Ekonometri-I Ödev

TC.No 26074735082

Uğur Dar

12 06 2020

Verinin Düzenlenmesi

```r
tablo$Maas <- tablo$Maas + 6       # TcNo 2. değer
tablo$Egitim <- tablo$Egitim + 7   # TcNo 4. değer
tablo$Deneyim <- tablo$Deneyim + 7 # TcNo 6. değer
tablo
```

```
## # A tibble: 20 x 3
##     Maas Egitim Deneyim
##    <dbl>  <dbl>   <dbl>
##  1  4751     19      16
##  2  5304     17      18
##  3  5757     23      14
##  4  5567     19      28
##  5  5586     19      25
##  6  5672     19      16
##  7  5710     19      23
##  8  5743     20      17
##  9  5768     19      23
## 10  5790     19      18
## 11  5790     21      17
## 12  5835     21      11
## 13  5850     19      14
## 14  5852     19      23
## 15  5864     19      21
## 16  5864     19      21
## 17  5864     20      23
## 18  5964     23      13
## 19  5884     20      12
## 20  5884     21      14
```

1.a) EKK

```
model <- lm(Maas~.,data=tablo)
summary(model)
```

```
##
## Call:
## lm(formula = Maas ~ ., data = tablo)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -874.55  -50.48   58.01  151.53  241.94
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3741.542   1103.083   3.392  0.00347 **
## Egitim        91.795     47.683   1.925  0.07112 .
## Deneyim        8.745     14.455   0.605  0.55319
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 258.1 on 17 degrees of freedom
## Multiple R-squared:  0.1868, Adjusted R-squared:  0.0911
## F-statistic: 1.952 on 2 and 17 DF,  p-value: 0.1725
```

- Çoklu doğrusal regresyon modeli çıktısında, eğitim değişkeni 0.05 düzeyinde anlamlı değil, 0.10 düzeyinde anlamlı gözükmektedir.
- Deneyim bağımsız değişkeni 0.05 düzeyinde anlamlı değildir.
- R-kare değeri 0.1868 ve düzeltilmiş R-kare değeri 0.09 olarak bulunmuştur. Kurulan modelde bağımsız değişkenlerin, bağımlı değişkenleri açıklama oranı çok düşüktür.

```
qf(p = 0.95 ,df1 = 2,df2 = 17)
```

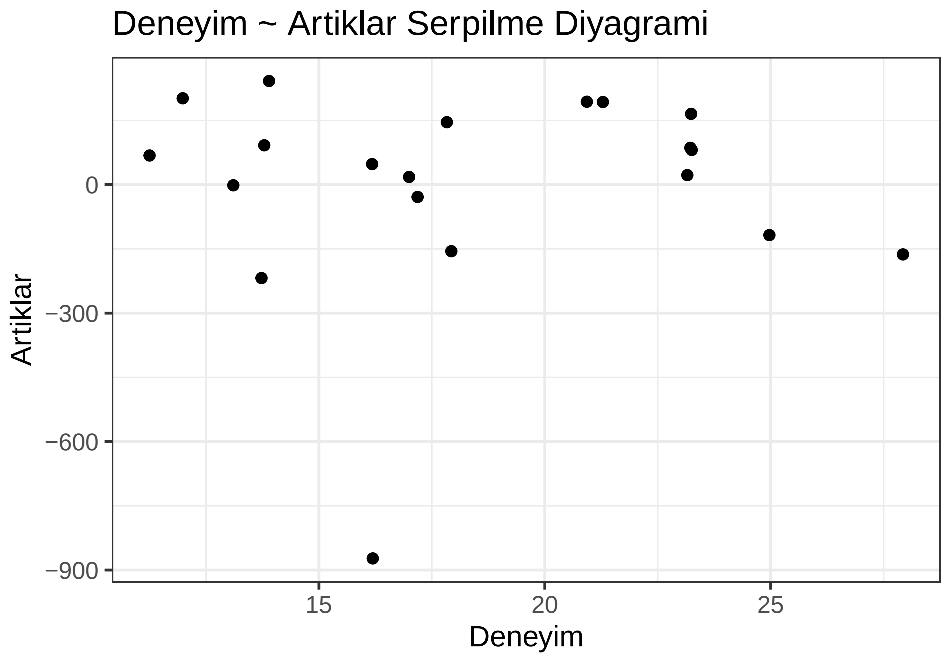
```
## [1] 3.591531
```

- F Tablo(alfa = 0.05) değer: 3.59 > 1.952 olduğundan, F testine göre model 0.05 anlamlılık düzeyinde, anlamlı değildir denilebilir.
- p-value = 0.1725 > 0.05 olduğundan model 0.05 anlamlılık düzeyinde anlamlı değildir denilebilir. Bağımlı değiken bağımsız değikenler tarafından %0.9 oranında açıklanmaktadır.

1.b) Artıkların, Bağımsız Değişkenlere Karşı Grafikleri

Artıklar - Deneyim Grafiği

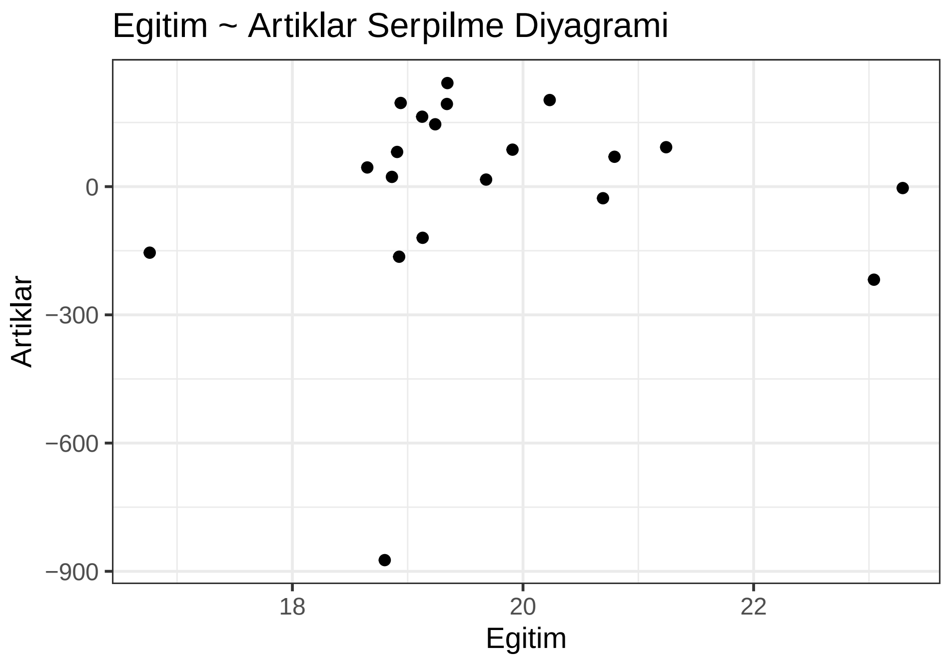
```
ggplot(tablo,aes(x=Deneyim,y=model$residuals))+
  geom_jitter() +
  labs(x = "Deneyim",y="Artıklar",title="Deneyim ~ Artıklar Serpilme Diyagramı")+
  theme_bw()
```



- Deneyim değişkeninin artıklara göre serpilme diyagramında noktalar rasgele dağılmıştır, değişen varyans sorunu olmayabilir. Aykırı değer olduğundan regresyon doğrusunun eğimi bu aykırı değerden çok fazla etkilenir. Yapılan yorum subjektif olup gerekli testler yapılarak varsayımlar incelenmelidir.

Artıklar - Eğitim

```
ggplot(tablo,aes(x=Egitim,y=model$residuals))+
  geom_jitter() +
  labs(x = "Eğitim",y="Artıklar",title="Eğitim ~ Artıklar Serpilme Diyagramı")+
  theme_bw()
```

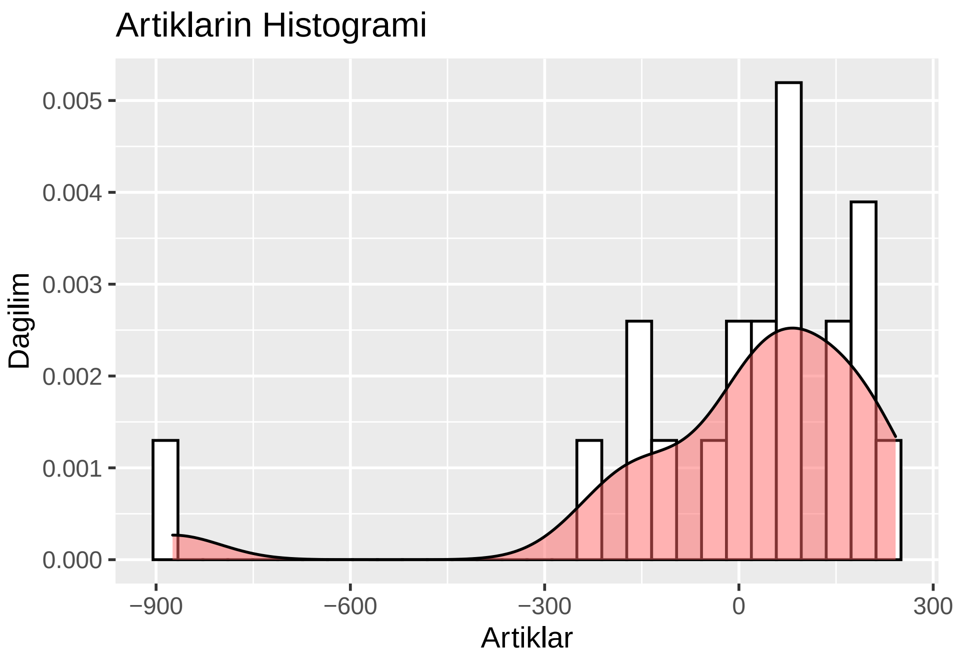


- Eğitim değişkeninin artıklara göre serpilme diyagramında noktalar rasgele dağılmamıştır, önce artan sonra azalan varyans gözükmektedir, değişen varyans sorunu olabilir. Deneyim değişkenin de olduğu gibi burada da bir aykırı değer gözükmektedir. Yapılan yorum subjektif olup gerekli testler yapılarak otokorelasyon incelenmelidir.

1.c) Artıkların Histogramı

```
artik <- as.data.frame(model$residuals)
ggplot(artik, aes(x=model$residuals)) +
 geom_histogram(aes(y=..density..), colour="black", fill="white")+
 geom_density(alpha=.5, fill="#FF6666") +
 labs(x="Artıklar",y="Dağılım",title="Artıkların Histogramı")
```

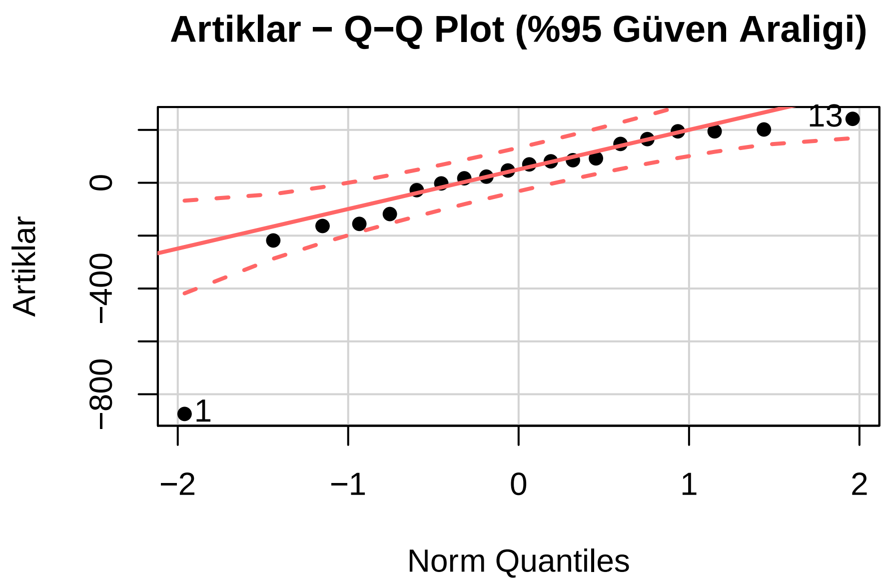
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



- Artıkların aykırı değerden çok fazla etkilendiği gözükmektedir. Eğer aykırı değerler olmasıydı sola çarpık bir dağılıma sahip olacaktı. Burada artıkların normal dağılmadığı söylenebilir ancak yapılan yorum subjektif olup normallik testleri yapılmalıdır.

1.d) Q-Q Plot

```
qqPlot(model$residuals,distribution = "norm",ylab="Artıklar",xlab="Norm Quantiles",
       main="Artıklar - Q-Q Plot (%95 Güven Aralığı)",
       pch = 16,col.lines = "#FF6666")
```




```
## [1]  1 13
```

- Artıkların teorik değerlerle karşılaştırıldığında bir çoğunun %95 güven aralığının içinde kaldığı gözükmektedir ancak aykırı değerler %95 güven aralığının dışında kalmıştır. Artıkların normallik varsayımı ihlal edilmiş görünüyor.

1.e) Normallik Testleri

```r
artik <- model$residuals
shapiro.test(artik)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  artik
## W = 0.74195, p-value = 0.0001327
```

```r
ad.test(artik)
```

```
##
##  Anderson-Darling normality test
##
## data:  artik
## A = 1.406, p-value = 0.0008767
```

```r
ks.test(artik,"pnorm",exact = FALSE)
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  artik
## D = 0.65, p-value = 9.151e-08
## alternative hypothesis: two-sided
```

Shapiro-Wilk, Anderson-Darling, Kolmogorov-Smirnov testlerine göre artiklar normal dağılmamaktadır.

1.f) Park Testi

n = 20 , anlamlılık düzeyi: 0.05 t tablo değeri : 2.09

Regresyon Modeli

```
data <- tablo
x1 <- data$Egitim
x2 <- data$Deneyim
y <- data$Maas
data <- cbind(y,x1,x2)
data <- as.data.frame(data)
lmfit <- lm(y~.,data=data)
ei  <- lmfit$residuals
ei2 <- ei^2
dt <- data.frame(lnei2 = log(ei2),x1 = log(x1),x2=log(x2))
lmfit1 <- lm(lnei2~x1,data=dt)
```

$$\ln\left(u_i^2\right) = \widehat{\beta}_0 + \widehat{\beta}_1 X_i + v_i \tag{1}$$

```
lmfit1 <- lm(lnei2~x1,data=dt)
summary(lmfit1)
```

```
##
## Call:
## lm(formula = lnei2 ~ x1, data = dt)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.7286 -1.2874  0.4337  0.9739  4.2117
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   56.353     21.808   2.584   0.0187 *
## x1           -15.881      7.314  -2.171   0.0435 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.279 on 18 degrees of freedom
## Multiple R-squared:  0.2075, Adjusted R-squared:  0.1635
## F-statistic: 4.714 on 1 and 18 DF,  p-value: 0.04354
```

$$\ln(u^2) = 56.353 - 15.881 X_1 + v_i \tag{2}$$

Şeklinde model tahmin edilmiştir. x1(Eğitim)'in katsayısı anlamlı 0.05 anlamlılık düzeyinde anlamlı olduğu için x1(Eğitim) değişkeninden kaynaklı değişen varyans vardır denilebilir.

```r
lmfit2 <- lm(lnei2~x2,data=dt)
summary(lmfit2)
```

```
##
## Call:
## lm(formula = lnei2 ~ x2, data = dt)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.7939 -0.7142  0.4675  1.3141  4.6610
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    5.393      6.368   0.847    0.408
## x2             1.260      2.205   0.572    0.575
##
## Residual standard error: 2.537 on 18 degrees of freedom
## Multiple R-squared:  0.01782,    Adjusted R-squared:  -0.03674
## F-statistic: 0.3266 on 1 and 18 DF,  p-value: 0.5747
```

$$\ln(u^2) = 5.393 + 1.260X_2 + v_i \tag{3}$$

Şeklinde model tahmin edilmiştir. x2(Deneyim)'in katsayısı 0.05 anlamlılık düzeyinde anlamlı olmadığı için x2'den kaynaklanan değişen varyans vardır diyemeyiz.

Sonuç: Park testine göre, ayrı ayrı kurulan modellerde, x2(Deneyim) değişkeninden kaynaklı değişen varyans gözükmemektedir ancak x1(Eğitim) değişkeninden kaynaklı değişen varyans sorunu vardır.

1.g) Goldfeld-Quandt testi

```
lm.fit <- lm(tablo$Maas~tablo$Egitim + tablo$Deneyim,data=tablo)
gqtest(lm.fit,fraction=3) # c = 3
```

```
##
##  Goldfeld-Quandt test
##
## data:  lm.fit
## GQ = 0.0062434, df1 = 6, df2 = 5, p-value = 1
## alternative hypothesis: variance increases from segment 1 to 2
```

H0:Sabit varyans varsayimi gecerlidir.

Hazır fonksiyon yardımıyla yapılan testte, H0 hipotezi reddedilemez. Yani sabit varyans yoktur denilemez. Her iki bağımsız değişken için de ayrı ayrı Goldfeld-Quandt testi yapılarak aşağıdaki sonuçlara ulaşılır.

Eğitim Değişkeni için Goldfeld-Quandt Testi

1. ve 2. adım

```
dt <- data.frame(Maas = tablo$Maas,Egitim = tablo$Egitim)
dt <- dt[order(dt$Egitim),] # Egitimin artan değerlerine göre sıraladık.
# Veri setini c=3 için 3 değeri atarak 2'ye bölersek
# (20-3)/2 =8.5  8 ve 9 birimden oluşan iki yeni grup yaparsak
dt1 <- dt[1:9,]   # 9 birimden oluşuyor
dt2 <- dt[13:20,] # 8 birimden oluşuyor
dt
```

```
##    Maas Egitim
## 2  5304     17
## 1  4751     19
## 4  5567     19
## 5  5586     19
## 6  5672     19
## 7  5710     19
## 9  5768     19
## 10 5790     19
## 13 5850     19
## 14 5852     19
## 15 5864     19
## 16 5864     19
## 8  5743     20
## 17 5864     20
## 19 5884     20
## 11 5790     21
## 12 5835     21
## 20 5884     21
## 3  5757     23
## 18 5964     23
```

3.,4. adım ve Karar

```
model1 <- lm(Maas~Egitim,data=dt1)# Veri setinin ilk kısmıyla kurulan model.
model2 <- lm(Maas~Egitim,data=dt2)# Veri setinin ikinci kısmıyla kurulan model.
e1 <- residuals(model1)
ss1 <- sum(e1^2)              # 1.model için SS_1.
e2 <- residuals(model2)
ss2 <- sum(e2^2)              # 2.model için SS_2.
sd <- (20-3)/2-2              # sd=(n-c)/2-k
lambda <- (ss2/sd)/(ss1/sd)   # Lambdanın hesaplanması.
Ftablo <- qf(1-0.05,sd,sd)    # F-Tablo değeri.
if(lambda > Ftablo){
  paste("Labmda =",round(lambda,3)," ve F Tablo değeri =",round(Ftablo,3),
        "Lambda > F_T olduğundan Sabit varyans rededilebilir.")
}else{
  cat("Labmda =",round(lambda,3)," ve F Tablo değeri =",round(Ftablo,3),"\n",
      "Lambda < F_T olduğundan Sabit varyans rededilemez.")
}
```

```
## Labmda = 0.043  ve F Tablo değeri = 4.012
##  Lambda < F_T olduğundan Sabit varyans rededilemez.
```

Deneyim Değişkeni için Goldfeld-Quandt Testi

1. ve 2. adım

```
dt <- data.frame(Maas = tablo$Maas,Deneyim = tablo$Deneyim)
dt <- dt[order(dt$Deneyim),] # Deneyim değişkenini artan değerlerine göre sıralandı.
# Veri setini c=3 için 3 değeri atarak 2'ye bölersek
# (20-3)/2 =8.5  8 ve 9 birimden oluşan iki yeni grup yaparsak
dt1 <- dt[1:9,]   # 9 birimden oluşuyor
dt2 <- dt[13:20,] # 8 birimden oluşuyor
dt
```

```
##    Maas Deneyim
## 12 5835      11
## 19 5884      12
## 18 5964      13
## 3  5757      14
## 13 5850      14
## 20 5884      14
## 1  4751      16
## 6  5672      16
## 8  5743      17
## 11 5790      17
## 2  5304      18
## 10 5790      18
## 15 5864      21
## 16 5864      21
## 7  5710      23
## 9  5768      23
## 14 5852      23
## 17 5864      23
## 5  5586      25
## 4  5567      28
```

3.,4. adım ve Karar

```
model1 <- lm(Maas~Deneyim,data=dt1)# Veri setinin ilk kısmıyla kurulan model.
model2 <- lm(Maas~Deneyim,data=dt2)# Veri setinin ikinci kısmıyla kurulan model.
e1 <- residuals(model1)
ss1 <- sum(e1^2)              # 1.model için SS_1.
e2 <- residuals(model2)
ss2 <- sum(e2^2)              # 2.model için SS_2.
sd <- (20-3)/2-2              # sd=(n-c)/2-k
lambda <- (ss2/sd)/(ss1/sd)   # Lambdanın hesaplanması.
Ftablo <- qf(1-0.05,sd,sd)    # F-Tablo değeri.
if(lambda > Ftablo){
  paste("Labmda =",round(lambda,3)," ve F Tablo değeri =",round(Ftablo,3),
        "Lambda > F_T olduğundan Sabit varyans rededilebilir.")
}else{
  cat("Labmda =",round(lambda,3)," ve F Tablo değeri =",round(Ftablo,3),"\n",
      "Lambda < F_T olduğundan Sabit varyans rededilemez.")
}
```

```
## Labmda = 0.034  ve F Tablo değeri = 4.012
##  Lambda < F_T olduğundan Sabit varyans rededilemez.
```

Sonuç: İki bağımsız değişken için de ayrı ayrı yapılan Goldfeld-Quandt testlerinde sabit varyans varsayımı reddedilememiştir. Goldfeld-Quandt testine göre sabit varyans varsayımı geçerlidir.

1.h) Glejser Testi

Artıkların mutlak değerini bağımlı, x1(Eğitim) ve x2(Deneyim) değişkenlerinin karekökünü bağımsız değişken olarak alarak regresyon modeli kurarsak;

```r
data <- tablo
x1 <- data$Egitim
x2 <- data$Deneyim
y <- data$Maas
data <- cbind(y,x1,x2)
data <- as.data.frame(data)
lmfit <- lm(y~x1+x2,data=data)
res <- lmfit$residuals
res <- abs(res) #Artıkların Mutlak değeri
x1 <- sqrt(x1); x2 <- sqrt(x2);
krk_data <- data.frame(res=res,x1=x1,x2=x2)
lm_glej <- lm(res ~ x1+x2,data=krk_data)
summary(lm_glej)
```

```
##
## Call:
## lm(formula = res ~ x1 + x2, data = krk_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -165.31  -89.20  -27.38   26.72  662.80
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2343.98    1612.88   1.453    0.164
## x1           -407.76     310.63  -1.313    0.207
## x2            -88.71      89.55  -0.991    0.336
##
## Residual standard error: 185 on 17 degrees of freedom
## Multiple R-squared:  0.09806,    Adjusted R-squared:  -0.008047
## F-statistic: 0.9242 on 2 and 17 DF,  p-value: 0.4159
```

H0:Sabit varyans varsayimi gecerlidir

Glejser testi ile H0 hipotezi reddedilemez,değişen varyans vardır diyemeyiz. Sabit varyans vardır.

Glejser testi birden fazla bağımsız değişken olduğunda uygulandığı gibi her bir bağımsız değişken için ayrı ayrı uygulanabilir.

h) 1- Glejser Testi(ayrı ayrı bağımsız değişkenlerle)

Sadece Eğitim değişkeni için model kurarsak

```
krk_data <- data.frame(res=res,x1=x1,x2=x2)
lm_glej <- lm(res ~ x1,data=krk_data)
summary(lm_glej)
```

```
##
## Call:
## lm(formula = res ~ x1, data = krk_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -153.09  -82.53  -40.53   18.39  698.23
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1249.1     1174.1   1.064    0.301
## x1            -246.1      264.2  -0.932    0.364
##
## Residual standard error: 184.9 on 18 degrees of freedom
## Multiple R-squared:  0.04599,    Adjusted R-squared:  -0.007006
## F-statistic: 0.8678 on 1 and 18 DF,  p-value: 0.3639
```

- Model anlamlı değil.

Sadece Deneyim değişkeni için model kurarsak

```
krk_data <- data.frame(res=res,x1=x1,x2=x2)
lm_glej <- lm(res ~ x2,data=krk_data)
summary(lm_glej)
```

```
##
## Call:
## lm(formula = res ~ x2, data = krk_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -170.90 -112.78  -13.40   38.59  711.80
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   270.62     332.91   0.813    0.427
## x2            -26.97      77.71  -0.347    0.733
##
## Residual standard error: 188.7 on 18 degrees of freedom
## Multiple R-squared:  0.006645,   Adjusted R-squared:  -0.04854
## F-statistic: 0.1204 on 1 and 18 DF,  p-value: 0.7326
```

- Model anlamlı değil.

Sonuç: Değişkenler tek tek incelendiğinde de H0 hipotezi 0.05 anlamlılık düzeyinde reddedilemez, sabit varyans yoktur diyemeyiz. Glejser testine göre sabit varyans varsayımı geçerlidir.

1.i) Spearman Sıra Korelâsyon Testi,

i) 1- Spearman Korelasyon Testi Maas-Eğitim

```r
cor.test(tablo$Maas,tablo$Egitim,method = "spearman")
```

```
##
##  Spearman's rank correlation rho
##
## data:  tablo$Maas and tablo$Egitim
## S = 661.34, p-value = 0.02386
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##       rho
## 0.5027488
```

Sonuç: Yapılan bu testte H0 hipotezi rho=0'dır. H0 hipotezi 0.05 anlamlılık düzeyinde reddedilir. Böylece, açıklayıcı değişken ve artıklar arasında sistematik bir ilişki olduğu söylenebilir, sabit varyans varsayımı ihlal edilmiş olunur.

i) 2- Spearman Korelasyon Testi Maas-Deneyim

```r
cor.test(tablo$Maas,tablo$Deneyim,method = "spearman")
```

```
##
##  Spearman's rank correlation rho
##
## data:  tablo$Maas and tablo$Deneyim
## S = 1800.2, p-value = 0.1262
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##        rho
## -0.3535697
```

Sonuç: Yapılan bu testte H0 hipotezi rho=0'dır. H0 hipotezi 0.05 anlam düzeyinde reddedilemez. Böylece, Deneyim açıklayıcı değişkeni ve artıklar arasında sistematik bir ilişki olduğuna dair kanıt yok ve değişen varyans olmadığı söylenebilir.

1.j) Breusch-Pagan-Godfrey Testi

```
lm.fit <- lm(Maas~.,data=tablo)
bptest(lmfit, varformula = NULL, studentize = TRUE, data = list())
```

```
##
##  studentized Breusch-Pagan test
##
## data:  lmfit
## BP = 1.3854, df = 2, p-value = 0.5002
```

Sonuç: H0 hipotezi: Sabit varyans vardır. Breusch-Pagan-Godfrey testi sonucu H0 reddedilemez. 0.05 anlamlılık düzeyinde Sabit varyans vardır.

Breusch-Pagan-Godfrey Testinin Algoritması

```
data <- tablo
x1 <- data$Egitim
x2 <- data$Deneyim
y <- data$Maas
data <- cbind(y,x1,x2)
data <- as.data.frame(data)
dt <- data.frame(y,x1,x2)
k <- 2
# Adım-1 Regresyon modelini tahmin ediniz. Artıkların bulunması.
lmfit <- lm(y~.,data=dt)
res <- lmfit$residuals
rss <- sum(res^2) # Artık kareler Toplamı
# Adım-2  sigma karenin Maximum Likelihood Tahmicisi
sigmasapka <- rss/length(x1)
# Adım-3 pi değerinin hesaplanması
p <- res^2/sigmasapka
dt2 <- data.frame(p,x1,x2)
# Adım-4 pi. regresyon modelinin kurulması
lmfit2 <- lm(p~.,data=dt2)
# Adım-5
essfit2 <- sum((lmfit2$fitted.values-mean(p))^2)
Theta <- essfit2/2

paste0("Theta = ",Theta)
```

```
## [1] "Theta = 5.77032912048304"
```

```
paste0("Ki-Kare ",k," serbestlik dereceli tablo değeri = ",qchisq(1-0.05,k))
```

```
## [1] "Ki-Kare 2 serbestlik dereceli tablo değeri = 5.99146454710798"
```

```
# Ki-Kare Tablo < Theta   -> Sabit varyans reddedilir
if(qchisq(1-0.05,k) < Theta) "Sabit varyans reddedilir" else
  "Sabit varyans vardır."
```

```
## [1] "Sabit varyans vardır."
```

1.k) White Testi

```r
# White testi
data <- tablo
x1 <- data$Egitim
x2 <- data$Deneyim
y <- data$Maas
data <- cbind(y,x1,x2)
data <- as.data.frame(data)
dt <- data.frame(y,x1,x2)
N <- length(x1)

# 1. Adım
# Asıl modelin tahmini
model1 <- lm(y~.,data=dt)
# Asıl denklemdeki artıkların kareleri
res2 <- model1$residuals^2


# 2. Adım (Yardımcı regresyon)

dt2 <- as.data.frame(cbind(res2,x1,x2,x1^2,x2^2,x1*x2))
colnames(dt2) <- c("res2","x1","x2","x1x1","x2x2","x1x2")
# Modelin Kurulması
model2 <- lm(res2~.,data=dt2)
k <- 5 # k değerimiz 5 açıklayıcı değişken olduğu için 5 oldu.
a <- summary(model2)
# 3. Adım
r2 <- a$r.squared #Yardımcı regresyonun R^2 değeri.
nr2 <- N*r2

paste("nR^2 = ",nr2)
```

```
## [1] "nR^2 =  1.72217924548321"
```

```r
paste("Ki-Kare Alfa=0.05 k = 5 için = ",qchisq(1-0.05,k))
```

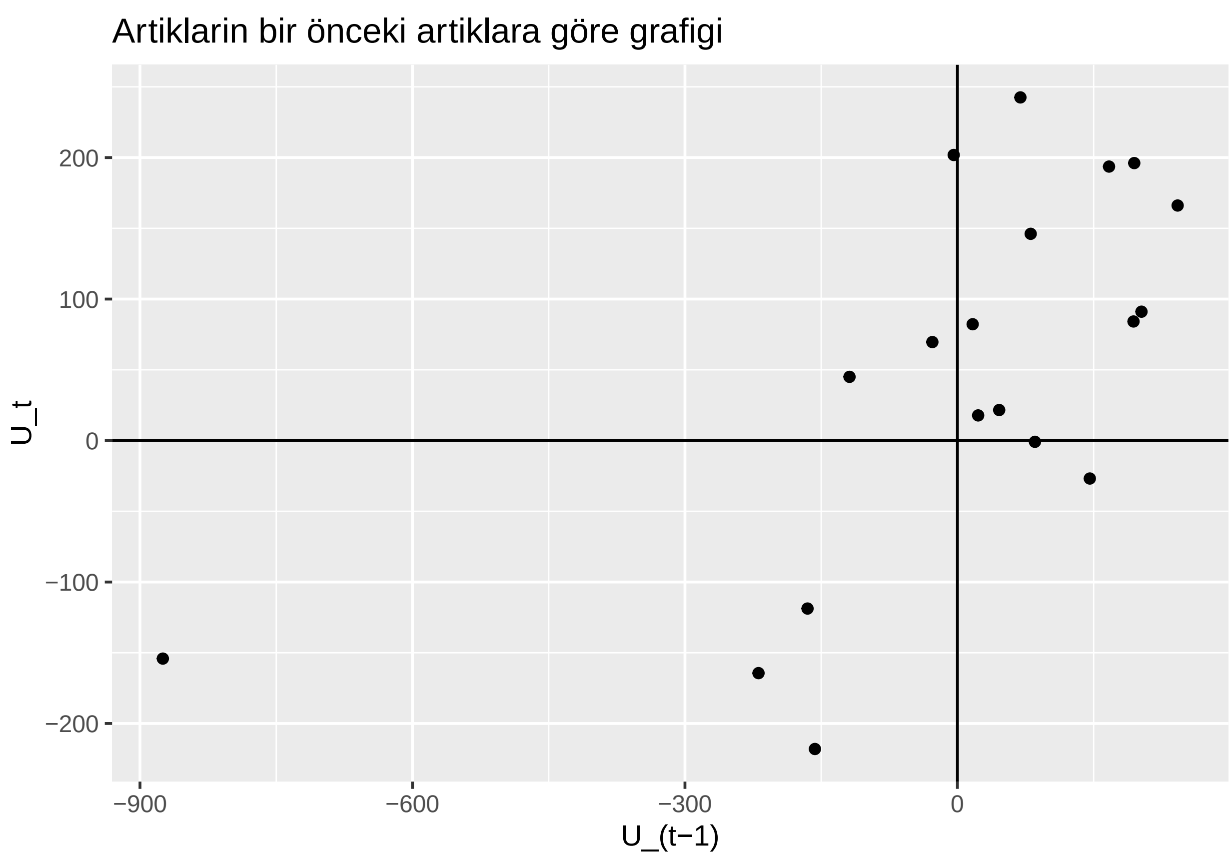
```
## [1] "Ki-Kare Alfa=0.05 k = 5 için =  11.0704976935164"
```

```r
# Karar:
## 0.05 anlalımlık seviyesinde - Ki Kare (k=5)

if(qchisq(1-0.05,k) < nr2) "Sabit varyans reddedilir, değişen varyans vardır denilebilir." else "Sabit
```

```
## [1] "Sabit varyans vardır"
```

1.l) Artıkların Gecikmeli Artıklara Göre Grafiği

```
mdl <- lm(Maas~.,data=tablo)
artik <- mdl$residuals
lagartik <- lag(artik)
df <- data.frame(artik,lagartik)
df <- df[-1,]

ggplot(df,aes(x=lagartik,y=artik))+
  geom_jitter()+
  geom_hline(yintercept = 0) +
  geom_vline(xintercept = 0) +
  labs(x="U_(t-1)",y="U_t",title="Artıkların bir önceki artıklara göre grafiği")
```



- Artıkların bir önceki artıklara göre grafiğinde pozitif yönlü otokorelasyon görülmektedir.

1.m) GEKK

Rho tahmini

```
model <- lm(Maas~.,data=tablo)
de <- lm(model$residuals~lag(model$residuals)-1)
summary(de)
```

```
##
## Call:
## lm(formula = model$residuals ~ lag(model$residuals) - 1)
##
## Residuals:
##     Min       1Q  Median      3Q     Max
## -164.35  -12.49   75.37  123.01  217.81
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## lag(model$residuals)   0.3467     0.1072   3.234   0.0046 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 113.6 on 18 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.3675, Adjusted R-squared:  0.3324
## F-statistic: 10.46 on 1 and 18 DF,  p-value: 0.004602
```

$$\hat{u}_t = \rho \hat{u}_{t-1} + v_t \tag{4}$$

$$\hat{u}_t = 0.3467 \hat{u}_{t-1} + v_t \tag{5}$$

$\hat{\rho} = 0.3467$ olarak tahmin edilmiştir.

Genelleştirilmiş EKK

```
rho <- as.numeric(de$coefficients)
yt <- tablo$Maas
xt1 <- tablo$Egitim
xt2 <- tablo$Deneyim
y_yildiz <- yt-rho*lag(yt)
x1_yildiz <- xt1-rho*lag(xt1)
x2_yildiz <- xt2-rho*lag(xt2)
df <- data.frame(y_yildiz,x1_yildiz,x2_yildiz)
df <- df[-1,]
gekk <- lm(y_yildiz~x1_yildiz+x2_yildiz)
summary(gekk)
```

```
##
## Call:
## lm(formula = y_yildiz ~ x1_yildiz + x2_yildiz)
##
## Residuals:
##    Min      1Q Median     3Q    Max
```

```
## -98.75 -52.00   9.15  46.44  85.40
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3455.078    163.722  21.103 4.17e-13 ***
## x1_yildiz     30.520     10.548   2.893   0.0106 *
## x2_yildiz     -5.256      3.496  -1.504   0.1522
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 60.96 on 16 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.5716, Adjusted R-squared:  0.5181
## F-statistic: 10.68 on 2 and 16 DF,  p-value: 0.001133
```

- Genelleştirilmiş EKK sonuçlarına göre x1_yildiz(Eğitim) bağımsız değişkeni 0.05 anlamlılık düzeyinde anlamlıdır. p-value=0.001 < 0.05 olduğundan model de genel olarak anlamlıdır. R^2 değeri 0.5716 bulunmuştur. Yani bağımsız değişkenler bağımlı değişkeni %57 oranında açıklamaktadır. GEKK modeli EKK modeline göre daha başarılı bir sonuç vermiştir.

1.n) Sabit Terimsiz EKK

```r
model2 <- lm(Maas~.-1,data=tablo)
summary(model2)
```

```
##
## Call:
## lm(formula = Maas ~ . - 1, data = tablo)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -667.64 -180.38   52.53  239.85  515.37
##
## Coefficients:
##         Estimate Std. Error t value Pr(>|t|)
## Egitim    249.82      12.78  19.550 1.43e-13 ***
## Deneyim    42.00      13.37   3.143  0.00563 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 324.8 on 18 degrees of freedom
## Multiple R-squared:  0.9971, Adjusted R-squared:  0.9968
## F-statistic:  3094 on 2 and 18 DF,  p-value: < 2.2e-16
```

Beta0 olan modelde bağımsız değişkenler 0.05 anlamlılık düzeyinde anlamlı değildi ancak Beta0'ı çıkarttıktan sonra bağımsız değişkenler 0.05 anlamlılık düzeyinde anlamlı çıktı. Modelimiz de genel olarak anlamlı oldu. Rˆ2 değeri 0.99'a çıktı. Sonuç olarak çok daha iyi bir model elde edildi. Beta0'ın olmadığı model daha uygundur.

1.o) Log-Log, İnverse, Quadratic Modeller

Log-Log Model Sabit Terimli

$$\widehat{\ln(Maas)} = \hat{\beta}_0 + \hat{\beta}_1 \ln(\text{Deneyim}) + u_i \tag{6}$$

```
model_log <- lm(log(Maas)~log(Deneyim),data=tablo)
summary(model_log)
```

```
##
## Call:
## lm(formula = log(Maas) ~ log(Deneyim), data = tablo)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.185670  0.000657  0.013902  0.026659  0.037538
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   8.70757    0.12956  67.209   <2e-16 ***
## log(Deneyim) -0.02012    0.04485  -0.449    0.659
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05161 on 18 degrees of freedom
## Multiple R-squared:  0.01106,    Adjusted R-squared:  -0.04388
## F-statistic: 0.2013 on 1 and 18 DF,  p-value: 0.659
```

$$\widehat{\ln(Maas)} = 8.68124 - 0.01396 \ln(\text{Deneyim}) + u_i \tag{7}$$

Log-Log Model Sabit Terimsiz

$$\widehat{\ln(Maas)} = \hat{\beta}_1 \ln(\text{Deneyim}) + u_i \tag{8}$$

```
model_log1 <- lm(log(Maas)~log(Deneyim)-1,data=tablo)
summary(model_log1)
```

```
##
## Call:
## lm(formula = log(Maas) ~ log(Deneyim) - 1, data = tablo)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.3137 -0.6756  0.1201  0.7912  1.5199
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## log(Deneyim)  2.98251    0.06173   48.32   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7974 on 19 degrees of freedom
## Multiple R-squared:  0.9919, Adjusted R-squared:  0.9915
## F-statistic:  2335 on 1 and 19 DF,  p-value: < 2.2e-16
```

$$\ln(\widehat{Maas}) = 3.5711 \ln(\text{Deneyim}) + u_i \tag{9}$$

İnverse Model Sabit Terimli

$$\widehat{\text{Maas}} = \hat{\beta}_0 + \hat{\beta}_1 \frac{1}{\text{Deneyim}} + u_i \tag{10}$$

```
inv_deneyim <- 1/tablo$Deneyim
Maas <- tablo$Maas

model_inv <- lm(Maas~inv_deneyim)
summary(model_inv)
```

```
##
## Call:
## lm(formula = Maas ~ inv_deneyim)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -974.93   -7.28   77.64  145.12  201.23
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   5566.3      245.1  22.711 1.06e-14 ***
## inv_deneyim   2554.6     4076.0   0.627    0.539
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 275.1 on 18 degrees of freedom
## Multiple R-squared:  0.02136,    Adjusted R-squared:  -0.03301
## F-statistic: 0.3928 on 1 and 18 DF,  p-value: 0.5387
```

$$\widehat{\text{Maas}} = 5607.9 + 941.6 \frac{1}{\text{Deneyim}} + u_i \tag{11}$$

İnverse Model Sabit Terimsiz

$$\widehat{\text{Maas}} = \hat{\beta}_1 \frac{1}{\text{Deneyim}} + u_i \tag{12}$$

```
model_inv1 <- lm(Maas~inv_deneyim-1)
summary(model_inv1)
```

```
##
## Call:
## lm(formula = Maas ~ inv_deneyim - 1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2543.4  -756.3   345.2  1717.4  2275.5
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## inv_deneyim    92162       5423   16.99 6.01e-13 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1458 on 19 degrees of freedom
## Multiple R-squared:  0.9383, Adjusted R-squared:  0.935
## F-statistic: 288.8 on 1 and 19 DF,  p-value: 6.014e-13
```

$$\widehat{\text{Maas}} = 43318\frac{1}{\text{Deneyim}} + u_i \tag{13}$$

Quadratic Model Sabit Terimli

$$\widehat{Maas} = \hat{\beta}_0 + \hat{\beta}_1 \text{ Deneyim } + \hat{\beta}_2 \text{ Deneyim }^2 + u_i \tag{14}$$

```
model_q <- lm(Maas ~ poly(Deneyim,2) ,data=tablo)
summary(model_q)
```

```
##
## Call:
## lm(formula = Maas ~ poly(Deneyim, 2), data = tablo)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -956.00    0.94   89.73  148.44  197.52
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        5714.95      63.16  90.490   <2e-16 ***
## poly(Deneyim, 2)1  -114.95     282.44  -0.407    0.689
## poly(Deneyim, 2)2   151.71     282.44   0.537    0.598
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 282.4 on 17 degrees of freedom
## Multiple R-squared:  0.02602,    Adjusted R-squared:  -0.08857
## F-statistic: 0.2271 on 2 and 17 DF,  p-value: 0.7992
```

$$\widehat{Maas} = 5708.95 - 114.95 \text{ Deneyim } + 151.71 \text{ Deneyim }^2 + u_i \tag{15}$$

Quadratic Model Sabit Terimsiz

$$\widehat{Maas} = \hat{\beta}_1 \text{ Deneyim } + \hat{\beta}_2 \text{ Deneyim }^2 + u_i \tag{16}$$

```
model_q1 <- lm(Maas ~ poly(Deneyim,2)-1 ,data=tablo)
summary(model_q1)
```

```
##
## Call:
## lm(formula = Maas ~ poly(Deneyim, 2) - 1, data = tablo)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##   4759   5716   5805   5863   5912
##
## Coefficients:
```

```
##                    Estimate Std. Error t value Pr(>|t|)
## poly(Deneyim, 2)1   -114.9     6030.3  -0.019    0.985
## poly(Deneyim, 2)2    151.7     6030.3   0.025    0.980
##
## Residual standard error: 6030 on 18 degrees of freedom
## Multiple R-squared:  5.535e-05,  Adjusted R-squared:  -0.111
## F-statistic: 0.0004981 on 2 and 18 DF,  p-value: 0.9995
```

$$\widehat{Maas} = -114.9 \text{ Deneyim } + 151.7 \text{ Deneyim }^2 + u_i \tag{17}$$

Sonuç: 6 Farklı regresyon modelinde en iyi sonucu veren sabit terimsiz Log-Log model olmuştur. Rˆ2 değeri 0.99 olduğundan ve model genel olarak anlamlı olduğundan yukarıdaki modeller arasında Log-Log modeli en mantıklı seçim olacaktır.

$$\ln(\widehat{Maas}) = 3.5711 \ln(\text{Deneyim}) + u_i \tag{18}$$

Verinin Düzenlenmesi

```
yillar$GELIR <- yillar$GELIR + 4 # TcNo. 5. rakamı
yillar
```

```
## # A tibble: 12 x 3
##      YIL TUKETIM GELIR
##    <dbl>   <dbl> <dbl>
##  1  1959   11378 11621
##  2  1960   13012 13301
##  3  1961   15260 15583
##  4  1962   16870 18019
##  5  1963   17765 19318
##  6  1964   18850 20202
##  7  1965   20074 20079
##  8  1966   21439 21434
##  9  1967   22833 22837
## 10  1968   24205 24209
## 11  1969   25307 25374
## 12  1970   27027 27004
```

2.a) EKK Modeli

```
model_ekk <- lm(TUKETIM~GELIR,data=yillar)
summary(model_ekk)
```

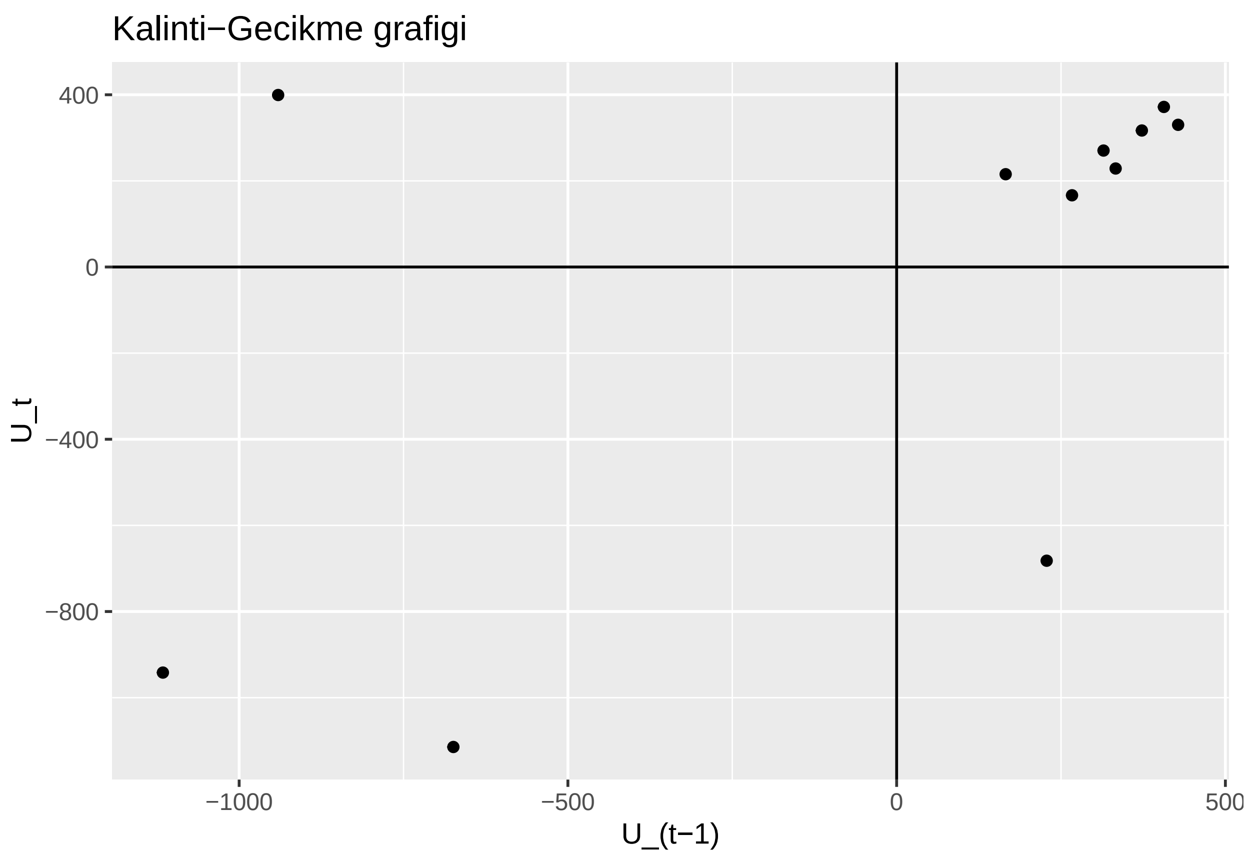
```
##
## Call:
## lm(formula = TUKETIM ~ GELIR, data = yillar)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1120.50   -39.77   250.53   344.30   435.45
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.050e+03  7.735e+02  -1.357    0.205
## GELIR        1.032e+00  3.788e-02  27.243 1.03e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 592.5 on 10 degrees of freedom
## Multiple R-squared:  0.9867, Adjusted R-squared:  0.9854
## F-statistic: 742.2 on 1 and 10 DF,  p-value: 1.028e-10
```

Modelde gelir değikeni 0.05 düzeyinde anlamlı çıkmıştır. Modelin R^2 değeri 0.98 çıkmıştır. EKK modeli genel olarak anlamlı gözükmektedir.

2.b) Artık Grafikleri

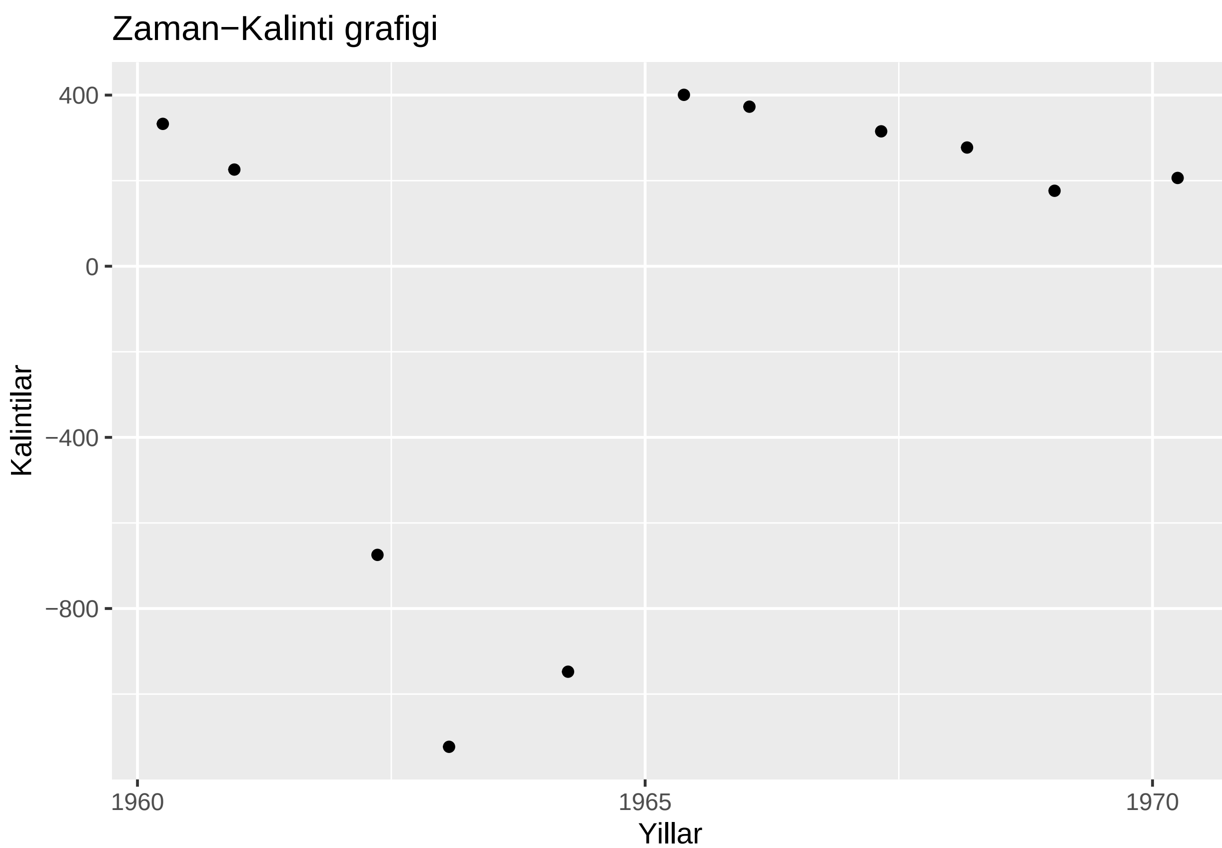
```
artik <- model_ekk$residuals
lagartik <- lag(artik)
```

```r
df <- data.frame(artik,lagartik)
df <- df[-1,]

ggplot(df,aes(x=lagartik,y=artik))+
  geom_jitter()+
  geom_hline(yintercept = 0) +
  geom_vline(xintercept = 0) +
  labs(x="U_(t-1)",y="U_t",title="Kalıntı-Gecikme grafiği")
```



- Artıkların bir önceki artıklara göre grafiğinde pozitif yönlü otokorelasyon görülmektedir. Yorumlar subjektif olup gerekli testler ile otokorelasyon test edilmelidir.

```r
artik <- model_ekk$residuals
lagartik <- lag(artik)
df <- data.frame(yil =  ymd(sprintf("%d-01-01",yillar$YIL)),artik,lagartik)
df <- df[-1,]

ggplot(df,aes(x=yil,y=artik))+
  geom_jitter()+
  labs(x="Yıllar",y="Kalıntılar",title="Zaman-Kalıntı grafiği")
```



- Artıkların zamana göre olan grafiğinde önce artıkların değerinin azaldığı sonra arttığı görülmektedir. Az sayıda veri olduğundan bu grafik üzerinden yorum yapmak daha güçtür. Gerekli testlerle otokorelasyon sınanabilir.

2.c) Durbin-Watson Testi

```r
durbinWatsonTest(model_ekk,alternative = "positive")
```

```
##  lag Autocorrelation D-W Statistic p-value
##    1       0.5524577     0.8285315   0.011
##  Alternative hypothesis: rho > 0
```

p-value = 0.005 < 0.05 olduğundan H0 hipotezi (rho=0) reddedilir. H1 hipotezi rho > 0 kabul edilir AR(1) durumu geçerlidir. DW testine göre 1. dereceden otokorelasyon vardır. EKK modeli R^2'si yüksek ve anlamlı çıksa da otokorelasyona sahiptir. Bu otokorelasyon regresyon varsayımlarını ihlal ettiği için EKK modeli kullanışlı değildir.

2.d) GEKK

Rho'nun Durbin-Watson d istatistiği ile tahmin edilmesi

```
dw <- durbinWatsonTest(model_ekk)
dw <- dw$dw # D değerinin alınması
rho_sapka <- 1- dw/2 # rho tahmini
paste("Tamin edilen Rho değeri = ",round(rho_sapka,4))
```

```
## [1] "Tamin edilen Rho değeri =  0.5857"
```

GEKK Modeli

```
yt <- yillar$TUKETIM
xt <- yillar$GELIR

y_yildiz <- yt-rho_sapka*lag(yt)
x_yildiz <- xt-rho_sapka*lag(xt)

df <- data.frame(y_yildiz,x_yildiz)
df <- df[-1,]
gekk <- lm(y_yildiz~x_yildiz)
summary(gekk)
```

```
##
## Call:
## lm(formula = y_yildiz ~ x_yildiz)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -777.1 -115.8  105.0  125.7  995.0
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -531.3357   912.0822  -0.583    0.574
## x_yildiz       1.0392     0.0959  10.837 1.83e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 497.6 on 9 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.9288, Adjusted R-squared:  0.9209
## F-statistic: 117.4 on 1 and 9 DF,  p-value: 1.826e-06
```

Sonuç: GEKK modelinin R^2'si EKK modelinin R^2'sine göre çok az bir farkla düşüktür. GEKK modeli otokorelasyondan arındılarak oluşturulduğu için, GEKK modelinin kullanılması daha doğru olacaktır.