# IST3002 Deney Tasarımı

## Rastgele Blok Tasarımı

Fatih Kızılaslan

Marmara Üniversitesi

2019-2020 Bahar VIII. Hafta

# Rastgele Blok Tasarımı: Giriş

- Rastgele blok tasarımında bir yönlü ANOVA gibi etkisi araştırılmak istenen faktör sayısı birdir.
- Gözlemler arasındaki farklılıklar hata varyansına önemli etkiler yapabilir. Ancak, bu etkiler çeşitli çevre koşulları (bir günün farklı zamanları, yılın mevsimleri, farklı günler, farklı sınıflar gibi) nedeniyle araştırma sırasında gözlemlenmeyebilir. **Gürültü değişkeni** olarak adlandırılan bu değişkeni kontrol altında tutmanın en basit yolu bu değişkeni sabit tutmaktır. Deney aynı gün, aynı saat, aynı yaş grubunda yapılacak biçimde tasarlanmalıdır [**Olmus, Erbas ve Nazman**].
- Bir diğer yöntem ise gürültü değişkenini faktörlerden biri olarak ele almaktır. Bu durumda gürültü değişkeni **bloklama** değişkeni olarak düşünülür.

Hülya Olmuş, Semra Oral Erbaş ve Ezgi Nazman, (2017). Araştırmacılar için SPSS uygulamalı İstatistiksel Deney Tasarımı, Gazi Kitabevi.

- Bloklar kendi içinde **homojen** ve bloklar kendi aralarında **heterojen** olarak oluşturulur.
- **Bloklama** ile gözlem hücreleri yardımcı etkenin düzeylerine göre alt hücrelere bölünerek daha homojen gözlemlerden oluşan hücreler ortaya çıkmaktadır ve modeldeki hata (hata varyansı) azalmaktadır [**Ozturk**].
- Bloklama deneysel hatanın azaltılması yoluyla deneyin hassaslığının artmasını sağlar [**Senoglu ve Acitas**].
- **Bloklamaya**, bağımlı değişken üzerinde etkilerini araştırmak istediğimiz etkenler dışında var olan, ancak bizim ilgilenmediğimiz ve etkilerini analiz etmek istemediğimiz etkenlerin eklenmesi olarak da bakılabilir [**Ozturk**].

Birdal Şenoğlu ve Şükrü Acıtaş (2014). İstatistiksel Deney Tasarımı Sabit Etkili Modeller, 3. Basım, Nobel Akademik Yayınevi.

Fikri Öztürk. İST306 İstatistik Deney Tasarımı Ders Notları: 9. Bloklama.

# Örnek

- Farklı tohum türlerinin verim üzerindeki etkisinin araştırıldığı bir deneyde her tohumun ekildiği tarlalar birbilerinden farklı özelliklere (toprak yapısı, iklim şartları gibi) sahip olabilir.
- Tarlaların bu farklılıkları dikkate alınmadığında varyans analizi sonucunda elde edilecek olan tohumun etkisi ile tarlaların özelliklerinin etkisi karışacaktır.
- Bu durumda tarlaların etkisini en aza indirmek için tarlalar özelliklerine göre olabildiğince homojen bloklara (gruplara, sınıflara) bölünür.
- Her tarla denenecek tohum sayısı kadar parçaya bölünerek her parçasında her tohum bir kez kullanılarak oluşturulan tasarıma **rastgele tam blok tasarımı** (randomized complete block design) denir.
- Her bir bloğun tüm faktör düzeylerini içerdiği durum **tam** tasarım olarak adlandırılır.
- Tarla tipi **bloklarımızı** oluşturur.
- Tohum türleri ise **faktör düzeyleridir**.

# Örneğin devamı

5 farklı tohum türünün, 4 farklı tarlada uygulanmasıyla elde edilen rastgele tam blok tasarım aşağıdaki gibi olacaktır.

| **Bloklar** | | | |
|---|---|---|---|
| **I. Tarla** | **II. Tarla** | **III. Tarla** | **IV. Tarla** |
| Tohum 1 | Tohum 2 | Tohum 5 | Tohum 4 |
| Tohum 2 | Tohum 4 | Tohum 3 | Tohum 2 |
| Tohum 3 | Tohum 1 | Tohum 2 | Tohum 5 |
| Tohum 4 | Tohum 5 | Tohum 4 | Tohum 1 |
| Tohum 5 | Tohum 3 | Tohum 1 | Tohum 3 |

Her tarla 5 eşit parçaya bölünüyor. Böylece 4 farklı blokta (tarlada) 5 farklı tohum türü **1 er kez** rastgele olarak uygulanır. **Rastgeleleştirme her blok içinde ayrı yapılıyor.**

- Yukarıda oluşturduğumuz gibi bir rastgele tam blok tasarımında bağımlı değişken üzerinde etkili birincil öneme sahip faktör tohum türleridir, ikincil öneme sahip olan faktör (yani tarla türü) ise **bloklama faktörü** olarak adlandırılır [**Senoglu ve Acitas**].
- Bloklama faktörü araştırmanın asıl konusu olmamakla beraber bağımlı değişkendeki değişimi açıklama potansiyeli bakımından önemlidir ve bu nedenle modele alınması gereklidir [**Senoglu ve Acitas**].

Ayrıca,

- Her bir faktör düzeyi her bir blok içinde birden fazla tekrar edebilir.
- İlgilenilen faktör ile bloklama etkeni arasında etkileşim olabilir.

# Rastgele Tam Blok Tasarımı: Matematiksel Model

$a$ faktör düzeyine sahip olan **bir faktör** ve $b$ farklı bloğun olduğu bir deneyde rastgele tam blok tasarımı için matematiksel model

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}, \;\; i = 1, ..., a, \;\; j = 1, ..., b \tag{1}$$

biçiminde ifade edilir. Burada,

| | | |
|---|---|---|
| $y_{ij}$ | : | $j$. bloktaki, $i$. faktör düzeyine (denemeye) ait olan gözlem değerini, |
| $\mu$ | : | genel ortalamayı, |
| $\tau_i$ | : | faktörün $i$. düzeyinin etkisini, |
| $\beta_j$ | : | $j$. bloğun etkisini, |
| $\epsilon_{ij}$ | : | $j$. bloktaki, $i$. faktör düzeyi için rastgele hata terimini |

gösterir. **Varsayım:** $\epsilon_{ij} \sim N(0, \sigma^2)$ biçiminde biribirinden bağımsız rastgele değişkenlerdir
Bu tasarımda her bir faktör düzeyi her bir blok içinde tam olarak **1 kez** kullanılmıştır.

Bu tasarım modeli için veri yapısı aşağıdaki gibi olur.

| **Düzeyler (Denemeler)** | **Bloklar** | | | | | | **Toplam** | **Ortalama** |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | . | . | . | $b$ | | |
| 1 | $y_{11}$ | $y_{12}$ | . | . | . | $y_{1b}$ | $y_{1.}$ | $\bar{y}_{1.}$ |
| 2 | $y_{21}$ | $y_{22}$ | . | . | . | $y_{2b}$ | $y_{2.}$ | $\bar{y}_{2.}$ |
| . | . | . | . | . | . | | . | . |
| . | . | . | . | . | . | | . | . |
| . | . | . | . | . | . | | . | . |
| $a$ | $y_{a1}$ | $y_{a2}$ | . | . | . | $y_{ab}$ | $y_{a.}$ | $\bar{y}_{a.}$ |
| **Toplam** | $y_{.1}$ | $y_{.2}$ | . | . | . | $y_{.b}$ | $y_{..}$ | |
| **Ortalama** | $\bar{y}_{.1}$ | $\bar{y}_{.2}$ | . | . | . | $\bar{y}_{.b}$ | | $\bar{y}_{..}$ |

- $y_{i.} = \sum_{j=1}^{b} y_{ij}$, $i = 1, ..., a$ ve $\bar{y}_{i.} = \frac{y_{i.}}{b} \rightarrow i$. faktör düzeyindeki gözlemler (satırlar) için ortalama
- $y_{.j} = \sum_{i=1}^{a} y_{ij}$, $j = 1, ..., b$ ve $\bar{y}_{.j} = \frac{y_{.j}}{a} \rightarrow j$. bloktaki gözlemler (sütunlar) için ortalama
- $y_{..} = \sum_{i=1}^{a} \sum_{j=1}^{b} y_{ij}$ ve $\bar{y}_{..} = \frac{y_{..}}{N} \rightarrow$ tüm gözlemler için genel ortalama

$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}$, $i = 1, ..., a$, $j = 1, ..., b$ modelinde ilk olarak faktör düzeylerinin ve blokların **sabit etkili** olduğu (yani özel olarak seçilmişler) durumu ele alacağız. Bu durumda

$$\sum_{i=1}^{a} \tau_i = 0 \text{ ve } \sum_{j=1}^{b} \beta_j = 0$$

olur. Ayrıca, $\tau_i$ ve $\beta_j$ birer sabit olduğundan $E(y_{ij}) = \mu + \tau_i + \beta_j$ ve $Var(y_{ij}) = \sigma^2$, $i = 1, ..., a$, $j = 1, ..., b$ olur.

Bu modeli kullanarak ilgilendiğimiz faktör düzeylerinin ortalamalarının eşitliğini test etmek isteriz. Bunun için

$$\begin{aligned} H_0 &: \mu_1 = \mu_2 = ... = \mu_a \\ H_1 &: \text{En az bir } (i, j) \text{ için } \mu_i \neq \mu_j \end{aligned} \tag{2}$$

hipotezlerini test ederiz.

Ayrıca,

$$\mu_i = \frac{1}{b}\sum_{j=1}^{b} y_{ij} = \frac{1}{b}\sum_{j=1}^{b}(\mu+\tau_i+\beta_j) = \mu+\tau_i+\underbrace{\frac{1}{b}\sum_{j=1}^{b}\beta_j}_{0} = \mu+\tau_i,$$

$i = 1, ..., a$ olduğundan

$$\begin{aligned} H_0 &: \mu_1 = \mu_2 = ... = \mu_a \quad \Leftrightarrow \quad H_0 : \tau_1 = \tau_2 = ... = \tau_a = 0 \\ H_1 &: \text{En az bir } (i,j) \text{ için } \mu_i \neq \mu_j \quad \Leftrightarrow \quad H_1 : \text{En az bir } i \text{ için } \tau_i \neq 0 \end{aligned}$$

biçiminde de yazılabilir.

Böylece,

$$\begin{aligned} H_0 &: \tau_1 = \tau_2 = ... = \tau_a = 0 \\ H_1 &: \text{En az bir } i \text{ için } \tau_i \neq 0 \end{aligned} \tag{3}$$

hipotezleri (2)'e denktir.

# Kareler Toplamının Parçalanışı

Tek yönlü ANOVA modelinde olduğu gibi burada da toplam kareler toplamı

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \overline{y}_{..})^2$$

uygun bir biçimde bileşenlerine parçalayarak (3) hipotezlerini sınamak için test istatistiği oluşturulur.

$$\begin{aligned}
SS_T &= \sum_{i=1}^{a}\sum_{j=1}^{b}\left[y_{ij} \underbrace{-\overline{y}_{i.} + \overline{y}_{i.}} \overbrace{-\overline{y}_{.j} + \overline{y}_{.j}} \underbrace{-\overline{y}_{..} + \overline{y}_{..}} -\overline{y}_{..}\right]^2 \\
&= b\sum_{i=1}^{a}(\overline{y}_{i.} - \overline{y}_{..})^2 + a\sum_{j=1}^{b}\left(\overline{y}_{.j} - \overline{y}_{..}\right)^2 \\
&\quad + \sum_{i=1}^{a}\sum_{j=1}^{b}\left(y_{ij} - \overline{y}_{i.} - \overline{y}_{.j} + \overline{y}_{..}\right)^2 \\
&= SS_{Deneme} + SS_{Blok} + SS_E
\end{aligned}$$

Böylece, toplam kareler toplamı (KT) $SS_T$

$$\text{Denemeler KT:} \qquad SS_{Deneme} = b\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2$$

$$\text{Blok KT:} \qquad SS_{Blok} = a\sum_{j=1}^{b}\left(\bar{y}_{.j} - \bar{y}_{..}\right)^2$$

$$\text{Hata KT:} \qquad SS_E = \sum_{i=1}^{a}\sum_{j=1}^{b}\left(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}\right)^2$$

olmak üzere

$$SS_T = SS_{Deneme} + SS_{Blok} + SS_E \tag{4}$$

biçiminde elde edilir.

(1) modelinde,

- toplam $N = ab$ gözlem olduğundan $SS_T$ : $N - 1$ serbestlik derecesine,
- $a$ faktör düzeyi olduğundan $SS_{Deneme}$ : $a - 1$ serbestlik derecesine,
- $b$ blok olduğundan $SS_{Blok}$ : $b - 1$ serbestlik derecesine
  sahiptir.
- Böylece, (4) eşitliğinden
  $$SS_E : (N-1)-(a-1)-(b-1) = \underbrace{N}_{ab} - a - b + 1 = (a-1)(b-1)$$
  serbestlik derecesine sahiptir.

Hataların normalliği varsayımı altında Cochran Teoremine göre

$$\frac{SS_{Deneme}}{\sigma^2} \sim \chi^2_{(a-1)}, \quad \frac{SS_{Blok}}{\sigma^2} \sim \chi^2_{(b-1)} \text{ ve } \frac{SS_E}{\sigma^2} \sim \chi^2_{(a-1)(b-1)}$$

biçiminde bibirinden bağımsız rastgele değişkenlerdir.

Ayrıca, **düzeylerin ve blokların etkisinin sabit olduğu (1) modeli** için

$$\begin{aligned}
E\left(MS_{Deneme}\right) &= E\left(\frac{SS_{Deneme}}{a-1}\right) = \sigma^2 + \frac{b}{a-1}\sum_{i=1}^{a}\tau_i^2 \\
E\left(MS_{Blok}\right) &= E\left(\frac{SS_{Blok}}{b-1}\right) = \sigma^2 + \frac{a}{b-1}\sum_{j=1}^{b}\beta_i^2 \\
E(MS_E) &= E\left(\frac{SS_E}{(a-1)(b-1)}\right) = \sigma^2
\end{aligned}$$

olarak elde edilir.

# Test İstatistiği

(1) modelinde, $H_0 : \tau_1 = \tau_2 = ... = \tau_a = 0$ hipotezi doğru olduğunda

$$F_{Deneme} = \frac{SS_{Deneme}/(a-1)}{SS_E/(a-1)(b-1)} = \frac{MS_{Deneme}}{MS_E}$$

test istatistiği $(a-1)$ ve $(a-1)(b-1)$ serbestlik dereceli $F$ dağılımına sahiptir: $F_{Deneme} \sim F_{(a-1),(a-1)(b-1)}$.

Eğer $F_{Deneme}$ test istatistiğinin değeri $F_{Deneme} > F_{(a-1),(a-1)(b-1),\alpha}$ olur ise

$$H_0 : \tau_1 = \tau_2 = ... = \tau_a = 0$$

hipotezi reddedilir.
Bu durumda, f*aktör düzeyleri arasında anlamlı bir farklılık vardır* denir.

Ayrıca, blok ortalamalarını karşılaştırmayı da düşünebiliriz. Bunun için hipotezlerimiz

$$\begin{aligned} H_0 &: \beta_1 = \beta_2 = ... = \beta_b = 0 \\ H_1 &: \text{En az bir } j \text{ için } \beta_j \neq 0 \end{aligned} \tag{5}$$

biçiminde olur. $H_0$ hipotezi test istatistiği

$$F_{Blok} = \frac{SS_{Blok}/(b-1)}{SS_E/(a-1)(b-1)} = \frac{MS_{Blok}}{MS_E}$$

kullanılır. Test istatistiğinin hesaplanan değeri için $F_{Blok} > F_{(b-1),(a-1)(b-1),\alpha}$ olur ise $H_0$ hipotezi reddedilir. Bu durumda, *bloklar arasında anlamlı bir farklılık vardır* denir.

- Blokların ortalamalarını karşılaştırmak genellikle çok tercih edilmez.
- Literatürde (5) hipotezlerinin test edilmesi ile ilgili bazı farklı düşünceler mevcuttur.
- Eğer $MS_{Blok}/MS_E$ oranı çok büyük ise bloklama faktörünün etkisinin önemli olduğunu söyleyebiliriz Bu durumda bloklama yapılarak faktör düzeylerinin karşılaştırılmasındaki hassasiyet iyileştirilmiş olur.

# ANOVA Tablosu

Şimdi yukarıda elde ettiklerimiz ile (1) modeli için ANOVA tablosunu oluşturalım.

| Değişim kaynağı | Kareler toplamı | Serbestlik derecesi | Kareler ortalaması | F test değeri |
|---|---|---|---|---|
| Denemeler | $SS_{Deneme}$ | $a-1$ | $MS_{Deneme} = \frac{SS_{Deneme}}{a-1}$ | $F_{Deneme} = \frac{MS_{Deneme}}{MS_E}$ |
| Bloklar | $SS_{Blok}$ | $b-1$ | $MS_{Blok} = \frac{SS_{Blok}}{b-1}$ | $F_{Blok} = \frac{MS_{Blok}}{MS_E}$ |
| Hata | $SS_E$ | $\underbrace{N-a-b+1}_{(a-1)(b-1)}$ | $MS_E = \frac{SS_E}{(a-1)(b-1)}$ | |
| Toplam | $SS_T$ | $N-1$ | | |

Ayrıca, $SS_T = \sum_{i=1}^{a} \sum_{j=1}^{b} y_{ij}^2 - \frac{y_{..}^2}{N}$, $SS_{Deneme} = \frac{1}{b} \sum_{i=1}^{a} y_{i.}^2 - \frac{y_{..}^2}{N}$ ve $SS_{Blok} = \frac{1}{a} \sum_{j=1}^{b} y_{.j}^2 - \frac{y_{..}^2}{N}$ biçiminde kolayca hesaplanabilir.

# Parametre Tahmini

$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}$, $i = 1, ..., a$, $j = 1, ..., b$ biçiminde verilen **sabit etkili rastgele tam blok tasarımında** bilinmeyen parametreler $\mu, \tau_i$, $i = 1, ..., a$ ve $\beta_j$, $j = 1, ..., b$ için *en küçük kareler (EKK)* tahmin edicileri

$$\begin{aligned} \widehat{\mu} &= \overline{y}_{..} \\ \widehat{\tau}_i &= \overline{y}_{i.} - \overline{y}_{..}, \ i = 1, ..., a \\ \widehat{\beta}_j &= \overline{y}_{.j} - \overline{y}_{..}, \ j = 1, ..., b \end{aligned} \tag{6}$$

biçiminde elde edilir.

Böylece, (1) modeli için yanıt değişkeni $y_{ij}$' nin tahmin edicisi

$$\begin{aligned} \widehat{y}_{ij} &= \widehat{\mu} + \widehat{\tau}_i + \widehat{\beta}_j \\ &= \overline{y} + (\overline{y}_{i.} - \overline{y}_{..}) + (\overline{y}_{.j} - \overline{y}_{..}) \\ &= \overline{y}_{i.} + \overline{y}_{.j} - \overline{y}_{..} \end{aligned} \tag{7}$$

olur.

Bu durumda, modelimiz için artıklar $e_{ij}$, $i = 1, ..., a$, $j = 1, ..., b$ olmak üzere

$$\begin{aligned} e_{ij} &= y_{ij} - \widehat{y}_{ij} \\ &= y_{ij} - \overline{y}_{i.} - \overline{y}_{.j} + \overline{y}_{..} \end{aligned} \tag{8}$$

biçiminde bulunur.

Tek yönlü ANOVA'da olduğu gibi $e_{ij}$, $i = 1, ..., a$, $j = 1, ..., b$ artıkları kullanılarak **modelin varsayımları kontrol edilir.**