**BÖLÜM 1**

**TEMEL KAVRAMLAR**

**1.1. Giriş**

Zaman serileri, bilimin her alanında uygulamaları bulunan, genellikle istatistiğin bazen de ekonometrinin önemli bir uygulama alanıdır. Bir zaman serisi, periyodik zaman aralıklarında gözlenen ölçümlerin bir dizisidir. Bir fabrikadan ihraç edilen aylık ürün miktarları, bir karayolunda meydana gelen haftalık kaza sayıları, bir göldeki saatlik su seviyesi yüksekliği, bir ülkenin yıllık ithalat ve ihracat miktarları, yıllık yatırım ve gayri safi milli hasıla gelirleri, yıllık işsizlik oranları, bir şehirdeki aylık yağış miktarları zaman serilerine örnek olarak verilebilir. Örnekler, iktisat, işletme, mühendislik ve temel bilimlerden verilebilecek uygulamalar ile genişletilebilir. Jeofizik, meteoroloji ve iktisadi verilerin incelenmesinde zaman serileri yoğun olarak kullanılmaktadır. Örneğin, geçmiş yıllara ait yıllık ihracat miktarları varsa, önümüzdeki yılın ihracat miktarı için iyi bir öngörü, bütçe hazırlıkları aşamasında önemlidir.

Bir araştırmacı faiz oranlarının enflasyonu nasıl etkilediğini öğrenmek isteyebilir. Bunun için ilk akla gelen, enflasyon ile faiz oranları arasındaki regresyon ilişkisini araştırmaktır. Böyle bir regresyon modeli başlangıçta uygun görünse de temel varsayımlar göz önüne alındığında uygun değildir. Hem faiz oranları hem de enflasyon oranları bağımlı değişkenlerdir. Oysa, regresyonda açıklayıcı değişken rolündeki değişken biliniyor (sabit) olmalı ve bağımlı değişken rolündeki değişken de bağımsız olmalıdır. Burada ise hem faiz oranları hem de enflasyon oranları bağımlı rasgele değişkenlerdir. Hemen hemen bütün istatistiki sonuç çıkarımlarda (tahmin, kestirim, öngörü gibi) olduğu gibi önce iyi bir modele ihtiyaç vardır. İyi bir model için model varsayımları kontrol edilmelidir.

Devlet, genellikle ekonomik değişkenlerle ilgili önceden kestirimleri kullanır. Başka bir deyişle, Gayri Safi Milli Hasıla (GSMH) ve bununla ilgili yatırım, istihdam, genel fiyat düzeyi gibi makro ekonomik değişkenlerin önceden kestirimi üzerinde durulur. Kısa, orta ve uzun dönemde uygulanacak ekonomik politikaların belirlenmesinde, örneğin faiz ve vergi oranlarının karşılaştırılmasında Devlet, GSMH’nın kestirim değerini temel alabilir. Bunun için de GSMH’ya etki eden parasal ve mali faktörlerin enflasyon, sermaye harcamaları, ithalat ve ihracat gibi makro değişkenlerin ileride alabileceği değerlere ihtiyaç duyulur.

Makro düzeydeki bir önceden kestirim, sadece bir değişken için yapılmaz, böyle bir önceden kestirim tek başına bir anlam ifade etmeyebilir. Makro düzeydeki önceden kestirimler, GSMH ve bunu etkileyen diğer faktörlerin birleşimi şeklinde yapılmaktadır. Önceden kestirimler, ekonomik politika değişkenlerinin ileriki dönemlerde alabileceği değerler hakkında bilgi sağlar. Bu bilgiler, Devletin ekonomiyi düzenlemek ve kontrol altına almak için düşündüğü seçeneklerden birini maksimum kılma amacı ile yapılır.

Bir perakendeci firma stok politikasını belirlerken, basit bir önceden kestirim yöntemine göre hareket eder. Geçmiş dönemlerdeki ürün satış miktarlarını göz önüne alır. Gelecek dönemlerin de geçmiş ile benzer olacağı varsayımından hareketle, ne kadar ürünün stokta bulunması gerektiğine karar verir. Böyle bir karar, önceden kestirimlerin stok kontrol politikası yapımında kullanılmasından başka bir şey değildir.

Özellikle orta ve büyük hacimli firmalar üretim, satış, bütçe, plan ve programlarını hazırlarken ekonomik koşulları göz önünde bulundururlar. Bu konuda yararlanabilecekleri kaynaklardan biri, belki de en önemlisi, devletin makro ekonomik önceden kestirimleridir. Bu ve buna benzer problemlerde zaman serisi teknikleri yoğun olarak kullanılmaktadır. Zaman serilerinde kullanılan bütün gözlemler birim zaman içinde elde edilmiş değerlerdir.

Zaman serileri, ortalamadan gösterdiği sapmalara göre, durağan ve durağan olmayan seriler olarak iki temel başlık altında incelenir. İncelenen zaman serisinin ortalaması ve varyansı simetrik bir değişme gösteriyorsa veya seri periyodik dalgalanmalardan arınmış ise bu tür seriler durağan zaman serileridir. Durağanlık zaman serilerinde önemli bir kavramdır. Bir çok istatistiki sonuç çıkarım serinin durağanlığı varsayımına dayanır. Seri durağan değilse, bazı teknikler kullanılarak (fark alma gibi) durağan hale getirildikten sonra analizler yapılır. Bir çok iktisadi veri (özellikle parasal veriler) durağan değildir. Durağan olmayan zaman serilerini de iki kısımda incelemek gerekir. Pratikte, durağanlığı bozan başlıca iki neden vardır. Birincisi, serinin ortalaması zamana bağlı olabilir. Yani, seri deterministik bir trend içerebilir. Diğeri de, serinin otokorelasyonların zamana bağlı olmasıdır. Deterministik trendi (ortalaması zamana bağlı ise) yok etmek basittir. Diğer durumda ise, serideki trend stokastik olup trendin ortadan kaldırılabilmesi için bazı teknikler (dönüşüm yapmak, fark almak gibi) kullanılır. Onun için serideki trendin deterministik mi, yoksa stokastik mi olduğu önceden kontrol edilmelidir. Bu konular ileride ayrıntılı olarak incelenmeye çalışılacaktır.

Zaman serisi modelleri genellikle regresyon modeline benzemesine rağmen, temel varsayımlarda birbirinden ayrılır. Zaman serisi modeli regresyon modeli gibi düşünüldüğünde, açıklayıcı değişken (bağımsız değişken) rolündeki değişkenler rasgeledir. Bu açıklayıcı değişkenler aynı zamanda bağımlı değişkenlerdir. Açıklayıcı değişkenler regresyon modelinde, rasgele olmayan değişken değildir. Bazı varsayımlar (durağanlık gibi) altında istatistiki sonuç çıkarımlar açısından yine regresyon teknikleri kullanılır. Aşağıda, bazı regresyon kavramları ayrıntılarına girilmeden kısaca özetlenecektir.

**1.2. Lineer (Doğrusal) Regresyon**

Bu kısımda, zaman serileri gibi bir çok alanda yoğun olarak kullanılan ve iki değişken arasındaki istatistiksel ilişkiden, yani regresyondan bahsedilecektir.  ve  gibi iki değişken arasında,  şeklinde bir ilişki,  fonksiyonu biliniyorsa deterministiktir. Örneğin,  gibi bir ilişki için  ise ,  ise  ve  ise  değerleri elde edilir.

|  |  |
| --- | --- |
| 1 | Böyle bir fonksiyonun grafiği yan tarafta verilmiştir. Diğer taraftan, iki değişken arasındaki ilişki doğrusal olmayabilir. Örneğin,  şeklinde parabolik bir ilişki de olabilir. Burada, ’in herhangi bir değeri için  nin değeri tam olarak bellidir. |

Gerçek hayatta değişkenlerden biri sabit tutularak (genellikle ) deney bir çok defa tekrarlanarak diğer değişkenin değerleri gözlenir. Aynı  değeri için  nin değerleri farklı olabilir. Yani,  değerleri belli bir hata ile gözlenir. Başka bir ifade ile,  ile  arasında,  gibi (bilinmeyen bir  fonksiyonuna bağlı) bir ilişkiden söz edilmektedir. Böyle bir ilişkiye istatistiksel (veya stokastik) bir ilişki denir.

Regresyonda önemli olan bazı koşullar altında bilinmeyen  fonksiyonunun belirlenmesidir. Bu koşullar (veya varsayımlar) genellikle  hata terimi üzerindedir.

 değişkenlerinin değerleri biliniyorsa  ler ile  arasında  gibi istatistiksel bir ilişkiyi göz önüne alalım.  bağımlı değişkeni rasgele olmak üzere,  ler açıklayıcı değişkenlerdir (reel sayılardır). Yani,  regresyon eşitliğinden bahsedilmektedir. Burada,  bağımlı değişken,  ler açıklayıcı değişkenler,  hata terimi,  de bilinmeyen parametre içeren ve ilişkinin yapısını gösteren bir fonksiyondur. Regresyonda amaç,  fonksiyonunu hataların kareleri en küçük (minimum) olacak şekilde tahmin etmektir.  ler  değişkenlerinin aldığı değerler olmak üzere,  nin  ler üzerine regresyonu,



şeklinde koşullu beklenen değerdir. Bu koşullu beklenen değer bazen lineer bazen de lineer değildir. Bu durumlara göre, regresyon denklemleri lineer ve lineer olmayan regresyon olarak iki gruba ayrılır.

**Örnek 1.2.1 *a*)**  ve  nin ortak olasılık yoğunluk fonksiyonu



şeklinde verilmiş ise  nin  üzerine regresyonu,



şeklinde lineer olmayan bir denklemdir. Burada ,  verildiğinde  nin koşullu olasılık yoğunluk fonksiyonudur.

***b*)**  ve  rasgele değişkenleri çok terimli binom dağılımına (multinomial) sahip olsun. Burada,  ve  nin ortak olasılık fonksiyonu,  için



şeklinde olup,



dir. Ayrıca,  olmak üzere,  verildiğinde  nin koşullu olasılık fonksiyonu da,



şeklindedir. Buradan  verildiğinde  nin koşullu olasılık dağılımı,



olarak elde edilir.  nin  üzerine regresyonu bu koşullu dağılımın beklenen değeri olduğundan regresyon denklemi  ve **** olmak üzere,

****

olarak bulunur.

***c*)**  ve  rasgele değişkenleri iki boyutlu normal dağılıma sahip olsun. Yani,  şeklinde olsun. Burada,

= 

olup, iki boyutlu normal dağılımın özelliklerinden koşullu dağılımlar da normaldir. Koşullu dağılım  dir. Burada, koşullu beklenen değer ve varyans

 ve 

dir. Buradan  nin  üzerine regresyonu,



şeklinde basit doğrusal regresyon denklemidir. Burada

 ve 

dir

Bilindiği gibi, regresyonda önemli olan bilinmeyen parametre içeren  fonksiyonunun tahminidir. Bu bazen doğrusal bazen de doğrusal olmayan bir fonksiyondur. Genellikle doğrusal durum ile karşılaşılır. Onun için aşağıda doğrusal regresyondan biraz ayrıntılı bahsedilecektir.

(ve ) açıklayıcı değişkenleri,  bağımlı değişkeni göstermek üzere,

 , 

şeklinde verilen eşitlik ,  ve  için  koşullarını sağlıyor ise denkleme lineer (doğrusal) regresyon denklemi denir.  için denklem basit doğrusal regresyon denklemi adını alır. İstatistiki sonuç çıkarım açısından hata terimlerinin normal dağılımlı olduğu varsayılır. Veri analizi yapılmadan önce bu varsayımların geçerliliği sınanmalı ve varsayımların bozulduğu durumlarda dönüşüm yaparak varsayımların geçerliliği sağlatılmalıdır. Amaç model parametrelerinin ( ve ) tahmin edilmesi ve eldeki verilere en uygun modelin belirlenmesidir.

Varsayımların geçerliliğini sınamak için literatürde değişik teknikler mevcuttur. Örneğin, hata terimleri kendi aralarında otokorelasyonlu olabilir (örneğin,  şeklinde bir ilişki olabilir). Yani, için  olabilir. Hata terimlerinin otokorelasyonlu olup olmadığını sınamak için Durbin-Watson testi kullanılabilir.

 ler  parametrelerinin en küçük kareler tahmin edicilerini (ileride açıklayacağız) göstermek üzere, kestirimler  ve artıklar da  şeklinde hesaplanır. Buradan, Durbin-Watson test istatistiğinin değeri (birinci dereceden otokorelasyon için),



formülü ile hesaplanır. Hata terimleri normal dağılımlı ve  için  ise  nin payındaki toplamın beklenen değeri yaklaşık  dir. Paydadaki toplamın beklenen değeri ise  dir. Dolayısı ile hata terimleri arasında bir korelasyon ilişkisi yoksa,  nin beklenen değeri yaklaşık olarak  olmalıdır.  nin alabileceği en küçük değer ise sıfırdır.  için tablolar düzenlenmiştir. Bu tablo değerleri kullanılarak hata terimlerinin otokorelasyonlu olup olmadığı araştırılabilir.

Hataların normal dağılıma sahip olduğunu sınamak için de değişik teknikler vardır. Artıkların histogramı, kutu çizitleri (box-plot) ve normal olasılık grafiği bunlardan bazılarıdır. Normal olasılık grafiği için önce regresyondan elde edilen artıklar küçükten büyüğe doğru sıralanır. Sıralanmış artıklar  olmak üzere, normal dağılım varsayımı altında,  yaklaşık olarak  dir (Neter, Wasserman ve Kutner, 1985 s.118-119). Bu değerler normal dağılım tablosundan bulunarak  değerlerinin  tablo değerlerine karşı grafiği çizilir. Bu grafikte bir doğrusallık gözleniyorsa, hata terimlerinin normal dağılıma uygun olduğu sezgisel olarak söylenebilir.

Şimdi, regresyon denklemindeki parametrelerin nasıl tahmin edileceğini görelim. Bunun için iki açıklayıcı değişkenli () regresyon modelini,



şeklinde yazalım. Burada,  ler bağımlı (bağımlılık,  lerin  açıklayıcı değişkenlere bağlılığı anlamında olup, bağımsız rasgele değişkenlerdir) rasgele değişkenler,  ve  bilinen (sabit tutulan, rasgele olmayan) açıklayıcı değişkenler,  ler beklenen değeri  varyansı  olan bağımsız rasgele değişkenler olup  ve  ler de model parametreleridir. Amaç, parametrelerin en iyi şekilde tahmin edilmesidir. Buna göre model,



olarak yazıldığında,  şeklinde lineer modele dönüşür. Buradan  nın en küçük kareler tahmin edicisi  dir. Basit doğrusal regresyonda (bir tane açıklayıcı değişkenin olması halinde) bu denklem sisteminin çözümü  ve  ve  olmak üzere,  dir.

Bu tahmin edicinin bazı özelliklerini kısaca özetleyelim. ,  boyutlu birim matrisi göstermek üzere,  ve  olduğu açıktır. Buradan aşağıdaki sonuçlar yazılabilir.

i)  gözlemlerin lineer birleşimidir. Ayrıca,  nın her bir bileşeni de  lerin lineer birleşimidir ().

ii)  nın beklenen değeri,



olduğundan nın en küçük kareler tahmin edicisi ()  için yansızdır.

iii) minimum varyanslıdır (en küçük varyanslı olması, herbir lineer birleşiminin varyansının herhangi bir lineer yansız tahmin edicinin lineer birleşiminin varyansından küçük olması anlamındadır).  nın varyansı



dir.

Bu üç özellikten,  nın en küçük varyanslı lineer yansız tahmin edici (**B**est **L**inear **U**nbiased **E**stimator, BLUE) olduğu söylenebilir.

iv) Hata terimlerin normal dağıldığını varsayalım. Yani,  lerin olasılık yoğunluk fonksiyonu,  için



şeklinde olsun. Buradan, çok değişkenli normal dağılımın özelliklerinden  da normal dağılır. Yani,  olup yine çok değişkenli normal dağılımın özelliklerinden,  nın her bir bileşeni de normaldir. Yani,  dir. Burada,  için  sayısı  matrisinin satır  sütun elemanıdır.

v) Diğer taraftan**, ** parametresinin en küçük kareler tahmin edicisi de



şeklindedir. ,  için yansız, normallik varsayımı ile en çok olabilirlik tahmin edicisidir. Burada,  açıklayıcı değişkenlerin sayısını,  de kestirimlerdir. Bu kestirimler,  şeklinde hesaplanır.

vi) Yukarıdaki özellikler kullanılarak modelin uygunluğunu (veya parametrelerin anlamlı olup olmadığını) sınamak için hipotez testleri yapılır. Örneğin,  hipotezini  lternatifine karşı test etmek için  istatistiğinin değeri tablo değeri ile karşılaştırılır. Eğer  ise,  yokluk hipotezi red edilir. Burada,  birinci tip hata olasılığını (testin anlam düzeyini),  standart hatayı,  de tablo değerini göstermektedir.

vii) Güven aralıkarı da  şeklinde hesaplanır.

viii) Regresyonda ANOVA tablosu önemlidir. ANOVA tablosu için,



değerleri ile ,  ortalama kareler toplamları hesaplandıktan sonra ANOVA tablosu aşağıdaki gibi oluşturulur.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Değişim  Kaynağı | Serbestlik Derecesi | Kareler Toplamı | Ortalama  Kareler Toplamı |  |
| Regresyon |  |  |  |  |
| Artıklar |  |  |  |  |
| Toplam |  |  |  |  |

Diğer taraftan, uygulamada çok karşılaştığımız açıklayıcı değişkenlerin bağımlı değişkeni açıklama oranı olarak da yorumlayabileceğimiz  istatistiğinin değeri,  oranı ile hesaplanır. İyi bir model için bu istatistiğin değerinin yüksek olması beklenir.

**Örnek 1.2.2.** Bir fabrikada çalışan işci sayısı () ile fabrikanın yıllık ihracat miktarları (, ton olarak)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 40 | 80 | 120 | 160 | 200 |
|  | 440 | 450 | 690 | 820 | 930 |

olarak gözlenmiştir. İşci sayısının ihjracata etkisini araştırmak isteyelim.

|  |  |
| --- | --- |
| 2 | Bu verilere basit doğrusal regresyon modelinin uygun olduğunu varsayalım ve modeli  için    şeklinde yazalım. Parametre tahminleri  ,  olarak hesaplanmıştır. |

Kestirim denklemi,  olup artıklar  den

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 396 | 531 | 666 | 801 | 936 |
|  | 44 | -81 | 24 | 19 | -6 |

olarak hesaplanmıştır. ANOVA tablosu için bazı özet bilgiler de

,  ve 

olarak hesaplanmış ve ANOVA tablosu

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Değişim  Kaynağı | Serbestlik Derecesi | Kareler Toplamı | Ortalama Kareler Toplamı |  |
| Regresyon | 1 | 182250 | 182250 | 57.73 |
| Artıklar | 3 | 9470 | 3156.7 |  |
| Toplam | 4 | 191720 |  |  |

şeklinde oluşturulmuştur. olup işci sayısının ihracat miktarını açıklama oranı yaklaşık 95% dir. Diğer taraftan,  ve  olduğu açıktır.

Hataların normal dağılıma uygunluğunu sınamak için normal olasılık grafiği kullanılan tekniklerden biridir. Bunun için artıklar küçükten büyüğe doğru sıralanır. Sıralanmış artık değerler ile -değerleri aşağıda verilmiştir (z-değerleri için normal dağılım tablosuna bakınız). Sıralanmış artık değerlerin -değerlerine karşı grafiğinden, tam olmasa bile doğrusal bir ilişki göze çarpmaktadır. Bu doğrusallık hata terimlerinin normal dağılıma uygunluğunu (sezgisel olarak) göstermektedir.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | -81 | -6 | 19 | 24 | 44 |
|  | -1.18 | -0.5 | 0 | 0.5 | 1.18 |

Normal olasılık grafiği aşağıdadır.

|  |  |
| --- | --- |
|  | Veri sayısının azlığı da dikkate alındığında hata terimlerinin normal dağılıma uygun olduğu söylenebilir. Kolmogorov–Simirnov istatistiğinin değeri kullanılarak da normallik sınaması yapılabilir. |

Varsayımlardan biri de hata terimlerinin ilişkisiz olmasıdır. Bu varsayımın geçerliliğini sınamak için Durbin-Watson istatistiği kullanılabilir. Hata terimleri arasında birinci dereceden otokorelasyonu sınamak için (yüksek dereceden otokorelasyonlara da bakılabilir) Durbin-Watson istatistiğinin değeri,



olarak hesaplanmıştır. Bu istatistik ile ilgili tablo değerleri oluşturulmuştur. Bu değer tablo değeri ile karşılaştırılarak hata terimlerinin otokorelasyonlu olup olmadığına karar verilir  nin değeri 0 ile 4 arasında olmalıdır.  ise negatif otokorelasyon vardır (Brocklebank ve Dickey, 1986, s. 8).

 parametresinin en küçük kareler tahmin değeri  olarak hesaplanmıştı. Şimdi  hipotezini  karşı test etmek isteyelim. Bunun için  in standart hatası,



olup  istatistiğinin değeri,



dir. Bu değer,  için  olup tablo değerinden küçük olduğu için  yokluk hipotezi red edilemez.

Ayrıca,  için 95% lik güven aralığı



bağıntısından yaklaşık olarak  şeklinde hesaplanmıştır.

Fabrika sahibi iş yerinde 250 işci çalıştırmak istediğinde yıllık ihracat miktarının ne olacağını merak edebilir. Yani, yıllık ihracat miktarı için 95% lik güven aralığına ihtiyaç duyabilir. Çalışan işci sayısı 250 ise kestirim değeri (denklemde  yazılarak)  dir. Buradan,  ve  olmak üzere, kestirimin standart varyansı  ve  nin standart hatası da = olarak hesaplanmış olur. Ayrıca,  için  olup  için 95% lik güven aralığı,  formülünde değerler yerine konulduğunda  olarak hesaplanır. Yani, yukarıdaki verilere göre fabrika sahibi işyerinde 250 işci çalıştırdığında 95% ihtimal ile 904 ton ile 1305 ton arasında ihracaat yapması beklenir

Regresyon analizinde, hata terimlerinin bağımsız ve sabit varyanslı olması önemli varsayımlardan biridir. Gerek regresyon analizinde gerekse zaman serilerinde hata terimlerinin varyanslarının zaman içinde değişmesi (değişen varyanslılık) çok karşılaşılan durumdur.  şeklindeki bir regresyon modeli için  varsayımı yapılmıştı. Kovaryans matrisi  yerine  şeklinde olabilir. Analizlere geçmeden  koşulu sağlanacak şekilde bir dönüşümden sonra istatistiki sonuç çıkarımlar yapılmalıdır. Bunun için  singüler olmayan bir matris olmak üzere, bu koşulu sağlamak için,

,  ve 

dönüşümleri yapılır. Buna göre yeni model,  şekline dönüşür. Dönüştürülmüş model için,



varsayımı sağlanmış olur. Bu modele göre  nın EKK tahmin edicisi



olup, bu tahmin edici  nın genelleştirilmiş EKK tahmin edicisidir. Burada hata terimleri üzerindeki varsayımlar değişmesine rağmen  tasarım matrisi üzerindeki varsayımlar hala korunmaktadır.

Basit doğrusal regresyon modeli  şeklinde verilmiş olsun. Parametrelerin EKK tahmin edicilerinin

 ve 

şeklinde olduğunu biliyoruz. Ayrıca,  ve



olduğundan parametrelerin en küçük kareler tahmin edicileri

 ve 

şeklinde de yazılabilir. Buradan  denirse,

 ve 

olup tahmin ediciler  ve  olarak da yazılabilir.  hata terimleri  ve  olacak şekilde bağımsız raasgele değişkenler olmak üzere regresyon modeli  şeklinde verildiğinde,

 ve 

dir. Buna göre,



ve



dir. Ayrıca,



olduğundan,



ve



dir. Şimdi, hata terimleri bağımsız normal dağılıma sahip olsun. Yani,  olsun.  olduğundan  da



şeklinde yazılabilir.  ler bağımsız normal dağılıma sahip rasgele değişkenler olup,  ve  de  lerin lineer birleşimi olduğundan her iki tahmin edici de normal dağılıma sahip rasgele değişkenlerdir. Yani,

 ve 

dir. Parametrelerin EKK tahmin edicileri yazıldıktan sonra kestirimler  şeklinde hesaplanır. Kestirimler hakkında istatistiki sonuç çıkarımlar için dağılımlarının bulunması gerekir.  nin normalliği,  nın  lerin lineer birleşimi olmasından açıktır. Diğer taraftan  ve





olup,  ve  dir. Buna göre,  için  güven katsayılı güven aralığı da,



şeklinde yazılır.

Şimdi,  verildiğinde *yeni* bir  değeri için güven aralığı yazmak isteyelim. Yeni , diğer  lerden bağımsızdır. Bu durumda,  için bir kestirim  şeklinde olup varyansı,



dir. Ayrıca,



olup,  olması halinde *yeni*  için  güven katsayılı güven aralığı  şeklindedir.

**Örnek 1.2.3.** Aşağıdaki verilere basit doğrusal regresyon modelinin uygun olduğunu varsayalım. Tabloda  nin  üzerine regresyonundan elde edilen kestirim ve artık değerler ile bazı özet bilgiler de bulunmaktadır. Verilere ait bazı özet bilgiler aşağıda verilmiştir.

, , ,

 ve 

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
| 2 | 15 | 12.25 | 13.8 | 1.2 | -52.5 |
| 3 | 17 | 6.25 | 15.8 | 1.2 | -42.5 |
| 5 | 19 | 0.25 | 19.8 | -0.8 | -9.5 |
| 6 | 20 | 0.25 | 21.8 | -1.8 | 10.0 |
| 4 | 18 | 2.25 | 17.8 | 0.2 | -27.0 |
| 5 | 18 | 0.25 | 19.8 | -1.8 | -9.0 |
| 6 | 22 | 0.25 | 21.8 | 0.2 | 11.0 |
| 7 | 23 | 2.25 | 23.8 | -0.8 | 34.5 |
| 8 | 26 | 6.25 | 25.8 | 0.2 | 65.0 |
| 9 | 30 | 12.25 | 27.8 | 2.2 | 105.0 |
| **55** | **208** | **42.5** | **208** | **0** | **85.0** |

Ayrıca,



ve



olup kestirim denklemi  şeklinde bulunmuştur. SAS’da

|  |  |
| --- | --- |
| data a; input x y ; cards;  2 15  3 17  ……  9 30  ;  proc reg; model y=x; run; | Açıklama: D:\DERYACA\ZAMAN SERILERİ(3)15.6.2012\Şekiller\UNTITLED-5.EPS |

kodları çalıştırılarak ANOVA tablosu aşağıdaki şekilde oluşturulmuştur. Buradan  olup  olmak üzere  için güven aralığı için  ,  ve



değerleri hesaplanmıştır. Buradan  için %95 lik güven aralığı  veya dir.

|  |
| --- |
| Sum of Mean  Source DF Squares Square F Value Pr > F  Model 1 170.00000 170.00000 87.18 <.0001  Error 8 15.60000 1.95000  Corrected Total 9 185.60000  \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*  Parameter Estimates  Parameter Standard  Variable DF Estimate Error t Value Pr > |t|  Intercept 1 9.80000 1.25815 7.79 <.0001  x 1 2.00000 0.21420 9.34 <.0001 |

 için yeni bir gözlemin kestirimi yine  dir. Bu önceden kestirim için güven aralığı için standart varyans,



olup,  için %95 lik güven aralığı  dir. Yani, aranan güven aralığı  dır.

**1.3. Zaman Serileri**

Bir zaman serisinin, periyodik zaman aralıklarında yapılmış gözlemlerin bir dizisi olduğunu söylemiştik. Yani, zaman serisi rasgele değişkenlerin bir kolleksiyonudur. Buna göre,  bir indis olmak üzere, zaman serisi  şeklinde ifade edilebilir.  indis kümesi genellikle  doğal sayılar kümesi seçilmesine rağmen,  tam sayılar kümesi de alınabilir. , reel sayılar kümesi veya  gibi sürekli aralıklar da indis kümesi olarak alınabilir.  İndis kümesi sayılamayan kümeler olarak seçildiğinde,  zaman serisine *sürekli zamanlı stokastik süreç* (continuous time stochastic process) denir.  indis kümesi  doğal sayılar kümesi veya  tamsayılar kümesi gibi kesikli (sayılabilir) kümeler ise  rasgele değişkenlerinin kolleksiyonuna bir zaman dizisi (veya zaman serisi) denir. Uygulamada zaman serisi denildiği zaman,  indis kümesi doğal sayılar kümesi anlaşılır. Aksi ifade edilmedikçe  indis kümesi doğal sayılar kümesi olacaktır.

Yukarıda söylendiği gibi, bir zaman serisi rasgele değişkenlerin bir kolleksiyonudur. O halde, zaman serisinin matematiksel tanımı için önce rasgele değişkenin ve dolayısı ile olasılık uzayının tanımına ihtiyaç vardır. Bu tanımlar, ayrıntıya girmeden aşağıda özetlenmiştir.

**Tanım 1.3.1 ** boş olmayan bir küme, **** nın bazı alt kümelerinin oluşturduğu bir sınıf da  olsun.  sınıfı aşağıdaki özellikleri sağlıyorsa  ya **** üzerinde bir *sigma cebir*,  ikilisine de *ölçülebilir bir uzay* denir.  nun her bir elemanına da bir *olay* adı verilir. Bu özellikler:

i) ****

ii) Her  için  ( nın tümleyenini göstermektedir)

iii)  için 

dir

**Tanım 1.3.2** ölçülebilir bir uzay ve  üzerinde tanımlı



 küme fonksiyonu aşağıdaki özellikleri sağlarsa  ye bir *olasılık ölçüsü* denir.  sayısına * olayının olasılığı*,  üçlüsüne de *olasılık uzayı* adı verilir. Bu özellikler:

i) 

ii) 

iii)  ler  da ayrık () olayların bir dizisi ise



dir

Bu kısa hatırlatmalardan sonra rasgele değişken ve zaman serisinin tanımını yazabiliriz.

**Tanım 1.3.3**  bir olasılık uzayı olsun.  örnek uzayından  reel sayılar kümesine tanımlı



 fonksiyonu,  özelliğini sağlıyorsa  fonksiyonuna bir *rasgele değişken* denir. Benzer şekilde



fonksiyonu her  ve her  için  koşulunu sağlıyorsa ’e *boyutlu rasgele vektör* denir

**Tanım 1.3.4**  bir olasılık uzayı,  de bir indis kümesi (genellikle doğal sayılar kümesi) olsun. Bir zaman serisi  çarpım uzayından reel sayılara giden bir fonksiyondur. Yani bir zaman serisi,



şeklinde tanımlanan bir fonksiyondur

Zaman serisi bazen , bazen de sadece  ile gösterilir. Tanıma göre zaman serisi, her sabit  için bir rasgele değişkendir. Sabit bir  için bir zaman serisi  nin reel değerli bir fonksiyonudur. Bu reel değerli fonksiyona zaman serisinin bir *realizasyonu* (veya bir *yörüngesi*) adı verilir. Bu yörünge gazetelerde, dergilerde ve kitaplarda görülen zaman serisi grafikleridir. Yani, gerçek hayatta görülen zaman serisi grafikleri aslında zaman serisinin bir yörüngesidir.

**Örnek 1.3.1** Bir zarın atılması deneyini göz önüne alalım. Bu deneyde örnek uzay,  olur (buradaki sayılar zarın üzerindeki nokta sayısına göre zarın şekillerini göstermektedir). Sigma cebir olarak kuvvet kümesini () alalım. Ayrıca  için  olarak tanımlanırsa, (**,**) üçlüsü bir olasılık uzayı olur (,  nın elemanlarının sayısıdır). Zarın üzerindeki noktaların sayısı , bir rasgele değişkendir.  in değerler kümesi  dir.  indis kümesi doğal sayılar kümesi olsun. Buna göre bir zaman serisi,



olarak tanımlanabilir.  için zaman serisi  değerlerini  olasılıkla alan bir rasgele değişkendir. Yani,  bir rasgele değişkendir.

 

Benzer şekilde  olarak sabit tutulursa, zaman serisi  gibi  nin reel değerli bir fonksiyonudur. Yani, zaman serisinin bir yörüngesidir. Bu yörüngelerden ikisi yukarıda verilmiştir

**Tanım 1.3.5** ***(Dağılım Fonksiyonu)***   bir olasılık uzayı,  de  üzerinde bir rasgele değişken olsun.  in dağılım fonksiyonu



dir

Dağılım fonksiyonu; sağdan sürekli ve azalmayan bir fonksiyon olup,  ve  özelliklerini de sağlar. İki boyutlu  rasgele vektörünün dağılım fonksiyonu



ve boyutlu  rasgele vektörünün dağılım fonksiyonu da,  için



şeklindedir. Diğer taraftan sonsuz boyutlu  rasgele vektörünün dağılım fonksiyonu, her bir sonlu  için



olarak verilir.

Buraya kadar temel olasılık kavramlarından bazıları kısaca tanıtılmaya çalışıldı. Bunlar, zaman serilerinde çok önemli bir kavram olan durağanlık özelliğini açıklayabilmek için yapıldı. Genellikle serinin durağanlığı bir varsayım olarak karşımıza çıkmaktadır. Hemen hemen bütün istatistiki sonuç çıkarımlar serinin durağanlığı varsayımına dayanır. Seri durağan değil ise, sonuç çıkarımlara geçmeden seri herhangi bir şekilde durağanlaştırılır. Pratikte, durağanlık denildiğinde zayıf durağanlık anlaşılır. Uygulamada varsayımların geçerliliğinin sınanması zor olduğundan pek kullanılmayan güçlü durağanlık (bazen kesin durağanlık, strictly stationarity) tanımı aşağıda fazla ayrıntıya girmeden verilmiştir.

**Tanım 1.3.6 (*Güçlü Durağanlık*)**  bir indis kümesi,  de bir zaman serisi olsun. Her , her , her  ve her  için



koşulu sağlanıyorsa,  zaman serisine *güçlü durağandır* denir

Daha basit bir gösterim ( aynı dağılıma sahip) ile, her  için



koşulu sağlanıyorsa  zaman serisi güçlü durağandır. Başka bir ifade ile, her  ve  için  rasgele vektörünün dağılımı ile herhangi bir öteleme ile elde edilen  rasgele vektörlerinin dağılımları aynıdır.

**Örnek 1.3.2**  güçlü durağan bir zaman serisi olsun.

***a*)**  için  zaman serisi aynı dağılımlıdır (identically distributed).  olup,  için de



olduğu kolayca görülür.  için ise



dir. Yani, güçlü durağan bir zaman serisi için, dizinin elemanlarının ileriye veya geriye doğru kaydırılması ile rasgele vektörün ortak dağılımı değişmez.

***b*)** Bağımsız aynı dağılımlı rasgele değişkenlerinin dizisi



özelliğini sağladığından güçlü durağandır. Benzer şekilde,  ler beklenen değeri , varyansı  olan bağımsız aynı dağılımlı rasgele değişkenlerin bir dizisi ise,  olarak verilen  zaman serisi de güçlü durağandır. 

***c*)** Şimdi,  ve  beklenen değeri  varyansı  olan bağımsız aynı dağılımlı rasgele değişkenler olmak üzere  zaman serisi,



şeklinde tanımlansın. Kolayca görüleceği gibi ,  ve



dir. Görüldüğü gibi bu zaman serisinin beklenen değeri, varyansı ve  ile  arasındaki kovaryansı zamana göre değişmiyor. Buna rağmen, dağılım hakkında herhangi bir bilgi olmadığından  zaman serisine güçlü durağandır diyemeyiz. Eğer,  ve  normal dağılımlı rasgele değişkenler olsaydı,  zaman serisi güçlü durağan olacaktı.

***d*)** ,  beklenen değeri  varyansı  olan bağımsız aynı dağılımlı rasgele değişkenler olmak üzere,  zaman serisi  olarak verilsin.  olmasına rağmen  olup serinin varyansı zamana bağlıdır. Dolayısı ile, serinin dağılımı da ye göre değişir. Yani,  olduğundan seri güçlü durağan değildir

Yukarıda, bir zaman serisinin güçlü durağanlık tanımı ile güçlü durağan olan ve olmayan serilere örnekler verildi. Bir zaman serisi verisinin güçlü durağanlık koşullarının sağlatılması kolay değildir. Onun için, güçlü durağanlık koşullarının zayıflatılması ile zayıf durağanlık tanımı aşağıda verilmiştir. Zayıf durağanlık için koşullar ilk iki moment üzerindedir. Onun için buna bazen ikinci dereceden durağanlık da denir. Pratikte durağanlık denildiği zaman hep zayıf durağan zaman serileri anlaşılır.

**Tanım 1.3.7 *(Durağanlık)***  zaman serisi

i)  (beklenen değer zamana göre değişmiyor),

ii)  sadece  nin bir fonksiyonudur

koşullarını sağlıyorsa  zaman serisine *zayıf durağandır* (*ikinci dereceden durağan*, *kovaryans durağan* veya sadece *durağandır*) denir

Tanımdaki (ii) koşuluna göre  kovaryansı sadece  nin bir fonksiyonudur. Bu fonksiyona serinin *otokovaryans fonksiyonu* denir ve  ile gösterilir.  serinin beklenen değeri olmak üzere otokovaryans fonksiyonu daha açık olarak



şeklinde yazılır. Bu fonksiyon, zaman serilerinin model derecelerinin belirlenmesinde ve serinin durağanlığının sezgisel olarak tespit edilmesinde çok kullanışlı özelliklere sahiptir. Bu fonksiyonun özellikleri ileride ayrıntılı olarak ele alınacaktır. Bu tanıma göre, Örnek (1.3.2c) de verilen seri güçlü durağan olmamasına rağmen zayıf durağandır.  ve  normal dağılıma sahip rasgele değişkenler ise, serinin güçlü durağan olduğunu biliyoruz. Yani, zayıf durağan bir seri normallik varsayımı ile güçlü durağandır. Diğer taraftan, bir seri güçlü durağan ise zayıf durağan olabilir gibi görünmesine rağmen güçlü durağanlık zayıf durağanlığı gerektirmez.

**Örnek 1.3.3**  bağımsız aynı Cauchy dağılımına sahip rasgele değişkenlerin dizisi olsun. Kolayca görüleceği gibi her  için  dir. O halde,  güçlü durağandır. Cauchy dağılımının momentleri yoktur. O halde, bu zaman serisi zayıf durağan olamaz.

Ayrıca,  bağımsız rasgele değişkenlerin dizisi,  ler  nin tek değerleri için beklenen değeri , varyansı  olan normal dağılıma sahip,  nin çift değerleri için de beklenen değeri  olan Üstel dağılıma sahip rasgele değişkenler olsun. Kolayca görüleceği gibi, bütün  ler için  ve



dir. Dolayısı ile, bu zaman serisi zayıf durağan olmasına rağmen  için  koşulu sağlanmadığından güçlü durağan değildir. Bununla beraber,  için,  iki boyutlu normal dağılıma sahip bir rasgele vektör,  dağılım fonksiyonu,



olan başka bir rasgele vektördür. Yani,  olup seri güçlü durağan değildir. Oysa seri zayıf durağandır

Güçlü durağanlık ile zayıf durağanlık birbirlerini gerektirmez. Ancak, bazı koşullar altında bu gerektirmeler sağlanır. Zayıf durağanlık normallik varsayımı ile güçlü durağanlığı, güçlü durağanlık da sonlu ikinci moment varsayımı ile zayıf durağanlığı gerektirir. Özet olarak;

a) Zayıf Durağanlık + Normallik  Güçlü Durağanlık

b) Güçlü Durağanlık + Zayıf Durağanlık

dır.