# Latin Kare, Greko-Latin Kare ve Tamamlanmamış Blok Tasarımlar Uygulama

Fatih Kızılaslan

08 05 2020

## Örnek 1: (Latin Kare Tasarım Uygulama)

Bir fabrikada çalışan bir mühendis 5 farklı ekip (A,B,C,D,E) tarafından üretilen ürün miktarları ile ilgili istatistiksel analiz yapmak istiyor. Üretilen ürün miktarında kullanılan 5 farklı malzeme türünün ve mesai gününün (hafta içi 5 gün) etkisi olduğu düşünülmektedir. 5 farklı ekibin 5 farklı malzeme ve 5 farklı günde elde ettiği ürün miktarları aşağıdaki gibi (data) elde edilmiştir.

```
y<-c(8,7,1,7,3,11,2,7,3,8,4,9,10,1,5,6,8,6,6,10,4,2,3,8,8)
gun<- factor(rep(1:5, times = 5)) #Günler Blok1
malzeme<- factor(rep(1:5, each = 5)) #Malzeme Blok2
ekip<-factor(c("A","B","D","C","E","C","E","A","D","B","B","A","C","E","D","D","C","E","B","A","E","D",


data<- data.frame(y,gun,malzeme,ekip)
print(data)
```

```
##     y gun malzeme ekip
## 1   8   1       1    A
## 2   7   2       1    B
## 3   1   3       1    D
## 4   7   4       1    C
## 5   3   5       1    E
## 6  11   1       2    C
## 7   2   2       2    E
## 8   7   3       2    A
## 9   3   4       2    D
## 10  8   5       2    B
## 11  4   1       3    B
## 12  9   2       3    A
## 13 10   3       3    C
## 14  1   4       3    E
## 15  5   5       3    D
## 16  6   1       4    D
## 17  8   2       4    C
## 18  6   3       4    E
## 19  6   4       4    B
## 20 10   5       4    A
## 21  4   1       5    E
## 22  2   2       5    D
```

```
## 23  3   3       5    B
## 24  8   4       5    A
## 25  8   5       5    C
```

$\alpha = 0.05$ anlamlılık düzeyinde bu veriyi kullanarak üretim miktarındaki farklılıkları istatistiksel olarak açıklayınız.

**ÇÖZÜM:**

Faktörlere göre ortalama üretim miktarları.

```
tapply(data$y,data$ekip,mean)
```

```
##   A   B   C   D   E
## 8.4 5.6 8.8 3.4 3.2
```

```
tapply(data$y,data$gun,mean)
```

```
##   1   2   3   4   5
## 6.6 5.6 5.4 5.0 6.8
```

```
tapply(data$y,data$malzeme,mean)
```

```
##   1   2   3   4   5
## 5.2 6.2 5.8 7.2 5.0
```

Bir ana faktör ve 2 bloklama faktörü olan Latin kare tasarımı için de "aov" fonksiyonunu kullanırız.

```
anova<-aov(y~ ekip + gun + malzeme , data = data)
summary(anova)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## ekip         4 141.44   35.36  11.309 0.000488 ***
## gun          4  12.24    3.06   0.979 0.455014
## malzeme      4  15.44    3.86   1.235 0.347618
## Residuals   12  37.52    3.13
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Bu sonuca göre, ana faktörümüz "ekip" için $p - value = 0.000488 < 0.05$ olduğundan $H_0$ hipotezi red edilir. Böylece, bu ekiplerin ürettikleri ortalama ürün miktarları arasında anlamlı bir farklılık vardır.

Ayrıca, bu modelin aşağıdaki gibi alt rastgele blok tasarımlarına da inceleyebiliriz.

```
anova1<-aov(y~ ekip + malzeme , data = data)
summary(anova1)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## ekip         4 141.44   35.36  11.370 0.000146 ***
## malzeme      4  15.44    3.86   1.241 0.333144
## Residuals   16  49.76    3.11
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova1<-aov(y~ ekip+ gun, data = data)
summary(anova1)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)    
## ekip         4 141.44   35.36  10.683 0.000207 ***
## gun          4  12.24    3.06   0.924 0.474092    
## Residuals   16  52.96    3.31                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

### Örnek 2: (Alıştırma 6.4-3 Şenoğlu ve Acıtaş) (Greko-Latin Kare Tasarım Uygulama)

4 farklı gübre türünün (A,B,C,D) çiçek tohumlarının filizlenmesine olan etkisi araştırılmak isteniyor. Bu amaçla, 4 farklı türde bitki $(B_1, B_2, B_3, B_4)$, 4 farklı çiçekçi $(C_1, C_2, C_3, C_4)$ ve 4 farklı marka saksı toprağı $(\alpha, \beta, \gamma, \delta)$ kullanılıyor. Bir ayın sonunda filizlerin boyları cm cinsinden aşağıdaki veride (data1) gösterildiği gibi ölçülüyor.

```r
y1<-c(7.75,6.93,6.82,5.18,5.76,6.54,6.19,4.67,4.48,7.02,5.24,8.07,3.77,6.44,6.00,5.24)
cicekci<-factor(rep(1:4, each = 4))
bitki<-factor(rep(1:4, times = 4))
Latin<-factor(c("A","C","D","B","D","B","A","C","B","D","C","A","C","A","B","D"))  #Gübre ana faktör
Greek<-factor(c("alfa","beta","gama","delta","beta","alfa","delta","gama","gama","delta","alfa","beta",

data1<- data.frame(y1,cicekci,bitki,Latin,Greek)
print(data1)
```

```
##      y1 cicekci bitki Latin Greek
## 1  7.75       1     1     A  alfa
## 2  6.93       1     2     C  beta
## 3  6.82       1     3     D  gama
## 4  5.18       1     4     B delta
## 5  5.76       2     1     D  beta
## 6  6.54       2     2     B  alfa
## 7  6.19       2     3     A delta
## 8  4.67       2     4     C  gama
## 9  4.48       3     1     B  gama
## 10 7.02       3     2     D delta
## 11 5.24       3     3     C  alfa
## 12 8.07       3     4     A  beta
## 13 3.77       4     1     C delta
## 14 6.44       4     2     A  gama
## 15 6.00       4     3     B  beta
## 16 5.24       4     4     D  alfa
```

$\alpha = 0.05$ anlamlılık düzeyinde

**a)** Gübre türleri arasında anlamlı bir farklılık olup olmadığını sınayınız.

**b)** Toprak markaları arasında anlamlı bir farklılık olup olmadığını sınayınız.

**c)** Çiçek türleri arasında anlamlı bir farklılık olup olmadığını sınayınız.

**d)** Çiçekçiler arasında anlamlı bir farklılık olup olmadığını sınayınız.

**ÇÖZÜM:**

```r
anova1<-aov(y1~  Latin + cicekci + bitki + Greek, data = data1)
summary(anova1)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## Latin        3  8.809  2.9365   5.212  0.104
## cicekci      3  3.761  1.2537   2.225  0.264
## bitki        3  3.592  1.1973   2.125  0.276
## Greek        3  3.530  1.1768   2.089  0.280
## Residuals    3  1.690  0.5634
```

Bu sonuca göre, her bir faktör için $p - value > 0.05$ olduğundan $H_0$ hipotezi kabul edilir. Böylece, hem ana faktör (gübre) hem de diğer faktörler için her birinin kendi düzeyleri arasında anlamlı bir farklılık yoktur.

Rastgele blok tasarım:

```r
anova2<-aov(y1~  Latin+Greek , data = data1)
summary(anova2)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## Latin        3  8.809   2.937   2.922 0.0926 .
## Greek        3  3.530   1.177   1.171 0.3736
## Residuals    9  9.043   1.005
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova3<-aov(y1~  Latin+bitki , data = data1)
summary(anova3)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## Latin        3  8.809   2.937   2.942 0.0913 .
## bitki        3  3.592   1.197   1.200 0.3641
## Residuals    9  8.982   0.998
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova4<-aov(y1~  Latin+cicekci , data = data1)
summary(anova4)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## Latin        3  8.809  2.9365   2.999 0.0878 .
## cicekci      3  3.761  1.2537   1.280 0.3389
## Residuals    9  8.813  0.9792
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Rastgele blok tasarımlar için de analiz sonuçları değişmiyor.

**ÖDEV:** Sizde bu veri için herhangi iki bloklama faktörünü seçerek oluşturulan Latin kare tasarımlar (toplam 3 tane) için varyans analizi yapınız.

## Tamamlanmamış Blok Tasarım Uygulama

## Örnek 3: (Örnek 10.1 Şenoğlu ve Acıtaş)

Dört farklı yemin (A,B,C,D) ineklerin süt verimine olan etkisi araştırılmak isteniyor. Bu amaçla her biri 3 inekten oluşan 4 farklı ırktan $(I_1, I_2, I_3, I_4)$ toplam 12 tane inek kullanılıyor. Bu deneye ilişkin veriler aşağıdaki gibi elde edilmiştir.

```
y1<-c(50,52,63,49,55,69,51,65,70,57,68,71)
yem<-factor(rep(1:4, each = 3))
irk<-factor(c("1","2","3","1","2","4","1","3","4","2","3","4"))

data1<- data.frame(y1,yem,irk)
print(data1)
```

```
##    y1 yem irk
## 1  50   1   1
## 2  52   1   2
## 3  63   1   3
## 4  49   2   1
## 5  55   2   2
## 6  69   2   4
## 7  51   3   1
## 8  65   3   3
## 9  70   3   4
## 10 57   4   2
## 11 68   4   3
## 12 71   4   4
```

Bu verileri kullanarak, $\alpha = 0.05$ anlam düzeyinde yemler arasında anlamlı bir farklılık olup olmadığını sınayınız.

**ÇÖZÜM:**

Verilere göre bu deney dengeli tamamlanmamış blok tasarımıdır. Bu veri için $a = 4, b = 4, r = 3, k = 3$ ve $\lambda = 2$ dir.

$H_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4 = 0$ ve $H_1$ : En az bir $\tau_i \neq 0$ hipotezlerini test ederiz.

*Dengeli tamamlanmamış blok tasarımında "aov" fonksiyonunun içinde modelin faktörlerini yazarken önce bloklama faktörü sonra ana faktör yazılmalıdır.* Çünkü, bloklama yaparak deney birimleri olabildiğince homojen olacak biçimde gruplandırılırak ana faktörün düzeylerinin karşılaştırılması istenir.

Böylece, doğru modelimiz ve sonucu aşağıdaki gibi olur.

```
anova1<-aov(y1~ irk + yem, data = data1)  #Doğru model "blok faktörü + ana faktör"
summary(anova1)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## irk          3  770.7  256.89 244.656 7.65e-06 ***
## yem          3   24.1    8.03   7.646   0.0258 *
## Residuals    5    5.3    1.05
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Bu sonuca göre, ana faktörümüz "yem" için $p - value = 0.0258 < 0.05$ olduğundan $H_0$ hipotezi red edilir. Böylece, yem türleri arasında anlamlı bir farklılık vardır.

Eğer modelde faktörlerin sırasını değiştirirsek aşağıdaki sonuçları elde ederiz.

```
anova2<-aov(y1~  yem+irk , data = data1)  #Yanlış model!!!
summary(anova2)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)    
## yem          3  188.7   62.89   59.89 0.000243 ***
## irk          3  606.1  202.03  192.41 1.39e-05 ***
## Residuals    5    5.3    1.05                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

**ÖDEV:** Modelin varsayımlarını test ediniz. İkili karşılaştırmaları yapınız.

## Örnek 4:

5 otomobil yarışçısının araç kullanım biçimlerinin 100km için ortalama yakıt (lt) tüketimi üzerindeki araştırılmak isteniyor. Bu amaçla 5 farklı yarış aracının her birinden 4 er tane olmak üzere toplam 20 tane araç kullanılıyor. Deneyin sonunda veriler aşağıdaki gibi elde ediliyor.

Bu verileri kullanarak $\alpha = 0.05$ anlam düzeyinde yarışçılar arasında anlamlı bir farklılık olup olmadığını sınayınız.

```
y3<-c(17,14,13,12,14,14,13,10,12,13,12,9,13,11,11,12,11,12,10,8)
yarisci<-factor(rep(1:5, each = 4))
araba<-factor(c("2","3","4","5","1","2","4","5","1","3","4","5","1","2","3","4","1","2","3","5"))

data3<- data.frame(y3,yarisci,araba)
print(data3)
```

```
##    y3 yarisci araba
## 1  17       1     2
## 2  14       1     3
## 3  13       1     4
## 4  12       1     5
## 5  14       2     1
## 6  14       2     2
## 7  13       2     4
## 8  10       2     5
## 9  12       3     1
## 10 13       3     3
## 11 12       3     4
## 12  9       3     5
## 13 13       4     1
## 14 11       4     2
## 15 11       4     3
## 16 12       4     4
## 17 11       5     1
## 18 12       5     2
## 19 10       5     3
## 20  8       5     5
```

**ÇÖZÜM:**

Verilere göre bu deney dengeli tamamlanmamış blok tasarımıdır. Bu veri için $a = 5, b = 5, r = 4, k = 4$ ve $\lambda = 3$ dür.

$H_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4 = \tau_5 = 0$ ve $H_1$ : En az bir $\tau_i \neq 0$ hipotezlerini test ederiz.

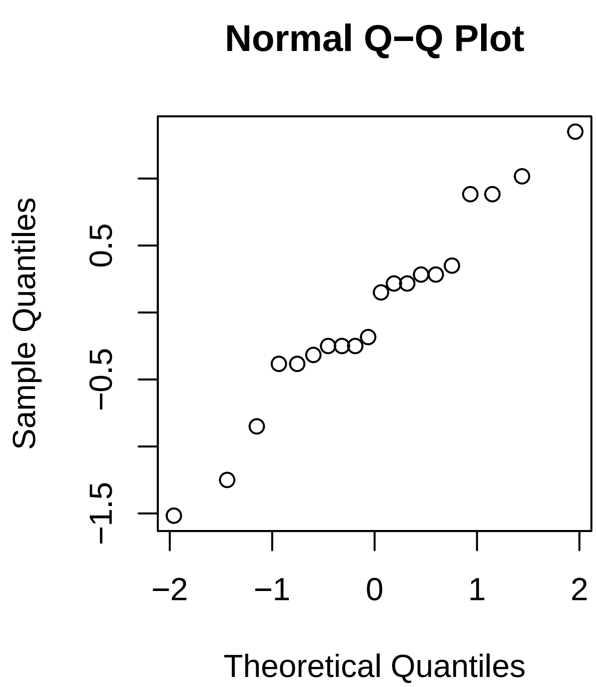
```
anova3<-aov(y3~ araba + yarisci, data = data3)  #Doğru model "blok faktörü + ana faktör"
summary(anova3)
```

```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## araba        4  31.20   7.800   8.566 0.00216 **
## yarisci      4  35.73   8.933   9.810 0.00125 **
## Residuals   11  10.02   0.911
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Bu sonuca göre, ana faktörümüz "yarisci" için $p-value = 0.00125 < 0.05$ olduğundan $H_0$ hipotezi red edilir. Böylece, yarışçıların araç kullanımları arasında anlamlı bir farklılık vardır.

Varsayım kontrolü: aşağıdaki sonuçlar göre normallik ve varyansların homojenliği varasayımları sağlanır.

```
qqnorm(residuals(anova3))
```




```
shapiro.test(residuals(anova3))
```

```
##
##  Shapiro-Wilk normality test
```

```
##
## data:  residuals(anova3)
## W = 0.96492, p-value = 0.6461
```

```r
ks.test(residuals(anova3),"pnorm",mean(residuals(anova3)),sd(residuals(anova3)))
```

```
## Warning in ks.test(residuals(anova3), "pnorm", mean(residuals(anova3)), : ties
## should not be present for the Kolmogorov-Smirnov test
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  residuals(anova3)
## D = 0.14877, p-value = 0.7679
## alternative hypothesis: two-sided
```

```r
library(goftest)
ad.test(residuals(anova3),"pnorm",mean=mean(residuals(anova3)),sd=sd(residuals(anova3)),estimated=TRUE)
```

```
##
##  Anderson-Darling test of goodness-of-fit
##  Braun's adjustment using 4 groups
##  Null hypothesis: Normal distribution
##  with parameters mean = -4.57994102712189e-17, sd = 0.726080561959993
##  Parameters assumed to have been estimated from data
##
## data:  residuals(anova3)
## Anmax = 1.246, p-value = 0.6819
```

```r
cvm.test(residuals(anova3),"pnorm",mean=mean(residuals(anova3)),sd=sd(residuals(anova3)),estimated=TRUE
```

```
##
##  Cramer-von Mises test of goodness-of-fit
##  Braun's adjustment using 4 groups
##  Null hypothesis: Normal distribution
##  with parameters mean = -4.57994102712189e-17, sd = 0.726080561959993
##  Parameters assumed to have been estimated from data
##
## data:  residuals(anova3)
## omega2max = 0.17225, p-value = 0.8053
```

**ÖDEV:** Bu analiz için tüm mümkün ikili karşılaştırmaları iki farklı yöntem ile yapınız. Sonuçlarını yorumlayınız.