

## **ÇOKLU KARŞILAŞTIRMA TESTLERİ**

Tarımsal alanda yapılan birçok çalışmada, incelenen işlemler arasında önemli bir farklılığın olup olmadığını bulmak için genellikle varyans analiz metodu kullanılır. Bunun sonucunda istatistiksel olarak belirlenen veri ortalamaları arasındaki farklılığın hangi işlem veya işlemlerden kaynaklandığını anlayabilmek için çeşitli karşılaştırma testleri kullanılır ki, bunlar varyans analizini tamamlayıcı tekniklerdir (Akyürek, 1991). Uygulanabilecek istatistiksel işlemler; regrasyon tekniği kullanılarak cevap fonksiyonlarının oluşturulması, ortalamalar veya ortalama grupları arasında karşılaştırmalar için çoklu karşılaştırma yöntemlerinden uygun birinin kullanılmasıdır. Çoklu karşılaştırma testleri, ortalamaların birbirinden olan farklılığının tespit edilmesi amacıyla kullanılır (Petersen, 1977).

### **HATA TIPLERİ**

Populasyondan alınan örneklerden elde edilen bilgilerle hipotezin kabul veya reddinin % 100 doğrulukla yapılabilmesi mümkün değildir. Bu ancak populasyonun bütün fertlerini ölçmekle mümkündür. Bu duruma göre örnekten populasyonu tahmin ederken mutlaka bir hata yapılacaktır. İşte önemli olan bu hatanın mümkün olduğunca küçük olmasıdır (Tosun, 1990).

Hipotez kontrollerinde yapılması olası iki tip hata vardır. Geçerli veya kabul edilmesi mümkün olan bir hipotezin kontrol sonucunda belirli olasılık sınırına göre red edilmesi halinde I. tip hata yapılmış olur. Geçerli olmayan test hipotezinin kabul edilmesi ile de II. tip hata yapılır. Başka bir ifade ile bir hipotezin yanlış olmasına rağmen doğru olarak kabul edilmesi riski II. tip hatadır. Bir de bunların dışında ortalamaların

karşılaştırılmasında yanlış sıralama ihtimali söz konusu olabilir ki bu da III. tip hata olarak bilinir.

Çoklu karşılaştırma testlerinde birden fazla hipotez birlikte test edildiği için I. tip hata seviyeleri değişik değerler almaktadır. I. tip hata seviyesinin alacağı değişik değerlerin bir sonucu olarak bazen kabul edilmesi gereken hipotezin red edildiği durumlar ortaya çıkmaktadır.

Hipotez kontrollerinde geçerli ve hatalı kararlar ile bunlara ait ihtimaller (Düzgüneş ve ark., 1987);

Gerçek Durum	Verilen Kararlar	
	$H_0$ kabul edilirse	$H_0$ red edilirse
$H_0$ geçerli	isabetli karar ( $1-\alpha$ ) ihtimalle	I. tip hata ( $\alpha$ ihtimalle)
$H_0$ geçersiz	II. tip hata ( $\beta$ ihtimalle)	isabetli karar ( $1-\beta$ ihtimalle)

İşte çoklu karşılaştırma testlerinin gayesi; alınan bir takım önlemlerle en doğru kararın verilmesine çalışmaktadır (Düzgüneş ve ark., 1987).

Hataların önlenmesi için çoklu karşılaştırma testleri 3 farklı gruba ayrılabilir:

1. Deneme hatasındaki büyümeyi önlemek için hiçbir tedbiri öngörmeyen, dolayısıyla karşılaştırma sayısı arttıkça hata oranı da önemli ölçüde artan testler (LSD testi gibi)

2. LSD'nin aksine deneme hatasının korunması için en ileri tedbirleri esas alan testler (Tukey, Scheffe ve Dunnet testleri). LSD testinde önemsiz olduğu halde önemli bulunabilecek bazı karşılaştırmaların varlığına karşılık, bu testlerde önemli bulunması gereken bazı ortalama farkları önemsiz bulunabilir. Çünkü birincisinde I. tip hata büyükken, ikincilerde I. tip hata küçük, II. tip hata büyüktür.

3. Bu iki ekstrem grup arasında yer alan ve test edilecek ortalamalar arasındaki farka göre deneme hatası oranına karşı tedbir alan testler (Duncan ve SNK testleri) (Yıldız ve Bircan, 1991).

### **LSD TESTİ (Least Significant Difference)**

Varyans analizi sonucu F değerinin önemli çıkması durumunda kullanılabilir. Testin uygulama kolaylığı ve daha küçük ortalama farkların önemli bulunması ihtimali yüksek olduğu için araştırmacılar tarafından yaygın olarak kullanılmaktadır. Karşılaştırılacak ortalama sayısının az veya ortalamalardan sadece birkaçının bağımsız olarak karşılaştırmak istenmesi halinde LSD testi güvenle kullanılabilir. Bunun aksine ortalama sayısının çok fazla ve tüm karşılaştırmaların yapılmak istendiği durumlarda, LSD testinin uygulanmasıyla I. tip hata seviyesi çok yükseleceğinden tavsiye edilmemektedir.

$$LSD = t_{\alpha} \times \sqrt{\frac{2 \times HKO}{r}} \quad t_{\alpha}: \text{tablo değeri (Hata SD; olasılık değeri)}$$

$r$  : tekrar sayısı

Bu test herhangi bir ortalama çifti arasındaki farka uygulanan ortak hata varyansını kullanan t testidir. LSD'nin 3'den fazla ortalama olduğu durumda kullanılması sakıncalıdır.

LSD testinin en zayıf tarafı birbirine yakın ortalamalar ile uzak ortalamaların aynı değer ile birbirinden ayırmaya çalışmasıdır.

### **KORUNMAMIŞ LSD TESTİ (Yeni LSD Testi)**

Varyans analiz sonucu F değerinin önemli çıkması gerekmez, fakat her bir karşılaştırma için hata oranı  $2/m$  olur ( $m$ : toplam karşılaştırma sayısı). Bütün karşılaştırmalar yapılırsa  $m = t_x(t-1)/2$  olur. Bunun için hazırlanmış t cetvel değerleri mevcuttur (Bek ve Efe, 1989).

Karşılaştırılacak işlem ya da grupların bağımsız veya ortagonal olması ve karşılaştırma sayısının  $i-1$  ( $i$ = işlem sayısı)'i geçmemesi arzu edilir. Örneğin A, B, C, D işlemlerine sahip bir denemede A ile D; C ile B ortalamaları, karşılaştırmada birer kez kullanılmıştır. Burada ortagonallik veya bağımsızlık kuralı geçerlidir ve her karşılaştırma % 5' lik seviyelere göre yapılmaktadır. Bazı hallerde A ile D ve A ile C'nin karşılaştırılması gibi durumda A iki karşılaştırmada da kullanıldığı için bağımsızlığı olamayacağından sonucun yorumunda biyolojik veya diğer göstergelere önem vermek gerekir.

### **TUKEY' in W TESTİ**

Tukey (1953) tarafından "studertized değişim aralığı" kullanımını sağlamak için önerilmiştir. İki den fazla ortalama kıyaslandığı zaman en büyük ve en küçük örnek ortalamaları kıyaslanacaksa;

$$\frac{Y \text{ en büyük} - Y \text{ en küçük}}{Sp}$$

Sp: genel populasyonun standart sapması ( $\sigma$ ) kullanılır (Ott, 1988).

Test W = HSD = TSD olarak da gösterilir (Carmer ve Walker, 1985). Bu testte bütün karşılaştırma çiftlerinde kritik olarak kullanılan aralık tektir. Bütün karşılaştırmalar için I. tip deneme hatası oranındaki korunma değeri sabit ve aynıdır.

$$W = q\alpha \times \sqrt{\frac{HKO}{r}}$$

$q\alpha$ : Tukey W testi için cetvel değeri (Hata SD; İşlem SD; olasılık değeri)

$r$  : tekrar sayısı (her bir örnekteki gözlem sayısı)

Kritik değerın çok yüksek olması bazen Tukey testinde aslında önemli görülen ortalama farkların önemsiz çıkması sonucunu verebilir. Bu

özelliğinden dolayı uygulaması kolay olmasına rağmen araştırmacılar tarafından yaygın olarak kullanılmamaktadır (Yıldız ve Bircan, 1991). Tukey testi bütün ortalamaların farkları için güven aralığı oluşturulmasında da kullanılabilir. Bu testte  $\alpha$ 'nın 0.1 civarında tutulması testin gücünü yükseltmek için tavsiye edilir. Test işlem ortalamaları arasında tüm çift yönlü kıyaslama gruplarına uygulanır. Tukey'de hata oranı muntazam olarak korunur ve araştırmacı gerçekten kullandığı hata oranının ne olduğunu bilir. Bu testin kullanılmasını savunanlar testin gücünü tekerrür sayısının artırılması ve uygun deneme planının seçimi ile zaten yükseltebileceğini ileri sürmektedir. Aynı zamanda II. tip hatanın çok önemli olduğu durumlarda araştırmacı I. tip hata seviyesini  $\alpha = 0.1$  veya  $0.2$  tutarak da testin gücünü yükseltebilir (Bek ve Efe, 1989).

### SCHEFFE TESTİ

Scheffe (1959) testi F dağılışını esas almaktadır. İşlem sayısı (i) kadar populasyon ortalamaları arasında mümkün olan tüm kıyaslamaları yapmak için kullanılabilir. Bu test çoklu karşılaştırmaların yanında, ortalamaların güven sınırlarının tahmininde de yaygın olarak kullanılmaktadır (Yıldız ve Bircan, 1991). Tukey'de olduğu gibi yapılacak bütün karşılaştırmalar için tek bir kritik aralık değeri hesaplanmaktadır.

Bu kritik değer Tukey'in testinden daha büyüktür. Bu işlem genel karşılaştırmalarda kullanılır. İkili karşılaştırmalar için Tukey testi, Scheffe testine tercih edilir. Hata oranı deneme başına hatadır. Bu nedenle Scheffe,  $\alpha = 0.1$ ' tercih etmektedir.

Test değeri;

$$SSD = S = \sqrt{(i-1) \times F_{\alpha} \times \frac{2 \times HKO}{r}}$$

i: işlem sayısı,

$F_{\alpha}$ : F cetvel değeri (Hata SD; İşlem SD; olasılık değeri)

Birinci tip hataya göre en muhafazakar test Scheffe testidir. Bu test işlemlerin ortalamaları arasındaki mümkün olan tüm linear mukayese gruplarına uygulanır. Kritik değeri Tukey'inkinden daha büyüktür.

### STUDENT-NEWMAN-KEULS TESTİ

SNK şeklinde gösterilir, Tukey testinin değişik bir şeklidir. Ancak daha az tutucu olup Tukey testinden önemli farklılıklar göstermektedir. Tüm çift yönlü kıyaslama gruplarında işlem görür, I. tip hata oranı LSD ve TSD (Tukey'in W değeri)'dekilerin arasındadır. Bu testte i-1 adet kritik değer hesaplanması gerekir (Carmer ve Walker, 1985). İki ortalama arasındaki fark sabit bir kritik değerle karşılaştırma yerine ortalamalar arasındaki kademe sayısına göre değişen değerlerle karşılaştırılır (Bek ve Efe, 1989). Testin kritik değeri;

$$SNK = q_{\alpha} \times \sqrt{\frac{S^2}{r}}$$

$q_{\alpha}$ : q cetvel değeri (Hata SD, i-1 adet karşılaştırma grup sayısı)

$S^2$ : hata kareler ortalaması

$r$  : tekrar sayısı

Kritik değer hesaplandıktan sonra, ortalamalar büyükten küçüğe doğru sıralanır ve en büyük ortalama ile en büyük SNK değeri çıkarılarak gruplandırma yapılır.

### DUNNETT TESTİ

İçinde kontrol ya da standart işlemi bulunan denemelerde, kontrol ile diğer işlem ortalamalarının karşılaştırılması için tavsiye edilmektedir (Bek ve Efe, 1989; Yıldız ve Bircan, 1991). F değeri önemli çıksın ya da çıkmasın bu test uygulanabilir. Bunun için kontrol dahil i tane işlem arasında (i-1) tane test yapılabilir. Bu test, kontrol ile diğer işlemler arasındaki tüm karşılaştırmaları yapmak için kullanılan  $i_D$  istatistiği için örnek dağılışı geliştirilmiştir (Bek ve Efe, 1989). Tek bir aralık değeri hesaplanır. Kontrol ile diğer işlemlerin tekrarları eşitse;

$$D = ti_D \times \sqrt{\frac{2xHKO}{r}}$$

$ti_D$ : t Dunnet cetvel değeri (Hata SD; İşlem SD; olasılık değeri)

$r$  : tekrar sayısı

Dunnett test değeri ile student-t değerleri  $i=2$  işlem için aynıdır. Fakat daha fazla sayıdaki işlem için Dunnett değeri daha büyüktür. 2'den fazla işlem için Tukey'in değeri daha büyüktür, dolayısıyla bütün ortalamaları karşılaştıran Tukey testi daha az farklılık verir.

Bu test daha önce de belirtildiği gibi kontrol yada standart işlemin diğerleri ile kıyaslanmasında kullanılır. Eğer kontrol dışındaki işlemler birbirleri ile kıyaslanacaksa diğer çoklu karşılaştırma testlerinden biri kullanılabilir (Chew, 1986).

## DUNCAN TESTİ

1951 yılında Duncan tarafından geliştirilen bu yöntemde, her bir işlem ortalaması diğer tüm ortalamalar ile ayrı ayrı test değerleri ile karşılaştırılır. Duncan testinin uygulanması için işlemlerin varyans analizi sonucu hesaplanan F değerinin istatistiksel olarak önemli olması şart değildir. Bu yöntem her iki durumda da sakıncasızca uygulanabilir, bunlar bu testin avantajlarıdır (Açıkgöz, 1993).

Deneme hatasından korunma oranı Duncan testinde SNK'dan daha küçüktür. Bu özelliğinden dolayı daha küçük gerçek ortalama farkların Duncan testinde önemli bulunması ihtimali daha yüksektir. İşlenen I. tip hata seviyesi SNK testinden daha yüksektir (Yıldız ve Bircan, 1991).

$$D = Q\alpha \times \sqrt{\frac{HKO}{r}}$$

$Q\alpha$ : Q cetvel değerleri (Hata SD; İşlem SD; olasılık değeri)

Araştırmacılar, test edilen işlem ortalamaları arasında fark bulmamayı değil bulmayı daha çok arzu ettikleri için SNK ve Tukey testlerine oranla daha fazla sayıda ortalama farkını önemli çıkaran Duncan'ı tercih ederler. Duncan değerleri ile güven aralığı tahmini yapılamaz, fakat testin gücü LSD, SNK ve Tukey'den daha yüksektir. Buna karşılık I. tip hata seviyesini koruyamamasından dolayı da birçok araştırmacı tarafından hatalı kullanılır.

Bazı durumlarda varılan sonuçları açıklamakta güçlükler çekilebilmekte, bazı durumlarda da açıklanamamaktadır. Örneğin; yapılan bir çalışmada koyunların sütlerindeki yağ miktarının yaş ilerledikçe arttığı görülmüştür. Duncan'a göre yapılan karşılaştırmada 2 yaşlı koyunların 3 yaşlılardan , bunların 4 yaşlılardan , bunların da 5 yaşlılardan farklı olmadıkları sonucuna varılmıştır. Buna göre mantıken yaşların birbirinden farklı olmamaları gerekir. Halbuki bir yaş grubu bir sonraki ile mukayese edildiğinde durum değişmekte 2 yaşlıların 4 yaşlılardan, 3'lerin 5 yaşlılardan önemli seviyelerde düşük yağlı süt verdikleri anlaşılmaktadır. Bu durumda araştırmacı bu hayvanlarla yapacağı başka bir araştırmada 2 yaşlılarla 3 yaşlıları bir grupta, 4 yaşlılarla 5 yaşlıları da ayrı bir grupta toplayamaz mı? Veya iki yaşlıları ayırıp diğerlerini yaş bakımından üniform bir materyal olarak kullanamaz mı? Duncan testi bu soruları cevaplamada yeterli değildir. Her ne kadar 5 yaşlılar hem 3 hem de 2 yaşlılardan farklı iseler de, 4 yaşlılar 3 yaşlılardan farklı olmadığından,



hayvanları 4-5 ve 2-3 yaşlılar diye iki gruba ayırmak doğru olmayabilir (Düzgüneş ve ark., 1987).

### **WALLER-DUNCAN-BAYESIAN TESTİ**

1969 yılında daha önce geliştirilmiş olan testlere bir alternatif olarak ortaya çıkmıştır. Daha önceki çoklu karşılaştırma metodunun hem geliştirilmiş hem de pratikleştirilmiş bir şeklidir. Bu test bazı bakımlardan Fisher'in LSD' sine benzemekte, hatta bazı araştırmacılara göre de LSD testinin geliştirilmiş halidir. Tek bir kritik değer kullanması açısından pratik ve diğer metotlarda dikkate alınmayan bazı unsurları değerlendirilmesi açısından da teorik üstünlükleri olan bir testtir. Kullanımı gittikçe yaygınlaşmaktadır (Akyürek, 1991).

Bu test için geliştirilmiş özel tablolar vardır. Bu testin uygulanmasında iki sınırlayıcı durum söz konusudur. İlki, tüm örnek sayı ya da genişliklerinin aynı olması gerekir. İkincisi ise, populasyon ortalamasının diğerlerinden çok farklılık olduğunu beklediğimiz durumdur ki böyle hallerde test kullanılmaz.

Bu teste I. tip hatanın II. tip hataya oranı olan " k, hata ağırlık oranı " belirlenir ki bu değer k= 50, 100 veya 500 olur. Bunlar klasik hesaptaki % 10, 5 ve 1' e karşılık gelir. Bu değer Waller-Duncan'ın minimum ortalama risk  $2\alpha$  değer cetveline bakılırken kullanılır.

İşlemin yapılışı:

$$LSD = tw\alpha \times \sqrt{\frac{2xHKO}{r}}$$

$tw\alpha$  değeri özel olarak hazırlanmış Waller-Duncan' ın minumum ortalama risk t değer tablosundan alınır (Akyürek, 1991).

### **SONUÇ**

Çoklu kıyaslama testleri belirli bir mantığa göre organize edilmemiş çeşit, herbisit, fungusit, insektisit, lokasyon, toprak tipi gibi kıyaslamalar için geçerlidir. Oysa aralarında anlamlı ortagonal kıyaslamalar yapılabilen, sıra aralığı, zaman, su, sıcaklık, gübre, ilaç vb. dozları ya da konsantrasyonları gibi kantitatif faktörler ve bunların seviyelerinin inceleneceği çalışmalarda çoklu kıyaslama testlerini uygulamak doğru değildir (Chew, 1986). Çünkü, kantitatif faktör seviyelerinin ele alındığı deneme konuları arasında, faktörün bir seviyesinden diğerine doğru bir süreklilik, geçiş vardır. Bundan dolayı faktör seviyeleri arasında bir ilişki söz konusudur. Faktör seviyeleri belirli bir minimum ve maksimum arasındaki çeşitli sayısal değerleri alabilir. Örneğin 0-30 kg N/da seviyeleri arasında 1 kg'lık artışla 29 adet 5 kg'lık artışlarla 6 adet seviye bulunabilir. Yani denemeye konu olarak alınacak faktör seviye sayısı teorik olarak sonsuzdur. Oysa, kurulacak bir denemede deneme tekniği ve ekonomik kısıtlamalar yüzünden ancak sınırlı sayıda ele alınabilir. Böyle bir denemede Duncan testi uygulanması halinde içlerinde fiziksel optimum konu seviyesinin bulunduğundan emin olmadığımız konu ortalamalarını birbirleriyle mukayese ediyoruz demektir. Zira araştırmacının optimum azot dozunu önceden bilmesine ve bunu deneme konuları arasına koymasına imkan yoktur. Eğer araştırmacı böyle bir bilgiye sahipse, deneme yapmaya da gerek yoktur. Duncan testinin böyle bir denemedeki azot dozlarına uygulanması halinde araştırmacı denemede ele alınan faktör seviyelerinin dışındaki seviyeler için herhangi bir yorum veya tahminde bulunamaz, yalnızca gerçek ekonomik optimum azot dozunu o da kabaca ve yaklaşık olarak tahmin edebilir.

Faktörlerin seviyeleri arasında süreklilik olan denemelerin amacı sadece en iyiyi seçmek değildir. Araştırmacı, minimum ve maksimum arasındaki tüm dozlarının bir süreklilik içinde verimi nasıl etkilediğini

merak eder. “Gübre dozları arttıkça verim artıyor mu ?” ya da “belirli bir doz uygulandığında verim ne olacaktır?” gibi sorulara cevap verecek en uygun istatistik analiz yöntemi, regrasyon teknikleri kullanılarak eldeki verilere rispons fonksiyonlarının çıkarılması ve grafiklerinin çizilmesidir. Bu yöntem araştırmacıya önemli ve yararlı bilgiler sağlar. Denemeye konu olarak alınsın veya alınmasın minumum ve maksimum dozlar arasındaki bütün seviyelere karşılık gelen verim miktarları tahmin edilebilir. Sonuç olarak bu konuyu şöyle özetlemek mümkündür (Petersen, 1977):

İstatistik testlerin seçiminde dikkatli olunmalıdır.

Çoklu kıyaslama testleri kalitatif işlemleri içeren denemeler için kullanılmalıdır.

Bu testlerin gelişi güzel kullanımı bilgilerin kaybına neden olabilir ve çok daha uygun testler varken etkinliği azaltır.

İşlemlerin faktöriyel olarak dağılımı veya kantitatif faktörlerin seviyelerini içeren denemeler için çoklu kıyaslama testleri yerine ortagonal kıyaslamalar ve regrasyon teknikleri kullanılmalıdır.

Çoklu kıyaslama testlerinin uygulandığı durumlarda da seçilecek testte dikkat edilmelidir.

Ortalama sayısının az veya sadece belirli bazı ortalama çiftlerinin karşılaştırmak istenmesi halinde LSD testi, ortalama sayısı fazla olduğu durumlarda eğer araştırmacı daha küçük ortalama farkların önemli çıkmasını istiyorsa Duncan, bunun aksine sadece daha büyük farkların önemli çıkmasını istiyorsa o taktirde SNK, Tukey, Scheffe ve eğer bir standart ya da kontrolle işlemleri kıyaslama düşünülüyorsa Dunnett testi seçilmelidir (Yıldız ve Bircan, 1991). Bu testlerden hangisinin daha hassas olduğunu söylemek zordur. Bu daha çok uygulamadaki duruma ve araştırmacılarca kabul edilebilir bir ayırım yapılıp yapılmadığına bağlıdır.