Bir-Yönlü ANOVA (Tamamen Rasgele Tasarım)

İstatistiksel Deney Tasarımı
Birdal Şenoğlu
Şükrü Acıtaş

İçindekiler

Varyans analizi, üç ya da daha fazla grup ortalaması arasında istatistiksel olarak farklılık olup olmadığını test etmek amacıyla kullanılan bir yöntemdir. Grup ortalamalarının karşılaştırılması, deneyin sonunda bağımlı değişkende meydana gelen değişkenliğin ne kadarının faktör(ler)den, ne kadarının hatadan, v.b. kaynaklandığının belirlenmesi, bir başka deyişle toplam varyansın bileşenlerine ayrılması yardımıyla yapılır.

Bir-yönlü ANOVA (one-way ANOVA), ANOVA nın özel bir halidir ve tasarımlar içinde en basit olanıdır. Etkisi araştırılmak istenen yalnız "bir" tane faktör (bağımsız değişken) olduğunda kullanıldığından dolayı, **bir-faktörlü ANOVA** (one-factor ANOVA) olarak da adlandırılır.

Bir-yönlü ANOVA, deney birimleri homojen olduğunda kullanılması önerilen en uygun tasarımdır. Buradaki homojen sözcüğü "mümkün olduğunca benzer" anlamında kullanılmaktadır, çünkü doğada aynı/özdeş deney birimleri bulunmaz, bkz. Hinkelmann & Kempthorne (1994).

Deney birimleri arasındaki bu küçük farklılıklar "rasgele" farklılıklar olarak adlandırılır. Deney birimleri homojen olarak kabul edildiği için rasgeleleştirme üzerinde hiçbir kısıt yoktur. Bu nedenle birçok kaynakta bu tasarıma **tamamen rasgele tasarım** (completely randomized design) adı da verilir.

Bir-yönlü ANOVA da rasgeleleştirme işlemi iki aşamada gerçekleştirilir. İlk aşamada parsellere 1 den 15 e kadar numara verilir, ikinci aşamada S1, S2 ve S3 denemeleri rasgele olarak bu parsellere atanır/uygulanır. Rasgeleleştirme işlemi sonucu aşağıdaki tabloya benzer bir sonuç elde edilebilir.

S1(1)	S1(2)	S3(3)	S2(4)	S2(5)
S3(6)	S3(7)	S2(8)	S1(9)	S3(10)
S2(11)	S2(12)	S1(13)	S1(14)	S3(15)

Bu tablodan da anlaşılacağı gibi S1 denemesi 1, 2, 9, 13, 14 numaralı, S2 denemesi 4, 5, 8, 11, 12 numaralı ve S3 denemesi de 3, 6, 7, 10 ve 15 numaralı parsellere atanmıştır/uygulanmıştır.

Bir-yönlü ANOVA için matematiksel model

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \quad i = 1, 2, \cdots, a; \quad j = 1, 2, \cdots n \tag{1}$$

şeklinde ifade edilir. Burada,

y_{ij},	$i-$inci denemedeki $j-$inci gözlem değerini,
μ,	genel ortalamayı,
τ_i,	$i-$inci denemenin etkisini ve
ε_{ij},	rasgele hata terimlerini

gösterir.

(1) modeli sabit etkili bir modeldir, bir başka deyişle $\sum_{i=1}^{a} \tau_i = 0$ olduğu varsayılır.

Bir-Yönlü ANOVA modeli için veri yapısı aşağıda gösterildiği gibidir.

	Gözlemler					
Denemeler	1	2	$\cdots$	n	*Toplam*	*Ortalama*
1	y_{11}	y_{12}	$\cdots$	y_{1n}	$y_{1\cdot}$	$\bar{y}_{1\cdot}$
2	y_{21}	y_{22}	$\cdots$	y_{2n}	$y_{2\cdot}$	$\bar{y}_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$	$\vdots$	$\vdots$
a	y_{a1}	y_{a2}	$\cdots$	y_{an}	$y_{a\cdot}$	$\bar{y}_{a\cdot}$
					$y_{\cdot\cdot}$	$\bar{y}_{\cdot\cdot}$

Burada,

$$y_{i\cdot} = \sum_{j=1}^{n} y_{ij} \quad \text{ve} \quad \bar{y}_{i\cdot} = \frac{y_{i\cdot}}{n} \quad , \quad i = 1, 2, \cdots, a \tag{2}$$

sırasıyla $i-$inci denemedeki gözlemlerin toplamını ve ortalamasını gösterir. Ayrıca, $N = an$ toplam gözlem sayısını göstermek üzere;

$$y_{\cdot\cdot} = \sum_{i=1}^{a} \sum_{j=1}^{n} y_{ij} \quad \text{ve} \quad \bar{y}_{\cdot\cdot} = \frac{y_{\cdot\cdot}}{N} \tag{3}$$

sırasıyla tüm gözlemlerin toplamı ve tüm gözlemlerin ortalaması olarak tanımlanır.

ANOVA tekniği kullanılarak yapılan analizler bazı temel varsayımlara dayanır. Bu varsayımlar aşağıda açıklanmıştır.

1. Varsayım ε_{ij} hata terimleri 0 ortalama ve σ^2 varyans ile normal dağılıma sahiptir.

2. Varsayım Hata terimlerinin varyansları homojendir.

3. Varsayım Hata terimleri birbirinden bağımsızdır.

Bu üç varsayım kısaca

$$\varepsilon_{ij} \sim NID(0, \sigma^2) \tag{4}$$

şeklinde gösterilir.

Bir parametrenin LS tahmin edicisi, modeldeki hata terimlerin kareleri toplamını bu parametreye göre minimum yapan değerdir. (1) modeli için hata kareler toplamı S ile gösterilsin; bir başka ifade ile

$$S = \sum_{i=1}^{a}\sum_{j=1}^{n}\varepsilon_{ij}^2 = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\mu-\tau_i)^2 \tag{5}$$

olsun. Bu durumda, model parametrelerinin LS tahmin edicileri

$$\begin{aligned}\frac{\partial S}{\partial \mu} &= (-2)\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\mu-\tau_i)=0,\\ \frac{\partial S}{\partial \tau_i} &= (-2)\sum_{j=1}^{n}(y_{ij}-\mu-\tau_i)=0\end{aligned} \tag{6}$$

denklem sisteminin çözümüdür.

(1) modelinin, sabit etkili bir model olduğu göz önüne alınarak, (6) denklem sistemi çözülürse, μ ve τ_i parametrelerinin LS tahmin edicileri, sırasıyla

$$\tilde{\mu} = \bar{y}_{..} \tag{7}$$
$$\tilde{\tau}_i = \bar{y}_{i.} - \bar{y}_{..} \tag{8}$$

olur.

Hatanın varyansı σ^2 nin LS tahmin edicisi,

$$\tilde{\sigma}^2 = \frac{\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \tilde{\mu} - \tilde{\tau}_i)^2}{N} \tag{9}$$

$$= \frac{\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..} - \bar{y}_{i.} + \bar{y}_{..})^2}{N} \tag{10}$$

$$= \frac{\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2}{N} \tag{11}$$

dir.

(11) denkleminde verilen LS tahmin edicisi yanlıdır. Gerekli yan düzeltmesi yapılırsa

$$\tilde{\sigma}^2 = \frac{\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2}{N - a} \tag{12}$$

olur.

Not

Benzer sonuçlar **en çok olabilirlik** (Maximum Likelihood-ML) yöntemi kullanılarak da elde edilebilir. Bir parametrenin ML tahmin edicisi, olabilirlik fonksiyonunu bu parametreye göre maksimum yapan değerdir. Normallik varsayımı altında, LS ve ML tahmin edicileri aynı sonucu verdiğinden, ML tahmin edicilerine bu bölümde ayrıca değinilmemiştir. Ancak, belirtmek gerekir ki; σ^2 nin ML tahmin edicisi de yanlıdır ve gerekli yan düzeltmesi yapıldığında bu tahmin edici (12) denklemindeki gibi ifade edilir.

(1) modeli, yeniden parametrelendirme yapılarak

$$y_{ij} = \mu_i + \varepsilon_{ij}, \quad i = 1, 2, \cdots, a; \quad j = 1, 2, \cdots n \tag{13}$$

olarak da ifade edilebilir. Burada $\mu_i = \mu + \tau_i$ dir. (13) modeli, ortalamalar modeli olarak isimlendirilmektedir (Montgomery, 2001).
(13) modeli için **Fisher bilgi matrisi** (Fisher Information matrix-$\boldsymbol{I}$)

$$\boldsymbol{I} = \begin{pmatrix} -E\left(\frac{\partial^2 \ln L}{\partial \mu_i^2}\right) & -E\left(\frac{\partial^2 \ln L}{\partial \mu_i \partial \sigma^2}\right) \\ -E\left(\frac{\partial^2 \ln L}{\partial \sigma^2 \partial \mu_i}\right) & -E\left(\frac{\partial^2 \ln L}{\partial (\sigma^2)^2}\right) \end{pmatrix} \tag{14}$$

olarak tanımlanır.

Gerekli işlemler yapıldığında, (*bkz. İstatistiksel Deney Tasarımı, sayfa 15-17.*)

$$\boldsymbol{I} = \begin{pmatrix} \dfrac{n}{\sigma^2} & 0 \\ 0 & \dfrac{N}{2\sigma^4} \end{pmatrix} \tag{15}$$

olur.
Fisher bilgi matrisinin tersi alınarak

$$\boldsymbol{I}^{-1} = \begin{pmatrix} \dfrac{\sigma^2}{n} & 0 \\ 0 & \dfrac{2\sigma^4}{N} \end{pmatrix} \tag{16}$$

bulunur.

Teorem

(13) *modelinde,* $\mu_i = \mu + \tau_i$ *olarak tanımlanan* μ_i *parametresinin LS tahmin edicisi* $\tilde{\mu}_i$*,* μ_i *nin yansız bir tahmin edicisidir.*

Teorem

(13) *modelinde,*

(i) $\mu_i = \mu + \tau_i$ *parametresi için RCLB* σ^2/n *dir.*

(ii) σ^2 *parametresi için RCLB* $2\sigma^4/N$ *dir.*

Teorem

(13) *modelinde,* μ_i *parametresinin LS (ML) tahmin edicisi* $\tilde{\mu}_i$ *nın varyansı,* $\dfrac{\sigma^2}{n}$ *dir.*

Tanım

Varyansı RCLB ye eşit olan yansız tahmin ediciye, ***minimum varyans sınırı*** *(Minimum Variance Bound-MVB) tahmin edicisi denir.*

Bu tanımın bir sonucu olarak,

$$Var(\tilde{\mu}_i) = RCLB(\mu_i)$$

olduğundan $\tilde{\mu}_i$ bir MVB tahmin edicisidir.

Sonuç

$\tilde{\mu}_i$, μ_i *ortalama ve* σ^2/n *varyansı ile normal dağılıma sahiptir.*

Teorem

$X_1, X_2, \cdots, X_n$, olasılık (yoğunluk) fonksiyonu $f(x|\theta)$ olan bir dağılımdan alınan örneklem ve bu örneklemin olabilirlik fonksiyonu

$$L(\theta|x) = \prod_{i=1}^{n} f(x_i|\theta)$$

olsun. $\mathcal{T}(X)$ tahmin edicisinin, beklenen değeri $\rho(\theta)$ ve varyansı $\frac{1}{m(\theta)}$ olan bir MVB tahmin edicisi olması için gerekli ve yeterli koşul; log-olabilirlik fonksiyonunun ilgili parametreye göre kısmi türevinin

$$\frac{\partial \ln L}{\partial \theta} = m(\theta)\left(\mathcal{T}(X) - \rho(\theta)\right) \tag{17}$$

şeklinde yazılabilmesidir.

Teoremin kanıtı için bakınız Kendall & Stuart (1967).

Teorem

(13) *modelinde,* μ_i *parametresinin LS tahmin edicisi* $\tilde{\mu}_i = \bar{y}_{i\cdot}$*, beklenen değeri* μ_i *ve varyansı* σ^2/n *olan bir MVB tahmin edicisidir.*

(1) modelinde, amaç denemeler arasında anlamlı bir farklılık olup olmadığını sınamaktır. Bu durum için sıfır hipotezi, üç farklı şekilde ifade edilebilir;

$$\begin{array}{lll} (i) & H_0 & : \text{Denemeler arasında anlamlı bir fark yoktur} \\ (ii) & H_0 & : \tau_1 = \tau_2 = \cdots = \tau_a = 0 \\ (iii) & H_0 & : \mu_1 = \mu_2 = \cdots = \mu_a = \mu. \end{array} \tag{18}$$

Açıktır ki, (i), (ii) ve (iii) hipotezleri birbirine denktir.

(18) hipotezini sınamak için kullanılan test istatistiği iki farklı yolla elde edilebilir:

1. Genel Kareler Toplamının Parçalanışı
2. İndirgenmiş Model-Tam Model Yaklaşımı

Deneme ortalamalarının birbirine eşit olduğunu ifade eden (18) hipotezi

$$\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{..})^2 \tag{19}$$

genel kareler toplamının, deneme kareler toplamı ve hata kareler toplamı olarak bileşenlerine ayrılmasıyla elde edilen test istatistiği yardımıyla sınanır.
(19) ifadesine $\bar{y}_{i.}$ terimi bir eklenip bir çıkartıldığında, bir başka deyişle

$$\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.}+\bar{y}_{i.}-\bar{y}_{..})^2 \tag{20}$$

olarak yazıldığında toplamın değeri değişmez.

(20) ifadesinde parantez içi $y_{ij} - \bar{y}_{i.}$ ve $\bar{y}_{i.} - \bar{y}_{..}$ olmak üzere iki gruba ayrılır ve karesel ifade açılırsa

$$\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2 = \sum_{i=1}^{a}\sum_{j=1}^{n}(\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2 \qquad (21)$$

elde edilir.

Eğer,

$$
\begin{aligned}
SS_{Toplam} &= \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{..})^2 \\
SS_{Deneme} &= \sum_{i=1}^{a}\sum_{j=1}^{n}(\bar{y}_{i.}-\bar{y}_{..})^2 = n\sum_{j=1}^{n}(\bar{y}_{i.}-\bar{y}_{..})^2 \\
SS_{Hata} &= \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2
\end{aligned}
\tag{22}
$$

denirse, (21) ifadesi kısaca

$$SS_{Toplam} = SS_{Deneme} + SS_{Hata} \tag{23}$$

olarak ifade edilir.

Dikkat edilirse, SS_{Toplam}, SS_{Deneme} ve SS_{Hata} sırasıyla gözlemlere, denemelere ve hata terimlerine ait varyansların paylarını ifade eder. SS değerleri, ilgili serbestlik derecelerine bölündüğünde varyanslar elde edilir.

Test İstatistiği

(1) modelinde, (18) hipotezini sınamak için

$$F_{Deneme} = \frac{SS_{Deneme} \Big/ (a-1)}{SS_{Hata} \Big/ (N-a)} = \frac{MS_{Deneme}}{MS_{Hata}} \tag{24}$$

test istatistiği kullanılır.

Teorem

(1) *modelinde,* H_0 *hipotezi altında,* F_{Deneme} *test istatistiği,* $a-1$ *ve* $N-a$ *serbestlik dereceli merkezi* $\boldsymbol{F}$ *dağılımına sahiptir.*

Teoremin kanıtı aşağıda ifade edilen ve doğrusal modeller teorisinde önemli bir yere sahip olan Cochran Teoremi yardımıyla yapılır.

Theorem (Cochran Teoremi)

$Y_1, Y_2, \cdots, Y_n$ iid $N(0,1)$ dağılımına sahip rasgele değişkenler olsun. Q_i ler, Y_i lerin doğrusal kombinasyonlarının kareler toplamlarını göstermek üzere

$$\sum_{i=1}^{n} Y_i^2 = Q_1 + Q_2 + \cdots + Q_k, \quad k \leq n$$

eşitliği yazılabilsin ve r_i, Q_i $(i = 1, 2, \cdots, k)$ rasgele değişkenlerinin serbestlik derecelerini göstersin.
Bu durumda, aşağıdaki ifadelerden herhangi ikisinin doğru olması üçüncü ifadenin de doğru olmasını gerektir.

(i) $n = r_1 + r_2 + \cdots + r_k$

(ii) $Q_1, Q_2, \cdots, Q_k$ *rasgele değişkenleri bağımsızdır.*

(iii) Q_i, $(i = 1, 2, \cdots, k)$ *rasgele değişkenleri* $\chi^2_{r_i}$ *dağılımına sahiptir.*

bkz. Cochran (1934).

Karar

F_{Deneme} test istatistiğinin değeri, α anlam düzeyinde $a-1$ ve $N-a$ serbestlik dereceli $\boldsymbol{F}$ tablo değerinden daha büyükse sıfır hipotezi reddedilir. Bir başka deyişle,

$$F_{Deneme} > \boldsymbol{F}_{\alpha;a-1;N-a}$$

ise "*Denemeler arasında anlamlı bir farklılık vardır*" denir. ♣

İndirgenmiş Model

(18) hipotezi altında, (1) modeli

$$y_{ij} = \mu + \varepsilon_{ij}, \quad i = 1, 2, \cdots, a; \quad j = 1, 2, \cdots, n \tag{25}$$

şeklinde ifade edilir.
(25) modeline *indirgenmiş model* denir.
İndirgenmiş modelin hata kareler toplamı $SS(R)$ ile gösterilir ve

$$SS(R) = (N-1)\tilde{\sigma}^2 = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2 \tag{26}$$

olarak ifade edilir.

Teorem

(26) *de verilen* $SS(R)$, *indirgenmiş model hata kareler toplamı olmak üzere, normallik varsayımı altında*

$$\frac{(N-1)\tilde{\sigma}^2}{\sigma^2} = \frac{\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{..})^2}{\sigma^2} = \frac{SS(R)}{\sigma^2} \tag{27}$$

ifadesi $N-1$ *serbestlik dereceli ki-kare dağılımına sahiptir.*

Tam Model

Tam model, (1) modelidir. Dolayısıyla $SS(F)$ ile gösterilen tam modelin hata kareler toplamı

$$SS(F) = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \tilde{\mu} - \tilde{\tau}_i)^2 = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2 \qquad (28)$$

dir.
Dikkat edilirse, $SS(F)$ ile SS_{Hata} birbirine denk iki gösterimdir.

Teorem

(28) de verilen $SS(F)$, tam modelin hata kareler toplamı olmak üzere, normallik varsayımı altında

$$\frac{(N-a)\tilde{\sigma}^2}{\sigma^2} = \frac{\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2}{\sigma^2} = \frac{SS(F)}{\sigma^2} \tag{29}$$

ifadesi $N-a$ serbestlik dereceli ki-kare dağılımına sahiptir.

Test İstatistiği

(1) modelinde, (18) hipotezini sınamak için

$$F_{Deneme} = \frac{[SS(R) - SS(F)] \Big/ (a-1)}{SS(F) \Big/ (N-a)} \tag{30}$$

test istatistiği kullanılır.

Teorem

(1) *modelinde,* H_0 *hipotezi altında,* F_{Deneme} *test istatistiği,* $a-1$ *ve* $N-a$ *serbestlik dereceli merkezi* $\boldsymbol{F}$ *dağılımına sahiptir.*

Yorum

Sıfır hipotezinin doğruluğu kısıtı altında yazılan $SS(R)$ ifadesi ile herhangi bir kısıt olmaksızın yazılan $SS(F)$ ifadesi arasındaki fark, denemeler arasında anlamlı bir farklılık olup olmadığını gösterir. $SS(R) - SS(F)$ farkı küçük ise denemeler arasında anlamlı bir farklılık yoktur; $SS(R) - SS(F)$ farkı büyük ise denemeler arasında anlamlı bir farklılık vardır, denir.

Yukarıda elde edilen bilgiler ışığında bir-yönlü ANOVA tablosu, aşağıda gösterildiği gibi oluşturulur.

Bir-yönlü ANOVA tablosu

Kaynak	df	SS	MS	F
Denemeler	$a-1$	SS_{Deneme}	MS_{Deneme}	F_{Deneme}
Hata	$N-a$	SS_{Hata}	MS_{Hata}	
Genel	$N-1$	SS_{Toplam}		

Deneme kareler ortalaması,

$$MS_{Deneme} = \frac{SS_{Deneme}}{a-1} = \frac{n\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2}{a-1} \tag{31}$$

dir. Burada, (1) modeli kullanılarak

$$\begin{aligned}
\bar{y}_{i.} &= \frac{\sum_{j=1}^{n} y_{ij}}{n} && (32)\\
&= \frac{\sum_{j=1}^{n}(\mu + \tau_i + \varepsilon_{ij})}{n} && (33)\\
&= \frac{n\mu + n\tau_i + \sum_{j=1}^{n}\varepsilon_{ij}}{n} && (34)\\
&= \mu + \tau_i + \bar{\varepsilon}_{i.} && (35)
\end{aligned}$$

ve

$$\bar{y}_{..} = \frac{\sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}}{N} \tag{36}$$

$$= \frac{\sum_{i=1}^{a}\sum_{j=1}^{n} (\mu + \tau_i + \varepsilon_{ij})}{N} \tag{37}$$

$$= \mu + \bar{\varepsilon}_{..} \tag{38}$$

bulunur.

(35) ve (38) eşitlikleri (31) de yerlerine yazılarak

$$
\begin{aligned}
MS_{Deneme} &= \frac{n}{a-1}\sum_{i=1}^{a}(\bar{y}_{i.}-\bar{y}_{..})^2 \\
&= \frac{n}{a-1}\sum_{i=1}^{a}(\mu+\tau_i+\bar{\varepsilon}_{i.}-\mu-\bar{\varepsilon}_{..})^2 \\
&= \frac{n}{a-1}\sum_{i=1}^{a}[\tau_i+(\bar{\varepsilon}_{i.}-\bar{\varepsilon}_{..})]^2 \\
&= \frac{n}{a-1}\left[\sum_{i=1}^{a}\tau_i^2+\sum_{i=1}^{a}(\bar{\varepsilon}_{i.}-\bar{\varepsilon}_{..})^2+2\sum_{i=1}^{a}\tau_i(\bar{\varepsilon}_{i.}-\bar{\varepsilon}_{..})\right]
\end{aligned}
$$

elde edilir.

Son ifadede beklenen değer alınırsa,

$$E(MS_{Deneme}) = \frac{n}{a-1}\left[\sum_{i=1}^{a}\tau_i^2 + \sum_{i=1}^{a}\underbrace{E(\bar{\varepsilon}_{i}. - \bar{\varepsilon}..)^2}_{Var(\bar{\varepsilon}_{i}.-\bar{\varepsilon}..)} + 2\sum_{i=1}^{a}\tau_i\underbrace{E(\bar{\varepsilon}_{i}. - \bar{\varepsilon}..)}_{0}\right] \quad (39)$$

olur. Burada,

$$\varepsilon_{ij} \sim N(0,\sigma^2) \quad (40)$$
$$\bar{\varepsilon}_{i}. \sim N(0,\sigma^2/n) \quad (41)$$
$$\bar{\varepsilon}.. \sim N(0,\sigma^2/N) \quad (42)$$
$$(43)$$

dir.

Bunların bir sonucu olarak,

$$E(\bar{\varepsilon}_{i.} - \bar{\varepsilon}_{..}) = 0 \tag{44}$$

ve

$$\begin{aligned}
Var(\bar{\varepsilon}_{i.} - \bar{\varepsilon}_{..}) &= Var(\bar{\varepsilon}_{i.}) + Var(\bar{\varepsilon}_{..}) - 2cov(\bar{\varepsilon}_{i.}, \bar{\varepsilon}_{..}) && (45)\\
&= \frac{\sigma^2}{n} + \frac{\sigma^2}{N} - \frac{2}{a} Var(\bar{\varepsilon}_{i.}) && (46)\\
&= \frac{\sigma^2}{n} + \frac{\sigma^2}{N} - \frac{2}{a}\frac{\sigma^2}{n} && (47)\\
&= \frac{(a-1)\sigma^2}{N} && (48)
\end{aligned}$$

bulunur.

Böylece,

$$\begin{aligned} E(MS_{Deneme}) &= \frac{n}{a-1}\left[\sum_{i=1}^{a}\tau_i^2 + \sum_{i=1}^{a}\frac{(a-1)\sigma^2}{N}\right] && (49) \\ &= \frac{n}{a-1}\sum_{i=1}^{a}\tau_i^2 + \frac{n}{a-1}a\frac{(a-1)\sigma^2}{N} && (50) \\ &= \sigma^2 + \frac{n\sum_{i=1}^{a}\tau_i^2}{a-1} && (51) \end{aligned}$$

elde edilir.
(51) denkleminden görülmektedir ki, sıfır hipotezinin doğru olması halinde MS_{Deneme}, σ^2 nin yansız bir tahmin edicisidir.

Hata kareler ortalaması,

$$MS_{Hata} = \frac{SS_{Hata}}{N-a} = \frac{\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2}{N-a} \quad (52)$$

ifadesinde, $y_{ij} = \mu + \tau_i + \varepsilon_{ij}$ ve $\bar{y}_{i.} = \mu + \tau_i + \bar{\varepsilon}_{i.}$ ifadeleri yerlerine yazıldığında

$$MS_{Hata} = \frac{1}{N-a}\sum_{i=1}^{a}\sum_{j=1}^{n}(\mu + \tau_i + \varepsilon_{ij} - \mu - \tau_i - \bar{\varepsilon}_{i.})^2 \quad (53)$$

$$= \frac{1}{N-a}\sum_{i=1}^{a}\sum_{j=1}^{n}(\varepsilon_{ij} - \bar{\varepsilon}_{i.})^2 \quad (54)$$

olur.

Beklenen değer alındığında,

$$E(MS_{Hata}) = \frac{1}{N-a}\sum_{i=1}^{a}\sum_{j=1}^{n}\underbrace{E(\varepsilon_{ij}-\bar{\varepsilon}_{i}.)^2}_{Var(\varepsilon_{ij}-\bar{\varepsilon}_{i}.)} \tag{55}$$

dir. Burada,

$$\begin{array}{lcl} \varepsilon_{ij} & \sim & N(0,\sigma^2) \\ \bar{\varepsilon}_{i}. & \sim & N(0,\sigma^2/n) \end{array}$$

olduğu hatırlanırsa

$$\begin{aligned} Var(\varepsilon_{ij}-\bar{\varepsilon}_{i}.) &= Var(\varepsilon_{ij}) + Var(\bar{\varepsilon}_{i}.) - 2cov(\varepsilon_{ij},\bar{\varepsilon}_{i}.) && (56) \\ &= \sigma^2 + \frac{\sigma^2}{n} - \frac{2}{n}Var(\varepsilon_{ij}) && (57) \\ &= \sigma^2 + \frac{\sigma^2}{n} - \frac{2\sigma^2}{n} && (58) \\ &= \frac{(n-1)\sigma^2}{n} && (59) \end{aligned}$$

bulunur.

Dolayısıyla,

$$E(MS_{Hata}) = \frac{1}{N-a}\sum_{i=1}^{a}\sum_{j=1}^{n}\frac{(n-1)\sigma^2}{n} \tag{60}$$

$$= \frac{1}{N-a}N\frac{(n-1)\sigma^2}{n} \tag{61}$$

$$= \frac{1}{a(n-1)}an\frac{(n-1)\sigma^2}{n} \tag{62}$$

$$= \sigma^2 \tag{63}$$

elde edilir.
(63) eşitliğinden görülmektedir ki, MS_{Hata}, σ^2 nin yansız bir tahmin edicisidir.

En az bir denemedeki gözlem sayısı diğerlerinden farklı ise bu tasarıma dengeli olmayan tasarım denir. Bu tanıma göre dengeli olmayan bir-yönlü ANOVA nın matematiksel modeli,

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \quad i = 1, 2, \cdots, a; \quad j = 1, 2, \cdots, n_i \tag{64}$$

şeklinde ifade edilir. Parametrelerin yorumları (1) modelinde olduğu gibidir. Bu modelde tek fark, $i-$inci denemede n_i gözlem olmasıdır. (64) modelinde

$$\sum_{i=1}^{a} n_i \tau_i = 0 \tag{65}$$

olduğu varsayılır.

- (64) modelinde, parametrelerin LS tahmin edicileri ve denemeler arasında anlamlı bir farklılık olup olmadığı hipotezini sınamak için gerekli test istatistiği dengeli tasarımda olduğu gibi bulunur.
- Formüller tamamen aynı olup gözlem sayılarının n_i olmasından dolayı küçük nüans farkları vardır.

LS tahmin edicileri

$$\begin{aligned} \tilde{\mu} &= \bar{y}_{..} \\ \tilde{\tau}_i &= \bar{y}_{i.} - \bar{y}_{..} \end{aligned}$$

olup, burada

$$y_{i.} = \sum_{j=1}^{n_i} y_{ij} \quad , \quad \bar{y}_{i.} = \frac{y_{i.}}{n_i} \quad , \quad i = 1, 2, \cdots, a$$

ve $N = \sum_{i=1}^{a} n_i$ toplam gözlem sayısını göstermek üzere

$$y_{..} = \sum_{i=1}^{a} \sum_{j=1}^{n_i} y_{ij} \quad , \quad \bar{y}_{..} = \frac{y_{..}}{N}$$

dir.

Denemeler arasında anlamlı bir farklılık olup olmadığını sınamak için

$$F_{Deneme} = \frac{SS_{Deneme} \big/ (a-1)}{SS_{Hata} \big/ (N-a)} = \frac{MS_{Deneme}}{MS_{Hata}}$$

test istatiği kullanılır. Burada,

$$\begin{aligned} SS_{Toplam} &= \sum_{i=1}^{a} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{..})^2 \\ SS_{Deneme} &= \sum_{i=1}^{a} \sum_{j=1}^{n_i} (\bar{y}_{i.} - \bar{y}_{..})^2 = \sum_{i=1}^{a} n_i (\bar{y}_{i.} - \bar{y}_{..})^2 \\ SS_{Hata} &= \sum_{i=1}^{a} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2 \end{aligned}$$

dir.
Dengeli tasarım için verilen diğer notasyonlar, yukarıdakine benzer şekilde dengeli olmayan tasarım için de yazılabilir.

(1) modelinde, çeşitli nedenlerden dolayı bir ya da birden fazla gözlem, kaybolmuş olabilir veya gözlemlenemeyebilir. Böyle bir durumda deneme etkilerinin toplamı, bir başka deyişle,

$$\sum_{i=1}^{a} \tau_i$$

toplamı sıfıra eşit olmayacağından genel kareler toplamı, deneme kareler toplamı ve hata kareler toplamı olarak bileşenlerine ayrılamaz (Hicks & Turner, 1999).

(1) modelinde, $i-$inci denemedeki, $j-$inci gözlemin (y_{ij}) kayıp olduğunu varsayalım. Bu kayıp gözlem m ile gösterilsin. Bu durumda veri yapısı, aşağıda gösterilen tablodaki gibi olur.

	Gözlemler						
Denemeler	1	2	$\cdots$	j	$\cdots$	n	Toplam
1	y_{11}	y_{12}	$\cdots$	$\cdots$	$\cdots$	y_{1n}	$y_{1\cdot}$
2	y_{21}	y_{22}	$\cdots$	$\cdots$	$\cdots$	y_{2n}	$y_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\vdots$	$\vdots$
i	$\vdots$	$\vdots$	$\cdots$	m	$\cdots$	$\vdots$	$y_{i\cdot}^{*} + m$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\vdots$	$\vdots$
a	y_{a1}	y_{a2}	$\cdots$	$\cdots$	$\cdots$	y_{an}	$y_{a\cdot}$

(1) modelinde, kayıp gözlem(ler) olduğunda öncelikle bu kayıp gözlem(ler) LS yöntemi ile tahmin edilir, daha sonra bu tahmin edici(ler)den yararlanılarak denemelerin anlamlılığı sınanır.
Kayıp gözlem m nin LS tahmin edicisi, gerekli işlemler yapıldığında

$$\tilde{m} = \frac{y_{i.}^{*}}{n-1} \tag{66}$$

bulunur. Burada,

y_{ij}^{*}, m dışındaki diğer gözlemleri
$y_{i.}^{*}$, $i-$inci denemedeki m dışındaki diğer gözlemlerin toplamını

gösterir.

- Bilinmeyen y_{ij} gözleminin yerine $\tilde{m}$ tahmin edicisi yazılır ve (22) eşitliklerinde verilen kareler toplamları hesaplanır.
- Daha sonra kareler ortalamaları da bulunarak ANOVA tamamlanır.
- Burada dikkat edilmesi gereken en önemli husus, hatanın serbestlik derecesinin $N - a - 1$ olduğudur; çünkü bir tane bilinmeyen gözlem vardır.
- Belirtmek gerekir ki, k tane kayıp gözlem olduğunda bir-yönlü ANOVA tablosunda hataya ait serbestlik derecesi $df = N - a - k$ olur.

Bu durumda ANOVA tablosu, aşağıda gösterildiği gibi oluşturulur.

Bir tane kayıp gözlem olduğunda bir-yönlü ANOVA tablosu.

Kaynak	df	SS	MS	F
Denemeler	$a-1$	SS_{Deneme}	MS_{Deneme}	F_{Deneme}
Hata	$N-a-1$	SS_{Hata}	MS_{Hata}	
Genel	$N-2$	SS_{Toplam}		