



Bölüm 8

Varyans Analizi



UZAKTAN EĞİTİM UYGULAMA VE ARAŞTIRMA MERKEZİ

telefon 0(312) 202 82 00 • eposta guzem@gazi.edu.tr • adres Gazi Üniversitesi Rektörlük Binası No:6/f

guzem.gazi.edu.tr • uzaktanegitim.gazi.edu.tr • lms.gazi.edu.tr

Tanım

Varyans analizi niin yapılır?

- ✓ İkiiden fazla veri grubunun ortalamalarının karşılaştırılması t veya Z testi ile yapılabilir.
- ✓ Ancak karşılaştırılacak ortalama sayısı arttıka test sayısı da ok fazla artmaktadır.

Örneğın:

Karşılaştırılacak ortalama sayısı 5 ise 10 adet t veya Z testi yapılmalıdır (${}_5C_2=10$). Bu durumda $\alpha=0.05$ ise $(1-\alpha)^{10}=(0.95)^{10}=0.60 \rightarrow \alpha=0.40$ olacaktır. Yani test sayısı arttıka kabul edilebilirlik ihtimali düşerken hata ihtimali artmaktadır (*istenmeyen bir durum*).

- ✓ Ortalamalar üzerindeki test, varyansların incelenmesi ile yapıldığından bu yöntemeye **varyans analizi** denir.
- ✓ Örneklerden hareketle anaktle ortalamaları karşılaştırılırken hipotezler anaktle varyansları üzerine değil, anaktle ortalamaları üzerine kurulur.

Örnek:

Bir atölyede aynı parçayı üreten 3 tezgah vardır. Her tezgah farklı bir operatör tarafından çalıştırılmaktadır. Altı saatlik bir üretim devresinde her bir saatte üretilen kusurlu birim sayıları belirlenmiş ve aşağıda verilmiştir. Bir saat içinde üretilen kusurlu parçaların normal dağıldığı varsayılmaktadır. %95 güven düzeyinde ($\alpha=5\%$ hata seviyesinde) tezgahların kusurlu üretim ortalamalarının aynı olup olmadığına karar veriniz.

	Tezgah Çeşidi			
	1.Tezgah	2.Tezgah	3.Tezgah	
Gözlemler (Çıktılar)	3	2	4	Genel toplam
	4	3	5	
	4	4	6	
	5	5	6	
	7	--	7	
	--	--	8	
Örnek Hacmi	5	4	6	15
Örnek Toplamı	23	14	36	73

İkiden fazla anakütle
bulunduğundan
Varyans Analizi
yardımı ile karar verilebilir.

Varyans analizinin esası

- ✓ Gözlenen değişkenliği meydana getiren kaynaklarına ayırarak test etmektir.
- ✓ Başka bir ifadeyle, toplam değişkenliği meydana getiren alt değişkenlik kaynaklarının belirlenmesi ve bu değişkenlik kaynaklarına karşı önlemlerin alınması araştırmalarda ulaşılmak istenen asıl sonuçtur.
- ✓ Önce problemin sebebini veya sebeplerini doğru teşhis edip, sonra da uygun önlemleri almak mümkündür.

Varyans analizi bu amaca hizmet eden yöntemlerden birisidir.

Varyans analizindeki tanımlar

- **Parametre (faktör):**
 - Deney sonucuna etkileyen değişken
- **Performans karakteristiği (çıktı veya cevap değişkeni):**
 - Deney sonucunda elde edilen değer
- **Seviye (düzey veya muamele):**
 - Parametrelerin deneyde kullanılan farklı değerleri

Örnek:

Parametre	Seviye	Performans Karakteristiği
Karbon Miktarı	% 0.8 %1.4 % 2	Çeliğin sertliği
Makine çeşidi	A B C	Kusurlu ürün oranı
Sıcaklık	20 35 50	Çözünürlük miktarı

- ✓ Tek parametre üzerine yapılan varyans analizine tek yönlü varyans analizi (ANOVA)
- ✓ iki parametre üzerine yapılan varyans analizine çift yönlü varyans analiz
- ✓ ikiden fazla parametre üzerine yapılan varyans analizine de deney planlaması denilmektedir

Varyans analizindeki dağılımlar

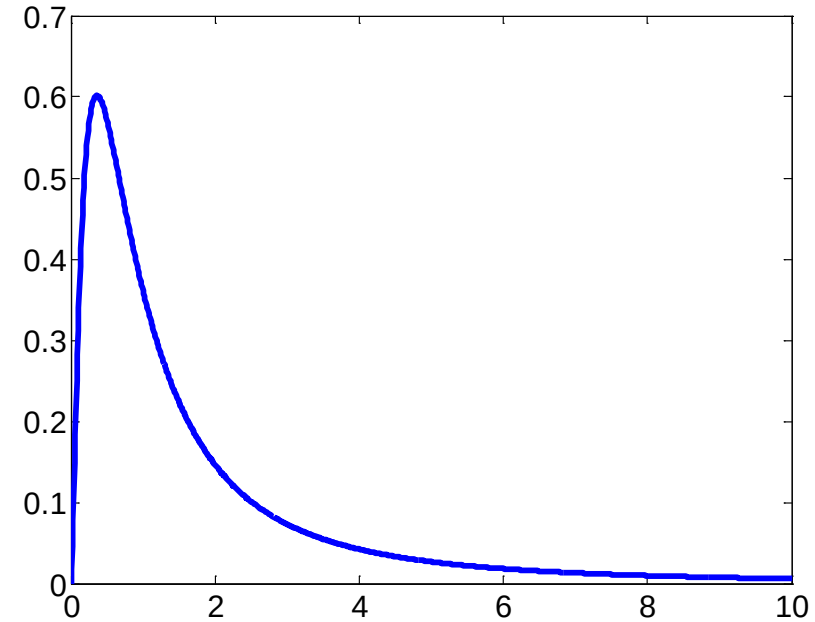
Varyanslar ile ilgili testlerin ve tahminlerin χ^2 dağılımı ile yapılmaktadır.

$$\chi_{n-1}^2 = \frac{n-1}{\sigma^2} S^2$$

Ancak iki ana kütleden varyansının birbirine eşit olup olmadığını belirlenmesinde **F dağılımı** kullanılmalıdır.

İki varyansın karşılaştırılmasında da v_1 ve v_2 serbestlik dereceli bağımsız iki χ^2 değişkeni oranının F dağılımına uyduğu bilinmektedir.

$$F_{v_1, v_2} = \frac{\chi_{v_1}^2}{\chi_{v_2}^2} = \frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2}$$



Tek yönlü varyans analizi

- ✓ Yapılan bir araştırmada elde edilen sonuçlar aşağıdaki gibi olsun

	Parametre A				
	A_1	A_2		A_k	
Gözlemler (Çıktılar)	X_{11}	X_{21}		X_{k1}	
	X_{12}	X_{22}		X_{k2}	
	X_{13}	X_{23}		X_{k3}	
					
	X_{1n}	X_{2n}		X_{kn}	Genel toplam
Örnek Hacmi	n_1	n_2		n_k	N
Örnek Toplamı	T_1	T_2		T_k	T

Bu gözlemlere dayalı olarak deneye alınan parametrenin seviye ortalamaları arasında farklılık olup olmadığı belirlemek için tek yönlü varyans analizi tekniğinden yararlanılabilir.

Tek yönlü varyans analizi



Tek yönlü varyans analizinde gözlenen toplam değişkenlik iki gruba ayrılabilir:

- ✓ Gruplar arası değişkenlik (*B alt indisi ile gösterilecek*)
→ Kontrol edilebilen değişkenlik
- ✓ Gruplar içi değişkenlik (*W alt indisi ile gösterilecek*)
→ Kontrol edilemeyen değişkenlik

Toplam değişkenlik (*T alt indisi ile gösterilecek*)

Tek yönlü varyans analizi

Gruplararası, gruplarıçi ve toplam değişkenlik değerlerinin, bunlardan hareketle grup varyanslarının ve bağlı olarak test istatistiğinin sistematik ve kolay bir şekilde hesaplanabilmesi için kullanılan formüller aşağıdaki varyans analiz tablosu (kısaca ANOVA tablosu → *analysis of variance*) üzerinde gösterilmiştir.

ANOVA Tablosu

Değişkenlik Kaynağı	Kareler Toplamı	Serbestlik Derecesi	Kareler Ortalaması	Test İstatistiği
Gruplar arası	$SS_B = \sum_{i=1}^k \frac{T_i^2}{n_i} - \frac{T^2}{N}$	$v_1=k-1$	$MS_B = S_B^2 = \frac{KT_B}{v_1}$	$F_h = \frac{S_B^2}{S_W^2}$
Gruplar içi	$SS_W = \sum_{i=1}^k \sum_{j=1}^{n_i} X_{ij}^2 - \sum_{i=1}^k \frac{T_i^2}{n_i}$	$v_2=N-k$	$MS_W = S_W^2 = \frac{KT_W}{v_2}$	-----
Toplam	$SS_T = KT_B + KT_W$	$v=N-1$	-----	-----

Tek yönlü varyans analizi

İkiden fazla veri grubunun ortalamalarının aynı anakütleden gelip gelmedikleri aşağıdaki hipotez testi ile belirlenebilir:

Hipotezler	Test istatistiği	Tablo Değeri ve Karar Kriteri
$H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$ H_1 : En az biri farklı	$F_h = \frac{S_B^2}{S_W^2}$	$\alpha = \alpha_0$ $F_{\alpha, (v1, v2)} = a_0$ $F_h > F_{\alpha, (v1, v2)} \rightarrow H_0$ reddedilir

Not: Yapılan işlem örnekleme olduğundan $\sigma_{\bar{X}}^2$ yerine $S_{\bar{X}}^2 = \sigma^2/n$ yazılabilir. Tahmin edilecek değer $\sigma^2 = n \cdot S_{\bar{X}}^2 = S_B^2$ ifadesi ile belirlenir. Buna anakütle varyansı σ^2 nin tümüyle örnek ortalamalarının dağılmasına dayandırılmış kestirimi denir. Öte yandan anakütle varyansı σ^2 nin tümüyle örnek varyanslarına dayandırılmış kestirimi ise S_W^2 dir ve aşağıdaki gibi hesaplanır.

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

$$\sigma^2 = \frac{\sum_{i=1}^k S_i^2}{k} = S_W^2$$

Eğer H_0 doğru ise $S_B^2 = S_W^2$ olacaktır. Bu durumda F_h değeri 1'e yakın çıkacak ve H_0 hipotezi reddedilemeyecektir.

Tek yönlü varyans analizi

Örnek:

Bir atölyede aynı parçayı üreten 3 tezgah vardır. Her tezgah farklı bir operatör tarafından çalıştırılmaktadır. Altı saatlik bir üretim devresinde her bir saatte üretilen kusurlu birim sayıları belirlenmiş ve aşağıda verilmiştir. Bir saat içinde üretilen kusurlu parçaların normal dağıldığı varsayılmaktadır. %95 güven düzeyinde ($\alpha=5\%$ hata seviyesinde) tezgahların kusurlu üretim ortalamalarının aynı olup olmadığına karar veriniz.

	Tezgah Çeşidi			
	1.Tezgah	2.Tezgah	3.Tezgah	
Gözlemler (Çıktılar)	3	2	4	Genel toplam
	4	3	5	
	4	4	6	
	5	5	6	
	7	--	7	
	--	--	8	
Örnek Hacmi	5	4	6	15
Örnek Toplamı	23	14	36	73

Tek yönlü varyans analizi

Örnek Çözüm:

Tezgahların kusurlu üretim ortalamalarının aynı olup olmadığı belirlenecektir. Bu durumda *karşılaştırılacak ortalama sayısı* $= 3 > 2$

parametre $\rightarrow$ *tezgah çeşidi* ve *performans karakteristiği* $\rightarrow$ *kusurlu üretim miktarı* olduğu için tek yönlü varyans analizi kullanılabilir

i. $H_0: \mu_1 = \mu_2 = \mu_3$

H_1 : en az biri farklı

ii. $\alpha = 0.05$ ve $v_1 = 3 - 1 = 2$ $v_2 = 15 - 3 = 12 \rightarrow F_{0.05, (2; 12)} = 3.88$

iii. F_h değerini hesaplayabilmek için ANOVA tablosunu oluşturalım.

$$SS_B = \frac{23^2}{5} + \frac{14^2}{4} + \frac{36^2}{6} - \frac{73^2}{15} = \frac{932}{60} = 15.5333$$

$$SS_W = 3^2 + 4^2 + 4^2 + \dots + 7^2 + 8^2 - \left(\frac{23^2}{5} + \frac{14^2}{4} + \frac{36^2}{6} \right) = \frac{1452}{60} = 2.20$$

UZA $SS_T = \frac{932}{60} + \frac{1452}{60} = \frac{2384}{60} = 39.7333$

Tek yönlü varyans analizi

Örnek Çözüm (devam):

ANOVA Tablosu

Değişkenlik Kaynağı	Kareler Toplamı	Serbestlik Derecesi	Kareler Ortalaması	Test İstatistiği
Gruplar arası	15.5333	3-1=2	7.767	3.851
Gruplar içi	24.2000	15-3=12	2.017	-----
Toplam	39.7333	15-1=14	-----	-----

iv. $F_h = 3.851 < F_{0.05, (2; 12)} = 3.88 \rightarrow H_0$ reddedilemez.

v. %5 hata (%95 güven) düzeyinde tezgahların kusurlu üretim oranlarının aynı olduğu söylenebilir.

Not: H_0 reddedilseydi farklılığın hangi tezgahtan kaynaklandığını belirleyebilmek için çoklu karşılaştırma (LSD, Duncan, Student Neuman Keuls, Tukey, vb) testleri yapılırdı.