



**T.C.**  
**ONDOKUZ MAYIS ÜNİVERSİTESİ**

**FEN EDEBİYAT FAKÜLTESİ**  
**İSTATİSTİK BÖLÜMÜ**

**İST.357 VARYANS ANALİZİ**

**DOÇ. DR. YÜKSEL ÖNER**

**1. Hafta**

## BÖLÜM 1

### 1.1 VARYANS ANALİZİNE GİRİŞ: TEMEL KAVRAMLAR

Varyans analizinde kullanılan temel kavramlardan bazılarının tanım ve açıklamaları bu kısımda verilecektir.

**Tanım 1.1** Kontrol altındaki çeşitli durumların/koşulların, deney birimlerinin bilinmeyen karakteristik özellikleri üzerindeki etkisini test etmek amacıyla uygulanan bir işleme veya sürece **deney (experiment)** denir.

**Tanım 1.2** Deney birimlerinin maruz kalacağı kontrol altındaki durumların/koşulların düzenlenmesine **deney tasarımı (experimental design)** denir

Bu durum/koşulların düzenlenmesi belirli ilkeler ve kurallar çerçevesinde geliştirilen tasarımlar yardımıyla yapılır.

**Tanım 1.3** Deney birimlerinin hakkında bilgi elde edilmek istenen karakteristik özelliklerine **bağımlı değişken (dependent variable)** denir. Bu değişken genellikle nicel türden ve en az eşit aralıklı ölçme düzeyine sahip olabileceği gibi, bazen nitel türden sınıflama yada sıralama düzeyinde ölçülen bir değişkende olabilir.

**Tanım 1.4** Deney tasarımında ilgilenilen durumlara veya koşullara ya da bağımlı değişkenin alacağı değerleri etkileyen genellikle kontrol edilebilir deneysel değişkenlere **bağımsız değişken (independent variable)** veya **faktör (factor)** denir. Bu değişken nitel veya nicel türden sınıflama ya da sıralama ölçme düzeyine sahip olan bir değişkendir.

Bağımsız değişken olarak adlandırılan faktör/faktörlerle ilgili olarak aşağıdaki açıklamalar önem arz eder.

i) Bir faktörün alabileceği değerlere **faktör düzeyi** veya **deneme (treatment)** adı verilir. Faktör düzey sayısı iki ya da daha fazla olabilir. Eğer deneyde tek faktör ve bu faktöründe  $k$  tane düzeyi varsa, o zaman bu deney  $k$  tane denemeden oluşmaktadır. Bu denemeler birbirinden bağımsız olan grupları meydana getirir.

ii) Bir faktörün düzey sayıları deneyi yapan kişinin kontrolü altında olabileceği gibi kontrolü dışında da olabilir. Eğer faktörün düzey sayısı deneyi yapan kişi tarafından belirleniyorsa, düzey sayısı kontrol altındadır. Aksi takdirde kontrol altında değildir. Örneğin; bir araştırmada sıcaklık faktör olarak alınmış ve deneme olarak 10 °C, 20 °C ve 30 °C düzeyleri alınmış ise, düzey sayısı kontrol altındadır. Ancak; cinsiyet bir faktör olarak alınmış ise, düzeyleri kadın” ve “erkek” olup, düzey sayısı kontrol altında değildir.

iii) Deneyde birden fazla faktör olduğu durumlarda, faktör düzeylerinin kombinasyonlarına deneme adı verilir. Örneğin; deneyde iki tane faktör ( $A$  ve  $B$  gibi) ve bu faktörlerden  $A$ 'nın 3 tane ( $A_1, A_2, A_3$ ) düzeyi,  $B$ 'nin de 2 tane ( $B_1, B_2$ ) düzeyi varsa, bu takdirde denemelerin/deneme kombinasyonlarının sayısı  $3*2=6$  tane olacaktır. Bu denemeler;  $A_1B_1, A_1B_2, A_2B_1, A_2B_2, A_3B_1$  ve  $A_3B_2$  şeklinde gösterilecek olan birbirinden bağımsız gruplar olacaktır. Bu tür deneylerde temel amaç, faktörlerin ana etkileri ile beraber etkileşim etkilerini de test etmektir.

iv) Araştırma konusu olan faktör veya faktörlerin istatistiksel olarak anlamlı olup olmadığını belirlemede kullanılan istatistiksel analize **Varyans Analizi (Analysis of Variance)** adı verilir.

**Tanım 1.5** Her bir faktörün etkisi, başka faktörlerin düzeylerine bağlı olduğunda ortaya çıkan ortak faktör etkilerine **etkileşim** denir. Etkileşimin önemsiz olması faktörlerin bağımlı değişken üzerine birbirinden bağımsız olarak etki yaptığı anlamına gelmektedir. Etkileşimin önemli olması ise faktörlerin bağımlı değişken üzerine etkilerinin birbirine bağlı olduğunu gösterir.

**Tanım 1.6** Denemelerin rastgele olarak uygulandığı canlı ya da cansız nesnelere/varlıklara **deney birimi**, üzerinde bağımlı değişken için ölçümlerin yapıldığı canlı ya da cansız nesnelere/varlıklara ise **gözlem birimi** denir.

Deney birimi ile gözlem birimi birbirinden farklı kavramlardır. Uygulama probleminin özelliğine göre deney birimleri ile gözlem birimleri aynı olabileceği gibi, farklı da olabilirler. Örneğin; bir araştırmacının buzağların doğum ağırlığı üzerinde anne ineklere verilen yem türünün etkisini araştırmak istediğini düşünelim. Araştırmacı deney için üç farklı yem türü belirlemiş ve bu yem türlerini hamile ineklere rastgele uygulayacaktır. Böyle bir deney için:

**Bağımlı değişken:** Buzağların doğum ağırlığı..... Nicel, sürekli ve ölçme düzeyi oranlama

**Faktör:** Yem türü..... Nitel ve ölçme düzeyi sınıflama

**Faktör düzeyleri:** 3 tane olup  $A_1$ ,  $A_2$ ,  $A_3$  ile gösterilebilir. Faktör düzey sayısını araştırmacı belirlediğinden kontrol altındadır. Bu faktör düzeyleri aynı zamanda denemeler olup, birbirinden bağımsız gruplardır.

**Deney birimi:** İnekler, çünkü yem türlerinin rastgele olarak uygulandığı varlıklar ineklerdir.

**Gözlem birimi:** Buzağlar, çünkü üzerinde ölçüm yapılan varlıklar buzağlardır.

Görüldüğü üzere bu deneyde deney ve gözlem birimleri birbirinden farklıdır.

Eğer; aynı deneyde temel amaç yem türlerinin ineklerin ağırlık artışına etkisini araştırmak olsaydı; yem türlerinin uygulandığı ve üzerinde ölçümlerin yapıldığı varlıklar aynı olacağından, inekler hem deney hem de gözlem birimi olacaktı.

**Tanım 1.7** Benzer çevresel koşullarda aynı deneye maruz kalan deney birimlerinden elde edilen ölçümlerin genellikle benzer olması beklenir. Bununla birlikte, bu durum deney birimlerinin doğası gereği imkânsızdır. Deney birimleri arasında ortaya çıkan ve kontrol edilemeyen bu farklılıklara **deneysel hata (experimental error)** denir.

Deney birimleri arasındaki homojenlik, deneysel hatanın küçülerek deneyin hassasiyetinin artmasını sağlar.

## 1.2 DENEYİN PLANLANMASI

Bir istatistiksel deney tasarımı belirlenmiş ve hipotez şeklinde ifade edilmiş bir araştırma probleminde söz konusu olan faktörleri göz önüne alıp en az hata ile çözüme/karara ulaşmak

için yapılan işlemlerdir. Bu durum dikkate alındığında bir deneyin planlaması aşağıdaki gibi özetlenebilir.

ARAŞTIRMA		
DENEY	TASARIM	ÇÖZÜMLEME
-Problemin ortaya konması -Bağımlı değişken(ler)in belirlenmesi -Faktör(ler)in belirlenmesi -Faktör düzeylerinin belirlenmesi. Düzeyler: Nitel/Nicel ve Özel seçimli/ Rastgele seçimli - Farklı faktör düzeylerinin kombinasyonlarının belirlenmesi (varsa) -Araştırılmak istenen hipotezlerin ifade edilmesi	-Denemelerdeki gözlemlerin sayısının belirlenmesi -Deney sırasının belirlenmesi -Kullanılacak rastgeleleştirme yönteminin belirlenmesi -Matematiksel modelin belirlenmesi (Genellikle doğrusal model kullanılır. Model faktör sayısına ve faktör düzeylerinin seçimine göre farklı adlarla ifade edilebilir.)	-Veri toplama ve işleme -İstatistiksel sonuç çıkarım -Sonuçların yorumlanması

Bir deney tasarımı için belirlenen matematiksel model, değişimin (varyansın) kaynağı ile bağımlı değişken arasındaki ilişkiyi ortaya koymalıdır. Deney verisinin analizinde kullanılan teknikler modele bağlı olacaktır. Araştırmacı tarafından faktör düzeyleri özel seçilmiş ise bu durumda bir faktörün etkisinin **sabit etkili** olduğu söylenir. Böyle bir faktörü kapsayan modele de **sabit etkili model** denir. Deneyde kullanılacak faktörün düzeyleri özel olarak belirlenmemiş, bir çok düzey arasından rastgele seçilmiş ise, bu faktöre **rastgele etkili** denir ve böyle faktörler kapsayan modele de **rastgele etkili model** denir. Bazı faktörlerin özel seçimli ve bazı faktörlerinde rastgele etkili olduğu modele de **karışık etkili modeller** denir.

### 1.3 BLOKLAMA, RASTGELELEŞTİRME VE TEKRAR

Deney tasarımının temel ilkeleri olan bloklama, rastgeleleştirme ve tekrar deneysel hatanın azaltılması için kullanılan ilkelerdir. Bu kavramlardan tekrar ve bloklama deneydeki kesinliği artırmaya yardımcı olurken, rastgeleleştirme, bir deneyin objektifliğini artırır.

**Tanım 1.8** Deneyin hassaslığını artırmak için aralarında sistematik farklar bulunan deney birimleri, kendi içinde homojen ve kendi aralarında heterojen olacak şekilde blok adı verilen gruplara bölünebilir. Bu işleme **Bloklama (blocking)** denir.

Örneğin: Faktör... Yem türü ve Faktör düzeyleri (denemeler):  $A_1$ ,  $A_2$ ,  $A_3$  iken

Blok: İnek ırkı ve Blok düzeyleri:  $I_1$ ,  $I_2$ ,  $I_3$  olabilir. Aynı ırktan olan inekler aynı blokta yer alır. Böylece kendi içinde homojen fakat kendi aralarında heterojen üç blok oluşturulmuş olur. Her deneme her blokta her bir deney birimine uygulanır. Böylece İrk'in bağımlı değişken üzerine olan etkisi, deneme etkilerinden arındırılmış olur.

**Tanım 1.9** Deney birimleri arasındaki farklılıkların, ölçüm değerleri üzerindeki sistematik etkisini kontrol altına alma işlemine **rastgeleleştirme (randomization)** denir.

Deney tasarımında deney birimlerinin olabildiğince homojen olması istenir. Bununla birlikte, deney birimleri arasında yaradılıştan gelen veya çevresel koşullardan kaynaklanan farklılıklar her zaman mevcuttur. Bu farklılıklar rastgeledir. Denemeler deney birimlerine rastgele uygulanmaz ise deneme etkileri arasındaki farklar ile hata varyansının tahmin değerleri yanlış olur. Rastgeleleştirme, deney birimlerinin denemelere atanma olasılıklarının eşit olmasını sağlar.

**Tanım 1.10** Her bir denemenin uygulandığı deney birim sayısına **tekrar (replication)** denir.

Denemelerin uygulandığı deney birim sayısı her bir deneme için aynı olabileceği gibi farklı da olabilir. Eğer; deney birim sayısı (tekrar sayısı) her bir deneme için eşitse bu tasarımlara ***dengeli (balanced) tasarımlar***, eşit değilse (en az bir deneme için) ***dengeli olmayan (unbalanced) tasarımlar*** adı verilir. Tekrar sayısı birden fazla olmalıdır, aksi takdirde deneysel hata hesaplanamaz. Tekrar sayısının artması, deneysel hatanın küçülmesine dolayısı ile deneme etkilerine ait tahminlerin ve testlerin hassasiyetinin artmasına yol açar. Optimum tekrarlama sayısı, istatistik teorisi hesapları (örnekleme tekniği ve örnek hacmi belirleme) dikkate alınarak belirlenmelidir. Deney birimleri arasındaki homojenlik ne kadar fazla ise tekrarlanma sayısı da o kadar az olmalıdır. Çünkü deneysel hata, deney birimleri arasındaki farklılıkların bir ölçüsü olarak ifade edilir.

## BÖLÜM 2 TEK YÖNLÜ VARYANS ANALİZİ

Varyans analizi üç ya da daha fazla bağımsız grup arasında ortalamalar bakımından istatistiksel olarak farklılık olup olmadığını test etmek amacıyla kullanılır. Her ne kadar **Student t** testini kullanarak grupları ikişer ikişer ortalamalar yönünden karşılaştırmak mümkün olsa da bu durum, I. tip hatanın( $\alpha$ ) büyümesine yol açar. Örneğin faktör düzey sayısı 4 iken mümkün olan ikili karşılaştırma sayısı  $\binom{4}{2} = 6$  dır. Bu ikili karşılaştırmaların her biri için gerçekte doğru olan  $H_0$  hipotezinin kabul edilme olasılığı  $1 - \alpha = 0,95$  alınır, tüm altı bağımsız test için  $H_0$  hipotezlerinin kabul edilme olasılığı  $(0,95)^6 \cong 0,74$  olacak ve I. tip hata önemli derecede artacaktır. Bu istenmeyen durumu gidermenin yolu tüm grupları birlikte aynı anda karşılaştıran yöntemler ortaya koymaktır. Grup ortalamalarının karşılaştırılması, deneyin sonucunda bağımlı değişkende meydana gelen değişkenliğin (varyansın) ne kadarının faktör(ler)den ne kadarının da hatadan v.s kaynaklandığının belirlenmesi, bir başka deyişle toplam varyansın bileşenlerine ayrılması yardımı ile yapılır.

Tek yönlü varyans analizi (ANOVA), deney birimleri homojen yada mümkün olduğunca benzer olduğunda kullanılması önerilen en uygun deney tasarımıdır. Doğada deney birimleri aynı/özdeş olamayacağı için, deney birimleri arasındaki küçük farklılıklar rastgele farklılıklardır. Deney birimleri homojen kabul edildiğinde rastgeleleştirme üzerinde hiçbir kısıt yoktur. Bu sebeple ANOVA tasarımına, tamamen rastgele kısıtlayıcısız tek faktör deney tasarımı da denir.

### 2.1 MODEL DENKLEMİ, VARSAYIMLARI ve VERİ DÜZENİ

Tamamen rastgele kısıtlayıcısız tek faktör deney tasarımı (Tek Yönlü ANOVA) için matematiksel model denklemi;

$$Y_{ji} = \mu_{.} + \tau_j + \varepsilon_{ji}, \quad j = 1, 2, \dots, k; i = 1, 2, \dots, n_j \quad (2.1)$$

şeklinde ifade edilir. Burada:

$Y_{ji}$  :  $j$ .nci deneme de  $i$ .nci gözlem birimi için bağımlı değişken değeri

$\mu_{.}$  : Genel kitle ortalaması (bilinmeyen fakat sabit bir parametre)

$\tau_j$  : Faktörün  $j$ .nci deneme etkisi (bilinmeyen fakat sabit bir parametre  $\tau_j = \mu_{j.} - \mu_{.}$  ve  $\sum_{j=1}^k n_j \tau_j = 0$ , (en az bir  $n_j$  farklı ise) veya  $\sum_{j=1}^k \tau_j = 0$  ( $n_1 = n_2 = \dots = n_k = n$  ise)

$\varepsilon_{ji}$  :  $j$ .nci deneme de  $i$ .nci gözlem birimi için bağımlı değişken değerine ait rastgele hata

dır. Eşitlik (2.1) ile verilen modelde:

i) Faktör düzeyleri özel seçimli ise modele sabit ekili model

ii) Faktör düzeyleri rastgele seçimli ise modele rastgele etkili model

iii) Tekrar sayıları her bir deneme için eşitse, yani  $n_1 = n_2 = \dots = n_k = n$  ise modele dengeli model

iv) Tekrar sayıları her bir deneme için eşit değilse, yani  $\exists n_j$  diğerlerinden farklı ise modele dengeli olmayan model denir.

Eşitlik (2.1) ile verilen modelin adı ne olursa olsun parametre tahminlerinde ve hipotez testlerinde izlenecek yol aynıdır. Bu modelin varsayımları şu şekildedir.

i) Verilen model, bağımlı değişkende meydana gelen değişimi etkileyen tüm kaynakları kapsar

ii) Deney, tüm denemelerden sadece ilgilenilenleri kapsar

iii)  $\varepsilon_{ji} \sim BND(0, \sigma_\varepsilon^2)$  dir.

iv)  $Y_{ji} \sim BND(\mu_{.j} + \tau_j, \sigma_\varepsilon^2), j = 1, 2, \dots, k$  dır.

v)  $N = \sum_{j=1}^k n_j$  (veya  $N = k * n$ ) tane homojen deney birimi,  $k$  tane denemeye tamamen rastgele dağıtılır.

Eşitlik (2.1) ile verilen, tamamen rastgele kısıtlayıcısız tek faktör deney tasarımı için kitleye ait veri düzeni Tablo 2.1 de verildiği gibidir.

**Tablo 2.1 Tek Faktör varyans analizi için kitle veri düzeni**

DENEMELER					
1	2	...	j	...	k
$Y_{11}$	$Y_{21}$	....	$Y_{j1}$	....	$Y_{k1}$
$Y_{12}$	$Y_{22}$	....	$Y_{j2}$	....	$Y_{k2}$
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
$Y_{1i}$	$Y_{2i}$	....	$Y_{ji}$	....	$Y_{ki}$
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
$\mu_{1.}$	$\mu_{2.}$		$\mu_{j.}$		$\mu_{k.}$
					$\mu_{..}$

Kabul edelim ki her bir denemeden  $n_j$  birimlik örneklem çekilsin. Bu takdirde örneklem veri düzeni Tablo 2.2 deki gibi olacaktır.

**Tablo 2.2 Tek Faktör varyans analizi için örnek veri düzeni**

DENEMELER							
	1	2	...	j	...	k	
	$Y_{11}$	$Y_{21}$	....	$Y_{j1}$	....	$Y_{k1}$	
	$Y_{12}$	$Y_{22}$	....	$Y_{j2}$	....	$Y_{k2}$	
	.	.	.	.	.	.	
	.	.	.	.	.	.	
	.	.	.	.	.	.	
	$Y_{1n_1}$	$Y_{2n_2}$	....	$Y_{jn_j}$	....	$Y_{kn_k}$	
$\bar{Y}_{j.}$	$\bar{Y}_{1.}$	$\bar{Y}_{2.}$		$\bar{Y}_{j.}$		$\bar{Y}_{k.}$	$\bar{Y}_{..}$
$n_j$	$n_1$	$n_2$		$n_j$		$n_k$	$N$
$T_{j.}$	$T_{1.}$	$T_{2.}$		$T_{j.}$		$T_{k.}$	$T_{..}$

Burada:

$T_{j.}$  :  $j$ .nci denemedeki gözlemlerden elde edilen değerlerin toplamı

$$T_{j.} = \sum_{i=1}^{n_j} Y_{ji} \quad , \quad j = 1, 2, \dots, k$$

$T_{..}$  : Deneydeki tüm gözlemlerden elde edilen genel toplamı

$$T_{..} = \sum_{j=1}^k \sum_{i=1}^{n_j} Y_{ji} = \sum_{j=1}^k T_{j.}$$

$\bar{Y}_{j.}$  :  $j$ .nci denemedeki gözlemlerden elde edilen değerlerin ortalaması

$$\bar{Y}_{j.} = \frac{T_{j.}}{n_j} = \frac{\sum_{i=1}^{n_j} Y_{ji}}{n_j} \quad , \quad j = 1, 2, \dots, k$$

$\bar{Y}_{..}$  : Deneydeki tüm gözlemlerden elde edilen değerlerin ortalaması (genel örnek ortalaması)

$$\bar{Y}_{..} = \frac{T_{..}}{N} = \frac{\sum_{j=1}^k \sum_{i=1}^{n_j} Y_{ji}}{N} = \frac{\sum_{j=1}^k T_{j.}}{N}$$

dir.