VARYANS ANALİZİ

İki örnek ortalaması arasındaki farkın önem kontrolü, örnek büyüklüğüne göre z veya t testlerinden biriyle yapılır. Bu testlerle, ikiden fazla örnek ortalamasını birlikte test etmek ve aralarındaki farkın önem kontrolünü yapmak mümkün değildir. İki veya daha fazla örnek ortalaması arasındaki farkın anlamlı olup olmadığını test ederken varyans analizi kullanılır.

TEK YÖNLÜ VARYANS ANALİZİ

$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \qquad i = 1,2,\dots,k; \quad j = 1,2,\dots,n_i$

y_{ij} : $i.$ inci denemede $j.$ inci gözlemin aldığı değer

μ : genel ortalama

τ_i : $i.$ inci deneme etkisi

ε_{ij} : $i.$ inci denemedeki $j.$ inci gözlemin yanıt değişkenine ilişkin hata

$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$

Gruplar (Denemeler)	Gözlemler						Toplam	Gözlem Sayısı	Ortalama
1	y_{11}	y_{12}	…	y_{1j}	…	y_{1n_1}	$T_{1.}$	n_1	$\bar{y}_{1.}$
2	y_{21}	y_{22}	…	y_{2j}	…	y_{2n_2}	$T_{2.}$	n_2	$\bar{y}_{2.}$
.	.	.	…	.	…	.	.	.	.
.	.	.	…	.	…	.	.	.	.
.	.	.	…	.	…	.	.	.	.
k	y_{k1}	y_{k2}	…	y_{kj}	…	y_{kn_k}	$T_{k.}$	n_k	$\bar{y}_{k.}$
	Genel Toplam						$T_{..}$	n	$\bar{y}_{..}$

$T_{i.}$: $i.$ inci denemedeki gözlemlerden elde edilen değerlerin toplamı

$$T_{i.} = \sum_{j=1}^{n_i} y_{ij}$$

$\bar{y}_{i.}$: $i.$ inci denemedeki gözlemlerden elde edilen değerlerin ortalaması

$$\bar{y}_{i.} = \frac{T_{i.}}{n_i}$$

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

$$T_{..} = \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}$$

$\bar{y}_{..}$: Tüm gözlemlerden elde edilen değerlerin ortalaması

$$\bar{y}_{..} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}}{\sum_{i=1}^{k} n_i} = \frac{T_{..}}{n}, n = \sum_{i=1}^{k} n_i$$

Yığın için denemelerin ortalamalarının eşitliğine ilişkin hipotez aşağıdaki gibidir.

$H_0: \mu_1 = \mu_2 = \cdots = \mu_k = \mu_{..}$

H_1: En az bir ortalama farklıdır.

Bu hipoteze denk olarak;

$H_0: \tau_1 = \tau_2 = \cdots = \tau_k = 0$

$H_1: Tüm\ \ \tau_i\ lerden\ en\ az\ biri\ sıfır\ değildir. (\ \tau_i \neq 0\ en\ az\ bir\ i\ için, i = 1,2,\ldots,k)$

hipotezi yazılabilir. Bu hipotezin testi için tek yönlü varyans analizi kullanılabilir. Varyans analizi demek, toplam değişebilirliğin parçalanması demektir. (toplam değişimin parçalanarak incelenmesi)

$$\underbrace{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{..})^2}_{\text{Genel Kareler Toplamı}} = \underbrace{\sum_{i=1}^{k}\sum_{j=1}^{n_i} n_i(\bar{y}_{i.} - \bar{y}_{..})^2}_{\text{GruplarArası Kareler Toplamı}} + \underbrace{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i.})^2}_{\text{Grup İçi Kareler Toplamı}}$$

$GnKT = GAKT + GİKT$

$$GnKT = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{..})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{T_{..}^2}{n}$$

$$GAKT = \sum_{i=1}^{k}\sum_{j=1}^{n_i} n_i(\bar{y}_{i.} - \bar{y}_{..})^2 = \sum_{i=1}^{k}\frac{(T_{i.})^2}{n_i} - \frac{T_{..}^2}{n}$$

$$GİKT = \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}^2 - \sum_{i=1}^{k}\frac{(T_{i.})^2}{n_i}$$

$GİKT = GnKT - GAKT$

$\frac{T_{..}^2}{n}$ düzeltme terimi

Tek Yönlü Varyans Analizinde Hipotez Testi

1) $H_0: \mu_1 = \mu_2 = \cdots = \mu_k$
H_1: En az bir ortalama farklıdır.

ya da;

$H_0: \tau_1 = \tau_2 = \cdots = \tau_k = 0$
$H_1: En\ az\ bir\ \tau_i \neq 0$

$$n_1 \neq n_2 \neq \cdots \neq n_k, \sum_{i=1}^{k} n_i = n$$

2) Varyans analizi tablosu

Değişim Kaynakları (DK)	***Serbestlik Derecesi (sd)***	***Kareler Toplamı (KT)***	***Kareler Ortalaması (KO)***	***Test***
Deneme (Gruplar arası)	*k-1*	$GAKT$	$GAKO = \frac{GAKT}{k-1}$	$F_H = \frac{GAKO}{GİKO}$
Hata…………. (Grup içi)	*n-k*	$GİKT$	$GİKO = \frac{GİKT}{n-k}$	
Genel	*n-1*	$GnKT$		

3) $F_H > F_T(\alpha, sd_1 = k-1, sd_2 = n-k)$ *ise* H_0 *hipotezi red edilir.*

En az bir μ_i *diğerlerinden farklıdır ya da en az bir* $\tau_i \neq 0$’dır.

Örnek: Dört farklı İngilizce kursunun öğrencilerinin YDS sınavındaki notlarına etkisinin olup olmadığı konusunda bir araştırma yapılmak isteniyor. Her bir kurstan 6'şar öğrenci rasgele seçiliyor. Puanlar aşağıda verildiği gibidir. Varyans analizi tablosunu hazırlayarak İngilizce kurslarının nota olan etkisinin önemliliği $\alpha = 0.05$ anlam düzeyinde test ediniz.

	$\boldsymbol{Toplam}$	**Gözlem sayısı**	$\boldsymbol{Ortalama}$	**KT**
K1: 70 65 75 72 74 68	$T_{1.}$ =424	6	$\bar{y}_{1.} = 70.67$	30034
K2: 50 52 70 65 74 55	$T_{2.}$ =366	6	$\bar{y}_{2.} = 61$	22830
K3: 80 82 77 88 83 85	$T_{3.}$ =495	6	$\bar{y}_{3.} = 82.5$	40911
K4: 90 92 82 86 88 72	$T_{4.}$ =510	6	$\bar{y}_{4.} = 85$	43612
	$T_{..} = 1795$	$n = 24$		137387

Model;
$Y_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad , i = 1,2,3,4\,; \quad j = 1,2,3,4,5,6$

1. $H_0: \tau_1 = \tau_2 = \tau_3 = \tau\alpha_4 = 0$
   $H_1: En\ az\ bir\ \tau_i \neq 0$

ya da;

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$
H_1: En az bir ortalama farklıdır.

2. $GnKT = \sum_{i=1}^{k} \sum_{j=1}^{n_i} y_{ij}^2 - \frac{T_{..}^2}{n} = 137387 - \frac{1795^2}{24} = 3135.96$

$$GAKT = \sum_{i=1}^{k} \frac{T_{i.}^2}{n_i} - \frac{T_{..}^2}{n} = \frac{424^2}{6} + \frac{366^2}{6} + \frac{495^2}{6} + \frac{510^2}{6} - \frac{1795^2}{24}$$

$$= \frac{818857}{6} - \frac{1795^2}{24} = 2225.125$$

$G\dot{I}KT = GnKT - GAKT = 3135.96 - 2225.125 = 910.835$

Varyans Analizi Tablosu

Değişim Kaynakları (DK)	***Serbestlik Derecesi (sd)***	***Kareler Toplamı (KT)***	***Kareler Ortalaması (KO)***	***Test***
Deneme	$k-1=3$	2225.125	*741.71*	$F_H=\frac{741.71}{45.54}=16.29$
Hata	$n-k=24-4=20$	910.835	*45.54*	
Genel	$n-1=24-1=23$	3135.96		

3. $F_H=16.29>F_T(0.05,3,20)=8.66$ olduğundan H_0 hipotezi red edilir.

Dört farklı İngilizce kursunun öğrencilerinin YDS sınavındaki notlarına olan etkisi önemlidir. (İngilizce kurslarından en az birinin etkisi farklıdır.)

En az biri farklı çıkarsa o zaman bu farkın hangisinden kaynaklandığına bakılır. Bunun için iki yol vardır.

Gerçek Önemli Fark Testi (GÖF)

$H_0: \mu_i=\mu_j \qquad (i\neq j)$
$H_1: \mu_i\neq\mu_j$

$D=\left|\bar{y}_{i.}-\bar{y}_{j.}\right|>GÖF=S_{\bar{Y}}Q_T$ ise H_0 red edilir. $(Q_{T(\alpha,k,n-k)})$

$\bar{m}=\frac{n-\sum_{i=1}^{k}\frac{n_i^2}{n}}{k-1}$, $S_{\bar{Y}}=\sqrt{\frac{GİKO}{\bar{m}}}$

En Küçük Önemli Fark Testi (EKÖF)

$H_0: \mu_i=\mu_j$
$H_1: \mu_i\neq\mu_j$

$D=\left|\bar{y}_{i.}-\bar{y}_{j.}\right|>EKÖF=t_T S_{\bar{Y}}\sqrt{2}$ ise H_0 red edilir.

($t_{T(\alpha/2,sd=n_i+n_j-2)}$,n_i *i*. grubun örnek hacmi, n_j *j*. grubun örnek hacmi)

$\bar{m}=\frac{n-\sum_{i=1}^{k}\frac{n_i^2}{n}}{k-1}$, $S_{\bar{Y}}=\sqrt{\frac{GİKO}{\bar{m}}}$

Çözülen örneği her iki yolla çözelim.

Örnekte H_0 red çıktığı için, ortalamalardan en az biri farklıdır. Şimdi farkın hangi kurstan kaynaklandığını bulalım:

EKÖF

1) $H_0: \mu_1 = \mu_2$
$H_1: \mu_1 \neq \mu_2$

$D = |\bar{y}_{1.} - \bar{y}_{2.}| = |70.67 - 61| = 9.67$

$EKÖF = S_{\bar{Y}} t_{T(0.025,\ sd=n_1+n_2-2=10)} \sqrt{2}$

$S_{\bar{Y}} = \sqrt{\frac{45.54}{6}} = 2.76 \longrightarrow$ *EKÖF*$=2.76(2.228)\sqrt{2} = 8.696$

D>EKÖF olduğundan H_0 red edilir. A ve B kursları arasında fark vardır.

2) $H_0: \mu_1 = \mu_3$
$H_1: \mu_1 \neq \mu_3$

$D = |\bar{y}_{1.} - \bar{y}_{3.}| = |70.67 - 82.5| = 11.83$

n'ler eşit olduğundan EKÖF değişmeyecektir.

$D > EKÖF = 8.696$ olduğundan H_0 red edilir. 1. ve 3. kurslar arasında fark vardır.

3) $H_0: \mu_2 = \mu_3$
$H_1: \mu_2 \neq \mu_3$

$D = |\bar{y}_{2.} - \bar{y}_{3.}| = |61 - 82.5| = 21.5$

D>EKÖF olduğundan H_0 red edilir. 2. ve 3. kurslar arasında fark vardır.

4) $H_0: \mu_3 = \mu_4$
$H_1: \mu_3 \neq \mu_4$

$D = |\bar{y}_{3.} - \bar{y}_{4.}| = |82.5 - 85| = 2.5$

D<EKÖF olduğundan H_0 red edilemez. 3. ve 4. kurslar arasında fark yoktur.

5) $H_0: \mu_1 = \mu_4$
$H_1: \mu_1 \neq \mu_4$

$D = |\bar{y}_{1.} - \bar{y}_{4.}| = |70.67 - 85| = 14.33$

D>EKÖF olduğundan H_0 red edilir. 1. ve 4. kurslar arasında fark vardır.

6) $H_0: \mu_2 = \mu_4$
$H_1: \mu_2 \neq \mu_4$

$D = |\bar{y}_{2.} - \bar{y}_{4.}| = |61 - 85| = 24$

D>EKÖF olduğundan H_0 red edilir. 2. ve 4. kurslar arasında fark vardır.

GÖF ile yapılırsa;

$GÖF = S_{\bar{Y}} Q_T = 2.76(3.96) = 10.93$ $\quad (Q_{T(\alpha=0,05, k=4, n-k=20)} = 3.96)$

1) $D = 9.67 < GÖF = 10.93$ olduğundan H_0 red edilemez. K1 ve K2 arasında fark yoktur.
2) $D = 11.83 > GÖF$ olduğundan H_0 red edilir. K1 ve K3 arasında fark vardır.
3) $D = 21.5 > GÖF$ olduğundan H_0 red edilir. K2 ve K3 arasında fark vardır.
4) $D = 2.5 < GÖF$ olduğundan H_0 red edilemez. K2 ve K4 arasında fark yoktur.
5) $D = 14.33 > GÖF$ olduğundan H_0 red edilir. K1 ve K4 arasında fark vardır.
6) $D = 24 > GÖF$ olduğundan H_0 red edilir. K2 ve K4 arasında fark vardır.