Rastgele Blok Tasarımı Uygulama

Fatih Kızılaslan

24 04 2020

Rastgele Blok Tasarım (Sabit Etkili Model)

Örnek 1:

Bir firmada satış temsilcisi olarak çalışan 4 kişinin (özel olarak seçiliyor) 2019 yılının ilk 6 ayı için yaptığı satış tutarları bin TL olarak aşağıda verilmiştir.

```
##         1. Ay 2. Ay 3. Ay 4. Ay 5. Ay 6. Ay
## 1. Kişi  90.3  89.2  98.2  93.9  87.4  97.9
## 2. Kişi  92.5  89.5  90.6  94.7  87.0  95.8
## 3. Kişi  85.5  90.8  89.6  86.2  88.0  93.4
## 4. Kişi  82.5  89.5  85.6  87.4  78.9  90.7
```

$\alpha = 0.05$ olmak üzere bu veri için

a) Rastgele tam blok tasarımını kullanarak ANOVA tablosunu oluşturunuz.

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ ve H_1 : En az bir (i, j) için $\mu_i \neq \mu_j$

hipotezlerini test ederek yorumlayınız.

b) Tek yönlü varyans analizi yaparak faktör düzeyleri arasında fark var mıdır test ediniz.

c) Varyans analizinin varsayımlarını kontrol ediniz.

d) Eğer a ve b'de yapılan analizlerin sonucunda düzeyler arasında farklılık var ise ikili karşılaştırmalalar yaparak farklı olan düzeyleri belirleyiniz.

ÇÖZÜM

```r
y<-c(90.3,92.5,85.5,82.5,89.2,89.5,90.8,89.5,98.2,90.6,89.6,85.6,93.9,94.7
     ,86.2,87.4,87.4,87,88,78.9,97.9,95.8,93.4,90.7)

temsilci<-factor(rep(1:4, times = 6))   # Satış temsilcileri
blok<- factor(rep(1:6, each = 4))   # 6 bloğumuz var: Aylar
data<- data.frame(y, temsilci,blok)
print(data)
```

```
##       y temsilci blok
## 1  90.3        1    1
## 2  92.5        2    1
## 3  85.5        3    1
## 4  82.5        4    1
```

```
## 5  89.2        1    2
## 6  89.5        2    2
## 7  90.8        3    2
## 8  89.5        4    2
## 9  98.2        1    3
## 10 90.6        2    3
## 11 89.6        3    3
## 12 85.6        4    3
## 13 93.9        1    4
## 14 94.7        2    4
## 15 86.2        3    4
## 16 87.4        4    4
## 17 87.4        1    5
## 18 87.0        2    5
## 19 88.0        3    5
## 20 78.9        4    5
## 21 97.9        1    6
## 22 95.8        2    6
## 23 93.4        3    6
## 24 90.7        4    6
```

```
tapply(data$y,data$temsilci,mean)  #Faktör düzeylerinin (kişilerin satış) ortalamaları
```

```
##        1        2        3        4
## 92.81667 91.68333 88.91667 85.76667
```

```
tapply(data$y,data$blok,mean)  #Blokların ortalamaları
```

```
##      1      2      3      4      5      6
## 87.700 89.750 91.000 90.550 85.325 94.450
```

a) Rastgele Tam Blok Tasarım

Bu veri için ilk olarak bloklalama yaparak Blok tasarım için varyans analizini yapacağız. "aov" fonksiyonunda hem ana (main) faktör "temsilci" hem de bloklama faktörü "blok" kullanırız.

```
anova1 <- aov(y ~ temsilci + blok, data = data)
summary(anova1)
```

```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## temsilci     3  178.2   59.39   8.107 0.00192 **
## blok         5  192.2   38.45   5.249 0.00553 **
## Residuals   15  109.9    7.33
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Bu sonuçlara göre Faktör düzeyleri için $p-value = 0.00192 < 0.05$ olduğundan $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ veya $H_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4 = 0$ hipotezi red edilir. Böylece, faktör düzeyleri arasında anlamlı bir farklılık vardır.

Ayrıca, bloklar için de $p-value = 0.00553 < 0.05$ olduğundan $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$ hipotezi red edilir. Böylece, bloklar arasında da anlamlı bir farklılık vardır.

b) Tek yönlü ANOVA

Aynı veriyi bloklama yapmadan tek yönlü varyans analizi ile analiz edelim.

```
#Aynı veri bloklama olmadan ANOVA
anova2 <- aov(y ~ temsilci, data = data)
summary(anova2)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## temsilci     3  178.2   59.39   3.931 0.0234 *
## Residuals   20  302.1   15.11
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

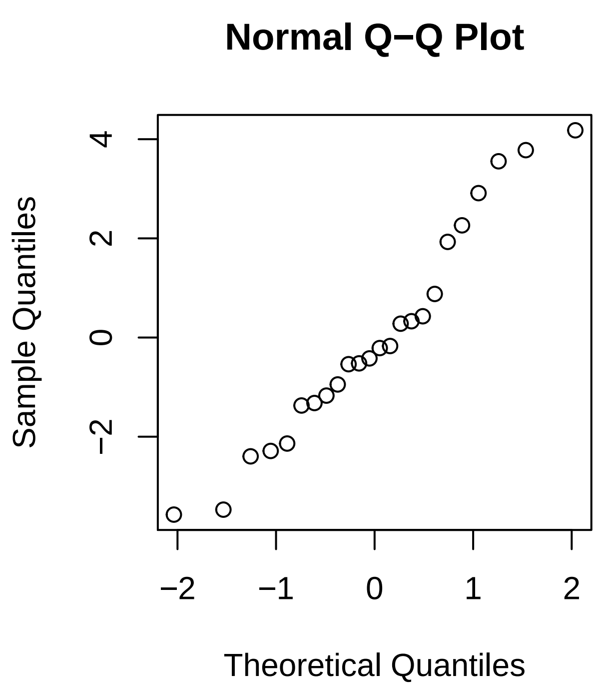
Yukarıdaki sonuçlara göre bu durumda da $p - value = 0.00234 < 0.05$ olduğundan $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ hipotezi red edilir, yani faktör düzeyleri arasında anlamlı bir farklılık vardır.

Ancak, bu modelde bloklamanın etkisi kaldırıldığı için $MS_E = 15.11$ yükselmiştir. Bloklama yapıldığında $MS_E = 7.33$ olarak bulundu. Görüldüğü gibi bloklama yapmak hata kareler ortalamasının azalmasına sebep olmaktadır. Dolayısıyla, bu veri için bloklama yapmak modelin hassaslığını arttırmıştır.

c) Varsayımların Kontrolü

1. Artıklar üzerinden normallik varsayımı kontrolü yapılır.

```
qqnorm(residuals(anova1))
```



```r
shapiro.test(residuals(anova1))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(anova1)
## W = 0.95631, p-value = 0.3689
```

```r
ks.test(residuals(anova1),"pnorm",mean(residuals(anova1)),sd(residuals(anova1)))
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  residuals(anova1)
## D = 0.1305, p-value = 0.7609
## alternative hypothesis: two-sided
```

Sonuçlara göre normallik varsayımı sağlanır.

2. Faktör düzeyleri (temsilciler) için varyansların homojenliğini kontrol edelim.

```r
bartlett.test(y ~ temsilci)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  y by temsilci
## Bartlett's K-squared = 1.2442, df = 3, p-value = 0.7424
```

```r
library(car)
leveneTest(y, temsilci) #medyana göre
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  3  0.7789 0.5195
##       20
```

```r
leveneTest(y, temsilci,mean) #ortalamaya göre
```

```
## Levene's Test for Homogeneity of Variance (center = mean)
##       Df F value Pr(>F)
## group  3  0.8542 0.4808
##       20
```

Sonuçlara göre faktör düzeylerinin varyansları homojendir.

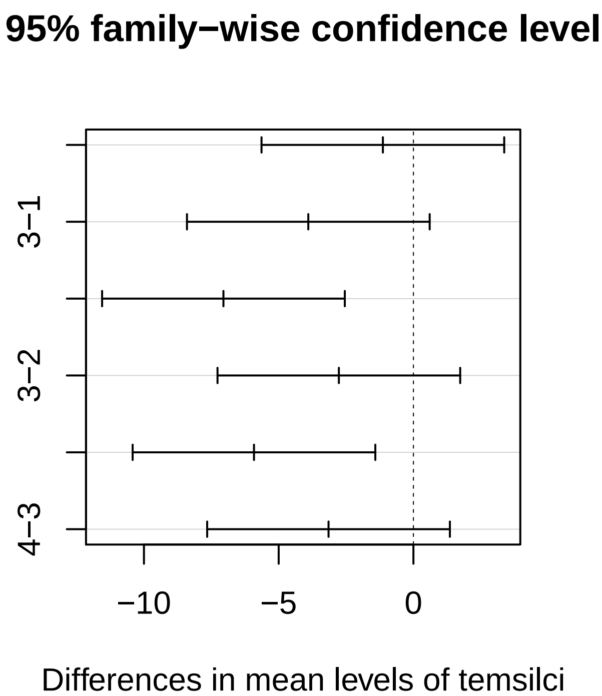
d) İkili Karşılaştırmalar

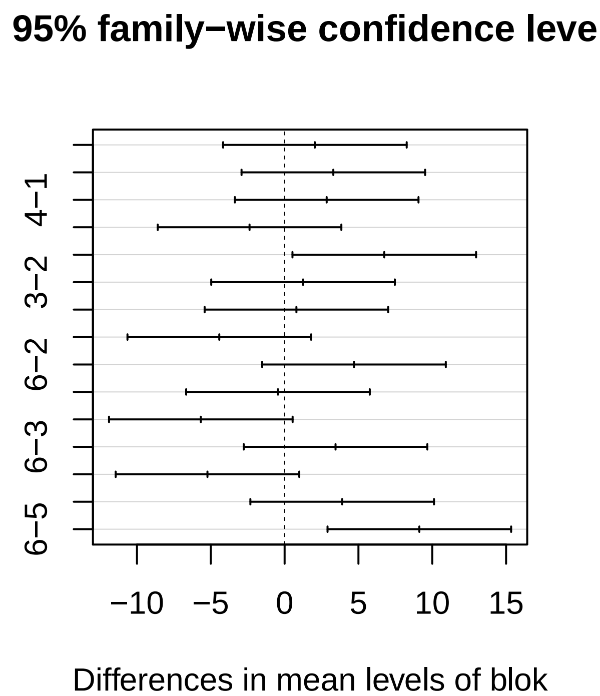
Rastgele blok tasarımı için Tukey testi ile ikili karşılaştırmaları yapalım.

```r
TukeyHSD(anova1)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = y ~ temsilci + blok, data = data)
##
## $temsilci
##          diff        lwr       upr     p adj
## 2-1 -1.133333  -5.637161  3.370495 0.8854831
## 3-1 -3.900000  -8.403828  0.603828 0.1013084
## 4-1 -7.050000 -11.553828 -2.546172 0.0020883
## 3-2 -2.766667  -7.270495  1.737161 0.3245644
## 4-2 -5.916667 -10.420495 -1.412839 0.0086667
## 4-3 -3.150000  -7.653828  1.353828 0.2257674
##
## $blok
##       diff         lwr        upr     p adj
## 2-1  2.050  -4.1680828  8.2680828 0.8853016
## 3-1  3.300  -2.9180828  9.5180828 0.5376297
## 4-1  2.850  -3.3680828  9.0680828 0.6757699
## 5-1 -2.375  -8.5930828  3.8430828 0.8105903
## 6-1  6.750   0.5319172 12.9680828 0.0297368
## 3-2  1.250  -4.9680828  7.4680828 0.9845521
## 4-2  0.800  -5.4180828  7.0180828 0.9980198
## 5-2 -4.425 -10.6430828  1.7930828 0.2483499
## 6-2  4.700  -1.5180828 10.9180828 0.1986961
## 4-3 -0.450  -6.6680828  5.7680828 0.9998784
## 5-3 -5.675 -11.8930828  0.5430828 0.0837504
## 6-3  3.450  -2.7680828  9.6680828 0.4925715
## 5-4 -5.225 -11.4430828  0.9930828 0.1263042
## 6-4  3.900  -2.3180828 10.1180828 0.3674672
## 6-5  9.125   2.9069172 15.3430828 0.0027838
```

```r
plot(TukeyHSD(anova1))
```





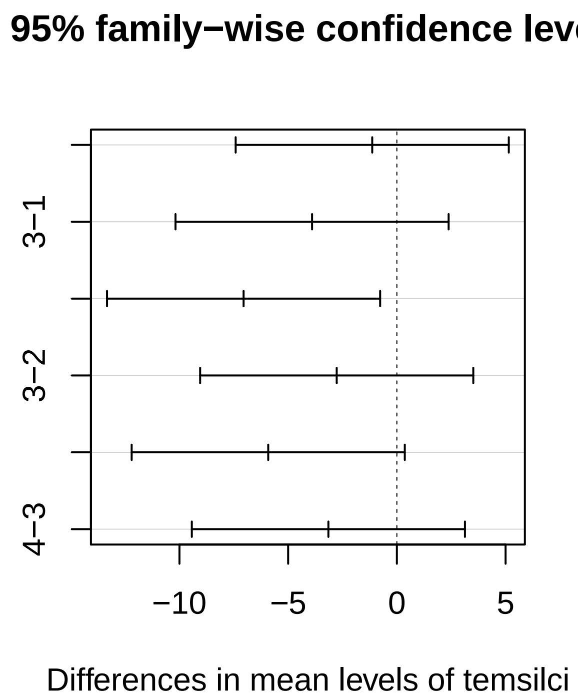
Yukarıdaki sonuçlara göre: **4. ve 1. düzeyler** arasında, **4. ve 2. düzeyler** arasında anlamlı fark vardır.

Eğer bu veri için **Bloklama yapmasaydık** tek yönlü ANOVA ile elde ettiğimiz sonuçlara göre Tukey testi ile ikili karşılaştırmalar aşağıdaki gibi olur.

```
TukeyHSD(anova2)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = y ~ temsilci, data = data)
##
## $temsilci
##          diff        lwr        upr     p adj
## 2-1 -1.133333  -7.414210  5.1475436 0.9569358
## 3-1 -3.900000 -10.180877  2.3808769 0.3312583
## 4-1 -7.050000 -13.330877 -0.7691231 0.0243604
## 3-2 -2.766667  -9.047544  3.5142102 0.6140278
## 4-2 -5.916667 -12.197544  0.3642102 0.0693182
## 4-3 -3.150000  -9.430877  3.1308769 0.5116013
```

```
plot(TukeyHSD(anova2))
```



Bu sonuçlara göre sadece **4. ve 1. düzeyler** arasında anlamlı fark vardır. Bloklama ile hem 4. ve 1. hem de 4. ve 2. düzeyler arasında anlamlı fark olduğu sonucuna ulaştmıştık.

Örnek 2: (Birdal Şenoğlu, Şükrü Acıtaş, İstatistiksel Deney Tasarımı, Örnek 5.2)

Y1, Y2, Y3 yemlerinin ineklerin süt verimne olan etkisi araştırılmak isteniyor. Bu amaçla, her biri 9'ar tane inekten oluşan 4 farklı ırktan (yerli kara (YK), boz ırk (BI), siyah alaca (SA), esmer İsviçre (Eİ)) toplam 36 tane inek kullanılıyor. 9'ar inekten oluşan 4 fraklı ırkın her biri için Y1, Y2, Y3 yemleri 3'er tane ineğe rastgele uygulanıyor. Bu deneye ilişikin veriler kg cinsinden aşağıda verilmiştir.

```
##      Y1   Y1   Y1   Y2   Y2   Y2   Y3   Y3   Y3
## YK  257  250  265  232  239  242  281  264  255
## BI 1600 1596 1609 1690 1698 1697 1717 1718 1721
## SA 5492 5497 5499 5589 5609 5588 5692 5688 5710
## Eİ 3993 3982 4008 4214 4206 4201 4291 4292 4312
```

Bu verileri kullanarak $\alpha = 0.05$ anlam düzeyinde,

a) Yemler arasında anlamlı bir farklılık olup olmadığını sınayınız.

b) Irklar arasında anlamlı bir farklılık olup olmadığını sınayınız.

c) Yem türleri ile ırklar arasındaki etkileşimin anlamlı olup olmadığını sınayınız.

ÇÖZÜM

Bu veride her bir blokta her bir düzey için 3'er tane gözlem yapılmıştır.

Bu analizde yem faktörünün süt üzerindeki etkisini ineklerin farklı özelliklerini göz önünde bulundurarak araştıracağız. Bu nedenle ana faktörümüz yem türleri Y1, Y2, Y3 ve inek ırkları bloklama faktörüdür. Ayrıca, bu iki faktörün kendi aralarında olan etkileşimini de modele ekleyeceğiz.

```r
y<-c(257,250,265,232,239,242,281,264,255,1600,1596,1609,1690,1698,1697,1717,
     1718,1721,5492,5497,5499,5589,5609,5588,5692,5688,5710,3993,3982,
     4008,4214,4206,4201,4291,4292,4312)

blok_irk<- factor(rep(1:4, each = 9)) # Bloklarımız inek ırkları
yem<-factor(rep(rep(1:3, each = 3)))  # Faktör düzeyleri yem türleri

data<-data.frame(y, blok_irk,yem)
print(data)
```

```
##       y blok_irk yem
## 1   257        1   1
## 2   250        1   1
## 3   265        1   1
## 4   232        1   2
## 5   239        1   2
## 6   242        1   2
## 7   281        1   3
## 8   264        1   3
## 9   255        1   3
## 10 1600        2   1
## 11 1596        2   1
## 12 1609        2   1
## 13 1690        2   2
## 14 1698        2   2
## 15 1697        2   2
## 16 1717        2   3
## 17 1718        2   3
## 18 1721        2   3
## 19 5492        3   1
## 20 5497        3   1
## 21 5499        3   1
## 22 5589        3   2
## 23 5609        3   2
## 24 5588        3   2
## 25 5692        3   3
## 26 5688        3   3
## 27 5710        3   3
## 28 3993        4   1
## 29 3982        4   1
## 30 4008        4   1
## 31 4214        4   2
## 32 4206        4   2
## 33 4201        4   2
## 34 4291        4   3
## 35 4292        4   3
## 36 4312        4   3
```

```r
tapply(data$y,data$yem,mean)
```

```
##        1        2        3
## 2837.333 2933.750 2995.083
```

```r
tapply(data$y,data$blok_irk,mean)
```

```
##         1         2         3         4
##  253.8889 1671.7778 5596.0000 4166.5556
```

Rastgele blok tasarımı kullanarak varyans analizi yapacağız. Etkileşimi modele "yem:blok_irk" biçiminde dahil ederiz. Bu modeli

$y_{ijk} = \mu + \tau_i + \beta_j + +(\tau\beta)_{ij} + \epsilon_{ijk}$, $i = 1, ..., a$, $j = 1, ..., b$, $k = 1, ..., n$

biçiminde ifade ederiz. Burada, $(\tau\beta)_{ij}$: i. faktör düzeyi ile j. blok arasındaki etkileşimi gösterir.

```r
anova<-aov(y~ yem + blok_irk+yem:blok_irk, data = data)
summary(anova)
```

```
##              Df    Sum Sq  Mean Sq  F value Pr(>F)
## yem           2    151772    75886    939.8 <2e-16 ***
## blok_irk      3 156429603 52143201 645736.2 <2e-16 ***
## yem:blok_irk  6     78891    13148    162.8 <2e-16 ***
## Residuals    24      1938       81
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

a) Bu sonuçlara göre Faktör düzeyleri yem türleri için $p - value < 0.05$ olduğundan $H_0 : \mu_1 = \mu_2 = \mu_3$ veya $H_0 : \tau_1 = \tau_2 = \tau_3 = 0$ hipotezi red edilir. Böylece, yem türleri arasında anlamlı bir farklılık vardır.

b) Bloklar için yani inek ırkları için $p - value < 0.05$ olduğundan $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ hipotezi red edilir. Böylece, inek ırkları arasında anlamlı bir farklılık vardır.

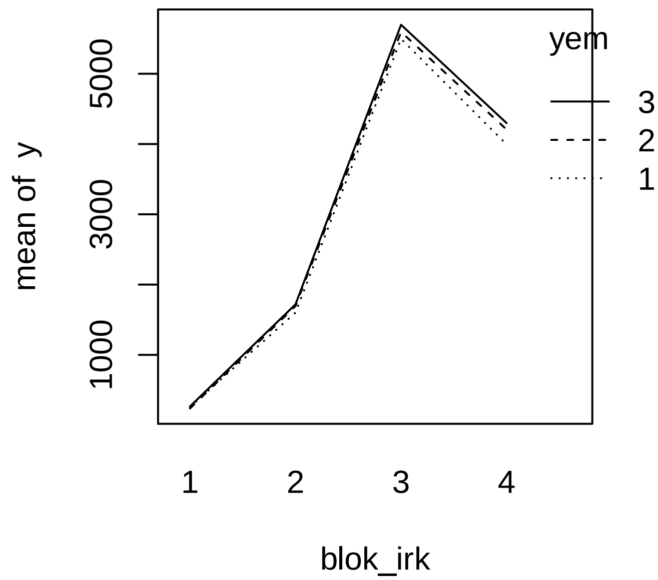
c) Yem türleri ile inek ırkları arasındaki etkileşim için sıfır hipotezi

$H_0 : (\tau\beta)_{11} = (\tau\beta)_{12} = ... = (\tau\beta)_{34} = 0$

biçiminde olur. Etkileşim için $p - value < 0.05$ olduğundan yem türleri ile ırklar arasındaki etkileşim anlamlıdır.

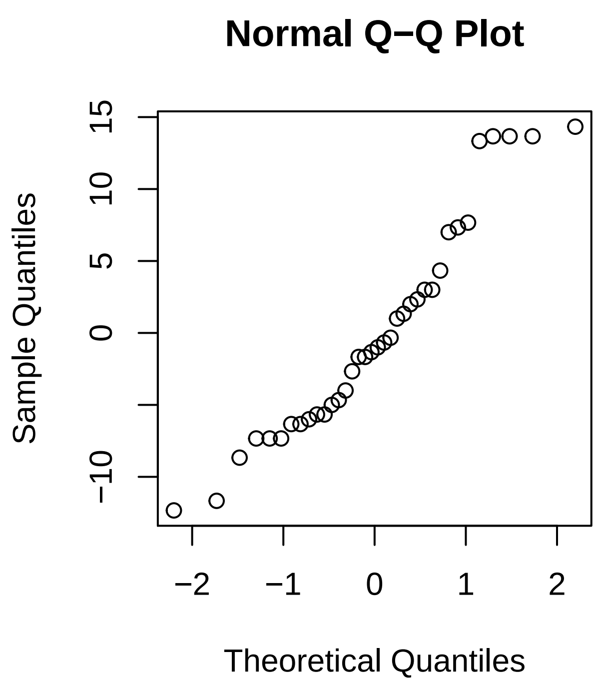
Ayrıca, yem türleri ile ırklar arasında etkileşim grafiği aşağıdaki gibidir. Eğer bu etkileşim grafiğinde birbirlerine paralel doğrular gözlenirse, bu faktörlerin etkilerinin birbirine bağlı olmadığı veya birbirleri ile etkileşimsiz olduğu yorumu yapılabilir. Ancak, etkileşim grafiği bu yorumları kesin olarak ifade etmez. Etkileşimin olup olmadığı hipotez testi ile kontrol edilmelidir.

```r
with(data, interaction.plot(x.factor =blok_irk , trace.factor = yem, response = y))
```



Ayrıca, aşağıdaki sonuçlardan varyans analizinin varsayımları hataların normalliği ve faktör düzeylerinin eş varyanslılığın sağlandığı görülmektedir.

```
qqnorm(residuals(anova))
```



```r
shapiro.test(residuals(anova))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(anova)
## W = 0.93539, p-value = 0.03656
```

```r
ks.test(residuals(anova),"pnorm",mean(residuals(anova)),sd(residuals(anova)))
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  residuals(anova)
## D = 0.10231, p-value = 0.8088
## alternative hypothesis: two-sided
```

```r
library(goftest)
ad.test(residuals(anova),"pnorm",mean=mean(residuals(anova)),sd=sd(residuals(anova)),estimated=TRUE)
```

```
##
##  Anderson-Darling test of goodness-of-fit
##  Braun's adjustment using 6 groups
##  Null hypothesis: Normal distribution
##  with parameters mean = 8.02309607639273e-17, sd = 7.44119806022037
##  Parameters assumed to have been estimated from data
##
## data:  residuals(anova)
## Anmax = 2.0867, p-value = 0.4111
```

```r
cvm.test(residuals(anova),"pnorm",mean=mean(residuals(anova)),sd=sd(residuals(anova)),estimated=TRUE)
```

```
##
##  Cramer-von Mises test of goodness-of-fit
##  Braun's adjustment using 6 groups
##  Null hypothesis: Normal distribution
##  with parameters mean = 8.02309607639273e-17, sd = 7.44119806022037
##  Parameters assumed to have been estimated from data
##
## data:  residuals(anova)
## omega2max = 0.51932, p-value = 0.1774
```

```r
bartlett.test(y ~ data$yem)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  y by data$yem
## Bartlett's K-squared = 0.019667, df = 2, p-value = 0.9902
```

```r
library(car)
leveneTest(y, data$yem) #medyana göre
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  2  0.0495 0.9518
##       33
```

```r
leveneTest(y, data$yem,mean) #ortalamaya göre
```

```
## Levene's Test for Homogeneity of Variance (center = mean)
##       Df F value Pr(>F)
## group  2  0.0496 0.9517
##       33
```