklemi ve dolayısıyla ana kütle hata terimi \( u_i \) bilinmediği için \( u_i \) ’nin varyansı hesaplanamaz. Örnek regresyon fonksiyonu mümkün olduğunca ana kütle regresyon fonksiyonuna yakın tahmin edilmelidir.

\[ \text{Şekil-2.3: Anakütle ve örnek regresyon doğruları} \]
Şekil-2.3 ‘de de görüldüğü üzere ana kütle regresyon doğrusu her alt ana kütle için hesaplanan bağımlı değişkenin beklenen değerlerinden geçerken, örnek regresyon doğrusu bağımlı değişkenin tahmini değerlerinden geçmektedir.

2.5. Hata Teriminin Kaynakları

1. Spesifikasyon (belirlenme) hataları
   - Dışlanmış değişken
   - Matematiksel biçimleme hatası: değişkenler arası fonksiyonel ilişki doğrusal olmadığı halde doğrusal olarak yapılan biçimleme hatası
   - Gereksiz değişken
   -Eşanlı modelin tek denklem ile ifade edilmesi

2. Ölçme ve birleştirme hataları

3. Aynı iktisadi olay için bireylerin davranışındaki farklılıklar.

EK

Excel Programı


Excel programı ilk açıldığında ekrana gelen Excel penceresi henüz kaydedilmemişse, Şekil 1’de de görüleceği üzere “Kitap1” şeklinde isimlendirilir.

Bu isimsiz dosyanın kaydedilmiş halinde, başlıka yer alan bu Kitap1 ifadesi de değişecektir.

Şekil 2’de Excel penceresinin içeriği görülmektedir. Harflerle sayfada yer alan sütunlar, sayılarla ise ilgili sayfada yer alan satırlar gösterilmektedir. Dolayısıyla satır ve
sütunların kesiştiği yer olan hücreler, bu sayı ve harflerin birleştirilmesiyle oluşan referans adresleri ile gösterilebilir. Örneğin B sütunu ile 5. Satırın kesiştiği hücre B5 referans adresi ile gösterilir.
Şekil 3. Bilgi Türleri

Excel programında bir hücreye dört farklı bilgi girilebilir. Bunlardan ilki, sadece rakamlardan oluşan ve üzerinde dört işlem (toplama, çıkarma, çarpma ve bölme) yapılabilen sayısal bilgi'dir. Şekil 3’de B5 hücreşinde bir sayısal bilgi örneği görülmektedir. Harflerden veya harf ve rakamlardan oluşan ve üzerinde işlem yapılamayan bilgi türüne ise alfanümerik bilgi denilmektedir. Şekil 3’te etkin hücre konumunda olan B6 hücreşinde bir alfanümerik bilgi örneği görülmektedir. Şekil 3’ten de görüleceği üzere, Ad Çubuğu’nda etkin hücrenin referans adresi, formül çubuğunda ise etkin hücrede yer alan bilgi görülmektedir. Excel programına girilebilecek diğer bir bilgi türü tarihsel bilgi'dir. Tarihsel bilgiye B7 hücreşinde yer alan bilgi türünün örnek olarak verilebilir. Bu veri, seçili hücreye 6/3 verisinin girilmesiyle elde edilmiş ve bu bilgi ile üzerine dört işlem yapılamaz. Mantıksal bilgi olarak adlandırılabilecek son bilgi türü, >, <, ≥, ≤, =, ≠ şeklinde sırasıyla büyük, küçüktür, küçük ya da eşittir, büyük yada eşittir, eşit değildir şeklinde mantıksal operatörler yardımcıyla oluşturulur ve sonucu DOĞRU veya YANLIŞ olarak Excel tarafından belirlenen bilgi türüdür. Şekil 3’te görülen B8 hücreşine =5>2 bilgisi girilerek DOĞRU sonucu elde edilmiştir ki, bu hücrenin etkin ali Şekil 4’te görüldüğü gibidir:
Şekil 4. Mantıksal Bilgi Örneği

Excel’de bir hücreye bilgi girilme yapabilmek için sadece ilgili hücreyi seçip klavyede gerekli bilgiyi girmek yeterlidir. Öte yandan hazır fonksiyonların ve çeşitli matematiksel operatörlerin hücreye girilmesi için, fonksiyonun yazımından önce “=” simgesinin kullanılması gerekmektedir. Şekil 4’de görülen mantıksal bilgi örneğinde, “>” operatörünün kullanılmasından önce “=” kullanılması buna örnek olarak gösterilebilir.

Excel programında operatör kullanımını sırasında dikkat edilmesi gereken bazı önemli noktalar bulunmaktadır:

- Bir sayısal bilginin yüzdesi basitçe % ifadesi ile alınabilir.
  = 258% değerinin herhangi bir hücreye girilmesi sonucunda 2,58 sonucu elde edilecektir.

- Bir sayının üssü ^ simgesiyle alınabilir:
  $3^2$ ifadesinin Excel’de kullanımı =3^2 olup, bu değer bir hücreye girilmesi sonrasında, işlem sonucu etkin hücrede belirecektir. Şekil 5’de B2 hücresına bu bilgi girilmiş olup, işlemin sonucu etkin hücrede görülürken, Ad Çubuğunda hücrenin referans adresi, formül çubuğunda ise bu hücreye yazilmiş olan bilgi görülmektedir.

- Excel programında dört işlem şu şekilde yapılabilimektedir:

- Şekil 6’da etkin hücre incelendiğinde sağ alt kısımda kare şeklinde bir simge bulunmaktadır. Doldurma kulpu adı verilen bu şekil bir diziyi genişletmek veya doldurmak amacıyla kullanılmaktadır. Doldurma kulpu adı verilen bu şekil bir diziyi genişletmek veya doldurmak amacıyla kullanılmakta olup, yukarıdaki şeklinde bu şekil fare imleciyle tutulup sağa doğru çekildiğinde A6 hücresinde girilmiş olan formüllerin diğer hücrelere de, hücre adreslerini güncelleyerek girildiği görülücektir.

- Şekil 8’de B6 hücresi etkin hücreyken, formül çubuğunda, doldurma kulpunun A6’da yer alan formülü, B sütununa göre güncellediği görülmektedir.
Şekil 8. Doldurma Kulpunun Hücre Adreslerini Güncellemesi

-Excel programında hazırlık fonksiyonlarından faydalanırken, ilgili adres aralığını klavye vasıtasıyla girmeye gerek yoktur. Hazır fonksiyon girildikten sonra klavyeden girilecek parantez işaretinden sonra işlem yapılması istenen hücre veya hücre aralığı fare imlecile yardımcıla seçildikten sonra, parantez kapatıp klavyeden “Enter” düğmesi tıklanırsa da etkin hücrede hazır fonksiyonun işlem sonucu görülebilecektir.

Uygulama 1.

Tablo 2.1.’de yer alan haftalık harcanabilir gelir ve tüketim harcaması verileri aşağıda görüldüğü gibidir.

Bu tabloda \( n \) frekansı, toplam ise her sütunun ayrı ayrı toplamını göstermektedir. Frekans değeri Excel programında hazırlık fonksiyonlarından faydalanarak hesaplanabilir.

Frekansı hesaplamak için Excel programında hazırlık bir fonksiyon bulunmaktadır. B10 adresinde yer alan hücre seçilip, aşağıdaki formül girilirse B sütununda yer alan haftalık harcanabilir geliri 80$ olan ailelerin sayısı öğrenilebilir:

\[=\text{BAĞ\_DEĞ\_SAY}(B3:B9)\]
Bu ifade verilen aralıktta (parantez içerisinde yazılı olan) yer alan sayısal bilgi bulunan hücrelerin sayısını verecektir.

Şekil 10’da B10 adresinde yer alan hücreye fonksiyonun girilmesi sonrasında Ad Çubuğu ve Formül Çubuğu’nda yer alan bilgiler görülmektedir. Fonksiyonun yazılıması sonrasında B10 hücresinde 5 sonucu yer almıştır.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>80</td>
<td>100</td>
<td>120</td>
<td>140</td>
<td>160</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>55</td>
<td>65</td>
<td>79</td>
<td>80</td>
<td>102</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>60</td>
<td>70</td>
<td>84</td>
<td>93</td>
<td>107</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>65</td>
<td>74</td>
<td>90</td>
<td>95</td>
<td>110</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>70</td>
<td>80</td>
<td>94</td>
<td>103</td>
<td>116</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>75</td>
<td>85</td>
<td>98</td>
<td>108</td>
<td>118</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>88</td>
<td>-</td>
<td>113</td>
<td>125</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td>115</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>n</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>7</td>
</tr>
</tbody>
</table>

Şekil 10. Frekans Hesaplama

B10’a hazır fonksiyon girildikten sonra, bu hücrenin doldurma kulpu sağa doğru çekilirse, bu hücrede yer alan adresler güncellenerek, ilgili sütunun frekansı diğer sütunlar için de hesaplanabilir.

Şekil 9’da yer alan sütun toplamlarını hesaplamak için B11 hücrene girilmesi gereken hazır fonksiyon aşağıdaki şekildeki şekildedir:

=TOPLA(B3:B9)

Bu ifade, ilgili aralıktta yer alan hücrelerdeki nümerik bilgileri toplamaya yaramaktadır. B11 hücresi doldurma kulpundan yana doğru çekilirse, diğer sütunlarda yer alan sayısal bilgiler için de gerekli toplama işlemi yapılmış olacaktır.

Uygulama 2

Bu uygulama da Tablo 2.2.’de yer alan koşullu olasılıkla Excel programında yer alan hazır fonksiyonlar aracılığıyla hesaplanacaktır.
Şekil 11. Koşullu Olasılıklar

Şekil 11’de, onuncu satırdada, ilgili sütunun koşullu ortalamaları (beklenen değerleri), on birinci satırdada ise koşullu varyansları görülmektedir.

Geliri 80 $ olan ailelerin tüketim harcamalarının koşullu ortalamalarını yani bekleme değerini hesaplamak için B10 hücresına şu ifade girilmiştir:

=ORTALAMA(B3:B7)

Bu ifade ilgili sütunda yer alan değerlerin ortalamasını hesaplamaya yaramaktadır. Önceki uygulama da olduğu gibi bu Excel uygulamasında doldurma kulpundan faydalanılmaz. Çünkü farklı gelir kalıplarında yer alan aile sayıları birbirinden farklıdır. Örneğin, geliri 100 $ olan 6 aile bulunurken, geliri 120 $ olan 5 aile bulunmaktadır. Bu nedenle bu fonksiyonun her sütuna teker teker ilgili sütunun aralığının dikkate alınarak girilmesi gerekmektedir. C10 hücreinde yer alan hazır fonksiyon =ORTALAMA(C3:C8) şeklindedir.

B11 hücreinde geliri 80 $ olan ailenin tüketim harcamalarının koşullu varyansları yer almaktadır. Daha önceden belirtildiği gibi burada yer alan koşullu varyans

\[
\sigma_i^2 = E\left[ (Y_i - E(Y_i|X = x_i))^2 \right] = \frac{\sum_{i=1}^{n} (Y_i - E(Y_i|X = x_i))^2}{N}
\]

Formülüyle hesaplanmaktadır. Geliri 80$ olan ailelerin tüketim harcamaları için koşullu varyans,
\[ E[Y_i - 65]^2 = \frac{(55 - 65)^2 + (60 - 65)^2 + (65 - 65)^2 + (70 - 65)^2 + (75 - 65)^2}{5} = 50 \]

seklinde hesaplanabilir. Excel programında ise bu formül hazır fonksiyonlardan yararlanarak şu şekilde yazılabilir:

\[ =((B3-B10)^2+(B4-B10)^2+(B5-B10)^2+(B6-B10)^2+(B7-B10)^2)/5 \]

Burada B10 hücreinde, geliri 80$ olan ailelerin tüketim harcamalarının beklenen değeri, B3,…B7 hücrelerinde ise bu ailelerin tüketim harcamaları bulunmaktadır. ^2 ifadesi ise ilgili değerin karesinin alındığını göstermektedir. Geliri 100$ olan ailenin tüketim harcamalarının koşullu varyansı ise şu şekilde hesaplanabilir:

\[ =((C3-C10)^2+(C4-C10)^2+(C5-C10)^2+(C6-C10)^2+(C7-C10)^2+(C8-C10)^2)/6 \]

Burada elde edilen değer, beş yerine altıya bölünmesinin nedeni, geliri 100$ olan 6 ailenin bulunmasıdır.
Uygulamalar
Uygulama Soruları
Bu Bölümde Ne Öğrendik Özeti
Bölüm Soruları

1. Tek denklemli regresyon modellerinden ve bu modellerin literatürdeki öneminden kısaca bahsediniz.

2. Basit ve çok değişkenli regresyon modellerini tanımlayarak, bu modellerle ilgili birer örnek veriniz.


4. Bir regresyon modelinin kurulabilmesi için hangi unsurların belirlenmiş olması gerekmektedir?

5. Ana kütle ve örnek kütle kavramlarını açıklayınız.

6. Ana kütle regresyon ve örnek kütle regresyon fonksiyonlarını şekil yardımcıyla açıklayınız.

7. Regresyon modellerinde nedensellik kavramını kısaca açıklayınız.

8. Regresyon analizi ile korelasyon analizi arasındaki farklılıkları maddeler halinde açıklayınız.


10. Doğrusal regresyon modeli ifadesindeki doğruluk kavramının ne anlama geldiğini belirtiniz.
3. KLASİK DOĞRUSAL REGRESYON MODELİNİN TEMEL VARSAYIMLARI
Bu Bölümde Neler Öğreneceğiz?

3.1.Klasik Doğrusal Regresyon Modelinin Temel Varsayımaları
Bölüm Hakkında İlgi Oluşturan Sorular
Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri
Anahtar Kavramlar
Giriş

Bu bölümde, klasik doğrusal regresyon modelinde bağımsız değişken ve hata terimi için öne sürülen varsayımlar detaylı olarak açıklanmıştır. Klasik doğrusal regresyon modelinin temel varsayımlarının anlaşılması, regresyon modelinin geçerliliğinin araştırılması ve varsayımlardan sapmaların test edilmesi açısından oldukça önemlidir. Bu nedenle sözü edilen varsayımların dayandığı temellerin açıklanması ve tahmin edilen regresyonun hangi koşullar altında geçerli olacağı bilgisinin kazanılması hedeflenmiştir.
3. KLASİK DOĞRUSAL REGRESYON MODELİNİN TEMEL VARSAYIMLARI

Regresyon modelinin geçerliliği, \( Y \)’i belirleyen \( X \) ve \( u \) ile ilgili varsayımların gerçekleşmesine bağlıdır. Örneklerin dayandığı ana kütlein bu varsayımlara uymaması, ekonometrik modelin aleyhine ciddi sorunlar doğurabilir. Klasik doğrusal regresyon modelinin bağımsız değişken( \( X \) ) ve stokastik hata terimi( \( u \) ) için yapılan varsayımlar aşağıdaki gibidir.

1. \( X \) veriyken \( Y \)'nin koşullu ortalaması, doğrusal regresyon modelidir.
   
   \[ E(Y | X) = \beta_0 + \beta_1 X \]

   Burada doğrusallıktan kasıt, regresyon modelinin parametreleri itibariyle doğrusal olmasdır, \( Y \) ve \( X \) değişkenleri doğrusal olmayabilir.

<table>
<thead>
<tr>
<th>Model parametreleri doğrusal mı?</th>
<th>Model değişkenlerde doğrusal mı?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evet</td>
<td>Hayır</td>
</tr>
<tr>
<td>DRM</td>
<td>DRM</td>
</tr>
<tr>
<td>Hayır</td>
<td>DORM</td>
</tr>
</tbody>
</table>

   Açıklama: DRM= Doğrusal Regresyon modeli

   DORM= Doğrusal olmayan Regresyon Modeli

   **Tablo-3.1:** Doğrusal ve doğrusal olmayan modeller

2. Ana kütle hata terimi beklenen değeri sıfır olan rastlantısal bir değişkendir.
   
   \[ E(u_i | X) = E(u) = 0 \]

   Ana kütle regresyon modelindeki sistematik kısm( \( \beta_1 + \beta_2 X \) ), \( X \) ile \( Y \) arasındaki ortalama bir ilişkiye ifade eder. Şöyledi;

   Alt örnekler itibariyle, sabit bir \( X \) değerine karşılık rastlantısal bir değişken olan \( Y \)’nin birden fazla değerinin karşılığı olduğu ve bunların bir olasılık dağılımı oluşturduğu üzerinde Bölüm 2’de durulmuştur. Her bir alt ana kütle için oluşan olasılık dağılımların beklenen değerleri hesaplayarak ve ana kütle regresyon doğrusunun bu noktalar üzerinden geçtiğini göstermişik. Gözlemlenen \( Y \) değerleri, ana kütle regresyon doğrusunun üzerinde ise hata terimi pozitif, altında ise negatif değer alacaktır. Varsayım gereği, sonuc ile sapmaların toplamı sıfır olacaktır. Bundan çıkartılabilecek sonuc ise hata terimi kaynaklarının \( Y \)’nin ortalaması üzerindeki etkisinin sıfır olduğu durur.

   Kısaca göstermek istersek,
\[ Y_i = \beta_0 + \beta_1 X_i + u_i \]

Bu özellikten yararlanarak ana kütle regresyon modelini \( X \) ve \( Y \)‘nin ortalamaları cinsinden ifade etmek mümkündür. Bunun için öncelikle anakütle regresyon modelinde her iki tarafın beklenen değeri alınır.

\[
E(Y_i) = E(\beta_0 + \beta_2 X_i + u_i)
\]

\( X \) değişkeni olasılıkli olmadığı için ve \( \beta_i \)’in bir sabit değer olması nedeniyle beklenen değer kendisine eşit olduğu için,

\[
E(Y_i) = \beta_i + \beta_2 X_i + E(u_i)
\]

şeklinde yazılabilir. \( E(u_i) = 0 \) olduğu varsayımı gereği

\[
E(Y_i) = \beta_i + \beta_2 X_i \text{ veya } \mu_{yX_i} = \beta_i + \beta_2 X_i
\]

sonucuna ulaşılır. \( \mu_{yX_i} \), \( Y \)‘nin koşullu ortalamasıdır. Yine aynı varsayım sonucu

\[
u_i = Y_i - E(Y_i)
\]

veya

\[
u_i = Y_i - E(Y_i)
\]

veya

\[
u_i = Y_i - \mu_{yX_i}
\]

şeklinde yazılabilen denklemin her iki tarafının beklenen değeri alınır ve

\[
E(u_i) = E(Y_i) - E\left[E(Y_i|X_i)\right] \quad E(u_i) = 0 \quad \text{ olduğu için}
\]

\[
E(Y_i) = E\left[E(Y_i|X_i)\right] \quad \mu_i = \mu_{yX_i} = \beta_i + \beta_2 X_i
\]

sonucuna ulaşılır. Aynı şekilde denklemdede \( X \) yerine \( X \)'ın ana kütle ortalamalarına( \( \mu_x \) ) yer verilirse, ana kütle regresyon modelini ana kütle ilişkisinin ortalamalar ile ifade etmek mümkün olacaktır.

\[
\mu_i = \mu_{yX_i} = \beta_i + \beta_2 \mu_x
\]

\( X \)'in ana kütle ortalaması( \( \mu_x \) ), alt ana kütle ortalamalarının tartılı ortalamasıdır.
3. $X$ değeri veriyleken, alt ana kütle hata terimlerinin $(u_i)$ koşullu varyansı bütün alt ana kütleler için aynıdır.

$$\text{Var}(u_i | X_i) = E\left[u_i - E(u_i | X_i)\right]^2$$
$$= E(u_i^2) = \sigma^2_u$$

$$E(Y | X_i) = \beta_1 + \beta_2 X_i$$

Şekil-3.1: Homoskedasite varsayımı

Şekil-3.1'de görüldüğü üzere her $X$ değerine karşılık gelen hata terimleri alt ana kütlelerin ortalamaları etrafında aynı dağılımı göstermektedir. Alt ana kütlelerin hata terimlerinin aynı değişim aralığında içinde yer almaları yanıvaryansların eşit olması, homoskedasite veya sabit varyans olarak adlandırılmaktadır.

Şekil-3.2: Heteroskedasite

Şekil-3.2'de $X$'in değeri artarken, alt ana kütle hata terimlerinin varyanslanküçülüğü görülmektedir.$X$'in değeri artarken, alt ana kütle hata terimlerinin varyanslarının büyüdüğü durum da söz konusu olması mümkündür.$X$'in aldığı değere göre hata teriminin varyansının
değişmesi ise heteroskedasite veya değişen varyans durumudur. Alt ana kütlelerin birim sayısı eşit olsa bile varyansları eşit olmaya bilir.

Regresyon doğrusu etrafında dağılmayı gösteren bağımlı değişken Y’nin koşullu varyansı diğer bir ifade ile ana kütle regresyon varyansı(\(\sigma_{ix}^2\)) hata teriminin varyansına(\(\sigma_u^2\)) eşittir.

\[
\sigma_{ix}^2 = \sigma_u^2
\]
eşitliğini aşağıdaki gibi elde etmek mümkündür. Bağımlı değişken Y’nin koşullu varyansı

\[
\sigma_{ix}^2 = E\left[ Y - E(Y|X_i) \right]^2 \text{e eşittir.}
\]

\[
u_i = Y_i - E(Y|X_i) \text{ olduğuundan}
\]

\[
\sigma_{ix}^2 = E(u_i^2) \text{'eşit olacak, böylece}
\]

\[
\sigma_{ix}^2 = E(u_i^2) = \sigma_u^2
\]
gösterimi mümkün olacaktır. Hata terimi ve bağımlı değişkenin dağılımlar aynı diğer bir ifade ile varyansları eşit olmakla birlikte ortalamaları farklıdır. Hatırlanacağı üzere:

Hata teriminin beklenen değeri (koşullu ortalaması):

\[
E(u_i|X_i) = E(u_i) = 0
\]

Bağımlı değişkenin beklenen değeri (koşullu ortalaması):

\[
E(Y|X) = \beta_0 + \beta X
\]

ile gösterilmiştir.

4. Hata terimleri arasında otokorelasyon(ardışık bağımlılık) yoktur. Bu varsayışa göre \(X_i\) ve \(X_j\) gibi iki \(X\) verilmişken \(u_i\) ile \(u_j\) arasında ilişki yoktur. Hata terimlerinin birbirleri ile ilişkisiz olması diğer bir ifade ile birbirlerinden bağımsız olmaları ancak \(Kov(u_iu_j) = 0\) olması ile mümkündür. Sembolik olarak

\[
Kov(u_iu_j|X, X_j) = E\left[\left[u_i - E(u_i|X_i)\right]\left[u_j - E(u_j|X_j)\right]\right]
\]

\[
i \neq j
\]

\[
= E(u_iu_j) = 0
\]

ile gösterilir.
Bu varsayımda gereği, $X_t$ veriye herhangi iki $Y$ değerinin ortalamalarından sapmaları belli bir kural izlemezler. Varsayının gerçekleşmemesi durumunda modelin aleyhine negatif veya pozitif otokorelasyon gerçekleşmekteidir.

![Pozitif Otokorelasyon](image1.png)  
![Negatif Otokorelasyon](image2.png)  
![Otokorelasyon Olmama](image3.png)

Ana kütle regresyon modelinde ($Y_t = \beta_0 + \beta_1 X_t + u_t$), $Y_t$ bağımlı değişkeninin değeri, $X_t$ bağımsız değişkenine ve $u_t$ hata terimine bağlıdır. Ancak $u_t$ ile $u_{t-1}$ arasında otokorelasyon varsa, örneğin $u_t$ ile $u_{t-1}$ pozitif olarak birbirleriyile ilişkilisiyse, $Y_t$ sadece $X_t$ ve $u_t$ değil, aynı zamanda $u_{t-1}$ ile de ilişkili olacaktır. Bu durumda bağımsız değişken ve hata teriminin, bağımlı değişken üzerindeki bireysel etkilerini tespit etmek mümkün olmayacaktır.

5. Hata terimi($u_t$) ile $X$ bağımsız değişkeni ilgisizdir. Bu varsayım hata terimi ile bağımsız değişkenin otokorelasyonsuz olduğunu ifade etmektedir.

$$Kov(u_t, X_t) = 0$$

$$Kov(u_t, X_t) = E \left[ (u_t - E(u_t|X_t)) [X_t - E(X_t)] \right] \quad E(u_t|X_t) = E(u_t) = 0 \text{dan}$$

$$= E[u_t(X_t - E(X_t))] = 0 \quad E(X_t) \text{ olasılıkli değişken değil} \quad X_t = \bar{X}$$

$$= E(u_t X_t) - E(X_t) E(u_t)$$

$$= E(u_t X_t) = 0$$

Daha önce ve 4. varsayımda da belirtildiği üzere bağımsız değişken(sistematik kısım) ve hata teriminin(tesadüfi kısım) bağımlı değişken üzerindeki etkisi ayrı ayrı ölçülebilir ve toplanabilir. Eğer hata terimi($u_t$) ile $X$ bağımsız değişkeni ilişkili ise bağımlı değişken üzerindeki bireysel etkilerini tespit etmek mümkün olamayacaktır. $X$ ile $u_{t}$ pozitif yönde ilişkilisiyse $X$ artarken $u_t$ de artacak, $X$ azalırken $u_t$ de azalacaktır. $X$ ile $u_t$ negatif yönde ilişkilisiyse $X$ artarken $u_t$ de azalacak, $X$ azalırken $u_t$ artacaktır.

6. $X$ bağımsız değişkeninin stokastik olmadığı yineleden örneklerde değerinin değişmediği, yani olasılık dağılımının olmadığı varsayılmaktadır.

7. Gözlem sayısı(n) anakütle parametreleri(k) sayısından fazladır (n>k).
Gözlem sayısı bağımsız değişken sayısından büyük olmalıdır. Bunu bir örnek üzerinde kısaca açıklayarak mümkün dır. X ve Y ile ilgili gözlemlerimizin sadece X=80, Y=55 gibi olduğunu düşünelim. Regresyon modelimiz iki bilinmeyen \( \beta_0 \) ve \( \beta_1 \) olduğuna göre, iki bilinmeyeni bulabilmek için en az iki noktaya ihtiyaç vardır. Verilerimiz ise tek bir noktayı belirlediğine göre, bu durumda \( \beta_0 \) ve \( \beta_1 \) parametrelerini tahmin etmek olanağız hale gelecektir.

8. Rastlantısal olmayan X bağımsız değişkeninin belli bir örnekteki değeri aynı olmamalıdır. X değişkeninin \( \text{Var}(X) > 0 \) olması, birden fazla alt ana kütleye olmalıdır.

\[
\text{Var}(X) = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu_x) > 0 \quad \mu_x \text{ veya } \bar{X}, \quad X \text{ değişkeninin ortalaması}
\]

Aksi takdirde ana kütleye parametrelerin tahmini olanağız hale gelecektir. X’ler aynı ise \( X_i = \mu_x \) olacak X’deki değişim sıfıra eşit olacaktır. Regresyon analizinde hem Y hem de X’te değişim zorunludur.


\[
Q = \alpha K + \beta L + u \quad (1)
\]

\[
Q = AK^{\alpha} L^{\beta} e \quad (2)
\]

\( Q = \text{Üretim} \)

\( K = \text{Kapital} \)
L=Emek
A=Teknoloji katsayısı

α ve β modellerin parametreleridir.

Model 1'e göre üretim faktörlerinden sadece emek veya sadece kapital ile üretimi gerçekleştirmek mümkün iken, Model 2'de üretim faktörleri çarpımdurumunda oldukları için birinin sıfır olması üretimin gerçekleşmemesine neden olacaktır. Dolayısıyla her iki üretim faktöründen belli bir oranda kullanılması zorunludur. Burada, araştırmacının doğrusal model(Model 1) veya doğrusal olmayan model (Model 2) konusunda tercihi önem kazanacaktır.

c. Modelde yer alan bağımlı değişken (Y), bağımsız değişken (X) ve hata terimi(u) ile ilgili olasılık varsayımının neler olduğunun bilinmesi gerekir.


Örnek:

Klasik doğrusal temel varsayımlarından elde edilen bilgilerle Bölüm 2’deki hipotetik verilerin kullanıldığı örnek için ana kütle regresyon modelini oluşturalım.

Öncelikle, hatırlanacağı üzere, ana kütle regresyon modeli,

\[ E(Y|X) = \beta_0 + \beta_1 X \]

veya

\[ \mu_Y = \mu_{\bar{Y}} = \beta_0 + \beta_1 \mu_X \]

gibi gösterilmektedir. Amacımız \( \beta_0 \) ve \( \beta_1 \)’in sayısal değerlerini bulmaktır. Hatırlanacağı üzere Bölüm 2’de her alt ana kütlenin koşullu ortalama(beklenen değeri) hesaplanmıştır.

Öncelikle bu alt ana kütle koşullu ortalamaları hesaplanarak ana kütle regresyonun doğrusunun eğimini veren \( \beta_1 \) hesaplanacaktır:

\[ \beta_1 = \frac{E(Y|X = 100) - E(Y|X = 80)}{X_2 - X_1} = \frac{77 - 65}{100 - 80} = \frac{12}{20} = 0.6 \]
veya başka iki noktadan;

$$
\beta_1 = \frac{E(Y \mid X = 220) - E(Y \mid X = 200)}{X_2 - X_1} = \frac{149 - 137}{220 - 200} = \frac{12}{20} = 0.6\
$$

veya,

$$
\beta_1 = \frac{E(Y \mid X = 120) - E(Y \mid X = 80)}{X_2 - X_1} = \frac{89 - 65}{120 - 80} = \frac{24}{40} = 0.6
$$

(Sizin seçeceğiniz başka noktalardan da aynı sonuca ulaşmanız mümkündür)

Bağımsız değişkenin($X$) koşulsuz ortalaması ($\mu_x$):

$$
\mu_x = \frac{5}{60}80 + \frac{6}{60}100 + \ldots + \frac{7}{60}260 = \frac{10420}{60} = 173.67
$$

olarak hesaplanır.

Bağımlı değişkenin($Y$) koşulsuz ortalaması ($\mu_y$):

$$
\mu_y = \mu_{yx} = \frac{\sum \sum Y}{N} = \frac{7272}{60} = 121.2
$$

sonucuna ulaşılır. Bu değer aynı zamanda $Y$'nin koşullu ortalamasına ($\mu_{yx}$) eşittir.

veya,

$$
\mu_y = \frac{5}{60}65 + \frac{6}{60}77 + \frac{5}{60}89 + \ldots + \frac{6}{60}161 + \frac{7}{60}173 = \frac{7272}{60} = 121.2
$$

hesaplanır. $Y$ ve $X$'in koşulsuz ortalamaları ve $\beta_1$ hesaplanทดığına $\beta_0$ göre aşağıdaki ana kütle regresyon modeli ile hesaplanabilir.

$$
\mu_y = \beta_0 + \beta_x \mu_x
$$

121.1 = $\beta_0 + 0.6 \times 173.67$
$\beta_0 = 17$
Böylece anakütle regresyon denklemi

\[ E(Y \mid X_i) = 17 + 0.6X_i \]

olarak bulunmuş olur. Anakütle regresyon denklemine göre gelir 1 birim değiştiğinde tüketim 0.6 birim artacaktır. Tanım gereği 0.6 marjinal tüketim meyline eşittir. Otonom parametre ise zorunlu tüketim harcamalarının 17 birim olduğunu göstermektedir.
Uygulamalar
Uygulama Soruları
Bu Bölümde Ne Öğrendik Özeti
Bölüm Soruları

1. Regresyon modelinin hangi varsayımlar altında geçerli olacağını, maddeler halinde kısaca yazınız.

2. Homoskedasite ve heteroskedasite kavramlarını tanımlayarak, şekil yardımıyla açıklayınız.

3. Hata terimlerinde otokorelasyon durumunu açıklayınız.

4. \( \text{Kov}(u_i, X_j) = 0 \) olduğunu varsayımı ispatlayınız. Bu varsayımı açıklamak için de değişken ile hata teriminin \( Y \) üzerindeki etkileri açısından anlamı gelmektedir?

5. \( \text{Var}(X) > 0 \) varsayımın geçerli olmadığı durumda ne gibi bir sorunla karşılaşılmaktadır? Açıklayınız.

6. Aşağıda yer alan boşlukları doldurunuz.

a. Ana kütte regresyon modelindeki sistematik kısımdan……olarak gösterilir.

b. Ana kütte regresyon doğrusu……..geçer.

c. Hata terimlerinin birbirleri ile ilişkisiz olması diğer bir ifade birbirlerinden bağımsız olmaları….olması ile mümkündür.

d. Klasik doğrusal regresyon modeli varsayımlarından biri, gözlem sayısının…dan fazla olduğunu öne sürmektedir.

7. \( \mu_{yi} = \mu_{x_i} = \beta_1 + \beta_2 X_i \) olduğunu gösteriniz.
8. Aşağıda bir malın fiyat ile talep verileri verilmiştir. (X=Fiyat, Y=Talep)

<table>
<thead>
<tr>
<th>Fiyat (Bağımsız değ.)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Talep (Bağımlı değ.)</td>
<td>45</td>
<td>44</td>
<td>40</td>
<td>35</td>
<td>36</td>
<td>32</td>
<td>32</td>
<td>31</td>
<td>28</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>46</td>
<td>45</td>
<td>42</td>
<td>38</td>
<td>39</td>
<td>35</td>
<td>34</td>
<td>32</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>47</td>
<td>46</td>
<td>44</td>
<td>42</td>
<td>40</td>
<td>37</td>
<td>36</td>
<td>33</td>
<td>32</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>47</td>
<td>46</td>
<td>44</td>
<td>42</td>
<td>38</td>
<td>38</td>
<td>34</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>48</td>
<td>48</td>
<td>46</td>
<td>43</td>
<td>39</td>
<td>40</td>
<td>35</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>47</td>
<td>42</td>
<td>36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>51</td>
<td>43</td>
<td>37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Toplam</td>
<td>336</td>
<td>230</td>
<td>220</td>
<td>252</td>
<td>200</td>
<td>266</td>
<td>180</td>
<td>238</td>
<td>160</td>
<td>90</td>
</tr>
</tbody>
</table>

a. Alt anakütle koşullu ortalamalarını (beklenen değerleri) hesaplayınız.

b. Alt anakütle koşullu varyansları hesaplayınız.

c. Anakütle regresyon modelini oluşturunuz.

9. X=Derse devam(%) , Y= Başarı notunu göstermek üzere, aşağıda anakütleyeilişkin bir alt anakütleye ait veriler yer almaktadır.

<table>
<thead>
<tr>
<th>X</th>
<th>0.60</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>60</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Yij</th>
<th>80</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>40</td>
</tr>
</tbody>
</table>

a. Yukarıdaki verileri kullanarak,  
\[ E[Y_i - E(Y_i|X_i)]^2 = E[u_i - E(u_i|X_i)]^2 \]

b. Derse devam %70 olduğunda, başarı notunun beklenen değeri 74 ise, derse devamın marjinal etkisini hesaplayarak yorumlayınız.
10. Aşağıdaki ifadelerden doğru olanları (D), yanlış olanları (Y) ile gösteriniz.

a. \( Y_i = \alpha_0 + \alpha_1 X_i \) (  )

b. \( \hat{y}_i = \hat{\alpha}_i x_i \) (  )

c. \( \hat{Y}_i = \alpha_0 + \alpha_1 X_i + u_i \) (  )

d. \( y_i = \hat{\alpha}_i x_i + \hat{u}_i \) (  )

e. \( E(Y_i | X = X_i) = \alpha_0 + \alpha_1 X_i \) (  )

f. \( Y_i = \hat{\alpha}_0 + \hat{\alpha}_1 X_i + \hat{u}_i \) (  )

Not: \( x = X - \bar{X}, y = Y - \bar{Y} \)
4. ANA KÜTLE REGRESYON MODELİNİN TAHMİNİ: EN KÜÇÜK KARELER YÖNTEMİ
Bu Bölümde Neler Öğreneceğiz?

4. ANA KÜTLE REGRESYON MODELİNİN TAHMİNİ: EN KÜÇÜK KARELER YÖNTEMİ

4.1. En Küçük Kareler Yöntemi

4.2. En Küçük Kareler Yönteminin Diğer Bir Uygulama Biçimi: Sapmalar Yöntemi

4.3. En Küçük Kareler Yönteminin Özellikleri

4.4. Uygulamalar
Bölüm Hakkında İlgi Oluşturan Sorular
Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri
Anahtar Kavramlar
Giriş

Bu bölümde ana kütle regresyon modeli parametrelerinin tahmini için kullanılan en yaygın yöntem olan En küçük kareler (EKK) yöntemi ile basit regresyon modelinin tahmin süreci açıklanacaktır. Buna ilaveten EKK yönteminin özellikleri de ele alınacaktır.

Bu bağlamda öğrencinin tek açıklayıcı değişken içeren regresyon modelleri için parametre tahmin sürecinin işletilmesi ve EKK yönteminin temel özellikleri konularında bilgi sahibi olması amaçlanmaktadır.
4. ANA KÜLTE REGRESYON MODELİNİN TAHMİNİ: EN KÜÇÜK KARELER YÖNTEMİ

4.1. En Küçük Kareler Yöntemi

Ana kütle regresyon modelinde yer alan unsurlar doğrudan gözlenemeyiz.

\[ Y_i = \beta_0 + \beta_1 X_i + u_i \]


Regresyon çözümlemesinde, örnek fonksiyonunu temel alarak ana kütle regresyon fonksiyonunu mümkün olduğu kadar doğruya yakın tahmin etmek temel hedeftir.

En küçük kareler yönteminde göre, n tane \( X_i \) ve \( Y_i \) veriye örnek regresyon modelinden elde edilen \( \hat{Y}_i \) ler, gözlemlenen \( Y_i \) lere olabildiğiince yakın olmalıdır. Bunun için kistas, gözlemlenen \( Y_i \) lere ile tahmin edilen \( \hat{Y}_i \) arasındaki fark olarak tanımlanan kalıntı (\( \hat{u}_i = Y_i - \hat{Y}_i \)) karelerinin toplamının minimum olmasıdır.

\[ \sum \hat{u}_i^2 = \sum (Y_i - \hat{Y}_i)^2 = \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 \Rightarrow \text{min} \]

Kalıntı kareler toplamının minimum (\( \sum \hat{u}_i^2 \rightarrow \text{min} \)) olması durumunda, tahmin edilen \( \hat{\beta}_0 \) ve \( \hat{\beta}_1 \) parametreleri en uygun, en güçlü ve en tesirli-BLUE- tahminleridir. Minimum yapma koşulu ise, kalıntı kareler toplamı fonksiyonunda \( \hat{\beta}_0 \) ve \( \hat{\beta}_1 \) parametrelerine göre birinci dereceden kısmi türevlerin sıfıra eşitlenmesi ile mümkündür. Kısaca,

\[ \frac{\partial}{\partial \hat{\beta}_0} \sum \hat{u}_i^2 = 0 \quad \text{ve} \quad \frac{\partial}{\partial \hat{\beta}_1} \sum \hat{u}_i^2 = 0 \]

olarak ifade edebiliriz. İlk olarak kalıntı kareler toplamı fonksiyonunda \( \hat{\beta}_0 \) parametresine kısmi türev alalım.
\[
\frac{\partial \sum \hat{u}_i^2}{\partial \beta_0} = \frac{\partial \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_i X_i)^2}{\partial \beta_0} = 0 \\
= \sum 2(Y_i - \hat{\beta}_0 - \hat{\beta}_i X_i)(-1) = 0 \\
= (-2) \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_i X_i) = 0
\]

Yukarıdaki eşitliğin her iki yanı (-2)’ye bölünirse,
\[
= \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_i X_i) = 0 \quad (*)
\]
sonucuna ulaşılır. Aynı işlem \( \hat{\beta}_i \) parametresi için tekrarlanır.
\[
\frac{\partial \sum \hat{u}_i^2}{\partial \beta_i} = \frac{\partial \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_i X_i)^2}{\partial \beta_i} = 0 \\
= \sum 2(Y_i - \hat{\beta}_0 - \hat{\beta}_i X_i)(-X_i) = 0 \\
= (-2) \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_i X_i)(X_i) = 0
\]

Yine eşitliğin her iki yanı (-2)’ye bölünür ve X parantez içine dağıtılrsa
\[
= \sum (Y_i X_i - \hat{\beta}_0 X_i - \hat{\beta}_i X_i^2) = 0 \quad (**)
\]
sonucuna ulaşılır.

(*) ve (**) eşitliklerini bir araya getirilip her iki denklem için de toplama işlemlerini yapildığında;
\[
\sum Y_i - n\hat{\beta}_0 - \sum \hat{\beta}_i X_i = 0 \\
\sum Y_i X_i - \hat{\beta}_0 X_i - \sum \hat{\beta}_i X_i^2 = 0
\]
elde ederiz (Not: \( \sum \)'ın sabit bir terim olan \( \hat{\beta}_0 \) ile çarpından "n \( \hat{\beta}_0 \) " elde edilir).
Nihayet bağımlı değişkeni içeren unsurlar eşitliğin sol tarafa bırakılarak,
\[
\sum Y_i = n\hat{\beta}_0 + \hat{\beta} \sum X_i \\
\sum Y_i X_i = \hat{\beta}_0 \sum X_i + \hat{\beta}_1 \sum X_i^2
\]
sonucuna ulaşılr.

Yukarıdaki iki denklem birlikte bir denklem sistemini oluşturur ve normal denklemler olarak adlandırılır. Burada n örneklem büyüklüğüdür. Bu denklem sistemi \( \hat{\beta}_0 \) ve \( \hat{\beta}_i \) için Cramer yöntemi ile çözülürse,
eşitlikleri elde edilir. Yukarıdaki eşitlikler “En Küçük Kareler Tahmin Edicileri” Normal denklemlerin çözümünden sağlanan \( \hat{\beta}_0 \) ve \( \hat{\beta}_i \) değerleri “En Küçük Kareler Tahminleri” olarak adlandırılır.

4.2. En Küçük Kareler Yönteminin Diğer Bir Uygulama Biçimi: Sapmalar Yöntemi

En küçük kareler yöntemi ile elde edilen normal denklemlere bağlı olan çözüm dışında aynı konuya başka bir yaklaşım daha vardır. Bu da değişkenlerin ortalama değerlerden olan farkları ile ifade edilmesidir. Yukarıda \( \hat{\beta}_0 \) ve \( \hat{\beta}_i \) ’nin hesaplanabilmesi için normal denklemlerden türetilmiş formüllerde \( X_i \) ve \( Y_i \) ’ye ait gözlemlenen örneklem değerleri yer almaktadır.

\( X_i \) ’nin kendi ortalamasından ( \( \bar{X} \) ) farklarını \( x_i \) ve \( Y_i \) ’nin kendi ortalamasından ( \( \bar{Y} \) ) farklarını \( y_i \) gösterelim:

\[
x_i = X_i - \bar{X} \\
y_i = Y_i - \bar{Y}
\]

Yukarıdaki denklemlerden \( \hat{\beta}_i \) ile ilgili olanın pay ve paydasını gözlem sayısı n’ye bölelim.

\[
\hat{\beta}_i = \frac{(n \sum X_i Y_i - \sum X_i \sum Y_i)/n}{(n \sum X_i^2 - (\sum X_i)^2)/n}
\]

\( \sum X_i/n = \bar{X} \) ve \( \sum Y_i/n = \bar{Y} \) eşit olduğu bilindiğine göre,

\[
\hat{\beta}_i = \frac{\sum X_i Y_i - n \bar{X} \bar{Y}}{\sum X_i^2 - n \bar{X}^2}
\]
elde edilecektir. Sonuç olarak,

\[ \hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2} \]

eşitliğe ulaştırılır. \( \hat{\beta} \) denkleminde yer alan değişkenler mutlak değerleri ile değil, mutlak değerlerin ortalamalarından farklılar ile yer almaktadır.

\[ \sum x_i^2 = \sum x_i^2 - n\bar{X}^2 \]
\[ \Sigma x_i^2 = \Sigma (x_i - \bar{X})(x_i - \bar{X}) = \Sigma x_i^2 - \bar{X}\Sigma x - \bar{X}\Sigma x + n\bar{X}^2 \]
\[ = \Sigma x_i^2 - 2\bar{X}\Sigma x + n\bar{X}^2 \quad \Sigma x / n = \bar{X} \quad \Sigma x = n\bar{X} \]
\[ = \Sigma x_i^2 - 2\bar{X}n\bar{X} + n\bar{X}^2 \]
\[ = \Sigma x_i^2 - 2n\bar{X}^2 + n\bar{X}^2 \]
\[ = \Sigma x_i^2 - n\bar{X}^2 \]

En küçük kareler yönteminin sapmalar kuralına göre \( \hat{\beta} \) den sonra \( \hat{\beta}_0 \) elde edelim. Bu amaç için normal denklemlerden ilki (\( \Sigma Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \Sigma X_i \)) kullanılır ve denklemin her iki yanı gözlem sayısına (n) bölünür.

\[ \sum Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum X_i \]
\[ \sum Y_i / n = n\hat{\beta}_0 + \hat{\beta}_1 \sum X_i / n \]

\[ \bar{Y} = \hat{\beta}_0 + \hat{\beta}_1\bar{X} \]

elde edilir. Eşitlik \( \hat{\beta}_0 \) için yeniden düzenlendiğinde;

\[ \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X} \]

sonucuna ulaşılmaktadır. Değişkenlerin gözlemlenen değerleri (\( X_i, Y_i \)) yerine, değişkenlerin ortalamalarından farklılarından (\( x_i, y_i \)) hareket edildiğinde \( \hat{\beta}_0 \) ve \( \hat{\beta}_1 \) için aynı sonuçlar elde edilecektir.
\[ \hat{\beta}_i = \frac{\sum x_i y_i}{\sum x_i^2} \]

eşitliğini aşağıdaki gibi farklı biçimlerde de yazmak mümkündür:

\[ \hat{\beta}_i = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\sum X_i y_i}{\sum X_i^2} = \frac{\sum x_i y_i}{\sum x_i^2} \]

Şöyle ki;

\[ \sum x_i = \sum (X - \bar{X}) = 0 \]
\[ \sum x_i y_i = \sum x_i (Y_i - \bar{Y}) = \sum x_i Y_i - \bar{Y} \sum x_i \]
\[ \sum x_i = \sum (X_i - \bar{X}) = 0 \]

olduğu için

\[ \sum x_i y_i = \sum x_i Y \]

olacaktır.

Benzer şekilde, \[ \sum x_i y_i = \sum X_i y_i \]
olduğunun göstermek mümkündür.

Açıklayıcı bir örnek

Bölüm 2’de 60 ailenin haftalık tüketim rakamları ve haftalık harcanabilir gelirlerine ait hipotetik verilerden “Örnek I” ve “Örnek II” olmak üzere iki örnek çekilmiş, Bölüm 3’te ise aynı hipotetik veriler kullanılarak ana kütle regresyon denklemi aşağıdaki gibi oluşturulmuş idi.

\[ E(Y \mid X_i) = 17 + 0.6X_i \]

Bu aşamada, ana külteden çekilen iki örnek için örnek kütle regresyon modelleri tahmin edilecektir.

**Örnek I için örnek kütle regresyon modelinin tahmini:**

- Normal Denklemler

\[ \hat{\beta}_0 = \frac{\sum Y \sum X}{n \sum X_i^2} \]
\[ \hat{\beta}_1 = \frac{\sum X_i Y_i}{n \sum X_i^2} = \frac{n \sum X_i Y_i - \sum X_i \sum Y_i}{n \sum X_i^2 - (\sum X_i)^2} = \frac{(1211 \times 322000) - (226020 \times 1700)}{(10 \times 322000) - (17000)^2} = 17.29 \]

- \[ \hat{\beta}_1 = \frac{n \sum X_i Y_i}{n \sum X_i^2} = \frac{n \sum X_i Y_i - \sum X_i \sum Y_i}{n \sum X_i^2 - (\sum X_i)^2} = \frac{(10 \times 226020) - (1700 \times 1211)}{(10 \times 322000) - (17000)^2} = 0.61 \]
\( \hat{Y} \hat{Y} = 17.29 + 0.61X, \) Örnek I regresyon modelinin tahmini

- Sapmalar Yöntemi
\[
\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\sum x_i y_i - \bar{X}\bar{Y}}{\sum x_i^2 - n\bar{X}^2} = \frac{226020 - (10 \times 170 \times 121.1)}{322000 - (10 \times 170^2)} = 0.61
\]
\[
\hat{\beta}_0 = \bar{Y} - \hat{\beta}\bar{X} = 121.1 - (0.61) \times 170 = 17.29
\]
\( \hat{Y} \hat{Y} = 17.29 + 0.61X, \) Örnek I regresyon modelinin tahmini

Beklenildiği üzere en küçük kareler yönteminin her iki uygulama biçimi de aynı sonucu vermiştir.

Örnek II için örnek kütle regresyon modelinin tahmini:

- Normal Denklemler
\[
\hat{\beta} = \frac{\sum Y \sum X}{n \sum x_i^2} = \frac{\sum Yx_i^2 - \sum Yx_i \sum X_i}{(\sum X_i^2 - (\sum X_i)^2)} = \frac{(1207 \times 322000) - (226800 \times 1700)}{(10 \times 322000) - (1700)^2} = 9.37
\]
\[
\hat{\beta}_0 = \frac{\sum Y}{n} - \hat{\beta}\frac{\sum X}{n} = \frac{n\sum Y - \sum X_i \sum Y_i}{n\sum x_i^2 - (\sum X_i^2)^2} = \frac{(10 \times 226800) - (1700 \times 1207)}{(10 \times 322000) - (1700^2)^2} = 0.65
\]
\( \hat{Y} \hat{Y} = 9.37 + 0.65X, \) Örnek II regresyon modelinin tahmini

- Sapmalar Yöntemi
\[
\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\sum x_i y_i - n\bar{X}\bar{Y}}{\sum x_i^2 - n\bar{X}^2} = \frac{226800 - (10 \times 170 \times 120.7)}{322000 - (10 \times 170^2)} = 0.65
\]
\[
\hat{\beta}_0 = \bar{Y} - \hat{\beta}\bar{X} = 120.7 - (0.65) \times 170 = 9.37
\]
\( \hat{Y} \hat{Y} = 9.37 + 0.65X, \) Örnek II regresyon modelinin tahmini

Böylece aynı ana küteden çekilen iki farklı örneğe ait regresyon model tahminlerinin
farklı oldu ğu görülmüştür.
**Alıştırma 1:**

Aşağıda bir spor giyim mağazasının 5 aylık satış gelirleri ile reklam harcamalarına ilişkin veriler yer almaktadır. Bu verileri kullanarak örnek regresyon denklemini tahmin ediniz.

<table>
<thead>
<tr>
<th>Aylar</th>
<th>Satış Gelirleri (Y) 1000 TL</th>
<th>Reklam Harcamaları (X) 100 TL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>5</td>
</tr>
</tbody>
</table>

**Tablo 1.1:** Satış Gelirleri ve Reklam Harcamaları

Örnek kütle regresyon modelini tahmin edelim. Öncelikle bağımlı ve bağımsız değişkenlerin belirlenmesi gerekir. Satışlar reklam harcamalarının bir fonksiyonu diğer bir ifade ile nedensellik reklam harcamalarından satışlara doğru olduğu için bağımlı değişken satışlar (Y) bağımsız değişken reklam harcamaları (X) olacaktır.

\[
\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i
\]

Örnek regresyon modeli, en küçük kareler yöntemiyle yukarıda anlatıldığı üzere iki şekilde oluşturulur. Bunlardan ilk X ve Y’nin gözlemlenen değerlerine diğer ise X ve Y’nin gözlemlenen değerlerinin kendi ortalamalarından farklarına EKK yönteminin uygulanmasıdır.

Öncelikle X ve Y’nin gözlemlenen değerlerini kullanarak örnek regresyon modelini tahmin edilecektir. Bu amaçla;

\[
\sum Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum X_i
\]

\[
\sum Y_i X_i = \hat{\beta}_0 \sum X_i + \hat{\beta}_1 \sum X_i^2
\]

normal denklemlerdeki değişkenlerle ilgili unsurların hesaplanması gerekecektir.
Tablo 1.2: Çözüm

Yukarıda hesaplanmış değerler denklemden yerine konursa normal denklemler;

\[ 23 = 5\hat{\beta}_0 + \hat{\beta}_1 15 \]
\[ 81 = \hat{\beta}_0 15 + \hat{\beta}_1 55 \]

\[ \hat{\beta}_0 = \frac{\Sigma Y \Sigma X}{n \Sigma X_i^2} = \frac{\Sigma Y \Sigma X_i^2 - \Sigma Y_i X_i \Sigma X_i}{n \Sigma X_i^2 - (\Sigma X_i)^2} = \frac{23.55 - 81.15}{5.55 - (15)^2} = 1.0 \]

\[ \hat{\beta}_1 = \frac{n \Sigma Y_i}{\Sigma X_i} = \frac{n \Sigma Y_i X_i - \Sigma X_i \Sigma Y_i}{n \Sigma X_i^2 - (\Sigma X_i)^2} = \frac{5.81 - 15.23}{5.55 - (15)^2} = 1.2 \]

\[ \hat{\beta}_0 = 1.0 \quad , \quad \hat{\beta}_1 = 1.2 \]

olan hesaplanmıştır. Buna göre örnek regresyon modeli;

\[ \hat{Y}_i = 1.0 + 1.2 X_i \]

Gözlemленen ve tahmini değerler ile örnek regresyon doğruşu aşağıdaki şekilde verilmiştir:
Şekil 1.1: Serpilme Diyagramı

Yukarıdaki verilere göre tahmin edilen regresyon modeli yorumlanır. Reklam harcamaları \( X \) sıfır olduğunda satış gelirleri 1000 $ olacaktır. \( \hat{\beta}_0 = 1.0 \) sabit parametredir ve geometrik açıdan regresyon doğrusunun dikey eksenini kestiği noktadır. \( \hat{\beta}_1 = 1.2 \) eğim katsayısıdır. \( X \) değişkeni \( \Delta X \) miktarı kadar değiştirirse \( Y \)’deki değişme \( \Delta Y = 1.2 \Delta X \) kadrardır. Reklam harcamaları 1 birim (100 $) artarsa satış gelirleri ortama 1,2 birim (1200$) artacaktır.

İkinci olarak en küçük kareler yönteminin diğer bir uygulama biçimi olan sapmalar yöntemiyle de aynı sonuca ulaşmak mümkündür. Bu amaç için aşağıdaki denklemler kullanılabılır.

\[
\hat{\beta}_i = \frac{\sum x_i y_i}{\sum x_i^2} \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_i \bar{X}
\]

Öncelikle yine denklemlerde yer alan unsurların hesaplanması gerekecektir. Ancak burada öncelikle \( \hat{\beta}_i \) tahminin (\( \hat{\beta}_i \)) yapılması gerekir.

\[
\bar{X} = \frac{\sum X}{n} = \frac{15}{5} = 3 \quad \bar{Y} = \frac{\sum Y}{n} = \frac{23}{5} = 4.6
\]

\[
\sum x_i y_i = \sum X_i Y_i - n\bar{X}\bar{Y} = 81 - 5.3(4.6) = 12
\]

\[
\sum x_i^2 = \sum x_i^2 - n\bar{X}^2 = 55 - 5.(3)^2 = 10
\]
\[ \hat{\beta}_1 = \frac{\sum x_iy_i}{\sum x_i^2} = \frac{12}{10} = 1.2 \]

ve

\[ \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} = 4.6 - (1.2)3 = 1.0 \]

olarak hesaplanır. Örnek regresyon modelinde tahmin edilen parametreler yer aldığında, \( \hat{Y}_i = 1.0 + 1.2X_i \) sonucuna ulaşılır.

Görüldüğü üzere en küçük kareler yönteminin değişkenlerin gözlemlenen değere \((X_i,Y_i)\) ve gözlemlenen değerlerin ortalamalarından farklarına \((x_i,y_i)\) göre uygulanması sonucu değiştirmemiştir.

Bu aşamada daha önce gördüğümüz özellikleri bu örnek üzerinde tekrarlayalım. Gözlemler itibariyle kalıntıları hesaplayalım. Kalıntı için;

\[ \hat{u}_i = Y_i - \hat{Y} = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i \]

esitliği verildiğine göre bu örnek için kalıntılar \( i=1,2,...,5 \) e kadar gözlem için \( \hat{u}_i = Y_i - 1.0 - 1.2X_i \) esitliğinden hesaplanacaktır. Dolayısıyla öncelikle \( \hat{Y}_i \) 'nin bulunması gerekir.

<table>
<thead>
<tr>
<th>X</th>
<th>X^2</th>
<th>XY</th>
<th>( \hat{Y}_i )</th>
<th>( \hat{u}_i = Y_i - \hat{Y}_i )</th>
<th>( \hat{u}_i^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1</td>
<td>3</td>
<td>2.2</td>
<td>0.8</td>
<td>(0.8)^2</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>8</td>
<td>3.4</td>
<td>0.6</td>
<td>(0.6)^2</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>6</td>
<td>4.6</td>
<td>-2.6</td>
<td>(-2.6)^2</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>24</td>
<td>5.8</td>
<td>0.2</td>
<td>(0.2)^2</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>40</td>
<td>7.0</td>
<td>1</td>
<td>1^2</td>
</tr>
<tr>
<td>( \sum Y_i = 23 )</td>
<td>( \sum X_i = 15 )</td>
<td>( \sum X_i^2 = 55 )</td>
<td>( \sum Y_iX_i = 81 )</td>
<td>( \sum \hat{Y}_i = 23 )</td>
<td>( \sum \hat{u}_i = 0 )</td>
</tr>
</tbody>
</table>

Tablo 1.3: Çözüm

1. gözlem için \( \hat{Y}_i \) aşağıdaki gibi hesaplayalım.
\[ \hat{Y}_i = 1.0 + 1.2X_i = 1.0 + 1.2(1) = 2.2 \]

Benzer şekilde örnek regresyon denkleminde \( X \) yerine sırasıyla 2, 3, 4, 5 rakamları verilir ve hesaplamalar yapılırsa diğer satırlarda yer alan gözlemlerdeki \( \hat{Y} \) değerleri de bulunacaktır.

Gözlemler itibarıyla \( \hat{Y} \) hesaplandırıguna göre kalıntılar;

\[ \hat{u}_i = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i = Y_i - 1 - 1.2X_i \]

Birinci gözlem için \( \hat{u}_i \)'yi aşağıdaki gibi hesaplayalım:

\[ \hat{u}_i = Y_i - \hat{Y}_i = 3 - 2.2 = 0.8 \]

Aynı şekilde diğer gözlemler için kalıntılar hesaplanır.

**4.3. En Küçük Kareler Yönteminin Özellikleri**

En küçük kareler yönteminin özellikleri aşağıdaki gibi sıralanabilir.

1. En küçük kareler tahmin edicileri (\( \hat{\beta}_0 \) ve \( \hat{\beta}_1 \)) \( X \) ve \( Y \) değişkenlerin gözlenen değerleri ile hesaplanabilir.

2. EKK tahmin edicileri nokta tahminlerdir. Örnek veriyken örnekten tahmin edilen parametre, ana kütledeki karşılığı için tek bir nokta tahmini verir.

3. Regresyon doğrusu \( A(\bar{X}, \bar{Y}) \) noktası sıradan geçer.

**Şekil 1.2:** Örneklemin Regresyon Doğrusunun Gösterimi

4. Bağımlı değişkenin tahmini değerlerinin ortalamaşı (\( \bar{Y} \)), gözlemlenen değerlerinin ortalamasına (\( \bar{Y} \)) eşittir. Buna göre,
$\bar{Y} = \bar{\hat{Y}}$

olduğunu aşağıda gösterelim.

Örnek regresyon modelinde $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_i X_i$ , $\hat{\beta}_0$ yerine aşağıdaki eşitini yazabiliriz:

$$\hat{\beta}_0 = \bar{Y} - \bar{\hat{X}}$$

$$\hat{Y}_i = (\bar{Y} - \hat{\beta}_i \bar{X}) + \hat{\beta}_i X_i = \bar{Y} - \hat{\beta}_i (X_i - \bar{X})$$

Tek bir gözlem için verilen bu eşitlik $i=1,2,\ldots,n$'ye kadar tüm gözlemler için kurulup her iki tarafın toplamı alındığında;

$$\sum \hat{Y}_i = \sum [\bar{Y} - \hat{\beta}_i (X - \bar{X})] = n\bar{Y} - \hat{\beta} \sum X_i \Sigma = \Sigma (X_i - \bar{X}) = 0$$

olduğundan

$$\sum \hat{Y}_i = n\bar{Y}$$

Her iki taraf gözlem sayısı $n$ ile bölündüğünde;

$$\frac{\sum \hat{Y}_i}{n} = \frac{n\bar{Y}}{n}$$

$$\bar{\hat{Y}} = \bar{Y}$$ sonucuna ulaşılr.

Bu özellik, modelde sabit parametre de olduğu sürece geçerlidir.


$$\sum \hat{u}_i = \sum Y_i - \hat{Y}_i = 0 \quad , \quad \bar{u} = 0$$

Bu özellik EKK yönteminin $\sum \hat{u}_i^2$ nın minimum olması varsayımına dayanmaktadır. Kalıntıların karelerinin toplamı fonksiyonunda $\hat{\beta}_i$ ‘e göre birinci mertebeden kısmi türevi alınarak $\sum \hat{u}_i = 0$ olduğu gösterilebilir:

$$\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_i} = \sum 2 (Y_i - \hat{\beta}_0 - \hat{\beta}_i X_i)(-1) = 0$$

$$= \sum [Y_i - \hat{\beta}_0 - \hat{\beta}_i X_i]$$

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

$$= \sum \hat{u}_i = 0$$

Eşitliğin her iki yanı n’ye bölündüğünde,
Bu özellik ile örnek regresyon modeli, sadece gözlemlenen \( X_i \) ve \( Y_i \) değerleriyle değil, \( X \) ve \( Y \)'nin ortalama değerlerinden sapmaları biçiminde de yazılabilmektedir. Bağımlı değişkenin gözlemlenen değerini;

\[
Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{u}_i \text{ ile ifade etmiştir.}
\]

Tek bir gözlem için yazılımış \( i=1,2,\ldots,n \) için tüm gözlemler için yazılıp her iki tarafın toplamı alınrsa,  

\[
\sum Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum X_i + \sum \hat{u}_i = 0 \text{ olduğundan denklem;}
\]

\[
\sum Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum X_i \quad \text{şekline indirgenir.}
\]

Her iki taraf gözlem sayısı \( n \)’ye bölündüğünde ( \( \sum Y_i / n = \bar{Y} \), \( \sum X_i / n = \bar{X} \) eşitliklerinden)

\[
\bar{Y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}
\]

elde ederiz. Sonraki aşamada \( Y_i - \bar{Y} \) oluştururulur.

\[
Y_i - \bar{Y} = (\hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{u}_i) - (\hat{\beta}_0 + \hat{\beta}_1 \bar{X}) = \hat{\beta}_1 (X_i - \bar{X}) + \hat{u}_i
\]

\[
y_i = \hat{\beta}_1 x_i + \hat{u}_i
\]

Ana kütle regresyon modeli elde edilir. Yeni elde edilen ana regresyon modelinde \( X_i \) ve \( Y_i \) yerini kendi örneklem ortalamalarından sapmalarına \( y_i (= Y_i - \bar{Y}) \) ve \( x_i (= X_i - \bar{X}) \) bırakmıştır. Bundan dolayı regresyon modelinin **sapmalı kalıbı** olarak adlandırılır. Bu durumda örnek regresyon modeli,

\[
y_i = \hat{\beta}_1 x_i \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \quad \text{tir ve} \quad y_i = \hat{\beta}_1 x_i + \hat{u}_i, \text{ve} \quad \hat{u}_i = y_i - \hat{\beta}_1 x_i \text{ yazabiliriz.}
\]


\[
\sum \hat{u}_i x_i = 0
\]

Bu durum, kalıntıların karelerinin toplamı fonksiyonunda \( \hat{\beta}_1 \) e göre birinci mertebeden kısımı türev alınarak gösterilebilir.

\[
\frac{\partial\sum \hat{u}_i^2}{\partial \hat{\beta}_1} = \sum 2(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)(-X) = 0
\]
= \sum [Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i)] (X_i) = 0

= \sum \hat{u}_i X_i = 0

7. Kalıntılar ile bağımlı değişkenin tahmin edilen değerleri (\hat{Y}) arasında ilişki yoktur.
\sum \hat{Y}_i \hat{u}_i = 0

Bu özellikle yukarıda elde edilen sapmalı kalıp yardımıyla açıklanabilir.
\sum \hat{Y}_i \hat{u}_i = \hat{\beta}_1 \sum x_i \hat{u}_i

= \hat{\beta}_1 \sum x_i (y_i - \hat{\beta}_1 x_i)

= \hat{\beta}_1 \Sigma x_i y_i - \hat{\beta}_1^2 \Sigma x_i^2 \hat{\beta}_1 = \frac{\Sigma x_i y_i}{\Sigma x_i^2} \text{ eşitliğinden}

= \hat{\beta}_1 \Sigma x_i y_i - \left( \frac{\Sigma x_i y_i}{\Sigma x_i^2} \right)^2 \Sigma x_i^2

= \hat{\beta}_1 \Sigma x_i y_i - \hat{\beta}_1 \Sigma x_i y_i = 0 \text{ elde edilir.}

4.4. Uygulama


İlk aşamada Tablo 1.1’de yer alan veri Excel’e girilir:

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Y</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Şekil 1: Veri Girişi
Veri girişinden sonra, Excel programında yer alan menülerden “Veri” menüsünden “Veri Çözümleme” tıklanır.

Şekil 2: Veri Menüsü

“Veri Çözümleme” tıklandığında ekrana aşağıdaki gibi yeni bir menü gelir:

Şekil 3: Veri Çözümleme Menüsü

Bu menüde yer alan “Regresyon” seçeneği seçip “Tamam” düğmesi tıklandığında ekrana gelen menü aşağıdaki gibi olacaktır:

Şekil 4: Regresyon Menüsü

Bu menüde yer alan Y giriş aralığı, bağımsız değişkene ait verinin referans adresinin girileceği yer, X giriş aralığı ise bağımlı değişken(ler)e ait verinin referans adresinin girileceği kısımdır. Bu giriş yerinin sağ tarafında yer alan düğmesi tıklandığında, ekrana
verisetinin fare imleci ile girilebileceği yeni bir menü gelir, veri seti fare imleci ile seçildikten sonra bu düğme yeniden tıklanıldığında verisetinin aralığı giriş aralığına girilmiş olur.

Örneğin; Şekil 1’deki A1:A5 aralığında yer alan veri bağımlı değişken olarak B1:B5 aralığında yer alan veri ise bağımsız değişken olarak seçildikten sonra, ilgili giriş aralıklarına veri girildikten sonra “Regresyon” menüsü aşağıdaki hâle gelir:

![Şekil 5: Regresyon](image)

Bu çıktının en altında yer alan kısımda en küçük kareler tahmin sonuçları görülmektedir. Kesişim (sabit) terimi 1, X bağımsız değişkeninin katsayısı ise 1,2 olarak bulunmuştur. Aynı zamanda tahmin çıktılarında her değişkene ait t-istatistikleri de yer almaktadır. Bu sonuçlarda yer alan “Düşük %95” ile “Yüksek %95” başlıklar ise güven aralıklarını göstermektedir.
Uygulamalar
Uygulama Soruları
Bu Bölümde Ne Öğrendik Özeti
Bölüm Soruları

1. Parametre tahmininde en küçük kareler yönteminin sıklıkla tercih edilme nedenleri nelerdir?

2. Basit regresyon modeli tahmin sürecinde, normal denklemlerin nasıl elde edildiğini gösteriniz.

3. Normal denklemlerden yararlanarak, basit regresyon modeli tahmincilerini elde ediniz.

4. En küçük kareler yönteminin temel özelliklerini maddeler halinde kısaca belirtiniz.

5.-7. sorularda ver alan boşlukları doldurunuz.

5. En küçük kareler yöntemi tahmin süreci, ..........nın minimize edilmesi temeline dayanmaktadır.


7. Regresyon doğrusu .... dan geçer. (Bu varsayımı şekil yardımcıla açıklayınız.)

8. \( \sum \hat{u}_i X_i = 0 \) varsayımını açıklayınız.

9. \( \bar{\hat{Y}} = \bar{Y} \) olduğunu ispatlayınız.

10. Aşağıda X ve Y değişkenlerine ait veriler verilmiştir.

\[
\begin{array}{cc}
  Y & X \\
  4 & 1 \\
  6 & 2 \\
  7 & 3 \\
  7 & 4 \\
  9 & 5 \\
 11 & 6 \\
\end{array}
\]
a) En küçük kareler yönteminin her iki uygulaması ile normal denklemler, sapmalar yöntemi- örnek regresyon denklemini tahmin ederek yorumlayınız.

b) Koordinat sisteminde bağımlı değişkenin gözlemlenen değerlerini, örnek regresyon doğrusunu, kalıntıları gösteriniz.

c) En küçük kareler kalıntıların toplamanın sıfır olduğunu gösteriniz

d) \[ \sum \hat{\alpha} \cdot x \] hesaplayınız.
5. EN KÜÇÜK KARELER TAHMİNCİLERİNİN STANDART HATALARI VE REGRESYON DOĞRUSUNUN UYGUNLUĞU
Bu Bölümde Neler Öğreneceğiz?

5. EN KÜÇÜK KARELER TAHMİNCELERİNİN STANDART HATALARI VE REGRESYON DOĞRUSUNUN UYGUNLUĞU

5.1. En Küçük Kareler Tahmincilerinin Varyans ve Kovaryansı

5.2. Regresyon Doğrusunun Verilere Uygunluğu

5.2.1. Belirginlik Katsayısı

5.2.2. Tahminin Standart Hatası

5.2.3. Uyum Katsayısı ve Uygunluk katsayları
Bölüm Hakkında İlgı Oluşturan Sörular
Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri
Anahtar Kavramlar
Giriş

Bu bölümde en küçük kareler tahmincilerinin varyans ve kovaryanslarının elde edilmesi, elde edilen bu varyans ve kovaryansların özelliklerinin açıklanması konuları detaylı olarak ele alınmıştır. Bunun yanında tahmin edilen regresyon doğrusunun uygunluğu; belirginlik katsayısı, tahminin standart hata, uyum ve uygunluk katsayıları kavramları çerçevesinde açıklanacaktır.
5. EN KÜÇÜK KARELER TAHMİNCİLERİNİN STANDART HATALARI VE REGRESYON DOĞRUSUNUN UYGUNLUĞU

5.1. En Küçük Kareler Tahmincilerinin Varyans ve Kovaryansı

Daha önce görüldüğü üzere $\hat{\beta}_0$ ve $\hat{\beta}_1$ ’in en küçük kareler tahmincileri örnek verilerinin bir fonksiyonu olup değerleri örnekten örnege değişmektedir. Ekonometrik modelin başarısı açısından $\hat{\beta}_0$ ve $\hat{\beta}_1$ ’in kesinliği ve güvenilirliği önemlidir. Öyleyse bu değişkenliği anlamak, bir tahmincinin güvenilirliğini ve hassasiyetini değerlendirmek açısından önemli bir ipucudur. Bu aşamada $\hat{\beta}_0$ ve $\hat{\beta}_1$ ’in varyans ve standart hataları elde edilecektir. Bilindiği üzere istatistikte bir tahminin kesinliği, standart hatasının ölçülmesiyle mümkündür.

Y ve X ana kütlesinden n gözlemli yinelemeli örnekler çekerek her örnek için $\hat{\beta}_0$ ve $\hat{\beta}_1$ parametreleri tahmin edilsin. Eğer mümkün örneklerin tamamı çekilmişse $\hat{\beta}_0$ ve $\hat{\beta}_1$ ’in ortalama değeri bekленen değere eşit olacaktır.

$$E(\hat{\beta}_0) = \beta_0$$
$$E(\hat{\beta}_1) = \beta_1$$

Kısaca örnekten hesaplanan $\hat{\beta}_0$ ve $\hat{\beta}_1$ lin beklenen değeri, ana kütüde $\beta_0$ ve $\beta_1$ ’e eşit olacaktır.

Şekil 5.1: Beklenen Değerlerin Gösterimi

$\hat{\beta}_0$ ve $\hat{\beta}_1$ için varyans ve standart hata aşağıdaki eşitlikler kullanılarak hesaplanır:
\[ Var(\hat{\beta}_0) = E\left[ (\hat{\beta}_0 - E(\hat{\beta}_0))^2 \right] = E\left( \hat{\beta}_0 - \beta_0 \right)^2 = \frac{\sum X_i^2}{n} \sigma_u^2 \quad E(\hat{\beta}_0) = \beta_0 \text{ eşit olduğundan} \]

\[ SE(\hat{\beta}_0) = \sqrt{\frac{\sum X_i^2}{n \sum x_i^2} \sigma_u^2} \]

\[ Var(\hat{\beta}_1) = E\left[ (\hat{\beta}_1 - E(\hat{\beta}_1))^2 \right] = E\left( \hat{\beta}_1 - \beta_1 \right)^2 = \frac{\sigma_u^2}{\sum x_i^2} \quad E(\hat{\beta}_1) = \beta_1 \text{ eşit olduğundan} \]

\[ SE(\hat{\beta}_1) = \sqrt{\frac{\sigma_u^2}{\sum x_i^2}} \]


\[ \hat{\beta}_0 \text{ in varyansı } Var(\hat{\beta}_1) \text{ cinsinden } Var(\hat{\beta}_0) = (\sum X_i^2 / n) Var(\hat{\beta}_1) \text{ şeklinde yeniden yazılabilir.} \]

\[ \hat{\beta}_0 \text{ ve } \hat{\beta}_1 \text{ in varyans ve dolayısıyla standart hata formüllerinde yer alan unsurlar } \sigma_u^2 \text{ dışındakiler, } X \text{ ve } Y \text{ ye ilişkin gözlemlenen veriler ile ilgilidir ve hepsi kolaylıkla hesaplanabilir. Dolayısıyla } \hat{\beta}_0 \text{ ve } \hat{\beta}_1 \text{ in varyanslarının hesaplanabilmesi } \sigma_u^2 \text{ nin bilinmesine bağlıdır.} \]

\[ \sigma_u^2 \text{ hata teriminin sabit varyansıdır. Hata teriminin varyansı, gözlemlenen } Y_i \text{ } \text{lerin, beklenen değer } E(Y_i) \text{ etrafındaki dağılımlarını göstermektedir. } \sigma_u^2 \text{ ne kadar büyükse dağılım o derece geniş ve } Y_i \text{ } \text{lerin } E(Y_i) \text{ etrafında nereye düştüğüne dair kesinlik azalır. Ana kütle yerine örnekten hareket edildiğinde ana kütle regresyonu ve dolayısıyla ana kütle hata terimi bilinmediği için hata teriminin varyansı ( } \sigma_u^2 \text{ ) da hesaplanamaz. Bu durumda örnek verilerinden yararlanarak ana kütle hata teriminin varyansı tahmin edilir. Ana kütle hata teriminin varyansı:} \]

\[ \hat{\sigma}_u^2 = \frac{\sum \hat{u}_i^2}{n - 2} \]
Ana kütle hata teriminin\( (\sigma^2_u)\), en küçük kareler tahminicisi\( (\hat{\sigma}^2_u)\) n-2 serbestlik derecesiyle \( \chi^2 \) dağılımlına uygundur. Varyans formülünün paydasındaki n-2 teriminde yer alan \( n \) gözlem sayısı, 2 ise modeldeki parametre sayısıdır. Payda ise kalıntı kareleri toplamı\( (\sum \hat{u}_i^2)\) yer almaktadır.

\( \hat{\sigma}^2_u \) regresyondan elde edilen bilginin bir ölçüsüdür. Böylelikle tahmin edilen regresyon doğrusunun uygulanığı hakkında bir ölçü olduğu ifade edilebilir. \( X_i \) veriyken hata teriminin koşullu varyansı \( \hat{\sigma}^2_y \) (kısaca \( \hat{\sigma}^2 \)) bağımlı değişkenin koşullu varyansına eşittir. \( \sigma_y^2 \) ise \( Y \)’nin koşulsuz varyansıdır ve \( \sigma_y^2 = \sum y_i^2 / (n-1) \) e (hatırlanacağı üzere \( \sum y_i^2 = \sum (Y_i - \bar{Y})^2 \) eşittir. \( \sigma_y^2 \), bağımlı değişkenin ortalama etrafındaki dağılmayı veren bir ölçüdür ve \( X_i \)’yi göz önüne almamaktadır. \( Y \)’nin \( X \)’e bağlı bulunduğu olup olmadığını önemli yoktur. Bağımlı değişkenin koşullu varyansı \( (\hat{\sigma}^2_{xy} = \hat{\sigma}^2_y) \) ise regresyon doğruşı etrafındaki dağılmayı vermeekte, bilgiyi \( X_i \)’den aldığı için belirsizlik azalmaktadır. Böylece \( Y \)’nin koşullu varyansı, koşulsuz varyansından küçüktür: \( \hat{\sigma}^2_u = \sigma^2 \leq \sigma_y^2 \)

Kalıntı kareler toplamı farklı yollardan hesaplanabilir. Bunlardan ilki,

\[
\sum \hat{u}_i^2 = \sum (Y_i - \hat{Y}_i)^2 = \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 \quad \text{dir.}
\]

Diğer bir yol ise,

\[
y_i = \hat{y}_i + \hat{u}_i
\]

\[
\sum y_i^2 = \sum (\hat{y}_i + \hat{u}_i)^2
\]

\[
\sum y_i^2 = \sum \hat{y}_i^2 + \sum \hat{u}_i^2 + 2 \sum \hat{y}_i \hat{u}_i
\]

\[
\sum y_i^2 = \sum \hat{y}_i^2 + \sum \hat{u}_i^2 \quad \text{çünkü} \quad \sum \hat{y}_i \hat{u}_i = 0 \quad \text{olduğundan}
\]

\[
\sum y_i^2 = \hat{\beta}_1^2 \sum x_i^2 + \sum \hat{u}_i^2 \quad \text{veya} \quad \hat{y}_i^2 = \hat{\beta}_1 x_i \quad \text{den}
\]

\[
\sum \hat{u}_i^2 = \sum y_i^2 - \hat{\beta}_1^2 \sum x_i^2
\]

devam edilirse ve

\[
\hat{\beta}_1 = \frac{\sum x_i y_i}{\sum x_i^2}, \quad \text{denklemde yer verilirse}
\]

\[
\sum \hat{u}_i^2 = \sum y_i^2 - \left( \frac{\sum x_i y_i}{\sum x_i^2} \right)^2 \sum x_i^2
\]

\[
\sum \hat{u}_i^2 = \sum y_i^2 - \left( \frac{\sum x_i y_i}{\sum x_i^2} \right)^2 \sum x_i^2
\]

\[
\sum \hat{u}_i^2 = \sum y_i^2 - \left( \frac{\sum x_i y_i}{\sum x_i^2} \right)^2 \sum x_i^2
\]

şeklindedir.
Ana kütle hata teriminin \( (\hat{\sigma}_u^2) \) artı değerli kare kökü tahminin standart hatasıdır.

\[
\hat{\sigma}_u = \sqrt{\frac{\sum \hat{u}_i^2}{n-2}} \quad \text{Tahminin standart hatası}
\]

Tahminin standart hatası, regresyon etrafındaki \( Y_i \) değerlerinin standart sapmasıdır ve örnek regresyonun standart sapması olarak da adlandırılır. Daha sonra görüleceği gibi, belirligin katsayısı gibi örnek regresyon doğrusunun verilere uygulunu gösteren bir ölçüdür. Çünkü \( \hat{\sigma}^2 \) aynı zamanda eğrisyondan elde edilen bilginin bir ölçüsüdür.

\( \hat{\beta}_0 \) ve \( \hat{\beta}_1 \) tahmin edicileri örnekten örnekle değiştiği gibi, aynı örneklemde birbiriyle bağımlılı olabilirler. Bu bağımlılık ise ortak varyansla ölçülür.

\[
Kov(\hat{\beta}_0, \hat{\beta}_1) = -\bar{X} Var(\hat{\beta}_1) = -\bar{X} \left( \frac{\sigma_u^2}{\sum X_i^2} \right)
\]

\( \hat{\beta}_0 \) ve \( \hat{\beta}_1 \)'in Varyansları ve Kovaryansının Özellikleri

1. Hata teriminin varyansı yukarıdaki formüllerin her birinde görülmektedir. Gözlemленen \( Y_i \)'lerin beklenen değer \( E(Y_i) \) etrafındaki dağılımını gösteren hata teriminin varyansı \( (\sigma_u^2) \) ne kadar büyüyse dağılım o derece genişleyecek ve \( Y_i \)'lerin \( E(Y_i) \) etrafında nereye düştüğine dair kesinlik azalacaktır. \( \sigma_u^2 \) büyükçe \( \beta_1 \) ve \( \beta_2 \)'ye ait sahip olduğumuz bilgi kesinliğini kaybeder, en küçük kareler tahmincilerinin varyansı büyür.

2. \( \sum x_i^2 = \sum (X_i - \bar{X})^2 \) ifadesi varyans ve kovaryans formüllerinin her ikisinde de bulunmaktadır. Bu ifade, bağımsız değişken \( X_i \)'nin kendi ortalamanasından ne kadar uzakta olduğunu ifade eder. Bu iki terim birbirinden uzaklaştırılacağı bağımsız değişkendeki değişimi gösteren \( \sum x_i^2 \) ifadesi büyüyecek, yaklaştıkça \( \sum x_i^2 \) ifadesi küçülecektir. Bağımsız değişkendeki değişim arttıkça diğer bir deyisle \( \sum x_i^2 \) ifadesi büyüyüşe en küçük kareler tahmincilerinin varyansları küçülür, daha güvenilir olurlar.

3. Örnek büyükluğu \( (n) \) arttıkça en küçük kareler tahmincilerinin varyans ve kovaryansları küçülür. Bu nedenle fazla gözlem ile çalışmak genellikle tercih edilir. Varyans ve kovaryans formüllerindeki her kareler toplam \( n \) gözlem için hesaplanmıştından varyans ve kovaryans formüllerinde her kareler toplamında \( n \) yer almaktadır. Gözlem sayısı \( n \) arttıkça \( \sum x_i^2 \) ifadesi büyür.

4. \( Var(\hat{\beta}_0) \) formülünde \( \sum X_i^2 \) ifadesi yer almaktadır. \( \sum X_i^2 \) büyüyüşe en küçük
kareler tahmincisi $\beta_0$'nin varyansı küçülmür. Sabit parametre $\beta_0$, $X=0$ olduğunda, bağımlı değişkenin beklenen değeri esittir. Veri seti $X=0$’dan ne kadar uzaksa kesme terimi sabit parametreyi hesaplamak o kadar zorlaşacaktır. $\sum X_i^2$ verilerin orijinden uzaklığını ölçer. Veri setindeki $X_i$ değerleri 0’a yakınsa $\sum X_i^2$ ve dolayısıyla $Var(\hat{\beta}_0)$ küçülecektir. Negatif veya pozitif olması fark etmeksizin X veri setinin değerleri büyükse $\sum X_i^2$ büyüyecek, buna bağlı olarak $Var(\hat{\beta}_0)$ da büyüyecektir.

5. $\beta_0$ ve $\beta_1$ için ortak varyans formüllерinde $X_i$'nin örneklem ortalamaları yer almaktadır. $\bar{X}$ 'nin değeri arttıkça ortak varyans mutlak değer olarak büyüyecektir. $\beta_0$ ve $\beta_1$ arasındaki ortak varyansın işareti, bağımsız değişken $X$'in ortalamanın $(-\bar{X})$ tersidir. $\bar{X}$ eksi ise, ortak varyans (+) işaretli, artı ise (-) işaretli olacaktır.

Şekil 5.2: Ortak Varyans

Örnek ortalama pozitif ise, eğim ve sabit parametre nin en küçük kareler tahmincileri arasında negatif ortak varyans vardır.

5.2. Regresyon Doğrusunun Verilere Uygunluğu

5.2.1. Belirginlik Katsayısı

Belirginlik katsayısı ($r^2$), örnek regresyon doğrusunun verilere uygulugunu gösteren bir ölçüsüdür. Örneklem verileri kullanılarak tahmin edilen regresyon modelinin belirlediği regresyon doğrusu, serpilme diyagramında gözlemlenen değerlerin arasından geçmektedir. Regresyon doğrusunun verilere ne kadar yakın olduğu, tahmin edilen modelin başarısı açısından önemli idi.

Belirginlik katsayısı, bağımlı değişkende meydana gelen değişmenin ne kadarnın bağımsız değişken ve/veya değişkenlerdeki değişim tarafından açıklanabildiğini göstermektedir. Belirginlik katsayısı $0 \leq r^2 \leq 1$ arasında yer almaktadır. $r^2$, l’e yaklaştıkça regresyon doğrusunun verilere uygulugunu artmaktadır.


**Şekil 5.3:** Toplam Değişme, Regresyonla Açıklanabilen Değişme, Regresyonla Açıklanamayan Değişme

Yukarıdaki şekilde yer alan unsurlar aracılığıyla $r^2$ katsayısını kolayca hesaplamak mümkündür.

Şekilde görüldüğü üzere, bağımlı değişkenin i. gözlemdeki değeri $(Y_i)$ ile bağımlı değişkenin ortalaması $(\bar{Y})$ arasındaki fark $(Y_i - \bar{Y})$ iki parçadan oluşmaktadır.

i. Gözlem için;

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

$$(Y_i - \bar{Y}) = y_i$$

$(Y_i - \bar{Y}) = y_i$, X’ten bilgi almadan önceki değişkenliği gösterir. Bütün örneklem için aynı değişkenliği hesaplanabilmesi için yukarıdaki denklemin her iki tarafının karelerinin toplamı alınır. $\sum y_i = \sum (Y_i - \bar{Y}) = 0$ olduğundan dolayı, $y_i$’lerin toplamı değil, karelerinin toplamı alınmaktadır.
\[ \sum y_i^2 = \sum (\hat{y}_i + \hat{u}_i)^2 \]
\[ \sum y_i^2 = \sum \hat{y}_i^2 + \sum \hat{u}_i^2 + 2 \sum \hat{y}_i \hat{u}_i \quad \sum \hat{y}_i \hat{u}_i = 0 \quad \text{olduğundan} \]
\[ \sum y_i^2 = \sum \hat{y}_i^2 + \sum \hat{u}_i^2 \quad \hat{y}_i^2 = \hat{\beta}_i x_i \text{'den} \]
\[ \sum y_i^2 = \hat{\beta}_i^2 \sum x_i^2 + \sum \hat{u}_i^2 \]

sonucuna ulaşılır. Denklemdeki bağımlı değişkenin gözlemlenen değerlerinden bağımlı değişkenin ortalamasının farkı “\[ \sum y_i^2 = \sum (Y_i - \bar{Y})^2 \]” bağımlı değişkendeki toplam değişme olarak adlandırılır. Bağımlı değişkenin tahmini değerlerinden bağımlı değişkenin tahmini değerlerinin ortalamasının (daha önce anlatıldığı \( \bar{Y} = \bar{Y} \)’dir) farkı “\[ \sum \hat{y}_i^2 = \sum \hat{Y}_i - \bar{Y} = \sum \hat{Y}_i - \bar{Y} = \hat{\beta}_i^2 \sum x_i^2 \]” regresyonla açıklanabilen değişmedir. Bağımlı değişkenin gözlemlenen değeri ile tahmin değeri arasındaki fark “\[ \sum \hat{u}_i^2 = \sum \hat{Y}_i - \hat{Y}_i \]” ise regresyon ile açıklanamayan değişimye eşittir.
\[ \sum y_i^2 = \sum \hat{y}_i^2 + \sum \hat{u}_i^2 \]
veya
\[ \sum (Y_i - \bar{Y})^2 = \sum (\hat{Y}_i - \bar{Y})^2 + \sum (Y_i - \hat{Y}_i)^2 \]
denklemlerinin her iki yanı toplam değişimye (\( \sum y_i^2 = \sum (Y_i - \bar{Y})^2 \)) bölünserde;

\[ 1 = \frac{\sum \hat{y}_i^2}{\sum y_i^2} + \frac{\sum \hat{u}_i^2}{\sum y_i^2} \]

sonucuna ulaşılır. Regresyon ile açıklanabilen değişmenin toplam değişimye oranı belirginlik katsayısına (\( r^2 \)) eşittir.
\[ r^2 = \frac{\sum \hat{y}_i^2}{\sum y_i^2} = \frac{\hat{\beta}_i^2 \sum x_i^2}{\sum y_i^2} = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2} \]
esittir. Belirginlik katsayısını daha farklı biçimlerde göstermek mümkündür.
\[ 1 = r^2 + \frac{\sum \hat{u}_i^2}{\sum y_i^2} \]
\[ r^2 = 1 - \frac{\sum \hat{u}_i^2}{\sum y_i^2} \]
$r^2$ için başka hesaplama yolları da önerilebilir.

$$r^2 = \frac{\hat{\beta}^2 \sum x_i^2}{\sum y_i^2}$$

Eşitliğin pay ve paydası $n$ veya $n-1$’e bölündüğünde

$$r^2 = \frac{\hat{\beta}^2 \sum x_i^2 / n}{\sum y_i^2 / n}$$

bulunabilir.

Açıklar ki $\sum x_i^2 / n$ terimi, X bağımsız değişkeninin koşulsuz varyansına ($\sigma_X^2$), $\sum y_i^2 / n$ terimi ise Y bağımlı değişkeninin koşulsuz varyansına ($\sigma_Y^2$) eşittir. Nihayet

$$r^2 = \frac{\hat{\beta}^2 \sigma_X^2}{\sigma_Y^2}$$

sonucuna ulaşılır. Yine

$$r^2 = \frac{\sum \hat{y}_i^2}{\sum y_i^2} \ \text{den} \ \sum \hat{y}_i^2 = r^2 \sum y_i^2$$

ve yine

$$r^2 = 1 - \frac{\sum \hat{u}_i^2}{\sum y_i^2} \ \text{den} \ \sum \hat{u}_i^2 = (1-r^2)\sum y_i^2$$

dır.

Böylece

$$\sum y_i^2 = \sum \hat{y}_i^2 + \sum \hat{u}_i^2$$
denklemi $r^2$ ve toplam değişme açısından yeniden yazılabilir:

$$\sum y_i^2 = r^2 \sum y_i^2 + (1-r^2)\sum y_i^2$$

$1-r^2$ belirsizlik katsayısidır ve toplam değişmenin ne kadarının regresyonda yer alan değişkenlerin dışındaki değişkenler tarafından açıklanabildiğini gösterir. $r^2$ 1’e yaklaştırısa, $1-r^2$ ise 0’a yaklaştırıca regresyon doğrusunun verilere uyumu artacaktır.
sonuçuna ulaşırız. Yukarıdaki denklemden anlaşılmalıdır ki basit regresyon modeli için hesaplanan belirginlik katsayısının \( r^2 \) karekökü, \( X \) ile \( Y \) arasındaki doğrusal ilişkinin varlığını gösteren basit korelasyon katsayısına \( r_{XY} \) eşittir ve belirginlik katsayısından korelasyon katsayısına geçilebilmektedir. Burada üzerinde durulması gereken nokta şudur: Belirginlik katsayısı \( 0 \leq r^2 \leq 1 \) iken korelasyon katsayısı \( -1 \leq r \leq 1 \) değerleri arasında yer alır. Belirginlik katsayısından korelasyon katsayısına geçerken değişim arasındaki ilişkinin yönünü gösteren korelasyon katsayısının işaretleri, regresyon modelindeki \( \hat{\beta} \)’nin işaretleri ile aynıdır.

\[
r^2 = \frac{\hat{\beta}^2}{\sum y_i^2} = \frac{\sum \hat{y}_i}{\sum y_i^2} = \frac{\left( \sum x_i y_i / \sum x_i^2 \right)^2 \sum x_i^2}{\sum y_i^2}
\]

\[
r^2 = \frac{\left( \sum x_i y_i \right)^2}{\sum x_i^2 \sum y_i^2}
\]

\( r^2 = 1 \) ise bağımlı değişkende meydana gelen toplam değişimden tamami (%100) bagimsız değişken(ler)deki değişim ile açıklanılmaktadır. Söz konusu durum, tam uyumu ifade etmektedir. \( r^2 = 0 \) olması ise eğim parametresinin \( \hat{\beta} = 0 \) olması koşuluna bağlıdır. Bu durumda regresyon modeli \( Y = \hat{\beta} X \), regresyon doğrusu ise \( X \) eksenine paralel yatay bir doğru biçimindedir.

Her ne kadar \( r^2 \)’nin \( 1 \)’e yakın olması regresyon doğrusunun verilere uygunluğunu göstermemekte ise de zaman serisi verilerinin kullanılıldığı modellerde trendin etkisiyle yüksek \( r^2 \)’ye, yatay kesit verilerinin kullanılıldığı modellerde ise düşük \( r^2 \)’ye rastlamanak mümkündür. Dolayısıyla regresyon doğrusunun uygunluğu konusunda \( r^2 \)’ye göre yorum yaparken kullanılan veri türünü dikkate almamalıdır.

Ayrıca \( r^2 \), hem \( X \) hem de \( Y \)’nin ölçü birimine tabi değildir. Bu özelliğiyle verilerin regresyon doğrusuna uyumunu gösteren diğer ölçü birimlerinden tahminin standart hatasına üstünlik sağlamaktadır.

* Basit korelasyon katsayı sıfır noktasından ve ölçekten bağımsızdır. Basit korelasyon katsayısı iki değişken arasındaki doğrusal ilişkinin yönü ve derecesini tespit için kullanılır. \( r = 0 \) ise iki değişken arasında ilişki olmadığı anlamına gelir, doğrusal ilişkinin olmadığını işaretir. Örneğin \( Y = X^2 \) ilişkisinde \( r = 0 \)’dır. \( r \)’nin (-) veya (+) değer alması iki değişken arasındaki örneklem ortak varyansını işaretine bağlıdır.
$r^2$ bir istatistiksel test amacıyla test edilmesi gerekir. Ana kütte belirginlik katsayısı $\rho^2_{yx}$ ile ifade edilmektedir.

### 5.2.2. Tahminin Standart Hatası

Regresyon doğrunun verilere uynuluğunun ikinci bir ölçütü tahminin standart hatasıdır.

$$\hat{\sigma}_u = \sqrt{\frac{\sum (Y_i - \hat{Y}_i)^2}{n-2}} = \sqrt{\frac{\sum \hat{u}_i^2}{n-2}} \text{ Tahminin Standart Hatası}$$

Hatırlanacağı üzere tahminin standart hatası, ana kütte hata terimi varyansını (\(\hat{\sigma}_u = \sum \hat{u}_i^2 /n-2 \)) karekökündür. Verilerin regresyon doğrusu etrafındaki dağılmaların ölçüsü olan tahminin standart hatası (\(\hat{\sigma}_u\)) büyüklüğü $Y_i$ ile $\hat{Y}_i$ arasındaki farkın büyüklüğüne bağlıdır. Bu farkın dolayısıyla \(\sum (Y_i - \hat{Y}_i)^2\) farkının küçük çıkması $\hat{Y}_i$'lerin verilere ($Y_i$)lere yaklaştığını, regresyon doğrusunun verilere uydugunu gösterecektir.

Ancak tahminin standart hatasını bu amaçla kullanmak aşağıda belirtilen nedenlerden dolayı sakıncalıdır.


- Tahminin standart hatasının belirginlik katsayında (0 ≤ $r^2$ ≤ 1) olduğu gibi her durum için geçerli kesin sınırları yoktur. Regresyon doğrusunun tam uyumu durumunda, diğer bir ifade ile örnek regresyon doğrusunun $Y_i$ değerlerinden geçtiği durumda $Y_i$ ile $\hat{Y}_i$ arasındaki fark 0 olacaktır. Dolayısıyla tahminin standart hatası için alt sınır 0'dır. Diğer uç bir durumda, uyumsuzlukta ise bir sınır getirilememektedir.

### 5.2.3. Uyum Katsayısı ve Uygunluk Katsayıları

Örnek regresyon doğrunun uyanılığı için bir diğer seçeneği $Y_i$ 'ler ile $\hat{Y}_i$ 'ler arasındaki ilişkinin saptanmasıdır. Buna göre uyum katsayısı;

$$r^*_{yy} = \frac{\sum (Y_i - \bar{Y})(\hat{Y}_i - \bar{\hat{Y}})}{\sqrt{\sum (Y_i - \bar{Y})^2} \sqrt{\sum (\hat{Y}_i - \bar{\hat{Y}})^2}} = \frac{\sum Y_i \hat{Y}_i}{\sqrt{\sum Y_i^2} \sqrt{\sum \hat{Y}_i^2}} \leq 1$$

$r^*_{yy}$, $|\parallel| 'e yaklaştıkça regresyon doğrunun verilere uygunluğu artmakta, 0'a yaklaştıkça azalmaktadır. $r^*_{yy}$ nin işaret X ile Y değişkenlerinin değişme yönünü göstermez.
Uygunluk katsayısı ise,

$$r^2_{yy} = \frac{\left[ \sum (Y_i - \bar{Y})(\hat{Y}_i - \bar{Y}) \right]^2}{\sum (Y_i - \bar{Y})^2 \sum (\hat{Y}_i - \bar{Y})} = \frac{\sum (Y_i - \bar{Y})}{\sum (Y_i - \bar{Y})^2} = r^2$$

'y eşittir.

EK

Bu derste, ikinci dersin 2.5. başlığı altında da yer almakta olan veriler kullanılarak Excel programı ile regresyon tahmini yapılacaktır.

**Uygulama 1**

Bahsi geçen bölümdede Örnek 1’de yer almakta olan veri seti aşağıdaki gibidir:

**Şekil 1:** Birinci Örnek İçin Veri seti


**Şekil 2:** Veri Çözümleme

Ekrana gelen menüden Regresyon seçeneği aktif hâle getirildikten sonra Tamam düğmesi seçilince aşağıdaki pencere ekrana gelecektir:

**Şekil 3:** Regresyon Penceresi


**Şekil 4:** Regresyon Çıktıları

Şekil 4’tede görüleceği üzere sabit terimin katsayısı 17, 29; bağımsız değişken katsayısı ise 0,61 olarak bulunmuştur.
Uygulama 2

Örnek 2’de yer alan veri seti aşağıdaki gibidir:

Şekil 5: İkinci Örnek İçin Veri Seti

Önceki uygulamada verilen yönergelerin izlenmesi hâlinde elde edilecek tahmini sonuçlar aşağıdaki gibi olacaktır:

Şekil 6: İkinci Örnek İçin Regresyon Çıktıları

Şekil 6’da tahmin sonuçları, “Sonuçlar” isimli yeni bir sayfada oluşturulmuş olup alt sekme de bu sayfanın ismi görülmektedir.
Uygulamalar
Uygulama Soruları
Bu Bölümde Ne Öğrendik Özeti
Bölüm Soruları


3. Toplam değişme, regresyonla açıklanabilen değişme ve regresyonla açıklanamayan değişme kavramlarını şekil üzerinde göstererek açıklayınız.

4. Aşağıdaki ara sonuçlardan
\[
\sum Y_i = 480 \quad \sum X_i = 36 \quad \sum x_i y_i = -440 \quad \sum x_i^2 = 20 \quad n = 12
\]

a) Örnek regresyon denklemini tahmin ederek yorumlayınız.

b) Tahmin edilen parametrelerin standart hatalarını hesaplayınız.

c) Belirginlik katsayısını hesaplayarak yorumlayınız.

d) Basit korelasyon katsayısını hesaplayarak yorumlayınız.

e) Tahminin standart hatasını hesaplayınız.

5. İhracatın, GSMH ile açıkladığı ekonomerik modelin ara sonuçları aşağıda verilmiştir.
\[
\sum (Y_i - \bar{Y})^2 = 45.413 \quad \sum \hat{u}_i^2 = 11.625 \quad n = 16
\]

a) Toplam değişme, regresyonla açıklanabilen değişme, regresyonla açıklanamayan değişmeyi hesaplayınız.

b) Belirginlik katsayısını hesaplayarak yorumlayınız.

c) Basit korelasyon katsayısını hesaplayarak yorumlayınız.

d) Tahminin standart hatasını hesaplayınız.

6. \( \hat{Y} = 3.08 - 0.86X \quad n = 20 \quad r^2 = 0.99 \quad \hat{u}_i^2 = 502 \)

Yukarıdaki verilerden yola çıkarak

a) Korelasyon katsayısını hesaplayarak yorumlayınız.

b) Toplam değişmeyi hesaplayınız.

c) Regresyonla açıklanabilen değişmeyi hesaplayınız.
7. \( \text{Kov}(\hat{\beta}_0, \hat{\beta}_1) \) neyi ifade eder açıklayınız.

8. \( \hat{Y} = 1207 - 1.25X \quad n = 12 \quad \sum x_i^2 = 8.07 \quad \sum y_i^2 = 49.82 \)

Yukarıdaki verilere göre

a) Belirginlik katsayısını hesaplayarak yorumlayınız.

b) Basit korelasyon katsayısı hesaplayarak yorumlayınız.

c) Ana kütle hata teriminin varyansını hesaplayınız.

d) Toplam değişme, regresyonla açıklanabilen değişme, regresyonla açıklanamayan değişmeyi hesaplayınız.

9. \( \hat{Y} = 0.81 + 0.013X \quad n = 16 \quad \sum (X_i - \bar{X})^2 = 83.71 \quad \sum (Y_i - \bar{Y})^2 = 52.075 \)

ise belirginlik katsayısı neye eşittir?

\[ \sum y_i^2 = \sum \hat{y}_i^2 + \sum u_i^2 \]

denkleminin sadece belirginlik katsayısı ve toplam değişme unsurları ile yazılabileceğini gösteriniz.
6. ELASTİKİYET VE HATA TERİMLERİNİN NORMAL DAĞILIM VARSAYIMI
Bu Bölümde Neler Öğreneceğiz?

6. ELASTİKİYET VE HATA TERİMLERİNİN NORMAL DAĞILIM VARSAYIMI

6.1. Elastikiyet

6.2. Hata Terimlerinin Normal Dağılımı
Bölüm Hakkında İlgi Oluşturan Sorular
Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri
Anahtar Kavramlar
**Giriş**

Bu kısımda, tahmin edilen parametrelerin elastikiyet hesaplamaları ve elde edilen hesaplamaların yorumları üzerinde durulacaktır. Buna bağlı olarak elastikiyet katsayısı, ortalama elastikiyet ve nokta elastikiyetlerinin elde edilmesi konuları hakkında bilgiler verilecektir. Bölüm sonunda ise konunun pekiştirilmesi amacıyla çözümlü örneklerle yer verilecek, marjinal verimlilik kavramı ve marjinal verimliliği elde edilme yöntemi açıklanacaktır.

Hata teriminin dağılımı üzerinde durulacak, hata teriminin dağılımı ile en küçük kareler tahmincilerinin dağılımı ilişkilendirilecektir.
6. ELASTİKİYET VE HATA TERİMLERİNİN NORMAL DAĞILIM VARSAYIMI

6.1. Elastikiyet

Örnek regresyon modeli; \( \hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_i X_i \) şeklinde tahmin edilmiş olsun. Örnek regresyon modelinin belirlendiği örnek regresyon doğrusunun eğimi \( \hat{\beta}_i \) ye eşittir. Matematiksel olarak \( \hat{\beta}_i = \frac{d\hat{Y}}{dX} \)’dir. \( \hat{\beta}_i, X \) ’teki küçük bir değişmeye karşılık \( \hat{Y} \) ’deki değişmeye gösterir. Amaç değişkenler arasındaki elastikiyet katsayılarını hesaplamak ve yorumlamak ise \( \hat{\beta}_i \), elastikiyet katsayısının bir bileşenidir.

Elastikiyet katsayısı aşağıdaki formülle gösterilir.

\[
\varepsilon = \frac{dY}{dX} = \frac{dY}{X} \frac{X}{dX} = \frac{dY}{X} \frac{X}{Y}
\]

Tanım gereği elastikiyet katsayısının bir parçası olan \( \frac{dY}{dX} \), \( \hat{\beta}_i \) e eşit olacaktır. Tahmin edilmiş regresyondan ortalama ve nokta elastikiyetleri hesaplamak mümkündür. Ortalamalara göre elastikiyet; \( \varepsilon = \frac{dY}{dX} \frac{X}{Y} = \frac{dY}{Y} \frac{X}{Y} = \hat{\beta}_i \frac{X}{Y} \) değerine eşit olacaktır. Örnek regresyon modelinden ortalama göre elastikiyet hesaplanabileceği gibi, gözlemlemiş herhangi bir değere göre de nokta elastikiyeti hesaplanabilir.

Nokta elastikiyeti; \( \varepsilon = \frac{dY}{dX} \frac{X}{Y} = \hat{\beta}_i \frac{X}{Y} \) ’den elde edilir. Denklemdeki \( \hat{Y} \), bağımsız değişken \( X \) \( \hat{X}_i \) değeriini aldığında bağımlı değişken \( Y \) alacağı değerdir.

6.2. Hata Terimlerinin Normal Dağılımı

Klasik doğrusal regresyon modelinde ve en küçük kareler yönteminin uygulamasında hata terimlerinin dağılımına ilişkin herhangi bir varsayımda bulunulmamıştı. Hata terimi için yapılan varsayımlar; aşağıda gösterildiği üzere beklenen değerinin sıfır olduğu, birbirleri ile ilişkisiz ve sabit varyanslı olduğu idi.

\[ E(u_i) = 0 \]

\[ E(u_i, u_j) = 0 \quad i \neq j \]
Eğer amaç sadece en küçük karelerin nokta tahminleri ise klasik doğrusal regresyon modelinin bu temel varsayımları yeterli olacaktır. Ancak ekonometrik araştırmada amaç; nokta tahminlerinden hareketle ana kütle değerlerine ilişkin çıkarsamalar yapmaktır. Bu amaç için hipotez (önsav) testleri uygulanır ve nokta tahminlerinin güven aralıkları oluşturulur. \( u_i \) 'nin olasılık dağılımı hipotez testlerinde son derece önemlidir.

Dolayısıyla hata terimi (\( u_i \)) ile ilgili olasılık dağılımini bilmemiz gerekmektedir. Klasik normal doğrusal regresyon modeli, \( u_i \) 'nin aşağıdaki değerlerle normal dağıldığını varsayar.

\[
\begin{align*}
  u_i & \sim N(0, \sigma^2) \\
\end{align*}
\]

Normal dağılımanın ortalama ve varyans olmak üzere iki parametresi vardır. \( u_i \sim N(0, \sigma^2) \), “Hata terimi 0 ortalama ve sabit varyansla normal dağılmaktadır.” şeklinde okunmaktadır.


Hata terimlerinin normal dağıldığını varsayıp en küçük kareler yönteminin uygulanabildiğini için gerekli délgenlik diğer bir parametre tahmin yöntemi olan en çok benzerlik yöntemi için gerekli.

Hata terimi normal dağılıyorsa bunun doğrusal fonksiyonu olan en küçük kareler tahmincileri (\( \hat{\beta}_0 \) ve \( \hat{\beta}_1 \)) de normal dağılacaktır. Şöyle ki bilindiği üzere bağımlı değişkenin gözlemленen değeri;

\[
Y_i = \beta_0 + \beta_1 X_i + u_i\]

olarak ifade edilmiştir. Daha önce de üzerinde çok durulduğu üzere bağımlı değişken sistematik kısm (\( \beta_0 + \beta_1 X_i \)) ve tesadüfi kısmın (\( u_i \)) doğrusal fonksiyonudur. Burada hata teriminin merkezi limit teoremine göre normal dağıldığını gösterdik. “Normal dağılan bir değişkenin doğrusal fonksiyonu da normal dağılacaktır.” Özellikle göre bağımlı değişken \( Y \) de normal dağılacaktır. Böylece hata terimi normal dağıldığı için bağımlı değişkenin de normal dağılıma uyduğunu göstermiş oldu.

Bu asama en küçük kareler tahmincilerinin (\( \hat{\beta}_0 \) ve \( \hat{\beta}_1 \)) dağılımlarının normal dağılıma uygunluğu açıklanacaktır.
Bu amaç için yine daha önceki konulardan bildiğimiz yukarıdaki eşitlikten hareket edeceğiz. Yukarıdaki eşitliğe göre en küçük kareler tahmincisi $\hat{\beta}_i$, bağımlı değişken $Y_i$’nin doğrusal fonksiyonudur. Yine, normal dağılan bir değişkenin doğrusal fonksiyonu da normal dağılacaktır özelliğinden $\hat{\beta}_i$ normal dağılır. Aynı özellik $\hat{\beta}_i$ için de geçerlidir.

Hata terimlerinin normal dağıldıği varsayıımı en küçük kareler yönteminin uygulanabilmesi için gerekli değil iken diğer bir parametre tahmin yöntemi olan en çok benzerlik yöntemi için gereklidır.

Örnek

Aşağıda bir firmaya ait üretim ile çalışma saatleri verileri yer almaktadır.

<table>
<thead>
<tr>
<th>Gözlem</th>
<th>Çıktı(Q)</th>
<th>Çalışma Saati(L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>7</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>9</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

Tablo 1.1: Örnek

Çıktı (Q) ve çalışma saatine(L) ait verilerden hesaplanan ara sonuçlar aşağıdaki gibidir.
Örnek regresyon modelini tahmin ederek yorumlayınız.

Örnek regresyon modeli

En küçük kareler yöntemini uygularken normal denklemleri, diğer bir ifade ile değişkenlerin gözlemlenen değerlerini değil; gözlemlenen değerlerin ortalama farklarının kullanıldığı sapmalar yöntemini kullanacağız. Öncelikle \( \hat{\beta}_i \) parametresinin tahmin edilmesi gerekecektir.

\[
\hat{\beta}_i = \frac{\sum x_i y_i}{\sum x_i^2}
\]

\[
\sum x_i y_i = \sum X_i Y_i - n\bar{X}\bar{Y} = 789 - 10.8 \times 9.6 = 21
\]

\[
\Sigma x_i^2 = \Sigma X_i^2 - n\bar{X}^2 = 668 - 10.8^2 = 28
\]

\[
\hat{\beta}_i = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{21}{28} = 0.75
\]

\[
\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X} = 9.6 - (0.75) \times 8 = 3.6
\]

\[
\hat{Y}_i = 3.6 + 0.75X_i
\]

Çalışma saatı 1 birim artarsa üretim 0,75 birim artar. Bu yorum, marjinal verimliliğin tanıma denktir. Böylece \( \hat{\beta}_i \) parametresi, ichtisat derslerimizden bildiğimiz emeğin marjinal verimliliğine eşittir. \( \hat{\beta}_0 \), emek 0 olursa 3,6 birim üretim yapılacağını göstermektedir. Dolayısıyla emeğin marjinal verimlilik (produktivite) elastikiyetini hesaplayabiliriz.

2. Çalışma saatı 5 birim iken emeğin marjinal verimlilik elastikiyetini hesaplayarak yorumlayınız.

\[
\varepsilon_{QL} = \frac{dQ}{dL} \frac{L_i}{Q_i} = \hat{\beta}_i \frac{L_i}{Q_i}
\]

\[
\hat{Q}_{L=5} = 3.6 + 0.75(5) = 7.35
\]

\[
\varepsilon_{QL} = \hat{\beta}_i \frac{L=5}{\hat{Q}_{L=5}} = 0.75 \times \frac{5}{7.35} = 0.51
\]
3. Emeğin ortalama marjinal verimlilik elastikiyetini hesaplayarak yorumlayınız.

\[ \eta_{QL} = \hat{\beta} \frac{\bar{L}}{\bar{Q}} = 0.75 \times \frac{8}{9.6} = 0.625 \]

Emekteki %1'lik değişme üretimi ortalama %0,625 arttırır.

4. Tahminin standart hatasını hesaplayın.

\[ \sigma = \sqrt{\frac{\sum \hat{u}_i^2}{n-2}} \]

\[ \sum \hat{u}_i^2 = \sum y_i^2 - \sum \hat{y}_i^2 \]
\[ \sum \hat{u}_i^2 = (\sum Y^2 - n\bar{Y}^2) - \hat{\beta} \sum x_i \]
\[ \sum \hat{u}_i^2 = (952 - 10 \times (9.6))^2 - (0.75)^2 \times 28 = \]
\[ \sum \hat{u}_i^2 = 30.4 - 15.75 = 14.65 \]

\[ \sigma = \sqrt{\frac{\sum \hat{u}_i^2}{n-2}} = \sqrt{\frac{14.65}{10-2}} = \sqrt{1.83} = 1.35 \]

5. Belirginlik katsayısını hesaplayarak yorumlayınız.

\[ r^2 = 1 - \frac{\sum \hat{u}_i^2}{\sum y_i^2} \]

veya

\[ r^2 = \frac{\sum \hat{y}_i^2}{\sum y_i^2} \]

Formüllerinden olmak üzere iki yoldan hesaplanabilir. Tahminin standart hatası hesaplanırken belirginlik katsayısıyla ilgili unsurlar hesaplandığı için burada yeniden hesaplanmazตร์.

\[ r^2 = 1 - \frac{\sum \hat{u}_i^2}{\sum y_i^2} = 1 - \frac{\sum \hat{y}_i^2}{\sum Y^2 - n\bar{Y}^2} = 1 - \frac{14.65}{30.4} = 0.52 \]

\[ r^2 = \frac{\sum \hat{y}_i^2}{\sum y_i^2} = \frac{\hat{\beta} \sum x_i^2}{\sum Y^2 - n\bar{Y}^2} = \frac{15.75}{30.4} = 0.52 \]

Üretimdeki değişmenin %52’si emek değişkeniyle açıklanmaktadır. Belirsizlik
katsayısı 1-\(r^2\)’den 0,48 olarak hesaplanır. Üretimdeki değişmenin %48’i emek değişkeni dışındaki modelde yer almayan diğer değişkenlerle açıklanmaktadır.

Yine belirginlik katsayısının karekökü alınarak üretim ile emek değişkenleri arasındaki basit korelasyon katsayısı hesaplanabilir.

\[ r_{qt} = \sqrt{r^2} = \sqrt{0.52} = +0.72 \]

Üretim ile emek değişkenleri arasında pozitif yönde güçlü doğrusal ilişki vardır.

Korelasyon katsayı işaretini \(\hat{\beta}_1\)’den almıştır.
Uygulamalar
Uygulama Soruları
Bu Bölümde Ne Öğrendik Özeti
Bölüm Soruları

1. Hata terimi hangi dağılıma uygundur?

2. Hata teriminin normal dağılıma uygun olması neden önemlidir.

3. En küçük kareler tahmincilerinin dağılım özelliğini, hata terimi ile ilişkilendirerek açıklayınız.


5. Aşağıda X malının talebine (Q:000kg) ve fiyatına (P:10TL) ait 10 aylık gözlemlerden hesaplanan ara sonuçlar verilmiştir.

\[
\sum q^2 = 3960.4 \quad \sum p^2 = 106.5 \quad \sum pq = -318 \quad \bar{P} = 7.5 \quad \bar{Q} = 52.6
\]

\[
q = Q - \bar{Q} \quad p = P - \bar{P}
\]

Talebin fiyat elastikiyetini hesaplayarak yorumlayınız.

6. En küçük kareler yönteminin uygulanabilmesi için normallik varsayımı gerekli midir?


8. Ortalamalara göre elastikiyet ve nokta elastikiyetinin formüllерini belirtiniz.

9. Ekonometrik bir modelde hata teriminin önemini ve temsil ettiği kavramları açıklayınız.

7. TAHMİN EDİLEN PARAMETRELER İLE İLGİLİ HİPOTEZ TESTLERİ ve GÜVEN ARALIKLARI
Bu Bölümde Neler Öğreneceğiz?

7. TAHMİN EDİLEN PARAMETRELER İLE İLGİLİ HİPOTEZ TESTLERİ ve GÜVEN ARALIKLARI

7.1. Aralık Tahmini ve Hipotez Testleri

7.2. $\beta_0$ ve $\beta_1$ İçin Hipotez Testi

7.3. $\beta_0$ ve $\beta_1$ İçin Güven Aralığı

7.4. Ana Kütle Hata Terimi Varyansının Güven Aralığı

7.5. Ana Kütle Hata Teriminin Varyansı için Ki-kare $\chi^2$ Testi
Bölüm Hakkında İlgi Oluşturan Sorular
Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri
Anahtar Kavramlar
**Giriş**

Bu bölümde, nokta tahminlerinin istatistiksel güvenilirliklerinin test edilmesi ve buna bağlı olarak modelin geçerliliğinin sınanması amacıyla bu tahminlere uygulanan hipotez testleri ile anlamlılık sınavlarından elde edilecek sonuçlar doğrultusunda oluşturulan güven aralıkları üzerinde durulacaktır.

Bu bağlamda hipotez testlerinin uygulanma süreci, anlamlılık düzeyleri, parametreler için hesaplanan test istatistikleri, çift ve tek kuyruklu sınama, parametrelerin ve ana kütle hata terimi varyansının güven aralıklarının oluşturulması konuları hakkında detaylı bilgiler verilecektir.
7. TAHMİN EDİLEN PARAMETRELER İLE İLGİLİ HİPOTEZ TESTLERİ VE GÜVEN ARALIKLARI

7.1. Aralık Tahmini ve Hipotez Testleri

En küçük kareler tahmincilerinin sayısal değerleri (\(\hat{\beta}_0\) ve \(\hat{\beta}_1\)) ana kütle regresyon modelindeki \(\beta_0\) ve \(\beta_1\) parametrelerinin nokta tahminleridir ve istatistik bakımından güvenilirliklerinin test edilmesi gerekir. Böylece örnek regresyon modelinin tahminini izleyen aşama, nokta tahminlerinin istatistik bakımından güvenilirliklerinin sınınaması kısaca bu tahminlere hipotez testlerinin uygulanmasıdır. \(\hat{\beta}_0\) ve \(\hat{\beta}_1\) parametrelerine uygulanan hipotez testleri sonucunda modelin geçerliliği konusunda karar verilir. Modelin istatistiksel açıdan geçerli olduğu sonucuna varılırsa ilgili parametreinin güven aralıklarını oluşturulur ve değişkenlere ilişkin öngörülerde ilişkin öngörülerde bulunulur.

7.2. \(\beta_0\) ve \(\beta_1\) için Hipotez Testi

En küçük kareler tahmin edicileri, örnek verilerinden elde edilmişlerdir. Örnekleme hataları bütün örnekler için kaçınılmaz olduğuna göre, tahminlerin geçerliliği konusundaki güven aralıklarını oluşturabilmek için anlamlılık sınavlarının uygulanması gereklemektedir.

Hipotez testleri \(\hat{\beta}_0\) ve \(\hat{\beta}_1\) parametreleri ile ilgilidir. İleri sürülen ön sava istatistik literatüründe sıfır hipotezi veya temel hipotez denir ve \(H_0\) ile gösterilir. Sıfır hipotezi \(H_1\) ile gösterilen karşı hipoteze karşı sınınanır. Karşı hipotez, genellikle alternatif hipotez olarak adlandırılır.

Hipotez testlerinde iki önemli nokta vardır. Bunlardan ilk; temel hipotezin ne şekilde kurulacağı ile ilgilidir. Bu konuya açıklık getirmek için aşağıdaki iki temel hipoteze yer verilmiştir:

\[H_0 : \beta_i = 0 \quad \text{ve} \quad H_0 : \beta_i = \beta_i^*\]

Temel hipotezlerden ilki, \(\hat{\beta}_i\) ’nin sıfırdan anlamlı bir farkın olup olmadığını, \(\hat{\beta}_i\) ’nin tahmin edildiği örneğin gerçek parametresi sıfır \((\beta_i = 0)\) olan bir ana kütleden geldiği savını ileri sürmektedir. Böylece hipotezin kabulü, bağımsız değişkenle bağımsız değişken arasında doğrusal ilişkinin olmadığını işaretidir. İkinci temel hipotez ise diğerinden farklı bir ilişkinin varlığı kabul edilmekte, \(\beta_i\) paramtresinin \(\beta_i^*\) gibi önsel bir değere eşitliği ileri sürülmektedir.

Diğer önemli bir nokta ise hipotez testinin tek taraflı (tek kuyruk) veya çif taraflı (çift kuyruk) kurulacağı ile ilgilidir. İktisadi açıdan bir zorunluluk olmadığı sürece hipotez testleri genellikle çif taraflı kurulur. Ancak hipotez testine konu olan parametre, marjinal tüketim
meyli ve marjinal tasarruf meyli gibiacağı işaret önceden belli olan bir iktisadi terime karşılık geliyorsa tek taraflı test uygulanmalıdır.

Tahmin edilen parametrelerin istatistik açıdan anlamlılık sınaması, normal dağılıma dayanan $Z$ sınamasıdır. Bu sınamada, ancak aşağıdaki koşullar geçerli ise kullanılabilir:

- Ana kütle varyansı biliniyorsa,

- Ana kütle varyansı bilinmediği halde örneklem yeterince büyükse ($n>30$). Çünkü ana kütle hata teriminin varyansının tahmini ($\hat{\sigma}_{u}^2$), ana kültün bilinmeyen varyansının ($\sigma_{u}^2$) büyük örnek için yeterince yakın tahminidir.

Aksi hâlde, Student $t$ sınaması kullanılır.

Ana kütle hata terimi varyansı ($\sigma_{u}^2$) biliniyorsa, normal dağılımın özellikleri ve $\sigma_{u}^2$ biliniyorsa, ortalama $\mu$, varyansı $\sigma^2$ olan normal dağılım normal eğrinin altında alanın, $\mu \pm \sigma$ aralığında yaklaşık $68\%$, $\mu \pm 2\sigma$ aralığında yaklaşık $95\%$, $\mu \pm 3\sigma$ aralığında yaklaşık $99.7\%$ yer alır.

Hata terimine ilişkin varsayımlar ($u_i \sim (0, \sigma^2)$) geçerli ise daha açık bir ifade ile hata teriminin normal dağılıdığı varsayımlı altında $\hat{\beta}_0$ ve $\hat{\beta}_1$ parametrelerinin örnekleme dağılımları aşağıda verilen ortalama ve varyans ile normal dağılıma uygundur.

$$E(\hat{\beta}_0) = \beta_0$$
$$E(\hat{\beta}_1) = \beta_1$$

$$Var(\hat{\beta}_0) = \sigma_{\hat{\beta}_0}^2 = \frac{\sum X_i^2}{n} \sigma_{u}^2$$
$$Var(\hat{\beta}_1) = \sigma_{\hat{\beta}_1}^2 = \frac{\sigma_{u}^2}{\sum x_i^2}$$

$$\hat{\beta}_0 \sim N\left(\beta_0, \sigma_{\hat{\beta}_0}^2\right) ; \hat{\beta}_1 \sim N\left(\beta_1, \sigma_{\hat{\beta}_1}^2\right)$$

$\hat{\beta}_0$ ve $\hat{\beta}_1$ parametrelerine ilişkin normal dağılımlar; ortalama 0 varyansı 1 olan standard normal değişken $Z$’nin birimlerine dönüştürülenerek standartlaştırılabilirler.

En küçük tahminleri olan $\hat{\beta}_0$ ve $\hat{\beta}_1$ için test istatistikleri aşağıdaki gibidir:

$$Z = \frac{\hat{\beta}_0 - \beta_0^*}{SE(\hat{\beta}_0)} = \frac{\hat{\beta}_0 - \beta_0^*}{\sqrt{\sigma_{u}^2 \sum X_i^2 / n \sum x_i^2}} \sim N(0,1)$$

$$Z = \frac{\hat{\beta}_1 - \beta_1^*}{SE(\hat{\beta}_1)} = \frac{\hat{\beta}_1 - \beta_1^*}{\sqrt{\sigma_{u}^2 / \sum x_i^2}} \sim N(0,1)$$

Böylelikle ana kütle parametresinin gerçek değeri ile ilgili hipotez sınamaları 148
yapılabilir.

\[ H_0 : \beta_i = 0 \]  \[ H_0 : \beta_i = \beta_i^* \quad i = 1,2 \]

\[ H_1 : \beta_i \neq 0 \]  \[ H_1 : \beta_i \neq \beta_i^* \]

Hipotezin sınanması için kullanlacak bir anlamlılık düzeyi (I.tür hata yapma olasılığı) seçilir. Geleneksel anlamlılık düzeyleri %1, %5 ve %10’dur. Örneğin; anlamlılık düzeyi %5 alınrsa kararımızı verirken 100 kerede 5 kere yanlış yapmayı, aslında doğru olan hipotezi reddetmeyi göze almış oluruz.

\[ H_0 : \beta_i = 0 \] Hipotezinin testinde, test istatistiği sadeleştirilerek

\[ Z = \frac{\hat{\beta}_i}{SE(\hat{\beta})} \] olarak yazılır.

**Şekil 7.1:** Normal Dağılım Kabul ve Red Bölgeleri

Ekonometrik uygulamalarda zorunluluk olmadığı sürece çift kuyruklu (çift taraflı) sınama yapılmaktadır. Çift kuyruk sınamasında, *standart normal dağılım* her iki kuyruğu seçilmiş, anlamlılık düzeyinin yarısı kadar olan bölümü eşik bölge olarak seçilmiştir. Örneğin %5 anlamlılık düzeyi seçilmişse her bir kuyruk 0,05/2’den 0,025’lik alanı (olasılığı) içerir. Eğrinin her iki ucunda 0,025 olasılığına karşılık olan Z’nin eşik değerleri “Z₁= -1,96 ve Z₁=1,96” standart normal dağılım çizelgesinden bulunur. Son aşamada test istatistiğinden hesaplanan değeri, Z’nin eşik değerleri ile karşılaştırılır.

Hesaplanan Z değeri eşik bölgesi içine düşüyorsa (\(Z_{hesap}>1,96\) ya da\(Z_{hesap}<-1,96\) ise), \(H_0\) hipotezi reddedilir. Diğer yandan hesaplanan Z değeri tablo değerlerinin (eşik değerlerin) arasına düşüyorsa (-1.96<\(Z_{hesap}<1.96\) temel hipotez reddedilemez.

%5 anlamlılık düzeyinde iki kuyruklu sınımlar için kıscaca,

- \(Z_{hesap}>2\) ise temel hipotez reddedilir.
Veya

\[ -SE(\hat{\beta}_i) < \hat{\beta}_i / 2 \] ise temel hipotez reddedilir.

Ancak ana kütte hata terimi varyansı \((\sigma_u^2)\) genellikle bilinmez, örnek verilerinden tahmin edilir. Eğer \(\sigma_u^2\) yerine \(\hat{\sigma}_u^2\) geçerse ve gözlem sayısı da 30’dan küçük ise \(Z\) istatistiğinin yerine \(Student-t\) istatistiği kullanılır. \(t\) istatistiği, \(Z\) istatistiğine benzer. Ancak \(t\) istatistiği serbestlik derecesine bağlıdır ve test istatistiğinde gerçek varyans yerine gerçek varyansın eğilimsiz tahmini yer alır.

\(Student-t\) dağılımı simetrik bir dağılımdır. Ortalaması 0 varyansı \((n-1)/(n-3)\) olup n büyüdükçe varyans 1’e yaklaşır. Böylece n artıkça \(Student-t\) dağılımı standart normal dağılıma yaklaşır: \(Z \sim N(0,1)\)

\(t\) testinin uygulanabilmesi için;

- Temel ve alternatif hipotezler tanımlanmalı,

\[
\begin{align*}
H_0 : \beta_i &= 0 \\
H_i : \beta_i &= \beta_i' \\
H_1 : \beta_i &\neq \beta_i' \\
\end{align*}
\]

- İstenen bir anlamılık düzeyi seçilmeli,

\(\alpha = 0.10, \alpha = 0.05, \alpha = 0.01\)

- Serbestlik derecesi belirlenmelidir.

Bu bilgilerle eşik bölgesini, bütün \(t\) değerleri kümesini kabul ve red bölgeleri olarak ikiye ayırın \(t\) eşik değerleri tespit edilir.

Tek bir örnekten hesaplanan \(\hat{\beta}_i\) için n-2 serbestlik derecesiyle test istatistiği şuına eşittir:

\[
t = \frac{\hat{\beta}_i - \beta_i}{SE(\hat{\beta}_i)} = \text{tahmin edici - Anakütle parametresi} / \text{tahmin edicinin standart hatası}
\]

Basit regresyonda iki parametre olduğu için serbestlik derecesi n-2’dir. Genel olarak serbestlik derecesi \(n-k\’dir.\) Genel olarak serbestlik derecesi \(n-k\’dir.\) modelde yer alan parametre sayısıdır. Yine temel hipotez \(H_0 : \beta_i = 0\) şeklinde ise test istatistiği;

\[
t^* = \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)}
\]

150
olur. Hesaplanan $t^*$ iki kuyruklu bir sınamada n-2(n-k) serbestlik derecesiyle eşik bölgesini tanımlayan tablo değerile karşılaştırılır. Eğer $t^*$ eşik bölgesinde yer alırsa sıfır hipotezi reddedilir. $\hat{\beta}_i$ istatistik bakımından anlamlı bir tahmindir. Eğer $t^*$, eşik değerler arasında yer alırsa $H_0$ sıfır hipotezi kabul edilir. $\hat{\beta}_i$ tahmininin verilen anlamlılık düzeyinde istatistik bakımından anlamlı olmadığı sonucuna varılır.

**Şekil 7.2:** t Dağılımı Kabul ve Red Bölgeleri

**Tablo 7.1:** t Anlamlılık Sınaması: Karar Kuralları

<table>
<thead>
<tr>
<th>Önsav türü</th>
<th>$H_0$: Sıfır Önsav</th>
<th>$H_1$: Karşı Önsav</th>
<th>Karal kural: $H_0$ Reddedilir</th>
</tr>
</thead>
<tbody>
<tr>
<td>Çift Kuyruk</td>
<td>$\beta_2 = \beta_2^*$</td>
<td>$\beta_2 \neq \beta_2^*$</td>
<td>$</td>
</tr>
<tr>
<td>Sağ Kuyruk</td>
<td>$\beta_2 \leq \beta_2^*$</td>
<td>$\beta_2 &gt; \beta_2^*$</td>
<td>$t &gt; t_{a, sd}$</td>
</tr>
<tr>
<td>Sol Kuyruk</td>
<td>$\beta_2 \geq \beta_2^*$</td>
<td>$\beta_2 &lt; \beta_2^*$</td>
<td>$t &lt; -t_{a, sd}$</td>
</tr>
</tbody>
</table>

Ayrıca: $\hat{\beta}^*$, $\beta_0$ nin varsayılan sayısal değerdir. $t$, $\alpha$ nin mutlak değeri. $t_{a/2}$ ya da $t_{a, sd}$ anlamlılık düzeyindeki eşik $t$ değerleridir. $t_{a}$ ya da $t_{a, sd}$ anlamlılık düzeyindeki eşik $t$ değerleridir. $t_{a/2}$, serbestlik derecesi, iki değişkenli modelde (n-2), üç değişkenli modelde (n-3), vb. Aynı sürece, $\hat{\beta}^*$ için önsav sınamasında da geçerlidir.

**Tablo 7.1:** t Anlamlılık Sınaması: Karar Kuralları

Temel hipotezin %5 anlamlılık düzeyinde iki kuyruklu bir sınaması için serbestlik derecesi (n-k) 8’den büyükse 2t kuralı olarak da bilinen kural uygulanabilir. Hesaplanan $t^*$ değeri 2’den büyükse temel hipotez reddedilir. Bu durumda tahmin edilen parametre, kendi standart hatasının iki katından fazladır.
7.3. $\beta_0$ ve $\beta_1$ İcin Güven Aralığı

Sıfır hipotezinin reddi, örnek küteden elde edilen $\hat{\beta}_0$ ve $\hat{\beta}_1$ tahminlerinin ana kütlenin gerçek parametreleri olan $\beta_0$ ve $\beta_1$'nin tek doğru tahmini olduğunu göstermez. Sıfır hipotezinin reddi ile ancak $\beta_0$ ve $\beta_1$ parametrelerinin sıfırdan farklı bir ana küteden çekilmiş bir örnekten geldiği sonucuna varılmaktadır.

Ana kütle parametrelerinin örnek verilerinden sağlanan tahmine ne kadar yakın olduğunu bilinmek ister. Bunun için ana kütlele ilişkin parametrelerin güven aralıkları belirlenmelidir. Ana kütle parametresi belli bir güven düzeyinde (1- $\alpha$), tanımlanan güven aralığından yer alacaktır.

Daha öncde belirtildiği üzere normal dağılım; ana kütle hata teriminin varyansı biliniyorsa ya da gözlem sayısı yeterince büyükse kullanılabılır. Öncelikle güven düzeyinin seçilen güven düzeyinde $\beta_1$ için güven aralığı;

$$
\text{Prob}\left( -Z_{a/2} \leq Z \leq Z_{a/2} \right) = 1 - \alpha
$$

gibidir.

$$
Z = \frac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_1)}
$$

ifadesi eşitsizlikte yerine konur ve yeniden düzenleme yapılırsa aşağıdaki sonucu ulaştırılır:

$$
\text{Prob}\left( -Z_{a/2} \leq \frac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_1)} \leq Z_{a/2} \right) = 1 - \alpha
$$

$$
\text{Prob}\left( \hat{\beta}_1 - Z_{a/2}SE(\hat{\beta}_1) \leq \beta_1 \leq \hat{\beta}_1 + Z_{a/2}SE(\hat{\beta}_1) \right) = 1 - \alpha
$$

$\beta_1$ eşitiği $\beta_1$ için yüzde 100(1- $\alpha$ ) güven aralığını gösterir.

Daha kısa, $\hat{\beta}_1 \pm Z_{a/2}SE(\hat{\beta}_1)$ şeklinde ifade edilebilir.

Buna göre $\beta_1$ için güven aralığını oluşturmak için $t$ dağılımı kullanılır.

$$
\text{Prob}\left( -t_{a/2} \leq t \leq t_{a/2} \right) = 1 - \alpha
$$

esitsizliği genişletilerek;

$$
\text{Prob}\left( -t_{a/2} \leq \frac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_1)} \leq t_{a/2} \right) = 1 - \alpha
$$

 şeklinde yazılabılır.

Eşitsizlikteki $t_{a/2}$; $\alpha/2$ anlamlılık düzeyinde ve n-2 serbestlik derecesinde t dağılığın bulunan t değeri olup $\alpha/2$ anlamlılık düzeyindeki t eşik değeridir.
Prob\left(\hat{\beta}_i - t_{a/2}SE(\hat{\beta}_i) \leq \beta_i \leq \hat{\beta}_i + t_{a/2}SE(\hat{\beta}_i)\right) = 1 - \alpha

\hat{\beta}_i \pm t_{a/2}SE(\hat{\beta}_i)

Yukarıdaki ifadeler \(\beta_i\) için yüzde 100\((1 - \alpha)\) güven aralığını gösterir.

Herhangi bir parametrenin güven aralığının genişliği, ilgili parametrenin standart hatası ile ilişkilidir. Tahmin edilen parametrenin standart hatası ne kadar büyükse güven aralığı da o nispete genişleyecektir. Güven aralığı sabit olup rastlantısal değildir. Dolayısıyla güven aralığının yorumu dikkatli yapılmalıdır. Yukarıda güven katsayısı \(1 - \alpha\) olduguna göre;

\((\hat{\beta}_i - t_{a/2}SE(\hat{\beta}_i), \hat{\beta}_i + t_{a/2}SE(\hat{\beta}_i))\) aralığı her 100 tanesinde 95’i ana kütleye ait gerçek \(\beta_i\) değerini içerir.

**7.4. Ana Kütle Hata Terimi Varyansının Güven Aralığı**

Ana kütle hata teriminin \((\sigma^2_u)\) normal dağılımı varsayımı altında, ana kütle hata teriminin varyansının tahmininin \((\hat{\sigma}^2_u)\) örnekleme dağılımı \(n-2\) serbestlik derecesiyle \(\chi^2\) (kı-kare) dağılımına uymaktadır.

Test istatistiği aşağıdaki verilmiştir:

\[\chi^2 = (n-2)\frac{\hat{\sigma}^2}{\sigma^2}\]

\[\hat{\sigma}^2 = \frac{\sum u_i^2}{n-2}\]

\[\sum u_i^2 = (n-2)\hat{\sigma}^2\]

\(\chi^2\) dağılımını kullanarak \(\sigma^2\) için güven aralığı oluşturulur. \(\sigma_u^2\) için güven aralığı;

\[\text{Prob}\left(\chi_{1-a/2}^2 \leq \chi^2 \leq \chi_{a/2}^2\right) = 1 - \alpha\]

şeklindedir. Buradaki iki eşitsizlik arasındaki \(\chi^2\) yukarıdaki test istatistiğidir. \(\chi_{1-a/2}^2\) ile \(\chi_{a/2}^2\), \(n-2\) serbestlik derecesiyle ki-kare tablosundan bulunan \(\chi^2\) eşik değerleridir.
Şekil 7.3: $\chi^2$ İçin $\%(1-\alpha)$ Güven Aralığı

Ana kütle hata terimi için oluşturulan güven aralığı simetrik değil, asimetriktr. Asimetrinin nedeni düşük serbestlik dereceli $\chi^2$ dağılımının çarpık olmasıdır. Serbestlik derecesi arttıkça merkezi limit teoremi gereğince $\chi^2$ dağılımı normal dağılıma yaklaşır. Test istatistiği, yukarıda $\sigma_u^2$ için güven aralığında yerine yerleştirilirse,

$$\text{Prob} \left( \chi^2_{1-a/2} \leq \frac{(n-2)\hat{\sigma}^2}{\sigma^2} \leq \chi^2_{a/2} \right) = 1 - \alpha$$

$$\text{Prob} \left( \frac{(n-2)\hat{\sigma}^2}{\chi^2_{a/2}} \leq \sigma^2 \leq \frac{(n-2)\hat{\sigma}^2}{\chi^2_{1-a/2}} \right) = 1 - \alpha$$

elde edilir ve bu eşitsizlik $\sigma_u^2$ için $100(1-\alpha)$ için güven aralığını verir. $\sigma_u^2$ için güven aralığını kalıntı kareler toplamını kullanarak da yazabiliriz.

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-2} \quad \text{den} \quad \sum \hat{u}_i^2 = (n-2)\hat{\sigma}^2 \quad \text{ye eşittir.}$$

Böylece,

$$\text{Prob} \left( \frac{\sum \hat{u}_i^2}{\chi^2_{a/2}} \leq \sigma^2 \leq \frac{\sum \hat{u}_i^2}{\chi^2_{1-a/2}} \right) = 1 - \alpha \text{ olur.}$$

Ana kütle parametreleri için oluşturulan güven aralıkları ile karşılaştırıldığında, hata teriminin varyansının güven aralıklarını varyansa ait olmasına dolaylı daha genişştir.

Varyans daima sıfırdan farklı ve pozitif bir değer olduğundan çift taraflı güven aralığı oluşturulması yerine sadece üst sınırın bulunması yeterlidir. Sadece üstten sınırın bulunması durumunda $\sigma_u^2$ için güven aralığı;
\[
\text{Prob} \left( 0 \leq \sigma^2 \leq (n-2) \frac{\hat{\sigma}^2}{\chi^2_n} \right) = 1 - \alpha
\]

veya
\[
\text{Prob} \left( 0 \leq \sigma^2 \leq \frac{\sum \tilde{u}_i^2}{\chi^2_n} \right) = 1 - \alpha \quad \text{şeklindedir.}
\]

**Açıklavıcı Örnek**

Aşağıda ki-kare eşik (tablo) değerlerinin bulunması için bir örnek verilmiştir.

![f(\chi^2)](image)

**Şekil 7.4:** 10 Serbestlik Dereceli $\chi^2$ İçin %90 Güven Aralığı

**7.5. Ana Kütle Hata Teriminin Varyansı İçin Ki-kare $\chi^2$ Testi**

Ana kütle hata terimi varyansı ($\sigma_u^2$) için $H_0 : \sigma^2 = 0$ gibi ana kütle hata terimi varyansının sıfıra eşitliğini ifade eden bir hipotez gereksizdir. Böyle bir hipotezin gereği, değişkenler arasındaki ilişki deterministiktir. Oysaki ekonometride değişkenler arasındaki ilişkiler kesin olmayan, rastlantısal ilişkilerdir.

\[
H_0 : \sigma^2 = \sigma_0^2
\]

Yukarıda verildiği üzere ana kütle hata teriminin belli bir değere eşitliğini, büyük veya küçük olduğunu test edebiliriz.
Tablo 7.2.: Hata Terimi Varyansının Testi İçin Hipotezler ve Karar Kuralları

Örnek

Öncesi bölümde başladığımız üretim ile emek arasındaki ilişkini araştıran uygulamanın bu konu ile ilgili olan devamı aşağıda verilmiştir.

1)Tahmin edilen parametrelerin anlamlılıklarını test ediniz.

Öncelikle \( \hat{\beta}_0 \) için anlamlılık sınavı yapılacaktır.

\[
H_0 : \beta_0 = 0 \\
H_1 : \beta_0 \neq 0 \\text{ için test istatistiği:}
\]

\[
t_h = \frac{\hat{\beta}_0 - \beta_0}{SE(\hat{\beta}_0)} = \frac{\hat{\beta}_0}{SE(\hat{\beta}_0)}
\]

Ancak parametrelerin standart hatasının bilinmemektedir. Standart hata, varyansının karekökü olduğuna göre öncelikle parametrenin varyansının hesaplanması gereklidir. Varyans şu şekilde hesaplanır:

\[
Var(\hat{\beta}_0) = \frac{\sum X_i^2}{n} \sigma^2 = \frac{668}{10.28} = 65.01 \\
\text{tir.}
\]
Standart hata ise

\[ SE(\hat{\beta}_0) = \sqrt{Var(\hat{\beta}_0)} = 2.09 \] 'a eşittir.

Test istatistiği;

\[ t_h = \frac{\hat{\beta}_0}{SE(\hat{\beta}_0)} = \frac{3.6}{2.09} = 1.72 \] 'ye eşittir.

Karşılaştırma yapabilmek için tablo değerine ihtiyaç vardır. Gözlem sayısı 30'dan küçük olduğu için Student- t dağılım tablo değerleri kullanılacaktır.

\[ t_i = t_{(\nu - 2)/2} = t_{(10 - 2)/0.05/2} = t_{8,0.025} = 2.306 \]

Karar aşaması:

\[ t_h = 1.72 < t_i = 2.306 \] olduğu için temel hipotez kabul edilir. Sabit parametre istatistiksel açıdan anlamsızdır. Bu modeli orijinden geçen bir model şeklinde yorumlamak da mümkündür.

Şimdi \( \hat{\beta}_1 \) için anlamlılık sınavı yapılacaktır.

\[ H_0 : \beta_1 = 0 \]
\[ H_1 : \beta_1 \neq 0 \] için test istatistiği:

\[ t_h = \frac{\hat{\beta}_1}{SE(\hat{\beta}_1)} \]

Standart hatanın hesaplanabilmesi için yine ilgili parametrenin varyansı hesaplanır.

\[ Var(\hat{\beta}_1) = \frac{\hat{\sigma}_u^2}{\sum x_i^2} = \frac{1.83}{28} = 0.0654 \]

Standart hata \( SE(\hat{\beta}_1) = \sqrt{Var(\hat{\beta}_1)} = 0.256 \) olarak hesaplanır.

Test istatistiği \( t_h = \frac{\hat{\beta}_1}{SE(\hat{\beta}_1)} = \frac{0.75}{0.256} = 2.93 \) ‘tür.

Karar aşaması:

\[ t_h = 2.93 > t_i = 2.306 \] ‘dır.
Böylece temel hipotez, alternatif hipotez lehine reddedilir. Emek, üretimi etkileyen istatistiksel açıdan anlamlı bir değişkendir.

\( \hat{\beta}_0 \) ve \( \hat{\beta}_1 \) arasındaki ortak varyansı da hesaplayabiliriz. Ortak varyans;

\[
Kov(\hat{\beta}_0, \hat{\beta}_1) = -\bar{X}Var(\hat{\beta}_1)
\]

\[
= -\bar{X} \left( \frac{\sigma^2}{\sum x_i^2} \right) \text{ dan}
\]

\[
Kov(\hat{\beta}_0, \hat{\beta}_1) = -\bar{X}Var(\hat{\beta}_1) = -(8)(0.0654) = -0.5232 \text{ ye eşittir.}
\]

2) Emeğin marjinal verimliliğinin 1'e eşit olduğu tezini test ediniz.

\[ H_0 : \beta_1 = 1 \]
\[ H_1 : \beta_1 \neq 1 \]

Test istatistiği;

\[
t_h = \frac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_1)} = \frac{0.75 - 1}{0.256} = 0.98 > -2.306 \text{dır. Bu sonuca göre temel hipotez reddedilemez.}
\]
Uygulamalar
Uygulama Soruları
Bu Bölümde Ne Öğrendik Özeti
Bölüm Soruları

1. İÜ Uzaktan Eğitim Merkezi İktisat Bölümü öğrencilerinden 20 kişinin verilerinden aşağıdaki regresyon modeli tahmin edilmiştir. C=Tüketim Y=Gelir

\[ C_i = 15.67 + 0.69Y_i, \quad r^2 = 0.72, \quad \sum \hat{u}_i^2 = 2.84 \]

\[ SE(\beta_i) = (2.81), \quad (0.07) \]

a) Tahmin edilen parametrelerin anlamlılıklarını test ediniz. (\( \alpha = 0.05 \))

b) Tahmin edilen parametrelerin güven aralıklarını oluşturunuz.

c) Marjinal tüketim meylinin 0,47 olduğu hipotezini test ediniz.

d) Ana kütle hata teriminin varyansını hesaplayarak güven aralığını oluşturunuz.


3. Tahmin edilen parametrenin anlamlılık sınamasının normal dağılıma uyan Z sınaması ile yapılabilmesinin şartlarını yazınız.

4. 100 gözlemli bir örnek regresyonunda (\( \hat{Y} = \hat{\beta}_0 + \hat{\beta}_iX_i \)) bağımsız değişkenin değişkenliği 3,6; regresyon ile açıklanabilen değişme 56,75 ve tahminin standart hata 1,82 olarak hesaplanmıştır. \( \beta_i \) parametresinin istatistiksel anlamlılığını test ediniz. (\( \alpha = 0.05 \))

5.-7. soruları aşağıda yer alan tabloyu dikkate alarak cevaplandırınız.

<table>
<thead>
<tr>
<th>( Y_i )</th>
<th>( X_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
</tr>
</tbody>
</table>

5. Davranışsal parametrenin (\( \beta_i \)) istatistiksel anlamlılığını test ediniz. (\( \alpha = 0.05 \))

6. \( \beta_i \) parametresinin güven aralıklarını oluşturunuz.
7. Ana kütle varyansının güven aralıklarını oluşturunuz.

Aşağıda X ve Y değişkenlerine ait veriler verilmiştir. 8.-10. soruları bu verileri dikkate alarak cevaplandırınız.

<table>
<thead>
<tr>
<th>Y</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>11</td>
<td>6</td>
</tr>
</tbody>
</table>

8. Tahmin edilen parametrelerin istatistiksel anlamlılığını test ediniz

9. Tahmin edilen parametrelerin güven aralıklarını oluşturunuz.

10. Ana kütle hata teriminin varyansının güven aralığını oluşturunuz.
8. EN KÜÇÜK KARELER TAHMİN EDİCİLERİN ÖZELLİKLERİ
Bu Bölümde Neler Öğreneceğiz?

8. EN KÜÇÜK KARELER TAHMİN EDİCİLERİN ÖZELLİKLERİ

8.1. Küçük Örnek Özellikleri

8.1.1. Doğrusallık
8.1.2. Sapmasız (Eğilimsiz, Sistematik Hatasız)
8.1.3. En Küçük Varyans
8.1.4. Etkinlik
8.1.5. Doğrusal En İyi Sapmasızlık
8.1.6. En Küçük Ortalama Kareli Hata
8.1.7. Yeterlilik

8.2. Büyük Örnek Özellikleri

8.2.1. Asimptotik Sapmasızlık
8.2.2. Tutarlılık
8.2.3. Asimptotik Etkinlik
Bölüm Hakkında İlgi Oluşturan Sorular
Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri
Anahtar Kavramlar
**Giriş**

Bu bölümde, parametre tahmininde tahmin edicilerin taşması gereken özelliklerden bahsedilerek en küçük kareler (EKK) tahmincilerinin küçük ve büyük örneklem özellikleri üzerinde durulacak ve farklı tahminciler arasında seçim yaparken dikkat edilmesi gereken noktalar belirilecektir.
8. EN KÜÇÜK KARELER TAHMİN EDİCİLERİN ÖzELLİKLERİ

Parametreler tahmin edilirken tahmin edicilerin bazı özellikleri taşması gerekmektedir. Klasik doğrusal regresyon modelinin varsayımlarının geçerli olması hâlinde, en küçük kareler (EKK) tahmincileri en uygun özellikleri taşımaktadır.

8.1. Küçük Örnek Özellikleri

8.1.1. Doğruluk

Regresyon modelindeki bağımlı değişken Y gibi rassal bir değişkenin doğrusal bir fonksiyonudur. Tahmin edilen parametre örnek gözlemlerin doğrusal bir fonksiyonu ise, diğer bir ifadeyle örnek verilerinin doğrusal bir bileşimiyle elde ediliyorsa doğrusal olduğu varsayılır.

\[ \hat{\beta}_1 = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\sum X_i y_i}{\sum x_i^2} = \frac{\sum x_i Y_i}{\sum x_i^2} \]

\[ \frac{\sum x_i}{\sum x_i^2} \text{ sabit bir değer olan } k \text{ 'ye, } \frac{\sum x_i}{\sum x_i^2} = k \text{ şeklinde eşitlendiği varsayılırsa; } \]
\[ \hat{\beta}_1 = k \sum Y_i \text{ olduğu sonucu elde edilir. } \hat{\beta}_1, Y_i \text{ 'nin doğrusal bir fonksiyonu olduğu için } \hat{\beta}_2 \text{ doğrusal bir tahmin edicidir. } k \text{ 'ler tartı olarak alındığı için } \hat{\beta}_1 \text{ 'nin } Y_i \text{ 'nin tartılı ortalaması olduğu ifade edilebilir.} \]

\( k_i \) Tartışının Özellikleri

1. \( X_i \) 'lerin olasılık dağılımı olmadığı için \( k_i \) 'lerin de olasılık dağılımı yoktur.

2. \[ \sum k_i^2 = 0 \rightarrow \sum k_i = \sum \left( \frac{x_i}{\sum x_i^2} \right) = \frac{1}{\sum x_i^2} \sum x_i = 0 \]

3. \[ \sum k_i^2 = \frac{1}{\sum x_i^2} \]

4. \[ \sum k_i X_i = \sum k_i x_i \rightarrow \sum \left( \frac{x_i}{\sum x_i^2} \right) x_i = \frac{\sum x_i^2}{\sum x_i^2} = 1 \]

8.1.2. Sapmasız (Eğilimsiz, Sistematik Hatasız)

Bir tahmin edicinin sapması; beklenen değerle ana kütü gerçek parametresi arasındaki fark anlamına gelmektedir. Beklenen değerin \( E \) ile gösterilmesi hâlinde sapmayı aşağıdaki gibi göstermek mümkündür:
Sapma = \( E(\hat{\beta}) - \beta \)

Aşağıdaki şekilde sapmalı bir tahminci görülmektedir. \( \beta \)’nin gerçek değeri ile \( \hat{\beta} \) tahmincisinin beklenen değeri arasındaki fark sapmayı göstermektedir.

Tahmincinin beklenen değerinin ana kütle parametresine eşit olması hâlinde, o tahmin edicinin sapmasız olduğu ifade edilir. Sapmasızlık, \( E(\hat{\beta}) = \beta \) anlamına gelmemekte, çok sayıda örnek alındığında doğru tahminin ortalama olarak elde edileceği anlamına gelmektedir. Diğer bir ifadeyle ana kütlede tekrar tekrar örnek çekilip bir örneklerle tahmincilerle ana kütle parametreleri tahmin edildiğinde elde edilen tahminciler ortalama olarak doğru olacaktır. Elbette bazı tahmincilerin çok yüksek, bazı tahmincilerin ise çok düşük olması olasıdır; fakat ortalamada parametrenin gerçek değerini verir. Sapmasızlık, gerekli fakat yeterli olmayan bir özelliktir. Sadece ortalama dayandığından ancak küçük bir varyansla birlikte önemli hâle gelir.

8.1.3. En Küçük Varyans

Bir tahminci başka bir tahminciye göre daha küçük varyansa sahipse o tahmincinin daha iyi bir tahminci olduğu ifade edilebilir. Diğer bir ifadeyle;

\[ \text{var}(\hat{\beta}) < \text{var}(\beta^*) \] ise \( \hat{\beta} \)’nın daha iyi bir tahminci olduğu söylenebilir. Bu durum aşağıdaki şekilde de görülmektedir. \( \beta^* \)’nin varyansı \( \hat{\beta} \)’nin varyansına göre daha küçük olduğundan bu tahmincilerden \( \hat{\beta} \)’nın daha iyi bir tahminci olduğu ifade edilebilir.

Düşük varyans; tek başına anlamlı olmamakta, ancak küçük sapmayla birlikte aranan bir özellik hâle gelmektedir. Varyansının küçük olması hâlinde, sapması büyük olan bir
tahminci ana kütle gerçek parametre değerine uzak bir değer etrafında değiştiği ifade edilebilir.

### 8.1.4. Etkinlik

Bir ana kütle parametresinin tahmini için birden fazla sapmasız tahminci olabilir. Bununla birlikte birden fazla küçük varyansa sahip tahminci de olabilir. Bu iki özellik bakımından bir tahmin edici dört farklı özelliğe sahip olabilir:

- \( \hat{\beta} \) sapmasız ve en küçük varyanslı tahmin edici olabilir.
- \( \hat{\beta} \) sapmasız fakat en küçük varyanslı tahmin edici değildir.
- \( \hat{\beta} \) sapmalıdır fakat ve en küçük varyanslı tahmin edicidir.
- \( \hat{\beta} \) sapmalı olmakla birlikte en küçük varyanslı değildir.

Bu dört durumdan en arzu edileni ilk, en istenmeyen durum ise dördüncü durumdur. Dolayısıyla sapmasız tahmincilerden en küçük varyansa sahip olan tahmincinin ana kütle parametresinin gerçek değeri etrafında yoğun olarak dağıldığı söylenir ve bu tahmincinin etkin olduğu ifade edilir. Aşağıdaki şekilde yer alan \( \hat{\beta} \), hem sapmasız hem de \( \beta^* \)’ya göre daha küçük varyanslardı. Bu nedenle etkin olduğu söylenir.

![Diagram](image)

Etkinliğin belirlenmesi için sapmasız tahmincilerin varyansları birbirine oranlanır. İki sapmasız tahmincinin varyansları şu şekilde oranlanır:

\[
\frac{\text{var}(\hat{\beta})}{\text{var}(\beta^*)}
\]

Bu oranı 1’den büyükse \( \beta^* \), küçük ise \( \hat{\beta} \) daha etkin tahminci olacaktır.

### 8.1.5. Doğrusal En İyi Sapmasızlık

Küçük örneklerde aranan bir diğer özellik, en iyi doğrusal sapmasız tahminci olma
özelliğidir. Bu özellik, seçilecek tahmincinin doğrusal sapması tahminciler arasında en düşük varyanslı olmasını göstermektedir. Bu tahminciler doğrusal, en iyi, sapmasız tahninci (DEST) olarak bilinir. Doğruluk koşulunun aranma sebebi, minimum varyansta tahmincilerin belirlenmesinde karşılaşılan sorundur. Doğruluk koşulu ile ilgili şu üç farklı durumdan bahsetmek mümkündür:

- Etkin tahminci, örnek gözlemlerinin doğrusal fonksiyonudur. Bu durumda en iyi doğrusal sapmasız tahninci aynı zamanda etkin tahnincidir.

- Etkin tahminci yaklaşık olarak doğrusaldır. Bu durumda en iyi tahminci, doğrusal sapmasız tahninci etkin değildir; fakat en iyi doğrusal sapmasız tahninciyle etkin tahnincilerin varyansları birbirine yakın olacaktır.

- Etkin tahminci doğrusal değilse, etkin tahnincisinin varyansı en iyi doğrusal sapmasız tahnincinin varyansından oldukça küçük olacağını için bu durum sorun yaratacaktır.

8.1.6. En Küçük Ortalama Kareli Hata

Sapmasızlık ve en küçük varyans özelliklerinin bir arada bulunmadığı durumlarda uygun tahmincinin seçimi ortalama kareli hata (OKH) kriteri dikkate alınarak yapılır. Sapmasızlık ve en küçük varyans özelliklerinin bileşimi olan ortalama kareli hata; tahmincinin ana kütle parametresi ile farklılarının karelerinin beklenen değerini almak suretiyle hesaplanabilir:

\[ OKH = E \left( \hat{\beta} - \beta \right)^2 \]

Sapmasız ama büyük varyanslı bir tahminci ile sapmalı fakat küçük varyanslı bir tahminci arasında seçim yaparken OKH’si daha küçük olan seçilebilir. Bununla birlikte modelin amacı bağımlı değişken için yapılan tahminlere kesinlik kazandırmak ise en küçük varyanslı, sapmalı tahmin edici seçilabilir.

8.1.7. Yeterlilik

Yeterlilik, tahmincinin tahmin edilecek parametre hakkında örnekte mevcut olan tüm bilgiyi kullanması özelliğidir. Bu durum, başka hiçbir tahmincinin ana kütle parametresi hakkında daha fazla bilgi veremeyeceği anlamına gelmektedir.

8.2. Büyük Örnek Özellikleri

8.2.1. Asimptotik Sapmasızlık

Eğer \( \hat{\beta} \) tahmincinisinin asimptotik ortalaması ana kütleinin gerçek \( \beta \) parametresine eşitse bu tahminci ilgili parametrenin asimptotik sapmasız tahmincisidir. Bir tahmincinin asimptotik olarak sapmasız olup olmadığını tespit edebilmek için örnek büyüklüğü (n)
sonsuza giderken limitini almak gerekmektedir:

\[
\lim_{n \to \infty} E(\hat{\beta}) = \beta \quad \text{durumunun sağlanması hâlinde } \hat{\beta}, \ \beta \text{'nün asimtotik sapmasız tahmincisidir.}
\]

Tahmincisinin küçük örneklerde sapmasız olması hâlinde aynı zamanda asimptotik olarak da sapmasız olduğu söylenebilir. Fakat bu durumun tersi geçerli değildir. Çünkü örnek büyüklüğü arttığında zaman sapmanın ve \( \beta \text{'nın örneklemeyi dağılımdaki değişkenliğinin azalmasıdır.}

8.2.2. Tutarlılık

Eğer bir \( \hat{\beta} \text{ tahmincisi, asimptotik sapmasızsa ve örnek büyüklüğü sonsuza giderken varyansı sıfıra yaklaşıyorsa bu tahmincinin tutarlı olduğu söylenebilir. Diğer bir ifadeyle örnek birim sayısının artarak sonsuza yaklaşması hâlinde } \hat{\beta}, \ \beta \text{'ya yaklaşıyorsa } \hat{\beta}, \ \beta \text{'nün tutarlı tahmincisidir:}

\[
\lim_{n \to \infty} E(\hat{\beta}) = \beta \quad \text{ve} \quad \lim \text{Var}(\hat{\beta}) = 0
\]

İkinci durum, tahmincisinin tutarlı olup olmadığını belirlemeye kullanılabilir. \( n \to \infty \) iken varyansın limitini alır ve bu değer 0'a eşit olduğu bulunursa tahmincinin tutarlı olduğu ifade edilir. Fakat bu koşulun gerçekleşmemesi tahmincinin tutarsız olduğunu göstermez.

8.2.3. Asimptotik Etkinlik

Tutarlı tahminciler arasında en küçük asimptotik varyanslı tahminci asimptotik etkin tahminci olarak ifade edilir. Asimptotik etkinlik, sadece son ortalama ve sonlu varyansı olan tahminciler için aranalabileceği bir özelliktir. Dolayısıyla bir tahmincinin asimptotik olarak etkin olması için üç koşul aranmaktadır:

- \( \hat{\beta} \) \text{'nin son ortalama ve sonlu varyanslı bir asimptotik dağılımı var ise,
- \( \hat{\beta} \) tutarlı ise,
- \( \beta \text{'nın tutarlı tahmincilerinden asimptotik varyansı en küçük olan tahminci } \hat{\beta} \text{ ise}

\( \hat{\beta}, \ \beta \text{'nın asimtotik etkin tahmincisidir.} \)
Uygulamalar
Uygulama Soruları
Bu Bölümde Ne Öğrendik Özeti
Bölüm Soruları

1) Aşağıdaki X değerleri için \[ \hat{\beta} = \sum \frac{x_i y_i}{\sum x_i^2} = \sum k_i y_i \] eşitliğinden hareketle \( k \) tartışının özelliklerini gösteriniz?

| X | 2 | 3 | 4 | 3 |

2) Küçük örnek özelliklerinden sapmasızlık ile ilgili olarak aşağıdaki ifadelerde boş kısımları doldurunuz.

a. Sapmasızlık, ............ anlamına gelmemekte, çok sayıda örnek alındığında .....................elde edileceği anlamına gelmektedir. Diğer bir ifadeyle ana küteden tekrar tekrar örnek çekip bu örneklerle yapılacak tahmincilerle ana kütle parametreleri tahmin edildikçe elde edilen tahminciler .............. doğru olacaktır.


3) Şekilde iki tahminci için grafikleri verilmiştir;

Bu tahmincilerden hangisinin daha iyi bir tahminci olduğu nasıl ifade edilebilir? Neden?

4) Küçük örnek özelliklerinden etkinlik ile ilgili olarak aşağıdaki ifadedeki boş kısımları doldurunuz.

................. tahmincilerden ............... varyansa sahip olan tahmincinin ana kütle parametresinin gerçek değeri etrafında yoğun olduğu söylenir ve bu tahmincinin ..........olduğu ifade edilir.
5) Aşağıdaki şekilde yer alan tahmincilerden hangisi daha etkindir? Neden?

![Diagram](image)


7) Sapmasız ama büyük varyanslı bir tahminci ile sapmalı fakat küçük varyanslı bir tahminci arasında seçim yaparken nasıl karar verirsiniz?

8) Büyük örnek özelliklerini açıklayınız?
9. ORİJİNDEN GEÇEN REGRESYON VE DOĞRUSAL OLMAYAN MODELLER
Bu Bölümde Neler Öğreneceğiz?

9. ORJİNDEN GEÇEN REGRESYON VE DOĞRUSAL OLMAYAN MODELLER

9.1. Orijinden Geçen Regresyon

9.2. Doğrusal Olmayan Modeller

9.2.1. Parametrelerin Özellikleri ile Doğrusallıktan Sapma

9.2.1.1. Tam Logaritmik Kalıp

9.2.1.2. Yarı Logaritmik Kalıp

9.2.2. Değişkenlerin Özellikleri ile Doğrusallıktan Sapma
Bölüm Hakkında İlgi Oluşturan Sorular
Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri
Anahtar Kavramlar
Giriş

9. ORIJİNDEN GEÇEN REGRESYON VE DOĞRUSAL OLMAYAN MODELLER

9.1. Orijinden Geçen Regresyon


Orijinden geçen regresyon modelleri, iktisat kuramının regresyon sabitine yer vermemesinden veya değişken dönüşümleri sonucunda otonom parametrein ortadan kalkmasından ortaya çıkabilir.

Örnek regresyon modeli;

\[ \hat{Y}_i = \hat{\beta}_1 X_i + u_i \]

ile ifade edilir.

Örnek regresyon modeline en küçük kareler yöntemi uygulanırsa;

\[ \sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 X_i)^2 \rightarrow \text{min} \]

\[ \hat{\beta}_1 = \frac{\sum X_i Y_i}{\sum X_i^2} \]

olacaktır. Kalıntı kareler toplamında \( \hat{\beta}_1 \)’e göre kısmi türev alınarak sıfıra eşitlenirse

\[ \frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_1} = -2 \sum (Y_i - \hat{\beta}_1 X_i) X = 0 \]

186
ve sadeleştirilmeler yapılarrsa

\[ \hat{\beta}_i = \frac{\sum X_i Y_i}{\sum X_i^2} \]

sonucuna ulaşılır. Orijinden geçen regresyon modelindeki \( \hat{\beta}_i \), sabit terim içeren regresyon modelerindeki \( \hat{\beta}_i \)‘in ortalama tahsisi yapılmamış hâlidir. Aralarındaki ilk fark; söz konusu parametrenin hesaplanmasında ortaya çıkmaktadır. Orijinden geçen ana kütle regresyon modeli ise;

\[ Y_i = \beta_i X_i + u_i \]

şeklinde ifade edilmektedir. Parametre ile ilgili özellikleri ise aşağıdaki belirtilmektedir:

Eğilimsizlik özelliği için,

\[ \hat{\beta}_i = \frac{\sum X_i Y_i}{\sum X_i^2} \]

\[ \hat{\beta}_i = \frac{\sum X_i (\beta_i X_i + u_i)}{\sum X_i^2} = \beta_i \frac{\sum X_i^2 + \sum X_i u_i}{\sum X_i^2} \]

\[ \hat{\beta}_i = \beta_i + \frac{\sum X_i u_i}{\sum X_i^2} \quad E(X_i u_i) = 0 \text{ dan} \]

\[ \hat{\beta}_i = \beta_i \]

Her iki tarafın beklenen değeri alınırsa;

\[ E(\hat{\beta}_i) = E(\beta_i) \]

sonucuna ulaşılır. Böylece tahmin edilen parametreler eğilimsizlik özelliğini korumaktadırlar.

\[ E(\hat{\beta}_i - \beta_i)^2 = E\left( \frac{\sum X_i u_i}{\sum X_i^2} \right)^2 \quad \hat{\beta}_i = \beta_i + \frac{\sum X_i u_i}{\sum X_i^2} \]

X bağımsız değişkeni olasılığı olmadığı ve u sabit varyanslı ve ardışık bağımsız olduğu için
\[ \text{Var}(\hat{\beta}_i) = E(\hat{\beta}_i - \beta_i)^2 = \sigma^2_u \sum X_i^2 \]

sonucuna ulaşırlar. Varyans formülündeki ana kütle hata teriminin varyansının tahmini ise

\[ \sigma^2_u = \frac{\sum \hat{u}_i^2}{n-1} \]

sekillerde hesaplanmaktaadır. Dikkat edilirse regresyon modelinde tahmin edilen bir parametre yer aldığından serbestlik derecesi 1’e eşittir.

Orijinden geçen regresyon modelinin önemli bir özelliği; kalıntıların toplamının her zaman 0 olmamasıdır (\( \sum \hat{u}_i \neq 0 \)).

\[ \sum \hat{u}_i \neq 0 \]

\[ \sum \hat{u}_i = 0 \quad E(u_i) = 0 \quad \text{varsayılır} \]

\[ \sum Y_i = \hat{\beta}_i \sum X_i + \sum \hat{u}_i \quad \sum \hat{u}_i = 0 \quad \text{olacağından} \]

\[ \hat{\beta}_i = \frac{\sum Y_i}{\sum X_i} = \frac{\bar{Y}}{\bar{X}} \]

sonucuna ulaşılır ki bu sonuç \( \hat{\beta}_i = \frac{\sum X_i Y_i}{\sum X_i^2} \) ye eşit değildir.

Orijinden geçen regresyonda sağlanabilen tek koşul hata terimlerinin bağımsız değişken ile ilişkisiz olduğudur. \( \sum \hat{u}_i X_i = 0 \), \( E(u_i X_i) = 0 \)

Bağımlı değişkenin gözlemlenen değerlerinin ortalaması, tahmin edilen değerlerin ortalamasına eşit değildir.

\[ \bar{Y} \neq \bar{Y} \quad \text{söyle ki} \]

i. gözlem için

\[ Y_i = \hat{Y}_i + \hat{u}_i \]

i=1…n bütün gözlemler için toplanırsa

\[ \sum Y_i = \sum \hat{Y}_i + \sum \hat{u}_i \]

188
Her iki taraf gözlem sayısı n’ye bölünirse
\[
\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{y}_i + \sum_{i=1}^{n} \hat{u}_i
\]
\[
\bar{Y} = \bar{Y} + \bar{\hat{u}}
\]
sonucu \( \bar{Y} = \bar{Y} \) geçerli olmadığını göstermektedir.

Orijinden geçen regresyon modelinde belirginlik katsayısı \( (r^2) \) negatif değer alabilir. Nedeni ise modelde sabit terim olduğunu varsaymasıdır.

\[
r^2 = 1 - \frac{\sum\hat{u}_i^2}{\sum y_i^2}
\]

\[
\sum\hat{u}_i^2 \leq \sum y_i^2 - \hat{\beta}_i \sum x_i^2 \leq \sum y_i^2
\]

Normal durumda, yukarıda görüldüğü üzere, kalıntıların karelerinin toplamı bağımlı değişkendeki toplam değişimden küçüktür. Böylece \( r^2 \) negatif değer almaz. Orijinden geçen regresyonunda ise

\[
\sum\hat{u}_i^2 = \sum y_i^2 - \hat{\beta}_i \sum x_i^2
\]
\[
\sum\hat{u}_i^2 < \sum y_i^2 - \hat{\beta}_i \sum x_i^2 < \sum NY^2
\]

olacağı garanti edilemez.

Orijinden geçen regresyon modelinde ham \( r^2 \) hesaplanır.

\[
ham\ r^2 = \frac{\left(\sum X_i Y_i\right)^2}{\sum X_i^2 \sum Y_i^2}
\]

Ham \( r^2 \) ile \( r^2 \) doğrudan karşılaştırılamaz. Çok güçlü bir neden yoksa sabit terimi modelin kullanılması önerilmektedir. Sabit terimin tahmini istatistiksel açından anlamsızsa sıfır noktasından geçen regresyon şekilde yorumlanır. Sabit terim olması gerekiyorken sabit terimsiz model tahmin edilirse spesifikasyon hatası yapılmıştır.

9.2. Doğrusal Olmayan Modeller

Doğrusal regresyon modellerinde bağımsız değişkenlerin bağımlı değişkeni etkilemeleri doğrusal ve toplanabilir bir matematiksel özelliği sahiptir.
Genel doğrusal regresyon modelinde bağımlı değişken ile bağımsız değişken arasında olduğu varsayılan

\[ Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \cdots + \beta_k X_{ik} + u \quad i = 1 \cdots n \]

şeklinde ifade edilen doğrusal ilişki, çoğu iktisadi ilişki için uygun olmayabilir. İktisadi ilişkiler veri ortamında çoğu iktisadi ilişkide doğrusal olmayan nitelikte ilişkilerin olması beklenir.

Sıklıkla rastlanan doğrusal olmayan ilişki biçimlerinden bazıları X çok terimlisinin yer aldığı regresyon modelleridir. Cobb-Douglas tipi üretim fonksiyonu yine doğrusal olmayan ekonometrik modeller arasında yer almaktadır.

\[ Q = AK^{\alpha}L^{\beta} \]

Burada \( Q \) üretimi, K kapitali, L ise emeği ifade etmektedir. A ise teknoloji katsayısıdır. Üretim pozitif bir değer olması gerektiğini modelde A>0, K>0 ve L>0 olmalıdır.

Bir modelin doğrusallıktan sapması iki biçimde ortaya çıkmaktadır:

I. Parametrelerin Özellikleri ile Doğrusallıktan Sapma

- Tam Logaritmik Kalıp
- Yarı Logaritmik Kalıp
- Logaritmik-Doğrusal
- Doğrusal-Logaritmik

II. Değişkenlerin Özellikleri ile Doğrusallıktan Sapma

9.2.1. Parametrelerin Özellikleri ile Doğrusallıktan Sapma

Parametrelerinin özelliklerinden dolayı doğrusallıktan sapmış modellerde parametreler değişkenlerin üssü şeklindedir ve birbiri ile çarpılabilir durumdadır. Söz konusu modelleri doğrusal hale getirmek için her iki tarafını logaritmasının alınması gerekmektedir. Bu şekilde elde edilmiş modeller tam logaritmik model ve yarı logaritmik model olmak üzere ikiye ayrılmaktadır.

9.2.1.1. Tam Logaritmik Kalıp

\[ Y = \beta_0 X_1^{\alpha_1} X_2^{\alpha_2} e^\nu \]

Yukarıdaki model, parametrelerinin özellikleri bakımdan doğrusal değildir. Hata
terimi toplanabilir nitelike olmayıp çarpar bir değer niteliğindedir. Eğer model;

\[ Y = \beta_0 X_1^{\theta_1} X_2^{\theta_2} u \]

şeklinde olursa hata terimi çarpar biçimdedir. Ancak bu durumda \( E(u_i) = 0 \) varsayımı geçerli değildir.

\[ Y = \beta_0 X_1^{\theta_1} X_2^{\theta_2} e^u \]

modelinde ise klasik regresyon modelinin temel varsayımları geçerliliğini korurlar.

\[ E(u_i) = 0 \quad E(u_i^2) = \sigma^2 \quad E(u_i u_j) = 0 \quad \text{for} \quad i \neq j \quad E(u_i X_i) = 0 \]

Ayrıca bu model monomial karaktere sahiptir. Değişkenlerden herhangi birinin değerinin sıfır olması, \( Y \)'nin değerini sıfır yapar. Bağımlı değişkenin pozitif olması için \( \beta_0 > 0 \quad X_1 > 0 \) ve \( X_2 > 0 \) şartı vardır. Bu modelin doğrusalına dönüştülebilir.
\[ \varepsilon_{Yx_i} = \frac{\partial Y}{\partial X_i} Y \]

\( Y \) fonksiyonunun \( X_i \)'e göre kısmi türevi alınır.

\[ \frac{\partial Y}{\partial X_i} = \beta_i (\beta_0 X_i^{\beta_i - 1} X_2^{\beta_i} e^u) = \beta_i X_i^{-1} (\beta_0 X_i^\beta X_2^{\beta_i} e^u) \]

\( Y = \beta_0 X_1^\beta X_2^{\beta_i} e^u \) bilinmekte

Böylece \( \frac{\partial Y}{\partial X_i} = \beta_i \frac{Y}{X_i} \)'e eşit olur.

\[ \varepsilon_{Yx_i} = \frac{\partial Y}{\partial X_i} Y \] \( \varepsilon \) eşliğinde \( \frac{\partial Y}{\partial X_i} Y \) yerine \( \beta_i \frac{Y}{X_i} \) konur ve gerekli sadeleştirmeler yapılar:

\[ \varepsilon_{Yx_i} = \frac{\partial Y}{\partial X_i} \frac{Y}{X_i} = \beta_i \frac{Y}{X_i} \frac{X_i}{Y} = \beta_i \]

\( Y \)'nin \( X_i \)'e göre sabit elastikiyetinin \( \beta_i \) parametresine eşit olduğu sonucuna ulaşılr.

Doğrusal olmayan modelde tahmin edilen parametreler doğrusal modeldeki gibi eğimi değil, elastikiyeti ifade ettiklerinden yorumları da buna bağlı olarak değişsecektir. Örneğin; \( \beta_i \) parametresi için \( X_i \)'deki %1'lik değişimye karşılık \( Y \% \beta_i \) kadar değişsecektir. Örnek kütlede elde edilen \( \hat{\beta}_i \) ve \( \hat{\beta}_2 \) tahminleri eğilimsizlik özelliklerini korurlar. Ancak \( \beta_0^* \) eğilimsiz olduğu hâlde, logaritmik dönüşümü \( \beta_0 \) eğilimli fakat tutarlı bir tahmindir.

**Açıklayıcı Örnek 1**

Cobb-Douglas tipi üretim fonksiyonu tam logaritmik kalıba örnek teşkil eden ekonomik bir modeldir.

\[ Q = AK^\alpha L^\beta e^u \]

Modelde;

\( Q \) = üretim hacmi

\( K \) = Kapital

\( L \) = Emek

değişkenlerini verir iken \( A \) Teknoloji seviyesini gösteren sabit parametredir. Modelin doğrusallaştırılmış biçimi; \( \ln Q = \ln A + \alpha \ln K + \beta \ln L + u \) biçimindedir.

Aşağıdaki örnek regresyon modelinden tahmin edilen
\( \dot{Q} = \dot{A} + \dot{\alpha}K + \dot{\beta}L \)  
\( \dot{Q} = \ln Q \quad \dot{A} = \ln A \quad \dot{K} = \ln K \quad \dot{L} = \ln L \)

parametreler(\( \hat{\alpha}, \hat{\beta} \));

\[ \hat{\alpha} = \frac{\partial Q}{\partial K} \frac{K}{Q} \]  
Kapitalin marjinal verimlilik elastikiyeti

\[ \hat{\beta} = \frac{\partial Q}{\partial L} \frac{L}{Q} \]  
Emeğin marjinal verimlilik elastikiyeti şeklinde elde edilmektedir.

Açıklayıcı Örnek 2

\( D_x = \beta_0 p_x Y^{\beta_1} e^{\mu} \) modelinde;

D= X malının talebini  
P= X malının fiyatını  
Y= X malını talep edenlerin ortalama gelirini ifade etmektedir.

Modelin parametreleri;

\[ \beta_1 = \frac{\partial D_x}{\partial P_x} \frac{P_x}{D_x} \]  
talebin fiyat elastikiyetine

\[ \beta_2 = \frac{\partial D_x}{\partial Y} \frac{Y}{D_x} \]  
talebin gelir elastikiyetine eşittir.

9.2.1.2. Yarı Logaritmik Kalıp

Yarı logaritmik kalıpta değişkenlerden biri geometrik değeriyle modele girerken diğeri aritmetrik değeriyle modelde yer almaktadır. Söz konusu modeller; kendi içinde logaritmik doğrusal ve doğrusal logaritmik olmak üzere ikiye ayrılır.

- Logaritmik-Doğrusal Model

\[ Y = e^{\beta_0 + \beta_1 X} \]

\[ \ln Y = \beta_0 + \beta_1 X \log-Doğ \]

\( \beta_1 \) parametresi, X’te meydana gelen mutlak bir değişmenin Y’de meydana getirdiği nispi değişmeye gösterir. \( \beta_1, \ln Y \)’nin X’e göre türevidir.
\[ \beta_i = \frac{d(lnY)}{dX} \quad \text{ve} \quad \frac{1}{\ln Y} \frac{dY}{dX} = \left( \frac{dY}{Y} \right) \frac{1}{dX} = \frac{Y^{'} \text{deki nisbi değişme}}{X^{'} \text{deki mutlak değişme}} \]

Yukarıdaki eşitlikten \( \frac{dY}{dX} = \beta_i Y \) elde edilir. Elastikiyet formülü yeniden yazılır ve \( \beta_i Y \) yerine yazılsa

\[ \epsilon_{xy} = \frac{\partial Y}{\partial X} \frac{X}{Y} = \beta_i Y \frac{X}{Y} \]

sonucuna ulaşırlar.

Sadeleştirmeler yapılar daha \( \epsilon_{xy} = \beta_i X \) olmaktadır.

**Açıklayıcı Örnek**

\[ Y_t = Y_0(1 + r)^t \]

Büyüme modelinde;

\[ Y_0 = \text{Başlangıç yılı gelirini} \]

\[ Y_t = t \text{ yılındaki geliri} \]

\[ r = \text{Büyüme oranını göstermektedir.} \]

Denklemin her iki tarafının logaritması alınır,

\[ \ln Y_t = \ln Y_0 + t \ln(1 + r) \]

elde edilir.

\[ \beta_1 = \ln Y_0 \quad \beta_2 = \ln(1 + r) \]

\[ \ln Y_t = \beta_1 + \beta_2 t \text{ Log-Doğ model} \]

Log-Doğ modelde \( \ln Y_t \) geometrik dizi iken \( t ( =1,2,3,\ldots) \) zamanı gösterdiği için aritmetik seri özelliği taşırlar.

\( Y^{'} \) nisbi değişme 100 ile çarpılsa bağımsız değişkendeki mutlak değişmeye karşılık \( Y^{'} \)deki yüzde değişme ya da büyüme oranı sağlanmış olur.

- **Doğrusal-Logaritmik Model**
Doğrusal-logaritmik formda model kalıbı;

\[ Y = \beta_0 + \beta_1 \ln X \]

şeklindedir. Burada \( \beta_1 \); X’teki nispi değişim Y’de meydana getirdiği mutlak değişimeyi göstermektedir. Modelden X’in Y’ye göre marjinal etkisi;

\[ \frac{dY}{dX} = \beta_1 \frac{1}{X} \]

olduğundan

\[ \varepsilon_{yx} = \frac{\partial Y}{\partial X} = \beta_1 \frac{1}{X} \frac{Y}{Y} = \beta_1 \frac{1}{Y} \]

şeklinde bulunabilmektedir. Burada elastikiyetin Y’ye göre değiştiği görülmektedir.

**9.2.2. Değişkenlerin Özellikleri ile Doğrusallıktan Sapma**

Bir model bağımlı ve bağımsız değişkenlerinden dolayı doğrusallıktan sapabilmektedir. Söz konusu bu form aşağıdaki gibi ifade edilebilir: 

\[ Y = \beta_0 + \beta_1 \frac{1}{X} \]

Bu modeli doğrusallıktan saptıran sebep X’in tersi ile modele girmesidir. X’in değeri sonsuz şekilde artarken Y’nin sınırlı bir değer aldığı durumda kullanılan bir model türüdür. Bu modelin aşağıdaki şekilde gösterilen üç hâli çok önemlidir.

**Şekil 1:** Ters Modelin Parametre Değerlerine Göre 3 Farklı Hâli

Karşılaştırıma yapmak açısından doğrusal, tam logaritmik, logaritmik-doğrusal, doğrusal-logaritmik modeller ile ters modele ait eğim ve elastikiyet formülleri aşağıdaki
tabloda özetlenmiştir.

<table>
<thead>
<tr>
<th>Model</th>
<th>Eğim</th>
<th>Elastikiyet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doğrusal Model</td>
<td>$Y = \beta_0 + \beta_1 X$</td>
<td>$\beta_1 \left( \frac{X}{Y} \right)^*$</td>
</tr>
<tr>
<td>Tam Logaritmik Model</td>
<td>$\ln Y = \beta_0 + \beta_1 \ln X$</td>
<td>$\beta_1 \frac{Y}{X}$</td>
</tr>
<tr>
<td>Logaritmik-Doğrusal Model</td>
<td>$\ln Y = \beta_0 + \beta_1 X$</td>
<td>$\beta_1 \left( X \right)^*$</td>
</tr>
<tr>
<td>Doğrusal-Logaritmik Model</td>
<td>$Y = \beta_0 + \beta_1 \ln X$</td>
<td>$\beta_1 \left( \frac{1}{Y} \right)^*$</td>
</tr>
<tr>
<td>Ters Model</td>
<td>$Y = \beta_0 + \beta_1 \frac{1}{X}$</td>
<td>$-\beta_1 \frac{1}{X^2}$</td>
</tr>
</tbody>
</table>

*İşareti, ortalamaya göre de elastikiyet hesaplanabileceğini ifade etmektedir.
Uygulamalar
Uygulama Soruları
Bu Bölümde Ne Öğrendik Özeti
Bölüm Soruları

1. Orijinden geçen regresyon modelinde \( \sum \hat{u}_i \neq 0 \) olduğunu gösteriniz.

2. Orijinden geçen regresyon modelinde hata terimleri ve açıklayıcı değişken ile ilgili ne tür bir varsayım yapılmaktadır?

3. \( \bar{Y} = \bar{Y} \) eşitliğinin orijinden geçen regresyon modelinde geçerli olup olmadığını gösteriniz.

4. Orijinden geçen regresyon modelinde determinasyon katsayısının hangi aralıktaki değerler alabileceği belirtir, nedenini açıklayınız.

5. Bir modelin doğrusallıktan sapmasının hangi şekilde ortaya çıkabileceğini açıklayınız.

6. Doğrusal olmayan bir regresyon modelinin **monomial** karakterde olması ne anlama gelmektedir?

7. \( \ln Y = \ln \beta_0 + \beta_1 \ln X_1 + \beta_2 \ln X_2 + u \) olarak gösterilen logaritmik doğrusal modelde yer alan \( \beta_1 \) ve \( \beta_2 \) parametreleri neyi ifade etmektedir? Açıklayınız.


9. \( Y_t = Y_0(1 + r)^t \) şeklindede gösterilen büyüme modelinde büyüme oranının nasıl hesaplanabileceği gösteriniz.

10. \( Q = AK^nL^n\alpha \) ile gösterilen Cobb-Douglas tipi üretim fonksiyonunda emeğin ve kapitalin marjinal verimliliklerinin nasıl hesaplanacağını gösteriniz.
10. ÇOK DEĞİŞKENLİ REGRESYON ANALİZİ
Bu Bölümde Neler Öğreneceğiz?

10. ÇOK DEĞİŞKENLİ REGRESYON ANALİZİ

10.1. Çok Değişkenli Regresyon Analizi

10.2. Çok Değişkenli Regresyonda Parametre Tahmini

10.3. Çok Değişkenli Regresyonda Tahmin Edilen Parametrelerin Varyans ve Kovaryansı

10.4. Çoklu Belirginlik Katsayısı (R2)

1.5. Varyans Analiz Tablosu Yaklaşımı: F testi
Bölüm Hakkında İlgı Oluşturan Sorular
Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri
Anahtar Kavramlar
Giriş

Doğrusal regresyon modelinde bir bağımlı değişken, çoğunlukla tek bir açıklayıcı değişkene bağlı olarak incelenemez. Buna bağlı olarak tek açıklayıcı değişkenli modellerin uygulamada yeterli olmadığını ifade etmek mümkündür. Bu bölümde birden fazla açıklayıcı değişken içeren, diğer bir ifade ile çok değişkenli doğrusal regresyon modelleri incelencektir.
10. ÇOK DEĞİŞKENLİ REGRESYON ANALİZİ

10.1. Çok Değişkenli Regresyon Analizi

Basit doğrusal regresyon modelinde bağımlı değişken sadece bir bağımsız değişken tarafından açıklanmaktadır. Örneğin; basit regresyon modelinde X malının talebi ($D_x$) sadece X malının fiyatına ($P_x$) ve rastlantısal hata terimine ($u$) bağlıdır. Böylece X malının talebine ilişkin basit regresyon modeli;

$$D_x = \beta_0 + \beta_1 P_x + u$$

şeklinde olur. Oysaki gerçekte bir malın fiyatı sadece kendi fiyatına değil, malı talep eden kişilerin gelirine ($Y$), eğer varsa tamamlayıcı malın fiyatına ($P_T$) ve rakip malın fiyatına ($P_R$) ve nihayet zevk ve alışkanlıklara ($Z$) göre değişiklik gösterecektir. Dolayısıyla regresyon modelinde saydığımız değişkenlere de yer verilmelidir. Bağımlı değişkeni açıklayan birden fazla bağımsız değişkenin yer aldığı regresyon modelleri, çok değişkenli regresyon modelleri olarak adlandırılır. Basit regresyona göre geçmişte daha yakından olan çok değişkenli regresyon modeli;

$$D_x = \beta_0 + \beta_1 P_x + \beta_2 P_T + \beta_3 P_R + \beta_4 Z + u$$

gibidir.

Regresyon modelinin mümkün olduğunca sade olması arzu edilmektedir. Ancak bağımlı değişken üzerinde etkili olan değişken(ler)in bilerek veya bilmeyerek model dışında bırakılması, daha ileriki konularda görülecek olan spesifikasyon hatalarından dışlanmış değişken sorunun istenmeyen sonuçlarını doğuracaktır.

Genel olarak çok değişkenli bir regresyon modeli;

$$Y_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \ldots + \beta_k X_{ik} + u_i \quad i = 1, 2, \ldots, n$$

 şeklini alır. Modelde $k$ sayıda parametre, $k-1$ sayıda bağımsız değişken vardır. Modelde $X_i$ değişkenine yer verilmemiştir. Sabit parametre $\beta_1$, bütün gözlemler için ($i=1\ldots n$) değeri 1 olan $X_i$ değişkeninin ($X_{i1} = 1$) katsaysıdır.

$$Y_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \ldots + \beta_k X_{ik} + u_i \quad i = 1, 2, \ldots, n$$

$X_{i1} = 1$ alınmasının sebebi, daha sonra işlenecek genel doğrusal regresyon modelindeki matris cebiri ile uyumlu olması içindir.

Bağımlı değişkenin beklenen değeri (koşullu ortalama) $X_2$ ’den $X_k$ kadar bağımsız değişkenlerin sabit değerlerine bağlıdır.

$$E(Y_i|X_{2i}, X_{3i}, \ldots, X_{ki}) = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki}$$
Modelde yer alan parametrelerden örneğin $\beta_2$; diğer değişkenler sabit iken $X_2$’deki bir birimlik değişimeye karşılık $Y$’nin beklenen değerindeki değişimyi ölçer. Matematiksel açıdan diğer değişkenler sabit iken $E(Y|X_{2i}, X_{k<i}, X_{k+1},...,X_k)$'ün $X_2$'ye göre eğimini verir. $\beta_1$, $X_2 = X_3 = .... = X_k = 0$ olduğunda $Y$'nin ortalama değeridir. Modelde yer alan değişkenlerin $Y$ üzerindeki etkisini ifade eder.

Analizlerimizde kolaylık olması açısından çok değişkenli regresyon modellerinin en basit hali olan iki bağımsız değişkenli model ile devam edilecektir.

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \quad i = 1, 2, ..., n$$

$$E(Y_i|X_{2i}, X_{3i}) = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i}$$

Bağımlı değişkenin değeri, basit regresyonda olduğu gibi bağımlı değişkenin beklenen değerine $E(Y_i|X_{2i}, X_{3i})$ ve yine rastlantısal bir değişken olan hata terimine bağlıdır.

Klasik doğrusal regresyon modelinin varsayımları

i. Hata teriminin beklenen değeri (koşullu ortalaması) sıfırdır.

$$E(u_i | X_{2i}, X_{3i}) = E(u_i) = 0$$

ii. Hata terimi normal dağılıma uygun dağılmaktadır.

$$u_i \sim N(0, \sigma_u^2)$$

iii. Homoskedasite söz konusudur.

$$Var(u_i) = E(u_i^2) = \sigma^2$$

iv. Otokorelasyon bulunmamaktadır.

$$Kov(u_i, u_j) = 0 \quad i \neq j$$

v. Hata terimleri bağımsız değişkenlerden bağımsızdır.

$$Kov(u_i, X_{2i}) = Kov(u_i, X_{3i}) = 0$$

vi. Modelin spesifikasyonu doğru belirlenmiştir.

vii. Bağımsız değişkenler arasında tam çoklu doğrusal bağlantılı yoktur. Çoklu doğrusal bağlantılı çok değişkenli regresyona özgü bir sorundur.

viii. Bağımsız değişkenlerde ölçme hatası bulunmamaktadır.
10.2. Çok Değişkenli Regresyonda Parametre Tahmini

Çok değişkenli örnek regresyon modeli;

\[ \hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} \]

şeklinde ve gözlemdeki \( Y \) nin değeri \( \hat{Y}_i \) ve \( u_i \) 'ye bağlıdır.

Buna göre;

\[ Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i \]

olaraktır. En küçük kareler yöntemi ile en uygun parametrelerin tahmini için yine kalıntı kare toplamının (\( \sum \hat{u}_i^2 \)) minimum olması ve bu koşulun sağlanabilmesi için ise kalıntı kare toplamı fonksiyonunda (\( \sum \hat{u}_i^2 \)) \( \hat{\beta}_1, \hat{\beta}_2 \) ve \( \hat{\beta}_3 \) ’e göre kısmi türevlerin sıfıra eşitlenmesi gerekecektir.

\[ \frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_1} = 0 \quad \frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_2} = 0 \quad \frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_3} = 0 \]

Gerekli işlemler ve sadeleştirilmeler yapılarak aşağıdaki denklem sistemi elde edilir.

\[ \sum Y_i = n\hat{\beta}_1 + \hat{\beta}_2 \sum X_{2i} + \hat{\beta}_3 \sum X_{3i} \]
\[ \sum Y_i X_{2i} = \hat{\beta}_1 \sum X_{2i} + \hat{\beta}_2 \sum X_{2i}^2 + \hat{\beta}_3 \sum X_{3i} X_{2i} \] Normal Denklemler
\[ \sum Y_i X_{3i} = \hat{\beta}_1 \sum X_{3i} + \hat{\beta}_2 \sum X_{2i} X_{3i} + \hat{\beta}_3 \sum X_{3i}^2 \]

Denklem sistemi üç bilinmeyenli üç denklemden oluşmakta ve en küçük kareler yönteminin normal denklemleri adını almaktadır. Denklemler eş anlı çözümlendiğinde modelin bilinmeyenleri \( \hat{\beta}_1, \hat{\beta}_2 \) ve \( \hat{\beta}_3 \) hesaplanır. Normal denklemlerde yer alan unsurlar değişken kare toplamlarıdır, orijin 0 noktasını olup değişkenlerin 0’a uzaklığı hesaplamaları katılmıştır.

Ancak bu normal denklemlerinin çözümü oldukça fazla zaman alır. Bundan dolayı üç değişkenli regresyon modelini tahmin edebilmek için değişkenlerin ortalamalarından farklarına (\( Y_i - \bar{Y} = y_i, X_{2i} - \bar{X}_2 = x_{2i}, X_{3i} - \bar{X}_3 = x_{3i} \)) göre normal denklemler yazılabilir.

Öncelikle normal denklemlerden birincisinin

\[ \sum Y_i = n\hat{\beta}_1 + \hat{\beta}_2 \sum X_{2i} + \hat{\beta}_3 \sum X_{3i} \]

her iki yanı \( n \) gözlem sayısına bölünür ve
\[ \bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}_2 + \hat{\beta}_3 \bar{X}_3 \]

sonucuna ulaşılır.

Daha sonraki aşamada \( \bar{Y} - \bar{Y} \) oluşturulur. \( \bar{Y} - \bar{Y} \) dan

i. gözlem için

\[ \hat{Y}_i - \bar{Y} = \bar{Y} - \bar{Y} = (\hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i}) - (\hat{\beta}_1 + \hat{\beta}_2 \bar{X}_2 + \hat{\beta}_3 \bar{X}_3) \]

eşitliğe ulaşılır. \( \hat{\beta}_i \) sadeleştir ve gerekli kısaltma yapılsa

\[ \hat{Y}_i - \bar{Y} = \beta_2 (X_{2i} - \bar{X}_2) + \beta_3 (X_{3i} - \bar{X}_3) \]

'ten

\[ \hat{y}_i = \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} \]

sonucuna ulaşılır. Daha önce \( Y_i = \hat{y}_i + \hat{u}_i \) eşitliğinin değişkenlerin kendi ortalamalarında farklarıyla \( y_i = \hat{y}_i + \hat{u}_i \) şeklinde yazılabileceğini görmüştük.

Buna göre;

i. gözlem için \( y_i = \hat{y}_i + \hat{u}_i \) ’den

\[ \hat{u}_i = y_i - \hat{y}_i = y_i - (\hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i}) \]

olarak yeniden yazılabilir.

\( \hat{u}_i \) ’nin \( i = 1 \ldots n \)’ye kadar her iki tarafın kareleri toplamı alınır ve hata terimi kalıntı kareler toplamında \( \hat{\beta}_2 \) ve \( \hat{\beta}_3 \) ’e göre kısmi türev alınarak sıfıra eşitlenir.

\[ \sum \hat{u}_i^2 = \sum \left( y_i - (\hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i}) \right)^2 \]

\[ \frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_2} = 0 \quad \frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_3} = 0 \]

Kısımsı türevlerin sıfıra eşitlenip çözümlenmesiyle aşağıdaki denklem sistemi elde edilir.

\[ \sum x_{2i} y_i = \hat{\beta}_2 \sum x_{2i}^2 + \hat{\beta}_3 \sum x_{2i} x_{3i} \]
\[ \sum x_{3i} y_i = \hat{\beta}_2 \sum x_{3i}^2 + \hat{\beta}_3 \sum x_{3i} x_{3i} \]

Yukarıdaki denklem sistemine Cramer kuralı uygulanarak \( \hat{\beta}_2 \) ve \( \hat{\beta}_3 \) hesaplanır.
\[\hat{\beta}_2 = \frac{\sum x_{2i}y_i \sum x_{3i}^2 - \sum x_{3i}y_i \sum x_{2i}x_{3i}}{\sum x_{2i}^2 \sum x_{3i}^2 - (\sum x_{2i}x_{3i})^2}\]

\[\hat{\beta}_3 = \frac{\sum x_{3i}y_i \sum x_{2i}^2 - \sum x_{2i}y_i \sum x_{3i}x_{2i}}{\sum x_{3i}^2 \sum x_{2i}^2 - (\sum x_{3i}x_{2i})^2}\]

\[\hat{\beta}_i\] ise yukarıda oluşturulan

\[\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}_2 + \hat{\beta}_3 \bar{X}_3\]

denklemde yerine konularak hesaplanır.

Basit doğrusal regresyon modeli uzayda bir doğru belirlerken çok değişkenli regresyon modelleri bir düzlem belirler.

10.3. Çok Değişkenli Regresyonda Tahmin Edilen Parametrelerin Varyansı ve Kovaryansı

Çok değişkenli regresyon modelinde parametrelerin beklenen değeri basit regresyonda olduğu gibi ana kütledeki gerçek değerine eşittir.

\[E(\hat{\beta}_1) = \beta_1 \quad E(\hat{\beta}_2) = \beta_2 \quad E(\hat{\beta}_3) = \beta_3\]

Parametre tahminlerinin varyansları ise aşağıdaki eşitliklerden hesaplanır.

\[\text{Var}(\hat{\beta}_1) = \left[\frac{1}{n} + \frac{\bar{X}_2^2 \sum x_{3i}^2 + \bar{X}_3^2 \sum x_{2i}^2 - 2 \bar{X}_2 \bar{X}_3 \sum x_{2i}x_{3i}}{\sum x_{2i}^2 \sum x_{3i}^2 - (\sum x_{2i}x_{3i})^2}\right] \sigma^2\]

\[\text{Var}(\hat{\beta}_2) = \frac{\sum x_{3i}^2}{\sum x_{2i}^2 \sum x_{3i}^2 - (\sum x_{2i}x_{3i})^2} \sigma^2 = \frac{\sigma^2}{\sum x_{2i}^2 (1 - r_{23}^2)}\]

\[\text{Var}(\hat{\beta}_3) = \frac{\sum x_{3i}^2}{\sum x_{2i}^2 \sum x_{3i}^2 - (\sum x_{2i}x_{3i})^2} \sigma^2 = \frac{\sigma^2}{\sum x_{3i}^2 (1 - r_{23}^2)}\]

\[\text{Kov}(\hat{\beta}_2, \hat{\beta}_3) = \frac{-r_{23} \sigma^2}{(1 - r_{23}^2) \sqrt{\sum x_{2i}^2} \sqrt{\sum x_{3i}^2}}\]

Varyans ve ortak varyans eşitliklerindeki hata teriminin varyansının tahmini (\(\hat{\sigma}^2\))

\[\hat{\sigma}^2 = \frac{\sum e_{ii}^2}{n - 3}\]

e eşittir. Modelde tahmin edilecek parametre sayısı 3'tür. Kalıntı kareler
toplamı 

$$\sum \hat{u}_i^2 = \sum y_i^2 - \hat{\beta}_2 \sum y_i x_{2i} - \hat{\beta}_3 \sum y_i x_{3i}$$ ile hesaplanır.

10.4 Çoklu Belirginlik Katsayısı ($R^2$)

Çok değişkenli regresyon modelinde belirginlik katsayısı $R^2$ ile gösterilir. $R^2$’nin hesaplanması aşağıdaki gibidir.

$$R^2 = \frac{\sum \hat{y}_i^2}{\sum y_i^2} = \frac{\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}}{\sum y_i^2 - n\bar{Y}^2}$$

Veya

$$R^2 = 1 - \frac{\sum \hat{u}_i^2}{\sum y_i^2}$$

ve yine belirginlik katsayısı 0 ile 1 arasında değer alır.

$$0 \leq R^2 \leq 1$$

$R^2$’nin değeri yükseldikçe regresyon düzleminin örnek gözlemlerine uyumu artar.

k-1 sayıda değişkenli çok değişkenli bir regresyon modeli için $R^2$ formülü genişletilirse

$$R^2 = \frac{\sum \hat{y}_i^2}{\sum y_i^2} = \frac{\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i} + \ldots + \hat{\beta}_k \sum y_i x_{ki}}{\sum y_i^2 - n\bar{Y}^2}$$ sonucuna ulaşılır.

Bağımlı değişken üzerinde etkili olsun veya olmasın modele giren her değişken belirginlik katsayısı $R^2$’nin değerini artıracaktır. Şöyle ki $R^2$ eşitliğindeki toplam değişme, ($\sum y_i^2$) bağımsız değişkenlerden bağımsızdır ve değeri sabittir. MODELE GIREN HER DEĞİŞKEN

regresyonla açıklanan değişmeyi ($\sum \hat{y}_i^2$) artıracaktır. Paydann ($\sum y_i^2$) değeri sabit iken payının ($\sum \hat{y}_i^2$) değerinin artması belirginlik katsayısının yanılıtıcı şekilde artmasına neden olur. Bu mahsuru ortadan kaldırmak için çok değişkenli regresyon modellerinde genellikle $R^2$ yerine düzeltilmiş $R^2$ hesaplanır. Düzeltilmiş belirginlik katsayısı $\bar{R}^2$ ile gösterilir ve

$$\bar{R}^2 = 1 - (1 - R^2) \frac{n-1}{n-k}$$ ye eşittir. Düzeltilmiş $R^2(\bar{R}^2)$; belirginlik katsayısının serbestlik derecesiyle yeniden düzenlenmiş hâlidir. Gözlem sayısı yeterince büyükse $\bar{R}^2$ ile $R^2$ birbirine yakındır. Ancak gözlem sayısı Küçükse düzeltilmiş belirginlik katsayısı, belirginlik katsayısından daha küçüktür hatta belirginlik katsayısı negatif değer almak zaten (orijinden geçen regresyon hariç) $\bar{R}^2$ negatif değer alabilmektedir.
Belirsizlik katsayısı yine 1-R²'ye eşittir.

Basit regresyon modelinde belirginlik katsayısının (r²) karekökü basit korelasyon katsayısı (r) Y ile X arasındaki ilişkinin derecesini gösterirken çok değişkenli regresyon modelinde belirginlik katsayısının (R²) karekökü çoklu korelasyon katsayısına eşittir. Çoklu korelasyon katsayısı da bağımlı değişken ile bağımsız değişkenler arasındaki ilişkinin derecesini gösterir. Ancak kiymet korelasyon katsaylarının işareti farklı olduğu için belirginlik katsayısından hesaplanan çoklu korelasyon katsayısına artı veya eksi işaret verilemez. Bundan dolayı, uygulamada genellikle belirginlik katsayısı hesaplanır ve yorumlanır.

### 10.5. Varyans Analiz Tablosu Yaklaşımı: F Testi

Varyans analizi tablosunun oluşturulmasındaki amaç, regresyonla açıklanabilen değişimnin anlamlılığını test etmektir.

\[
\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i}
\]

Yukarıdaki regresyonun genel anlamlılığını sınavması;

\[H_0 : \beta_2 = \beta_3 = 0\] temel hipotezinin \[H_1 : \beta\] ların hepsi birden sıfır değildir

alternatif hipotezine karşı sınavmasıdır. Temel hipotez kabul edilirse bağımlı değişkenle bağımsız değişkenler arasında doğrusal ilişki yoktur.

Ayrıca varyans analizi yaklaşımı, özellikle çok değişkenli regresyon modellerinde, aynı verilere uygulanan alternatif modellerin seçimi açısından yol gösterici özelliği ile ayrı bir öneme sahiptir. Ayrıca belirginlik katsayısı (R²), anca kütle hata teriminin varyansı (\(\hat{\sigma}_i^2\)) gibi ölçülerin varyans analizi tablosundan hesaplanması mümkündür. Varyans analiz tablosu, ANOVA tablosu olarak da adlandırılır. Varyans analiz tablosunun hazırlananabilmesi için aşağıdaki kareler toplamlarının hesaplanmış olması gerekir:

\[
\sum \hat{y}_i^2 = \sum \hat{y}_i^2 + \sum \hat{u}_i^2
\]

\[
\sum (y_i - \bar{y})^2 = \sum (\hat{y}_i - \bar{y})^2 + \sum (y_i - \hat{y}_i)^2
\]

\[
TSS = RSS + ESS
\]

\[
TSS = \text{Toplam değişim}
\]

\[
RSS = \text{Regresyonla açıklanan değişim}
\]

\[
ESS = \text{Regresyonla açıklanamayan değişime}
\]
Daha sonraki aşamada hesaplanan kareler toplamlarının serbestlik dereceleri bulunur. \(\sum \hat{y}_i^2\) için serbestlik derecesi \(k-1\)'dir. Burada \(k\), sabit parametre de dâhil olmak üzere toplam parametrelerin sayısıdır. \(\sum \hat{u}_i^2\) için serbestlik derecesi \(n-k\)'ye eşittir ve \(n\) gözlem sayısını ifade eder. Nihayet kareler toplami \(\left(\sum y_i^2\right)\) için serbestlik derecesi \((k-1)+(n-k)\)'den \(n-1\)’e eşittir.

Varyans Analiz Tablosu (ANOVA)

<table>
<thead>
<tr>
<th>Değişimin Kaynağı</th>
<th>Kareler Toplamı</th>
<th>Serbestlik Dereceleri</th>
<th>Ortalama Hata Karesi</th>
</tr>
</thead>
<tbody>
<tr>
<td>RSS (X_2, X_3)</td>
<td>(\sum \hat{y}_i^2)</td>
<td>(k-1)</td>
<td>(\sum \hat{y}_i^2 / (k-1))</td>
</tr>
<tr>
<td>ESS</td>
<td>(\sum \hat{u}_i^2)</td>
<td>(n-k)</td>
<td>(\hat{\sigma}_u^2 = \sum \hat{u}_i^2 / n-k)</td>
</tr>
<tr>
<td>TSS</td>
<td>(\sum y_i^2)</td>
<td>(n-1)</td>
<td></td>
</tr>
</tbody>
</table>

Varyans analiz tablosundaki verilerden, \(F\) istatistiği;

\[
F = \frac{\sum \hat{y}_i^2 / (k-1)}{\sum \hat{u}_i^2 / (n-k)}
\]

ile hesaplanır. Hesaplanan \(F\) istatistiği verilen bir anlamılık düzeyi için \(v_1= k-1\) ve \(v_2= n-k\) serbestlik derecelerindeki eşik değeri ile karşılaştırılır. Hesaplanan \(F\) istatistiği, tablo değerinden büyükse, temel hipotez reddedilir. Bu sonuca göre regresyonun anlamılı olduğu, \(\beta\)ların hepsinin birden \(0\)'a eşit olmadığı sonucuna varılır. Regresyonla açıklanan değişme arttıkça hesaplanan \(F\)’nin değeri de artar, böylece yüksek \(F\) değeri bağımlı değişken ile bağımsız değişkenler arasında anlamlı bir ilişkinin varlığını işaret etmektedir.

Bir regresyonun genel anlamılığına ilişkin \(F\) istatistiği, \(R^2\) cinsinden de yazılabilir. Daha önce de değinildiği üzere \(R^2\) bir istatistiktir ve test edilmesi gerekir. \(R^2\) nin testi için \(F\) sınaması yapılır.

Yukarıdaki hipotezleri şu şekilde göstermek de mümkündür:

\[
H_0 : R^2 = 0
\]
\[
H_1 : R^2 \neq 0
\]

Buna göre, \(R^2\) yi barındıran \(F\) istatistiği kolaylıkla elde edilebilir.
Öncelikle \( F = \frac{\sum \hat{y}_i^2}{\sum \hat{\hat{u}}_i^2 / n-k} \) terimi \( F = \frac{\sum \hat{y}_i^2}{\sum \hat{\hat{u}}_i^2 / k-1} \) şeklinde yazılır.

Pay ve payda toplam değiştme \( (\sum y_i^2) \) bölünenek

\[
F = \frac{\sum \hat{y}_i^2 / \sum y_i^2}{\sum \hat{\hat{u}}_i^2 / \sum y_i^2} = \frac{n-k}{k-1}
\]
elde edilir.

Daha önceki konulardan aşağıdaki eşitlikler bilinmektedir:

\[
\sum \hat{y}_i^2 = R^2 \quad \text{ve} \quad \sum \hat{\hat{u}}_i^2 = 1 - R^2
\]

Bunlar denklemde yerlerine konursa \( F \) istatistiği \( R^2 \) cinsinden aşağıdaki gibi yazılabilir.

\[
F = \frac{R^2 / (k-1)}{(1 - R^2) / (n-k)}
\]
Uygulamalar
Uygulama Soruları
Bu Bölümde Ne Öğrendik Özet
Bölüm Soruları

1. \( \hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} \) şeklindeki örnek regresyon modeli için EKK normal denklemlerin nasıl elde edileceğini gösterin.

2. \[
\begin{array}{ccc}
Y & X_2 & X_3 \\
2 & 2 & 3 \\
4 & 3 & 2 \\
9 & 4 & -2 \\
\end{array}
\]

Yukarıda yer alan verileri kullanarak çok değişkenli modeli \( Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \) şeklinde tahmin edin.

3. \( \hat{y}_i = \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} \) olduğunu gösterin.

4. \[
F = \frac{\sum \hat{\varepsilon}_i^2 / (k-1)}{\sum \hat{u}_i^2 / (n-k)}
\]

şeklinde ifade edilen F istatistiğini, determinasyon katsayısı cinsinden yazınız.


6. 5. Soruda verilen cevaba bağlı olarak basit doğrusal regresyon modelinde varyans analiz tablosu oluşturulmaya gerek var mıdır? Açıklayınız.

7. Çok değişkenli regresyon modeli için klasik doğrusal regresyon modeli varsayımlarını yeniden belirtiniz.

8. Gözlem sayısının küçük olduğu durumda düzeltilmiş belirginlik katsayısı, belirginlik katsaysından daha ........ bir değer alacaktır.


10. \( \hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} \) şeklinde bir regresyon modeli için aşağıda yer alan varyans analiz tablosunu tamamlayın.

<table>
<thead>
<tr>
<th>Değişim</th>
<th>Kareler</th>
<th>Serbestlik</th>
<th>Ortalama Hata Karesi</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X_2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X_3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kaynağı</td>
<td>Toplamı</td>
<td>Dereceleri</td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>---------</td>
<td>------------</td>
<td></td>
</tr>
<tr>
<td>RSS ((X_2, X_3))</td>
<td>185.17</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>ESS</td>
<td>25.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TSS</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
11. MODELE YENİ BİR BAĞIMSIZ DEĞİŞKENİN EKLEME GEREKLİLİĞİNİN TESTİ
Bu Bölümde Neler Öğreneceğiz?

1. MODELE YENİ BİR BAĞIMSIZ DEĞİŞKENİN EKLEME GEREKLİLİĞİNİN TESTİ
Bölüm Hakkında İlgı Oluşturan Sorular
Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri
Anahtar Kavramlar
Giriş

Bu bölümde, yeni bir bağımsız değişkenin modele alınıp alınmamasının gerekli olup olmadığı tespiti için yapılması gereken testler işlenecektir.
11. MODELE YENİ BİR BAĞIMSIZ DEĞİŞKENİN EKLEME GEREKLİLİĞİNİN TESTİ

Yeni bir bağımsız değişkenin modele alınmasının gerekli olup olmadığını tespiti için t testi kullanılmaktadır. t testi sonucunda bir bağımsız değişkenin parametresi anlamlı çıkıysorsa o değişkenin modelde yer alması gerekliyor; aksi durumda, bağımsız değişkenin modelde yer alması gereksizdir.

Modele eklenen açıklayıcı değişkenlerin anlamlılığını ve gerekliliğini araştırmanın bir başka yolu ise F testinin uygulanması ile mümkündür.

Modele yeni bağımsız değişkenler ekledikçe \( R^2 \) değeriin yükselmesi ve dolayısıyla regresyonun gözlemelere uyumunun daha da iyileşmesi beklenir. Eklenen yeni bir değişkenin modele anlamlı bir katkısının olup olmadığını, yani, uyumda meydana gelen iyileşmenin istatistik bakımdan anlamlı olup olmadığı F testi ile araştırılır. Burada kullanılan F testi, regresyonun genel anlamlılığını değerlendiriren F testinden farklıdır.

Zira;

\[
Y = \beta_0 + \beta_1 X_1 \rightarrow F \text{ testi anlamlı ise} \\
Y = a_0 + a_1 X_1 + a_2 X_2 ' \text{de anlamlı bulunmaktadır.}
\]

Aşamalar,

1. Hipotezlerin Kurulması

\( H_0 \): Yeni değişken modele alınmamalıdır.

\( H_1 \): Yeni değişken modele alınmalıdır.

2. Hipotezlerin Kurulması

Seçilen anlamlılık seviyesine göre F tablo değeri bulunur.

Yeni eklenen değişken sayısı: \( v_1 \)

Serbestlik derecesi: \( v_2 = n-k \) (k: Yeni modeldeki parametre sayısı)
3. Test istatistiği (Kritik Oran) Hesabı

\[ F_{\text{hesap}} = \frac{\left( \sum \hat{y}_{\text{yeni}}^2 - \sum \hat{y}_{\text{ESKI}}^2 \right)}{\sum \hat{u}_{\text{yeni}}^2 / v_1} \]

veya

\[ F_{\text{hesap}} = \frac{\left( R_y^2 - R_e^2 \right)}{\left( 1 - R_y^2 \right) / v_2} \]

4.

\[ F_{\text{hesap}} > F_{\text{tablo}} \rightarrow \text{Gerekli} \]

\[ H_0 \text{ reddedilir} \quad H_1 \text{ Kabul edilir.} \]

\[ F_{\text{hesap}} < F_{\text{tablo}} \rightarrow \text{Gereksiz} \]

\[ H_0 \text{ Kabul edilir} \quad H_1 \text{ reddedilir.} \]

Örnek:

\[ \hat{Y} = -0.0970 + 0.2036X_1 \quad R^2 = 0.99945 \quad n = 10 \]

\( (0.190) \quad (0.0014) \)

\[ \hat{Y} = -0.0139 + 0.2065X_1 - 0.0003X_2 \quad R^2 = 0.99946 \]

\( (0.280) \quad (0.0071) \quad (0.0007) \)

1. \( H_0 \): \( X_2 \) değişkeni modele alınmamalıdır.

\( H_1 \): \( X_2 \) değişken modele alınmalıdır.

2. \( v_1: 1 \quad v_2: n-k = 10 - 3 = 7 \quad F_{\text{tab}} = 5.59 \)

3. Test İstatistiği hesabı;
\[
F_{\text{hesap}} = \left( \frac{R_y^2 - R_e^2}{1 - R_y^2} \right) / v_1 \\
\]
\[
F_{\text{hesap}} = \left( \frac{0.99946 - 0.99945}{1 - 0.99946} \right) / 7 \\
F_{\text{hesap}} = 0.1296
\]

4.

\[ F_{\text{hesap}} < F_{\text{tablo}} \]

\[ 0.1296 < 5.59 \]

\( H_0 \) kabul edilir. \( X_2 \) değişkeninin modele eklenmesi modele önemli bir katkı sağlamamıştır. \( R^2 \)'de anlamlı bir artış olmamıştır.

F testi, bir değişkenin modele eklenip eklenmemesi kararını vermeke kullanılan biçimsel bir yöntem sağlar. Aynı bağımlı değişkeni içeren, ancak farklı açıklayıcı değişkeni olan rakip modeller için ne yapılmalıdır? Düzeltilmiş \( R^2 \)‘si büyük olan tercih edilmelidir.

Bir değişken \( \bar{R}^2 \)‘yi artırıyorsa KKT’yi azaltmaya da modelde kalır.

\( \bar{R}^2 \) ne zaman yükselir?

Eğer yeni eklenen değişkenin \( t \) değeri mutlak olarak 1’den büyükse modelde kalır.

\[
Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_m X_m \quad k = m + 1
\]

<table>
<thead>
<tr>
<th>Değişim Kaynağı</th>
<th>Kalıntılar Toplamı</th>
<th>Serbestlik Derecesi</th>
<th>Ortalama Kareler Toplamı</th>
</tr>
</thead>
<tbody>
<tr>
<td>( X_1, X_2, \ldots, X_m )‘deki değişim</td>
<td>( \sum \hat{y}_{\text{eski}}^2 )</td>
<td>Parametre Sayısı-( l = m-1 )</td>
<td>( \sum \hat{y}_{\text{eski}}^2 / m )</td>
</tr>
<tr>
<td>( X_1, X_2, \ldots, X_m, X_{m+1}, \ldots, X_k ) Bütün X’lerdeki değişim</td>
<td>( \sum \hat{y}_{\text{yeni}}^2 )</td>
<td>( k )</td>
<td>( \sum \hat{y}_{\text{yeni}}^2 / k )</td>
</tr>
<tr>
<td>Değişkenlerden ek değişim</td>
<td>Հ yeni - Հ eski</td>
<td>k-m</td>
<td>(Հ yeni - Հ eski) / k-m</td>
</tr>
<tr>
<td>--------------------------</td>
<td>-----------------</td>
<td>-----</td>
<td>------------------------</td>
</tr>
<tr>
<td>Y=f(X₁,..,Xₖ) olan kalıntı değişim</td>
<td>Σ uy 2</td>
<td>n-k+1</td>
<td>Σ uy yeni / n -(k+1)</td>
</tr>
<tr>
<td>Σ uy yeni</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Toplam Değişim</td>
<td>Σ y 2</td>
<td>n-1</td>
<td></td>
</tr>
</tbody>
</table>

\[ F_{hesap} = \frac{(\sum \hat{y}_{yeni}^2 - \sum \hat{y}_{eski}^2)}{\sum \hat{u}_{yeni}^2 / v_2} \]

Örnek

n = 30

\[ Y = \alpha_0 + \alpha_1 X_1 \Rightarrow Y = f (X_1) \]
\[ R^2 = 0.864 \]

\[ \sum \hat{y}_{eski}^2 = 845 \]

\[ \sum \hat{u}_{eski}^2 = 133 \]

\[ Y = f (X_1, X_2) \]
\[ R^2 = 0.892 \]

\[ \sum \hat{y}_{yeni}^2 = 873 \]

\[ \sum \hat{u}_{yeni}^2 = 105 \]
<table>
<thead>
<tr>
<th>Değişimin Kaynağı</th>
<th>Kalıntılar Toplamı</th>
<th>Serbestik Derecesi</th>
<th>Ortalama Kareler Toplamı</th>
</tr>
</thead>
<tbody>
<tr>
<td>( X_1 )</td>
<td>( \sum \hat{\gamma}_{eski}^2 = 845 )</td>
<td>m-1=2-1=1</td>
<td>( \sum \hat{\gamma}_{eski}^2 / m )</td>
</tr>
<tr>
<td>( X_1, X_2 )</td>
<td>( \sum \hat{\gamma}_{yeni}^2 = 873 )</td>
<td>k-1=3-1=2</td>
<td>( \sum \hat{\gamma}_{yeni}^2 / k )</td>
</tr>
<tr>
<td>( X_2 )den değişim</td>
<td>( \sum \hat{\gamma}<em>{yeni}^2 - \sum \hat{\gamma}</em>{eski}^2 = 28 )</td>
<td>k-m=3-2=1</td>
<td>( \frac{\sum \hat{\gamma}<em>{yeni}^2 - \sum \hat{\gamma}</em>{eski}^2}{k - m} )</td>
</tr>
<tr>
<td>( Y = f(X_1, ..., X_k) ) olan kalıntı değişim</td>
<td>( \sum \hat{\gamma}_{yeni}^2 = 105 )</td>
<td>n-k=30-3=27</td>
<td>( \frac{\sum \hat{\gamma}_{yeni}^2}{n - k} = \frac{105}{27} )</td>
</tr>
<tr>
<td>Toplam Değişim</td>
<td>( \sum \gamma^2 = 978 )</td>
<td>n-1</td>
<td></td>
</tr>
</tbody>
</table>

\( F_{0.05} = 4.21 \), \( F_{hesap} = 7.18 \)
Uygulamalar
Bu Bölümde Ne Öğrendik Özeti
Bölüm Soruları

1. Aşağıda verilmiş olan bilgilerden yola çıkarak $X_2$ değişkeninin modele dâhil edilip edilmeyeceğini F testi yaklaşımıyla belirleyiniz.

$n = 40$
$Y = \alpha_0 + \alpha_1 X_1 \Rightarrow Y = f(X_1) \quad R^2 = 0.884$
$\sum \hat{y}_{\text{eski}}^2 = 945$
$\sum \hat{u}_{\text{eski}}^2 = 233$
$Y = f(X_1, X_2) \quad R^2 = 0.898$
$\sum \hat{y}_{\text{yeni}}^2 = 973$
$\sum \hat{u}_{\text{yeni}}^2 = 105$

2. Aşağıdaki boşlukları doldurunuz


ii. $t$ testi sonucunda bir bağımsız değişkenin parametresi .......... çıkıyorsa, o değişkenin modelde yer alması .......... aksi durumda, bağımsız değişkenin modelde yer alması ..........

3. Eklenen yeni bir değişkenin modele anlamlı bir katkısının olup olmadığı , yani, uyumda meydana gelen iyileşmenin istatistiki bakımdan anlamlı olup olmadığı F testi ile araştırılır. F testinin aşamalarını yazınız.
12. ÇOK DEĞİŞKENLİ REGRESYON MODELLERİYLE İLGİLİ BİR UYGULAMA
Bu Bölümde Neler Öğreneceğiz?

12. ÇOK DEĞİŞKENLİ REGRESYON MODELLERİYLE İLGİLİ BİR UYGULAMA
Bölüm Hakkında İlgi Oluşturan Sorular
Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri
Anahtar Kavramlar
**Giriş**

Bu bölümde, şu zamana kadar işlenen tekniklerle çok değişkenli regresyon modelleriyle ilgili bir uygulama yapılacak.
12. ÇOK DEĞİŞKENLİ REGRESYON MODELLERİYLE İLGİLİ BİR UYGULAMA

<table>
<thead>
<tr>
<th>Yıllar</th>
<th>Y (İthalat, milyar TL)</th>
<th>X1 (GSMH, milyar TL)</th>
<th>X2 (İthal Mallar Fiyat İndeksi)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1986</td>
<td>1</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>1987</td>
<td>3</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>1988</td>
<td>5</td>
<td>16</td>
<td>2</td>
</tr>
<tr>
<td>1989</td>
<td>7</td>
<td>22</td>
<td>3</td>
</tr>
</tbody>
</table>

\[
\overline{Y} = 4 \quad \overline{X_1} = 15 \quad \overline{X_2} = 2
\]

\[n = 4\]

\[
\begin{array}{ccc}
Y^2 & YX_1 & YX_2 \\
1 & 10 & 2 \\
9 & 36 & 3 \\
25 & 80 & 10 \\
49 & 154 & 21 \\
\end{array}
\]

\[
\sum Y^2 = 84 \quad \sum YX_1 = 280 \quad \sum YX_2 = 36
\]
\[ X_1^2 \quad X_2^2 \]

\begin{align*}
100 & \quad 4 \\
144 & \quad 1 \\
256 & \quad 4 \\
484 & \quad 9 \\
\sum X_1^2 & = 984 \quad \sum X_2^2 = 18
\end{align*}

\[ X_1X_2 \]

\begin{align*}
20 \\
12 \\
32 \\
66 \\
\sum X_1X_2 = 130
\end{align*}
Ara Sonuçlar

\[ \sum X_1 X_2 = 130 \]
\[ \sum y x_1 = \sum YX_1 - n \bar{X}_1 \bar{Y} = 40 \]
\[ \sum y x_2 = \sum YX_2 - n \bar{X}_2 \bar{Y} = 4 \]
\[ \sum y^2 = \sum Y^2 - n \bar{Y} = 20 \]
\[ \sum x_1^2 = \sum X_1^2 - n \bar{X}_1^2 = 84 \]
\[ \sum x_2^2 = \sum X_2^2 - n \bar{X}_2^2 = 2 \]
\[ \sum x_2 x_1 = \sum X_1 X_2 - n \bar{X}_1 \bar{X}_2 = 10 \]

1. Regresyon modelini tahmin ederek yorumlayınız.

\[ \sum y x_1 = \hat{\beta}_1 \sum x_1^2 + \hat{\beta}_2 \sum x_2 x_1 \]
\[ \sum y x_2 = \hat{\beta}_1 \sum x_2 x_1 + \hat{\beta}_2 \sum x_2^2 \]

\[ 40 = \hat{\beta}_1 84 + \hat{\beta}_2 10 \]
\[ 4 = \hat{\beta}_1 10 + \hat{\beta}_2 2 \]
\[ \hat{\beta}_1 = 0.58 \]
\[ \hat{\beta}_2 = -0.94 \]
İthalat malları fiyat indeksi sabit iken GSMH’deki 1 milyar TL’lik artış, ithalatı ortalama olarak 0,58 milyar TL artıracaktır.

GSMH sabit iken ithal malları fiyat indeksindeki 1 birimlik artış, ithalatı 0,94 milyar TL azaltır.

GSMH ve ithalatindeki sifirken ortalama -2,82 milyar TL ithalat yapılır.

2) GSMH ve fiyat göre ortalama ithalat elastikiyetini hesaplayıp yorumlayınız.

\[ \beta_0 = \bar{Y} - \beta_1 \bar{X}_1 - \beta_2 \bar{X}_2 = -2.82 \]

\[ \hat{Y} = -2.82 + 0.58X_1 - 0.94X_2 \]

GSMH’deki %1’lik artış, ithalatı ortalama olarak %2,2 arttırmaktadır.

Fiyatlardaki %1’lik artış, ithalatı ortalama olarak %0,47 azaltmaktadır.

3) \( X_1=16 \) ve \( X_2=2 \) olduğu noktada, GSMH ve fiyat için nokta elastikiyetini hesaplayarak yorumlayınız.

\[ \hat{Y} = -2.82 + 0.58 \times 16 - 0.94 \times 2 = 4.58 \]

\[ GSMH \rightarrow \varepsilon = \frac{\partial Y}{\partial X_1} \cdot \frac{X_1}{\bar{Y}} = (0.58) \cdot \frac{16}{4.58} = 2.02 \]

\[ Fiyata \mbox{ göre } \rightarrow \varepsilon = \frac{\partial Y}{\partial X_2} \cdot \frac{X_2}{\bar{Y}} = (-0.94) \cdot \frac{2}{4.58} = -0.41 \]

4) Belirginlik katsayısını hesaplayarak yorumlayınız.
İthalattaki toplam değişimin yaklaşık %92’si GSMH ve ithal malları fiyat indeksi değişkenleri ile açıklanmaktadır.

\[
\bar{R}^2 < R^2
\]

5) Tahmin edilen parametrelerin varyans ve standart hatalarını hesaplayınız.

\[
\hat{\sigma}^2 = \frac{\sum \hat{u}^2}{n-k} = \frac{0.56}{4-3} = 0.56
\]
\[
\sum \hat{u}^2 = \sum y^2 - (\hat{\beta}_1 \sum x_1 y + \hat{\beta}_2 \sum x_2 y)
\]
\[
\sum \hat{u}^2 = 20 - (0.58 \times 40 - 0.94 \times 4) = 0.56
\]

\[
Var(\hat{\beta}_1) = \hat{\sigma}^2 \frac{\sum x_2^2}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}
\]

\[
Var(\hat{\beta}_1) = 0.01647
\]

\[
se(\hat{\beta}_1) = 0.128
\]
(247)

6) Parametrelerin %5 anlamlılık seviyesine göre güven aralıklarını hesaplayarak yorumlayınız.

\[
Var(\hat{\beta}_2) = \sigma^2 \frac{\sum x_1^2}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1x_2)^2}
\]

\[
Var(\hat{\beta}_2) = 0.69176
\]

\[
se(\hat{\beta}_2) = 0.832
\]

\[
Var(\hat{\beta}_0) = \sigma^2 \left[ \frac{1}{n} + \frac{\bar{X}_1^2 \sum x_1^2 + \bar{X}_2^2 \sum x_2^2 - 2 \bar{X}_1 \bar{X}_2 \sum x_1x_2}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1x_2)^2} \right]
\]

\[
Var(\hat{\beta}_0) = 150.9
\]

\[
se(\hat{\beta}_0) = 12.28
\]

\[
\hat{Y} = -2.82 + 0.58X_1 - 0.94X_2
\]

\[
se (12.28) (0.128) (0.832)
\]

\[
P(\hat{\beta}_0 \pm se(\hat{\beta}_0) \times t_{tablo}) = 1 - \alpha
\]

\[
\alpha = 0.05
\]

\[
sd = n - k = 4 - 3 = 1
\]

\[
n = 4 < 30 \quad t_{tablo} = 12.706
\]

\[
\hat{\beta}_0 \rightarrow \hat{\beta}_0 \pm se(\hat{\beta}_0) \times t_{tablo}
\]

\[
\hat{\beta}_0 \rightarrow (-2.82) \pm 12.28 \times 12.706
\]

\[
P(-158.879 \leq \beta_0 \leq 153.209) = 0.95
\]
\[ \hat{\beta}_1 \rightarrow \hat{\beta}_1 \pm se(\hat{\beta}_1) \times t_{table} \]
\[ \hat{\beta}_1 \rightarrow (0.58) \pm 0.128 \times 12.706 \]
\[ P(-1.046368 \leq \beta_1 \leq 2.206) = 0.95 \]

\[ \hat{\beta}_2 \rightarrow \hat{\beta}_2 \pm se(\hat{\beta}_2) \times t_{table} \]
\[ \hat{\beta}_2 \rightarrow (-0.94) \pm 0.832 \times 12.706 \]
\[ P(-11.511 \leq \beta_2 \leq 9.631) = 0.95 \]

Ana kütleden çekilmesi mümkün örneklerin %95’inin katsayıları -158.879 ile 153.209 sınırları arasında yer alacaktır.

Ana kütleden çekilmesi mümkün örneklerin %95’inin \( \hat{\beta}_1 \) katsayıları -1.046368 ile 2.206 sınırları arasında yer alacaktır.

Ana kütleden çekilmesi mümkün örneklerin %95’inin \( \hat{\beta}_2 \) katsayıları -11.511 ile 9.631 sınırları arasında yer alacaktır.

7) %5 anlamlılık seviyesine göre parametrelerin sıfırdan farklılığını test ediniz.
\[ H_0 : \beta_0 = 0 \]
\[ H_1 : \beta_0 \neq 0 \]

\[ t_{hesap} = \frac{\hat{\beta}_0}{se(\hat{\beta}_0)} = \frac{-2.82}{12.28} = -0.2296 \]

\[ |t_{hesap}| < |t_{tablo}| \]

\[ 0.2296 < 12.706 \]

\[ H_0 \text{ kabul, } H_1 \text{ red} \]

\[ H_0 : \beta_1 = 0 \]
\[ H_1 : \beta_1 \neq 0 \]

\[ t_{hesap} = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)} = \frac{0.58}{0.128} = 4.53125 \]

\[ |t_{hesap}| < |t_{tablo}| \]

\[ 4.53125 < 12.706 \]

\[ H_0 \text{ kabul, } H_1 \text{ red} \]
\[ H_0 : \beta_2 = 0 \]
\[ H_1 : \beta_2 \neq 0 \]

\[
t_{hesap} = \frac{\beta_2}{se(\hat{\beta}_2)} = \frac{-0.94}{0.832} = -1.129
\]

|t_{hesap}| < |t_{tablo}|

1.129 < 12.706

\[ H_0 \text{ kabul}, \ H_1 \text{ red} \]

8) %5 anlamlılık seviyesine göre parametrelerin anlamlılığını topluca test ediniz.

\[ H_0 : \beta_1 = \beta_2 = 0 \]
\[ H_1 : \beta_1 \neq \beta_2 \neq 0 \]

\[
F_{hesap} = \frac{R^2 / (k-1)}{(1-R^2) / (n-k)}
\]

\[
F_{hesap} = \frac{0.972 / (3-1)}{(1-0.972) / (4-3)}
\]

\[ F_{hesap} = 17.357 \]
\[ F_{tablo} = 200 \]

\[ F_{hesap} < F_{tablo} \]

\[ H_0 \text{ kabul}, \ H_1 \text{ red} \]

veya
9) X1'in Y'ye toplam etkisinin doğrudan ve dolaylı etkilerinin toplamına eşit olduğunu gösteriniz.

\[ F_{hesap} = \frac{\sum \hat{y}^2 / (k - 1)}{(\sum \hat{u}^2) / (n - k)} \]

\[ F_{hesap} = \frac{19.44 / (3 - 1)}{(0.56) / (4 - 3)} = 17.357 \]

\[ \hat{\beta}_{YX_1} = \hat{\beta}_1 + \hat{\beta}_2 \hat{\beta}_{21} \]

0.47619 = 0.58 + (-0.94)(0.119)

0.47619 = 0.461

\[ \hat{\beta}_{YX_1} = \frac{\sum yx_1}{\sum x_1^2} = \frac{40}{84} = 0.47619 \]

\[ \hat{\beta}_{21} = \frac{\sum x_2x_1}{\sum x_1^2} = \frac{10}{84} = 0.119 \]

\[ \hat{\beta}_{YX_2} = \hat{\beta}_2 + \hat{\beta}_1 \hat{\beta}_{21} \]
10) Varyans Analiz tablosunu düzenleyiniz.

<table>
<thead>
<tr>
<th>D.K.</th>
<th>K.T.</th>
<th>sd</th>
<th>OKT</th>
</tr>
</thead>
<tbody>
<tr>
<td>RSS</td>
<td>$\sum y^2 = \hat{\beta}_1 \sum x_1 y + \hat{\beta}_2 \sum x_2 y$</td>
<td>k-1</td>
<td>$\sum y^2 / k-1$</td>
</tr>
<tr>
<td>ESS</td>
<td>$\sum \hat{u}^2$</td>
<td>n-k</td>
<td>$\sum \hat{u}^2 / (n-k) = \hat{\sigma}^2$</td>
</tr>
<tr>
<td>TSS</td>
<td>$\sum y^2$</td>
<td>n-1</td>
<td></td>
</tr>
</tbody>
</table>

n=4, k=3

<table>
<thead>
<tr>
<th>D.K.</th>
<th>K.T.</th>
<th>sd</th>
<th>OKT</th>
</tr>
</thead>
<tbody>
<tr>
<td>RSS</td>
<td>19.44</td>
<td>2</td>
<td>19.44/2=9.72</td>
</tr>
<tr>
<td>ESS</td>
<td>0.56</td>
<td>1</td>
<td>0.56/1=0.56</td>
</tr>
<tr>
<td>TSS</td>
<td>20</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>
Uygulamalar
Uygulama Soruları
Bu Bölümde Ne Öğrendik Özeti
Bölüm Soruları

1.

<table>
<thead>
<tr>
<th>Gözlem</th>
<th>Y (Kira Tutarı)</th>
<th>X (Gelir)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>14</td>
</tr>
<tr>
<td>5</td>
<td>12</td>
<td>17</td>
</tr>
</tbody>
</table>

A. Regresyon modelini tahmin ederek yorumlayınız.

B. Ortalama gelir elastikiyetini hesaplayıp yorumlayınız.

C. \( X_1=16 \) ve \( X_2=2 \) olduğu noktada GSMH ve fiyat için nokta elastikiyetini hesaplayarak yorumlayınız.

D. Belirginlik katsayısını hesaplayarak yorumlayınız.

E. Tahmin edilen parametrelerin varyans ve standart hatalarını hesaplayınız.

F. Parametrelerin \( %5 \) anlamlılık seviyesine göre güven aralıklarını hesaplayarak yorumlayınız.

G. \( %5 \) anlamlılık seviyesine göre parametrelerin anlamlılığını topluca test ediniz.

H. \( X \)’in \( Y \)’ye toplam etkisinin doğrudan ve dolaylı etkilerinin toplamına eşit olduğunu gösteriniz.

K. Varyans Analiz tablosunu düzenleyiniz.
13. REGRESYON MODELLERİNİN YAPISAL KARARLILIKLARININ TESTİ
Bu Bölümde Neler Öğreneceğiz?

1. REGRESYON MODELLERİNİN YAPISAL KARARLILIKLARININ TESTİ

1.1. Yapısal Kararlılığın Chow Testi ile Sınanması

1.2. Yapısal Farklilaşmanın Gölge Değişken Yöntemiyle Belirlenmesi
Bölüm Hakkında İlgi Oluşturan Sorular
Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri
Anahtar Kavramlar
Giriş

13. REGRESYON MODELLERİNİN YAPISAL KARARLILIKLARININ TESTİ

13.1. Yapısal Kararlılığın Chow Testi ile Sınanması


Örnek 1
\[ \hat{Y}_{1i} = \hat{\alpha}_1 + \hat{\alpha}_2 X_i \quad u_i \sim N(0, \sigma^2_{u1}) \quad i = 1 \cdots n_1 \quad sd = n_1 - k \]

Örnek 2
\[ \hat{Y}_{2i} = \hat{\beta}_1 + \hat{\beta}_2 X_i \quad u_2 \sim N(0, \sigma^2_{u2}) \quad i = 1 \cdots n_2 \quad sd = n_2 - k \]

\( u_{1i} \) ile \( u_{2i} \) bağımsızdır. Chow testi uygulanmadan önce örnek regresyon modellerinin hata teriminin vartıyanlarını eşitliği test edilmeliidir. Örnek regresyonlarının hata terimi vartıyanlarının (\( \sigma^2_{u1} \) ve \( \sigma^2_{u2} \) eşit olmadığını sonucuna varılması Chow testi uygulanamaz.

\[ H_0 : \sigma^2_{u1} = \sigma^2_{u2} \]
\[ H_1 : \sigma^2_{u1} \neq \sigma^2_{u2} \]

\( \sigma^2_{u1} \) için \( \hat{\sigma}^2_{u1} = \frac{\sum \hat{u}_{1i}^2}{n_1 - k} \)

\( \sigma^2_{u2} \) için \( \hat{\sigma}^2_{u2} = \frac{\sum \hat{u}_{2i}^2}{n_2 - k} \)

Vartıyanların eşitliği ile ilgili hipotezler aşağıdaki gösterildiği gibidir.

\[ H_0 : \sigma^2_{u1} = \sigma^2_{u2} \]
\[ H_1 : \sigma^2_{u1} \neq \sigma^2_{u2} \]
Test istatistiği ise;

\[ F^* = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_v^2} = \frac{\sum \hat{u}_i^2/n_1 - k}{\sum \hat{u}_i^2/n_2 - k} \]
ye eşitir.

\[ v_1 = n_1 - k \text{ ve } v_2 = n_2 - k \]
serbestlik dereceleriyle F tablo ile karşılaştırılır. \( F^* < F_0 \) ise varyansların eşit olduğu savını ileri süren temel hipotez reddedilemez. Böylece Chow testi uygulanabilir.

**Chow Testinin Aşamaları:**

1. **Aşama**

Hipotezlerin kurulması aşamasıdır.

\( H_0: \) İki denklem birbirinin ayındır.

\( H_1: \) İki denklem birbirinden farklıdır.

2. **Aşama**

İki örnek bir araya getirilerek \((n_1 + n_2)\) gözlemli bir örnek oluşturulur. \(N\) gözlemli model tahmin edilir.

\[ \hat{Y}_i = \hat{\lambda}_1 + \hat{\lambda}_2 X_i \\ u \sim N(0, \sigma_u^2) \quad i = 1 \ldots n \quad sd = n - k \]

3. **Aşama**

İki örneğin açıklanamayan değişmeleri toplanarak yeni bir açıklanamayan değişme bulunur.

\[ \sum \hat{u}_i^2 + \sum \hat{u}_i^2 \]

Hesaplanan açıklanamayan değişmenin serbestlik derecesi, tahmin edilen iki modelin serbestlik derecelerinin toplamıdır.

\[ (n_1 - k) + (n_2 - k) = n_1 + n_2 - 2k \]

Ayrıca gözlemlerin tamamından tahmin edilen \( n - k \) serbestlik derecesiyle regresyonun açıklanamayan değişmesi
4. Aşama

Kalıntı değişimleri toplamından bir araya getirilmiş regresyonun kalıntıları çıkarılır.

\[ \hat{u}_i^2 - (\sum \hat{u}_{in}^2 + \sum \hat{u}_{in}^2) \]

Serbestlik derecesi;

\[ (n - k) - (n_1 + n_2 - 2k) = (n_1 + n_2 - k) - (n_1 + n_2 - 2k) = k \quad (n - k) = (n_1 + n_2 - k) \]

olmaktadır.

Buradaki k; tahmin edilen ana kütle katsayılarının sayısıdır.

5. Aşama

Test istatistiği hesaplanır.

\[ F^* = \frac{\left[ \sum \hat{u}_i^2 - (\sum \hat{u}_{in}^2 + \sum \hat{u}_{in}^2) \right]/k}{(\sum \hat{u}_{in}^2 + \sum \hat{u}_{in}^2)/(n_1 + n_2 - 2k)} \]

6. Aşama

Hesaplanan \( F^* \) oranı, v1=k ve v2= n1+ n2- k serbestlik dereceleriyile \( \alpha \) anlamılık düzeyi için kuramsal F tablo değeri ile karşılaştırılır.

\[ F^* > F_\alpha \] ise, temel hipotez reddedilerek iki denklemin anlamlı bir şekilde farklılaştırıldığı sonucuna varılır.
Örnek 1

Milli gelir (Y) ve tüketim (C) ile ilgili iki farklı döneme ait veriler aşağıdaki gibi verilmiştir (milyar TL).

<table>
<thead>
<tr>
<th>Yıllar</th>
<th>Yt</th>
<th>Ct</th>
<th>Yıllar</th>
<th>Yt</th>
<th>Ct</th>
</tr>
</thead>
<tbody>
<tr>
<td>1968</td>
<td>17.5</td>
<td>12.4</td>
<td>1978</td>
<td>22.8</td>
<td>15.4</td>
</tr>
<tr>
<td>1969</td>
<td>18.2</td>
<td>12.7</td>
<td>1979</td>
<td>23.7</td>
<td>16.1</td>
</tr>
<tr>
<td>1970</td>
<td>18.9</td>
<td>13</td>
<td>1980</td>
<td>24.9</td>
<td>16.7</td>
</tr>
<tr>
<td>1971</td>
<td>19.1</td>
<td>12.8</td>
<td>1981</td>
<td>25.8</td>
<td>17.1</td>
</tr>
<tr>
<td>1972</td>
<td>19.5</td>
<td>12.9</td>
<td>1982</td>
<td>25.9</td>
<td>17.5</td>
</tr>
<tr>
<td>1973</td>
<td>20.4</td>
<td>13.4</td>
<td>1983</td>
<td>27.1</td>
<td>18.4</td>
</tr>
<tr>
<td>1974</td>
<td>21.2</td>
<td>14</td>
<td>1984</td>
<td>28.7</td>
<td>19.1</td>
</tr>
<tr>
<td>1975</td>
<td>21.9</td>
<td>14.5</td>
<td>1985</td>
<td>29.4</td>
<td>19.4</td>
</tr>
<tr>
<td>1976</td>
<td>22.3</td>
<td>14.7</td>
<td>1986</td>
<td>30</td>
<td>19.8</td>
</tr>
<tr>
<td>1977</td>
<td>22.7</td>
<td>15</td>
<td>1987</td>
<td>30.5</td>
<td>20.2</td>
</tr>
</tbody>
</table>

Bu veriler kullanılarak aşağıdaki model tahminleri elde edilmiştir. Buna göre 2 örnek dönem denklemlerinin eşitliğini (yapısal değişme olup olmadığını) test ediniz. \( (a = 0.05) \)

\[
\begin{align*}
C &= 0.85 + 0.63Y & C &= 3.31 + 0.51Y & C &= 1.54 + 0.61Y \\
R^2 &= 0.992 & R^2_i &= 0.958 & R^2_i &= 0.994 \\
\hat{\alpha} &= 1062 & \hat{\alpha} &= 323.3 & \hat{\alpha} &= 128.5 \\
\hat{\alpha}_i &= 10 & \hat{\alpha}_i &= 10 & \hat{\alpha}_i &= 10 \\
n &= 20 & n_i &= 10 & n_i &= 10
\end{align*}
\]

1) \( H_0 \): İki denklem birbirinin aynıdır.

\( H_1 \): İki denklem birbirinden farklıdır.

2) \[
F_{\text{heapp}} = \frac{[1062 - (323.3 + 128.5)]/2}{(323.3 + 128.5)/(10 + 10 - 4)} = 10.8
\]
3) \( F_{table(2,16)} = 3.63 \)

4) \( F_{hesap} \geq F_{table} \) H₀ hipotezi ret, H₁ hipotezi kabul. Bu sonucu göre 2 denklem arasındaki farkın anlamlı olduğu, diğer bir ifade ile tüketime yapılanın 2. dönemde değiisme uğradığı söylenebilmektedir.

### 13.2. Yapısal Farklılaşmanın Gölge Değişken Yöntemiyle Belirlenmesi

Ekonometrik modellerde kullanılan değişkenler, ölçülebilir nicel değişkenler olabileceği gibi nitel değişkenler de olabilmektedir. Nicel değişkenler süreklilik arz ettikleri hâlde nitel değişkenler birden fazla farklı değer alabilen değişkenlerdir. Modelde nitel değişkenler kullanılmak gerek yaygın görünmektedir. Ayrıca zamanda, ekonometrik bir modelin kapsadığı dönemde yapısal bir farklılaşma varsa gölge değişken yöntemi kullanmak yapısal değişimin kaynağı (sabit ve / ve ya eğim) tespit etmemize yardımcı olacaktır.

\[
Y_i = b_0 + b_1 D_i + u_i
\]

şeklinde ifade edilen modelde \( Y_i \) nicel (sürekli) değişken iken \( D_i \) sadece iki değer alabilen nitel bir değişkendir. Değeri 0 ya da 1 olabilen söz konusu bu değişken aşağıdaki gibi oluşturulmaktadır.

\[
D_i = \begin{cases} 
1 & 
\text{iş} \\
0 & 
\text{aş} 
\end{cases}
\]

Örneğin; modele cinsiyeti belirten bir gölge değişken ilave edileceke kadınlar için 1, erkekler için 0 değeri verilebilmektedir. Bu durumda söz konusu değişkenin aldığı değere bağlı olarak model yapısı değiştirilmektedir.

\[
E(Y_i) = b_0 + b_1 D_i
\]

Gölge değişken yönteminde dikkat edilmiş gerekken şey; gölge değişkenler belirlenirken mevcut durumun 1 eksiği kadar sayıda gölge değişkenin modele ilave edilmesidir. Örneğin; cinsiyeti ifade eden gölge değişkeni modele ilave ederken kadın ve erkek durumlarından birinin ele alınması gerekmektedir. Benzer şekilde mevsimleri ifade etmek üzere gölge değişkenler belirlenirken de mevcut 4 (ilkbahar, yaz, sonbahar, kış) durumun bir eksiği kadar yani (4-1)=3 tane gölge değişken modele ilave edilmelidir.

Tahmin edilen modellerin sabitinde ve / veya eğiminde meydana gelebilecek değişimlere göre farklı durumlar ele alınabilmektedir.
1. Durum: Modelin eğiminin değişmediği fakat sadece sabitin değiştiği durum

\[ Y_i = b_0 + b_1X_i + b_2D_i + u_i \]

Bu modelde \( Y_i \) ve \( X_i \) nicel değişkenler, \( D_i \) ise 1 ve 0 değerlerini alan nitel bir değişkeni ifade etmektedir.

\[
E(Y_i | X_i, D_i = 1) = Y_i = (b_0 + b_2) + b_1X_i \\
E(Y_i | X_i, D_i = 0) = Y_i = b_0 + b_1X_i
\]

Yukarıdaki iki modelde eğimler aynıdır, fakat sabitler farklı değerler almaktadır. Bu durumu şeklen aşağıdaki gibi görselleştirmek mümkündür.

2. Durum: Regresyon sabitinin değişmediği fakat eğimin değiştiği durum

\[ Y_i = b_0 + b_1X_i + b_2(X_iD_i) + u_i \]

ile ifade edilen modelde \( X_i \) değişkeni ile \( D_i \) değişkeni çarpımsal olarak modelde yer almaktadır.

\[
E(Y_i | X_i, D_i = 1) = Y_i = b_0 + (b_1 + b_2)X_i \\
E(Y_i | X_i, D_i = 0) = Y_i = b_0 + b_1X_i
\]

Dikkat edileceği gibi yukarıdaki 2 modelin sabitleri aynı olması rağmen eğimleri farklıdır. Aşağıdaki şekil durumu daha ayrıntılı ifade etmektedir.

3. Durum: Modellerin hem eğimlerinin hem de sabitlerinin değiştiği durum
\( Y_i = b_0 + b_1X_i + b_2D_i + b_3(X_iD_i) + u_i \) ile ifade edilen model gölge değişkenin aldığı değerlerine göre 2 farklı fonksiyonel forma dönüşecektir.

\[
E(Y_i \mid X_i, D_i = 1) = Y_i = (b_0 + b_2) + (b_1 + b_3)X_i
\]

\[
E(Y_i \mid X_i, D_i = 0) = Y_i = b_0 + b_1X_i
\]

Şekilden de anlaşılacağı gibi söz konusu modellerin hem sabitleri hem de eğimleri birbirinden farklıdır.
Uygulamalar
Uygulama Soruları
Bu Bölümde Ne Öğrendik Özeti
Bölüm Soruları

1. Yapısal değişmenin belirlenmesinde kullanılan yöntemler nelerdir?

2. Chow testi hangi durumlarda uygulanmamaktadır?

3. Chow testi sonucunda temel hipotezin reddedilmesi neyi ifade etmektedir?


\[ \hat{IH} = 0.17 + 0.30GSMH \]
\[ R^2 = 0.87 \]
\[ \hat{\alpha} \hat{\beta}^2 = 1700 \]
\[ n = 31 \]


\[ \hat{IH} = 0.41 + 0.11DK \]
\[ R_i^2 = 0.93 \]
\[ \hat{\alpha} \hat{\beta}^2 = 230.5 \]
\[ n_i = 11 \]

Elde edilen sonuçlara göre yapısal değişme olup olmadığını test ediniz. \( a = 0.05 \)

5. Nitel değişkenleri nicel değişkenlerden ayıran özellik hangisidir?

6. Gölge değişken yöntemi ile yapısal değişimın araştırılmasının Chow testine göre avantajı nedir?

7. Modele ilave edilecek golge değişkenleri oluşturulurken dikkat edilmesi gereken unsur nedir?

8. \[ Y_i = b_0 + b_1X_i + b_2(D_i) + u_i \]

ile ifade edilen modelde golge değişkenin aldığı değerlerle göre \( (D_i=0, D_i=1) \) model yapısında meydana gelen değişmenin nereden kaynaklandığını açıklayınız.

9. \[ Y_i = b_0 + b_1X_i + b_2D_i + u_i \]

ile ifade edilen modelde golge değişkenin aldığı değerlerle göre \( (D_i=0, D_i=1) \) model yapısında meydana gelen değişmeleri şekil yardımcıla açıklayınız.
14. NORMALLİK VARSAYIMI
Bu Bölümde Neler Öğreneceğiz?

14. NORMALLİK VARSAYIMI

14.1. Jarque-Bera Testi ile Normallık Varsayımı
Bölüm Hakkında İlgı Oluşturan Sorular
Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri
Anahtar Kavramlar
Giriş

Bu bölümde, regresyon analizinin temel varsayımlarından biri olan normallik varsayımının üzerinde durulacaktır. Öncelikle normallik varsayımı açıklanacak ve söz konusu varsayımı test etme yöntemlerine değinilecektir.
14. NORMALLİK VARSAYIMI

Regresyon analizinin temel varsayımlarından biri olan normallık varsayımı ana kütle hata teriminin normal dağıldığını varsayar.

Klasik doğrusal regresyon modelinde hata terimi, sıfır ortalama ve sabit varyansla normal dağılmaktadır:

\[ u_i \sim N(0, \sigma^2) \]

\[ u_i \sim IID(0, \sigma^2) \]

\[ E(u_i) = 0 \]

\[ E(u_i^2) = \sigma^2 \]

\[ E(u_i | u_j) = 0 \]

EKK yöntemi uygulanırken tahmin edilen parametrelerin doğrusal, eğilimsiz ve tesirli olabilmesi için hata terimleri ile ilgili bu varsayımlar gereklidir.

Bir ana kütleden çekilebilecek bütün örnekleri çektiğimizi farz edersek bu örneklerin her birinden \( \hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2 \) elde edilecek, böylece söz konusu parametre tahminlerinin her biri için ayrı ayrı örnekleme dağılımı olarak adlandırılan farazi birer bölünmeyi oluşturacaktır.

Bu durumda, \( \hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2 \) için farazi bölünmelerin hangi dağılıma uydukları ve ortalama ile varyanslarının ne olduklarını sorunları ortaya çıkmaktadır.

EKK yönteminin uygulanabilmesi için hata teriminin normal dağılım varsayımı gerekli değildir. Amaç nokta tahmin olsa idi EKK varsayımları yeterliydi.

Ancak amaç; örnekten \( \hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2 \) elde etmek değil, ana kütle hakkında çıkarsamalar yapmak, tahmin ettigimiz parametreler ile ilgili hipotez testlerini uygulamak, parametrelerin güven aralıklarını oluşturmaktır.

Bunun için hata teriminin dağılımnın bilinmesi zorunluluğu vardır.

EKK tahmin edicileri, hata terimlerinin doğrusal fonksiyonudur. Dolayısıyla EKK tahmin edicilerinin örneklem (olasılık) dağılımı, hata teriminin olasılık dağılımasına ilişkin varsayımı dayanır.
Hata terimi normal dağılır.

\[ u_i \sim N(0, \sigma^2) \]

\[ Y_i = \beta_0 + \beta_1 X_i + u_i \]

\[ (1) \quad (2) \]

\[ Y \text{, hata teriminin doğrusal fonksiyonudur. Dolaysıla } Y \text{, hata terimiyle aynı dağılıma sahip olacak yani normal dağılıacaktır.} \]

\[ \hat{\beta}_1 \text{ de } Y \text{'nin doğrusal bir fonksiyonu olduğu için } \hat{\beta}_1 \text{'de normal dağılıacaktır.} \]

\[ \hat{\beta}_1 = k_i Y_i \]

\[ \hat{\beta}_1 \sim N(\beta_1, \sigma_{\beta_1}^2) \]

- Normal dağılımın iki parametresi vardır.
  1) Ortalama

\[ u_i \sim N(0, \sigma^2) \]

2) Varyans

Normal dağılan iki değişkenin ortak varyansı sıfır ise değişkenlerin bağımsız olduğu anlamına gelir.

\[ E(u_i, u_j) = 0 \quad i \neq j \]

\[ u_i \sim IID(0, \sigma^2) \]

Normal bağımsız ve benzer dağılımaktadır.

Hata teriminin normal dağıldığı varsayımı, çoğu durumda gerçekleşmesi beklenen bir varsayımdır. Bunun nedeni; hata terimlerinin genellikle modele dâhil edilmemeyen çok sayıda bağımsız değişkenin etkisini temsil etmesidir.
Kısaca hata terimleri modelde yer almayan değişkenlerin bileşik etkisini yansıtır ve bu etkinin rassal olması beklenir. Çok sayıda bağımsız ve aynı şekilde (benzer) (IID) dağılım rassal değişken varsa bu değişkenlerin sayısı sonsuza yaklaştıkça bunların toplamlarının dağılımı normal dağılıma uyar.

Merkezî limit teoremine göre rastlantısal değişkenlerin sayısı artarken dağılımları ne olursa olsun toplamları normal dağılıma yaklaşacaktır.

Merkezî limit teoremine göre örnek birim sayısının artması hâlinde, Y’ler normal dağılmasa bile, parametre tahminleri asimptotik olarak normal dağılacaktır.

Normallik varsayımı altında EKK Tahmin Edicileri;

• Eğilimsiz
• En Küçük Varyanslı
• Tutarlı

\[ \hat{\beta}_0, \hat{\beta}_1 \text{ normal dağılır} \]
\[ \beta_0, \beta_1 \quad \tilde{\sigma}^2 \text{den bağımsız değildir.} \]
\[ E(Y_i) = \beta_0 + \beta_1 X_i \quad \text{Var}(Y_i) = \sigma^2 \quad Y_i \in N(\beta_0 + \beta_1 X_i, \sigma^2) \]

Kısaca bu varsayım,

Bağımlı değişken Y’yi ve onun doğrusal kombinasyonları olan

\[ \hat{\beta}_i \text{'leri de normal dağıttığından} \]
\[ \beta_i \text{'ler için hipotez testleri yapılabilirmekte,} \]
\[ güven aralıkları oluşturulabilirmektedir ve} \]
\[ \text{En Çok Benzerlik yöntemi minin uygulanmasına imkan vermektedir.} \]

14.1. Jarque-Bera Testi (Normallik Varsayımı)

En küçük kareler kalıntılarına dayanan bir testtir. Bu testte en küçük kareler artıklarının eğiklik ve basıklık ölçülerinin hesaplanması ile elde edilen test istatistiğinin ki-kare dağılımı ile karşılaştırılması sonucu normallik varsayıımının geçeri olup olmadığını inceler.
Normal dağılım simetrik bir dağılım olduğu için çarpıklık=S=0 olup basıklık=K=3’tür.

Eğiklik ve basıklık momentler yoluyla hesaplanır.

\[ S = \frac{\hat{\mu}_3}{\sigma^3} = \alpha_3 \]

\[ K = \frac{\hat{\mu}_4}{\sigma^4} = \alpha_4 \]

Momentler;

\[ \mu_r = E(X - \mu)^r \quad r = 0,1,2,3,4 \]

\[ \mu_0 = E(X - \mu)^0 = 1 \]

\[ \mu_1 = E(X - \mu)^1 = E(X - \mu) \]

\[ \mu_2 = E(X - \mu)^2 = \sigma^2 \]

\[ \mu_3 = E(X - \mu)^3 \]

\[ \mu_4 = E(X - \mu)^4 \]

Momentler X’in ölçü birimine bağlı olduklarından çarpıklık ve basıklık ölçüleri olarak ölçü birimlerinin etkisi giderilmiş (S ve K) momentler kullanılmaktadır.

E(u)=0 hata terimlerinin beklenen değeri sıfır olduğu ve aynı zamanda hata terimlerinin yerine kalıntıların ikame edilebileceği göz önüne alınacak olursa;
Ho: Kalıntılar normal dağılır.

H₁: Kalıntılar normal dağılıma uygun değildir.

\[ H_0 : S = 0 \text{ ve } K = 3 \]
\[ H_1 : S \neq 0 \text{ ve } K \neq 3 \]

- **Test İstatistiği**

\[
JB = n \left[ \frac{S^2}{6} + \frac{(K - 3)^2}{24} \right]
\]


- **Jarque-Bera Normallık Testine bir örnek:**
Kalıntıların normal dağıldığı hipotezini test ediniz.

Ho: Kalıntılar normal dağılıma uymaktadır.

H1: Kalıntılar normal dağılıma uygun değildir.

\[
\sum \hat{u}^2 = 24.649 \\
\sum \hat{u}^3 = 89.908 \\
\sum \hat{u}^4 = 414.816 \\
n = 13
\]

\[
JB = n \left[ \frac{S^2}{6} + \frac{(K - 3)^2}{24} \right]
\]

eğiklik \(\rightarrow S = \frac{\hat{\mu}_3}{\sigma^3}\)

basıklik \(\rightarrow K = \frac{\hat{\mu}_4}{\sigma^4}\)

\[
\sigma^2 = 1.896 \\
\sigma^3 = 2.611 \\
\sigma^4 = 3.595
\]

\[
\hat{\mu}_2 = \frac{\sum \hat{u}^2}{n} = \sigma^2 = \frac{24.649}{13} = 1.896
\]

\[
\hat{\mu}_3 = \frac{\sum \hat{u}^3}{n} = \frac{89.908}{13} = 6.916
\]

\[
\hat{\mu}_4 = \frac{\sum \hat{u}^4}{n} = \frac{414.816}{13} = 31.909
\]
Tam logaritmik modellerde JB testi uygulandırsa kalıntıların antilogaritmları alınır.

**Örnek 2**

\[ \hat{Y}_i = 52.35 + 0.138X_i \quad n = 14 \quad \sigma = 39.023 \quad r^2 = 0.82 \]

(37.285) (0.0187)

\[ S = 0.989512 \]

\[ K = 3.5452 \]

Kalıntıların normal dağılıdığını hipotezini test ediniz.

Ho: Kalıntılar normal dağılıma uymaktadır.

H1: Kalıntılar normal dağılıma uygun değildir.
\[ S = \frac{\hat{\mu}_3}{\sigma^3} = \alpha_3 \]

\[ K = \frac{\hat{\mu}_4}{\sigma^4} = \alpha_4 \]

\[ H_0 : S = 0 \text{ ve } K = 3 \]

\[ H_1 : S \neq 0 \text{ ve } K \neq 3 \]

\[ JB = n \left[ \frac{S^2}{6} + \frac{(K-3)^2}{24} \right] \]

\[ JB = 14 \left[ \frac{(0.989512)^2}{6} + \frac{(3.5452 - 3)^2}{24} \right] \]

\[ JB = 2.458 \]

\[ JB = 2.458 < \chi^2_2 = 5.99 \]

\[ H_0 \text{ reddedilemez.} \]

Kalınlılar normal dağılır.
Uygulamalar
Uygulama Soruları
Bu Bölümde Ne Öğrendik Özeti
Bölüm Soruları

1. Normallik varsayıımı ne için gereklidir?

2. Normal dağılımın söz konusu olması için çarpiylık ve basıklık katsayıları hangi değerleri almalıdır?

3. Normal dağılım varsayıımı altında en küçük kareler tahmin edicileri hangi özellikleri taşırlar?

4. Aşağıda sonuçları verilmiş modelde $\alpha = 0.05$ için kalıntıların normal dağılıp dağılmadığını test ediniz.

$$\hat{Y}_r = 10.520 + 0.255X_r \quad n = 20 \quad \sigma = 19.7 \quad r^2 = 0.77$$

(25.10) (0.052)

$s = 0.70$

$K = 5.21$

$$\mu_r = E(X - \mu)^r \quad r = 0, 1, 2, 3, 4$$

$$\mu_0 = E(X - \mu)^0 = 1$$

$$\mu_1 = E(X - \mu)^1 = E(X - \mu)$$

$$\mu_2 = E(X - \mu)^2 = \sigma^2$$

$$\mu_3 = E(X - \mu)^3$$

$$\mu_4 = E(X - \mu)^4$$
KAYNAKÇA


Davidson, R., MacKinnon, J.G., Econometric Theory and Methods, Oxford Universi

Davidson, R., MacKinnon, J.G., Estimation and Inference in Econometrics, Oxford


Hill, R.C., Griffiths, W.E., Lim, G.C., Principles of Econometrics, Third Edition,

Hill, R.C., Griffiths, W.E., Judge, G.G., Undergraduate Econometrics, Second Edition,

Maddala, G.S., Introduction to Econometrics, Third Edition, John Wiley and Sons Ltd,


Selahattin Güriş, Ebru Çağlayan, Burak Güriş, EViews ile Temel Ekonometri, Der
Yayınları, 2011.