t-testi ve Varyans Analizi İlişkisi

Örnek: İki farklı gübrenin domates verimine etkisi araştırılıyor. Buna göre iki gübre arasında verime etkileri bakımından fark var mıdır? Test ediniz ve sonucu yorumlayınız.

Amonyum Sülfat (A)	8	12	15	9	12	10	$\sum y_A = 66$	$\bar{y}_A = 11$	$S_A^2 = 6.4$
Üre (B)	8	10	7	6	9		$\sum y_B = 40$	$\bar{y}_B = 8$	$S_B^2 = 2.5$

Çözüm:

$$Ho: \mu_A - \mu_B = 0 \Rightarrow \mu_A = \mu_B$$

$$H\imath: \mu_A - \mu_B \neq 0 \Rightarrow \mu_A \neq \mu_B$$

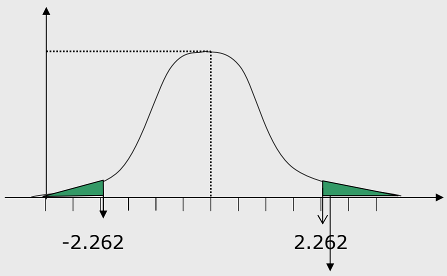
Populasyon varyansı σ^2 bilinmiyor ve $n_A<30$ ve $n_B<30$ olduğundan,

$$t = \frac{(\bar{y}_A - \bar{y}_B) - (\mu_A - \mu_B)}{S_{\bar{x}_A - \bar{x}_B}} = \frac{(11-8)-0}{\sqrt{S_{ort}^2\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}} = \frac{3}{\sqrt{4.67\left(\frac{1}{6} + \frac{1}{5}\right)}} = \frac{3}{1.31} = 2.29$$

$$S_{ort}^2 = \frac{(n_A - 1)S_A^2 + (n_B - 1)S_B^2}{n_A + n_B - 2} = \frac{(6-1)*6.4 + (5-1)*2.5}{6+5-2} = \frac{42}{9} = 4.67$$

α=0.05/2=0.025 alınırsa

$$t_{n_A+n_B-2,\alpha/2} = t_{6+5-2,0.05/2} = t_{9,0.025} = 2.262$$



H_0 hipotezi RED edilir. İki gübre arasında domates verimine etkileri bakımından istatistik olarak önemli fark vardır (P<0.05).

Varyans analizi (ANOVA, **AN**alysis **O**f **VA**riance) tekniği, iki vaya daha fazla (t adet) bağımsız yada bağımlı grup ortalamaları arasındaki farklılığın istatistiksel olarak önemli olup olmadığının test etmek amacıyla kullanılır. İki grubun karşılaştırılması yönteminin kullanılabildiği her durumda varyans analiz yöntemini de kullanılabileceği gibi ikiden fazla grubu da aynı anda karşılaştırma olanağı verir.

Genel yaklaşımı, gözlemler arasındaki değişimi, bu değişimi meydana getiren unsurlara parçalamaya veya analiz etmeye dayanır. Gübrenin domates verimine etkisi ile ilgili örnek üzerinde durumu gösterelim…

t-testi ve varyans analizi ilişkisi

Amonyum Sülfat (A)	8	12	15	9	12	10	$\sum y_A = 66$	$\bar{y}_A = 11$	$\bar{y}_{..} = 9.64$
Üre (B)	8	10	7	6	9		$\sum y_B = 40$	$\bar{y}_B = 8$	

Tablodaki herhangi bir gözlemin genel ortalamadan farkı , bu gözlemin kendi grubu ortalamasından farkı ile aynı gözlemin grup ortalamasının genel ortalamadan farkının toplamına eşittir. Yani,

(15-9.64)=(15-11)+(11-9.64)
5.36 = 4 + 1.36
5.36 = 5.36
Bu eşitlik tüm gözlemler için benzer şekilde geçerlidir.

Bu durum genelleştirilirse…

t-testi ve varyans analizi ilişkisi

	i \ j	Tekerrür 1	2	3	…	n	Σ	Ort.	Ort.
Gruplar	1	y_{11}	y_{12}	y_{13}	…	y_{1n}	$y_{1.}$	$\bar{y}_1$	$\bar{y}_{..}$
	2	y_{21}	y_{22}	y_{23}	…	y_{2n}	$y_{2.}$	$\bar{y}_2$	
	3	y_{31}	y_{32}	y_{33}	…	y_{3n}	$y_{3.}$	$\bar{y}_3$	
	.	.	.	.	…	.	.	.	
	.	.	.	.	…	.	.	.	
	.	.	.	.	…	.	.	.	
	t	y_{t1}	y_{t2}	y_{t3}		y_{tn}	$y_{t.}$	$\bar{y}_t$	

Tablodaki herhangi bir gözlemin (örneğin y_{11}), genel ortalamadan ($\bar{y}_{..}$) farkı (y_{11} - $\bar{y}_{..}$), bu gözlemin kendi grubu ortalamasından ($\bar{y}_1$) farkı (y_{11} -$\bar{y}_1$) ile aynı gözlemin grup ortalamasının genel ortalamadan farkının ($\bar{y}_1$ - $\bar{y}_{..}$) toplamına eşittir. Yani,

$$(y_{11} - \bar{y}_{..}) = (y_{11} - \bar{y}_1) + (\bar{y}_1 - \bar{y}_{..})$$

Eşitliğin sağında ve solundaki terimlerin kareleri alınır ve tüm gözlemler için bu farkların kareleri toplanırsa,

$$\sum_{i=1}^{t}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{..})^2 = \sum_{i=1}^{t}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2 + n\sum_{i=1}^{t}(\bar{y}_{i.}-\bar{y}_{..})^2$$

Genel Kareler Toplamı (GKT)	Grup İçi Kareler Toplamı (GİKT) (Hata Kareler Toplamı) (HKT)	Guruplar Arası Kareler Toplamı (GAKT)

elde edilir.

Varyans analizinin bu uygulanış şekline tek yönlü varyans analizi denir ve genel varyasyonu oluşturan iki unsur vardır. Bunlar, grup içi varyasyon ve gruplar arası varyasyondur.
Kolaylık sağlaması bakımından bu terimler üzerinde gerekli kısaltmalar yapılarak uygulamada aşağıdaki formüllerin kullanılması daha faydalı olur.

$$GKT = \sum_{i=1}^{t}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{..})^2 = \sum_{i=1}^{t}\sum_{j=1}^{n}y_{ij}^2 - y_{..}^2/nt$$

$$GAKT = n\sum_{i=1}^{t}(\bar{y}_{i.}-\bar{y}_{..})^2 = \sum_{i=1}^{t}y_{i.}^2/n - y_{..}^2/nt$$

$$GIKT(HKT) = \sum_{i=1}^{t}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2 = \sum_{i=1}^{t}\sum_{j=1}^{n}y_{ij}^2 - \sum_{i=1}^{t}y_{i.}^2/n \quad \text{veya}$$

$GIKT(HKT)=GKT-GAKT$

Yukarıdaki eşitlikler kullanılarak elde edilen değerler Varyans Analiz Tablosu (VAT) şeklinde gösterilir.

Varyasyon Kaynakları (VK)	Serbestlik Derecesi (sd)	Kareler Toplamı (KT)	Kareler Ortalaması (KO)	F Test İstatistiği (F)	Olasılık Düzeyi (P)
Gruplar Arası	t-1	GAKT	GAKO	GAKO/HKO	?
Grup İçi (Hata)	N-t	HKT	HKO	-	-
Genel	N-1	GKT	-	-	-

Varyans Analiz Tablosunda hesaplanan F test istatistiği değeri (F_{hesap}), (t-1),(n-t) serbestlik dereceli önem düzeyindeki F cetvel değeri ($F_{(t-1),(N-t),\alpha}$) ile karşılaştırılır. Eğer $F_{hesap} > F_{(t-1),(N-t),\alpha}$ ise Ho red edilir ve bu durum **"grup ortalamalarından en az birisi diğerlerinden istatistik olarak önemli düzeyde farklıdır"** anlamına gelir.

Tablodaki olasılık düzeyi (P) sütunu, genellikle bir bilgisayar programı kullanılarak yapılan analizlerde verilir ve Ho hipotezinin red edilebilmesi için gereken olasılık düzeyini gösterir. Diğer bir ifadeyle Ho hipotezinin red edilebilmesi için bu değer en fazla 0.05'den (α=0.05 için) veya 0.01'den (α=0.01 için) daha küçük olmalıdır. Eğer olasılık düzeyi 0.05'den büyük ise Ho red edilmez ve **"grup ortalamaları arasındaki fark istatistik olarak önemsizdir"** anlamına gelir.

Varyans Analiz Tekniğinin Varsayımları: Birçok parametrik istatistik analiz tekniğinde olduğu gibi, varyans analiz tekniğinin uygulanabilmesi için de bazı varsayımlar vardır. Bunlar;

- Toplanabilirlik varsayımı,
- Eşit varyans (varyansların homojenliği), sıfır korelasyon varsayımı,
- Hataların tesadüfi ve normal dağılışa sahip olması varsayımı,
- yij'lerin, Y şans değişkeninin birer gözlenen değerleri olması varsayımıdır.

Varyans analizi sonuçlarının geçerli olabilmesi için bu varsayımların analizde kullanılan veri seti için tutması gerekir. Varsayımların tutup tutmadığının kontrolü için ayrı ayrı testler vardır. Uygulamada genellikle bu varsayımlar arasından **varyansların homojenliği testi** yapılır. Bu amaçla;

F testi, Fmax testi, Cochran testi Levene testi ve Bartlett testi

en sık kullanılan testlerdir. Eğer bu varsayım tutuyorsa, varsayımın diğer varsayımlarla olan sıkı ilişkisi nedeniyle diğer varsayımların da tuttuğu varsayılır. Ancak, varsayım tutmuyorsa veride yapılacak uygun transformasyonlar (karekök, logaritmik, açı transformasyonu gibi) ile varsayım sağlanmaya çalışılır. Yine tutmuyorsa, bu durumda varyans analizi kullanılamaz, bunun yerine etkilerin testi için bu tip varsayımlara ihtiyaç duymayan parametrik olmayan istatistik analiz teknikleri (Kruskal-Wallis, Brown ve Forsythe'nin Modifiye F istatistiği, Box'un normal F istatistiği, Friedman testi gibi) uygulanmalıdır.

Örnek: İki farklı gübrenin domates verimine etkisinin araştırıldığı veriler varyans analizi tekniği ile analiz edilirse

Ho: İki gübre arasında fark yoktur
Hı: İki gübre arasında fark vardır

Amonyum Sülfat (A)	8	12	15	9	12	10
Üre (B)	8	10	7	6	9	

$$DK = y_{..}^2/n = 106^2/11 = 1021.45$$

$$GKT = \sum_{i=1}^{t}\sum_{j=1}^{n} y_{ij}^2 - DK = (8^2 + 12^2 + ... + 9^2) - 1021.45 = 1088 - 1021.45 = 66.55$$

$$Güb.KT = \sum_{i=1}^{t} y_{i.}^2/n - DK = (\frac{66^2}{6} + \frac{40^2}{5}) - 1021.45 = 1046 - 1021.45 = 24.55$$

$$GIKT(HKT) = GKT - GübKT = 66.55 - 24.55 = 42$$

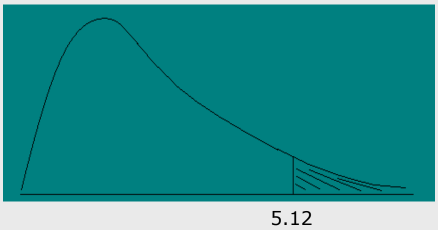
Varyans Analiz Tablosu (VAT)

VK	SD	KT	KO	F
Gübre	2-1=1	24.55	24.55	5.26
Hata	11-2=9	42	4.67	
Genel	11-1=10	66.55	-	

Gübre için test

$F=24.55/4.67=5.26 > F_{1,9,0.05}=5.12$

H_0 hipotezi RED edilir. İki gübre arasında domates verimine etkileri bakımından istatistik olarak önemli fark vardır ($P<0.05$).



Ayrıca iki grup için $t^2=F$ ilişkisi söz konusudur. Yani $2.262^2=5.12$ dir