

İSTANBUL TEKNİK ÜNİVERSİTESİ
FEN EDEBİYAT FAKÜLTESİ
MATEMATİK MÜHENDİSLİĞİ BÖLÜMÜ



ZAMAN SERİLERİNDE ANOMALİ TESPİTİ

BİTİRME ÖDEVİ

Tarık YILDIRIMLI 090120425

Sezgin PASLIOĞLU 090120451

Tez Danışmanı:Doç. Dr. Atabey KAYGUN

Teslim Tarihi:29.05.2017

ÖNSÖZ

Bu çalışmayı hazırlamamızda bize yol gösteren, yardımını ve bilgisini hiçbir zaman esirgemeyen Sayın Hocamız Doç. Dr. Atabey KAYGUN'a sonsuz teşekkürlerimizi sunarız.

Mayıs, 2017

Tarık YILDIRIMLI

Sezgin PASLIOĞLU

İÇİNDEKİLER

	<u>Sayfa</u>
ÖZET	1
1. GİRİŞ	2
2. ZAMAN SERİLERİ	3
2.1. Trend	3
2.2. Mevsimsellik	4
2.3. Konjonktür	5
2.4. Düzensiz Bileşen	6
2.5. Zaman Serilerinde Durağanlık ve ADF Testi	6
2.6. Dekompozisyon Fonksiyonu(Decompose)	7
2.7. Otokorelasyon Fonksiyonu(ACF)	8
2.8. Kısmi Otokorelasyon Fonksiyonu(PACF)	8
2.9. Arima Modeli	9
3. İŞSİZLİK	10
4. UYGULAMA ve SONUÇLAR	11
KAYNAKLAR	31

ÖZET

Bu çalışmada Amerika Çalışma Bakanlığı işsizlik verilerinin üzerine bir veri analizi çalışması yapılmıştır. Bu uygulama ile yıllara göre işsizlik değerleri bilinen söz konusu veriler üzerinde anomali tespiti yapılmıştır.

Bu amaçla Amerika Çalışma Bakanlığı resmi sitesinden “(Unadj) Employment Level (1948 – 2017/03)” verileri alınmış, veriler düzenlenerek grafikler oluşturulmuştur. Veriler üzerinde analiz yapılmaya uygun hale getirilmiştir.

Anomali tespitini gerçekleştirebilmek için RStudio programı kullanılmıştır. Daha sonra RStudio programı yardımıyla grafikler oluşturulmuştur ve bu grafikler istenilen sonuç doğrultusunda incelenip yorumlanmıştır.

1. GİRİŞ

Günümüzde üretilen veri hacmi internet sayesinde çok büyük rakamlara ulaşmıştır. Bu yüzden veri analizi artık günümüz teknolojisinde oldukça önemli bir konumdadır. Veri analizi yararlı bilgileri keşfetmek, sonuç çıkarmak ve karar vermeyi desteklemek amacıyla verileri incelemek, temizlemek, dönüştürmek ve modellemek için kullanılan bir süreçtir. Veri analizi, finans, ekonomi, bilim ve sosyal bilim alanlarında çeşitli isimler altında çeşitli teknikleri kapsayan çok yönlü bir süreçtir. Neredeyse tüm çalışma alanlarında istatistiksel veri analizi büyük öneme sahiptir.

Bu bitirme projesi kapsamında veri madenciliği uygulaması ile Amerika Çalışma Bakanlığı kaynaklı işsizlik verilerinin değişimi üzerine grafik ve modeller oluşturulmuş ve analiz yapılmıştır.

2. ZAMAN SERİLERİ

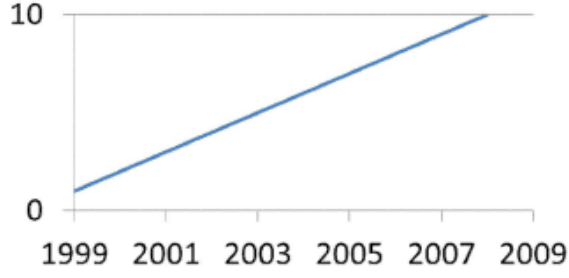
Bir verinin bir dönemden diğerine (günlük, haftalık, aylık vs.) ardışık şekilde gözlendiği sayısal değerlere "zaman serisi" denilmektedir. Her seri zaman serisi değildir, bir serinin zaman serisi olabilmesi için zamana bağlı bir durum olmalıdır. Örneğin borsa değeri bir zaman serisidir, borsa değeri hesaplanırken bir önceki günün kapanış değeri bir sonraki günün değerini etkilemektedir. Bu tarz veriler saatlik olabileceği gibi aylık, 3 aylık, yıllık da olabilir. Zaman serileri ekonomi, finans, mühendislik, sağlık, eğitim gibi çok farklı alanlarda toplanabilir. Bu farklılık sonucunda veri özellikleri değişim gösterir. Bir zaman serisi analizi için uygun olan yöntemin veya modelin seçimi, serinin oluşumunu etkileyen unsurların belirlenmesindeki başarıya bağlıdır. Bir zaman serisi 4 bileşenden oluşur;

- 1) Trend
- 2) Mevsimsellik
- 3) Konjonktür
- 4) Düzensiz Bileşen

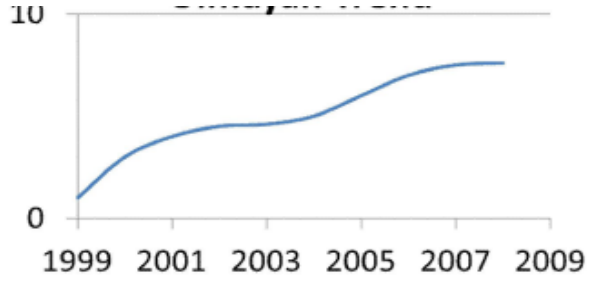
2.1 Trend

Bir zaman serisi uzun dönemde kararlı bir şekilde yükselme veya alçalma eğilimleri gösterebilir. Bu alçalma veya yükselme eğilimine trend denir. Trendin var olabilmesi için yaklaşık 15 yıllık bir döneme ihtiyaç vardır. Bir seride trend doğrusal veya doğrusal olmayan şekilde ortaya çıkabilir. Nüfusun artmasıyla birlikte işsizlik değerinin artması doğal bir sonuç olduğundan trend yardımıyla serideki reel değişimlerin ortaya çıkarılması amaçlanır. Daha sağlıklı uzun dönemli planlar için öncelikle zaman serisinin nüfus değişimi etkisinden arındırılması gerekmektedir.

Doğrusal Trend

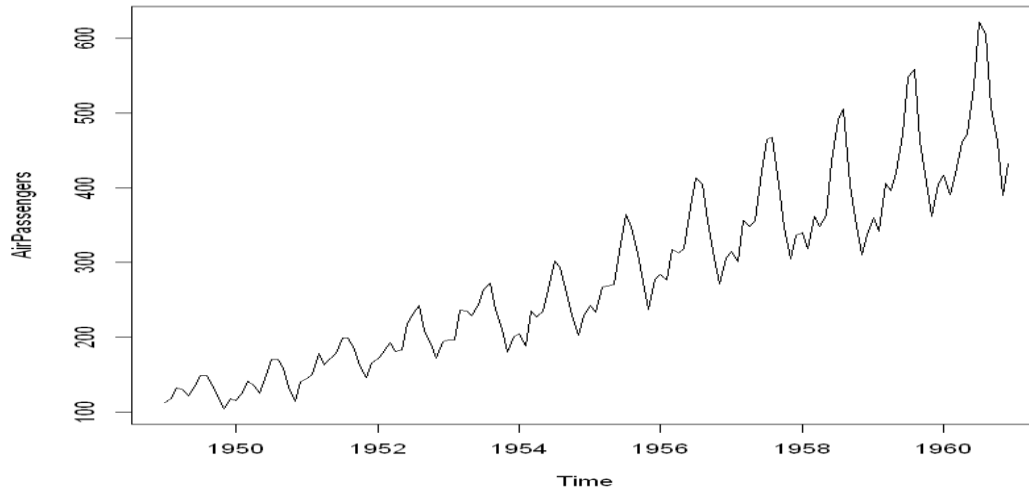


Doğrusal Olmayan Trend



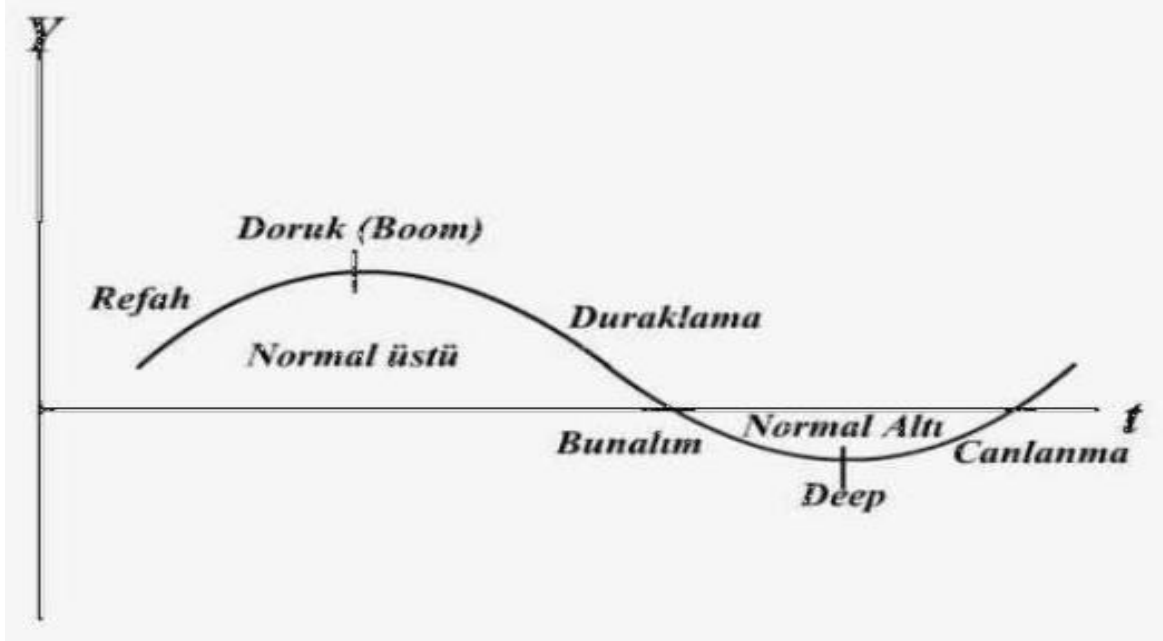
2.2 Mevsimsellik

Zaman serisi değerlerinin alçalış veya yükseliş eğiliminin belirli zaman aralıklarında periyodik olarak kendini tekrar etmesidir. Bu periyotlar günlük, aylık veya yıllık olabilir. Örnek vermek gerekirse yılın bazı dönemlerinde doğal gaz kullanımının artması veya azalması veya yaz mevsimlerinde tarlada çalışan işçilerin artması nedeniyle bu dönemlerde işsizlik oranının azalması mevsimselliktir. Mevsimselliğe örnek olarak hava yolunu kullanan yolcuların zamana bağlı değişimi aşağıdaki şekilde gösterilmiştir.



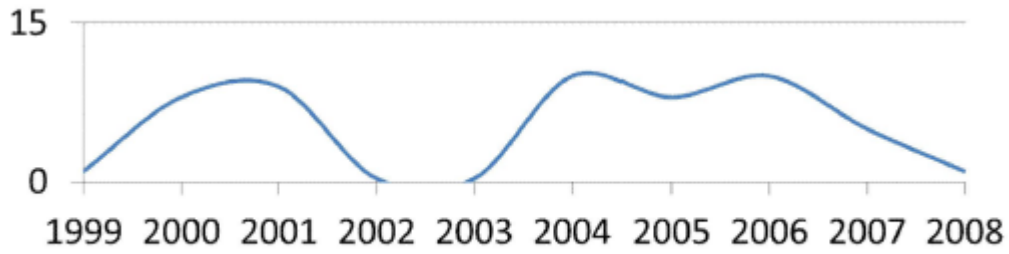
2.3 Konjonktür

Zaman serisinde meydana gelen dalgalanmalar 1 yıldan daha uzun süreci kapsayan bir şekilde devam ediyorsa bu gidişat konjonktür olarak adlandırılır. Konjonktürel değişimler genelde ekonominin veya sektörlerin refah ya da kriz dönemlerini içeren değişimlerdir. Refah dönemlerinde üretim, gelir gibi ekonomik göstergeler bir süre için artış gösterir. Mevsimsel hareketlerde dönemler düzenli ve periyodik bir salınım gösterirken, konjonktürel hareketlerde dönemler düzensiz ve periyodik olmayan bir yapıdadır. Ayrıca konjonktürel hareketlerin ortalama uzunlukları mevsimsel dalgalanmalardan daha uzundur.



2.4 Düzensiz Bileşen (Hata Terimi)

Ani nedenlerle geçici olarak ortaya çıkan nedenler düzensiz bileşenlerdir. Büyüklüğü ve yönü önceden tahmin edilemez. Düzensiz hareketlerin analiz edilebilecek bir seyirleri yoktur. 11 Eylül saldırısını örnek olarak verebiliriz. Bir modelin ele alınan seriye uygun olup olmadığının tespiti hata teriminin incelenmesi ile yapılabilmektedir. Eğer model doğru oluşturulmuşsa hata serisinin ortalamasının sıfıra yakın olması beklenir.



2.5 Zaman Serilerinde Durağanlık ve ADF Testi

Bir zaman serisi trend ve mevsimsellikten arındırılmış ise durağan seri olarak adlandırılır ve analiz yapmaya hazır hale gelmiştir. Trend ve mevsimsellikten arındırılan bir zaman serinin ortalaması zaman boyunca sabit kalır. Bu da serinin durağan olduğunu gösterir. Oluşturulan zaman serisi modelinde nedensellik testinin yapılması için öncelikle modelde kullanılan serilerin durağanlığı bilinmelidir. Bu nedenle serinin durağanlığını tespit etmek için Dickey-Fuller tarafından geliştirilen ADF Birim Kök Testi uygulanmıştır. ADF Birim Kök Testini uygulamak için oluşturulan hipotez;

$$H_0: \rho=1 \quad H_1: \rho<1$$

Hipotezdeki ρ gecikmeli değeri ifade etmektedir. Çıkan sonuç 1'e yaklaşıyorsa birim kökün var olduğunu yani modelin durağan olmadığını gösterir. Sonuç 1'den uzaklaşmışsa bu modelin durağan olduğunu gösterir. Eğer modelde durağanlık yoksa modelle elde edilen sonuçlar tutarsız çıkacaktır. Bu nedenle modelin durağan olmasını beklenir.

2.6 Dekompozisyon Fonksiyonu(Decompose)

Mevsimsel bir zaman serisi 3 bileşenden oluşur. Bunlar mevsimsel, trend ve random bileşenlerdir. Decompose fonksiyonu yardımıyla zaman serisi bu 3 bileşene ayrıştırılır. Dekompozisyon işlemini yapmadan önce serinin “linear additive”(katma) veya “multiplicative”(çarpımsal) olup olmadığına karar verilmelidir. Başarılı bir dekompozisyon elde etmek için, katma veya çarpımsal model arasında seçim yapmak önemlidir. Doğru modeli seçmek için zaman serisine bakılması gerekir.

Kullanılan katma değer modeli şöyledir:

$$Y[t] = T[t] + S[t] + e[t]$$

Kullanılan çarpımsal model şöyledir:

$$Y[t] = T[t] * S[t] * e[t]$$

Y=Gözlenen zaman serisi

T=Trend bileşeni

S=Mevsimsel bileşen(Seasonal)

e=Hata bileşeni(error)

Son olarak, hata bileşeni, orijinal zaman serilerinden trend ve mevsimsellik kaldırarak belirlenir.

2.7 Otokorelasyon Fonksiyonu (ACF)

ACF ile deęerler arasında korelasyon olup olmadıęı test edilir. Korelasyon, olasılık kuramı ve istatistikte iki rassal deęiřken arasındaki doęrusal iliřkinin ynn ve gcn belirtir. Bununla birlikte, yatay kesit verilerinde de otokorelasyona rastlanabilir. Zaman serilerinde otokorelasyon, zaman periyodunun byklę veya kklęne gre deęiřebilir. Periyot, bir aylık veriye dayanıyorsa, otokorelasyon byk,  aylık biraz daha kk ve yıllıksa daha da kktr. zet olarak ACF zaman serisini geriye kaydırarak deęerlerin nceki aylarla korelasyonu olup olmadıęını test eder. Otokorelasyonun nedenlerinden bazıları bazı aıklayıcı deęiřkenlerin modele alınmaması veya aıklanan deęiřkende lme hatası olarak sıralanabilir.

2.8 Kısmi Otokorelasyon Fonksiyonu (PACF)

Zaman serisi analizinde, kısmi otokorelasyon fonksiyonu (PACF), zaman dizilerinin kendi gecikmiř deęerleri ile kısmi korelasyonunu verir ve zaman serilerinin deęerlerini daha kısa gecikmelerde kontrol eder. Dięer gecikmeler iin kontrol etmeyen otokorelasyon fonksiyonu (ACF) ile ters orantılıdır. Kısmi otokorelasyon fonksiyonu ilk korelasyon ıkarıldıktan sonra kalanına bakıp, kalanın bir sonraki deęerler ile arasında iliřki olup olmadıęını kontrol eder.

2.9 Arima Modeli

Arima Birleştirilmiş Otoregresif Hareketli Ortalama Modeli anlamına gelir ve model 3 parametreden oluşur (p,d,q) . $ARIMA(p,d,q)$, zaman serisinin p dereceden kendisinin gecikmesi ile ilişkisini ifade eden süreci içerdigini, q ile ifade edilen ve hata terimlerinin geçmiş değerleri ile ilişkisini ifade eden ve rassal süreci yansıtmaları açısından hata terimlerinin düzeltilme sürecini ifade etmektedir. Ayrıca eğer zaman serilerinde genelde olduğu gibi bir durağan olmama durumunun olması ya da durağan bir zaman serisi ile karşılaşılması d ile ifade edilir ve zaman serisi hangi düzeyde durağan ise belirtilir. Üstel yumuşatma ve ARIMA modelleri, zaman serileri tahmininde en yaygın kullanılan iki yaklaşımdır ve soruna tamamlayıcı yaklaşımlar sağlar. Üstel yumuşatma modelleri, verideki trend ve mevsimsellik tanımına dayansa da ARIMA modelleri, verinin içindeki otokorelasyonları tanımlamayı amaçlıyor. Farklı serilerin gecikmeleri "otoregresif" olarak, tahmini verilerdeki gecikmeler "hareketli ortalama" olarak adlandırılır.

3. İŞSİZLİK

Bir ekonomide çalışmak isteyen fakat çalışacak iş bulamayan kişilerin olmasına işsizlik denir. Diğer bir ifade ile çalışmaya hazır, iş arayan kaynağa gerekli istihdamın yaratılamamasıdır. Potansiyel iş gücü ifadesi önemlidir, çünkü bir kişinin işsiz sayılabilmesi için çalışacak durumda olması ve iş arıyor olması gerekmektedir. Bir diğer önemli nokta ise bu kişilerin mevcut ücretler üzerinden çalışmayı kabul etmesi gerekliliğidir. 15-65 yaş aralığı çalışma yaşı olarak kabul edilmektedir. Örneğin ev hanımları, öğrenciler ya da hükümlüler işsiz sayılmamaktadırlar. Belirtilen yaş aralığında ve çalışan, çalışma isteği olan kişilerin toplamına bir ülkenin iş gücü ya da aktif nüfusu denilmektedir.

İşsizlik oranı yukarıdaki tanımda olduğu gibi çalışmak isteyen kişilerin toplam işgücündeki payını ifade eder:

$$\text{-İşsizlik Oranı} = (\text{İşsiz sayısı} / \text{İşgücü}) * 100$$

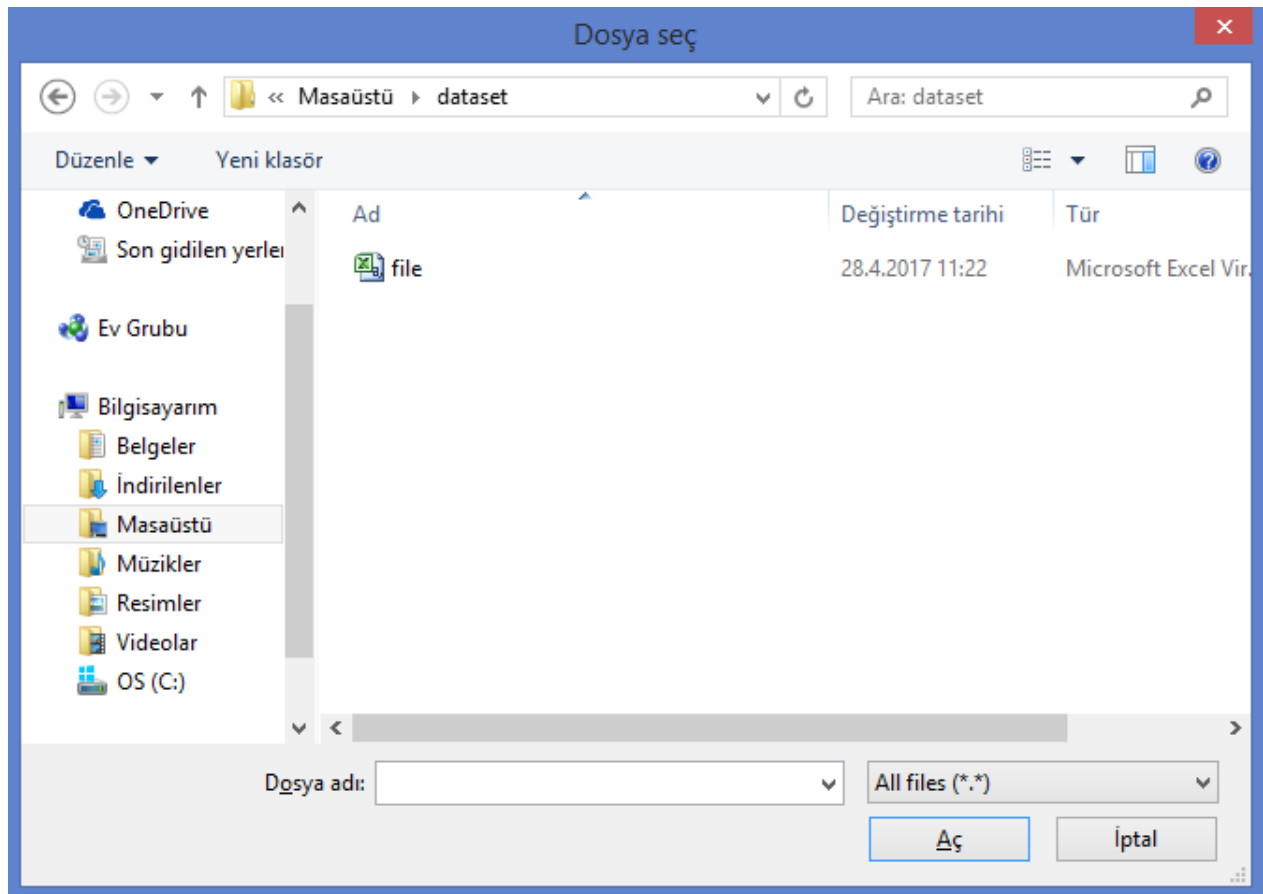
İşgücüne katılım oranı ise toplam işgücünün bir ekonomide çalışma yaşında olan nüfusa, yani çalışma çağındaki nüfusa oranını gösterir.

$$\text{-İşgücüne katılım} = (\text{İşgücü} / \text{Çalışma çağındaki nüfus}) * 100$$

4. UYGULAMA ve SONUÇLAR

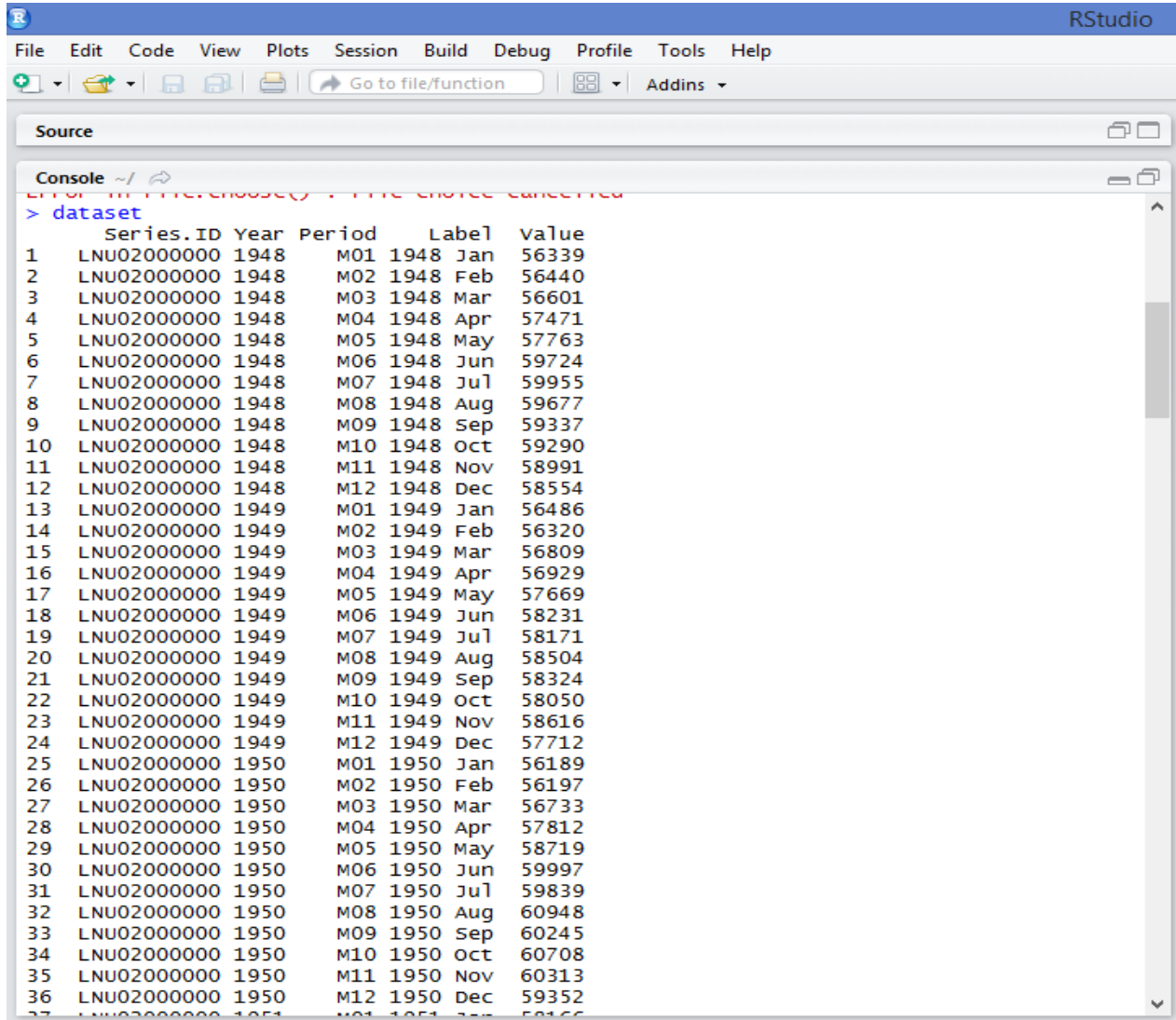
```
dataset<-read.csv(file.choose())
```

Bu kodun sonucunda gelen ekran aşağıdaki gibidir. İndirilen verinin seçilip RStudio programına aktarılması için bu komut kullanılmıştır.



dataset

Verinin RStudio programına aktarıldıktan sonraki çıktısı aşağıdaki gibidir



Source

Console

```
> dataset
```

	Series.ID	Year	Period	Label	Value
1	LNU02000000	1948	M01	1948 Jan	56339
2	LNU02000000	1948	M02	1948 Feb	56440
3	LNU02000000	1948	M03	1948 Mar	56601
4	LNU02000000	1948	M04	1948 Apr	57471
5	LNU02000000	1948	M05	1948 May	57763
6	LNU02000000	1948	M06	1948 Jun	59724
7	LNU02000000	1948	M07	1948 Jul	59955
8	LNU02000000	1948	M08	1948 Aug	59677
9	LNU02000000	1948	M09	1948 Sep	59337
10	LNU02000000	1948	M10	1948 Oct	59290
11	LNU02000000	1948	M11	1948 Nov	58991
12	LNU02000000	1948	M12	1948 Dec	58554
13	LNU02000000	1949	M01	1949 Jan	56486
14	LNU02000000	1949	M02	1949 Feb	56320
15	LNU02000000	1949	M03	1949 Mar	56809
16	LNU02000000	1949	M04	1949 Apr	56929
17	LNU02000000	1949	M05	1949 May	57669
18	LNU02000000	1949	M06	1949 Jun	58231
19	LNU02000000	1949	M07	1949 Jul	58171
20	LNU02000000	1949	M08	1949 Aug	58504
21	LNU02000000	1949	M09	1949 Sep	58324
22	LNU02000000	1949	M10	1949 Oct	58050
23	LNU02000000	1949	M11	1949 Nov	58616
24	LNU02000000	1949	M12	1949 Dec	57712
25	LNU02000000	1950	M01	1950 Jan	56189
26	LNU02000000	1950	M02	1950 Feb	56197
27	LNU02000000	1950	M03	1950 Mar	56733
28	LNU02000000	1950	M04	1950 Apr	57812
29	LNU02000000	1950	M05	1950 May	58719
30	LNU02000000	1950	M06	1950 Jun	59997
31	LNU02000000	1950	M07	1950 Jul	59839
32	LNU02000000	1950	M08	1950 Aug	60948
33	LNU02000000	1950	M09	1950 Sep	60245
34	LNU02000000	1950	M10	1950 Oct	60708
35	LNU02000000	1950	M11	1950 Nov	60313
36	LNU02000000	1950	M12	1950 Dec	59352
37	LNU02000000	1951	M01	1951 Jan	58166

class(dataset)

Verinin class kontrolü yapıldı. Aşağıdaki çıktıda görüldüğü gibi veri dafa.frame sınıfına ait. Data.frame çoklu veri vektörlerini depolamak için kullanışlıdır. Ancak veriyi grafiğe dökkebilmek için yılları gösteren kolonların satır olarak bulunması gerekmektedir.

```
> dataset<-read.csv(file.choose())  
> class(dataset)  
[1] "data.frame"
```

dataset<-ts(dataset\$Value, start=c(1948,1), freq=12)

Veri ts komutuyla time serie sınıfına çevrildi. Start komutu verinin 1948 yılının ocak ayında başlamasına göre düzenlendi. Bir yıl 12 aydan oluştuğu için frekans 12 olarak alındı.

class(dataset)

Verinin class kontrolü tekrar yapıldı. Verinin time serie("ts") sınıfına çevrildiği gözlemlendi.

```
> class(dataset)  
[1] "ts"
```


dataset

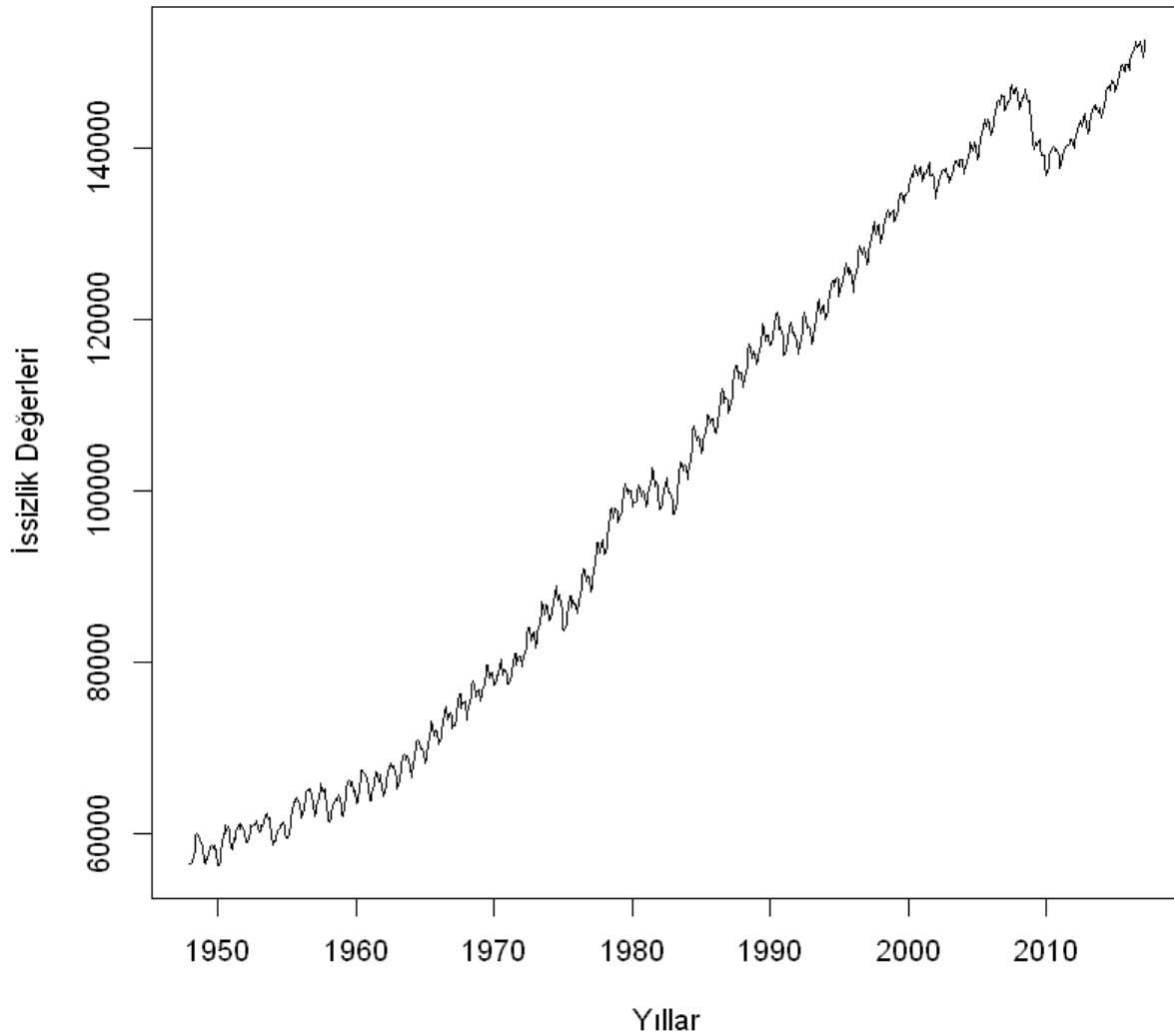
Aşağıdaki çıktıda görüldüğü üzere veri istenilen formata getirildi. Artık veri grafiğe dökülmeye hazır hale getirildi.

```
> dataset
```

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov
1948	56339	56440	56601	57471	57763	59724	59955	59677	59337	59290	58991
1949	56486	56320	56809	56929	57669	58231	58171	58504	58324	58050	58616
1950	56189	56197	56733	57812	58719	59997	59839	60948	60245	60708	60313
1951	58166	58102	59366	59206	60219	60373	60968	61128	60408	60906	60464
1952	58884	58834	58912	59232	60250	60988	60775	60872	61162	60992	61394
1953	60134	60271	60874	60757	61061	62166	62186	62271	61529	61805	61302
1954	58645	59059	59119	59537	60020	60497	60523	60858	60952	61210	60901
1955	59354	59336	59850	60861	61780	62568	63497	63876	63676	64138	63840
1956	62049	61773	62172	63002	64045	64707	64940	65085	64831	65074	64310
1957	61974	62512	63134	63512	64213	65127	65726	65009	64769	65112	64129
1958	61508	61265	61567	62116	63098	63652	63810	64018	63766	64480	63890
1959	62052	62015	63091	64241	65036	65924	66193	65897	65414	65891	64877
1960	63375	63871	63607	65450	66342	67288	67239	67004	66892	66563	66394
1961	63797	63869	64700	64957	65831	67151	66911	67028	66036	66786	66348
1962	64215	64872	65421	65957	67066	67852	67849	68096	67621	67850	67046
1963	65168	65519	66329	67240	67984	68844	69225	69052	68567	68964	68471
1964	66468	67197	67695	68947	69952	70448	70839	70676	69849	70147	69892
1965	68235	68690	69385	70220	71298	72278	73093	72695	71408	72112	71824
1966	70368	70691	71090	72066	72619	74037	74655	74665	73248	73744	73995
1967	72161	72505	72560	73445	73638	75393	76220	76170	74632	75180	75218
1968	73272	74114	74517	75143	75931	77273	77748	77431	75939	76365	76608
1969	75357	76180	76520	77077	77265	78958	79615	79646	78026	78671	78716
1970	77313	77489	77957	78408	78374	79416	80343	79965	78341	79019	78860
1971	77389	77427	77675	78404	78929	79717	80942	80897	79589	80375	80527
1972	79459	79734	80578	81027	81646	83073	83911	83993	82537	83223	83231

```
plot(dataset, type="l", ylab="İssizlık Deęerleri",  
xlab="Yıllar")
```

Bu komut ile işsizlik değerleri ve yılları gösteren grafik çizdirildi. X ve Y koordinatlarına xlab, ylab komutlarıyla istenilen isimler verildi.

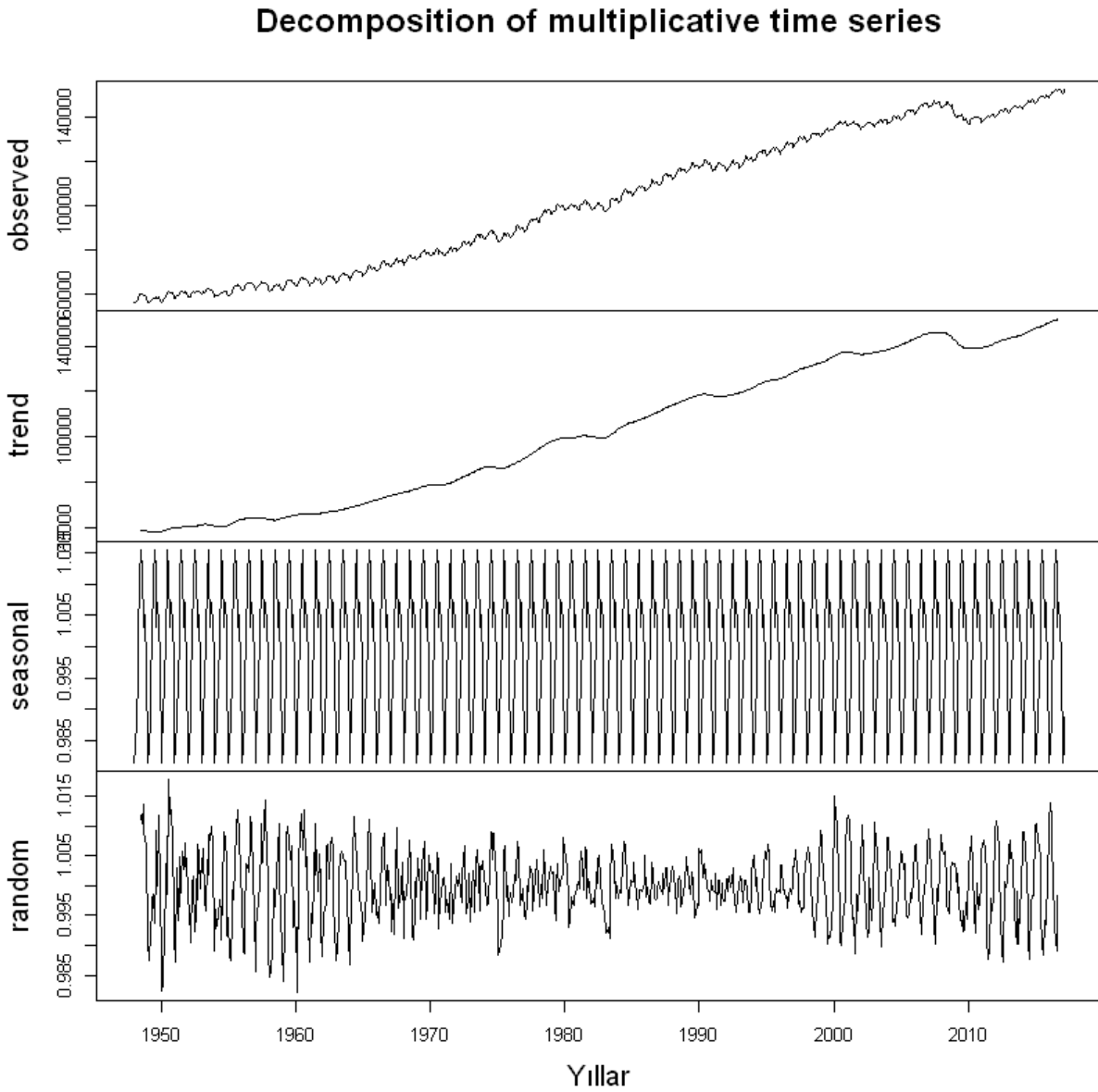


```
datasetdec<-decompose(dataset, type=c("m"))
```

Mevsimsel bir zaman serisi 3 bileşenden oluşur. Bunlar mevsimsel, trend ve random bileşenlerdir. Decompose fonksiyonu yardımıyla zaman serisi bu 3 bileşene ayrıştırılır. Dekompozisyon işlemini yapmadan önce serinin “linear additive”(katma) veya “multiplicative”(çarpımlı) olup olmadığına karar verilmelidir. Başarılı bir dekompozisyon elde etmek için, katma veya çarpımlı model arasında seçim yapmak önemlidir. Doğru modeli seçmek için zaman serisine bakılması gerekir. Üzerinde işlem yapılan seri “multiplicative” yani çarpımsal artan bir seridir. Bunun sebebi işsizlik değerlerinin her yıl bir önceki seneye bağlı olarak yüzdesel artmasıdır. Örneğin işsizlik değerleri her sene %5 olarak artıyorsa, her sene 1.05 ile çarpılması anlamına gelir. Bu da demek oluyor ki seri çarpımlı olarak artmaktadır. Bu yüzden “m” komutu seçildi.

```
plot(datasetdec, xlab="Yıllar")
```

Dekompozisyon yapılan seri plot komutuyla bileşenlerine ayrılmış bir şekilde grafiğe döküldü.

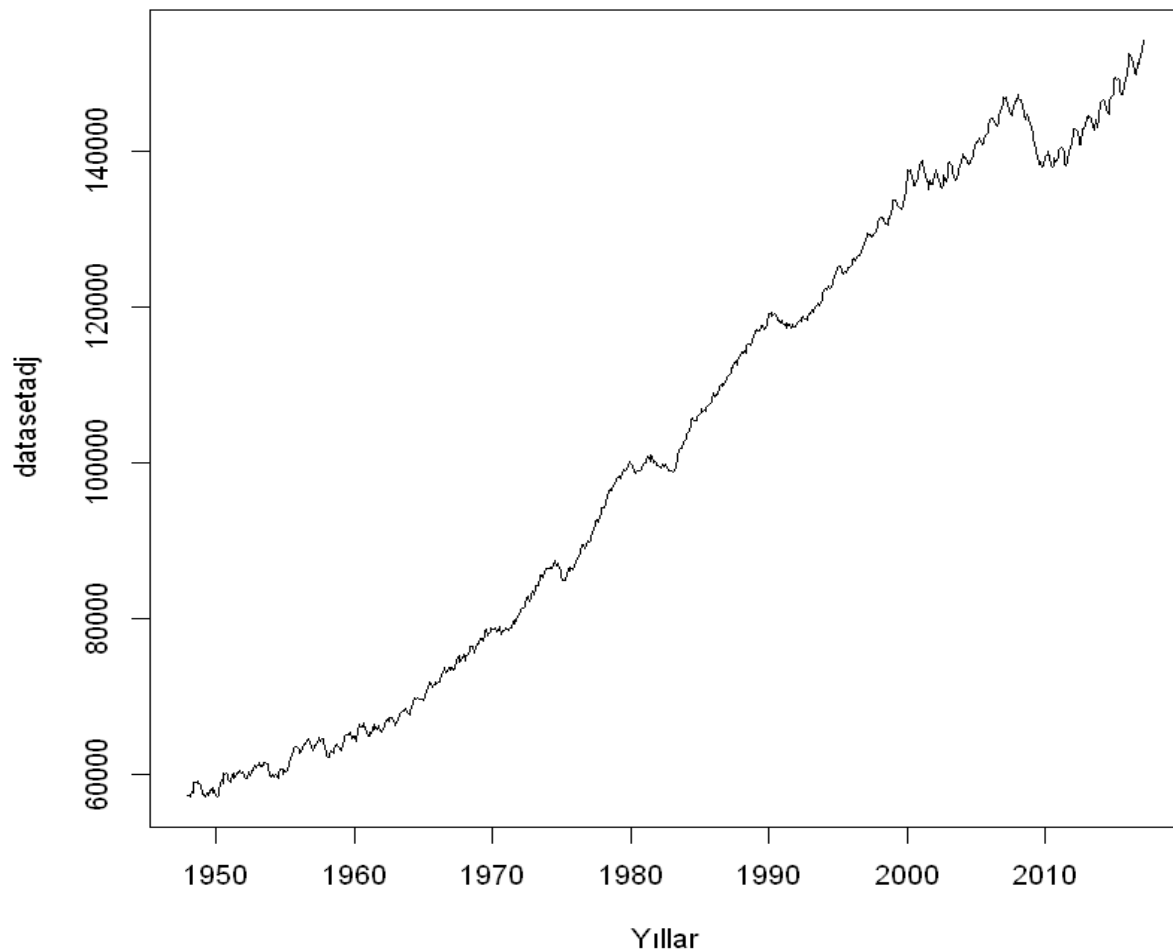


```
datasetadj<- dataset / datasetdec$seasonal
```

Bu komut ile seriden mevsimsel bileşen çıkarıldı.

```
plot(datasetadj, xlab="Yıllar")
```

Dekompozisyon yapılan seriden mevsimsel bileşenin çıkarılması ile oluşan grafik aşağıdaki gibidir.

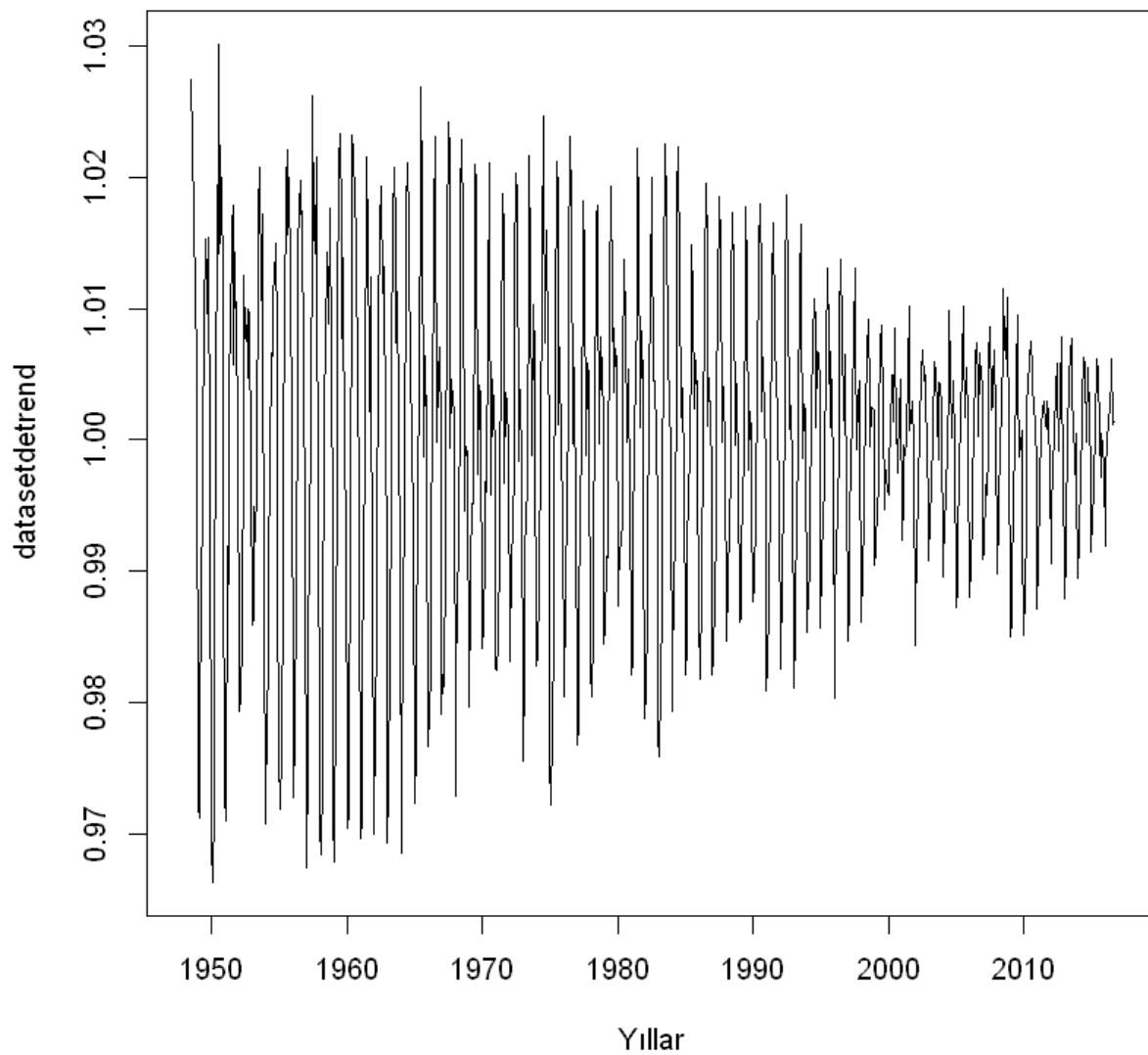


```
datasetdetrend<- dataset / datasetdec$trend
```

Bu komut ile seriden trend bileşeni çıkarıldı.

```
plot(datasetdetrend, xlab="Yıllar")
```

Dekompozisyon yapılan seriden trend bileşeninin çıkarılması ile oluşan grafik aşağıdaki gibidir.

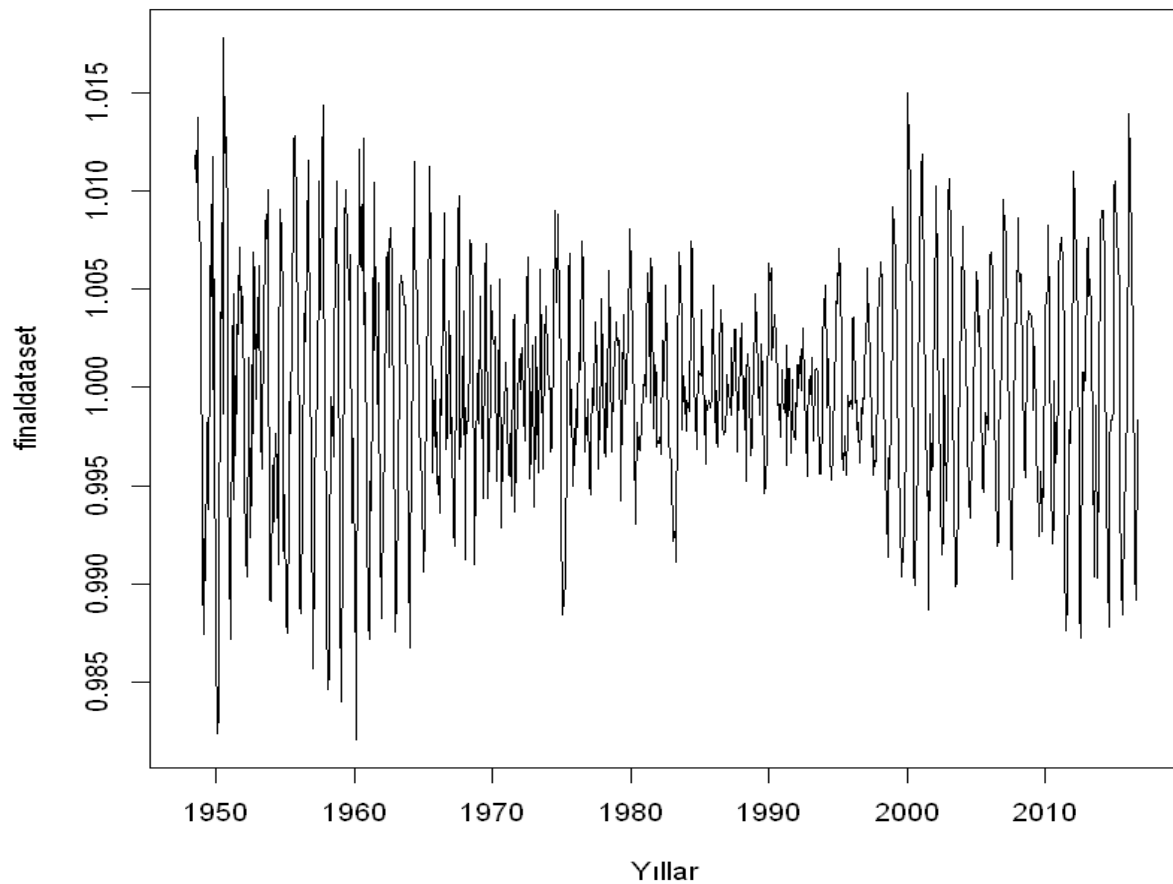


```
finaldataset<- dataset / (datasetdec$seasonal *  
datasetdec$trend)
```

Seriden mevsimsel ve trend bileşeninin çıkarılması işlemi yapılmıştır.

```
plot(finaldataset, xlab="Yıllar")
```

Dekompozisyon yapılan seriden mevsimsel ve trend bileşeninin çıkarılması ile oluşan grafik aşağıdaki gibidir.

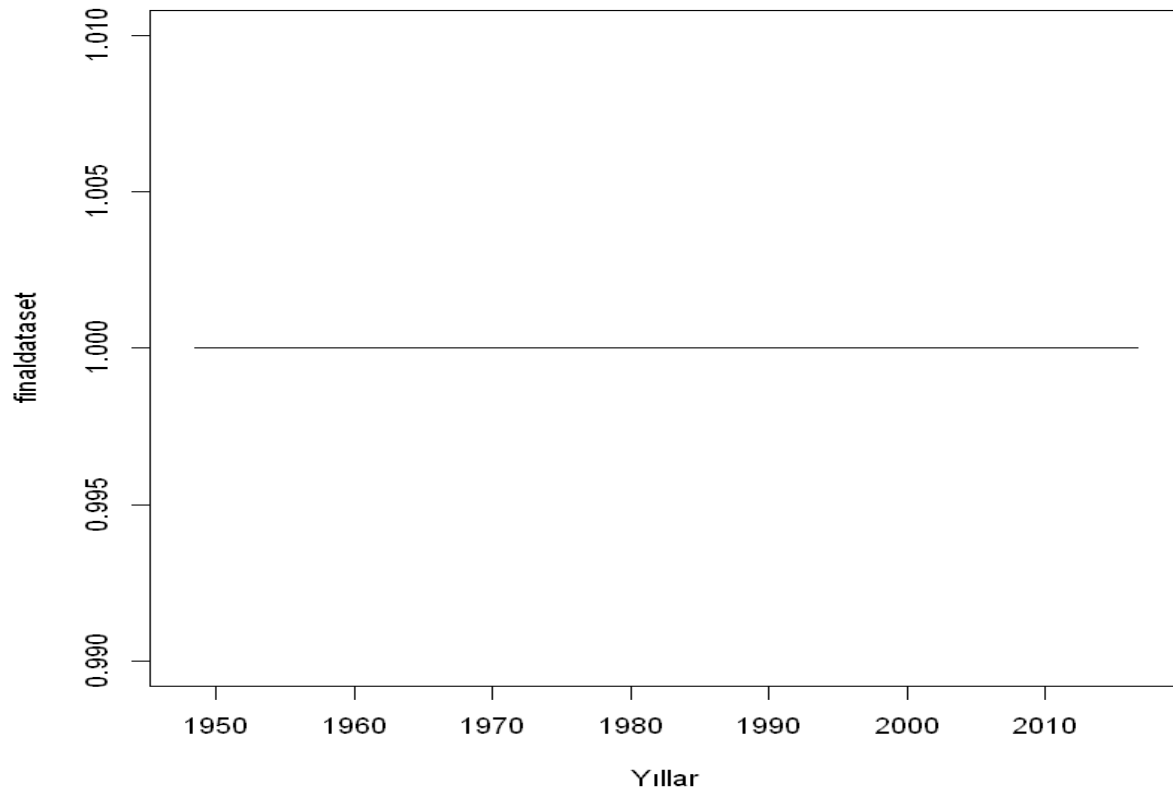


```
finaldataset2<- dataset / (datasetdec$seasonal *  
datasetdec$trend * datasetdec$random)
```

Bu komut ile seriden mevsimsel, trend ve random bileşeni çıkarılmıştır.

```
plot(finaldataset2, xlab="Yıllar")
```

Dekompozisyon yapılan seriden mevsimsel,trend ve random bileşeninin çıkarılması ile oluşan grafik aşağıdaki gibidir.

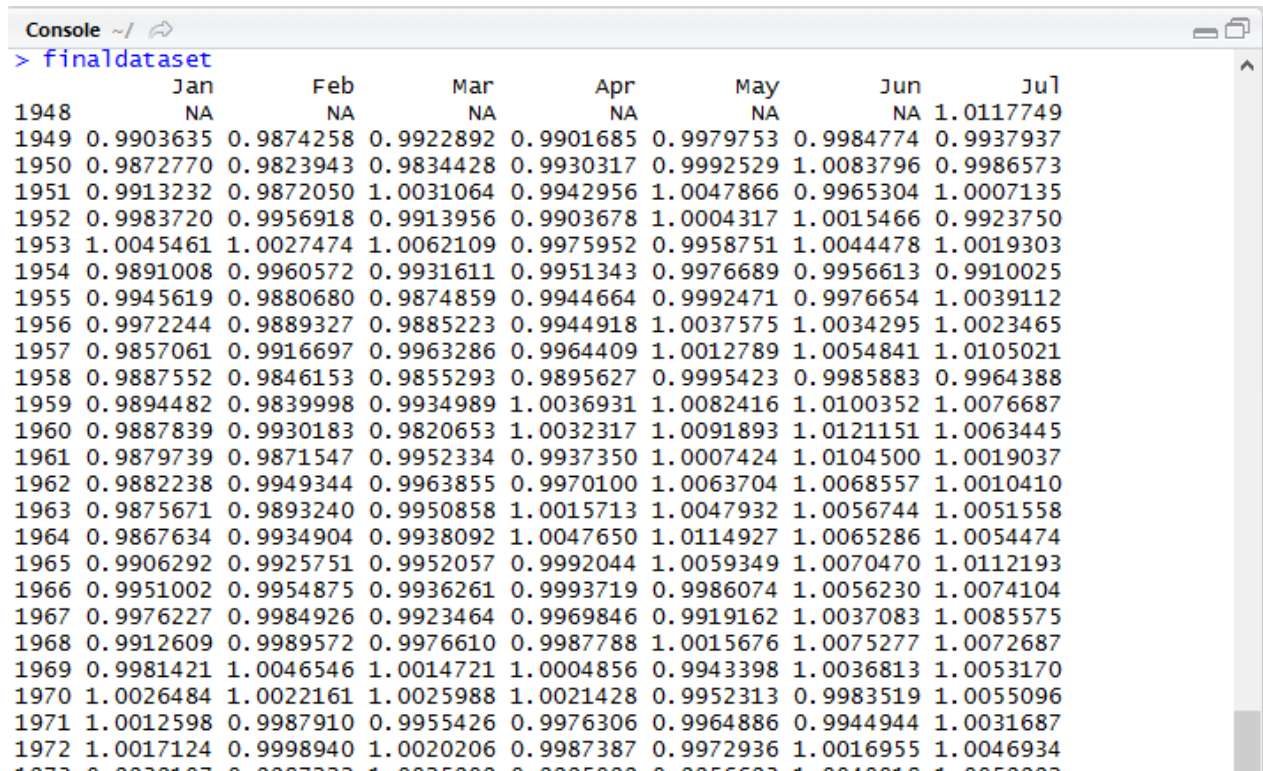


library(tseries)

Bir sonraki adımda kullanılacak olan `adf.test` komutunun çalışabilmesi için RStudio kütüphanesinde mevcut olan “tseries” paketinin aktif hale getirilmesi için yazılmıştır.

finaldataset

`adf.test` yapılmadan önce veride eksik değerler(NA) olup olmadığı kontrol edilmiştir.



```
Console ~/
> finaldataset
```

	Jan	Feb	Mar	Apr	May	Jun	Jul
1948	NA	NA	NA	NA	NA	NA	1.0117749
1949	0.9903635	0.9874258	0.9922892	0.9901685	0.9979753	0.9984774	0.9937937
1950	0.9872770	0.9823943	0.9834428	0.9930317	0.9992529	1.0083796	0.9986573
1951	0.9913232	0.9872050	1.0031064	0.9942956	1.0047866	0.9965304	1.0007135
1952	0.9983720	0.9956918	0.9913956	0.9903678	1.0004317	1.0015466	0.9923750
1953	1.0045461	1.0027474	1.0062109	0.9975952	0.9958751	1.0044478	1.0019303
1954	0.9891008	0.9960572	0.9931611	0.9951343	0.9976689	0.9956613	0.9910025
1955	0.9945619	0.9880680	0.9874859	0.9944664	0.9992471	0.9976654	1.0039112
1956	0.9972244	0.9889327	0.9885223	0.9944918	1.0037575	1.0034295	1.0023465
1957	0.9857061	0.9916697	0.9963286	0.9964409	1.0012789	1.0054841	1.0105021
1958	0.9887552	0.9846153	0.9855293	0.9895627	0.9995423	0.9985883	0.9964388
1959	0.9894482	0.9839998	0.9934989	1.0036931	1.0082416	1.0100352	1.0076687
1960	0.9887839	0.9930183	0.9820653	1.0032317	1.0091893	1.0121151	1.0063445
1961	0.9879739	0.9871547	0.9952334	0.9937350	1.0007424	1.0104500	1.0019037
1962	0.9882238	0.9949344	0.9963855	0.9970100	1.0063704	1.0068557	1.0010410
1963	0.9875671	0.9893240	0.9950858	1.0015713	1.0047932	1.0056744	1.0051558
1964	0.9867634	0.9934904	0.9938092	1.0047650	1.0114927	1.0065286	1.0054474
1965	0.9906292	0.9925751	0.9952057	0.9992044	1.0059349	1.0070470	1.0112193
1966	0.9951002	0.9954875	0.9936261	0.9993719	0.9986074	1.0056230	1.0074104
1967	0.9976227	0.9984926	0.9923464	0.9969846	0.9919162	1.0037083	1.0085575
1968	0.9912609	0.9989572	0.9976610	0.9987788	1.0015676	1.0075277	1.0072687
1969	0.9981421	1.0046546	1.0014721	1.0004856	0.9943398	1.0036813	1.0053170
1970	1.0026484	1.0022161	1.0025988	1.0021428	0.9952313	0.9983519	1.0055096
1971	1.0012598	0.9987910	0.9955426	0.9976306	0.9964886	0.9944944	1.0031687
1972	1.0017124	0.9998940	1.0020206	0.9987387	0.9972936	1.0016955	1.0046934

na.omit(finaldataset)

na.omit komutuyla verideki eksik değerler(NA) giderildi. Çıktıda görüldüğü üzere NA yazan yerler yok edildi.

```
> na.omit(finaldataset)
```

	Jan	Feb	Mar	Apr	May	Jun	Jul
1948							1.0117749
1949	0.9903635	0.9874258	0.9922892	0.9901685	0.9979753	0.9984774	0.9937937
1950	0.9872770	0.9823943	0.9834428	0.9930317	0.9992529	1.0083796	0.9986573
1951	0.9913232	0.9872050	1.0031064	0.9942956	1.0047866	0.9965304	1.0007135
1952	0.9983720	0.9956918	0.9913956	0.9903678	1.0004317	1.0015466	0.9923750
1953	1.0045461	1.0027474	1.0062109	0.9975952	0.9958751	1.0044478	1.0019303
1954	0.9891008	0.9960572	0.9931611	0.9951343	0.9976689	0.9956613	0.9910025
1955	0.9945619	0.9880680	0.9874859	0.9944664	0.9992471	0.9976654	1.0039112
1956	0.9972244	0.9889327	0.9885223	0.9944918	1.0037575	1.0034295	1.0023465
1957	0.9857061	0.9916697	0.9963286	0.9964409	1.0012789	1.0054841	1.0105021
1958	0.9887552	0.9846153	0.9855293	0.9895627	0.9995423	0.9985883	0.9964388
1959	0.9894482	0.9839998	0.9934989	1.0036931	1.0082416	1.0100352	1.0076687
1960	0.9887839	0.9930183	0.9820653	1.0032317	1.0091893	1.0121151	1.0063445
1961	0.9879739	0.9871547	0.9952334	0.9937350	1.0007424	1.0104500	1.0019037
1962	0.9882238	0.9949344	0.9963855	0.9970100	1.0063704	1.0068557	1.0010410
1963	0.9875671	0.9893240	0.9950858	1.0015713	1.0047932	1.0056744	1.0051558
1964	0.9867634	0.9934904	0.9938092	1.0047650	1.0114927	1.0065286	1.0054474
1965	0.9906292	0.9925751	0.9952057	0.9992044	1.0059349	1.0070470	1.0112193
1966	0.9951002	0.9954875	0.9936261	0.9993719	0.9986074	1.0056230	1.0074104
1967	0.9976227	0.9984926	0.9923464	0.9969846	0.9919162	1.0037083	1.0085575
1968	0.9912609	0.9989572	0.9976610	0.9987788	1.0015676	1.0075277	1.0072687
1969	0.9981421	1.0046546	1.0014721	1.0004856	0.9943398	1.0036813	1.0053170
1970	1.0026484	1.0022161	1.0025988	1.0021428	0.9952313	0.9983519	1.0055096
1971	1.0012598	0.9987910	0.9955426	0.9976306	0.9964886	0.9944944	1.0031687
1972	1.0017124	0.9998940	1.0020206	0.9987387	0.9972936	1.0016955	1.0046934
1973	0.9939107	0.9987222	1.0025909	0.9995099	0.9956693	1.0040018	1.0059992
1974	1.0013043	0.9999268	0.9999570	0.9967091	0.9969768	1.0030245	1.0090034
1975	0.9938392	0.9884232	0.9891569	0.9902675	0.9917371	0.9970010	1.0056047

`adf.test(na.omit(finaldataset))`

`adf.test` komutuyla serinin durağan olup olmadığı test edildi. Oluşturulan zaman serisi modelinde nedensellik testinin yapılması için öncelikle modelde kullanılan serilerin durağanlığı bilinmelidir. Bu nedenle serinin durağanlığını tespit etmek için Dickey-Fuller tarafından geliştirilen ADF Birim Kök Testi uygulanmıştır.

ADF Birim Kök Testini uygulamak için oluşturulan hipotez;

$H_0: \rho=1$

$H_1: \rho<1$

Hipotezdeki ρ gecikmeli değeri ifade etmektedir. Çıkan sonuç 1'e yaklaşıyorsa birim kökün var olduğunu yani modelin durağan olmadığını gösterir. Sonuç 1'den uzaklaşmışsa bu modelin durağan olduğunu gösterir. Eğer modelde durağanlık yoksa modelle elde edilen sonuçlar tutarsız çıkacaktır. Bu nedenle modelin durağan olmasını beklenir.

```
> adf.test(na.omit(finaldataset))
```

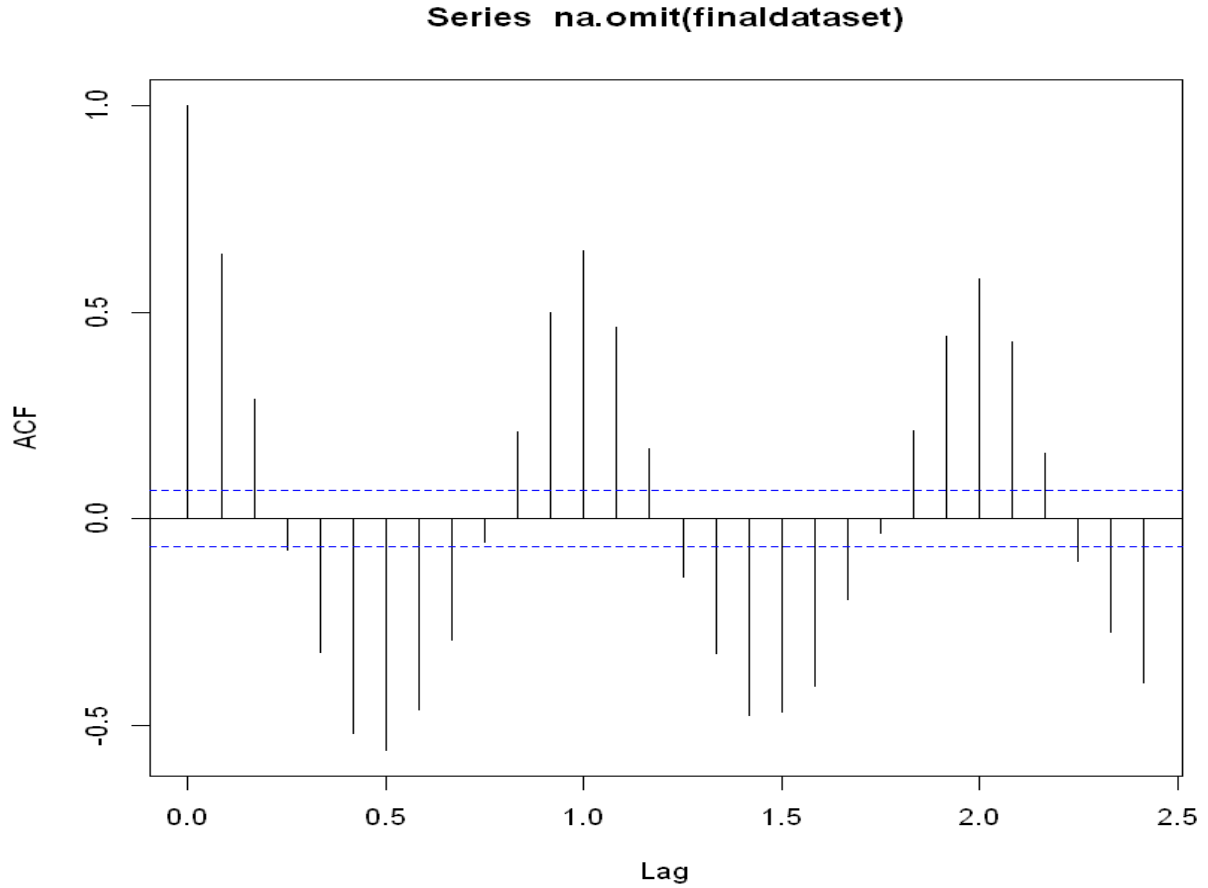
```
Augmented Dickey-Fuller Test
```

```
data: na.omit(finaldataset)
Dickey-Fuller = -13.801, Lag order = 9, p-value = 0.01
alternative hypothesis: stationary
```

Augmented Dickey-Fuller Test'i yaptığımız zaman prob değerimizi 0.01 olarak görmekteyiz. P değeri 1'den uzak bir değer olduğu için serinin durağan olduğu sonucuna varılır.

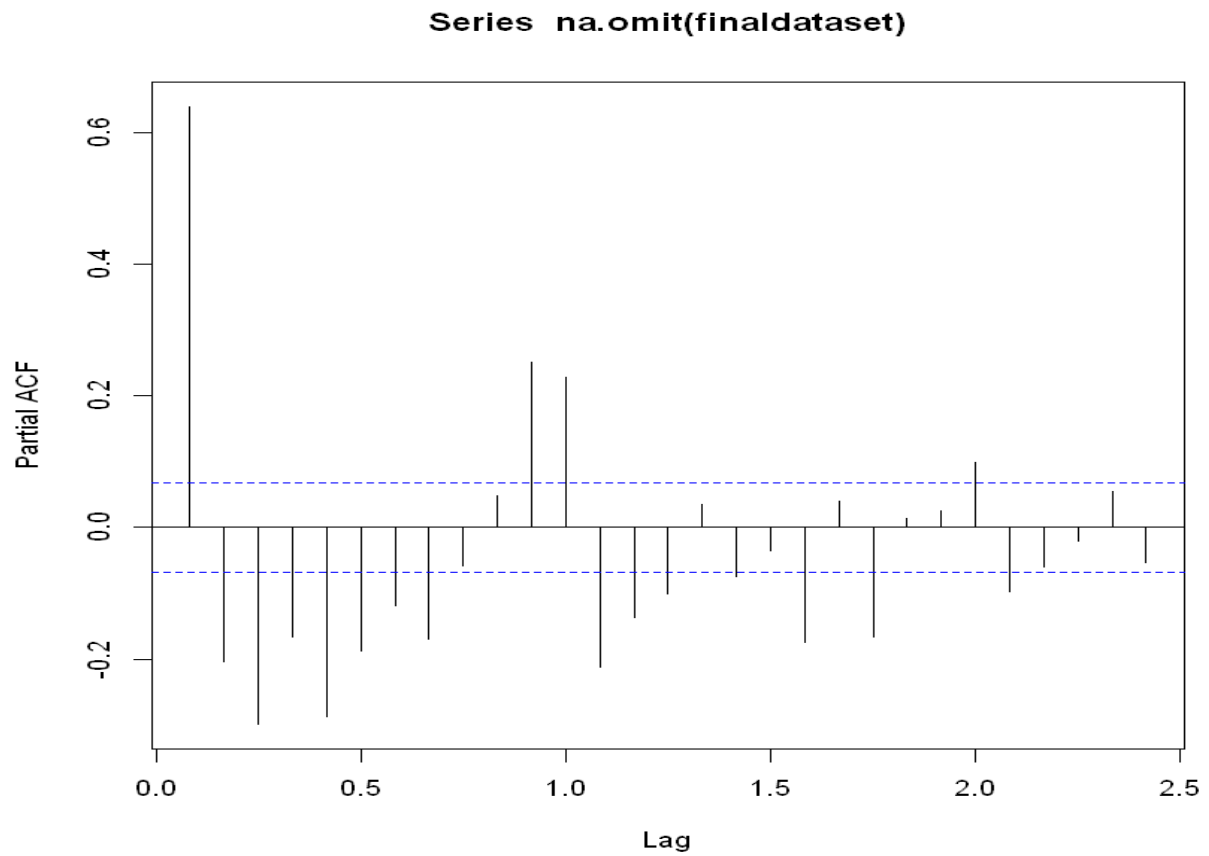
acf(na.omit(finaldataset))

ACF komutu zaman serisini geriye kaydırarak değerlerin önceki aylarla korelasyonu olup olmadığını test eder. Aşağıdaki çıktıda görüldüğü üzere değerlerin bazıları hata payının dışında kaldığı için bu test verinin tam olarak temizlenmediğini gösterir.



`pacf(na.omit(finaldataset))`

Kısmi otokorelasyon fonksiyonu ilk korelasyon çıkarıldıktan sonra kalanına bakıp, kalanın bir sonraki değerler ile arasında ilişki olup olmadığını kontrol eder.



library(forecast)

Bir sonraki adımda kullanılacak olan auto.arima komutu için RStudio kütüphanesinde mevcut olan forecast komutu aktif hale getirildi.

auto.arima(finaldataset)

Bu komut otomatik olarak en iyi arima modelini kullanıcıya sunar. Buradaki ARIMA(p,d,q) değerleri otomatik olarak atanır. Bu değerlerin anlamları şu şekildedir:

p: otoregresif terimlerin sayısı

q: tahmin denklemindeki gecikmeli tahmin hatalarının sayısı

d: durağanlık için alınan fark sayısıdır.

```
> auto.arima(finaldataset, stationary=FALSE, seasonal=FALSE)
Series: finaldataset
ARIMA(5,0,0) with non-zero mean

Coefficients:
      ar1      ar2      ar3      ar4      ar5    mean
    0.6093 -0.0201 -0.1725  0.0215 -0.2942  1e+00
s.e.  0.0335  0.0397  0.0393  0.0397  0.0336  2e-04

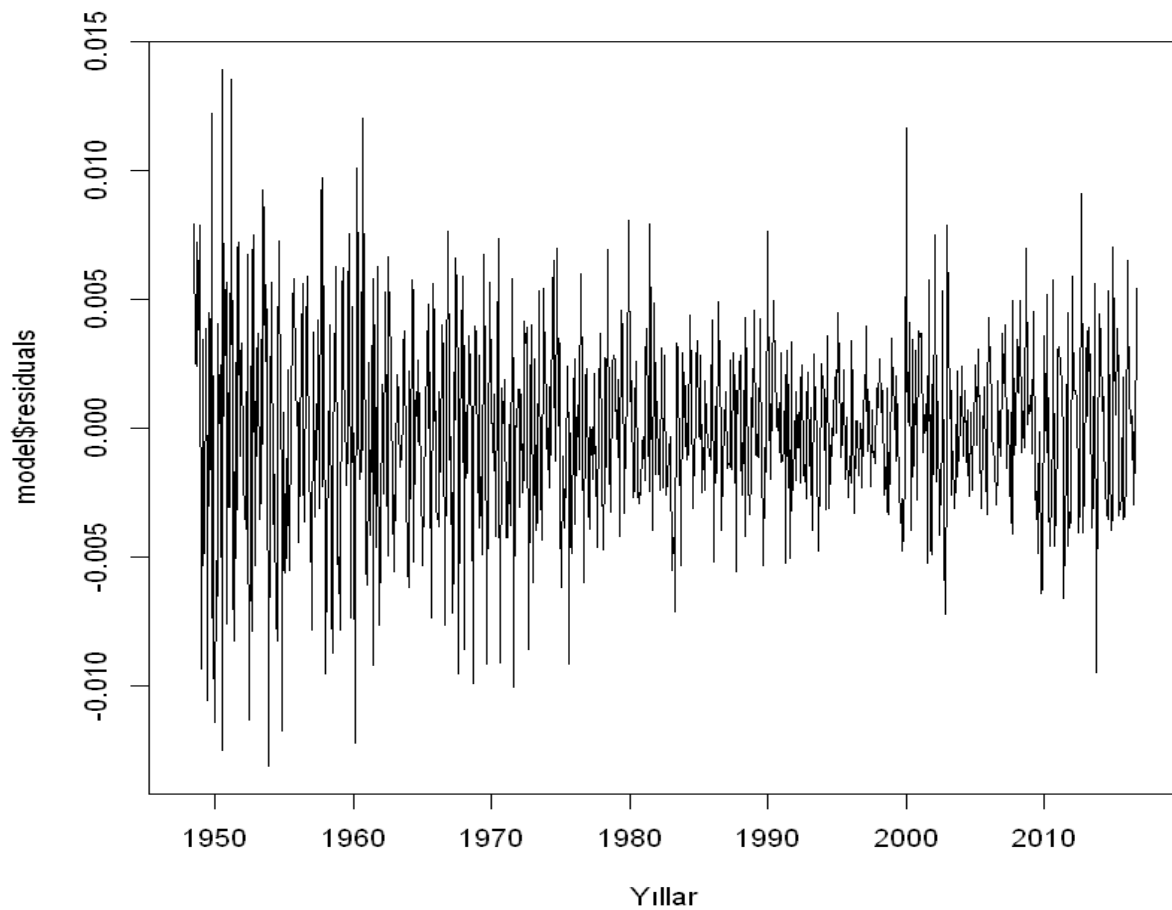
sigma^2 estimated as 1.473e-05:  log likelihood=3390.03
AIC=-6766.06  AICC=-6765.92  BIC=-6733
>
```

```
model<-arima(finaldataset, order=c(5,0,0))
```

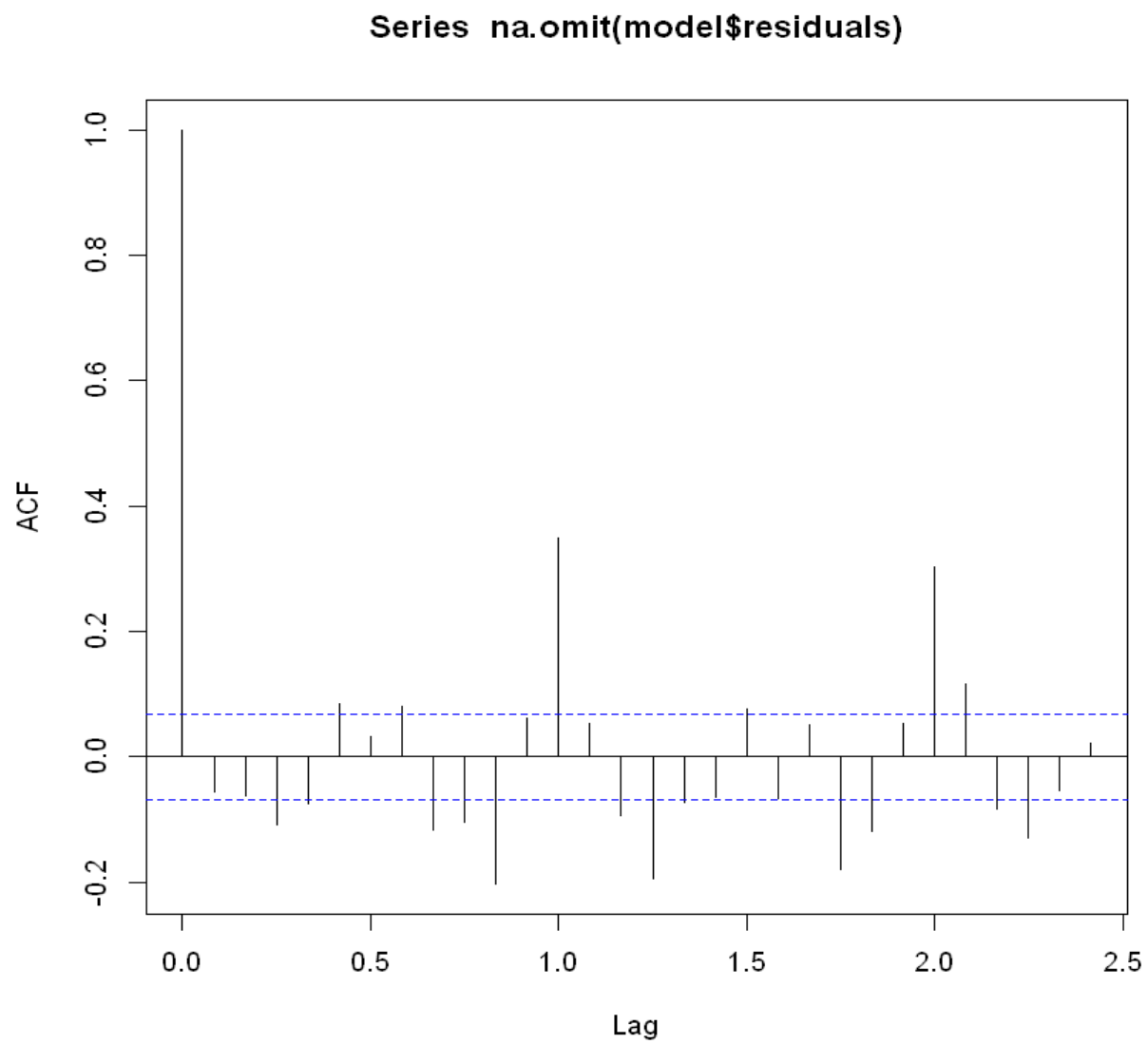
Oluşturulan otomatik arima modeline uygun (p,d,q) değerleri ile model oluşturuldu.

```
plot(model$residuals, xlab="Yıllar")
```

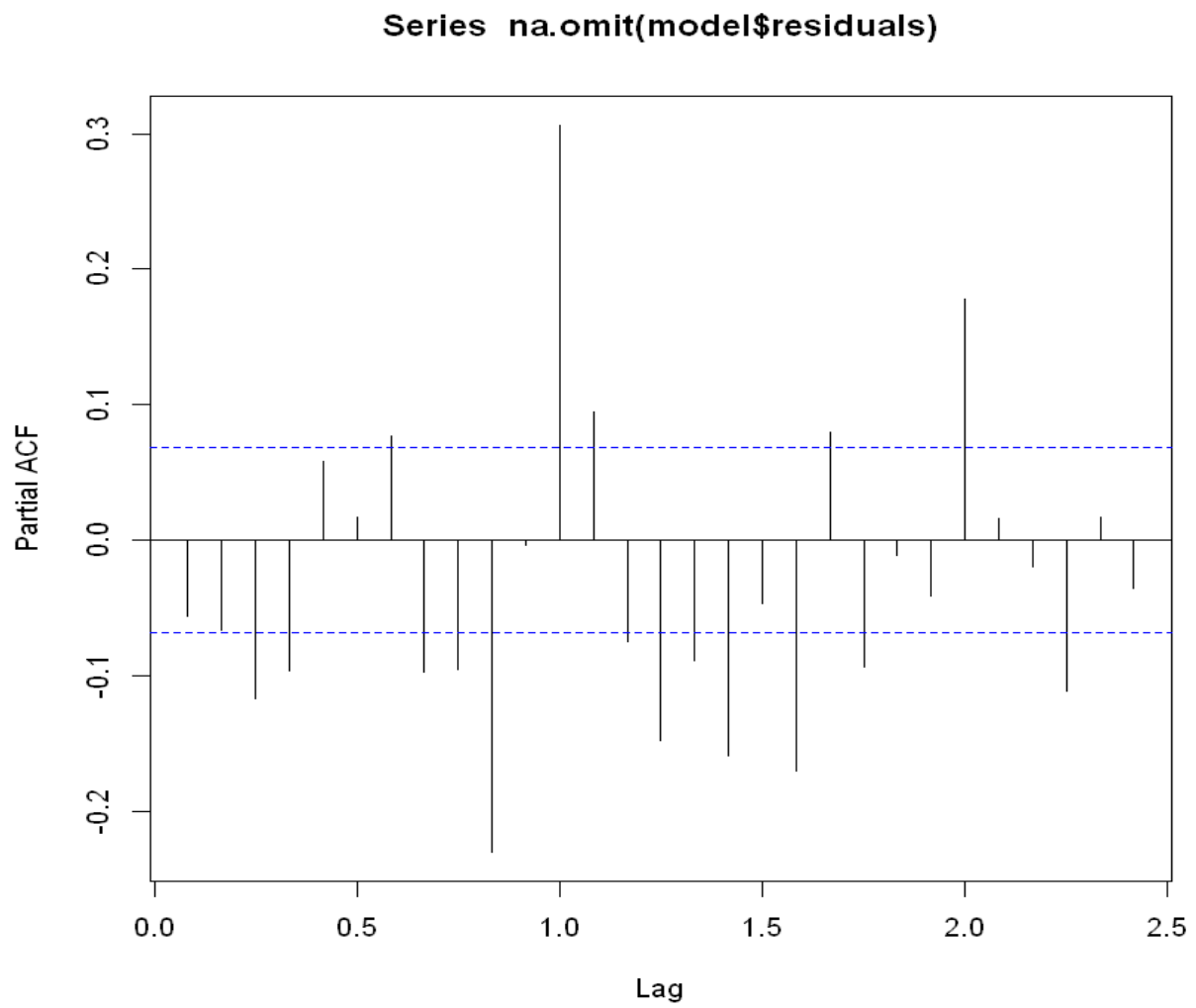
Oluşturulan arima modelinden kalanlar(residuals) grafiğe döküldü. Çıktıda görüldüğü üzere serinin zaman boyunca ortalaması sıfıra yaklaşıyor. Bu yüzden durağandır ve analize uygundur.



`acf(na.omit(model$residuals))`




```
pacf(na.omit(model$residuals))
```



KAYNAKLAR

- [1]Arslan İ., 2015. *R ile İstatiksel Programlama*
- [2] **United States Department of Labor**, (*Unadj*) *Employment Level*,
<https://beta.bls.gov/dataViewer/view/timeseries/LNU02000000>
- [3]Shumway R.H.&Stoffer D.S. , 2016. *Time Series Analysis and Applications*,
pp.5,54,88-99
- [4]Ven G., 2010. *Removal of Trend & Seasonality Handout 4*, p.2.
- [5]Magakian M., 2015. *Extract Seasonal&Trend: using decomposition in R*
- [6]Hyndman R.&Athanasopoulos G., *Classical Decomposition*,
<https://www.otexts.org/fpp/6/3>
- [7]Kırman O., 2016. *R ile Analiz /Bölüm4: Birim Kök Testi -1 : Paketler*,
<http://www.verianalitigi.org/programlama/r-ile-analiz-bolum-4-birim-kok-testi-1-paketler/>
- [8]PennState Eberly College of Science, *Applied Time Series Analysis*,
<https://onlinecourses.science.psu.edu/stat510/node/69>