VARYANS ANALİZİ

İki grubun bir özelliğini karşılaştırmak için izleyebileceğimiz yöntem şu şekilde olabilir. İki grup için 10 ar tane eleman seçip bu elemanların meydana getirdiği deney sonucu ele geçen verilerle bir istatistik çıkartabiliriz.

Bir çift ana kütlenin ortalamasının sıfır olduğunu savunan H_0 hipotezini T dağılımına uydurarak hesaplayabilmiştik.

Bu kısımda anlatacaklarımız için birer örnek üzerinden gidelim. A ve B arabalarının yakıt tüketimini incelemek istersek 10 ar tane örneklem ele alarak μ ve S değerlerini hesaplayıp (n_1+n_2-2) serbestlik dereceli T dağılımı ile H_0 hipotezinin doğruluğunu test edebiliriz. C arabasının da bu hesaplamaya dahil edildiğini düşünelim. Bu 20 sürücüyü $A \to 7$, $B \to 7$, $C \to 6$ olacak şekilde alalım.

Ortalamalar arasında karşılaştırma yapacağımız için öncelikle grubların ortalamalarını bulalım.

$$\mu_A = \frac{\Sigma X_a}{7} \qquad \mu_B = \frac{\Sigma X_B}{7} \qquad \mu_C = \frac{\Sigma X_C}{6}$$

Doğal olarak ortalamalar birbirinden farklı çıkacaktır. Hipotez seçilirken bu tür farklılıklar şans eseri ortaya çıkabilir fikrini savunuruz. Bir başka deyişle Gruplar arasında fark yoktur hipotezini H_0 olarak seçeriz. Eğer şans eseri ortaya çıkamayacağı ile ilgili sonuca varırsak H_0'ın doğruluğundan şüphe duyarız.

Ana kütlelerin ortalamalarından grup aynıdır yada farklıdır demek imkansızdır. Bu bilginin yanısıra ortalamalardan sapmalar da önemlidir. Eğer varyans küçükse ve ortalamalar da farklıysa ~~ve~~ bunun üzerine $(\mu+\sigma, \mu-\sigma)$ aralıkları üç grup için çakışmıyorsa gruplar farklıdır diyebiliriz. Aksi takdirde aynı çıkma sonucuna pek şaşırmayız.

Bu tür analizlere varyans analizi denir.

TEK YÖNLÜ VARYANS ANALİZİ

Varsayalım her birinin varyansı eşit olan K tane anakütlenin ortalamasını karşılaştırmak istiyoruz. Bu anakütlelerden $n_1, n_2 -- n_k$ tane gözlemli bağımsız rassal örnekler alınmış olsun. Örneklem ~~değerlerin~~ x ile gösterelim.

$X_{ij} \longrightarrow$ i. anakütleden j. gözlemin değeri

Tablo 1. K anakütleden alınmış örneklemler.

Anakütle		
1	2	K
X_{11}	X_{21}	X_{K1}
X_{12}	X_{22}	X_{K2}
X_{1n_1}	[illegible]	X_{Kn_K}

Şimdi Gruplararası sapmaları bulalım. Temel hesap, tekil grup ortalamalarının genel ortalamadan sapmalarına dayanır.

$$(\bar{X}_1-\bar{X})^2, \quad (\bar{X}_2-\bar{X}), \ldots, (\bar{X}_k-\bar{X})$$

$$G_AKT = \sum_{i=1}^{K} n_i (\bar{X}_i - \bar{X})^2$$

Bir kare toplamları daha hesaplanır. Bu da Genel kareler toplamıdır.

$$G_nKT = \sum_{i=1}^{K} \sum_{j=1}^{n_i} (X_{ij}-\bar{X})^2$$

yada

$$G_nKT = G_iKT + G_AKT$$

Anakütle ortalamalarını hesaplamamız, K tane anakütlenin ortak varyansı olduğu varsayımına dayanır. Ho hipotezi doğruysa herbir karetoplamı ortak varyansı tahmin etmeye yardımcı olur. Bu tahminleri elde etmek için uygun serbestlik dereceleri seçilmelidir.

İlk olarak anakütle varyansının sapmasız tahmin edicisini elde etmek için GiKT nın (n-K) ya bölünmesi gerekir. Buna Gruplçi Kareler ortalaması denir.

$$G_iKO = \frac{G_iKT}{n-K}$$

Ana kütle ortalamaları eşitse, başka bir sapmasız tahmin ediciyi bulmak için G_AKT'nin (K-1)'e bölümesi gerekir. Böylece Gruplararası Kareler ortalamasını buluruz.

$$G_AKO = \frac{G_AKT}{K-1}$$

Anakütle ortalamaları eşit değilse, G_AKO, ortak anakütle varyansının sapmasız bir tahmin edicisi olmaktan çıkar. Şu tahminlerin birbirine yakın olmalarını beklemek akla uygundur. Bu iki tahmin arasındaki açıklık ne kadar büyükse, Ho hipotezinin doğruluğu üzerine kuşkular artar.

Ho hipotezinin testi ortalama kareler arasındaki orana dayanır.

$$F = \frac{G_AKO}{G_iKO}$$

Bu oran 1'e çok yakınsa anakütle ortalamaları aynıdır diyen Ho hipotezi doğrudur diyebiliriz.

Ho hipotezi doğruysa, anakütle dağılımı normal dağılımdır varsayımı altında ortalama kareler oranına karşılık gelen rassal değişkenin, pay serbestlik derecesi $(K-1)$, payda serbestlik derecesi $(n-K)$ olan F dağılımına uyar.

$$P\left(F_{K-1,n-K} > F_{\alpha,K-1,n-K}\right) = \alpha$$

$$\frac{G_A KO}{G_i KO} > F_{\alpha,k-1,n-k}$$ ise Ho red edilir.

Tablo 2. Tek yönlü varyans analizi tablosu

Değişim Kaynağı	Kareler Toplamı	Serbestlik dereceleri	Kareler ortalaması	F oranı
Gruplararası	$G_A KT$	$K-1$	$G_A KO = \frac{G_A KT}{K-1}$	$\frac{G_A KO}{G_i KO}$
Grupiçi	$G_i KT$	$n-K$	$G_i KO = \frac{G_i KT}{n-K}$	
Toplam	$G_n KT$	$n-1$		

İKİ YÖNLÜ VARYANS ANALİZİ

Tek yönlü varyans analizinde tekbir bağımsız değişken ve bir bağımlı değişken vardı. İki yönlüde ise iki bağımsız değişken ve bir bağımlı değişken vardır.

Bağımsız örneklemler için iki yönlü varyans analizi ile bağımsız değişkenlerin bağımlı değişken üzerinde ortak etkileri belirlenir. Aynı zamanda ayrı ayrı her iki değişkene ilişkin grupların bağımlı değişkene göre ortalamalarının karşılaştırılarak farkın belirli bir güven düzeyinde (%95, %99) anlamlı olup olmadığı araştırılır.

İki yönlü varyans analizindeki temel amaç, bağımsız değişkenlerin bağımlı değişken üzerindeki ortak etkisini ölçmektir. Örneğin, bir işletmede çalışan personelin performansını ölçelim. Bağımsız değişkenlerimiz de cinsiyet ve çalıştığı bölüm olsun. Buradaki temel sorumuz: Cinsiyetin bir bölüme olan etkisi diğer bölüme olan etkisinden farklı mıdır?

Bu analizde asıl ölçmek istediğimiz bağımsız değişkenin yanında bir değişken daha bulunuyor. Buna bölüntü değişkeni denir. Mesela, 3 tip aracın yakıt tüketimlerini ölçmek için 6 tane sürücü tarafından bu araçlar kullanılıyor. Burada araçların yakıt tüketim ortalamaları aynı mıdır sorusuna cevap arıyoruz. Bunun yanısıra sürücü yaş gruplarını da incelecek olursak, yani her yaş grubundan sürücü için yakıt tüketimleri ortalaması aynıdır dersek, o zaman yaş grupları bölüntü değişkeni olur. H ile gösterelim.

K grup ve H bölüntü olsun.

$X_{ij} \longrightarrow$ i. grup j. bölüntü değeri.

Tablo 3. K gruplu, H bölüntülü örneklem

		Grup			
		1	2	— — —	K
Bölüntü	1	X_{11}	X_{21}	— — —	X_{K1}
	2	X_{12}	X_{22}	— — —	X_{K2}
	⋮	⋮			⋮
	H	X_{1H}	X_{2H}	— — —	X_{KH}

i. grup ortalaması $\overline{X}_{iB}$

$$\overline{X}_{iH} = \frac{\sum_{j=1}^{H} X_{ij}}{H} \qquad i=1,2,\dots K$$

j. bölüntü ortalaması $\overline{X}_{Kj}$

$$\overline{X}_{Kj} = \frac{\sum_{i=1}^{K} X_{ij}}{K} \qquad j=1,2,\dots H$$

Tüm örneklemlerin genel ortalaması $\overline{X}$

$$\overline{X} = \frac{\sum_{i=1}^{K}\sum_{j=1}^{H} X_{ij}}{K.H} = \frac{\sum_{i=1}^{K} \overline{X}_{iH}}{K} = \frac{\sum_{j=1}^{H} \overline{X}_{Kj}}{H}$$

X_{ij} değerine aşağıdaki 4 bileşenin toplamı olarak bakılabilir.

1) Bir genel ortalama μ
2) i. grubun genel ortalamadan farkını ölçen katsayı G_i
3) j. bölüntünün " " " " " B_j
4) Deneysel hatayı temsil eden ε_{ij}

$$X_{ij} = \mu + G_i + B_j + \varepsilon_{ij}$$

ε_{ij} nin çoklu regresyon modeline uyduğu kabul edilir. Bu yüzden bağımsızlık ve eşit varyanslı varsayımlarını yaparız.

$(X_{ij} - \overline{X})$ kadar ortalama farklılık ... G_i kadarı asıl bağımsız değişkenden, B_j kadarı ikincil bağımsız değişkenden kaynaklanmıştır. geri kalanı da diğer etmenlerden

Şimdi son bulduğumuz eşitliğin her iki yanının karelerini alıp n örneklem gözlemi için toplarsak aşağıdaki gibi olur.

$$\sum_{i=1}^{K}\sum_{j=1}^{H}(x_{ij}-\bar{x})^2 = H.\sum_{i=1}^{K}(\bar{x}_{iH}-\bar{x})^2 + K.\sum_{j=1}^{H}(\bar{x}_{Kj}-\bar{x})^2 + \sum_{i=1}^{K}\sum_{j=1}^{H}(x_{ij}-\bar{x}_{iH}-\bar{x}_{Kj}+\bar{x})^2$$

Burada, toplam örneklem değişkenliği sırayla gruplar arası farkların, bölüntüler arası farkların ve hata değişkenlerinin toplamı olarak okunur. İşte bu tür, gruplar ve bölüntülere ayrılmış şekilde yapılan analize iki yönlü varyans analizi denir.

İki yönlü varyans analizinde Ho hipotezi, Grupların anakütle ortalamaları eşittir der. ve Gruplararası kareler ortalamasının hata kareler ortalamasına bölünmesiyle test edilir.

Gruplararası Kareler ortalaması $G_A KO = \dfrac{G_A KT}{K-1}$

Bölüntülerarası " " $B_A KO = \dfrac{B_A KT}{H-1}$

Hata " " $HKO = \dfrac{HKT}{(K-1)(H-1)}$

Hataların normal dağıldığı varsayımı ile α anlamlılık düzeyinde $(K-1)$ ve $(K-1)(H-1)$ serbestlik dereceli F dağılımına uyar.

$$\frac{G_A KO}{HKO} > F_{\alpha,(K-1),(K-1)(H-1)}$$

ise K grubun ana kütle ortalamaları aynıdır diyen Ho red edilir.

$$\frac{B_A KO}{HKO} > F_{\alpha,\ [illegible]}$$

ise H bölüntünün ana kütle ortalamaları aynıdır diyen Ho red edilir.

Tablo 4. iki yönlü varyans analizinin genel görünümü

Değişim kaynağı	Kareler Toplamı	Serbestlik dereceleri	Kareler ortalaması	F oranları
Gruplararası	$G_A KT$	$K-1$	$G_A KO = \dfrac{G_A KT}{K-1}$	$\dfrac{G_A KO}{HKO}$
Bölüntülerarası	$B_A KT$	$H-1$	$B_A KO = \dfrac{B_A KT}{H-1}$	$\dfrac{B_A KO}{HKO}$
Hata	[illegible]	[illegible]	$HKO = \dfrac{HKT}{[illegible]}$	
[illegible]	$G_n KT$	$n-1$		

ETKİLEŞİMLİ VARYANS ANALİZİ

İki yönlü varyans analizindekine ek olarak bu kez bir gruptaki gözlem sayısı birden fazla olması durumudur. Aynı deneyin L tane gözlem yapıldığını düşünelim.

Tablo 5. K gruplu, H bölüntülü, L gözlemli örneklem

Gruplar

[redacted]

Modelde varsayılan modelimiz

$$X_{ijl} = \mu + G_i + B_j + I_{ij} + \varepsilon_{ijl}$$

I_{ij} = ij gözündeki l. gözlem.

Kareler toplamı		Serbestlik derecesi
Genel	$G_A KT = \sum_i \sum_j \sum_l (X_{ijl} - \bar{X})^2$	$KHL-1$
Gruplararası	$G_A KT = HL \sum_{i=1}^{K} (\bar{X}_{iHL} - \bar{X})^2$	$K-1$
Bölüntülerarası	$B_A KT = KL \sum_{j=1}^{H} (\bar{X}_{KjL} - \bar{X})^2$	$H-1$
Etkileşim	$EKT = L \sum_{i=1}^{K} \sum_{j=1}^{H} (\bar{X}_{ijL} - \bar{X}_{iHL} - \bar{X}_{KjL} + \bar{X})^2$	$(K-1)(H-1)$
Hata	$HKT = \sum_i \sum_j \sum_l (X_{ijl} - \bar{X}_{ijL})^2$	$KH(L-1)$

Bu durumda Genel kareler toplamı şu şekilde yazılır

$$G_n KT = G_A KT + B_A KT + EKT + HKT$$

Test istatistiği $G_A KO$, $B_A KO$, EKO nun HKO lara bölünerek ve kendi pay ve payda serbestlik dereceleri kullanılarak F dağılımına uyar.

$$\frac{G_A KO}{HKO}, \frac{B_A KO}{HKO}, \frac{EKO}{HKO} > F_{\alpha, sd1, sd2} \text{ ise Ho red edilir.}$$

Tablo 5. Her gözde L gözlemin bulunduğu iki yönlü varyans analizi

Değişim Kaynağı	Kareler Toplamı	Serbestlik derecesi	Kareler ortalaması	F oranları
Gruplararası	$G_A KT$	$K-1$	$G_A KO = \frac{G_A KT}{K-1}$	$\frac{G_A KO}{HKO}$
Bölüntülerarası	$B_A KT$	$H-1$	$B_A KO = \frac{B_A KT}{H-1}$	$\frac{B_A KO}{HKO}$
Etkileşim	EKT	$(K-1)(H-1)$	$EKO = \frac{EKT}{(K-1)(H-1)}$	$\frac{EKO}{HKO}$
Hata	HKT	$KH(L-1)$	$HKO = \frac{HKT}{KH(L-1)}$	—
Toplamlar	$G_n KT$	$KHL-1$		—

Ho hipotezi	test istatistiği	F değeri
Gruplararası fark yoktur.	$\frac{G_A KO}{HKO}$	$F_{\alpha, K-1, KH(L-1)}$
Bölüntülerarası fark yoktur	$\frac{B_A KO}{HKO}$	$F_{\alpha, H-1, KH(L-1)}$
Etkileşim yoktur.	$\frac{EKO}{HKO}$	$F_{\alpha, (K-1)(H-1), KH(L-1)}$

Test istatistiği F değerinden büyükse Ho red edilir.

Varyans analizi örnek:

Bilgisayara giriş dersini alan 3 farklı grub öğrencilerin aldığı notlar aşağıda verilmiştir.

A	B	C
7,0	10,0	3,6
5,4	4,1	10,0
4,5	5,4	10,0
4,6	10,0	10,0
6,4	5,1	9,5
6,0	7,5	9,7
7,5	2,0	6,0
2,5	9,8	4,5
5,9	8,8	3,8
10,0	7,8	2,6

Bu grub öğrenciler aynı bölümde okumaktadır. Hepsi 1. sınıf öğrencilerdir.

Bu üç grub öğrenci arasında fark var mıdır?

$\alpha = 0,05$ güven aralığı alınız.

Ho: Gruplar arasında fark yoktur.
Ha: Gruplar arasında fark vardır.

- Kareler toplamını bulalım.

Genel kareler toplamı: $G_n KT : \sum x^2 - \frac{(\sum x)^2}{\sum n} = 1535,1 - \frac{200^2}{30} = 201,77$

Gruplararası kareler " : $G_a KT : \sum_{j=1}^{n} \left(\frac{(\sum x_j)^2}{n_j} \right) - \frac{(\sum x)^2}{\sum n} = \frac{59,8^2}{10} + \frac{70,5^2}{10} + \frac{69,7^2}{10} - \frac{200^2}{30} = 7,105$

Grupiçi kareler " : $G_i KT : G_n KT - G_a KT = 201,77 - 7,105 = 194,665$

- Serbestlik dereceleri toplamı:

Genel serbestlik derecesi : $G_n SD : n - 1 = 30 - 1 = 29$
Gruplararası " " : $G_a SD$: Grupsayısı $- 1 = 3 - 1 = 2$
Grupiçi " " : $G_i SD$: $n -$ Grupsayısı $= 30 - 3 = 27$

- Kareler ortalamasını bulalım:

Gruplararası kare ortalaması: $G_a KO = \frac{G_a KT}{G_a SD} = \frac{7,105}{2} = 3,55$

Grupiçi " " : $G_i KO = \frac{G_i KT}{G_i SD} = \frac{194,665}{27} = 7,21$

~~Test istatistiği~~ olarak

$\frac{G_a KO}{G_i KO}$ $G_a SD$, $G_i SD$ serbestlik dereceli F dağılır.

$F = \frac{3,55}{7,21} = 0,49$

$F_{0,05,\,2,27} = 2,35$

$F_{Tablo} > F$ olduğundan

Ho kabul edilir.

Gruplararası fark yoktur.