Zaman Serileri Analizi

Ramazan Erduran, İlkay Ş. Baytar

Hacettepe Üniversitesi

Özet:

Bu ödev raporlamasında veri seti olarak, İstanbul Büyükşehir Belediyesi'nin İkitelli'de kurulan güneş panellerinden elde edilen enerjinin ölçüldüğü zaman serisi kullanılmıştır. Bu veri seti 15 dakikalık ölçümler içeriyor olup, saatlik olacak şekilde R programında ön işleme yapılmıştır.

Analiz sürecinde ön işleme olarak zaman serisi analizi yapıldıktan sonra serinin mevsimselliğe ve trende sahip bir zaman serisi olduğuna karar verilmiştir. Buna ilişkin; toplamsal ayrıştırma, çarpımsal ayrıştırma, basit doğrusal regresyon, Holt-Winters Üstel düzleştirme yöntemi ve Box-Jenkins modelleri kullanılmış olup, hataları ak gürültü serisi, hata kareler ortalaması (HKO) ve Baessian Yanlılığı (BIC) değeri en düşük olan model ARIMA(0,1,0)(2,1,0) seçilmiş olup, öngörüler bu modellerin tahminleriyle oluşturulmuştur.

Analiz sonucunda gelecek zamana ait enerji üretimini tahmin edebilen modeli elde ettik.
Modelimizin BIC değerini 1542.19 RMSE değerini ise 99.38836 olarak elde ettik.

Temel Zaman Serisi Analizi:

Verilerimizi XLSX formatından ilk okuttuğumuzda elde ettiğimiz veriler *Şekil 1* deki gibi 15 dakikalık olarak ölçülmüştür. Verilerimize ön işleme yaparak şekil 2 de görülen saatlik hale getirdik. Analizimizde ham veriyi Şekil 2 deki veri kabul ettik.

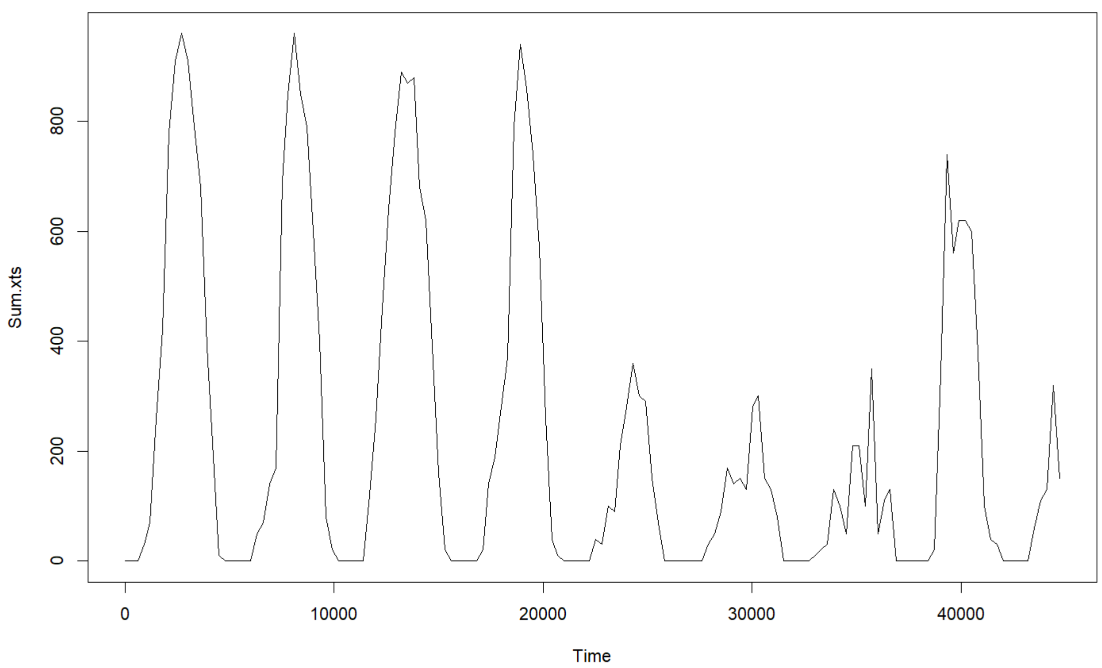
Tarih	Üretim (kWh)
2018-05-01 09:30:00	40
2018-05-01 09:45:00	70
2018-05-01 10:00:00	80
2018-05-01 10:15:00	80
2018-05-01 10:30:00	100
2018-05-01 10:45:00	130
2018-05-01 11:00:00	110
2018-05-01 11:15:00	160

Şekil 1

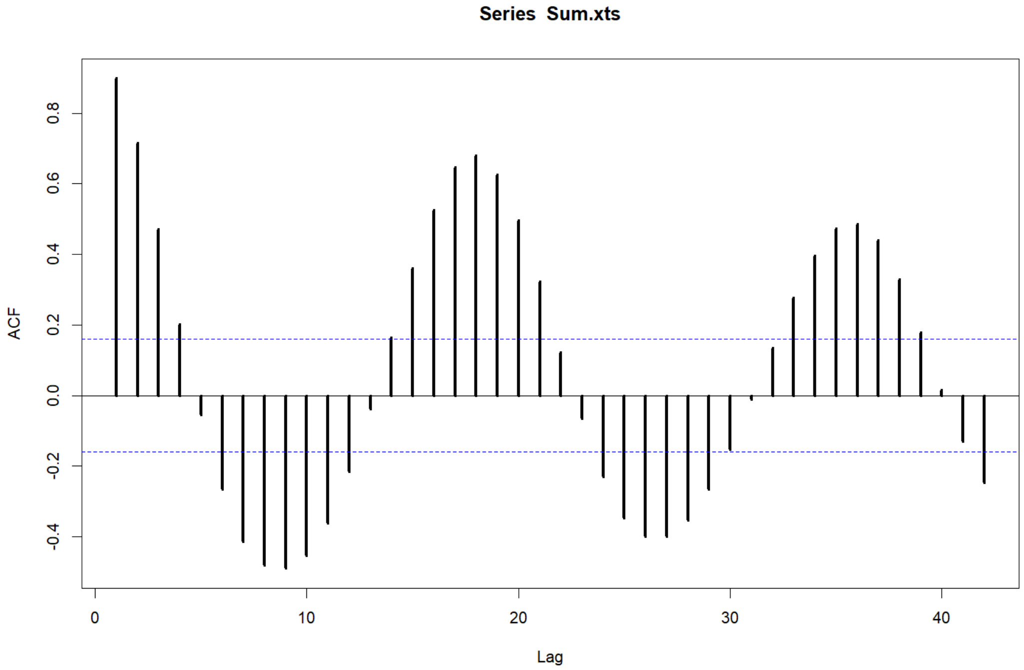
	V1
2018-05-01 09:59:59	260
2018-05-01 10:59:59	420
2018-05-01 11:59:59	780
2018-05-01 12:59:59	910
2018-05-01 13:59:59	960
2018-05-01 14:59:59	910
2018-05-01 15:59:59	790

Şekil 2

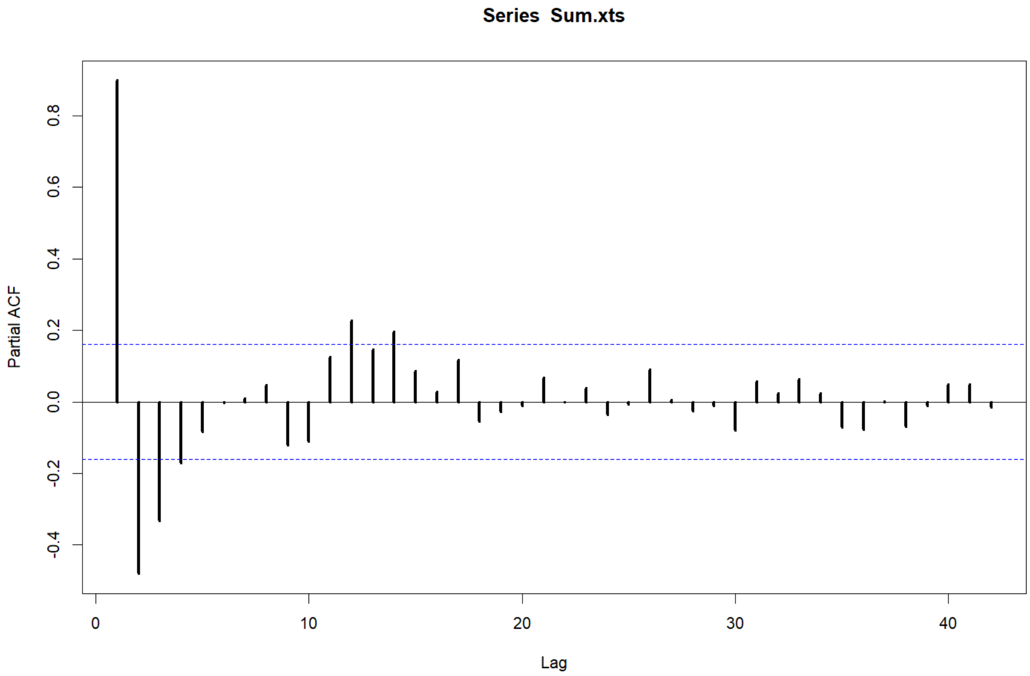
İlk olarak verimizin zaman serisi grafiğini çizdirerek analizimize başlayalım. Şekil 3 te gözüktüğü üzere mevsimsellik net olarak bellidir. Ancak trendin varlığından tam olarak söz edemeyiz. Bunun için zaman serimizin 42 gecikmeli ACF ve PACF grafiklerine bakmamız gerekir.



Şekil 3



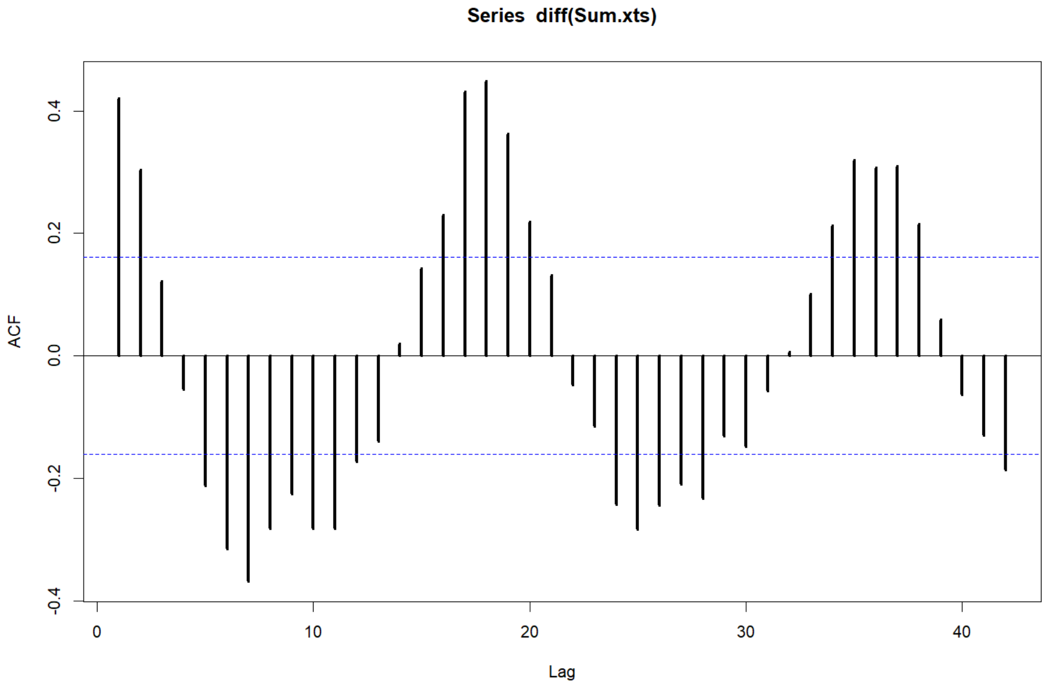
Şekil 4



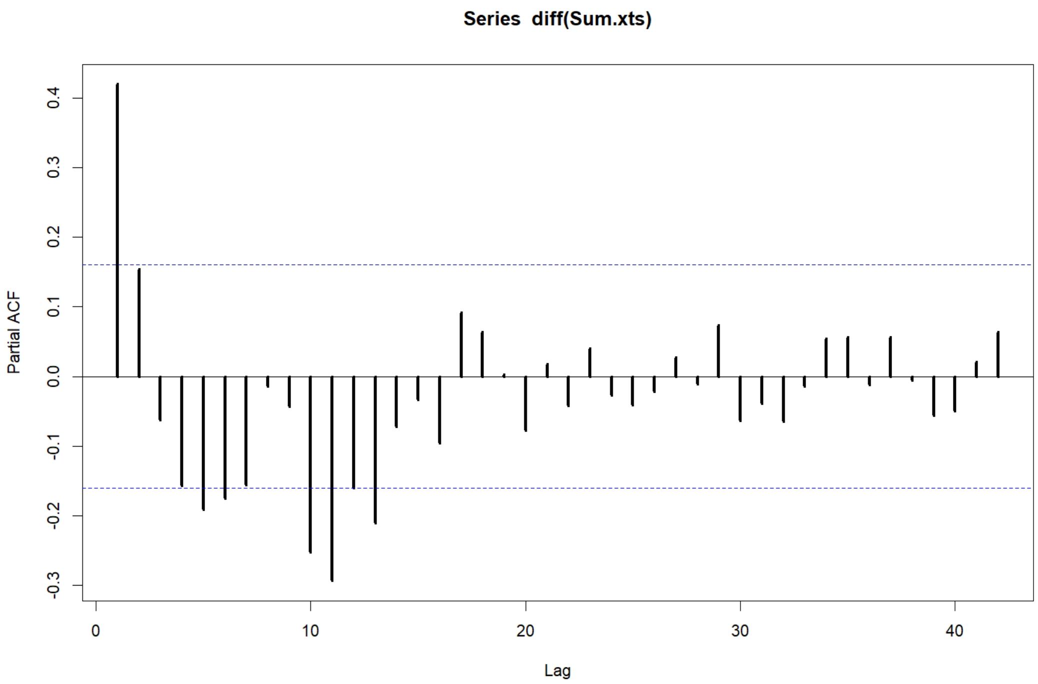
Şekil 5

Şekil 4 de zaman serisinin ACF grafiğini görmekteyiz. ACF grafiğinden çıkarımımız şunlar olacaktır:

1. İlk 4 gecikme CI (Güven Aralığı) sınırları dışında yer aldığı için trend vardır şeklinde yorum yapabiliriz.
2. Her 18 gözlemde bir verilerimizde artış gözlemlendiği için mevsimsellik vardır ve periyodumuz 18dir diyebiliriz.
3. Durağanlaştımak için serimizin birinci farklarını almamız gerekecektir. Durağanlaştırdıktan sonra eğer mevsimsellik devam ediyor olursa mevsimsel farklarını da almamız gerekecektir.

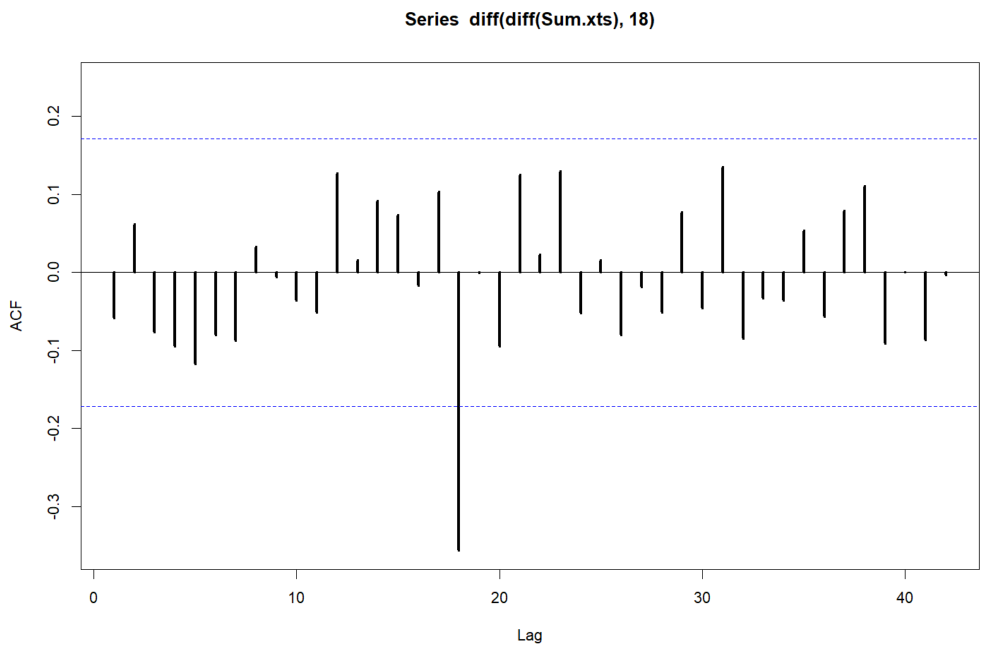


Şekil 6



Şekil 7

Serimizin 1. farklarını aldıktan sonra, trendden arındırdık. Artık serimiz durağan, ancak baskın bir mevsimsellik var bu nedenle serimizin mevsimsel farkını da almamız gerekecek.



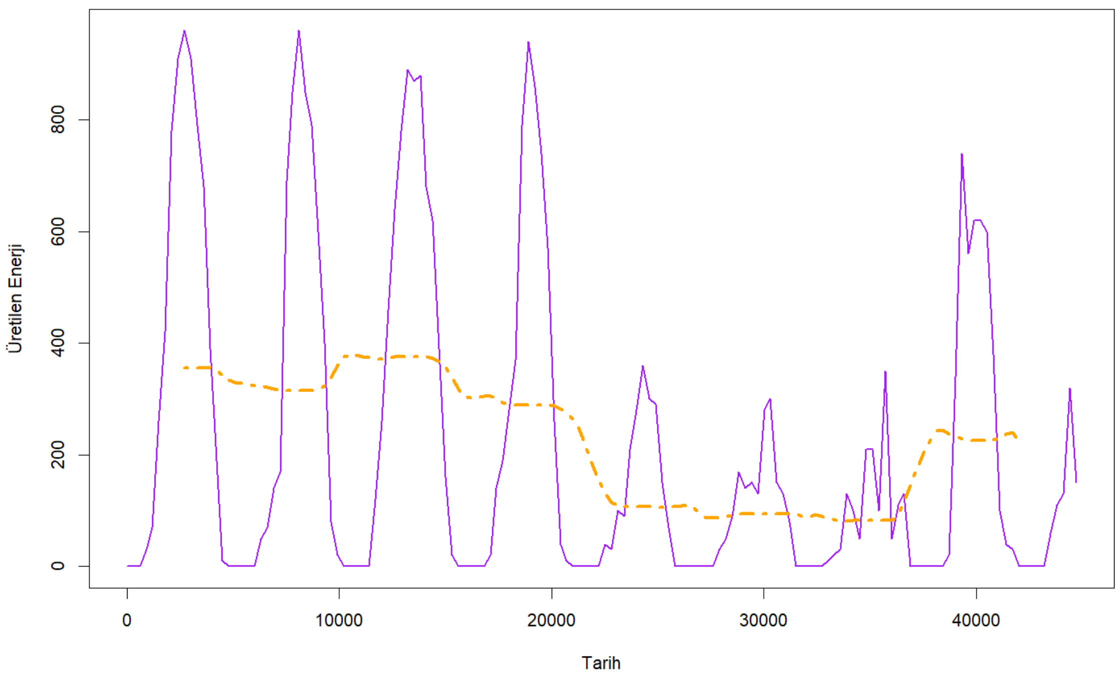
Şekil 8

Şekil 8 de görüldüğü üzere mevsimsellikten de arındırdık ve görüyoruz ki serimiz akgürültü serisi değil. Serimizin 1. Mevsimsel ve 1. Farkı alındıktan sonra durağan hale geldiği söylenebilir bu nedenle serinin 2. Farklarının alınması anlamlı değildir.

Buraya kadar olan kısımdan not etmemiz gereken hususlar:

1. Serimizi durağanlaştırmak için kaç defa fark işlemi uyguladık: 1
2. Serimizi baskın mevsimsellikten arındırmak için kaç defa mevsimsel fark işlemi uyguladık: 1
3. Bu kısımlar not etmemiz bize ileride Box-Jenkins modellerini kurarken yardımcı olacak.

Merkezsel Hareketli Ortalama (MHO) serisi ile orijinal serinin birlikte bulunduğu grafiği Şekil 9 da inceleyebiliriz. Merkezsel Hareketli Ortalama Serisini hesaplarken germe katsayısını, periyodumuz ne ise (18) ona eşitleyerek hesaplamalıyız.



Şekil 9

Toplamsal Ayrıştırma Modeli:

Toplamsal Ayrıştırma Modeli'ni direkt uygulamadan önce bazı bileşenler elde edilmelidir. Ondan da önce ACF ve PACF grafiğine bakılıp trent varsa trendden arındırmalı mevsimsellik baskında mevsimsel fark uygulanmalıdır. Bir önceki başlıkta bunları yaptığımız için bu kısımda tekrar tekrar bunları yapmayacağız.

Öncelikle ihtiyacımız olan değişkenleri oluşturalım, sonrasında ise bu oluşturduğumuz değişkenleri kullanarak Toplamsal Ayrıştırma modelini uygulayıp bir tahmin serisi elde edelim. Eğer tahmin serisi anlamlı hatalar akgürültü serisi ise bu model kullanılabilir şeklinde yorum yapacağız. Bu işlem süresince şu şekilde ilerleyeceğiz:

1. Trent bileşeni
2. Mevsimsel bileşen serisi
3. Merkezsel Hareketli Ortama MHO
4. Mevsim bileşeni
5. Mevsimlerin ortalaması
6. Mevsim endeks değerleri
7. Hataya sahip trent bileşeni
8. Hatadan arındırılmış trent bileşeni
9. Tahmin serisi
10. Hata Serisi
11. Modelin Geçerliliği

1. Trent Bileşeni

Trent bileşenini oluştururken orijinal serimizi zaman serisi "time series" olarak tanımladığımızdan emin olmalıyız.

```
> df_trent <- tslm(df_ts ~ trend) ; df_trent

Call:
tslm(formula = df_ts ~ trend)

Coefficients:
(Intercept)        trend
    372.954       -1.962
```

Şekil 10

Şekil 10 'da görüldüğü üzre trent bileşenimizi oluşturduk.

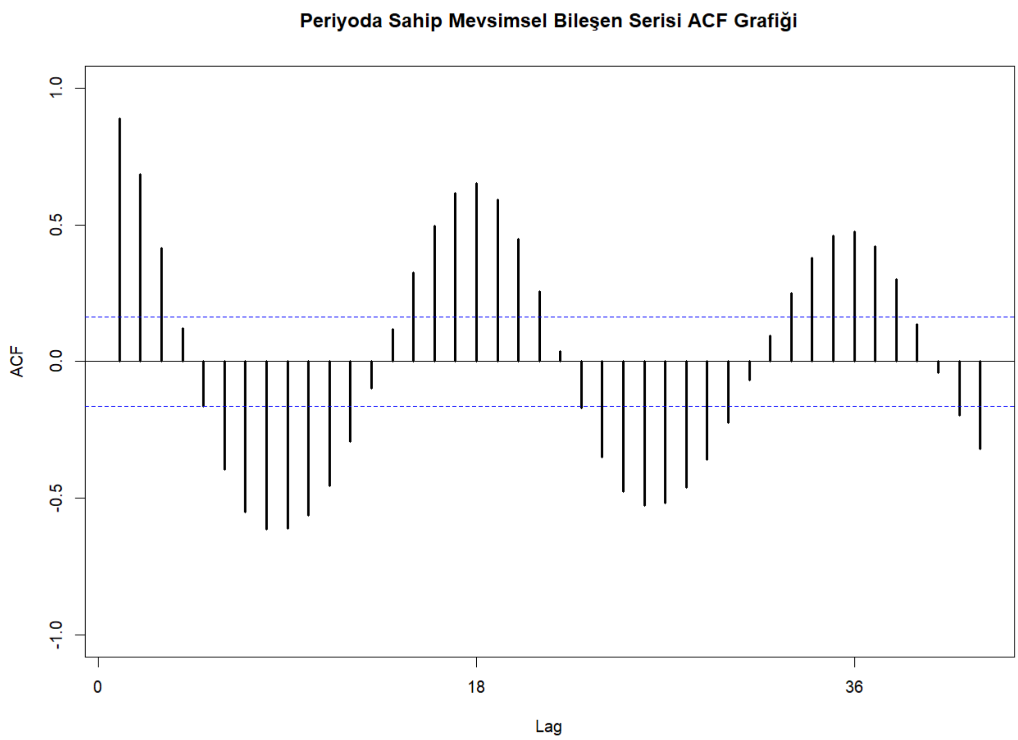
2. Periyoda Sahip Mevsimsel Bileşen Serisi

Periyoda sahip mevsimsel bileşen serisi orijinal seriden trent bileşenini çıkararak elde edebiliriz. Elde ettiğimiz mevsimsel bileşen serisinin ilk 5 gözlemi ni Şekil 11 ‘deki gibi elde ettik.

```
Time Series:
Start = c(1, 1)
End = c(8, 18)
Frequency = 18
             df_ts
  [1,] -370.99167
  [2,] -369.02937
  [3,] -367.06707
  [4,] -335.10478
  [5,] -293.14248
```

Şekil 11

Mevsimsel bileşen serisinin oto korelasyon grafiğini (ACF) çizdirdiğimizde ise periyodu 18 olarak belirledik. Mevsimsel bileşenin ACF grafiği Şekil 12 ‘de görülebilir.



Şekil 12

3. Merkezsel Hareketli Ortalama (MHO)

Merkezsel Hareketli Ortalama bize "Mevsim" serisini bulmada yardımcı olacak bir eleman. Merkezsel Hareketli Ortalama, belirli bir süre boyunca bir güvenliğin ortalama değerini gösteren değer tabanlı, gecikmeli (veya reaktif) bir göstergedir. Merkezsel Hareketli Ortalama, momentumu ölçmenin yanı sıra trendleri doğrulamanın ve destek ve direnç alanlarını tanımlamanın iyi bir yoludur.

Orijinal serimizin MHO ile birlikte olan çizgi grafiğine Şekil 9 'dan ulaşılabilir.

4. Mevsimsel Bileşen Serisi

Mevsimsel bileşen serisi, orijinal seriden Merkezsel Hareketli Ortalamanın çıkartılmasıyla elde edilen bir seridir.

Mevsimsel bileşen serisi elde edildikten sonra bu seri kullanılarak mevsimsel endeks serisi hesaplanır.

5. Mevsimlerin (Dönem) Ortalamaları

Mevsimsel endeks değerlerinin bulunabilmesi için, serideki periyodun 1,2,…,18 değerleri için ayrı ayrı mevsim serisinin ortalamaları bulunmalıdır.

```
> colMeans(donemort, na.rm = T)
 [1] -233.48413 -229.95238 -223.68254 -176.97619  -84.57937   50.60317
 [7]   74.25397  195.72222  328.57937  401.34921  365.63492  331.34921
[13]  171.38889   17.50397 -138.32540 -229.35714 -237.84921 -236.06349
> mean(colMeans(donemort, na.rm = T))
[1] 8.117504
```

Şekil 13

Şekil 13 'te dönemlerimizin ortalamaları görülmüştür. Örnek bir yorum olarak. Her periyodun ilk dönemindeki "mevsim" serisinin verilerinin ortalaması -233.48413 'müş. Bu 18 değerin ortalaması ise 8.117504 olarak elde edilmiş.

6. Mevsimsel Endeks Değerleri

Mevsimlerin ortalamalarından hareketle, mevsimsel endeks değerleleri Şekil 14 'teki gibi bulunur.

```
> endeks
 [1] -241.601631 -238.069885 -231.800044 -185.093695  -92.696869
 [6]   42.485670   66.136464  187.604718  320.461861  393.231702
[11]  357.517416  323.231702  163.271384    9.386464 -146.442901
[16] -237.474647 -245.966711 -244.180996
```

Şekil 14

Bulunan mevsimsel endeks değerleri seri boyunca yazılmalıdır. Bu durumda oluşturulan endeks serisine “inkeds” ismi verilir.

7. Hataya Sahip Trent Bileşeni

Trent bileşeni ilk elde edildiğinde hata değerlerini de içinde barındırı. Bu bileşeni oluşturmak için orjinal seriden endeks serisi çıkartılır ve trent serisi bulunur:

```
> trenthata
Time Series:
Start = c(1, 1)
End = c(8, 18)
Frequency = 18
        Series 1
  [1,]  241.60163
  [2,]  238.06989
  [3,]  231.80004
  [4,]  215.09369
  [5,]  162.69687
```

Şekil 15

Ancak istenilen şey hata değerlerini de barındıran trent serisi değildir. Bu nedenle trent serisini hatadan arındırmak gerekir.

8. Hatadan Arındırılmış Trent Bileşeni

Hataya sahip trent serisinden hareketle elde edilen bir seridir. “trenthata” serisine doğrusal regresyon uygulanması ile elde edilir.

```
> trent <- tslm(trenthata~trend)
> trent

Call:
tslm(formula = trenthata ~ trend)

Coefficients:
(Intercept)        trend
    377.408       -2.024
```

Şekil 16

Şekil 16 ‘da doğrusal regresyon modeil görülebilir. Ayrıca bu kısımda trent serisinin içinde bulun “fitted.values” orijinal serinin trent bileşenidir.

9. Tahmin Serisi

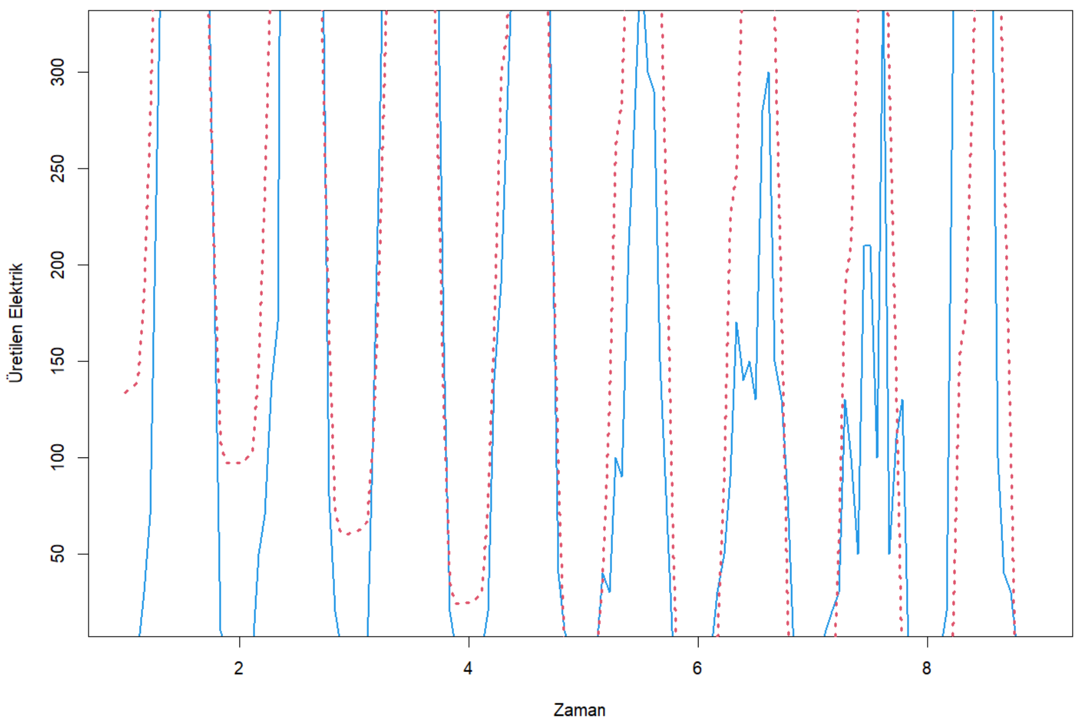
Tahmin serisini bulmak için indeks serisi ile trent serisinin “fitted.values” olarak adlandırılan orijinal serinin trent bileşenlerini toplamayla elde edilir. Tahmin serisinin ilk 5 gözlemini Şekil 17 ‘de görebiliriz.

```
> tahmin
Time Series:
Start = c(1, 1)
End = c(8, 18)
Frequency = 18
            [,1]
  [1,]  133.782359
  [2,]  135.290378
  [3,]  139.536492
  [4,]  184.219114
  [5,]  274.592212
```

Şekil 17

10. Hata Serisi

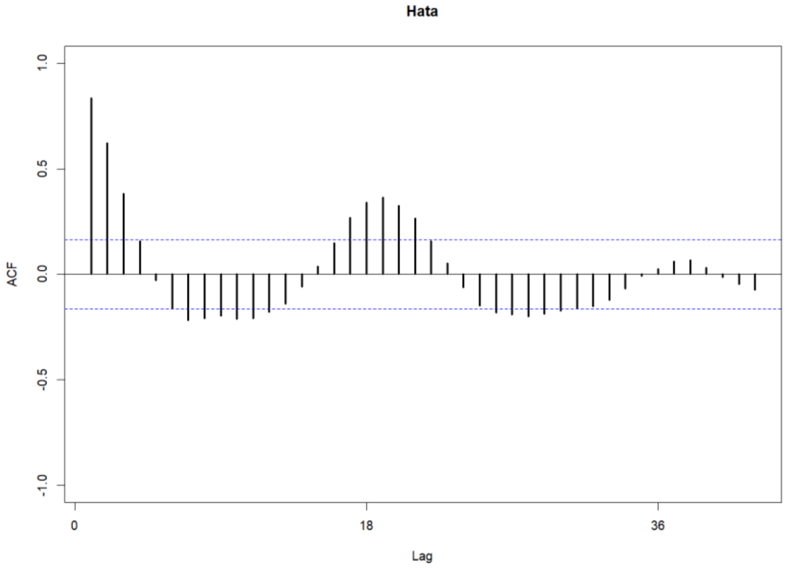
Oluşturduğumuz modelin hatalarını inceleme hususunda yapılması gerekilen şey hataların “akgürült” serisi olup olmadığına bakmaktır. Ama öncesinde modelimizin orijinal seri ile uyumuna bi göz atalım.



Şekil 18

Grafikten hareketle pek iç açıcı bir model olmadığını orijinal seri ile çok uymayan bir model olduğunu düşünüyorum ancak bu düşüncem modelin geçerliliğini etkileyemez. Hataları incelememiz gerekiyor.

Hataların ACF grafiğini çizdirilerim ve inceleyelim:



Şekil 18 ‘den hareketle yorumlarımızı yapmaya başlayabiliriz ki yapılabilecek pek yorum da yok. Çok açık bir şekilde görülüyor ki modelimizin hata serisi “akgürültü” serisi değil. Ancak bir umut belki modelimiz geçerlidir fikrine sarılıp Box-Ljung testi yapıp nihai sonucu almalıyız.

11. Modelin Geçerliliği

Modelin geçerlili konusunda şunu söyleyebiliriz ki hataların “akgürültü” olması bize modelin geçerli olup olmadığını söyleyecekti. Bir önceki kısımda ACF grafiği ile hataların “akgürültü” serisi olmadığına yönelik çıkarımda bulunmuştuk. Ancak Box-Ljung testi kesin sonucu verecektir:

H0: Hata serisi ile “akgürültü” serisi arasında fark yoktur.
H1: Hata serisi ile “akgürültü” serisi arasında fark vardır.

```
> Box.test(hata, lag = 42, type = "Ljung-Box")

        Box-Ljung test

data:  hata
X-squared = 376.29, df = 42, p-value < 2.2e-16
```

Şekil 19

Şekil 19 ‘dan hareketle üzülerek söylemek zorundayız ki yokluk hipotezini reddediyoruz. Yani %95 güvenle söylenebilir ki hatalar akgürültü serisi değildir. Modelimiz kullanılamayacak olan bir modeldir çöp oldu. Sağlık olsun…

Çarpımsal Ayrıştırma Yöntemi:

Çarpımsal ayrıştırma yönteminde ilk olarak serinin periyodu (s) germe sayısı alınarak merkezsel hareketli ortalama işlemini seriye uygularız.

```
> df_ts
Time Series:
Start = c(1, 1)
End = c(8, 18)
Frequency = 18
       Series 1
 [1,]        0
 [2,]        0
 [3,]        0
 [4,]       30
 [5,]       70
 [6,]      260
 [7,]      420
 [8,]      780
 [9,]      910
[10,]      960
[11,]      910
[12,]      790
[13,]      680
```

```
> MHO
Time Series:
Start = c(1, 1)
End = c(8, 18)
Frequency = 18
           [,1]
 [1,]        NA
 [2,]        NA
 [3,]        NA
 [4,]        NA
 [5,]        NA
 [6,]        NA
 [7,]        NA
 [8,]        NA
 [9,]        NA
[10,] 355.55556
[11,] 355.55556
[12,] 355.55556
[13,] 356.11111
```

```
> mevsimsel_bilesen
Time Series:
Start = c(1, 1)
End = c(8, 18)
Frequency = 18
            df_ts
 [1,]          NA
 [2,]          NA
 [3,]          NA
 [4,]          NA
 [5,]          NA
 [6,]          NA
 [7,]          NA
 [8,]          NA
 [9,]          NA
[10,] 2.70000000
[11,] 2.55937500
[12,] 2.22187500
[13,] 1.90951638
```

Orijinal seriyi merkezsel hareketli ortalama serisine bölerek hata terimini içeren mevsimsel bileşen serisini elde ederiz. Hata terimini yok edebilmek amacıyla her bir periyottaki dönemlerin ortalama değeri hesaplanır.

```
> donemort
 [1] 0.00000000 0.00000000 0.01664355 0.21639681 0.57536295
 [6] 1.29173289 1.36275552 1.76428306 2.46908182 2.65426238
[11] 2.50656242 2.74205171 1.57519291 1.07694928 0.47489070
[16] 0.02573615 0.00000000 0.00000000
```

Sonrasında dönem ortalamalarının da ortalamasını alırız ve dönem ortalamalarını, en son elde ettiğimiz genel dönem ortalamasına bölerek 1 buluruz. Bu da bize mevsimsel endeks serisini verir.

```
> donemort_son
[1] 1.041772
> mean(donemort/donemort_son) # 1 e eşit olmalı
[1] 1
> mevsimsel_endeks_serisi
 [1] 0.00000000 0.00000000 0.01597619 0.20771986 0.55229240
 [6] 1.23993779 1.30811260 1.69353993 2.37007810 2.54783341
[11] 2.40605583 2.63210263 1.51203180 1.03376643 0.45584882
[16] 0.02470420 0.00000000 0.00000000
```

Hata terimini de içeren trend serimizi, genel serimizi mevsimsel endeks serimize bölerek elde edebiliriz. (NaN'lar 0 yapılmıştır)

```
> trent_serisi
Time Series:
Start = c(1, 1)
End = c(8, 18)
Frequency = 18
         Series 1
  [1,]   0.00000
  [2,]   0.00000
  [3,]   0.00000
  [4,] 144.42528
  [5,] 126.74446
  [6,] 209.68794
  [7,] 321.07328
  [8,] 460.57373
  [9,] 383.95359
 [10,] 376.79073
 [11,] 378.21234
 [12,] 300.14027
 [13,] 449.72599
```

Hata terimini yok etmek için doğrusal regresyon yapalım. Doğrusal regresyondan elde ettiğimiz \`fitted.values\` değişkenleri, orijinal serinin trend bileşenini oluşturur.

```
> trent_bileseni

Call:
tslm(formula = trent_serisi ~ trend)

Coefficients:
(Intercept)        trend
    267.682       -1.332
```

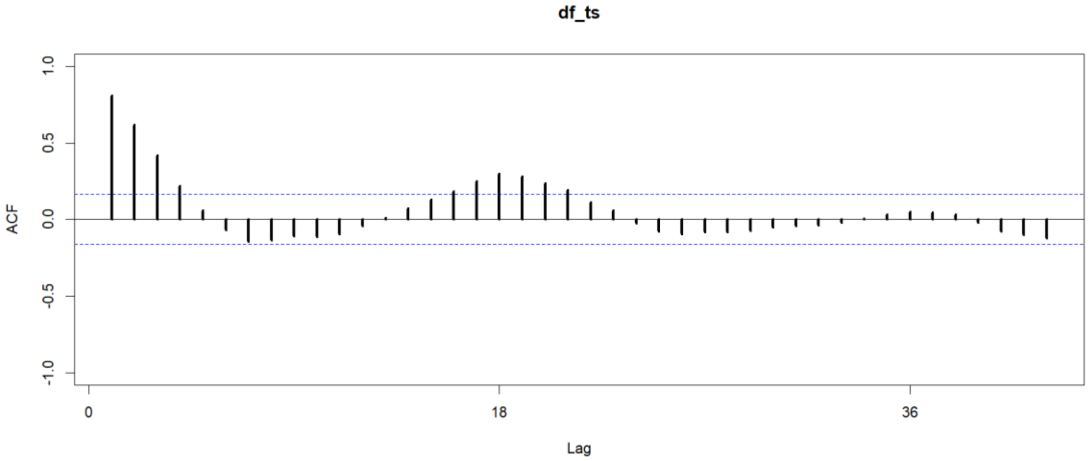
Orijinal serinin tahmin serisini bulmak için trend bileşeni serisi ile mevsimsel endeks serisi birbiri ile çarpılır.

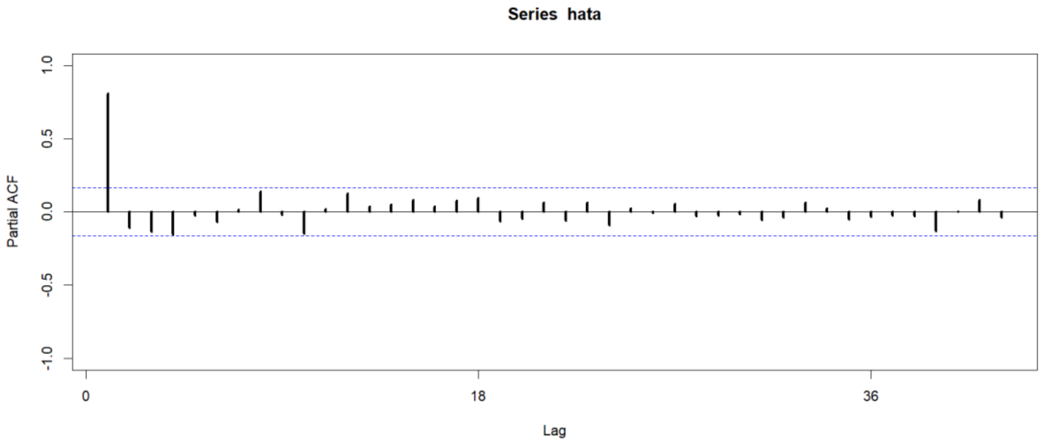
```
> tahmin_serisi
Time Series:
Start = c(1, 1)
End = c(8, 18)
Frequency = 18
         1          2          3          4          5
  0.000000   0.000000   4.212681  54.495829 144.159385
         6          7          8          9         10
321.996425 337.957546 435.278058 606.005594 648.060945
        11         12         13         14         15
608.792719 662.481037 378.552632 257.436534 112.911602
```

Orijinal seriden tahmin serisini çıkartarak hata serisini elde ederiz.

```
> hata
Time Series:
Start = c(1, 1)
End = c(8, 18)
Frequency = 18
                df_ts
  [1,]     0.0000000
  [2,]     0.0000000
  [3,]    -4.2126810
  [4,]   -24.4958291
  [5,]   -74.1593850
  [6,]   -61.9964251
  [7,]    82.0424537
  [8,]   344.7219419
  [9,]   303.9944061
 [10,]   311.9390553
 [11,]   301.2072807
 [12,]   127.5189631
 [13,]   301.4473675
```

Çarpımsal ayrıştırma yönteminin kullandığımız veri setinde etkili olup olmadığını hata serimizi inceleyerek anlarız. Çarpımsal ayrıştırma yönteminin geçerli olması için hata serimizin ak gürültü olması gerekmektedir.





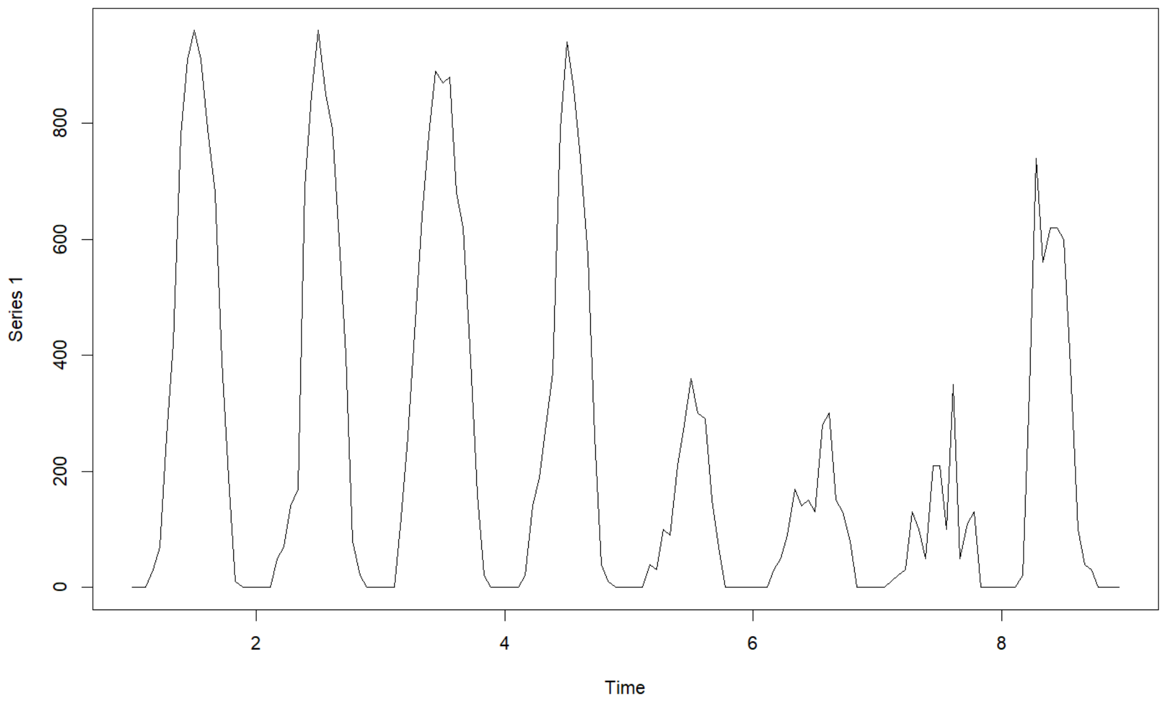
ACF grafiğinde görülebileceği üzere hata serimiz belirlenen sınırların dışına taştığı için ak gürültü olduğunu söyleyemeyiz. Ancak PACF grafiğinde sınırların dışında sadece bir tane var. Kesin sonuç için Box-Ljung testini yapalım.

```
> Box.test(hata, lag = 42, type = "Ljung-Box")

        Box-Ljung test

data:  hata
X-squared = 281.07, df = 42, p-value < 2.2e-16
```

Box-Ljung testi sonucuna göre p değerimiz çok düşük çıkmaktadır. Bundan dolayı çarpımsal ayrıştırma yönteminin bu veri üzerinde etkili olmayacağını söyleyebiliriz.

Holt-Winter Düzleştirme Yöntemi:

Zaman serimizin grafiğine baktığımızda mevsimsel dalgalanma olduğunu söyleyebiliriz. Bundan dolayı Holt düzleştirme yerine Winters düzleştirme yöntemini kullanabiliriz.



Ancak trendimiz olup olmadığına da bakmamız gerekir. Acf grafiğimizde ilk 4 gecikme sınırlar dışında olduğu için serimizin trende sahip olduğunu biliyoruz. Winters düzleştirme yöntemi ile devam edelim. Winters üstel düzleştirme yöntemi önce serinin ortalama düzeyine eğimine ve sonra mevsimsel bileşenine uygulanmaktadır.

```
> summary(winters1)
ETS(A,A,A)

Call:
 ets(y = df_ts, model = "AAA")

  Smoothing parameters:
    alpha = 0.9994
    beta  = 3e-04
    gamma = 2e-04

  Initial states:
    l = 373.9446
    b = 0.3433
    s = -244.1957 -246.305 -236.571 -146.5569 9.1296 163.9213
           323.6085 357.2771 393.1685 320.0845 187.2999 66.5692 4
2.8 -93.0293 -184.852 -232.5992 -237.9329 -241.8166

  sigma:  104.8107

     AIC     AICc      BIC
2077.600 2086.800 2145.906

Training set error measures:
                    ME     RMSE      MAE MPE MAPE      MASE
Training set -1.202264 96.47267 64.34656 NaN  Inf 0.6491846
                  ACF1
Training set 0.1480704
```

R çıktısına göre ortalama düzeyin başlangıç değeri 373.9446, eğimin başlangıç değeri 0.3433 mevsimsel terimin başlangıç değerleri ise sırasıyla çıktıdaki s değerleridir. Bu başlangıç değerlerinde optimal düzleştirme katsayıları α = 0.994, β=0.0003, γ= 0.0002 olarak bulunmuştur. HKO değeri ise 96.47267'dir.

Seriye bir de Çarpımsal Winter üstel düzleştirme yöntemi denenmelidir. Ancak verimizde 0 lar olduğu için R programı aşağıdaki çıktıda görüleceği üzere Çarpımsal Winter düzleştirme yöntemine izin vermemektedir.

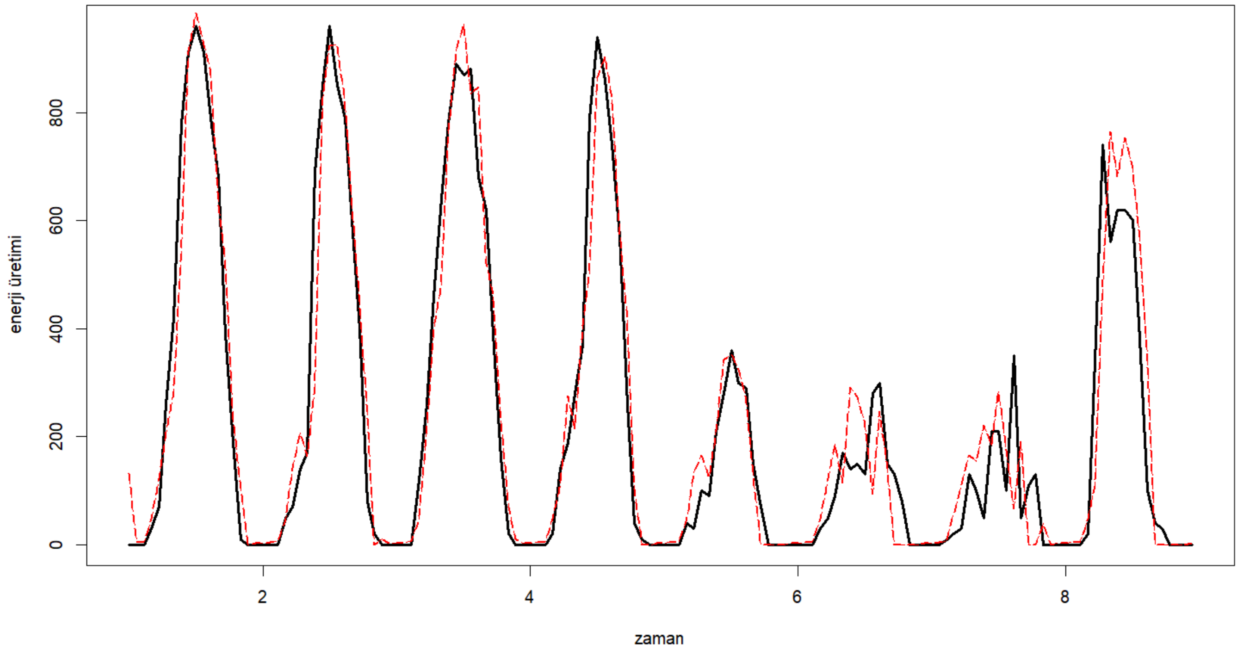
```
> winters2 <- ets(df_ts, model = "MAM")
Error in ets(df_ts, model = "MAM") :
  Inappropriate model for data with negative or zero values
```

Bunun için tüm gözlemlerimize küçük bir sayı ekleyip sonrasında bunu tahmin serisinden çıkartabiliriz.

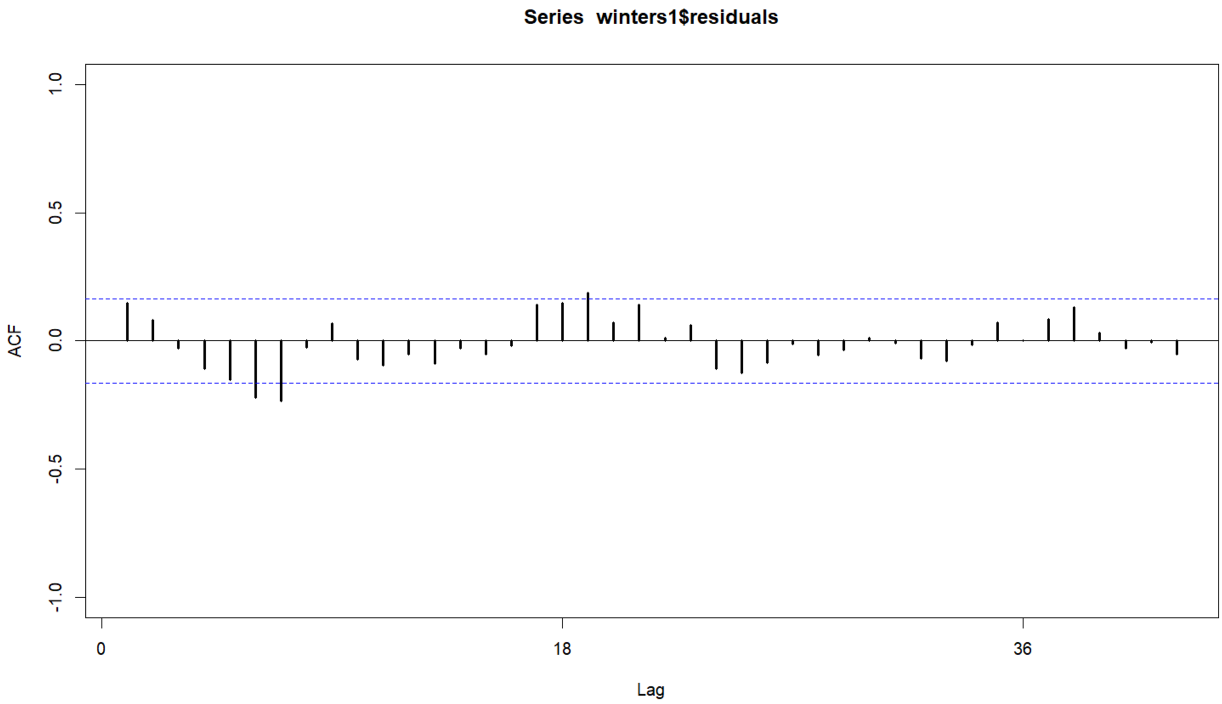
```
> winters2 <- ets(df_ts +0.0005, model = "MAM")
> summary(winters2)
ETS(M,Ad,M)

Call:
 ets(y = df_ts + 5e-04, model = "MAM")

  Smoothing parameters:
    alpha = 0.082
    beta  = 0.0012
    gamma = 0.0501
    phi   = 0.9785

  Initial states:
    l = 180.8062
    b = 20.5458
    s = 0 3e-04 0.0033 0.0263 0.0577 0.0846
           0.1453 0.1329 0.1393 0.2719 0.3762 0.073 0.0699 0.049
0.0221 0.0363 0 16.5119

  sigma:  1.559

     AIC     AICc      BIC
1873.965 1884.050 1945.241

Training set error measures:
                    ME     RMSE      MAE        MPE      MAPE
Training set -1776.852 9221.005 1932.895 -363373210 363373247
                 MASE        ACF1
Training set 19.50075 -0.03803342
```

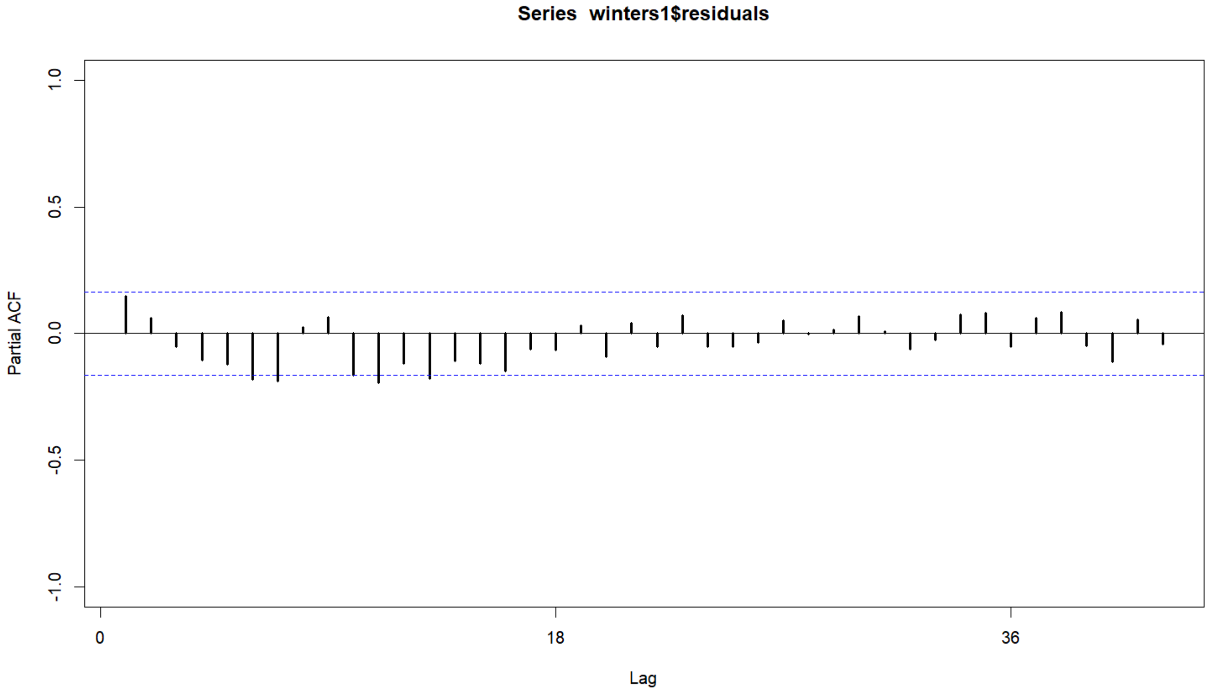
Winter’s çarpımsal üstel düzleştirme yönteminin HKO’su 9221.005 > 96.47267 olduğundan. Toplamsal winters üstel düzleştirme yöntemi serimize daha uygundur.

Toplamsal modelin tahmin ettiği verileri normal verilerimizle kıyaslayalım. Ancak tahmin verilerinde negatif değerler de var. Elektrik üretimi negatif olamayacağı için bu negatif verileri 0 yapmamız gerekir. Öngörü değerlerimiz yeni winters verimizideki ‘fitted’ değerleridir. Bunları normal verimiz ile aynı grafikte inceleyebiliriz.



Görüleceği üzere winters öngörü değerlerimiz normal serimiz ile paralellik göstermekte. Şimdi hataların Acf ve Pacf Grafiklerini inceleyelim.





Acf grafiğinde sınırı geçen gecikme değerleri mevcut. Serinin akgürültü olup olmadığı hakkında net bir yorumda bulunamıyoruz. Box-Ljung testi ile akgürültü olup olmadığı hakkında net bir cevap alabiliriz.

```
> Box.test(winters1$residuals, lag = 42, type = "Ljung-Box")

        Box-Ljung test

data:  winters1$residuals
X-squared = 64.049, df = 42, p-value = 0.01578
```

P değerimiz 0.05 den küçük olduğu için hatalar ak gürültü serisi değildir. Bu zaman serisi verisi için winters üstel düzleştirme anlamsızdır.

Mevsimsel Serilerde Regresyon Analizi

Bu kısımda mevsimselliğe sahip olan serimiz için tahminlerde bulunabilecek bir regresyon m odeli kuracağız. Bakalım her şeyin sonunda modelimiz anlamlı çıkıp da öngörüler yapabilecek miyiz?

Tüm bu sorular kafanızı meşgul mü ediyor, sabredin hepsi birazdan karşınızda…

Toplamsal Regresyon Modeli

Şimdi bir adet toplamsal regresyon modeli kuracağız eğer anlamlı çıkarsa değişken sayısı arttırılmış 2. Bir toplamsal regresyon modeli kuracağız ve bu şekilde ilerleyeceğiz. Ta ki regresyon modelimiz anlamsız çıkana kadar.

Örneğin 1. 2. Ve 3. modellerimiz anlamlı çıktı ve 4. Kez regresyon model ikurduk anlamsız çıktı o zamna 3. Regresyon modeline döneceğiz ve anlamlılığına bakacağız o da anlamlı çıkarsa bizim kullanılabilir modelimiz bu diyeceğiz.

1. t, sin1, cos1 değişkenlerini tanımlama:

Bu kısımda "t" serimizin başından sonuna kadar giden bir zaman göstergesi, bir index görevi görecek bir nevi,

Sin1 değişkenimizi ise;

$$\sin 1 = \sin\left(\frac{2 \times 3.1416 \times t}{periyot = 18}\right)$$

Şelinde tanımlıyoruz. Cosinus1 değişkenimiz de Sinus1 değişkeninden pek farklı olmayacak şekilde;

$$\cos 1 = \cos\left(\frac{2 \times 3.1416 \times t}{18}\right)$$

Şeklinde tanımlıyoruz. Tanımalamalarımızı yaptıktan sonra regresyon modelimizi kurabiliriz.

Bu aşamada orijinal serimizi bağımlı değişken yapıp diğer ürettiğimiz değişkenler (t, sin1, cos1) bağımsız olacak şekilde doğrusal regresyonumuzu kuruyoruz. Regresyon sonuçları Şekil x 'de görüldüğü gibidir.

```
> regresyon.model1 <-lm (y~t+sin1+cos1)
> summary(regresyon.model1)

Call:
lm(formula = y ~ t + sin1 + cos1)

Residuals:
    Min      1Q  Median      3Q     Max 
-364.26 -129.50  -23.45  139.45  548.73 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  363.049     29.770  12.195  < 2e-16 ***
t             -1.722      0.342  -5.036 1.38e-06 ***
sin1        -101.060     20.814  -4.855 3.06e-06 ***
cos1        -286.143     21.083 -13.572  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 181.2 on 146 degrees of freedom
Multiple R-squared:  0.6216,	Adjusted R-squared:  0.6138 
F-statistic: 79.95 on 3 and 146 DF,  p-value: < 2.2e-16
```

Regresyon modelimizi yorumlamadan önce hipotezlerimizi kuralım ve geçerliliğine bakalım:

H0: Modelin anlamlılığı = 0
H1: Modelin anlamlılığı ≠ 0

Yukarıdaki çıktımızda p-value < 0.05 e dayanarak;

%95 Güvenle söylenebilir ki kurulan model anlamlıdır. Ayrıca değişkenlerimizin p-değerleri de 0.05 'ten küçük olduğu için kurduğumuz bu 3 bağımsız değişken de anlamlıdır.

Modelimiz anlamlıysa değişken sayımızı arttırıp tekrar bi model kuralım ve onu deneyelim. Bu şekilde sürekli deneyeceğiz, ta ki anlamsız bir model kurana kadar.

2. Modele sin2 ve cos2 dahil etme:

Bu kısımda "sin2" değişkeni ile "cos2" değişkenini modele dahil ederken bir tanımlama yapmalıyız. Ancak yukarıdaki tanımlamadan bağımsız bir tanımlama değil. Tek bir tane ufak fark var.

$$\sin 2 = \sin\left(\frac{2 \times 3.1416 \times t \times 2}{periyot = 18}\right)$$

$$\cos 2 = \sin\left(\frac{2 \times 3.1416 \times t \times 2}{periyot = 18}\right)$$

Şeklinde tanımlamalarımızı yaptık. Şimdi deneyeceğimiz regresyon modelinde orjianl serimiz yine bağımlı ancak bu sefer bağımsızlarda "t", "sin1", "cos1", "sin2" ve "cos2" var:

```
> regresyon.model2 <- lm(y~t+sin1+cos1+sin2+cos2)
> summary(regresyon.model2)

Call:
lm(formula = y ~ t + sin1 + cos1 + sin2 + cos2)

Residuals:
    Min      1Q  Median      3Q     Max 
-437.87  -98.02  -12.73  116.09  617.18 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  360.6553    28.7117  12.561  < 2e-16 ***
t             -1.6914     0.3299  -5.127 9.32e-07 ***
sin1        -100.2886    20.0781  -4.995 1.68e-06 ***
cos1        -288.5647    20.3391 -14.188  < 2e-16 ***
sin2          57.6916    20.1763   2.859  0.00488 ** 
cos2          44.1387    20.2075   2.184  0.03056 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 174.7 on 144 degrees of freedom
Multiple R-squared:  0.6531,    Adjusted R-squared:  0.641 
F-statistic: 54.21 on 5 and 144 DF,  p-value: < 2.2e-16
```

Evet yine bir öncekinde olduğu gibi hipotezlerimizi kurup inceleyelim:

H0: Modelin anlamlılığı = 0
H1: Modelin anlamlılığı ≠ 0

Modelimizin çıktısında gördüğümüz üzre p-value < 0.05 olduğu için yokluk hipotezini reddedebiliriz.

Yani %95 güvenle söyleyebiliriz ki modelimiz anlamlı bir model. Ayrıca modelimizideki tüm değişkenlerin p değerleri 0.05 'ten küçük olduğu için tüm değişkenlerimizin modele katkısı anlamlıdır.

Bu modelimiz de anlamlı olduğu için 3. Regresyon modelini oluşturup onun anlamlılığını kontrol etmeliyiz.

3. Modele sin3 ve cos3 dahil etme:

Yine bir önceki maddede olduğu gibi, "sin3" ve "cos3" değişkenlerini oluşturup ekleyeceğiz.

$$\mathrm{sin3} = \sin\left(\frac{2 \times 3.1416 \times t \times 3}{periyot = 18}\right)$$

$$\mathrm{cos3} = \sin\left(\frac{2 \times 3.1416 \times t \times 3}{periyot = 18}\right)$$

Şeklinde tanımlamalarımızı yaptık.

Regresyon modelini kurduk ve çıktı aşağıdaki gibi sonuçlandı:

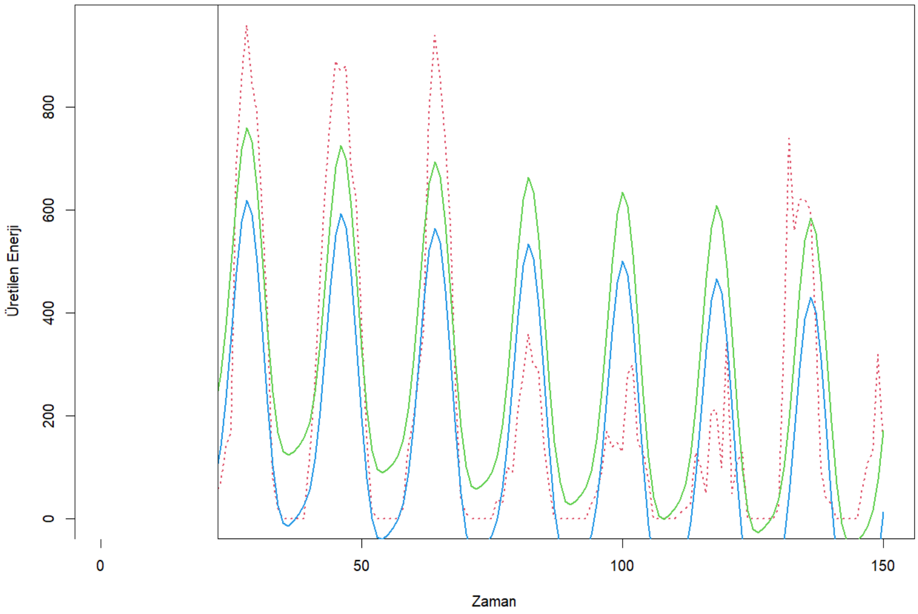
```
> regresyon.model3 <- lm(y~t+sin1+cos1+sin2+cos2+sin3+cos3)
> summary(regresyon.model3)

Call:
lm(formula = y ~ t + sin1 + cos1 + sin2 + cos2 + sin3 + cos3)

Residuals:
    Min      1Q  Median      3Q     Max 
-439.99 -104.47  -10.57  117.62  607.88 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  361.1591    28.8807  12.505  < 2e-16 ***
t             -1.6981     0.3319  -5.117 9.91e-07 ***
sin1        -100.3704    20.1893  -4.971 1.89e-06 ***
cos1        -288.2421    20.4568 -14.090  < 2e-16 ***
sin2          58.1424    20.2986   2.864  0.00481 ** 
cos2          44.3556    20.3285   2.182  0.03076 *  
sin3          -8.4328    20.3118  -0.415  0.67865    
cos3          10.4141    20.2985   0.513  0.60872    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 175.7 on 142 degrees of freedom
Multiple R-squared:  0.6541,	Adjusted R-squared:  0.6371 
F-statistic: 38.36 on 7 and 142 DF,  p-value: < 2.2e-16
```

Çıktımızı yorumlarken mavi ile işaretlenmiş kısıma bakacak olursak sonradan eklediğimiz "sin3" ve "cos3" değişkenlerinin modele katkısı anlamsız bulunmuş bu demektir ki 2. Regresyon modelimize dönüp onu kullanmalıyız.

Şimdilik her şey güzel, 2. Regresyon modelimiz gayet anlamlı üstelik tüm değişkenlerin modele katkısı anlamlı. Ancak bu modeli kullanabilmemiz için modelimizin hatalarının akgürültü serisi de olması gerekiyor, hadi gelin bunu inceleyelim.

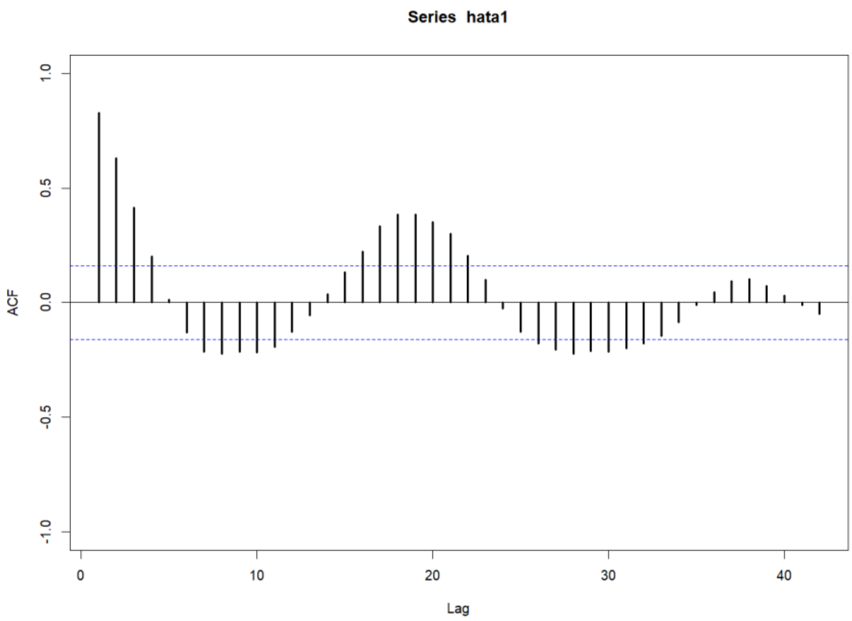
4. Modelin Geçerliliği:

Modelimizin geçerli olması için regresyon modelinin anlamlı olması gerekirdi ki bu şartı sağladı ancak hatalar da “akgürültü” serisi olmalı. Hatalara bakmadan bi orijinal seri ile uyumuna bakalım mı:



Nispeten uyumlu bir grafik ancak bakalım hataları da akgürültü serisi mi?

H0: Hatalar ile “akgürültü” serisi arasında fark yoktur.
H1: Hatalar ile “akgürültü” serisi arasında fark vardır.



Hataların ACF grafiğine göre akgürültü serisi değil gibi duruyor bir de Box-Ljung şahsına soralım bakalım bu şahıs bize ne diyecek. Zira Box-Ljung ‘ın söyleyecekleri model için nihai karar olacaktır.

Bakalım Box-Ljung 2. Regresyon modelinin kalemini kıracak mı?

```
> Box.test(hata1, lag = 42, type = "Ljung")

        Box-Ljung test

data:  hata1
X-squared = 438.83, df = 42, p-value < 2.2e-16
```

Maalesef ki Box-Ljung 2. Regreson modelimizin kalemini kırıyor ve diyor k;

%95 güvenle 2. Regresyon modeli kullanılabilir bir model değildir, tez kellesi vurula bu modelin, hemen çöpe atın.

Çarpımsal Regresyon Modeli

Bu tip regresyon modellerinde toplamsalda olduğu gibi bir süreç izleyeceğiz. Ufak farklar var formüllerde ve birkaç yerde. Hadi çarpımsal doğrusal regresyon modeli kurup test edelim.

Bakalım bu bölümde karar mercihi Box-Ljung çarpımsal doğrusal regresyon modelinin akıbeti hakkında neler söyleyek? Ya da regresyon modeli hiç anlamlı olamayıp Box-Ljung huzuruna bile mi çıkamayacak? Hepsi az sonra …

1. t, sin1, cos1 değişkenlerini tanımlama:

Bu kısımda toplamsaldan farklı bir formülasyon izleyeceğiz:

$$\sin 1 = \mathrm{t} \times \sin\left(\frac{2 \times 3.1416 \times t}{18}\right)$$

$$\cos 1 = \mathrm{t} \times \cos\left(\frac{2 \times 3.1416 \times t}{18}\right)$$

Şeklinde tanımlanır.

Tanımlamalarımızı yapıp 1. Çarpımsal regresyon modelimizi kurduk ve sonucunda elde ettiğimiz çıktı aşağıdaki gibi oldu:

```
> regresyon.model4 <- lm(y~t+s1+c1)
> summary(regresyon.model4)

Call:
lm(formula = y ~ t + s1 + c1)

Residuals:
    Min      1Q  Median      3Q     Max 
-353.76 -182.49  -66.89  139.46  596.12 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 357.7002    40.1264   8.914 1.83e-15 ***
t            -1.6625     0.4621  -3.598 0.000438 ***
s1           -0.3853     0.3201  -1.204 0.230635    
c1           -2.2867     0.3296  -6.938 1.20e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 244.2 on 146 degrees of freedom
Multiple R-squared:  0.3128,    Adjusted R-squared:  0.2986 
F-statistic: 22.15 on 3 and 146 DF,  p-value: 7.064e-12
```

Çıktımızı yorumlamadan önce hipotezlerimizi kuralım ve modelin anlamlılığına bakalım:

H0: Modelin anlamlılığı = 0
H1: Modelin anlamlılığı ≠ 0

P-value < 0.05 için %95 güven düzeyiyle söylenebilir ki kurduğumuz regresyon modeli anlamlıdır.

Ancak değişkenlerin significant değerlerine baktığımızda "a1" (sin1) değişkeninin modele katkısının anlamsız olduğunu görüyoruz bu nedenle modelimizin bağımsızlarını "t" ve "cos1" olarak alabilriz.

Bu konuda 2. Bir regresyon modeli kurup "sin2" ve "cos2" değişkenlerinin eklenmiş haliyle de test etmek ve modelden "sin1" değişkenini çıkarıp 1. modelde kalma fikirleri arasında kaldım. Ancak tüm değişkenlerimiz anlamlı olmadığından bu kurduğumuz anlamlı modelde kalma fikri daha mantıklı geldi.

2. Modelin Geçerliliği:

Modelimizin geçerli olabilmesi için modelin anlamlı olması şartının yanında hatalarının birbiri ile ilişkisiz olması ve aynı zamanda hatalarının "akgürültü" olması gerektiği şartı da var.

O zaman şartların sağlanıp sağlanmadığının testine başlayalım:

2.1 Modelin Anlamlılığı:

Bunu tekrardan test etmeyeceğim zira yukarıda bu şartı sağladığını görmüştük.

2.2 Hataların Bağımsızlığı:

Modelin hatalarının bağımsız olup olmadığını araştıralm.

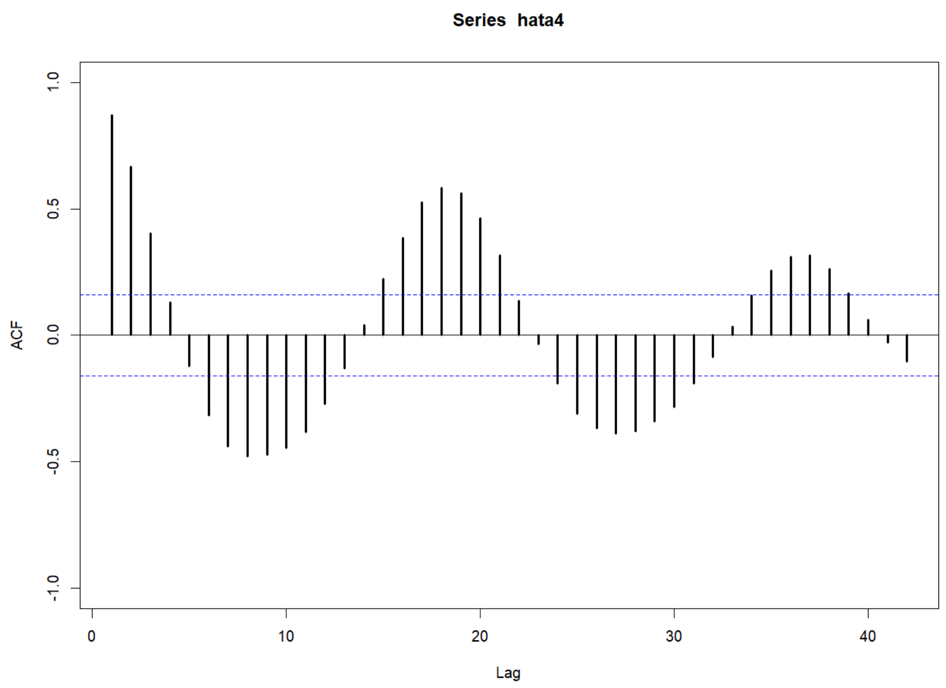
H0: Kurulan regresyon modeli hatalarının arasında ilişki yoktur.
H0: Kurulan regresyon modeli hatalarının arasında ilişki vardır.

Bu kurduğumuz hipotezleri Durbin-Watson testi ile test edelim:

```
> dwtest(y~t+c1)

        Durbin-Watson test

data:  y ~ t + c1
DW = 0.24408, p-value < 2.2e-16
alternative hypothesis: true autocorrelation is greater than 0
```

Durbin-Watson testi sonucunda P-value < 0.05 için %95 güvenle söyleyebiliriz ki kurulan regresyon modelinin hataları arasında ilişki vardır.

Yani aslında bu modelimiz aslında hataların bağımsızlığı şartını bile sağlayamıyor ama dur bakalım hataları "akgürültü" serisi miymiş?



Yukarıdaki ACF grafiğinden çok net anlaşılacağı üzere hatalarımız “akgürültü” serisi değil. Bu yargıyı Box-Ljung da söylesin de tastiklenmiş olalım:

H0: Hata serisi ile “akgürültü” serisi arasında fark yoktur.
H1: Hata serisi ile “akgürültü” serisi arasında fark vardır.

```
> Box.test(hata4, lag = 42, type = "Ljung")

	Box-Ljung test

data:  hata4
X-squared = 886.85, df = 42, p-value < 2.2e-16
```

Box-Ljung testi sonucunda P-value < 0.05 için %95 güvenle söylenebilir ki hatalar “akgürültü” serisi değildir.

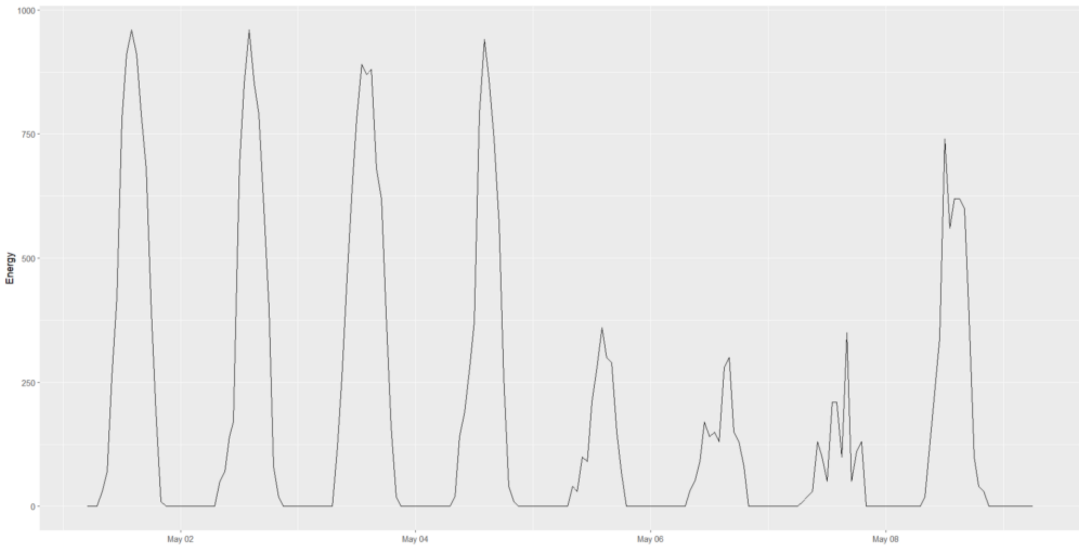
Yani buradan yapacağımız çıkarım şudur ki, kurduğumuz bu çarpımsal regresyon modeli kullanılamayacaktır, doğruca çöpe atılıyor.

Evet arkadaşlar dememiz o ki regresyon modeliyseniz eğer anlamlı olmanızın yanında hatalarınızın ilişkisi olmamalı, yani hatalarınız bağımsız olmayı. Ayriyeten Box-Ljung huzuruna çıkmalı ve hatalarınızın “akgürültü" serisi olduğunu doğrulatmalısınız. İşte ancak o zaman kullanışlı bir model olursunuz ve sizinle öngörüler yapabiliriz.

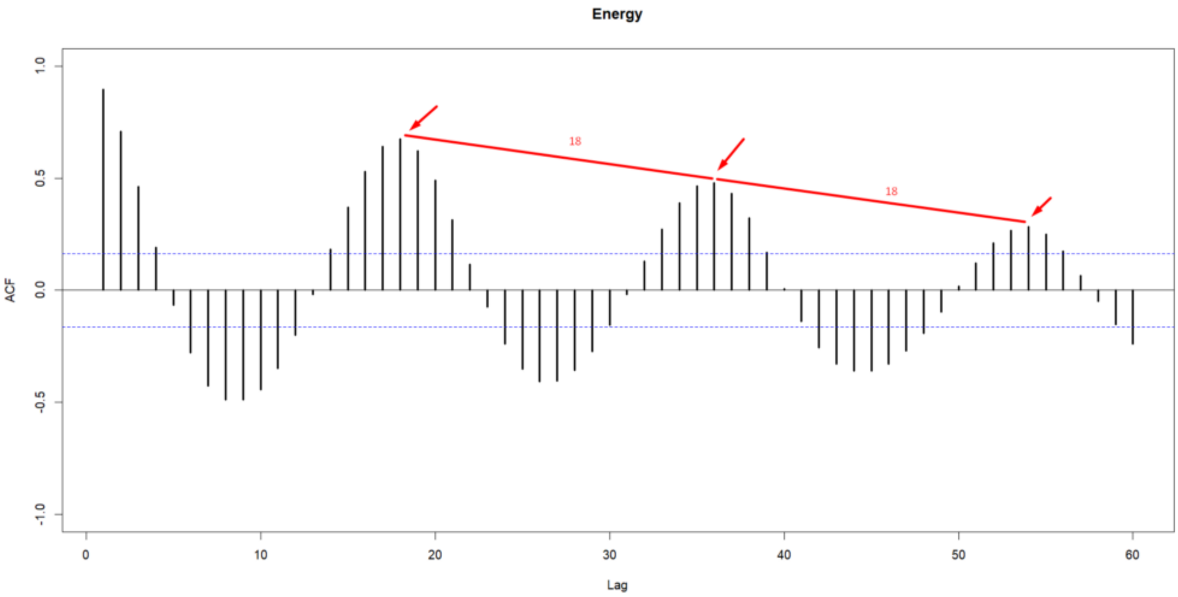
Box-Jenkins Modelleri

Box-Jenkins modelleme, bir tahmini modelleme yöntemidir ve genellikle zaman serilerinde tahmin etme amacıyla kullanılır. Bu yöntem, Arthur Box ve George Jenkins tarafından 1970'lerde geliştirilmiştir ve günümüzde hala zaman serisi tahmini modellerinin inşasında yaygın olarak kullanılmaktadır.

ARIMA modelleri kurmadan önce verimizi inceleyelim:



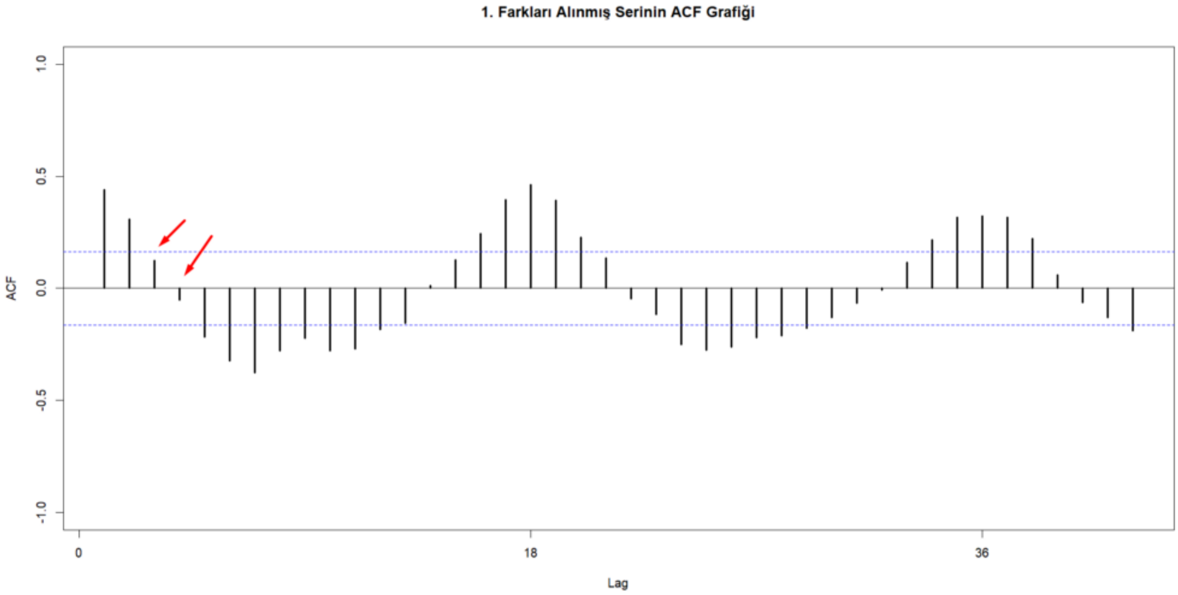
Yukarıdaki zaman serisi grafiğine baktığımızda trendi olmayan ancak mevsimselliği olan bir zaman serisi gibi duruyor ancak buna bu şekilde karar veremeyiz. ACF grafiğine bakıp trentve mevsimsellik var mı bunları bulmalıyız.



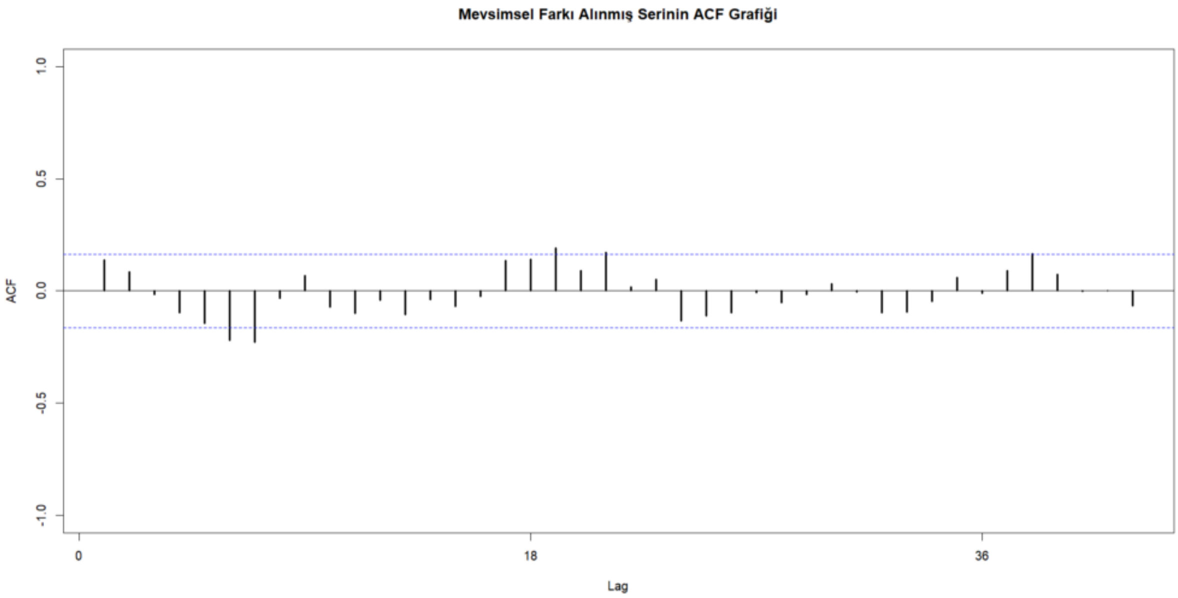
Yukarıdaki ACF grafiğinin ilk 4 gecikmesi güven aralığı dışına çıktığından dolayı trent vardır diyebiliriz. Ayrıca her 18 gecikmede bir artış gözlemlendiğinden mevsimsellik de var deriz.

ARIMA modeli kurmak için serimizi trentten ve mevsimsellikten arındırmalıyız:

- Zaman serisine 1. Fark işlemi uyguluyoruz ve sonucunda trentten arındırılmış bir seri elde ediyoruz. O zaman ARIMA modelindeki "d" parametremiz 1 olmuş oluyor.

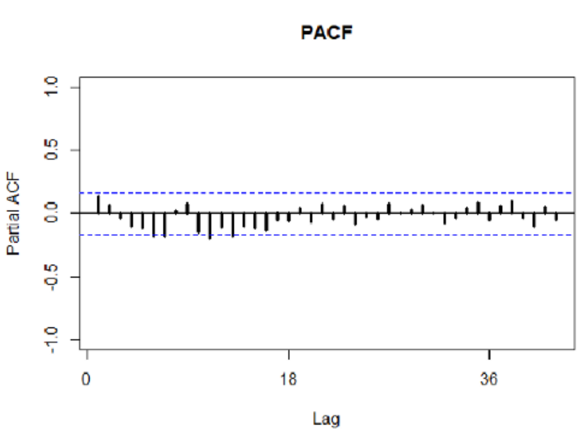


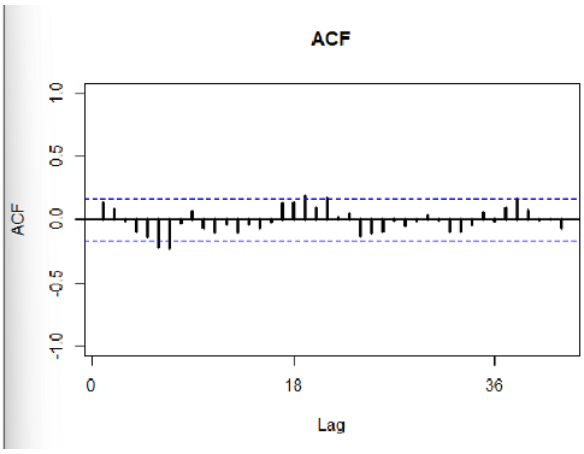
- Mevsimsellikten arındırmak için serinin 1 kez mevsimsel farkını alıyoruz ve mevsimsellikten de arındırmış oluyoruz. O zaman ARIMA modelimizin "D" parametresi de 1 olmuş oluyor.



Şimdi sırada ARIMA modelimizin p,q,P ve Q parametrelerini bulmakta.

Öncelikle “p” ve “q” parametrelerinden başlayalım:



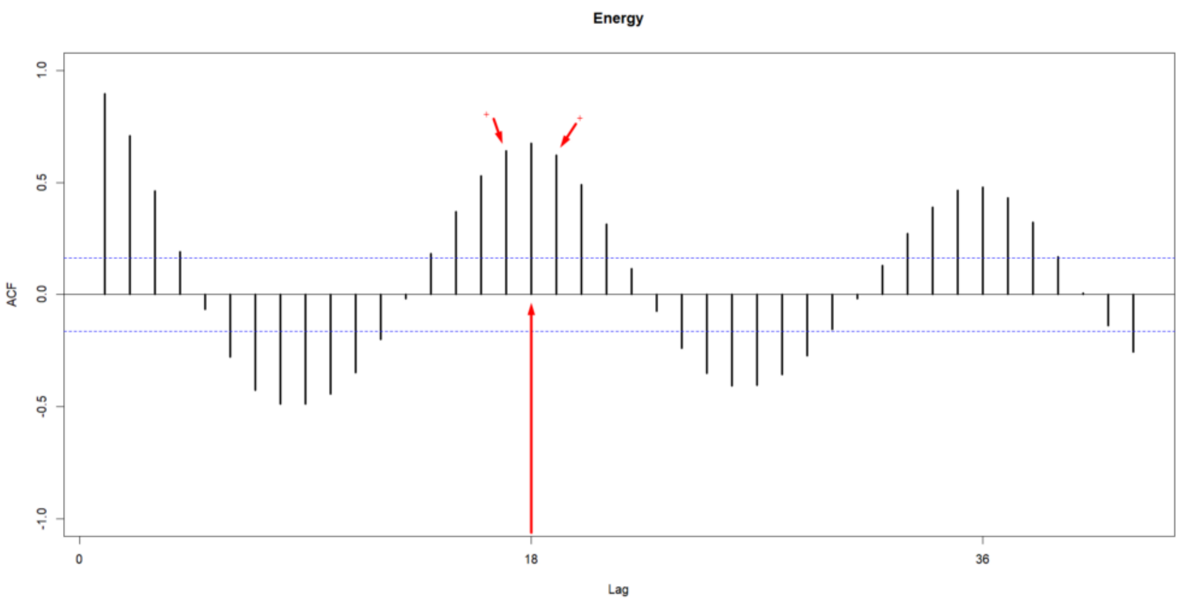


ACF ve PACF grafiklerini yanyana getirip baktığımızda ikisinin de aynı hızda azaldığını söylüyoruz. Bu nedenle modelimiz ARMA(p,q) modeli oluyor.

“P” değerini PACF grafiğinin ilk 3 gecikmesinin kaç tanesinin CI dışına çıkmış olduğuna bakarak buluyoruz ki 0 (sıfır) olarak belirliyoruz.

“q” değerini ise ACF grafiğinin ilk 3 gecikmesinin kaç tanesinin CI dışına çıkmış olduğuna bakarak buluyoruz ki bu değişken de 0 (sıfır) oluyor.

Şimdi sırada P ve Q parametreleri var. Bu parametreleri 1’den başlayarak 3’e kadar ya da 3’ten başlayarak 1’e kadar deneyeceğiz. Bu başlama sırasını belirlemek için ise orijinal serimizin ACF grafiğinden belirlenen periyodun bir önceki ve bir sonraki değerlerinin aynı yönlü olup olmadığına bakıyoruz.



Yukarıdaki grafikte 17. Ve 19. Gözlem aynı yönlü olduğundan dolayı 1’den başlayıp 3’e kadar deneme yapacağız.

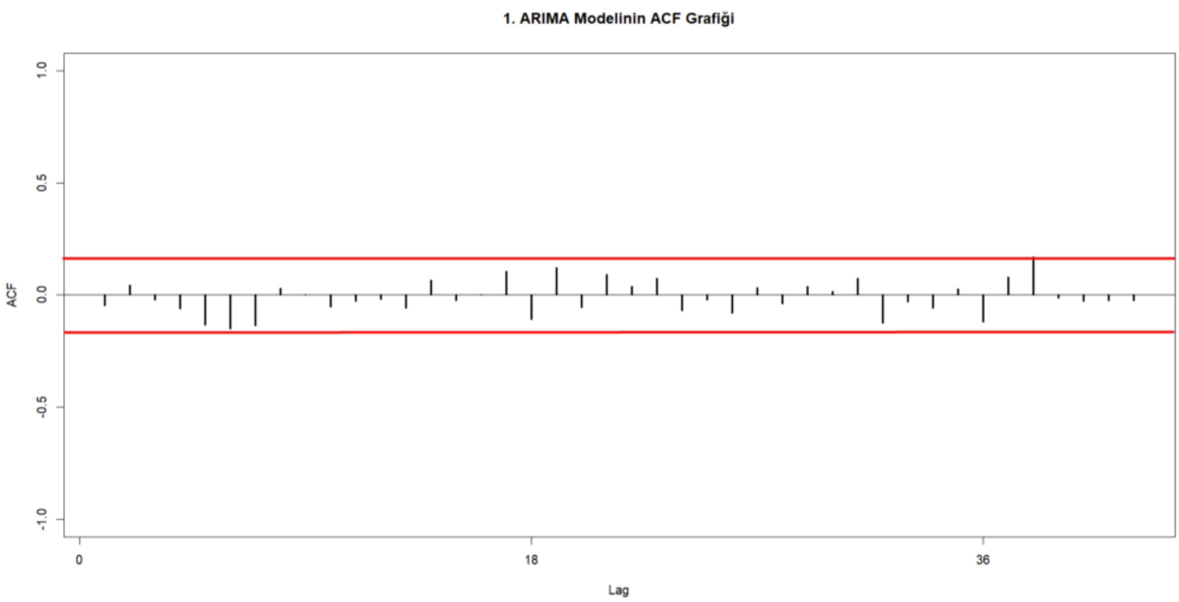
Bu dene kısmında tek tek her modelin çıktısını göstermeyeceğim. Örnek olarak ilk modeli çıktıları ile yorumlayıp daha sonra denediğim tüm modelleri eğer anlamlılarsa BIC ve RMSE değerleri ile yazacağım.

1. Model : ARIMA(0,1,0)(1,1,0)

```
> coeftest(ARIMA1)

z test of coefficients:

      Estimate Std. Error z value  Pr(>|z|)    
sar1 -0.418521   0.092636 -4.5179 6.245e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

H0: Model anlamlı değilir.
H1: Model anlamlıdır.

Yukarıdaki çıktıdan modelin anlamlı olduğunu %95 güvenle söyleyebiliriz.



ACF grafiğinden ise modelin hatalarının akgürültü serisi olup olmadığına bakıyoruz. Ancak sanki 38. Gözlem hafiften CI dışına çıkmış gibi. Bunu net bir şekilde doğrulamak için Box-Ljung testi yapıyoruz.

```
> Box.test(ARIMA1$residuals, lag = 42, type = "Ljung")

        Box-Ljung test

data:  ARIMA1$residuals
X-squared = 39.142, df = 42, p-value = 0.5971
```

H0: Modelin hata serisi ile akgürültü serisi arasındaki fark = 0
H1: Modelin hataları akgürültü serisi değildir.

Box-Ljung testi sonucunda ise %95 güvenle söylenebilir ki modelin hata serisi ile akgürültü serisi arasında fark yoktur. Yani hatalar akgürültü serisidir.

Tamamdır, bu modelimiz anlamlı olma ve ve hata serisinin akgürültü serisi olma şartlarını sağladı. Artık gelecekte diğer modeller ile karşılaştırılabilmesi için BIC ve RMSE değerlerine bakalım:

```
> summary(ARIMA1)
Series: df
ARIMA(0,1,0)(1,1,0)[18]

Coefficients:
         sar1
      -0.4185
s.e.   0.0926

sigma^2 = 12142:  log likelihood = -766.38
AIC=1536.75   AICc=1536.85   BIC=1542.41

Training set error measures:
                      ME     RMSE      MAE MPE MAPE      MASE
Training set 0.005061814 102.2535 62.54135 NaN  Inf 0.6309721
                    ACF1
Training set -0.04731722
```

Yukarıdaki çıktıdan 1. ARIMA modelinin BIC ve RMSE değerleri sırasıyla 1542.41 ile 102.2535 olarak elde edilmiştir.

Bu model anlamlı olduğu için ARIMA(0,1,0)(2,1,0) modelini de denemeliyiz.

Ben bu modeli de denedim. Hem model anlamlı çıktı hem de hataları akgürültü serisi çıktı. O modelin BIC ve RMSE değerlerini şimdi yazmayacağım. Her model için tek tek yazıp kalabalık etsin istemiyorum. En sonda anlamlı çıkan modellerin hepsini yazıp karşılaştıracağız.

Bu 2. Model de anlamlı çıkınca ARIMA(0,1,0)(3,1,0) modelini denedim ancak o model anlamlı bile çıkamadı bu nedenle kıyaslamaya koyamayacağız.

Denediğim tüm modelleri aşağıdaki tabloda bulabilirsiniz.

MODEL	ANLAMLILIK	BIC	RMSE
ARIMA(0,1,0)(1,1,0)	Anlamlı ve Hataları Akgürültü	1542.41	102.2535
ARIMA(0,1,0)(2,1,0)	Anlamlı ve Hataları Akgürültü	1542.19	99.38836
ARIMA(0,1,0)(3,1,0)	Anlamsız	X	X
ARIMA(0,1,0)(0,1,1)	Anlamlı ve Hataları Akgürültü	1538.64	100.1105
ARIMA(0,1,0)(0,1,2)	Anlamsız	X	X

Yukarıdaki tablodan BIC (Bayesian Bilgi Kriterleri) ve RMSE (Hata Kareler Ortalaması Karekökü) değerleri en düşük olan modeli kendimiz için asıl model olarak belirlemek en doğrusu olacaktır ki bu durumda 2. Modeli asıl modelimiz kabul ediyoruz ve öngörülerimizi oluştururken bu modeli kullanacağız.

Seçilen Model -> **ARIMA(0,1,0)(2,1,0)**

Bu modeli kullanarak 5 dönemlik öngörü yapalım:

```
Time Series:
Start = c(9, 1)
End = c(9, 5)
Frequency = 18
[1] -1.262177e-29 -1.454498e-16  2.592970e+00  2.247717e+01  1.872869e+02
```

Yukarıdaki çıktıdan hareketle kırmızı şerit arasındaki değerler önümüzdeki 5 döneme ilişkin öngörülerdir.

Kullanılan Kodlar:

Temel Zaman Serisi Analizi:

```
library(tidyverse)
library(readxl)
library(zoo)
library(xts)
library(forecast )
enData <- read_xlsx('./ikitelli-gune-enerjisi-santrali-elektrik-
uretim-miktarlar.xlsx')
View(enData)

data.xts <- xts(enData$`Üretim (kWh)`, as.POSIXct(enData$Tarih)-1)
Sum.xts <- period.apply(data.xts, INDEX = endpoints(data.xts,
"hours", 1), FUN = sum)
Sum.xts <- Sum.xts[1:150,]

# hourly.apply(data.xts, sum)
View(Sum.xts)
typeof(enData$Tarih)

Acf(diff(diff(Sum.xts), 18), lag.max = 42, lwd =3)
nrow(Sum.xts)

ma18 <- ma(Sum.xts, order = 18, centre = T)

ts.plot(window(Sum.xts), xlab = "Tarih", ylab = "Üretilen Enerji",
lty = 1, col="purple", lwd=2)
par(new=T)
lines(ma18, lty=10, col="orange", lwd=3)


legend("topright", c(expression(paste(Sum.xts)),
expression(paste(MA(Sum.xts, 18)))))
ts.plot(Sum.xts)
```

Toplamsal Ayrıştırma:

```
# install.packages("fpp")
# install.packages("stats")
rm(list = ls())

# libs
library(fpp)
library(stats)
library(xts)
library(ggplot2)
library(forecast)
```

```
# set directory
setwd("./Toplamsal_Ayristirma")

# load data
df <- read.csv("./dataset.csv")
names(df) <- c("date", "electric")
head(df);tail(df)

# tarih-saat belirleme
df[,1] <- as.POSIXct(df[,1], format = "%Y-%m-%d %H:%M:%S")
df_xts <- as.xts(df[,-1], order.by = df[,1])
class(df_xts) ; head(df_xts)

# grafikler
ts_plot <- ts.plot(df_xts, xlab="Zaman", ylab="Üretilen Enerji")

acf_plot <- Acf(df_xts,lag.max = 42,  ylim=c(-1,1), lwd=3)
pacf_plot <- Pacf(df_xts,lag.max = 42, ylim=c(-1,1), lwd=3)

# zaman serisi
df_ts <- ts(df_xts[1:144], frequency = 18)

# trent bileşeni
df_trent <- tslm(df_ts ~ trend)

# Periyoda sahip mevsimsel bilesen serisi
periyot <- df_ts - df_trent[['fitted.values']]
Acf(periyot,lag.max = 42,  ylim=c(-1,1), lwd=3) # periyot=18
belirledik

# Merkezsel Hareketli ortlama MHO
MHO <- ma(df_ts, order = 18, centre = TRUE)

# Mevsim bileşeni
mevsim <- df_ts-MHO

# Mevsimlerin ortalaması
donemort <- t(matrix(data=mevsim, nrow = 18, ncol=8))
colMeans(donemort, na.rm = T)
sum(colMeans(donemort, na.rm = T))
mean(colMeans(donemort, na.rm = T))

# mevsim endeks değerleri
endeks <- colMeans(donemort, na.rm = T) - mean(colMeans(donemort,
na.rm = T))

# seri boyunda endeks değerlerini yerine koyma
indeks <-  matrix(data = endeks, nrow = 144)

# hatadan arındırılmamış trent bileşeni
```

```
trenthata <- df_ts - indeks

# seriyi hatadan arındırmak için doğrusal regresyon modeli
kullanalım
trent <- tslm(trenthata~trend)
# trent serisinin #fitted.values -> orjinal serinin trent
bileşenidir.

# tahmin serisini bulalim: (mevsimsel endeks + saf trent serisi)
tahmin <- indeks + trent$fitted.values

#hata serisini bulalim:
hata <- df_ts - indeks - trent$fitted.values

### MODELIN GUVENILIRLIGI ###
#orijinal seri ile tahmin serisinin uyumu
plot(window(df_ts), xlab = "Zaman", ylab="Üretilen Elektrik",
      lty=1, col=4, lwd=2, ylim=c(19,320))
lines(window(tahmin), lty = 3, col = 2, lwd = 3)
legend("topleft", lwd = c(2,2), lty = c(1,3),
      cex = 0.6, col = c(4,2))

# hatalar akgurultu serisi mi?
Acf(hata,main="Hata", lag.max = 42,  ylim=c(-1,1), lwd=3)
Pacf(hata,main="Hata",lag.max = 42, ylim=c(-1,1), lwd=3)
# ACF grafiğine göre hatalar akgürültü serisi değil
# Ancak nihai kararı Box-Ljung testi verir:
# H0: hatalar serisi ile ak-gürültü serisi arasında fark yoktur.
# Hs: hatalar serisi ile ak-gürültü serisi arasında fark yoktur.
Box.test(hata, lag = 42, type = "Ljung-Box")
# test sonucuna göre H0 reddedilir. Yani hatalar akgürültü serisi
olmadığı için model kullanılamaz.
```

Çarpımsal Ayrıştırma:

```
# libs
library(fpp)
library(stats)
library(xts)
library(ggplot2)
library(forecast)

# load data
df <- read.csv("./Toplamsal_Ayristirma/dataset.csv")
names(df) <- c("date", "electric")


# tarih-saat belirleme
df[,1] <- as.POSIXct(df[,1], format = "%Y-%m-%d %H:%M:%S")
df_xts <- as.xts(df[,-1], order.by = df[,1])
class(df_xts) ; head(df_xts)
```

```r
#periyodu 18 olan zaman serisi
df_ts <- ts(df_xts[1:144], frequency = 18)

#germe sayısı s = 18 olan merkezsel hareketli ortalama işlemi seriye
uygulanır
MHO <- ma(df_ts, order = 18, centre = TRUE)

#orijinal seriyi mho ya bölerek serinin mevsimsel bileşenlerini
bulmaya çalışalım. (hata terimini de içeren)

mevsimsel_bilesen <- df_ts/MHO

#hata terimini yok edebilmek amacıyla her periyottaki ortalama değer
hesaplanır.
donemort <- t(matrix(data=mevsimsel_bilesen, nrow = 18, ncol=8))
donemort <- colMeans(donemort, na.rm = T)
donemort_son <- mean(donemort)

mean(donemort/donemort_son) # 1 e eşit olmalı

mevsimsel_endeks_serisi <- donemort/donemort_son
##https://stackoverflow.com/questions/18142117/how-to-replace-nan-
value-with-zero-in-a-huge-data-frame
trent_serisi <- df_ts/mevsimsel_endeks_serisi
##nan ları 0 yapalım
trent_serisi[is.nan(trent_serisi)] <- 0

#doğrusal regresyon
trent_bileseni<- tslm(trent_serisi~trend)

#tahmin serisini bulalım
tahmin_serisi <- trent_bileseni$fitted.values *
mevsimsel_endeks_serisi
#hata serisi
hata <- df_ts - tahmin_serisi

#HATA Serisinin Analizi
#Hata serisi akgürültü olmalıdır.
View(hata)
Acf(hata, lag.max = 42, lwd=3, ylim=c(-1,1))
Pacf(hata, lag.max = 42, lwd=3, ylim=c(-1,1))

Box.test(hata, lag = 42, type = "Ljung-Box")
##Hatalar akgürültü değildir. Demekki serimiz çarpımsal ayrıştırmaya
uygun bir seri değildir.
```

Holt-Winter Düzleştirme Yöntemi:

```r
library(forecast)
library(xts)
```

```
df <- read.csv('./Datasets/dataset.csv')

# tarih-saat belirleme
df[,1] <- as.POSIXct(df[,1], format = "%Y-%m-%d %H:%M:%S")
df_xts <- as.xts(df[,-1], order.by = df[,1])
class(df_xts) ; head(df_xts)

#periyodu 18 olan zaman serisi
df_ts <- ts(df_xts[1:144], frequency = 18)

#zaman serisi grafiği
plot.ts(df_ts)

#toplamsal
winters1 <- ets(df_ts, model = "AAA")
summary(winters1)

# hw1 <- HoltWinters(df_ts)
# hw2<- HoltWinters(df_ts, alpha = 0.99, beta = 0.0003, gamma =
0.0002)
# plot(df_ts, ylab="enerji üretimi", xlab="zaman", lwd=3)
# lines(hw1$fitted[,1], lty=2, col = "blue", lwd=2)
# lines(hw2$fitted[,1], lty=2, col = "red", lwd=2)

winters1$fitted[winters1$fitted <0 ] <- 0 #0 dan küçük değerlerin 0
a eşitlenmesi.
plot(df_ts, ylab="enerji üretimi", xlab="zaman", lwd=3)
lines(winters1$fitted[,1], lty=2, col="red", lwd=2)
winters1$fitted

#Acf Pacf grafikleri
Acf(winters1$residuals, lag.max = 42, lwd=3, ylim=c(-1,1))
Pacf(winters1$residuals, lag.max = 42, lwd=3, ylim=c(-1,1))
Box.test(winters1$residuals, lag = 42, type = "Ljung-Box")
```

Mevsimsel Regresyon Analizi:

```
# clear environment
rm(list=ls())

# libs
library(fpp)
library(forecast)
library(haven)

raw_df <- read.csv("./Datasets/dataset.csv")
raw_df[,1] <- as.POSIXct(raw_df[,1], format = "%Y-%m-%d %H:%M:%S")
rownames(raw_df) <- raw_df$X
raw_df$X <- NULL
View(raw_df)
```

```
# önceden periyodumuzu 18 olarak belirlemiştik ACF grafiğinden
periyot_acf = 18

# t değişkeni veri setimizin başından sonuna kadar 1'er artan bir
seri
t <- 1:1:nrow(raw_df)


####################################
### Toplamsal Regresyon Modelleri ###
####################################

# 1. Regresyon Modeli:
# Eğer 1. regresyon modelinin tüm değişkenleri (t, sin1, cos1)
anlamlı
# çıkarsa, 2. regresyon modelini deneyeceğiz. Anlamsız çıkana kadar
# deneyip anlamsız çıkan ilk regresyon modelinin bir öncekine dönüp
onu
# kullanacağız.

# sin, cos tanımlama
sin1 <- sin(2*3.1416*t/periyot_acf)
cos1 <- cos(2*3.1416*t/periyot_acf)

# ham veri seti ile birleştirme
df <- as.data.frame(cbind(raw_df, t, sin1, cos1))
names(df)<- c("y", "t", "sin1", "cos1")
attach(df)
View(df)

# 1. regresyon modelini kurma
regresyon.model1 <-lm (y~t+sin1+cos1)
summary(regresyon.model1)
# t, sin1 ve cos1 anlamlı 2. modeli denemeye geçiyoruz
#####################################################

# sin, cos tanımlama
sin2 <- sin(2*3.1416*2*t/periyot_acf)
cos2 <- cos(2*3.1416*2*t/periyot_acf)

# ham veri ile birleştirme
df2 <- as.data.frame(cbind(raw_df, t, sin1, cos1, sin2, cos2))
names(df2) <- c("y", "t", "sin1", "cos1", "sin2", "cos2")
attach(df2)
View(df2)

# 2. Regreson Modeli
regresyon.model2 <- lm(y~t+sin1+cos1+sin2+cos2)
summary(regresyon.model2)
#tüm değişkenler anlamlı 3. regresyon modelini denemeye geçiyoruz.
################################################################
```

```
# sin, cos tanımlama
sin3 <- sin(2*3.1416*3*t/periyot_acf)
cos3 <- cos(2*3.1416*3*t/periyot_acf)

# ham veri ile birleştirme
df3 <- as.data.frame(cbind(raw_df, t, sin1, cos1, sin2, cos2, sin3,
cos3))
names(df3) <- c("y", "t", "sin1", "cos1", "sin2", "cos2", "sin3",
"cos3")
attach(df3)
View(df3)

# 3. Regresyon Modeli
regresyon.model3 <- lm(y~t+sin1+cos1+sin2+cos2+sin3+cos3)
summary(regresyon.model3)
# eklediğimiz 3. sin ve cos değerleri anlamsız çıktı
# bu nedenle 2. modele geri dönüyoruz.

# %95 güven düzeyinde seçilen model 2. Regresyon Modeli
summary(regresyon.model2)
dwtest(y~t+sin1+cos1+sin2+cos2)

# H0: kurulan regresyon modeli hatalarının arasında ilişki yoktur.
# H1: kurulan regresyon modeli hatalarının arasında ilişki vardır.
# p<0.05 için söylenebilir ki hatalar arasında ilişki vardır.

# 2. model icin tahmin serisi, hata serisi ve tahminin alt ve üst
sınırlarına ait seriler
tahmin1 <- predict(regresyon.model2)
sinir1 <- predict(regresyon.model2, interval = 'confidence' ,level =
.95)
hata1 <- resid(regresyon.model2)

plot( window(y),
      xlab="Zaman", ylab="Üretilen Enerji", type="l", lty=3, col=2,
lwd=2)
lines(window(sinir1[,2]) ,type="l",lty=1,col=4,lwd=2)
lines(window(sinir1[,3]) ,type="l",lty=1,col=3,lwd=2)
legend("topleft",c(expression(paste(x18)),
                   expression(paste(Altsinir)),
                   expression(paste(Üstsinir))),
       lwd=c(2,2,2),lty=c(3,1,1), cex=0.7, col=c(2,4,3))

#Hatalar akgurultu mu?
Acf(hata1, lag.max = 42,  ylim=c(-1,1), lwd=3)
# hatalar akgürültü serisi olarak görünmüyor, çünkü tüm değerler CI
dışında
# ancak nihai kararı Box-Ljung testi verecek

#Box-Ljung
```

```
Box.test(hata1, lag = 42, type = "Ljung")
# p<0.05 için %95 güvenle söylenebilir ki modelin hataları akgürültü
# serisi olmadığı için model kullanılabilir bir model değildir.


######################################
### Çarpımsal Regresyon Moddelleri ###
######################################
s1<-t*sin(2*3.1416*t/9)
c1<-t*cos(2*3.1416*t/9)

df4 <- as.data.frame(cbind(raw_df, t, s1, c1))

names(df4)<- c("y", "t", "s1", "c1")
attach(df4)

regresyon.model4 <- lm(y~t+s1+c1)
summary(regresyon.model4)
# çarpımsal regresyon modelinin sinus ve cosinus değerleri anlamlı
değil
# ancak t değişkeni anlamlı bu şekilde model kurup bir test edelim

regresyon.model4 <- lm(y~t)
dwtest(y~t)
# H0: kurulan regresyon modeli hatalarının arasında ilişki yoktur.
# H1: kurulan regresyon modeli hatalarının arasında ilişki vardır.
# p<0.05 için söylenebilir ki hatalar arasında ilişki vardır.

# 4.. model icin tahmin serisi, hata serisi ve tahminin alt ve üst
sınırlarına ait seriler
tahmin4 <- predict(regresyon.model4)
sinir4 <- predict(regresyon.model4, interval = 'confidence' ,level =
.95)
hata4 <- resid(regresyon.model4)

#Hatalar akgurultu mu?
Acf(hata4, lag.max = 42,  ylim=c(-1,1), lwd=3)
# hatalar akgürültü serisi olarak görünmüyor, çünkü tüm değerler CI
dışında
# ancak nihai kararı Box-Ljung testi verecek

#Box-Ljung
Box.test(hata4, lag = 42, type = "Ljung")
# p<0.05 için %95 güvenle söylenebilir ki modelin hataları akgürültü
# serisi olmadığı için model kullanılabilir bir model değildir.
```

Box-Jenkins Modelleri:

```
# clear environment
rm(list=ls())
```

```
# libs
library(fpp)

# load data
path = './Datasets/dataset.csv'
df <- read.csv(path)
df <- df[1:144,]
df[,1] <- as.POSIXct(df[,1], format = "%Y-%m-%d %H:%M:%S")
rownames(df) <- df$X
df$X <- NULL
View(df)

# acf plot
Acf(df, lag.max = 42,  ylim=c(-1,1), lwd=3)
# trend: var
# mevsimsellik: var
# periyot: s=18

# trendden arındırmak:
df <- ts(df, frequency = 18)
df_1 <- diff(df)
Acf(df_1, lag.max = 42,  ylim=c(-1,1), lwd=3)
# ilk 4 çubuğun hepsi CI dışına çıkması lazım ancak hepsi çıkmamış
trent yok
# trent farkı sayısı: d=1

# mevsimsellikten arındırma konusunda şuna başvurulabilir.
# ACF grafiğinden periyodu 18 olan bir mevsimsellik bulduk
# ancak PACF grafiğinde bu mevsimsellik önemli ise (ilgili gözlemler
CI dışında)
# tamam mevsimsellik var ve periyodu 18 derken
# PACF de ilgili gözlemler CI dışında değil ise
# mevsimellik farkı alınmasına gerek yok o kadar da önemli değil
şeklinde yorum da yapılabilir.

# mevsimsellikten arındırmak:
season <- rep(1:18, length.out = 144)
df_2 <- df_1 - ave(df_1, season, FUN = mean) #veya
# diff(diff(zaman_serisi),periyor)
Acf(df_2, lag.max = 42,  ylim=c(-1,1), lwd=3)
Pacf(df_2, lag.max = 42,  ylim=c(-1,1), lwd=3)
# mesimsellik farkı sayısı: D=1
# PACF grafiği ACF grafiğinden daha hızlı azalıyor: MA(q) modeli
# ACF -> PACF 'den hızlıysa p=0 verilip q'nun kaç değer aldığına
bakmak için
#       ACF grafiğinin ilk 3 değerinin kaç tanesi CI dışında ona
bakılır.
# PACF -> ACF'den hızlıysa q=0 verilip p'nin hangi değeri aldığını
bulmak için
#       PACF grafiğinin ilk 3 değerinin kaç tanesi CI dışında ona
bakılır.
```

```r
# Hem PACF hem ACF ikisi de yavaş azalıyorsa
# p değeri için PACF 'ye
# q deperi için ACF'ye bakılır. ilk 3 gözlemin kaçı CI dışında.

# PACF -> ACF 'den hızlı o zaman q=0 ve p için PACF bakarız.
# trendden ve mevsimsellikten arındırılmış serinin
# PACF ilk 3 gözleminin hiçbiri CI dışında değil: p=0 :

# (P,Q) için ilgili grafikte periyot kaç ise 1 öncesi ve 1
sonrasındaki gözlem
# aynı yönde ise 1'den başlayıp 3'e kadar dene değilse tam tersinden
başla 1'e doğru dene
Acf(df, lag.max = 42,  ylim=c(-1,1), lwd=3)
# aynı yönlü o zaman 1 den başlayıp 3'e doğru deneyeceğiz.
# Modeller -> ARIMA(p,d,q)(P,D,Q)

# Model-1: ARIMA(0,1,0)(1,1,0)
ARIMA1 <- Arima(df, order=c(0,1,0), seasonal=c(1,1,0),
include.constant = TRUE)
coeftest(ARIMA1) # P<0.05 Model anlamlı
summary(ARIMA1) # BIC:1542.41 RMSE:102.2535
Acf(ARIMA1$residuals, lag.max = 42,  ylim=c(-1,1), lwd=3) # Bence
akgürültü serisi
# Hataların dağılımı ile akgürültü serisi arasında fark yoktur
# Hataların dağılımı ile akgürültü serisi arasında fark vardır
Box.test(ARIMA1$residuals, lag = 42, type = "Ljung") # Box-Ljung
testine göre de akgürültü serisi


# Model-2: ARIMA(0,1,0)(2,1,0)
ARIMA2 <- Arima(df, order=c(0,1,0), seasonal=c(2,1,0),
include.constant = TRUE)
coeftest(ARIMA2) # P<0.05 Model anlamlı
summary(ARIMA2) # BIC:1542.19 RMSE:99.38836
Acf(ARIMA2$residuals, lag.max = 42,  ylim=c(-1,1), lwd=3) # Bence
akgürültü serisi
# Hataların dağılımı ile akgürültü serisi arasında fark yoktur
# Hataların dağılımı ile akgürültü serisi arasında fark vardır
Box.test(ARIMA2$residuals, lag = 42, type = "Ljung") # Box-Ljung
testine göre de akgürültü serisi


# Model-3: ARIMA(0,1,0)(3,1,0)
ARIMA3 <- Arima(df, order=c(0,1,0), seasonal=c(3,1,0),
include.constant = TRUE)
coeftest(ARIMA3) # P>0.05 Model anlamlı değil
summary(ARIMA3)


# Model-4: ARIMA(0,1,0)(0,1,1)
ARIMA4 <- Arima(df, order=c(0,1,0), seasonal=c(0,1,1),
```

```
include.constant = TRUE)
coeftest(ARIMA4) # P<0.05 Model anlamlı
summary(ARIMA4) # BIC:1538.64 RMSE:100.1105
Acf(ARIMA4$residuals, lag.max = 42,  ylim=c(-1,1), lwd=3) # Bence
akgürültü serisi
# Hataların dağılımı ile akgürültü serisi arasında fark yoktur
# Hataların dağılımı ile akgürültü serisi arasında fark vardır
Box.test(ARIMA4$residuals, lag = 42, type = "Ljung") # Box-Ljung
testine göre de akgürültü serisi


# Model-5: ARIMA(0,1,0)(0,1,2)
ARIMA5 <- Arima(df, order=c(0,1,0), seasonal=c(0,1,2),
include.constant = TRUE)
coeftest(ARIMA5) # P>0.05 Model anlamlı değil
summary(ARIMA5)

##################################
### Seçilen Model ve Öngörüler ###
##################################

# Model-2: ARIMA(0,1,0)(2,1,0)
ARIMA2 <- Arima(df, order=c(0,1,0), seasonal=c(2,1,0),
include.constant = TRUE)
coeftest(ARIMA2) # P<0.05 Model anlamlı
summary(ARIMA2) # BIC:1542.19 RMSE:99.38836

# 5 dönemlik tahminde bulunalım
ongoru <- forecast(ARIMA2, h=5)
ongoru[['mean']]
```