IST3002 Deney Tasarımı

Faktöriyel Tasarımlar

Fatih Kızılaslan

Marmara Üniversitesi

2019-2020 Bahar XII. Hafta

Faktöriyel Tasarımlar: Giriş

- Şimdiye kadar yanıt değişkenini etkileyen sadece bir ana faktörün (main effects) olduğu tasarımları inceledik.
- Veriyi daha homojen bir yapıya getirmek için bloklama faktöründen faydalandık.
- Rastgele blok tasarım (ana faktör ve bir bloklama faktörü), Latin kare tasarım (ana faktör ve iki bloklama faktörü), Greko-Latin kare tasarım (ana faktör ve üç bloklama faktörü)
- Karşılacak olduğumuz gerçek verilerde yanıt değişkenini etkileyen ana faktör sayısı **iki veya daha fazla** olabilir.
- Faktör düzeylerinin tüm olası kombinasyonlarının araştırıldığı tasarıma **faktöriyel tasarım** (factorial design) denir.
- İki veya daha fazla faktörün ana etkilerini ve etkileşim etkilerini aynı anda araştırmak için kullanılan en yaygın tasarım **faktöriyel tasarım**lardır [**Senoglu ve Acitas**].

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

- 1. faktörün a tane düzeyi, 2. faktörün b tane düzeyi,..., k. faktörün m tane düzeyi var ise bu tasarım $axbx...xm$ faktöriyel tasarım olarak adlandırılır. Bu tasarımda toplam $axbx...xm$ tane deneme kombinasyonu olur.
- Bazı özel faktöriyel tasarımlar uygulamada sıklıkla kullanılır.
  1. Örneğin, k tane ana faktörün **ikişer** tane düzeyinin olduğu faktöriyel tasarım 2^k **faktöriyel tasarım** olarak adlandırılır. Burada 2 düzey sayısını, k ise faktör sayısını gösterir. Bu tasarımda $2x2x2x....x2 = 2^k$ tane düzey kombinasyonu vardır.
  2. k tane ana faktörün **üçer** tane düzeyinin olduğu faktöriyel tasarım 3^k **faktöriyel tasarım** olarak adlandırılır.
  3. Her biri 2 düzeye sahip iki faktörümüz var ise 2^2 faktöriyel tasarım olur. ($2^2 = 4$ tane düzey kombinasyonu olur.) Eğer her biri 2 düzeyli 3 faktörümüz var ise 2^3 faktöriyel tasarım olur.
     Benzer olarak farklı faktör ve düzey sayıları ile faktöriyel tasarımlar tanımlanabilir.
- Örnek bir veri yapısını inceleyelim.

Örnek 1: Bakır plakaların eğilmelerini araştırmak için sıcaklık ve levhaların bakır oranı olmak üzere iki faktörlü bir deney tasarlanmıştır. Bu deneyde her bir faktör düzeyi kombinasyonunda (her bir hücrede) 2 gözlem yapılmıştır. Bu deney sonucunda elde edilen veriler aşağıda verilmiştir.

	Bakır oranı (%)							
Sıcaklık (C°)	40		60		80		100	
50	17	20	16	21	24	22	28	27
75	12	9	18	13	17	12	27	31
100	16	12	18	21	25	23	30	23
125	21	17	23	21	23	22	29	31

Bu veri için 2 faktör ve her faktörün 4 düzeyi vardır. Bu tasarımda toplam 4x4=16 tane farklı düzey (deneme) kombinasyonu vardır. Bu tasarım 4^2 faktöriyel tasarımdır.

- İlk olarak iki ve üç ana faktörlü tasarımları arkasından 2^k faktöriyel tasarımları inceleyeceğiz.

İki Faktörlü Faktöriyel Tasarım: Matematiksel Model

A faktörünün a tane düzeyi ($i = 1, ..., a$), B fakörünün b tane düzeyi ($j = 1, ..., b$) ve her bir düzey için n tekrar ($k = 1, ..., n$) olduğu iki faktörlü faktöriyel tasarım için matematiksel model

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

biçiminde ifade edilir.

Burada,

y_{ijk} : A faktörünün i. ve B faktörünün j. düzeyindeki k. gözlem değeri
μ : genel ortalamayı,
τ_i : A faktörünün i. düzeyinin etkisini,
β_j : B faktörünün j. düzeyinin etkisini,
$(\tau\beta)_{ij}$: τ_i ve β_j düzeylerinin etkileşim etkisini,
ϵ_{ijk} : rastgele hata terimini

gösterir. **Varsayım:** $\epsilon_{ijk} \sim N(0, \sigma^2)$ biribirinden bağımsız r.d..

- (1) modelinde toplam *abn* tane gözlem vardır.
- (1) modelinde iki faktöründe sabit etkili olduğu varsayımı altında $\sum_{i=1}^{a} \tau_i = 0$, $\sum_{j=1}^{b} \beta_j = 0$ olur. Bu durumda etkileşim etkileri de $\sum_{i=1}^{a} (\tau\beta)_{ij} = \sum_{j=1}^{b} (\tau\beta)_{ij} = 0$ olur.

(1) modeli için veri yapısı aşağıdaki gibi olacaktır.

$$
\begin{array}{ll}
 & \overbrace{\hspace{28em}}^{B \text{ faktörü}} \\
A \text{ faktörü} & \left\{
\begin{array}{cccccccc}
 & 1 & 2 & . & . & . & b \\
1 & y_{111}, y_{121}, \dots, y_{11n} & y_{121}, y_{122}, \dots, y_{12n} & . & . & . & y_{1b1}, y_{1b2}, y_{1bn} \\
2 & y_{211}, y_{212}, \dots, y_{21n} & y_{121}, y_{122}, \dots, y_{12n} & . & . & . & y_{2b1}, y_{2b2}, y_{2bn} \\
. & . & . & . & . & . & . \\
. & . & . & . & . & . & . \\
. & . & . & . & . & . & . \\
a & y_{a11}, y_{a12}, \dots, y_{a1n} & y_{a21}, y_{a22}, \dots, y_{a2n}\cdot & . & . & . & y_{ab1}, y_{ab2}, y_{abn}
\end{array}
\right.
\end{array}
$$

Hipotezler

(1) modeli için aşağıdaki hipotezleri test edebiliriz.

- A faktörünün anlamlı olup olmadığı (veya faktörün düzeylerinin etkileri arasında fark olup olmadığı) için

$$\begin{aligned} H_0 &: \tau_1 = \tau_2 = \ldots = \tau_a = 0 \\ H_1 &: \text{En az bir } i \text{ için } \tau_i \neq 0 \end{aligned} \tag{2}$$

- B faktörünün anlamlı olup olmadığı için

$$\begin{aligned} H_0 &: \beta_1 = \beta_2 = \ldots = \beta_b = 0 \\ H_1 &: \text{En az bir } j \text{ için } \beta_j \neq 0 \end{aligned} \tag{3}$$

- A ve B faktörleri arasında etkileşimin anlamlı olup olmadığı için

$$\begin{aligned} H_0 &: (\tau\beta)_{11} = (\tau\beta)_{12} = \ldots = (\tau\beta)_{ab} = 0 \\ H_1 &: \text{En az bir } (i,j) \text{ için } (\tau\beta)_{ij} \neq 0 \end{aligned} \tag{4}$$

Kareler Toplamının Parçalanışı

(1) modeli için kareler toplamları

$$
\begin{aligned}
SS_T &= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk}-\bar{y}_{...})^2 = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}^2 - \frac{y_{...}^2}{N} \\
SS_A &= bn\sum_{i=1}^{a}(\bar{y}_{i..}-\bar{y}_{...})^2 = \frac{1}{bn}\sum_{i=1}^{a} y_{i..}^2 - \frac{y_{...}^2}{N} \\
SS_B &= an\sum_{j=1}^{b}\left(\bar{y}_{.j.}-\bar{y}_{...}\right)^2 = \frac{1}{an}\sum_{j=1}^{b} y_{.j.}^2 - \frac{y_{...}^2}{N} \\
SS_{AB} &= n\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\bar{y}_{ij.}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y}_{...}\right)^2 \\
&= \frac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij.}^2 - \frac{y_{...}^2}{N} - SS_A - SS_B
\end{aligned}
$$

$$SS_E = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(y_{ijk} - \overline{y}_{ij.}\right)^2 = SS_T - \left(\frac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij.}^2 - \frac{y_{...}^2}{N}\right)$$

olmak üzere $SS_T = SS_A + SS_B + SS_{AB} + SS_E$ biçiminde bileşenlerine ayrılır.

Kareler toplamında,

$$\begin{aligned}
y_{i..} &= \sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}, \quad \overline{y}_{i..} = \frac{y_{i..}}{bn}, \ i = 1, ..., a \\
y_{.j.} &= \sum_{i=1}^{a}\sum_{k=1}^{n} y_{ijk}, \quad \overline{y}_{.j.} = \frac{y_{.j.}}{an}, \ j = 1, ..., b \\
y_{ij.} &= \sum_{k=1}^{n} y_{ijk}, \quad \overline{y}_{ij.} = \frac{y_{ij.}}{n}, \ i = 1, ..., a, \ j = 1, ..., b \\
y_{...} &= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}, \quad \overline{y}_{...} = \frac{y_{...}}{N}, \quad N = abn
\end{aligned}$$

biçimindedir.

Test İstatistikleri ve Kurallar

1. Eğer $F_A = \frac{MS_A}{MS_E} = \frac{SS_A/(a-1)}{SS_E/ab(n-1)} > F_{(a-1),ab(n-1),\alpha}$ olur ise

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

hipotezi reddedilir. Bu durumda, A faktörünün düzeyleri arasında anlamlı bir farklılık vardır.

2. Eğer $F_B = \frac{MS_B}{MS_E} = \frac{SS_B/(b-1)}{SS_E/ab(n-1)} > F_{(b-1),ab(n-1),\alpha}$ olur ise $H_0 : \beta_1 = \beta_2 = ... = \beta_b = 0$ hipotezi reddedilir. Bu durumda, B faktörünün düzeyleri arasında anlamlı bir farklılık vardır.

3. Eğer $F_{AB} = \frac{MS_{AB}}{MS_E} = \frac{SS_{AB}/(a-1)(b-1)}{SS_E/ab(n-1)} > F_{(a-1)(b-1),ab(n-1),\alpha}$ olur ise

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

hipotezi reddedilir. Bu durumda, A ve B faktörleri arasındaki etkileşim etkisi anlamlıdır.

ANOVA Tablosu

Değişim kaynağı	Kareler toplamı	Serbestlik derecesi	Kareler ortalaması	F test değeri
A faktörü	SS_A	$a-1$	$MS_A = \frac{SS_A}{a-1}$	$F_A = \frac{MS_A}{MS_E}$
B faktörü	SS_B	$b-1$	$MS_B = \frac{SS_B}{b-1}$	$F_B = \frac{MS_B}{MS_E}$
AB Etkileşim	SS_{AB}	$(a-1)(b-1)$	$MS_{AB} = \frac{SS_{AB}}{(a-1)(b-1)}$	$F_{AB} = \frac{MS_{AB}}{MS_E}$
Hata	SS_E	$ab(n-1)$	$MS_E = \frac{SS_E}{(N-ab)}$	
Toplam	SS_T	$N-1$		

Not: Burada $N = abn$ olmak üzere $ab(n-1) = N - ab$ dir. Faktör düzeyleri arasında farklılık var ise her iki faktör için de çoklu karşılaştırma testleri yapılarak farklı olan düzeyler belirlenebilir.

Parametre Tahmini

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

biçiminde iki faktörlü faktöriyel tasarım modelinde bilinmeyen parametreler $\mu, \tau_i, \; i = 1, ..., a, \; \beta_j, \, j = 1, ..., b$ ve $(\tau\beta)_{ij}$ için *en küçük kareler (EKK)* tahmin edicileri

$$\begin{aligned}
\widehat{\mu} &= \overline{y}_{...} \\
\widehat{\tau}_i &= \overline{y}_{i..} - \overline{y}_{...}, \;\; i = 1, ..., a \\
\widehat{\beta}_j &= \overline{y}_{.j.} - \overline{y}_{...}, \;\; j = 1, ..., b \\
(\widehat{\tau\beta})_{ij} &= \overline{y}_{ij.} - \overline{y}_{i..} - \overline{y}_{.j.} + \overline{y}_{...}, \;\; i = 1, ..., a, \;\; j = 1, ..., b
\end{aligned} \tag{5}$$

biçiminde elde edilir.

Varsayım Kontrolü

Böylece, yanıt değişkeni y_{ijk}' nin tahmin edicisi

$$\begin{aligned} \widehat{y}_{ijk} &= \widehat{\mu} + \widehat{\tau}_i + \widehat{\beta}_j + (\widehat{\tau\beta})_{ij} \\ &= \overline{y}_{...} + (\overline{y}_{i..} - \overline{y}_{...}) + (\overline{y}_{.j.} - \overline{y}_{...}) + (\overline{y}_{ij.} - \overline{y}_{i..} - \overline{y}_{.j.} + \overline{y}_{...}) \\ &= \overline{y}_{ij.} \end{aligned} \quad (6)$$

olur. ij hücresindeki k. gözlem için tahmin değeri o hücredeki gözlemlerin ortalamasıdır, $\overline{y}_{ij.} = \frac{1}{n}\sum_{k=1}^{n} y_{ijk}$.

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

$$e_{ijk} = y_{ijk} - \widehat{y}_{ijk} = y_{ijk} - \overline{y}_{ij.} \quad (7)$$

biçiminde bulunur.
Tek yönlü ANOVA'da olduğu gibi e_{ijk} artıkları kullanılarak **modelin varsayımları** kontrol edilir.

Örnek 1: Bakır plakaların eğilmelerini araştırmak için sıcaklık ve levhaların bakır oranı olmak üzere iki faktörlü bir deney tasarlanmıştır. Bu deneyde her bir faktör düzeyi kombinasyonunda (her bir hücrede) 2 gözlem yapılmıştır. Bu deney sonucunda elde edilen veriler aşağıda verilmiştir.

	Bakır oranı (%)							
Sıcaklık (C°)	40		60		80		100	
50	17	20	16	21	24	22	28	27
75	12	9	18	13	17	12	27	31
100	16	12	18	21	25	23	30	23
125	21	17	23	21	23	22	29	31

$\alpha = 0.05$ anlamlılık düzeyinde

a) Her iki faktörün eğilme miktarını etkilediğini gösteren kanıt var mıdır?

b) Etkileşim grafiğini çiziniz. Faktörler arasında herhangi bir etkileşim var mıdır?

Örnek 2: Bir mühendis yüzey pürüzlülüğünün kullanılan boya türünden ve kuruma süresinden etkilendiğinden şüphelenir. 15, 20 ve 25 dakikalık üç kuruma süresini seçip iki boya kullanır.

	Kuruma süresi (dk)		
Boya	15	20	25
1	75	73	78
	64	60	85
	50	44	90
2	92	95	66
	86	73	45
	70	88	85

a) İki faktörlü faktöriyel tasarım kullanarak bu deney ile ilgili hipotezleri yazınız ve $\alpha = 0.05$ anlamlılık düzeyinde test ediniz?

b) Artıkları kullanarak varsayımları kontrol ediniz.

İki Faktörlü Etkileşimin Olmadığı Faktöriyel Tasarım

A faktörünün a tane düzeyi ($i = 1, ..., a$), B fakörünün b tane düzeyi ($j = 1, ..., b$) ve her bir düzey için n tekrar ($k = 1, ..., n$) olduğu iki faktörlü faktöriyel tasarımda faktörler arasında etkileşim olmadığında matematiksel model

$$y_{ijk} = \mu + \tau_i + \beta_j + \epsilon_{ijk}, \ \ i = 1, ..., a, \ \ j = 1, ..., b, \ \ k = 1, ..., n \qquad (8)$$

biçiminde ifade edilir.

Eğer faktörler arasında etkileşim olmadığı düşünülüyorsa bu model kullanılabilir. Faktörler arasında etkileşim varken bu modeli kullanırsak sonuçları yanlış yorumlama gibi bir durum ile karşılaşabiliriz.

Her Hücrede Bir Gözlemin Olduğu İki Faktörlü Faktöriyel Tasarım:

A faktörünün a tane düzeyi $(i = 1, ..., a)$, B fakörünün b tane düzeyi $(j = 1, ..., b)$ ve her bir düzey için **sadece** 1 **tekrar** olan iki faktörlü faktöriyel tasarım için matematiksel model

$$y_{ij} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ij}, \;\; i = 1, ..., a, \;\; j = 1, ..., b, \tag{9}$$

biçiminde ifade edilir.

- Burada tekrar sayısı $n = 1$ olduğundan $n > 1$ için verilen ANOVA tablosunda hatanın serbestlik derecesi 0 olacaktır.
- Ancak, bu mümkün olamayacağından bir tekrarlı faktöriyel tasarımlarda en yüksek dereceli etkileşim genellikle hata terimi olarak alınır. Çünkü yüksek dereceden etkileşimlerin genellikle önemsiz olduğu varsayılır.
- Bu durumda etkileşim etkisini test etmek için Tukey'in Toplamsallık Testi kullanılır.

ANOVA Tablosu

(9) modelinde faktörler sabit etkili olduğunda ANOVA tablosu aşağıdaki gibi olacaktır.

Değişim kaynağı	Kareler toplamı	Serbestlik derecesi	Kareler ortalaması	F test değeri
A faktörü	SS_A	$a-1$	$MS_A = \dfrac{SS_A}{a-1}$	$F_A = \dfrac{MS_A}{MS_E}$
B faktörü	SS_B	$b-1$	$MS_B = \dfrac{SS_B}{b-1}$	$F_B = \dfrac{MS_B}{MS_E}$
Hata veya AB	SS_E	$(a-1)(b-1)$	$MS_E = \dfrac{SS_E}{(a-1)(b-1)}$	
Toplam	SS_T	$ab-1$		

Not: Burada $SS_A = \frac{1}{b}\sum_{i=1}^{a} y_{i.}^2 - \frac{y_{..}^2}{ab}$, $SS_B = \frac{1}{a}\sum_{j=1}^{b} y_{.j}^2 - \frac{y_{..}^2}{ab}$, $SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij}^2 - \frac{y_{..}^2}{ab}$ ve $SS_E = SS_T - SS_A - SS_B$ biçimindedir.