İstatistik 2
Bölüm 9

Tek Faktör Varyans Analizi (Parametrik)

VARYANS ANALİZİ (ANOVA)

Genel Açıklamalar

- **Daha önce hipotez testleri konusunda yapılan açıklamalarda iki bağımsız populasyondan rassal örnekleme yöntemi ile elde edilen örnekler için ortalamaların farklı olup olmadıkları test edilirken varyansların eşit olduğu varsayılmaktaydı.**

Genel Açıklamalar

- **Bu bölümde söz konusu varsayımın nasıl test edildiği de açıklanacak. İkiden fazla populasyon ortalamasının farkları için hipotez testleri ve birden fazla bağımsız populasyonun varyanslarının eşitlik testleri için F dağılımı kullanılacaktır.**

Genel Açıklamalar

- **Populasyonu alt gruplara bölen fark değişkenler veya fark faktörleri bulunmaktadır. Populasyonu oluşturan alt gruplara ait bağımlı değişkenlerin ortalamalarının farklı olup olmadığı bu yöntemle analiz edilebilmektedir.**

Neden ANOVA?

- *Z* yada *t* testi kullanılarak ortalamaların farklılıkları test edilebilir.
- *Her bir Z* ve *t* testi I. Tip hata içerir
- *k* sayıda çiftten oluşan bir gözlem seti için toplam 1. Tip hata yapma olasılığı: $1-(1-\alpha)^k$
  - Eğer 5 ortalama varsa ve $\alpha = .05$ ise
    - 10 karşılaştırma yapılmalı
    - 1. Tip hata $1-(.95)^{10} = .40$
    - Bunun anlamı %40 olasılıkla boş hipotez reddedilecektir.

TEK FAKTÖR VARYANS ANALİZİ

- Tek faktör varyans analizinde amaç, k sayıda (k= iki veya daha fazla) bağımsız populasyondan elde edilen tek bir faktörü test ederek k sayıda populasyon ortalamalarının birbirlerinden farklı olup olmadıklarını belirlemektir. ***Faktör***, varyans analizinde bağımsız değişkene verilen isimdir. Bağımsız değişkenin aldığı değerlerde ***faktör düzeyleri*** olarak isimlendirilirler..

TEK FAKTÖR VARYANS ANALİZİ

- Örneğin farklı marka otomobil lastiklerinin ortalama ömürlerinin birbirinden farklı olup olmadığı, tek faktör varyans analizi yöntemi ile test edilebilir. Bu örnekte farklı marka otomobil lastikleri faktör, otomobil lastiğinin ortalama ömrü ise faktör düzeyidir.

Sabit Etki Modeli

- Sabit etki modelinde, analiz edilecek konunun homojen olduğu varsayılarak faktörler tamamen rassal olarak belirlenirler. Sabit etki modelinde tek faktör (yukarıdaki örnekte, otomobil lastiğidir) ve iki veya daha fazla sayıda faktör düzeyi (yukarıdaki örnekte lastik ortalama ömrüdür) bulunmakta olup tek faktör varyans analizi yöntemi ile çözümleme yapılmaktadır.

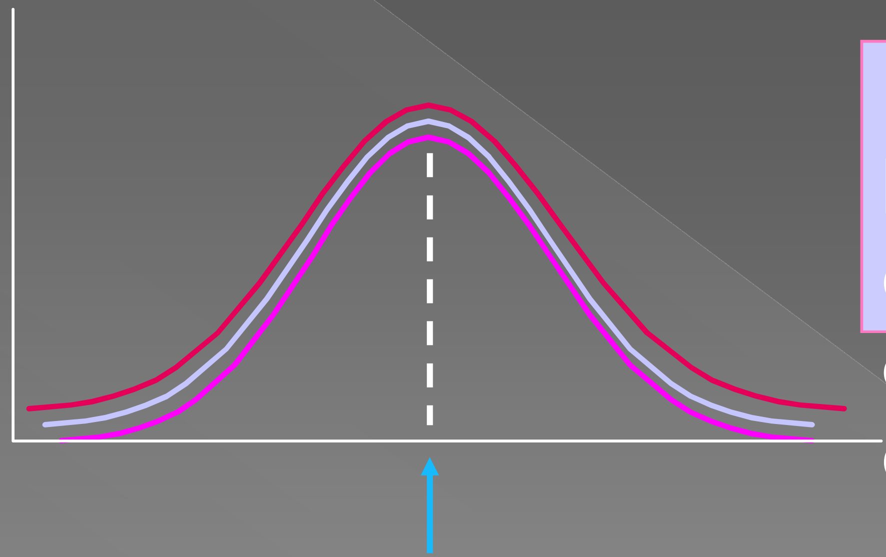
Varsayımlar

- 1. Örnek setleri rassal örnekleme yöntemi ile ve bağımsız olarak oluşturulmaktadır. Bu koşulun sağlanması zorunludur.
- 2. Örnek setlerinin seçildiği populasyonların normal dağılıma sahip olmaları gerekmektedir.
- 3. Populasyon varyansları birbirine eşittir.

Tek Faktör Varyans Analizinde Hipotezlerin Yapısı

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_c$$

$H_1 :$ **Bütün ortalamalar birbirine eşit değildir**



$$\mu_1 = \mu_2 = \mu_3$$

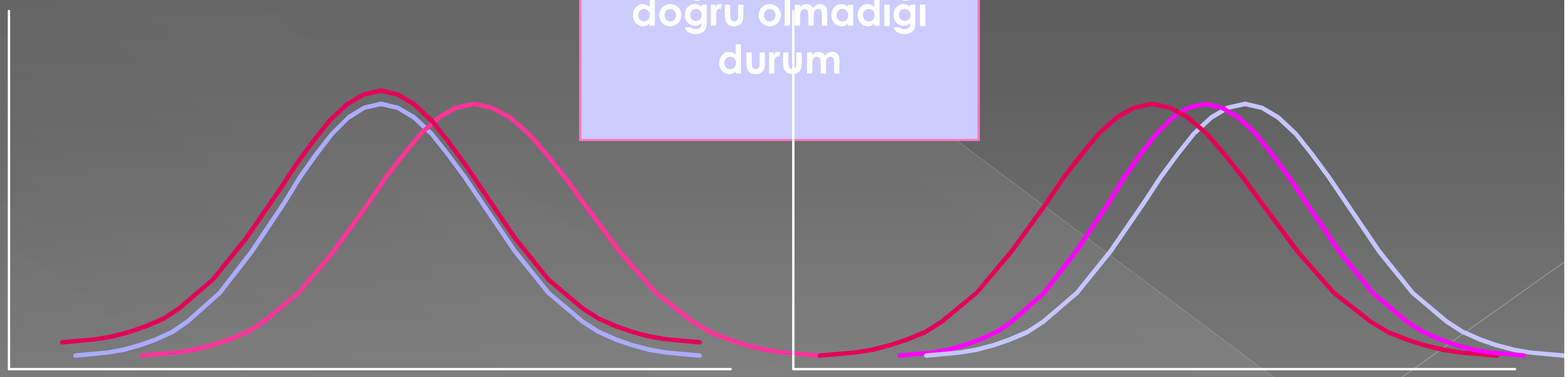
Boş hipotezin doğru olduğu durum

Tek Faktör Varyans Analizinde Hipotezlerin Yapısı

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_c$$

$H_1 :$ **Bütün ortalamalar birbirine eşit değildir**

Boş hipotezin doğru olmadığı durum



$$\mu_1 = \mu_2 \neq \mu_3 \qquad\qquad \mu_1 \neq \mu_2 \neq \mu_3$$

Genel Kareler Toplamı

- Farklı faktör düzeyleri için her bir gözlemin değişebilirliğinin toplamım genel değişebilirlik veya genel kareler toplamı olarak isimlendirilir. Genel kareler toplamı (GKT) aşağıdaki formülle hesaplanır:

Genel Kareler Toplamı

$$GKT = (x_{11} - \bar{\bar{x}})^2 + (x_{21} - \bar{\bar{x}})^2 + \ldots + (x_{n_k k} - \bar{\bar{x}}_k)^2$$

$$GKT = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (x_{ij} - \bar{\bar{x}})^2$$

Formülde,

k **= toplam grup sayısı,**

nj **= populasyon j'den elde edilen örnek gözlem sayısı,**

x*ij* **= j inci örnekteki i inci gözlem,**

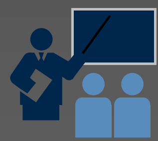
$\bar{\bar{x}}$ **= veri setinde yer alan bütün gözlemler için genel ortalamadır.**

Genel Kareler Toplamı

- Genel ortalama formülü:

$$\bar{\bar{x}} = \frac{\sum_{j=1}^{k} \sum_{i=1}^{n_j} x_{ij}}{n}$$

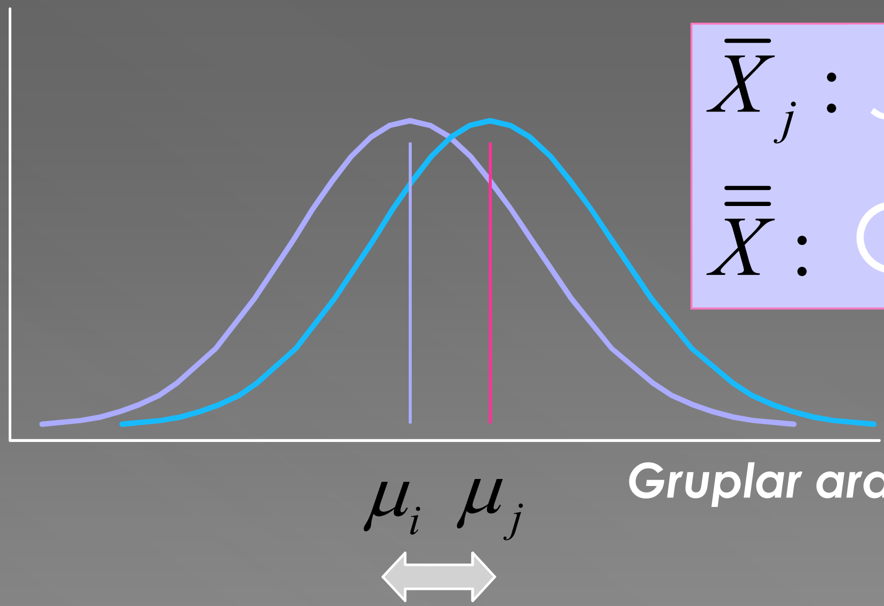
Rassal olarak düzenlenen örnek

	Faktör (eğitim yöntemleri)		
Faktör düzeyleri			
Rassal belirlenen birimler			
Bağımlı değişken (cevap)	21 hrs	17 hrs	31 hrs
	27 hrs	25 hrs	28 hrs
	29 hrs	20 hrs	22 hrs

Gruplar arası kareler toplamı

Gruplar arası kareler toplamı aşağıdaki formülle hesaplanır.

$$GAKT = \sum_{j=1}^{k} n_j (\bar{x}_j - \bar{\bar{x}})^2 \qquad GAOK = \frac{GAKT}{k-1}$$



$\bar{X}_j$: J inci grup ortalaması

$\bar{\bar{X}}$: Genel ortalama

Gruplar arası farklılıktan dolayı ortaya çıkan sapma

Gruplar arası kareler toplamı

$$GAKT = n_1(\bar{x}_1 - \bar{\bar{x}})^2 + n_2(\bar{x}_2 - \bar{\bar{x}})^2 + \ldots + n_k(\bar{x}_k - \bar{\bar{x}})^2$$

$$\bar{x}_j = \frac{\sum_{i=1}^{n_j} x_{ij}}{n_j}$$

Gruplar İçi Kareler Toplamı

$$GIKT = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2$$

$\bar{X}_j$ J inci grup ortalaması

X_{ij} : J inci grupta i inci gözlem değeri

Gruplar İçi Kareler Toplamı

$$GIKT = (x_{11} - \bar{x}_1)^2 + (x_{21} - \bar{x}_1)^2 + \ldots + (x_{n_k k} - \bar{x}_k)^2$$

$$GIOK = \frac{GIKT}{n-k}$$

$$GIOK = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2 + (n_3 - 1)S_3^2 + \bullet\bullet\bullet + (n_k - 1)S_k^2}{(n_1 - 1) + (n_2 - 1) + (n_3 - 1) + (n_k - 1)}$$

F DAĞILIMI

- F dağılımı, z ve t-dağılımları ile benzer özelliklere sahip olduğu gibi farklı özelliklerde göstermektedir. F dağılımının özellikleri aşağıdaki gibidir:
- 1. F dağılımlarından oluşan bir grup vardır. Belirli bir grubu tanımlayan iki değişken bulunmaktadır. Bunlardan birincisi payın serbestlik derecesi, ikincisi ise paydanın serbestlik derecesidir. F dağılım eğrisinin şekli, pay ve paydanın serbestlik derecelerine bağlı olarak farklılaşmaktadır.
- 2. F dağılımı sağa çarpıktır. Bu nedenle negatif değerler almaz.

F DAĞILIMI

- 1. F dağılımı sürekli bir dağılımdır.
- 2. F dağılımının aldığı değerler sıfır ile artı sonsuz aralığındadadır
- 3. Her bir F dağılım eğrisinin pay ve paydası için serbestlik derecesi farklıdır. Eğrinin tepesi dikleşdikçe serbestlik derecesi artmakta, eğri basıklaştıkça serbestlik derecesi azalmaktadır.

F DAĞILIMI

- 1. Boş hipotezin doğru olması durumunda F istatistiği 1'e yakın olacaktır. Boş hipotezin doğru olmaması durumunda ise F >1 olacaktır.
- 2. F istatistiğini hesaplamada kullanılan ve aşağıda verilmiş olan formülde payın paydadan büyük olması beklenir.

F İstatistiği

- Tek faktör varyans analizinde test istatistiği şu formülle hesaplanır:

$$F = \frac{GAOK}{GIOK} = \frac{GAKT / k-1}{GIKT / n-k}$$

- **k-1= payın serbestlik derecesi ve**
- **n-k= paydanın serbestlik derecesidir.**

F İstatistiği

Red Bölgesi $F > F_{\alpha, k-1, n-k}$

Olması durumunda boş hipotez reddedilir

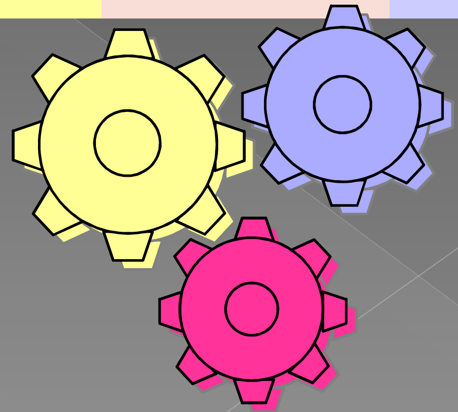
Tek Yönlü Varyans Analizi İçin ÖZET ANOVA TABLOSU

Değişim Kaynağı	Serbestlik derecesi	Kareler Toplamı	Ortalama Kare)	F istatistiği
Gruplar arası	$k-1$	GAKT	$GAOK=\dfrac{GAKT}{k-1}$	$\dfrac{GAKT/k-1}{GIKT/n-k}$
Gruplar içi	$n-k$	GIKT	$GIOK=\dfrac{GIKT}{n-k}$	
Toplam	$n-1$	GKT= GAKT + GIKT		

Örnek

Üretim sürecinde 3 adet makine kullanılmakta. Üretim müdürü olarak makinelerin ortalama dolum süreleri arasında fark olup olmadığını öğrenmek istiyorsunuz. Her bir makineye 5 deneyimli eleman atadınız. 0.05 anlamlılık düzeyinde ortalama dolum sürelerinin farklı olup olmadıklarını test ediniz

M1	M2	M3
25.40	23.40	20.00
26.31	21.80	22.20
24.10	23.50	19.75
23.74	22.75	20.60
25.10	21.60	20.40



Örnek

M1	M2	M3
25.40	23.40	20.00
26.31	21.80	22.20
24.10	23.50	19.75
23.74	22.75	20.60
25.10	21.60	20.40

$$\bar{X}_1 = 24.93 \qquad \bar{X}_2 = 22.61$$

$$\bar{X}_3 = 20.59 \qquad \bar{\bar{X}} = 22.71$$

Örnek

M1	M2	M3
25.40	23.40	20.00
26.31	21.80	22.20
24.10	23.50	19.75
23.74	22.75	20.60
25.10	21.60	20.40

$$\bar{X}_1 = 24.93 \qquad n_j = 5$$

$$\bar{X}_2 = 22.61 \qquad c = 3$$

$$\bar{X}_3 = 20.59 \qquad n = 15$$

$$\bar{\bar{X}} = 22.71$$

GAKT= $5\left[(24.93-22.71)^2+(22.61-22.71)^2+(20.59-22.71)^2\right]$

$= 47.164$

GIKT $4.2592 + 3.112 \; + 3.682 \; = \; 11.0532$

GAOK = GAKT/k-1 $47.16/2 \; = \; 23.5820$

GIOK , = GIKT/n-k $= \; 11.0532/12 \; = \; .9211$

Tek Yönlü Varyans Analizi İçin ÖZET ANOVA TABLOSU

Değişim Kaynağı	Serbestlik derecesi	Kareler Toplamı	Ortalama Kare)	F istatistiği
Gruplar arası	$k-1$ $3-2=2$	GAKT 47.1640	$GAOK = \frac{GAKT}{k-1}$ 23.5820	$\frac{GAKT \;/\; k-1}{GIKT \;/\; n-k}$ 25.60
Gruplar içi	$n-k$ $15-3$	GIKT 11.0532	$GIOK = \frac{GIKT}{n-k}$ 0,9211	
Toplam	$n-1$ $15-1$	GKT= GAKT + GIKT 58.2172		

Tek Faktör Varyans Analizi: Örneğin Çözümü

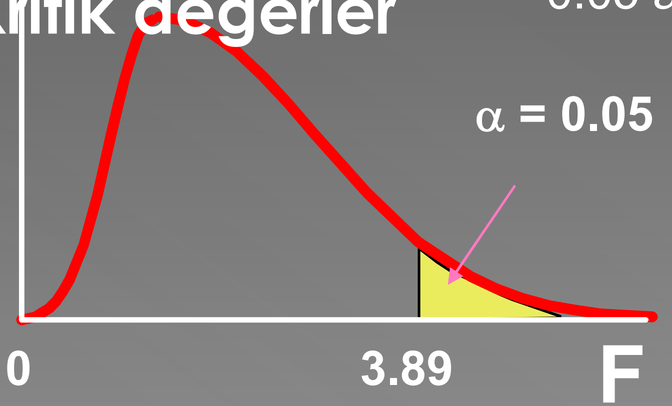
H_0: $\mu_1 = \mu_2 = \mu_3$

H_1: eşit değil

$\alpha = .05$

$SD_1 = 2$ $SD_2 = 12$

Kritik değerler



Test istatistiği:

$$\frac{GAKT/k-1}{GIKT/n-k} = \frac{23.5820}{.9211} = 25.6$$

Karar:

0.05 anlamlılık düzeyinde boş hipotez red

sonuç:

Ortalamalardan en az birinin diğerlerinden farklı olduğuna ilişkin elimizde kanıt var